

科學月刊  
第十八卷第三期  
中文資訊的處理（二）

## 中文資訊處理的科際關係

### 一、與語文學的關係

做中文資訊處理的研究固然必須具備計算機科學的知，然而，在中國語文方面的知識也是不能缺少的。數千年來，語文是世界上任何一個民族文化延綿的主要工具，也是人與人溝通的主要橋梁。自從計算機發明了以，語文開始兼作人與計算機之間溝通的媒體。這現象明顯地指出語文和計算機有密不可分的關係。然而，由於時下計算機的能力有限，無法像人一樣處理和應用我們的語言，所以只得設計一套更簡明的語言，讓計算機使用。這種語言就稱為人工語言或計算機語言，以便和我們日常使用的「自然語言」區別。

目前的計算機語言使用了自然語言所用的符號：譬如，英文字母（中文字）、標點符號、阿拉伯數字等等，和自然語言沒有什麼兩樣，然而在語法結構上卻比自然語言簡明許多，這是為了要配合時下計算機的能力來處理計算機語言的語法的緣故。當我們學習一種計算機語言時，常常覺得它刻板之至，所有的規定不容絲毫偏差，這不僅不通情理，而且顯得固執和愚蠢，這是計算機語言的語法太欠缺彈性所造成的。另一個原因是：在語意的處理上，計算機語文能做的更少；除了必要設定的語意在事先就硬性規定以某種形式表達以外，其他都忽略掉了，或是留給應用程式自己去看著辦！

雖然目前的計算機語言和自然語言之間有這麼大的差距，然而科學家們正努力在減少它們的距離，希望有一天，計算機能具有相當的語文程度，用自然語文和我們溝通。賦與計算機處理自然語文能力的努力，統稱為自然語文處理的研究，這是近年來十分熱門的研究方向之一。對我國而言，由於研究的對象是我國的語文，自然有其獨特的問題。這方面的研究可名為「中國語文處理」，應該是中文資訊處理研究中極重要的一環。

語文與計算機的結合是件大事，在國外已行之有年。像計算機語言學（computational linguistics）、計算機詞彙學（computer morphology）、文獻處理（text processing）以及自然語文處理等等新學科，皆在此結合下蓬勃發展。可是，國內的情形就顯得完全不一樣。在大學中，即使偶而開了上列的課程，也

用英語作為對象，這和我們的社會、文化以及現實生活脫了節。為了要使計算機有處理中文語文的能力，或是要利用計算機來推動語文知識的應用，沿著上述的發展方向，針對中文語文的特色來發展中文的計算機詞彙學、中文的全文處理技術以及中文的自然語文處理等學科是必要的。

## 二、與人工智慧的關係

計算機使用得越普遍，對語文處理技術的要求就越殷切，而且對處理能力和品質的需求就越高。要言之，語文處理的需求來自二方面：其一是藉以改善人與機器溝通的方式，其二是藉以提升對文獻和事務處理的能力。在日本第五代電腦計畫想發展的智慧型計算機中，自然語文處理和智慧型人機介面都是必要的功能。而智慧型人機介面所做的事，除了處理語音和字形丟表達的媒體以外，還是要依賴語文的處理來做好認別及產生的工作。

早期計算機的使用大多是處理一些表格化的語文資料。例如，處理各種單據、查詢一些欄位化的資訊等等。目前，這樣的使用方式已經無法滿足接踵而來的要求；人們要求處理完整的原始文獻，如信件、公文、法律條文與判例、新聞、會議紀錄、專利文獻、史籍檔案與文獻.....等等。處理的媒體亦多元化，包括聲音、影像及圖表等等。對處理的基本功能也起了變化，不再囿於計算、邏輯判斷等，而是擴大到一些較智慧型的能力，譬如：推理、計劃、識別、學習、理解等等。這也就是在上期中，本文所說的由資訊處理發展至知識處理的過程。其實，語言、知識、智慧三者是相互交織密不可分的，當計算機有些處理自然語文的能力時，它自然將擁有一些知識以及具備某種程度的智慧能力，而這正是人們心目中追求的理想。再說，人類累積的知識和資料，絕大部分是用語文表達的，計算機和語文的結合，自自然然地增加了處理知識的能力，包括知識的表達、取得、組織以及應用。對計算機來說，用機器來做這些事，就是人工智慧研究的主題。

再者，在研究中文的語意時，已經涉及意念和知識的表達。為了要認定什麼是知識？它的範圍和界定是什麼？知識如何產生？如何取得？如何分類？認知的過程如何？如何運用知識？.....等等問題時，將涉及知識論（哲學）、認知科學（心理學）、分類學（圖書館學），乃至於腦神經醫學等相關的知識。而目前在國外發展的趨勢，也是朝這個科際大給合的方向邁進。

## 三、與其他科學的關係

前文已經提到，改進人機溝通的方式是做中文資訊處理研究的主要原動力之一。在考慮人機溝通因素時，不可避免地要用到一些專門學科的知識。例如：心理學、人機工學（erg-onomics）、傳播學等。為了使機器表達資訊的方式

能為工作人員接受和喜愛，並且要維持一個良好的工作環境，不要讓工作人員受到傷害，這些專門學科的介入是必須的。有許多關於輸入輸出的工業標準，是經由人機工學的設計以及心理學對接受程度和使用效率的測試而訂定出來的。譬如：在不同媒體中字體點陣的大小標準；字的間距、行距和畫面的標準；鍵盤上符號的安排標準以及輸入方法的評估等等均是。

由以上的討論，我們已經了解：中文資訊處理的研究涉及許多專門學科，是相當複雜的一門科際整合形態的新研究領域。在這樣的情形下，我們還需要一些知識來駕御這些錯綜複雜的科際關係，才能做好中文資訊處理在科學上的研究和工程方面的發展。這些知識包括模控學（cybernetics）、系統科學以及管理科學，此外在工程發展方面則須重視系統工程（system engineering）的學養。模控學和系統科學都是觀察和歸納複雜的系統現象，並以之推導出對系統的了解（知識）的學問。雖然它們之間有對自然系統和人工系統的分野和差異，然而對中文資訊處理的研究而言，二者的觀念和素養均為必需：因為中文資訊處理的研究不只涉及機器系統（人工的），更涉及到「人」以及在「人的社會組織形態下」，如何使這機器系統能夠良好的運作以充分發揮其效能。是故它涉及與人有關的系統（自然的）和管理科學的知識了。此外，由工程的立場來看，要開發這麼複雜的產品，系統工程的知識當然重要。

綜合以上所述，中文資訊處理的研究不僅是計算機科學和語文學的深入結合，還經常涉及下列三類的專門學科，是典型的科際整合的形態。

- 一、心理學、人機工學、傳播學（科技、語文傳播）；
- 二、智識論（哲學）、認知科學、圖書館學（分類學、資訊系統）、腦神經醫學；
- 三、模控學、系統科學、管理科學、系統工程。

### 基本研究問題的分類

中國語文有其特色。做中文資訊處理時依語文之特性不同，而演繹出各種的處理方法，所以中文資訊處理可以據此分類。依文獻結構的元素來分，處理的對象可分為：字、詞彙、片語、句子、段落、文章等。由文法結構的層次來說，則可分為語法和語意兩大部分。目前的商用系統，還停留在只能處理「字」的階段。詞彙或更複雜的形式，則無理論的模式可用，只能依應用問題的性質寫些依附資料（date dependent）的程式作特殊的解決之道。這樣的作法使得一個程式只能解決一個問題，而無法與同樣性質的問題共享，其投入之成本自然高漲，且對複雜的問題則無法做有系統的解法。至於以語法與語意二者，還只是

研究中的對象。在研究室裡，上述各層面的問題雖均已有小小的涉獵，可惜計畫之規模較小，且久缺長期的恆定性，以致於目前的成就不大，離實用仍有距離。

另一個角度的分法是由語文表達所用的媒體來區分。可分為三類：形、音以及碼。「形」是指人類以視覺功能處理的文字外觀，它包括各種印刷、顯示、或書寫的形態。以「音」表達者就是以人類聽覺可以處理的語文形態，稱為語音。「碼」是指以計算機可以閱讀的方式所作的表達。它包括數位化的字形點矩陣、數位化的語音訊號、字的認別碼、交換碼、檢索碼等等。

第三個角度的做法是以資訊處理的基本功能來分。計算機本質上是自動機（*automata*），所以由自動機的基本功能來說，研究的問題可分為產生和識別二大類。若由應用的角度來細分其功能，則定義二中之各項運作都可視為資訊處理的基本功能。

上述的三個分類角度是各自獨立的，可以組成一個三度空間。在此空間中一個點則可代表一種中文資訊處理的基本研究問題。例如做字形產生的研究是位於「字」、「形」、以及「產生」的交會點；又如語音識別為「字、詞、句子」、「音」以及「識別」等交會處之總稱。此結構表現如附圖。圖中的三個坐標分別是：語文結構、語文表達的媒體、和資訊處理之基本功能。以此分類，可以對中文資訊處理面臨的種種基本問題作一綱領式的了解。

以目前研究的情形而言，許多基本問題尚待努力，而研究範圍則局限於「句子」以下的簡單語文結構。若是我們把功能概略納入產生和識別兩類，把語文結構約略分為目前常見的處理對象——字、詞、句，則附圖可化簡為分別以形、音以及碼之三個平面，而每個平面上則有「產生、認別」與「字、詞、句」等組成之六類問題。這個約略的分類可以含蓋了目前所有中文資訊處理研究的基本問題。（待續）

科學月刊第十一卷第十期  
中華民國 75 年 10 月  
計算速率的追求平行處理

「人多好做事」——這就是「平行處理」的直覺觀念。如果一部計算機要花  $X$  秒才能算完的問題，用  $N$  部計算機「大家一起來」做，是不是能把時間縮短到  $X/N$  秒算完呢？平行處理的基本信念就是這麼平庸、單純。這種想法非但不新穎，簡直古老，而且做事的人，人人都會這麼想。

然而，近五年來，有關「平行處理」的研究卻大紅大紫；美、歐、日等工業國家在官方、學術界、企業界的通力合作之下，投入了數百計的研究計畫來開發「平行處理」的技術。為什麼有這種現象？為什麼要這麼做呢？

就近因來說，這一場競爭——結合了學術與工業的研究和開發的競爭，是日本第五代計算機計畫所誘發出來的。根據第五代計算機計畫所誘發出來的。根據第五代計算機的計畫，所謂第五代計算機是智慧型的計算機，它不僅將擁有一些解決問題的智慧來幫我們做事，而將以人與人溝通的方式——用自然語言和文字——來和我們交往，在這個命題下，計算機要做的事將遠比目前所能做的多得太。根據研究的數據估計，要達到上述的目標，第五代計算機的運算速率至少要比目前同型的機種快數千倍、數萬倍，乃至數十萬倍。換句話說，如果第五代計算機無法比目前的快萬倍以上，它根本就成功不了。因此，計算速率的追求成為第五代計算機的命脈和成天的必要條件之一。也正因如此，激起了近五年來追求計算速率突破的狂熱。

電子計算機在本質上是一具電子儀器。所以，由儀器系統的觀點它的計算速率可由下列三個層次來增加：一在元件層次上，儘量用快的元件；二在包裝（packing）層次上，儘量減短訊號流經的途徑；三在系統結構層次上，尋求更好的架構和演算方法（algorithm）以減少運算的時間。

綜觀這三個層次，在過去三十多年以來，計算機各種性能的改進，包括：容量、能力、效率、對能源的消耗、可靠性以及計算速率等等都得歸功於前二者：元件和包裝技術，而此二者之突飛猛進實與微電子技術的發展息息相關。然而，目前有些元件製作技術已趨飽和，有些元件的速率已趨近理論值之極限，今後要求速率上大幅突破，其情勢早已今非昔比，不可能再循舊途改進。由另一角度來看，計算機的結構在過去三十幾年間一直沒有什麼重大的改變，

也該是變一變的時候了。可是，怎麼變呢？如何突破傳統的結構？這個問題引起了所有研究人員的關心，並導致對計算的本質作更深入了解的探求。

## 馮紐曼模式

傳統的計算機結構是所謂的馮紐曼模式（von Neumann model），如圖一所示。圖一（a）的機器裡，只有一個處理機，一個記憶單元和一個收發通道（input/output channel）。因為只有一個處理機控制全部機器運作，故又稱為單處理機系統（uniprocessor system）。它做事的方法是：一個命令對應一組數據資料；命令加諸於資料則產生一個基本的運算動作；命令成串排列，一脈單傳地一個一個依次執行。這種操作的方式稱為單序操作（monosequencing operation）。這種結構和人做事的方法類似；人做事，也是以單序操作的形態為主。傳統的解題方法也依附這個結構，人們把它寫成程式，上計算機解問題。所以，馮紐曼模式是相當「自然」的模式，人們不知不覺地已用了三十幾年。它和人們配合得相當好，以致於寫程式時，幾乎忘了機體結構的存在，用不著擔心所寫的程式是否能配合計算機的構作有效的運算——反正它們是配得差不到那裡去。

要突破這樣一個窠臼，實在不是件容易的事。溯本清源來探究計算的本質是必要的。要言之，馮紐曼模式是單處理機單序操作的典型，在此情況下，無論所解的問題為何，計算性質的也好，非計算性質的也好，都已無再變化之餘地。因此，為突破速率的極限，就自然而然地走上「人多好做事」的多處理器結構，也就是所謂「平行處理」的研究領域中了；人們希望月成成千上萬的處理器堆砌出速度來！

## 基本概念與運用

根據上述的發展情形，「平行處理」一詞目前是用來泛指不只一個處理器（或處理元素，processing element）的計算機系統結構（參考資料 12），其研究重點在尋找於多處理器情況下，什麼是理論上共同特徵？如何根據這些特徵來建立其理論模式？以及如何依據此理論模式來發展適當的演算方法（亦即平行演算法）？這些問題都觸及計算科學的核心問題，譬如：什麼是計算？以及如何計算？什麼是可算的和不可算的？如何評定一個演算方法的好壞等問題。而這些問題，到目前為止都還沒有滿意的答案。也正因為沒有答案，所以吸引大批的研究者一探寶山的決心。這是平行處理變成熱門研究問題的主因和遠因。

前文已說過，平行處理的概念實在不是什麼新東西。設計硬體的工程師，老早就用過它了，這幾乎是每一位設計電路工程師都有過的經驗。在上一期的

「計算機和語文」專欄裡，我們談「記憶與速率」的問題時，交織式的記憶庫結構和隱藏記憶（cachememory）的層次結構，就是兩種基本平行結構的典型。前者是將硬體加多（重複相同的硬體），並同時工作（存與取）以達到速率加快的目的。而後者是一種管道型式（pipe-lined）的安排，其目的在加大某工作流程中單位時間的流量。比方說，有一條生產線的工作，如圖二所示。此工作以一個人來做時要十九分鐘（a），如果分為甲乙丙三人來做，雖然增加了些工作分派及管理上的工作，使得總共的時間為廿一分鐘，但其流量（throughput，即在重複性工作之供應不斷的情形下，單位時間的生產量）反而增加，亦即每九分鐘可做完一件事（b）。若是將（b）中乙的工作再分為二，如（c）中所示，則更可將流量提高至每六分鐘可完成一件事。在圖三中，若甲乙丙均為電路，則在甲乙丙各級間必須加暫存器，以安置算好的或等待下一級運算的資料，這種暫存器名為管道式暫存器（pipe-lined registers）是設計高速運算電路常用的法寶之一。而 cache 及記憶體的層次結構的構想，也和此管道式的想法相關聯。又如，在計算機中指令執行之同時，預先將下一個指令先行解碼的做法（instruction look-ahead），亦一典型之管道式平行結構。

雖然平行處理不是新觀念，而且，已往在設計計算機硬體時也常用，可是，目前我們對平行處理的了解還是極膚淺的，而且多半將它應用在一些特殊結構上。除了上述的例子以外，數位系統（計算機）用於軍事系統，生產控制以及各種儀表的設計裡，均可找出無數個特殊平行結構的例子。換言之，如果針對某一特殊問題，已經有一個解決問題的流程（它通常包含一些同時處理的步驟），而據此流程所做的硬體設計都是些特殊的平行結構（special purpose parallel structure）。而這種做法早已行之有年。

### 高度平行的結構

所以，目前平行處理研究的主力，並不放在這些特殊平行結構上，而是著重在對一般性質（generalpurpose）平行結構的了解，和對平行運算本質的了解。這些通用的平行結構又稱為高度的平行結構，在圖三裡有二個代表性的架構。由圖三中可知，考慮平行處理的結構時，主要的系統元素是處理器和記憶體。而設計平行結構的目的是：集合一群循序處理機並使之能在相互合作和溝通的運作下，很快速地解決複雜的大問題（參考資料 3）。

根據上面的說法，研究平行處理系統涉及三項特徵：

- 一、共通性（generality）：也就是對系統整體能力，如：處理流量之大小，或反應時間之快慢等，應具有可測的性質。

二、功能度（ scalability ）：是指增加處理機數目時，對整體系統功能增減的單位幅度（以每處理機為準）應為可測。

三、單機形象（ single system image ）：無論此平行處理系統用了多少處理機，對使用者而言，應如同他在使用一個極高速的計算機一般，而無視於這此處理機的存在。

而這三項特徵也正是目前作平行處理研究的重要前提假設。

## 定義和基本問題

人多雖好做事，可是也有「三個和尚沒水吃」的可能。平行處理的概念雖是老生常談，然而在執行時卻有太多的變因而使得問題變成複雜難解。為了解這些複雜的變因，讓我們用前述平行結構的定義，來做個條列式的說明。

一、「集合一群循序處理機.....」，相關的變因有：

- 多少個？有人做以十計的，有以千計的，甚至有用超大型積體電路以百萬個處理器計的。
- 多大一個？大多以 16 位元或 32 位元的微處理機為參考。
- 什麼樣的處理機？它的指令集是些什麼？
- 記憶單元有多大？將依處理機功能及資料之區域性（ data locality ）而定。
- 這些處理機的收發功能如何？將依處理機之功能以及處理機之間通訊的負荷而定。

二、「....並使之能在相互合作和溝通的運作下....」，相關的變因有：

- 如何將之連接？有用縱棋式交換機（ cross-bar ）、樹（ tree ）、矩陣（ arrays ）、超空間方塊（ hyper-cube ）或者洗牌（ shuffles ）網路等連接者。
- 如何溝通？那些訊息需要溝通？
- 如何同步（解決工作順序與工作時差之間的問題）？
- 工作如何劃分（使之合作）？單位工作的大小（ grain size ）為何？
- 操作系統如何設計或溝通？
- 用單指令束或複指令束結構？（關於此項，下面將作詳細解釋。）

三、「....很快速地解決複雜的大問題」，相關的變因有：

- 那些是「複雜的大問題」？那此應用有這類問題？
- 計算模式為何？
- 如何訂定平行運算的規格？



- 上述的事情由誰來做？人或是系統？
- 有任何工具可用嗎？譬如描述平行處理演算方法的語言？或是有平行處理特性的其他工具？
  - 上述之語言有編譯程式嗎？
- 平行演算的法則和演算方法是什麼？
- 有多快？
  - 這樣的系統，其可靠性、可用性及維護性如何？以上這些問題，沒有一個好回答。第一類中的問題，主要是涉及硬體結構。在第三類中的問題，則主要涉及平行演算方法、軟體和理論模式。第二類中的問題涉及通訊，並與軟硬體均有牽連。目前，有關硬體架構之問題並不大，而平行處理的理論模式和演算方法是整個研究亟待解決的癥結所在。

### 系統內的連接方式與計算模式

弗林教授（ M.J. Flynn ）是最早嘗試為各種處理的連接方法做分類的人。在 1972 年，他依據資料流束（ data stream ）和指令流束（ instructionstream ）的數目來分類，此分類法至今仍常採用，詳如表一（見下頁）。

在表一中，最簡單的結構是 SISD（ single instruction stream and single datastream 以下簡稱類推），亦即單指令束和單資料束的狀況。這就是傳統的馮紐曼模式。管道模式是 MISD 的典型，它只有一資料束，可是每一個處理機對此資料束做的指令均不同，所以是複指令束。管道模式曾用於 Cray-1 及 Crber 205 等超級計算機中來做向量的運算。SIMD 的單指令束複資料束的結構則出現於早期號稱第一台超級計算機的 Illiac -IV 中，它一共用了 64 個處理機來分享一個指令，也就是同時對 64 組數據只能做同一種運算。這種結構適合作矩陣或向量的計算。MIMD 是較通用型的結構，如圖三的例子。

弗林的分類法提供了很好的參考點，它使我們有系統的了解各種不同的內部結構。然而，它久缺具體的計算模式，對於系統特性的分析方面功用並不太，因此，近年來又有許多模式的出現，茲將重要者介紹如下：

一、洗牌模式（ shuffer-exchange model ）：此模式的處理機數與記憶單元數目相等，而彼此間之連接是以排列的匹配函數（ permutation function ）來連接的。這個模式的優點是簡潔，容易在此模式下導出平行運算的演算方法，有許多很好的排序（ sort ）及搜尋（ search ）的平行演算法均在此模式下導出。它的毛病是以積體電路的技術來做時相當複雜，不易付諸實用。然而在理論研究上，有其肯定的價值。

二、超空間方塊模式 (hyper - cube model)：這是加州理工學院在解某物問題時所發明的模式。之後，此專利為 Intel 公司收購，並於去年推出上市，是唯一的平行處理元件商品。它的機種叫 iPSC，分為 32、64 以及 128 個處理機的三種型號。它的處理機是 80286/80287 每個附帶有 512KBytes 的記憶單元。其主控部分是 Intel 310 微計算機，用 Zenix 操作系統，並提供 C 及 Lisp 語言編譯器。這種系統仍較適合 SIMD 方式的運作（雖然不一定要如此），它可將 N 次的運算減低到  $\lceil \log_2 N \rceil * \text{次}$ （N 為處理器的數目）。註： $\lceil \log_2 N \rceil$  表示大於  $\log_2 N$  之最小正整數。

三、樹狀機模式 (tree machine)：在資料結構中，樹狀結構是取常用到的一種。因此，為配合資料結構乃有樹狀機的產生。這種機種自然是適合樹狀資料皂運算，在做數學運算 (SIMS 方式)，及在樹狀結構的資料中搜查時，它均可將 N 次的運算降為  $\lceil \log_2 N \rceil$  次。

四、數據流模式 (data-flow model)：這是歐洲、日本和美國麻省理工學院熱衷研究的模式，它的特徵是具有結構的可塑性 (re-configurable)。換言之，其內部的接線是可依程式的不同而改變的，這一點和上述的三種模式大不相同。它和傳統的模式也大異其趣：它沒有傳統的控制單元和程式記數器 (program counter) 作程式執行控制，取而代之的是對數據皂可用狀態作偵測，一旦數據準備妥當時，便能立刻做相關的運算。在理論上，此模式有更健全的計算模式，因此在計算理論的角度上看來，它的發展前途均較前三者為佳。然而，其理論模式甚為複雜，描述不易，因而使其使用之語言呈現出圖形的結構（詳情參閱圖四），不易使用。

## 平行的演算方法

演算方法是真正平行處理研究的成敗關鍵。如果沒有有效的平行演算方法，有再多的處理機也無用武之地。反之，有了好的平行演算方法，用硬體做也好，用軟體模擬也好，總可以充分利用即有的計算設備把問題解出來。目前，研究平行演算法的人很多，題目也很多，因為以前發展的所有單序演算法，全部可以考慮重新來過改為平行演算法。

然而，根據上一節的討論，演算法亦將受限於其硬體連接的模式。譬如說，在樹狀機模式下的演算法一定與洗牌模式下的演算法迥異。因此，它幾乎又多了一維 (dimension) 的變化。

