

中文字根式輸入之簡化

A Simplifications for the Input of the Chinese Radical System

許桂敏 Kuei-Min Hsu

Department of Computer Science

(Received January 19, 1976)

ABSTRACT — According to the Chiao-Tung Radical System [1,2], a Chinese character is represented by a component sequence, which consists of radicals and operators. In this paper, we will find that a Chinese character can be represented solely by a radical sequence. From a given radical sequence, by using the algorithm described in this paper, the component sequence can be constructed by generalizing the required operators and inserting them between radicals. This way of representation will improve the input-speed more than 1.58 times. In addition, it is easier for the user to type-in a correct radical sequence than a component sequence.

摘要——交大曾提出以 499 個字根，配合九種結合符號，可將 48,713 個中文字應用於電子計算機(1,2)。基本上每二字根均應有結合符號連接(1)，但經過適當處理，則可將定位符號省去近 93%(4)。

現本文再就使用之結合符號，作更進一步的探討。根據 9,060 個基本用字(基本用字 8,533 個，再加參照字形 527 個)分析各字根所接用之結合符號，知可利用計算機之程式設計，將結合符號全部省去。當機器組字時，只要收到連續之字根，即可將結合符號自動補入字根間或字根之最前與最後，因而得到造字的字根結構式。

由統計結果知：平均每一中文字使用 1.98 個字根及 1.16 個結合符號(4)。故如省去結合符號時，平均每字只按鍵 1.98 次，則中文輸入之速率將增加 1.535 倍，且省去結合符號，由計算機程式處理，可得標準之結構式，因而中文在計算機處理上將方便甚多，而有助於電子計算機之文化。

一、前言

中文字本身即具有優美之外形結構，故在利用電子計算機處理中文資料時，除非當作圖案，否則實難趨於理想，其主要之困難在中文資料輸入輸出受相當大的限制，若將每一字視為圖案，則必佔用相當大量的計算機儲存空間才能容納夠多的中文字，因而無法與英文之拼字方式相比，但交大提出之中文字根式處理方法却可補足相當大的缺點。其方法為由中文字之結構原理析出 499 個字根(1,2)，按其字形之比重組合字根，再形成中文字。此種方法可組合 48,713 字，幾含所有使用之中文字，而其優點為不需將每字存入計算機中，如此可節省極大的儲存空間，因而適用於小型計算

機，此正適合國人購置之機種。且其字體之組合方法，在使用上與書寫方法幾乎完全一致，故相當實用。但由字根之組合，必須有結合符號配合使用。

在組字方面除使用 499 個字根外，結合符號共有九種，分別為：

1. 左右括號二種：決定結合之優先順序。
2. 定位符號三種：橫連 ($\triangleleft$)，直連 ($\triangleup$) 及包含 ($\triangle$)。
3. 方便符號四種：品形 (ϕ)，炎形 (1)，林形 (2) 及 《》 形 (3)。

此三類中以第二類定位符號使用最多，關於定位符號之使用有以下兩點值得考慮：

(一) 某些字根在組成時有固定之位置，如此則將使用固定的結合符號與之配合。如「亻」字根必為接橫連之符號，而其後無法用直連的定位符號，「宀」則必跟直連符號，而不能跟橫連符號。又如「乚」其後必跟包含符號，諸如此類之字根，常跟固定之定位符號，如此其後使用定位符號實顯多餘。

(二) 某些字根之組合，對某字而言，常隨人之主觀而異。如「祭」，視「示」在「夂」之內，則為「夂」後用包含符號，如視「示」在「夂」之下，則為「夂」後用直連符號。又如「飈」亦可寫作「風台」，如此，前者為風包含台，而後者却為風橫連台。此項差別影響字形美觀甚大，而且亦會給軟體造字設計工作增加相當的困擾。

因此，如將其結合符號省去，由設計系統程式時加上，在使用時只以字根送入，如此與一般書寫法則更為接近。對電子計算機之中文化，必有很大的助益。

二、字根與結合符號之分析

本節以 9,060 個中文字為基礎，分析字根與結合符號之間的關係。

根據統計，平均每一中文字約含 1.98 個字根及 1.16 個結合符號。在結合符號中，以定位符號最多約佔 90.4%，左、右括號佔 6.3%，剩下方便符號也只有 3.3% (4)。

如果一中文字平均有 1.16 個結合符號，而只有 1.98 個字根，則在字根結構式中，兩字根間之結合符號，多為二個以上組合運用的。此種組合之結合符號，實可當作一種新的結合符號。

經由 9,060 個字之結構式之統計，得到全部結合符號，共有 53 種 (見附錄一)。其中 (, ((, 二個必為用於全字根式之最前者，而用於最後者共有 14 個，即),))， ϕ , 1, 2, 3,) ϕ ,)1,)2, 1), 2), 21, 3) 及 2) ϕ 。故在整個結構式中，實際上有三十七種結合符號是用在兩個字根間的。

利用此三十七種新結合符號，補入字根間，則每兩字根間必為只有一個結合符合，其結構式除最前之括號及最後之結合符號外，其餘必成字根與結合符號交替出現，最後仍為字根之規則形式，亦即 n 個字根必有 $n-1$ 個結合符號。

