

汉字符底码键盘输入法

——兼论 25 种汉字基本构成单元和汉字构成全过程

喻 宜 著

汉字符底码

原子能出版社

汉字符底码键盘输入法

——兼论 25 种汉字基本构成单元和汉字构成全过程

喻 宜 著

原 子 能 出 版 社

北 京

图书在版编目(CIP)数据

汉字符底码键盘输入法:兼论 25 种汉字基本构成单元和汉字构成全过程
/喻宜著. —北京:原子能出版社,1997.4

ISBN 7-5022-1608-1

I. 汉… II. 喻… III. ①汉字信息处理②汉字编码 IV. TP391

中国版本图书馆 CIP 数据核字(96)第 17997 号

内 容 简 介

在本书中,著者第一次提出了一个系统的现行汉字构成学说,发现了 25 种汉字基本构成单元——符底(组字构件落底笔画),证明了汉字构成是有规律的,并在此基础上编制了汉字符底码键盘输入法。符底码是完全的形码,适用于全国各个方言区和世界各个汉字文化区。其主要特点是把一般所谓“字根”分成两部分:一是作为取码单元的 393 个组字构件;二是创建了具有可识别特征的 25 种符底。符底码可以成为一种易学、好用、高效的汉字字形输入法。著者运用汉字字形系统论解开了“汉字之谜”,证明了符底码就是理想的汉字键盘输入法。符底码的产生是形码的真正突破,为结束目前万“码”混战的局面提供了一个可选择的统一码方案。

本书可作为学习汉字输入法和汉字学的教材和参考书。适用于汉语汉字工作者、大中小学语文教师、学习汉语汉字的中外大学生、家用电脑使用者、从事文字处理工作的专业和非专业微机操作人员以及一切关注于探索“汉字之谜”的海内外人士。

符底码已申请中国专利。申请号:96 1 06336.X。符底码征求技术转让。

©原子能出版社,1997

原子能出版社出版发行

社址:北京市海淀区阜成路 43 号 邮政编码:100037

北京地质印刷厂印刷 新华书店经销

开本:787×1092mm1/16 印张 6.0 字数 147 千字

1997 年 4 月北京第 1 版 1997 年 4 月北京第 1 次印刷

印数:1—2000

定价:16.00 元

目 录

解开“汉字之谜”,合乎汉字构成规律的编码自然产生	(1)
汉字符底秩表	(6)
未列入《汉字符底秩表》的其它传统组字构件取码一览表	(7)
汉字符底码解说	(9)
第一章 几个名词解释	(9)
第二章 关于 25 种汉字符底和与之相对应的全部组字构件的说明	(10)
第三章 关于汉字字形编码的两个问题:字的拆分和组字构件的分类	(21)
第四章 汉字构成全过程(汉字字形系统论)	(27)
国标字基本集(6763 字)汉字符底码码本	(37)
汉字符底码取码规则	(37)
A 一点底首符字(440 字)	(38)
B 二点底首符字(258 字)	(40)
C 三点四点底首符字(226 字)	(42)
D 横一底首符字(207 字)	(44)
E 横二底首符字(268 字)	(46)
F 横三底首符字(400 字)	(48)
G 横端附点底首符字(270 字)	(51)
H 出头竖底首符字(391 字)	(53)
I 不出头竖底首符字(307 字)	(55)
J 侧竖弧竖底首符字(237 字)	(58)
K 二竖底首符字(635 字)	(60)
L 三竖四竖底首符字(53 字)	(63)
M 简单双分底首符字(224 字)	(64)
N 复合双分底首符字(143 字)	(66)
O 交叉双分底首符字(160 字)	(68)
P 竖斜双分底首符字(71 字)	(70)
Q 简单三分底首符字(453 字)	(71)
R 复合三分底首符字(130 字)	(74)
S 折底三角框底首符字(203 字)	(75)
T 空方框底首符字(303 字)	(77)
U 实方框底首符字(214 字)	(79)
V 特位特形框底首符字(384 字)	(81)
W 撇底捺底首符字(562 字)	(84)
X 左斜钩底首符字(163 字)	(87)
Y 右斜钩底首符字(61 字)	(89)

解开“汉字之谜”， 合乎汉字构成规律的编码自然产生

汉字*是一个谜。

汉字是世界上历史最悠久，使用人数最多，也是最具有生命力的文字之一。但是，在科学上，汉字还有个根本问题未能解决：其一，拼音文字（如英文）是有字母的。汉字不是拼音文字，有没有类似或接近字母的基本构成单元？如果说有，有多少？在哪里？哪些是汉字的基本构成单元？其二，英文的构成方式是以字母为构字单元，自左而右作线性排列，一步完成。汉字的基本构成单元是怎样排列成字的呢？也是一步完成的吗？汉字有没有不同于英文的、其它的构成方式？这是一个问题的两个方面，前者是关于汉字的构成单元方面，后者是关于汉字的构成方式方面，两方面总合起来就是汉字构成规律问题：汉字是由哪些基本构成单元，采取哪些构成方式构造出来的？人们提出和研究汉字构成规律问题是为了更好地理解和使用汉字。首先是为了制订出更简便的、能直接翻页的查字法。时间一年年过去，“汉字之谜”未能解开。中国人照样识字、写字、造字。计算机的出现对汉字提出了严重的挑战。世界上其它主要文字的电脑输入问题均已解决，只有汉字输入还是一个问题。汉字要进入计算机，迫切需要找到汉字的基本构成单元。