綜合和歸納各種演算法，將有希望導出新的計算模式。雖然目前尚無此方面的具體結果，然而，這將是未來研究的一線希望。

根據目前研究的成績，我們已經發現有些問題的演算法適合用平行的方式來解；有些則不適合，只宜用單序演算法來做。然而如何判斷何類問題適合用何種解法仍是一未解的難題。

此外，能描述平行處理的語言將是推展平行處理應用的成敗關鍵，可惜目前此方面的研究仍待努力。

有關平行演算法部，就討論至此，有興趣一探究竟的讀者可參閱本期介紹平行演算法的「平行處理的計算方法」一文。

### 平行處理之前途看好

平行處理的研究，固然是為了對計算速率的追求，然而，其盛行的緣起亦有環境的客觀因素存在：那就是由於超大型積體電路技術的發展，充分提供了「硬體不值錢」的環境，使得一次的設計投資可以長期的享用，而硬體的直接生產成本幾乎變得和電路的複雜度無關。這種環境不只使得平行處理的研究如火如荼地展開，也改變了數位系統設計的整個理念和價值觀；設計人員不再追求省一些電路以降低成本，而反過來研究如何多用一些電路來加強系統的可靠性、可測性及對錯誤的容忍度等等。這些都是硬體昂貴時不願去碰的問題。因此，大型積體電路不僅是發展平行處理的必要元件，更是發展更複雜的大型系統必要的技術。

平行處理的研究雖然目前極為熱門，然而，其研究的成果究竟將如何仍是個未知數。主要的原因是因為人不習慣於平行處理的方式，能有老頑童周伯通「分心神功」的人，少之又少。所以發展這種不甚「自然」的平行（處理）演算法，似乎仍有一段長長的路要走。同樣的理由，要訓練使用計算機的人用平行處理的語言，並且時時要記得將演算法配合特有的平行硬體結構來寫程式，是件極苦的差事，也極不容易做好。是不是能發展出有系統的科學方法，將單序演算法自動轉換為平行演算法呢？目前此方面的研究成果仍然有限得很，只能做些最簡單的轉換而已（如數據流模式即已有些此功能）。

由上觀之，我們對平行運算的本質了解得還不夠。如果在計算理論上對平

行處理的了解沒有突破性的進展，那麼對平行處理的知識和應用將滯留在某一程度上也是顯而易見的。

平行的演算法在本質上是屬於各個擊破法（**divide and conquer**）解題策略中的一種。由於問題被擊破的程度不同，在其平行結構上亦將產生極大的差異。如何在此間求得一最佳大小的擊破（問題分解），目前仍是以系統的計算成本作為選擇的主要依據。這方面的研究直接影響到平行處理的結構，也是值得注意的發展領域。

總而言之，平行處理的研究是突破現有計算速率的希望所在，雖然目前看來仍然有許多問題，可是前途依然看好！

科學月刊第十八卷第二期  
中華民國 76 年 2 月  
中文資訊的處理（一）

中文資訊處理的研究是一個獨特的研究領域，它是一個經複雜的科際整合而誕生的新學科。十五年來，此領域從無到有，由理論上的探討、工程上的開發，而綿延到社會上廣泛的應用。如今，在國內，幾乎沒有一種計算機無法處理中文資料；在國際上，我們的研究工作一直穩定的保持在研究的第一線上，受到相當的重視和肯定。這個研究領域和它的發展過程，應該是國人珍重和自豪的，因為它完完全全是土生土長的。

由本專欄來介紹中文資訊處理的研究是很恰當的，它用到計算機，處理的是我們中國語文所表達的資料。

## 誕生

民國六十年的冬天，在國科會工程組兼職的馬志欽教授動了一項非常有意義的研究工作：「如何利用計算機做中文資料處理？」在此之前，計算機是毫無辦法處理中文資料的。這個倡議在當時獲得了空前熱烈的響應；幾乎所有的相關系所都先後加入了這個行列。自此以後，每年的碩士論文至少有 10% 以上在做中文資訊處理相關的研究，迄今依然如此。因而替這個領域紮下了萌芽和成長的基礎。

最早期的研究集中在：「如何使計算機能夠接受中文資料，和顯示出中文字形」的問題方面。這也就是所謂的「中文輸入和輸出」的問題。在這個階段中，發展了許多中文輸入的方法，這些成果經多次的測試、改良，而逐漸商品化。雖然，目前各種的中文輸入方法並不是很好用，仍然值得我們去改良，可是對於專業的打字員來說，已經可以用它有效地承擔資料輸入的工作了。

如何產生中文字形和如何顯示或列印中文字形是另一個焦點問題。各種中文字形產生器，無論是軟體或是硬體，陸續地開發成功。目前這些產品不僅物美而且價廉，是日本、中共、香港、美國等地區望塵莫及的。

早期的研究十分辛苦。在當時能供中文使用的周邊設備甚少。因此，各種中文鍵盤、螢幕顯示裝備、印字設備等等都經過無數次的試作。只可惜這些成

品並沒有妥為收藏保存，否則倒是博物館陳列的好材料。

從民國六十五年，陸續有中文資訊處理的商品問世。之後的五年間，各種商品如雨後春筍，蔚蔚然已成為「面」的發展，各種商業上的志用，亦次第展開。譬如：各種應用軟體的開發，包括：清單列印、文書處理以及處理人事、會計、出納、採購、庫料、財產和一些特殊行業的應用軟體等。又如軟體工具之開發；包括各種計算機語言的編譯器、檔案系統、資料庫、編輯程式、列表程式等等。這些發展的系統雖非至善，可是它們所發揮的功效，的確已經提升了我們社會處理事務的能力。

民國六十一年，林樹所著的「中文電腦用字的研究」提供了常用字集的分類、字形的整理以及字的使用頻率等資料。此書一直是設計中文資訊處理系統非常重要的參考資料。其實，在中文資訊處理的研究剛開始的時候，就已經牽涉到很深入的文字學問題。譬如：中文字一共有多少？如何檢索中文字？如何處理異體字和破音字？類似這些的問題有一大籬筐！而迄今仍待解決。這類問題，若無文字和語文學者的協助，輕則所做的研究或發展工作品質不好，重則對我們的文字產生破壞、扭曲、誤導等現象，其後果實難逆料。時下的系統中，仍有許多弊病是因此而起的。譬如：有錯字、字蒐集得不全，將字胡亂拆解、亂造新字、以及有關字的屬性資料久缺、無法排序、無法檢索和許多不合理的功能限制等等。這些都是欠缺完好的科際合作所造成的後果。

目前的研究工作已明顯地擴大了科際合作的層面。較積極參與的是文字學、詞彙學、語言學、心理學、人體工學( human factor's study, 或 ergonomics ) 等等。而處理的對象亦由文字擴充到詞彙、文句、文章、書籍的處理；處理的媒體則延伸到語音、字的影像等；而研究的範圍則揉合了各個學科中所關心的各種問題。這些問題雖然表面上看來是千頭萬緒，而且分布很廣，然而它們之間仍有其共同性與相關性；這是建立在語言和文化獨特的性質上面，是建立在處理資訊的共同方法上面，也是建立在和人發生關係的層面上。也就是由於這些共同性與相關性的緣故，構成了這個科際整合型態的「中文資訊處理」研究領域。

本文嘗試著將此領域作一整體性的介紹。我們將由中文資訊處理的定義談起，釐清研究之範疇，討論其特性，分論主要的研究方向與題目。由於篇幅較長，全文將在此專欄分段陸續刊登。

## 中文資訊處理的定義

用計算機處理中文資訊時，會遇到一些與計算機、中文資訊或與使用人有

關的問題。為解決這類問題所做的研究工作，統稱為中文資訊處理的研究。具有某些中文資訊處理能力的計算機系統，我們稱之為中文資訊處理系統。在通俗的報導中，常稱這類系統為「中文電腦」。我們對「中文資訊處理」一詞的界定是廣義的。如果要說得更明確些，可以參考下列的義。

#### 定義一：中文資訊

凡是以中國語文表達其原始形態的訊息皆稱為中文資訊。中文資訊依其表達的媒體和呈現的物理現象不同，分為自然形態與人工形態兩類。自然形態是指依語文的表徵，以語音與字形所表現者。自然形態的中文資訊經過物理量或數值符號的轉換後，以各種機器可以處理的形式所表現者，稱為人工形態的中文資訊。

由定義一可知，在計算機中所表達的中文資訊都是人工形態的。其實，語文表達的資訊可粗分為二個層面；其一是它呈現於外的物理現象，如：語音訊號、字形等是。這些資訊不難轉換為計算機可以讀取的形式(以下簡稱為機讀形式或機讀式)。其二是它孕育於內的抽象含義。例如：音調代表的情緒，詞、句、文章之含義等是，這一部分不容易在計算機中表達"(圖片省略)"，因為它們涉及知識的表達和智慧的处理。到目前為止，這仍是人工智慧研究的熱門的課題。我們所說的中文資訊包含上述兩個層面的訊息。

#### 定義二：處理

處理是泛指對資料的任何運作。例如：表達、轉換、計算、儲存、收發、傳送、檢查、排序、合併、分類、搜尋、查詢、產生、識別、判斷、分析、推理、了解、譯釋、學習、評估等等。

要做好定義二中「處理」所指的各項功能，必須要對處理的對象——資訊，作較深入的了解。一般而言，資訊依表達的方式可粗分為下列五類：

- (一) 以音頻訊號表達者：包括語言，樂音等。
- (二) 以數學語言表達者：包括數據、公式等等。
- (三) 以文字表達者：又分為
  1. 以語文敘述方式表達者；如文章、詩詞等。
  2. 以結構化欄位表達者：如表格及欄位化的檔案等。
- (四) 以圖案表達者：如統計圖、地圖等是。
- (五) 以影像表達者：又分為
  1. 靜態影像。如電視、傳真、照片、繪畫等是。
  2. 動態影像。如錄影片、一般影片、卡通影片等是。

當然，以上的各種表達方式，對於以資訊發生的實際過程來比較，都或多或少有些失真的。因此，在我們日常生活中所常見到的表達方式，多半是上述各種方式來綜合運作。例如，圖案或影像多配有文字說明或語音來表示，又如，一本書中所包涵的訊息，很可能有圖表、文字、照片、公式等等混合使用著。

## 語文資訊的處理

根據以上的分類，前三項都與語文有密切的關係。無論是自然語言或人工語言，它們可以說是前三項表達方式的基礎——理論模式。又因後二者亦經常配合語文的表達方式使用，因此，無論中文西文，語文的理論與結構深深地影響到資訊的表達與處理。

目前，雖然我們由報章雜誌上得知：計算機對於上述的各種資訊形態，均可作某種程度的處理，然而其中仍有許多處理的技術並未成熟到運用自如的地步，因此多未能普及大量使用。一般的計算機能處理得較好的，仍局限於以人工語言為礎的表達方式；譬如，數學運算或各種計算機語言所表達的程式。對於與自然語言有深切關係的各項，僅局限於結構化的欄位資料而已。譬如目前計算機中的檔案系統、資料庫等工具均不適用於處理散文、信件或公文的原始文獻等等。

在目前這種情形下，計算機處理資訊的能力可說是仍處於剛剛起步的幼稚階段，與自然語文的關係並不十分密切。因此，它才能成為世界各國「通用」的工具。在這種水準下，我們可將原來設計給英語使用的系統，稍事修改，添加中文的輸入、輸出能力，就可改造為「處理」中文資訊的系統了。市面上所有的「中文電腦」全是依這樣的模式發展出來的。也正因為如此，一般人的感覺是：中文資訊處理的能力似乎不是一種必要的事。可是，當計算機處理資訊的能力逐漸成熟之後，其處理的技術勢必對自然語文產生更多的依賴性，到那時才會明白這個事實，少努力「急起直追」或「迎頭趕上」時則為時已晚矣。

## 資訊處理的境界

對於資訊處理的各項功能而言，自然是有些容易有些難的情形。因此，有人試將處理的境界劃分為如附圖中的四個層次。圖中，最容易的工作是處理數



據。數據經過有系統有條理的整理之後，便產生了資訊。其實，此處所指的「資訊」亦可重作馮婦，再如上述的方式作為「數據」再經整理而產生另一階段的「資訊」。所以，數據與資訊二者，實是目前對於「處理」工作所定義的相對名詞：在處理之前則視為數據，在處理之後則視為資訊。然而，也因為這個緣故經常使二者含義混淆，每每使初學者不易掌握其確實之含義。

資訊處理較數據處理複雜。其複雜之處表現在其結構上，而其提升處理層面的主要技術是結構化的整理。知識處理是資訊處理更高境界的層面，其主要提升之技術在於一些時下列入人工智慧的功能，如推理、識別等。今年，日本已發表了以推理功能為硬體設計重點的計算機上市了，這對於知識處理的發展有很大的助益。所以目前計算機處理訊息的能力，就介於資訊處理與知識處理之間。至於智慧處理，目前則仍是一個夢想。若進步到這個地步，則計算機將擁有擬似人類的判斷能力與學習能力。至於如何才能到達此一境界，目前還沒有明顯的發展跡象。（待續）

資訊與電腦第五期

69年11月

論大學的資訊教育

## 一、前言

大學中資訊教育的目的是●重的。狹義的說，是培育資訊科學和資訊工程方面的專才，並從事於研究發展工作，以提高學術與工業水準。廣義的說，是訓練學生，都有能力運用自動處理資訊的工具，充分享用累積的智慧、知識、和專業的資訊，以協助他們未來的工作。

狹義的資訊教育，一如其他科學或工程教育，是屬於某科系的教育工作，然而，廣義的資訊教育就如同國文、英文等語言教育，是全面性、普及性的，也是一個現代大學生面對未來社會必備的基礎教育。

在過去的二十年的日子裡，我們資訊教育的推展幾乎全在大專學校裡，靠從事的教育家們，以他們的遠見、毅力，來推動資訊教育的發展。政府在政策上是支持這個工作的，但是坦誠地說，以往對資訊教育支持的具體措施，較之資訊教育對我們國家前途的重要性而言，是不成比例的。自去年起，我們很興奮地看到資訊工業的推動，如今，「普及電腦教育及使用」和「發展電腦、資訊工業」已成為施攻的重點，這是使我們國家有資格步入已開發國家的歷史關鍵和決策。

當此重要的時刻，感奮餘，讓我們了解一下資訊教育的本質和目前遇到的困難，以便確實地把資訊教育辦好，來配合資訊工業的全面發展。

資訊教育的目的是提供資訊建設所需的人力。因此，可以依其分為二類：其一是如何使社會上的在其位者和當其事者有能力從事資訊建設，其二是如何教育國家的新血，使他們在各級學校畢業，步入社會之時，就有能力參與資訊建設的工作。大學的資訊教育，即屬後者中最重要之一環。也是本文詩論的主要對象。

資訊教育發展的成敗，涉及的因素很多，大致可分為師資、課程、設備和

教育行政等四大類。本文先對資訊教育的發展過程作節要的回顧，然後就及課程師資和它們所涉及的教育行政問題分別探討。關於設備的問題，請閱本期「大學計算機中心縱橫談」。

## 二、大學資訊教育的發展

資訊科學和資訊工程發展的時期很短，迄今不過三十餘年，資訊教育的興起更是近二十年的事，但是進展之神速，是有史以來任何學科所不及的，而且方興未艾。由目前跡像顯示，到公元二千年以前，只有更加速的進步，絕不會緩和下來。因此，現在推動資訊教育，正是時機！

較具規模的資訊教育，發●於一九五〇年代末期，大致由美國的麻省理工學院、哈佛、賓州及伊利諾等大學肇其端。初期的教育形態分為兩類，其一是以邏輯設計的課程為主，傳授電腦電路的設計為目的。這類教育多發源自電機、電子相關的科系（以下簡稱電子科系）。其二是以偏重於組合語言的程式設計課程為主，傳授設計電腦中的控制程式和解決應用上的數學問題為目的。這類教育多由電腦中心協同數學系安排。此後的數年中，進展驚人，每年都更新深程的內容，且新添了許多課程。源於電子方面的，逐漸形成資訊工程（或稱計算機工程）的教育體系，而源於數學方面的，則逐漸形資訊科學（或稱計算機科學）的教育體系。資訊教育的兩大主流由此而生。

到一九六〇年代中期，計算機科學系首先自理學院的數學系中脫穎而出，成為獨立的學系，如伊利諾（at Urbana）、史丹佛、西北、德州（at Austin）等大學是。然而資訊工程則仍然留在電子科系中，但是由於資訊工程越來越多，影響越來越大，致使許多電子科系更名，目前多改稱為電機及計算機工程系（Computer Science and Electrical Engineering Dept. 或 Electrical and Computer Engineering Dept.）。諸如麻省理工學院、加州（at Berkeley）、密西根、南加州、哥倫比亞、威斯康辛等大學。

計算機科學和計算機工程究竟有什麼區別呢？我們最好由他們授課的重點上來判斷。計算機科學以傳授演算及程式的理論和應用為主，多有：自動機理論、語言理論、數值分析、計算理論、演算法、資料結構、和程式結構等課程，涉及硬體的課程不多。而計算機工程則著重於資訊系統的規格、設計、實作、實施和應用，它不只涉及硬體，亦涉及軟餅，而且目前軟體課程已較硬體課程為多，不論計算機工程涉及多少軟、硬體，他是工程中的一支，是研究資訊系統的結構、組織、設計和應用的科學，他使手的主要工具是數位處理的電子系統，例如電腦，就是最常用的數位處理工具，但是，別忽略了其他的數位

電子設備，例如數位儀器、數位通訊、數位控制等等系統均是，其重要性不亞於電腦。

我們的資訊教育，肇始於民國五十年交大電子研究所，民國五十二年，台大電機系繼之。大學四年的完整教育，則始於民國六十年，由淡江、東吳、交大、逢甲等校相繼設立計算機科學系，開始推廣。目前，計有：計算機科學、計算機工程、資訊科學、資訊工程等四種學系暨研究所的名稱。此外，許多電子科系中，仍有資訊方面的課程。在國內，常將「資訊」、「計算機」、「電腦」三個名詞互相交替使用著，略可認為是同義詞。就其含義及範圍而言，「資訊」是較恰當的，在本文中除科系名稱外，多用此語。國內的所系，大致上冠以科學為名的多重視資訊科學的課程，而冠以工程式在電子科系中的，多重視資訊工程的教育。

目前，國內每年約造就一位博士，一百位碩士，五百五十位學士，以及四百位專科畢業的人才。就發展的趨勢及需求看來，應有計劃地全面增加人數，其中特別重要的是碩士及博士的訓練，這是帶動發展的種子。基於精密高級工業的需求，碩士與學士的比例應於近年內達到一比二的結構，始可應付未來的需要。

一般科系的資訊教育究竟始於何時，已難考證，大致在一九六〇年代中期以後，已經意識到全面推廣資訊教育的重要性，之後，陸續地要求各科系的學生，必修資訊方面的課程。

在國內，如交大、清大、淡江等校亦自動於六、七年前，要求全校學生必修資訊的基礎課程，但是，今天仍有許多大學沒能做到，這是很大的缺失。

### 三、資訊教育的課程

在資訊教育發展的過程中，最重要的是課程的發展。此所謂的課程，包括授課的科目、內容、教材和實驗的安排，以及全部科目的規劃、結構和彼此間的配合關係。課程可以具體地顯示教育的目的和品質，反映出資訊學術和工業的水準，是教育行政中規劃師資和設備等要求的主要依據、它的發展過程，足以代表資訊工程和資訊科學的成長和內涵。所以，良好的課程規劃和執行，是大學資訊教育成功最基本的必要條件。

資訊課程的規劃，有三個重要的里程碑：第一個是民國五十七年發表的

ACM Curriculum 68. 為便於稱呼，以下簡稱為 ACM 課程,其次是 COSINE Committee 自民國五十七年玉六十一年間，一系列的課程設計，簡稱為 COSINE 課程。再次是 IEEE 的 Computer Society 自民國六十三年以後的努力，簡稱為 IEEE 課程。自民國五十七年迄今的十二年中，課程設計方面研究的論文和報告，有二百五十篇以上，目前尚且持續不斷。這說明了資訊科技發展的神速，迫使課程不斷的更新，以反映資訊科技的實質和配合實際的需要。這現象呈現出一個重要的事實，那就是：資訊科技的機動性極大，要辦好資訊教育，課程務必要日新又新，適時調整。這是教育行政上必需要配合的。

茲將上列三種課程之要點介紹如次：

#### (一) ACM 課程

ACM 課程是為四年制大學及研究所的教學所設計的課程,除了資訊本科的課程外,有關的學科，例如數學課程也包括在內。此課程雖然陸續受到其他研究的影響而有修改，在資訊教育的基本課程安排上，仍有其權威性，許多學校仍以它為課程之骨幹。

#### (二) COSINE 課程

這是美國國家工程學院（National Academy of Engineering）在一九六〇年代中期開始研究及設計的一系列課程。主要的對象是在電子科系中資訊教育的課程安排和內容。是目前資訊工程的課程主要依據。所涵蓋的範圍大致在計算機結構、數位系統、各種實驗課程、迷你電腦等領域之內。

#### (三) IEEE 課程

這是 IEEE 的 Computer Society 在一九七四年籌組的教育委員會設計的課程，又稱為模式課程（Model Curriculum），這個課程的主要構想是協調過去十餘年所成長的資訊科學與資訊工程間的關係，以期相輔相成。他們把所有課程分為四個主要系列：1.數位邏輯，2.計算機組統與結構，3.操作系統和軟體工程，4.計算理論。詳細課程請分別參考表一至表四。"（圖片省略）"

ACM 課程和 IEEE 課程的精神是有重要區別的，在 ACM 課程中，並不明顯地劃分資訊科學和資訊工程，認為他們是一個整體。而 IEEE 課程則認為他們是兩個主流，強調彼此之間交織著的關係。車施 IEEE 課程時，可以根據它的四個主要系列，依照自己所訂定的教育目的，選擇重點配合執行。由另一角度看來，IEEE 課程有意彌補日益加大的硬體和軟體之間間隙，也是減少資訊科學和資訊工程之間的差距，使資訊發展成為均衡、健全的一門學問。

由課程發展的趨勢看來，今後大學的資訊教育是朝著將理論（工程及科學）、硬體（工程）、軟體（工程及科學）、等合而為一的教育，以期更能培育出具有整體資訊系統設計、研究、發展能力的人才，不是往一方向教育而產偏頗現象。

由國內各大學目前所提供的課程實質來評，仍主要停留在 ACM 的課程的精神上，由於師資質與量的不足，在執行上能夠全面照顧妥善，保持良好的教育水準者實為少數。依 IEEE 課程的精神來評的話，則顯得不夠均衡，涵蓋不夠廣泛，要培養有能力從事資訊系統整體規劃、設計、應用等方面的人才，有基因難之處。導致此情形的理由之一固然是師資欠缺，另外的原因是系所的編制太小，不足以容納。這些問題都是我們今後共同期勉，努力的要點。

從一九七五年以後，有幾篇論文是呼籲在資訊教育的課程中，加入資訊系統管理、電腦中心管理，以及資訊研究發展專題管理等科目。國內教育界人士對於在資訊課程中加入相關管理科學之科目的做法，亦多有相同之見解，業者亦盼如是，這也是今後值得參考的做法。

關於一般科系資訊教育的課程內容，請參閱同期內「大專院校非資訊科系的資訊教育」一文，在此不贅述。

有一件事，我們應該在此特別指出的，那就是：許多國家已將資訊的基礎教育推廣到中學及小學去了。如果我們預定在五年內將資訊教育推到高中高職，十年久到小學的話，就會產生一個持續的課程向下推移的現象，也就是說：資訊的基礎教學的部份內容，會自大學而專科而高中而初中而小學，第次的推移。因此，課程的科目及內容，勢必持續地變化，才能夠順利的銜接每個階段的教育以構成一貫的資訊教育系統。由此觀之，這是第二個理由，使課程的修訂成為一件長期持續的工作，它應該作為一個長期研究計劃來處理。

有關課程設的一般參考資料請參閱文後附表。

#### 四、師資

目前在推展資訊建設的國家，都普遍的呈現極嚴重的資訊人力缺乏現象，這情形預估在十年之內都難以消失。造成缺乏的主要原因是資訊自動化的應用程度成長過速和應用面擴展太廣。其次是資訊人力培養的量過少。例如，去年美國市場資料顯示需要一千五百位博位，而當年的供應量只有三面人而已。

