

论中国语言资源有声数据库的建设*

李 宇 明

提要 中国语言资源有声数据库,是国家语委正在开展的一项语言工程,旨在用现代信息技术、遵循统一的工作规范和技术规范、将中国各县域的语言实态(也包括方言和地方普通话)记录下来,归档建库,永久保存。通过对该库的学术开发和行政开发,可以起到全面了解语言国情、科学制定国家语言规划、科学保存和开发国家语言资源、促进普通话的推广、促进语言文字的信息化、促进语言科学的发展等作用。

关键词 中国语言 有声 数据库

用现代信息技术将我国语言的现实状况采录下来,建成可满足社会多方需求、可不断维护更新的有声数据库,是利在当代、惠及后人的事业。我国语言学的百年发展、方言学家和民族语言学家的研究积累、现代信息技术的水平和国家的经济力量,使中国语言资源有声数据库的建设具有了现实可行性。

国家语言文字工作委员会对中国语言资源有声数据库建设筹划多年,设立多个项目进行论证性的专题研究^①,制定了一系列工作规范和技术规范,并在江苏省的五个城市前前后后进行了一年多的试点工作。本文根据国家语委的筹划、研究和试点,谈谈中国语言资源有声数据库设计的基本情况、特点及可以发挥的作用。当然,中国语言资源有声数据库的建设还没有全面展开,有些技术规范(比如:方言字的规范、少数民族语言的调查方案)还在进一步完善中。发表本文的目的,是希望得到学界和社会对这一工作的关心,也希望得到更多的指教与帮助。

一 中国语言资源有声数据库的基本情况

中国语言资源有声数据库是国家语言资源建设工程之一,它用现代信息技术采录语言数据,经转写、标记等加工程序将相关的文本文件、音频文件及视频文件整理入库,以数据库、互联网、博物馆、语言实验室等形式向学界和社会提供服务。现根据《中国语言资源有声数据库调查手册》^②,将中国语言资源有声数据库设计的基本情况,概述如下:

* 本文为中国社会科学院语言研究所建所60周年而作。60年来,语言研究所为国家的语言文字事业做出的贡献有口皆碑。1955年的现代汉语规范问题学术会议,奠定了我国语言规范化的理论基础。根据会议精神编写的《现代汉语词典》,在我国的语言生活中发挥着“教师”的作用;开展的方言研究,为普通话推广和方言学发展建功立勋。1955年与中央民族学院联合召开的民族语文科学讨论会,树立了民族语言文字工作的里程碑。创办的《中国语文》《方言》等刊物,培养了代代学人,推动着学术发展。值此之际,谨致敬意,并祝语言研究所为国家的语言文字工作和语言科学的发展做出更大贡献。

(一) 语言和方言调查

1) 调查点

根据县级行政单位设置调查点,原则上“一县一点”,特殊情况下可以增点或减点。本调查重在反映当下语言的实态,因此调查点选择县城等在当地影响较大的地方。

2) 调查对象

每个调查点根据性别因素和年龄因素选择有代表性的4名发音合作人,其中男女各2人,老年青年各2人。

3) 调查内容

调查内容分语言结构调查和话语调查两大部分。语言结构调查是基础,用1000字调查语音系统,用1200词调查基本词汇系统,用50个句子调查主要的语法现象。话语调查是重点,分为讲述和对话两部分。讲述包括“规定故事”和“自选话题”。规定故事已经选定了《牛郎和织女》^③,它具有中国文化的特点,分布地域也比较广泛。自选话题尽量反映当地文化特色,例如:当地口耳相传的民间故事;当地的童谣、谚语、歇后语、顺口溜;当地的旅游景点和土特产;当地的风俗习惯和传统节日;个人和家庭的情况;时事热点评论等。发音合作人从这些话题中自选若干个进行讲述。对话是4名发音合作人在上述话题中自选话题进行对话。

4) 调查方法

采用规定的录音设备、软件进行录音,辅之以录像和照片。并且,对调查内容还要按照一定要求进行必要的国际音标转写和汉字转写。最后,对一个调查点的所有调查资料和电子文件进行命名、分类、归档。

