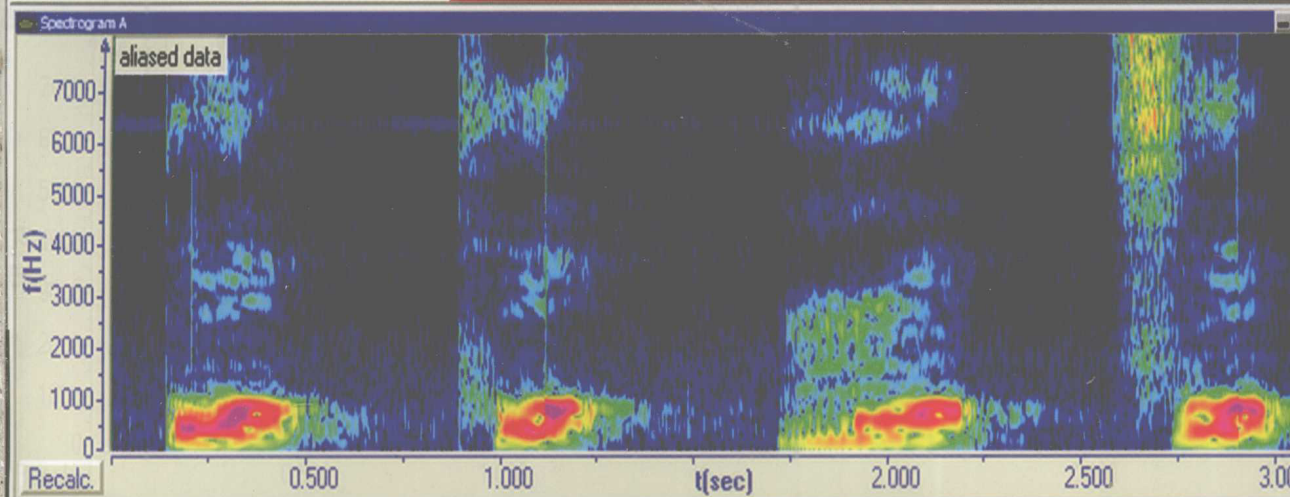


- 云南师范大学语言学及应用语言学重点学科系列丛书
- 云南师范大学重点学科建设经费资助

语言信息处理丛书



语言信息处理 与普通话水平测试

王渝光 杨瑞鲲 李洪平 著

云南大学出版社
Yunnan University Press

云南师范大学语言学及应用语言学重点学科系列丛书
云南师范大学重点学科建设经费资助

语言信息处理丛书

语言信息处理与普通话水平测试

王渝光 杨瑞鲲 李洪平 著

 云南大学出版社

图书在版编目(CIP)数据

语言信息处理与普通话水平测试/王渝光著. —
昆明:云南大学出版社,2010
(语言信息处理丛书)
ISBN 978 - 7 - 5482 - 0024 - 6

I. ①语… II. ①王… III. ①普通话—水平考试—研究 IV. ①H102

中国版本图书馆 CIP 数据核字(2010)第 026150 号

语言信息处理与普通话水平测试

王渝光 杨瑞鲲 李洪平著

责任编辑:邓立木

封面设计:猎鹰创想

出版发行:云南大学出版社

印 装:昆明市五华区教育委员会印刷厂

开 本:787mm × 1092mm 1/16

印 张:15.5

字 数:310 千

版 次:2010 年 1 月第 1 版

印 次:2010 年 1 月第 1 次印刷

书 号:ISBN 978 - 7 - 5482 - 0024 - 6

定 价:38.00 元

社 址:云南省昆明市翠湖北路 2 号云南大学英华园内

邮 编:650091

网 址:[http://: www. ynup. com](http://www.ynup.com)

E - mail: [market@ ynup. com](mailto:market@ynup.com)

电 话:(0871)5031071/5033244

序

语言信息处理,是用计算机对语言信息(包括口语和书面语)进行编码、转换、传输、解码、存储、分析等进行加工处理的科学。语言信息处理是一门与语言学、计算机科学、心理学、数学、控制论、信息论、声学、自动化技术等多学科相联系的交叉性边缘学科。人类利用语言进行交际的过程,实际就是人自身进行语言信息处理的过程。

就汉语而言,语言信息处理又叫中文信息处理,包括汉字信息处理和汉语信息处理两个部分。中文信息处理,是自然语言信息处理的一个分支。现阶段的中文信息处理技术,虽有了很大的发展,但离现代化的要求还有较大的差距。许多语言信息处理技术,既需要计算机技术的不断发展,更需要从语言信息处理的角度加强对汉语本身的研究。