实践的目的决定问题的提法。问题是这样提出的：在汉字字形中，可以成为编码输入根据的根本特征在哪里？字母是英文的最小书写单位，笔画是汉字的最小书写单位，人们首先想到的是笔画。一般认为，一（横）、丨（竖）、丿（撇）、丶（点，或捺）、㇚（折）是“五种基本笔画”（即“札”字五笔）。然而，“五种基本笔画”既不能象英文字母组成英文那样直接组成汉字，也不能方便地输入汉字。于是，人们扩大了寻找范围。有人认为在字的四角或三角；有人认为在字义或频序；有人认为在按书写顺序的首笔、次笔区分为“五种基本笔画”的若干“字根”；有人认为在汉字或组字构件与英文字母的“形似”或“音似”；有人认为在这样一批他所“精选”出来的组字构件，有人认为在那样一批他所“精选”出来的组字构件；还有人认为汉字的本质在于它中间隐藏着不断旋转的英文字母，并宣布破除了汉字不是拼音文字的“传统观念”。名称叫法也不一致，如“字根”、“字元”、“汉字拼形字母”等。其数量也大相悬殊，从二三十个（加上变体实际上接近100个）到300多个不等。显然，相当一部分编码人并不认为只有汉字的基本构成单元才是可据以编码输入的根本特征；那些坚持寻找汉字基本构成单元的人也只是模仿英文（单一的构字单元和单一的构字方式）的模式，撇开对汉字构成方式的研究，希图直接找到汉字的基本构成单元。良好的出发点迅速地走向自己的反面。由于某些编码人任意地拆字选根，违背了人们的识字传统和识字习惯，破坏了汉字的完整性和统一性，以致有人惊呼汉字有被肢解的危险。结果是汉字编码尽管成百上千，却没有一个“理想的”汉字编码方案，也就是没有一个为广大汉字使用者普遍接受和共同使用的，易学、好用、高效的字形编码。

造成以上现象的根本原因在于不了解汉字构成规律。著者认为，必须把汉字基本构成单元

* 本书所称汉字即现行汉字（楷书），不包括隶书以前的古汉字。

和汉字构成方式两方面统一起来研究,同时考虑到汉字构成单元的多层次性和多种构成方式,才可能找到汉字构成规律,解开“汉字之谜”;只有从理论上解开了“汉字之谜”,合乎汉字构成规律的编码就会自然而然地产生。

符底码的基本设计思想是著者关于汉字构成全过程的思想,即汉字字形系统论。著者认为,全部汉字主要是由三个系列的构字单元以其特有的三种构字方式依次运作的结果。其中第二个系列是 25 种组字构件落底笔画,其名称是:

代码	组字构件落底笔画	组字构件举例
A	一点底	丶 冫 彳 ㄣ ㄤ
B	二点底	八 人 ㄥ ㄨ ㄨ
C	三点四点底	ㄣ ㄤ ㄨ ㄨ ㄨ ㄨ
D	横一底	一 ㄣ
E	横二底	二 工 土 土 上 止
F	横三底	三 王 五 正 且 里
G	横端附点底	一 虫 一 ㄣ ㄨ ㄨ
H	出头竖底	丨 十 丰 丰 巾 中 申
I	不出头竖底	丨 丁 干 干 手 手 甲 甲
J	侧竖弧竖底	乚 乚 口 口 了 了
K	二竖底	ㄥ ㄥ ㄥ ㄥ ㄥ ㄥ
L	三竖四竖底	川 雨 用 用 用 用 用
M	简单双分底	人 入 大 夫 夫 夫
N	复合双分底	几 九 尤 入 入 足
O	交叉双分底	乂 丈 史 冬 女 又
P	竖斜双分底	才 牙 长 食 戌 我
Q	简单三分底	小 东 个 本 未 来
R	复合三分底	水 水 水 水 爪 儿
S	折底三角框底	乚 世 ㄣ ㄣ
T	空方框底	口
U	实方框底	口 日 日 目 目
V	特位特形框底	白 由 画 言 甘 母 凹
W	撇底捺底	丿 丿 丿 丿 广 广 之
X	左斜钩底	乚 乚 乚 乚
Y	右斜钩底	乚 乚 心 乙 电 乚

著者认为,这 25 种组字构件落底笔画就是汉字中类似或接近英文字母的东西,就是汉字的基本构成单元。汉字基本构成单元就在组字构件的脚下!只有汉字的基本构成单元才是汉字字形编码的理想的取码标志。组字构件落底笔画简称“符底”。符底码认定 25 种符底为取码

标志,以对应国际通用键盘上 A~Y 25 个字母键,故本码取名“符底码”。

符底码同现有的其它形码的主要区别在于,把一般所谓“字根”分为两部分:一是认定 393 个结构单一的组字构件为取码单元,以决定字如何拆分;二是确定 25 种具有可识别特征的符底为取码的标志,决定取什么码。——“符底”是符底码的独创,为其它形码所无有——从而解决了其它形码难记难背、难拆难分的问题。

首先,现有的其它形码不为广大汉字使用者接受的主要原因是难学,要“背字根”,必须强制记忆那些关于“字根”键位的口诀和助记歌。这是由于他们认定的“字根”是不可识别的,“字根”在键位上的分配是主观的、随意的,“字根”之间是没有关联的。符底码不要求“背字根”。符底码确定的 25 种符底是具有可识别特征的,它是客观存在的,是构成图像的,互相之间是有秩序的。它存在于一切汉字所包含的组字构件之中,任何一个多少认识一些汉字的人只要掌握了这 25 种符底特征及其代码,就可以用它去识别任何一个汉字所包含的组字构件,决定它该取什么码。其数量只有 25 个,比任何一种“字根”最少的形码的“字根”还要少,因而符底码是易学的。

