中 文 字 根 之 分 析

謝 清 俊（交通大學工學院）
黃 永 文（中原理工學院電子工程系）
林 樹（交通大學工學院）

（1972年12月16日收到）

摘要——本文討論之主要內容，爲分析中文字形結構之方法，其目的在減省計算機輸出中文字形所需之字形儲存空間。此法係將中文視爲由字根拼成之二度空間字形，進而提取其共同點，減少重複，並加系統之整理，以求得一類似語文模式（Linguistic Model）之結構。經統計結果，此模式可產生 48,713個已有之中國字，若以九千通用字計，可將現用之字形儲存空間減少至二十分之一以下，且對於中文輸入計算機，提供一種新方法。

I. 基本用字研究

我國文字，自殷商時代，已有甲骨文約三千字，至漢說文解字書，收九千餘字，清康熙字典，集字四萬二千餘，民國五十七年，印中文大辭典，得四萬九千餘字。然於茲四萬數千字中，今日能靈活運用者，祇占極少數，餘者可謂死字或垂死之字。古今重要典籍，其使用單字，罕有逾三千者，如易經，有單字1,595 字，[1]三民主義爲十六萬餘字鉅著，僅有 2,134 單字。[2]至於時下常用字數，亦三、四千字足矣。聯合報中文自動鑄排機，採常用字彙2,376 字，羊汝德先生編訂新聞常用字彙三千字。然爲應社會上各方面之需要，計算機所需字彙，應由現代社會通用文字中，盡量擴大蒐集範圍。因此，本文所述研究工作，採用中文電腦基本用字研究一書，列9,129字，略歸爲四類，計收「最常用字」1,857字，「次常用字」2,068字，「間用字」2,182字，「罕用字」2,425字，「異體字」597 字，已較一般常用字彙超出甚多，足供應用。

再者，此書中字彙及其出現頻率係集蔡樂生先生著常用字選等十一種資料統計而成，統計中，共取字樣（Samples）數爲 2,022,604，其中最常用字占全部字樣 97.34%，次常用字占 2.27%，間用字占 0.27%，罕用字占 0.12%。各字均附有其在全部字樣中之出現次數。可供中文字根分析之參考。[3]

II. 以往之字根研究

二十世紀以還，由於計算機科學之飛躍進展，自第一部計算機 ENIAC 問世，迄今雖祇二十餘年之歷史，而其應用範圍日益廣大，應用項目日漸增多，儼然成爲歐美各國促進科學、經濟、文化發展之重要工具；然而我國由於文字之字數太多，字形複雜，以致難於利用計算機處理中文資料。近年來，國內外不少專家學者，曾致力於中文資料處理自動化之研究，嘗試中文輸入、輸出之各種方法：或以一字一碼，或取字之部份筆畫爲碼，或以字音爲碼……等，然都難以推廣應用，究其原因，均碍於所需儲存空間太大，機器製作困難，操作費時費事。爲謀此等問題之解決，遂有中文字根之研究。此法係將每一中國字用其所含之字根（Radicals）以及指示字根關係位置之定位符號（Operators）表示，作爲中文資料輸入輸出之藍本。茲將其較重要者，略述如下：

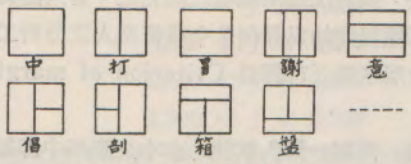
Fujimura 及 *Kagaya*:[4] 彼等計用字根二十一個，每個字根上有二至三個連接點；以不同之定位符號表示字根相連接之關係。此法固然有其優點，但其表示法相當冗長，而且爲遷就所定義之定位符號，而犧牲傳統筆順，尤其是兩個不相連接之筆畫，諸如"二"，"彳"等，無法加以描述，更是嚴重缺憾。

Toshiyuki 等[5,6] 或 *Haruhisa* 及 *Shigeo*[7]，其所用方法大致相似。利用許多定位符號，以表示字根間相關位置。依其觀點，用於數量較多之中國字，惟有增加定位符號數目，或定位符號下不能表示之字視爲一字根。因此，難免淪於定位符號數目繁多，規則瑣碎，以及字根增加，應用困難之弊。例如 *Toshiyuki* 使用五種不同之定位符號 S. K. L. J. G. 分別表示困、字、進、司、原五類結構字形之字根相關位置[5]；其實，此種相關位置即"木"放在"口"中，"子"放在"宀"中，"隹"，放在"辶"中，"言"放在"冂"中，"厤"放在"厂"中的位置，在中文字根本身性質中已具備，今用五種不同之定位符號加以區別，豈非徒增不必要之記憶與乎操作之煩難。

