

百种语文

小丛书 曹先擢 主编

# 汉字整理的 昨天和今天



HANZI ZHENGLI DE ZUOTIAN HE JINTIAN

费锦昌 著

语文出版社  
[http // www. ywcbs. com](http://www.ywcbs.com)





# 百种语文小丛书

## 首批20种书目

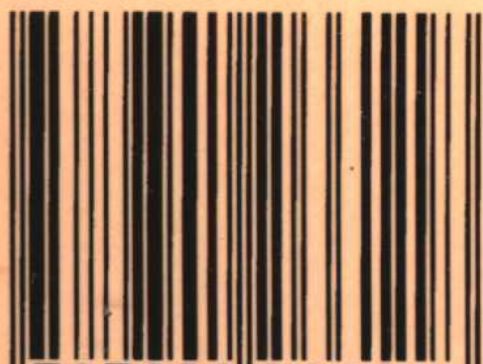


汉语与外语	季羨林	著
人类文字浅说	周有光	著
普通话和北京话	林 焘	著
古代白话说略	江蓝生	著
新语词	陈 原	著
普通话语音史话	唐作藩	著
词汇与写作	许威汉	金 甲 著
古书里的假借字	刘又辛	著
字形的规范	傅永和	著
从甲骨文到简化字	董 琨	著
一门新学科：现代汉字学	苏培成	著
汉字整理的昨天和今天	费锦昌	著
俗字里的学问	张涌泉	著
汉字与计算机	顾小凤	著
汉语新词语与社会生活	陈建民	著
汉语词汇新讲	周 荐	著
术语浅说	冯志伟	著
谚语新谈	鲁启华	著
现代汉语缩略语	凌远征	著
汉语拼音用处多	李乐毅	著

责任编辑／南保顺

封面设计／正 阳

ISBN 7-80126-611-0



9 787801 266118 >

ISBN7-80126-611-0/H

定价：3.00元

FJC

F124.1



百种语文

小丛书 曹先擢 主编

# 汉字整理的 昨天和今天

HANZI ZHENGLI DE ZUOTIAN HE JINTIAN

汉字整理

费锦昌 著

语文出版社

[http // www. ywcbs. com](http://www.ywcbs.com)

~~~~~  
**图书在版编目(CIP)数据**

汉字整理的昨天和今天/费锦昌著. - 北京:  
语文出版社, 2000.7

(百种语文小丛书)

ISBN 7-80126-611-0 / H·172

I. 汉… II. 费… III. 汉字-整理 IV. H124.1

中国版本图书馆 CIP 数据核字(1999)第 50358 号  
~~~~~

百种语文小丛书

HANZI ZHENGLI DE ZUOTIAN HE JINTIAN

**汉字整理的昨天和今天**

\*

**语 文 出 版 社 出 版**

100010 北京朝阳门南小街 51 号

Email: ywp@public2.east.net.cn

新华书店经销 北京印刷二厂印刷

\*

850 毫米×1168 毫米 1/48 1.5 印张 38 千字

2000 年 7 月第 1 版 2002 年 6 月第 2 次印刷

印数: 2,001-5,000 定价: 3.00 元

---

本书如有缺页、倒页、脱页, 请寄本社发行部调换。

# 《百种语文小丛书》编委会

主 编 曹先擢

副主编 苏培成 高文元

编 委 (按音序排列)

曹先擢	董 琨	高文元	顾士熙	侯精一
季恒铨	江蓝生	李建国	李守业	隆 林
吕宏伟	孟吉平	南保顺	宋绍年	苏培成
佟乐泉	王 宁	熊正辉	于根元	赵 曾

# 说 明

在信息时代、知识经济时代,知识的普及工作比以往任何时候都更为重要和迫切。语言文字是信息的载体,是人类文明发展的不可须臾缺少的基本手段,在信息时代其发挥的作用更为巨大。语言文字知识既有传统的,也有新的。语言文字知识的普及工作有待加强。我们编辑这套《百种语文小丛书》,正是想在这方面尽绵薄之力。小丛书的篇幅较小,只有几万字。我们力求少而精,以便于读者购买和阅读。

语言文字知识的内容十分广泛,你如果看一下《中国大百科全书·语言文字卷》的分类目录,便会有一个大概的了解。我们这套小丛书,不是对语言文字知识作全面系统的介绍,而是偏重于人们在学习、生活中会遇到的各种语言文字知识。

书的质量很大程度上取决于作者水平,取决于作者对相关知识了解和掌握的程度。我们希望这套丛书是高质量的。因此,在约稿的时候采取的办法是,谁对某方面问题有研究,就请谁撰写某方面的书



稿。作者素有研究,自然能够驾轻就熟,保证所介绍的知识的科学性,而有的还具有知识的前沿性。把最新的研究成果用普及的形式介绍给读者,也是本丛书编纂目的之一。这只是个总的思路,做起来并不那么简单。作者在写书时,要考虑整个丛书的编写原则,写什么,怎么写,从谋篇到动笔,是一个新的创作过程。本丛书是为普及语言文字知识服务的,然而书中有些内容事属专门;有的见解为作者一家之言,与社会通常的见解有的不完全一致。我们认为这些内容,对读者广见闻是有裨益的。小丛书的作者都是学有专长的语文工作者,其中有的是享有盛誉的著名学者。正是由于得到了这些作者的大力支持,小丛书才能够顺利编辑出版,奉献给社会公众。

本丛书的出版,采取成熟一批出一批的方式,一时难以完全显示学科的系统性。我们打算在出版到一定数量后,再来分类配套,如语音的、语法的、词汇的、文字的,等等。敬希读者谅解。

《百种语文小丛书》编辑委员会

# 目

# 录

一	历史上的汉字整理工作 .....	( 1 )
1.1	先古的汉字整理 .....	( 1 )
1.2	秦代的“同文字” .....	( 2 )
1.3	许慎的《说文解字》 .....	( 3 )
1.4	唐代的“字样”学 .....	( 3 )
1.5	历史上的简化汉字工作 .....	( 5 )
1.6	历史上的审音定音工作 .....	( 7 )
1.7	历史上的汉字定序工作 .....	(11)
1.8	《国音常用字汇》的历史作用 .....	(12)
二	现行汉字“四定”任务的提出 .....	(13)
三	现行汉字的定量 .....	(15)
3.1	汉字的字数 .....	(15)
3.2	常用汉字高频化 .....	(20)
3.3	如何实现现行汉字的定量 .....	(24)
四	现行汉字的定形 .....	(33)
4.1	整理、规范现行汉字的笔画 .....	(34)
4.2	“现行汉字规范字形”的具体涵义 .....	(38)
4.3	整理异体字 .....	(39)
4.4	推行规范的简化字 .....	(41)
4.5	统一计量单位名称用字 .....	(44)



4.6 整理、统一印刷体字形 .....	(45)
五 现行汉字的定音 .....	(47)
5.1 推广“普通话”任务的提出 .....	(47)
5.2《普通话异读词三次审音总表初稿》 的研制 .....	(48)
5.3《普通话异读词审音表》的研制 .....	(49)
六 现行汉字的定序 .....	(51)
6.1 汉字排序的难点 .....	(51)
6.2 音序查字法 .....	(52)
6.3 部首查字法 .....	(54)
6.4 笔画笔形查字法 .....	(56)
6.5 四角号码查字法 .....	(56)
6.6 书刊由直排改横排 .....	(58)
主要参考书目 .....	(60)

汉字是世界上迄今为止历史最悠久而且至今还富有活力的文字符号体系。现在和可以预见的将来它仍然是记录汉语最重要的文字工具。

源远流长的汉字,像横贯中国历史的大江大河一样,千百年来,用它宽广的胸膛承托着、传送着中华民族悠久而灿烂的文化。

这条文字长河之所以能够满载古老而又新鲜的中华文化不知疲倦地破浪前行,这条古老的文字长河之所以能够经历千年百年而不遭淤塞,没有干涸,除了它自身的应变能力以外,一个重要的原因就是千百年来中国人民对它的维护,对它的疏浚,为它清淤,替它开拓。这项“清淤开拓”工程的总名称就是我们常说的“汉字的整理”。

在长期的使用过程中,社会交际会不断地向汉字提出新的要求,社会发展也会不断地给汉字提供新的条件,促使汉字在字量、字形、字音、字义等方面发生或多或少、或大或小的变化。为了使汉字适应社会发展的需要,胜任记录汉语的任务,人们要不定期地整理汉字,使汉字经常保持着“明”和“简”辩证统一的良好状态。

## 一 历史上的汉字整理工作

### 1.1 先古的汉字整理

汉字的整理工作,从它产生之后不久就开始进

行了。原始的汉字有了一定数量的积累以后,在它形成相对成熟的文字符号体系的阶段,一定会由经常跟汉字打交道的人,如传说中的黄帝史官“仓颉”们,凭着成天跟文字接触而获得的经验和知识,对汉字进行整理和修葺的工作。没有“仓颉”们的整理,汉字不可能由原始走向成熟;没有“仓颉”们的修葺,汉字不可能从散落的符号组合成完整的体系。正如章太炎说的:

一二三诸文,横之纵之,本无定也。马牛鱼鸟诸形,势则卧起飞伏,皆可则象也;体则鳞羽毛鬣,皆可增损也。字各异形,则不足以合契。仓颉者,盖始整齐画一,下笔不容增损。由是率尔著形之符号,始为约定俗成之书契。(《检论·造字缘起说》)

根据《周礼·春官·外史》记载,“外史”之官“掌书外令,掌三皇五帝之书,掌达书名于四方”。“书名”指文字,“掌达书名于四方”应该就是统一全国文字的一种措施。有人认为就是由京都向各地推广文字教学,使各地同其音义。文字学界指出,这是继仓颉之后又一次重要的汉字整理工作,时间在春秋之前的周宣王中兴时代,主持人物还是一位史官,他的名字可能就叫史籀,代表作品是《史籀篇》。

## 1.2 秦代的“同文字”

战国时代,各国“文字异形”,给书面交际带来极



大的不便,社会上出现了统一文字的强烈要求。秦始皇统一中国,由李斯等成功地进行了著名的“同文字”的工作,“罢其不与秦文合者”,废除了六国文字中跟秦文不符的异体。同时,李斯、赵高、胡毋敬分别编写了《仓颉篇》《爰历篇》和《博学篇》,作为推行规范文字的范本。政府还用规范字形书写大量诏版,随同权、量等器物颁发各地,在日常生活中扩大规范文字的影响。秦始皇二十八年(公元前 219 年)琅邪刻石上“同书文字”四个大字宣告了这次汉字整理工作的成功。

### 1.3 许慎的《说文解字》

东汉初年,今文经学者根据当时通行的已经不再象形的隶书字形随意演绎汉字字义,曲解经书,什么“马头人为‘长’”“人持十为‘斗’”“‘虫’者屈‘中’也”,不一而足,甚至引用这些错误的说解歪曲法律条文的内容。为了“理群类,解谬误,晓学者,达神旨”,古文经学大家许慎用了 22 年的时间编著了中国文字学的宝典《说文解字》,以规范的小篆为字头,作为解释字义的根据,把籀文、古文等附在正体小篆之后,并“举一形以统众形”,发明用部首统领部属字的编排方法,对汉字的规范、字义的说解、字序的排列和字典的编著都起到了重要的示范作用。