5) 建库与开发

国家组织专人进行验收,将调查材料统一建档入库。并及时组织专家对数据库进行开发,例如:编写《中国语言国情报告》,其内容包括中国语言、方言、地方普通话的基本数据,急需保护的方言、方言目录,与社会稳定和国家安全相关的“关键语言”和“关键方言”目录等。绘制详细的多媒体语言地图。通过网站、开放实验室、博物馆等方式对社会提供服务,最大限度地让这些数据造福国家与学界。

(二) “地方普通话”调查

地方普通话是由方言向普通话发展过程中呈现的各种中介语(inter-language),它在当前跨地区交际中发挥着重要作用,也是一种重要的语言资料。了解地方普通话的状况,探索地方普通话由低级向高级的发展规律,总结地方普通话的“板结”特点,对于推广普通话、研究语言学习规律等都具有重要意义。

地方普通话调查与语言和方言调查一同进行。其选点要求也与语言和方言的调查相同,原则上是一县一点。每调查点选择3名发音合作人,这3名发音合作人的普通话水平处在不同的等级:1名相当于《普通话水平测试等级标准》^④规定的三级甲等,1名仅次于三级乙等,1名普通话水平最差,接近方言。

地方普通话的调查内容是:1.用普通话讲述规定故事《牛郎和织女》2.用普通话朗读两篇短文《诚实与信任》《大学生村官》。调查得到的音频文件等经整理验收,建档入库,形成地方普通话语料库。地方普通话语料库是中国语言资源有声数据库的一个分库。

二 中国语言资源有声数据库的若干特点

我国的语言、方言研究,取得了许多标志性成果,如《中国语言地图集》《现代汉语方言大词典》《普通话基础方言基本词汇集》《现代汉语方言音库》《汉语方言地图集》《中国新发现语言研究丛书》等。中国语言资源有声数据库,充分借鉴我国以往的语言调查方法和研究成果,同时也形成了自己的一些设计特点。

1) 具有语言普查的性质

重视田野语言调查是我国现代方言学和民族语言学的传统,但过去的调查,选择调查点一般是着重“典型性”,多是选取民族语言、汉语方言的代表点。“典型性”选点,对于了解一种语言或一种方言的基本情况是适宜的,但是对于了解全国的语言状况常有缺憾。时至今日,我国还有许多县域的语言没有调查过或全面调查过,民族地区许多地方的汉语方言并未得到很好调查,东北、内蒙地区的汉语方言因被认为不具典型性而调查得很不深入,边境地区的语言状况并未做到县县清楚。显然,只把眼光聚焦在民族语言、汉语方言的代表点上,并不利于对一个语言或方言的深入了解,也不利于准确了解一个语言或方言的地域变化,画不出详细到县域单位的语言地图。

中国语言资源有声数据库的数据采录,原则上是“一县一点”,估计实际选点会达到4000以上,这就使调查具有了语言普查的性质。可以相信,这种“普查”会发现许多新现象,得到不曾注意到的语言或方言的新线索。孙宏开、胡增益、黄行主编的《中国的语言》记录了很多新发现的语言,这些新发现语言多是在突破了“典型性”选点的情况下才能被发现的。此外,普查式选点可为每一个县域单位留下一份21世纪初年的语言资料,为以后的语言国情研究树立了一个永恒的测量点。

2) 重视年龄因素和性别因素

年龄和性别是与语言和语言活动最为相关的两大因素。不同年龄、不同性别的人,在语音、词汇、语法、语用等方面会表现出不同的差异,其中年龄形成的老派、新派的语言差异,还往往预示着语言发展的方向。传统的语言调查常以老年男性为发音合作人,以求得到最为“纯正”、古老的语言面貌。这是一种学术旨趣。

中国语言资源有声数据库在着重采集老年男性的语言数据的同时,还采集老年女性、青年男性、青年女性的语言数据。这些数据将来在研究中可以处理为不同的分库,从而推进性别因素、年龄因素的语言研究。资料长期积累下来,通过不同年龄的语言差异研究,还可以对语言发展走向进行科学推演。

3) 重视语法

传统的汉语方言研究,精力基本集中在语音和词汇上。近年来,方言语法的研究兴趣大增,甚至还进行了跨方言的专题研究。中国语言资源有声数据库的语法调查,根据类型学的研究成果和我国语言的实际情况,设计出50个句子,以调查汉语和民族语言的若干重要的语法点。这些数据隐含着类型学的背景,因此不仅可以对汉语各方言进行比较研究,而且也可以对我国的各语言进行比较,甚至可以同世界上的许多语言进行比较,进而获得类型学上的成果。这种类型学上的语法调查,对于我国语法学的发展有积极意义^⑤。

