

教育部人才培养模式改革和开放教育试点教材

现代汉字学纲要

(增订本)

苏培成 著



北京大学出版社

PEKING UNIVERSITY PRESS

现代汉字学纲要

(增订本)

苏培成 著

北京大学出版社

北 京

图书在版编目(CIP)数据

现代汉字学纲要·增订本/苏培成编著. —北京:北京大学出版社, 2001. 12.

ISBN 7-301-02597-1

I. 现… II. 苏… III. 汉字学-概论 IV. H12

书 名: 现代汉字学纲要(增订本)

著作责任者: 苏培成 著

责任编辑: 胡双宝

标准书号: ISBN 7-301-02597-1/H·0254

出版者: 北京大学出版社

地 址: 北京市海淀区中关村北京大学校内 100871

网 址: <http://cbs.pku.edu.cn>

电 话: 出版部 62752015 发行部 62754140 编辑部 62752028

电子信箱: zpup@pup.pku.edu.cn

排 印 者: 北京大学印刷厂

印 刷 者: 北京飞达印刷有限责任公司

发 行 者: 北京大学出版社

经 销 者: 新华书店

850毫米×1168毫米 32开本 12印张 300千字

2001年12月第2版 2003年5月第5次印刷

定 价: 18.00元

前 言

近十多年来，汉字学的一个新分支——现代汉字学悄然兴起，人们似乎突然认识到这是一块大有开发前途的处女地。现代汉字学的研究对象是现代汉字。查一查语言学文献，早在1952年8月号的《中国语文》上就有丁西林先生的《现代汉字及其改革的途径》，这篇文章中已经使用了“现代汉字”这个术语。可是在其后的很长时间里，现代汉字的研究并没有引起人们的注意。直到1980年，周有光先生在《语文现代化》丛刊第二辑上发表了《现代汉字学发凡》之后，随着改革开放大潮的迅速发展和汉字信息处理的迅速发展，现代汉字学才找到了适宜的土壤，正如周先生所说的那样“嫩黄新绿渐见于今日”。1991年8月，在北京举行了“海峡两岸汉字学术交流会”，我提交给会议的论文是《现代汉字研究简述》，意在向与会的海峡两岸的学者介绍现代汉字学这个新生儿。这几年，研究现代汉字的论文多起来了，大学用的“现代汉字学”专业教材也出版了，真是令人高兴。不久前我读到台湾大学黄沛荣教授的新著《汉字的整理与统合》，在这本研究报告中三次出现了“现代汉字”这个术语。黄教授参加了1991年的“海峡两岸汉字学术交流会”和1992年在北京举行的“海峡两岸汉字学术座谈会”。这说明现代汉字的研究已经引起了他的注意。

什么是现代汉字？有人理解为目前正在使用的这套汉字，也有人理解为记录现代汉语用到的字。这两种解释不完全相同，但是都表明它和我们的现实文字生活密不可分。随便想一想就可以提出一连串的问题：出版印刷要用多少字，都是哪些字？九年义务教育要教多少字，都是哪些字？汉字的笔画有多有少，哪几种笔画

负载的汉字数量最多,为什么?汉字在学习和书写时,要对字形做出分析,应该怎么分析?多数字是由几个小的构字单位组成的,组合时有没有理据可言?要把成千上万个汉字排成一个系列,根据什么原则,是字音、是字形,还是什么别的?用什么办法可以比较方便地把汉字输入电脑?如何评价汉字简化的利弊得失?如何使汉字变得更加合理更加规范,使学习和使用都比较容易?汉字的前途是什么,要不要改为拼音文字?等等,等等。现代汉字学要对这些问题做出科学的回答。只要现实生活中这些问题还存在,现代汉字学就有存在的价值,就有用武之地。现代汉字学的研究已经做了许多工作,可是还很不够,因为它太年轻,不成熟。和传统汉字学相比,它的底子很薄,基本理论和基本方法还不很完备,需要有更多的学者关心它扶植它。相信它一定有远大的前途。

