

開放文學－漢文樂園－中文電腦漫談

附錄：中文電腦的心臟.....中文字形產生器的製作方法

這篇附錄是在校對完成以後才決定加的，原先我打算把它單獨出版，由於同事們看完本書後，紛紛表示「內容」不足要改寫。在目前忙碌的情況下不大可能，把兩本合為一本才是上策。作為附錄還有一層意義，前面我「高調入雲」，可是並沒有提供年輕人一個可以入門的方向，罪過！罪過！

因此，我願意把我目前賴以「維生」的「中文向量組字法」的獨家技術公開出來，希望年輕朋友們一齊來共同研究。

不過有一點聲明，我公開技術是一回事，我的商品權利是另一回事，商品已經上市，已經有了專利權，若是有任何人想不費腦筋，「翻版」盜用，請不妨試試看，我會不惜傾家以拼（雖然多年來為了研究，我早已傾家蕩產，至今負債六百多萬元），爭的是個「理」，爭的是所有研究發明人的「權益」。

至於提供技術，讓你能研究，希望你做得更好，那是我的心願。如果你加入了新的構想，做得比我的更好，我會第一個向你道賀，向你學習。別忘了，中華文化不是個人的私利，也不是某人的責任，只有大家共同奉獻出各人的精力及智慧，才能繼續往開來。

向量組字法耗費了我八年的心血，也投下了我們公司十幾位同仁無數寶貴的時間及智慧，它之能有今天，該感謝的還有更多默默的無名英雄，他們和你我一樣，都是中國人，都熱愛自己的文化傳統。今天我把這些技術公開，絕非作無謂的犧牲，而是拋磚引玉。

話說在一九七二年，我在巴西一個龐大的文化出版公司CUTURAL ABRIL 做美工完稿的工作。那時，我已浪跡海外達八年之久，從事過三十多種不同的行業。孤家寡人，形單影隻，每天我問自己：「又過了一天，我為什麼活著？生活的意義是什麼？」當然有答案，但是卻可望而不可及。

任何人，所追求的都是生活幸福，然而對一個中國人說來，生活幸福不僅僅只是物質的享受而已。因為，廣大的領土上，還有十億同胞，只因為少數人的野心，史無前例的浩劫降臨了。而在台灣，地狹人稠，廿年前經濟條件很不理想。因此，我背井離鄉，滿以為過不多久就可以揸一面「歸國僑領」的旗子，衣錦還鄉！

問題不那麼簡單，北美、南美，我所到之處，「中國人」所能從事的行業，都是些洗衣店、餐館、華洋雜貨。更由於隔閡，華人都被視為固執、神秘、不思進取的怪物。

由於國人謀生不易，省吃儉用地攢小錢為大錢。擺脫了貧窮，立刻千方百計與自己同胞隔離，羞與為伍。華人圈中多少年來，就是這樣的「鯉魚躍龍門」的方式，新舊交替著，跳進龍門的不再是魚了，引得魚兒們羨慕不已！

我作過很深入的觀察，體驗過每個階層的心態，得到唯一的結論是：「中國人的教養，本質都不錯，但限於「知識」，尤其是現代社會的各種專業知識，以及與外國人交際應酬的「介面知識」太缺乏了，以致於先求自保而因之自閉，更由於自閉而自卑。最後即使發了財，他們仍然沒有安定感、沒有幸福、沒有成就。」

這種情形和巴西早期的社會有很多類似之處，巴西是個地大物博的樂園，面積和我國大陸相近，但是可耕地卻比我國多出50%（我國約17%，巴西67%），年平均溫攝氏24度，除極南端外，一年如春。

巴西人種混雜，是移民的天堂，由於移民知識水準較高且比較勤儉，所以都有所成，頗引起巴西人的「側目」而自暴自棄。

這種現象直到一九六四年四月，一批年輕軍官推翻了左傾的總統姜古拉，厲行改革，他們認定要使國家現代化，必須從教育及知識水準提高做起，因之成立了這家文化出版公司。

我在該公司服務的那一年，公司的營業額佔巴西全國各種企業中排行第三十位，公司員工有五千人，每週固定出版98種專業雜誌，以及不定期書籍刊物無數！

這種企業化傳播知識的手段確實有效，近年來，巴西經濟的成長在我國之上，你無法想像十多年前他們曾被視為蠻荒，國民有80%都是文盲！

有一天，我那個小單位的主管出缺，由於工作的關係，編輯部門交給我一份「急件」，是剛在美國出版而且暢銷的一本小說，翻成葡文後（巴西用葡萄牙語），指定要當天上市。

我楞住了：「今天上市？」

「怎麼？你是新來的？」編輯部的人一臉狐疑。

我不敢多問，反正按標準作業做就是了，心中有一萬個懷疑，那是基於在國內的經驗，出版一本書少說也要一兩個月！

我先作了大樣，然後將稿件送交「打字間」，記下時間是早上八點四十分，我告訴領班：「今天要上市啊！」

領班是個金髮女生，她說：「好的，中午前一定打好！」我怕她聽錯了我的中式葡語，又強調一次：「一共有廿多萬字哩！」

她笑了笑，挑戰似的說：「怎麼樣？十一點交給你。」

我不敢置信，跟她走進打字間，裏面有廿台「終端機」，只有十五人操作。領班進去隨意拈了拈稿紙，然後拆開分做十五份，分在打字架上，然後對打字員們說：「女士們！給中國人看看！我們兩小時內交卷！」。公司中只有我一個中國人，所以變成了我的名字。

十幾個小女生們一聽，誰都喜歡接受挑戰，莫不振作精神，十指如飛，每個人以平均每分鐘一百字的職業速度，果然在十一點鐘前全部打入電腦。

緊接著是拿著電腦印的文稿，送校對組，接著下來是我們的完稿工作，到下午四時我已把完稿全部送進製版部，我的責任全部達成！

為了想知道公司效率究竟如何，晚上我特別找一個較大的書報攤，由七點鐘等起，看著錶，大約七點四十五分，一輛漆著我們公司標誌的大卡車疾駛過來，連停都未停，一大包東西由車上擲下，我趨前一看，一點不錯！正是那本暢銷小說！

這是個新奇的經驗！假如一本最流行的暢銷書，能在一天內便由另一個國家介紹到自己國家來，那麼各種新奇的知識，有意義的知識，都能在最短的時機內，讓我們國人在同一時間，與他們同享！如果我們確知所獲得的是最新的，最有價值的一手資料，我們還落後什麼？我們還自卑什麼？

人性有個共通之處，比如說是籃球比賽吧！兩隊如果比數過於懸殊，彼此鬥志都會受到影響。輸得太多的人只求趕快下台，贏的人或許想多灌幾分，也可能顧全幾分面子，不論如何，賽程令人不忍卒睹。但是當差距在幾分之間，那就緊張了，誰也不會認輸，比賽完了，仍然是磨拳擦掌：「下次看我的！」。

如果我們請大量的留學生替我們在美國或全世界搜尋資料，翻成中文（我甚至希望有部電腦翻譯機），立刻透過這種類似的文化公司，用最快速的方法，用最合理的價格，把這些新知傳輸到全國每一個角落。知其一就想知道其二，一系列的、各行各業的、各種程度的、趣味的、益智的、源源不絕，接二連三地送到手中，有誰不好奇？有誰不願接受一些對他而言並不是完全陌生的新事物呢？

人所排斥的是他不瞭解的，人拒絕接受他所懷疑的，這是人之常情。一個陌生人到你面前，拿出一顆藥丸對你說：「吃下去你就會長生不老！」，如果你吃，你就是白癡！誰不希望長生不老？可是我們必須瞭解原理，進而証實其可行性。

