

# 第一章

## 汉字输入方法的基础知识

汉字输入方法简单地讲就是如何把汉字输入到计算机中。目前实现汉字输入的手段多种多样,如利用汉字识别技术、语音识别技术等,但在本书中我们所要讨论的是如何利用标准键盘来进行汉字输入,这也是最常见、使用面最广、最重要的汉字输入手段。在本书的以后章节中,如无特殊说明,则汉字输入指的就是利用标准键盘来进行的汉字输入。

标准键盘除了各种控制键外,标明用来进行符号输入的键位共有 47 个,包括每个键位的上、下档键共可直接输入 94 个字符。利用标准键盘来进行的汉字输入实际上就是通过某种规则建立汉字与字符串之间的对应关系。例如,拼音是汉字与字符串之间的最直接、最自然、最简单的对应关系,但这种对应关系并不是一一对应的。当这种对应关系建立起来以后,至于如何将码串对应到相应的汉字,又如何将相应的汉字存储到计算机中,这一系列技术问题都属于各录入软件和汉字操作系统的事,在本书中就不详细讨论了。

要掌握汉字输入方法,首先必须了解基本的计算机操作知识。另外,所有的汉字输入方法在建立码串与汉字之间的对应关系时都有各自特定的规则,而这些特定规则的建立都是基于对汉字的各种特性的认识,有的利用了汉字的形,有的利用了汉字的音。因此,读者还应该对汉字有进一步的认识。汉字输入方法作为一个专门的研究领域,当然有其基本的概念和术语,下面就逐一向读者介绍。

### 1.1 汉字编码集合

汉字历经几千年的发展演变,积累起来的总字数大约有 56000 多个。在这么多汉字中,绝大多数(约 4 万多个)已经淘汰不用,只在古籍书中才能遇到。若以现代编撰的《辞海》和《现代汉语词典》为参考,现代可能用得着的汉字(包括异体字和繁体字在内)约有 11000—14000 个。实际上,在这 1 万多个汉字中,约有一半左右在现代文献中已经罕见。1975 年和 1985 年,我国曾进行过两次规模很大的字频统计,统计范围包括时事政治、社会科学和自然科学的各个领域。统计结果表明,在几千万字文献资料中,实际用到的汉字单字总数为 6335 个。在这 6 千多个汉字中,有 3 千多个字的累计使用频度达到 99.9%,另外 3 千多个字的累计使用频度不足 0.1%。

目前我国汉字编码集都以国家标准(“信息交换用汉字编码字符集,基本集”即 GB2312-80)所收集的 6763 个汉字为准,而繁体字的汉字编码集则采用台湾制定的 BIG5 码,BIG5

码共收集繁体字 13500 余字。

通常简体汉字输入方法的编码集至少应该覆盖 GB2312-80 所收集的全部 6763 个汉字,而繁体汉字输入方法则应该覆盖 BIG5 码所收集的全部 13500 多个汉字。

## 1.2 汉字的字形特征

汉字的数量成千上万,每个字不仅寓意不同,而且形态也各异,每个字的字形结构更是繁简不一。除了极少数几个笔画最简单的字以外,绝大多数的汉字在字形上都是可以分解的,即由单字可以拆分成部件,而部件可以进一步拆分成笔画。显然,单字、部件和笔画这三者之间的构成规律应当是:由笔画构成部件、由部件构成单字。了解了汉字字形上的这种层次结构,有助于我们理解和掌握各种汉字输入方法,也为我们创造新的汉字输入方法创造了条件。

下面我们对字形结构的笔画、部件和单字三个层次来分别加以讨论。

### 1.2.1 汉字的笔画和笔顺

笔画是汉字字形结构的最低层次。通常人们根据写字的运笔方向将笔画归纳为以下五种:

横(一) 竖(丨) 撇(丿) 捺(㇏) 折(乙)

即所谓的“札字五笔”。这种归类法常见于各种字典的部首排序及一些常用的笔画编码法中。也有将汉字笔画归纳为六类的,即把“折”进一步分成“左折”(乚,又称“弯”)和“右折”(㇚,又称“拐”)两类。还有的在上述六类笔形的基础上,又增加了“叉”(十)和“方”(口)两种复合笔形,这样就构成了八类笔形。人们将笔画分得细一点有利于简化编码及减少重码。

在所有笔画归类中,通常都把“挑”(㇀)归入“横”笔,“点”归入“捺”笔。这么做是有一定道理的,首先它们的运笔方向相同,“横”和“挑”都是从左到右,“捺”和“点”都是从左上到右下;其次,它们在许多情况下常可以互相转化。例如,“土”字的末笔为“横”,但作为偏旁时末笔就变为“挑”了,类似的情况还出现在“工、王、立”等字中。

笔顺是指人们写字时的笔画先后顺序。按照一般的书写习惯,有关笔画顺序的规律大致可以归纳为以下十条:

- (1) 先上后下。如“宣”、“会”等。
- (2) 先左后右。如“理”、“作”等。
- (3) 先中后旁。如“水”、“小”等。
- (4) 先外后内再封口。如“同”、“田”等。
- (5) 先横后竖。如“干”、“王”等。
- (6) 先整后零。如“书”、“戈”等。
- (7) 先横后撇。如“不”、“夫”等。
- (8) 先撇后点。如“丸”等。
- (9) 先撇后捺。如“人”、“义”等。
- (10) 先钩后挑。如“扎”等。