4) 立足话语

传统的语言调查基本上是语言结构调查,较少涉及话语;即便做了话语调查,那也是作为

语言结构调查的附属进行的。中国语言资源有声数据库,则把话语作为基本数据进行采集,因为它的基本理念是保存当下的语言实态。最能反映语言实态的是话语,保存语言样本的最好方式是保存话语。话语是语言的“实态样本”,蕴含着语言结构的各种要素,还有包括语流音变在内的语言结构各种要素的语流变化;有语用学感兴趣的相关内容,如独白与对话的差异、话轮转换、话题关联、合作原则与礼貌原则的具体体现等;话语还负载着社会生活、风俗文化和当地人的喜怒哀乐等,这些口语文化是中华民族文化的重要组成部分,而且调查较少,保存不易。语言调查的重心放在话语上,是一种颇有意义的尝试。

5) 开展地方普通话调查

地方普通话是现实语言生活的一种实态,是语言学习产生的一种样态。近年来对地方普通话有些讨论,地方普通话开始被纳入研究的视野,但是,总体上看对它的研究还十分薄弱。地方普通话的语言性质、语言地位还有待确定,地方普通话的调查方法还有待探讨,地方普通话的面貌还有待描写,甚至“地方普通话”这一名称也还有待确定。地方普通话的调查是一项填补空白性的工作。

6) 充分运用信息化成果

使用录音手段辅助进行语言调查,并不新鲜,甚至一些先行者还为一些方言、语言建立了音档,但总体来看,多数语言调查还是以笔头记录为主,录音多是“备忘”性质的。中国语言资源有声数据库的数据采集,基本采取录音的形式,补之以照相录像,笔录反成了“备忘”性质的。录音质量是有声数据库关注的核心技术指标,为了保证录音质量,对录音的器材、软件、环境等都做了统一要求。同时,为了方便调查、验收和建库等工作,还专门设计了系列的工作软件。每一个参与中国语言资源有声数据库调查的人员,都需要进行信息技术的专门培训。这有望推进语言调查迈上信息化的新台阶。

7) 统一标准

中国语言资源有声数据库建设强调统一标准,为此花大气力制定了工作规范和技术规范,并不断通过试点进行完善。工作规范包括选点、遴选发音合作人、调查人员的学术素养、调查的程序、数据采集的环境、验收、建库等一系列规范。技术规范主要是调查用的字表、词表、语法表、话语方式、话语调查使用的话题和篇章材料、国际音标的应用、方言字的规范、录音技术、建库技术等一系列规范。采用同一的工作规范和技术规范,可以保证工作质量,可以使各调查点的数据合榫对接,可以对数据进行统计分析,可以最大限度地共享各地的调查成果。同时,这些工作规范和技术规范一旦被其他研究项目采用,这些项目的语言数据便可以合成到中国语言资源有声数据库中,实现数据库的更新扩展。

当然,这些标准是国家有声数据库的标准。各地在保证用统一标准完成国家库数据采集的同时,还可以根据各地情况增加语言和文化方面的更多内容,形成各有特色的语言数据库。

三 中国语言资源有声数据库的作用

用现代信息技术大规模采集中国语言数据,建成中国语言资源有声数据库,对我国语言生活规划和语言科学研究将产生极其重要的影响。下面简述中国语言资源有声数据库可能发挥的一些作用。

(一) 全面掌握语言国情,制定科学的语言规划

语言生活是社会生活的重要内容,语言生活状况是一种基本国情。语言是人类用于交际

和思维的最为重要的符号系统,同时也是国家的重要资源,是影响社会安定和国家安全的重要因素。促进国家通用语言文字的国内推广和国际传播,自觉保护和充分开发利用国家的语言资源,妥善处理中国境内各语言(方言)之间错综复杂的关系,维护公民的语言权利,向社会提供高质量高效率的语言服务,构建和谐的语言生活,这些都是21世纪中国语言规划必须考虑的内容。而要制定符合国情的语言规划,必须对语言国情有全面而深入的了解。

