

譯翻腦電

樂秀章

林明岳 周乾康 江 鑾

著

M
A
C
H
I
N
E

T
R
A
N
S

O
N



香港中文大學電腦翻譯研究組出版

73.8793
5

電 腦 翻 譯

樂秀章

林明岳 周乾康 江 璧

著

香港中文大學電腦翻譯研究組出版

內容簡介

本書簡明扼要地講述了“電腦翻譯”的基本知識。並介紹了
中文大學電腦翻譯系統的研究過程及其應用。

2P96/13

電 腦 翻 譯

樂秀章 林明岳 周乾康 江 鑾 著

*

香港中文大學電腦翻譯研究組出版

*

1975 年 12月 第一次版

*

這本書早已在一九七二年底就寫成了。一直擱着沒有付印。當初的目的是想將我們剛完成的研究結果發表出來，給大家知道“中文大學語言翻譯器”的設計原理和應用。同時也想解答一些常被問到的有關“電腦翻譯”的問題，比如，什麼是“電腦翻譯”？為什麼和怎麼樣來進行“電腦翻譯”等等。

“中文大學語言翻譯器”雖然是在一九七二年設計成功了，同時也公開示範了它的功能，不過這只是通過翻譯十篇短的科技論文的結果。所譯得的英文，雖然是非常滿意，但是因為沒有經過嚴格和有規律的試驗，所以不能代表和保證“中文大學語言翻譯器”的實際性能和功效。

自從一九七五年一月，我們因為得到亞洲基金會的資助和鼓勵，才正式利用這翻譯器來翻譯中國科學院出版的數學學報和物理學報，借此對它作嚴格的審查、對它的潛力作更進一步的認識。

經過差不多一年不斷地應用翻譯器來翻譯科技雜誌，得到的結果出乎我們意料地美滿。目前這些譯文經已公開出版並得到歐美各大學和研究院的訂閱，同時也帶給我們更大的信心在不久將來會應用“中文大學語言翻譯器”去進行其他科學論文的翻譯。

現在我們對“中文大學語言翻譯器”具有很大的信心和對它在將來的廣泛應用于科技雜誌的翻譯工作上抱有無限的希望，因此我們覺得有把這本書出版的需要。

我們的電腦翻譯研究工作，本書的編纂和印刷，有賴亞洲基金會的資助，謹在此表示謝意。同時對陳偉慈、盧長蒂、歐鄧和平、黃敏芬、杜源浩、梁寶銅等同事的協助均表謝意。

一九七五年十月

目 錄

前 言

第一章 爲什麼要進行電腦翻譯	1
第二章 電腦是怎樣工作的	5
第三章 電腦翻譯簡史	11
第四章 香港中文大學電腦翻譯系統的簡介與特點	25
第五章 儲藏於電腦系統中的漢英辭典之結構及應用	35
第六章 輸入方法	51
第七章 電腦翻譯的技巧	65
第八章 中大電腦翻譯系統的設計及執行	82
第九章 未來的發展	90

附 錄

第一章 爲什麼要進行電腦翻譯

一 語言是人類交流思想的工具

語言的產生和發展是一個頗有趣的故事。我們的祖先最初只能以打手勢和作不同的面部表情來表達自己的思想，久而久之，由於動作之時所發出的聲音便形成了口語。人類既然可以說話，以文字來記錄口語也就是必然之結果了。

不同的民族有不同的生活習俗和思想方式，因而產生了不同的語言。古時由於交通不發達，人類的生活只能局限於一個地域之中。群居生活是古人唯一對抗大自然的生活方式。人與人之間的接觸也僅局限於所屬的個別群體之中。既然不和外界接觸，也就不會引起因語言不同而產生的種種麻煩。

科學昌隆的今天，人與人之間的“實際距離”相對於時間而言已大大地縮短。雖然國與國之間的界限仍然存在，人們之間的互相往來已是一件極平凡的事。尤其是噴射式飛機的出現，由地球的一端飛到另一端也只需十幾小時。這樣一來，由於語言的不同所引起種種的不便也就愈來愈受人注意。尤其當一個人到了另一個國土，言語不通所造成正如“鸛同鴨講”之難堪情形此時人們就會覺得語言翻譯之需要。

