

百种语文

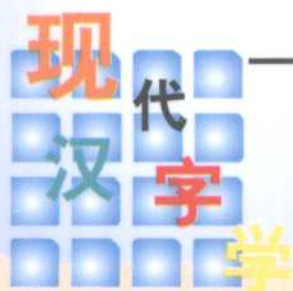
小丛书

曹先擢 主编

一门新学科： 现代汉字学

YI MEN XIN XUEKE XIANDAIHANZIXUE

苏培成 著



语文出版社

<http://www.ywcbs.com>

百种语文小丛书

YI MEN XIN XUEKE XIANDAI HANZIXUE

一门新学科：现代汉字学

苏培成 著

YUWEN CHUBANSHE

语文出版社

目 录

一	现代汉字和现代汉字学	(1)
	(一)现代汉字	(1)
	(二)现代汉字学	(3)
二	现代汉字的字量	(5)
	(一)汉字的总字数和现代汉字的总字数	(5)
	(二)汉字效用递减率	(8)
	(三)通用字的研究	(9)
	(四)中文信息处理中的字量问题	(10)
三	现代汉字的字形	(12)
	(一)现代汉字的字形标准	(12)
	(二)现代汉字字形分析	(13)
	(三)新字形和旧字形	(17)
	(四)正确使用简化字	(19)
	(五)异体字的规范	(21)
	(六)异形词的规范	(23)
四	现代汉字的字音	(25)
	(一)字音的确定	(25)

(二)同音字	(28)
(三)异读字	(31)
(四)多音字	(34)
五 现代汉字的字序	(41)
(一)字序标准化	(41)
(二)部首法	(42)
(三)音序法	(44)
(四)号码法	(47)
(五)笔画笔形法	(49)
(六)汉字输入的双轨制	(50)
六 现代汉字的构字法	(52)
(一)独体表意字	(53)
(二)会意字	(53)
(三)形声字	(55)
(四)半意符半记号字	(57)
(五)半音符半记号字	(58)
(六)独体记号字	(58)
(七)合体记号字	(60)
参考书目	(61)

一 现代汉字和现代汉字学

(一)现代汉字

汉字是汉族祖先独立创造出来的文字，至今已有三千多年的历史。在古代，用汉字书写的典籍浩如烟海，负载着灿烂的中国文化；在现时，它仍是记录汉语的正式文字，在社会生活的各个方面起着十分重要的作用。

当前，用汉字记录的汉语有两大类，一类是古代汉语，一类是现代汉语。它们的书面语，分别是文言文和白话文。不论是古代汉语还是现代汉语，在书写时用的都是楷书和行草，在印刷出版时用的都是宋体、仿宋体等字体，这是共同点；可是从另一方面说，由于古代汉语和现代汉语有许多不同，所以在用字上也有许多不同。例如，“佻圯窳筮𦰩”（这些字都见于《论语》），只用于古代汉语，不用于现代汉语；“啤叨掰榭𦰩𦰩𦰩𦰩𦰩𦰩𦰩𦰩”，都是近代现代产生的字，只用于现代汉语，不用于古代汉语。自然也有不少字是古代汉语和现代汉语都用的，例如“人半前兼舒僵建刃阶懈笔写”。其次，古代汉语是已经死去的语言，它用的字是个封闭的系统，字量不会增加，每个字负载的音义不会改

变；而现代汉语正在使用之中，它的用字也在不断地发展变化，字量有增减，每个字负载的音义有发展。从这个意义说，古代汉语用字是死字，现代汉语用字是活字。从学习和应用上说，现代汉语用字比古代汉语用字更重要，使用的范围更广。一个生活在现代的人，每日每时都离不开现代汉语用字，而古代汉语用字只在读古书时才要用到。根据以上的差别，站在现在的立场，我们把古代汉语专用字叫做文言古语用字，把古今通用字和现代汉语专用字叫做现代汉语用字，简称为现代汉字。古代汉语的书面语是文言文，现代汉语的书面语是白话文，因此文言古语用字也就是文言文专用字，现代汉字也就是白话文用字。文言古语用字和现代汉字不是根据汉字本身所处的时代先后而分出来的类，而是根据汉字所记录的语言的不同性质而分出来的类。

现代汉字还有另外一个意义，是从历史演变着眼指现代所用的字，和古代汉字、近代汉字相对而言的。根据汉字所处的不同历史时期，古代使用的汉字是古代汉字，如甲骨文、金文等；近代使用的汉字是近代汉字，如明清时期的汉字；现代使用的汉字是现代汉字。作为现代汉字学研究对象的现代汉字，指的现代汉语用字，也就是现代白话文用字，而不是现代使用的汉字。