### 1.4 唐代的“字样”学

从东汉末年,经魏晋南北朝,到隋唐统一,近四

百年间,频繁的战乱加剧了文字的混乱,再加上民族杂居、不同语言的频繁接触加速了新字的产生,整理文字的任务愈益繁重。我国第一部以楷书为收字对象的顾野王的《玉篇》,是南北朝负有整理汉字重任的代表性字书。今本《玉篇》卷末还附有“分毫字样”,共收形体近似容易混淆的字 248 个。到了初唐,学界出现了整理异体、辨别俗讹、统一字体、刊正经籍的强烈呼声,产生了一系列类似于今天正字法字典一类的“字样”之书。唐玄宗开元二十三年(公元 735 年)颁布的《开元文字音义》,据说对楷书字形的统一起了不小的作用。颜师古的《字样》则开了唐代正字法的先河。颜的四世重孙颜元孙,又以《字样》为底本,编定《干禄字书》,根据文字应用的不同层次分别规定了俗、通、正三体的应用范围,体现了整理、规范汉字工作中原则性和灵活性的统一:

所谓“俗”者,例皆浅近,唯籍账、文案、券契药方,非涉雅言,用亦无爽,倘能改革,善不可加;所谓“通”者,相承久远,可以施表、奏、笈、启、尺牍、判状,固免诋诃;所谓“正”者,并有凭据,可以施著述文章、对策、碑碣,将为允当。

属于同一类型的整理汉字的字书,还有杜延业的《群书新定字样》、张参的《五经文字》、唐玄度的《新加九经字样》,延及后代,则有宋郭忠恕的《佩觿》、张有的《复古编》,辽时僧人行均的《龙龕手鑑》,元李文仲的

《字鉴》，明焦竑的《俗书刊误》和清龙启瑞的《字学举隅》等。

### 1.5 历史上的简化汉字工作

字形是文字的物质形式。汉字的字形随着书写工具的更换和使用范围的扩大、使用人数的倍增而演变。汉字的基本单位先是“线条化”，后是“笔画化”，汉字的形体由象形意味浓重的金文、甲骨文，通过减弱图画性加强符号性、简省重复部分、简省次要部分和归并义近或形近部件，发展为线条圆转的小篆。进入隶楷阶段以后，带有图绘性质的象形表意字符演变成笔画个性鲜明、结体渐趋方正的符号，大大便利了书写。单个的简体字也出现得更多了，在汉魏六朝的碑刻中，在唐代的写经里，都保存了不少当时民间流行的简体字。宋代以后，简体字由碑刻、手写转入雕版印刷的书籍上，从而扩大了它的流通范围，数量也有所增加。刘复、李家瑞合编的《宋元以来俗字谱》收集了宋元明清 12 种民间刻本中出现的 6240 个俗字，其中绝大多数是简体字。到了近代，太平天国在其文书甚至玉玺上都使用简体字。可见，汉字的整体面貌经历了几次规模较大的“简化”和“规范化”交替滚动前进的历程。宋代以后，由于印刷业的发达和科举考试对文字书写的要求严格化，汉字字形，就总体而言，不断地趋向稳定。到了近代，随着印刷业的进一步发展以及新式学校教育



的兴起和普及,字形稳定的趋势愈益明显。但是,由于使用人数的增长、使用范围的扩大,汉字字形结构繁复和一字多形等状况仍不能适应汉字教学和汉字使用的需要。

1840年鸦片战争以后,开发民智、普及教育的呼声日益高涨。于是,作为政府行为实施具有规模效应的简体字工程被一次又一次地提了出来。把简体字跟普及教育明确联系起来,自觉、积极地加以倡导的,首推陆费逵。1909年(宣统元年),他在《教育杂志》创刊号上发表题为《普及教育应当采用俗体字》的文章,论述教学中采用简体字的好处。把汉字简化运动具体化并推向高潮的则是钱玄同。1922年,他在国语统一筹备会上提出“简省现行汉字的笔画案”:

汉字改用拼音是治本的办法,简省现行汉字笔画是治标的办法。那治本的事业,关系重大,不是一朝一夕就能达到目的的。但现行汉字在学术上、教育上的作梗,已经到了火烧眉毛的地步,不可不亟图补救的办法!我们决不能等拼音的新汉字成功了才来改革,所以治标的办法,实是目前最切要的办法。

钱玄同的提案获得国语统一筹备会第四次大会议决通过。该会正式成立了“汉字省体委员会”。这表明提倡简体字已经由民间人士呼吁阶段进入政府

机构研究阶段。1934年,钱玄同又在国语统一筹备会提出“搜集固有而较适用的简体字案”。1935年,国语统一筹备会的一个工作组在钱玄同的主持下编成《简体字谱》草稿。同年春天,上海陈望道等发起组织“手头字推行会”,选定第一批手头字300个,并用来排印杂志。1935年8月,国民党政府教育部从《简体字谱》草稿中选取324个简体字编成《第一批简体字表》予以公布。1936年2月又被废止。抗日战争期间,简体字在解放区广泛流行,油印书报采用和创造了许多简体字。

### 1.6 历史上的审音定音工作

汉字不是拼音文字,表音功能原本就弱,再加上古音、异读音、读书音、口语音,以及方言读音的存在,加大了正确掌握汉字字音的难度。

从历史记载看,为了沟通各方言区的共同交际,人们往往选择一种使用较广的方言作为共同语。这种共同语往往就是当时政治中心所在地的方言。在西周,因为建都镐京(今陕西西安西南),就以秦音为标准音,这个标准可能到东周也没有改变。

前面曾引到《周礼·春官·外史》载:“掌达书名于四方。”郑玄注:“古曰文,今曰字。使四方知书之文字,得能读之。”有人认为,这是由京都向各地推广文字教学,使各地通其音义。这里就涉及到定音和推广统一读音的问题。

《论语·述而篇》说：“子所雅言，《诗》、《书》执礼皆雅言也。”后人认为“雅言”就是当时在一定区域范围内通行的“共同语”。

从先秦的“雅言”、汉代的“通语”，到元代的“天下通语”、清代的“官话”，都说明我国共同语的基础在北方方言，只是语音标准有所改变，从古代的镐京逐渐东移到近代的北京。

通行于汉代的“读若”“读曰”“直音”也有审定和统一读音的作用。

公元三世纪，中国有了反切注音的办法。反切的发明说明当时人已经能够根据汉语语音特点把一个音节分为声母和韵母两部分。到南北朝，学界又注意到“平、上、去、入”四声的分别，从此奠定了汉语音节分为声韵调三部分的基础。随着反切的发明，从三国开始出现韵书。“盖有反语，则类聚之即成韵书，此自然之势也。”（陈澧《切韵考》）通常认为，最早的韵书是三国魏代李登的《声类》和晋代吕静的《韵集》（均已亡佚）。韵书开始是为了适应诗赋的需要而编写的，后来则在定音和规范读音上也有重要作用，所以，有人认为，“韵书的编制，要兼及审音和检韵这双重的目的”。（胡奇光《中国小学史》，上海人民出版社 1987，第 146 页）如，由于南北朝长期政治分裂，语言交际上也难有一种统一的语音规范。随着隋文帝一统天下大业的开展，定音和规范读音也



提上日程。那时,方音分歧严重,要统一全民的口音,显然不可能,惟一可行的是先在文人的读书音上确定一个大体通行的标准。“榘而量之,独金陵与洛下耳。”在这样一个原则的指导下,编成了《切韵》这本重要的韵书,确定了当时南北读书音的准则。后来在《切韵》的基础上扩编成《广韵》,凝聚了唐宋许多学者的心血。这是我国也是世界上至今仍然完整地保存着的最早的一部韵书。元代周德清的《中原音韵》,在隋唐中古音向元明近代音转变的关键时刻,选中了以大都为主的北方口语音系,则直接关系到以北京音为标准音、以北方话为基础方言的现代汉民族共同语的形成。有人称之为“普通话音系的历史源头”。到明初又有乐韶凤、宋濂等奉诏撰成《洪武正韵》。

总之,从汉末的反切、齐梁的四声,隋代的《切韵》,到唐末宋初的字母和韵图,元代的《中原音韵》、明代的《洪武正韵》,表明我国学者从不同的角度逐一揭示出汉语字音的特征,为历史上的审音定音工作提供了重要的依据。到了清朝末年,以北方官话为基础的汉民族共同语实际上已经形成。

由于受日本明治维新的影响,在清末的后十年就已经明确提出统一国语的问题。1911年,清朝学部召开“中央教育会议”,通过了“统一国语办法案”。这个提案建议在京城成立“国语调查总会”,各省设

立分会,进行语词、语法、音韵的调查;根据调查结果,审定“国语”标准,编辑国语课本、国语辞典和方言对照表等,为统一国语提出了不少具体措施。

民国成立后,1912年召开“中央临时教育会议”,决定先从统一汉字的读音做起。1913年召开“读音统一会”,议定了6500多字的读音,就是俗称的“老国音”,还确定了拼切国音的“注音字母”(又称“国音字母”)。1916年北京教育界人士组织“中华民国国语研究会”,敦促北洋政府公布注音字母和改学校“国文”科为“国语”科。1918年11月北洋政府教育部公布“注音字母”。1919年该部成立政府的推行机构“国语统一筹备会”(国民党政府改称“国语统一筹备委员会”),并于1920年通令全国国民学校改“国文”科为“国语”科。1919年出版《国音字典》,1921年经国语统一筹备会校订后再次出版,定名为《教育部公布校改国音字典》,共收13000多字。我们可以把它看作20世纪中国政府第一次正式公布的“现行汉字表”。1923年国语统一筹备会成立国音字典增修委员会。1924年修改“读音统一会”所定的“老国音”,决定采用北京语音为标准音,称为“新国音”,并在1932年教育部公布发行的《国音常用字汇》中采用、推行。“国语统一筹备会”和“国语统一筹备委员会”这两个机构先后做的跟定音有关的工作主要有:(1)分期举办国语讲习所和国音字母

讲习所,并协助各地举办讲习班;(2)修订注音字母方案;(3)修改国音标准;(4)制订“国语罗马字拼音法式”; (5)规定注音字母单用和“词类连书”办法;(6)促进国民党政府大学院公布“国语罗马字拼音法式”,作为注音字母第二式;(7)编辑出版《国音常用字汇》,并促请国民党政府教育部公布“新国音”标准;(8)设计铸造汉字旁加注注音字母的注音汉字连体字模;(9)调查全国语言分布状况;等等。

### 1.7 历史上的汉字定序工作

古代的字书,包括识字课本,主要是按字义用韵语编排的。以现存的《急就篇》为例,它的开头五句是:

急就奇觚与众异:罗列诸物名姓字,分别部  
居不杂厕,用日约少诚快意,勉力务之必有喜。

全书正文分为三部分:一是“姓氏名字”,如“宋延年,郑子方”之类;二是“服器百物”,如食物类的“稻黍秫稷粟麻稭,饼饵麦饭甘豆羹”之类;三是“文学法理”,如“春秋尚书律令文,治礼掌故砥砺身”之类。