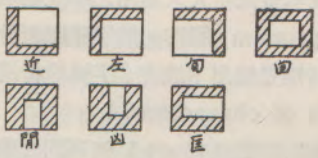
又以字合成之美觀言之，兩個字根所用定位符號相同，但其被包含字根位置之大小却不一定相若，如"問"與"岡"二字，*Toshiyuki* 同用定位符號K，但"口"在"門"內與"冈"在"冂"中所佔位置之大小，却未加注意；其實，"冈"所占位置應較"口"爲大，方能符合字形美化原則。

綜上所述，以往之字根研究，泰半由於未能深切了解中國文字之結構特性，更未能考慮字形之美化問題，以致成效欠佳，難於推廣應用；因此，吾人從事字根分析，勢非着重於國字結構特性之探究不爲功。

III. 中文結構之形式



圖一　直根與橫根



圖二　包　含　根

圖一爲楷書文字可區分爲數部份之常見形式，當然尚有若干較複雜之形式（倪耿先生共列出68種）[8]，但在應用計算機數學模式時均無問題。須特別注意者，端爲圖二所示之包含根。所謂包含根係指一字根之位置有包含其他字根之勢，此種字根，依其包含之形狀與位置，可概分爲七大類，（如圖二所示），茲各舉一字爲例，以說明之，如「近、左、旬、回、閒、凶、匡」等字中之「斤、工、日、口、月、×、王」均爲被包含之字根，其餘部份即分別依序爲圖二所示包含根之例。包含根與被包含根常有一定之相關位置，不容輕易更動，此特性對於定位符號之省略，有極大助益，被包含根之地位及大小與字形之美觀甚有關聯，因此字形組合之設計，必須注意及此。

圖一中，字根分爲橫向或直向排列，可稱之爲橫根或直根，甚多字根只作橫根或直根用，且可分爲左橫根、右橫根、上直根及下直根，其縱橫比例亦多有一定者，可減省字之復合工作。

IV. 重複原理與字根

各國文字均有其重複（Redundancy）之處，中國文字之重複處在於字之構形。多數字可以析分爲數部份。**若一字可分爲數部份，每部份各有其他字之部份與之相同者，則各部份稱爲字根。**若取字不加析分時，則此字亦可稱爲字根。字根實即重複之源由。

在諸字中共含之字根爲數甚衆，設若一概漠視之，不加利用，而視字彙爲整個之個體，則無異不解中文之外籍人士，視中文如圖畫，而以處理圖畫之方式處理中文，用之於傳眞，固無不可，用之於計算機，則頗有商榷餘地。

設採一字所占方格劃分爲32×32小格，小格中有墨者（或墨多者）以黑點或「1」代表，無墨

者（或墨少者）以空白或「 0 」代表，如此一字所需之記憶容量爲32×32＝1024 bits＝128 bytes，以貯一萬字計，共需記憶容量1280 Kbytes! 此項數字，頗爲驚人，非中、小型計算機所足以勝任者。此外，鍵盤上將有極多之字數，則打字工作又非專業不辦矣。

究其原因，實因視國字如圖畫所致，造字原則如形聲、會意等，導致字形重複之處，絲毫未加利用之故。其實國字雖多衍變，大部仍井然有序，字根之數甚爲有限，並非任何形狀之圖案均可能出現者，若然，則前述巨額之記憶容量固無法減省矣。

復有一種構想，謂中國文字不過點橫直撇捺等而已，歸併爲八類，良可應付，賦以編號，設八鍵之鍵盤足矣，窮所有國字，未有出此範疇者。法似甚善，然欲計算機能辨認分別一字，則一字一碼，一碼一字（One-to-one Correspondence）爲必要條件，在此分筆法中，不易作到，此點與注音符號分類法之困難相同。倘加入人爲編號（一如電碼），則有强記背誦與專業訓練之苦事，難以推廣應用。此外，尚有筆順次序，簡、異字體及輸入費時等主要困難。