(二) 现代汉字学

以汉字为研究对象的学科叫做汉字学，在我国有很长的历史。传统汉字学也叫历史汉字学，主要研究汉字形体的演变和古代汉字的造字规律，对现代汉字没有给予足够的重视；而人们在现实生活中遇到的汉字问题，绝大多数属于现代汉字，而不是文言古语用字。新中国建立后，国家对汉字进行了大规模的简化和整理，简化和整理的对象是现代汉字。进入70年代，中文信息处理技术崛起，给现代汉字的研究带来了新的活力。1980年周有光先生发表了《现代汉字学发凡》（载《语文现代化》丛刊第二辑），提出了建立现代汉字学的主张。他指出：“现代汉字学研究现代汉字的特性和问题，目的是为今天和明天的应用服务，也就是为四个现代化服务，减少汉字在现代生活中的不方便。‘现代汉字学’是个新名词、新事物。它播种于清末，萌芽于‘五四’，含苞于解放，嫩黄新绿渐见于今日。”他还指出：“现代汉字学研究的问题和研究方法跟历史汉字学很不相同。它是以语言学为基础而结合信息论、统计学、心理学等的边缘科学。这决不是抛弃或背叛历史汉字学。在汉字学的领域里应当厚今而不薄古、厚古而不薄今。”周先生的这些话为现代汉字学勾画出了轮廓，为现代汉字学的建

立打下了基础。周先生说的“现代汉字的特性和问题”，主要指的是现代汉字在数量、形体、读音、排序上的特性以及在这几个方面存在的问题。现代汉字学要研究并解决这些问题，最终的目的是要减少汉字学习和使用中的困难。

周先生的论文反映了汉字研究的新发展，具有鲜明的时代特征，得到了许多语言文字学家的赞同：汉字学的分支学科——现代汉字学就应运而生了。自那时到现在还不到 20 年，在这短短的时间内，这门学科得到了较快的发展。近年来出版的、有关现代汉字研究的通论性著作就有四种：（1）高家莺、范可育、费锦昌合著《现代汉字学》（高等教育出版社，1993），（2）苏培成著《现代汉字学纲要》（北京大学出版社，1994），（3）尹斌庸、John S. Rohsenow（罗圣豪）合著《现代汉字》（英文本，华语教学出版社，1994），（4）李禄兴《现代汉字学要略》（文津出版社，1998）。同时还发表了一批质量较高的研究论文。有的高校中文系在现代汉语课中加进去现代汉字的内容，有的高校开设了现代汉字的专题课或选修课。我们应该看到，现代汉字学毕竟年轻，研究队伍很弱小，研究成果还不多，还不能充分满足社会提出的需要，它迫切需要得到各方面的关心和扶植。

二 现代汉字的字量

(一) 汉字的总字数和现代汉字的总字数

自古至今汉字一共有多少字？这个问题很难回答。下面把几种重要字书的收字数做一介绍，从中可以看出个大概：

《说文解字》，东汉许慎著，公元 121 年成书，有 9353 字，还有重文 1163 字，共收 10516 字。

《玉篇》，南朝梁顾野王著，公元 543 年成书，收 16917 字。

《广韵》，宋陈彭年等著，公元 1008 年成书，收 26194 字。

《集韵》，宋丁度等著，公元 1039 年成书，收 53525 字。

《康熙字典》，清张玉书等著，公元 1716 年成书，收 47043 字。

《中华大字典》，民国陆费逵等著，1915 年成书，收 44908 字。

《汉语大字典》，徐中舒主编，1990 年出齐，收 54678 字。

对上述字数，要说明三点。第一，字书收字数并不能完全反映字书编写时社会实有的字数，一定

会有许多遗漏。甲骨、钟鼎、简帛、玺印、货布、碑碣等上面的字很难完全包括在内。第二，字书计算收字数时，异体字是作为独立的单字分别计算；韵书计算收字数时，多音字是按音项分别计算。如果去掉异体字、合并多音字，总字数就没有这么多。例如《集韵·二冬》“農”有 5 个异体字，《六脂》“𪗇”有 4 个异体字，《七之》“箕”有 8 个异体字。仅这 3 个汉字就有 17 个异体字。第三，字书收字数是历代积累起来的总字数，每个时期实际使用的字数没有这么多。裘锡圭指出：“甲骨文所用的单字约有四五千个。”“从商代后期到周末，一般使用的文字的数量，很可能一直在四五千左右徘徊。直到现在，据近年的统计，一般使用的汉字的数量也还是四五千的样子。”（《文字学概要》，商务印书馆 1988 年版第 31 页）