符底码的易学程度绝不是现有的其它形码所能望其项背的,只有全拼码可以与之相比。一个会讲普通话的人,如果不是在小学时代学会了普通话拼音,要掌握全拼码也有 26 个拼音字母要记住,还有 417 个基本音节要熟习。符底码要记住的也不过是 25 个符底及其代码,要熟习的也只有 393 个取码单元。对于操作者来说只须记住这 393 个组字构件是取码单元,不拆不分,取一码;其它组字构件不是取码单元,要拆,取两码以上就够了。至于每个取码单元的键位,在其它形码那里是最难记的;在符底码这里是不用记的,因为它可以识别。这就是说,拿一个一般的汉字使用者学会使用符底码,同一个会讲普通话的人学会全拼码相比,其难易程度大致相当。但是,要识别符底特征是每一个汉字使用者都能做到的,要学会普通话,就不是所有的人所能做到的了。

其次,现有的其它形码不为广大汉字使用者接受的另一原因是不好用,主要的是拆分混乱。汉字的拆分是有不明确的地方,但谈不上混乱。汉字输入中的拆分混乱在很大程度上是由于某些编码人为了追求高速输入、认定“字根”太少造成的。因此也是不难解决的。为了避免拆分混乱,必须做到在一个组字构件内相连相交的笔画不拆,不搞特例编码字,这就需要有统一的、足够多的取码单元(“字根”),其数必须在 300 个以上。问题是这样做结果,“字根”多了会增加操作者的记忆负担。符底码找到了组字构件中的可识别特征,无需顾虑取码单元太多会增加操作者的记忆负担;全部汉字按《汉字符底秩表》认定为取码单元的 393 个组字构件取码,无一特例编码字。393 个取码单元,这个数字比认定“字根”最多的形码的“字根”还多。因此,在目前条件下,符底码的拆分是最规范的,也是最好用的。

符底码也有自己的缺点,字输入重码率 33.3%。同现有的其它形码相比,符底码的重码率是够高的了。大家知道,国际通用键盘是为英文打字设计的。英文字母 26 个,国际通用键盘 26 个字母键。一个字母一个键位,不存在重码问题。汉字则不同。汉字组字构件有 600~700 个,国际通用键盘上可利用的键位只有三四十个,平均每个键位要承担 20 来个组字构件。因此,只要是利用国际通用键盘,汉字编码中的重码就是不可避免的。现有的其它形码的低重码率是人为的,是以增加学习难度为代价换来的。办法就是不顾组字构件的客观特征,把不同的“字根”在不同的键位上调来调去,同时设置特例编码字,搞非规则编码。符底码对全部汉字和组字构件按其客观特征分类,按规定的码元取码,无一非规则编码字,其重码率自然会高于其它形码。

但是,符底码的重码率同全拼码比则是很低很低的了(全拼码的重码率几乎是100%)。同全拼码相比,符底码每个同码对应的字数也少得多(全拼码同码对应的字数最多的超过100个)。符底码同码对应字数最多的16字,一码。同码对应15字、13字、11字者各一码,对应12字者二码,对应6~10字者30码。其它同码均在2~5字之间。这就是说,在一般软件技术条件下,采用高频先见,屏底显示选择技术(每页以10字计),除六码外,其它重码均无需翻页。加上词输入、自设定词等技术,非专业的电脑使用者操作起来比全拼码要方便得多。近年来,开发出一种能够代替人选择重码字的的人工智能软件。依靠这种软件技术,已经产生了若干种高效的拼音码;依靠这种软件技术,符底码也一定可以成为易学、好用、高效的形码。

符底码之所以是好学的,是因为它不过是提供了一个新的视角:“汉字的基本构成单元就在组字构件的脚下!”只要人们掌握了这个视角,就可以用以识别组字构件本身具有的可识别特征。符底码是好学的,也需要学习。世界上就没有不需要学习的文字。汉字需要学习,汉字的使用方法——符底码也需要学习。至少25种符底及其代码必须记住吧!还有393个取码单元需要熟习。熟习过程也是学习过程。只不过符底码的学习成本低一些,有一定文化基础的人可以自学,勿需培训。可以预料,25种汉字符底既经发现,一旦它之为汉字的基本构成单元的地位得到确认,就一定会像普通话拼音一样纳入中小学语文教学内容,逐步成为每一个有文化的中国人的常识。只有到了那时,25种汉字符底就像是每个人不学自会的(因为在中小学里已经学过了)一样,才能做到识字、查字和汉字输入的完全统一。

中国要尽快地实现现代化,电脑要成为全民使用的普及性工具,全国计算机要联网运行,迫切需要一个汉字输入统一码。目前这种万“码”奔腾的混乱局面必须尽快结束。结束这一局面的最好办法就是把汉字统一码确定下来。

有人主张以音码为汉字统一码。他们说,汉字是没有规律的,形码是汉字编码的“死路”。与此相反,著者认为,汉字统一码非形码不可。首先,字形是字音和字义的载体,字音和字义离开了字形就无所依附。汉字键盘输入是一种最基本的输入方式,应该按规范的现行汉字字形进行输入。如果键盘输入失去了字形,其结果是:汉字可以拼音输入,可以语音输入,也可以手写输入,唯独无法按规范的汉字字形进行输入,那将大大淡化人们对识字重要性的认识,带来无穷的“后遗症”,贻害子孙。其次,中国历来只有“书同文”,没有“语同音”。只有形码才能突破地区的隔阂,解决音码输入中普通话同地方音、方言的矛盾,以及多音字和生字输入问题,使电脑真正成为全国统一的,各个地方通用的普及性工具。再次,只有形码才能克服不同国家汉字形、音、义的差异,适用于世界各个汉字文化区,把不同国家和不同地区的汉字编码统一起来,发挥汉字在世界信息交换中的作用,弘扬汉字文化传统,加强汉字在世界华人中的凝聚力。假若确定音码为汉字统一码,只能有两种解释:要么是承认汉字没有规律,要么是承认中国人也解决不了汉字规律问题。二者必居其一。