二 人類之互相了解可以促進和平與社會的進步繁榮

因語言之不同所引起的誤解，常被人們作為茶餘飯後之談笑資料。這種情形可以進而推廣至國家和國家間的關係。過去人們只可以通過本地報紙的通訊以及爲數甚少之譯文去了解他國居民的生活。因此當權者可以控制報業煽動人民對外作戰以達到某種政治目的。如果人民可以直接通過新聞的譯文去了解其他的國家內部的一切情況，相信大部份的戰爭是可以避免的。

社會進步至如今的地步，很多事情是需要各個國家之間的通力合作。暫勿論及聯合國的機構是否健全，它的建立便是這種需求的自然產物。阻止人口販賣、提倡世界和平、改善大自然的美化和保護等等便是實例。

三 技術交流

科學技術的不斷發展，資訊學也隨之而產生。每一個國家每年都有成千上萬篇科學論文相繼發表，但能被翻譯到其它文字的科學論文只是為數有限的一部份。這種情形我們冠之以“資訊爆炸”的現象也不至於太過火。既然不能知悉他國之科學成就，那種「閉門造車」的現象往往造成了某項研究工作不必要的重複。一個國家動員了大量的物力和人力去進行某項研究工作，幾年後才發現他們所遵循的道路正是某一國家的科學家所走過的，注定失敗的路綫。更不幸的是他們鼎力而為的研究工作在幾年前，已被別國的科學家研究成功。這種情形所引起的人力、物力上的損失還是其次，時間上的損失和技術上的停滯所帶來的損失才是無可估計的，再說現在的任何一項科學成就都再不是個人天才的發現而是群體努力的結晶。於是人們才開始認識到技術交流對推動社會科學進步所具有的偉大意義，因此科學技術論文的翻譯就變得更加迫切了。

四 統一的國際語

很久以前，曾有科學家提倡在世界各國推廣一種統一的國際語言，以致可以解決多民族多語言所造成的隔膜而形成的世界性問題。但這種呼聲很快便消失了，因為直到目前為止還沒能找到一種語言，可以適合成為國際語。他們的提議是一種理想，可惜提得不合時宜，將來是否能成功，將有待各國科學家及語言學家的共同努力。

五 翻譯人才的培養

在目前而言，翻譯是解決語言隔膜的唯一辦法。許多國家都要求他們的大學生修讀外國語言，甚至在有些國家中，小學與中學的學生也需選讀外國語言。但能同時做到讀寫兩國文字的人畢竟很少。即使一個人能說寫二種文字他也未必能成為一個好的翻譯人員。有些人能讀寫很好的英文，是由於他們能習慣用英文去構思。如果他們同時也懂得中文，但當他們應用中文時仍然以英文去思考他們想要講或寫的內容，那麼即使他們中英文的基礎都很好，仍難成為一個好的翻譯人員，其理由在於他們不能將中英二種文字融會貫通。再看一個就讀於中文學校的學生是怎樣學英文的。當

他每看見一句英文時，首先在腦海中將英文翻譯為中文，然後他才了解那句英文的意思。這種很自然的思想過程便是翻譯的難形。

要作為一個好的翻譯人員，他不但對語言本身要有較透澈的認識，而且對所要進行翻譯之專業範圍也應有相當的了解。一個好的中英文翻譯專家，如果他對數學一竅不通，如要他將一篇華羅庚之大作翻譯成英文，恐怕他就束手無策了。如此可見一個專業的翻譯人員却是三專合一的人才。例如一個中、英、數學翻譯專才就必須是一個同時通曉中文、英文和數學的專家。

培養一個具有一技之長的人，不是一朝一夕所能成功的事，那麼要培養一個三專合一的人才也就變得更困難了。即使一個翻譯專家，他所能翻譯的範圍仍然顯得十分狹窄，而且數量上仍趕不上需求。現在還沒有一種速成的方法可以大量培養翻譯人才，去適合社會的需要。