人力供應不足第一個受責難的自然是學校，為什麼不多教些學生呢？學校也有苦衷，那就是，學校本身就人力不足：師資不夠。

造成這種情形的原因之一是工作過重，資訊科技進展過速，教師們除了本身教學及研究的工作以外，還要不停學習，否則至多四、五年，他們就落伍

了。其次是新課程不斷地添加、選讀資訊的學生人數、班數不斷加多、非資訊科系的資訊課程不斷增加這種原因。自然造成學校中教師大夠的現象。但是，更糟的是：師資供應的來源幾乎斷絕了，很難找到新老師。大學中的師資一直是由大學本身培養的，由於市場需求過大，畢業生就業情況太好，願意就低薪教職者，百難選一，尤有堪者，有些教師甚至為了高薪改行到工商業界去了。造成一個標準的惡性循環。

社會上覺得缺乏人力的另一個原因是現職人員的能力不足以應付所需。否則就不必向學校要那麼多的人。這完全是科技發展過速所造成前所未有的問題。以國內的情況為例，大學四年的資訊科系成立於民國六十年以後，算上二年兵役，畢業生在就業時應是民國六十六年。試問大部份在六十六年以前就業的人如何解決職務上遇到資訊自動化的難題？再者，全國各校開設「微電腦」(微電腦)課程大部份在民六十四年以後，就算課在四年級，等他們就業時，也到了六十七年(二年兵役)，一則是：六十七年以前就業的人遇到了微電腦怎麼辦？二則是：即使是在六十四年學過微電腦，但是，六十七年微電腦的機種由己與六十四年的大不相同了，微電腦自民國六十二年問世以來，每二年就有一次重大的技術突破，至今非正式的算法已歷四代，即便是學過的學生，待服完兵役後再就業，就會感到落伍了。這情形中外皆然，根據民國六十六年電子雜誌( *Electronics* )的調查，約有一半的學生不滿意美國當時的資訊教育內容，豈非無因？這現象使得社會上認為學校供應的人力，不只是量不夠，連質也有問題。

綜合以上所述，資訊人力的缺乏，不是學校可以獨自解決的問題，其問題包括了一個龐大的在職人員再教育問題，絕非學校中薄弱的師資可以負擔的，亦非任何一個公民營機構可以承擔的，應該是社會各界全面配合，共同努力，才能解決的問題。在資訊科技發展和應用未趨飽和之前，這個需求一直會存在著，再教育的工作將永無休止，而且是資訊建設成敗的直接關鍵。這是一個教育的問題，大學應該負擔一部份的責任，但是以大學中自願不暇的師資，有心無力，目前能夠做的實在不多。

由以上所談的兩方面來看，學校要師資，再教育要師資，師資便成為整個資訊人力問題的焦點，而師資需要培育，此事實是當務之急。我們建議此事宜有全盤的規劃，先解決能培育師資的師資問題，以專案方式解除不合實際需要的名額限制，以可行的方案協助學校充實師資，也就是先要解決目前各校資訊師資缺乏的現況，並且要特別輔導國內的博士班，用以長期加張師資的供應，師資若能供應充份，則可收到以滾雪球方式加速人力供應的效果。

本文坦率地談到了許多大學中資訊教育的如題，每個問題都交織著許多複

雜的因素，要解決這些問題，人與人之間，機關與機關間的互相諒解、互助合作、掌握住我們共同的大目標是必要的條件。

環顧過去的廿年之間，我們國家曾經歷過多少各式各樣的困難，也擺平過多少荊棘。這一切，都在別人讚許我們進步的聲中，走了過來，雖然，目前推動資訊教育仍舊存在有許多困難的問題，雖然，以後的工作一直會艱鉅辛苦，但是，我們有信心，我們過去在進步，如今在進步，以後還是會艱鉅辛苦，但是，我們有信心，我們過去在進步，如今在進步，以後還國會進步。而在這進步的輪軸當中，一定有我們大家以血汗推動的痕跡。

#### 參考資料

- 1.ACM. Curriculum Committee on Computer Science: Curriculum 68; Comm ACM 11, 3 ; March 1968, P151-197
- 2.M.C. Mulder,model Curricula for 4-year Computer science and engineering program; Computer Vol. 8 No. 12, DEC. 1975 p28-34
- 3.IEEE;Special Supplement:Computer Science and Engineering Education(includes 11 papers) Computer, Vol.10,No.12,DEC.1977.
- 4.IEEE;The Oregon's Report:Computing in the 1980's (includes 7 papers), Computer, Vol.11, No.9, Sept 1977.
- 5.IEEE;Special Issue:Micro processors and Education(includes 8 papers), Computer Vol.10,No.1 Jan.1977.
- 6.A.I. Wasserman & P.Freeman;Software Engineering Concepts and Computer Science Curricula, Computer, Vol. 10, No.6, June, 1977.P85-92.
- 7.P.HSIA & F.E.Petry;A framework for Discipline in Programming, IEEE Trans on Software Engineering Vol.SE-6 No.2, March. 1980. P22-231.
- 8.IEEE;Special Issue:Computer Education (includes 4 papers); Computer, Vol. 13, No.6, June, 1980
- 9.IEEE;Special Issue:using Computers in Education (includes 5 papers) Computer, Vol. 13, NO.7, July. 1980.
- 10.A.Ailken; Computer Science Education in the 1980's Computer, Vol.13, No.7, July. 1980
- 11.W.Lin;Micro-Processor Education:Bottom Up Academic Course Series for Teardown System Design;Dept. of E.E & Applied physics, Case Western Univ.1979.
- 12.C.V.Ramamoorthy;Computer Science and Engineering Education,IEEE Trans. on Computers Vol.c-25 No.2, Dec. 1976 P.1200-1206
- 13.R.Austing, B.Barua and G.Engel;A survey of the literature in Computer Science Education Science Curriculum 68, Comm.ACM. Vol.20, No.1,Jan, 1977; p13-21



- 14.F.McFarlan & R. Nolan; Curriculum Recommendations for graduate professional programs in Information Systems: Recommended Addendum on Information Systems: Recommended Addendum on Information Systems: Administration; Comm. ACM. Vol. 16, No.7; July 1973, p439-441
- 15.S.Conte & M. Halstead; The Technology of Computer Center Management: A proposed course for graduate professional Programs in Computer Science or information systems Comm. ACM. Vol.19 No.7 July 1976, P369-370
- 16.M.E. Solam; Survey of Electrical Engineering and Computer Science Departments in the U.S; Computer. Vol.8 No.12 1975, P35-42

科學月刊

第十七卷第九期

記憶與速度——談計算機中的記憶層次結構

「儲存」是計算機很重要的功能。不僅是數據資料，程式也是儲存在計算機中的。這個性質，以目前大家對計算機的了解而言，似乎是理所當然的常識。可是，在計算機的發展過程中，儲存程式的觀念（store-program concept）是發展的里程碑之一。試想，若計算機不能儲存數據或程式，會變成什麼樣兒？那就沒有「計算機」了！它充其量不過是一具複雜的電子儀器罷了。

「儲存」靠一些「記憶體」來執行。在計算機中，記憶體的應用無所不在，並不限於專門做儲存的「記憶單元」、「磁碟」及「磁帶」等等。在計算機裡，凡是資料經過任何一過運作過程的先後，都要用一些暫時儲存的裝置，來存放待處理或已處理的資料。譬如，數學運的前後，數據必須妥為保存。又如在計算機的進出管道（輸入、輸出單位）上，資料亦須留置以待運輸或取用。這些都是儲存的功能，且必須靠些記憶體臨時性安置資料的工作，記錄器（register）或緩衝記憶體（buffer memory）是常用來做這些事的設備。這些記憶體的容量通常都很小，速率非常快。高速率的要求是必然的，因為它必須能匹配以電子速度進行的其他部門，如計算或傳輸，才能獲得較好的資料處理能力。這種速率匹配的問題和記憶體與生俱來，一直到現在還是設計工程師必須小心應付的重要問題。

大量儲存資料的記憶體，可概略分為主記憶體與次級記憶體兩類。主記憶體是計算機工作時主要的資料儲存場所，而次級記憶體則多以週邊設備的形式出現，作為大量儲存資料的倉庫來用。例如：說到計算機結構時所指的記憶單元是主記憶體，而磁碟、磁帶等是次級記憶體。

這篇文章將介紹計算機裡主記憶體和次級記憶體如何組成一個有系統的記憶層資結構（memory hierarchy），以及如何解決它們之間速率匹配的問題。這些討論將有助於對計算機結構的了解。

### 記憶體簡介

目前的主記憶體均由電子電路構成。由於其結構相當規律，是最適合以積體電路技術製作的電路。它是各種記憶體中速率最快的一類，也是單價最貴的。在積體電路發展以前，延時線（delay lines）及磁蕊（magnetic cores）都曾

是風光一時的主記憶體元件。以積體電路所做的記憶元件，它的存取速率和電子訊號須經過的途徑長短有密切的關係。所以它的速率約略與晶片的體積成反比。近十幾年來，積體電路製造的密度雖然一直增大，約每二、三年加大四倍，然而每塊晶片上的記憶容量也隨此比例加大，這使得晶片的大小沒有太大的變化。因此，積體電路記憶元件的容量雖然不停地增大，可是主記憶體存取速率的改進就顯得少的多。

記憶體有二個基本功能，是「存入」和「取出」資料。在專業上常作「寫入」和「讀出」。為了使用和設計上的方便，主記憶體存取資料時有一個固定大小單位，稱為「字」(word)。這個字的長度經常是和中央處理機處理資料的單位一致，因此，它代表計算機特性的一個重要指標——處理的能力。我們常見的所謂八位元、十六位元電腦，即指此字長。可是，word 被譯為字，不甚委當。譬如 computer 是一個 word，譯作中文時「計算機」是一個三個字的「詞」，不是字。此譯名沿用已久，積習難改。然而明瞭此間之差異將使你深得三昧。

主記憶體做存取工作花的時間是固定的，稱之為記憶週期。在早期的設計中，一個週期包含一次取出和一次存入的動作，而且一定是依先取後存的次序。當計算機開機以後，主記憶體即不停的重複期記憶週期。所以，處理機要存取資料時，必須等待到一個新週期的開始時刻，才能使用記憶體。在這種設計的安排之下，處理機和主記憶體必須依照記憶週期來協調同步，故稱之為「同步型」之主記憶。由於此設計較為簡單，迄今仍有部分機種沿用它。

較新的設計就不這樣做了。由於「取」、「存」的二個動作不一定是非連在一起不可，而且根據統計，「存」的機會比「讀」的機會小得多。因此記憶週期被分為「存」與「取」二個週期，可單獨工作。更由於「存」較「取」略快，所以同步的形態也沒有了。主記憶體空閒時，隨時可以聽令做「存」或「取」的動作，這種設計的方式稱為「非同步型」的設計。新式的主記憶體多用這種方法。然而為了討論上的方便，下文中我們不再分「存」與「取」之差異，一概以  $T$  表示非同步型主記憶體之記憶週期。譬如，讀取  $r$  個字串時，則  $rT$  的時間。

主記憶體是屬於隨取記憶(random access memory)的性質。簡稱為RAM，它的特徵(已如上述)是無論存取什麼位置的字，其花費的時間是一定的，與該字的「地址」無關。與RAM相對的是SAM，是順取記憶(Sequential access memory)的簡稱。這種記憶存取的時間是存放地址的函數，因為其資料的存取必須按一定的次序。譬如磁帶就是典型的例子。磁碟的結構是介於二者之間的，其儲存軌道的尋取是類似RAM的結構，而軌道內的資料仍然是SAM的

方式安排。這種結構稱為直接取用，而此類設備稱為直接取用的儲裝置（direct access storage device, 簡稱 D A S D）。

本文要談的是處理機、主記憶體和磁碟類記憶裝置（D A S D）之間的關係。磁帶在早期曾扮演磁碟的角色，可是目前漸淪為作備分資料或離線交換資料的媒體，與計算機處理資料的工作關係，逐漸疏離了。

磁碟和磁帶都是以電磁特性儲存資料的裝置，所以它的速度較慢，約在幾十分之一秒的範圍內，而主記憶體的記憶週期則以千萬分之一秒的單位計算，因此，此二者之速率有數十邁倍之差距。再者，處理器內運算的速率更以億分心一秒來算，又約較主記憶體快將十倍，如何安排一個有效率的整體記憶體架構，以使得記憶體的整體效率像 R A M，而價格像 S A M，一直是設計者面臨的課題，也是計算機內記憶結構安排的中心思想。

### 記憶體的層次結構

目前，計算機裡的記憶體結構分為三個層次：第一層為隱藏記憶（cache memory），這是直接和處理機相連的；第二層是主記憶體，它介於隱藏記憶與次級記憶體之間；第三層是資級記憶體，其主要的設備是直接取用記憶裝置（D A S D）。此結構如圖一所示。

隱藏記憶 cache 一詞是由法文而來，原意指隱藏東西的地方。由於使用者用計算機時，對有關記憶體的問題之思考仍是以主記憶體為對象，對 cache 視而不見（像是透明的），故稱之為藏記憶。

cache 是一個小而高速的 R A M，通常只有數千至數萬位元組（bytes）大小，用的是最快速的電子記憶元件，約比主記憶體快十倍上下。通常，一個處理機裝置一個 cache。c-ache 的構想最由見於英國劍橋大學韋爾克斯教授（M. W ilkes）1965 年的研究報告，它的主要角的是解決處理機與主記憶體之間的速率匹配問題。

cache 和主記憶體之間存取資料之單位較字為大，稱為框（frame），通常框之大小在 4~128 位元組之間，而常用的大小是 32 位元組。

當處理機向 cache 拿資料時，要是資料已在 cache 中，則直接取用，若不在 cache 中，則必須到主記憶體中，將該資料所有的那一框資料一起搬到 cache 中來，再予取用。因此在電路上較為複雜，它必須維持一個 cache 的目錄（directory），以便隨時查閱所要的資料是否在 cache 中。同時還要些管理的邏

輯，以決定在不同情況時，採取適當的尋取或存入資料的步驟。請參考圖二中的例子說明。

主記憶體和次級記憶體之間，以頁（page）為存取資料的單位，頁又較框大，這樣的系統稱為分頁系統（paging system）。通常，頁的大小配合磁碟的結構，將一段（sector）貝料作為頁心大小，其數值常在一千位元組左右。雖然此二者心間速率差在數十萬倍，但是很少見到利用另一個 cache 調協二者間之差距。然而在去年（1985），IBM公司宣布推出在DASD中加Cache的結構，在想的情況下，可將磁碟的平均存取速率自35ms降下10ms以下。相信這類系統，不久便會被廣泛採用在各種不同的磁碟系統中。

### 無 cache 的系統

讓我們思考一種最單純的記憶架構：處理機和主記憶體直接連接的狀況，沒有 cache，也沒有次級記憶。

通常，主記憶體是由許多完全相同的記憶庫（banks）構成的，而不是整塊的RAM。每個RAM做的記憶庫都有獨自的存取功能。這些記憶庫又稱記憶模組（memory module），通常以數目編號來識別之。例如，若有M個模組，則由0, 1, ..., 編至(M-1)號。為有效利用電路，M常為2的整數倍數。這樣模組化的目的有二，其一是加快速度以配合中央處理機心需。因為M個模組可以同時工作，所以存取速度可以有快約M倍之潛力。其二是當記憶系統出毛病時，有問題的模組可以不用，在中央處理機的控制之下，可以重新將沒有毛病模組編組使用，使計算機能繼續工作。

資料存取的地址是經過電路的選擇來控制的。換言之，地址將被選擇電路轉換為記憶模組裡的位置。這轉換的方式有二種：一、交織式（interlaving）；二、分割式（partitioning）。在交織式的方式下，連續的址依次指定在不同的模組中。例如：第0、M、2M.....等地址放在第0個模組中，第1、M+1、2M+1、.....等地址放在第一個模組中，餘此類推。在這樣的安排下，一個地址的指令，就可以取出M個連續的「字」。所以，平均每個字的記憶週期就降至約 $T/M$ 。這是一個平行的結構。

在分割式的方式下，一群連續的地址放在同一個模組中。若每個模組容量是C個字，則可能第0個模組的位置是由0至(C-1)，第一個模組的位置是C至(2C-1)，.....餘此類推。由此觀之，交織式的安排有較高之效率，若處理機要處連續r個地址資料時，在分割式的記憶下約需 $rT$ 的時間，而在交織式的結構下只需 $\lceil r/M \rceil \cdot T$ （註一）的時間。然而在電路上，交織式的將

較分割式的複雜得多。

為了建立處理機與記憶模組間的溝通，需要一個管理的電路稱為記憶控制器（memory controller），它將處理機存取的要求傳至記憶模組，並將記憶模組之反應和結果送回給處理器。此外，它還需應付對同一記憶模組幾乎同時提出存取要求的情況。處理器與記憶模組之間的連接通常有二種方法：一、用交換機（switching mechanism），二、用匯流排（bus）。用交換機可將任何一個處理器和任何一個記憶模組直接相連，它可使任何一個處理器借此通道，與任何一個記憶模組完成一次存取的動作。而這個匯流排是大家共用的。匯流排的控制部分，將負責匯流排使用權的分配工作。成熟的匯流排設計，並不用匯流排把一個處理器與對應的記憶模組完全連接在一起，而是將處理器的要求和記憶器的答覆分別處理。

當一次存取工作在匯流排中占用的時間遠較  $T/M$  為小時，匯流排的效率將趨近於交換機的。若有  $N$  個處理器， $M$  個記憶模組，交換機的複雜程度將與  $MN$  成正比，而匯流排之複雜度僅與  $M+N$  成正比。然而當  $M$  與  $N$  都很大時，這二個系統都會性生問題，對交換機而言將需太多的電路，而匯流排的反應時間將會加長。這種  $N$  個處理器分用  $M$  個記憶模組的安排方式，已是平行處理的結構，這種想法在硬體系統裡早已存在。

$N$  個處理器能使  $M$  個記憶模組忙碌到什麼程度呢？這是個困難的問題，因為許多處理器若要同時利用同一個記憶模組，會產生相當大的干擾。

因為一個程式（經常）會產生連續的地址，所以，若是  $N=1$ ，可選擇  $T/M$  之值較處理器產生地址之間隔時間略短，就可匹配處理器之速度和記憶體的資訊處理流量（throughput）。

當  $N > 1$  時，處理器之間性生地址的情形將是混合的隨機，而近似於亂數尋址的狀況。在此情形下，根據白斯凱特（F. Baskett）和史密斯（A. Smith）在 1976 年的分析，記憶體使用率效率為

上式的  $U$  是使用率，它表示每記憶週期中平均有效的使用程度。若  $N = M$ ，則  $U = 0.586$ ，即約有 60% 的使用率。若  $N = 2M$ ，則使用率為 76%。若處理器產地地址的間隔是  $T/N$ ，則處理器與記憶模組之間可得一完全之匹配。

這情況似乎不壞，然若使用 cache，則會更好。

## 使用 cache 之分析

如果在上述的例子中，在主記憶體與處理機中間加了一個 cache，情形會變成怎樣呢？當處理機向 cache 要資料時，若是該地址已在 cache 中，則直接由 cache 中取用。如果地址不在 cache 裡，則須至主記憶體中尋取，找到該地址後將之存入 cache 中，並將該字送到處理機去。由於 cache 與主記憶體之間是以框為單位傳輸，而框中心地址又是連續的，因此使用效率將提高。

設若框大小為  $B$  位元組，而主記憶體為  $M$  個交織式的記憶模組，則需  $(M-1)/2$  次來尋找所需取用的第一個位元組，再加上  $B$  次的時間來取其餘，因此，其使用效率為

若  $B$  為 32 位元組， $M = B$ ，則  $U$  約為 90%。故 cache 可提高記憶模組的使用率， $N$  之數目無關。

除上述的理由外，cache 還有許多其他好處。舉些數字，就可明白

一個典型的主記憶體約在 1 M byte ~ 64 M bytes 之間。常用的 cache 是 4 ~ 32 K bytes，典型的主記憶體記憶週期是 500ns (註二)，而 cache 是 50ns。在一般的應用中，不中的機會 (miss ratio，也就是指處理機向 cache 要求，而 cache 中無該記憶位置的情形) 是 5%，這是在 cache 有 20 K bytes 之估計，然而針對某些應用時，此比例可以少到 1%。在此情形之下，平均的取存週期是  $0.59 * 50 + 0.05 * 500 = 72.5$  ns。cache 雖貴，但由於量少，僅占全記憶體價格之 1/10 以下。然而整個記憶體的效能，卻幾乎與 cache 之效能相差無幾。

以上的討論說明一個不錯的記憶結構模式是：在處理機外的記憶體是交織式記憶模組與 cache 之綜合，並且可以調整 cache 的大小，使達到 95% 以上的使用率。

存在 cache 中的資料，我們可稱之為程式的工作集 (working set)。通常它們是程式中最常用的一群數據，無論它是指令或是 data。為了減少尋不中的比例，程式設計人員必須了解工作集的觀念，在程式中大要經常更易其常用數據。這觀念稱之為程式的區域化 (programming for locality)，這是很容易做到的，亦能使程式人員充分發揮計算機的威力。

註一： $\lceil r/M \rceil$  表示較  $r/M$  大的最小整數。

註二：ns 為 10 秒 (nanoseconds) 。



近年來，介紹計算機的書籍和文章到處都是。若是有心一探計算機的奧秘，不難找些適當的資料來了解它的構造、性能、原理和程式設計等等的基本知識。各種應用計算機的報導也是俯首可得。有時候，一些極富創意的佳作，不早得令人擊節之嘆。然而，見到的次數一多，也就不再那麼引以為奇了。自從政府推動資訊工業及各種自動化的運動以來，計算機的觀念、知識和使用經驗都大為普及。可是，如果我們問：「計算機科學是什麼？」，會得到什麼樣答案呢？

也許有人說，劣聽過這個名詞。也許有人認為它就是電腦，就是資訊，就是自動化。當然，也有人會小同意這說法，或許有人會反問道：「計算機是科學嗎？是工程？還是僅僅一部機器？」這些意見，說不定會引起一精彩的唇槍舌戰。總之，你將發現，不只不易解釋清楚，而且意見紛云莫衷一是。

其實，計算機科學一詞在世界各國已經用得很久了。如果暫且認為它是一門科學，那麼，像下面的這些問題：它是自然科學嗎？是基礎科學嗎？和社會科學有什麼關係？計算機科學和計算機工程有什麼不同？如同化學之於化工嗎？……等等。這一大籬筐的問題，那一個是好回答的？這些為難之處都指向一個癥結，那就是：計算機科學究竟是怎麼樣的一門學如呢？

如前所述，介紹計算機有關的文獻雖多，可是，介紹計算機科學的通俗作品，無論在國內外都難得一見。因為，介紹計算機這個「機器」容易，介紹計算機科學這個「科學」就不簡單了。計算機科學，這門四十多歲年輕的學問，由於成長過於迅速，變化得太多，無論在科學界裏，或是在一般大眾的心目中，都還未能樹立起清晰簡明的共識。話雖如此，它在過去的二十多年裏，對每一門科學的發展都產生了巨大無比的影響，這是不爭的事實。因此，在此時刻，以科學的態度對計算機科學的本質作一番探討，應該是一件很有意義的事。此舉不僅可以增進我們對計算機科學的了解，還可以藉此機會從更廣的角度來看看自然科學的本質。這些理早就是促使者撰寫本文的原動力。

非常歡迎讀者閱後提出問題討論或是批評指教。若是本文能引發大家探討計算機科學的興趣，就不枉爬格子一場了。

## 定義的初探

計算機科學一詞已經正式使用了很長的一段時間。譬如：在圖書的編目規則和分類法中；在學術界的學科分類上；以及用於檢索科學資料的詞彙（thesaurus）裏。在分類時，它是屬於自然科學的一支。雖然它的定義不容易簡明扼要的表達，但是人們見到它時都可以想像到它大概和計算機有關的種種。因此，就不再窮究它的定義和界說。套句科學的「八股」來給它下個定義的話，計算機科學可說是研究計算機有關的種種現象和問題的學問。但是，如果追問：「什麼是關於計算機種種的現象和問題呢？」我們就會發現，這個定義的缺點是不夠清楚和具體。若要一個更清楚的定義，就必須追根究柢，從自然科學的一些基本性質談起。

自然科學是一門自成系統的知識。它有自己的模式（models）和理論（theories）以解釋其內涵。它有嚴謹的研究方法。研究的重點是發現（discovery），也就是對自然事物觀察和分析的結果。由新的發現，可以擴大它的知識體系或是修正原有的使之更趨完善。這種由觀察、分析至建立模式和理論的路線，是一般自然科學共同的表徵。自然科學觀察的對象是自然現象的本質，也就是以人為方式無法改變的基本性質，亦即自然界之常性（invariants）。在說理時，以演繹推理的方式（deductive inference）最為普遍。