以上十条只是反映了笔顺的大致规律,在实际书写时不难发现例外或不统一的情况,况且目前关于笔顺并无统一的标准。

利用汉字的笔顺属性,可以很方便地把方块汉字拆成若干基本笔画的线性排列,由此也产生了许多基于汉字笔画的输入方法,如五笔画等。另外,许多输入方法在决定部件的拆分顺序时也遵循了汉字的笔顺规律,如台湾的大易输入法等。

1.2.2 汉字的部件

“部件”又称“字根”、“字元”、“构件”等等,是汉字字形的基本构成单元。据统计,构成《辞海》全部汉字的部件总数约为 675 个,构成国标基本集汉字的部件总数约为 564 个。这些部件在组字时的使用频度差别是很大的。

大多数的汉字输入方法为了做到以尽可能少的部件组成尽可能多的汉字,都对部件进行了精心的优选和归并,将使用频度高的部件挑选出来,将不常用的部件进行归并或进一步的拆分。

所有的汉字部件从结构上来看可以分成单、散、连、交、混五种结构类型。详见表 1.1。

表 1.1 汉字部件的结构

| 部件类型  | 笔画结构方式 | 列子            |
|-------|--------|---------------|
| 单     | 单笔结构   | 一   丿 ㇏ 乙     |
| 散     | 离散结构   | 三 八 ㄥ ㄣ 心 ... |
| 连     | 相互连结结构 | 工 一 人 刀 弓 尸 白 |
| 交     | 相互交叉结构 | 十 乂 卅 扌 力 车   |
| 混 (合) | 有连有散   | 纟 彳 见 讠 ...   |
|       | 有交有散   | 寸 乂 戈 米       |
|       | 有连有交   | 土 大 木 王 禾     |
|       | 有连有散有交 | 舟 豕 酉 雨       |

从完整性来看,部件又可以分为成字部件和不成字部件两种。所谓成字部件,是指部件本身就是一个汉字,如“日”、“月”、“金”、“木”、“水”、“火”、“土”等。不成字部件本身不能单独成字,只作为一个组字的部件,如“亻”、“彳”、“讠”等。

由于汉字字形结构上的复杂性,再加上笔形变化的多样性,部件的划分和确定目前尚无一个公认的统一的标准,因此部件划分的灵活性还是比较大的。各输入方法在部件的选择上都具有各自的特点,这是读者在掌握各输入方法时需细心体会的。

1.2.3 汉字的字型特征

每个汉字从形态上来看都是一个拓朴图形。若以构成单字的部件排列位置来观察,汉字大体可以分为四种类型,即单体型、左右型、上下型和内外型,后三类又称为合体型。

举例如下:

| 字 型 | 例 字             |
|-----|-----------------|
| 单体型 | 三 月 天 电 必 重 事 乘 |
| 左右型 | 的 仆 江 结 到 湘     |
| 上下型 | 字 召 另 花 照 曼     |
| 内外型 | 囚 旭 区 司 这 网 履   |

在许多汉字输入方法中都需要对汉字进行正确的拆分。显然合体字的分解比较容易,而单体字由于其字形结构紧密,分解起来则比较麻烦。通常可以把单体字看作是已知部件的嵌套结构,分解时可按书写顺序首先抽取出包括首笔笔画在内的最大部件,然后依次类推进行拆分。

### 1.3 汉字的字音特性

汉字的音具有“声、韵、调”三种属性。大多数汉字的读音含有一个声母和一个韵母,只有极少数的字音只有韵母而没有声母。

声母是大多数汉字字音开头的音。按照我国现行拼音方案规定,汉字字音的声母共有 23 个,例举如下:

|       |       |       |      |
|-------|-------|-------|------|
| b(玻)  | p(坡)  | m(摸)  | f(佛) |
| d(得)  | t(特)  | n(讷)  | l(勒) |
| g(哥)  | k(科)  | h(喝)  |      |
| j(基)  | q(欺)  | x(希)  |      |
| zh(知) | ch(蚩) | sh(诗) | r(日) |
| z(资)  | c(雌)  | s(思)  |      |
| y(一)  | w(屋)  |       |      |

韵母是大多数汉字字音除去声母以后的音。韵母共有 34 个,根据其不同的发音特点及其在拼音中的位置,可分为单韵母、复韵母、鼻韵母和介音四类。

单韵母单纯而发音响亮,发音时必须颤动声带。单韵母共有 6 个,它们是:

|      |      |      |
|------|------|------|
| a(啊) | o(喔) | e(鹅) |
| i(衣) | u(乌) | ü(迂) |

复韵母是由两个单韵母复合而成,共有 8 个,它们是:

|       |       |       |       |
|-------|-------|-------|-------|
| ai(哀) | ei(欸) | ao(熬) | ou(欧) |
| ie(椰) | üe(约) | iu(优) | u(威)  |

鼻韵母是在单韵母后面带一个鼻音,共有 9 个:

|        |           |        |        |       |
|--------|-----------|--------|--------|-------|
| an(安)  | en(恩)     | in(音)  | un(温)  | ün(晕) |
| ang(昂) | eng(亨的韵母) | ing(英) | ong(翁) |       |

