

科学决策 辉煌成就*

——纪念《汉语拼音方案》颁布50周年

2008年4月21日，伴着春步，沐着春雨，纪念《汉语拼音方案》颁布50周年学术研讨会在京举行。会议由教育部语言文字信息管理司主办，中国应用语言学会（筹）承办，中国语言学会、中国语文现代化学会、世界汉语教学学会、中国文字学会、全国汉语方言学会、中国辞书学会、中国语文报刊协会、中国中文信息学会、中国民族语言学会、中国音韵学研究会协办。会议主题是总结《汉语拼音方案》发布实施之经验，探讨《汉语拼音方案》及其实施中需要完善之方面，以使其在新世纪语言生活中充分发挥作用。会议气氛热烈，畅所欲言，高见迭出，富于建设性。一些论文在本期结集刊出，论文精神概言如下：

一 历史集成 千案聚粹

古代汉字注音，或用“直音法”以甲字注乙字，或用“反切法”以两字拼一字。明朝末年（1605年）西来传教士利玛窦

* 本文为纪念《汉语拼音方案》颁布50周年而作，发表时使用的笔名是余信斯。

(Matteo Ricci)，为汉字设计了拉丁式拼音方案，成为汉语拼音上溯之始点。1867年，英国驻华公使威妥玛(Thomas Francis Wade)创造的拼音方案影响深远，其身影至今犹在，如Peking(北京)、Tsinghua(清华)。1892年，卢戢章发表《一目了然初阶》，拉开了切音字运动大幕，开启了中国语文现代化征程。至1912年“读音统一会”召开，拼音方案计有数十种之多，探讨了方案设计的多种可能性。“读音统一会”总结清末20年切音字运动之精要，终在1913年制定了首个官方主持的拼音方案“注音字母”，但拖至1918年才正式公布实施。1928年教育当局公布“国语罗马字”。1921年，瞿秋白在苏联远东地区潜心研究中国各种文字改革方案，拟就《拉丁化中国字》。1929年在莫斯科出版的《中国拉丁化字母方案》，逐渐在国内传播，影响广泛，1940年陕甘宁边区政府给予拉丁化新文字以合法地位。

斗转星移，汉语拼音伴着时代的脚步不断进步。1949年10月中国文字改革协会在北京成立，在百废待兴、百事待举之时已着手汉语拼音方案的研制。1952年中国文字改革研究委员会成立，内设拼音方案组专司其事。1955年成立拼音方案委员会。至此，除专家设计的拼音方案近10种外，各界还提交方案655种，社会智慧高度凝聚。1956年1月，在毛泽东的提议下，汉语拼音方案决定采用拉丁字母。又经广泛讨论，数次修改，《汉语拼音方案》终在1958年获得第一届全国人民代表大会第五次会议批准。

《汉语拼音方案》采用拉丁字母，实行音素制，符合世界文字拉丁化的发展方向。字母名称音主要参考法文、德文等，字母音值尽量照顾国际惯例，更增强了方案的国际性。巧妙利用汉语声韵的互补特点，并在方案设计上考虑到隔音等问题，为拼写汉语留下了充足空间。《汉语拼音方案》科学而实用，它充分借鉴了350年的历史经验，充分吸收了切音字运动、注音字母运动、国语罗马字运

动、拉丁化新文字运动的成果，汲取了近千拼音方案之精华，集大成者必优异。

二 信息津渡 文化桥梁

《汉语拼音方案》公布以后，当时就在许多领域发挥了巨大作用。譬如：识字、扫除文盲、学习普通话；创制与改进少数民族文字；手语、盲文、电报、旗语、灯号等设计；产品型号等代码编制；排序和信息文献检索；外国人学习汉语等。

1977年9月，联合国第三届地名标准化会议在雅典举行，《汉语拼音方案》被认定为中国地名罗马字母拼写法的标准。1982年8月，国际标准化组织正式发出ISO 7098—1982《文献工作——中文罗马字母拼写法》，《汉语拼音方案》成为拼写汉语的国际标准。1988年国家教委和国家语委联合发布《汉语拼音正词法基本规则》，使汉语拼音在拼写汉语方面有明确规范，并于1996年以国家标准名义重新颁布。