國家的前途也是一樣的，有人說如此如此國家就富強了，如果他說的我不懂，我若相信我也是白癡！同理，我要說服某人某事如何如何，我也必須先讓他瞭解這一切的相關知識，如果我們面對的是全國同胞，要怎樣才能讓這麼多人瞭解呢？

這就是我所看到的答案，一個有效率的文化出版公司。不過，還有點技術問題，我們怎樣才能做到二小時打出廿萬字的一本書來？

找到了一條光明大道，不僅自己的生存有了意義，也能造福人群。我因此而離開了該公司，全部作業瞭然於胸，關鍵在於中文輸入。

在巴西沒機會及環境，於是我於一九七三年（民國六十二年）返抵國門，滿心以為會有人支持我的計劃。結果不然，為了謀生，也實地體驗一下出版狀況，我寫了兩本「冷門」小說，從頭盯到尾，一本書至少要二個半月，足足比巴西人的速度慢了七十五倍！（當然他們是有組織的做法！）

人不能靠理想維生，我一方面打工，一方面開始了「中文輸入長跑」，直到今天，很遺憾我的理想還沒有全部實現。人人視我為瘋子，即使這個瘋子一再証明了他沒胡思亂想，但在一百個成功之後（只是工作的成功，而非權威的成功），第一百零一個構想仍被別人視為「沒那麼簡單」。

更糟糕的是我不會賺錢，也不願發文化財，因之我永遠註定只能帶著我們公司這一批「不知名利為何物」的傻瓜，辛苦地默默耕耘！

閒話少提！現在我來談從開始的構想到今天，有關中文輸入的技術問題，以及其解決之道。如果你對枯燥的技術問題沒有興趣，我們就此說聲再見吧！已經辛苦你太久了！

在康熙字典上，中文有四萬多字，每個字都是一獨特的個體，字與字之間並沒有明顯的關係存在。

現在所傳衍下來的文字，是歷代的學者不斷地整理歸納詮釋，到今天才能有這樣一個規模。從東漢許慎的說文解字（其書共十四篇，收九千三百五十三字，以小篆為主，按字形義類分立五百四十部，每一字下先釋其義，次解其形，其下又引不同於小篆之古文、籀文、小篆等體及通行俗書，共一千一百六十三字。），到宋朝，有鄭樵的六書略，依六書分類；歐陽修的集古錄，趙明誠的金石錄，都對文字的整理分類，有很大的貢獻。清朝由於政治因素，研究文字學的風氣達到鼎盛期，有阮元作積古齋鼎款識，考識甚詳。此外，段玉裁的說文解字注，朱駿聲說文通訓定聲都是文字學史上貢獻卓著之書。民國以後，以羅振玉、王國維、章炳麟等亦均有其著述。到了今天，中國文字面臨新的挑戰---把文字用在資訊工具上---中文輸入的問題。

我在前面曾強調過，要解決中文輸入必須利用小鍵盤，小鍵盤上文字只佔26鍵，26鍵怎麼夠用？如果想要多用幾鍵可以考慮：1、標點鍵，2、數字鍵，3、大小寫字母分辨鍵。

標點鍵不宜用，理由很簡單：如果輸入一篇文章，最常用的就是標點。標點鍵要是被佔用了，用標點時，就必須用其他的控制鍵來辨識在什麼情況下是文字。那時，最常用的符號反而成為最麻煩的。故，此路不通，絕不能佔用標點鍵。

其次，數字鍵也有其困難，因為資料處理時，數字往往比文字多，與標點鍵不能使用的理由相同，此路不通。

再如將大小寫字母都應用上，又增加了26個符號，看來理想得多，尤其是用中文時，沒有大小寫的道理，為何不順便利用呢？這裏牽涉到一個問題，中文輸入是個根本的大計，關係到中文資訊處理的遠景。我們要知道，在遠距電傳打字機上，是沒有小寫字母的，我們既然要遷就現成的系統，又不能放棄利用中文電傳的功能，因此，最好不要用大小寫鍵。此外，大小寫鍵字母的轉換又要多按一鍵，用起來多麼的不方便！因此，只有一條最艱險的路可走了！不論用什麼方法，一定要用26鍵來完成。

英文是拼音系統，我國也有注音符號，且不論注音字母有多少，用注音輸入能不能考慮呢？

當然可以，電信局已經用得很成功，可是用在什麼地方與其成敗有相當大的關係。電信局查號非常適合，你一接通104，告訴服務小姐一個名字，你透過電話傳到她那邊的是聲音，只要讀音八九不離十，她立刻就完成了輸入的手續。但是由於國語的音只有一千三百個，而文字有幾萬個，重複率太高，所以你必須再加以說明那個名字的相關定義，服務小姐可以作第二次的選擇。

問題在一般性的資料處理，我們也提過「盲目按鍵」是先決條件之一，就以一萬五千字的應用來說吧！一千二百個音，其重複率已超過了百分之一千，再以「一」字為例，這個字的同音字約有一百，用什麼方法作第二次選擇？用眼睛？那就不符合「盲目按鍵」的原則了。

更嚴重的問題是，除非是專業人員，鮮有人能正確地拼出一萬六千個中文的讀音！不瞞你說，「別」字加上慣用的「鄉音」字，我個人所認識的五千個中文字中，有三千個讀音都不正確，你比我強多少？就算你正確拼出八千字吧！（專家不過如此！）還有一半的字你無法使用。再如希望能將所有中文全用上，四萬多！天曉得世上有沒有這個人？

很顯然，如果注音可行，中文輸入早解決了，別人不是傻瓜，動腦筋的中國人可多的是！

四角號碼呢？同理，如果可能，用不著等到今天。

天下只有一條路最實在，就是下死功夫，千萬別想找現成的、簡單的，你想想吧！難道只有你一個人認為中文輸入重要？

我的看法就是這樣的不通情理，我相信有不少專家學者已經下了功夫，已有了不少成就，但事實是沒有一種方法宣告成功。我在能力上，學識上遠不如他人，再要去參考他們的，學一些技巧，到時只會走上他們已開發出來的大道上，永遠走在他們後頭。

因此，我一口氣買了五本小字典，把裏面的字一個一個剪下來，不管部首，也不管筆劃，我存心要走一條新路子出來。

最初的構想是，文字既然是人用的，人要用在什麼地方呢？不外乎大自然的事物、人造的工具器械、人際的關係以及無法歸類的雜類。

很幸運的，一點歷史知識幫助我釐定一組最重要的「字母」，那就是漢儒所強調的「陰陽五行」。這是中國哲學，理化的基礎，對文化的影響極為深遠。我當時也希望這種影響能在文字的結構上呈現顯著的反應，那麼我就能減少一些未知因素。

這點幸運影響之大，簡直可以視為我成敗的關鍵，在其後幾年的努力中，幸而有了這七個字母恆定不移的信心，就如同天象中的七斗，居其所而眾星拱之，我才能用各種不同的組合去試驗、去分析，漸漸地，一個一個地，找出其他的字母來。

這七個字母也有點巧合，就是「月」字，月在文字中用得不多，而我把它視為陰陽的陰字的代碼，是否老祖先們早有遠見不得而知，竟然有個肉字旁寫成與月字一樣（至少演進到今天完全一樣了），於是日字成為陽的代碼，其他金木水火土分別就位。

有了七個字母，我們開始分類，先將與這七種「元素」意義上有關的字歸併在一起，大約已「解決」了四分之一。再下來就找不到任何線索了。每天把這些字搬來搬去，剪了幾十本字典，又排列組合了幾千幾百次，最後勉強訂出了第一套字母表，並定名為「中文形意檢字法」，於民國六十七年五月出版並於內政部登記在案。

