

## 第五章 词典的生成和修改

### 5.1 系统配置的词典

#### 5.1.1 系统的词典文件名

在系统硬盘的 C:\LXPC 目录中, 可以看到这样一些文件名:

DICT1.TBL	-----	
DICT2.TBL	↑	
DICT2.2ND		编码词典文件 (代
DICT3.TBL		表汉字输入方案)
.....		
DICT12.TBL	↓	
-----		
LX.EXE		联想字典文件
CZ.EXE		联想词典文件

这些文件都是系统提供的词典文件, 它们的文件名是法定的, 不可随意更改。

编码词典文件名有统一的规定格式: DICT<sub>n</sub>.xxx

其中  $n$  = 词典文件所对应的输入方案的编号,  $n$  的有效值为 1...19。(各输入方案的编号可以从“输入方案配置表”查到)

xxx = 文件附加名。

因为输入方案的大小不同, 因此与它对应的词典文件的个数也可能不同。

若某方案只对应一个词典, 则 xxx = TBL;

若某方案有多个分词典, 则各分词典的文件名相同 ( $n$  值相同), 附加名不同。附加名最多可以有 6 个, 顺序为: TBL、2ND、

3RD、4TH、5TH、6TH。

例如：“拼音”方案的方案编号为 2，它有二个分词典，文件名为 DICT2.TBL 和 DICT2.2ND

### 5.1.2 词典文件的作用

输入汉字时，提示区显示的汉字都是从词典文件中取出来的：

编码字取自字编码词典  
编码词取自词编码词典  
联想字取自联想字典  
联想词组取自联想词典

如果你不满意系统提供的词典，可以自己生成新词典，也可以修改原有词典，以适合你的需要。

## 5.2 词典生成

### 5.2.1 词典源文件

如果要生成新词典，必须首先按下列规定格式编辑词典源文件。

#### 1. 编码词典源文件格式

##### 1) 字编码格式

每行的左边是单个汉字，右边是编码，汉字与编码之间空一格。允许一个汉字对应多个码（各码之间以空格隔开），也允许多个汉字对应一个码。

例如:                   啊 a  
                              阿 a  
                              行 xing hang  
                              .....

由这种格式的源文件生成的词典称字编码词典或字编码输入方案。(字编码也可以采用词编码格式,目前机器采用词编码格式。)

## 1) 词编码格式

每行左边是编码,右边是词(词由一至多个汉字组成),编码与词之间空一格。允许一个编码对应多个词(各词之间以空格隔开),也允许多个编码对应一个词。

例如:                   rcsmln 人民  
                          people 人民  
                          ta 他 它 她 踏 塔  
                          ylyl 意义 一亿 意译 疑义 一一  
                          1985 一九八五年  
                          kxy 科学院  
                          .....

由这种格式的源文件生成的词典称词编码词典或词编码输入方案。

## 2. 联想字典源文件格式

每行左边为联想引导词(由1到4个汉字组成),右边是该词的联想字,引导词与联想字之间空一格,联想字之间不空格。

例如：

我 们 国 家 的 军 方  
时 间 候 差 代 期 事  
控 制 论 流 机

.....

每个引导词最多可以联想 128 个汉字。

由这种格式的源文件生成的词典称**联想字典**。

### 3. 联想词组词典源文件格式

每行左边为联想引导词（由 1 到 4 个汉字组成），右边跟着它的一串联想词，引导词与联想词之间、各联想词之间都要空一格。

例如：

汉 字 信 息 处 理 系 统 语 拼 音 字 输 入 方 案  
国 家 元 首 际 形 势 民 经 济  
中 国 科 学 院 人 民 大 学 银 行  
联 合 国 秘 书 长 安 全 理 事 会

.....