著者也认为现有的其它形码不具备成为汉字统一码的条件。因为这些编码都是在不了解,甚至不承认汉字构成规律的思想基础上编制的。但是,如果从中得出汉字构成没有规律的结论,那就大错特错了。在本书中,著者从汉字字形的根本特征在组字构件落底笔画这一观点出发,发现了25种汉字的基本构成单元(符底),完成了对全部组字构件的科学分类,并一步步地描述了以三个系列的构字单元和相应的三种构字方式为主要内容的汉字构成全过程,充分证明了汉字构成的规律性;并把25种汉字基本构成单元(符底)定位在国际通用键盘的键位上,

这种定位是顺乎自然的、别无选择的,也就是理想的。这样,也就有理由认为在上述思想基础上编制的符底码是合乎汉字构成规律的。只有合乎汉字构成规律的编码才是理想的汉字字形编码。因此,符底码是唯一有条件成为汉字统一码的汉字编码方案。

符底码起步很晚,现在刚刚形成包括国标字基本集(6763字)的技术方案。汉字编码“热”已经过去,已经涌现出成百上千种汉字编码,其中有的已有多发展的基础,有的已经形成强大的经济后盾,有的已经产业化并占领了市场。面对如林的强手,著者不由得不感到自身的软弱和稚嫩。特别值得一提的是,形码的难记、难背、难拆、难分和某些字形编码人的不实宣传已在人们心中造成了很深的负面印象,要人们相信符底码就是理想的汉字字形输入法,必须作加倍的努力。但是,著者知道,对于目前这种万“码”奔腾的混乱局面,群众是不满意的,领导也是不满意的;编码的使用者是不满意的,广大的要求使用汉字编码却无码可用的人更是不满意的。最不满意的是,编码成百上千,却没有一个令大家满意的编码方案。著者不揣冒昧,自以为符底码可以赢得广大汉字使用者的满意,把目前这种混战的局面统一起来。人们拒绝接受的只是那些无理的、随意的编码;合乎汉字构成规律的编码是会受到欢迎的。符底码终将为广大汉字使用者所接受,尽管这个道路是漫长而艰巨的。

汉字符底秩表

类别	代码	符底 (可识别特征)	组字构件 (取码单元 393 个)	首字
点底类	A B C	一点底 二点底 三点四点底	丶ノミくマシム专 ハワハ《ハハハ ①ハハハハハ②ハ(一)魚	专八魚
横底类	D E F G	横一底 横二底 横三底 横端附点底	一(ノ)一 ニユエ土士上止立血 三ヨ王丑主五主正正且並且其曲里豆聖里金(金)重 ①一虫②一ム③一ム血	一上主虫
竖底类	H I J K L	出头竖底 不出头竖底 侧竖弧竖底 二竖底 三竖四竖底	丨十千丰巾中巾巾丰巾申韦申申申帝韋串車由事 イ丁丁干于于手手手手甲甲甲卑 ①トトリリ十十リリリリ手手手耳聿 ②了子子子子子 ①川井廿升日井月同丹丹冉弗 ②口門内内内内兩禹禹③斤片片 ①川卅用南雨雨雨(雨)肅②冊而而冊册	中于了肉用
双分底类	M N O P	简单双分底 复合双分底 交叉双分底 竖斜双分底	人(人)入大火夫央夫央央央夷夷 ①儿几凡九尤龙无尤无无无 ②スススス③止正正(正)足(足) ①义又文文史史史史及及④女⑤父 ①才才牙弟弟身②匕(匕)氏长食(食) ③比氏氏戊我戎戎	大九及长
三分底类	Q R	简单三分底 复合三分底	①小ナ(山)东乐 ②个木未未来束東果乘 ①小(山)兼②水水水永承求秉承③収禾禾农 ④水水水⑤瓜(西)⑥儿小本东汎	个本
折框底类	S T U V	折底三角框底 空方框底 实方框底 特位特形框底	①レリロセヒヘフ止世 ②コヨ工出毒③山山白④ム云么 口 □日田目四田四 ①白白由石自曲曲曲西西曲言(言) ②々甘母甘母母瓦母凹凸呂	山口日自
斜底类	W X Y	撇底捺底 左斜钩底 右斜钩底	①丿厂广广尸尸尸尸②リリナナナナテ尹尹尹尹尹尹 ③夕夕夕夕夕④夕⑤之之 ①丁(フ)刀(力)力力②丁③夕勿④夕与马乌弓⑤乃乃 ①乚戈戈戈戈戈②心③し乙七七也毛也屯电电④飞气	之力也

未列入《汉字符底秩表》的其它传统组字构件取码一览表

(320个,按书写顺序取码,按末符分类)

