

第7章 编码规则

《郑码》的编码规则有三条，给汉字和词语编码须遵循。

7.1 第一条 汉字拆分成基本字根才能编码

给汉字编码，首先看这个汉字是由哪几个基根组成的，现在将这一过程俗称为“拆分汉字”。为什么要拆分？因为只有基本字根和笔画才有代码，所以必须将汉字拆分为有代码的基本字根才能进行编码，在没有合适的基根可以表述时，要进一步拆分为笔画才能编码。

7.1.1 汉字拆分的要领

绝大部分的字，例如“汉、码、章、辑”等很容易拆分，但见到“尊、首、街、灸、姊、督”等字，如何拆分？大家会感到没有把握，为此要给大家介绍一下汉字拆分的要领。

1、记住基根和笔画才能正确拆分汉字：

因为基根和笔画有代码，所以它们是拆分汉字的法宝，我们必须熟知自己的百宝箱——基本字根表中有哪些法宝，才能准确地运用它们给汉字编码。

● 熟悉基根不需要死记硬背：

因为85%的规范部首都是《郑码》的基根，大家对部首是熟悉的，不需要再去背。因此，见字后先按自己熟悉的部首拆分汉字，通常都会迎刃而解。如果行不通，再作进一步分解。例如：“恤”先拆分为“亻血”，但在“撇一竖”的基根中找不到“血”，于是再进一步分解为“亻丿皿”。将“辣”先拆分为“辛束”，但“束”也不是基根，于是要进一步分解为“辛木口”（请参阅第11章的“分解示例”）。

● 只有记住了基根才能正确地拆分汉字：

例如：“尊”字，必须拆分为“丩酉寸”，如果先拆出一个“亼”，下边的“酉”缺少“一（一横）”就不成字根了。“街”字拆分为“行土土”（按嵌套拆分）。

2、没有合适的基根才能拆分成笔画：

如：“灸”拆为“㇇火”；“姊”拆为“女一与丨”；“叔”拆为“丨一小又”。

3、从汉字中拆分出的基根和笔画必须是最少的：

就是说，能拆分成两个的决不拆成三个；能拆成三个的，决不拆成四个。例如：“首”字必须拆分成“亼自”，而不能拆分为“丩一自”，因为有“亼”这个基根。

汉字拆分的示例：

跟 → 足艮	甜 → 舌甘	榭 → 木身寸	靴 → 革亻七	鸮 → 贝贝女鸟
纸 → 纟氏	配 → 酉己	滤 → 彳卢心	散 → 艹月攵	鳍 → 鱼彡匕日
基 → 其土	尚 → 亻回	靠 → 亻口非	最 → 曰耳又	举 → 丩一八丰

美 → 羊大 券 → 类刀 商 → 宀夕冂 奖 → 丩夕大 毫 → 宀口冂毛
 扎 → 扌乚 旧 → 丨日 仍 → 亻冫丨 突 → 穴大、 藕 → 艹耒曰冂山
 鹭 → 宀口小尤鸟 凹 → 丨凵丨冂一 襄 → 宀口口圭衣
 懿 → 士冂一口彡欠心 凸 → 丨一丨冂一 赞 → 艹儿儿儿儿

示例中用到的形近根有：匕（乚）；廿（艹）；小（宀）；冂（冂）；牛（𠂇）；工（丰）；羊（羊）；半（类）；甘（圭）；氏（衣）；厶（山）。

7.1.2 对部分独体字要作变通的拆分处理

部分独体字如“丰、车、且、手、身、九、也”等，已是《郑码》的基本字根，直接输入它们的代码这些字就能上屏。其余的独体字如“久、束、史、聿”等，看做是由笔画与基根组成，或是由基根与基根组成，要按其结构作变通的拆分处理，拆分方法将在“第11章”中讲述。

7.2 第二条 拆分的汉字按基本字根的根序取码

根序是指基本字根在汉字中排列的先后顺序(包括笔画在内)。

《郑码》确定汉字根序的方法与《现代汉语通用字笔顺规范》一致。但需要说明的是根序是以字根为单位的排序，因此，与规范一致就是说《郑码》是按照规范的笔顺来确定字根的顺序，笔画先写的字根排列在前。