以上僅就輸入而言，若輸出則分筆法幾不能適用，蓋上一筆與下一筆之相對位置及長度變化，過於繁複，處理極端不易，更遑論字之整齊與否矣。

V. 字根分析之邊際效用原則

綜上所述，關於中文字根之重複性，如利用不足，則記憶容量與鍵盤規模龐大，成本過昂，影響中文電腦化之推廣；如利用過度，則輸入速率過緩，且軟體程式過繁，成本亦增，倘有字形過劣之問題不易解決。前舉兩例恰爲二極端情形，以一字爲一圖案，絲毫不加析分者，利用不足（毫無利用），以一筆爲一單位，將字全部分解至極限者，則利用過度。

欲求一成本至低而人人易用之中文處理辦法，須於上述兩極端之間覓之。吾人認爲兩極端間存在一最佳（Optimum）解答。國字究應析分至何種程度始爲最佳？今爲使吾人之分析工作有一理論根據可循，特探究提出下述之「字根分析邊際效用原則」（可譯爲 Criterion of marginal utility for analysis of characters）：

假設一輸入裝置具有 $n+1$ 個不同之輸入鍵，而每一輸入訊號（在本文應用中即每一中文字）平均需按鍵 s 次。

今考慮其中一鍵應否取消，設此鍵所相應之輸入訊號，即中文字，可分爲二部份（即分爲二字根），若其中有一部份與其餘 n 鍵所代表者均不相同，則此訊號字如分解爲二部份爲無意義，因必須新設一字根鍵，故雖原鍵因分解而取消，並未減少總鍵數，而徒然增加平均按鍵次數。

若一鍵所對應之字可分解爲二字根，而此二字根均已包含於其餘之 n 鍵中，則如將此鍵取消，可收減少鍵數之利，其所代表之字仍可經其他鍵而輸入（如何將字根再組合成字見[9]），鍵數則自 $n+1$ 減少爲 n。

但此時按鍵次數即增加，原按一次即可，今則需按二次（另尚需按組合鍵一次，以使兩部份復合爲一，惟組合鍵極少且常用，故其費時可略而不計），至於其影響平均按鍵次數如何，須視此一字（或字根）之常用程度而定。設 f_i 爲字之出現頻率，此字在全部字樣 $\sum_{i=1}^{N} f_i$ 中出現 f 次（N爲全部字數），則此字出現之機會爲 $f/\Sigma_i f_i$，當此字（或字根）出現時，按鍵次數則增1，故平均按鍵次數應自 s 次增加至 $s+f/\Sigma_i f_i$ 次。

在應用輸入裝置時，鍵數愈少愈方便，又每一訊號平均按鍵次數愈少亦愈方便，故使兩者之乘積爲極小值，頗具意義（其實際意義，在後文中將再討論）。

第 $n+1$ 鍵如不取消，則此乘積爲 $(n+1)s$；如取消，則積爲 $n(s+f/\Sigma_i f_i)$，此鍵取消與否，當視兩乘積之大小而定，若

$$(n+1)s > n\left(s+\frac{f}{\Sigma_i f_i}\right),$$

則當取消之。上式即

$$s > nf/\Sigma_i fi,$$

或

$$f < \frac{\Sigma_i fi}{n} \qquad (1)$$

此處 $\Sigma_i f_i$ 爲取全部字樣數，按基本用字研究一書，卽 2,022,604。s 本隨 n 而變，理論上每一 n 確定後，可統計一新平均按鍵次數 s，但此統計工作甚爲繁巨，故可以嘗試誤差法 (cut and try) 決定之。事實上，s 爲一變化甚緩慢之函數，當 $n=8,532$ 時，$s=1$, $n=496$ 時，經統計 s 之值爲 1.9。計算得 f 諸值如下（通常 n 甚大於 1，故 $n+1$ 與 n 值近似）：

鍵盤鍵數	分解爲二部份之字 其出現次數應小於(約值)
50	76,000
100	38,000
200	19,000
300	12,000
400	9,500
500	7,600