A 一点底末符(46个)					
为 AXA	豕 DRA	豕 DSAA	弋 DYA	土 EA	
玉 FA	书 HA	冂 寸 JA	冂 斥 KA	兵 KDA	
戕 KYA	甫 LA	犬 MA	令 MAA	凡丸尤龙 NA	
乂 乂 OA	氏 戍 我 戍 戍 戍 PA		戍 PAA	成 PXA	
术 QA	求 RA	彳 SA	身 夕 WA	发发 WOA	
刃 勾 习 勾 XA	羽 XAXA	兔 XNA	匆 XWA	飞 YA	
B 二点底末符(15个)					
立 EB	黄 EVB	五 FB	穴 GB	肅 HKB	
平 乎 IB	夹 夾 MB	兆 NB	※ OB	来 來 QB	
办 办 XB					
C 三点四点底末符(5个)					
黑 熏 FC	鸟 XAC	马 乌 XC			
D 横一底末符(19个)					
乚 BD	业 CD	亚 DRD	互 DVD	虫 出 HD	
垂 ID	卫 JD	韭 JJD	丘 阝 KD	厶 MD	
無 MLD	业 丞 RD	尸 WD	重 WHD	与 刁 XD	
E 横二底末符(5个)					
回 KE	至 SE	止 壬 血 WE			
F 横三底末符(7个)					
兰 BF	佳 IF	亘 HF	雀 JF	缶 MF	生 皿 WF
G 横端附点底末符(9个)					
𠂇 CG	𠂇 DBG	𠂇 𠂇 HG	𠂇 𠂇 KG	𠂇 QG	
瓜 RG	𠂇 WG				
H 出头竖底末符(20个)					
乚 半 BH	巾 DH	辛 EH	車 HUH	𠂇 IH	
𠂇 JJH	伞 MBH	𠂇 午 年 MH	竹 MHMH	革 VH	
千 𠂇 牛 巾 𠂇 𠂇 WH		𠂇 YH			
I 不出头竖底末符(8个)					
予 AI	𠂇 羊 BI	𠂇 DI	臣 SI	𠂇 WI	
争 XI					
J 侧竖弧竖底末符(6个)					
斗 AJ	下 卞 DJ	非 JJ	𠂇 MJ	耳 XJ	
K 二竖底末符(15个)					
𠂇 DBK	𠂇 开 丙 两 再 DK	𠂇 DOK	𠂇 EK	青 FK	
骨 GK	鼻 VUDK	𠂇 升 舟 WK	面 WUK		
L 三竖四竖底末符(7个)					
州 CL	𠂇 DTL	𠂇 MDTL	𠂇 RDL	鼎 UL	
角 WL	角 XL				

M	简单双分底末符(19个)				
	头 AM	关美 BDM	类兴 BM	天灭 DM	亥 DWM
	买 GAM	欠 GM	建 JM	贝 KM	矢 MM
	以 SAM	尺 SM	页 WKM	夭失 WM	奂 XM
N	复合双分底末符(10个)				
	羌 BN	兀无 DN	元走 EN	见 KN	光 QDN
	久 WN	虎 WYN	免 XN		
O	交叉双分底末符(16个)				
	义 AO	父 BO	必 CO	文更 DO	麦 FO
	支 HO	艮支 JO	卤 JUO	父 MO	风受 NO
	妻 SO	皮爻 WO			
P	竖斜双分底末符(8个)				
	矛 AP	良良 ASP	𠂔 DP	喪 EBP	民艮目 SP
Q	简单三分底末符(11个)				
	米 BQ	不禾 DQ	示 EQ	糸 SQ	𠂔 UQ
	采 WBQ	禾朱 WQ	麻 WQQ	系 WSQ	
R	复合三分底末符(10个)				
	兼 BDR	衣亦豕 DR	𠂔 DSR	赤 ER	辰 FR
	彖 VDR	黍 WQMR	象 XWR		
S	折底三角框底末符(24个)				
	𠂔 BQS	𠂔 DKS	亡𠂔玄 DS	𠂔 EMS	云去 ES
	甚 FBS	𠂔 FS	出喜 HS	𠂔 JUS	𠂔 MHS
	鬼 NS	𠂔 巨 SS	𠂔 WES	𠂔 WHS	𠂔 正 么 WS
	𠂔 WXS	𠂔 XMS			
T	空方框底末符(5个)				
	谷 BMT	高 DTKT	舌 WHT	司 XDT	司 XT
U	实方框底末符(6个)				
	首 BWU	音 EU	旧 HU	香 WQU	𠂔 WU
V	特位特形框底末符(1个)				
	𠂔 WV				
W	撇底捺底末符(23个)				
	户 AW	𠂔 BDW	𠂔 BW	𠂔 DRDW	𠂔 DW
	产𠂔步 EW	𠂔 FSWW	𠂔 HMMW	𠂔 JW	𠂔 KDW
	𠂔 QMOW	少 QW	𠂔 𠂔 SW	𠂔 𠂔 𠂔 WW	𠂔 𠂔 𠂔 XW
X	左斜钩底末符(9个)				
	𠂔 DJX	𠂔 𠂔 DX	𠂔 EX	𠂔 FX	𠂔 ISX
	今 MAX	𠂔 WX			
Y	右斜钩底末符(16个)				
	𠂔 JY	𠂔 KY	𠂔 𠂔 𠂔 𠂔 SY	𠂔 TY	𠂔 WSSY
	𠂔 𠂔 𠂔 𠂔 WY	𠂔 𠂔 XSY	𠂔 𠂔 XY		

汉字符底码解说

汉字符底码是一种以字形为取码根据的计算机键盘输入方法。符底码是完全的形码,按组字构件取码,不含任何“形托”或“音托”的因素。

由于实现了对汉字的严格定序,符底码还可以用于汉字检索(直接翻页)、电报通讯、排版检字、图书分类等方面。在没有计算机的条件下,可以先学习符底码查字法,作为学习符底码键盘输入法的准备。

