

# 阿里投入 1 亿元对汉语方言进行保护和开发,此举能否化解上海年轻人不会说标准沪语的尴尬

# 没有语言环境使用方言,现在的保存就是冷冻?

■本报见习记者 王倩 刘雪妍

700元月薪,不限性别、年龄和学历,只要能说一口地道的四川话,即可担任“声音模特”,成为四川方言的“代言人”。3月中旬,智能音箱天猫精灵的这则招聘启事引发热议。

3月20日,天猫精灵研发部门、阿里 AI labs(人工智能实验室)宣布,将成立方言保护专项小组,投入1亿元对汉语方言进行保护和开发。上周,阿里宣布最新工作进展,四川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。

在长期使用中,汉语方言已成为识别身份的“符号”。把一个地方的方言记录下来,就相当于有了一本关于当地的百科全书。使用人工智能等新技术挽救方言文化,是否可以对方言保护带来新的可能?

## 现状▶▶▶

### 平均每个月有两种语言消失

世界上已知的语言有近7000种,但96%的语言目前只由占全球总数不到3%的人使用,它们消失的速度比物种灭绝还要快。

纪录片《海豚湾》导演路易·西霍尤斯这么形容物种灭绝:“就像把全球所有知识放在一座图书馆中,在读这些书之前,我们就放火烧图书馆了。”随着栖息地大幅减少,物种消失的速度,要超过人类记录物种存在的速度。

世界上已知的语言有近7000种,但有96%的语言目前只由占全球总数不到3%的人使用,它们消失的速度比物种灭绝还要快,平均每个月就有两种语言消失。

据《中国濒危语言志》汉语方言系列双主编之一、陕西师范大学文学院教授邢向东介绍,汉语方言的分区是分层次的,官话、粤语、赣语、吴语等是大方言区,大方言里还分次方言、小方言。方言消失有两种情况:一种是城市化水平越来越高、农村空心化严重,方言受普通话和强势方言的影响,特点被消磨。新生代虽说可以听懂方言,但日常不说或根本不会说。学校采用普通话教学,广电媒体也使用普通话,在家庭环境中也是普通话交流,方言出现了代际传承危机,随之慢慢消亡。

另一种是一些弱势的小方言或方言岛成了濒危方言。在南方方言区,特别是在方言复杂的地区,“十里不同音”是很常见的,比如湖南的乡话,分布只有湘西三个县的偏僻农村,是典型的濒危方言。“方言的消失速度很快,濒危方言更是如此,有的地方找濒危方言的发言人都很困难了。”

“一种语言的消失就是一种文化的消失,上海滩早期的曲调还有几个人知道?连上海川剧院招演员都要去外地找,很多年轻人第一语言已经是普通话了。全世界目前尚有7000种语言,再过50年,可能只有一半语言能成活。”复旦大学中国语言文学系教授陈忠敏说,说话的环境就像种花的土壤,没有了土壤,花就无法生存。下一代上海话可能就面临这样的危险境地。

2011年,陈忠敏曾和80多位学者一起倡议,在推广国家通用语言文字的同时,科学保护上海话。近二三十年城镇化进程加剧,上海城市方言由于语言接触而导致的语言变异越来越多,“90后”出生的新一代上海人说的上海话,与四五十年岁以上的上海人所说的上海话,

已大不一样。由陈忠敏领衔的复旦一批语言学专家,目前正在从事城市方言的变异研究,寻找上海话演变的脉络。

方言是具有鲜明地域文化特征的话语符号,在声韵调的搭配、句式的特点、词汇的变化上拥有独特魅力。陈忠敏以上海话中的“哇塞”举例,这个词是心情郁闷、难受的意思,在汉字中无法找到对应的字。

## 希望▶▶▶

### 方言保护工作开始露出曙光

语保工程投入专业技术人员超4500名,参与高校和科研院所超过350所。汉语方言调查了全国34个省市、1495个调查点。

“现在国家层面逐渐认识到方言保护的文化和方言存在的价值了,认识到对方言对地方文化的承载功能。”邢向东说,“虽然现状堪忧,但有曙光露出来,还是有希望的。最近,联合国教科文组织在北京发布首个以‘保护语言多样性’为主题的重要永久性文件《岳麓宣言》,相信会对中国的语言文化保护事业起到积极的推动作用。”