其中第一行是前鼻音,第二行是后鼻音。

介音介于声母和韵母之间,介音字母由 i、u、ü 充当。介音共有 11 个,它们是:

|         |         |        |        |
|---------|---------|--------|--------|
| ia(呀)   | iao(腰)  | ian(烟) | ang(央) |
| iong(雍) | ua(蛙)   | uo(窝)  | uai(歪) |
| uan(弯)  | uang(汪) | uan(冤) |        |

汉语的声调是指字音高、低、升、降的“调子”。汉语语音共分四个声调,它们是:

阴平声(ˉ) 阳平声(ˊ) 上声(ˇ) 去声(ˋ)

汉字中大量的字的读音都是相同的,字音相同的字称为“同音字”,除此之外还有不少多音字。

## 1.4 汉字输入法的基本概念和术语

汉字输入技术的研究自 70 年代末至今已有一二十年的历史,在这期间,国内外涌现了大量的汉字输入方法。到目前为止,仅国内登记在案的汉字输入方法就有五六百种之多。在不断的研究和发展过程中,汉字输入技术也逐渐形成了一整套的理论和规律性的东西,掌握这些对于了解和掌握各汉字输入方法都是至关重要的。下面我们就分别加以介绍。

### 1.4.1 汉字输入方法的分类及其特点

按照各汉字输入方法所依据的汉字属性的不同,大致可将它们分为形码、音码、音形码和序号码四种类型。

#### 1. 形码输入方案

所谓形码输入方案,顾名思义,就是完全依照汉字的字形来对汉字进行编码。该类方法通常依照某种规则,将汉字分解成若干笔画或部件的排列序列,以此来对汉字进行编码。属于形码的汉字输入方案很多,根据它们对字形拆分的不同层次,又可分为基本笔画编码和部件编码。

基本笔画编码是将字形拆分到预先规定的几种基本笔画,并根据对这些基本笔画的编号,将它们按一定的顺序排列成码串,以此来进行汉字输入。“五笔画”输入方案就是一种典型的基本笔画编码。

一般说来,基本笔画输入方案的规则比较简单,需要用到的码元数量(实际上就是基本笔画数量)比较少,学习和掌握比较容易。但该类方法的码串较长,且重码率较高,这就严重影响了输入效率。该类方法只适用于非专职录入人员。

部件编码是将汉字按照某种规则拆分成若干基本部件的排列序列,并以此来对汉字进行编码。在该类方法中基本部件的选择是最关键而又最困难的,基本部件过少不足以表示成千上万个汉字,基本部件过多则又会增加记忆和学习的负担。长春双凸、大易、郑码、五笔字形等输入法都属于部件编码输入法。

部件编码输入法的编码规则比较复杂,基本部件也较多,这就给学习和掌握带来了一定的困难。但平均码长较短,一般少于 4 码,输入效率较高,实用性较强,尤其适用于专业录入人员。

## 2. 音码输入方案

按照汉字的读音来进行汉字编码及输入的汉字输入方法就属于音码输入方案。常见的有拼音输入法、双拼输入法、智能 ABC 等。

该类方法的特点是学习比较容易,输入时不影响创作思绪。但输入单字时重码太多,因而影响输入效率。但若以词组输入为主,再加上智能化的输入手段,输入效率上会得到极大的改善,智能 ABC 就是一个很好的例子。另外,音码输入法的一个最大的弱点就是遇到不认识的字就无法输入了,读音不准在输入时也会遇到一些困难。该类方法只适用于非专职录入人员,也可以作为辅助录入手段。

## 3. 音形码输入方案

音形码输入方案在对汉字进行编码时既考虑了汉字的形,又兼顾了汉字的音。

音形码输入方案在学习难度、效率等方面都介于音码和形码之间,重码率也很低,几乎可以和形码相媲美。但由于在输入汉字时既要观其形,又要辨其音,因此大脑负担过重,长时间输入容易疲劳,也影响到输入速度。该类输入方法适合于非专职输入人员。近年来出现的五十字元、自然码、四笔声形等都属于音形码。

## 4. 序号码输入方案

所谓“序号码”就是将待编码的汉字集以一定规则非序以后,依次逐个赋予相应的编号数字(或字母)作为汉字输入代码。电报码和区位码就是两种最典型、最常见的序号码。

电报码又称邮电码,是广泛应用于民用通信的汉字传输代码。它以四个数码组成一组,每一组电码代表一个汉字。由于历史悠久,已经约定俗成,加之行业的特点,因此,电报码在邮电行业普遍使用。

GB2312-80 代码表中的汉字和图形符号还可以用它们所在的“行”和“列”的位置——即“区号”和“位号”来表示,例如,“啊”字被安排在表中的第 16 行第 1 列的位置上,也就是 16 区、01 位,因此“1601”就是“啊”的区位码。

由于区位码可以输入 GB2312-80 中的全部符号及汉字,因此区位码常用来作为辅助输入手段。它的装机率是非常高的,几乎所有的汉字操作系统都支持区位码。

序号码输入方案的共同特点是它们都有标准的代马表或码本,汉字与码组之间有严格的一一对应关系,在熟记常用字的代码以后,可以有很高的单字输入速度。但序号码输入方案其编码缺乏规律性,很难记忆。