我于1988年秋开始为北京大学中文系汉语专业讲授“现代汉字学”这门课程,其后又讲了几次,每讲一次都要对讲稿做些补充和调整。1993年下半年,用了几个月的时间把历来所用的讲稿做了全面的补充和改写,就成为现在这本教材。

我在讲课和写作这本教材时,参考了许多前辈的著作,并从中引用了一些观点和材料,在此向这些先生表示深深的感谢。我在讲课和写作这本教材时,还得到许多师友的关注和帮助,在此向他们致以诚挚的敬意。

这本书讨论现代汉字,涉及已经停用的繁体字、异体字和旧字形以及古字,为排版增加了不少麻烦。这里特向北京大学印刷厂激光照排车间的同志表示感谢。

我自己水平不高,对现代汉字的研究刚刚起步。书中定有不妥之处,敬请读者批评指正。书末所附的参考文献疏漏也很多,敬希各位专家见谅。

作者

1993年岁末于北大

增订本说明

《现代汉字学纲要》这本小书1994年10月出版。出版后的几年中又印行了三次，共印行了14000多册。在同类的书里面，这个印数是比较多的。这说明人们对它有兴趣，社会对它有需求。不久前我收到一位德国学者寄来的信，这位学者在信中说：“《现代汉字学纲要》对我来说，是一本很有意思的书，里面有新的思想和大量的信息，它是我所读到过的中文书籍当中少数有作者自己的新思维和新内容的好书之一。”感谢读者的鼓励，我愿意继续努力去开拓这块新领域。这次有机会做些增订，出版这个增订本，我感到很高兴。这次增订，一方面是改正初印本里的讹误，另一方面是增补近年来现代汉字研究取得的新成果，充实书的内容。

促成这次增订有一个原因，是中央广播电视大学与北京大学中文系合作，在中央电大开设中文专业，进行专科升入本科层次的学历教育。根据教学计划，要开设现代汉字学；北大中文系指定我来承担这门课的教学工作，教材就采用我编写的《现代汉字学纲要》。电大的教学和北大的教学有一些差别。为了适应电大教学的需要，教材也必须做些调整。在内容方面，我把讲解字形的部分由原来的一章改为三章，使各章的篇幅大体平衡，便于教学。此外我还去掉了字义部分，以便突出汉字的“四定”。在每章的最后，我增加了“思考与练习”和“参考文献”（论文部分）。练习有巩固性的，更多的是思考性的，希望能促使学员进行思考，对教学内容提出自己的看法。为了使学员做练习比较方便，本书增加了附录，附录提供了政府主管部门公布的几个字表。根据我多年来从事教学的经验，要学好大学的课程，只读教材是不够的，一定要阅读参考

文献。考虑到学员的时间和精力有限,这里列出的是最低限度的参考文献。电大的学员分布在全国各地,有些地方可能找不到参考文献。现已取得北京大学出版社的支持,他们同意把本书开列的参考文献汇编成册出版,书名定为《现代汉字学参考资料》,供电大学员和其他读者选购。

增订本一定会有讹误有疏漏,敬请读者批评指正。

作者

2001年9月10日

重印附言

利用重印的机会,我把全书校读了一遍。除了改正排印的错字外,在71页、136页、185页和234页都略有调整,增加了新的资料。

作者

2002年4月1日

目 录

第一章 汉字概说	1
第一节 汉字的性质.....	1
第二节 汉字的特点.....	5
第三节 汉字的总字数.....	9
第四节 汉字和汉文化.....	11
第五节 汉字文化圈的过去和现在.....	15
第二章 现代汉字和现代汉字学	21
第一节 现代汉字.....	21
第二节 现代汉字学.....	24
第三章 现代汉字的字频统计与分析	32
第一节 综合性的字频统计.....	32
第二节 字频统计的应用.....	38
第四章 现代汉字的字量	43
第一节 现代汉字的总字数.....	43
第二节 通用字和常用字.....	48
第三节 专门用字.....	52
第四节 专业用字.....	55
第五节 限制和减少汉字的字数.....	56
第五章 现代汉字的构形法	64
第一节 笔画.....	65
第二节 部件.....	74
第三节 整字.....	86
第六章 现代汉字的构字法	92