语言文字是一个国家主权的象征,也是民族自信心的体现。现在世界上的各个国家、民族都比以往任何时候更加珍视自己的语言文字所具有的标志性意义。

语言文字是交流的工具,也是文化的载体,承载着丰富、厚重的社会文化信息,服务于社会的经济、政治和文化,是能够产生社会效益和政治、经济、文化、科技等效益的特殊社会资源,是国家最宝贵的财富。语言是人类重要的文化资源和经济资源。当今时代,语言已经进入到经济和高科技领域,成为经济发展的重要资源。随着信息时代的发展,语言作为经济资源的性质会体现得越来越明显,其经济意义越来越显著。

民族共同语是指一个民族内用作共同交际工具的语言。标准语是经过加工规范后的民族共同语,是民族共同语的规范化形式。确立民族共同语的标准,推广民族共同语是政治、经济、文化发展的需要,是一个国家实现工业化、信息化、现代化的需要。新中国成立后,中国共产党、中央人民政府对汉民族共同语的确立和推广十分重视。1955年进行了现代汉语规范问题的讨论。1956年2月6日国务院发布了《关于推广普通话的指示》,明确提出并规定“汉语统一的基础已经存在了,这就是以北京语音为标准音,以北方话为基础方言,以典范的现代白话文著作作为语法规范的普通话。”1958年周恩来总理在《当前文字改革的任务》中又把推行汉语拼音方案、推广普通话作为两项重要的任务。国家根据当时的情况,制定了“大力提倡,重点推行,逐步普及”的推广普通话的方针。1982年《中华人民共和国宪法》明确规定“国家推广全国通用的普通话”。实施普通话

水平测试是信息化、现代化社会发展的必然要求。1994年10月30日，国家语言文字工作委员会、国家教育委员会、国家广播电影电视部作出了《关于开展普通话水平测试工作的决定》，发布了《普通话水平测试大纲》和《普通话水平测试等级标准(试行)》，对普通话水平测试提出了具体的实施办法，对汉语普通话水平的等级体系进行了细化，明确提出了三级六等的客观的评判标准。通过几年的实践，《普通话水平测试等级标准(试行)》经国家语言文字工作委员会再次审定作为部级标准，于1997年12月5日正式颁布。2000年10月31日九届人大常委会通过了《中华人民共和国国家通用语言文字法》，第六条规定“国家颁布国家通用语言文字的规范和标准，管理国家通用语言文字的社会应用，支持国家通用语言文字的教学和科学研究，促进国家通用语言文字的规范、丰富和发展”；第十九条规定“以普通话作为工作语言的播音员、节目主持人和影视话剧演员、教师、国家机关工作人员的普通话水平，应当分别达到国家规定的等级标准”；第二十四条规定“国务院语言文字工作部门颁布普通话水平测试等级标准”。根据《中华人民共和国国家通用语言文字法》，2003年，教育部、国家语委颁布了修订后的《普通话水平测试大纲》，进一步确认了这个标准。《普通话水平测试等级标准(试行)》成为普通话水平测试三级六等的法定标准。这些重大举措，对于贯彻“大力推行，积极普及，逐步提高”的新时期推广普通话的方针，对于推广普通话工作的科学化、规范化、制度化，对于各级各类学校的普通话教学改革，有着极为重要的意义，已经并将继续对我国社会信息化、现代化的发展产生深远的影响。

随着21世纪的到来，人类已经进入信息化的时代。迅猛发展的互联网成为人们获取信息的主要方式。面对信息网络时代和信息技术的不断进步，语言文字应用研究要与时俱进，要借助信息技术手段实现语言资源的数字化和资源的整合与共享，从而提升语言研究的科学性和研究效率。要从深度和广度两个方面充分挖掘语言资源。对于推广普通话、推行规范汉字、推行《汉语拼音方案》、实施普通话水平测试这些语言文字工作，应提升到语言信息处理的高度来认识和实行，以求能够适应不断发展的时代要求。

语言文字的规范化、标准化和信息化，是实现社会信息化的基础。一方面要把促进语言文字的规范化、标准化和信息化作为应用语言学研究的重要任务，作为国家语言文字工作的重要任务。另一方面也要将现代信息技术成果引进到语言研究中，引进到社会语言文字的应用中，引进到语言文字的工作中，实现语言研究手段的现代化和语言文字工作手段的现代化。

目前，国家通用语言文字的普及程度和规范化、信息化水平，还不能完全适应经济社会发展的需要。加强国家信息化建设，需要加强语言文字基础研究和基础工程建设，提高语言文字的规范化、标准化水平，注重国家语言战略研究，利