### 1.4.2 码元和码长

标准键盘上全部可以利用的符号键位(包括各个键位的上、下档键)共有 94 个,利用标准键盘来进行汉字输入实际上就是按照某种规则建立汉字与这 94 个字符形成的字符串之间的对应关系,这些字符串称为相应汉字的码串。组成所有码串的字符所组成的集合称为编码集。编码集中的字符称为码元。所有码串的最大长度称为该输入方法的最大码长,所有码串的平均长度称为该输入方法的平均码长。

从理论上讲,若某种输入方法的码元总数为  $m$ ,最大码长为  $n$ ,则该输入方法的编码

空间就为  $m^n$  个,也就是说该输入方法最多可以覆盖  $n^n$  个汉字。若  $m=26, n=4$ , 那么  $26^4=456976$ , 即用 26 个码元最大码长为 4 共可对 456976 个汉字进行编码, 这个数目远远大于 BIG5 码中所含的 13500 这一汉字总数。因此, 一般来讲, 汉字输入法的最大码长为 4 就足够了。目前多数汉字输入法的最大码长都为 4 码, 只有少数输入法(如仓颉输入法)的最大码长为 5。

### 1.4.3 重码率

在对汉字进行编码时会出现若干汉字码串相同的情况, 这种现象称为重码。当在进行汉字输入遇到重码时, 需要根据屏幕提示选出特定的汉字, 因而大大降低了输入速度, 严重影响了盲打。

设编码集中汉字的总数为  $W$ , 则总的重码数  $C$  可以表示为:

$$C = \text{重码字数} - \text{重码组数}$$

故静态重码率  $P_0$  可以用下式表示:

$$P_0 = \frac{C}{W} \times 100\%$$

由上式可以看出, 静态重码率实际上是重码总数占全部编码汉字总数的百分比。

动态重码率是指在实际使用中出现重码的概率。

在本书中如无特殊说明, 重码率是指静态重码率。

### 1.4.4 汉字输入方法的输入效率

汉字输入方法的输入效率可以用每分钟所能输入的汉字总数来衡量。显然这一技术指标是因人而异的, 不可能给出一个准确的数字。

汉字输入方法输入效率的高低取决于以下三个方面:

- (1) 输入一个汉字的平均击键次数, 即平均码长。
- (2) 完成每次击键的动作所需的时间。
- (3) 从看见汉字到获得其编码的反应思考时间。

由此可见, 输入效率是由人为因素和输入方法本身两方面来决定的。

### 1.4.5 汉字输入方法的难字

所谓“难字”是指其编码不符合规定的编码规则, 需要专门记忆的字。

有一些输入方法(如仓颉输入法)为了尽量避免出现重码情况, 不得已对一些字采用非常规编码处理, 以增加记忆量来降低重码率。这些输入方法往往会专门提供一个难字列表, 供用户学习和查询, 用户需要逐一记住这些难字的编码。

### 1.4.6 简码

在汉字中有许多常用字, 其使用频率相当高。为了进一步提高这些常用字的输入效率, 各输入方法常常将它们列为简码字, 按一定规律分别赋予其简码以进一步缩短码长, 简码字又称高频字。

简码字的输入既可以利用其简码, 也可以利用其正常编码。因此不要求用户必须记忆简

码字及其简码。但由于简码通常比正常编码短,且简码字的使用频率又很高,记住这些简码字对于提高效率大有好处。

简码字通常又可以分为一级简码字、二级简码字和三级简码字,有时也分别称之为一级高频字、二级高频字和三级高频字,其使用频率通常也是一级简码字最高,二级简码字次之,三级简码字再次,其简码长度分别为 1、2、3。

简码字的输入可以直接利用其简码,由于简码不足 4 码,因此利用简码输入简码字时必须以空格键结束输入。

#### 1.4.7 模糊键

为了方便用户学习和查询,大多数输入方法都专门安排了一个键位作为模糊键,如五笔字型输入法中的 Z 键。

在进行汉字输入时经常会遇到这样的情况,即只知道码串中某些码元,而对其它码元则拿不定主意。此时,对无法确定的码元可以用模糊键来代替。例如,对于模糊键为 Z 的五笔字型输入法,如果输入 abzc,也就是说对码串中第三个位置上的码元拿不定主意,此时屏幕下部提示行内就会列出所有符合条件的汉字,这些汉字所对应码串中第一、二、四位置上的码元分别为 a,b,和 c。

再如,如果输入 abz,实际上也就是希望查询码长为 3,且第一、二个码元分别为 a,b 的码串所对应的汉字。

使用模糊键可帮助用户学习汉字输入方法;也能有效地进行个别疑难字的查询和输入。

#### 1.4.8 汉字输入方法的选择

面对众多的汉字输入方法,所有用户都希望能够选择一种适合自己的方法。在进行选择时,用户首先应考虑以下几个问题:

- (1) 希望达到的输入速度。
- (2) 准备投入多少时间和精力来学习汉字输入。
- (3) 记忆力和理解力如何。
- (4) 是否经常进行文字输入工作。

现有的编码方案可分为两大类。第一类是面向广大非专业录入人员的普及型方案,典型的有拼音、自然码、表形码、智能 ABC 等。这类方案的特点是规则比较简单,易学、易记,但输入效率不高。因受输入方法本身的限制,要想进一步提高输入效率也比较困难。第二类是面向广大专业操作员的快速型方案。这类方法以形码为主,典型的有五笔字型、郑码、层次四角、大易、仓颉等。这类方法的特点是编码规则比较复杂,重码率低,宜于高速盲打,但难于学习和掌握。熟练掌握该类方法后能够达到很高的输入效率。

“易学的打不快,打得快的不易学”基本上反映了目前我国汉字输入技术的现状,且在相当长的一段时间内还将维持下去。

图 1.1 所示为汉字输入方法的学习时间与输入速度之间的关系。

图中的拐点 a 和 b 是这类输入方法学习期和使用期的分界点, $t_1$  和  $t_2$  是学习期的时间,该参数是衡量输入方法易学程度的很重要的标志。 $g_1$  和  $g_2$  是这两类输入方案所能达到的极限速度。当然,不同的输入方法  $t$  和  $g$  这两个参数的值是不同的。

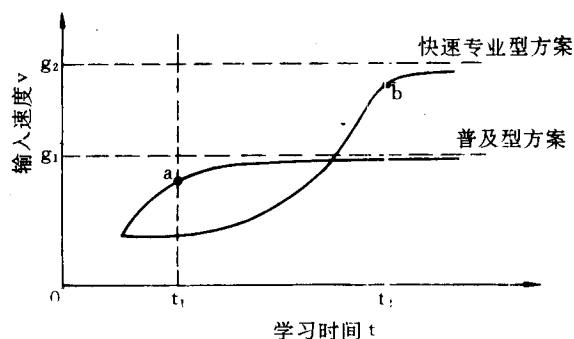


图 1.1 汉字输入方法的学习时间与输入速度之间的关系示意图

## 1.5 键盘操作要领

键盘操作的熟练程度直接影响汉字输入的效率,除了需要经常练习外,非常重要的一点是要掌握计算机操作的要领。

### 1.5.1 正确的姿势

初学键盘输入时,首先必须注意击键的姿势。如果姿势不当,就不能做到准确快速地输入,也容易感到疲劳。正确姿势的要点是:

- (1) 身体应保持笔直。
- (2) 应将全身重量置于椅子上,坐椅要旋转以便于手指操作的高度,两脚平放。
- (3) 两肘轻轻贴于腋边,手指轻放于规定的字键上,手腕平直。人与键盘的距离可通过移动椅子或键盘的位置来调节,调节到能保持正确的击键姿势为好。
- (4) 显示器宜放在键盘的正后方。在放置输入原稿前,先将键盘右移 5cm,再将原稿紧靠键盘左侧放置,以便阅读。

### 1.5.2 正确的指法

#### 1. 基准键与手指的对应关系

基准键共有 8 个,如图 1.2 所示(G,H 两键除外)。

#### 2. 字键的击法

- (1) 手腕要平直,手臂要松弛,全部动作仅限于手指部分(上身其它部位不得接触工作台或键盘)。
- (2) 手指略微弯曲,指尖后的第一关节微成弧形,除大拇指之外的 8 个手指分别轻轻地放在各基准键的中央。

