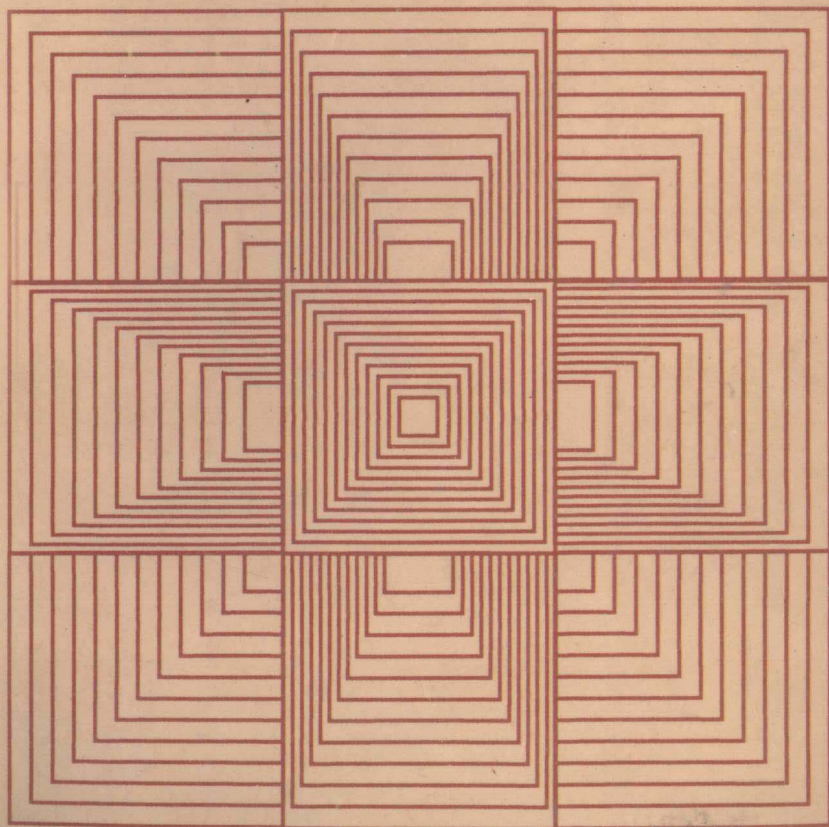


汉字认知心理与汉字编码

何克抗 宋继华 编著



汉字认知心理与汉字编码

何克抗 宋继华 著

北京师范大学出版社

责任编辑 倪 花

封面设计 孙 琳

图书在版编目 (CIP) 数据

汉字认知心理与汉字编码/何克抗, 宋继华编著, -北京: 北京师范大学出版社, 1996. 4

ISBN 7-303-04152-4

I. 汉… II. ①何…②宋… III. 汉字信息处理-编码-文字识别-基本知识 IV. TP391

中国版本图书馆 CIP 数据核字 (96) 第 07123 号

北京师范大学出版社出版发行

(100875 北京新街口外大街 19 号)

冶金部保定华泰公司激光排版

北京师范大学印刷厂印刷 全国新华书店经销

开本: 850×1168 1/32 印张: 15.625 字数: 390 千
1996 年 7 月北京第 1 版 1996 年 7 月北京第 1 次印刷

印数: 1—1000 册

定价: 21.00 元



何克抗, 1937 年 8 月出生于广东省大埔县, 1963 年于北京师范大学无线电电子学专业研究生毕业。1986 年 6 月晋升教授, 1993 年被国务院学位委员会评为教育技术学博士生导师。现为北京师范大学无线电电子学系主任, 现代教育技术研究所所长。曾先后五次获国家教委和北京市科技进步奖, 一次获北京国际发明展览金奖。1992 年被国家人事部授予有突出贡献专家称号, 1994 年入选英国剑桥世界名人录。