用语言维护国家的文化安全。普通话水平测试是国家语言文字工作信息化的重要体现,要不断提高这项测试的现代化水平。语言信息处理是提高普通话水平测试标准化、现代化水平的核心内容。本书根据国家有关的语言文字政策,根据《普通话水平测试大纲》精神,以云南省改革开放三十多年来的普通话水平测试实践为基础,从语言信息处理的角度,分析探讨普通话水平测试的理论和具体方法,以求针对普通话水平测试的特点,在广泛吸取现代语言学各有关学科、教育测量学、计算机科学的理论和研究方法的基础上,建构普通话水平测试的理论框架,探寻测试的优化方法,提高普通话水平测试信息化、科学化的水平,完善普通话测试体系,满足新时期社会对语言文字发展的需求,构建和谐的语言文字环境。

云南是全国最早开展普通话水平测试的省份,早在20世纪80年代中期,吴积才教授等就编写了《普通话训练和标准化测试手册》,20世纪90年代初期又编写了《标准化水平考试·普通话教材》(语文出版社,1992年)。王渝光作为这两本教材的主要设计者之一,在其中许多地方已经体现出了语言信息处理的思想方法。从20世纪90年代初起,被云南省普通话培训测试中心长期聘为普通话水平测试科研负责人,完成多项国家语委课题。时任国家语委主任的许嘉璐教授、时任国家语委副主任的王均教授、仲哲明教授、傅永和教授和陈章太教授、刘照雄教授等都对云南的普通话水平测试研究给予了大力的支持和肯定,在此谨向这些前辈表示崇高的敬意。现任国家语委领导、国家普通话水平测试中心领导一如既往地云南的普通话水平测试研究给予了大力的支持和肯定,谨向他们表示崇高的敬意。历任云南省普通话培训测试中心领导的戴梅芳主任、傅牧主任、袁耘毅主任给予了连续不断的支持。与先为同事,后为领导的姚一斌先生、李竹屏先生、陈典红女士、金程先生、李宏坤女士一直合作愉快,谨向他们表示衷心的感谢。云南全省的各民族普通话测试员、普通话教师和普通话应试者为本书数据的取得做了基础性的工作,谨向他们表示诚挚的谢意。

云南师范大学历届语言学及应用语言学硕士研究生,很多直接参与了普通话水平测试的研究,做了大量的基础性工作,其中有杨万兵、蒋利平、李国强、熊子瑜、刘华、吴昌银、江友辉、张鹏、杨铮琳、杨端如等。杨瑞鲲、李洪平同志成绩突出,做了大量的工作。本书的具体分工为:王渝光撰写第一章、第六章,杨瑞鲲撰写第三章、第五章,李洪平撰写第二章、第四章,普通话水平测试题库系统自动出题程序由王渝光编写,计算机辅助普通话水平测试人机对话评分系统主程序由杨瑞鲲编写,最后由王渝光修改定稿。本书是从语言信息处理的角度对普通话水平测试进行深度研究的专著,可供广大语言文字工作者、语文教师和普通话测试员做深入研究时参考,也可供广大应试人员了解普通话水平测试参考。本书可做语言学及应用语言学研究生、全日制本科生语言信息处理课程的教材。本书在撰写时,参考了前辈、时贤及网上的一些资料,限于篇幅,未能一一列

出，谨向他们表示崇高的敬意。

本书的出版要感谢云南大学出版社的领导和邓立木女士，正是在他们的大力支持下，本书才能有这样短的时间内以这样精美的形式出版。本书得到云南师范大学语言学及应用语言学重点学科建设经费资助，谨向有关领导和管理人员表示衷心的感谢。本书既是作者多年来从信息处理的角度对普通话教学和普通话测试的总结，又是对新时期推普工作如何与时俱进的思考，谨以本书的出版献给中华人民共和国六十周年华诞！