每个引导词可以联想的词总长度达 128 个汉字，每个联想词的最大长度为 26 个汉字。

由这种格式的源文件生成的词典称**联想词组词典**。

## 5.2.2 词典生成的基本步骤

### 1. 编辑源文件

调编辑程序(如PE), 按规定格式编辑词典源文件。

## 2. 生成词典

调主清单运行DICT-GEN程序(见第二章2.3.4节)。

DICT-GEN是一个通用词典生成程序, 它可以将各种类型的词典源文件生成为对应的新词典, 并将新词典以文件名DICT.TBL存盘。

如果生成的是编码词典, 而且源文件非常大, 生成后可能会得到若干个分词典, 分词典的个数取决于源文件大小, 最多可以有6个分词典。生成后按顺序以DICT.TBL、DICT.2ND、DICT.3RD、DICT.4TH、DICT.5TH、DICT.6TH等文件名分别存盘。其中分词典DICT.2ND的容量为56K, 其余分字典为64k。

联想字典和联想词典的容量不能超过56KB, 因此没有分词典。

## 3. 新词典改名

新词典的文件名(DICT.TBL)不是系统法定的词典文件名, 不被系统认可, 因此必须改名。

编码词典: 将 DICT.TBL 改成 DICT<sub>n</sub>.TBL  
其中,  $n=1, 2, \dots, 19$

如果生成了多个分词典, 各分词典要逐一改名: 将DICT 改为DICT<sub>n</sub> (各分词典取相同的  $n$  值), 附加名不变。

编码词典改名后即装入“汉字输入方案配置表”，调主清单运行METHODS 程序可以看到。

联想字典：将 DICT.TBL 改成 LX.EXE

联想词典：将 DICT.TBL 改成 CZ.EXE

注：由于系统中已有联想字、词典，因此新的联想字、词典改名前必须先系统将原有的字、词典改名。

#### 4. 新词典装功能键或装汉卡

任何词典（包括编码词典和联想字、词典）都必须装入汉卡才能使用。

编码词典在使用时按功能键 Alt-F1..F5 调入汉卡（代表同一输入方案的所有分词典将同时调入）。新生成的编码词典改名时若取 n=11..19，使用前须修改方案配置，将新方案装入功能键。

联想字、词典在执行 LDNEW命令时调入汉卡（系统启动时自动执行 LDNEW，因此启动后联想字、词典即装入了汉卡）。

新生成的联想字、词典虽经改名但仍留在磁盘上。使用前必须执行 LDNEW命令（或重新启动系统），将它们装入汉卡。

注：大型编码词典源文件的另一种生成方法

为了便利于词典修改和维护，对于大型编码词典

源文件，也可以在编辑时将它分割成几个小型源文件，然后逐个调DICT-GEN程序将它们分别生成小词典。

但是，为了使这些分别生成的小词典能够代表同一个输入方案，必须在生成过程中输入相同的方案名称，相同的编码特点，并且在回答屏幕询问“编码表是否还有续表”时给出各个小词典之间的链接关系。这时，生成程序才能知道这些小词典是代表同一输入方案的，它会按生成顺序自动给它们命名：DICT.1ST、DICT.2ND、DICT.3RD... DICT.6TH。

这些词典生成后也要按上述方法分别改名、装功能键，用时按功能键便可将它们同时调入汉卡。

### 5.2.3 词典生成举例

#### 1. 生成编码词典

例题：生成一个新的编码方案，取方案名为“学习”，源文件名为CODE，源文件内容为：

- a 北 南 天 地
- b 大 学 工 学 院 文 学 院
- c 方 极 京 安 门
- d 汉 满 蒙 回 藏 高 山
- e 族 民

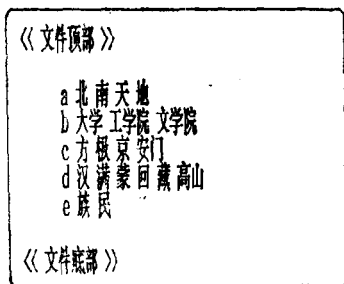
1)用TE程序编辑源文件CODE

首先在提示符 C:\LXPC> 后键入命令：

```
pc code <Enter>
```

回车后出现 PE 编辑屏幕。

在 PE 的文本区写文件内容：



最后，按 F3 键并回车，将文件存盘。

## 2) 将源文件生成为词典

调主清单运行 DICT-GEN 程序。

命令执行后，屏幕提示：

请输入生成的词典类型：(1...3)  
 1-编码词典 2-联想字典 3-联想词组词典

按： 1 <Enter>

屏幕提示：



请输入词典源文件名:

在 ASCII 或 (ASC) 输入方式下按:

code <Enter>

屏幕提示:

请输入生成的方案名称  
(用二个汉字或四个 ASCII 符表示)

在汉字输入方式下按: 学习 <Enter>

屏幕提示:

请输入码元类型: (1..5)

- 1-码元为 10 个数字: 0...9
- 2-码元为 26 个字母: a...z
- 3-码元为 26 个字母加上 10 个数
- 4-码元为右边小键盘 (按下 NumLock 键)
- 5-码元为 47 个下档字符

按: 2 <Enter>

屏幕显示:

编码方案是否处理词? (1...2)  
1--按词编码 (词由一个或者多于一个汉字组成)  
2--按字编码

按 1, 生成词编码输入方案;

按 2, 生成字编码输入方案。

源文件是词编码格式, 按: 1 <Enter>

屏幕显示:

编码输入到最大码长且无重码时是否自动输入对应的词  
1--是 (自动输入唯一的词)  
2--否 (必须按空格键或大写字母键才输入对应的词)

“自动输入”指按完编码以后, 所需的字 (或词) 能自动输入到屏幕光标处。但是要求输入方案的码长一致, 且重码少 (遇重码不能自动输入)。

按: 2 <Enter> (若选 1, 屏幕还将提示你输入最大码长。)

屏幕提示:

编码表是否还有续表? (1...2)  
1--只有一个表, 或者已经是多个编码表的最后一个表  
2--编码表还有续表 (编码表不止一个, 并且当前处理的不是最后一个表)

“编码表”指编码词典的源文件，“多个码表”指由一个大型源文件分割而成的多个小源文件。如果只有一个源文件，不管其大小，都按 1；如果源文件分割成了几个小文件，在前几个小文件生成时按 2，在最后一个文件生成时按 1。

例题仅一个源文件，按： 1 <Enter>

屏幕提示完毕，开始生成词典。屏幕显示：

生成: DICT.TBL

表示开始生成词典文件 DICT.TBL。  
生成完毕，自动返回操作系统，屏幕显示系统提示符：

C:\LXPC>

此时，在 C:\LXPC 目录中增加了一个文件名为 DICT.TBL 的编码词典。

### 3) 词典改名

将新词典改名为 DICT13.TBL。

在提示符 C:\LXPC> 后键入命令：

ren dict.tbl dict13.tbl <Enter>

命令执行后，系统中增加了一种新的汉字输入方案“学习”，它的方案编号=13（调主清单运行

METHODS 程序可以看到)。

#### 4) 词典装功能键

将新词典装入功能键 Alt-F4。

调主清单运行 METHODS 程序, 将新词典装入功能键 Alt-F4。

此后, 按 Alt-F4 键就能使用新方案。

## 2. 生成新的联想词组词典

例题: 生成一个联想词组词典, 取源文件名为 LXCD,  
内容为:

计 策 谋 量 器 算 机 算 技 术 程 车  
电 子 工 业 部 对 抗 信 息 协 会  
科 学 院 计 算 所 物 理 所 化 学 所

### 1) 编辑源文件

在提示符 C:\LXPC> 后键入命令:

pe lxcd <Enter>

在编辑屏幕上写文件内容:

```
<< 文件顶部 >>  
计 策 谋 量 器 算 机 算 技 术 程 车  
电 子 工 业 部 对 抗 信 息 协 会  
科 学 院 计 算 所 物 理 所 化 学 所  
<< 文件底部 >>
```