第一套字母是這樣的：「日月金木水火土，人心手足口耳目，王石山虫魚犬馬，衣言絲草竹」，除字母外尚有輔助字形120個。用這些字母，我們把字典的8,000字經過編碼列序。如若每字僅取三碼，重複字高達8%，再若將輔助字形用輔助鍵（即等於增加26鍵），重複字則降0.5%。

至於字母的安排，我也曾考慮過與使用頻率的配合，（在後面將進一步分析）但「使用頻率」本身就很難定義。譬如說，據交大統計，「的」字使用頻率最高，那是根據報紙上的文章而來，如果用電話號碼簿做對象，怕找不到幾個人用「的」字做名字。同時在資料處理上，文章只是很小的一部份，絕大多數是人名、物料、數字，那麼，使用頻率應該以那一種為標準呢？於是，在初期我只能按分類來安排。

不論如何，我自知離理想的目的尚遠，用52鍵不能接受，用26鍵，8%的重複率太高，而使用頻率我也無從下手。充其量，到現在為止，只能證明這條路似乎可行，結果如何還有待繼續努力。

民國六十七年七月，經過好友林俊甫的推介，大周建設公司答應允支持我做進一步的研究工作，並為我聘請台大中文系的沈紅蓮來協助我。

這一次是為了實用，我們選定了國語日報大辭典的字彙為準，先將字彙編卡，再選取字母組碼。

這時，由於一個極偶然的機會，我在中華書局看到了王安電腦公司的三角號碼，其對我的啟示：編碼絕對可行，但是僅僅只為輸入的編碼便制定一套方法，而這套方法與該字本身的組成並無實際關係，這樣只能解決一半問題，另一半組字的問題還要頭痛一次。

怎樣能把中文字形組合起來呢？我假設有一套基本字模，如同英文打字一樣，將這些字模裝在槓桿的一端上，只要我能用輸入碼控制這槓桿，就能依序把字拼湊在一起。假定上述條件成立，那我應該如何開始？

這個假設很大膽，到最後我終於放棄了，因為困難太多無法實行。但是這個假設對後來發展出來的「向量組字法」有決定性的影響，因為如果沒有這種假定，我永遠不會找到組字的法則。同時，如非這個假定在先，我的取碼方式也不可能與組字緊密地結合在一起。

這是我在經驗中學得的一個定律：「要解決無數難以解決的問題，唯有先假設已解決了某些問題，再來專心考慮剩餘的其他問題。」

假定字模已有，多少字模是最少而最必需的呢？這個問題又牽涉到用字多少？以什麼為取捨標準了。

我再作一種設定，如果以一台打字機來考慮，常用字三四千個應該夠用了，比較困難的字暫時不去管它，先就組合率高的下手。

有了方向，問題就單純了，我先取第二代組碼出來的字研究，才發現當時的歸類方式大有斟酌的餘地。

因為要組合字形，就必須瞭解中文字形的結構，中文是方塊字，原由象形字轉變而來，那是因為工具演進與使用者的習慣所致。

象形字恰如其名，是先民根據事物的形象，以單純的線條簡化而得，誰都知道，畫圖不是人人能勝任的，需要觀察力及技巧的控制運用。因此，當文字的應用日漸普及時，就令象形中優雅美觀的圖形，慢慢地僵化、符號化，圓的變成了方的，曲線成為直線（你不妨試試，直線遠比曲線更容易畫）。

再其次文化的層次提昇，文字的表達面更廣，新字的需求應運而生，古人歸納了一套造字法則，即是六書：「象形，指示，會意，形聲，轉注，假借」。其中對後世造字貢獻最大的就是形聲字了。

形聲是一半用形，一半以音的功能結合字。由於象形字體本身已具備讀音，所以形聲字並沒有增加「字形」本身的複雜性，而僅僅提供了字形的組合原則。

時到今天，我國文字絕大多數是形聲字，根據這個原則，我也找到一個方向，那就是利用形聲字的結合原理，將字形分為「字首」及「字身」兩個部份。

這是一個新的突破，根據最初的分類，在四千個字中，我找了大約一百多個字首，三百個字身。以四百個字根可以組成四千個字，雖然還嫌麻煩，但已經小有進步了。更令我興奮的，這一百個字首與三百個字身所組合的字，實際上約有一萬多個已見諸於字典中，而我沒有收集在研究的資料裏。

受到這個鼓勵，我們再努力整理，設法找出其中最佳的組合。

有一點已經是肯定的，就是以廿六個鍵作為輸入的代碼，那麼在取碼時怎樣才能得到最佳的組合？要配合組字的規律，要顧及取的碼數少而又能避免重複字，又要考慮組字的要求，因為根據多年經驗，只要有充份的耐性、有人力、有時間，總有一天會找到一種最理想的字母代碼。可是組字問題如不先行解決，決定了代碼而不能將字形配合輸出，其結果還是一場空。

當我們最後用電報明碼作了一次分析後，我終於有了決定的根據。分析的結果是，字首240個，字身有1200個（很多字身只用一兩次），不能分為字首及字身的字多半就是字身本身，它也同時是完整的文字，這類字有2500個。

數據上證明，既然字首有400個，假若多收字，字首必然增加，但估計不會超過400個。我們已有26鍵，每個鍵一碼，26鍵只能代表26個字。因此，240個字必須是 26^2 ，換句話說，必需用26鍵取二次，二碼的組合有676種，足夠包含400個字首了。

我再解釋一下，若字首有400個，字首取碼必須大於400個，才能避免重複，因此字首取一碼至二碼。

再看字身，字身有1200個，同理，二碼又不夠了，只有增加到三碼，組合高達17576種，看來有些浪費，但是不這樣做就解決不了問題。

字身取一至三碼還有一個妙處，因為很多字身也是一個完整的字，其中仍然可以分為字首及字身。我們且假定為第二字首及字身吧！第二字首也可以符合字首取一至二碼的規定，在組字時妙用無窮，藉著這方法，我們不僅可以組合出現有的所有文字，還可以依循老祖先六書造字的法則，組合出幾萬種創新的文字來。至此，中國字不再是「死字」，它可以和拼音文字一樣活用，而且更具力量。

至於組字的技術問題，我們將會進一步地說明，現在我們必須回到輸入的原則上去，這樣你才會有個完整的概念。

大原則確定了，字首取一至二碼，字身取一至三碼，一個完整的字形，則是一碼至五碼。

在編碼技術上，最重要的是要容易學習及應用，要容易學習必須根據一套簡單的方法，絕不能靠死記。然而中國文字已經存在，不容改變，不論用什麼方法整理，都避免不了空碼的事實，唯一的例外便是回頭採用電報明碼的方式，按部首筆劃順序編碼。

我們且看看這種死記的方式有什麼優劣點？其優點是收集一萬字，就是一萬個號碼，絕對不會重複。電報明碼已施行了幾十年，如果能解決問題，豈要等到今天來頭痛中文資訊問題？

有人提出了一個醫治頭痛的方法，就是重新編碼。電報明碼字太少，加字不就得了？加多少呢？加到一萬六千字好了！至於一萬六千字夠不夠用呢？根據戶政資料統計，目前國內一千六百萬人口，人名用字是一萬八千多字，而且最妙的是，每月會「新生」字彙十餘個！

「新生」字彙？你沒聽說過吧！理由很簡單，新生的嬰兒需要命名，有些家長是飽學之士，拿出康熙字典來找字，誰能說這不是「中國字」？只是你沒見過而已！有些家長不識字，但他有權給他兒「畫」一個字，今天的社會一切依法行事，誰敢說這樣不合法？再加上鄉鎮公所工作繁忙，一切文件都要用手抄，一不小心，一個字多一點或者漏了一橫，白紙黑字，原始文件在此，別人誰敢改？一改說不定便是偽造文書！