王渝光 杨瑞鲲 李洪平
2009年9月6日于银河星辰

目 录

第一章 语言信息处理与普通话水平测试概述	(1)
第一节 语言信息处理	(1)
第二节 普通话水平测试	(5)
第二章 普通话水平测试的语言统计研究	(10)
第一节 统计语言学的基本原理和方法	(10)
第二节 汉语普通话的语言概率	(13)
第三节 普通话测试内容的语言统计研究	(26)
第三章 普通话水平测试的语音实验研究	(31)
第一节 实验语音学的基本原理和方法	(32)
第二节 普通话元音的声学特征与正误辨析	(37)
第三节 普通话辅音的声学特征与正误辨析	(48)
第四节 普通话韵律的声学特征与正误辨析	(56)
第四章 普通话水平测试的教育统计与教育测量研究	(64)
第一节 教育测量学的基本原理和方法	(64)
第二节 普通话水平测试的教育测量理论	(71)
第三节 普通话水平测试的教育测量标准	(76)
第五章 普通话水平测试数据库建设	(80)
第一节 数据库与数据库设计	(80)
第二节 普通话水平测试各等级标准语言特征数据库系统	(92)
第三节 普通话水平测试题库系统	(113)
第四节 计算机辅助普通话水平测试评分系统研究	(130)
第六章 普通话水平测试的质量评析与改进实验研究	(137)
第一节 普通话水平测试的质量评析与改进	(137)
第二节 普通话水平测试的改进实验研究	(153)
参考书目	(173)
附录	(175)
一 普通话水平测试等级标准 (试行)	(175)
二 普通话水平测试大纲	(176)

三	普通话水平测试样卷及答案数据·····	(183)
四	普通话水平测试题库系统自动出题程序·····	(198)
五	计算机辅助普通话水平测试人机对话评分系统主程序·····	(235)

第一章 语言信息处理与普通话水平测试概述

第一节 语言信息处理

一 信息与信息技术

信息是事物运动的状态与方式，是物质的一种属性，是人们关于事物运动状态和方式的知识。观念性的信息要通过具体的物质形态表现出来，因而信息可以有多种物质载体。这些物质载体必须能作用于人的感官，因而信息的具体表现形式可以是听觉形式的声音，也可以是视觉形式的图像，还可以是触觉、嗅觉、味觉等可感知的形式。同样的信息可以用不同的物质形态表现出来。

信息的产生、传递、存储及接收等过程必须有物质能量的支持、投入和转换。信息只有通过一定的物质介质为中介，才能成为人们可以感知的对象。

信息技术是研究信息的获取、传输、处理和存储的技术，是在信息科学的基本原理和方法的指导下扩展人类信息功能的技术。

信息系统是对信息源输入的信息进行收集、存储、传输、分析、变换的处理，并输出处理后信息的系统。信息系统由“信源—编码器—信道—解码器—信宿”构成。信息系统要解决的问题是在一定的传输距离和传输媒体之内，把影响信息传递的各种噪声降到最低限度。信源只有在利用最少的物质能量的支持、转换和投入的情况下，传递出有用的、关键的、必需的信息，并有效地克服内部噪声和外部噪声的干扰，信宿的信息量才会达到最高值。

人的信息功能有信息获取功能、信息传递功能、信息认知功能、信息执行功能。信息技术是扩展人的信息功能的技术。信息具有可度量——信息可以度量、可识别、可转换、可存储、可处理、可传递、可复制、可压缩等特征。

媒体，是信息的载体，根据具体的物质形态和功用可以分为感觉媒体、显示媒体、存储媒体和传输媒体等。多媒体技术是一种把话语、文本、图形、图像、动画和声音等形态的信息结合在一起，并通过计算机进行处理的综合性信息技术。计算机只有达到能够实时处理视频数据流和声音数据流的水平，才可能产生多媒体技术。多媒体技术能够按人的要求以多种媒体的形式表现出来，同时作用于人的多种感官，具有集成性、控制性、交互性、实时性、信息使用的方便性和信息结构的动态性等特点。

二 语言是信息的载体

语言是音义结合的符号系统。语言符号的形式是声音，语言符号的内容是意义。信息是人们对事物运动状态和方式的认识，从语言学的角度来看，信息就是语义。语言以声音为具体的表现形式，表达了具体的语义内容，也就是表达了具体的信息内容。语言的本质就是信息。语言是信息的载体，是人类生产信息、承载信息、传播信息、接收信息、储存信息、检索信息、处理信息的最主要的信息工具。

语言信息具有物质性、抽象性、社会性等特点。信息是事物运动的状态与方式，是人们判断事物、控制能源、了解情况、处理事情的根据。交换信息是人类社会存在的必然要求。人类通过进化获得多种信息手段，并在进化过程中对这些信息手段进行不断的选择、发展。语言的形成是人类祖先在漫长的进化过程中对信息手段的自然选择、积累、修正、补充的必然结果。语言能力是人类最重要的信息能力。人类历史上的许多次信息能力的革命，都大大提高了人类的信息能力，同时也大大地提高了人类的社会生产力。随着社会的发展和进步，人类的信息手段也不断发展进步和飞跃，然而每一次信息革命都与语言这个最重要最基本的信息载体密切相关。