前 言

众所周知，计算机在我国的推广应用要比西方发达国家落后许多年，除了经济基础和科技发展条件的制约以外，一个很重要的原因就是汉字输入计算机要比拉丁文字困难得多。十多年来，经过海内外炎黄子孙的共同努力，已经研制出的汉字输入编码方案有 600 多种，已在计算机系统上实现的方案也有 100 多个，其中有些还得到较大范围的应用。但是由于汉字是大字符集（一二级常用汉字就在 6000 个以上），再好的编码方案也需要对几千个汉字进行编码，因此，总是要有一段专门学习或培训的时间，与只有几十个字母的拉丁文字相比，学习和使用的难易程度是不可同日而语的。这就是为什么西方国家的计算机早已进入中小学校和家庭，成为教师编写讲义和家庭中书信往来的文字处理工具，而我国计算机在学校仍主要是用于教学、科研，在家庭则主要用作游戏机、学习机的根本原因。显然，这种状况将严重影响计算机在我国的推广和普及。

有的专家预言：“如果有更方便友好的接口，使计算机更容易掌握，则可以使计算机达到像手表那样人手一个的普及水平”。随着“膝上型”、“笔记本型”等袖珍微机和单片巨型机的出现，从世界范围看，计算机像手表那样普及到每一个人已不再是天方夜谭。在西方一些发达国家这已是即将到来的现实。反观我们国内，要达到这样的计算机普及应用水平，从目前看，还是一个不切实际的幻想。原因何在？是缺乏硬件设备吗？是国民文化素质低下吗？我们认为，这些都是有关原因，但是最根本的原因在于，在中国要把母语输入计

计算机是要经过专门学习或职业培训才能获得的专门技能；而在拉丁语系国家，把自己母语输入计算机则是任何一位受过中小学基础教育的国民都能具有的基本技能。很显然，要从根本上解决这个问题，使我国计算机普及率真正赶上西方发达国家水平，出路只有一条——把计算机汉字输入问题与我国中小学的基础教育，特别是语文教育紧密结合起来。让中小学的语文教育和计算机教育、编码学习融为一体，做到识字、查字、编码和打字四者相结合，使学生经过小学和中学阶段的学习，既获得基础教育所要求的基本知识，又获得将汉字输入计算机所需要的编码方法与打字技能。由于这种方法与技能的教育是与中小学的语文教育紧密结合的（如音码可以与汉语拼音教学相结合；形码则可以与汉字的笔画笔顺、部件、间架结构等教学相结合，有条件的地方最好从小学一年级就开始这样的结合），因此可以成为学生认知结构中基本知识 with 基本技能的有机组成部分。这种通过基础教育所获得的基本知识与基本技能是特别牢固，难以遗忘的。这样掌握的汉字输入方法将会和我们书写汉字一样运用自如。但是，在此之前我们还未看到有任何一种汉字输入方案是完全按照这样一种社会需求（即与中小学基础教育紧密结合的社会需求），是严格遵循中小学语文教学的规律，并且是在中小学做过认真的试验后研制出来的。这就是为什么国家教委要将“全国中小学教学用汉字编码规范与计算机汉字输入系统研究”项目列入“八五”重点攻关项目的理由。本书正是在总结此攻关项目研究成果的基础上写成的。

邓小平同志关于“电脑要从娃娃抓起”的指示是高瞻远瞩、富有战略眼光的。开展“全国中小学教学用汉字编码规范与计算机汉字输入系统研究”正是全面贯彻落实这一重要指示的根本性措施之一。青少年是我们民族的希望，祖国的未来；中小学是中文信息处理大普及的基本阵地。研究出适合于中小学教学用的汉字编码规范及计算机汉字输入系统，对我国广大劳动后备军素质的提高以及计

算机在我国的推广应用将产生极为重大而深远的影响，其意义是难以估量的。

本书的第一、二、三、六章以及附录一、二由何克抗执笔；第四、五章以及附录三、四主要由宋继华执笔。其中，第七章利用了於国荣、杨开成、邢文华等同志在开发认知码汉字输入系统过程中所取得的成果，在此一并致谢！