下面舉出三個例子，分別說明之：

(1) 劇：

劇字之結構式為

𠂔△(一△𠂔)△𠂔。

此處△(，)△△即為新的結合符號，在此式中有𠂔，一，𠂔，𠂔四個字根，故必有三個結合符號加入字根間即△(，△，)△△。

(2) 關：

關字，其結構式為

門△(幺2△𠂔)

此處字根有門，幺，𠂔三個，而結合符號有△(，2△，)三個，其中“)”為最後之括號。

(3) 𠂔：

𠂔之結構式為

(白△水)3。

有白，水二字根，有(，△，)3三個結合符號，其中“(”為最前者，而“)3”為最後者。

三、字根與字根間結合符號之分析

在字根中有許多只跟一種定位符號，但經組合後會產生兩種以上的結合符號。如“イ”字根只跟△△，但“傑”字之結構式為

イ△△(夕△△𠂔)△木

因而“イ”之後除△△結合符號外，尚有“△△(”之組合符號。又“𠂔”之後應跟△。如“迫”為𠂔△白，但如“遑”則為𠂔△(白△王)又產生一種△(之組合符號。

因此如前節所述，將字根間之全部結合符號，視為一種新的結合符號，則對新的結合符號而言，二字根間只有一個結合符號。兩字根間之結合符號可能是惟一的，如虫與心為△[惠]，尸與目為△[眉]，乘與𠂔為△△[刺]，有些是兩個的，如矛與木可為△△[楸]亦可為△[柔]，木與口可為△△[楫]亦可為△[杏]，厶與羽可為△△[翹]，亦可為△[翁]。也有三個的，如山與人即有△△[嶺]，△[峇]及△([嶺])三種。

對口與木而言則更有△△[啞]，△[呆]，)△[架]及0△[澡]等四種。

由此可知字根與字根間使用之結合符號可能是很多種的。

四、結合符號省去之可能性分析

本節首先將討論字根間之結合符號欲省去的可能性。

由前節知：由組合而成的結合符號共有53種，而在兩字根間者，共有37種。綜

觀全部 499 個字根，有 114 個後面不再接字根，亦即是可獨立使用或作字尾之字根。只跟一種結合符號的有 169 個，跟兩種的有 128 個，跟三種的有 54 個，跟四種的有 18 個，跟五種的有 8 個，跟六種及七種的分別有 4 個及 3 個，最多的是跟十種結合符號，只有一個（見附錄二）。

為了便於說明，將以上字根再分成四類。第一類即不跟字根或僅獨立使用者共 114 個。第二類為只跟一種結合符號的，有 169 個，其他跟兩種及兩種以上的區分兩類：即第三類為跟兩種及兩種以上的結合符號，但每種結合符號均接用不同的字根，此類字根共有 152 個。剩下尚有 64 個均為不同之符號可接用相同之字根此為第四類（見附錄三）。

以下將分類討論其省略結合符號之可能性。

第一類實際上並無結合符號之可能，但這是因其後無字根，故後面仍須在結尾符號中討論之。此處如只討論字根間的結合符號，則可不必處理。

第二類，因其只限接用一種結合符號，故如省去時當可很正確的將其結合符號補入。如“肩”必跟△△，“才”只跟△“卩”只跟△△等等，當字根送入為肩及支時，很顯然可補上△△因其不跟其他結合符號之故。同樣如送入字根才及子，則其間必須使用△之結合符號，故此類處理亦無問題。

第三類，因一字根可能跟不同之結合符號，但一字的結合符號必跟某些固定的字根，故此種只要將每一結合符號所能跟的字根列表出來，即可查對，也就是說，當我們送入沒有結合符號之字根式時，可由字根間之關係利用列出的表而決定使用的結合符號。

如“冂”字根其可跟之結合符號有△及△（兩種，但接用△者為“口”，“人”，“小”，“𠔁”，“古”及“卅”；而接用△（者則為“禾”，“土”，“士”，“、”，“母”，“一”及“入”。兩者之字根均不相同，故送入口及𠔁必為冂△，而送入口及士，則必為冂△（士。故此類字根當可省略其結合符號。

根據以上的分析可知有 435 個字根均可很容易的完成省去的工作，下面再討論剩下的第四類。

第四類的字根，因其後接用的字根並不能唯一決定所跟之結合符號，但經統計結果可以看出，極大多數均可由其前後之相關字根決定之。茲以可接用七種結合符號的字根“儿”為例說明之。

“儿”後可跟之結合符號有七種，分列如下：

- (1) △△：跟卩（元），女（微），丩（元），一（輝），頁（頑），及羽（耀）

共六個字根。

- (2) △△：跟龜（龜）一個字根。

- (3) △：跟厶（羗），、（右下點）（寬）共二字根。

(4) $\triangleleft$ ：跟木(樂)一個字根。

(5) $\triangleleft$ ：跟寸(尅)，丿(尅)，力(勉)，厂(𡵓)，虫(虺)，羽(翹)，毛(彪)共七個字根。

(6) $\triangleleft$ ：跟一(尅)，人(尅)，丿(上撇)(尅)，丿(寇)共四個字根。

(7) $\triangleleft$ ：跟臣(𨾏)一個字根。

其中丿與羽均可與儿配合接用 $\triangleleft$ 及 $\triangleleft$ ，儿與丿之間取 $\triangleleft$ 則為刂，取 $\triangleleft$ 則為刂，儿與羽之間取 $\triangleleft$ 則為耀，取 $\triangleleft$ 則為翹，如此儿後雖可接用七種結合符號，但由字根序列仍可決定其結合符號。