在一字分解爲三部份或以上，則（1）式應修改推廣。設一字分爲 r 部份，則此字出現時增加按鍵 $r-1$ 次，故平均按鍵次數爲 $s+(r-1)f/\Sigma_i f_i$，得分解條件爲

$$f < \frac{\Sigma_i f_i}{n(r-1)} \qquad (2)$$

在 $n=500$ 時，計算得下表：

$f > 7,600$	不分解
$3,800 < f < 7,600$	可分解爲二部份
$2,500 < f < 3,800$	可分解爲三部份
$1,900 < f < 2,500$	可分解爲四部份

前文中述及使用方便之程度，此詞似甚爲抽象，不易確定，惟細推使用時吾人之主要考慮條件，無非學習速率（卽學習之難易），輸入速率及成本等項，今分別論之。

學習速率與鍵數 n 有關，因如欲操作純熟，須記憶各鍵之位置，如記憶之事物在人腦一般之容量以內，吾人認爲記憶所需之學習時間大致與欲記事物之量（此處卽鍵數）成正比；如此則可假定學習速率與 n 大致有線性關係。

輸入速率以尋鍵時間及平均按鍵次數兩者爲主要決定因素。尋鍵時間有二情況，一爲熟譜者所需，包括記憶之讀出及移動手指時間，一爲生手所需，包括鍵面文字之逐行尋找，兩者均不難理解爲大約與 n 成正比。至於平均按鍵次數卽 s。

成本復可分爲機件費用、打卡員薪資及計算機租金。機件費用包括按鍵開關數及線路材料以及裝置人工，易知其均與 n 成正比。打卡員薪資視輸入速率成正比，與平均按鍵次數最有關聯。至計算機租金亦視輸入速率而定，由於計算機運用之快速，吾人可預見輸入過程將成爲瓶頸，故租金一項，主要亦與 s 有關。

綜上所述，使用之不方便程度，其中各項因素均大致與 n 及 s 成正比。若成正比，則(1)(2)二式爲必然結果。但二式之成立不以正比爲必要條件，事實上，若不方便程度爲 ns 之任意遞增函數，如 $\phi(ns)$，均可得相同結論。

爲探討起見，設此項不方便程度 $\phi(n^p s^q)$ 爲與 n 之 p 次方及 s 之 q 次方有關，則一鍵應予取消分解之線性近似條件爲

$$\frac{\partial \phi}{\partial n}\Delta n > \frac{\partial \phi}{\partial s}\Delta s, \qquad (3)$$

式中 $\Delta n=1$, $\Delta s=(r-1)f/\Sigma_i f_i$ 故得

$$pn^{p-1}s^q\phi^1 > qn^p s^{q-1}\phi^1 \frac{(r-1)f}{\Sigma_i f_i},$$

或

$$f < \frac{p}{q} \frac{\Sigma_i f_i}{n(r-1)}, \qquad (4)$$

即分解爲 r 部份時各 f 限值之比例不變，僅相差一 (p/q) 因數。

VI. 字根之分析及頻率統計

由上理論，理想之字根選擇，爲每分解一字重新計算頻率分配一次，逐字進行，求其邊際效用，至無再應分解者爲止，但事實上由於繁複程度，難以作到，實際選字根之過程如下：

中文以左右及上下可分爲二部份之字數最多，尤以左右可分者居多。第一步先將左右可明顯分爲二部份之字分解，相同之字根歸併，並將其所屬原字頻率相加，如此可將 9,129字減少爲 3,256字及字根。第二步復將上下可明顯分爲二部份與包含結構之字分解，結果字及字根數減少至 621，經簡單分解，已將原有字數減少約十四倍。