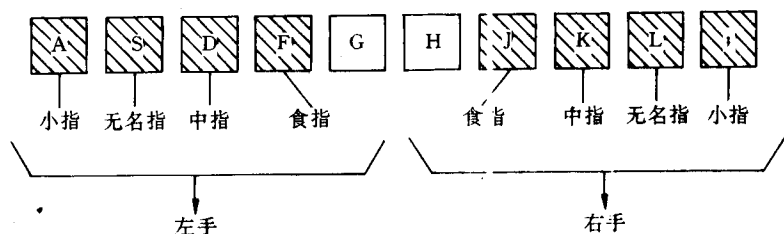


图 1.2 基准键位图

(3) 手指在击键时象弹琴,每击一次键后立即将手退回基准键位,不可停留在已击的字键上。

(4) 输入过程中击键要均匀,不可用力过猛。

### 3. 空格键的击法

右手大拇指横着向下击空格键并立即退回。

### 4. 换行键的击法

需要换行时,用右手小指击 RETURN 键,击后右手立即退回原基准键。在右手回归过程注意不要把“;”号带入。

## 1.5.3 键盘指法分区

前面我们讲了基准键位与手指的对应关系,读者必须牢记,否则,基准不对,后患无穷。

对于其它字母、数字、符号的输入都以基准键为准。键盘的指法分区如图 1.3 所示。凡在斜线范围内的字键都必须由规定的同一手指来管理。这样既便于操作,又便于记忆。

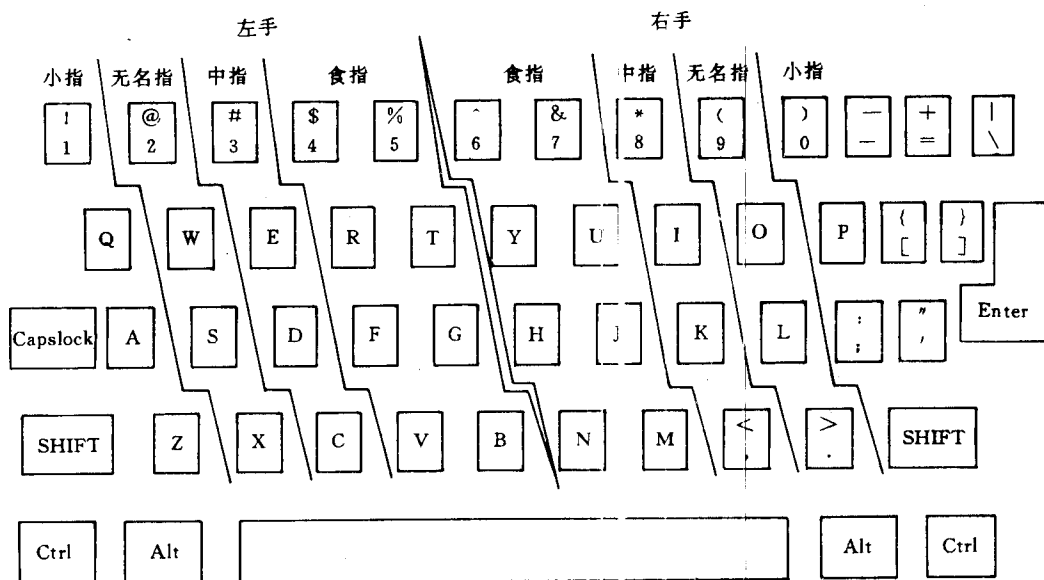


图 1.3

# 第二章

## 智能 ABC

“智能 ABC”(简称“ABC”)是长城微机通用的标准汉字输入软件。它由长城集团和北京大学合作共同开发完成,是一种以拼音为基础、以词输入为主的智能化输入法。它智能化程度高,能够自动分词,简单易学,和其它拼音输入方案相比具有较高的输入效率。为了适应各种不同类型、不同地区用户的需要,智能 ABC 除设置了标准方式、双打方式二种输入方式外,还支持全拼、简拼、混拼、全拼加笔形、简拼加笔形、混拼加笔形、纯笔形等多种输入模式,而且在标准方式下可同时使用这些输入模式,无须切换。另外,它还具有记忆功能,并且允许用户对词组进行重定义和扩充,使用起来非常方便。目前,该输入方法应用十分普遍。

方法类型: 音码

适合对象: 非专职录入人员

易学程度: 容易。只要会汉语拼音,立刻就能进行汉字输入

输入效率: 熟练以后一般每分钟平均可输入 30 个字以上

运行环境: 长城系列微机, IBM-PC 及其兼容机, 操作系统以 MS-DOS 2.10-5.00 均可, ABC 的 DR-DOS 6.0 版可在 MS-DOS 6.0 上运行

### 2.1 ABC 的安装及运行

#### 2.1.1 ABC 的安装

将 ABC 的系统盘插入 A(或 B)驱动器中,然后敲入:

A: INSTALL↵

即进入安装过程。在安装过程中,安装程序会自动识别用户显示器类型,用户不用做任何选择。

安装成功后,安装程序会自然在 C 盘上建立 CVGA24 子目录,并且会将如下的文件拷贝到该目录下:

|           |        |
|-----------|--------|
| ABC.EXE   | 执行文件   |
| ABC.OVL   | 覆盖模块   |
| ABC.CWD   | 基本词库   |
| CCLIB     | 字模库    |
| CMODE.EXE | 汉字显示模块 |

|              |              |
|--------------|--------------|
| ABC1. BAT    | 批处理文件        |
| WS. COM      | 中文 WORD STAR |
| WSMSG5. OVR  |              |
| WSOVLY1. OVR |              |

说明:

- (1) C 盘中必须有 800K 以上的空间。
- (2) 如果一次安装不成功,可以重复几次。
- (3) 如果没有硬盘,用不着安装,可直接将系统盘插入高密驱动器上运行。

## 2.1.2 ABC 的运行

在 C:\CVGA24 目录下有一个批命令文件 ABC1. BAT,用户只要敲入:

```
C>CD\CVGA24
```

```
C>ABC1
```

即可启动汉字输入模块并进入 ABC 输入状态。

为了使用方便,用户可以将 ABC1. BAT 文件拷到根目录下,并将其内容修改如下:

```
echo off
```

```
cls
```

```
cd\cvga24
```

```
cmode /e
```

```
abc
```

```
cd\
```

```
echo on
```

这样用户就可以在根目录下运行该系统了。

为了满足不同用户的需要,特作以下两点说明:

### 1. 显示模块的启动

显示模块具有如下四种启动方式:

#### (1) CMODE

在这种情况下,所有字库均装入 640K 基本内存中,此时虽然显示速度会大大加快,但由于占用了过多的系统基本内存,不可避免地会给运行 DBASE III 等应用软件造成一定的困难。

#### (2) CMODE/O

在这种情况下,所有字库均放在硬盘上,虽然这么做节省了内存,但却大大降低了显示速度,并且由于系统频繁地读硬盘,会减少硬盘的使用寿命。

#### (3) CMODE/1

在这种情况下,只将一级字库放在了内存中,在显示速度和占用内存之间作了一个折衷。

#### (4) CMODE/E

在这种情况下,显示字库全部放入了扩展内存中,如果计算机中具有扩展内存,那么这

是最好的选择。如果计算机中不带扩展内存,则用户可根据自己的需要选择前三种启动方式。

用户一旦根据自己的需要确定了 CMODE 的启动方式后,可直接修改 ABC1. BAT 文件。

## 2. ABC 的撤消

当用户想把 ABC 输入系统从内存中撤消时,不用重新启动机器,只须执行如下两个步骤:

### (1) 撤消输入模块

敲入如下命令:

ABC/Q↵

此时屏幕上即会出现如下提示:

“已经有一个 ABC 模块,撤消吗[Y/N]?”