同樣可以接用七種結合符號的字根虫比儿更難，其中配合字根日，西，丿(上撇)可接用 $\triangleleft$ 及 $\triangleleft$ ，配合田，厶可接用 $\triangleleft$ 及 $\triangleleft$ ，而配合匚亦可接用 $\triangleleft$ 及 $\triangleleft$ ，但我們知道：

1. 虫 $\triangleleft$ 日只有蜴一字。

虫 $\triangleleft$ (日亦只一字蜴。

2. 虫 $\triangleleft$ 田共有螺，蛸，螺三字。

虫 $\triangleleft$ (田只有蛸一字。

3. 虫 $\triangleleft$ 厶只有蛭。

虫 $\triangleleft$ (厶只有蛭。

4. 虫 $\triangleleft$ 丿(上撇)只有蛭。

虫 $\triangleleft$ (丿(上撇)只有蛭。

5. 虫 $\triangleleft$ 匚只有蠃。

虫 $\triangleleft$ (匚則只有蠃。

得知如此情況及其他字根均只配合接用一種結合符號，則亦可由字根序列決定其結合符號了。

再如“人”，“口”之間亦可接用“ $\triangleleft$ ”及“ $\triangleleft$ ”兩種結合符號，而其組合字的個數各為 27 及 19。故如在兩字根間欲找出正確的結合符號，可將該等組合字列出，由其前後字根之關係，即可決定。

以上的討論，為限於其所成組合字均不相同，故其字根亦不相同。但有些字在將字根取出省去結合符號時，其順序完全相同，如：吧，邑；唄，員；瞭，景；盱，旱；暉，暈；旭，咎；群，羣；崐，崑等，其中群及羣；崐及崑為相同之異體字，故不加討論，因可由處理時取定一種即可，但其他各組，因其字形字義均不相同，必須能區別之。

此類字如只有字根時，實在不易插入適當的結合符號，因必須要考慮其前後字辭的文意，才能決定。而這種方法，實不可取。故本文提出簡單的處理方法，即將此數字先定好，在上下的口改為□(作包含用的字根)，而偏旁的仍舊，同樣在上下的日

改為曰，而偏旁仍舊，如此處理實很簡單，但應將此類字之結合符號轉至口及曰之下。

另外尚有一字患，因其字根序列為中心，與忠完全相同，而其又無結尾符號，故應加以處理，簡單的方法是將患之字根序列特定為中中心，而與忠加以區別。

下面本節將討論在字根式前後所帶之結合符號。

因最前面只是二個結合符號，即（及（（，故只須將後面之結合符號處理好，再計算對應之左右括號，即可將前面所缺者補入。故此處討論最後結尾的結合符號。

根據統計，有 47 個字根可做為一字的最後，而其後跟有除），））以外之結合符號。其中跟一種的有 33 個，跟二種的有 10 個，跟三種的有 2 個，跟四種的有一個，最多是五種，也只有一個（見附錄四）。

因最後接用的）或））決不會是與字頭的（或（（對應的，故這兩個可由設計者處理之。

如果字根最後只跟一種結尾符號的，則可查出其應附之結尾符號即可。如最後字根為“夕”，則因夕如有結尾符號，必跟 1，故補入即可。如其字根可做結尾之字根，而所跟之結尾符號是二個以上，則當特別處理之。以下共有 14 個字根其接受之結尾符號有兩個或兩個以上，分別說明之：

1. 口：

口後可跟 ϕ ，2，）2，3，3）五種結尾符號，但跟 3，3），及）2 均只一字分別為嚙，嚙及詰。而跟 2 者有器，囁及囁三字，故除此六字外如有結尾符號，必為跟 ϕ 。