人為翻譯的作品，一般讀起來感到十分流暢，尤其是採用意譯的技術，但它的缺點是準確度較差。因此翻譯後的作品往往受譯者個人意向和思想之支配，而且譯文中用的詞彙不能趨於統一，以致使一篇同樣的作品因譯者的不同，其譯文的格式與意思也就不同了。假如再將這些譯文翻譯成原來的文字，那麼真是會變成「牛頭不對馬嘴」了。

六 電腦翻譯之可能性

電腦自一九四五年以來便被廣泛地應用於科學研究及工商業管理等方面，甚至在日常生活之中也不能缺少它。由那時開始科學家便開始研究用電腦去翻譯語言。電腦之高速，準確性及其龐大的記憶系統，再加上可儲藏程式計劃的特性奠定了電腦翻譯的可能性。

電腦翻譯是進行大量翻譯的唯一方法。它不但能將翻譯的時間大大縮短而且可以不停地工作。用這種方法翻譯出來的作品較注重原文的「真實性」而不帶有任何個人的偏見，而且所用的詞彙統一，尤其是科學術語。但電腦畢竟不是人腦，它無法感受到原作品之感情，所以對文藝性文章的翻譯仍然無能為力。即使以科學論文的翻譯來講，它也無法在質量上和文筆之流暢上與人為翻譯一決長短。

七 電腦翻譯成功之意義

如果說第一次工業革命使機器代替了人類的體力勞動，使文明社會邁

進了一大步，那麼電腦廣泛地被應用便是第二次工業革命的先聲了。由於電腦的介入使人類腦力勞動的負擔逐漸地減輕了。故它的影響是深遠和難以估計的。因此我們有必要對人與機器，人與人之間的關係作一個全面性的調整與研究。

電腦翻譯的成功不但解決了大量翻譯科學技術論文的當前任務，而且還可以使科學家對各國語言作進一步的研究，從而促進各國人民之間的互相了解和信任。進一步來說，科學家們正待研究建立一套完整兼統一的電腦資訊系統，如果這項研究是世界性的話，那麼電腦翻譯便是必經之道。

（任何一項科學研究的進行都不應太注重立時可見的實用價值。過於斤斤計較研究成果的得失，也就不會產生像今天那樣先進的文明世界。電腦翻譯對一般人而言其實用價值並不太大，但對作為一項學術研究，對加強國際間互相了解與合作，對世界繁榮與進步來講，它的意義是重大且深遠的，也是未來電腦世界必須解決的難題之一。）

第二章 電腦是怎樣工作的

(衆所周知電腦的發明被譽爲第二次工業革命的開端。第一次工業革命使機器代替了人類繁重的體力勞動，而第二次工業革命却以截然不同的姿態出現——電腦的發明減輕了人類的腦力勞動，使工作效率大爲提高。電腦最初只被應用於科學研究和統計工作上，現在它已被廣泛地應用到工商業的每一部門，甚至普及到我們日常生活中的每一個角落。)

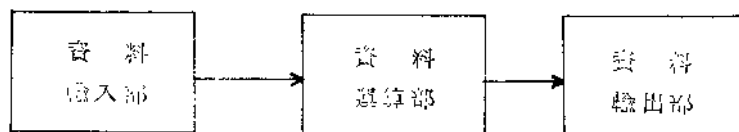
電腦的科學名稱是電子數字計算機 (Electronic Digital Computer)，由於電腦和人腦有許多類似之處，人們往往以人腦的工作過程來比喻電腦，故「電腦」這個俗名由此而得。其實人腦的結構和機能要比電腦奧妙和複雜得多。「電腦」這個俗名的命名是值得商榷的。但作者爲了應用方便起見以電子數字計算機簡稱爲電腦也無所不可。但電腦是否爲一般人所認識的那樣一個萬能的怪物呢？電腦到底是如何萬能法，它的工作原理又是怎樣呢？要回答這些問題，讓我們對這個所謂萬能的怪物進行一次剖解。

一 資料運算 (Data Processing)

所謂資料運算就是將一種原始資料加以分析，運算而得出最後的結果。運算當然包括計數，整理資料，分類以及資料形式轉換等等。

任何的資料如要對它進行運算都要經過三個過程，即資料的輸入、資料的運算和資料的輸出。如果我們能設計一副機器去履行這種運算過程的話，那麼它就是一副資料運算機 (Data Processor)。