从字型图(44页表7)可以看出汉字根序的配置有一定的规则，虽然字型信息不参与编码，但字型图可作为确定根序的蓝图。正确认识汉字的结构和根序不但利于编码，而且能帮助我们纠正书写的错误笔顺，利于汉字认知水平的提高。

7.2.1 左右字与上下字的根序

左右字与上下字的根序很容易确定。因为，左右字与上下字的字根之间有明显的“分隔沟”，书写时，按“从左到右”、“从上到下”的顺序，写完一个字根再写另一个。因此，左右字与上下字的根序与书写顺序是相同的，与我们已有的知识背景一致，因此很容易确定。通过7.1.1节的汉字拆分举例或表7的字型图，大家能体会到这一点。

7.2.2 包围字与包孕字的根序

包围字或包孕字(即半包围字)的根序不太明确，因为书写时往往两个字根的笔画交替进行。比如写“式”字，是先写一横，然后再写“工”，最后才写“弋”的折笔和点，在确定根序时容易产生二义性。但是，用《郑码》就不会有二义性，因为《郑码》把那些在字型中起着包围或包孕作用的字根，例如“口、冂、凵、广、疒、户、弋、戈、匚、毛、讠”等全部选做了基本字根，然后以基根为单位来拆分汉字，并按笔画先写的字根排列在前的原则排列根序，这样就不会有二义性。利用这种方法还可以将包孕字视为上下字或左右字来处理。

比如将“函”视为上下字，根序就是“冫水凵”（从上到下）。这种“以简驭繁”的处理方法，简单、规范、明确，利于见字快速给码。

7.2.3 确定包围字与包孕字的根序的方法：

● 包围字的根序：

将包围字视为外内字处理，根序是“从外到内”。例如：

包围字 (囗)： 困 → 口木 园 → 口二儿

● 包孕字的根序：

将以下四种包孕字视为上下字处理，根序是“从上到下”。例如：

上包孕字 (凵)： 套 → 大扌 周 → 冫土口 阔 → 门讠舌 夙 → 儿又
 下包孕字 (凵)： 凶 → 乂口 函 → 冫水凵 画 → 一田凵 幽 → 亼厶厶山
 左上包孕字 (凵)： 虚 → 虍业 质 → 厂十贝 局 → 尸凵口 扉 → 户非
 右上包孕字 (凵)： 氧 → 气羊 贰 → 弋二贝 匆 → 勹久勹 戴 → 戠田艹八

将以下两种包孕字视为左右字处理，根序是“从左到右”。例如：

左包孕字 (凵)： 医 → 匚矢 区 → 匚乂 巨 → 匚凵 甌 → 匚车九
 左下包孕字 (凵)： 起 → 走己 彪 → 虎彡 魁 → 鬼凵十 毯 → 毛火火
 达 → 讠大 建 → 廴聿二 断 → 凵米斤 随 → 卩讠辶工月

示例用到的形近根有： 氏 (攴)； 甘 (艹)； 冫 (凵)； 凵 (聿)； 大 (ナ)； 彡 (彡)。

需要说明的是：构形有“讠”“廴”“凵”的字，根序虽然与书写笔顺不完全一致，但这种字在常用字中仅占2%左右。而且作为一种检索法，按这种根序编码可使首根是“讠”的字集中在一起，增强检索的划一性。计算机输入这类字，如“达”字按书写顺序“达→大讠”或按《郑码》根序“达→讠大”都能打出。但词语必须按“达→讠大”的根序输入。

总之，只要掌握了这一方法，即使遇到一些多单元的复合型字，不论它们的结构多么复杂，我们都能顺利正确地认识它们的根序。例如：

匈 → 勹乂口 摺 → 扌冫冬口 瞻 → 目夕厂八言 榭 → 木厂一口艹寸
 廛 → 匚户冫 凵 → 彡凵口示 髓 → 骨讠辶工月 蠲 → 艹八皿凵勹虫

7.3 第三条 单字与词语的编码不得超过4个英文字母

每个单字和每条词语的编码不得超过4个英文字母。输入的单字或词语若不足4码，要加敲空格键作为结束信号，这样，字词才能上屏。

对于基根较多的单字，若代码总数超过了4个字母，要按取码方法对基根的代码加以取舍。取舍的方法见“第8章 单字的取码方法”和“第9章 词语的取码方法”。