为科学保护各民族语言文字,2015年由教育部和国家语言文字工作委员会牵头的中国语言资源保护工程正式启动,这是目前世界上最大规模的语言资源保护项目,分为汉语方言调查、民族语言调查和建立中国语言资源保护工程采录展示平台三个部分。

语保工程投入专业技术人员超4500名,参与建设的高校和科研院所超过350所。汉语方言调查了全国34个省市、1495个调查点,寻找了5000多个发音人。许多像邢向东一样的专家和专业技术人员投入语保工程,大批专业工作者调查员课题组根据《中国语言资源调查手册》,对方言进行语言结构和口头文化两大方面调查,按照统一的技术标准,摄像、录音、保存、上交、入库。

“每一个方言调查点,我们要采集1000个字的读音,1200个词组和50个句子,还有口头文化语料”,对于那些濒危方言做了更系统、更大量的调查,“有的专家拿上来的书稿有五六十万字,在出版阶段又被压到35万字。”邢向东认为这个书可以写得更多,容纳更多内容。据悉,第一批10册濒危汉语方言志即将出版。

作为濒危方言调查项目的负责人,他认为这个工程的系统性非常强,“从技术力量、专业力量看,几乎把全国从事方言工作,甚至跟方言有关系的语言学专家,全都动员起来。”

截至目前,中国语言资源保护工程的汉语方言调查和少数民族语言调查记录的百万条视频、音频的储存总容量超过40TB。

清华大学的团队负责中国语言资源保护工程采录展示平台的搭建和更新维护,这是语保工程中的技术支撑。据项目参与人、清华大学计算机软件研究所高级工程师张鹏介绍,语保工程采录展示平台建设的目标在于对数字语言资源进行收集、存储、管理和服务,并进行语言资源的分析与研究,进而投入社会应用和建设多级语言系统。

除了“国家队”,一些个人也参与到行动中。主持人汪涵希望自己能以“语言保护者”的身份被记住。2015年他发起“響應”(响应)计划,个人投入500万元,计划用5至10年时间,对湖南57个调查地的方言进行搜集研究,用声像方式保存方言资料。

## 聚焦脱贫攻坚 聚焦特殊群体

## 聚焦群众关切

(上接第1版)国务委员王勇参加会见并在会上讲话指出,要深入学习贯彻习近平总书记关于民政工作的重要论述和指示精神,认真落实李克强总理重要讲话精神,牢固树立以人民为中心的发展思想,全面扎实完成脱贫攻坚兜底保障任务,加强和完善各类困难群体基本生活保障,不断提升基层治理和社会基本服务水平,努力推进民政事业改革发展上新台阶。

肖捷、何立峰参加会见。

第十四次全国民政会议的任务是,以习近平新时代中国特色社会主义思想为指导,认真学习贯彻习近平总书记重要论述,总结党的十八大以来民政事业改革发展取得的成就,研究部署今后一个时期的民政工作。

各省市区和计划单列市、新疆生产建设兵团,中央和国家机关有关部门负责同志,全国民政系统先进集体、先进工作者和劳动模范、“孺子牛奖”获得者代表等出席会议。

## 十九届中央第三轮巡视全面展开

## 今年再增百项个人事项全市通办

(上接第1版)围绕推动全面从严治党纵深发展,检查落实“两个责任”情况;围绕贯彻新时代党的组织路线,检查落实党建工作责任,加强组织建设和选人用人情况;围绕强化整改主体责任和监督责任,检查整改落实情况。加强对共性问题的调查研究,为完善体制机制、深化改革提供参考,发挥巡视标本兼治战略作用。

据悉,本轮中央巡视,15个中央巡视组将在被巡视单位工作2个半月,巡视期间设专门值班电话和邮政信箱,主要受理反映被巡视党组织领导班子及其成员、下一级党组织领导班子主要负责人和重要岗位领导干部问题的来信来电来访,重点是关于违反政治纪律、组织纪律、廉洁纪律、群众纪律、工作纪律和生活纪律等方面的举报和反映。巡视组受理信访时间截止至2019年6月10日。

(上接第1版)若根据传统“货到申报”模式,货物抵港后再安排理货、换单、申报、赋税、放行等一系列流程,一般要5个工作日后才能真正提高,“提前申报”却让企业以海运成本实现了空运效率。钱莉透露,公司过去进口票数中的空运比例为80%,为之付出了3倍于海运的成本,但在首票试单成功后,公司旋即空运比例降至50%,最保守估算,可为公司每年减少数十万美元支出。据悉,目前上海口岸“提前申报”比例已提升至21%,为全面推广该制度,上海海关还配套建立了进口货物“提前申报”容错机制。

