

人工智能视域下的林州方言建档

文 / 武瑞丰

摘要：方言建档不仅需要采集方言的纸质、语音资料，而且需要把方言档案更好地保存下来，这些都是难题。文章对林州方言的特征进行了分析，探讨了林州方言建档的必要性，进而分析了林州方言建档中音档制作质量不高、保存不良、检索困难的问题，提倡以视频、录音的方式对方言建档，以利于后期林州方言档案能够进行智能化的检索。

关键词：林州方言；人工智能；方言建档

中图分类号：G272;G275.9

文献标识码：A

文章编号：1005-9652 (2018) 01-0062-03

林州在地理位置上隶属于河南省，但林州方言却非河南官话；从语言学和社会学的角度看，林州方言是由林州这一区域的社会成员共同创造的能记录并反映当地文化的载体。从方言地图上看，林州方言实质上是一种移民语言，是山西晋方言的变异。由于林州和山西之间太行山及中条山的阻隔，普通话的强力推广和河南中原官话的影响使得林州方言在语言的各个方面，尤其在语音、词汇、语调等方面都表现出了自己独有的特征，与其母方言——晋方言的关系并不完全吻合。无论在声调还是词汇方面，林州方言和河南话之间存在很大差异。例如，在林州方言里“妈妈”的发音并不是官话的/mama/，而是类似于/momo/的发音。许多入声声尾都以o的形式被保留，这一点在安阳县西部、林州市（原林县）和汤阴县的部分农村都有所体现。林州话里保留了官话中早已消失的许多词汇，比如林州人会把“酱油”叫做“清酱”，“头发”称为“maoyi”，将马车和人力车等唤作“轱辘马”，还有一些特有的词汇“颗星（小雨）、屙囊（脏）、砸（咋，如何）”等。随着林州现代化建设步伐日益加快，人群之间的交流也日渐频繁，加之普通话的强制推广，林州方言面临着巨大的挑战和压力。因此，林州方言的保护和建档工作也提上了议事日程。

一、林州方言建档的必要性

城市化进程的大力推进使得很多方言正逐步濒临

灭亡。以林州市为例，1980年代以前，男女老幼都会使用林州市方言交谈，农村还可以听到走街串巷做小买卖的林州市方言的吆喝声。现在，在林州市区已经很难听到正宗的林州市口音了。以一家祖孙三代为例，爷爷辈的方言口语用词，到父母那一辈最多传承接纳80%，因为从父母开始留居城市后，方言逐渐和城市普通话融合。结果是，父母这一辈对家乡方言能够熟练使用，对城市的方言也比较熟悉。而到了孙子这一辈，从小就进入幼儿园，接触的大都是普通话，并且可能完全融入城市，剩下的方言词汇更是少之又少，甚至听不懂方言词汇。

（一）传统地方文化

语言是反映文化的一面镜子，方言是语言变体，具有稳定性和历史传承性。方言所记录的风俗习惯、戏剧戏曲等文化现象虽然正在逐渐走向消亡，但通过建立“方言语音档”，就可以作为镜像长期保存。因此，方言可以当作反映历史人文的“活化石”，能够推动社会和谐发展。

从历时语言学的角度看，方言建档能够为未来方言的研究与传承提供宝贵的资料，方言档案用音频、视频的格式保存，其价值不可估量。中国文化源远流长，但是很多历史文化只有文字资料而缺乏声像资料，致使语言研究中的许多问题难以解决。方言建档除了能给一般方言研究提供可靠的资料外，还具有重要的汉语语音史研究的意义。众所周知，林州人从山西洪洞县移民而来。对林州方言的深入调查和系统