半个多世纪以来,我国开展过数次规模不等的语言和方言调查,为国家语言政策的科学制定与有效实施,发挥了重要作用。1956年,根据国务院指示开展了汉语和民族语言普查。共普查了1849个县市的汉语方言;并组成七个民族调查队,调查了主要民族地区的语言。这次语言普查,对于推广普通话和汉语规范化,对于少数民族文字的改革与创制,对于民族身份的认定等,都起了重要作用。1999年,教育部、国家语委等11部(委)联合开展了中国语言文字使用情况调查,调查采用入户问卷的调查方式,涉及全国1063个县(市、区),直接被调查对象47万多人^⑥。这次调查获得了我国语言文字使用的一些基本数据,为当今的语言决策提供了重要支撑。除了这两次大的语言调查之外,我国学者还持续进行了汉语方言、民族语言、海外华语的调查研究,取得了许多重要成果。

但是语言生活异常丰富复杂且与时而变,半个多世纪以来所进行的各种语言调查,或因时间已久,或因调查理念、调查目的、调查方式(包括调查技术)、调查领域等限制,难以较好反映语言国情。可以说,今天对语言国情的掌握还是粗线条的,许多方面是模糊不清的。语言国情不明,语言决策就会更多地依赖历史惯性,更多地依赖相关领域决策的迁移,更多地依赖“大脑实验室”的推演奇迹,这当然不适应日行千里的国家发展形势,难以驾驭纷繁日新的语言生活。随着中国社会经济的快速发展和语言状况的急剧变化,语言国情的调查已成为一项紧迫任务。中国语言资源有声数据库采取“一县一点”的普查性数据采集,调查境内的各种语言、方言的面貌和地方普通话的情况,可以获取语言国情的基础数据,有效支撑国家的语言规划。

(二) 促进普通话的科学推广

自清末开始提倡国语统一^⑦,现代汉民族共同语逐渐形成并推广开来,进而成为国家通用语言。现在全国60%以上的人口能够使用普通话^⑧,青少年人群中能够使用普通话的比率更大。在普通话不胫而走的新时期,推广普通话需要有新的方略。这新的方略就是:调整工作重点;提高推普效率。

首先,要重点关注一些特殊地区和特殊群体,比如农村和西部地区,比如学前至小学阶段的儿童等。农村和西部地区是普通话推广较为薄弱的地区,而农村的发展,西部的大开发,都要求让那里的人民掌握国家通用语言。学前至小学阶段是普通话学习的关键时期,这是语言学习规律告诉我们的。在这一时期打好了普通话(特别是普通话语音)基础,就不需要在成人期花费很高的社会成本推广普通话。

其次,是提高普通话学习的效率,包括普通话学习的速度和质量。中国语言资源有声数据库,有助于了解汉语各方言、各民族语言的特点,有助于了解各地“地方普通话”的特点及由方言学习普通话的规律,这无疑可以大大提高普通话推广的效率,提高普通话学习的质量。

(三) 保存语言实态

语言及其方言是文化的重要载体,又是文化的重要组成部分。中华语言及其所负载的文化,构成了中华文化的基础。中国语言资源有声数据库将当今的语言实态记录下来,就是为中华民族留下了一部数字化的可永久保存的口语样本,留下了一幅可以进行数据加工的“中国

语言地图”。在现代信息技术尚未发展起来的时代,人们只能用笔记记录语言和方言,损耗了大量的语言信息,也不可能对语言进行实态保存。而今现代信息技术和数据库技术等已经获得长足的发展,有条件将这些技术用于语言调查,使语言能以语音这种“原态”方式永久保存。

口语的特点是:保留着许多古老的语言成素和文化内容,但是口语在代代相传中又发展变化迅速,它的许多语言成素和文化内容会在较短的时间内流失。就此而言,对口语的保存永远都是具有“抢救”性质的工作。特别是近几十年来,随着政治经济、文化教育、大众传媒、通讯和交通事业的迅速发展,我国的少数民族语言和汉语方言已经发生了许多变化,语言的结构要素(特别是词汇)在变化,语言的使用功能在变化,使用语言的人群在变化。特别是亿万农民进城务工,我国的城市化进程在以加速度的方式前进,中国的“语言地图”在不长时间内将会大幅度地改写,许多小语言、小方言会急剧地萎缩甚至消亡,中华语言资源面临着急剧流失的危险。语言濒危已经引起国际社会的广泛关注,人类文化的多样性因语言濒危而受到前所未有的威胁,有人警告到21世纪末,人类将失去90%的语言!^④在这样的国际国内背景之下,中国语言资源有声数据库的建设,显然具有抢救记录和保存语言资料的功能,具有保护中华语言文化遗产的功能。

