

中国语言资源论丛

ZHONGGUO YUYAN ZIYUAN LUNCONG

主编 ● 张普 王铁琨



商务印书馆

中国语言资源论丛

(一)

主 编：张 普 王铁琨

商 務 印 書 館

2009 年 · 北京

图书在版编目(CIP)数据

中国语言资源论丛(一)/张普,王铁琨主编. —北京:
商务印书馆,2009

ISBN 978-7-100-06683-9

I. 中… II. ①张…②王… III. 社会语言学—中国—
文集 IV. H1-53

中国版本图书馆 CIP 数据核字(2009)第 093890 号

所有权利保留。

未经许可,不得以任何方式使用。

ZHONGGUÓ YǔYÁN ZīYUÁN LùNCóng

中国语言资源论丛

(一)

主 编: 张 普 王铁琨

商务印书馆出版

(北京王府井大街36号 邮政编码 100710)

商务印书馆发行

北京瑞古冠中印刷厂印刷

ISBN 978-7-100-06683-9

2009年9月第1版

开本 850×1168 1/32

2009年9月北京第1次印刷

印张 12 1/4

定价: 26.00 元

编辑委员会

主 编：张 普 王铁琨

执行主编：杨尔弘

编辑委员会成员：（按音序排列）

侯 敏 苏新春 王铁琨

杨尔弘 张 普 周洪波

目 录

序言.....	周有光(1)
前言.....	张普 王铁琨(3)
论语言资源.....	陈章太(7)
国家语言资源发展战略研究	徐大明(18)
语言使用实态考察研究与语言规划	
——发布年度语言生活状况报告的思考	王铁琨(32)
语言资源整合与对比研究	
——以两岸词汇对比研究为例	黄居仁(52)
国家语言资源建设和现代汉语规范化及汉语教学 ...	胡明扬(67)
作为第二语言教学的汉语教学必须重视书面语教学 ...	陆俭明(70)
喀卓人的双语关系	
——兼论双语关系在语言国情研究中的地位	
..... 戴庆厦 蒋颖 邱月 常俊之 赵燕珍	(82)
论汉语方言的资源及其开发利用.....	李如龙(104)
两岸四地语言文字使用现状与汉语国际推广.....	陈荣岚(117)
新加坡双语教育政策的挑战与新机遇.....	吴英成(131)
试论在新马两国推进华语文传播的策略.....	汪惠迪(154)

论新词语的短期消亡·····	王宁(168)
共时、同题语料库和社会文化·····	邹嘉彦 邝蔼儿(174)
论语言的稳态·····	张普(188)
关注语言文字差错,营造国民用好母语氛围·····	华绍和(207)
教材语言的性质、特点及研究意义 ·····	苏新春 杜晶晶 关俊红 郑淑花(213)
语言监测需要多功能、易维护的自动分词标注系统 ·····	侯敏 胡凤国 滕永林 刘伟(224)
蒙古语语言资源平台架构设想·····	那顺乌日图(236)
维吾尔语语音语料库的设计研究 ·····	吾守尔·斯拉木 那斯尔江·吐尔逊(249)
论面向信息处理现代维吾尔语资源库构建 ·····	玉素甫·艾白都拉 潘伟民 阿布都热依木·沙力(256)
动名兼类词及其词性标注规则 ·····	黄昌宁 李玉梅 新光瑾(268)
基于计量研究的现代汉语常用词库的构建 ·····	俞士汶 朱学锋 支流(289)
自然语言处理中的理性主义和经验主义·····	冯志伟(302)
关于语文教科书编选问题的思考·····	顾之川(323)
现代汉语词汇通用度及其计算方法研究初步 ·····	赵小兵 张普 唐宁 田寄远(332)

对外汉语初级教材义类分布研究

——以“性质与状态”义类为例

…… 柯丽芸 杨艳 杜晶晶 李安 田立宝 苏新春(345)

基于动态流通语料库通用词语的字词语关系考察

…………… 韩秀娟(354)

基于词汇时间分布信息的未登录词提取…………… 何伟 侯敏(368)

关于语言资源的几点思考(闭幕词)…………… 崔希亮(380)

序 言

2007年9月,教育部语言文字信息管理司和北京语言大学共同主办了一个高峰论坛,讨论“国家语言资源与应用语言学”的问题。主题是“国家语言资源”,提出:语言,包括文字,是人文社会资源,是国家资源,是人类资源,需要加以保护、开发和利用,进行调查、测量和研究。应用语言学应当研究语言资源。这些概念和观点,使人耳目一新。