## 样品还在实验室,矿石已放行

获得感之二,来自“先放后检”。我国是资源性矿产品进口大国,以铁矿石为例,由于国内高品质铁矿石资源相对匮乏,我国铁矿石对外依存度常年保持在85%以上。作为一家综合性矿石专业公司,上海贝优能实业有限公司日前收获了“天大的利好”——过去,公司到上海口岸办理矿石船舶靠岸事宜,要等海关实验室45天后出具相关检测报告后才能进行货物提离。而至少45天的码头仓储费及场地中转等费用,对企业而言是笔不小的金额,也无疑增加国内矿石产业链的整体成本。去年8月和10月,海关

卡塔国能源部部长、卡塔尔石油公司总裁兼首席执行官萨阿德·谢里达·阿尔卡比,特立尼达和多巴哥共和国能源部部长富兰克林·卡恩出席。中国海洋石油集团有限公司董事长杨华、中国石油天然气集团有限公司董事长王宜林致辞。

4月1日至5日,第19届国际液化天然气会议在上海举行,主题为“LNG助力可持续发展的能源未来”。



世界上已知的语言有近7000种,但96%的语言目前只由占全球总数不到3%的人使用,平均每个月就有两种语言消失。

语保工程每一个方言调查点,要采集1000个字的读音,1200个词组和50个句子,还有口头文化语料。

绘图:狄斐

## 记者手记

## 莫把能活下来的东西变化化石

今年2月21日“国际母语日”,联合国教科文组织在北京发布首个以“保护语言多样性”为主题的永久性文件《岳麓宣言》,号召国际社会就保护和促进语言多样性达成共识。

目前国内正在用来保护方言的科技手段,在部分学者眼中依然不能有效解决方言使用热情的问题。在复旦大学中国语言文学系教授陈忠敏看来,语言是活的,人与人之间的交流形式多样,一个言语社团充满变异。同样一句话,在不同的背景噪音、不同的心情、语气、环境等因素下,会产生不同的变异,但是不妨碍人际沟通,而机器可能就无法分辨。

“机器语言识别的方法是获取大样本提高算法精度,从而提高语言的识别率。现在的人工智能是大样本匹配,与人类的认知仅需小样本就能达到高识别率完全不同。目前,语言运作的脑机制研究还处在盲人摸象阶段。”

陈忠敏把通过AI或其他先进科学方式识别记录、保存方言资源的过程,“不恰当”地比喻为:“好像为一个生育能力下降的高龄女子冻卵。没有语言环境去使用方言,现在的保存就是一种冷冻。”对此,阿里巴巴集团公关部公关经理朱珠回应,“我们的产品(天猫精灵)是想推动大家日常使用方言,营造一个方言氛围。首先,天猫精灵肯定会

做方言的版本,起码以后在四川地区,使用产品时会倾向用方言。”

拿一份报纸,能用方言流畅读下去,是陈忠敏判断方言掌握情况的标准。看似简单,但其实“现在很多年轻人做不到,说几句上海话就会冒出普通话。报纸上有很多书面的词汇,现在电台和电视台也很难招到沪语播音员了。”

陈忠敏80多岁的母亲不会说普通话,去三甲医院看病,许多医生听不懂她的上海话,往往要带着陈忠敏去做翻译。他认为,地方方言既有传承地方文化、弘扬地方特色的功能,也有语言服务、语言经济效益的功能。他更希望充分利用好这一无形资产,方言保护从娃娃抓起,不要把能活下去的东西变成化石。

# 人工智能会成为濒危语言的救命稻草吗

## 如果使用到位,至少能避免“一个人带走一种语言”重演

两年前,讯飞输入法也曾推出方言保护计划,呼吁“为世界留下多彩乡音”。2018年春节返乡时,北京地铁4号线上的讯飞输入法乡音广告让很多人拿起手机,录下几句家乡话。目前,讯飞输入法宣称支持粤语、四川话、上海话等23种方言语音识别。

数字技术的发展和人工智能的应用,为语言保护工程带来新的可能。和传统手段相比,新技术不仅让语言存档速度和能力提升,并且可以通过互动学习为方言库的完善做出贡献。

### 四川话声音模特还没找到合适的

人工智能会成为濒危语言的救命稻草吗?《人民日报》曾评论称,“从某种程度上说是可以的,如果使用到位,至少能避免重演‘一个人带走一种语言’的尴尬。”