(四) 推进我国语言科学的发展

中国语言资源有声数据库的建设,将对我国语言科学的发展起到推进作用。就语言调查而言,它重视了年龄因素和性别因素对语言的影响,把年龄、性别列入遴选发音合作人的基本条件,在语言观上这是不把语言看作匀质的系统,而是看作有差异的、时刻处于发展变化之中的系统。它使语言结构的调查圆满起来,把语法数据的收集放在适当的位置,补充了以前侧重于语音、词汇的做法;而且这些语法项目是在语言类型学的参数指导下设计的,保证了语法数据的系统性和可比性。它把话语调查作为重点,兼顾独白和对话,兼顾语言和文化,从而使语言调查由抽象的结构要素层面进入到具体的话语实际层面。

中国语言资源有声数据库,以现代信息技术作为数据采集、加工、存储的基本手段,为语言调查遴选出、开发出系列的软硬件,使语言调查在技术上迈到一个新阶段。参加数据库调查的人员,虽然需要很好的听音记音能力,但是更需要掌握相关的现代信息技术,通过数据库的建设,也为学界培养了一支新型的学术队伍。

中国语言资源有声数据库存储着海量的原始数据,学界在很多方面可省却实地调查之劳;学界共同对这一数据库做多方面的长期开发利用,会得到大量的直接和间接的成果。语言学的发展,在很大程度上依赖于语言事实的采录与加工。中国语言资源有声数据库在语言数据采录方面走在了时代前列,如果根据共享原则做好开发利用,这个数据库就是一个学术宝库。

中国语言资源有声数据库把语言调查从结构推进到话语,把调查技术从笔头推进到现代信息技术,把海量原始数据积聚起来供学界长期开发,并在实践中培养起一支新型的学术队伍,这明显地能够把我国的语言科学向前推进一步。

(五) 其他方面的作用

中国语言资源有声数据库不只是为语言学而作,对这些资源可以进行学术开发,更需要进行行政开发。它也不只是为语言文字工作而作,其作用是多方面的,面向全社会的。比如:用有声数据库训练机器,可以帮助机器识别各地的方言、各种语言和各地普通话,由此可以大大提高语言信息化的水平。语言信息化的这些成果,可以用于语音识别、语音合成、人机语音交换的多种领域,促进生产和生活的信息化。这些成果用于军事、公安、边防等领域,可以为

公安侦破、国家安全做贡献。

再如:我国有 30 多种跨境语言,如朝鲜语、赫哲语、鄂伦春语、鄂温克语、达斡尔语、蒙古语、维吾尔语、哈萨克语、柯尔克孜语、塔吉克语、塔塔尔语、乌兹别克语、俄罗斯语、藏语、门巴语、珞巴语、景颇语、傣语、布依语、哈尼语、傈僳语、佤语、拉祜语、怒语、独龙语、德昂语、壮语、瑶语、京语等。迄今为止,对这些语言的跨境分布情况和各种使用情况的了解还很不充分,而深入了解这些语言,科学运用这些语言,是和边睦邻的关键,也是为国境铺设的一条新的安全线。中国语言资源有声数据库有助于对跨境语言的了解。

又如:中国语言资源有声数据库是依照一定话题采集话语数据的,采集的这些话语也可以看作是文化素材,因此,中国语言资源有声数据库也可以看作是中华口语文化库,可以对它进行文化开发。比如将各地讲述生老病死的话语素材标记出来,就可以获得各地关于生老病死的民风民俗,将各地讲述农事活动的话语素材标记出来,就是中国各族人民的农事风景画。如果在此基础上,一些有条件有兴趣的地方将当地口头文化(如戏曲、歌谣、民间传说、特色文物等)再行搜罗,并配以图片、录像等视频材料,便形成当地的语言文化数据库。这一语言文化库以当地的语言为内核,以话语为中轴,以本地文化为表层,会成为地方用于文化保护和博物展览的文化建设的重要工程。