随着信息时代的到来，汉语拼音在计算机和网络世界里更显神通。计算机输入法有百千种，但用汉语拼音编制的音码使用最为普遍。音码在通用键盘上几乎不用编码、不需花费多少学习时间、不易遗忘，对一般输录者优势显而易见。2001年国家语委发布《汉语拼音方案的通用键盘表示规范》，统一了音码的键位布局。此外，在数据库标注、计算机网络的信息交换、机器自动翻译等方面，也常用到汉语拼音。在手机的数字键盘上，拼音输入法更占优势。

历史不能改变但可假设。试想若无汉语拼音，若无拉丁化的汉语拼音方案，中文信息化真不知会遇多大困难，会处何样境地，中国或难有现在的语文生活。

三 与时俱进 日臻完善

《国家通用语言文字法》确立了《汉语拼音方案》的法律地位：“国家通用语言文字以《汉语拼音方案》作为拼写和注音工具。”“《汉语拼音方案》是中国人名、地名和中文文献罗马字母拼写法的统一规范，并用于汉字不便或不能使用的领域。”这一定位鉴史据实，科学可行，并为汉语拼音的应用提供了广阔空间。

《汉语拼音方案》发布实施历50载，语言生活发生了巨变，语言理念大幅嬗变，50年实践也积累了经验、发现了缺失。新形势下进一步完善汉语拼音方案及其应用规范，既有条件又意义重大。

总体而言，汉语拼音方案本身并无大憾，只是字母表中字母读音不宜推广，zh、ch、sh、ng的省写方式 \hat{z} 、 \hat{c} 、 \hat{s} 、 η 很少应用，或为庸余。为汉字注音，汉语拼音方案已相当称职，但要拼写汉语还有完善空间。汉语拼音正词法许多细节需深入研究，如用字母表示声调的方法，同音词在拼式上的分化，文雅语句的拼写，成语、熟语的拼写，“抬起头来”“东张张、西望望”等特殊语法结构的拼写，“被称为”“满足于”等前附成分、后附成分的拼写，较为复杂的地名拼写，少数民族人名的拼写等等。

应用领域也有不少课题。例如：汉语拼音缩写的方式与读音，汉语拼音作为代号时的读音，个别字母（如ü）的大写问题，汉语拼音教学法的进一步改进，汉语拼音在通用键盘、手机键盘上的技术编码，相关领域贯彻《汉语拼音方案》作为中国人名、地名和中文文献罗马字母拼写法的统一规范问题等。

拼音方案的完善与科学应用，关键在实践。当前，注音实践不少但缺乏科学总结，且一些益民惠学的优秀传统丢掉了，如出版物上难字、易错读字、多音字跟注汉语拼音的传统就没有了。汉语拼音的电脑输入编码，“万码奔腾”，但缺乏合理高效的选优提升机

制。汉语拼音在拼写汉语方面的实践园地渐趋荒芜，拼音报刊已是当下罕物，“注音识字、提前读写”的教学试验田已耕耘乏人。在城市建设步伐加快、中外交流日益频繁的新时代，中国人名、地名拼写越来越重要，但其间遇到的问题却难以及时而科学地解决。要汉语拼音在“汉字不便或不能使用的领域”很好发挥作用，看来尚有一段崎岖长路。

汉语拼音方案是400年、特别是近百余年、尤其是近60年来研究与实践的结晶，是中国人过好现代语言生活必需的文化利器，是虚拟世界中的高速公路，是具有世界声望、国际公用的拼写汉语的方案。纪念《汉语拼音方案》发布实施50周年之际，更感到方案制定者的远见卓识，更明白传播它、完善它的重大意义与历史使命，更清醒认识到千年汉字和汉语拼音双翼高翥，才能带动中国现代文明寰宇腾飞。“路漫漫其修远兮，吾将上下而求索。”

[原载《语言文字应用》2008年第3期(专刊)]

关于辞书现代化的思考

中国是最早编纂辞书的国家。曹先擢先生（1986）曾经指出，汉代之前虽然还未有辞书出现，但是辞书编纂的条件已在酝酿。春秋战国时期古籍中已有不少训诂资料，周秦时的《史籀篇》及秦代《仓颉篇》《爰历篇》《博学篇》等识字课本，已将汉字作了初步编排，实乃字典之滥觞。两汉之《尔雅》^{【1】}《方言》《说文解字》，是中国辞书的奠基之作。由汉至清，古代辞书可谓成就辉煌，出现了《永乐大典》《古今图书集成》《康熙字典》这样的煌煌巨著（参见林玉山，1996）。