把計算機科學和上述自然科學的性質作一對比，我們發現首先應該弄清楚的是：計算機科學觀察的對象是什麼？計算機這個機器顯然不是正確的答案。試想，我們為什麼要創造一個複雜的機器，然後把它常作自然界的常性去觀察？這不只沒有必要，更無意義。這個問題對使用計算機有經驗的人來說，並不困難，答案也很單純。那就是：計算機科學觀察的對象是和程式有關的現象。在計算機裏，程式之間的關係是非常複雜的。因此，要了解整個系統行為的話，唯一可行之道是觀察這個系統共同的特性，再加以分析。而這種分析的方式和原系統裏的各個程式之間，並無必然的關聯性。這觀察到的性質，也具有常產的特質。在計算機科學中，由觀察至建立模式和理論的過程，完全因襲自然科學。這也許是它被納入自然科學的原因之一。但是，在計算機科學裏所做的研究都和計算機有關，換句話說，它受到計算機的限制。因此，有關計算機本身的知識是必需具備的。這一點是計算機科學與其他科學最重要的不同心處，也是計算機科學一個明顯的特徵。那麼，什麼是和程式有關的現象呢？談到這裏，就需要加一些計算機的觀念才能說明得清楚了。

在許多人的印象裏，計算機只是一個很會做算術的機器。這麼說，有其簡潔的好處。但是，只說對了一部分，一個更正確的說法是：它是處理「符號」的機器。也就是說，凡是和我們日常生活起居、工作等等任何有關的「消息」，能

夠有系統的編碼，例如變成 0 與 1 的形態，而能為計算機所接受的，就叫做「符號」。計算機主要的功能就是會處理這種符號。計算機如何處理這些符號呢？我們必需教它，而教計算機如何處理符號的指示就是程式。因此，在程式有關的現象中，首要的問題，就是如何將日常生活及工作中有關的種種「消息」，表達為計算機可以接受的形態。這就是所謂的「表達問題」（the problem of representation）。譬如：我們將十進位的數字以二進位法表示，就是一個淺顯的例子。各種數字，如整數、小數、負數、分數、合理數、虛數等的表達方式也不難理解。而向量、矩陣、各種方程式以及三角、幾何的量的量，他們的表示法就複雜了許多。至於物理現象，像是聲音、影像、能量等，就難了些。更進一步說，有關抽象事物及知識的表達就更困難了。到目前為止有關常識、知識等的表達法仍然是計算機科學中，在人工智慧的領域裏，最重要的研究題目。換句話說，這是計算機科學裏尚未解決的基本問題之一，其實，以目前我們自己對計算機了解的程度，去想想看什麼能表達給計算機知道，什麼又不能，就是一個很好玩的嘗試。

和程式相關的問題太多了，我們無法一一窮舉，讓我們再以一個例子說明。就談談程式本身吧。在我們生活或工作遇到的各種問題中，有許多是性質相近的。因此，在特定的工作環境下，我們可以為某一類型的問題，研究出一種共同可用的制式方法和程序來解決它們。這種解題的制式方法和程序稱為繹解方法（algorithms）。若從這個觀點出發，程式可視為是繹解方法的實施。換句話說，將繹解方法以計算機能懂的方式表達出來，以便加計算機替我們工作，這就是程式。因此，繹解方法和程式有親密不可分的關係，甚至可說是一體的兩面。然而，繹解方法和程式有親密不可分的關係，甚至可說是一體的兩面。然而，繹解方法所涉及的問題可就太多了。例如，若以計算機為工具，那些問題能解？諸如此類的問題的計算機科學的領域中仍然沒有完全解決。

上面所說的，都是計算機科學觀察的對象。這些對象是抽象的，在本質上，如上所述，都有其有自然現象中常產的特質，若是將之視為自然現象亦有其理由。那麼，將計算機科學作為自然科學的一支，更是順理成章的事。

和程式有關的現象，就說明到此為止。若想進一步了解程式間複雜的關係，只能沿著計算機科學主要的幾個分支，逐一去了解。總而言之，計算機科學發展得相當廣泛、分散，且多變化。若想一言以蔽之，不是失之過簡，就是有所偏頗，這也是不容易介紹計算科學的原因之一。

難為的名字

許多學者都同意，計算機科學這個名字取得不好，在字面上，它太強調計

算機這個「機器」了，以致於不能清楚地反映這門科學的內涵。坦白說，由上一節的討論，我們可以說這個名字不只不好，若是望文生義的話，它實在還有些矛盾、滑稽的成分在內。也難怪一直到今天，仍然有人建議給它改個名字。但是山於沿用已久，而且目前也沒有更好的名值得做天翻地覆的改鸞，所以只好仍舊用它。

許多學者認為，用計算科學（*computing science* 或是 *the science of computing*）遠較計算機科學妥當。筆者也有同感，所以本文就以「計算的科學」為名。計算一詞，以其含義作為此科學的名科，原甚妥貼（參閱科學月刊 74-9 月號拙作「由 *compute* 一詞談科學中的專門詞彙」）。它不僅減少了機器的影響成分，而且它的含義較廣；計算一詞和繹解方法在語意上有若干的重複，這些都是比以「計算機」為名好的地方。但是，由另一角度來看，它還是無法忠實地達北門科學的內涵（如前述），所以雖有此改名的呼聲，但是並未施行。

上山節的討論，或許有些讀者認為計算機科學所觀察的對象多少與資訊有關，那麼改名叫「資訊科學」豈不是很好嗎？誠然，有許多專家也持同樣的看法。資訊一詞和上文中所說：生活或工作中的「消息」這一詞，很難有清晰的界限，而且二者的含義有許多重複。「資訊科學」一詞是比「計算機科學」來得好且更接近原意。可惜的是，「資訊科學」由己名花有主，已經為兩個地方採用了。其一是在文學院裏，用在圖書館學和大眾傳播學中。這兒所指的資訊科學是指一群整理好的資訊以及其相關的系統，在其結構、運作、傳遞、應用、維護及影響等方面所涉及的知識。坦白說，這門科學若以「資訊系統的科學」為名將更為貼切。但是資訊科學這個名稱已沿用多年了，其次是在電機和電子的領域中，用於電子加碼訊號的處理方面，例如：消息理論（*information theory*）。因此，對於資訊科學這個名稱，就只有割愛了。

談到這兒，也許有人會反對以上的說法。你看，不是有些國言不學以資訊科學為學系的名稱嗎？目前的確是如此。國內的情況是很特別的。要言之，國內似乎沒有人介意為這門科學，取一個統一的名稱和下一個嚴註的定義。因此，國內的現況是：資訊、計算機、電腦等名詞交替使用著，另外一方面是：科學、工程與工業的區別也不用心思去分辨，再加上自動化、機器人…等等時髦名詞結構！例如：電腦科學、計算機科學、資訊科學、電腦工程、計算機車程、資訊工程、資訊工業、電腦工業、自動化工程、自動化工業、…等等，都沒有清晰的定義和界說。這現象發生在社會上的商業行為裏已經不太好了，似乎有欺負消費者的嫌疑。若是在大學裏和學術界也是如此，就不可原諒，這是違反科學的原則和科擗的道德的，在科學的操守上是大事饒恕的過失。

在歐洲，有一個很有趣的名詞「*informatics*」甚為流行。在該地區這一個名

詞幾乎可以做計算機科學的同義詞。**informatics** 是由資訊 (**information**) 和數學 (**mathematics**) 兩個詞彙分段拼湊而成。它字面的含意是指處理資訊的數掙。由於計算機科學的理論部分，像許多自然科學一樣，是以數學為基礎，因此，這個詞在字表面上對計算機科學的理論部分有很好的註釋。目前它還沒有適當的譯名，暫譯為「資訊數理學」吧！這個名詞取代計算機科學的機會似乎也不太大，但是，相信它會越來越流行。

目前，在找不到理想名稱的情況下，只有此著歷史的包袱，繼續用計算機科學這個名字。其實，若輔以清晰的定義建立起大家的共識，也就沒有什麼不好了。試想，許多語文中的名詞不也是這樣的嗎？重要的是，研讀科學時，對專有名詞不可以用望文生義的方式去理解，一定要切切實實去了解這些專門詞彙的定義。這是現代國民應具備的基本科學素養。

### 計算機科學與數學

計算機科學和數學不只有極深厚的淵源，而且還有許多相似的屬性。比方說，他們觀察的都是抽象的事物，就是一端。早期的科學發展過程中，有人主張把數學納入藝術的範疇，理由是它觀察的對象不是自然界的事物而是抽象的「數」，因而推論數學不是自然科學。然而目前數學卻有「科學之母」(**queen of sciences**) 之譽，許多科學的理論都建立在數學上。這一點，計算機科學也有類似的趨勢。也就是說，計算機科學所發展出來的成果，越來越為各門科學所採用，作為研究上必需的方法和工具。這也是前文所提到的，計算機科學對每一門科學的發展都產生巨大影響的根由。

若是探索計算機科學的起源，我們發現它源於三個很深的根：即數學、邏輯和工程。在計算機裏，邏輯的形態是以數理邏輯的方式表示。若是也數理邏輯視為數學，那麼，計算機科學可視為是數學和工程（尤其是電子工程）所結合的產物。再者，如前文所述，計算機科學的理論部分也還是建立在數學的基礎上。因此，有時候還真不容易將它們二者之間的關係分辨清楚。

有人這麼說：「數學研究的對象是『可化簡的複雜問題』 (**reducible complexity**)；而計算機科學是對付『不可化簡的複雜問題』 (**irreducible complexity**)。」其實，這是不對的。也許是計算機科學面臨的問題太複雜了，才有這樣的感覺。計算機科學面臨的問題雖然複雜，但如果無法之分解為較小的單元而是個別處理的話，就永遠無法了解這麼複雜的系統了。計算機科學所用的許多繹解方法中，有一個很有名也很有效的策略叫做「個個擊破法」(**divide and conquer**)，就是這種論點不能成立的最好例證。正確的說法是：在計算機科學中，解析問題時所用的繹解方法和數學中所慣用的分析方法性質不同。許多

計算機科學裏用的方法是不能為傳統的數學所接受的。

由於受到計算機限制的緣故，計算機科學中所用的數學有二個特色。第一是過於複雜而不易明瞭（*incomprehensible*），這主要是由於環境的限制及所用的方法所造成的。第二、關於這些複雜現象的識別（*identifying*）、量化（*quantifying*）和管理（*managing*）等的技術顯得十分重要；這種現象在計算機科學的每一個分支中都有。以上二點，也和數學有顯著的差異。譬如，在計算機科學中所發生的資源分配問題（*resource sharing/management*）、安排工作時段的問題（*task scheduling*）就是很好的例子。

在系統的合成方面，這二門學問處理的過程雖然類似，但是強調的重點卻不相同。數學是以定理（*theorems*）、無窮的程序（*infinite process*）和討論靜態的關係（*static relationships*）作為研究的重點。而計算機科學強調的是繹解法，有限的架構（*finitary constructions*）和動態的關係。數學裏，有句常用的話：「因為它是有限的，所以它是瑣碎的小事（...*the system is finite, therefore trivial*）。」這句話的意思是說：這個系統既然是有限的，就可以用窮舉法證明，因此它是瑣碎的事，不需費神再予證明了。如果這句話也適用於計算機科學的話，那麼它就割去了計算機學裏一大塊豐腴的江山，而使計算機科學所剩無幾了。在計算機中，由於環境的限制，窮舉法未必能行得通。於是講求效率的繹解方法應運而生，成了計算機科學中的要臣。

除了上述的不同點外，在計算機科學中所用的數學還涉及組織（*organization*）、政策（*policy*）和由於太複雜而產生的非固定的性質（*non-deterministic aspects*），這也是它和數學分家的本錢。

這一節或許比較生澀，讓我們再以一個故事來說明吧。故事是這樣的：據說有一個外星人偷偷地到地球來訪問了一段時間。在這段日子裏，他像 E.T. 一樣結交了一位地球上的好朋友。當這位外星人要返回他的星球時，很想把地球上的文明帶回去。他的地球朋友打算送他一套大英百科全書作為紀念。但是他的朋友卻因為這套書太重了，太空船無帶走，而為此發愁。當外星人知道此事時，哈哈大笑道，這不是問題。說著就拿出一支權杖，向他的朋友說道：「不必擔心。我只要在這權杖上點一下，就可以把這百科全書中的所有資料帶走，而不必帶那些書了。」他的朋友聞言感到大惑不解。只見他把百科全書上的字母和符號用一連串的「1」來編碼，例如 A 是「1」，B 是「11」，C 是「111」等等。不同的字母和不同的符號都用不同長度的「1」表示，詞之間用「00」來表示句之間用「000」來表示，餘此類推。當他用這個方法編碼後，全本百科全書的資料就變成了一長串的 0 與 1。然後他把這串 0 與 1 看成小於 1 的小數，也就是把小數點加在數字串的左端。這個時候，他問他的朋友說道：「你看，整套書

的資料都忠實地表示在這個小數裏啦。現在，我在這權杖上點上一點當做分段的記號，使得兩段長度的比例正好等於百科全書那個小數的值，如此，這一點就代表所有的資料啦。所以，書不用帶了。謝謝你，地球上的朋友。我要走啦，再見！」

這個故事其實是一則有名的謎（paradox）。只因為這個謎正好可以說明計算機科學和數學在做法上的差異，所以特揭於此。在故事裏，外星人的做法就是數學家的做法，在理論上是可行的，而且該數字是有限的，所以這根本是一個瑣碎的問題。但是，計算機科學定就不能這麼做，因為他找不出適當的表達方法來表示那麼精確的小數因此，他面臨的問題是一個變通的辦法，在計算機有限的環下，如何有效地儲存這麼精確的小數。換句話說，他在找一個繹解方法以應付這個問題（也可說是在找一個表達的方法）。

### 計算機科學與計算機工程

對大多數的自然科學而言，工程就是應用科學。因此工程和應用科學可視為同義詞。但是，也有例外。譬如：化學、應用化學和化學工程此三者卻有明顯不同的界定。科學和工程之間的關係是一個老掉牙的爭論。此二者之間實不易以工作的性質來區分。以工程來說吧，有的包括許多很困難的科學工作在內，譬如太空航空車程。但也有如家用水電這般，只需了解一些材料特性和手藝（skill）的工程。因此，一般區別工程和科學的方法是山工作的目的來劃分，而避免從工作的性質來判斷。如果一定要早產質來區分的話，我們只能說：工程是比較偏重實效的（pragmatic），是較源於經驗的（empirical），是比較追求實務（i-mplemcmtational）。但是，這些條件也僅是在程度上的差異而已。以目標而言，就可明顯地看出二者的別了。工程的目的是要做些東西給我們用，也就是以利厚生。而科學則以了解自然現象為主，其方是求知。因此，有人簡潔地說：「科學下去明白為什麼（know why）；而工程是去知道怎麼做（know how）。」實大失為雋語。

自然科學和它相關的工程之間均有一定的關係。例如工程的定義多半從它相關的自然科學的定義延伸而來。此二者的內涵難免有些重複，但常以分享理論的部分居多。在工程裏，「比較」和「選擇」是二項極重要的功能。因為，通常要達到一個工程的目標，其做法不只一種。所以，如何「比較」各個不同的做法，以「選擇」最佳者，是很其工程的代表性的。但是計算機車程之於計算機科學就不是這個樣子。在計算機科學中，「比較」和「選擇」的問題的處都是，使得計算機車程和計算機科學有更多重複的內涵，這也是計算機科學不同於其他自然科學的特色之一。也正因如此，有少數的學者仍堅持計算機學應該稱為計算機車程，並且應歸入工程類。

導致上述情況的根源，仍然是計算機的限制所引起的。由於計算機風可用的資源有限，因此程式的效益一直是令人關切的問題。換句話說，在計算機科學裏，幾乎隨時都會引發比較和選擇的挑戰。這情形使得計算機車程與科學在基本的問題上，經常糾纏在一起而難以區分。譬如說，在前文提到的表達問題中，為了要考慮系統的效率，而導致資料結構的研究。這種因果關係，在其他自然科學中是少見的。

## 結語

總而言之，計算機科學和其他的自然科學一樣，有自己的知識系統和規律（discipline），包括理論、實驗方法和有關的工程。就其內涵而言，自立為一門自然科學，應該是無疑問的。但是它太年輕，成長得太快，什麼時候才能像物理、化學一樣的定型，目前尚無人可以回答。

要深入了解計算機科學，最好的方法莫過於鑽到裏面去，看看它各個分支在做些什麼。下表是它目前主要的分支：

1. 計算理論（theory of computation），1940
2. 數值計算（numerical computation），1945
3. 硬體系統（hardware systems/architecture），1950
4. 程式語言（programming languages），1960
5. 解繹方法與資料結構（algorithms and data structure），1968
6. 操作系統（operating systems），1971
7. 軟體方法（software methodology），1975
8. 網路（networks），1975
9. 人機介面（human interface），1978
10. 資料庫系統（data-base systems），1980
11. 平行處理（concurrent computation/processing），1982
12. 人工智慧（artificial intelligence），1985?

表中的這些研究領域，將陸續在本專欄中介紹。但是介紹的次序和名稱並不一定依照上表。上表中，右邊有一行年代，是表示該項研究最盛行的年代，也可以說是奠定基礎的時候。由此可看出計算機科學有多麼年輕，變化的有多麼快、多麼大。



科學月刊第十八卷第五期

中華民國 76 年 5 月

中文資訊的處理（三）

在末期中，我們將討論中文資訊處理的基本符號集的問題。它是中文資訊系統要處理的基本對象。在討論此問題時，我們將發現：中文資訊處理系統已是一個中英文的雙語系統；中文字集是一個開放的集合；中文字集有許多語言文字學上的問題，包括字數、字形、字序等，使得中文資訊處理系統所要處理的基本符號集變得十分複雜。

### 符號的世界

計算機裡的世界是符號的世界：任何不用的訊息，無論其原始形式是以文字、數字、公式、圖表、聲音所表達的，或者是自然界中任何的物理量，或是文明世界中的抽象概念等，當它們置於計算機中時，都是用一群基本符號（symbols）來表達及處理的。所以，計算機可說是一個處理符號的機器，而這群基本符號集也因之顯得十分重要。

在任何一種計算機系統中，這個基本符號集都是有限的，而且它包含兩種符號：一種是控制符號（control symbols），另一種是圖形符號（graphicsymbols）。控制符號是用來操作機器用的，通常不用它來表達被處理的資訊。一般而言，控制符號包括控制周邊設備和通訊設備的符號，以及表達資料格式（format）及資料間隔的符號（information separators）。圖形符號是表達資訊用的，它和各國的語言有關，包括字母、數字、標點符號，以及一些常用的數學及特殊符號等。

為了使計算機心間可以分享資訊，並且也為了使週邊設備及通訊設備可以適用於不同廠牌的機種，上述的基本符號集需要有個共同的標準。這個標準直到 1973 年，才由國際標準組織（ISO）設定：編號為 ISO 646，全名是：7 位元資訊處理標準交換碼。

在 ISO 646 中，用了 7 位元的空間，所以共有  $2^7=128$  個位置可供安置符號；其中控制符號 34 個，圖形符號 94 個（圖表省略）。

ISO 646 是一個母法標準，各國可以依據其規定，設計自己的國家標準。為了達到資訊與設備共享的目的，控制符號是不可改變的。換言之，根據 ISO

646 設計自己的國家標準時，只可重新定義其 94 個圖形符號。這個標準對拉丁語系的國家是十分適用的，只要適當地修改圖形符號中的字母部分，就可言刻適用於該國，儀且可與其他各國的設備和資訊互容。

## 中文資訊處理的環境

對於使用漢字的國家而言，就覺得此一標準不能適用了，因為 94 個圖形符號的空間無論如何也容納不下成升上萬的漢字。於是 ISO 又制定了 ISO 2022 的標準來擴充。ISO 2022 允許將信元碼擴充至 8 位元，或是擴充為許多個位元碼一起來用。譬如：二個 7 位元碼的圖形符號就可以有  $94*94*94$  個，這已超過了 83 邁個符號所需的位置，即使要容納全世界的字母和漢字也是足足有餘的。

在這兒必須注意的是 8 位元的擴充方法。雖然 ISO 2022 提供了這種擴充，可是 8 位元的碼和許多通訊設備無法相容，因此也就無法利用或通過這些通訊設備。譬如：非同步通訊介面就是一個例子。再者，由於 ISO 646 是母法，以 ISO 646 為基礎而孳生的 ISO 標準就有幾十個，其中包括磁碟及磁帶的存錄標準在內，當使用 8 位元碼時亦無法與這些生之標準相容。所以，8 位元碼的使用還是有許多限制的。

由於美國是計算機王國，而英語又是最通用的世界語，根據 ISO 646 所訂定的美國標準交換碼（American Standard Code for Information Interchange，簡稱 ASCII）就成為計算機界最用的基本符號集。幾乎所有的計算機都必須用它，甚至在我國設計的計算機也不例外。

以上所敘述的，就是目前中文資訊處理所面對的環境。為了與世界各國分享資訊與設備，我們必須遷就 ISO 646 的標準。如何在這樣的限制下有效地處理中文資訊，就成為非常值得研究的課題。

## 不平等的待遇

常有人說：「計算機的硬體是很公平的，它只處理 0 與 1 的訊號，所以無分英文、中文，都沒有歧視！」這句話只能說是部分正確的。說它正確的理由是：計算機的基本符號集是早操作系統來經營、管理和運用的，不是由硬體來執行，因而主要的罪過不在硬體。譬如說，當鍵盤和主機溝通時，操作系統會依據基本符號集檢視每一個收到的訊息，當發現有控制符號時，便依當時的情況立刻做適當的處置，遇到圖形符號時，則留待其他程式處理。由此觀之，似乎不要修改硬體，只需早操作系統改起，便可使計算機適用於中文資訊的處理了，其實並不盡然。

並不盡然的理由是：關於設備的控制部分是依據 ISO 646 的規定設計的，而這一部分卻是依賴硬體執行的。根據 ISO 646，所有的控制符號都是以 7 位元為單位，因而在硬體設計上即以此為本。這樣的安排使得 34/128 的空間用於機器控制，壓小了表示訊息部分所能使用的空間，使得它無法容下中國字。因此，不得不擴充，而擴充的結果是：操作系統無法直接有效地處理代表中國字的符號，必須依賴應用層次的軟體來做，這樣的安排對計算機資源的利用是很不好的。

由另一角度來看，設若在訂定基本符號集時，即已考慮到中文，那麼此符號集的空間就可設計得大一些，毋需像目前用 7 位元的結構。例如用 16 位元的空間就可以有 65,536 個位置，足可供控制符號與中英文合併使用。這樣的話，處理中文資訊便不會像喻如此地委由和辛苦了。

然而，這樣更新事實上是極大易成功的，因為它改變了處理控制符號的過程，也影響到一些控制用的硬體規格。若要更新，目前在使用中的設備均將被波及，此代價實在太大。

#### 中文資訊處理的本符號集

看了以上的背景資料，再讓我們看看中文資訊處理系統需要那些符號。可以考慮的項目

如下：

- 一、處理中文資訊所需要的控制符號
- 二、可處理的中國文字集
- 三、發音符號
- 四、標點符號
- 五、ASCII 集
- 六、其他日常常用符號

關於控制符號部分，相當於 ISO 646 控制符號中格式攔制與資訊分隔符號的擴充。這是由於中文文獻的格式、中文文字之處理方式以及中文編碼結構等，都與英文的不盡相同的緣故，所以必須增加一些控制用的符號以應處理時的需要。譬如，中文有直印與橫印的區別，中文輸入時修改的動作複雜，中文碼的控制亦較複雜，而橫印與直印所用的標點符號不盡相用……等等，這些因素都是增加控制符號的原因。

關於發音符號則包括注音符號、國語注音第二式等。在國外的系統中並不包括發音符號，然而在我們的應用中卻少不了它。譬如：教學用、輸入用等，

所以也應收容。標點符號應略作改良，例如：書名號、私名號等與文字重疊印出的表現方法宜以其他方式表現為僅，這是便於機器處理的緣故。再者，為了應付橫直兩種印刷方式，標點符號亦應分為兩組使用。