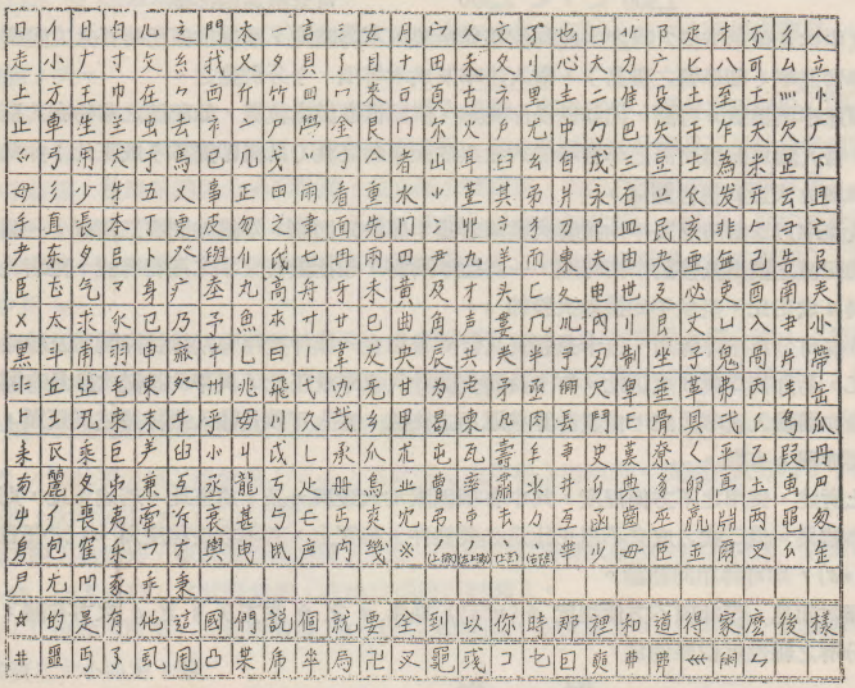
須注意者，如前論，應保留出現次數大於 (2) 式條件之字，不予分解，如此共保留25字。此外，一字分解後，其字根之一，無在他字中共用者，爲孤獨之字根，亦不予分解。

再將前述原則應用於字根，作第三步之分解，並檢查第一、二步中不合原則之分解，予以恢復，如此共得字根448，原保留之25字以及罕用字根23，合計字根共爲496。（見表一，表二）此數雖仍較吾人所樂意之數爲多，然如再加分解，則不免支離破碎，燕雜而輸入費時，已如前述。

496字根中，含有常用熟字數305（爲整體之字），其出現次數超過所有字數之半。當出現時，僅需按鍵一次。

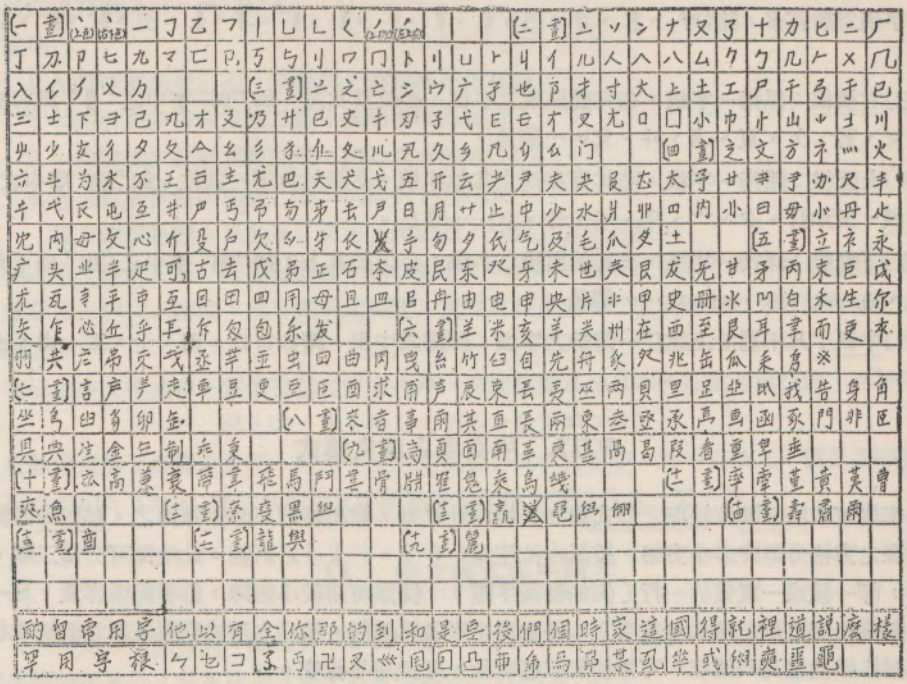
字根之使用率累計百分比如下：按常用次數之序排列，最前五字根占 11.3%，最前25字根占30%，最前50字根占49%，最前100字根占 66.7%，前200字根占 84.9%，前300字根占95%，前368字