清末到民国，我国辞书进入现代阶段。1911年上海国学扶轮社出版的《文科大辞典》、作新社出版的《普通百科新大词典》（参见李迪、郭世荣，2002），是我国现代辞书的萌芽。1915年出版的《中华大字典》《辞源》，以及此后的《辞海》《国语辞典》，是早年现代辞书的重要著作。然而同西方相比，就现代辞书领域而言，20世纪上半叶我国实已沦落为“辞书小国”。

1949年至今50余年，我国的辞书事业获得了巨大发展。编纂了

【1】曹先擢（1986）：“《尔雅》相传为周公所作（还有别的说法），实为西汉初年学者缀集旧文而成，是我国第一部词典。”

《新华字典》《现代汉语词典》《汉语大字典》《汉语大词典》《中国大百科全书》《现代汉语方言大词典》《普通话基础方言基本词汇集》《故训汇纂》等精品辞书，香港、台湾等地也出版了不少很有价值、很有分量的工具书。经半个多世纪努力，无论从辞书出版的种类、发行数量，还是从家庭辞书拥有量^[1]等方面看，我国确实可称得上“辞书大国”了。

但是，中国显然还不是“辞书强国”，精品辞书的比例还不高，在国际上有影响的辞书还不多，特别是辞书编纂手段、观念等方面，与先进国家还有距离。促进中国辞书现代化，是缩短、弥合同先进国家的辞书差距，使我国由辞书大国成长为辞书强国的重要战略举措。

一 努力促进辞书编纂手段现代化

信息时代的重要特征是：数字技术成为处理信息的主要手段，数据库成为信息的重要载体。辞书编纂手段现代化，其现实含义就是在辞书整个编纂过程中充分使用数字技术和各种数据库，具体说，就是开发功能强大的语料库、知识库和适用于辞书编纂的软件系统。

（一）语料库

李宇明（2004）曾将现代意义上的语料库定义为：“机器可以处理的有一定规模、结构和知识标记的自然话语材料的集合。”世界第一个现代意义上的语料库，是1964年布朗大学建成的布朗语料

【1】李志江、吕京、王伟等1987年8月至1998年6月，通过问卷调查发现，没有字典、词典的家庭很少，“语文辞书几乎是每个具备一定文化水平的读者必备的工具书”。大多数家庭拥有多部字典、词典，其中拥有4部以上字典的家庭为32.80%，拥有4部以上词典的家庭为26.64%。详情见词典调查课题组（1999）。

库。此后接连诞生了一批语料库，例如：

LOB语料库（LOB Corpus），1970年开始建设，1978年完成；

美国传统中级英语语料库（The American Heritage Intermediate Corpus），建于1971年；

伦敦—隆德语料库（London-Lund Corpus），1973年建成；

戈尔哈布印度英语语料库（Kolhapur Corpus of Indian English），建于1978年；

康奈尔语料库（The Cornell Corpus），1980年建成；

加拿大书面英语语料库（Corpus of English-Canadian Writing），1984年建成；

新西兰书面英语威灵顿语料库（Wellington Corpus of Written New Zealand English），1986年建成；

澳大利亚英语语料库（Australian Corpus of English），建于1986年。

这些语料库的规模都是十万级或百万级^[1]，一般称为第一代语料库。随着计算机处理能力的快速发展，20世纪80年代开始建构规模在千万级以上的第二代语料库，进入90年代以后出现了亿级语料库。而今语料库仍处于快速发展期，数量飞速增加，规模急速加大，加工更为精细，用途不断扩展。^[2]

语料库的用途是多方面的，但是辞书编纂是其重要用途之一。借助语料库编纂辞书已有诸多成功范例。1980年，英国柯林斯公司的COBUILD语料库开始运转，以此为基础，1987年出版了《柯林斯COBUILD英语词典》（*Collins COBUILD English*

【1】语料库的规模依照语料库所包含的词量计算，词量在10万以上、100万以下，为“十万级”语料库；词量在100万以上、1 000万以下，为“百万级”语料库，余类推。