第一章 几个名词解释

1. 组字构件和偏旁、部首

汉字组字构件(偏旁、部首)是由笔画组成的,具有特定构字功能的构字单元。一般称为部件,本书称为组字构件。据著者所知,零件、部件和构件都是机器或设备中的构成单元。零件是单个的个体,部件是由若干零件组成的集合体,构件既包括单个的零件,也包括由若干零件组成的部件。偏旁、部首就是这样,有的由单笔画组成,有的由若干笔画组成,故称组字构件比部件更合适。

1985年4月中国文字改革委员会和武汉大学统计:汉字末级部件648个。1986年科学出版社出版的《汉字信息字典》载:汉字部件623个。两家取法和统计不尽相同。本书著者把两家相同的组字构件算一个,不同的算两个,繁体、简体和变体分别计数,其总数为700多个,大于两家单独的统计。减去据义划分是单体字、据形划分不是单体字的组字构件(如孔、扎、札、乱、礼等),减去无关的组字构件,剩下与国标字基本集(6763字)有关的传统组字构件669个。其中大约有一半的组字构件是单一结构的,也包括一些显然由几个构成单元组成的部首(如麻、鼻、香、青、鹿、齿、象等),还有许多单体字是不是由几个构成单元组成的,人们有不同的看法(如伞、丧、州、产、半、赤、办、币等)。

2. 组字构件和取码单元

《汉字符底秩表》包括符底码认定为取码单元的组字构件393个,其中349个来自传统组字构件,另有在编码过程中新产生的“人工组字构件”44个。符底码认为这些组字构件的结构是相对单一的。结构的单一性——任何组字构件只能有一种落底笔画,其落底笔画不可拆分——应该是组字构件的根本特征,只有结构单一的组字构件才是严格意义的组字构件。在汉字输入时,《汉字符底秩表》认定为取码单元的组字构件不能拆分,取一码。其中较复杂的组字构件也不能拆分成较简单的组字构件。如“金”字不能拆分为人、王、丿;“言”字不能拆分为一、二、口。如果你弄不清一个组字构件是否应该拆分,只须在《汉字符底秩表》中查阅,表中有的就是取码单元,不拆;表中没有的就不是取码单元,必须拆分。

在《未列入〈汉字符底秩表〉的其它传统组字构件取码一览表》中,包括的是符底码没有认定为取码单元的传统组字构件320个。符底码认为这些组字构件的结构是多元的,实际上是合体字(多符字),在汉字输入时必须拆分,取两码以上。而且要按照《汉字符底秩表》中最大的组

字构件进行拆分。如“露”字不能拆分成雨、亠、口、止、久、口，只能拆成雨、显、久、口。这叫做“取大优先”。著者注意到：其中相当一部分多元结构的传统组字构件容易被人误认为是结构单一的取码单元，按其最低部位的笔画特征取码。这里，我们将错就错，按其书写顺序的末符分类，以便查阅。意思是说，如果你弄不清一个多元结构的组字构件应该如何拆分，可以先按其最低部位的笔画特征试取一码，在该表中按末码核查，然后再予确定。不过书写顺序和部位高低毕竟不完全一致，有些以一点、二点为末符的组字构件，如犬、书、尤、来、平，仍须按其书写顺序取码，才能在该表中找到。其实，《汉字符底秩表》规定了393个取码单元，也就规定了其它传统组字构件应该拆分和应该如何拆分。

3. 符底和“字底”

“符底”是组字构件落底笔画的简称（因为叫“件底”实在太别扭）。“符”字作为组字构件的单字简称，除符底、符底码外，本书还在其它地方使用，如首符、次符、一符、二符、末符等。

符底，即组字构件落底笔画有种种不同的情况：有的组字构件没有次要笔画，如各种起始构件、丶、ハ、ㄣ、一、丨、ㄥ、川、人、小、口、丿等，其全部笔画都是落底笔画；有的组字构件上部有次要笔画，如鱼、虫、而、食等，其最低部位的笔画是落底笔画；有的组字构件的次要笔画在构件中部，它是由自上而下、贯穿到底的笔画落底的，如十、升、事、中、申、屯、曳、帝、电等；还有的组字构件是由自上而下、贯穿到底的笔画和最低部位的笔画一起落底的，如大、夫、尤、木、未、兼等。由此可见，在组字构件中，不落底的笔画是次要笔画，落底笔画是决定组字构件性质的主要笔画。组字构件的落底笔画可以是单一笔画，也可以是由若干单一笔画组成的复合笔画。无论落底笔画最多是少——一般不超过四画——这些笔画必须结成单一的结构。只有结构单一的落底笔画才有可能准确取码。

“字底”是一个不确定的概念。汉字由不同数量的组字构件组成。由一个组字构件组成的字叫单体字（一符字），字底就是符底。由两个以上的组字构件组成的字叫合体字（多符字），由于组字构件数不同，组合方式不同，可以是一种落底笔画，也可以是多种落底笔画（如横向组合的字：明、柳、涨等），不存在统一的字底。符底则是确定的概念。无论一个字包含几个组字构件，无论组合方式如何，也无论组字构件在字中处于什么地位，每一个组字构件都有一个结成单一结构的落底笔画，即使该组字构件处于字的最高部位也是如此（如草、文、关、宣、笠等字中的艹、宀、㇀、宀、㇀）。故本书不用“字底”一词。

第二章 关于25种汉字符底和与之相对应的全部组字构件的说明

学习符底码的最基本材料是《汉字符底秩表》。现在，我们对《汉字符底秩表》中25种符底和它所对应的、认定为取码单元的393个组字构件作以下的说明，同时说明未列入《汉字符底秩表》的、必须拆分的320个其它传统组字构件。