表一：中文字根表甲（依頻率高低排列）



說明：1. 本表依字根出現頻率之高低由左而右，由上而下順序排列。
2. ☆爲酌留“常”用字“井”爲罕用字根。
3. 本表計收字根 448 個，酌留常用字25個，罕用字根23個，總計 496 個。

表二：中文字根表乙（依筆畫多少排列）



說明：本表按字根筆畫多少排列；筆畫相同者，按具起筆點，橫、直、撇、捺順序；起筆相同者，按頻率高低排列。

根占99%，其次50字根占 0.8%。

VII. 字根之分類

由前述中文結構之形式，可知每一中文字根與另一字根間，常有一定的相關位置，且其縱橫比例亦多有一定，此一特性，對於字根組合所用定位符號之省略，有極大助益，因此若能將字根按其特性，妥爲歸類，則表示字根間相關位置之定位符號，似可省略，如此必可使操作方便快速。茲將前節分析所得字根，略加歸類如下，以供進一步字根分類之參考。

(一)包含根：係指一字根之位置包含有其他字根之勢。如“間”字中，“門”爲包含根。可概分爲七類：（下圖所示，設一單字之形狀爲口，斜線部份爲包含根位置，空白部份爲被包含根所佔位置。）七類之中，尚有程度之不同，茲舉例如下：

1.  ：辶、廴、九、走、毛、尢、夂……
2.  ：𠂇、广、厂、尸、戶、尸、疒、尹、……
3.  ：勹、冂、气、戋、弋、戈、弋、幾、鳥……
4.  ：口
5.  ：門、閂、扈、戊
6.  ：凵
7.  ：匚

(二)左橫根：係指一字根隨其後出現之其他字根，必接於其右。如“位”字中之“亻”爲左橫根，此種字根爲數甚多，諸如氵、扌、彳、糹、礻、衤、忄、⻊、犭、孑、亻、爿、彡、镸、豸……等是。

(三)右橫根：係指一字根，在其前出現之字根，必位於其左，如“到”字中之“刂”爲右橫根，又如：攵、頁、欠、東、艮、斗、殳……等是。

(四)上直根：係指一字根，隨其後出現之其他字根，必接於其下一區位、如“箱”字中之“竹”爲上直根，又如艹、⺈、亠、⺕、丷、廿、⺊……等是。

(五)下直根：係指一字根，在其前出現之字根，必位於其上，如“第”字中之“弔”爲下直根，又如氏、电、儿、爭、艮、疋……等是。

(六)中根：凡不屬於上述任一情形之字根、統歸爲中根，諸如口日、田、十、王、山、幺……等是。

VIII.　中文字之語文模式

各字根可以點矩陣或線段之形式儲存資料於計算機，輸出時則以點矩陣爲之。點矩陣即前述將字根所在之方格分爲許多小方格，按黑白依序爲「０」、「１」之訊號。此法尙可分爲二類，一爲單線字，即一橫或一直僅佔一行(例如藝術字體)，優點可用較小矩陣，節省記憶容量，另一爲粗體字，每筆粗細可變(例如毛筆字)，優點爲字形逼眞美觀。線段之形式則視筆劃爲直線或折線，以其起點終點位置資料儲存，可更節省記憶容量，此法用單線字。

字根儲存之形式，所佔空間之大小，及其分類等之資料結構 (Data Structure)在[9]中詳細討論。

輸入時另制定四種基本定位符號 (Relational Operator)，即(　)◬，厶，及厸，其優先次序即如所列。◬表包含，隨於◬後之字根爲◬前字根所包含。厶及厸表上下及左右組合，括號之作用一如數式中者，爲最優先。因包含根以局部居多(即僅包含一根最爲常見)，故令居優先。又圖一第二行前二形較後二形爲常見，故令厶較厸優先，在後二形中使用時則加括號。

舉例以明之，「休」「類」「盟」「葡」諸字之按鍵法如下：

休	亻厸木
類	米厶犬厸頁
盟	(日厸月)厶皿
葡	艹厶勹◬甫

當括號中僅有二字根時，前後括號之一尙可能省略。

二字根作縱向或橫向復合時，其相對大小比例即同於長或寬之指標數之比例[9]。例如「休」字二部份作橫向復合時，若「亻」根寬指標爲１，「木」根寬指標爲３，則「亻」占寬度¼「休」占寬度¾。