东汉许慎的《说文解字》依据字形用 540 个部首统率 9353 个汉字,把形体相似或意义相近的部首排在一起。

历代的韵书则按字音编排,用声韵为经纬,把收录的汉字排到它所在的位置上。



也有综合运用几种办法的,如宋元间戴侗的《六书故》是按字义结合字形编排的。全书的分部先定为数、天文、地理、人、动物、植物、工事、杂、疑(形体有怀疑的)九种,再立 479 目,每目之下又以“六书”编次词条。

明末清初梅膺祚的《字汇》在排序法上有所革新。一是进一步简化《说文解字》的部首,减为 214 部,纠正了分部过于繁琐的毛病;二是凡部首的排列和各部中汉字的排列都按笔画从少到多为次序;三是首卷后附“检字”,按笔画数排列不易辨明部首的难检字,还在每卷卷首列一个表,载明该卷所有各部首和每部所在的页数。

1918 年公布“注音字母”(1930 年改称“注音符号”)以后,许多工具书就按“注音字母”的顺序编排收录的汉字,如杨树达的《词诠》、中国大辞典编纂处的《国语词典》。

### 1.8 《国音常用字汇》的历史作用

国语统一筹备委员会的编纂处于 1926 年完成《增修国音字典稿》。后又几经修订,于 1932 年 5 月定名为《国音常用字汇》,由当时的教育部正式公布。据黎锦熙《国语运动史纲》评述,《国音常用字汇》收正字 9920 个、“别体重文”(异体字)1179 个、“变音重文”(异读字)1120 个,共计 12219 字;重新确定了以北京语音为标准音的“新国音”;加收了当时通行的

简体字形；按“注音符号”音序排列正文，从而在字量、字形、字音和字序等方面初步建立了现代汉字的规范标准。《国音常用字汇》的出版可以看作中华人民共和国建立以前我国整理现代汉字的一块里程碑。

综观汉字发展的全过程，几乎历代都曾进行过汉字的整理和规范工作，其差别仅在于规模的大小和意识上的自觉自发而已。

## 二 现行汉字“四定”任务的提出

中华人民共和国成立以后，要在旧中国的破烂摊子上建设一个崭新的理想的国家，经济建设和文化建设的任务都异常繁重。年轻的共和国想用最快的速度扫除文盲、普及教育、提高全民的文化素质，于是在语言文字领域提出了文字改革的任务，一方面积极探索用拼音文字代替汉字的可能性和具体途径，另一方面大刀阔斧地整理汉字，“以利目前的应用”。这种整理工作是全方位的，既有字量方面的，又有字形方面的、字音方面的，还有字序方面的。如果说，建国的初期，在汉字整理工作中还有“以利目前应用”这种临时思想的话，那么，随着对汉字改革艰巨性的认识愈益深化，对汉字整理工作长期性的认识也越来越加深，特别是用计算机处理汉语汉字的任务提上日程以后，我国汉字整理工作的规模和

力度都明显地得到加大和强化。1980年5月20日,充实和加强后的中国文字改革委员会召开第一次全体委员会议,会上原则通过了王力、叶籁士、倪海曙、周有光四位委员提出的《关于研究和制订标准现代汉字表的建议》和《制订标准现代汉字表的科研计划(草案)》。这份《建议》和《计划》的目标是,对于“五四”以来的现代汉语用字进行全面的、系统的、科学的整理,针对现行汉字仍然存在字数多、笔画繁、读音乱、检索难的问题,提出要做到“字有定量、字有定形、字有定音、字有定序”,为今后我国的语文教学、出版印刷、新闻通讯、各种文字机器和电子计算机中文信息处理等,提供用字的规范。建国以后,特别是进入八十年代以后,我国的汉字整理工作主要就是在定量、定形、定音和定序四个方面展开的。目前,实现现代化的宏伟目标、整个民族进一步开发智力和培养人才的迫切要求、利用电脑进行语言文字信息处理的飞速发展和科教兴国的坚实步伐,把促进汉字的规范化、标准化提到了比以往任何时期都更为重要的地位。1986年的全国语言文字工作会议把促进语言文字规范化、标准化作为我国新时期语言文字工作方针任务的主要内容。1997年的全国语言文字工作会议指出,世纪之交这一重要历史时期对语言文字工作提出了更新更高的要求,语言文字工作面临着严峻的挑战,更面临着前所未有的发

展机遇。会议在汉字方面提出了:2010年以前,“汉字的社会应用基本规范,社会用字混乱现象得到有效的遏制,出版物用字、影视屏幕用字和计算机用字达到较高的规范水平”;二十一世纪中叶以前,“语言文字规范标准和各项管理制度更加完善”“语言文字规范化、标准化水平显著提高”“中文信息技术产品在语言文字规范标准方面实现较高水平的优化统一”等奋斗目标。当前,字形整理和汉字规范工作的对象是现行汉字,其目标就是要在以前汉字整理工作的基础上全面实现“字有定量、字有定形、字有定音、字有定序”。

### 三 现行汉字的定量

#### 3.1 汉字的字数

汉字总共有多少字?到目前为止,谁也答不上来。原因甚多:一是,汉字是“语素文字”,语素的数量本来就多,一个语素至少要用一个汉字甚至几个汉字来记录,汉字的字数怎么可能不多?就以化学元素为例。在近代以来新造的汉字中间,化学用字占了很大比例。八十年代中期,主管语言文字工作的部门跟化学界达成共识,明确宣布不再给105号以后的元素造新字,而改用序号表示。这一决定引起语文界一片喝彩声,以为从此以后创造新字的势头将会受到遏制。谁曾想到,仅仅过了十来年,化学



界又搬用“金字旁”加谐音的老办法,分别加上“杜”“喜”“波”“黑”和“麦”给 105 ~ 109 号元素造了五个新字:“𨮒、𨮓、𨮔、𨮕、𨮖”。此禁一开,只要新元素不断发现,带“金字旁”的新字就还会不断出现。二是,汉字是“表意兼标音的文字”,造字者可以从不同的角度去表达字义。喝水的“bēizi”,“仓颉甲”着眼于它是一种器皿,于是用“皿”作形旁,加上声旁“不”组成一个形声字;“仓颉乙”着眼于它当时使用的材料,于是用“木”作形旁,加上声旁“不”组成另一个形声字。汉字中充当标音符号的又是现成的汉字,造字者也可以挑选同音或近音的汉字来出任声旁。于是记录同一个语素“kù”(下衣)的形声字,“仓颉甲”在“衤”的右边选用了声旁“库”,“仓颉乙”在“衤”的右边选用了声旁“夸”;再加上多数汉字是多部件字,造字者又可以根据自己的喜爱确定该字的结构方式。记录家禽“é”的形声字,“仓颉甲”组合成左“我”右“鸟”,“仓颉乙”倒了个个儿,组合成左“鸟”右“我”,“仓颉丙”又把“我”放到“鸟”的上头,组成了上下结构。异体纷呈是汉字字数众多的另一个原因。三是,古今并用。汉字是一个开放的体系,又绵延不绝地使用了几千年。有些汉字后来不用了,在社会书面交际中已然消失,成了“死字”,但不必像“死人”那样要去注销户口,而且,它们在古书中还保存着;有些新的汉字被创造出来了,它们也不必像新生婴儿

那样要去申报户口,很多人甚至根本不知道它们的存在。四是专用字多。除了化学字以外,还有动植物名称专用字、古代学术和文物专用字、民族和宗教专用字、人名和地名专用字、行业专用字、译音和象声专用字、方言专用字和外族汉字,等等。凡此种,使汉字的总数要以万计。于是,汉字总共到底有多少字,成了一笔糊涂账。

遇到这个难题的时候,很多人常用的办法是列举历代字书的收字量。下面这张表开列的是从公元一世纪到公元二十世纪中国主要字书的收字数量:

年代(公元)	字 书	编 者	收字数
100 年 东汉	说文解字	许 慎	9353
230 年 魏	声 类	李 登	11520
400 年 晋	字 林	吕 忱	12824
500 年 北魏	字 统	阳承庆	13734
534 年 南梁	玉 篇	顾野王	22726
601 年 隋	切 韵	陆法言	12150
751 年 唐	唐 韵	孙 愐	15000
753 年 唐	韵海镜源	颜真卿	26911
997 年 辽	龙龕手鑑	行 均	26430
1008 年 宋	广 韵	陈彭年等	26194
1039 年 宋	集 韵	丁 度等	53525
1066 年 宋	类 篇	王 洙等	31319
1615 年 明	字 汇	梅膺祚等	33170
1675 年 明	正字通	张自烈	33440
1716 年 清	康熙字典	张玉书等	47043
1915 年 民国	中华大字典	陆费逵等	44908
1990 年中华人民共和国	汉语大字典	徐仲舒主编	54678
1994 年中华人民共和国	中华字海	冷玉龙等	85000 多

现在收集到的字数还在增加,因而,以收字齐全为特色的汉字字典的字种数量仍呈上升趋势。

如果学习和使用汉字真的要掌握七八万个汉字的形音义的话,那汉字将是世界上没人能够也没人愿意学习和使用的文字工具,它可能早就像西亚的钉头字、北非的圣书字和美洲的马亚文一样成为世界文字的历史化石。幸好《中华字海》一类字书里收录的汉字绝大部分是“死字”,就是历史上存在过而今天的书面语里已经废置不用的字。比如记写毛色不同的马,咱们今天记作“黑白杂毛的马”,而在古代书面语里就有个专用的字“𩇛”(bǎo);咱们今天记作“青黑色的马”,而在古代书面语里又有个专用的字“𩇛”(xuàn)。显然,今天采用语言手段代替古代大量专用字的办法要高明得多。另外,如果同一个字有几个不同的形体,字书也算作几个字,就像同一个人,先是留了胡子,后来戴了眼镜,再后来又镶了金牙,就算作四个人一样。比如“门 chuāng”的 chuāng,在汉字历史上至少有六种写法:“窗、窻、窓、憲、牕、牕”。如果把“死字”去掉,把重复多余的异体字去掉,那么同时在社会上通行的汉字(周有光先生称为“字种”)大约只有五六千到六七千个左右。据文字学家考证,中国历朝历代通行的汉字,虽然在具体的字种上略有差别,有新的字在不断产生,有旧的字在陆续隐退,但在总体上维持着通行字总量的相对稳

定。这里显然有汉字记录汉语和人们认知汉字的能力(周有光先生称为“记忆的心理极限”)等规律在起作用。

上面说的是有关汉字字数总的情况,下面再来看看现行汉字的字数。

现行汉字大致包含四部分字:

一是沿用古代汉字的字,学术界称为“传承字”。如:人、牛、父、母、天、地、山、木、水、火……这些字是汉字中最稳定的部分,约占 75%。

二是沿用古代汉字的字形,但音或义已经发生变化的字。如:“胺”,古代汉字的字义是“肉腐败变臭”,现行汉字则用来记录一种有机化合物;“她”,古代汉字的字义是“同‘姐’”和作为女子人名用字,现行汉字则用来记录女性第三人称代词。这类字约占 1.3%。