刚成立不到一个月的阿里AI labs(人工智能实验室)方言保护专项小组,现在是虚拟小组状态,语音识别和语音合成两块业务正在推进。据阿里巴巴集团公关部公关经理朱珠介绍,目前四川话的语音识别工作已经完成。

反应热烈的四川话声音模特招募,正是为了语音合成做准备。朱珠表示,虽然报名四川话声音模特的人数众多,但目前还没找到合适人选。“声音模特跟大家意义中的配音不一样,在技术上我们要求声音模特录制基本的语料,这些语料可以自由组合,合成新的句子。即使没有录制过这句话,也能‘说’出来。声音模特人确定后,技术开发还需要3个月左右的时间。”

四川话是西南官话,使用者众多且一致性高。与学界优先保护濒危语言的做法不同,阿里AI labs方言保护始于四川话,正是看准了语音交互产品在方言优势地域的巨大商业价值,这背后是三四线城市以及农村地区。朱珠承认,“现阶段,我们做不到像语保工程那么广的覆盖度,但我们重在积累厚度,同时也要考虑方言和产品的结合,选择从有大量使用者的方言起步。”

### 方言保护不可能瞬间用钱砸出来

在陕西师范大学文学院教授邢向东眼中,企业参与方言保护,在技术和资金方面有强大的优势。“不过也有困难,方言文化资料的记录、保存在

语保工程里都有明确、严格的规范和标准,企业做的动态方言库中,如何收集方言材料?方言材料的质量有没有一个下限式的标准?如何判断是什么方言?人工智能对方言材料依照什么原则分类和存储?这些都需要在启动、设计阶段就考虑到。”

对于学者提出的疑问,朱珠回应称,“产品开发还在进行中,我们需要和学界多沟通,共同研究出更科学的方式。”

阿里AI labs的方言保护小组还没来得及和他们的顾问见面。作为第一位顾问,暨南大学华文学院教授郭照得知项目投入1亿元后说,“估计不够”。

朱珠表示,在初步阶段,学界对于这个项目是很好奇的,自己接触的专家中还是鼓励的多。“郭老师也说,他相信肯定有一些改变,至于我们能做成什么样,他也很想看一看。”朱珠强调,方言保护不是一个人或一家公司能做到的事情,也不可能指望能瞬间用钱砸出来,“记录和保留即将消失的语言,我不知道能不能做得到,这是一个很终极的目标。希望能借此提高大家对于方言保护的关注度,更重视方言,更愿意使用方言。”

## 打造面向亚太的天然气管道和定价中心

(上接第1版)我们要充分利用国际国内两个市场、两种资源,构建更加多元的能源供给格局,要坚持绿色发展理念,大力推广清洁能源利用,加快形成以清洁能源为主的能源消费结构,要支持创新驱动发展,突破一批核心能源装备和关键共性技术,深度应用大数据、人工智能等新技术,加快建设能源互联网,要坚持完善市场体系,打造面向亚太地区的天然气管道和定价中心,相信本届会议将为世界清洁能源创新发展贡献智慧,成为实现能源可持续发展的风向标。

章建华在致辞时说,中国和企业近年来积极推进能源转型,大力推动包括天然气在内的清洁能源高效利用,液化天然气基础设施不断完善,市场逐步发育,机制政策日臻完善,产业发展基础进一步夯实。中国现已成为全球最具活力的天然气和液化天然气市场之一,为世界液化天然气产业发展提供了长期稳定的

市场保障。希望全行业共同努力,持续优化液化天然气产业链,加强合作,推动创新,共同破解行业发展存在的难题。

姜周明在致辞时说,作为液化天然气行业的“奥林匹克”盛会,国际液化天然气会议一直致力于发挥天然气在可持续性能源解决方案中的重要作用。本届大会让我们齐聚上海,一起推动液化天然气的贸易,共同应对所面临的挑战,在努力让更多人提供能源的同时,降低成本并保护环境。

卡塔尔国能源部部长、卡塔尔石油公司总裁兼首席执行官萨阿德·谢里达·阿尔卡比,特立尼达和多巴哥共和国能源部部长富兰克林·卡恩出席。中国海洋石油集团有限公司董事长杨华、中国石油天然气集团有限公司董事长王宜林致辞。