根據以上之分析，可將任意一中文字分解成爲一個字根與定位符號之表示式，此結構即爲一個有優先序的文法結構 (Precedence Grammar)，可稱之爲中文字之語文模式或中文模式，用 Bakcus Normal Form 可將之表示如下：

<中文用字>::=<文字>/<符號>

<符　號>::=＋/－/×/÷/？/・/，/：/；/(/)/{/}/〔/〕/*/$/#/「/」/……等等
並加上大小寫英文字母以及希臘字母數學符號等，多少不拘。

<文　字>::=<字根形>/(<字根形>)/<文字><定位符號><字根形>/<文字><定位符號>(<文字>)

<字根形>::=<字根>&/<字根>∞/<字根>8/<字根>∞/<字根>

<字　　根>::=496個（如表一）

<定位符號>::=ㅿ/亼/◬

說明

1. ∴符號表示成「品」形重疊，例如車∴即成轟。

2. ∞符號表示橫向重複二個，例如木∞即成林。

3. ooo符號表示橫向重複三個，例如口ooo即成□□□。

4. 8符號表示縱向重複二個，例如火8即成炎。

以上四個符號，只是爲了使用方便而設，故稱之方便符號。

以上中文模式，是一種再生形式（Recursive form），寫成此形式，純爲表示簡潔方便，事實上，並非再生形式之結構，因爲表示式長度，實爲有限長度，在倪耿論文中，曾統計此形式約有六十八種之多(8)。

在以上中文模式中，定位符號有優先次序，若將括號及∴8∞ooo符號考慮在內，可排列爲

∴，∞，8，ooo>（>◬>亼>ㅿ>）

此結構提供了必備之資料，可將中文表示式組合成中文單字，其中方法之一便是用結構樹分析法（Structure Tree）詳如本期中文字根的貯存和中文字的合成一文。(9)

IX. 中文模式與中國字

本文研究之中文模式，雖僅有 496 個字根，然若用以組合中國字，則除本研究所採用之 9,129 個單字毫無困難外，尚可組合成歷來所有中國字之97.61%!!

我國文字之總字數，歷經數千年之累積，迄今已達五萬字之譜。張其昀先生等編纂之中文大辭典，係歷來所有字書中單字數之最多者，該書網羅經史子集，旁採類書、叢書、輯佚書、字書、辭典……等資料，全書計收單字 49,905 字，堪稱集古今中國字之大成(10)。以本研究之中文模式，逐字加以核試，結果於該49,905個單字中，計可組合48,713字，僅有1,192字難以組合。然於此吾人必須特加留意者，端爲此類難以組合之字，泰半已成死字或爲異體字(3)。諸如："禮、弌" 等字，其義未詳；"𣥐、甪、禞" 等字，各爲 "蹋、用、神" 之本字；"𠁈、壽、冈、酋" 等字，各爲 "龜、壹、宮、齒" 之古字；"[illegible]、乇" 等字，各爲 "變、中"之籀文；"肖" 係 "𡉀" 之篆字；"圙" 爲古器；"飞" 爲飛之簡體；"凲" 與 "滂" 同；"𣗊" 爲 "椒" 之俗字……等是。

再者，誠如前述，中文模式具有再生性質，因此除可組合已有之 48,713個中國字外，尚可產生新字，對於日後爲應時代、社會需要而增加之新字，未嘗不無裨益！

X. 鍵　盤　設　計

前文中所得 496 字根各以一鍵代表，作成一鍵盤，以爲輸入之始端機器。鍵盤中字根之排列法與使用之方便與否有極重要之關係，其主要之考慮包括記憶之方便，尋找之方便及手指移動平均距離之減少等，以增加速率並降低輸入者之勞累，庶幾爲大衆所樂用而推廣流行。

設計時考慮之因素可列舉如下：

(一)根之類別：凡屬包含根者，極易區別，可分列之；有爲明顯之左橫根、右橫根、上直根及下直根者，亦可分列之；此外，有左右上下中俱可明顯應用，筆劃甚少而極常用者，可列中間；尙有獨字常單獨應用，極少組合爲他字之根者，可列於較偏僻處。