今天如此，明天也不會好到那裏去，除非國家有套文字標準，否則字彙只有一天一天地增多。如果訂標準，誰來訂？訂多少字？康熙字典中的字我們認不認帳？要認帳，很簡單，最少也要四萬三千多個！

因此重新編碼雖然勢在必行，但絕非憑幾個主觀地選定幾個字，就可以視為標準。萬一位仁兄突然變成新聞人物，而他的名字未編入碼，你總不能叫他X X X吧！再往遠一步想，有一天我們要用電腦來整理固有文化，卻發現一大半的古字，電腦都不認識，那時怎麼辦？難道怪我們的文化太不識相？

此外，死記絕對行不通，目前八千字的電報明碼的專業訓練是半年，不僅半年是段漫長的時間，那份死記的苦功怕也沒有幾個人受得了。假若將四萬個字連續編碼，老天！要學會它恐怕不是電報明碼的五倍，而是五十倍了！

有專家認為，至少該有套「交換碼」，不用人來記，只需放在電腦記憶體中，不論各家電腦用什麼碼，有了交換碼為標準就可以互相通聯、交換資料。

這也是中文資訊的奇妙現象之一，同在一個國家體系之中，用同一種文字，可是彼此之間卻要用「交換碼」，你或許聽說過這個名詞，也可能沒有。如果我們用一個例子來說明一下，可能你也會啼笑皆非。

且假定每一個電腦代表一個「區域」吧，外國「區域」很單純，全部都用英語，可是我們大中華民族，地域觀念特別濃厚。因此，每個「區域」要用一種語，也就是說，每個區域間的人民，都無法用彼此能懂的言語溝通。

這就是目前中文電腦世界，每家都有其「特定」的輸入方法。輸入的資料，除了同種的電腦，任誰也無法共用...，這是無謂的浪費。同樣的工作、同樣的資料，在不同的系統上，就要重複地做一次，甚至無數次，今天國家的人力財力都有限，怎能這樣下去？因此，遂有了交換碼之議。

可是「交換碼」行得通嗎？國科會努力過，結果卻因某位自命高人一等的人士，堅持要用他的方法一統天下而告流產。其實，交換碼只是把一般用的文字，依某種順序排列一下而已。既不能當輸入碼用，放在電腦中也徒增空間。但，如果有它，至少在輸入方法未能統一之前，不失為可行的「臨時通道」。不過，它必須有個先決條件，就是要有包羅所有中文字彙的「雅量」，只憑在字典上抄16,000字就號稱交換碼，請想想，其他的三萬個字難道不是中國字？我們不認識是我們學得不多，我們不用是我們用不到，但是，代表國家的交換碼，怎能只顧目前的需求，只因暫時能力不及，就斷然否定所有的傳統文字？

當然，我們今天的生存不能忽視，可是，難道沒有更好的方法嗎？如果大家肯捐棄私見，不計名利，共同公開討論研究，公道自在人心，我相信必然有個合理的結果！

不論如何，我決定採用自然編碼法，取五碼的好處是它的排列組合有一千多萬種，任何中文都可以容納進去。又有人認為五碼太浪費，我不同意，我計算過三萬五千個字彙的平均取碼是4.1碼，英文據一般統計是4.5碼，為什麼沒有人說英文浪費？同時，節省儲存及傳輸的方式很多，片語、縮寫字都是最有效的方法，在一千多萬種組合中，我們正可以充分應用這種功能（假如你已用過我的倉頡輸入法，會發現我實際上只用了24個字母，另外兩個就是用來做這種工作的，我還在發展一種片語資料庫，有24^4種片語功能，卻只佔64KB！）。

根據我的理想，中文應和英文一樣，輸入碼、輸出碼、內碼及傳輸都是一體的，這樣才會發揮其最大的功能，也才是一勞永逸治本的良策。

這些原則確定，只剩下一件事要做了，就是選出能代表字首字身的代碼來，最後挑出字母。

這一步不難，尤其是現在我有了電腦，可以節省大量的時間、人力，但在當時，每假設一字母後，所做的工作如次：「將一萬字的卡片一一編碼、校對，編妥再將卡片依字母順序排好，再將同碼的所有卡片找出，研究其同碼的原因，以及避開這種同碼字的方法，最後再訂一套字母。」

週而復始，一個月最多能整理1.5次，同碼重複字並不難減少，難的是所有字母必須有分類上的意義，不能只為避免重複字而決定。除了字母必須有意義外，輔助字形必須與字母能扯上令人容易聯想的關係，否則不易學習，難為人接受，而減低其價值。

有人認為重複字的組碼法才最理想的方法，如果能透過很合理的規則將重複字完全避免，那當然好，但只是為了消除重複字而削足適履就值得商榷了。以「晚」及「晃」為例，在文字學上都是「日」，取碼相同，只有看位置，一個在左，一個在上可以分辨。為了這個理由把所有在左側的「日」都加個「左」鍵顯然是浪費，只加「取碼相同」的字，那又與「重複字」的定義有何分別？

為了節省篇幅，其中的過程就不用提了，我只在此把我曾印書出版的第二代，第三代與現在正在使用的，列表做個比較：

第一代 日月金木水火土，人心手足口耳目，
王石山虫魚犬馬，衣言絲草竹。

第二代 日月金木水火土，人心口耳手衣竹，
交叉紐斜點縱橫，廿田山十卜。

第三代 日月金木水火土，人心手口田草竹，
交叉縱橫紐斜點，衣西山又卜。

現在 日月金木水火土，斜點交叉縱橫鉤，
人心手口，側並仰紐方卜。

由上表中可以看出，除了日月金木水火土及人心手口外，其他都曾有不同選擇。同時，由第二代起，一個新的構想加入了，就是筆劃，到了最後一代，筆劃更擴大為成為字形。

這種轉變是經過統計分析得來的，但也恰好與理論配合，成為一種很自然的結果。

在理論上說，文字是人用符號藉以表達思考概念的，人的思考概念首為哲學思想，次為人生，對於這兩點，我們整理出日月金木水火土，人心手口作為字母。另一方面，中文是以筆劃及形狀來表達其差異的，在筆劃中，我們選擇了斜點交叉縱橫鉤，在形狀上，我們選擇了側並仰紐方卜，應用時也頗差強人意。

現在的26個字母，留下X，Z二鍵留待擴充之用，在我的構想中，有三種擴充的方式：一是重碼字，我以X鍵（在宏基的天龍系列上是用的Z鍵）來處理，目前我們重複字約1%，但若來者不拒全部收齊，可能會增至2%，一鍵可夠應付。其次為縮寫字，譬如說「金屬氧化半導體」這七個字在應用上極不方便，我們可以視為金氧半的縮寫，寫成「金氣」，簡單明瞭。「金屬結晶」寫成「金晶」也是順理成章又有何不可？（當然我們不能越俎代庖，這只是舉例而已！）

前述這些字是我們的字形產生器第二代功能，它能組合二百至三百萬個「怪」字，但絕對符合取碼規則，這些字一點也不佔空間，用不用隨你。

第三種擴充的方式就必須增加一片64KB的模板，我們要預先選定所有的專有名詞，常用的片語等字數在二字以上的，多多益善，用最精簡的資料結構方式，我估計至少可以存24^4種，而在使用時當作一個單字用，只是首碼為Z而已！

這一來，中文的傳輸及儲存將是精簡無比！

再談談字母各鍵的安排與人體工學的關係，由於所收的字很難有個標準，有人主張以最常用字（據新聞報紙用字統計，約有三千字），也有人主張以學生字典（約八千字）。不論一種，都有缺點！因為對象的不同，字彙的使用頻率就互異。一種編碼方法，要用各種情況，就不能以偏概全，我認為面面俱到是不可能的，只要能符合大前提的要求，已經是難能可貴的了，我的前提是這樣的：