按 F1 键后回车，将文件存盘。

## 2) 生成词典

调主清单运行 DICT-GEN 程序。  
命令执行后，屏幕提示：

请输入生成的词典类型：(1..3)  
1-编码词典 2-联想字典 3-联想词组词典

按： 3 <Enter>

屏幕提示：

请输入词典源文件名：

按： lxcd <Enter>

开始生成词典。屏幕显示：

生成：DICT.TBL

生成完毕，自动返回操作系统。

此时，在 C:\LXPC 目录中有了一个名为 DICT.TBL 的联想词组词典。

### 3) 词典改名

要将 DICT.TBL 改名为 CZ.EXE，须在 C:\LXPC> 后键入二条命令：

```
ren cz.exe cz.bak      <Enter>  
ren dict.tbl cz.exe    <Enter>
```

命令执行后，在C:\LXPC目录下有了一个新的联想词组词典 CZ.EXE。

### 4) 词典装汉卡

在提示符 C:\LXPC> 后键入命令：

```
ldnew f      <Enter>
```

命令执行后，新词典装入汉卡，可以立即使用。  
(也可重新启动系统，将新的联想词典装入汉卡)

注：联想字典的生成过程与联想词典的生成过程是相同的，这里不再举例。

## 5.3 词典修改

若要修改系统中已有的词典，有两种方法：

1. 利用系统提供的通用词典还原程序 (DICT-TAB.EXE) 修改词典。这种方法通常用在需要大量修改的场合。
2. 利用系统的自学习功能 (Alt-0) 修改词典。这种方法经常用于少量修改或现场修改。

### 5.3.1 利用还原程序修改词典

#### 1. 基本步骤

1) 运行 DICT-TAB 程序将系统中的词典还原成源文件, 并将源文件存盘。

注: 系统中的“国标”和“区位”方案不能使用 DICT-TAB 程序还原。

2) 调编辑程序 (如 PE) 修改源文件, 再将修改后的源文件存盘。

3) 运行 DICT-GEN 程序将修改后的源文件重新生成词典, 然后将新词典改名、装汉卡或功能键。

下面只讨论如何将词典还原为词典的源文件。

## 2. 词典还原举例

各类词典 (包括编码词典、联想字典和联想词典) 的还原过程完全一样, 因此下面仅举一例说明。

例题: 将系统提供的“英中”方案还原成源文件, 用文件名 ENGLISH 存盘。

首先调主清单运行 DICT-TAB 程序,

命令执行后, 屏幕立即显示:

请输入词典文件名:

输入要还原的词典的文件名。

注意: 必须使用系统规定的词典文件名。因为“英

中”输入方案装在功能键Alt-F3上，所以方案编号为3，词典文件名为 DICT3.TBL。

按入：                    dict3.tbl                    <Enter>

屏幕提示：

选择： 1.显示码表  
         2.码表存盘

若选择“显示码表”，可以立即在屏幕上看到码表，但是只显示一次，不能保留。

若选择“码表存盘”，还原以后码表将直接存盘，不能立即显示。

按入：                    2                    <Enter>

屏幕提示：

请输入文件名：

要求输入码表存盘的文件名。（可以随意取名）

根据例题要求，按入： english                    <Enter>

还原后，在C:\LXPC目录中增加一个新文件ENGLISH，它就是“英中”词典的源文件。



如果输入方案有多个分词典，则需多次使用还原程序将每个分词典分别还原，分别存盘，才能得到完整的词典源文件。

### 5.3.2 利用自学习功能修改词典

进入汉字系统以后，只要按Alt-0键，随时都能修改系统配置的各类词典。

#### 1. 要点说明

- 1) 按一次Alt-0键，只能增加或减少一个字或一个词。
- 2) 用Alt-0增加的字（或词）显示在提示区最前面。
- 3) 按Alt-0键修改的是汉卡上的词典，修改后新内容可以马上使用。
- 4) 要想保留修改结果，须在关机前按Alt-9键。

#### 2. 修改编码词典举例

例题：在“拼音”方案中增加一个词：

用编码:ki表示“中国科学院计算所”

操作步骤：

按Alt-0键

(进入词典修改程序)

按:3

(修改编码词典)

按:1

(增加词)

选择一种ASCII输入方式