附 注

① 承担这些项目的学者主要有:曹志耘、戴庆厦、郭龙生、何瑞、黄行、李如龙、刘丹青、潘悟云、乔全生、魏晖、谢俊英、徐大明、张振兴等先生。组织开展这些项目的主要人员有李宇明、王铁琨、陈敏等先生。

② 中国语言资源有声数据库建设领导小组办公室编写。《调查手册》的汉语方言部分(曹志耘执笔),已由商务印书馆出版。除了项目承担者和项目组织者之外,顾黔、侯精一、刘俐李、孙茂松、汪平、杨尔弘、赵晓群等先生,为调查手册和工作规范的完善从不同的方面做出过贡献。

③ 《牛郎和织女》的故事文本,主要由汪平先生整理。

④ 见国家语委普通话培训测试中心编制的《普通话水平测试实施纲要》,商务印书馆,2008年。

⑤ 语法部分的主要设计者是刘丹青先生,50个例句的设计参考了不久前他出版的《语法调查研究手册》。该手册以科姆里(Bernard Comrie)、史密斯(Norval Smith)编制的《Lingua版语言描写性研究问卷》为提纲,对问卷的内容进行详细的注释、例示、补充和分析,同时也指出了其不足或不适合中国语言的地方。《Lingua版语言描写性研究问卷》,声言为各地人类语言的调查描写提供了一个尽量客观、全面和包容性强的语法框架,让调查研究者可以尽可能摆脱语种局限和学派成见,调查到尽可能多的语法事实。

⑥ 详情参看中国语言文字使用情况调查领导小组办公室编《中国语言文字使用情况调查资料》,语文出版社,2006年。

⑦ 见《清末文字改革文集》(文字改革出版社,1958年)载吴汝纶《东游丛录》《学部中央教育会议议决统一国语办法案》。

⑧ 20世纪末的统计数据,全国能够使用普通话的人数为全民的 53.06%(见《中国语言文字使用情况调查资料》)。而今又过了 10年,能够使用普通话的人数应当超过 60%。

⑨ 参看徐世璇《濒危语言研究》,中央民族大学出版社,2001年。

参考文献

薄守生 赖慧玲 2009 《当代中国语言规划研究》,中国社会科学出版社。

曹志耘主编 2009 《汉语方言地图集》,商务印书馆。

陈章太 李行健主编 1996 《普通话基础方言基本词汇集》,语文出版社。

陈章太 2008 《论语言资源》,《语言文字应用》第 1 期。

- 达·巴特尔 2007 《论语言资源保护》,《内蒙古社会科学》第6期。
- 戴庆厦主编 1993 《跨境语言研究》,中央民族学院出版社。
- 主编 2009 《中国少数民族语言研究60年》,中央民族大学出版社。
- 邓晓华 王士元 2009 《中国的语言及方言的分类》,中华书局。
- 范俊军 肖自辉 2010 《国家语言普查刍议》,《语言文字应用》第1期。
- 国家语言文字工作委员会普通话培训测试中心 2008 《普通话水平测试实施纲要》,商务印书馆。
- 侯精一主编 1994-1999 《现代汉语方言音库》,上海教育出版社。
- 李荣 熊正辉 张振兴主编 1987,1990 《中国语言地图集》(中国社会科学院和澳大利亚人文科学院合作编纂),香港朗文(远东)有限公司出版。
- 李如龙 2008 《汉语方言资源及其开发利用》,《郑州大学学报(哲学社会科学版)》第1期。
- 李宇明 2009a 《中国语言规划论》,商务印书馆。
- 2009b 《中国语言规划续论》,商务印书馆。
- 刘丹青 2008 《语法调查研究手册》,上海教育出版社。
- 孙宏开 胡增益 黄行主编 2007 《中国的语言》,商务印书馆。
- 王辉 2007 《语言规划的资源观》,《北华大学学报(社会科学版)》第4期。
- 王世凯 2009 《语言资源与语言研究》,学林出版社。
- 文字改革出版社 1958 《清末文字改革文集》,文字改革出版社。
- 徐大明 2008 《语言资源管理规划及语言资源议题》,《郑州大学学报(哲学社会科学版)》第1期。
- 徐世璇 2001 《濒危语言研究》,中央民族大学出版社。
- 张普 王铁琨主编 2009 《中国语言资源论丛(一)》,商务印书馆。
- 中国语言文字使用情况调查领导小组办公室编 2006 《中国语言文字使用情况调查资料》,语文出版社。
- 中国语言资源有声数据库建设领导小组办公室编 2010 《中国语言资源有声数据库调查手册(汉语方言)》,商务印书馆。
- 周洪波 2007 《加大语言资源的开发力度》,《长江学术》第1期。