随着当今世界信息科学的高速发展，信息载体将会以更多的形式出现并为人所广泛使用。这些载体以及由此产生的信息传输方式，也必将在人类的社会实践活动中发挥重要的作用。语言现已成为人类和计算机共同的交际工具。随着信息时代的到来，以计算机为中介的信息传播正深刻地改变着人类信息传播的方式，并创造了一种新的信息文化形态。语言文字这种抽象的听觉、视觉媒体，如果与其他具体形象的听觉、视觉媒体结合起来就可以迸发出巨大的信息能量。

三 语言信息处理

语言信息处理，是用计算机对语言信息(包括口语和书面语)进行编码、转换、传输、解码、存储、分析等进行加工处理的科学。

语言信息处理是一门与语言学、计算机科学、心理学、数学、控制论、信息论、声学、自动化技术等多学科相联系的交叉性边缘学科。

人类利用语言进行交际的过程，实际就是人自身进行语言信息处理的过程。

在人们利用语言信息进行交际的过程中，从始至终都存在着噪声的干扰。语言信息处理的过程同时也是一个力图把噪声干扰限制在最小的范围，力求用最少的能量达到最大效益的信息输出的过程。

在人们的主客观世界中，存在着大量的人们尚未认识或认识不够的事物。信息量是指人们认识某个事物性质状态的不确定性的消除。不确定性消除的越多，信息量就越大。信源只有在最少的物质能量的支持下，传递出清晰的关键的信息，信宿的信息量才会达到最高值。在语言交际中，信源为了保证关键的信息不

受干扰,使信宿获得最大的信息量,就必须加大关键信息的冗余度,从而避免信宿被错误引导或信息产生歧义。

语言的概率,指的是构成整个语言系统的各种结构要素在使用频率方面所呈现出来的随机性与出现频率相关的可能性。语言符号在使用频率上既有很强的随机性,又有一定的必然性。充分利用语言概率进行语言信息处理,可以使我们在较短的时间内达到尽可能大的语言交际、语言教学和语言测试效果。

就汉语而言,语言信息处理又叫中文信息处理,包括汉字信息处理和汉语信息处理两个部分。中文信息处理,是自然语言信息处理的一个分支。现阶段的中文信息处理技术,虽有了很大的发展,但离现代化的要求还有较大的差距。许多语言信息处理技术,既需要计算机技术的不断发展,更需要从语言信息处理的角度加强对汉语本身的研究。目前,汉语信息处理在机器翻译、汉语的人机对话和中文情报检索等方面,已经取得了一定的成果。汉语信息处理,还包括言语统计、计算机辅助语言教学(编制各种语言教学软件)、汉语语音识别和言语合成、汉语照排系统、方言研究、修辞学研究(用统计的方法研究风格特征,并进行风格比较)等。

四 语文化现代化与语言信息处理

语言文字是一个国家主权的象征,也是民族自信心的体现。语言文字是一个民族团结的象征。语言文字是一个民族文化传统的标志。现在世界上的各个国家、民族都比以往任何时候更加珍视自己的语言文字所具有的标志性意义。

语言文字是交流的工具,也是文化的载体,承载着丰富、厚重的社会文化信息,服务于社会的经济、政治和文化,是能够产生社会效益和政治、经济、文化、科技等效益的特殊社会资源,是国家最宝贵的财富。语言是人类重要的文化资源和经济资源。当今时代,语言已经进入到经济和高科技领域,成为经济发展的重要资源。随着信息时代的发展,语言作为经济资源的性质会体现得越来越明显,其经济意义越来越显著。

语言是人类最重要的交际工具,也是人类最重要的信息载体和信息处理工具。人类历史上的每一次信息革命都与语言息息相关,都使语言的表现产生了新的形式。从语言的本体来看,有声语言的产生是人类第一次信息革命,这是一次人类信息手段的根本性的革命。文字的产生是第二次信息革命,文字的出现使语言除了口语形式外又多了书面语的形式。造纸和印刷术的产生是第三次信息革命,这次革命的成果使得语言能够以纸的形式记录下来并得到传播,从而产生了中华民族光辉灿烂的物质文明和精神文明。通信技术的产生和发展是第四次信息革命,这次革命的成果使得语言能够以光、电、磁的形式记录下来,并得到广泛的传播,有力地推动了现代科学技术的发展。计算机的出现和发展是第五次信息革命,这次革命对语言进行数字化加工,极大地扩充了语言的各种功能。

语文现代化就是语言文字通过自身的规范与发展，与时俱进，适应现代社会的要求。语文现代化以信息处理技术为基础，主要是语言文字的信息化，包括语言文字的标准化，语言文字的数字化，语言文字的存储、加工、保护和使用的信息化。