研究,便于我们分辨出山西移民带来的晋语成分,从中提取晋语的语音残存。

(二) 推动社会和谐发展

林州方言代表的是林州地域的文化,而千千万万不同区域的文化构成了华夏文明。各地方言对同一现象有不同的表达法,体现出不同地区、不同族群的不同认知模式。说什么样的话就是什么样的人,方言正好体现了言语说话者的不同身份。例如,“老乡见老乡,两眼泪汪汪”,“老乡”一词,蕴含着特殊的情感价值,说同一种方言的老乡们在各种交往和交流中,能够先人一步地建立良好的互动关系,甚至会带来超出理智的一种信任。社会的和谐发展恰恰是建立在良好的互动关系和彼此的信任之上。

二、林州方言建档过程中存在的问题

方言是一种软实力,承载着传统文化,记录了具有主流意识形态与核心价值观的群体语言,不仅影响人群的思维习惯,而且影响人群的行为习惯。方言建档的独特人文价值毋庸置疑,而它在政治和经济领域中的作用也日渐被认知。令人遗憾的是,方言的建档或多或少还存在一些不足。

首先,当前林州政府对文化软实力、方言的濒危地位、方言的文化载体性质认识不够,导致林州方言建档工作没有提上办事日程。林州政府应该从保护本地非物质文化遗产的高度来认识方言,充分意识到城市化背景下方言和地域文化濒危的严重性,通过方言建档及时保存、有效传承地方文化。

其次,林州方言档案在制作中,技术原因、摄像机的质量、录音笔的效果以及建档工作人员的工作态度都对方言建档造成一定的阻碍。比如,视频影像清晰度、音档的响度、音色等都会有不理想的问题。还有许多采录回来的语音资料由各个研究人员分散管理,没有进行及时地汇总研究,也没有形成一个语音资源交流平台和方言资料集合,大大降低了研究水平,难以充分体现其研究价值。

再次,在方言的音档保存方面,档案部门仅仅对纸质版的传统档案有着严格的管理制度,而音档资料的管理制度尚不明确。再加上执行力的薄弱,管理制度变成了一纸空文。另外,音档工作人员随意、无序地对音档进行管理,对于摄像和录音工作重视不够,并认为产生的音档资料仅仅是本部门活动的一种常规记录,就存放在本部门即可,没有将其提升到真正的档案高度予以保管。

从目前看,档案信息管理人员素质有待提高,许多单位档案管理部门缺乏高技术人才,而档案信息处理的复合型人才就更寥寥无几了,大部分档案人员是档案管理的“老人”,虽然经验丰富,但是应用现代电子计算机的技术水平偏低,对高科技工具的使用无从下手。尽管引进了一些现代化设备,不会使用,更谈不上档案信息电子化了。实现档案信息电子化,不能仅仅满足于一般的计算机操作。方言人工智能档案库的建立,要有具备较高知识层次和先进技术水平现代化管理人员,在专家系统知识库建立后,能自主、快速地对档案保管期限、保管密级、档案自动标引做出判定,也能对相关档案资料进行智能检索,以便将来二次开发利用。

最后,当前方言档案的借阅和检索主要以人工操作为主,需要通过人力翻阅、查找,工作量大,这不仅耗费大量的精力,反复查找可能会出现错误的信息,档案也会不同程度地受损。

三、林州方言建档的策略

根据方言建档存在的问题,“人工智能”的林州方言档案系统主要包括:方言档案案卷管理系统、方言文档的检索和方言文档保存。其功能主要是档案的录入与归档、档案的手工检索和计算机检索、纸质档案和电子档案的保存。

(一) 方言档案的录入和归档

为了处理好普通话的推广和林州方言的保护之间的关系,制定出科学的语言保护政策,领导部门需要在方言使用群体范围内宣传方言建档工作的重要性,鼓励全民参与方言建档。在制作调查文本、档案录入等各工作环节都应该有明确的规定。

第一,确定合作录制的方言合作人:50岁以上的林州男性,发音清晰,口齿清楚。性别要求男性,主要是因为他们对林州当地的风俗习惯、风土人情比较了解,与从外地嫁入的女性相比更能体现方言全貌;世居当地,即三代内都在林州本地居住;文化程度虽说要求不高,但须认识较多的汉字,阅读不能有障碍,对当地的民间故事、传说也要十分了解。