第一节	字符和字符的分类	92
第二节	现代汉字的新六书	93
第三节	现代汉字的理据性	102
第七章	现代汉字的简化和整理	107
第一节	现代汉字的简化	107
第二节	现代汉字的整理	125
第八章	现代汉字的字音	141
第一节	现代汉字的读音标准	141
第二节	异读字	146
第三节	多音多义字	152
第四节	同音字	160
第九章	现代汉字的字序	166
第一节	中国古代的字序法	166
第二节	中国现代的字序法	172
第十章	熵和多余度	187
第一节	信息量	187
第二节	汉字的熵	189
第三节	汉语书面语的多余度	192
第十一章	汉字和中文信息处理	196
第一节	中文信息处理	196
第二节	汉字的交换码和内部码	199
第三节	汉字的输入	203
第四节	汉字的输出	212
第十二章	现代汉字的规范化	216
第一节	现代汉字在人际界面的应用	216
第二节	现代汉字的正字法	218
第三节	现代汉字的“四定”	225
第十三章	海峡两岸的书同文	230

第一节	台湾地区汉字使用的情况·····	230
第二节	海峡两岸汉字异同的研究·····	235
第三节	海峡两岸书同文的展望·····	238
第十四章	汉字的评价与前途·····	246
第一节	汉字的评价·····	246
第二节	新中国的汉字政策·····	259
第三节	汉字的前途·····	266
参考文献 ·····		278
附录		
	现代汉语通用字表·····	280
	关于重新发表《简化字总表》的说明·····	313
	简化字总表·····	314
	第一批异体字整理表·····	337
	普通话异读词审音表·····	348

第一章 汉字概说

第一节 汉字的性质

语言是人类最重要的交际工具。人类自身在漫长的形成过程中,同时也逐渐地形成了语言,语言的历史和人类的历史一样的漫长和古老。只有到了距离现在大约一万年左右,才开始出现原始文字。文字的历史和语言的历史相比,要短得很多。作为交际工具,语言存在于口耳之间,稍纵即逝,受到时间和空间的限制。借助于文字的帮助,语言可以传给后世传到远方,克服了时空限制,扩大了交际的范围。在现今的世界上,有的民族有语言可是没有文字;相反的情况,就是有文字而没有语言的民族,却从来没有出现过。所以,先有语言,后有文字;语言是第一性的,文字是第二性的;文字是记录语言的符号,是最重要的辅助性的交际工具。语言文字学的这些基本观点,我们必须坚持,如果背离了这些观点,就谈不上文字研究。

某种具体文字的性质,比如说英文的性质、日文的性质或汉字的性质,指的是这种文字区别于其他文字的本质特征。这是文字类型学要研究的问题。正确地认识一种文字的性质,是正确地解决有关这种文字的种种问题的基础。文字是用来记录语言的,所以要联系语言来研究文字。要确定某种具体文字的性质,就是要看这种文字的基本单位记录的是什么样的语言单位。文字的基本单位是指文字体系中能和语言成分相对应的最小的单位。英文的基本单位是字母,日文的基本单位是假名,汉字的基本单位是单字。下面再说语言单位。语言从它的内部结构说是个层级体系,

分为上下两层。下层是语音层,包括音素和音节两级。这一层为语言符号准备了形式的部分。“语言的上层是音义结合的符号和符号的序列,这一层又分为若干级。第一级是语素,意义在这里被装进形式的口袋,成了音义结合的最小的符号。第二级是由语素构成的词,第三级是由词构成的句子。词和句子都是符号的序列。”^① 世界上的文字有许多种,从文字的基本单位记录的语言单位来看,主要有三种,就是音素文字、音节文字和语素文字。英文的字母记录的是英语的音素(或者说是音位),所以英文是音素文字(或者说是音位文字);日文的假名记录的是日语的音节,所以日文是音节文字;汉字的单字记录的是汉语的语素,所以汉字是语素文字。