1、字母的分類牽涉到原理、記憶與使用，絕不能因為任何單獨的理由而令之支離破碎。在分類中，基本類有哲理及筆劃，次元的有人體及筆形。要變動只能在各類的字母中，前後調整。

2、要使字母朗朗上口，就必須顧及音韻，尤其在未來使用時的字母排序上，要充分利用字母的背誦性，才能順利地發揮其功能。舉例而言，如果你不會背九九乘法表，對做算術一定很不方便。同樣地，如果你不會背英文字母，查英文字典也會增加不少困擾。因此，中文字母一定要會背誦，要能背誦就必須顧及其音韻的抑揚頓挫。

3、在人體工學而言，食指用得最方便，其次為中指，最難控制的是小指。對左右手而言，絕大多數人慣用右手。因此，右手手指又比左手手指方便。在鍵盤位置而言，文字鍵有三排，中排最佳，因為打字時，係以中排定位，字鍵使用的頻率高低應以前述因素作通盤的考慮。

有了前面三個前提，再去根據組碼規則，選擇取碼的樣品，一一分析比較，其過程之艱辛，不足為外人道。更由於我選擇的樣

品是康熙字典中的35,000字，而這35,000字的取捨也有待進一步的斟酌。雖然最後確定了字母的順序，我仍然覺得不夠理想。

為了測試字母順序與人體工學上的配合是否妥當，現在且用教育部頒訂的4,803 字為準，編碼統計於後，以供參考。（註：由於編碼小有疏忽，多收了九個字，致成為4,812 字，但無礙全部取碼的正確性，故未加更正。）

在表中，比較不符人體工學的有「N」鍵（右手食指下排）以及「K、L」兩鍵相反，這是基於前述第二個前提的考慮所致，「O、P、Q、R」四鍵係分類的限制，其餘尚差強人意。

字母	一碼	二碼	三碼	四碼	五碼	小計	使用頻率
日A	96	137	139	121	30	523	2.70%
月B	220	235	336	209	85	1085	6.61%
金C	128	116	101	135	106	586	3.03%
木D	194	191	164	143	38	730	3.77%
水E	267	78	62	124	73	604	3.12%
火F	116	293	100	111	96	716	3.70%
土G	126	173	159	158	52	668	3.45%
竹H	366	347	425	216	39	1393	7.21%
戈I	272	283	229	219	85	1088	5.63%
十J	186	247	158	133	74	798	4.13%
大K	91	142	152	180	44	609	3.15%
中L	132	196	224	161	34	747	3.86%
一M	308	314	415	270	57	1364	7.06%
弓N	165	186	206	119	36	712	3.68%
人O	313	323	255	165	78	1134	5.87%
心P	124	123	177	109	37	570	2.95%
手Q	235	83	63	58	24	463	2.39%
口R	264	299	226	256	115	1160	6.00%
尸S	156	154	170	120	39	639	3.30%
廿T	271	177	179	121	38	786	4.06%
山U	55	174	130	263	115	737	3.81%
女V	233	71	98	133	52	587	3.03%
田W	69	132	172	106	18	497	2.57%
重X	87	24	20	23	7	161	0.83%
卜Y	338	289	224	81	26	958	4.95%
小計：	4812	4787	4584	3734	1398	19315	
每字平均取碼數： 4.01							

至此中文輸入的說明暫告一段落，否則我們的篇幅就不夠了。

我前面說過，輸出之假定是用槓桿打字頭，很明顯的其中技術困難重重，幾百種字鍵倒可以解決，如何使字鍵放大縮小，改變形狀，改變位置呢？

我當初用的只是一種工作的方法--- 各個擊破--- 並不是說將未來的希望全寄托在打字的技术上。

我另外有一個構想，是用光學透視的方法，可以將各種字形經過反射透視鏡組合起來，用在打字機並不困難，後來發現採用縮底片，幾千個字都可以濃縮在一片底板上，再經輸入碼改換成座標位置，即可將字形輸出，技術上確無困難。事實上我也開始動手設計，但問題出在輸出之印字頭，必須用半導體技術來解決，也就是利用半導體將光轉化為熱，再用熱印在熱帶上，熱色帶受熱會將字熔在普通紙上。

研究本身是相當大的投資，我費盡心力希望做出便宜而經濟的打字機來，要達到這個目的，我研究出來還是不行。必須能工業化大量生產。我先做好實驗，再找人投資，一百個投資者有一百個人要求專利保障。因之，我拿熱字頭及熱色帶去申請專利保障，不幸被批駁了。理由是樣品做得不夠好，不夠成熟。

大家都知道熱印字頭目前是最便宜最快速且無噪音的印字方法，但是卻無法普遍被人接受。原因無他，只為了熱字頭必須用熱感紙。假如一台打字機只能用一種特殊紙，它的性能再好，用起來也不方便。因此熱色帶如果得到專利，它本身就是一件極有市場的商品，它可以使熱打字頭取代目前大量被採用的衝擊式打字頭。

朋友們！我已心力交瘁，我申請專利的目的是希望不被一些唯利是圖的商人壟斷，但是既然得不到專利，我寧願你們快想法做出來，至少不要落在外國人手中，再來賺我們的血汗錢！

熱色帶的試驗很簡單，當然商業用品要考慮其配方及多方面的性能，但試驗其可行性卻簡單不過：「用最薄的塑膠紙（最理想是鋁箔，是熱的良導體），大約0.01MM至0.02MM厚，上面塗一層易熔的色素如同腊，或者用奇異墨汁（那種易燃的），用時可將鉛字加熱，在背面輕輕一壓，字形就印在紙上了。」

如果你有錢，希望你快先申請國外專利（你有了專利，不妨賣給我），國外的有了，國內或許有希望。不過你要注意，一般熱印字頭只有攝氏60度的瞬間溫度，你必須多方試驗，降低色素熔點。再不然就得提高熱印字頭的溫度，那就要用到半導體的技術了。再談中文字形吧！由於前述專利受阻，無人敢投資，我又無力生產，因此，中文打字機一直無問世。幸而微電腦適時大行其道，我便將全力放在這方面。

英文的字形出現在電腦上，是利用座標點的掃描完成的。一般說來，電腦顯示器上有512*256點，或者是小一點的256*192點，再大一點640*512點，這代表些什麼呢？我們叫它顯示密度，也就是指我們所看到的光點數。以512*256為例，512點是指顯示幕上從左到右可以顯示512個光點，相當於有512個小燈泡，可以控制其明暗。256則是指上下共有256排燈，換句話說，就是一共有256排，每排512個小燈的顯示幕。

那麼英文字怎樣控制呢？最普通的就是用5*7的燈陣來表示，我們再用前面用過的0與1的方式，寫一個「E」字，只是為了讓你看得清楚，我們用「.»代表「0」，用「鬱」字代表「1」，其結果是：

```
鬱鬱鬱鬱鬱
鬱 . . . . .
鬱 . . . . .
鬱鬱鬱鬱鬱
鬱 . . . . .
鬱 . . . . .
鬱鬱鬱鬱鬱
```

我相信你一定看得出這整體的視覺反應，是個英文字母「E」字，在顯示幕上，表示1的地方燈就會亮，0則不亮（當然也有反過來的做法）。

由於英文字母很簡單，5*7的點陣就能把明暗的特徵表達出來。其原理是，每一筆劃與另一筆劃之分辨，全賴二者之間的反差現象。明確地說，字形之辨認，端視對比色的排列訊號與我們視覺感應所產生的效果。英文字母之特徵在於不論橫向縱向，其對比差異不超過五次，以筆劃論則不超過三劃，因此，他們只要5*5的點陣就夠了。