2. 匕：

匕後可跟 1，2，1），2） ϕ 四種，跟 1 者有能，櫛二字，跟 1）者有罷，罷，及罷三字，跟 2） ϕ 者只一字𠂔。其他如跟結尾符號，必為 2。

3. 人：

人後可跟 ϕ ，）2，）2）三種，跟 ϕ 者只一字𠂔，跟）2）亦僅一字𠂔，其他如有結尾符號，必為 2。

4. 幺：

幺後可跟 2，2），）2 三種，跟“）2”及“2）”均只跟一字即𠂔及繼，其他則均為 2。

5. 木：

木後可跟 ϕ 及 2，跟 ϕ 者只一字森。

6. 又：

又後可跟 2 及 21，跟 2 者僅双及𠂔二字。

7. 土：

土後亦可跟 ϕ 及1，跟 ϕ 者僅𡗗一字。

8. 止：

止後可跟 ϕ 及)2，跟 ϕ 者有二字，𡗗及𡗘。

9. 虫：

虫後可跟 ϕ 及2，跟 ϕ 者只有一字蟲。

10. 火：

火後可跟 ϕ 及1，跟 ϕ 者有二字，𡗗及𡗘。

11. 戈：

戈後可跟1及1)，跟1)者有一字𡗗。

12. 水：

水後可跟 ϕ 及) ϕ ，均只限一字即森及𡗗。

13. <：

<後可跟2及3，亦各只跟一字即𡗗及巡。

14. 艸：

艸後可跟2及)1，跟2者只有一字艸。

故由上面分析結果，最後如已知跟結合符號，則均可確定所跟之結合符號。但最後有無跟結尾符號，則必由輸入資料時供給，否則即無法補入。如泉及𡗗。其字根式均由“白△水”組成，但泉為“白△水”，𡗗則為(白△水)3，故單由字根式，實無法區別之，另鹿及麤亦有此種情形。故應能確知最後有無結合符號始可決定此字之結構式。

最後討論關於簡便符號之使用。

簡便符號如存在於兩字根中間混合使用時，則當作一新的結合符號，處理即無問題，但是如為在一字之最後時，即發生問題，此因必須知其有無結尾符號，蓋因字根中並無此提示。如泉與𡗗，其字根式均為白，水，但前者為白△水而後者為(白△水)3，因(，)3均在字根之外，無法由字根決定有無此結尾符號，而在最後使用結尾符號之字一共有304個，所佔比例不小，故必須要處理，因同類字甚多，如水，森，牛，犇；金，鑫；鹿，麤等均是。

如將簡便符號省去，當然相當簡單，但本文認為在字根間如有簡便符號亦不用，則非常不經濟，故本文提出下列方法。

因在使用中文時，如不用結合符號，必只剩一連串之中文字根，因而在二字之間必有一字間符號分離之，否則無法知其字根之歸屬。

由統計結果顯示，除了結尾之簡便符號外，字根之間的結合符號均可省去。故遇到在304個有結尾符號中的字時，可多按一字間鍵，在程式處理時，可查其是否有兩個字間鍵而決定是否須加結尾符號。其缺點是操作的人員應稍微了解一點中文字之結

構式。但此法不失為一相當好的方法。

在含結尾符號中有麤，𠂇，森，𣎵四字之字根序列為𠂇，人，木，火，分別與鹿，从，林，炎四字之字根序列相同，為求區別起見，特定麤，𠂇，森，𣎵四字之後，再加一字間鍵，即以三字間鍵結尾。如此即可區別之。

故在一般情況外，尚有三點應加注意。其一是邑，員之口改作包含根之□，以與吧，頃區別；旱，暈，景，省四字之日改用日，以與肝，暉，瞭，旭四字區別。其二是患字之輸入字根序列改為中中心，使與忠區別。其三是麤，𠂇，森，𣎵四字之後以三個字間鍵結尾。

五、結合符號省去之方法及驗證

依據上節之分析，本節將進一步提出以中文字根式輸入中文之簡化——即在使用字根式輸入電子計算機時，只要將一連串的字根序列送入，就可利用程式將標準的中文結構式列出。

下面將對此項工作分兩部份，做一敘述，第一部份將說明各種表的建立與使用方法，第二部份為說明此項工作之執行步驟，將以流程圖方式說明。

在執行過程中，前節所述四類字根之性質不同，須各別處理，故必須將其分類列表備用。

第一類因屬於獨立使用或做字尾之字根，故其後應無字根，否則即有錯誤，如為新的組成字，因已超出 9,060 字之範圍，等於修改此一系統，處理較煩，故本文將不作討論。屬於第一類的字根共有 114 個，故只需將此 114 個字根逐字列於表中，以備查驗是否屬此類即可（見表 1）。

第二類字根，因只跟一種結合符號，故將字根與結合符號配合列於表中，當查知屬此類時，即可直接取出結合符號加於其後，此類字根共 169 個（見表 2）。

如“子”必跟“△△”，“疋”必跟“△△”。

第三類因不同之結合符號配合使用不同之字根，故此種配合列表，將各字根配合之結合符號列出即可，當查知屬於此類時，可由其後之字根查出所接之結合符號。此類字根共有 152 個，列於表 3。如“白”字根中可接用三種結合符號，但跟“）△△”者只有一字根“石”，跟“△△”者有儿，小，乚，王，巾，水，𠂇等 7 個，而跟△△者有 12 個，故只將跟）△△及△△者列出，其他如有結合符號，必跟△△無疑。

第四類為麻煩的一類，因在相同二字根間，可用不同之結合符號，故若要決定何者正確，必須查其前或後或前後兩者之組合字根始可決定。因而必須將其能夠分辨出結果的字根全部列出而形成樹狀，而此項列表工作亦選擇最多的一種做固定處理，其他分別列出，此類共有 64 個，列於表 4。如字根亻後接夕，則其間可使用△△（及△△二種結合符號，但如其後再接一字根牛時，必使用△△（，而其他均跟△△，故由此表