由于作者经验不足，书中的缺点错误在所难免，欢迎广大读者批评指正。

作 者

1995 年 10 月

目 录

前言	(1)
第一章 汉字编码的性能指标分析	(1)
§ 1.1 汉字编码的性能指标体系	(1)
§ 1.2 如何做到符合规范性	(4)
§ 1.3 如何实现易学性	(11)
§ 1.4 如何提高快速性	(14)
§ 1.5 如何达到兼容性	(15)
§ 1.6 如何保持编码规则的一致性	(15)
§ 1.7 如何争取编码方案的完备性	(16)
第二章 汉字认知心理与认知模型	(17)
§ 2.1 汉字编码的主要矛盾	(17)
§ 2.2 汉语阅读的认知模型与汉字识别的认知模型	(19)
1. 汉语阅读过程的认知模型	(20)
2. 汉字识别过程的认知模型	(22)
§ 2.3 汉字输入与认知模型的选择	(24)
第三章 基于认知心理的形码方案设计	(25)
§ 3.1 部件的归并与选取	(26)
§ 3.2 部件分类系统的建立	(29)
1. 联想记忆的设计考虑	(29)

2. 均衡分配的设计考虑	(31)
§ 3.3 汉字拆分规则的制订	(33)
1. 认知心理学的考虑	(34)
2. 小学识字教学规律的考虑	(36)
§ 3.4 汉字编码规则的制订	(37)
1. 认知心理学的考虑	(38)
2. 扩充输入软件功能的考虑	(41)
第四章 汉字编码的计算机辅助设计及其评测	(42)
§ 4.1 汉字编码辅助设计的基本概念	(42)
1. 汉字编码辅助设计的定义与分类	(42)
2. 汉字编码辅助设计的现状与目标	(43)
3. 汉字编码辅助设计目标的评价	(45)
§ 4.2 汉字编码方案的评测	(46)
1. 汉字编码评测现状	(46)
2. 汉字编码方案的评测指标体系	(47)
§ 4.3 汉字编码规范性指标的评测	(49)
1. 规范性的目的和意义	(49)
2. 汉字编码规范性的理论依据	(51)
3. 规范性指标的评测	(53)
§ 4.4 汉字编码易学性指标的评测	(55)
1. 汉字编码易学性的理论依据	(56)
2. 易学性指标的评测	(58)
§ 4.5 汉字编码快速性指标的评测	(59)
1. 汉字编码快速性的正确含义	(60)
2. 快速性指标的评测	(61)
第五章 认知码的计算机辅助设计及其优化	(63)
§ 5.1 认知码机助设计系统的逻辑框图	(63)
§ 5.2 汉字间架结构及相应部件数的统计	(69)

§ 5.3	认知码所用部件的统计分析·····	(70)
§ 5.4	部件与部件组合频率对编码方案性能指标的 影响·····	(71)
§ 5.5	汉字大字符集认知码性能指标的分析与优化·····	(72)
第六章	认知码的编码方法 ·····	(77)
§ 6.1	基本概念·····	(77)
§ 6.2	认知码的汉字拆分规则·····	(79)
§ 6.3	认知码的汉字编码规则·····	(82)
§ 6.4	认知码的词语编码规则·····	(89)
§ 6.5	认知码的不成字部首编码规则·····	(90)
第七章	认知码的上机实践 ·····	(92)
§ 7.1	系统简介·····	(92)
1.	系统的安装 ·····	(92)
2.	系统的加载和卸载 ·····	(93)
§ 7.2	认知码汉字输入系统的辅助功能·····	(94)
1.	功能键定义 ·····	(94)
2.	模糊键 V 的功能 ·····	(94)
3.	拼音输入查认知码 功能 shift+F2 ·····	(95)
4.	国标区位输入查认知码 shift+F1 ·····	(95)
5.	快速输入英文字符 ·····	(95)
6.	设置认知码词组处理策略 ·····	(95)
7.	快速输入制表符 ·····	(96)
8.	用户自定义词组 ·····	(96)
9.	用热键移去本外挂系统 ·····	(98)
10.	系统信息 ·····	(98)
§ 7.3	键位动态定义技术·····	(98)
§ 7.4	认知码的编码练习字 ·····	(102)
1.	独体字编码实例 ·····	(102)