你會問，能用5*5為何還要用5*7呢？那是由於美觀及避免誤認所致，由於英文字母中有斜線也有弧形，這些用單純的橫直點來表示不夠美觀。再加上還要用阿拉伯數字，其他符號等，太簡單易生誤會，5*7的變化多了三分之二，辨識性就理想得多。

當然，5*7並非最理想的，只是最經濟的而已，點陣密度愈大愈美觀，辨識性也愈強。現今美日高級的電腦，也已逐漸走高密度點陣的路了。尤其是日本，他們為了要與漢字共用，英文字母已用到11*24。

中文呢？中文該用多大的點陣？這又得分幾方面來說了。首先我們要知道，最經濟而又能分辨大多數中文的點陣，以多少最理想？其次，我們把全部中文收羅進來，又應該是多少？這樣我們才能做詳盡的分析。

最經濟的條件是要配合電腦的結構，我們前面說過，電腦的最基本辨識單位是字元，也就是八組的開關，在此我們將之正名為八個點位或位元。電腦顯示幕上的光點，是電子束在陰極射線管中，受到水平向及垂直向的控制後，所產生的座標位置。正常情況都是先水平掃描再一行行地向下移（也有將陰極射線管橫放，以致先見到垂直動作的），由於電腦以八個點位作為一單位，因之橫向最好用八的倍數，縱向雖比較不重要，但如能用八的倍數，則更方便。

為什麼呢？由於電腦設計之初，以英文為主，英文只要5*7的點陣就可以了，而且英文及符號只要94個，全部放在一起，直接用點陣儲藏，其空間要多大呢？94*5*7是3290個點，而八點為一位元，那麼3290/8為412個位元，也就是相當於0.4千筆可以放到一片小小的積體電路器中。

這片積體電路器都是放在顯示器線路中的，一旦某一字母的訊號傳到，立刻即可將相關位置上的光點訊號送出，直接了當。中文不然，我們無法用英文的這種方法，原因是字太多了，顯示器本身是個附屬機構，一切作不了主，主系統那邊要遙控的話，也有一定限度。像中文儲存，一萬個字最起码要320千筆，誰也控制不了。

因此，還得想別的辦法來處理，方法之一，也是目前最理想的方法，就是利用繪圖的功能，把中文當作一個圖形來處理，這一來，就必須是個能繪圖的電腦了。（注意市面上有很多電腦都號稱能繪圖，其中有很多是假的。這種稱為SEMI GRAPHIC，他們是先設計好一些圖形碼，視作英文符號之一，然後可以藉積木方式拼出圖形來，而實際上，並不是真正的圖形處理。）

有繪圖功能的電腦，在能力上就比普通電腦要強，當然成本就高，其中必須加一片「螢幕記憶板（SCREEN MEMO）」。它的功能是先要把圖形畫在其中，然後再用掃描方式傳到螢光幕上。

很顯然，把記憶中的圖像搬到螢幕上，就必須配合電腦傳遞的方式，電腦既然用位元為基礎，你用一個，兩個或三個位元都沒有多

大問題。但是如果用1.3 個或2.5 個位元，那可就麻煩了。

舉個例子來說，你上學時總要揹書包吧！書包是根據書本的大小設計的，你今天如果設計一種書，當然要選就書包大小，如果你設計的書比書包大，別人帶起來不方便，除非你不在乎別人用不用，否則不是你替別人改書包，就是你改書本的尺寸。

其次，我們還要考慮中文字的筆劃到底要用多少點才適合？我們只能這樣說：「上下十六點可以分辨的字約佔80%，十八點可以辨識的字佔85%（上下可以不採八的倍數），廿四點可以分辨的字佔99%。左右八點能分辨的字佔30%，左右十六點可分辨的字佔93%，廿四點可分辨的佔99%（以上數據僅憑我們發展各種字體所得的實際經驗，尚有待學者們整理）。」

點數愈多自然愈理想，但是成本相對提高，在我個人的考慮中，我選了16*16 作為低成本的用途，24*24 作為一般高級電腦之用，還有32*32，可供自動排版印刷用。

現在我就來解釋16*16 的作法與相關的技術原理。

16*16 實際上不能真正地令左右打十六點。各位知道，中文字形有很多是左右對稱的（這完全符合人的生理學，因為人的眼睛一左一右，形狀大小相若而方向相反，但上下則不然，所以你如果深入研究下去，會發現美感之產生與生理息息相關。），我們以一個「十」字為例，「橫」高一點、低一點影響不太大，但「直」卻一定要在中間。

如果直在中間，我們且假定直佔了橫向中間一點，左邊的橫再假定為七點，右邊亦然，總共是十五點，還剩下一點怎麼辦？加在右邊？左邊？或者把直加粗？

單線體是不允許任何一劃比別的筆劃粗的。因此，我們為了左右對稱，只能用十五點，至於上下，我們仍用十六行，因為橫在正中間反而不好看。一般說來，在中文的美學觀念上，上密下疏較為好看，你看看人的面孔就知道，如果眼睛眉毛橫在臉的正中間，是個什麼樣兒？

市面上有幾種電腦字體，左右還是用十六點，那些字是美術專家寫的，一個一個地寫的，一筆一劃都很講究。但是你看了會不知為何，有些「怪怪的」感覺，說穿了，就是因為左右不對稱，你不習慣！

一行多出了一點，十六行就十六點，這不僅不是浪費，卻正是我能用最空間，收容幾萬個中文字形的最大秘密。把這一點學去了，我這八年心血也就變成你的了。

在武俠小說中，常有什麼「武術秘笈」的爭奪，無論成名多年的大俠，獨霸一方綠林，以及後生小輩，莫不搏命以求。我年輕時也做過夢，滿以為得到了一個訣竅，立刻可以學貫古今，其實那真是夢，而且是癡人說夢！訣竅是有的，懂了可以節省很多無謂的重複浪費。但是，不去下苦功，把這個訣竅與其他的相關知識融會貫通，那麼訣竅還是訣竅，它不會自己變成學問的。

因此，你如果真有心，我希望你看了下面的介紹後，還要多下功夫，進一步追求各種相關的知識，功夫下得愈深，你所得的就愈多，才能靈活運用。

前面介紹過排列組合的概念，不過還不十分清楚，我們再簡單解釋一下：「當我們設定每行有十六行時，我們就有16*16 個點，也就是256 點。換句話說，我們有了256 個位置。由於我們要在這256 個位置中畫圖、要寫中文，我們就必須賦與每個位置一個名稱，以便處理。」

命名的方法很多，顯然在這裏用數字最適合，而且要用座標的方式先取X 軸的點位，再取Y 軸的點位，我們可以順次命名為第一行第一點，第一行第二點...。

再把文字省略掉，我們可以寫11、12、13...，不過你看得出來超過10的點位就變成110、111、112...，我們無法知道11是指X 軸或Y 軸，更好的辦法是，統一取二位數，如0101、0102、0103...0111、0112...0201、0202...。