21 世纪的到来，人类已经进入信息化的时代。互联网的迅猛发展成为人们获取信息的主要方式。面对信息网络时代和信息技术的不断进步，语言文字应用研究要与时俱进。要借助信息技术手段实现语言资源的数字化和资源的整合与共享，从而提升语言研究的科学性、效率。要从深度和广度两个方面充分挖掘语言资源。对于推广普通话、推行规范汉字、推行《汉语拼音方案》、实施普通话水平测试这些语言文字工作，应提升到语言信息处理的高度来认识和实行，以求能够适应不断发展的时代要求。

我国的语言文字应用研究取得了很多成果。规范汉字表的研制，普通话水平测试的实施，汉字应用水平等级及其评测系统的研制，智能语音技术在普通话测试中的应用等有较强的科技含量，有力促进了我国语言文字的规范化、信息化建设，对现实语言生活具有重大影响。

语料库是存储于计算机中并可利用计算机进行检索、查询、分析的语言素材的总体。基于语料库的分析方法是对传统的基于规则的分析语言的方法的一个重要补充。语料库是最理想的语言知识资源，是直接服务于语言文字信息处理等领域的基础工程。我国从 1990 年开始建立大型的国家级语料库——现代汉语语料库。现代汉语语料库作为国家级语料库，在语料可靠、标注准确等方面应具有权威性，在汉语语料库系统开发技术上应具有国际领先水平，主要服务于语言文字的信息处理、语言文字规范标准的制订、语言文字的学术研究、语文教育和语言文字的社会应用等方面。

语言文字的规范化、标准化和信息化，是实现社会信息化的基础。一方面要把促进语言文字的规范化、标准化和信息化作为应用语言学研究的重要任务，作为国家语言文字工作的重要任务。另一方面也要将现代信息技术成果引进到语言研究中，引进到社会语言文字的应用中，引进到语言文字的工作中，实现语言研究手段的现代化和语言文字工作手段的现代化。

目前，国家通用语言文字的普及程度和规范化、信息化水平，还不能完全适应经济社会发展的需要。加强国家信息化建设，需要加强语言文字基础研究和基础工程建设，提高语言文字的规范化、标准化水平，注重国家语言战略研究，利用语言维护国家的文化安全。当前，《通用规范汉字表》征求意见，引起了社会热烈反响。信息化时代赋予的任务包括：树立科学的语言发展观和规范观，积极普及国家通用语言文字，完善现行汉字和普通话语音的基本标准体系，建成语言文字规范标准的评测认证系统，建成汉语应用能力的立体测试系统。加快规范标

准建设,审定汉字属性、汉语语音、语汇以及外语应用等系列规范,探索语言文字规范标准新的审定和发布模式,及时整合、修订已发布的规范标准;继续开展语言文字基础标准、信息处理和文化教育等领域语言文字规范标准的研制发布工作,加速语料库、知识库等语言工程建设,建立“普通话口语标本库”和“普通话口语语料库”,完善普通话语音规范,制定普通话声学参数标准,完善语言文字标准体系,采用智能语音技术在辅助普通话学习与测试,对国家语委现代汉语语料库进行深度开发利用并提供给社会使用,加快中文信息处理规范标准建设步伐,为国家标准化发展战略服务。

第二节 普通话水平测试

一 实施普通话水平测试是信息化、现代化社会发展的必然要求

共同语是一个言语社团内用作共同交际工具的语言。民族共同语是指一个民族内用作共同交际工具的语言。在我国古代,人们把这种共同语称作“雅言”、“通语”、“官话”,辛亥革命以后又称作“国语”。标准语是经过加工规范后的民族共同语,是民族共同语的规范化形式。确立民族共同语的标准,推广民族共同语是政治经济文化发展的需要,是一个国家实现工业化、信息化、现代化的需要。

新中国成立后,中国共产党、中央人民政府对汉民族共同语的确立和推广十分重视。1955年进行了现代汉语规范问题的讨论。1956年2月6日国务院发布了《关于推广普通话的指示》,明确提出并规定“汉语统一的基础已经存在了,这就是以北京语音为标准音,以北方话为基础方言,以典范的现代白话文著作为语法规范的普通话。”1958年周恩来总理在《当前文字改革的任务》中又把推行汉语拼音方案、推广普通话作为两项重要的任务。国家根据当时的情况,制定了“大力提倡,重点推行,逐步普及”的推广普通话的方针。1982年《中华人民共和国宪法》明确规定“国家推广全国通用的普通话”。1986年国家语委《新时期的语言文字工作》报告中提出:“重点应当放在推行和普及方面,在普及方面应当更加积极一些。”报告提出“普通话的标准只有一个”,“从实际出发具体要求可以不同”,“可以分为以下三级:第一级是会说相当标准的普通话,语音、词汇、语法很少差错。第二级是会说比较标准的普通话,方音不太重,词汇、语法较少差错。第三级是会说一般的普通话,不同方言区的人能够听懂”。