4月1日至5日,第19届国际液化天然气会议在上海举行,主题为“LNG助力可持续发展的能源未来”。

## 今年再增百项个人事项全市通办

的将来,带着一大沓证件去窗口办事的场面,就将成为过去。

如果说孤立的数据只是应用场景有限的“元素”,当这些“元素”经过汇聚、分析,就可以发生奇妙的“化学反应”,创造“1+1=N”的倍增效应。

通过数据汇聚,“一网通办”总门户还为每个人和企业“画像”,将个人和企业由政府不同部门每次办事所填写的信息、提交的申请材料以及办事的过程结果等信息数据进行归集、清洗、存储和动态更新,形成了个性化的“一人一档”和“一企一档”。

有了这些档案,用户就按照人生事件、人群、行业、职业等类别,被贴上了一个个“标签”,系统也会越来越“善解人意”,通过人工智能学习的方式不断进行动态更新。如同网购时能收到系统根据用户购买习惯推荐的商品信息一样,我们在享受政务服务

## 货物在海上,报关已进行

总署分别下发《海关总署关于调整进口铁矿检验监管方式的通知(试行)》和《关于调整部分进口矿产品监管的方式》,对进口铁矿石、锰矿石、铬矿石、铅矿石及其精矿等矿产品,采取“先放后检”的监管方式。

上海海关综合处处长徐枫介绍,根据新政,这些矿产品在码头完成检验检疫工作且符合要求后便可提离,企业几乎“零等待”,而实验室检测、综合评定及证书签发等工作均在货物提离后实施,平均每批次货物较原先降低了83%的费用。据统计,去年10月19日至今年1月19日这3月间,上海港进口需检验矿产品平均验放时长为4.2天,较2017年的平均47.6天缩短了91.2%。

## 美食冻品下了船,一天上餐桌

获得感之三,来自“边检边放”“随到随检”。目前,从上海口岸进口的肉类产品约占全国进口总量的25%;进口三文鱼等冰鲜水产品约占全国进口量的33%;全国经空港口岸的鲜活货中有50%自上海口岸入境;还有俄罗斯鲜活帝王蟹已实现了常态化供应,变更前“一年才来一船”到而今“每月都来”……而这些看得见的变化,不过两三年间,背后谁

时,也能收到系统的“推送”。

比如,当机动车行驶证将要到期时,用户就能收到年检提醒;又比如,当孩子一天天长,接种疫苗、办理社保、报名入学等提醒就会传送到父母的手机上……

“在个人的每个成长阶段,企业的每个发展阶段,主动、精准地提供个性化、智能化的服务推荐,使用户了解什么时间可以办什么‘事’,由被动服务向主动服务转变。”市大数据中心副主任池龙说。

按计划,今年上海还将实现100件事务流程革命性再造,新增100项个人事项全市通办,逐步实现对数据的分析应用。“经过努力,要逐步实现企业、市民到政府办事像‘网购’一样方便。”不久前召开的市政府常务会议上也要求,要加快政府部门网站和信息系统集成,实现更大范围的信息数据汇聚、互联、共享,努力为主体和市民办事提供更多便利。

是推手?受益者保正(上海)供应链管理股份有限公司道出缘由。公司总裁曹娟介绍,得益于上海口岸的贸易便利化措施,使公司的“胆大”客户明显增多。有一位客户,今年开始进口产自新西兰、保质期仅15天的巴氏牛奶,流程可谓环环相扣——在新西兰,周一挤奶,随即进入工厂加工,当晚上飞机,周二抵达上海空港,开始清关配送,周三便可送至市民餐桌;另一位客户则致力于三文鱼进口,把加拿大渔民捕捞的新鲜三文鱼在12小时内送达温哥华加工厂,在6小时内完成加工和包装,随后坐12小时飞机直航上海,落地2小时内完成清关及冷链配送,确保一线城市居民在三文鱼上岸后36小时内便能享用。为此,浦东机场海关对进口鲜活货开辟绿色通道,实行7x24小时,365天全天候通关查验制度,确保鲜活货随到随检。

上海跨境贸易营商环境的变化,正娇惯着上海“吃货”们的胃口。利乐包装,保质期长达6个月以上的常温奶曾红极一时,而今已然失宠,赏味期不超过2周的进口鲜奶更受欢迎。随着常温奶一同“过气”的,还有零下18℃保存、保质期长达2年的进口冻牛肉,取而代之的是保质期仅2个月的冰鲜牛肉,以及下了飞机还在爬的波士顿龙虾。目前,沪上进口牛奶、肉类与海鲜界均已出现如此“短保”现象。