敲入“Y”即可。

### (2) 撤消显示模块

键入如下命令:

CMODE/Q

此时屏幕上会出现提示:

“Already Has an External INT10H,Delete? [Y/N]”

此时键入“Y”即可。

## 2.1.3 ABC 的功能设置

### 1. ABC 的功能设置

|         |                 |
|---------|-----------------|
| ALT+F1  | 区位码输入状态         |
| ALT+F2  | 笔画查字(即字典)工作方式   |
| ALT+F3  | 标准变换输入工作方式      |
| ALT+F4  | 双打变换输入工作方式      |
| ALT+F5  | 英文输入工作方式        |
| ALT+F6  | 全角/半角状态切换       |
| ALT+F7  | 双打键盘布局及笔形代码提示   |
| ALT+F8  | 符号、外文字母输入(动态键盘) |
| ALT+F9  | 设备、状态控制         |
| ALT+F10 | 强制记忆功能          |

### 2. 状态及设备控制

- (1) 标点转换(开/关)      允许/不允许中文标点符号转换
- (2) 纯笔形输入(双打/标准)      纯笔形输入方式选择
- (3) 调词频(开/关)      调整词的频率,以适应不同文类的需要

- (4) 光标(开/关)      显示或隐去光标
- (5) 打印设置      设置打印字库的点阵、字体、字型和纸宽等。

## 2.2 ABC 的标准输入方式

按 ALT+F3,即可进入 ABC 的标准输入方式。在标准输入方式下,用户可直接同时使用全拼、简拼、混拼、全拼加笔形、简拼加笔形、混拼加笔形、纯笔形等多种输入模式,无须切换。下面我们对这些输入模式分别加以详细说明。

### 2.2.1 全拼输入模式

用全拼输入模式进行输入时,所有的字和词都使用其完整的拼音。下面对字和词的输入分别进行介绍。

#### 1. 单字的输入

##### (1) 用单字的全拼进行输入

| 单字 | 输入码           |
|----|---------------|
| 我  | wo            |
| 想  | xiang         |
| 女  | nv(ü 用 v 来代替) |
| 城  | cheng         |
| 震  | zhen          |

ABC 输入法用空格键表示输入码的结束,如输入“wo ”即表示代码“wo”已经输入完毕,此时屏幕下端就会出现所有拼音为“wo”的重码字列表。在屏幕下端的重码列表中,当尾部出现字符“→”时,说明下面还有内容,此时可通过“↓”键来翻看下一行的提示内容;当尾部出现字符“!”时,说明提示内容已经结束;还可以通过“↑”键来向上翻看提示内容。当找到需要的单字时,可通过按相应的数字键来完成对该字的输入。

在所有输入模式下,都可遵循以上操作来进行输入,即:

- ① 空格键表示码输入结束。
- ② 通过“↑”和“↓”两键来上、下翻屏查找重码字或词。
- ③ 选相应的数字完成输入。

##### (2) 以词定字的单字输入法

用全拼输入单字重码很多,不容易查找。但如果用与之相联系的词来输入该字,则输入速度要快得多,因为双字词或多字词的重码率要低得多。

以词定字的方法是:在词输入码的末尾加上“↑”或“↓”。“↑”表示取词的前一个字,“↓”表示取词的后一个字。例如“键盘”一词的输入码为“jianpan”,因此,“jianpan ↑”就会输入“键”

字,而“jianpan ]”则会输入“盘”字。

2. 词组的全拼输入

将词组中每个单字的全拼合在一起就构成了该词的全拼输入。

| 例词  | 全拼输入码       |
|-----|-------------|
| 计算机 | jisuanji    |
| 长城  | chang cheng |
| 金沙江 | jinshajiang |
| 西安  | xi'an       |
| 单个  | dan'ge      |

在以上例子中使用了隔音符号,这主要是为了有助于进行音节划分,避免二义性。例如,既可以将“xian”理解成“献”的全拼,也可以将其理解为“xi an”(西安的全拼),因此,我们借助于隔音符号“'”以使得音节划分更加明确。

全拼输入模式虽然码长较长,但出现重码的机会非常少。

2.2.2 简拼输入模式

简拼输入模式只适合于输入词组。简拼码的取码规则是:取组成词组的每个单字拼音的第一个字母依次排列所组成的码串来作为该词的简拼码,对于声母为 zh,ch,sh 的单字,也可以取前二个字母。

| 例词   | 全拼码                                           | 简拼码              |
|------|-----------------------------------------------|------------------|
| 计算机  | j <u>i</u> suan <u>j</u> i                    | jsj              |
| 长城   | <u>c</u> hang <u>c</u> heng                   | cc,cch,cbc,ch ch |
| 中华   | <u>z</u> hong <u>h</u> ua                     | z'h,zhh          |
| 愕然   | <u>e</u> <u>r</u> an                          | e'r              |
| 和风细雨 | <u>h</u> e <u>f</u> eng <u>x</u> i <u>y</u> u | hfxxy            |