(李宇明 北京 教育部语言文字信息管理司 100010)

黎锦熙先生诞辰120周年纪念暨学术思想研讨会在京召开

由北京师范大学文学院与教育部人文社会科学重点研究基地北京师范大学民俗典籍文字研究中心联合举办的黎锦熙先生诞辰120周年纪念暨学术思想研讨会,于2010年4月10日至11日在北京师范大学隆重举行,参加会议的有来自全国各地近40所高等院校和科研院所的百余位学者。

此次会议的话题除了传统的语法之外,比较集中的还有语文现代化、修辞学、母语以及第二语言教学等。微软亚洲研究院的黄昌宁教授结合自然语言处理的实际情况和需求,认为句本位和中心词分析法具有普遍语法的意义。中国社会科学院语言研究所的刘丹青教授在当前语法理论和汉语语法研究不断深化的背景下,重新解读了黎氏语法中补足语、(实体词的)位、代名词以及指示形容词等重要术语。安徽大学曹德和教授对黎氏语篇分析思想和实践进行了总结与探究,认为1962年出版的《汉语语法教材》是汉语语篇学的开山之作。

与会学者认为,黎锦熙先生的所有语文研究和实践都围绕中国语文的现代化。黎先生的学术思想特别是语法研究理论,应当花大气力去进行总结和发掘。

此次会议论文将结集出版。

(刁晏斌)

other historical paths of MC first and second grades of *Xie* group in Wu dialects and the necessity of giving a distinction between the external phonological superposition and the internal phonological diffusion. Finally, it provides clues to distinguish both phonological changes.

Key words external phonological superposition, internal phonological diffusion, strata, steps-broken, Wu dialects, MC first and second grades of *Xie* group

CHANG Kuang-yu The comparative method in Chinese historical phonology

The comparative method as was practiced by the Swedish Sinologue Karlgren in Chinese historical phonology is a great departure from the mainstream represented by Meillet in Europe and Hoenigswald in North America. In the area of written records, his overly subjective interpretations are also, in many fundamental ways, at odds with his colleagues in China. These factors converge to derive a picture of reconstructions without a concomitant linguistic continuity.

Key words comparative method, West Europe, North Europe, North America, China

LU Bingfu, JIN Lixin On the two explanatory models of implicational description: A case study of the correspondence between description and explanation

This paper explores the implicational description and proposes two explanatory models: one is based on likeliness hierarchy and the other on the two-factor interaction. If the implicant and implicatum constitute a minimal pair in properties, the explanatory model of likeliness hierarchy is preferred; otherwise the model of two-factor interaction is better. The former is simpler than the latter in nature.

Key words implication, likeliness hierarchy, interaction, minimal pair, tetrachoric table

TAN Jingchun Semantic relations between nouns and noun modifiers and their roles in dictionary definition

Firstly, this paper discusses semantic relations between nouns and their noun-modifiers. With the transformational analysis and the concept of self-reliance in word sense, a noun and its noun modifier can be regarded as a simple combination by the connection with implied predicates, which leads to the complication of semantic relations of the noun phrases. Twelve types of semantic relations are summarized to explain that the word meaning, contexts and common knowledge are the main factors in semantic relation interpretations. On the basis of the semantic relations, a model of dictionary definition can be proposed and current inaccuracies in dictionary definition identified.

Key words noun modifier, semantic relation, implied predicate, dictionary definition

LI Yuming On the establishment of the audio database of Chinese language resources

The audio database of Chinese language resources is part of the language engineering project of the National Committee of Language, which aims to record, file and preserve the original states of dialects and regional lingua franca in each county in China. This database may play an important role in language situation providing, language planning, language resource preserving and developing, Mandarin popularization, and linguistic studies.

Key words Chinese, language resource, audio database