字 根
也
乍
用
子
...
共 114 個

表 1

字根	結合符號
子	△△
疋	△△
不	△△
彳	△△
人	△△
...	...
共 169 個	

表 2

字根	後接字根	後跟結合符號
白	儿	
	小	
	匕	
	王	△△
	巾	
	水	
	冫	
人	石	)△
	其他	△△
	木	
	一	
	十	
	八	2△△
	戈	
	心	
	火	
	火	)2△△
	川	
	毋	
	卩	△△
	頁	
文	鳥	)△△
	且	1△△
	其他	△△
	門	△△
	王	
走	彳	△△
	一	△△
	其他	△△
	又	
	佳	
	干	
	旦	
	卜	△
	巳	
	日	
走	戌	
	ン	△((
	其他	△(
...	...	...
共 152 個		

表 3

4, 可知其應接之結合符號, 再如字根日, 其後可接用△△, △△, 0△△三種結合符號, 如其後為字根木時, 可用△△及△△, 但後如再接字根斤時, 必用△△而構成晰, 其他則使用△△。如日之後字根為立, 中間亦可使用△△及△△, 但如立後再接字根, 必使用△△, 因使用△△者只組成一字即翌。如日後跟一, 中間亦可使用△△及△△, 但一之後跟日, 則取△△, 其他則取△△。如日後為大, 大之後再跟电則取△△, 構成𡗗, 大之後跟其他字根則取△△。日之後接冫, 其後再接且則取0△△構成壘, 不是接且則取△△, 另外日後除跟虫, ヹ, 正, 九, 辰, 共, 子等取△△外, 其他字根均取△△。

在儿字根中, 如接リ, 其間可用△△及)△△二種結合符號, 各別均只有一字即利及剋, 因字根リ為最後一個, 無法向後查, 故需向前查, 如儿前為二則取△△, 儿前為古則取)△△, 剩下的與字根日相似。如此, 第四類的組成很煩, 且佔用較多的表列位置, 但由字根序列, 必可查出字根間的結合符號則無疑問。

字根	後一字根	後二字根	結合符號
イ	夕	𠂇	△△(
		其他	△△
	其他		△△
日	木	斤	△△
		其他	△△
	一	日	△△
		其他	△△
	大	𠂇	△△
		其他	△△
	立	空白	△△
		其他	△△
	冫	且	0△△
		其他	△△
	虫		
	ノ		
	正		
	九		△△
	辰		
	共		
	子		
	其他		△△
儿	リ	" =	△△
	篆向前查	" 古	△△△
	羽	佳	△△
		空白	△△
	邇		△△
	厶		△
	丶(下點)		
	木		△△
	冫		
	乂		
	冫		△△
	頁		

表 4 (I)

字根	後一字根	後二字根	結合符號
儿	寸		
	厂		)△
	力		
	虫		
	其他		)△(
之	...		...

表 4 (II)

字根	結尾符號
日	φ
儿	) 2
女	φ
...	...
瓜	2

表 5

其次將決定結尾符號，根據分析統計，有 47 個字根可作結尾使用而其後可接結尾符號。在這 47 個中有 33 個只跟一種結尾符號，此可將其字根與結尾符號配合列表，當知需結尾符號時，查表即可取出使用，此類列於表 5，如結尾字根為儿，並有結尾符號，其結尾符號必為) 2，即兢或競字。而有無結尾符號，則由其後之字間鍵個數決定之，如二字根序列之間只一個字間鍵，則無結尾符號，否則即有結尾符號。剩下 14 個，為相同結尾字根，而結尾符號可以不同者。此類結尾符號之確定，必有賴

字根	前一字根	前二字根	前三字根	結尾符號
口	雨	乚		3)
		木		3
	士			)2
	犬			2
	頁			
	臣			
	其他			φ
乚	月	厶	四	1)
			其他	1
	其他			2
⋮	⋮			⋮

表 6

字根	結尾符號
乚	(尸 厶 乚 2) φ
木	木 φ
人	人 φ
犬	犬 φ

表 7

其字尾前面的字根，故列表時必形成樹狀（見表 6）。其中有麤，众，森及焱等四字需特別處理，因其字根序列分別與鹿，从，林及炎相同，即分別為鹿乚，人，木及火。此處本文以最簡單之方法解決之，即此四字送入時，再加一字間鍵，即與下一字之間有三個字間鍵，因共只有四字，應甚易記。此四字列於表 7，備三字間鍵之查用。

如結尾字根為口，其前面字根為臣，頁，犬時，則結尾符號為 2，如其前為士則取)2，如為雨則要再取前面一字根才能辨別，如為乚則取 3)，如為木則取 3，在口之前為其他字根時則取 φ。