三是新造的字。字数不多,其中很多是象声字(如:嗨、嘹、咚、哎……)和化学用字(如:氨、氮、碘、钙……)。还有一些日常用字和方言字。这类字约占 2.7%。

四是简化字。从 20 世纪五十年代开始成为我国内地和马来西亚、新加坡等国的规范字形。约占 20% 强。(参考丁方豪《现代汉字造字法探索》,《语文现代化》丛刊第 10 辑,知识出版社,1990 年)

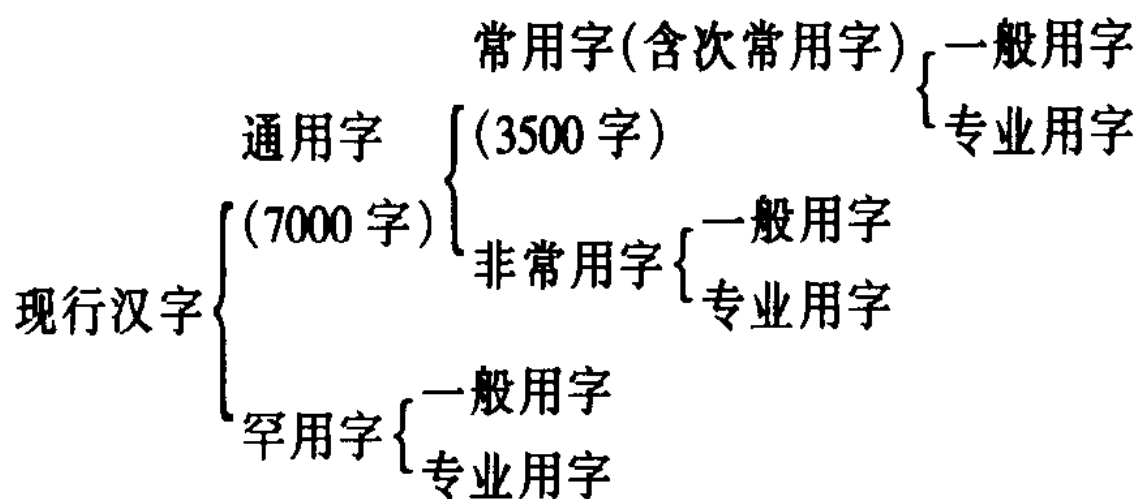
现行汉字的总字数现在也还没有一个准确的说

法,连多数语文工具书也不习惯提供准确的收字数,如《新华字典》1998年修订本肯定已用电脑录入,但“修订说明”还是习惯于提供收字的约数:“修订后的字典计收单字(包括繁体字、异体字)10000余个”(1992年重排本的“说明”是“本字典所收单字,包括异体、繁体字在内,共计一万一千一百左右”,似乎更接近确数)。有的语言文字学家认为,《新华字典》收录的字种数量大致接近现行汉字的总量。

### 3.2 常用汉字高频化

现行汉字可按使用频率分为通用字和罕用字。通用字指一般书刊报纸上共同用来书写现代汉语的字。1988年国家语言文字工作委员会和新闻出版署联合发布的《现代汉语通用字表》收了7000字。罕用字指使用频率很低的字,大部分是生僻的、仅仅适用于某一专业或某一领域的字。这部分字的数量不容易统计,所以,现行汉字的总字数也就难于确定下来。通用字中又可以主要根据使用频率分为常用字和非常用字。1988年国家语言文字工作委员会和国家教育委员会联合发布的《现代汉语常用字表》收了2500个常用字和1000个次常用字,共3500字。现行汉字的总体格局大致如下表:





有人可能要问,现行汉字总数在一万个上下,常用字却只有 3500 个,通用字也只有 7000 个,用它们来记录现代汉语够用吗?我们先请大家看几个数字:

司马迁的《史记》,全书 518284 字,只用了 4987 个不同的汉字;

孔尚任的《桃花扇》,全书 80121 字,只用了 3315 个不同的汉字;

曹雪芹的《红楼梦》,前 80 回 501113 字,只用了 3264 个不同的汉字;后 40 回 237132 字,只用了 2589 个不同的汉字;全书 738245 字,只用了 3384 个不同的汉字;(以上数字是[美]伊利诺大学邱锦华教授 1997 年 7 月在语言文字应用研究所作“常用汉字数”学术报告时提供的)

叶圣陶的《倪焕之》,全书 138000 字,只用了 3039 个不同的汉字;

老舍的《骆驼祥子》,全书 15 万字,只用了 2413 个不同的汉字;

赵树理的《三里湾》，全书 124000 字，只用了 2069 个不同的汉字；

毛泽东的《毛泽东选集》(1~4 卷)，全书 660273 字，只用了 2981 个不同的汉字。

这些数字告诉我们，人们在记写现代汉语时经常用到的不同汉字的数量是有限的。经过对大量书面语材料的统计(语料选自政论、文艺、新闻、科技等)，得出了下面一组非常有规律的数字：

不同汉字数	覆盖率(%)	不足率(%)
1000	90	10
2400	99	1
3800	99.9	0.1
5200	99.99	0.01
6600	99.999	0.001

这组数字告诉我们，在被统计的语料中，综合起来看，如有 1000 个高频字反复出现，它们就可占一般书面语总字数的 90%；如有 2400 个高频字反复出现，它们就可占一般书面语总字数的 99%；如有 3800 个高频字反复出现，它们就可占一般书面语总字数的 99.9%……这就是说，一个人如果掌握了位列高频字表前 1000 个汉字，他读一般的现代书报的时候，平均在 10 个字里会有 1 个字不认识；如果掌握了位列高频字表前 2400 个汉字，他读一般的现代书报的时候，平均在 100 个字里有 1 个字不认识；如

果掌握了位列高频字表前 3800 个汉字,那他读一般的现代书报的时候,平均在 1000 个字就只有 1 个字不认识了……老舍的《骆驼祥子》,全书总字数 107360 字,其中,光是“的”就出现了 4423 次,“他”就出现了 2573 次,“不”就出现了 2417 次。“的、他、不”三个字加起来的出现次数竟高达 9413 次,几乎占了全书总字数的 9%,也就是说几乎每 10 个字里就有 1 个“的”或“他”或“不”。可见,汉字字数繁多的缺点被常用字高频化这一规律缓解了。人们只要掌握了两三千个常用字,就可以读懂通俗的书报,就可以书写浅显的文章,就可以进行最一般的书面交际了。两三千这个数字已经不少,但汉字有它自身的种种特点和优点,使得人们在学习和使用的时候,通过努力,完全能够掌握,而且,学会了一批基本字和常用部件以后,识记汉字的速度会大大加快。这就是汉字字数虽多,但千百年来绵延不绝,一直是记录汉语的有效工具的重要原因。

记录现代汉语,据估计,大约至少需要六七千汉字,依据是:1965 年发布的《印刷通用汉字字形表》收录 6196 字,1981 年发布的《信息交换用汉字编码字符集·基本集》收录 6763 字,1988 年发布的《现代汉语通用字表》收录 7000 字。有人估计,掌握了 5200 个比较常用的汉字就能识读一般书刊用字的 99.99%。

### 3.3 如何实现现行汉字的定量

实现现行汉字的定量可以提高字典词典收字的科学性,可以帮助各级教育用字确定科学的数量和字种,可以促进语言文字信息处理的电脑化、自动化。

为了实现现行汉字的定量,首先必须根据记录对象,区分古代汉语用字和现代汉语用字,区分普通话用字和现代汉语方言用字。周有光先生称之为审定“字性”。

审定字性,要划清两个界线。第一,分清现代汉语用字和古代汉语用字的界线。现代汉语用字包括古今通用的字和现代汉语专用字,不包括古代汉语专用字。但是,现代汉语普通话在发展过程中会不断地吸收古代汉语中的有用成分。这就使得有些古代汉语用字进入了现代汉语书面语,变成了现行汉字。这说明在区分现代汉语用字和古代汉语用字的时候,有一个难于分清的中间地带,要根据所记录的词语在现代汉语中的稳定程度和参考它们的出现频率来仔细判断。第二,分清普通话用字和现代汉语方言用字的界线。如,有些方言词语已经进入普通话,记录它们的汉字就应该计入现行汉字(“尴”“尬”“甬”“掰”等);有些方言词至今还没进入普通话,记录它们的字就不应计入现行汉字(记录粤方言词的“冇”,记录吴方言词的“罳”等)。但也有若干字,因

为记写的语词是否已经进入普通话还不易判定,它们的身分也不易认定。

其次是根据使用范围,区分现代汉语通用字和现代汉语专业用字。这两类字都属于现行汉字的范围,区分的目的是为了便于分层定量。区分通用字和专业用字主要依据字的主要意义。专业用字按照字义主要可以分为六类:(1)史地用字。历史用字包括记录古国名、古器物名、古典章制度名等的字,如“邶、笏、卣、铨、朕”等。地理用字包括专用于记录山、水、湖、县、乡、村等名称的字,如“邛、汨、沔、岷、邨、邨、濮、埔、峒”等。(2)动植物名称用字。如“黑、犀、兕、獾、狷、獬、鼬、犴、狒、狨、麋、麂、獐、獭”和“杉、枸、椴、楸、栎、箬、樨、蕈、苕、柚、芎、苈”等。(3)科技用字。如“氢、氦、钙、锰、脘、酶、硫、羟、哌、冪、熵”等。(4)宗教用字。如“匍、稣、嗒、禱、醺、乩、筮”等。(5)民族用字。如“佉、傣、侗、畚、傈、傈、佻”等。(6)译音用字。如“啤、咖、啡、唛”等。

合理控制专业用字的数量是现行汉字定量工作的重要内容。如 20 世纪五六十年代我国曾成功进行更改部分生僻地名字的工作。

我国幅员辽阔,历史悠久,现有的地名中县级以上的就有三千多个。由于历史的原因,这些地名用字中,有的读音异于通常字音,如江西省东北部在上饶市和鹰潭市之间有个铅山县,“铅”不读 qiān,而要



读作 yán;有的地名用的不是现行的常用字、通用字,而是生僻字、罕用字,如陕西省西安市西南有个鄠县。《汉语大字典》注:“古邑名。本夏之扈国,秦置鄠邑,汉改鄠县。”这个古地名用字在现代汉语用字中是个生僻字、罕用字,除了记录地名以外,没有别的用处。一般读者不认识这个字形,读不出它的字音。对于多数读者来说,这个字已经失去了标记地名的作用,并给信息的传递和交流带来了麻烦。

为了消除这个困难,使地名用字尽可能易认易读,1955年11月23日,中华人民共和国教育部和中国文字改革委员会建议由内务部和文改会同这些有关的地方人民委员会协商,请他们提出用常用的同音字或简字来代替地名中的生僻字。我国少数民族的汉文名称所用的字亦有不少生僻或笔画繁复的,建议由民族事务委员会和文改会同本民族协商,改用较简易的字。1956年6月28日,中华人民共和国内务部和中国文字改革委员会联合发出征求对生僻地名字简化意见的函件。同年10月17日中国文字改革委员会又给各省、市、区人民委员会发出关于建议更改生僻地名字问题的函件。