關於納入 ASCII 集的理由已很明顯，國語注音第二式就非用到它不可，此外中英混雜的文獻比比皆是。將 ASCII 納入中文資訊處理的本符號集中的意義是很重要的，這樣做的結果是等於在發展雙語系統。而事實上，目前幾乎所有的中文資訊處理系統都是雙語系統。關於其他符號沒有什麼問題，只是選擇的工作而已。

### 中文字集的問題

在上列的項目中，最難決定的還是中文字集的問題。首先遇到的難題是：究竟有多少中文字？很不幸的沒有人知道有多少。目前最大的字書（即字典、辭典）是中華大辭典，共蒐集 49,950 個字。然而文建會的國字組在編輯中文資訊交換碼時，目前已發表了 53,940 個字，另外約還將增加 20,000 字將於年底發表。所以，已知的最多字數已經超過 74,000 字，可是仍可能有遺珠之憾。

在這麼許多字中，常用的字只占極少數。時下一般的字書約蒐集 8,000 至 13,000 字。國小和初中的所有課本總共用了 5,404 個字，新聞常用字（羊汝德先生輯）只有 3,000 字。根據林樹教授的分析，最常用的 4,000 字其使用頻率之百分比已略超過 99.6%，而擴充到最常用的 6,000 字時，使用頻率之百分比約為 99.88%。由此可知，中文字的使用頻率分布是極不均勻的。這種情形並不是缺點，反而可以利用它來設計一個效用高的系統。

正由於這種情形，有許多人建議只將一部分的中文字納入基本符號集內。目前我們的國家標準通用漢字交換碼，就只包含了約 13,000 字。可是這種做法造成了文字間的歧視，不在符號集中的文字將永遠無法以計算機來處理，這種情形自非吾人所樂見。於是通用碼只好允許使用人可以自己增加新字。這種方式似乎解決了問題，可是卻留下兩個大難題：其一是這種做法違反了分享資訊的原則，自己加的字是難以和別人交換的。其二是這樣的中文字集變成了「開放」集合，而不是封閉有限的集合（ISO 646 即是如此）。在計算機的設計實務中，開放系統的結構比有限封閉系統的結構複雜得多，而目前國內要做到這樣的系統是十分不容易的事。

除了字數的問題以，是否需包容滿、蒙、藏等少數民族的文字也是爭議的問題。

再其次就牽涉到中國文字的特性，中國文字有許多異體字。例如：「十」元與「拾」元的「十」、「拾」通用，此外如「群、●」、「●、鄰」、「栖、棲」等在一般的應用中皆視為同一個字。可是「路不拾遺」不可作「路不十遺」，「梧棲」不可作「梧栖」、「唐三藏」不可作「唐參藏」。也就是說當用於人名、地名及作為專有名詞時，異體字就不可視為同一個字。簡體字亦有此特性。這種情形使得中文字集的蒐集更為複雜。

再次就是字形的標準問題。標準字集當然需要標準字形，否則就不成其為標準。可是，目前只有教育部公布的 48,172 個標準楷書字形，對於常用的仿宋體、粗體等字形則沒有標準，這些工作都有待努力。

最後，讓我們再談一固字序的問題。所謂字序是說字的排列順序。試想成升上萬的字放在一堆，若是沒有規則的話，是很難找出來的。所以，中文字排列的方法涉及到中國文字的檢索。然而中文字的檢索一直是一個研究中的問題，根據社學知教授的估計，自民國元年到五十一年止，已發表的檢字法已達百種以上，如果加上近二十年來為中文計算機系統所發展的方法當超過 150 種。可是，仍然沒有一種大家公認為滿意的方法。中文字的檢索問題，又涉及文字的形、音、義三要素以為共定義的問題，關於這些問題將在下一期本專欄中討論。（作者保留著作權）

科學月刊第十七卷第八期

中華民國 75 年 8 月

光波上的訊息——一叢光值百萬言

經過百年成長，人們終於對無線電通訊的觀念習以為常了。可是，當初這個觀念曾是那麼神秘和難以想像得可怕。現在通訊界卻將注意力移轉到新近發展中的光學技術上。談到光的通訊應用，我們立刻聯想到古時綿延山野間的烽火，或是海上明滅閃爍的信號燈光。其實，我介入人天天用的「照子」又何嘗不是由光而帶來訊息的呢！

若將光纖傳輸訊息的容量完全開發出來，那麼，只需一根光纖就可以應付目前全美國電話的話務量。若經由光纖，收藏在美國國會圖書館內全部典藏的內容，只需幾秒鐘就可以傳送完畢。真的，我們已經發現了這種幾乎用心不竭的通訊媒體。目前，由於光纖技術大多數用於作為取代電子設備的同級品，所以它似乎還沒有把我們居住的土作明顯的改變。

利用光波通訊的研究肇始於 1971 年代初期，當時康寧玻璃公司（Corning Glass）的研究人員，製成了一條可以傳遞光線的光導管——一條約頭髮粗細的玻璃纖維。這個成就立刻引起了通訊業界的注意，因為與電子傳輸比較，這條光導管——光纖——幾乎有難以●計的「頻寬」。

一個通訊波道的頻寬是以它能負載的頻率範圍來計算的。譬如：由最低至最高音鍵所產生的頻率範圍，就是鋼琴能提供音樂訊息的範圍，也就是它的頻寬。頻寬的大小決定了資訊負載量的多寡。由於光線頻率遠較無線電的高出許多，它能負載的頻寬也就大得多，約是無線電波道的一百萬倍！

經過十年的努力，無論在負載能力及傳輸速率上，光纖均有驚人的進展。以矽土（二氧化矽）做的光纖而言，其透明度幾乎已近理論值的極限，使光脈波在經過 100 公里以上的光纖後仍然清晰可測。最近的實驗紀錄是：約十億位元（bits）的資訊——相當於 30 冊大英百科全書內容的分量——在一秒鐘內成功地被送到 117 公里以外。雖然此紀錄已足令人驚訝，然而它只及理論上大容量的五分之一，也許這差距就是未來十年努力的目標吧。

雖然在研究室裡成就斐然，可是以十年的光陰來開發整套的光傳輸科技是顯然不夠的。光纖構成的高速大道並不能獨自成為一個完整實用的通訊系統。

如何集結待傳的資訊，如何分配輸送的途徑，在難以想像的高速下如何做分封（package）和解封（unpackage）資料的工作等，都是待解決的「光交換」和「光邏輯」（optical switching and logic）的問題。

以目前的光纖科技而言，下一步應該是發展微光波管道模組，包括放大器、開關、偶合器、濾波器、絕緣子等，然後將它做在微積體迴路晶片上。一個光波管道應該能容下各種不同波長光線混合組成的訊號，就像無線電通訊系統處理混合波長訊號一樣的順暢，譬如：在調幅收音機頻帶中含有許多電台，各有各的頻率。目前，已上市的光纖系統通常只能藉單一波長傳輸資料，就像是只有一個無線電台的系統。由於處理許多頻道的技術尚未建立，使得光纖寬度的頻寬無法被開發出來充分享用。

由於研究人員仍然被果於上述的問題中，光纖技術的全面開發是不可能一蹴而成的。然而目前資訊的流量已將令電子系統不堪負荷，對我們有極大貢獻的電子終將會被突然發跡的光子取代。

最初，商用的光纖系統只用於連接都市裡的電話系統上。之後，長程的光纖系統出現在美國的東北部和加州。在 1990 年以前，美國將布滿這種大容量、小體積的光纖超級大道。

第一條連接美國、英國和法國的光纖海底電纜預定在 1988 年久開始運作。另一個類似的系統玄將架設在太平洋區，以供各島嶼間的連繫。

再下一步，就是把光纖接到每一石住家。這其實是一個價錢比技術更難解決的問題。像是目前的高速公路沒有急速的彎道一樣，也許光纖大道的銜接點會離住家還有里許之遙。以目前情況而言，光纖的價格仍然無法與電話線或有線電視所用的同軸電纜相較。然而，把光纖接到住家的殷切盼望，終將使以百計的電視頻道、雙向的視像電話能變得物美價廉而無所不在。這情形會使得影像資訊的利用非常方便，我們可以像瀏覽圖書資料一樣的從容閱讀螢光幕上的資料。

光纖之於目前的世界，就像高速公路之於一個蠻荒心地。發展光纖一定會帶動些新興的企業，開發出各種不同的應用。已往，人類從未虛擲任何一種通訊的科技，當越過開發期的挑戰以後，我們深信我們有充分開發光纖通訊的潛力。

科學月刊第十七卷第三期  
中華民國 75 年 3 月  
再談名詞的翻譯問題

談名詞翻譯問題實在是舊調重彈。然而，自從近幾期在本刊上登了幾篇文章以來，不少關心的朋友「問」我這件事，所以只得再公開談談，一則表示我自己的看法，二來希望讀者們與朋有們指教，以提升大家對這個問題的共識。

朋友們的意見表達方式雖不一樣，然而骨子裡的一句話都是：「老謝，你這傢伙真不守規矩，明明有標準的譯名，為何不用？是找麻煩？還是標新立異？」坦白說，這幾篇文章裡的確有些譯名是我自撰的，可是，我不是想找麻煩，也不是要標新立異，只是覺得，我這樣譯法較好——更能忠實地傳達科學的訊息」是寫通俗科學作品的首要之務，是最優先的原則。違反這個原則的任何限制，我都不想要「守」。所以，我就這麼做了。

引起些議論是意料中的事情，也是件好事。也許，有些老朋友又要笑罵道：「老謝真狡滑，不守規矩時就談原則！」是的，我正是如此。試想想看：我們有多少規章典制在執行時還顧念著制定它們的原意？八百年前就拋到九霄雲外去了吧！機械化的照本宣科不產生官僚式的品質偏差才怪——尤其是遇到沒水準的人在執行的時候。

上面所指的規矩或「標準」譯名是針對著國立編譯館所發行的譯名而說的。不是科月的——因為科月根本就沒規矩。也正因為沒規矩，才可提供一片學術自由的園地。若是科月有這樣一項規定：所有譯名必須遵照國立編譯館的文獻，那麼，咱這幾篇「作文」早就被老編在「照章行事」之際一腳踢出刊外去也。

即使你同意上面的說法，你也許還會責怪我不服從「大多數」不守「行規」——作為一個學者為什麼對職業道德和操守（CODE OF ETHICS AND GOODCONDUCTS）「脫序」？當然，盜亦有道，這點我同意。可是，我不太明白我這麼做是如何「脫序」？而更重要的事是忠於學術的精神——忠實地介紹科學的原來面目。因此，我不會去沿用錯的，或是不良的譯名。當然，這所謂的「錯」或「不良」都是相對的，又來了，它還是出自我盡己的自信。

如果我的譯名比原來的更好，那真對不起啦！我會沾沾自喜地好好樂上一樂。陶醉於自以為對科學有些小小的貢獻裡——提高了一點點作品的品質。若是我錯了，那更妙。希望你本著學術良心在大家談上痛斥我一頓，把我說得啞



口無言心悅誠服。如此一來，不僅對大家有貢獻——讓真理更門，而對我，好處可就更大啦——讓我佑道錯在那裡就可取你之長以補我學識之不足，我的「道行」不就更深一些了嗎？

什麼？慢點，慢點，你說我會因之而丟臉？錯啦！朋友，凡是源於盡己的自信是沒有等臉的事！若是不曾盡己地知無不言，言無不盡才是丟臉的事。根據上述的理由，我對名詞翻譯的做法是：

- 一、凡是錯的或是不良的譯名另譯之並加定義及原文。可能的話，將現有之譯名亦列出供讀者參考。
- 二、從俗，以便流傳。
- 三、不會譯的，或譯不好的用原文，不勉強譯。但仍在文中註釋。

啥？有人不同意第三點的做法？有違科學中文化嗎？開玩笑！別亂扣帽子。你以為弄些不三不四的譯名就是科學中文化人嗎？這樣做只怕適得其反吧！

唉！君不見 012,ABC,+-\*均係舶來品？sin、cos、log、●、Au、Hg、又何如？凡此種種已是我文化中心血肉，何能割捨？那麼何獨對 CPU,APAR,AI,……等等排斥呢？放開胸懷讓我們同化它豈不更好？試想，若不准大學教授上課用原文名詞會是什麼局面？再看看日本人的做法和他們今日的成就，就知道科學中文化最重要的是增進國人對科學的素養，而不是故步自封地守住一套既定的不良譯名，你說是不是？最後，讓我告訴你一件秘密，那就是：此文是剛看完龍應台的野火集寫的。「直」了些，希望你受得了。

近年來，計算機素養的教育，自大學，中學而小學，如火如荼地在國內外展開。甚至有些幼稚園也趕時髦；相關的補習班、訓練營亦如雨後春筍應運而生。因此，「學電腦」就蔚然成為流行風尚。

在國內，大學各科系均已到了非修電腦學分不可的地步，高職和高中設有計算機課程亦行之有年，而國中和小學則躍躍欲試。這些教育應有別於計算機科學的本科教育，故名之為計算機素養（computer literacy）的教育。可是，什麼是計算機素養？內容為何？恐怕難期共識。

在一般報導中看到的計算機素養教育，幾乎全是鼓吹它的好處，包括實施計算機教育的計畫在內，似乎所有文獻只有一個聲音：要計算機教育。在這種沒有反對意見、沒有深入探討的情形下，所帶來的訊息不是客觀、持平和完整的。我們應該了解執行這種教育有那些負面的影響，若不然，至少應該明瞭執行這種教育時應具備什麼樣的環境，執行時有那些限制和困難，以及了解執行後是否能達到預期的目的。凡此種種問題均需建立在對計算機素質的深入了解上；沒有深入的知識是無法把計算機教育辦好的。

目前，我們國內的計算機素養教育在心態上難脫趕流行之譏——咿人有了我也要；在內涵上，對計算機素養了解的不夠深入。譬如，在執行前有許多該做的調查和研究，卻付之闕如。又如，事後欠缺追蹤評估。在此情形下，產生偏差是意料中的事。目前一個很普遍的現象是：當你問學生「你修過計算機課程嗎？」，「是」；再問「計算機有用嗎？」，「很有用！」；若再問：「計算機能幫你（主修的科系）做些什麼事呢？」，答案是令人訝異的「不知道！」。

其實，目前學者對計算機素養的教育並無一致的看法。有人認為，計算機只不過是一種工具，實在不必大張旗鼓實施特殊教育計畫，以致妨礙了正常的教學。他們認為，要使用計算機並不一定要對計算機的理論和結構做深入的了解，也不需要會寫程式，所需的只是操作的經驗與技術；而這些並不足以構成開一連串課程的條件。事實上，使用計算機的人，絕大部分是「不知亦能行」的這一類。例如：玩電動遊戲、使用收銀機和售票機、使用計算器、操作自動化的生產設備、用個人電腦做文書處理（WP）工作，甚至於設定錄放影機預錄節目的時間、使用電子表設定鬧鈴等等，凡此種種都是在使用電腦，但卻均無需深入的計算機知識。在專業上，使用各種套裝軟體，如薪資、庫存等等，或如

各種計算機輔助之工程設計系統(CAD)亦復如是。這些學者並不反對在目前課程中稍許加入些介紹計算機的實務課程。不過，這也只是過渡時期的做法，理想中的教育情景是：在每一門專業課程中，如果會用到電腦化的工具，那麼都將這些工具的使用納入該科目課程中。這才是教育的理想境界，而不是像目前這樣單獨開一些課程的做法。

贊成實施計算機素養教育的學者，大致均同意此教育應包含：一、計算機的知識；二、使用計算機的技能；及三、和計算機有關的人文和道德面的論題等三大類內容。可是，這三類內涵應介紹到什麼程度？涉及多廣？卻難獲一致的看法。例如：程式設計是否應列入？就是一個長久以來相持不下的爭議。尤有甚者，目前資訊技術與設備的發展一日千里，因此整個關於計算機的環境不斷地變化，而素養課程亦難免隨波逐流，難以定形。因為益增執行上之分歧。處此情況，益顯得建立資訊素養教育的了解和共識為當務心急：唯有確實掌握此教育心目標和原則，並對國內之教育環境有深入的了解，才能在此動態的環境中因事制宜地做好資訊素養教育的工作。基於這樣的理今，權做下列之建議：

- 一、實施計算機素養教育應該以不影響本科教育之品質為前提。如果實施後反而降低了本科教育之水準，則無異於捨本逐末、自毀教育根基。由長期社會發展的趨勢來看，在資訊化的社會中，知識分子需要比現在更多的專業知識來駕馭資訊化的設備，這趨勢是無爭議的。因此，理想中的教育原則應是：提升本科教育品質為第一要務，其次才是介入計算機素養課程。
- 二、應立刻積極檢討目前實施的成效，以資改進。談到計算機的知識，脫離不了數學、邏輯和資訊相關的知識。若是這些相關學習心智並未成熟，就冒然插入計算機素養的課程，恐將白白浪費了投下的教育資源。例如，學生的年齡（心智的成熟）與相關課程的配合等，都應考究。由此觀之，高中以下的計算機素養教育之實施，尤其需要仔細規劃和評估。
- 三、應加強在資訊社會中，人文、道德與計算機（或資訊系統）三者間交互影響之探討。這一主題是目前實施的教育中所欠缺的。在計算機越來越普遍的情形下，它對人文和道德方面的影響將越來越密切。這種關係影響到未來社會的發展，實不可等閒視之。譬如：相關之法律，如智慧財產權問題與個人操守之行為準則之間如何協論？說到這裏，也許有些讀者認為筆者陳義過高，不切實際。然而，讓我們看看四周的

親友：有多少父母花了錢為兒女買電腦，卻不願花錢替小孩買軟體而任其抄襲（COPY）他人的軟體？這種教育得到的是什麼——助長仿冒，不誠實的行為？能訓練出怎樣的下一代？未來社會的行為規範要靠這一代建立，未來社會美景也要靠這一代規劃，了解電腦對社會文化的衝擊，才能勾勒出美好未來社會的藍圖。

資訊與電腦第五期

69 年 11 月

談大專院校非資訊系的資訊教育課程

在大專院校的資訊教育中，非資訊科系的資訊教育是推動全面資訊教育中重要的一個關鍵，在日本的大專院校中大部份科系的共同必修科目只有二個，一是日文、其次就是計算機方面的課程，這種安排足以說明一般科系中計算機基礎課程的重要性。

在一般的文獻中，討論此類課程的論文不少，在美國會開過多次會議討論此事，也制定過若干課程的規範，我們把蒐集到的資料整理出來，在此文中介紹，以為推動資訊教育的參考。

在美國三千人以下的人文科學，社會科學、自然科學或藝術的大專學校（Small Liberal Arts Colleges）多半有資訊方面的課程，他們的教學課程設計約分為八種不同的類型，是：

最低程度的服務課程以語言為主的訓練課程

A C M 68 課程

A C M 68 小型學校課程

A C M 一、二年級學生資訊課程

W H E A T O N 會議設計的課程

C U P M 課程

非常態的課程

茲將上述課程分別介紹於次

一、最低程度的服務課程：

這類型的課程是指僅有一門簡介課程的電腦教學，或者是有統計、數學或是物理等其他課程中，加入了電腦使用的教學方式，在 A C M 一九七七年發表的抽樣調查中，約有 1/5 的學校採取這種教學方式，在國內大專院校中，這種方式甚為普遍。

二、以語言為主的訓練課程

這種課程的安排是指以訓練計算機語言及應用程式為主的一種方式，通常均提供有二門以上的語言課程。根據上述統計提供這類課程的學校，約佔 1/4 左右，在國內大專院校中，這種情形也甚普遍。

三、A C M 68 課程

這類課程是指根據 ACM Curriader68 所設計的課程。當然，在非資訊相關科系中，無法完全照 ACM68 所設計的立課程，而包括以下之課程者均歸入此類之中：

(一) 必需具備的課程：

Introduction to Computing;Computers and Programming; Introduction to Discrete Structures;Numerical Calculus;Data Structures; Programming languages;Computer Organization;System Programming;

以及：

(二) 下列課程中必需有二門以上、Compiler Construction Switching Theory; Sequential Machines; and Numerical Analysis I and II. 根據上述統計，執行此類課程的學校約佔 12%，在國內尚無此類課程出現。

#### 四、ACM 小型學校課程

在 1973 年 ACM 的計算機課程委員會專門為了小型學校設計了四門資訊的課程，這四門課程並設有名字，他們的主要安排如下：

(1) 第一門課：介紹演譯法 ( Algorithms ) 以及利用一種高語言，來教導良好的程式設計技術。據統 Fortran 用的最多，依次是 BASIC 及 COBOL。

(2) 第二門課：介紹計算機在各行各業的應用；討論資訊自動化對社會的影響；如何以分析的立場探討電腦解決問題之能力；授以更多的演譯法及程式技巧

(3) 第三門課：是計算機結構的課程。其要點是注重軟體、硬體間相互的關係。

(4) 第四門課：是注重非計算性問題的應用與解法，包括檔案結構、資料結構，和相關的非計算演譯法等。

執行此類課程的學校約佔 10%。在國內無類似的安排。

#### 五、ACM 一、二年級學生資訊課程：

這是 ACM 在 1974 年發表的設計，主要是依據 ACM68 更新而來，專為一、二年級學生安排的資訊基礎課程，這個系列中包括五門課，他們是：Computer Programming I and II; Assembly Language Programming; Introduction to File Processing; and Introduction to Computer Organization.

這些課程大致和 ACM68 的基礎課程類似，唯一明顯的不同是加了檔案處理這門課 ( File processing )，這門課是以傳授更新、刪減、增加、插入等檔案處理的技術為主，輔以相關的資料結構觀念，這是值得注意的新趨勢。

執行此類課程的學校約佔 7%。國內亦無如此安排者。

#### 六、Wheaton 會議設計的課程：

這是 1972 年在伊里諾州 Wheaton 召開的研討會所制定出專為小型文理院校(包括藝術、自然科學、社會科學、人文科學等) liberal arts Colleges 傳授資訊科學所設計的課程，參與人員包括 ACM 68, Cosine 及 Cupm 等委員會之成員；此會議中建議在 Liberal Art Colleges 中的學生，應該修習下列五門資訊教育的課程：

**Introduction to Computing; Machine Structure and Programming; Information Structures; Survey of Applications; and Systems Analysis**

此課程之前三門課約相當於 ACM 68 的前三門課，第四門課略相當於 ACM 小型學校課程中的第二門課。而第五門課，系統分析是這課程設計的特徵所在，這門課的目的是授予學生如何去用邏輯及計量的方法來解決實際的問題。委員們認為這五門課是一個整體，應該修全才能收到預期的效果。

執行此類課程的學校約有 10%。在國內有些學校僅有系統分析的課程，所授內容則不一致，但未全照此課程系列安排。

#### 七、CUPM 課程

CUPM 是「大學中電腦教學課程教學研究委員會」的簡稱( The committee on the Undergraduate Program in Mathematics for a computational mathematics Course of studies )，這個課程是特別設計給數學科系的學生，其內容包括三門計算機有關的課程，他們是： Introduction to Computing, Computer Organization and Programming 和 Programming languages and Data Structure.