2. 准独体字编码实例	(104)
3. 合体字编码实例	(105)
附录一 认知码的基本部件表	(109)
附录二 认知码基本部件简表 (小学低年级适用)	(112)
附录三 认知码所用部件统计分析表	(115)
附录四 认知码所需汉字属性表	(125)
参考文献	(489)

第一章 汉字编码的性能指标分析

§ 1.1 汉字编码的性能指标体系

通过社会需求分析、文献资料调查研究和中小学的现场试验，我们认为一个优秀的、适合中小学教学用的汉字编码方案及相应的计算机汉字输入系统应当满足下列性能指标的要求：

1. 规范性：音码应符合国家颁布的《汉语拼音方案》和国家教委制定的《全日制小学语文教学大纲》中有关拼音教学的要求；形码应符合汉字结构规律，符合国家语委颁布的语言文字规范，符合国家教委制定的《全日制小学语文教学大纲》中有关识字教学的要求；不论音码还是形码，在计算机系统中实现时，均应遵循国家技术监督局（前国家标准总局）发布的《信息交换用汉字编码字符集》的有关规定。

2. 易学性：在中小学使用的汉字编码及相应的汉字输入系统应易学、易记、易用，能促进中小学的拼音识字教学，提高学习效率，而不能加重学生的课业负担，不能像培养专业操作员那样来要求中小學生。因此，规范编码的规则应尽可能简明扼要，记忆量应减至最小，并有方便友好的人机界面。

3. 快速性：面向中小学教学用的汉字编码及相应的汉字输入系统，显然不应像面向专业操作员的输入方法那样，以快速性作为最主要的追求目标，但是既然作为一种汉字输入方法，如果达不到较高的输入速度，缺乏应有的效率，也是没有意义的，这就是早期“全拼音方案”尽管有很好的规范性和易学性，却难以为广大用户所接受的根本原因。所以快速性尽管不应作为规范编码最主要的追求

目标，但却是不容忽视的一项基本指标。

4. 兼容性：在计算机系统中实现规范汉字编码时应考虑三方面的兼容性，即：与国际通用键盘兼容，与多种显示终端兼容以及与通用操作系统兼容。当今计算机软、硬件产品已日益趋向国际化，过分强调汉字输入系统的专用性，不考虑与国际上流行的软、硬件相兼容的做法，不仅在国外无法应用，不利于国际间的交流，而且在国内也难以普遍推广。

5. 一致性：编码规则应前后一致，贯彻始终，不能自相矛盾，同类结构的汉字应使用相同的编码规则，尽量避免随意性。

6. 完备性：信息处理用汉字集通常分三个层次：基本集（含国标一、二级汉字共 6763 个），大字符集（包括基本集和辅助集，约 21000 个汉字）和全汉字集（包括全部汉字约 60000 个左右）。一个编码方案的完备性是指，该方案不仅能用其规则对基本集中的汉字进行编码，而且还能用同样的规则对大字符集乃至全汉字集中的所有汉字进行编码。

上述六项性能指标的重要性并不是均等并列的，而是针对不同的社会需求有不同的侧重。作为面向中小学教学用的汉字编码及汉字系统，我们认为，把“规范性”要求放在首位加以强调是必要的，这对于保持祖国语言文字的纯洁和统一，促进中小学语文教学质量的提高将起到积极的作用。与此同时，为了满足实用和普及推广的需要，对易学性、快速性和兼容性等三项指标也应给予足够的重视，即应把它们作为系统研制所必需达到的基本指标来追求。至于一致性和完备性两项指标则可作为附带目标，在保证达到前面四项基本指标要求的基础上努力加以争取，但是不一定要把它们列为系统研制所必需达到的基本指标。