近年来,教育部语言文字信息管理司,建立专题中心,进行语言资源的调查、研究和分析,每年把成果编成《中国语言生活状况报告》;提供共时的描写资料,便利历时的比较探讨。教育部做的这些工作,实际开辟了一个语言资源研究的新园地。

本书中的论文和资料,都是专门研究的成果,应用了新的理论和新的方法。研究的语种包括:汉语、少数民族语言、华侨地区语言、濒危语言等。研究重视语言生活的实际状态,分别动态和稳态,动态反映语言的演变,稳态传承语言的功能。这些研究,有理论意义,有实用价值。语言文字学呈现了新气象。

现在是21世纪,21世纪是全球化时代,全球化是人类社会发展的新阶段。在新阶段中,一切事物都要“重新认识、从头超越”。厚今而不薄古,重中而不轻外,既需继承,更要创新。时代在前进,

语言文字学不能不“与时俱进”。

我将追随在语言文字学界同行们之后，共同学习，共同前进。

周有光

2008-08-08

时年 103 岁

前 言

由教育部语言文字信息管理司和北京语言大学共同主办,国家语言资源监测与研究中心平面媒体分中心(北京语言大学应用语言学研究所)承办的“国家语言资源与应用语言学高峰论坛”于2007年9月12—14日举行。本论文集是该论坛的特约报告和论文报告的精选,还收入了几位青年学者的研究报告。

召开本次高峰论坛的目的是为了更好地推进国家语言资源监测与研究的实践,进一步推动中国应用语言学学科建设的深入发展。

2004年到2005年期间,教育部语言文字信息管理司先后与国内五所大学签署了共建国家语言资源监测与研究中心平面媒体分中心(北京语言大学)、有声媒体分中心(中国传媒大学)、网络媒体分中心(华中师范大学)、教育教材分中心(厦门大学)、海外华语研究中心(暨南大学)的协议。五个分中心都分别建立了学术委员会,制定了相关章程,学术委员会对申报的监测项目分别进行了评议,有的中心的项目还进行了中期考核、结题与验收。但是五个分中心学术委员会还没有一次机会坐在一起,认真对国家语言资源、语言资源的监测、监测的理论、方法、结果等问题进行一次专门的充分的集中的讨论。这次“国家语言资源与应用语言学高峰论坛”,是五个分中心的学术委员会(与会委员共计53人,其中中国

大陆地区 46 人,香港地区 2 人,海外 5 人)首次一起开会,并特约了语言学及应用语言学、“语言资源联盟”等有关国内外专家,就国家语言资源与应用语言学科发展等重大问题进行专门研讨。

国家语言资源监测与研究,是一项面向语言应用的实际状态(可称为语言的“实态”)的监测与研究。监测的规模是空前的,监测的实践是动态的,监测的方法和技术都是需要不断探索的,监测结果的连续发布在国际上也是没有先例的。国家语言文字工作委员会以“中国语言生活绿皮书”的形式连续发布了《中国语言生活状况报告 2005》、《中国语言生活状况报告 2006》,《中国语言生活状况报告 2007》也将于 2008 年发布。国家语言资源监测与研究,中心每年监测的语料约 10 亿汉字符次(12 亿字符次)的数据量,无论监测的理论还是监测的方法,以及监测的汉语信息处理技术和设备,都随着监测的实践的推进而不断发生变化,我们的认识也在不断深化。因此,召开一次“国家语言资源与应用语言学高峰论坛”的研讨会,延揽国内外的相关专家学者,深入研讨一些问题,已经是国家语言资源监测科学、持续、深化发展的需要。正是在这样的背景下,教育部语言文字信息管理司和北京语言大学共同发起并主办了这次高峰论坛,也得到了国家语言资源监测与研究中心的各分中心的响应与支持。

本次会议以“国家语言资源与应用语言学”为主题,具体议题主要包括以下内容:

1. 国家语言资源建设、监测与研究
2. 国家语言资源与语言服务
3. 国家语言资源与语言规划
4. 国家语言资源与语言文字规范

5. 国家语言资源与语言文字社会应用
6. 国家语言资源与母语教学
7. 国家语言资源与国际汉语传播
8. 国家语言资源与民族地区汉语教学
9. 国家语言资源与语言信息处理
10. 其他与会议主题相关的论题

把语言资源作为一种国家资源,是从一个新的视角来看待语言,自然就产生了许多新的研究内容。召开此论坛的目的就是请专家们这样一个新的视野下,探讨其中的新问题、新发现、新方法、新技术和新理论,进一步推动国家语言资源的监测研究与应用语言学的发展。