如結尾字根為木，其後有二字間鍵，則知有結尾符號，且必取 2 成林字形，如其後有三個字間鍵，則取 φ，因此時送入者必為森。

此七個表建立之後（因限於篇幅，未將每表全部列出），即可配合各表之使用方法，作填入結合符號之處理工作。以下將以流程圖說明其執行步驟。圖中使用之符號說明如下：

(A)一表開始接收一字根序列（含字間鍵） $C_1 C_2 \cdots C_n$ 。

(B)一表一字根欲加結合符號之開始。

(C)一查驗有無結尾符號。

(D)一計算左右括號，查驗是否在字首字尾要補加括號。

②一表字根屬第二類，故查表 2。

③—表字根屬第三類，故查表 3。

④—表字根屬第四類，故查表 4。

⑤—表有結尾符號，屬表 5，故查表 5。

⑥—表有結尾符號，屬表 6，故查表 6。

⑦—後有字間鍵三個，表特別四字；麤，众，森及焱。故查表 7，查其結構式即可。

K—表所查字根或字間鍵在字根序列中之位置。

以上②③④項之分類可由各字根之分類作一對應表查出，⑤⑥兩項亦可如此處理。有關各表之使用方法，請參看本節前段之說明。

以下將以五個例子說明使用方法，其中 * 表字間鍵。

1. 彳人乚 *

彳屬表 2 之字根，只跟 $\Delta\Delta$ ，人之後可用 6 種結合符號，但屬表 3，查表知與乚配合使用者為 2 Δ ，故加入其結合符號而成

彳 $\Delta\Delta$ 人 2 Δ 乚。

因其無括號，故此即為此字（從）之結構式。

2. 亼王几 *

亼之後非夕，故接 $\Delta\Delta$ ，王之後可跟 5 種結合符號，但與几配合者只有 Δ ，故填入結合符號後為

亼 $\Delta\Delta$ 王 Δ 几。

此時有括號，且只有右括號，故按括號之處理方法，應在字首加上左括號（，故結果應為

（亼 $\Delta\Delta$ 王 Δ 几。

3. 禾日大电 *

禾與日配合可用 $\Delta\Delta$ ， Δ 及 2 Δ 三種結合符號，但取 $\Delta\Delta$ 只能合成秝，取 2 Δ 只能合歲曆，其他均取為 Δ ，故禾後取 Δ ，日之後可用三種結合符號，與大配合者亦有 $\Delta\Delta$ ， Δ 兩種，但大之後為电者則在日後取 $\Delta\Delta$ ，大之後可用 $\Delta\Delta$ ， Δ 兩種結合符號，但與电配合只能取 Δ ，其中無括號，其後無二字間鍵，故其字（發）所產生的結構式為

禾 Δ 日 $\Delta\Delta$ 大 Δ 电。

4. 木雨口 **

木之後可用 6 種結合符號，屬表 4，但與雨配合只有 $\Delta\Delta$ ，雨之後可跟 Δ ， Δ （兩種結合符號，但與口配合只有 Δ ，將其填入應為

木 $\Delta\Delta$ 雨 Δ 口，

因其後有二字間鍵，知有結尾符號，查表 6 知口後可有五種結尾符號，口前為雨時亦有兩種，但雨前為木時，其結尾符號即知為 3，故此字（樞）之結構式為

木△△雨△△口3。

5. 𠂔 𠂔 ***

此字使用三個字間鍵且以𠂔結尾，查表7知其結構式為

(𠂔△𠂔2)φ。

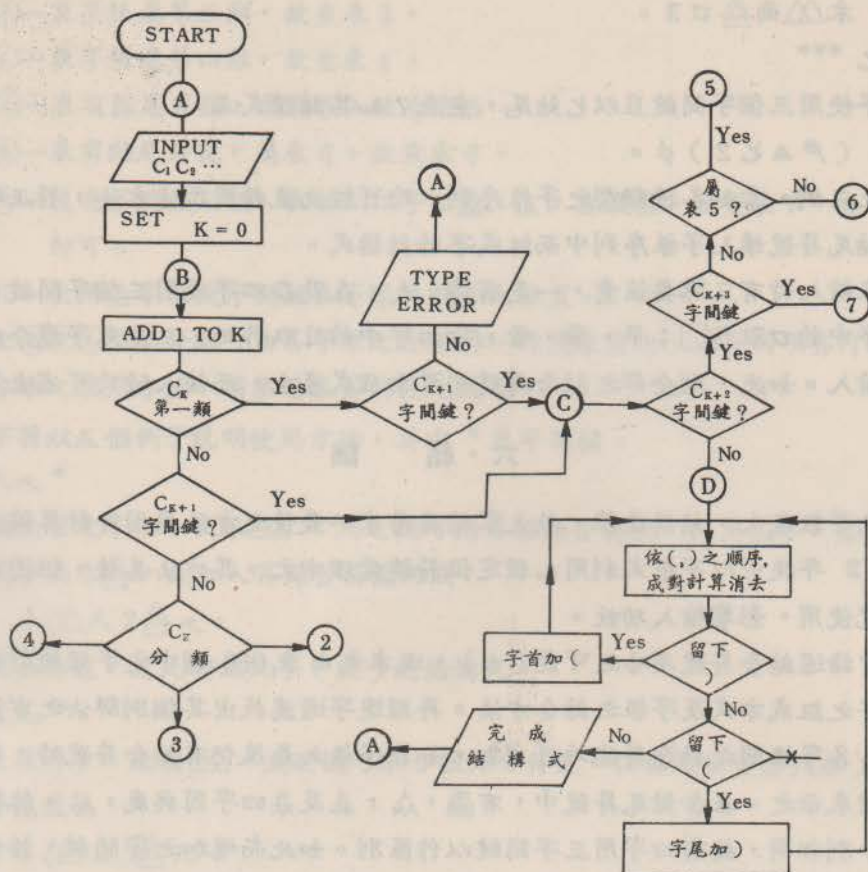
由此五例，當知各種類型之字根序列，均可按此流程圖所述方法，將正確的結合符號或結尾符號填入字根序列中而組成字的結構式。