第二,规定调查文本由语音、词汇、语法和短文四部分来构成。语音主要包括声母、韵母、声调的调查;词汇、语法调查应选择能够体现林州当地方言特色的风土人情以及天文、地理、风俗等方面的内容;短文应该具有地方特色的一些内容,主要包括民间故事、谜语、歇后语、童谣、山歌或当地的戏曲等。

第三,汇总统计工作人员不仅要进行录音、摄像,而且要对这些声音进行文字和音标记录,确保录音清楚、摄像图像清晰。利用“电生磁”原理把声音、图像信息的电信号变成磁信号而记录在磁带上,再利用“磁生电”的原理将磁记录材料上的磁信号转变成电信号释放出来,也就是我们平时说的磁带、光盘和录像。但是,此类声像档案往往直观性不强,文字说明材料成为有益的补充。按照文字说明,我们可以明确声像档案形成的时间、目的和内容,易于分类整理。

第四,“人工智能”的林州方言档案的归档应该按照归档制度的要求,将具有保存和利用价值的磁带内容和相关的文字材料同时移交方言档案管理部门归档;要对声像档案说明该档案的内容、时间、制式;还应该归档一套可借阅的备份。

(二) 方言档案的保存

大数据背景之前的方言音像档案主要有光盘版和文字版两种。光盘版CD一般来说,语音真实清晰,音质较好。CD所配备的文字版是其书面材料,是对光盘上的每一个音的记录。音像档案不同于传统的纸质档案,对保存的环境如温度、湿度、电磁辐射有着相当高的要求。但是现在的档案管理人员,由于工作马虎,思想麻痹,通常将音像档案和传统纸质档案混合放置,导致声像档案受损而不能正常使用,破坏了声像档案资源的完整性,从而丧失了方言档案的价值。

“人工智能”的方言档案恰好能解决这一弊端,其基本流程包括,方言档案案卷整理、目录建库、数据处理和信息存储。案卷整理是对需要扫描的方言档案案卷进行整理,对方言档案的行号、架号、列号、全宗号、分类号和案卷号等做出标识。为了便利“人工智能”的检索,必须建立必要的目录数据库。数据处理需要对扫描图像进行校对,确保图像完整,并对文件格式进行转换、分类处理、添加说明性文件和备份。

(三) 方言档案的检索

方言档案的检索体系主要包括两个方面:存贮系统和检索系统。存贮系统的任务是建立检索工具体系或数据库。检索系统包括两个方面:手工检索和计算机检索。无论是组织手工检索体系,还是建立机检数据库,都必须对方言档案进行检索前处理工作,即录入工作。录入工作要求对纳入检索系统的每一份档案给出检索标识,以简练的方式标识档案的特征。没有检索标识的档案,不能存贮在检索系统中,当然也不能对其检索。

为了满足用户的需要,必须灵活、全面、准确和迅速地建立方言档案检索体系。否则,“人工智能”的检索体系也就失去了意义。“人工智能”的检索是通过对电脑存储的数据库搜寻实现的。其结果体现为数据库的建立,而数据库的建立必须依赖于档案著录项目的输入,没有这种输入,“人工智能”检索便成了无米之炊。

参考文献:

-
- [1]徐越.浙江方言音档的构建及预期价值[J].档案文化,2012(1):39-42.
 - [2]李锦芳.中国濒危语言认定及保护研究工作规范[J].广西大学学报:哲学社会科学版,2015(2):112-119.
 - [3]王桂宏.方言语音档案建设刍议[J].中国档案,2008(4):50-51.
 - [4]陈章太.当代中国的语言规划[J].语言文字应用,2005(1):3-13.
 - [5]胡元潮.档案管理理论与实践[M].杭州:浙江工商大学出版社,2015.

作者简介:

武瑞丰(1982-),男,河南安阳人,安阳学院外国语学院讲师,研究方向:理论语言学、句法学。