会议收到特邀报告、论文报告 50 篇,高峰论坛后又经作者认真修改,精选了其中的 28 篇,收入本论文集,并依据论文的内容,大致作了分类。对于语言资源最宏观的、战略的论述,放在第一类。

高峰论坛于 9 月 12 日上午开幕,北京语言大学崔希亮校长、教育部语信司李宇明司长作为主办方领导致开幕词,胡明扬先生代表专家讲话,然后进入特约报告和论文报告的研讨。值得特别提出的是,本次论坛于 9 月 12 日晚专设了青年学术骨干及学生报告专场,来自中央民族大学、香港浸会大学、中国传媒大学、厦门大学、北京语言大学、第二外国语学院等从事国家语言资源监测、语言教学与研究的 11 位青年学者报告了论文,引起热烈的关注。受条件限制,我们不能将这些论文悉数收入本论文集,希望创造其他的机会让这些研究成果能与广大读者见面。

尤其值得一提的是 9 月 13 日下午的“自由论坛”,这是与其他的学术研讨会完全不同的一种学术上的自由论坛,是本次高峰论坛

坛专门安排的集中的长达三个小时的交互讨论。在通常的学术报告过程中,每篇报告后都有简单的提问和讨论,一些新鲜的热门的大家关注的问题常常有很多人想提问、发言或讨论,但交互时间不够,往往就已经有按捺不住的感觉。因此大家早就盼望在“自由论坛”的集中交互时间,能够充分对一些重大的问题、感兴趣的问题进行深入讨论,甚至于交锋。会议提倡学术平等,人格尊重,观点交锋。受国家语言资源与应用语言学高峰论坛两位主席的委托,“自由论坛”由王铁琨、苏新春、侯敏、周洪波和张普五位主持,每人主持半小时,主持人做开场发言 5—8 分钟,交互发言 3—5 分钟。自由论坛发言热烈,讨论激烈,发人深思,频频有思想的火花碰撞出来,全部发言均由现场两位速记员实录,并经本人修改。我们希望这个实录的“自由论坛”另外有机会单独发表。

自由论坛结束后,教育部语言文字信息管理司王铁琨副司长作大会总结,北京语言大学崔希亮校长致闭幕词。2007 国家语言资源与应用语言学高峰论坛在热烈欢快的气氛中,在紧张有序的节奏下谢幕了。

为了与更多的专家学者分享这次会议的成果,也希望有机会向更多的同仁求教,我们整理出版了此论文集。

论文集的编辑委员会成员是:侯敏、苏新春、王铁琨、杨尔弘、张普、周洪波。各篇入选论文都经作者授权,文责自负。

张普 王铁琨

2008 年 3 月 29 日

论语言资源

陈章太 教育部语言文字应用研究所

引言

0.1 语言是不是资源,怎样看待语言资源,过去国外学界谈论多一些,国内学界讨论不太多。近些年来,这个问题逐渐引起我国语言学界和有关部门的关注,邱质朴(1981、2000)、黄行(2000)、李宇明(2003、2005)、张普(2006、2007)、孙宏开(2006)、徐大明(2007)等教授对语言资源问题进行了较多的研究与论述,发表了很好的意见。教育部和国家语委对这一问题十分重视,从2005年起相继建立了“国家语言资源监测与研究中心”及多个分中心,对我国语言资源进行监测与研究,先后发表了多部重要翔实的监测研究报告,取得了很好的绩效。现在,学界正在讨论这个问题,2007年10月在北京语言大学举行的国家语言资源与应用语言学高峰论坛的主要议题,就是国家语言资源的建设与应用。

0.2 我国学界讨论的“语言资源”有广义和狭义之分,广义的“语言资源”是指语言本体及其社会、文化等价值;狭义的“语言资源”是指语言信息处理用的各种语料库和语言数据库,以及各种语言词典等。随着我国现代化、信息化的加速发展,社会对语言实际需求的加大,以及语言功能的变化,人们对语言的认识正在发生深刻的变化,从视语言为问题逐渐转向将语言看作资源,进而发出保

护建设和开发利用语言资源的呼吁,这有重要的意义,对我国语言生活和语言文字工作将产生积极影响。本文讨论的“语言资源”是广义的“语言资源”,还包括文字。具体讨论三个问题:语言是一种特殊的社会资源;我国的语言资源状况;语言资源的保护建设与开发利用。