惟在輸入時有三點要注意，一是有麤，众，森及焱四字要用三個字間鍵，二是邑，員二字中的口取作□；旱，暈，景，𠂔四字中的日取作日，三是患字應分成中中心三字根輸入。如此，則全部之結合符號均可由程式產生，而輸入時均可省去了。

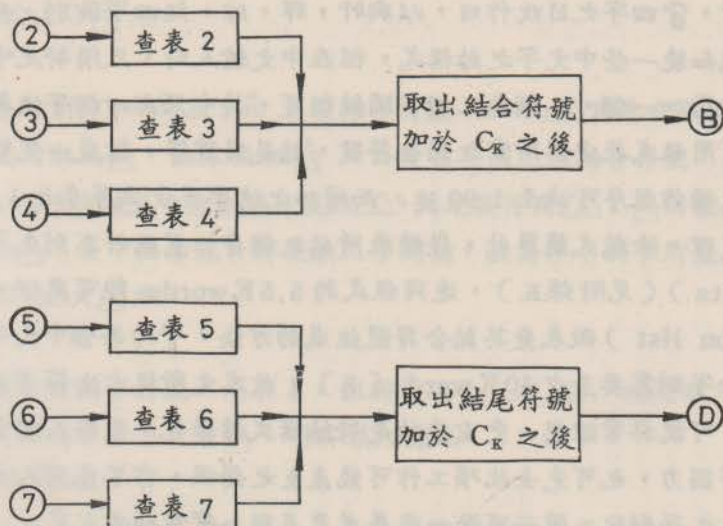
六、結 論

中文字數龐大，結構複雜，故大家均在尋求一最佳之方法應用於計算機上。交大曾於1972年提出以字根式利用九種定位符號處理中文。其方法甚佳，但因必須配合定位符號使用，影響輸入功效。

前曾論述結合符號省略之可能〔4〕，現本文由9,060個中文字經統計分析後，得知各字之組成方式及字根之結合方法。再經逐字過濾找出其個別解決之方法。利用此方法，各字根間之結合符號均可省去，如在字根之最後仍有結合符號時，則可多按一字間鍵表示之。在含結尾符號中，有麤，众，森及焱四字因與鹿，从，林及炎之字根序列分別相同，故前四字用三字間鍵以作區別。如此而增加之字間鍵，按頻率計算，平均每字影響不到0.01鍵。另有二點應加注意，其一是患字輸入時其字根序列應為中中心，以與忠區別。其二是邑，員二字之口改作□（包含根），以與吧，唄區別，旱，暈，景，𠂔四字之日改作日，以與肝，暉，瞭，旭四字區別。如此處理，其缺點為使用人應知曉一些中文字之結構式，但在中文輸入時，只須將此中文字分解之字根順序送入，再加一個，二個或三個字間鍵即可，其中只有四個字使用三個字間鍵。在處理時，可用程式找出其所需之結合符號，結尾符號等，構成一完整的字根結構式。如此每字之操作程序可減至1.99鍵，而增加之速率至少為原來之1.58倍。依本文討論之方法處理，除程式簡單外，幾種表所佔之儲存位置總計不到4.7 K words (1 word = 16 bits) (見附錄五)，連同程式約5.5 K words 即可處理，而以字根序列 (Description list) 做表查其結合符號組成的方法，平均每個中文字約需1.2至1.5 words，九千字則需要至少10 K words〔5〕，故本文所提方法節省甚多，對小型迷你電腦而言，可說非常理想。更重要的是將結構式標準化，使用人不需在分解組合字根上化費太多腦力，也可免去此項工作可能產生之錯誤，亦不慮因人主觀之文字組合，得到不同之文字形狀，因而可進一步尋求更美觀，實用的中文字。



流 程 圖 (I)



流 程 圖 (II)

七、誌 謝

本文之完成，得謝清俊、蔡中川兩位教授之鼓勵及建議甚多，特此致謝。

八、參考文獻

1. 謝清俊等：中文字根之分析。交大學刊六卷一期（1973）
2. 謝清俊等：字根之貯存和字的合成。同上。
3. 林 樹：中文電腦基本用字研究。國立交通大學科技研究報告 CC-601。(1972)
4. 許 桂 敏：中文字根之分類及使用程序之簡化。國立交通大學科技研究報告 CS-006。(1973)
5. M. W. Du, J. C. Tsay, C. C. Hsieh: The Design of the CHIPS — A Chinese Information Processing System.