当时确定的有关更改地名生僻字的原则是:

1)凡不容易读出音来的生僻地名用字,改用读音明确的常用字,如“盩厔”可改作“周至”;

2)同音代替字最好和原来的字同音同调,如“瓊

璿”可改作“爱辉”；

3)地方读音与书面读音(即普通话读音)不相符合的地名用字,可按书面读音来更换新的同音字,如“璿源”不作“媿源”而作“门源”；

4)为了使新旧地名用字在推行和使用上易于联想,在同音代替时,应尽可能使新的字形和原字形相近似,如“盱眙”可用“吁怡”；

5)有些生僻地名用字,原来是不大通用的异体字或通假字,这些字可改用音义相同或相近的常用字,如“忻县”改作“欣县”；

6)有些生僻地名字是后来加上不必要的形旁的,可省去形旁,如“嵯县”可改为“乘县”等；

7)某些地名用字,由于行政区域的变更,根据该地历史或地理上的条件,也可改用其他名称,不受上述原则限制,如山西的“崞县”不改作“郭县”而改称“平原县”等；

8)少数民族地区地名字的读音,可以根据当地的具体情况,按普通话音节作准确而通俗的译音定名。

当时规定更改地名生僻字的程序是:县以上地名用字的更改,由各省、市、自治区提出更改意见,报内务部和文改会组织成立的地名审改组审核同意后再报请国务院审批;县以下地名用字的更改,由各省、市、自治区审批,报内务部备案,同时抄送中国文

字改革委员会。截至 1964 年 8 月 29 日,经国务院批准更改的县以上的地名用字共有 35 处、36 字。

再次,是主要根据使用频率,区分现代汉语常用字和现代汉语非常用字。

现代汉语常用字指在日常语体的现代汉语书面语中使用频率高的字。

历代积累的汉字总数数以万计。近年出版的汉字字典收字总数从五万多逼近八万多,但历代日常书面语常用的汉字一般都控制在三四千字。这是字数繁多的汉字千百年来能为人们接受、成为汉族人民记录汉语的主要文字工具的一个重要原因。也正因为这样,选取当代的常用字作为学习和使用汉字的重点,是历代历朝整理和教学汉字的一个重点内容。历史上的《三字经》《千字文》都是用当时的常用字编写的识字课本。

现代常用字的研究和选取工作,较有影响的有:1928 年陈鹤琴编制的《语体文应用字汇》,收 4261 字;1952 年中央人民政府教育部公布的《常用字表》,收 2000 字。考虑到《常用字表》公布已有三十多年,社会用字的情况发生了许多变化,为了适应当前语文教学、词书编纂以及汉字机械处理和信息处理等各方面的需要,国家语言文字工作委员会组织力量从 1986 年起开始研制《现代汉语常用字表》,并于 1988 年 1 月 26 日由国家语言文字工作委员会和

国家教育委员会联合发布。

《现代汉语常用字表》是在过去同类工作及其取得的成果基础上编制的。所用的统计资料包括从1928年到1985年发表的共15种。选取的原则主要有四条：

1)统计原则——确定一个字是否常用,主要的依据是该字在书面语中出现的频率。为了提高统计的科学性,要保证统计语料的足够数量。在字量适度的基础上尽量扩大语料的覆盖面。

2)分布原则——“分布”就是字在各组语料中出现的次数。甲字只在一组语料中出现,而乙字在多组语料中出现,乙字的分布率就大于甲字。分布应兼顾语料类型和语料时间跨度,即要注意统计语料的类型,保证选取的字不仅频率高而且分布在各种学科的语料中;同时注意适当拉长语料的时间长度,纵观各个不同时期的用字情况,以此来衡量该字的使用情况是否稳定。尽量选取稳定的字可以避免选字的偶然性。《现代汉语常用字表》语料选材的时间范围是从1928年到1986年。在此时间区域内采用不等密度抽样,即抽样量按时间顺序递增,以近期的资料为主,使选取的字有较大的实用性。

3)构词构字原则——各个汉字的构词能力是不一样的。一般说来,构词能力强的字,在书面语中出现的次数就多,它的使用频率就高。但不能把构词

能力的强弱作为确定该字是否常用的惟一标准,因为,有一些字的构词能力很弱,甚至等于零,但它却是地地道道的常用字,如“又”“很”“也”等等。许多汉字除了构词能力以外,还有构字能力,常常作为构字部件跟别的部件组合成新字。如“非”,既能独立出现,还能构成“匪、啡、菲、翡、诽、翡、悱、绯、樨、腓、扉、斐、痱、蜚、篚、霏、鲱”等字,这二者就保证了它的常用字的地位。

4)常识原则——这一原则是为了弥补统计原则的不足而确定的。有些字虽然在现代汉语书面语中出现的次数不多,但它记录的语词确实是日常生活中不可或缺的,如“厕”,如果单纯根据统计数字把它排除在常用字之外,实在是悖常理。有了常识原则,就可以把这类字列入常用字的行列。

综合运用以上原则,《现代汉语常用字表》选收了 2500 个常用字、1000 个次常用字,共 3500 字。

经过 10 年的使用,《现代汉语常用字表》在语文教学和其他方面发挥了积极的作用。但实践中又发现有少数常用字使用机会甚少,又有少数比较常用的字在《现代汉语常用字表》中缺收。这就提出了一个在经过若干年实践的基础上,像其他规范标准一样,正常地进行修订的问题。

现代汉语通用字指一般报纸书刊上流通使用的记录现代汉语的字。

1956年8月,中国文字改革委员会曾印发《通用汉字表草案(初稿)》。该表收5390字,分常用字1500个、次常用字2004个和不常用字1886个。不常用字分为六栏:文言成分、姓名、史地、动植物、科学技术和其他(含宗教、民族、方言、译音等)。1959年12月,文化部根据“革新铅字字体座谈会”的意见,委托中国文字改革委员会、教育部和中国科学院语言研究所联合组成汉字字形整理组,对中国文字改革委员会提出的《汉字字形整理方案(草案)》进行修订。1960年提出《通用汉字字形表(草案)》。根据各方面的意见,决定把《通用汉字表(草案)》和《通用汉字字形表(草案)》合并成《印刷通用汉字字形表》。1965年1月发布《印刷通用汉字字形表》,共收字6196个,是我国五六十年代进行的汉字简化和整理工作成果的集大成,反映了那一个时期现代通用汉字的面貌。

6196这一字量符合常人对通用汉字约在6000至9000字范围内的估计。北京新华字模厂编集的《字目表》收录6805字,反映了印刷一般书刊用字的数量;中国科学院印刷厂编集的《印刷字表》收录9353字,反映了印刷专业书刊用字的数量;70年代邮电部编印的《标准电码表》正编收录7921字,补遗收录1396字(主要是姓名、地名、行业、计量、方言等用字);80年代国家标准局发布的《信息交换用汉字



编码字符集·基本集》收录 6763 字。这几个字表的收字量可以证明《印刷通用汉字字形表》收录的字数基本符合现代汉语通用汉字的实际。

考虑到社会用字近年来发生了较大的变化,六十年代发布的《印刷通用汉字字形表》已不能完全适应实际的需要。为了促进语言文字规范化,满足出版印刷、信息处理以及其他各方面的需要,语言文字主管部门组织力量编制了《现代汉语通用字表》,由国家语言文字工作委员会和中华人民共和国新闻出版署于 1988 年 3 月 25 日联合发布。

《现代汉语通用字表》是在《印刷通用汉字字形表》的基础上制订的。另外还统计了从 1928 年陈鹤琴的《语体文应用字汇》(4261 字)到 1987 年新华通讯社技术研究所研制的《1986 年度新闻信息流通频度》(6001 字)等 19 种资料。选取通用字时综合运用以下原则:

1)根据汉字的使用频率,选取使用频率较高的字;

2)在使用频率相同的情况下,选取学科分布广、使用度较高的字;

3)根据汉字的构词能力,选取构词能力较强的字;

4)根据汉字的实际使用情况斟酌取舍,书面语中很少使用,进行用字统计时往往统计不到,但在日

常生活中却比较常用,像这类字,也应适当选取。

《现代汉语通用字表》收录 7000 字,其中包括了《现代汉语常用字表》收录的 3500 字。字形标准依据《印刷通用汉字字形表》,未作新的调整。

实践证明,《现代汉语通用字表》基本符合社会实际用字的需要,但由于制订过程中,追求 7000 这个整数,在使用频率失去作用的时候,又主要根据编制者自己的字感选定,在字性的审定上还欠火候,致使一部分罕用字列入通用字表。这就提出了在经过若干年实践以后对《现代汉语通用字表》及时地、审慎地、科学地加以修订的任务。

通过以上的区分、选择,就可以把现行汉字分成几个层次,比如根据使用频率分为常用字、次常用字、通用字、罕用字,根据使用范围分为扫盲用字、小学用字、一般用字和专业用字等,进而科学地确定各个层次汉字的数量和具体的字种,逐步积累,不断调整,最后制订出“现代汉语用字全表”。

#### 四 现行汉字的定形

作为社会重要交际工具的文字,总是具有两面性:一方面是个人的书写中为了追求快速方便而出现的难以避免的随意性,另一方面是社会交际为了提高效率而反复要求的字形的统一性。这是一对方向不同的分力。文字字形就是在这样两种矛盾力量的

交替作用下向前演进的。就社会而言,为了提高交际效率,就要在社会公共交际活动中遏制文字字形的个人变异、地区变异和行业变异,提倡字形的规范化、标准化,以保证社会书面交际活动的畅通和高效。

中华人民共和国成立以后,为实现现行汉字的定形做了大量工作。如整理规范了汉字的笔画形状和常用字通用字的笔画数目;整理淘汰了部分异体字形;公布、推行了具有“正体”身分的规范简化字形;整理了部分计量单位名称用字;删汰了旧字形,推行了新字形,等等。

#### 4.1 整理、规范现行汉字的笔画

笔画是构成汉字字形的最小单位。现有的笔画是在长期书写实践中约定俗成的。整理、规范现行汉字的笔画形状和规定常用字、通用字的笔画数目是汉字定形的基础。

现行汉字的笔画数,少的只有一笔(如“一”“乙”),多的可以到二三十笔(如“轟”24笔,“𪚩”36笔)。识字教学、查检索引和机器处理汉字,都需要准确的笔画数。要能准确计算笔画数,得有两个条件,一是统计的对象必须是规范字形(如“鬼”,规范字形是9笔,旧字形是10笔;“象”,规范字形是11笔,旧字形是12笔),二是要遵守汉字书写的基本规则(至少有两条:在同一笔画上,笔尖只能走一次,不

能走回路；横笔的笔向只能从左到右，竖笔、撇笔、捺笔的笔向只能从上到下）。有了这两条，一般字的笔画数都能准确地数出来。当一个笔画的头跟另一个笔画的头相接的时候，有时会遇到麻烦。判别的方法大致有这么几条：