此中、前二門課與 ACM 68 相似，第三門課的精神是授予學生程式語言的基礎以及將存在電腦中的資訊組織及連接起來的重要技術，除了以上三門課以外，尚有幾門演算數(Computational Math)方面的課程。此類課程執行的學校不詳，在國內無類似之安排。

#### 八、非常態的課程

所謂非常態的課程是指除了能提供一些基礎課程的安排以外還有一些特殊課程，如：人工智慧、模擬，自動機理論等等。這類的安排多半有他們特殊的目的和理由。如此執行的學校約有 12%，在國內尚無此類安排。

當然也有些 Liberal Art Colleges 設有資訊方面的課程，根據前述的調查，約有 5% 沒有任何資訊課程的安排，不過這比例可能不止 5%，因為沒有寄回

調查表的學校設有此類課程安排的比例應該是不不少的。

由以上的情況來看，國內大專院校執行的情形還有待加強，但是在國內執行時除了設備、師資、教材，等等問題外，還有一個很特殊的問題是：一般大專院校系所的本身專業課程已經很多了，如果要加入四門到五門該料系與資訊相關的課程，實在是不容易安插，是否應該在此時機對現行的課程及教育方針做全面性的檢討呢？有待您的指教。

#### 參考資料

- 1 · R.H.Austing & G. Engel: A Computer Science Course Program for small colleges; Comm. ACM, Vol. 16, No. 3, March. 1973; P139-147
- 2 · J.L. Mckenney & F.M. Tonge: The state of computer Oriented Curricula in Business Schools 1970; Comm. ACM. Vol. 14, No. 7, July 1971; P443-448
- 3 · E.Horowitz, H. Morgan & A.Shaw, Computers and Society: A proposed Courses for Computer Scientistis; Comm. ACM. Vlo. 15 No. 4, April 1972; P257-26
- 4 · R.Austing & E. Engel: Curriculum recomm. andations for freshman and sophomon level computer science courses. ACM. S Committee Report, Virginia Inst Marine Science, July 1974.
- 5 · J. LaFrance & R. Roth: Computer Science for liberal arts colleges. Workshop Report SIGCSE Bulletin(ACM) 5.1, Feb. 1973. P.70-76
- 6 · J. Nerison: Computing in the liberal arts college; Science Vol. 194. Oct. 1976: P396-402
- 7 · Lopez, R. Reymond & R. Tardiff: A surve of Computer Science Offerings in SmLiberal Arts Colleges; Comm. ACM, Vol. 20, No. 12; Dec. 1977, P902-9



資訊與電腦第七期

70年1月

國字輸入方法之評估中文資訊處理系統的評估之二

### 一、前言

當我們接觸一個電腦化的資訊系統時，首先遇到的就是輸入的問題。假如你不給電「輸入」一些命令（**Command**），你和電腦之間就一點兒也無法溝通，那個呆頭鵝一電腦決不會心有靈犀地揣摩你的心意，自動地做些事以博取你的歡心。此外，凡是程式、數據、文字等等，一切由電腦處理的原始資料，也必需經由「輸入」的通路，才能存入電腦中，留待以後適時取用。

能輸入電腦中的資料當然不止前述的這四種，像圖形、聲音以及各種由複雜的儀錶所測知的訊號，都可以輸入到電腦中去處理，但是前述的命令、程式、數據以及文字卻是最基本、最常用的形態，也和電腦能處理的文字和基本符號（**Symbol Set**）有直接的關係。（請參閱上期本文。）由於每一類資料都略可歸納出他們的特徵，除了可以直接用文字和基本符號的組合與一定的格式去表示他們以外，也可以匠心獨運地利用他們共同的特徵，為他們設計方便、有效的輸入方式。例如在一個中文資訊處理系統中，一些命令就可以由人機交談的方式，像做選擇題似的以數字輸入，如此則可以免除使用的人以複雜的中文輸入命令。又如在大量的中文資料要輸入時，有些單位用人工先將之轉換為電報號碼，再借用現有的輸入設備，輸入這些以數字表示的中文，在中文資料輸入系統（**Data Entry System**）尚未發展完備以前，這未嘗不是一個可行的辦法。

在本文中，我們將不去討論上述的中文輸入方法，但是在評估中文資訊處理系統時，我們應將之列入考慮。

為了不使討論的範圍太廣，在本文中我們只討論以鍵盤直接輸入國字的問題，因為這是一個中文資訊處理系統必要的基本功能，而且中文鍵盤是目前最通用的中文輸入設備。至於以其他方式和利用其他設備的輸入問題，留待以後再談吧。

由本刊上一期中本文的討論，我們知道一個中文資訊處理系統，除了要能處理相當數目的國字以外，還要能處理許許多多其他的字母和符號，這些字母和符號也有輸入的問題，像是：英文字母，注音符號、標點符號，數學符號．．．．．等等也需要遵照中文的輸入方式輸入嗎？若由上述的兩種方法輸入同樣的字母或符號時，電腦內可以相通互用嗎？抑或變成兩個不相容的字母或符號呢？這類的種種問題，也請容許我們以後再討論，因為這類的問題和

本文中所討論的國字直接由鍵盤輸入所涉及的問題關連不大。但是，這些符號和字母的輸入問題，的確是目前許許多多中文資訊處理系統的盲點，處理的方式和效果都不盡理想，在評選時，宜特別小心查證。

## 二、中文輸入的根本問題：中文的檢索方法

中文的檢索是個古老的問題，自民國以來，已發表的中文檢索方法，有案可查的不下百種，但是能夠達到單一選擇而且運用方便的，一種也沒有，這就造成了中文輸入電腦的障礙，因為，很不幸的，中文輸入電腦時，一定要透過一種檢索的方法。

有些人以為一字一鍵的大鍵盤不需要用中文的檢索方法，只要找到那個字的鍵，按一下就可以輸入到電腦裡去了，最方便不過。事事上，成千上萬的字如何安排在鍵上呢？總有它的規則和次序吧！這就是大鍵盤用的中文檢索方法。操作人員要依字鍵安排的規劃、次序和所在位置去「找」那個字的鍵，就是檢索的工作，只是這工作全由操作人員的腦力去做的，並沒有特別的動作惹人注意而已。

中國文字的定義分為形、音、義等三部份，檢索方法則絕大多數是根據字形、字音來設計的。原則上說，如果我們只需知道一個字越少的資料，就可以檢索到這個字，那麼這種檢索方法就越好。到目前為止，除了某些大鍵盤以外，沒有一個中文資訊處理系統，是用傳統的檢索方法——部首和筆劃來作輸入的，這種傳統方法固然是無法達到單一選字的目標，但是真正不用它的理由，卻在於傳統的部首分類法需要文字學中有關字源的知識，使用此法時，嫌它需要了解文字的資料太多，雖然它的流傳又外又廣，卻無人願意用它來設計為電腦的中文輸入方式。

如果一種檢索的方式，只需知道一個字較少的資料，那麼，其對於操作人員要救的「水準」也就較低，操作也比較簡單，較易為一般大眾所接受。因此，評估中文資訊處理系統的輸入部份時，詳細評估它所用的中文檢索方法，比較檢索時所需的文字知識和資料，是很重要的。

## 三、深入分析的基礎：中文鍵盤的操作過程

不論是那一種的中文鍵盤，輸入中文的細部操作過程，可以用一個通用的流程圖來表示，如圖一。在圖中，矩形表示體力或腦力的工作，菱形表示一件判定的工作，圓形表示當時主要的相關資料，而單線及箭頭表示工作的過序。在方塊的左上角，以字母或數字來做該方塊的簡稱。

操作的過程是這樣子的：

首先，在中文資料中以位置或其他線索，去找當時要輸入的那個字（A）。找到那個字以後要去認識他（B）。然後找出這個字的索引作為按鍵輸入的依據（C）。對每一種輸入的方法而言A和B的工作都是相同的，但是C卻大不相同。例如一個照發音順序排列的大鍵盤，操作人員就要在腦子裡找出那個字的發音，又如用三角號碼的鍵盤，就要將這個字的三角號碼在腦子裡產生出來。操作至此，已將認識的國字轉換成一個或多個輸入的索引了，其次，便是要用腦子判定一個一個索引在鍵盤上的位置（D）。再根據此判斷的位置去尋找該按的鍵（E）。找到了鍵後，按鍵輸入一個索引（F）。此時，要決定一下是否已將此字的所有索引都按入了（G），如未完，則經由H尋找下一個索引的鍵，或回到原來的字去找一下個索引如點線所示，重複前述之重作直到將此字索引完全輸入為止。例如：用三角號碼的輸入時，每個字都需將DEFGH的工作重複六次，是分別打入六個數目字的工作量。前述的大鍵盤只需一次，但是D和E的工作卻複雜了許多。

當一個字的索引完全輸入以後，則至I從螢幕上看看輸入的字是否正確（J）。這部份工作大體上每個系統都是相同的，當操作熟練時，或可省去，如核閱無誤，則至K和L，L繼續工作。由以上的分析，一個正常無誤的輸入工作是由ABCDEFGHIJKL才能輸入一個字，在此中ABIJKL對任何一個輸入的方式都是相同的，各種輸入方法的優劣主要表現在CDEFGH這段路途中，而其中又以C和環路中的D及E最為重要，輸入方法的快慢，消耗腦力的多寡，使用是否方便等因素主要是由CDE而定的。

NOP的工作是處理一些輸入錯誤和需要在電腦中加入新字時的過程，而M的工作是為處理重碼字而設計的，如果操作的人沒有螢幕可以看到輸入的字形，那麼，這一部份的功能就全沒有了，但將增加以後更正檔案的許多工作。

表一 專業人員與非專業人員選擇中文鍵入方法的比較

項目	專業人員	非專業人員
應用泛例	大量資料輸入	人機交談式操作
輸入資料量	大	小
輸入速度	要快	比較不重要
工作形態	最好以勞力為主，並能訓練至反射動作式的盲目操作。宜儘量避免勞心之工作。	只要易學不易忘，並避免太專業化的動作。
On-line/off line	均可	以 On-line 為宜
交談式的操作	不宜全部使用，因速度慢，最適	

輸入之索引方法	宜但可為難字索引之輔助工具 一種已足	希望有許多種，以應付不同的使用人和各種工作之需
螢幕立刻顯示輸入之字形	Off-line 時可省去，但有此特色將方便許多	必要
編輯能力	需要基本的編輯能力	需要較成熟的編輯能力
創新字的能力	需要	需要
鍵入有重碼現象	宜避免，以免不規則的操作	可用
輸入索引之彈性 (如省去其中一部份)	並非必要	最好有
用戶自定的片語鍵 (Soft keying)	最好有	最好有
鍵盤大小、種類	不宜太大，以能盲自操作為宜	宜取適行者，避免特殊鍵盤，以減少學習之時間
初次學習時間	較長無妨，以三個月以內為宜	越短越好，最好在數小時以內
長久不用後再使用時需要之溫習工作時間	可較長	越短越好

#### 四、使用環境的考慮：誰是主人？

輸入的工作是人和電腦交往的首要通道，因此受使用環境影響很大，例如工作之性質，資料的種類，工作量和資料量．．．．．等，都會對輸入產生些特別的要求，然而其中影響最大的是人，誰是操縱中文鍵盤的主人？

使用中文鍵盤和電腦溝通的人，可分為專業性的操作人員和非專業性的用戶二種，專業性的人員多半是做資料輸入的工作者，間或做電腦控制之操作，而非專業的人員則以中文鍵盤為線上終端機的一部份，作資料檢索、查詢的工作居多。這兩類的操作要救差別很大，請參閱表一。此表中的項目，只是一般的考慮重點，還是要根據自己業務上的需求，將特殊要求列入一齊考慮才能做圓滿的評選。

#### 圖表省略

圖一的流程也大致上適用於英文的輸入，只是在輸入英文時，CD 及 E 的

工作很簡單，幾乎可以完全以反射動作來做，但是，處理中文字就完全不同了，CD 及 E 全是需要腦力的工作。如果某種中文輸入方式中，CD 及 E 的工作太重了，則不適宜做大量資料輸入的工作，因為無論把操作人員訓練到多麼熟練的地步，長時工作下來，必定會引起疲勞現象而影響工作的積效。

評估中文輸入的方法時，將每一種方法在 BCD 的工作量仔細列出，再加權比較，是很具體、精確的做法。

中華民國74年11月

計算機和語文(二)

漫談人工智慧

人工智慧是麥卡錫( J.McCathy )在1956年達特茅斯的一個學術會議中首先提出來的，至今年正好是三十年。在1980年以前，這類學術研究工作並沒有受到特別的關注。由於第五代計算機標榜的是人工智慧型的計算機，霎時間，人工智慧像是突然引爆的煙火，噴出了眩目的光芒，風靡了全球。

近二年來，教育部和國科會推動了些人工智慧的密集課程和研討會。資訊工業策進會和工研院電子所也已將它列入研究發展的目標。甚至民間的圍棋教育基金會也響應：舉辦「電腦圍棋比賽」。由此看來，人工智慧應該是日前國內外計算機科學發展重點之一。然而它是什麼樣的一門科學呢？現在，就讓我們來談談吧。

日本的灰姑娘

近幾年來，有「灰姑娘」(註一)這般奇遇的學術研究工作有二；就是生物工程和人工智慧。比較此二者，生物工程發展得平實穩健，中規中矩。而人工智慧就顯得頑皮；它散漫、多變以及欠缺條理。譬如說，它的研究範圍就是許多學者爭論的問題之一，從麥卡錫迄今一直沒有明顯的界定。再者，它的理論基礎也不厚實，學者們意見紛云，莫衷一是。因此，又換來數不清的爭辨。從學術成就上來說，成績是有的，但是並沒有什麼值得鼓舞慶幸的突破，仍然有許多基本的問題等待解決。照這麼說來，它變成日本的灰姑娘豈非異數麼？

說實在的，人工智慧之所以有今日的光彩，實是拜日本人之賜。日本人看中的是它發展的潛力而不是它今天的實力。換言之，第五代計算機計畫正是日本人以人工智慧的潛力做為籌碼，為爭奪世界科技領導地位所下的賭注。自從日本人下了賭注以後，美歐各國也陸續斥鉅資跟進，一時蔚為風尚。這場豪賭將如何收場呢？

人工智慧是計算機科學的家族中，最活潑的角色。也是計算機與哲學、心理學、語言學、系統科學、制衡學( cybernetics )和神經醫學等等之間溝通的橋梁。有人說，它目前是隻醜小鴨，總有一天會變成天鵝—為科學界大放異彩。也有人說日本人的賭注是賠定了，大家沒什麼好緊張的。我們關心的問題是：人工智慧究竟有什麼潛力值得大家這麼重視？

## 難解的情結

自從十七世紀，巴斯卡（ B.Pascal ）和萊布尼茲（ G.W.von Leibniz ）分別發表了機械的計算器以來，人們開始夢想著具有人類智慧般的機器。這心情是種複雜的情結—又則望又害怕。在歷來的科幻小說中，充分表露出這樣的情懷。這也就是人工智慧被看好的原因。

現代人很難想像在十七、十八世紀時，當那些盛裝的學者，來到一大堆奇形怪狀的齒輪前，看到它們湊在一堆轉動著，居然能正確的做出很複雜的加法和乘法運算的時候，心中是何等的感觸！做算術應該是很有智慧的果實來—如人們思考的效果。在想像中，機器似乎已變成了有思想的怪物，張牙舞爪地飛撲到人間來。機器能擁有智慧嗎？這可能嗎？

至此，哲學家和科學家就面對了這樣一個謎樣的問題和難解的情結。當巴倍奇（ C.Babbage ）的分析機（ analytical machine ）和程式觀念在 1833 年公開討論時，機器能做些什麼？這個問題再引起了科學家，哲學家和政客們濃厚的興趣。羅芙蕾絲夫人的觀點：「機器只會做我們教它做的事（註二）。」，至少仍有許多人持同樣的看法。為了了解機器能做些什麼，當時討論的箭頭曾指向用機器下棋和作曲。因為下棋和作曲一向是智慧型工作的典型。用句今日的口語說，它們是智慧密集形態的活動；外界的干擾少，環境獨立易於掌握，便於觀察和分析。因此，選它們做討論的範例是很適當的。此外，能改變自己的程式（ self modifying code ）也提出討論過。以上這些問題，都是人工智慧的濫觴。可是，至今它們仍然是擺在研究桌上的問題。

## 無形的巨影

羅芙蕾絲夫人的話雖是對的，但是她並沒有真正回答機器能做些什麼事這個問題。她只是把問題換了個形式而已。換言之，若無法審慎的界定什麼是人類能教機器做事的極限，我們還是不能真正地明白機器能為牠們做些什麼。比方說：「人類能不能教機器『學習』的本領，使它能學著做些需要有智慧才能做的事？」這個問題就很不容易回答。

其實，再追究下去，真正的難結是如何對智慧下定。義歷來的爭辨皆由此而起。若是機器可以擁有智慧，那麼，如何去了解它所擁有的智慧能到什麼程度？如何分辨機器的智慧和人類到底有什麼異同？知識是智慧嗎？常識又何如？那麼資料呢？…這無窮盡的問題會越問越玄，越問越不好回答，而終究會僵在人類所知的邊緣外，淪入哲學知識論的巨影之中。已往如此，此後亦復可能如此。換言之，所有人工智慧的努力，可說只是知識論的一個實驗罷了。

由另一個角度觀察，目前已經有許多系統，表現得蠻有智慧的，已往的研究絕不是沒有成就。只是若不能在知識論的巨影中出頭，人工智慧所追求的終極目標恐將是水中之月，似乎我們可以努力地更接近它，但永遠到不了！因此，較切實際的態度是去追求機器智慧的極限，而不是以人的智慧為標的，如果我們把智慧定義為人類獨特的行為；是那麼玄妙、神秘和難以理解，那麼，沒有機器能做得得到。如果我們允許以觀察和實驗的方法，比較人與機器的行為來界定智慧、知識、思想…等等，那麼，應該可以獲得相當程度的成果的。

因此，人工智慧的研究工作可以這麼說：它是研究如何設計「智慧型」的計算機系統，以使此系統具有模擬人類智慧的行為。所謂智慧型的行為，目前研究的項目有：推理、解題、規劃、學習、認別、了解、表達…等。這些當然還只是智慧的一部分，但即使如此，其研究的領域已相當分散了。

### 情到濃時情轉薄

今天計算機所做的事，遠比做加法和乘法多得太多。可是，人們並不覺得它是多麼不可思議。前述「智慧之磨」的情懷，似乎今天都不存在了。這是不是有些了盾呢？有個戲稱為泰思洛定理的（**Tesler's Theorem**）是這樣說的：「人工智慧就是指那些目前還沒能做到（自動化）的事。」（註三）當然，此話是誇大了些，然而卻對人工智慧研究對象—永遠在變的情形，做了很好的詮釋。

在1950年代，剛開始有計算機下棋的程式時，雖然下得很「糗」，而當時認為它就是人工智慧了。如今，數百元台幣就可買到的下棋電玩，比當時下棋程式高明得多，然而，我們並不覺得它有多聰明和多了不起。在1960年代初期，有些會證明一些數學定理或以符號處理的方式解些微分方程的程式，也是很了不起的。而今天，有更強的程式用在計算機輔助系統中，似乎也是理所當然的事。由這些事實，我們不難明白：人類對某些問題了解得越多，就越覺得解決它是例行公事，不再是很有「學問」的事，智慧的成分似乎變得淡薄了許多，而且越來越淡，直到看不到為止。

去年，明斯基（**M.Minsky**）也幽了人工智慧一默。他說：「如果一個超越我們的外星人來研究人類的行為，他很可能認為我們的外星人來研究人類的行為，他很可能認為我們愚蠢得連一點點智慧都沒有！」人工智慧就是這樣——一個處於科技進步的尖鋒上，不斷地向前挺進，但卻又永遠沒有固定目的物——的一門學問。

### 機器會思考嗎？



回顧十九世紀以前的這些努力，雖然它們促使計算機科學和人工智慧的萌芽，但是並沒有引起社會大眾廣泛的注意。真正首先引起震撼的是塗林（ A.Turing ），接下來是先前提到的第五代計算機計畫。

塗林被尊為人工智慧之父。也和丘吉（ A.Church ）所建立的計算理論（ computation theory ），對人工智慧和計算觀念的建立有劃時代的貢獻。計算理論所討論的是「計算」的本質和天性（註五）。它是架設在計算機、程式結構以及制式（ formal ）的推理法則之間的橋梁。在塗林的計算理論中，最有名的是他發明了一個簡明通用的非數值計算的模式，以說明計算的本性。此模式就是有名的塗林機（ Turing machine ）。塗林曾以它直接討論人工智慧的範疇。

根據他和丘吉的論點（ Church-Turing thesis ）：凡是可以用塗林機表達的問題，如果塗林機解不出來，那麼，人也解不出來。反之，對於人能解的問題或能做的智慧型工作，總有一天機器也能做到相同的程度。之後，有許多學者曾嘗試著希望能用它去控導出：「人類能『知道』的和能『思考』的就是能『計算』的」這樣的結果出來。然而，這論點終究不是科學上的定理，它只不過是將前述知識論中爭論的問題，以另一種方式表現出來罷了。

驀然回首

事實上，探討邏輯、數學和計算等的極限的努力，可追溯到巴倍奇時代。當時研究邏輯的實證學派（ logic positivism ）就曾以嚴謹的態度，用他們的理論去詳查知識論中的各個論點。知識論是探討知識的定義、知識的起源、獲得知識的方法和探求知識的極限…等問題的學問。當時，學者心目中根本不知近代計算機為何物。他們試著以數學來為知識的一些基本性質下定義和立模式。其中，布爾（ G.Boole ）和第摩根（ A.DeMorgan ）在學術上有永恆的貢獻。他們所發表有關思想的法則（ laws of thought ）的論著，是啟開數理邏輯的鑰匙，也是今日命題演算（ proposition calculus ）的前身。

這些研究對以後人工智慧的發展有極深遠的影響。例如，命題演算至今仍是在計算機中常用的知識表達方式之一。然而知識太複雜了，不是只憑邏輯就可以完整表達的。雖然到今天，除了邏輯以外，還有別的嘗試，諸如：以程序、以語意網路、以制式的產生系統（ production system ）和以揮灑表現自如。換言之，到今天在人工智慧的研究中，知識的表達法還是叫每一個系統頭大的問題。

到了1930年代，哥德（ K.Godel ）的不完整定理（ incomplete

theorem )更對邏輯、數學和計算的能力極限做了非常重要的詮釋。要言之，它們的能力是有限的。比方說，此定理可證明所有的制式系統 ( formal systems )，如果不是太簡單的話，都會產生一些無法斷論的命題 ( propositions )。這個定理被荷夫史達德 ( D.R.Hofstadter ) 稱為數學中非常重要的成就 ( the most important in all mathematics )。

這些研究和塗林的研究有薪火相傳的因果關係，他們的努力不只對計算的極限有重大的貢獻，同時也是近世計算理論的啟蒙者。

計算理論是計算機科學中討論些基本數學問題的學科。也就是研究關於計自機、程式、釋解方法 ( algorithms ) 及一般資訊處理系統中有關可算性、效率性、複雜性等等的問題。從 1967 年起，它的發展相當可觀，也更正了許多已往的錯誤觀念。目前，我們對於那些問題可用程式解，那些不能，已推導出了些可用的理論。譬如，對一般性通用的程式系統，我們已經了解對計算能力的極限有了較清晰的了解。但是，這些了解仍舊不足以明確指出，機器智慧的極限究竟在那種境界裡。

## 燈火闌珊

現在，讓我們看看人工智慧的研究究竟做了些什麼？前文已談過，在這行裡的研究對象與推理、解題、規劃、學習、認別、了解、表達…等人類的智慧行為一定有某種程度的關連。以下的這些項目都曾是人工智慧研究的熱門題目。它們和上面列舉的智慧行為有什麼關連呢請你想想吧！

- 下棋、打牌、其他局戲 ( game )
- 繪畫、作曲
- 證明幾何、代數中的定理
- 以類似人的做法，處理符號的方式，去解題，化簡數學式，求無窮級數的和
- 去認別手寫或印刷的字母、符號、中國字等，認識稿件中的段節和標題
- 認別照片中的特定物，將景觀依物分解
- 認人 ( 由像片 )，認指紋，認別飛機、船艦的類型，認幾何圖形
- 走迷宮，解謎
- 聽得懂人在說什麼，把文章讀給人聽
- 了解自然語言，回答人的問話，分析句子結構
- 協助診斷病情並擬處方或安排治療與檢查的步驟
- 協助探勘油礦，規劃計算機系統，修護儀器設備
- 依學習者的性向和程度，規劃適當教材
- 機器翻譯，用程式設計程式

上列的每一個單項，它的水準都無法與傑出的人相比。可是，要到什麼程度才算得上智慧呢？從學術的角度來看，真正重要的是這些擬具智慧的程式，它們的知識的表達方式、程式設計的技術以及對計算觀念之建立等等的貢獻，才是學術研究追求的重心。

## 結語

在計算機科學開始的時候，人工智慧就已同時誕生了。在成長的歲月中，人工智慧研究所得到的觀念對計算機科學發展的影響是既深且遠。例如，早期記憶單元、處理機、控制單元的結構設計就有明顯痕跡。此外，對各階層的程式語言的結構、程式設計技巧、軟體工具的發展都有不可磨滅的貢獻。