音素文字有许多种,如国外的英文和俄文,国内的藏文和壮文。音素文字使用的字母有许多种,如英文用拉丁字母,俄文用斯拉夫字母,阿拉伯文用阿拉伯字母,藏文用梵文的天城体字母,壮文用拉丁字母。现在使用得最广的字母是拉丁字母,有 100 多个国家都使用它。一种语言的音素数量是有限的,不过几十个,音素文字的字母数量因而也是有限的,也只有几十个。代表音素的字母按照一定的规则拼合起来,就能拼出它所记录的语言的声音。你如果懂得这种语言,阅读这种文字时,从字母拼合表示出来的声音就知道它表示的是什么语言成分,也就知道这些语言成分所表示的是什么意义。有些历史悠久的音素文字,由字母拼合所表示出来的语音和现今实际语言有差别。例如英文的 night(夜晚),其中 gh 现在不发音。这是因为文字一经形成之后有稳固性,变化缓慢,而活的语言在发展,两者的变化不同步,于是造成了脱节。不过就音素文字来说,这种情况是局部的,并不改变这种文字的本质。

音节文字的代表是日文的假名。假名也是一种字母,一个假

^① 叶蜚声、徐通锵著《语言学纲要》,北京大学出版社 1981 年版第 30 页。

名代表日语里的一个音节。例如か、き、く、け、こ五个假名代表日语里的 ka、ki、ku、ke、ko 五个音节。语词是由音节构成的,用假名来记录音节,就可以表示出相关的语词。例如日语的“山”读 ya-ma,有两个音节,就用やま两个假名来表示。や读 ya,ま读 ma。日语的音节较少,音节结构较简单,比较适合使用音节字母。

语素(morpheme)也叫词素,是最小的语音语义结合体。例如汉语的“伟大”是由“伟”和“大”两个语素构成的。每个语素都有音和义:“伟”的音是 wēi,意思是大;“大”的音是 dà,意思是跟“小”相对。汉字的一个个的单字记录的是汉语里的一个个语素,因此汉字是语素文字。古代两河流域的丁头字、古代埃及的圣书字和形成于公元前不久的玛雅字,都是语素文字。但是这三种文字早已不用,自古至今一直在使用的语素文字只有汉字。汉语语素有同音的,也有不同音的。不同音的语素一般用不同的汉字来表示。例如汉语的“中”和“国”是两个不同音的语素,书写时就用“中”和“国”这两个不同的汉字。而同音语素也常常用不同的汉字来表示。例如“衣、医、揖”这三个语素都读 yī,而书写时用三个不同的汉字,“衣”是衣服的“衣”,“医”是医生的“医”,“揖”是作揖的“揖”。

每个汉字除了有形体和读音外,还有意义。这和英文字母、日文假名很不相同。汉字的音和义来自它所记录的语素的音和义。例如我们用“人”这个字表示汉语里“人”这个语素,因为语素“人”读 rén,意思是能制造工具并能使用工具进行劳动的高等动物,于是汉字“人”就读 rén,意思就是能制造工具并能使用工具进行劳动的高等动物。汉字的形和它所表示的语素的音和义紧密结合在一起,形成了一个形音义的统一体。汉语语素大多数是单音的,但也有一些是多音的。多音语素有的是联绵词和重叠词,例如“踟蹰、徘徊、翩翩、潺潺”;有的是译音词,例如“咖啡、柠檬、巧克力、奥林匹克”。记录多音语素的每个汉字,例如“踟”“蹰”“咖”“啡”,只有形体和读音,而没有意义(“踟蹰”“咖啡”才有意义),因而不是形音