這是十進位的寫法，對電腦而言，十進位非常浪費，一個位置要四個字元來表示！

其實，有種方法，在16*16 的點陣中，每個位置剛剛好只要一個字元就可以表示，要節省就要充分利用其有利的特徵，這正是我們採用16*16 的最大理由！

這種方法是二進位的延伸，叫做十六進位，是瞭解電腦最重要的一步。因此，我不厭其煩地把十進位，二進位及十六進位的對照表列在後面，以供比較及參考：

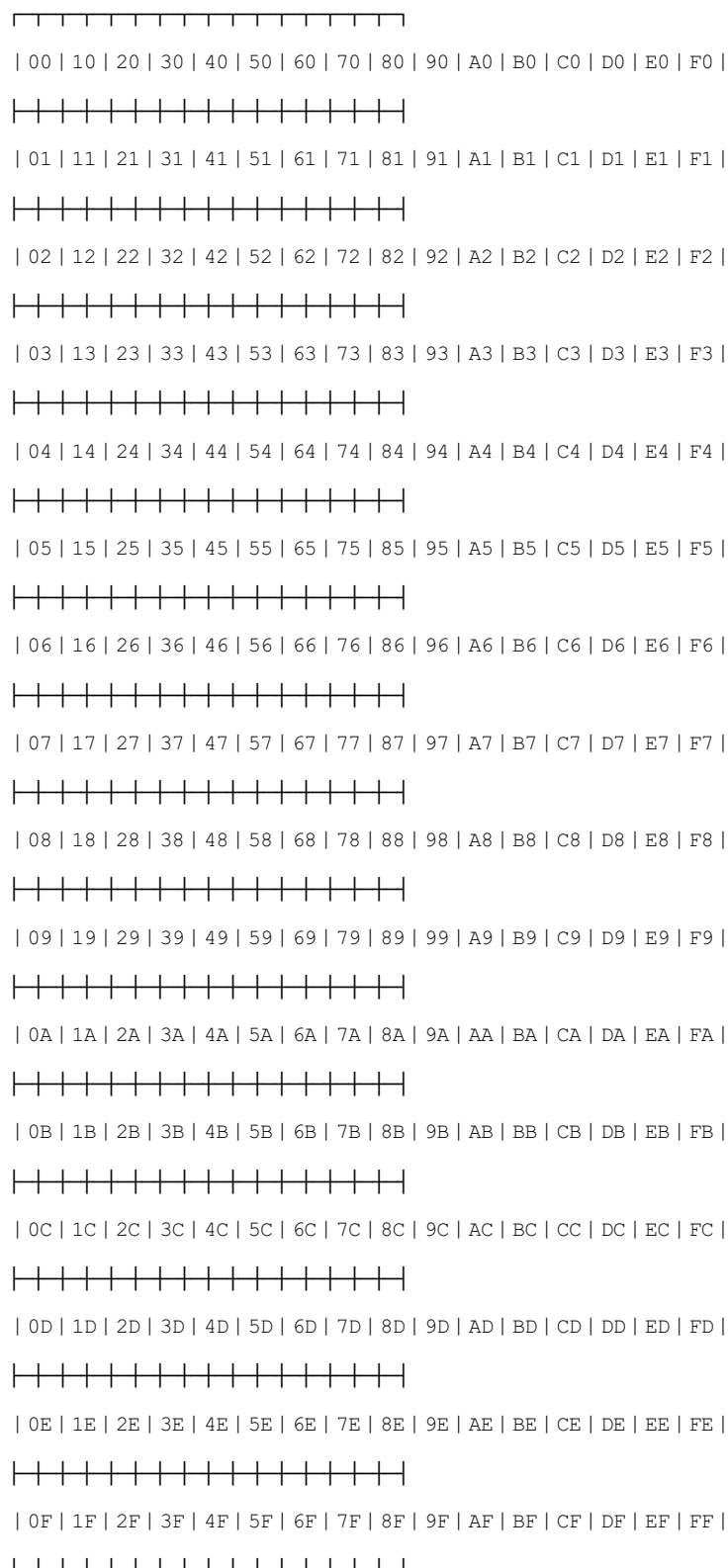
十進位	二進位	十六進位
0 0	0 0 0 0	0
0 1	0 0 0 1	1
0 2	0 0 1 0	2
0 3	0 0 1 1	3
0 4	0 1 0 0	4
0 5	0 1 0 1	5
0 6	0 1 1 0	6
0 7	0 1 1 1	7
0 8	1 0 0 0	8
0 9	1 0 0 1	9
1 0	1 0 1 0	A
1 1	1 0 1 1	B
1 2	1 1 0 0	C
1 3	1 1 0 1	D
1 4	1 1 1 0	E
1 5	1 1 1 1	F

你不難發現，在0～15，二進位只佔四個位置，或稱四位元，十六進位是一個數字，而十進位則要兩個數字。（實際上，十進位通常要佔一個位元即八個位置來表示，這又牽涉到其他枝節，這裏恕我不多做介紹。）

電腦雖然都是用二進位的開關方式操作運算的，但是一大堆的0與1，由人去辨識很容易混淆（記得我們前面提到的中文與英文字母之比較吧？變化太少不見得適合人類複雜的頭腦），所以人就把十六進位給自己用，而讓電腦它們去玩0與1的開關把戲！

各位也看出來了，前面我們說的16點，正好可以用0到F來表示，16行也用0到F表示，那麼我們就可以寫成01、02、03...0F、FE、FF。每對數字一個位置，同時由於每個16進位數字只佔四點，所以每對數字正好八點。換句話說：「16*16的點陣上，每個位置正好以一個位元來表示。」這時，一切都明白了，我們若要劃一橫，可以指定由某個位置到某個位置，直也可以，斜也可以，只是要注意：「斜只能畫對角線，否則不好看，也畫不出來。」（有種畫法是把斜分成幾段橫或直，見仁見智，但我沒有這樣做。）

你不妨用下面的附圖，按其中的位置寫出你要寫的字來，然後把座標值讀出，就是電腦字！



你照這個方法，會發現一個字平均要費大約卅位元，如果以儲存二百個字首，一千五百個字身來算，大約要51KB，而這51KB可以

組合出 $200 \times 1500 = 300,000$ 個不同的字形哩！

以上這種方法稱為「座標法」。

假如你還想節省一點，方法也有，你得花很多功夫，把各種字形中，形狀大小相同的各個「零件」找出來，集中放在一起，要用時就拿出來，那樣又可以省出30%的空間來。原理很簡單：「以一個「零件」（我們姑且稱之為共用字形吧）佔廿位元計，如果有十個用到它，你必須耗費二百位元（這種字形在我的統計中，平均每個共用字形約八位元，每字平均共用四次，約有500字），如果這種字多了，你的空間就很浪費。」假如你採用共用的方法，該字在應用時只需寫出其位址（在記憶體中，一般是以64K為一頁，64K正好可以用二個位元來定址，如同256中用一個定址一樣，因為 $2^8 = 256$ ，而 $2^{16} = 64K$ ）即可。因此，該字需20位元，用十次，每次只需2位元，共20位元， 20×20 為40位元。本來需要200位元的字形，現在用40位元，已經節省了160位元。

如果你還想節省，也有辦法，你要怎樣節省都可以，不過有一點要考慮清楚，空間與時間往往是兩個極端，空間的節省除了技術因素外（意思是說，技術上必需不能作無謂的浪費），只有用時間來交換。因之空間小了，時間也就相對慢下來了。速度太慢無法供電腦處理應用，這其間的微妙關係不是幾句話可以說明，這必須你能完全掌握住所有細節後才能明白。

我還用了一種指令的節省法，就是用一些指定的指令，來處理一些常用而特殊的字形，不過這種方法，你必須對基礎組程式語言相當熟練才行，因為你必須知道如何「定義」指令。

有一種方法我可以介紹給你，是一種效率最高最容易處理的，而且正是我前面所提到的武林秘笈……九天玄陽真(无,音:氣)……！當 16×16 之點陣可以用一位元來定址時，也就是說一位元可以代表256點中之任何一個位置，那麼這一位元視為 16×16 點陣中字形的結構資料。結構資料這個名詞是我暫定的，我們必須用基礎結構程式來處理它，才能把它變成我們肉眼能辨識的字形。