由是觀之，日本之賭也不無道理。第五代計算機這場世紀之賭，說不定會大幅改變人類乙往的生活形態，這個結局不會等太久，應在五年之內，便知分曉。總之，無論由那個角度來看人工智慧都可說是最多采多姿的科學。牛威爾（A.Newell，註六）曾說：「科學的面目隨科學家凝煉自然的現象而顯露，科學只能發現不能創造。」這也是人工智慧忠實的寫照。

有人說，物理學家會問：為什麼這個專欄叫「計算機與語文」？到此為止，我們並沒有談到任何語文上的問題呀！是的。但是，你可曾想過：我們人類是用什麼來表達知識和智慧的呢？

註二：巴倍奇有「計算機之父」之譽，他所設計的分析機雖未完成，但其設計之規格直至滿百年之後的工業水準去試仍覺困難。關於分析機的資料甚多出自羅芙蕾絲夫人的譯著。此句之原文為：“The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform.” 美國軍方的標準計算機語言 Ada，即以她的名字命名。她實是討論程式功能的第一人。

註三：這句話是泰思洛（L.Tesler）說的，故名。其實，它不是學術上的定理，它和管理科學圈中戲稱的「莫菲定理」（Morphy's law）性質相類似。它們只是以幽默的口吻強調某些事情的特質罷了。

註四：塗林的名句：“Can machine think?”。事實上，若深究這句話的解答，就引至人工智慧的問題癥結所在。故此問句傳聞頗廣，為點睛之雋語。

註五：文中「計算」（computation）一詞之含義，請參考科月74—9月號大家談拙著「由compute一詞談科學中的專門詞彙」。請勿望文生意，以到對本文產生誤解。

註六：本文所舉之人名；MacCathy、Tesler、Minsky、Turing、Church以及Newell均係研究人工智慧大師級的學者。

科學月刊  
第十六卷第十二期  
計算機和語文專輯

## 卷首語

數千年來，語文是世界上任何一個民族文化延綿的主要工具，是人與人交往的橋梁，也是人之所以異於萬物的重要表徵。可是，自從計算機發明以後，語文卻兼作人與計算機之間溝通的媒體。這現象不僅相互指出語文和計算機方重要性，更明白的告訴我們語文和計算機密不可分的關係。

語文是人類社會的產物，不是天生成的。為了區別人用的語文和機器用的，我們稱前者為自然語文，後者為人工語文或計算機語文。目前，此二者有明顯的差異。然而，科學家們正努力消除它們的差距，希望有一天，計算機能有相當程度的語文能力，用自然語言和我們溝通。

如果這一天真的來臨了，就代表著計算機和文化的結合，其影響之深遠將較以前任何一個計算機帶給我們社會的衝擊為大。屆時，「計算機為人人所用」的理想才能實現；我們才有機會將計算機的功用發揮到極致；再者，計算機將擁有一些知識並具有某種程度的智慧，因語文、知識與智慧實是交織不可分的。

本期的專輯就是想把語文、知識、智慧和計算機的關係作入門性的介紹。但是，這是一個科技整合性質的主題，涉及的學科很多。上期的專輯「神經科學」就是其中之一。神經科學的知識有助於對人類認知行為的了解。此外，和哲學、語言學、心理學、計算機科學等等都有密切的關係。因此，我們安排了前二篇（蘇元良「認知科學—心理學與人工智慧的結合」以及朱愷「知識及知識的表達」）來介紹什麼是認知的行為、什麼是知識以及它在計算機中如何表達等作為了解主題的基礎。後二篇（曾士熊「了解自然語言—計算機處理自然語言的簡介」和陳克健「電腦可以和人類交談嗎？—談電腦如何分析中文」）著重在介紹如何用計算機處理自然語文和中國語文的問題。關於文詞表達方面，容或和已往科月的文章有些不同，但希望讀者能體會不同層次知識的表達方式。

由於篇幅和籌備時間上的限制，本專輯在協調上不是很周延的。然而，本人深深感謝四位作者的合作與努力；在一個月內能有這樣的成績，實已難能可貴。如果有所缺失，警請指正，並且希望有機會在以後的科月中彌補。

最後，我們要指出：計算機處理自然語言的研究著濃厚國家意識的，因為語文本具有民族的色彩。如果這件事——用計算機處理中國語文——是有用的，那就只有靠我們自己的力量來努力了。

科學月刊  
第十六卷第十二期  
功業彪炳的機器

計算機是二次大戰期間英美兩國追求科技創新狂熱風潮下的產物。英國的科學家塗林（A.Turing）曾參與一項極機密的計畫：製造一個名叫 Colossus 的計算機，以破解納粹密碼機 Enigma 所產生的密碼。Colossus 於 1943 年正式啟用。它該是世界上第一個用真空管的計算機（約用了 2,000 支 { C 許多軍事分析家以為它是導致希特勒戰敗的關鍵之一，因為德國人始終不知道 Colossus 的解碼速度幾乎和 Enigma 的編碼速度一樣快！

由於此計畫極為機密，美國在設計第一架通用型的計算機 Eniac 時，對此事毫無所知，還一直深信 Eniac 是創世界紀錄的。Eniac 是 1943 年開始建造的。主要的設計人是物理學家毛希理（J.P.Ekert）。1945 年秋，Eniac 在賓州大學試車成功，次年二月起正式服役。它的主要任務是計算炸彈和火箭的軌跡。當時，這個工作是要用二百人以桌上型的計算器來做的。Eniac 共用了 17,468 只真空管，工作時產生的熱量足以使室溫升高到氏 120 度，並因此經常熱壞了真空管。每當機器故障時，就有成群的技工穿梭在機架間覓找燒壞了的真空管。不過，雖然如此，這怪物還真管用，用了近十年的時間。除計算軌跡外，後期還做些有關氣象預測的計算。

Eniac 之後的四十年，計算機的改進始終沒有間斷過。它的基本元件由真空管、電晶體、積體電路、微處理晶片、記憶晶片等等，戲劇化地大幅減少了它的體積及消耗的能量，並加快了它的速度。

除了上述硬體的因素外，計算機的功能隨軟體之發展而擴增。由早期數值計算的工作到處理複雜的視訊及音訊的符號。要言之，它可以處理任何能以符號表示的資訊。例如設計、生產、資料蒐集、…等工作，它都可以幫忙。

計算機的應用實在是不勝罄書。讓我們以它在醫療方面的應用為例來說明吧，這個例子可視為計算機在世界上千奇百怪各種應用的縮影。

目前，用計算機處理病歷和醫療費用，已經相當普遍了。它取代了既占空間又難尋水的舊式檔案。心急的等待器官移植的病人，可以透過通訊網路很快地找到捐贈者。我們可以根據疾病發生的地區、病人的職業、年齡、或其他特殊因素來做分類統計和分析，以確定疾病流行根源。凡此種工作，都在保護你我並增加我們的福祉。

更重要的一項改變是，計算機帶來了千百倍生活品質的提升。不只使我們使用的物品有大幅量的增加，而且更有精緻的質的提升。這些提升是無形的。譬如說：汽車的速度比人跑步快了十倍左右，而飛機的速度又較汽車快了十倍左右，這些提升是有形的，我們易於察覺。但是，計算機大量使用的結果，其所帶來生活品質的提升已遠較上例為大，只緣身在此山中，不易察覺罷了。

例如，在例行的樣品化驗和分析的工作上，計算機不僅替我們做了許許多多令化驗人員感到痛苦和壓煩的工作，而且還做得更正確、更準確和更可靠。醫師們可由桌上的終端機上直接閱取需要的資料，比任何一個信差或飛毛腿送的都快。

由於計算機的使用，大大地減低了許多診斷的危險性，發展成所謂的無侵害性診斷技術（noninvasive diagnostic techniques）。在1970年代初期的電腦化軸斷層掃描設備（computerized axial tomographic scanner），它可以提供身體剖面的圖像，尤其適用於體內的器官、柔軟的組織，甚至細胞上。在此後十年間，斷層掃描設備進步得更成熟了。最先進的核磁共振掃描裝置（nuclear magnetic resonance scanners）甚至可標示單一細胞的鈉含量（sodium level），可提供細胞是否健康的重要資料。

可預見的，幾乎每一個行業，醫師、工程師或從事管理的人員，每人都將擁有一台電腦的工作站，而每個醫院都會這麼做。

在過去這麼多年，計算機一直往更便宜、體積更小和功能更多的方向前進。大略地說，每兩年其功能就增至雙倍而價格卻只有原來的一半。更可喜的是它變得越來越好用。譬如在1960年代初期發展的高階程式語言，像FORTRAN、LISP等最近已變得越來越口語化，接近常用的英語。計算機會將這些高階語言的程式，自動轉換為較易、較快的碼（Code）去執行。近日來，一些新的程式語言更像英語。學起來實在不花很多工夫。這樣發展的最終目的，最希望機器能用我們日常用的語言和我們溝通。除此以外，現在流行的營幕菜單式操作方式，也給我們許多方便。而微電子技術使得廉價的微處理機到處都用，也給我們帶來許多意外的功能和方便。

計算機最怪異的發展，恐怕要算是它涉入智慧的領域了。發展人工智慧就必須確實明白什麼是智慧。自從計算機使用之後，心理學中的認知科學有相當大的改變；包括使用的儀器、實驗的模式以及近年來的發現等等。

在智慧相關的領域中，專家系統的發展是一件重大的事。專家系統是許

多學者，如 Edward Feigenbaum, Joshua Lederberg, Bruce Buchanan 和他們的同僚共同研究奠定的基礎。一個專家系統，就是一個摹仿「專家」執行專業知識或技術的程式。它可以在這個專業的領域中，通常都是較專精狹窄的專門知識，根據已得的資料和情況作推理和判斷，並做出專家水準的建議。由於計算機有遠優於人類快速處理大量資料的能力，有時候，專家系統較人類的專家（他的老師）更勝一籌而青出於藍。

在史丹福大學的醫學中心有一個叫 **Oncocin** 的專家系統。它可以協助醫師為某些癌症病人作適量的化學療法配方。它的建議是根據大量類似病人的臨床資料統計和分析的結果，一般醫師是無法做到的。其他正在用的專家系統，還包括探礦用的，設計計算機系統用的等等。專家系統的發展是朝著「會學習的程式」這個目標前進。日本的第五代計算機計畫就有這個野心，這計畫已進行了三年，他們希望在 1990 年以前，能推出會由經驗中學習的計算機。

計算機科學的理論部分對人類的知識也有貢獻。資訊可以量化並用數學處理，此情形導致人類對於智慧行為更多的發現和了解。譬如，**Michael Rabin**

證明了在一個決策過程中，若是有大量的隨機因素，無論決策者是人或是計算機，這系統會因之而更有秩序，而不是更混亂。這是出人意表不合常識的結果。未來的計算機系統很可能會遇到這種情形。例如，許多個人計算機以網路結合在一起，它確可以單獨運作，亦可經由網路分享資料及交換情報。上述的理論將有助於此一系統的統一、控制以及維持良好的秩序。

計算機的工作是如此的活躍，新的不斷地在孕育、在發生，我們幾乎無法預言一來會是什麼情形。像它在醫療方面的應用一樣，我們只能說它會不斷地提升我們生活的素質，協助我們解決更多的和更困難的難題。對計算機而言，無論是藉由符號、或語言、或數學的處理，它一直在加速擴大我們對智慧的掌握。計算機實在是功業彪炳的翹楚！



科學月刊

第十八卷第十期

寓知於字—談中文構字中蘊含的知識

「字」由「文」生

古時候，「文」和「字」有不同的定義。東漢許慎「說文解字敘」中云：「倉頡之初作書，蓋依類象形，故謂之文；其後形聲相益，即謂之字。」由此可知，所謂「文」是指漢字在最初造字時所產生的一群最原始的基本符號，它們是中國文字最基本的元素（*primitiv-es*），亦具有「字母」的功能，可用以複合而孳生出許許多多的「字」。

「依類象形」的方式有兩種：一是象「具體」之形，謂之「象形」；一是象「抽象」之形，謂之「指事」。換言之，六書中「象形」和「指事」是造「文」的法則。當「文」緣事物之形象構形之後，再以「音」、「義」相附，便完整地擁有了文字「形、音、義」的要件。漢字中的「文」的數目不多，根據「說文解字」記載的只有489個而已。這些造字之初的文，我們有時特別名之曰「初文」以資識別。

在表一和表二中，分別舉了些象形和指事的例子。”（圖表省略）”舉這些例子的目的，是讓讀者體會體會先賢是如何從事這麼富於智慧的工作：對自然界事物詳實的觀察，精闢地體會並掌握它們的「特徵」（*features*），然後再用幾筆簡單的圖繪，將事物「特徵」傳神地轉變為文字。而這個過程和目前人工智慧中，做識別（*recognition*）類工作時的第一步驟「特徵抽取」（*feature abstraction*）的程序和要旨，是不謀而合的。若是對事物主要的特徵不能把握住，那麼識別的效果就不會很好。讀者不妨把我們的文字暫時拋開，自己造造這些字試試，是不是能造得更好？若能這樣的嘗試，當會更深切的體認先民造文字之睿智與優美。

由這些例子裡，我們也可發現近代認知科學中歸納「聯想」的四個主要法則：接近律、類似律、（圖表省略）反對律和因果律，亦在此全部呈現。在象形的例子中多合類似律，而「旦」可解釋為因果律，「叵」則為依反對律所造之形。

到了宋朝，鄭樵對「文」下了另一個定義，也說：「獨體為文，合體為字。」這個定義的精神和許慎的並無不同，都是持「字」由「文」生的觀點，然而做法和結果上卻截然不同。根據相關資料對「象類書」（該書已失傳）之描述，鄭樵

定義了870個以聲為主的「文」和330個以形為表的「文」，合計共1200「文」，是為所有「字」孳生的基礎。顯然，馮樵將重點放在字的孳乳上，才得到這樣個結果。當然，文字在字體演變過程中經歷筆劃的嬗變亦可能是導致做法不同的原因之一。很可惜的是我們已無法一睹此1200「文」的面目了。然而，「字」是怎麼產生的呢？

根據許慎的說法是「形聲相益」，其實就是六書中「形聲」和「會意」二者。會意是形與形的相益；形聲是形與聲的組合。關於文字孳乳的過程，鄭樵在「六書略」序中云：「象形為本；形不可象，則屬諸事（指事）；事不可知，則屬諸意（會意和形聲）；意不可會，則屬諸聲（形聲、轉注和假借），聲則無不諧矣。」這段話把文字孳乳綿延寫得十分明晰。

在表三和表四中，分別舉了些會意和形聲的例子。由表可知，抽象含義的字越來越多，這說明了文字孳乳由實而虛的發展方向。在文字學中對六書有四體二用之說：「象形」、「指事」、「會意」和「形聲」是造文字的法則，是體；而「轉注」和「假借」是文字運的法則，是用。「轉注」是指互訓的字，約略可視為一群近義字之間，其義「相轉相為」的關係。「假借」是指一字數用的方法，約可視為一個字兼具數音數義時，其「依聲記事」而引申其義的關係。由表三與表四觀之，在中文裡對於一些不認識的文字，可以「望文生意」是有道理的，因為中文構字之中，已經蘊含了許多知識，由此可略窺字義，而字與字之間也頗有形、音、義交互的關係可尋。

## 構字的發展

為了彌補古書失傳的遺憾和因應中文資訊處理的需要，自民國六十年以降，近人對中文構字的問題上也花費不少心血研究。民國六十二年謝清俊等人發表了496個「字根」，這組字根可組成4萬8千餘漢字。這個工作並未考量到文字的源流與孳乳，它純是在為了能根據字形檢索文字和能產生中文字形的需求下，用制式系統的技術所推導出漢字字形構成的基本元素。之後，這種做法引起了許多人的興趣，發展出許多「字根集合」（雖然所用之名稱各異）。這些理論與基礎上的探討，是今日許多新的輸入方法和設計字形產生器的基礎，同時亦引導著以標準鍵盤取代了已往大鍵盤的發展歷程。

由文字學方面著手研究這個問題的有：民國七十三年周何的「中文字根孳乳表稿」為代表。周教授依據「中文資訊交換碼」第一、二集中共2萬2千餘字，歸納出文字孳乳的「字根」共計1134個，其中「聲母」869個，「形母」265個。這個數字和失傳的鄭樵「象類書」中記載的頗為一致。

以上的討論說明了一件很重要的事情：漢字的數目雖然很多(目前已蒐集到的 8 萬個，尚不包括書法之變異)，而且字數與日俱增〔此字之集合是開放的 ( open set ) 〕，然而漢字字形構成的基本元素(無論從「初文」、「文」，或「字根」的角度來分析)，卻是有限且恆定的 ( closed set ) 。這個現象對中文資訊處理技術的發展有極其深遠的影響。

由以上的討論和舉例，我們已經瀏覽了中國文字的孳乳和構字的法則，在構字中所蘊含的知識究竟可不可以用既有的人工智慧技術來表達呢？在下文中且讓我們用些題例加以說明。

### 構字中的知識表達

由於文字和語言是依約、定、俗、成的法則演進的，文字的含義自然亦隨時間的步伐而轉變。因此，我們將以造字之初，文字孳乳和近代字書等三種資料為例，說明構字與文字間知識表達的結構。此所指的知識表達的結構是由人工智慧的角度來看的，和知識論、認知科學等所討論者不同(有心一探人工智慧中知識表達法的讀者，請參閱參考資料 7 ) 。

冰字由「冫」和「水」組成，是水凝凍狀(見表二)。冰這個字含有表示由水凍結成冰的過程的意思。這是先民對自然現象觀察的心得所造的字。水凍結成冰的現象可以用下列的語意網路表示：讀者將發現表示冰「字根式」的(1)和表示水和冰關係的語意網路(2)，實是極為相似。換言之，若將構字中含的知識結構以語意網路 ( semantic net ) 表達，則該字的字根式至少可作為它語意網路之索引，或作為該知識庫控制功能之用。圖二表示一個較複雜的字，「愛」的語意網路，或更可助於讀者了解。若是我們用制式系統中的產生系統

( production system, rule based ) 作為知識表達的工具，則可寫一條規則：若 ( if ) 水在 0 ° C 以下則 ( then ) 凝固 ( 3 ) 來表示水結冰的現象，而 ( 1 ) 式亦可為 ( 3 ) 式之索引。讀者試比較此 ( 1 ) 、 ( 2 ) 、 ( 3 ) 式間的關係與英文中 ice 和 water 之關係，自不難發現我們文字中蘊含之知識，實較西方拼音文字不可同日而語。由些例亦可知，若當我們有一如 ( 3 ) 式成的物理知識系統時，( 1 ) 式或 ( 2 ) 式仍然可以與之配合使用。

### 〔例二〕字根「缶」之孳乳字系

現在，讓我們看一個更複雜的例子。在圖三中是由「缶」孳乳出一系列的示意圖，缶之本義是上面有蓋、有頸、有把手，下端為容體的容器，是一個象形字。由於缶是瓦器，因此缶、瓦、和●是屬可通義之字。

缶有六個主要的性質，其一是●酒「漿」用，此「漿」之意被借用而孳生出「萄」字，它是「●之含漿者」之意。由於「缶」需開蓋使用，故覆以●（見表三），而「●」成為「缶」之重文。又缶可「握持」，「握持」之意借於「●」字則為持刀之意。

缶之顏色有「灰」有「黃」，借這些顏色而產生「●」字，表示色如「缶」之馬也。

又缶為瓦器，故「●」象燒缶之「窯」，因而「●」亦與「缶」通。在窯中缶多「堆積」而燒，故生「陶」字，表示色如「缶」之馬也。

又缶之形狀渾圓宏張，囊之大者如缶，故有「囊」字。再者蟻封形狀如缶，合「囊」與「虫」產生「囊」字，即通「蟻」字。

缶中空，因而擊之有「聲」，借「聲」之意而生「●」字，彭●之聲也。又因有「聲」故可擊節為「歌」，借「歌」之意而生「謠」字。為歌則需「張口舒氣」，借此意生「●●」之詞表示出氣貌。又聞歌則「喜」，借而生「●」字，喜也。又歌則●譁吵人，故生「●」字。

缶在古時甚為貴重，故「寶」字中亦有「缶」焉。

這個缶字孳乳的字系是依據杜學知「文字孳乳考」而引來的，雖然杜教授的論點在文字學界尚有爭論，即使圖三中有些關連不盡符合文字孳生之真實過程，它還是很有用的，適合人工智慧的處理。

這個孳乳之字系，可用人工智慧中的語意網路，或是框架結構（frame）來表示。除了字的意義所含的知識以外，像燒●時坯是成堆放置的、中空之容器擊之有聲、唱歌時須張口舒氣…等等知識，盡含於此結構之中。若能設計有效之譯出程序（interpretive procedure），則可在此資料結構中尋出上述之知識。

另一個相似的例子是「不」字的字系（見圖四）。因限於篇幅，不再贅述（若有不明可參考資料6）。（圖表省略）。

〔例三〕缶部的字（國語日報辭典）

根據國語日報辭典缶部共10字。其中2字，「缺」與「●」表示「狀態」或「性質」。其餘8字為器物名稱，其間關如圖五所示。圖五亦可用語意網路或框架結構來表達。

根據圖五，我們可一眼看出，此圖與圖三（同是缶）所含之知識差異甚大，在此圖中，應可推導出瓦器皆有易「缺」及易生「●」之性質，若不然此二現象（狀態）之字不會從「缶」孳生。同理，瓦器可盛酒漿和固體，亦可為樂器，此等性質可由缶部器物類之 8 字中導出。

在表五中列有中華大辭典中木字部 336 字之分類表。此分類表雖現甚為粗略，然而木材之主要功用卻不言可知。此外，根據字義，關於木製器物之中，有關建築者 39 字，有關交通者 18 字（其中 13 字與舟楫有關），此外家具 13 字、容器 12 字、文具 9 字、敲打工具 9 字、棍棒 7 字、刑具及棺具各 6 字及武器 4 字等為較大宗者。由此可推知木材之主要用途。

在木字部中有「柴」字，其意為作為燃料之木材，（圖表省略），是否由此可推知木之可燃性呢？是否可由此推知「有些家具可燃」的知識呢？當然，這些知識均已蘊含在構字和字典的部首結構中，我們人可以明瞭，是不是電腦也可以聰明地推導出這些知識呢？則是尚得努力的研究題目。

總之，由上面的例子，我們可確信中國文字中蘊含的知識成分遠較拼音的文字為多，而且它有一套系統化的架構，若能將之電腦化，則其蘊藏之知識實可組成一極佳的常識知識庫（common sense knowlege base），而常識知識庫，正是國外學者為解決其他人工智慧問題：如自然語文的處理及機器翻譯等研究工作，而急於想獲得的。

### 最佳的知識表達系統

先賢造字，肇始於對自然現象之觀察，此中涉及人體、天文、物理、化學、動物、植物等等科學領域。以現代科學的水準來觀察千百年前文字的演進，我們幾乎找不到違反科學方痕跡。這是多麼令人驚訝的事！由此可知，先賢造字均能屏除巫迷，在文字的結構裡表現出對事物真象客觀正確的體認。這種具體而微，詳盡周延的觀察功夫，深深契合科學方法的要領。對於人倫、文化與社會方面的觀察，也充分表現出持正不阿、善體天心的心本文化精神。由是觀之，亦無怪乎春秋時期有諸子百家的蔚蔚風流。在文字孳乳方面，則思路明晰，舉凡推論之邏輯、觀念之延申、聯想之法則、重點之把握，以及由實物進而引申到抽象意義的創意過程等等，精闢絕倫，無不表現出卓越的睿智和極其科學化的內涵。這種成就，發生在數千年前的上古時代，實在令人難以想像！

有些語言學者認為「音」與「義」是語言最主要的兩大要素，至於文字「形」則是次要的，它只是記錄語言的符號罷了。文盲亦可流利的使用語言，就是常

用的佐證。這種說法，在西方拼音語系中尤顯得有力，因為拼音文字的「形」與「音」的關係是直接的，而「形」與「義」的關係就比較疏遠，「形」常須經由「音」的媒介與「義」相關連。然而，我們文字「形」與「義」的關係就不盡然如此。在中國文字結構中蘊藏了極豐富的知識，這些知識和字「義」緊密的結合在一起。換言之，我們的文字，「形」、「音」、「義」三者是密切相關無分親疏之分。其實，世界上任何一種語言都有豐富的知識內涵，然而中西語言表達知識的層次卻不相同。在我們的語言裡，即使在構字的層面上，就已經把知識和語言緊密地結合在一起，更不用說構詞、構句和文章了。

