

中文資訊處理系統的選擇

資訊與電腦第六期

69 年 12 月

謝清俊

語云：「工欲善其事，必先利其器」國內目前銷售的電腦多半有相當好的中文能力，而且各個系統都有其特色，究竟那一種最好呢？答案見仁見智，對一個企業而言，應以適咳自己用途的中文電腦為最佳。

前言

經過了八、九年的努力，目前銷售的電腦多半有相當好的中文資訊處理能力，便宜的更在新台幣四十萬元以內。語云：「工欲善其事，必先利其器」，如此利器，自宜善加利用為是。

但是，研究單位和大學不算，目前能提供中文電腦（註一）的廠商都有十七家之多（日本廠商尚不在內），而且各個系統都有其特色。我們常聽到有人問：「究竟那一種最好呢？」，當找不出答案時，就有要「統一」中文電腦的想法。選擇電腦系統原本就不是容易的事，需要相當多的專業常識，再加上中文的問題就更為複雜了。

近年來，有關中文電腦的論文和報導很多，但是始終沒有較完整且深入的評估文章。事實上，評估的工作做得很少，此類報導也自闕如。如果要用戶自行蒐集技術資料，待全盤消化後再列出評估的條件重點做一選擇，是十分艱鉅的工作。

想要做一個全面通用的評估，選出「第一名」來，是不可能的，主要的原因是任何的應用，都有它主觀的成份—本身業務的性質存在，因此，也就難以擇一標準放諸四海皆準了。選擇一個中文電腦適合自己的業務是一個醫配的問題

題，所以不應想「如何選出一個最佳的中文電腦」，而是該去做：「如何選擇一個中文電腦最適合自己的用途」這件事。

※註一：「中文電腦」一詞所受指責甚多，本文內此詞是指「能夠處理中文資訊的電腦系統」而言，取其簡潔而用之。

要適合自己的用途，首先要了解自己的業務。如孫子所云：「知己知彼，百戰不殆」吧！事實上，這是選擇電腦的第一要務，並不僅限於中文電腦。目前電腦的應用雖甚普遍，但是，要由他本身產生「智慧」，如科幻電影中的那些電腦一般，能了解這花花世界的情形，自動地設法解決問題的這種使用方式，還差得太遠，也許永遠沒有這麼一天。目前的用法是：由人們澈底地了解業務的每一細微末節，然後規劃出一套可行的處理方法和步驟（至少不能有錯），將之轉換成電腦能了解的命令，也就是所謂的程式，再令電腦記下這些命令去奉命行事。因此，做一個好選擇，是必需了解本身業務的。以上的步驟，正是開發所謂「應用軟體」的步驟，電腦主要的價值亦盡在於此，順便也說明了為什麼軟體值錢的原因。如果有些業務，無法歷此過程，那麼，電腦也就無能為力了。

選擇中文電腦的過程，始終是圍繞著「自己業務上的需求」這個主題的，本文以下亦時時刻刻離不了這個主題。事實上，需要了解那些業務上的資料，才足以做好選擇中文電腦的事呢，也是本文想討論的問題。

由於此問題所涉甚廣，勢將陸續分段刊出以竟全篇，敬祈見諒，再者，一己之見難免有所疏漏，筆者由衷地希望能以此文拋磚引玉，無論是設計的工程師或學者，應用中文電腦的人，文字學家，通訊專家甚至於心理學，社會學專家...等等，凡有關者，大家由不同的角度來談談這個問題，相信此舉必有助於中文電腦的評估和選擇，進而促進國內資訊工業之推展。這就是本文的期望。

中文資訊處理系統的選擇之一—文字和符號的選擇

任何一個電腦系統，都只能處理有限的文字符號，西方由字母拚成的文字處理起來比較方便，例如早期的 Fortran 語言，能處理的範圍只限於 26 個大寫字母，10 個阿拉伯數字和 13 個控制符號，總共只有四十九個，除此以外，如小寫的字母，就無法辨認了。其餘的訊息，就必需由這四十九個符號以文法組織的方式來表示了。因此，選擇中文系統的第一件事，就是要了解這個系統能處理多少文字及相關的符號。

中文電腦相關的符號和文字可分類如下：

- 1．可處理的中國文字
- 2．中國文字發音所用的符號
- 3．中文標點符號
- 4．外語的字母及符號
- 5．專業所用的特殊符號

茲就各類分述如后：

一、可處理的中國文字

目前的中文電腦系統可處理的中文字數約在八千至一萬二千左右，多的可至一萬五千餘字。由於一般常用字在五千字以內，例如教育部頒的標準常用字形表共 4808 個字，世界中文報業協會的新聞常用字彙亦僅有三千字。以約一萬字的能力而言，應均可滿足一般的用途了，但是全部中文的字數，遠超過此數目，目前最大的蒐集是中華大辭典，共集四萬九千九百零五字，如考慮各式竹異體字及地方性的文字，如廣東字等，則尚不止於五萬字。由於各行專業應用的字彙差異甚大，列如人名、地址等用的字彙就和化工業者用的字彙差異

甚大。因此，在選擇中文電腦時，對於能處理中文的數目，以及它們究竟是那些字，必需小心查證。

很多中文電腦是根據一些字書來訂定他們的字彙的，例如，林樹先生著作的「中文電腦用字」便是其一，也有些是根據較通行的字典來訂定的。因此，如果您的專業上需要一些冷僻字，目前的中文電腦中不一定有。