义的统一体。不过这样的字数量不多,不反映汉字的本质,不影响汉字是语素文字的论断。

汉字是语素文字,这是许多学者经过多年的探索得出来的结论。赵元任把汉字叫做词素文字。他说:“在世界上通行的能写全部语言的文字当中,所用的单位最大的文字,不是写句、写短语的,是拿文字一个单位,写一个词素,例如我们单独写一个‘毒’的字形,来写‘毒’这个词素。”“用一个文字单位写一个词素,中国文字是一个典型的最重要的例子。”“它跟世界多数其他文字的不同,不是标义标音的不同,乃是所标的语言单位的尺寸不同。”赵元任说的“所标的语言单位的尺寸不同”,也就是所记录的语言单位的大小不同。赵元任说的“词素”就是我们说的“语素”。“词素”和“语素”都是英语 morpheme 的意译,所以赵元任说汉字是“词素文字”,和我们说汉字是“语素文字”,意思是一样的。^①

吕叔湘说:“世界上的文字,它的形式是多种多样的,但是按照一定的原则来分类,也就是按照文字代表语言的方式来分类,可以分成三类。一类是音素文字,一个字母代表一个音素(又叫做音位)。英语、法语等等所用的拉丁字母(罗马字母),俄语、保加利亚语所用的斯拉夫字母,都是音素文字。第二类是音节文字,一个字母代表一个音节,就是辅音和元音的结合体。日语的字母(假名)、阿拉伯语的字母,都属于这一类。音素文字和音节文字都是拼音文字,拼音文字的字母原则上都是没有意义的,有意义是偶然的例外。第三类文字是语素文字,它的单位是字,不是字母,字是有意义的。汉字是这种文字的代表,也是唯一的代表。汉字以外的文字都只是形和音的结合,只有汉字是形、音、义三结合。”^②

① 赵元任著《语言问题》,商务印书馆1980年版第142、144页。

② 吕叔湘著《汉语文的特点和当前的语文问题》,载《语文近著》,上海教育出版社1987年版第142页。阿拉伯字母一般认为是音素文字。

朱德熙说：“文字是记录语言的。就汉字跟它所要录的对象汉语之间的关系来看，汉字代表的是汉语里的语素。例如[tàn kāu]的[kāu]、[kāu iàu]的[kāu]和[t' iàu kāu]的[kāu]读音相同，意思不一样，是3个不同的语素，分别由3个不同的汉字‘糕、膏、高’（蛋糕、膏药、跳高）来表示。从这个角度看，汉字可以说是一种语素文字。”^①

文字是记录语言的符号，是最重要的辅助性的交际工具，这是文字的性质；汉字的单字记录的是汉语的语素，所以汉字是语素文字，这是汉字的性质。

第二节 汉字的特点

汉字的特点指的是汉字与其他文字的不同点，主要是与拼音文字的不同点。文字的性质和它的特点有密切的关系，文字的性质是根本，特点是由性质派生出来的。汉字作为语素文字，有以下几个特点。

第一，汉字和汉语基本适应。汉字的基本单位是一个个的字，这些字都是单音节的。和汉字相对应的汉语单位是语素，汉语语素以单音节为主。用单音节的汉字来记录单音节的汉语语素，两者基本适应。这是汉字的第一个特点。印欧语有形态变化，而汉语缺少形态变化。印欧语的形态变化主要用音素来表示，反映到文字上就是要用字母来表示。例如，英语名词的复数一般是在单数的后面加-s，如book→books, pupil→pupils。这种变化用字母来表示十分方便，如果用单音节的汉字来表示就非常困难。汉字和汉语基本适应，这是汉字富有生命力的根本原因。汉字延续使用了几千年，至今仍然充满了活力，能够为汉民族很好地服务。相

^① 朱德熙著《汉语》，载《朱德熙文集》第三册，商务印书馆1999年版第211页。

反,如果汉字和汉语基本不适应,使用非常困难,汉字就不会使用到现在。当然,汉字和汉语之间的配合只能说是基本适应,而不能说是完全适应,两者之间也有不适应的一面。语言的组成单位是词而不是字。现代汉语词汇以双音节和多音节的为主,而单音节的汉字不能够从形式上把词的界限显示出来:词和词之间既不分写,双音词和多音词的内部也不连写。这就会给写作和阅读带来麻烦。