由人工智慧的角度來看，我國文字的結構是一極佳的知識表達系統。雖然它是給人用的，然而它的結構卻與目前人工智慧中所了解的知識表達技術，一呼應。這是多麼令人拍案驚奇的因果關係！莊子與「得魚得筌」之語，他以「筌」喻語言，以「魚」喻語言要傳達的意念。我們今天來看看「筌」，卻有意料之外的收穫：若能善用我們文字的特質，對於計算機處理中國語文的研究，或將有獨樹一幟的發展。

#### 參考資料

- 1·謝清俊談中國文字中的知識結構民國七十六年五月三十一日於中研院史語所語言組例會之講稿
- 2·林尹 文字學概說 正中書局印行 民國六十年初版
- 3·齊佩□ 訓詁學概論 漢京公司印行 民國七十四年台初版
- 4·王初慶 中國文字結構析論 文史哲出版社印行民國七十二年再版
- 5·周何等 中文字根孳乳表稿 文建會國字整理小組印行 民國七十三年初版
- 6·杜學知 文字孳乳考 世界書局印行 民國六十八年再版
- 7· A Barr & E.A.Feigenbaum, The Handbook of Artificial Intelligence, Vol.1, William Kaufman, Inc.,1981
- 8·謝清俊等 中文字根之研究 交大學刊 民國六十二年
- 9·王申倍 Intelligence in chinese character 中文計算機學會年會論文集 June,1987. (作者保留著作權)

資訊與電腦第六期

69 年 12 月

中文資訊處理系統的選擇

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，評詁的工作做得很少，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，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者費心。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，各系統對增添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、韋氏拚音（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（中國文字多當作圖形之單體處置）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。

科學月刊  
第十七卷第六期  
教育問題的省思

哲學家金恩（S.Keen）曾批評當時美國的教育：

「自我走向校門，發現自己竟毫無熱忱；  
我雖擁有職業，卻無安身之命的目標；  
有知識卻缺乏智慧；  
有觀念卻缺乏感受；  
技能精煉豐富，但言仰貧弱；  
接受了相當的『教育』，卻喪失了認同！」  
看看人家，想想自己：我們可有這份面對現實的胸襟和勇氣？

中國人說：「活到老，學到老。」杜威（J.Dewey）說：「社會即學校。」說法不一，而寓意則同：教育之於個人，在供應每一個人自由成長和發展的能力。

所以，教育不是孤立學術追求；教育是動態的，它隨著個人的成長發展和社會的變遷而變化。

由健全的個人成長而導致和諧的社會發展，是教育的社會意義。所以，杜威又說：「教育是經驗不斷地重組或重健。」

目前的教育，在專業領域中的內容已太偏狹，而在執行上又未能啟承高素質的理想。長久以來，早已種下破壞社會和諧運行的危機。時下接踵而至丟盡中國人顏面的種種社會事件，又豈是事出無因？

通識教育是一種教育哲思的理念，不是通識課程。通識教育之論點固然是針對太專業化、太偏狹的教育而來，然而其終極目的還是在於導引個人健全的成長。

所以，要執行通識教育，就應該本此理念，檢討目前教育制度、課程流程、授課內容、教學方法、師資、教材、設備和教學環境等等，作通盤的改革。在此改革的同時，尚需對未來社會的變遷和未來每一個個人一生的成長和發展，有相當正確的體認，才能規劃出目前我們需要的通識教育。

時下教育的弊病，豈是增加幾門冠以「通識教育」名稱的課，就可以治好？

只會開退燒藥的醫生，是治不好病的！

科學月刊十六卷第九期

74年9月

由 COMPUTE 一詞談科學中的專門詞彙

compute 一詞是由拉丁語 compute 演化而來。原意是指在做算術（ arithmetic ）時有關的「思考和演算」的行為。此所謂「思考」的行為與另一較古的詞 reckon 含義相近，有將各種解決問題的方法和手段均包涵在內的意思。譬如：說理、估計、假設、分析、核對、推論…等均是。而「演算」是指計數、枚舉或算數的行為。

自從計算機（ computer ）使用以來，compute 的語意隨著計算機的功能而延伸。在這算機中，除了算術運算以外，還有非算術型的運算工作。例如：資料之儲存、擷取、傳送、轉換、程式指令之了解、執行之控制…等等。而且，以工作量而言，通常非算術型之工作遠較算術型之工作為重。因此，compute 一詞之含義自然延伸至非算術工作有關的思考和演算的行為了。至此，在計算機科學的領域中，computer 一詞已與口語中的不盡相同了。

將以上 compute 之定義用於 computer 一詞，十分妥當。將 compute 譯成「計算」，實亦佳作。「計算」一詞在我國古老的語意中，本就包含計略、謀略等非算術型工作之含義。如在韓非子一書中就有：「故父母之於子也，猶用計算之心以相待也，而況無父母之澤乎。」此外史記、漢書、孫子兵法及韓愈的作品中亦多如是用者。可惜時下一般人的理解將「計算」一詞的主位放在「算」上面，太強調演算，因而忽略了更重要的「計」的部分，以致產生語意之扭曲。更糟的是時下的辭典中找不到如前述的解釋。因此若只從字面上來看「計算」和「計算機」的譯詞就不很妥貼。對次專門詞彙，最主要的是它的定義，而不是它詞面上的字義。如果有妥貼的定義，字面上妥不妥當就並不重要了。

其實，我們學習新知時，最主要的先決條件，莫過於確實了解在這專門知識領域中，「專門詞彙」的真正含義。通常，它們都與日常口語的語意不同，且更精確和細緻。一般而言，自然詞彙是有限的，創造新詞又回於種種限制而不一定可行。因此，習慣上，當找不到妥貼的詞彙時，就找一個近義詞作為代替品。此時，必須重新定義。這種情形在翻譯時也會發生。如果不仔細了解這些科學詞彙或譯詞的原意，會經常造成失真和曲解，使所學的知識似是而非，甚至於自以為是。

上述的情形中外皆有，而我們因涉及翻譯問題而更形嚴重罷了。坦白說，這都是由於欠缺正確的觀念和態度而引起的。這也是最基本的科學素養。在科學的論著裡，首要的事就是為所用的「專門詞彙」訂定安當的定義。若是犯了這條錯誤，就會被判以極刑——被判決為缺乏科學的素質而沒有做科學研究的資格。當然，這些論著也就是毫無價值的垃圾了。

語文是活生生的，會隨我們生活的環境而逐漸轉變。在一日千里的科技領域中，專門詞彙的語意和口語的語意差別就更大、更快了。切忌望文生義，套句孫子兵法的話說：「不可不查也！」

### 定理的檢驗

一個天真無邪的小精靈看到了上期的「愛你」之謎，以為此謎實在簡單不過！立刻擬了一個信心十足的完美步驟，希望能一舉破解成功。這兒就是他的法寶：

- 第一步：將他所有的字串（此時只有一個「愛我」，分別應用四條規則來產生新字串。此時，他擁有的新字串是：「愛我你」、「愛我我」。

- 第二步：將他所有的字串（「愛我」、「愛我你」、「愛我我」），分別應用四條規則來產生新字串。此時，新字串是：「愛我我你」、「愛我你我你」、「愛我我我我」。

- 第三步：將他所有的字串（是些什麼？），分別應用四條規則來產生新字串。此時，新字串。

準此以降，每一個「愛你我」系統能產生的定理，遲早總會出現的吧！他頑皮的說：就這樣一直做下去，等到「愛你」出現了，就證明它是定理。如果它永不出現，那麼，它就不是定理。

這未免太天真些了吧？雖說神仙無歲月，不在乎如失流年，也不能沒完沒了地這樣做下去呀！所以這類型的定理檢驗方法不能算做判定程序。換言之，鑑別字串是否是定理的判定程序必須是有限的步驟、保證可停的方法。

無論如何，尷尬的小精靈沒能破解這個謎。坦白說，此謎不易解。以下，且讓我們先看另一個字串之謎，並增加一些對制式系統的了解以後，再回頭來討論。

### 「風馬牛」系統

這個謎的名字叫「風馬牛」系統，它只用了三個字—風、馬、牛。為了避免像在「愛你」之謎中無法找到判定程序的尷尬情形，我們一開始就把公理的特性界定清楚：

- 定義：任何合於「X馬風牛X風」形式的字串均為公理。此中X表示任意長度只「風」字的字串。

由此定義，公理的特性非常清楚。換言之，此定義亦表明了公理的判定程序。根據此定義，此系統有無窮多個公理。

此系統只有一條產生規則：

·規則：令 X、Y、Z 表示三個只含「風」字的字串，若「X馬Y牛」是一定理，則「X馬Y風牛Z風」亦為一定理。

例如：令 X 為「風風」，Y 為「風風風」，Z 為「風」。若「風風牛風風風牛風」為一定理。則「風風馬風風風牛風」亦為一定理。像這樣的系統，你能找出判定它定理的程序嗎？請先試試再看下文。

柳暗花明

假設你試過了，你一定發現在「風馬牛」系統中的字串裡有許多「風」字，它們被「馬」和「牛」分成三段，而「馬」和「牛」各只出現一次，並且是「馬」在「牛」前。這現象排除了許多其他組合的字串成為定理的可能，也提供了很好的線索來設計判定程序。

如果你仔細觀察，會發現在此系統中所有的字串前二串「風」字數目之和等於第三串「風」字的個數！例如：「風風馬風風牛風風風風」是一定理，因為  $2 + 2 = 4$ 。當然，這是可以證明的（此證不難，你不妨試試）。據此，此系統之判定程序已解出來了。

這答案以乎是突然冒出來的。其實，如果你用「+」取代馬，「=」取代牛，再從頭看看此「風馬牛」系統就會恍然。

比得？

在前期中，曾建議將「愛你我」系統改作「咬人狗」系統來觀察。相信你對這二種不同的表示方式會有不同的感受。在上面的例子裡，若我們做下列的轉換：

馬<=====>+  
牛<=====>=  
風<=====>1  
風風<=====>2  
風風風<=====>3  
等等

你將發現「風馬牛街系統事實上就是一個兩個整數的加法系統。我們稱這兩個系



統為同結構（ isomorphism ）的關係同結構體的觀念在科學研究中常常用到。像上面的例子，我們就用它來詮釋「風馬牛」系統的物理意義。

在日常生活中，也常用到同結構體的觀念，而將寓意延伸。譬如以兩虎相●必有一傷，來比喻一件爭奪的事件和可能的後果。又如：螳螂捕蟬，黃●在後等等均是。用口語來說，這種用法便是比方，而我們不時也見到些「比不得」的臉紅局面。

談到這兒，我們可以了解，由於同結構的關係而導致系統意義的延伸，而且此現象不僅在制式系統中有，在日常生活上也用，可是此二者之間，還是有差別的。在制式系統中（或是其他的學科裡），同結構形體的認定，必須經過嚴格證明。譬如：「風馬牛」系統與二位整數加法系統之間，其元素必須互相一一對應，而兩個系統的運作規則也必須一一對應，才能證明此二系統是同結構體。證明是同結構體後，才能將各系統所推導出的特性或是意義，延伸到另一系統上去。

可是日常生活中的比方，就沒有這麼嚴謹，而且不易定義什麼是同結構體的條件。通常，比方的條件是由較高層次對事物的「理解」而來，摘其重要的特徵，取其相似之處作為依據，因此就難免有主觀和感性的因素在內。也正如此，常導致對事務認知不同而有「比得」與「比不得」之爭執。我們稱這類型的同結構特性為高層次的，而制式系統中的為低層次，這是以人類智慧層面作參考的分法。由另一角度來看，比方是測試聰明智慧是絕佳方法，我國古代的八股舉試；選拔出的人才都是聰明絕頂的，其根據的重點之就是「比方」的能力。

### 制式系統的詮釋

在制式系統裡，以其同結構體來詮釋其意義是很重要的特性。許多制式系統的特性經由此詮釋而豁然明朗。對一個意義不太清楚的制式系統，一旦找到一個同結構體來解釋，真有如一聲霹靂，震散滿天陰霾。由另一個角度看，制式系統是一個抽象的結構，它代表著某一類型系統的共同特徵。這就是它可愛、有趣和有威力的地方。當然，像「愛你我」系統這樣的個案，若於找不到我們熟知的同結構體來做進一步的了解，也的確令人苦惱。

制式系統的詮釋是只能被動的。譬如：雖然  $2 + 2 + 2 = 6$ ；可是「風風馬風風馬風風牛風風風風風風」就不是在「風馬牛」系統中的一個定理。這說明了什麼是「被動式」的詮釋，其理不難明白。

一個制式系統也可能有許多不同的詮釋。譬如：若將「風馬牛」系統作下列

的對比：

「馬」<=====>「二」

「牛」<=====>「右列中扣除」

「風」之轉變與上例相同

也是可以的。可是其詮釋就不同了。

比不得！

談到這兒，引起一個有趣的問題：我們能不能將宇宙中所有實體的現象，都用些制式系統來表示呢？粗看之下，似乎是可以。試想，宇宙間一切實體都是由基本粒子組成，而這些粒子的行為又都膺服物理學上的「制式規則」。因此，粒子每一瞬間的行為都存有固定的因果關係，而唯一的公理就是宇宙誕生時那一瞬間的狀態。照這樣推測下來，這豈不是一個未來命運都定好了的（deterministic）宇宙嗎？該不是這樣的吧？那麼問題出在那兒？

若我們不談以上那麼大且近乎哲學的問題，就以數學來說吧。我們能用制式系統來作為某些數學系統的模式嗎？如果我們不是對這友數學百分之百了解的話，我們能這樣做嗎？如果我們研究制式系統的目的，是為了在此學科中增加些新的了解，那麼，除非我們能證明同結構體是百分之百的全等，否則我們又怎麼能證明每一個定理的詮釋都是正確的？而且，如何證明同結構體是百分之百的全等呢？均將有助於對科學的了解。

由同結構的關係而導出一個系統延伸的意義，是很重要的觀念。一般在學的朋友常常很有本領推導公式，可是一但涉及數學（或某種抽象表達法）的物理意義即望而怯步，這實在是求學上的障礙。希望本文能提供你一個新的思考方式，在你的思考的清流裡注入一束源頭活水。

科學月刊第十七卷第一期

中華民國 75 年 1 月

制式系統淺說(上)

制式系統是由 formal system 譯來的。也有人譯作：形式系統、正規系統或是正系統。在計算機系統裡，無論軟體硬體，制式系統用得非常多。它在不知不覺中影響我們思考的方式，因為許多結構模式都是製式系統的形態。譬如，各種計算機語言的結構就是典型的制式系統。若能掌握制式系統的觀念和共通性，對於計算機科學實事半功倍的效果。

本文是為從未接觸過制式系統的朋友所做的介紹。  
本文主要內容改寫自 D.R.Hofstadter 所著的：Godel, Escher, Bach: An Eternal Golden Braid 一書中的第一、二章。為了引起你的興趣，讓我們從一則點題的遊戲——「愛你」之謎開始。

### 「愛你」之謎

這是一個字串的遊戲。字串是依順序排列的一串字。在這遊戲裡，只准用三個字（這是此系統制式的規定之一），我們姑且用：「愛」、「你」、「我」。換言之在此遊戲中，所有的字串都是以「愛」「你」「我」組成的，譬如：

「愛你」

「你我愛」

「愛你你愛你你」

「你我我你愛我你你我愛你我我你」

等都是符合制式規定的字串。

現在，我們提供你一個字串：「愛我」。這是你目前擁有的唯一字串。然後，我們再提供四條規則。這些規則可以隨你選擇重複地使用，但是不可以自己發明規則（這也是制式的要求）。給你的題目是：看你能不能變出「愛你」這個字串！這四條規則是：

規則一：若你擁有一個字串的末尾一字是「我」，你可以在此字串後加一個「你」字，成為你擁有的一個新字串。

規則二：若你擁有一個字串「愛 X」，那麼你可擁有「愛 X X」的新字串。此中“X”代表任何的字串。譬如：

· 由「愛我你」可得「愛我你我你」，X = 「我你」

- 由「愛你愛」可得「愛你愛你愛」， $X = \text{「你愛」}$
- 由「愛你」可得「愛你你」， $X = \text{「你」}$

請注意：X不是此系充中的一員。它像是代數中的變數（或函數），用以表示某一字串罷了。換言之，X只是一個代表任何字串的符號，我們可藉以便利地說明規則。

規則三：若字串中含有「我我我」，則此「我我我」可為「你」取代，並成為你的新字串。譬如：

- 由「你愛我我我愛你」可得「你愛你愛你」
- 由「愛我我我我」可得「愛我你」或「愛你我」
- 由「愛我我我」可得「愛你」
- 由「我我愛我我」則什麼都得不到，因為隔中沒有連續的「我我我」。

以上這些規則都是不可逆的。例如：由「愛你」變成「愛我我我」就錯了，制式的規則就是這麼僵硬的，不可權通變達。

規則四：若字串中含有「你你」，可以把它去掉以得到新字串。譬如：

- 由「你你你」可得「你」
- 由「愛你你你我我我」可得「愛你我我我」

至此，所有的規則已介紹完畢，你可以放手一試了。最好先試試再看下文。希望你玩得愉快

。

註：為了便於閱讀，所有的字串均以「」括出。

定理、公理、規則

嘗試過了嗎？其實，找不找得到答案並不重要。重要的是你自己親身的體驗。相信你已做出了不少的字串。對制式系統已有些體會了。

為了便於稱呼，先讓我們界定些名詞。這此經由規則所產生的字串叫作「定理」（theorem）。當然，此定理與數學中的定理不同。數學中的定理是經過嚴謹的推理，並經證明為真的一些陳述。在這個系統中的定理，並未經過上述的證明過程，而是由系統中既有的規則所產生的。但其「正確性」一如數學中之定理。這個產生的過程非常機械化，因為所用的規則都是很刻板的。

在剛開始的時候，我們送給你一個字串「愛我」。這種不經產生就有的字串稱之為「公理」（axiom）。當然，它和數學中的公理也不一樣。但是二者皆毋需證明就可存在是互致的。在一個制式系統中，公理的個數並沒有限制，可

以從無到有、到無窮多個。

規則是用來變換字串形態用的。它可以產生新字串，由新字串更可帶來新的訊息。所以稱它們為產生的規則（ rules of production ）或是推理的規則（ rules of inference ）。產生新字串的過程稱為推導（ derivation ）。推導和證明實似表親關係，它們分別在制式系統和數學中扮演非常類似的角色。

下面就是個推導的例子

公理 「愛我」 ..... ( 1 )  
由 ( 1 ) 用規則二得 「愛我我」 ..... ( 2 )  
由 ( 2 ) 用規則二得 「愛我我我我」 ..... ( 3 )  
由 ( 3 ) 用規則一得 「愛我我我我你」 ..... ( 4 )  
由 ( 4 ) 用規則三得 「愛你我你」 ..... ( 5 )  
由 ( 5 ) 用規則二得 「愛你我你你我你」 ... ( 6 )  
由 ( 6 ) 用規則四得 「愛你我我你」 ..... ( 7 )

由上可知，推導是一板一眼很制式化的。在制式系統中，如何運用既有的規則才是容許自由變化之處，但也經常是問題的所在。

名詞介紹至此，前即遊戲的問題可以用這麼一句話來說：在上述的遊戲中，「愛你」是定理嗎？

圈內、圈外

許多人在剛開始試「愛你」之謎的時候，都是隨意的先試著產生幾個定理。但不久，就會注意到這些定理間有些重要的共同性質。這現象是你的智慧參與工作的表現。例如，你或午已發現所有的定理都是以「愛」作第一個字。起先，你或許對這現象還有些懷疑。此時很可能為了對付這個疑問，你會暫時放下產生「愛你」的嘗試，先去解決心中那個疑團。因此，你檢視規則，逆著定理產生的方向追蹤，直到正本清源地回到「愛我」的公理為止。至此，你相信你已證明了任何一個定理都是以「愛」為出發點的現象。於是，心中疑雲盡消，你的注意力又回到原來的問題：如何產生「愛你」上面來了。

以上的現象說明了一件非常值得注意的事：人和機器的區別。我們可以輕而易舉地寫一程式，令計算機逐一產生定理，當「愛你」出現時便停止。但是它決不會主動去觀察每個定理是什麼字起頭的。然而，即使是一個非常遲鈍的人，在一些嘗試之後，會對面臨的問題做些觀察和較深入的體認。換言之，人的智慧很不「安分」，它在做一件事的時候，經常會做到一半，就跳出了做事的圈子，在圈外看看圈內，回顧一下他做了些什麼，蒐集一些資料後，再跳回圈子裡去繼續未完的事。

在處理制式系統時，分清什麼工作是圈內的和什麼是圈外的是一件很重要的事。你可以如此試試看，在解「愛你」之謎時，準備二張紙；一張記錄圈內的工作—屬於「愛你」之謎這個系統內的，一張記錄圈外的工作—是你對此系統的觀察和心得。你將發現你會經常往返二張紙之間。這樣做的好處之一是可以更有條理地觀察和了解相關的問題，其次是使你較不易犯規。除非你是個訓練有素的研究人員，否則經常會不自覺的把制式系統以外的東西帶到制式系統中，而這是不允許的。

為了便於稱呼，我們且稱圈內工作為機械形態（ **mechanical mode** ），圈外的為智慧形態（ **interligent mode** ）。其實，我們在求知時，亦常在此二者之間往返。做習題過程就是個好例子。將圈內和圈外的事務混為一談更是出錯的泉源，比方說在雨天觀察玻璃窗，一滴水滴加上加一滴水滴還是一滴水滴，你能說這是  $1 + 1 = 1$  嗎？

#### 判定的程序

上文所談的「愛你」之謎實是一種制式系統。這種制式系統是1920年代一位美國邏輯學者波斯特（ **E.Ppst** ）發明的。故又稱波氏產生系統（ **Post production system** ）。我們且稱這「愛你」之謎為「愛你我」系統，因為系統中所有的字串全由這三個字組成。

然而，根據以上的討論，並非所有僅含「愛」「你」「我」的字串均在「愛你我」系統之中。譬如，「你」「我」這些字串就不是，因為它們不是以「愛」起頭的。那麼，「愛你」是不是呢？記得嗎？我們還沒有把這個問題解決呢！說到這兒，引起了一個基本的問題，那就是：這個制式系統中的字串究竟有什麼特性？或說：我們如何判定一個字串是或不是在這個系統中？若是我們能清楚地了解系統中字串的特性，我們就有方法依據這些特性來判別一字串是否在系統之中。反之，我們若能判斷任一字串是否在系統之中，就可依判斷的條件歸納出系統中字串的特性了。所以上面二個問題實是一個問題的二種問法。

在制式系統中，判定一字串是否是定理（在系統中）的方法和過程，就叫做此系統的「判定程序」（ **decision procedure** ）。如果一個系統有判定程序，那麼，此系統中定理的特性就可明白地刻畫出來。若否，則定理的特性將隱含有公理和規則之中。即使是簡單如「愛你我」系統；只有一個公理和簡單的四條規則，但是其定理的特性即不易明。

判定程序也可說一個「識別過程」（ **recognition process** ）。通常，制式

系統的問題可歸入二大類；一是產生（ generation ），其次就是識別。這二類問題解決的交會之處，就是認清定理特性的判定程序。

暫停

討論至此，我還是捨不得把「愛你」之謎的答案現在告訴你。真是很把歉，且留待下回分解吧！在結束之前，讓我們再做個小小的試驗：在「愛你我」系統中，請將「咬」取代「愛」，「人」取代「你」，「狗」取代「我」，也就是把「愛你我」變成「咬人狗」，而所有字串也都變成「咬」、「人」、「狗」三個字的組合。此時，再回頭重新看看本文中的系統（「咬人狗」系統），你是否感覺很滑稽呢？在一笑之餘試問「愛你」之謎和「咬人」之謎有什麼不同嗎？你願意花點時間想想這個問題嗎？