(二)字形：字形相似者並列，如「己、巳、已、㔾」等，對於記憶極有助益。

(三)筆劃：筆劃極少以及極多之字，一目了然，爲極易分別之事，惜筆劃適中者，不易分別。

(四)首筆分類：取一字之首筆及第二筆等，視其爲點、橫、直、撇等加以分類。

(五)部首：因字典之通用，熟習部首者極衆，但有難查之字。對於字根，則不易適用。

(六)注音：注音可一如英文字母排列，消除許多不便，但有同音字問題及不諳注音者。對於不成字之字根無法應用。

(七)字義：同義字、反義字及相似詞等可並列，減少記憶困難，但僅能局部應用。

(八)詞彙：考慮字與字間之關聯，卽連詞之應用，如並列，則可減少移動距離及尋找時間，亦只能局部應用。

(九)頻率：以熟字僻字分列，可增輸入效率，鉛字盤中已有應用，但本法在實用上僅能作粗疏分隔。

(十)歌訣：可助記憶，例如與地命相之類，藉歌訣傳千百年之久。如將無可分類之處，編爲歌訣，或亦可助益。

XI. 結　語

(一)本研究係根據中文電腦基本用字研究一書所列單字 8,532字及其異體字 597 字，合計 9,129字，經分析統計而求得 496 個字根，由此，字形之儲存空間將可減少約 20 倍；再者，字根之筆畫數遠較單字爲少，應用資料壓縮法 (Data Compression Schemes) 亦可使其儲存空間大幅減少，[9]合而觀之，其儲存空間總計至少可縮小40倍以上。

(二)該儲存空間，已縮小到適合使用 Mini-Computer 或 Special Wired Read Only Memory 實際製造，故有望利用 LSI 技術，大量生產，以利國人使用。

(三)此組字根尙可於計算機中再分，使用時以軟體組合，更可節省儲存空間，此點亦頗有研究之價值。

(四)由於文中之中文模式具有再生性質，故用此496個字根，不僅可組成 48,713 個已有單字，而且尙可利用語文模式，創造新字，此一特點乃爲其他方法之所無。

(五)加碼 (Encoding) 之方法可與其他中文加碼方法互相轉換，僅需一對照表而已，而此對照表亦遠較本文第三節中所述之儲存空間爲小。

(六)用此方法組合單字應不困難，訓練容易，使用方便；至於如何將字根組合成單字以及字形之美化問題，本期另有專文論述。[9]

(七)目前每字之按鍵次數，若符號不計，平均僅爲 1.9；而厶，ㄙ，△等符號之省略，或可用字根分類法爲之，乃爲尙待研究之問題。

(八)應用中文字根編排索引，對於中文資料之檢索以及中文字典之編排，提供一種新式方法。

(九)以此字根設計之中文鍵盤，其鍵數較一字一鍵法減少約20倍，如此不僅可使鍵盤之製造大幅減少，操作亦大爲簡便，對於中文打字機及計算機中文輸入裝置之應用，頗具價值。

誌　謝

本研究之完成，承蒙杜敏文教授提供許多寶貴意見；關於字根之分析統計，復蒙劉冠軍、柳肇輝、林明達、黃玉美、林家榮、曾七雄、秦　順諸先生協助，對本研究貢獻良多，謹此一併致謝。

參 考 文 獻

1. 見燕京大學索引。
2. 彭仁山：教育研究二十三期，中山大學，民國十九年。
3. 林　樹：中文電腦基本用字研究，交通大學工學院，民國六十一年。

4. O. Fujimura & R. Kagaya, "Structural Patterns of Chinese Characters", Univ. of Tokyo, 1969.
5. Toshiyuki, Sakoto Nagao, Hidekazu Terai, "A Discription of Chinese Character Using Sub-Patterns.," Dep. of E. E., Kyoto Univ.
6. Toshiyuki Sakai, Shigehara Sugita, Hideo Fujita, "Some Experiments on Kanji I/0 System" Dep. of E. E. Kyoto Univ.
7. Haruhisa Ishida, Shigeo Furukawam, "Synthesis Display of Kanjis by Unit Construction". Univ. of Tokyo.
8. 倪　耿：中國字之結構模式及其分析 p. 55-59. 交通大學工學院碩士論文，民國六十一年。
9. 謝清俊等：中文字根的貯存和中文字的合成，交大學刊六卷一期，交通大學工學院，民國六十二年。
10. 張其昀等：中文大辭典，中國文化研究所，民國五十七年。