(1)笔画与笔画在左上角相接时，分作两笔，如：厂、几、口、日。

(2)笔画与笔画在左下角相接时，有两种情况：全包围结构的字，分作两笔，如：回、田、囚、四；不是全包围结构的字，连为一笔，如：山、区、臣、画。

(3)笔画与笔画在右上角相接时，连作一笔，如：司、刁、句、月。

(4)笔画与笔画在右下角相接时，分作两笔，如：由、自、省、雷。

现行汉字中，9笔到12笔的字比较多。笔画数的分布面不是随意形成的，而是现行汉字字形既要便于区别，又要字形简易，兼顾“明”和“简”的产物。

笔画形状简称笔形。现行汉字的笔形可以分为基本笔形（有的称平笔笔形）6种和派生笔形（有的称折笔笔形）25种：

[illegible]

现行汉字笔形中以横和竖出现的次数最多,这就决定了现行汉字的方块形式和以直线为主很少弧线的特点。

笔形有时要受到左右邻笔笔形的制约。为了保持全字的紧凑匀称,按照规范标准,有的笔画在特定场合要改变形状,如“工、土、马、王”等部件作左偏旁时,它们的末笔“横”改“提”(巧功项劲、地坏坦塔、馱驼骏骗、玩球璫璜);“七、己、光、毛”等部件作左偏旁时,它们的末笔“竖弯钩”改“竖提”(切沕、改凯、辉耀、橈橈)。在一个字中,邻近的两笔(甚至三笔)都是“捺”时,必须把其中的一“捺”或两“捺”变成“点”(这在书法上叫做“避重捺”),比如“奏秦”是下边的“捺”避上边的“捺”,“达灸”是上边的“捺”避下边的“捺”,“黍餐”是上下的“捺”避中间的“捺”。

目前,现行汉字笔画的分类、名称还不统一;有些笔画形状是“刻意求简”那个时代的产物(如“专”的第三笔“竖折撇”、“及”的第二笔“横折折撇”等),应该在适当的时候予以分化整理;把二三十种笔形归入“横、竖、撇、点、折”五种基本笔画类型时也有个别笔形归类不当的问题(如把“竖钩”归入“竖”);为了呈现内聚、匀称的风格,在有些字中有的笔画要发生形变,我们的规范字形中笔画形变任意性的现象不是一处两处,所有这些表明我们在笔画这个层次上整理工作做得还不细,规范水平还有待提高。



## 4.2 “现行汉字规范字形”的具体涵义

国家主管部门推行的规范字形是就中国内地日常的书面交际而言的。所谓汉字规范字形,指经过有关部门研究整理,由政府主管部门公布的现行汉字字形,主要包括:

(1)1986年10月10日由国家语言文字工作委员会重新发布的《简化字总表》所收录的简体字字形;

(2)1988年3月25日由国家语言文字工作委员会、中华人民共和国新闻出版署联合发布的《现代汉语通用字表》所收录的汉字字形。

相对而言,日常书面交际中现行汉字不规范的字形,主要包括:

(1)《简化字总表》中被简化的繁体字字形,如:豐龍;

(2)1986年国务院废止的《第二次汉字简化方案(草案)》收录的简体字字形,如:尸(展)亡(宣);

(3)经过调整后的《第一批异体字整理表》淘汰的异体字字形,如:邨疋;

(4)1977年由中国文字改革委员会、国家标准计量局联合发布的《部分计量单位名称统一用字表》中淘汰的译名用字字形,如:氩湮;

(5)社会公共交际中出现的自造的简体字字形,如:埴(增)忒(德);

(6)1965 年因中华人民共和国文化部、中国文字改革委员会联合发布《印刷通用汉字字形表》而淘汰的相应的旧字形,如:直(直)吴(吴)。

### 4.3 整理异体字

由于现行汉字由意符、音符和记号组合而成,意符选取的角度可以因人而异,音符又不同于拼音文字中与音素或音节固定对应的字母,所以,一字多形的现象在汉字的历史上比比皆是。如果从严掌握,我们只把音同义同形不同的字称为异体字。为了精简汉字的字数,20 世纪五十年代,中国进行了整理异体字的工作。1955 年 12 月 22 日,中华人民共和国文化部、中国文字改革委员会联合发布《第一批异体字整理表》,并通知从 1956 年 2 月 1 日起在全国实施。《第一批异体字整理表》列异体字 810 组,每组最少收异体字 2 个,最多收异体字 6 个,全表共收录异体字 1865 个。每组中选定一个字形作为规范字形,共 810 个,俗称“选用字”,严密的说法应是“汉字规范的异体字”,如:“决、決”这一组异体字中,选定“决”为规范的异体字;“窗、窻、窓、窻、牕、牕”这一组异体字中,选定“窗”为规范的异体字(“选用字”)。

《第一批异体字整理表》收录的字是宽泛意义上的异体字,包括三类四种:

- (1)音义全同类,如:“翅”——“𪔐”,“稗”——“稗”;
- (2)音义包孕类。又可分两种:

1)大包小,如:“痲”①féi 中风病。②fèi a.  
痲子,b.同“腓”,病,枯萎。

“痲”fèi 痲子;

选用的“痲”,其音义都多于被淘汰的“痲”。  
这类选用字跟被淘汰字的音义关系,俗称  
“大包小”。

2)小包大,如:“熔”róng 以高温使固体物质  
转为液体。

“熔”róng a. 铸铁的模型,b.  
销熔,c.熔化,d.喻陶冶,e.矛  
属。

选用字“熔”的字义少于被淘汰字“熔”。这  
类选用字跟被淘汰字的音义关系,俗称“小  
包大”。

(3)音义交叉类,如:“蝶”①dié 蝴蝶,②tiē“蝶  
螯”,虫名。

“螯”dié a. 同“蝶”,b. 一  
种海蟹。

选用字“蝶”的音义跟被淘汰字“螯”的音义,谁  
都包孕不了谁,是相互交叉的关系。

1956年3月23日,该表的发布单位发出修正通  
知,恢复“阪”“挫”规范字形的身分。1986年10月  
10日,国家语言文字工作委员会重新发布的《简化  
字总表》中恢复了“沂、谯、晔、薈、诃、鲚、绌、划、鲢、

诨、讎”等 11 字的规范字形身分。1988 年 3 月 25 日,国家语言文字工作委员会、中华人民共和国新闻出版署联合发布的《现代汉语通用字表》恢复了“翦、邱、於、澹、骼、徬、菰、溷、徽、薰、黏、桉、愣、暉、凋”等 15 字的规范字形身分。1993 年 9 月 3 日国家语言文字工作委员会文字应用管理司发出的《关于“镕”字使用问题的批复》恢复了“镕”字的规范字形身分。经过上述调整,《第一批异体字整理表》(见《国家语言文字政策法规汇编》,语文出版社 1996 年)由原来的 810 组减少为 794 组,应淘汰的非规范异体字字形由原来的 1055 个减少为 1024 个,选定的规范的异体字字形(“选用字”)由原来的 810 个减少为 794 个。

#### 4.4 推行规范的简化字

汉字的简化字,跟“繁体字”相对应,指同一汉字所具有的结构较简、笔画较少的字形。如记录“yìshù”这个词,现在中国内地用的字形是“艺术”,原先用的字形是“藝術”。对于“藝”来说,“艺”是简化字;对于“術”来说,“术”是简化字。

中华人民共和国成立以后,一个普及教育、提高人民文化水平的高潮迅速出现,于是,选定常用字和简化汉字的工作很快被提上日程。1950 年中央人民政府教育部社会教育司曾经编制《常用简体字登记表》,征求意见。1951 年拟出《第一批简体字表》,

收常用简化字 555 个。1952 年中国文字改革研究委员会开始进一步研究并草拟简化汉字笔画和精简汉字字数的方案。1954 年底编成《汉字简化方案草案》。该草案于 1954 年 12 月在中国文字改革委员会举行的第一次全体会议上获得通过,并于 1955 年 1 月公开发布征求意见。这个草案包括三个字表:“798 个汉字简化表”“拟废除的 400 个异体字表”“汉字偏旁手写简化表”。1955 年 7 月国务院成立以董必武为首的汉字简化方案审定委员会。同年 9 月中国文字改革委员会根据群众意见修订了草案:“拟废除的 400 个异体字表”修正后由有关部门另行公布实施;“汉字偏旁手写简化表”由于只简化手写偏旁,不便推广,只能作为一种提倡和介绍;“798 个汉字简化表”减为 512 个,再增加 56 个简化偏旁。审定委员会同意上述修订,决定把留下的内容分成“简化字表”和“简化偏旁类推表”,逐字审查,投票表决。1955 年 10 月全国文字改革会议讨论《汉字简化方案草案》,并提出修正意见,最后通过了包括 515 个简化字和 54 个简化偏旁的《汉字简化方案》,建议中国文字改革委员会把《方案》提请国务院审定公布。1956 年 1 月国务院全体会议第 23 次会议通过《汉字简化方案》。1 月 31 日《人民日报》发表国务院“关于公布《汉字简化方案》的决议”和《汉字简化方案》。《方案》共分三部分:“汉字简化第一表”收 230 个简

化字；“汉字简化第二表”收 285 个简化字；“汉字偏旁简化表”收 54 个简化偏旁。第一表的简化字自公布之日起立即在全国印刷和书写的文件上通用。第二表的简化字和第三表的简化偏旁先行公布试用。1964 年 5 月，中国文字改革委员会根据国务院有关类推简化的批示，编制出版了《简化字总表》。《总表》由三个表组成：第一表收不作简化偏旁用的简化字，共 352 个，第二表收可作简化偏旁用的简化字，共 132 个，还收录简化偏旁 14 个，第三表收应用第二表所列简化字和简化偏旁类推得出来的简化字，共 1754 个。《总表》除去在第一表和第三表重复收录的“签”“须”二字，实收简化字 2236 个。

1977 年 12 月 20 日，《第二次汉字简化方案(草案)》发表。《草案》分两个表：第一表收 248 个简化字，自发表之日起陆续在图书报刊上先行试用；第二表收 605 个简化字，还要在群众中广泛征求意见。两表共收 853 个简化字。由于《草案》中的简化字很多还不够成熟，推行步骤又过于急促，新公布的简化字在群众中引起许多不同意见。1978 年 4 月教育部发出通知，要求教材停止使用新公布的简化字。7 月《人民日报》《解放军报》也相继停止使用。从此，《第二次汉字简化方案(草案)》进入讨论、修改阶段。1986 年 6 月 24 日，国务院决定废止《第二次汉字简化方案(草案)》。