附 錄 一

以 9,060 個中文字所形成之組合符號共有 53 種，分列如下，其中 + 表橫連，— 表直連，* 表包含， ϕ 表品字形，1 表炎形字，2 表林形字，3 表《 》形字。

1. 字首使用者有 2：(，((。

2. 字尾使用者有 14：)，))，) ϕ ，)1，)2，0，1，1)，2，2)，21，2) ϕ ，3，3)。

3. 字根間使用者有 37：+，—，*，)+，)—，)*，+(，—(，*(， ϕ +， ϕ —，1+，1—，2+，2—，3+，3—，)) +，))—，))*，)—(，)* (，)1+，)1—，)2+，)2—，+((，—((，21+，21—，*(((，1)—，2)+，2)—，2) ϕ ，2—(，))* (，))* ((。

附 錄 二

字根後接用結合符號之情形：

1. 跟十種的有 1 個：

口。

2. 跟七種的有 3 個：

儿，乚，虫。

3. 跟六種的有 4 個：

木，人，又，夕。

4. 跟五種的有 8 個：

一，目，力，王，土，干，九，丨。

5. 跟四種的有 18 個：

女，糸，心，八，厶，隹，乚，止，火，弓，戈，山，×，魚，丰，肉，※，、（右下點）。

6. 跟三種的有 54 個：

日，白，乚，月，文，□，卩，走，夕，貝，十，田，禾，广，立，斤，四，古，工，車，艸，尸，艮，戶，矢，欠，厂，米，乡，少，石，衣，丁，勿，刀，皿，彡，夕，夫，己，犮，又，卩，巳，曰，双，弋，甘，尺，丰，乚，艸，風，丁。

7. 跟二種的有 128 個：

イ，門（门），言，シ，宀，艸，才，小，寸，父，久，リ，大，可，上，方，巾，西，竹，口，云，ネ，里，主，二，巾，生，去，ネ，一，金，冂，中，勹，巴，犬，馬，儿，人，耳，白，自，戊，豆，足，母，义，雨，重，水，其，艸，云，且，皮，先，民，非，亡，呂，然，兩，而，由，血，告，臣，气，マ，高，舟，牙，冂，又，必，酉，求，巳，乃，予，廿，几，儿，リ，艮，冂，入，耂，小，黑，甫，羽，麻，子，鬼，丘，亞，毛，兆，无，庀，矛，弗，岳，凡，束，戈，門，瓜，巨，羊，由，壽，弓，心，业，氷，井，厂，尤，羸，包，フ，廔，爾，尤，冂。

8. 跟一種的有 169 個：

子，疋，不，イ，人，ナ，我，了，在，ク，來，頁，彳，至，繼，天，心，ソ，フ，者，幺，士，才，五，事，正，四，小，董，弟，片，永，丹，手，直，更，聿，面，シ，卩，亡，牙，下，亥，人，少，卜，血，小，氏，乚，冉，四，尹，羊，央，辰，身，壺，丸，未，黃，才，电，世，南，疾，水，曲，角，声，婁，内，文，申，巾，し，章，央，辰，共，炎，半，制，坐，片，止，束，尸，办，衤，卑，垂，革，丙，ト，少，毋，川，久，乡，甲，曷，凡，長，E，骨，戈，冂，鳥，來，乘，リ，爪，尤，屯，彳，事，萸，蔡，く，乙，丹，麗，兼，丞，龍，册，鳥，率，肅，身，典，豸，卯，鳳，虫，尸，甚，彳，市，夕，齒，巫，奴，肩，霍，才，幾，ノ（上撇），一（左上撇），丶（上點），臣，又，人，缶，卅，是，有，他，就，你，又，フ，弗，𠂇，𠂇，人，局。

附 錄 三

按字跟所接結合符號分類，共有四種。

1. 只接一種結合符號者有 169 個，同附錄二(8)。

2. 可接兩種以上結合符號，但均不跟相同字根者，有 152 個：

白，人，文，走，寸，貝，目，父，リ，力，可，立，上，巾，西，斤，四，云，古，ネ，里，主，二，工，…，巾，止，生，艸，去，金，艮，冂，勹，巴，矢

C：存該字根之類別，佔 1/2 words。

A：佔 1 word 如屬第一類此欄空白，如屬第二類，此處存此字根應接結合符號之代號，如屬第三、四類，此處則存此字根在表 3，表 4 中之位置。

此表每一字根佔 2.5 word，有 499 個字根，共佔 1,248 words。

3. 表 3。

此表所屬之字根，如其後接不同之字根則配合不同之結合符號。故列出後接字根及其配合之結合符號即可。每組佔 1.5 word，有 580 組，共佔 870 words。

4. 表 4。

表 4 中分三類，其結合符號只由一個後接字根即可決定者，共有 645 個，屬第一類，由後兩個或前一個始能決定者有 180 個，屬第二類，由後三個或前二個始能決定者，有 270 個，屬第三類。由此三類形成表 4 之三個副表，其組成型式除列出其後接之字根及其結合符號之代號外，並在一、二類字根之前增加一個指標欄，指標欄型式如下：

0	150	7
IND	C	N

IND：佔 1 word，如 C 為 0 則無作用，如 C 為 1 則此欄存本字根在下一副表中之位置。

C：佔 1 bit，0 表其後不再接下一副表，1 表其後尚有下一類之副表。

N：存該字根在本表中之個數，做檢查之限制。

第一類每字根使用 1.5 word，第二類每組使用 2.5 words，第三類每組使用 3.5 words，此表共佔 2,363 words。

5. 結尾符號表。

含表 5，表 6，及表 7，分別佔 51，94，9 words 共 154 words。

總合五種表共需佔 4,770 words，約近 4.7K words。