如果编码方案和汉字输入系统的研制是针对社会上的专业操作员和打字员，则方案的设计应把快速性指标放在首位，在此前提下，考虑易学性、兼容性、完备性等项要求，并把它们作为系统研制所

必需达到的基本指标，而对一致性和规范性则可视为附带目标，甚至不作要求。

对于教师、作家和管理干部这类非专业操作员来说，显然“易学性”指标是最重要的，在此前提下再考虑快速性、兼容性、一致性和规范性等项要求，而对完备性则可以不予考虑。

对于某些较专门的应用领域，如大型图书资料检索、历史文献与古籍整理、古汉语研究……等等，由于其特定的社会需求，这时对编码方案的设计必然会把“完备性”指标放在首位，在此前提下再考虑快速性、易学性、规范性等项指标，而兼容性、一致性则被视为附带目标，甚至可不作要求。

由此可见，对编码方案和相应汉字输入系统的优选或研制，离不开社会需求分析，离不开具体的应用领域和使用对象。使用的对象不同，优选的原则和标准也不同，在某个应用领域或某个社会范畴内属于优秀的汉字编码方案，在中小学教学领域内未必是优秀方案；反之，在中小学教学领域被优选的最佳方案对于某些特定领域的应用也未必是最优方案。因此，那种认为可以把目前社会上专业打字员中使用最广的汉字输入方法直接在中小学推广的看法是片面的；那种认为可以用中小学教学用汉字编码及相应汉字输入系统，代替社会上所有其他汉字输入方案及输入系统的看法也是不全面的，因为这两种看法都忽视了对不同的应用领域和使用对象采用不同的评价标准这一基本原则。

面向中小学教学用汉字编码方案及汉字输入系统的设计目标是要实现一个符合规范性要求，并有良好的易学性、兼容性和很高输入速度的编码方案和相应的汉字输入系统，在达到上述四项基本指标要求的基础上，努力争取使系统的编码规则有较好的一致性和完备性。下面我们围绕如何实现上述设计目标，对汉字编码性能指标体系作进一步的分析。

§ 1.2 如何做到符合规范性

如上所述,基于形码的汉字输入系统的规范性体现在三个方面:

- * 应符合汉字结构规律和国家语委颁布的语言文字规范;
- * 应符合语文识字教学规律和国家教委制定的《全日制小学语文教学大纲》中有关识字教学的要求;
- * 应遵循国家技术监督局发布的《信息交换用汉字编码字符集》中的有关规定。

在这三个方面的规范要求中,前两个方面是与汉字编码设计密切相关的,第三个方面主要与汉字输入系统的软、硬件实现有关,而与编码方案设计关系不大。因此,我们在考虑形码方案的设计时可以只考虑前两个方面的规范要求。

根据汉字造字法和汉字演变历史,以及我国目前小学语文识字教学中,主要通过汉字部件教学生识记汉字的基本事实,我们认为,规范形码的研制应以部件码为主,笔画码为辅。下面我们在这个指导思想的基础上,从汉字构形理论和小学识字教学规律这两个方面对形码的规范性问题作进一步的探讨(关于音码,由于有汉语拼音方案作依据,其规范性容易保证,所以对音码规范性问题这里不予讨论)。我们认为,形码的规范主要涉及部件、笔画和简化字、异体字等几个方面的规范问题。

1. 关于部件的规范

部件也称字根或字元。关于部件,傅永和先生曾下过一个比较科学的定义〔14〕:“部件是构成合体字的最小笔画结构单位,其下限必须大于基本笔画,上限小于复合偏旁。从功能看,部件并不一定具有音义;从存在形式看,它是一个独立的书写单位,不管笔画多么复杂,凡是笔画中连在一起的,都作为一个部件看待。”有关部件的规范是指部件如何选取,部件如何构字以及部件如何拆分的规