1994年10月30日,国家语言文字工作委员会、国家教育委员会、国家广播电影电视部做出了《关于开展普通话水平测试工作的决定》,发布了《普通话水平测试大纲》和《普通话水平测试等级标准(试行)》,对普通话水平测试提出了具体的实施办法,对汉语普通话水平的等级体系进行了细化,明确提出了三级六

等的客观的评判标准。通过几年的实践,《普通话水平测试等级标准(试行)》经国家语言文字工作委员会再次审定作为部级标准,于1997年12月5日正式颁布。

2000年10月31日九届人大常委会通过了《中华人民共和国国家通用语言文字法》,第六条规定“国家颁布国家通用语言文字的规范和标准,管理国家通用语言文字的社会应用,支持国家通用语言文字的教学和科学研究,促进国家通用语言文字的规范、丰富和发展”;第十九条规定“以普通话作为工作语言的播音员、节目主持人和影视话剧演员、教师、国家机关工作人员的普通话水平,应当分别达到国家规定的等级标准”;第二十四条规定“国务院语言文字工作部门颁布普通话水平测试等级标准”。根据《中华人民共和国国家通用语言文字法》,2003年,教育部、国家语委颁布了修订后的《普通话水平测试大纲》,进一步确认了这个标准。《普通话水平测试等级标准(试行)》成为普通话水平测试三级六等的法定标准。

这些重大举措,对于贯彻“大力推行,积极普及,逐步提高”的新时期推广普通话的方针,对于推广普通话工作的科学化、规范化、制度化,对于各级各类学校的普通话教学改革,有着极为重要的意义,已经并将继续对我国社会信息化、现代化的发展产生深远的影响。

普通话水平测试,作为一个国家的测试标准,作为国家职业的标准,用法律形式来规定,许多部门已经把普通话测试作为从业的必要条件。普通话水平测试是国家语言文字工作信息化的重要体现,要不断提高这项测试的现代化水平。普通话测试和汉字测试的健全与完善,对建立一个立体的语言文字测评体系有着重要意义。

二 普通话水平测试的性质

我国的普通话水平测试(缩写为PSC)是一项较新的语言测试,它的测试理论和具体方法都还很不完善。本书根据国家有关的语言文字政策,根据《普通话水平测试大纲》精神,以云南省的普通话水平测试实践为基础,分析探讨普通话水平测试的理论和具体方法,以求针对普通话水平测试的特点,在广泛吸取现代语言学各有关学科、教育测量学、计算机科学的理论和研究方法的基础上,探寻测试的具体方法,构拟普通话水平测试的理论框架,完善普通话水平测试。

作为语言测试的普通话水平测试,应该遵循有关的语言测试的理论和方法,但它与国内外的第二语言的测试又不大一样。普通话水平测试在测试的性质、内容、重点、难点、对象、方法等方面都有自己的特点。

《普通话水平测试大纲》(修订)明确了普通话水平测试的性质,指出“普通话水平测试是一种目标参照性考试,是对应试人员掌握普通话规范程度的检验,是对应试人员运用普通话进行交际的能力水平的测试和评价”。普通话水平测试

是一种目标参照性测试。目标参照性测试也称“标准参照性测试”、“水平测试”，它是在一个比较明确的相对限定的范围内对应试人员进行的普通话所达到的等级标准的考试。这个标准就是国家《普通话水平测试等级标准》，该标准对普通话三级六等的标准作了具体的描述。

目前国内通行的语言水平测试大多是外语水平测试，而普通话水平测试对大多数中国人来说，是第一语言的测试，是对以汉语为第一语言的使用汉语方言的人们进行的一种对汉语标准语——普通话掌握程度的语言测试。这一特性使普通话水平测试区别于把汉语作为第二语言的 HSK 考试。

普通话水平测试是一种测试对象相当广泛的测试，它适用于不同的性别、年龄、民族、文化水平、职业、阶层、社会地位的有关人员，测试对象的语言文化背景可以有很大的差别。这一特性使得普通话水平测试的测试对象的数量远远超过其他考试。

从语言信息处理的角度看，普通话水平测试与其他语言测试的主要区别在于它是完全的口语测试，对测试成绩的评判主要取决于应试者发出的语音信号及其表达的内容信息，因而对语音信号的质量及其所载信息的评价极为重要。只有认清普通话水平测试的特点，把可能产生的各种误差减小到最低限度，普通话水平测试才有科学的前提，测试的严肃性、公平性、权威性才有保证。