第二,汉字是形音义的统一体。汉字的字不但有形、有音,而且还有义,这是汉字的另一个特点。正因为汉字有义,“衣”和“医”虽然同音但是意义不同。脱离开上下文,我们仍然知道“衣”是衣服的“衣”,“医”是医生的“医”,这给阅读带来方便。如果写成了汉语拼音,只有形和音,而没有义。脱离了上下文,单独的一个 yī,就不知道是什么意思了。汉字的字有意义,因此它的形体负载着较多的文化因素,这也是与拼音文字不同的地方。汉字作为语素文字,在每个时期使用的字数多达几千个,而且每个字都有义(没有义的是极少数),必然增加学习的负担。

第三,汉字有较强的超时空性。现代的英国人读 600 百年前诗人乔叟(Geoffrey Chaucer, 1340? —1400)写的诗比较困难,感觉像外国的作品,有许多地方难于理解。而汉人读两千多年前的《论语》《孟子》反倒比较容易。“学而时习之,不亦说(悦)乎?”“叟,不远千里而来,亦将有以利吾国乎?”稍加指点就容易读懂。这是因为汉语语素意义的变化比语音的变化慢,也就是字义的变化比字音的变化慢。这是汉字的超时代性。北京人、上海人、福州人、广州人,要是不说普通话,彼此又不懂对方的方言,就很难用语言沟通,可是书信往来并无困难,可以阅读同一份报纸。这是因为方言间语素意义的差别小而语音的差别大,也就是说他们所用的字意义大致相同,只是读音不同。这是汉字的超方言性。汉族文化悠久,古籍很多,需要阅读古籍的时候比较多,同时方言差别也比

较大。在这种条件下,汉字的超时空性又给学习和使用带来了方便。不过我们还要知道,汉字的这种超时空性的形成,是因为汉字写的是汉语共同语,而不是方言,这种共同语是学过汉字的人大体都了解的。相反,如果用汉字写方言,那么对不懂那个方言的人来说,照旧读不懂汉字写的文章。

第四,汉字字数繁多,结构复杂,缺少完备的表音系统。在共时的平面上,一种语言的语素有几千个,如果语素和文字大体上是一个对一个,那么语素文字的字数也要有几千个。汉字有繁体和异体,数量就会更多。如果计算历代积累的总字数,就要有几万个。现代汉语的通用字就有 7000 字,《中华字海》所收的古今汉字竟多达 86000 多字(其中有些是不该收的)。而拼音文字的字母一般只有几十个,比汉字少得多。数量繁多给学习和使用都增加了困难,对机械处理和信息处理也十分不利。汉字字数繁多带来许多困难。从形体说,文字都是由线条构成的。汉字的线条构成笔画,笔画组合为部件,部件再组合为单字。笔画和部件的种类都很多,笔画的组合方式和部件的组合方式也很多。而拼音文字的字母不但数量少,而且结构简单。所以我们说汉字的结构复杂。除此之外,作为文字最好能够做到见字知音,而汉字恰恰缺少完备的表音系统,做不到这一点。形声字的声旁应该是能表音的,可是事实上并不如此。用“者”作声旁构成的字,如“奢诸猪煮渚著箸翥楮褚暑署都睹堵屠绪”都不读“者”。据统计,《新华字典》收 8075 字(不算繁体和异体),其中有可用作声旁的字 1348 字。在这 1348 字中,能够准确表音的只有 386 字,占声旁字数的 29%,不到 1/3。这 386 字只占字典收字数 8075 字的 5%。例如“惶惶惶隍徨徨”都读声旁“皇”。可见现代汉字中声旁的表音功能很低。

第五,汉字用于机械处理和信息处理比较困难。世界进入现代化时期后,文字工作实现了机械处理和信息处理,这对科技、教育和经济的发展至关重要。使用拼音文字的国家早在 100 年前就