现代汉字的总字数有多少？早在 20 世纪 50 年代，周有光就提出了研究现代汉语用字全表的课题。他说：“汉字随时随地增加，‘出生不报’；又随时失去作用，‘死而不葬’。字典里死字和活字并立，不加区别。书写现代汉语究竟要用多少汉字？谁也说不出来。这种现象不应当长此任其自然。整理汉字的一项基本工作就是清点汉字的‘家底’，编出一份‘现代汉语用字全表’的清单来。”“有一天我们能够告诉人家现代汉语究竟要用多少个汉

字，把‘不可知数’变成‘可知数’，这将是汉字史上的一件大事！”（《汉字改革概论》，文字改革出版社 1961 年版第 312 ~ 313 页）他还说：“现在有没有一份‘现代汉字表’？这有两个答案：（1）一本今天有代表性的字典（例如《新华字典》1971）里的全部‘正字’就是一份现代汉字表。这是权宜的答案。（2）现代汉字表就是‘现代汉语用字全表’，尚待研究选定。”（《现代汉字学发凡》，《语文现代化》丛刊第二辑）周有光在《现代汉语用字的定量问题》里进一步讨论了现代汉语用字全表的问题。他说：“现代汉语的‘用字全表’跟‘常用字表’是不同的概念，必须明确区别开来。‘用字全表’是根据‘字性’来审定的。‘常用字表’是根据概率来选择的。”“区分古今是‘定量’的基本要求。”“‘定量’要定‘总量’，要得到一个‘用字全表’。什么是‘全’？能不能‘全’？要想一步达到百分之百是不可能的。可能的办法是步步逼近，‘逐步逼近’百分之百。‘现代汉语用字表’的草稿要在实践中检验、复查、补充、订正，最后达到‘最佳逼近’。‘全’就是‘最佳逼近’。”（《辞书研究》1984 年第 4 期）根据有关资料的分析，现代汉字总字数不超过一万字。

汉字字数繁多，而且字数无定，是造成汉字难学难用的重要原因。现代汉字字量研究首先要做到

字有定量，在此基础上才谈得上限制和减少字数。因此，研制现代汉语用字全表就具有重要的意义。

（二）汉字效用递减率

从存在的状态来考察，汉字可以分为两类，就是储备状态的字和使用状态的字。储备状态的字是静态的字，就是字典里的字；使用状态的字是动态的字，就是出现在文章里的字。字频统计的对象是使用状态的字。在字典里每个字构成一个条目，地位是相等的；文章里不同的字出现频率不同，有的甚至非常悬殊。根据《现代汉语频率词典》（北京语言学院出版社 1986 年版）中的《汉字频率表》，“的”出现 75306 次，频率为 4.16489%，而“乾、薔、薇”等 400 多字，每个字只出现一次，频率均为 0.00011%。按照降频排列，前 100 个字的累计频率是 47.33584%，前 2418 字的累计频率是 99.00023%。

由于每次字频统计所用的语料和所用的方法不尽相同，得出的结果也有差异。周有光对各家统计的结果进行分析归纳，抽象出汉字出现频度不平衡规律，也叫汉字效用递减率（《中国语文纵横谈》，人民教育出版社 1992 年版第 156 页）：

字种数	增加 字数	合计 字数	覆盖率	欠缺率
1000		1000	90.000%	10.000%
1000	+ 1400 =	2400	99.000%	1.000%
2400	+ 1400 =	3800	99.900%	0.100%
3800	+ 1400 =	5200	99.990%	0.010%
5200	+ 1400 =	6600	99.999%	0.001%

这个规律可以表述为：最高频 1000 字的覆盖率大约是 90%，欠缺率是 10%。以后每增加 1400 字，欠缺率减少 9/10，只有原来的 1/10。根据这个规律，我们可以知道：（1）汉字常用字非常集中。字频统计中前 2400 字，覆盖率高达 99%。（2）字频统计中序号在 6600 以后的字，不管有多少字，总的覆盖率不超过 0.001%。根据这个规律，还可以确定现代汉字的分级定量：最常用字约 2400 字左右，是小学识字量；常用字约 3800 字左右，是九年制义务教育识字量；通用字约 6600 字左右，是中学识字量。