范。

(1) 部件如何选取

目前国内较流行的形码方案中，对部件数量的选取，少的取 50 个左右（如五十字元码和安徽声形码），中等的取 100—200 个（如全息码取 100 个，王码取 150 个，郑码取 170 多个，部首码取 189 个），多的取 300 个以上（如表形码取 380 多个）。对部件的选取原则也各不相同，如郑码的部件 90% 选自《汉字统一部首表》，其余 10% 是惯用的组字构件；部首码的全部部件选自新华字典的 189 个部首；有的编码作者则主张形码部件应由通用部首中的独体字、呈相交笔画的偏旁、简化偏旁以及通用字表中三笔画以内的字等四部分组成；还有的编码作者干脆根据自己处理重码的需要任意选取部件。由于国家语委并未颁布过文件对汉字部件进行规范，所以目前要回答“形码部件应如何选取才符合规范？”这个问题是很困难的。但是这个问题又不能不加以解决，否则，就会造成语言文字使用的混乱，并对小学识字教学造成严重干扰。我们认为，对这个问题可以依据中国文改会（国家语委前身）正式发表过的一个文件和一个材料为基础，结合字源与汉字结构规律的研究来加以解决，这里说的文件和材料是：

* 1983 年公布的《汉字统一部首表》；

* 1985 年发表的《汉字末级部件组字频度表》。

第一个文件给出了规范的 201 个部首，其中 60% 左右是独体字，另有 23% 左右是合体字，其余为基本笔画和复笔部件；第二个材料则根据文改会和武汉大学对《辞海》所收全部单字（16296 个）以及《辞海》未收而 GB2312—80 收入的 43 个字（合计 16339 个汉字）的统计分析给出了 675 个末级部件。若不考虑其中的繁体字，则末级部件数为 648 个。所谓末级部件是指不能再进行切分的部件，由这 600 多个末级部件就可以组合出全部的汉字。上面所述的第一个文件是由中国文改会作为建议正式公布的，并且与我国中小学语文

教学中一直使用的《189 部首表》基本吻合,而广大中小学师生对 189 个改良部首是很熟悉的。尽管部首主要用于排检和查字,不等同于部件,但由于绝大多数部首构字能力强,因此以规范部首作为形码部件选择的基础和主要组成部分是可行的。上述第二个材料尽管不是具有法律效力的正式文件,但也是经过许多著名专家论证,经过科学的统计分析并以中国文改会名义公开发表的,因此是具有权威性的。为了缩短码长,减少重码,以提高输入速度,对形码部件的数量应选多一些;但是为了易学、易记,又希望形码部件的数量应尽量选少一些。直接取 201 个部首或 648 个末级部件作为形码部件尽管比较规范,但并不符合理想编码的要求。考虑到《201 部首表》中有些合体字构字能力很弱,对编码作用不大,我们认为选取规范形码部件的较好做法应当是:

①从《201 部首表》中选取所有独体字和复笔部首作为部件;

②从《201 部首表》中选取部分构字能力强,且为小学用间架结构进行识字教学所需要的合体字部首(如穴、足、齿、麻等)作为部件;

③从《648 末级部件表》中选取构字能力强,且为小学用间架结构进行识字教学所需要的若干独体字和复笔部件如“农”、“曲”、“曳”、“重”、“秉”、“毋”、“𠂔”、“曳”、“永”以及“惠头字(𠂔)”、“庸字底(肅)”、“御字中(𠂔)”、“杨字边(𠂔)”等以增补形码部件。

④不生造违反小学识字教学规律的部件。

(2) 部件如何构字

如何构字指部件的构字方式,也就是合体字的结构方式。按照汉字构形理论,为了分析合体字的结构可以采用平面分析法或层次分析法。平面分析法是将合体字一次分析出其全部构字部件,如分析“灌”字,一次分析得出的全部构字部件为“氵、艹、口、口、亻、圭”,其结构框图如图 1.1。层次分析法是将合体字逐级予以拆分,依次分析到部件为止。仍以“灌”字为例,第一级拆分,得出的部件