【2】关于语料库的发展状况，详见章宜华（2004：61—79）。

Language Dictionary)。1991年该语料库改称“英语库”(Bank of English),并不断扩充,2002年1月已成为具有4.5亿词的巨型语料库。COBUILD语料库在词典编纂方面的革命性成就,引发牛津大学、朗文、剑桥大学、麦克米伦等出版社群起效仿,造成势不可遏的滔滔洪流(曾泰元,2005:83;章宜华,2004:65-66)。

利用语料库编纂辞书,是词典编纂方法的根本性变革。程荣(1997)、解建和(1997)、王小海(2001)、钱厚生(2002)、章宜华(2004)等,都曾讨论过语料库对辞书编纂的作用。梳理这些宏论,结合已有实践,语料库在辞书编纂中的主要作用可总结为四方面:

1. 帮助发现新词语、词语的新意义和新用法。社会阅历的有限性,必然造成个人语言经验的有限性;且语感的悟觉,常需语言材料、语言环境或语言知识的触发。语料库是社会语言应用的缩影,能够在相当程度上延伸个人的语言经验,触发语感的悟觉。通过语料库得到的词语表,有助于发现新词语,可使辞书收词更为完善与及时。通过语料库较易发现词语的新组合,这种新组合往往包含着词语的新意义或新用法,新意义、新用法的发现可以使词语解释更完满。

2. 统计词语、词语的各义项和各用法(语法的和语用的)的使用频度和流通频度。频率统计为语料库所擅长。利用语料库统计词频,可以帮助确定辞书的收词量,如中小型辞书,可以只收词频较高的词语。对词语各义项、各用法的频度统计,可以帮助确定词语的哪些义项需要解释,以及义项的排列顺序,可以帮助确定词语的哪些用法需要说明,以及用法的说明顺序等等。

3. 帮助选择或剪裁例句。选裁例句是辞书编纂的重要工作,也是辞书质量的一种体现。语料库是大量真实的语言用例集合,在语料库中选择合适用例,或是在语料库一定例句的基础上进行剪裁

加工，不仅方便，而且比编造例句更有真实感。

4. 利用语料库进行知识挖掘。语言是人类知识的结晶，同时也是人类知识的载体。因此语言学不仅是在研究语言，而且也是在研究知识的表达方式，故而成为当今信息处理、知识发掘工程中的重要方面军。语言与知识的关系决定了语料库不仅是语言的集合，同时也是知识的集合；语料库不仅可用来研究语言，而且也可用来挖掘知识。知识挖掘的领域非常广泛，辞书编纂者最感兴趣的知识挖掘，也许是发现词语的内涵，发现词语之间的相似与不同。这种知识挖掘，可以帮助定义词语，帮助确定同义词、近义词、反义词、上下位词等词语之间的关联，帮助梳理“词际关系”。

语料库对于词目选定、词语关联、词语释义、例句选裁等都具有重要作用，因此拥有高品质的语料库，就等于拥有了高品质辞书的基础。^[1]曾泰元（2005）在讨论汉英词典语料库时指出，建设语料库是件非做不可的事，“因为语料库的时代早已来临，若不急起直追，恐怕就会被淘汰”。为辞书编纂建设适用的高品质语料库，是辞书界最应关注的问题。

关于语料库建设的论述已经很多，不必赘述，但仍愿强调四点：

1. 科学选取语料入库，并及时更新。网络上有大量语料，可以用来做些检索，也可以很方便地下载（参见张铁文，2004），但是当前网络语言的质量还不尽如人意，特别是精品辞书的编纂，还需要认真选取语料建库，或是购买权威的语料库。

2. 语料库的规模要达到足够的量级，“没有数量就没有

【1】王馥芳、马兰梅（2003）指出，语料库与语言生活中词语使用的实际有差距，语料库的这种“不完全代表性”，会影响到以语料库为基础的词典的科学性。王、马的提醒值得注意，的确应认识到语料库尽管功能强大，但在当前只是辞书编纂的辅助手段。不过另一方面，丝毫不能因语料库尚不完善便低估甚至否定语料库的重要作用，应该做的是加快语料库建设，促其尽快完善，逐渐接近语言生活。