一、点底类

在汉字中，有单点，有点与点的组合（复合点），还有点与线的组合。单点和组合点只能成为组字构件而不能成字；只有点和线形笔画的组合既能成为组字构件，又能成字。点的组合有横向组合和纵向组合。点在落底部位的横向组合变化是基本的变化，是不同符底的变化；点的纵

向组合变化,包括在点的上部增添各种次要笔画,是非基本的、同一符底内的变化。在一个组字构件内,无论是什么样的组合,只有点居于组字构件落底部位,才属于点底类构件;也无论组字构件总的点数有多少,只有在落底部位的点数,才是其落底笔画。

代码 A 一点底构件

一点底构件是一点居于落底部位的组字构件。首先,一点是一个独立的组字构件,它可以不增添任何笔画,就是一个完整的构字单元和取码单元。其次,它又是一个可以纵向扩张的落底笔画。在它的上中部不断地增添次要笔画或改变其书写形态,可以形成其它的一点底构件——它是一点底构件的起始构件。请注意,各种起始构件的纵向扩张有一个特点,它是单向的向上扩张,其落底笔画不变。冫(冰两点)、氵(三点水)都属于一点底构件。再次,一点又是一个可以横向扩张的落底笔画。其它点底起始构件,ハ、ㄣ、ㄥ(横二点、横三点、横四点)都是、(一点底)横向扩张的结果——它是点底类的起始构件。《汉字符底秩表》认定为取码单元的一点底构件有 8 个(括号内为例字):

丶(户)、冫(冰)、ㄣ(冬、斗)、亠(飞)、マ(勇、令)、氵(江)、ㄥ(监)、专。

未列入《汉字符底秩表》的、以一点底构件为末符的传统组字构件有 46 个(后附取码,括号内为例字,下同):

为 AXA	豕(豕)DRA	宀(率)DSAA	弋(武)DYA	土(压)EA
玉 FA	书 HA	卵(卵)寸 JA	冂(丽)斥 KA	兵 KDA
戕 KYA	甫 LA	太犬 MA	令 MAA	凡丸尤龙 NA
又(夜)OA	氏(越)我戊(戩)戩(戩)PA	戍 PAA	成 PXA	
术 QA	求 RA	卩(以)SA	夕(卵)夕(然)WA	发(拔)发 WOA
刃勾习勾 XA	羽 XAXA	兔 XNA	匆 XWA	飞 YA

代码 B 二点底构件

二点底构件是二点居于落底部位的组字构件。ハ(横二点)是、(一点)横向扩张的结果,又是二点底构件的起始构件。它本身是一个组字构件,又可以在它的上部增添次要笔画或改变其书写形态而形成其它二点底构件。在汉字辞书中,八字和ハ(横二点)是不分的,我们也把它归在一起。《汉字符底秩表》认定为取码单元的二点底构件有 7 个:

ハ(六)、ㄣ(关)、八、𠂇(𠂇)、𠂇(脊)、𠂇(𠂇)、𠂇(竺)。

未列入《汉字符底秩表》的、以二点底构件为末符的传统组字构件有 15 个:

𠂇(丧)EB	黄 EVB	𠂇(釜)FB	穴 GB	肅 HKB	平乎 IB
𠂇(𠂇)MB	兆 NB	𠂇(𠂇)OB	来束 QB	办(梁)办 XB	

代码 C 三点四点底构件

三点四点底构件是三点或四点居于落底部位的组字构件。ㄣ、ㄣ(横三点、横四点)是、(一点)进一步横向扩张的结果,又是三点四点底构件的起始构件。它本身是组字构件,又可以在它的上部增添次要笔画或改变其书写形态而形成其它三点四点底构件。由于字少,两种组字构件共用一个代码。在简化字中,大多数ㄣ、ㄣ(横三点、横四点)已经简化为一(一横)。为了保持汉字符底序列的完整性和繁简体共用,彡、彡等组字构件和马、鸟、鸟等字中的一(横),仍按三点四点底取码。与、丝字中的下横按横取码。《汉字符底秩表》认定为取码单元的三点四点底构件有 8 个:

①三点底构件:

ㄣ(州)、ㄥ(必)、ㄨ(兴)、ㄣ(红)、ㄣ。

②四点底构件:

灬(杰、黑)、一(马、鸟)、鱼。

未列入《汉字符底秩表》的、以三点四点底构件为末符的传统组字构件有 5 个:

黑熏 FC 鸟 XAC 马鸟 XC

二、横底类

一(一横)是横底类构件的起始构件,又是横一底构件的起始构件。一(一横)的横向扩张是横端附点,属横底类构件的符底变化。一(一横)的纵向变化,即在底横之上增添次要笔画(包括横画),本不属符底的变化。但其组字能力强,组字构件形态多。为了分流,特以底横之上的横画数区分之,称之为横一底、横二底、横三底,意思是:横底之一、横底之二、横底之三。

代码 D 横一底构件

横一底构件的特征是以单一笔画一(一横)为底,其底横两端不连不接其它笔画,底横之上,其它笔画可有可无,但不得再有横画,或有横画参与组合的混合笔画(如横折竖、竖折横等)。一(一横)是横一底构件的起始构件,它本身是一个组字构件,又可以在它的上部增添次要笔画(非横画)或改变其底横书写形态而形成其它横一底构件;还可在底横之上增添横画或横端附点而形成其它横底起始构件——它是横底类的起始构件。《汉字符底秩表》认定为取码单元的横一底构件有 3 个:

一、丿(刁、丿挑是横的特殊书写形态)、一(文)。

未列入《汉字符底秩表》的、以横一底构件为末符的传统组字构件有 19 个:

ㄣ(关)BD	ㄣ(兴)CD	ㄣ DRD	互 DVD	ㄣ(贵)ㄣ(蚩)HD
垂 ID	卫 JD	韭 JJD	丘ㄣ(舆)KD	ㄣ(合)MD
ㄣ(舞)MLD	业丞 RD	尸(后)WD	重 WHD	与刁 XD

代码 E 横二底构件

二(二横)是一(一横)纵向扩张的结果,是横二底构件的起始构件。二(二横)本身是组字构件,又可以在它的上中部增添次要笔画或改变其底横书写形态,而形成其它横二底构件。横二底构件的特征是以单一笔画一(一横)为底,底横两端不连不接其它笔画,底横之上其它笔画可有可无,但只能有、也必须有一横画,或一由横画参与组合的混合笔画(如横折竖、竖折横等)。其它笔画则在底横之上,与上横相连、相交或相接。《汉字符底秩表》认定为取码单元的横二底构件有 10 个:

二、ㄣ(侯)、工、土、士、上、ㄣ(共)、止、立、皿。

未列入《汉字符底秩表》的、以横二底构件为末符的传统组字构件有 5 个:

ㄣ(冒)KE 至 SE ㄣ(先)壬血 WE

代码 F 横三底构件

三(三横)是一(一横)进一步纵向扩张的结果,又是横三底构件的起始构件。三(三横)本身是组字构件,又可以在它的上中部增添次要笔画或改变其底横书写形态,而形成其它横三底构件。横三底构件的特征是以单一笔画一(一横)为底,底横两端不连不接其它笔画,底横之上其它笔画可有可无,但不得少于二横画,不论这些横画的长短、左右,是混合笔画还是单一笔画。其它笔画在底横之上,附着于上部诸横的上下左右之间。《汉字符底秩表》认定为取码单元的横

三底构件有 23 个:

三、ㄣ(录)、王、丑、圭(表)、五、主、正、步(卸)、且、𠂔(寒)、𠂔(影)、圭(佳)、且(具)、其(其)、𠂔(典)、里、豆、𠂔(黑)、𠂔(盖)、金、𠂔(铁)、重(熏)。

未列入《汉字符底秩表》的、以横三底构件为末符的传统组字构件有 7 个:

𠂔(盖)BF 佳 IF 𠂔(彭)HF 雀(鹤)JF 金(卸)MF 生且(寡)WF

代码 G 横端附点底构件

横端附点是一(一横)的横向扩张。一(一横)因自身宽阔,占地太广而无横向扩张的余地,只有横端附点尚有可为。横端附点底构件是以底横的一端或两端与点相连或相接而形成的组字构件。横端附点有一端附点、两端附点之分。一端附点又有两种:一为横尾接点,写作一;一为横尾连点,写作一(横下钩)(著者认为,汉字笔画中的钩是线形笔画尾端连点,出锋而形成的混合笔画)。两端附点者写作一(点接横首,横尾连点,称平宝盖)。其它横端附点底构件是在附点后的底横上部增添次要笔画或改变其底部具体的笔画组合的结果。《汉字符底秩表》认定为取码单元的横端附点底构件有 7 个:

①横尾接点底构件:一(惠)、虫。

②横尾连点底构件:一(买)、人(尔)。

③平宝盖底构件:一(军)、一(安)、𠂔(骨)。

未列入《汉字符底秩表》的、以横端附点底构件为末符的传统组字构件有 9 个:

𠂔(学)CG

𠂔(帝)DBG

𠂔(惠)𠂔(囊)HG

𠂔(蒙)𠂔(曼)KG 𠂔(常)QG

瓜 RG

𠂔(風)WG

三、竖底类

丨(一竖)是竖底类构件的起始构件,𠂔、川、𠂔(二竖、三竖、四竖)是丨(一竖)的横向扩张。丨(一竖)又是一竖底构件的起始构件。汉字中以合体字居多,合体字中以左右组合居多。一竖底构件体态修长,适合左右组合,其组字能力极强,体态繁多。为了分流,我们按符首或符身的特征,把一竖底构件分为几种形态:出头竖底构件是一竖底构件的主要形态,不出头竖底构件和侧竖弧竖底构件是一竖底构件的变化形态。各立代码称之。

代码 H 出头竖底构件

出头竖底构件是一竖底构件的主要形态,其特征是一竖落底,主竖上端高于同一组字构件的其它笔画。其主竖上下贯通,穿天透地,次要笔画多与主竖相交,紧贴于主竖两侧。《汉字符底秩表》认定为取码单元的出头竖底构件有 24 个:

丨(引)、十(李、博)、丰(奉、半)、扌(打)、𠂔(舛)、𠂔(出、蚩)、巾(币、师)、𠂔(书)、𠂔(年)、丰、𠂔(肺)、中、𠂔(唐、肃)、𠂔(韦、半(革)、𠂔(囊)、𠂔(申、𠂔(制)、𠂔(聿、𠂔(惠)、𠂔(串、𠂔(害、重)、𠂔(叟)、事。

未列入《汉字符底秩表》的、以出头竖底构件为末符的传统组字构件有 20 个:

𠂔半 BH

𠂔(师)DH

𠂔 EH

𠂔(朝)HUH

𠂔(攸)IH

𠂔 JJH

𠂔 MBH

𠂔(竹)午年 MH

竹 MHMH

𠂔 VH

𠂔(在)𠂔(牛)𠂔(牛)𠂔(壮)𠂔(制)WH 𠂔(迅)YH

代码 I 不出头竖底构件

不出头竖底构件是一竖底构件的另一形态,其主竖可以看成是一竖的纵向扩张。同其它组