三 普通话水平测试是系统工程

普通话水平测试是一种综合教育测量学与现代语言学研究成果的多学科综合应用技术。它涉及语言学的各有关领域和新兴学科，涉及教育测量学、计算机科学、实验语音学、言语声学、统计语言学、语言教学与测试等许多新兴学科，是一个开放性的系统工程。

普通话水平测试研究总的指导思想和研究试验方法为：以现代语言学和现代教育测量学为测试的理论基础，以实验语音学、音位学的理论和研究成果为评判正误的尺度，研究制定普通话水平测试的教育测量学标准，建立并完善普通话水平测试题库，实施规模性的普通话水平考试。

普通话水平测试的基本实施步骤包括：确定普通话水平测试的理论基础；制订普通话水平测试大纲，根据普通话水平测试大纲确定测试的范围、题型、题量及测试方法；拟定编题计划；命制题目；审定题目；拼组试卷；实施测试；评定并复查成绩；分析试题、试卷质量；搜集、整理、分析有关数据；确立规范卷标准；根据需要，组配命制若干套等值的规范试卷；检测教学效果，进行信息反馈，指导教学等。在测试的内容、题型、题量、测试方法和规范卷标准都确定以后，普通话水平测试的关键环节，就是克服成绩评判的主观性问题，明确并实施测试评判的标准，并在理论与实践的结合中不断修正、补充和完善这个标准。

四 语言信息处理是提高普通话水平测试标准化、现代化水平的核心内容

普通话水平测试是一个新兴的较为复杂的语言测试系统工程，既需要宏观的政策导向和理论指导，也离不开微观的调查研究和统计分析，随着普通话测试与教学的不断探索与实践，以及各种测试理论的引进与介入的不断加深，普通话水平测试在理论和实践上都需要进一步深入、完善与提高。

《普通话水平测试大纲》作为指导我国普通话水平测试的总纲，在测试理论、测试覆盖面等方面需要统计语言学相关数据的有力支持。统计语言学的理论和研究成果对普通话水平测试的内容和标准的确定有着重要意义。只有应用统计语言学的理论和方法，通过对统计数据的研究分析，推测普通话水平测试的代表性，才能提高普通话水平测试的客观性、准确性，维护国家普通话水平测试的权威性，更好地推动汉语普通话的推广，促进汉语走向世界。只有通过统计语言学的理论和方法，以量化的手段，通过实验分析，对普通话水平测试进行研究，才能得出确切而非经验的各项量化指标，才能进行精确的测试，才能进行科学的、更大领域的判断和预测，提高普通话水平测试的科学性、准确性。对普通话水平测试进行微观的统计语言学研究，是普通话水平测试发展的必然趋势和要求。

《普通话水平测试等级标准(试行)》是普通话水平测试三级六等的法定标准。这是一个表述性的标准。现代汉语语音的声学特征在该标准中占有极为重要的地位。作为国家的一个法定的标准，只有表述性的阐释是不够的，需要有科学、准确、客观的声学数据作为确定各等级标准声学特征的基础。各个等级标准需要有客观、具体的语言样本和科学、客观的语言学、声学语音学的具体实验数据和有关的理论支撑。从信息处理的角度来看，普通话水平测试的语音实验研究势在必行。普通话水平测试的语音实验以已有的言语声学、实验语音学理论和研究成果为依据，通过语音取样，绘制、测量、分析符合普通话口语实际的波形图、频谱图和语图，考察不同普通话等级语音的声学特征，得出相关的声学数据，通过语音声学的分析研究，确定区分普通话水平测试不同等级的客观数据，为国家制定的《普通话水平测试等级标准》的表述性阐释提供有力的客观的声学数据。做到普通话水平测试的客观公正，最大限度地避免测试的主观性，加强和维护普通话水平测试国家法定标准《普通话水平测试等级标准》的客观性、准确性和权威性，有效提高普通话水平测试质量。语音实验的研究成果还可以用在普通话的教学中。对不同地区不同测试对象的语音问题进行科学的分析，提出科学的纠正方法，以求达到较好的教学效果。

从语言信息处理的角度看，普通话水平测试同样需要教育统计和教育测量的理论和数据的支持。教育统计与教育测量是采用定量分析的理论和方法，对各种教育现象和教育问题进行客观、科学的研究的一门应用科学。《普通话水平测试大纲》作为指导我国普通话水平测试的总纲，它在测试理论、测试内容、测试数