為了應付這個問題，多半的中文電腦設計了有增添新字的方法，以應賭者之需，在這一點而言，各系統之間難易的程度就差得很大。簡單的可以不需修改任何原系統的機件（硬體）和程式（軟體），只需鍵盤上以一定的格式，將新字的資料告訴電腦即可，甚至新字的字形均自動產生不需用戶費心。而另一極端是，必須做硬體及軟體之修正，此時要停機做工程上的變更，是很麻煩的。

目前中文系統中處理中文字形的方法可分二大類，字數的限制也以此分為二種情況，其一是採取直接貯存的方式，也就是把每一個中文字形均存起來，採用此法者，通常遇新增字彙時，處理起來較困難，然而巧拙不同，在方便程度上仍有很大差異。其二是採用結構橫式處理中文字形的，如以字根，字基產生字形的系統等是，這類系統處理新字較易，然而其設計是否完善，仍然不盡相同，本必都好。

一般的廠商都能提供一個完整的字表，告訴客戶他們的系統能處理那些字，但有些廠商提不出這資料，此時宜特別加以查證。

在增添新字時，必需要多佔用原系統內的一些記憶空間，因此，各系統對添增新字時多有一個數目上的限制，在超過此數目時，可能要將原系統做一些工程上的修改才能應付所需，如果您的應用涉及許多新字。此點宜加查明。

以上所談的是涉及中文字數的一些問題，有關中文處理其他的一些問題，例如如何選擇適用的輸入方式，字形和字體的問題等等，將陸續在後文中討論。

二、中國文字發音所用的符號

中國文字是由其形、音、義三者定義而得的，因此，在應用時也許會涉及發音的問題，此時可能用到注音、韋氏拼音（Wade-Jile's System）或是耶魯拼音（Yale's System），目前廠商中約有一半能提供注音符號的處理能力，至於各種拼音系統中的特別符號，如U、A等，就需特別加以查明了。

三、中文標點符號

目前廠商能提供的中文標點符號很不統一，這是一個很值得注意的問題，在文書、排版方面的應用，標點符號是極重要的。目前關於標點符號的資料流傳不多，其中國語日報發行的標點符號的用法，是很好的參考資料。

中文有橫、直兩種通用的印法，由於位置的關係此二者所需之標點符號不盡相同，中文電腦能否涵蓋此二者，是值得注意的事。

統一中文電腦中的標點符號，應非難事，值得有關當局考慮執行的。

四、外語字母及符號

在許多應用裡中文英文很難分開，如果要電腦分成二個系統，分別處理中英文資料，實不是好的做法。目前一般的做法是將英文電腦中用的符號，實不是好的做法。目前一般的做法是將英文電腦中用的符號，包括字母、數字及控制符號等，納入中文系統之中，中文電腦能處理的符號就包涵了原電腦能處理了外語字母及符號了。這些外語字母及符號通常是A S C I I 集合（American Standard Code for Information Interchange）的全或一部分。目前各個中文電腦並沒有包涵A S C I I 中的全部符號，通常只是一部份，且完全相同者甚少。在此情形下，是否能滿足您的應用是在選擇時應注意的事。

這兒涉及一個問題：在中文系統中這些 A S C I I 符號的電碼是否和原 A S C I I 的電碼相同呢？事實上，沒有一家能做到相同的地步，在中文系統中這些符號都另有電碼。因此，在中文系統和英文系統之間資料和數據如何轉換，以及彼此間在控制符號的應用上是否會引起系統內的衝突？是值得關心和探究的問題。我們仍將這些問題稍延後再討論，在此，僅先談談該有那些符號的問題。

五、專業所用的特殊符號

各行業中多半有些特殊的專業符號，在英文電腦系統中，處理起來很困難，因為 A S C I I 電碼的空間有限，不可能兼容並蓄，除非是一個有能力處理圖形的特殊系統，勿有餘裕處理這些符號；即使如此，電腦中亦不把這些專業符號認為是和字母一般的基本符號，多半是以字母組合的片語來表示之，在處理圖形的軟體系統之下，方能發揮功能。但是中文系統由於本質上就有處理圖形的能力（中國文字多當作圖刑之單體處置）是故對付這些專業符號，較英文系統直接而簡勿。

一般而言，特殊的數字符號，數學運算符號等，各個中文系統中多半有一些，甚至有些外文，如日文的假名、韓國的拚音字母等等，也能提供。專業符號中文系統中是很難包涵完備的，但是在採購之前，要求廠商提供所需的特殊符號，應是可行的，可以減少了許多日後添加時作業上的困擾。

目前，還沒有一個通行的中文資訊交換碼。中文電腦中這些文字及符號的電碼在各個系統中均不盡相同，是故系統間在交換資訊時，將發生一些困難。此上所談到的新字和新符號的加碼工作，廠商下否有一個有系統的且統一的做法，是值得注意的，如果廠商對各個用戶增加的碼不能有系統的協調統一，將產生以後轉變為通用交換碼的困難，需對各個用戶分別處理，或要求用戶自己去做，這也是可能要注意的事。