在简拼输入模式下,进一步扩大了隔音符号的作用。在以上例子中,“中华”的简拼应该是“z'h”或“zhh”,若取“zh”就错了,因为我们无法分辨它是双声母“zh”还是二个字母“z”和“h”的简拼,因此,使用了隔音符号来帮助系统进行识别。

简拼输入模式虽然码长很短,但出现重码的情况非常多。

2.2.3 混拼输入模式

所谓的混拼输入模式,就是指在根据组成词的每个单字的拼音来对该词进行编码时,有的字取得是其全拼,而有的字取得是其拼音的第一个字母或完整声母,即在词的编码中既有简拼,又有全拼。

| 例词  | 混拼码                  |
|-----|----------------------|
| 金沙江 | jīn s j              |
| 历年  | lǐ'n, lnian          |
| 单个  | dān'g, dge           |
| 计算机 | jisj, jsuanj, j s ji |

在以上例子中,使用了隔音符来帮助系统进行音节划分。当我们对词中某个字的拼音拿不准,而只能肯定它的声母时,就可以采用混合输入模式进行输入。

### 2.2.4 纯笔形输入模式

为了弄清楚该输入模式,首先必须明白 ABC 对笔形的一些规定。ABC 将汉字的基本笔画按其形状共分为如下八类:

| 笔形   | 笔形名称   | 实例  | 编码 |
|------|--------|-----|----|
| 一(丿) | 横(提)   | 工,政 | 1  |
| 丨    | 竖      | 同,少 | 2  |
| 丿    | 撇      | 担,斤 | 3  |
| 丶(丶) | 点(捺)   | 写,忙 | 4  |
| 乚(乚) | 折(竖弯钩) | 刀,队 | 5  |
| ㇏    | 弯      | 么,以 | 6  |
| 十(乂) | 叉      | 也,希 | 7  |
| 口    | 方      | 国,是 | 8  |

只要记住了以上8种笔画及其编码,即可对汉字按纯笔形模式进行编码。编码的总原则是:按照笔顺(即汉字的书写习惯)最多取6笔,汉字笔画不足6笔的,有几笔取几笔;多于6笔的,将多余笔画省略。

下面分三种情形加以说明。

#### 1. 简单汉字可顺序取码

简单汉字即指独体字,可按笔画顺序一一取码,如:

小 534    力 53    串 882  
 事 1851    我 315    重 3781  
 舟 33514    兼 4315    乎 34315

#### 2. 合体字的取码

合体字即指上下、左右或内外结构的字。对于合体字,可将其分为上下、左右或内外两部分,分别对其进行编码。若第一个字块不足三码时,第二个字块可顺延取码,第二个字块仍可

一分为二,按每部分取三码顺延取码。

第一个字块多于三码的例子,如:

筠 314 713    船 335 36    装 412 413  
氧 311 431    肇 451 57    飒 414 467  
敲 418 217

第一个字块不足三码的例子,如:

传 32 1154    薛 72 358 4  
蓟 72 358 2    国 8 1714  
花 72 323    做 32 78 3

### 3. 一些特例

对于一些特殊的偏旁部首及字,为了避免二义性,硬行规定了如下编码:

耳    非    卄    火    女    卅    卅    开    井    弗  
122    211 424    433    631    72    132    1132    1132    51532  
凸    凹  
25    26

纯笔形输入模式只适合于输入单字,和其它输入模式相比,输入速度较慢。但对于不懂拼音者来说可选用该输入模式。另外,对于个别不知道读音的字,也可选用该输入模式。

在刚进入 ABC 时,系统自动将纯笔形输入模式设在了双打方式下,如果要使标准变换也具备笔形输入功能,必须先按 ALT+F9,改变纯笔形的输入状态设置。

## 2.2.5 拼音+笔形输入模式

拼音加笔形输入模式可分为全拼加笔形、简拼加笔形、混拼加笔形三种输入模式,加笔形特征的目的是为了尽量减少重码。

音形组合的一般形式如下:

(拼音+[笔形描述])+(拼音+[笔形描述])+...+(拼音+[笔形描述])  
|-----N-----|

其中 N 为词长,其值可以是1—9。“拼音”既可以是全拼,也可以是简拼。“笔形描述”可有可无,并且笔形描述的码长可以为1—6。

| 例词 | 全拼码      | 全拼加笔形     | 简拼加笔形 | 混拼加笔形 |
|----|----------|-----------|-------|-------|
| 预言 | yuyan    | yu5 yan   | y54y4 | yu5y4 |
| 市场 | shichang | shichang7 | sc71  | shi4c |

## 2.3 自动构词过程以及人为干预

ABC 是以词为主的输入方法,但汉语的词不但数量繁多,而且千变万化,致于人名、地名等不成其为词的专用名词就更是数不胜数。ABC 为我们提供了自动构词手段,一旦构词