當然，在此我必需假設你已瞭解基礎結構程式，否則我的困難便多了。

前面也提過，在 16×16 的點陣中，我實際上只用了 15×16 ，那多出來的16點……也就是16位置座標……怎麼辦呢？正好，我們恰恰需要16個特別的法寶，每個法寶有它獨到的功能。譬如說：「甲代表一種形狀，乙又代表另外一種，丙後面跟的資料需當作住址用，丁表示資料的位置要移動……。」

這一來，甲乙丙丁等就可以和結構資料放在一起使用了，這個方法又將空間縮小一倍！

你會說：「這樣該夠了，不到20K就可以存這麼多字，64K都用不完嘛！」

別急！這些只是資料，資料還要經過處理，在處理的過程中，還要增加不少東西的！

要增加的最重要一項是例外字，我們談了半天字首及字身的組合，有些字並不是那樣輕易地就可以用字首字身來界定的，這些字也正是最討厭的一部份。在四千個常用字中，這類字約佔60%，而在一萬字中，這類字降低到30%，到了三萬字時，這類字只佔15%。

你會覺得奇怪，為什麼選字愈多，例外字愈少？理由很簡單，前面也提過，在英文文法中，幾乎所有不規則動詞都是最常用的動詞。同理，愈是歷史悠久的文字，愈是沒有規則。此外，我們也提到過，象形字是我國文字的主體，後人根據象形發展的形聲字，才是我們採用的準則。因此，幾乎由象形字蛻變而成的字都是例外字！這種例外字，大約要佔25K。

在組合字之時，我們根據取碼規則，輸入了字碼，程式立刻就要先去找例外字。如果是例外字，立刻要按照例外字的規定處理；如果不是例外字，這才分析輸入碼為何為字首、為何為字身。由於字首字身早已確定，已經存在記憶體中，所以不難分辨。

現在到了最複雜的一個步驟，每一個字首有它不同的位置以及不同的點位大小，同一個字身怎麼知道應該如何配合呢？

我的方法是：「把每個字首先分析，寫好字形，然後把剩下來的空位用指令表示之。」也就是說，每一種字首有一個特殊的位置指令，用來指示字身應該放在那裏，應該有多大？

而字身則每個字身又配有一種刪減指令，每一個刪減指令都要經過測試，不能把筆劃刪掉了，也不能把筆劃刪得重疊在一起，千方百計，要刪得均勻、好看。有時為了刪減的便利，連原字形也要改變，以便適合刪的要求。

舉個例子，「十」字很好刪，當它和水字旁組合時，三點水在左右向只佔五點，為了要組合必須刪掉五點。可是「十」字要對稱，只能刪四點或六點，刪四點很好看，但「十」又可以和金字旁組合，金字旁也是五點，如果只刪四點，必然會有一點重合，那就難看了。因此，為了統一，「十」字只有刪六點，正好在字首及字身之間留一行空。

刪的時候，可以把左邊1、2、3行及右邊的13、14、15行去掉，便只剩下九點的一個瘦十字架了。這時再把瘦十字移到右邊，加上水字旁或金字旁即可。

如果我們要刪「贛」字以便和三點水結合時，那可麻煩多了。「贛」字非常複雜，所以在寫這個字的時候，要特別小心，不僅要刪得好，不破壞字形，還要考慮它用的刪減指令能不能與別的字共用。如果刪減指令不共用的話，問題也很大，因為有一千多個字要刪，字首的種類又多，如果每個字有三種刪法，就將近有四千多種，加上每個刪減指令至少要佔三個位元（第一個是指令，第二個及第三個位元是各刪減的位置。）那又要花費12K的空間。

在我的記憶中，那是幾年痛苦的經驗，刪減指令最難節省，在第一代的字形產生器中，刪減指令佔了8K的空間，第二代繼續努力，字增加了不少，但刪減指令只佔4K，這原是一件無法討巧的工作，只有靠一步一步的嚐試改進。

字形組合完成了，又面臨輸出的問題，這雖只是技術上的小枝節，可是也影響到速度及效率。我們常建議電腦生產廠商，把中文字形產生器放在顯示器及印字機中，這樣這可以避免傳輸速度的影響。

一般在輸出介面中，不可避免地要考慮輸出入通道的選擇，訊號的安排與控制。這好比在一個工廠中，生產部門把產品造好了，他該知道把產品送到什麼地方去，是送去倉庫（磁碟機）？或送去再加工（系統記憶）？是送去外埠（印字機）？還是送到門市部（顯示器）？

送到不同的地方去並不是一句話，工廠有很多手續要辦，首先要經過品管合格（這由應用程式來負責）、審查訂貨單（由操作系統控制）、準備運輸工具（即輸入通道，也由操作系統控）、發貨、收貨、驗貨（都由操作系統統籌處理）。

這些工作都需要繁雜的手續，費時費事，要想做到迅速確實，就牽涉到很多既有的硬體功能與設計。以目前而論，一般微電腦的能力，多半是每秒9600個位元，也就是1200個字元。說具體一點，就是在傳輸資料時，每秒鐘最多不能超過1200個字元。

我們前面說過中文用 16×16 的點陣，正好是32個字元。因之，每秒鐘最多能送30個中文字形！

每秒鐘30個字，如果一個畫面有600字，就要20秒鐘才能傳送完畢，豈不把頭髮等都白了？

英文為什麼快得多呢？並不是英文字母點數少（其實以一個「字」而言，差不了多少），而是英文字形產生器（一片積體電路器）就放在顯示器中，在傳輸時只送代碼，一個代碼一字元。因此，英文每秒鐘可以送1200個字母，相當於240個字彙。

如果中文用代碼傳輸，把字形產生器放在顯示器裏，也可以達到每秒鐘240個字彙（當然，先決條件是中文字形產生器有這個速度。目前，我們第二代的字形產生器，每秒鐘只能組合120字）。

說了這麼多，相信你也煩了，可是還有一點不能不提，就是組字程式。由於這種程式屬於中文電腦的基本結構之一，使用者沒有必要去變動它。而且，其性能的要求特別高，所以需用基礎結構程式語言來寫。

我們目前的發展系統用的是Z80 CPU，當然就要用Z80結構語言了。除了Z80外，還有8080、6502、6800等。更上層樓還有8086、28,000，以及68,000等十六位元的結構語言，由於它太複雜，決非幾句話可以說明。因此，程式的內容，恕我在此不加以介紹。

我們要強調的，只是在寫作技巧上，很多人以為結構語言是外國人創始的，他們寫的一定好，這種觀念我不能苟同，我們曾經分析研究過許多高價買來的外國原始程式，發現其技巧不見得比我們強。原因很多，最基本的一項是需求問題，因為用結構語言寫程式

的人太少，而要做的工作太多，沒有時間及必要去一而再，再而三的修改。

對中文組字程式而言，我個人認為不僅要不斷地改，要改得精簡、容積小、速度快，還要改得功能強、應用範圍擴大。因為，既然中文電腦比英文多一片中文字形產生器，成本就高了一點，如果我們不充分去利用它，用到它鞠躬盡瘁，那就是浪費！

一般寫應用程式的標準做法，是先做系統分析，畫好流程圖，再畫細部流程，最後才依照細部流程來寫程式，這是完全正確而且極有效率的方式。可是，在寫中文組字程式就不太適宜。因為應用程式的要求是在最少的人工下，以最有效的速度，來完成一個程式。這個程式是一種商品，商品並不能要求完美，只要求能合用，能賺錢。中文組字程式不然，它是中文電腦的心臟，也可能是千秋萬世的工具，它必須完美，必須毫無缺點。