1986年10月10日,根据国务院指示,由国家语言文字工作委员会重新发布《简化字总表》。《总表》共分三个表:第一表收“不作简化偏旁用的简化字”350个;第二表收“可作简化偏旁用的简化字”132个和“简化偏旁”14个(“讠”“亻”“纟”等);第三表收“应用第二表所列简化字和简化偏旁得出来的简化字”1753个,全表共收录简化字2235个。跟1964年5月由中国文字改革委员会发布的《简化字总表》相比较,主要的调整内容为:(1)“叠”“覆”“像”“囉”不再作“迭”“复”“象”“罗”的繁体字处理。因此,原第一表的“迭[叠]”“象[像]”删去,“复”字字头下的繁体字[覆]删去;原第二表“罗”字字头下删去繁体字[囉];在第三表新增“啰[囉]”。(2)“了[瞭]”加注:“瞭:读 liǎo(了解)时,仍简作了,读 liào(瞭望)时作瞭,不简作了。”作了上述调整以后,《简化字总表》收录的简化字总数由2236个减为2235个。

#### 4.5 统一计量单位名称用字

1977年7月20日中国文字改革委员会、国家标准计量局联合发出《关于部分计量单位名称统一用字的通知》,后面附了《部分计量单位名称统一用字表》。表中规定的计量单位用字,长度单位用字有:(方括号前为规范计量单位用字,方括号内为被淘汰的计量单位用字。下同)海里[浬、海浬],英里[哩],英寻[呎、浔],英尺[呎],英寸[吋];面积单位有:英



亩[嗽、亩];容量单位有:升[公升、蚌],蒲式耳[喇],加仑[呷、罄];重量单位有:英担[咄],英石[哂],盎司[啊、英两、温司],格令[喱、英厘、克冷];其他各科单位有:功率单位-千瓦[瓩],压力单位-托[乇],响度级单位-方[叻],响度单位-宋[唻],音调单位-美[嘆],纤度单位-旦[紫]、特[纟]。

#### 4.6 整理、统一印刷体字形

从20世纪五十年代中期开始,中国内地又用了十来年时间进行印刷字形的整理、统一工作。在这以前,出版物上的字形混乱现象比较严重。由于字体、笔画、部件或者整字结构的不一致,造成一字异形甚至多形,既有印刷宋体与楷体的分歧,也有宋体内部的不一致,给学习和使用带来不便。中国文字改革委员会于1955年提出进行汉字印刷字模标准化工作,根据“从简、从俗”的原则,消除印刷宋体字形的不一致,并使印刷宋体向楷体靠拢。1956年拟订了《标准字形方案(草案)》。1957年拟出修订草案,并改名为《汉字字形整理方案(草案)》。1959年12月,文化部召开关于革新铅字字体座谈会,并委托中国文字改革委员会、教育部和中国科学院语言研究所等单位组成汉字字形整理组。经过深入的研究、对比,对选取的六千多字的笔画形状、笔画数目、笔画顺序和结构方式作出规定,编成《通用汉字字形表(草案)》,再向全国有关部门征求意见。经过多次

修订,于1964年5月拟定《印刷通用汉字字形表》,为6196个通用汉字规定了规范的印刷字形,包括笔画数目(如:印刷宋体字形的折笔,由原来的断笔改为连笔,作一笔处理,于是“牙”由五笔改为四笔、“印”由六笔改为五笔等)、笔画形状(如:把印刷宋体的长方点、斜方点、横点、竖点、撇点合并为侧点)、笔画顺序(如:“象”先写扁框,再写第六笔“撇”)和间架结构(如:“𠂔”由原来的左下包围结构改为左右并列结构,“惑”由原来的上右包围结构改为上下结构)。1965年1月文化部和文字改革委员会联合发布了《印刷通用汉字字形表》。整理、统一印刷体字形的成果后来又反映在1988年发布的《现代汉语常用字表》和《现代汉语通用字表》中。

经过多年的推行,基本上消除了出版物上字形混乱的现象。但是,由于制订《印刷通用汉字字形表》时,限于当时的研究水平,所定的规范字形中存留不少字模技工的习惯性和主观随意性的痕迹,影响了规范字形的标准化水平,如:独体字“美”的最后一个横笔短于上一个横笔,但“羹”中“美”的最后一个横笔长于上一个横笔;又如:“辱”为上下结构,“褥”“蓐”等字中的“辱”成了半包围结构(“辰”的第二笔“撇”包孕下边的“寸”)。1988年发布的《现代汉语通用字表》照搬《印刷通用汉字字形表》的字形,使我国的规范字形停留在1965年的水平。近年,主

管部门的工作没能跟上中文信息处理飞速发展的步伐,抓电脑用字的规范化标准化明显滞后,力度不够,目前电脑软件用字还相当混乱。“继续加快中文信息处理急需的规范标准的制定”已作为“当前的主要任务”在 1997 年的全国语言文字工作会议上被提了出来。

## 五 现行汉字的定音

### 5.1 推广“普通话”任务的提出

1954 年 12 月,作为国务院的直属机构,中国文字改革委员会成立,任务是推行各项文字改革的具体工作。当时提出的 1955 年的任务是(1)制定《汉字简化方案》;(2)制定《汉语拼音方案》;(3)研究和推行标准音(普通话)的教学。

1955 年 10 月,教育部和中国文字改革委员会联合召开了全国文字改革会议。教育部部长张奚若在会上作了题为《大力推广以北京语音为标准音的普通话》的报告。会议讨论并通过了张奚若的报告。代表们对什么是普通话取得了基本一致的认识,解决了语言学界长期以来对这个问题的争论。普通话的定义,由 1955 年全国文字改革会议的“以北京语音为标准音的普通话”,改为 1955 年现代汉语规范问题学术会议的“普通话以北方话为基础方言,以北京语音为标准音”,最后修订为 1956 年国务院“关于

推广普通话的指示”中的“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”。

## 5.2 《普通话异读词三次审音总表初稿》的研制

1955年10月,继全国文字改革会议之后,中国科学院又召开了现代汉语规范问题学术会议。会议提出了组织普通话审音委员会、组织词典计划委员会、拟订方言普查计划等开展具体工作的建议。

根据国务院的规定,中国科学院语言研究所负责普通话语音、词汇和语法规范的研究、宣传。1956年1月,语言研究所成立了负责审订普通话常用词汇读音的非常设性学术机构“普通话审音委员会”,聘请丁西林、丁声树、舒舍予、罗常培、吴文祺、吴晓铃、周有光、周祖谟、徐世荣、高名凯、陆志韦、陆宗达、欧阳予倩、齐越、魏建功等15人为委员,开始审订普通话异读词的读音。1957年到1962年,普通话审音委员会分三次发表了《普通话异读词审音表初稿》,审订了1800多个异读词的读音,比如:“脓”有nóng/néng两种读法,审定为“一律念nóng,不取néng音”;“波浪”有bōlàng/pōlàng两种读法,审定为“‘波’一律念bō”。并于1963年辑录成《普通话异读词三次审音总表初稿》,由文字改革出版社出版。《初稿》公布后,受到文教、出版、广播等部门广泛重视,对现代汉语的语音规范和普通话的推广起了积极作用。

### 5.3 《普通话异读词审音表》的研制

进入 20 世纪八十年代,距《普通话异读词三次审音总表初稿》发表已过了二十来年。随着语言的发展,《初稿》中原审的一些词语的读音需要重新审定;同时,作为语音规范化的标准,《初稿》也亟须定稿。因此,1982 年 3 月,中国文字改革委员会决定重建普通话审音委员会,对《初稿》进行修订工作,并同意王力、李荣、周祖谟、徐世荣、夏青为审音委员会五人小组成员,王力任主任。委员会由 19 位委员组成,除五人小组成员外,还有:曹乃木、方伟、傅兴岭、高景成、蒋仲英、梁容若、陆宗达、孙德宣、孙修章、吴青、徐仲华、俞敏、张志公、周有光。普通话审音委员会以符合普通话语音发展规律为原则,以便利广大群众学习普通话为着眼点,采取约定俗成、承认现实的态度进行修订工作。经过三年时间的征询、集纳、研究、审议,完成《普通话异读词审音表》的研制工作。仍用原来的名称,形式则改为以字系词,其定音、举例,尽量保持原貌;在原来基础上,也作了一些增删、更改、补充,如:“呆板”,《初稿》审定为 áibǎn,改定为 dāibǎn,“取消 ái 音,统读 dāi”/“驯服”,《初稿》审定为 xúnfú,改定为 xùnfú,“取消 xún 音,统读 xùn”,共修订《初稿》词语 41 条;“嗟叹”有 jiētàn/juētàn 两种读法,《审音表》“取消又音 juē”,审定为 jiētàn;“装帧”原读 zhuāngzhèng,《审音表》“取消 zhèng

音,统读 zhēn”,审定为 zhuāngzhēn,共增补 16 条;还删汰了《初稿》中的部分词条,有的是“本无异读”的,如“队伍”(duìwǔ)、“爆竹”(bàozhú),有的是“罕用词语”,如“定子”(发电机上的一种机件,dìngzǐ)、“氈氈”(一种毛织品,pǔlu),有的是“方言土音”,如“翘辫子”(死,qiàobiànzi)、“归里包堆”(guīlibāozuī,“堆”在别的地方念 duī),有的是“不常用的文言词语”,如“不啻”(bùchì)、“伧夫”(cāngfū),有的属于“轻声不审”,如“屏息”(bǐngxī)、“糊涂”(hútu),有的属于“音变现象,不必审订”,如“胡里八涂”(húlibātū)、“伙计”(huǒji),等等。1985 年 12 月 27 日,国家语言文字工作委员会、国家教育委员会、广播电视部联合发出《关于〈普通话异读词审音表〉的通知》,正式发布普通话审音委员会拟定的《普通话异读词审音表》,规定自公布之日起,文教、出版、广播等部门及全国其他部门、行业所涉及的普通话异读词的读音、标音,均以此表为准。普通话异读词的审音工作减少了汉字的异读,减轻了学习和使用汉字的难度。

《普通话异读词审音表》审订的主要是普通话有异读的词和有异读的作为“语素”的字。不列出多音多义字的全部读音和全部义项。由于轻声问题比较复杂,除《初稿》涉及的部分轻声词以外,《审音表》一般不予审订。人名、地名的异读审订,除原表已涉及的少量词条外,留待以后再审。

《普通话异读词审音表》经过十多年的推行,在实践中也发现了一些不如人意的地方,比如:对所审订的词语原来有哪些异读,为什么取此舍彼,没有交代;多数字音注为“统读”,有的“统”得过宽,如“胜”(shèng 统读),没照顾到有一种有机化合物“胜”,应该读 shēng;等等。这表明汉字字音的定音工作还有待改进。

## 六 现行汉字的定序

### 6.1 汉字排序的难点

两个或两个以上的汉字列在一起,就有一个排序问题。汉字字数的众多使汉字的科学排序显得分外重要。编制工具书索引、编排图书档案资料、排列姓名次序,以及研制电脑输入汉字的编码方案等,都需要有一个好的排序方法,以方便查找。

汉字有形、音、义三方面的属性,与之相应的就有形序、音序和义序三类排序法。汉字的复杂性和使用汉字人员的广泛性决定了汉字的排序方法不可能单一化。从汉字本身来说,汉字的字形结构复杂,用形序法有些字不便查检;汉字表音功能弱,用音序法常常会遇到不会读的字;汉字表义系统脉络错综,不易编排严格的序列,于是造成形序法、音序法和义序法并用的复杂局面。从使用者来说,普通话说得比较正确又熟练掌握汉语拼音的人,使用音序法比