（三）通用字的研究

现代汉字根据使用频度等参数，可以分为通用字和罕用字两大类。通用字是出版印刷用字，是当今社会一般要用到的字。研究通用字，要制定出现

代汉字通用字字表。本世纪 60 年代以来，正式公布的通用字字表有四个：（1）《印刷通用汉字字形表》（1965）收 6169 字。（2）GB2312 - 80《信息交换用汉字编码字符集·基本集》（1980）收 6763 字。（3）《标准电码本（修订本）》（1981）收 7080 字。（4）《现代汉语通用字表》（1988）收 7000 字，其中包括 3500 个常用字。在这四个字表中，《现代汉语通用字表》水平最高，社会影响最大，但是它也有些不足。这主要表现在两个方面。第一，字性审查不严。据研究，它收了 400 多个文言古语用字，如“嗑犴獐邶蹕狴𡩂”；还收了一些方言字，如“俺忒浜爿盆幽”；还收了部分罕用字，如“砵孳灵”。这些字都不属于现代汉语通用字的范围，应该去掉。把这些字去掉后，字表收字数还可以减少。第二，收字的规范性不够，收了一些异体字，例如“堃”和“垧”，“荳”和“蕒”，“碱”和“硷”。

（四）中文信息处理中的字量问题

中文电脑字库要收多少字，都是哪些字，这是现代汉字字量研究应该解决的问题。现行的中国制造的电脑字库，执行的是中华人民共和国国家标准《信息交换用汉字编码字符集·基本集》（GB2312 - 80），其中有汉字 6763 个，分为两级。第一级是常

用汉字 3755 个，汉语拼音音序排列；第二级是次常用汉字 3008 个，部首排列。代码表分成 94 个区，每个区有 94 位。区的编号从 1~94，由第一字节标志；位的编号也从 1~94，由第二字节标志。代码表中的任何一个图形字符的位置用它所在的区号与位号标志，区号与位号之间用连字符连接。例如“啊”的区位号是 16-01，“阿”的区位号是 16-02，“埃”的区位号是 16-03。经过多年的使用，人们发现《基本集》中大约有 1100 多字是不用或很少用的，而另外又有 950 多字经常需要使用可是又不在《基本集》之中。这说明《基本集》的收字需要修订。

为了满足用字量大的用户的需要，国家标准局于 1987 年 3 月 27 日发布了两个“辅助集”。一个是《信息交换用汉字编码字符集·第二辅助集》(GB7589-87)，收简化汉字 7237 个；另一个是《信息交换用汉字编码字符集·第四辅助集》(GB7590-87)，收简化汉字 7039 个。

电脑用字为限制和减少汉字的字数提供了机遇，因为在信息处理时，字库里有的字畅通无阻，字库里没有的字寸步难行。这些年，每当遇到电脑字库的字不敷应用的时候，人们就想到增加字库的字数，能不能换个思路，找到一条不增加字数而又能满足需要的新路。

三 现代汉字的字形

(一) 现代汉字的字形标准

字形指文字的形体，是文字存在的形式。任何一种实际使用中的文字，都要有明确的形体标准，也就是形体使用的规范。使用的人必须遵守这个标准，语文教学中要用这个标准来纠正种种误用。

在解放前的中国，汉字的字形规范化程度不够高，存在着一定程度的繁与乱。新中国建立后，针对着字形的繁与乱进行了简化和整理，形成了新的用字标准。新的字形标准是通过三个字表建立起来的。第一，1955年12月文化部和文字改革委员会发布的《第一批异体字整理表》。这个字表对810组异体字进行了整理，每组确定一个为标准字，其余的形体作为异体加以淘汰。这是有关异体字的国家标准。第二，1964年文字改革委员会公布的《简化字总表》。这个表1986年重新发表时，内容做了个别调整。《简化字总表》共收简化字2235个，每个简化字后面列出了相应的繁体字。这是有关简化字的国家标准。第三，1965年1月文化部和文字改革委员会发布的《印刷通用汉字字形表》。这个表收印刷铅字6196个，规定了每

个字的字形结构、笔画数和笔顺。这是有关印刷铅字字形的国家标准。综合上述三个字表的规定，就形成了现代汉字字形的标准。1988年3月国家语言文字工作委员会和新闻出版署发布的《现代汉语通用字表》，收通用字7000个。在字形上，它综合反映了上述三个字表的内容，完全符合上述字表的规定，因此它的字形就是现代汉字字形的标准。至于非通用字的字形，权威的语文工具书，如《新华字典》《现代汉语词典》等，依据上述三个字表的精神，提供了字形标准。

有关现代汉字字形的标准和写法方面的问题，还可以查阅费锦昌、黄佑源、张静贤合著的《汉字写法规规范字典》，上海辞书出版社1992年出版。

（二）现代汉字字形分析

汉字研究和汉字教学都要对字形进行分析。分析现代汉字的字形，要以标准字形为对象。现代汉字字形的构成，从小到大有三个层次，就是笔画、部件、整字。

1. 笔画是汉字构形的最小单位。写字时笔每次和纸从接触到离开，在纸上留下的痕迹就是一画。对于笔画，我们要注意以下几个问题：

（1）笔画数。一个汉字所包含的笔画的多少，就是这个字的笔画数。汉字中笔画最少的字只有1