一 语言是一种特殊的社会资源

1.1 所谓“资源”,是指有价值、可利用、出效益、能发展的事物。资源学通常将资源分为自然资源和社会资源两大类,下面再各分成若干小类。语言是一种符号系统,又是一种社会交际工具,它存在、依附于人类社会,与社会紧密相结合,为社会交际服务,因此被认为离开了社会无所谓语言。语言既是物质的又是社会的,而更多的是社会的。语言以它的物质结构系统,承载着丰富、厚重的社会文化信息,为社会所利用,能够产生社会效益和政治、经济、文化、科技等效益,所以是一种有价值、可利用、出效益、多变化、能发展的特殊的社会资源。

1.2 语言资源主要由语言本体和语言社会应用两部分构成。语言本体包括语音系统、词汇系统、语法系统和语义系统,是语言资源的物质基础;语言应用包括语言在社会各领域的应用,是语言资源价值和可利用性的具体体现。语言资源的价值有隐性的也有显性的,其隐性价值是语言本体的价值,具体包括语言地位,规范程度,语言历史,承载的文化、信息,记录的文献、资料等;显性价值是语言本体价值的具体体现,具体包括语言社会交际作用,使用人口,使用领域,应用效益等。语言资源的隐性价值要通过显性价值

加以体现,而所有语言资源价值的显现一般是逐渐、缓慢的,只有在语言功能、语言地位、语言作用发生大的变化之后,某些语言的资源价值才会有较快的显现。例如希伯来语,原来只是犹太人在宗教活动中使用的一种“死了的语言”,并不是犹太民族的日常用语。以色列建国以后,因为国家、民族、社会的需要,实行一个国家一种语言的政策,确定希伯来语为国语,从此改变了希伯来语的地位、功能和作用,希伯来语成了以色列法定的官方语言和犹太人的日常通用语,这样希伯来语的语言资源价值才从隐性转为显性,很快显现出来。又如马六甲海峡的一种廖内——柔佛方言(Riau-Johor),在1972年印度尼西亚与马来西亚协议共同采用以这种方言为基础的标准马来语为两国共同语以后,这种方言的地位、功能和作用等发生了重大变化,随之,这种方言的语言资源价值也才凸显出来。菲律宾的他加禄语(tagalog)在成为菲律宾国语之后,它的语言资源价值也很快显现。

1.3 语言资源价值的高低,与语言功能、语言活力和社会需求、国力状况等有密切关系。就语言功能和语言活力来说,凡是语言功能完备或比较完备,语言活力强盛或比较强盛的语言,如一般国家的官方语言,多数民族共同语和许多区域通用语,其语言资源价值就较高;反之其语言资源价值则较低或很低,如世界上和我国不少弱势或濒危的语言、方言。就社会需求和国力状况来说,凡是社会处于稳定、发展时期,社会对语言的需求就要加大,或是国家综合国力处于上升、鼎盛状态,这都会促使该国家社会主要语言的语言资源价值升高,如我国近期的崛起、发展与汉语普通话语言资源价值的攀升;反之,国家、社会发展得缓慢或停滞,其主要语言的语言资源价值则降低,如亚洲、非洲的有些国家与主要语言。

1.4 评价语言资源价值,可以参考语言活力评价指标体系,我以为具体指标大致有这样几项:(1)语言在社会中的地位与声望情况;(2)语言规范程度及表现力;(3)语言使用人口及其年龄结构与社会分布;(4)语言使用范围与应用领域;(5)语言记录文献和承载信息情况等。依据这个语言资源价值评价指标体系,大致可以将我国的语言资源分为以下五类:语言活力超强的超强势语言资源;语言活力强盛的强势语言资源;语言活力减弱的弱势语言资源;语言活力缺乏、处于濒危状态的超弱势语言资源;语言活力没有、已经消亡但仍有一定价值的语言资源。超强势语言资源和强势语言资源,一般都达到评价指标体系中各项或多项指标的较高标准,如国家、民族、地区的共同语,通用语言,官方语言和法定语言等。然而有的超弱势语言,甚至是消亡了的语言,因为具有评价指标体系中某些指标的高标准,如记录文献和承载文化信息较多,这样的语言也具有一定的资源价值,如我国历史上的契丹语、女真语和当代的满语,外国的梵语、巴利语、哥特语等。

二 我国的语言资源状况

2.1 上文说过,我国的语言资源大致有这样几类:超强势语言资源,强势语言资源,弱势语言资源,超弱势语言资源,消亡但仍有一定价值的语言资源。

2.2 语言活力超强的超强势语言资源。汉语普通话由于它形成的历史悠久,记录了极其丰富、珍贵的文献资料,现在使用人口众多,应用地域广阔,使用领域广泛,规范程度很高,语言功能完善,语言活力超强,应用效益显著,语言声望很高,语言影响广泛,