较顺手;不会汉语拼音的人,使用部首法、笔画笔形法或四角号码法比较方便;小型字典,按音序排列字头比较适用;大型字典,生僻字多,按读音查字就很不方便。长期以来,一直有人在研究汉字的排序法,20世纪五十年代有关部门也组织过汉字排序法的讨论,取得不少成果。1961年,文化部、教育部、中国文字改革委员会和中国科学院语言研究所曾联合组成汉字查字法整理工作组,审核各类查字法方案,并在1964年4月提出“拼音字母查字法”“部首查字法”“四角号码查字法”和“笔形查字法”四种草案,推荐给文化、教育和出版界试用,以促进汉字查字法趋向统一。1983年6月又由上海辞书出版社、汉语大词典编纂处、汉语大字典编纂处、商务印书馆和中国社会科学院语言研究所等五个单位组成工作组拟订统一的部首排检法方案,并于同年由中国文字改革委员会和国家出版局联合发布了“汉字统一部首表(草案)”,共立201个部首,但似乎到现在还没能达到预期的规范目的。排序法或称查字法的规范问题尚需继续努力。

## 6.2 音序查字法

音序查字法就是按照字音来排列一个个汉字,建立所收全部汉字的序列。早年用1918年公布的注音字母作为注音的工具,于是就用注音字母的顺序编排工具书中汉字的序列。1958年发布了《汉语

拼音方案》，从此以后，我国出版的用音序法编排的工具书就都按汉语拼音字母的顺序排列汉字。

汉语拼音查字法按照汉字的汉语拼音字母顺序编排单字。声母韵母相同的，再按声调阴平、阳平、上声、去声、轻声的顺序排列。同声同韵同调的字再附加别的条件排列。有的按字的笔画数，如“青”“清”“轻”“倾”都读 qīng，按笔画数排成“青”“轻”“倾”“清”的序列。笔画数也相同的，再按笔画形状区分先后，顺序是横、竖、撇、点、折。比如“办”“半”“伴”“扮”“瓣”都读 bàn，先按笔画数的多少排成“办/半/伴扮/拌/瓣”。“伴”“扮”都是 7 笔，因为“扮”的第一笔是横，“伴”的第一笔是撇，所以再排成“办半扮伴拌瓣”。有的按偏旁(音符)组合，各组按第一字的笔画数排列组与组之间的顺序，同组内也按笔画数排列先后。比如《新华字典》(1998 年版)中读 lǚ 的 12 个字，按音符分成 5 组，排列顺序是“吕侶铝𨮒/捋/旅膂/倮屨缕楼/履”。

对多字条目，有两种排列法。一种是纯字母顺序法，不考虑汉字的字形，并只在拼音完全相同的时候才考虑声调。比如《汉语拼音词汇》(语文出版社，1991)是这样编排的：guǐbāguà 鬼八卦/guǐbài 跪拜/guǐbǎn 龟板/guǐbǎo 瑰宝/guǐbǎxi 鬼把戏/guǐbì 规避/guǐbǐ 鬼笔/guǐbiàn 诡辩/guǐbiànjiā 诡辩家/guǐbiànshù 诡辩术/guǐbiǎo 圭表/guǐbīn 贵宾/guǐbìng 归并。另

一种是音节·汉字·字母顺序法,先按照第一个字的音节排列,比如先排 guī 音节的字(圭/归/龟/瑰/规),再排 guǐ 音节的字(鬼/诡),再排 guì 音节的字(跪/贵)。第一个字同声同韵同调的汇集在一起,再按汉字的排序法排列,比如在 guī 音节那一组中,按笔画数先“归”再“圭”再“龟”再“规”,最后是“瑰”。第一个字字形相同的词条再按第二、第三个字的音序排列,比如第一个字都是“诡”的三个词就排成“诡辩、诡辩家、诡辩术”。后一种方法更适合一般读者使用。

音序查字法的局限是,首先必须知道字的读音,其次必须掌握注音工具,如《汉语拼音方案》。

### 6.3 部首查字法

“部首”是东汉许慎创造的。为了说明字的本义与字形、字音之间的关系,许慎把 9353 个字的意符进行分析和归类,得到 540 个小类,每一类字都有一个相同的意符,就把这个意符立为部首。许慎所编《说文解字》的部首是按照文字学的原则建立起来的。后来,编写工具书需要给成千上万个汉字编排次序,便在许慎 540 部的基础上,按照检字法的原则加以增删、改造,逐渐形成了部首查字法。

使用部首查字法,要分两个步骤:先查到一个字所属的部,再计算除去部首以外的笔画数,根据所得笔画数到那个部内相应的位置上去查字。

中华人民共和国成立后,对部首查字法的改进作过多次探索。《新华字典》建立的部首数目先后有126部、187部、191部,1966年以后的版本都是189部。对某些不易查检的字,在“部首检字表”中还采用了“多开门”的变通办法,即按照一个字可能分出的部首把这个字分别归入几个部中。《辞海》(1979年修订本)采用了六十年代汉字查字法整理工作组推荐的《部首查字法(草案)》,依据字形定部,一般采取字的上、下、左、右、外等部位作部首,其次是中坐和左上角,按照以上七个部位都无从确定部首的,另立单笔部首,共得250部。依形定部的方便处是摆脱义的束缚,只从字形着眼就能确定部首,缺点是部首数目太多,有的部首过分生僻。1983年,中国文字改革委员会和文化部出版局联合召开了统一汉字查字法座谈会,成立了统一部首查字法工作组,提出了《汉字统一部首表(草案)》。这个部首表依形定部,设201部。《汉字统一部首表(草案)》推出已经十多年,但由于没有制订具体的单字归部原则,至今没有达到预期的目标。八十年代后期先后推出的大型工具书《汉语大字典》和《汉语大词典》都采用部首法编排正文。立部时,以《康熙字典》的214部为基础,共立200部。由上可知,部首查字法直到目前为止,无论在立部还是归部上,还没能统一,尚待进一步探索研究。

## 6.4 笔画笔形查字法

这是一种按照笔画数目和笔画形状查字的方法。笔画少的在前,笔画多的在后。笔画数目相同的,再按笔画形状排列次序。常见的是把笔形分成五类,顺序是:横一(包括提㇀)、竖丨、撇ノ、点丶(包括捺㇏)、折乚(包括一丁乚等笔形)。先按第一笔笔形分类,第一笔相同的,再按第二笔,依此类推。“上”“三”“大”“与”“工”“万”都是3笔。后五个字的第一笔都是横,只有“上”的第一笔是竖,“上”排在那五个字的后面。再看那五个字的第二笔,“三”仍是横,应在前;“工”是竖,应排第二;“大”是撇,应排第三;“与”和“万”的第二笔都是折,分不出先后。再看这两个字的第三笔:“与”是横,在前;“万”是撇,在后。这样就排定了这六个字的次序:“三、工、大、与、万、上”。有的工具书,在第一、二笔笔形相同的情况下,根据字形结构排列先后:左右结构的在前,上下结构的在后,最后是整体结构的字。如“哭”“圆”“峪”都是10笔,第一笔都是竖,第二笔都是折。最后按字形结构排定的次序是“峪、哭、圆”。

## 6.5 四角号码查字法

这也是一种形序法。1925年王云五提出。这种查字法把汉字四个角上可能出现的笔画形状分为10类,并用0-9这10个数字代表:㇇(头)-0,一(横)-1,丨(垂)-2,丶(点)-3,十(叉)-4,丰(插)

-5,口(框)-6,丿(角)-7,八-8,小-9。为了便于记忆,有人编了“四角号码歌”:

横一垂二三点捺,叉四插五方框六,七角八八九  
是小,点下有横变零头

此外,还规定了取四个角的顺序:左上角、右上角、左下角、右下角,如:

端 0212      烙 9786      截 4325      跑 6711

为了区分同码字,又取右下角上方的笔形为第五个号码,通称“附号”,如:

芒 4471<sub>0</sub>      先 2421<sub>2</sub>      佳 2421<sub>4</sub>      仇 2421<sub>7</sub>

1964年汉字查字法工作组推荐的《四角号码查字法(草案)》有所改进,习惯上称为“新四角号码查字法”。

四角号码查字法的优点是由字形一次就可以转换成代码,按代码就可以查到需要的字。它不必像音序法那样要先知道字音,也不必像部首法那样要先找出部首,也不必像笔画笔形法那样要先数清笔画数并知道笔形和笔顺,但是,笔形与代码之间没有理据,取码的补充规定比较复杂,同码字也较多,如《现代汉语词典》(1983年版)的“四角号码检字表”中,号码是4422的就有59个字,加上“附号”7的也有47个字。

研究汉字的排序法、检字法,在中国已经有了相当长的一段历史,直到今天,研究更理想的汉字排序

法、检字法的还大有人在。

## 6.6 书刊由直排改横排

附带要说一说的是汉字进入交际过程以后的排列问题。长期以来,在书面上汉字是从上到下、从右到左排列的。据有人考证,这种格式的形成可能跟古代书写的工具(毛笔、竹木材料的纹理和竹简狭窄的形式)有关。19世纪末20世纪初,先有卢懋章的《一目了然初阶(中国切音新字厦腔)》(1892年)所附读物,后有严复《英文汉诂》(1904年)尝试实行横排,但其他印刷物仍沿用“右起直行”。直到1955年1月,《光明日报》首先实行横排,随后,《新华月报》《新观察》《中学生》《时事手册》等纷纷仿效,国家机关的公文和书籍也全面改用横排。在阿拉伯数字、数学公式和罗马拼音文字大量进入书面汉语的情况下,由直排改为横排带来了很方便,可以看作是书面语汉字排列的国际化。

目前,实现现代化的宏伟目标、整个民族进一步开发智力和培养人才的迫切要求、利用电脑进行语言文字信息处理的飞速发展和科教兴国的坚实步伐,把促进汉字的规范化、标准化提到了比以往任何时期都更为重要的地位。1986年全国语言文字工作会议把促进语言文字规范化、标准化作为我国新时期语言文字工作方针任务的主要内容。1997年

全国语言文字工作会议回顾了 12 年来的语言文字工作,分析了面临的机遇和挑战,提出了跨世纪的语言文字工作的指导思想和奋斗目标,这一切使我们清楚地看到现行汉字定量、定形、定音和定序的工作依然相对滞后。汉字的整理工作、汉字的规范化标准化工作都任重而道远。



## 主要参考书目

1. 国家语言文字工作委员会政策法规室编《国家语言文字政策法规汇编》，语文出版社，1996。
2. 高家莺、范可育、费锦昌《现代汉字学》，高等教育出版社，1993。
3. 苏培成《现代汉字学纲要》，北京大学出版社，1994。
4. 周有光《中国语文的时代演进》，清华大学出版社，1997。
5. 黎锦熙《国语运动史纲》，商务印书馆，1934。
6. 胡奇光《中国小学史》，上海人民出版社，1987。
7. 费锦昌主编《中国语文现代化百年记事》，语文出版社，1997。
8. 刘叶秋《中国字典史略》，中华书局，1983。