資料運算簡圖

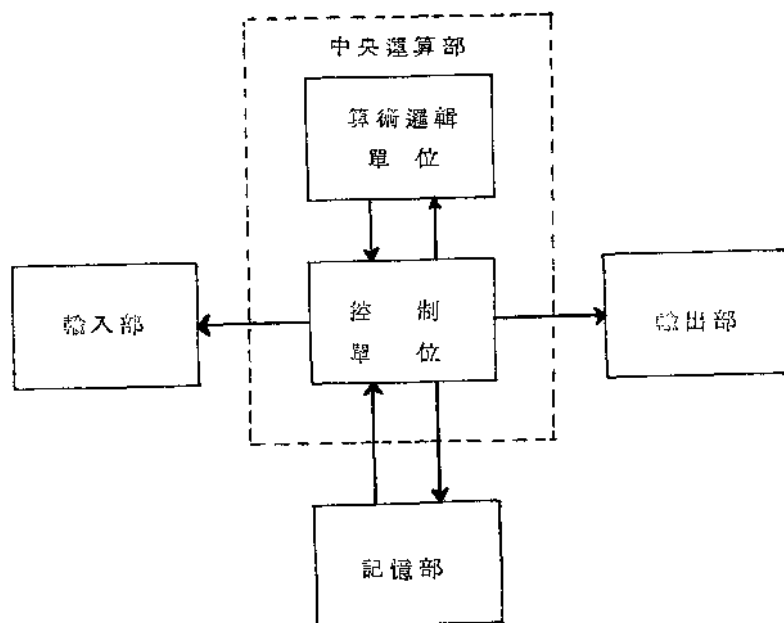


人類向來就善於運算資料，人的觸覺器官例如眼、耳、鼻、舌和皮膚就是輸入器，人腦是最複雜和敏感的運算器，而在大腦的指揮下，人類還可以用動作、表情、聲音、文字以及其他方式來輸出資料。

電腦之基本部分

電腦包括四個基本部分－輸入部 (Input)、中央運算部 (Central Processing Unit)、記憶部 (Memory) 和輸出部 (Output)。而中央運算部是由控制單位 (Control Unit) 和算術邏輯單位 (Arithmetic and Logical Unit) 所組成。

電腦之基本部份



1. 記憶部

電腦的記憶部是受電腦的中央運算部直接控制。凡是由輸入部輸入的那些需要電腦進行運算的資料或相關的程式計劃以及那些經過運算後的輸出資料都是貯藏在電腦的記憶部中。在運算過程中，電腦可以隨意取用儲藏於記憶部中的原始資料，也可以將運算得到的中間結果或最後答案再度儲藏起來，這種提取或儲存資料的時間之短，往往能配合算術及邏輯單位的運算速度。儲藏在記憶部中的資料一般都是暫時性的，當另一個新的運算開始時，舊的一個運算所需的原始資料或結果將不復在記憶部中存在，換言之它所儲藏的內容已被新的資料所取代。

電腦的記憶部是由許多記憶單位(Memory Location)所構成。每一個單位可以儲藏定量的資料，且標以號碼作為位址(Memory Address)以便尋找。這些記憶單位正如我們熟知的那種設在郵政局中的私人郵政信箱一樣，每個信箱上的信箱號碼就相當於我們所要尋找的資料的編號。電腦就是應用這種原理去取得它在運算過程中所需的資料。

電腦的記憶部所能容納的資料，一般以它所具有的記憶單位的多寡而定，換言之，記憶部的容量大小是取決於記憶單位之數目。通常電腦的主要記憶部的容量計有4K、8K、16K、32K、64K、96K和128K等等，它的容量之遞增是基於2的倍數。每一個K代表1024個記憶單位即是2的10次方， 2^{10} 。

每一個記憶單位是由若干個套在金屬綫上的磁圈(Magnetic Core)所組成，磁圈被磁化後所產生的磁場方向有賴於通過金屬綫中的電流的方向。所以電腦內部資料的儲存一般是採用二進位數字系統(Binary Number System)即[0]與[1]兩種狀態。

一般而言儲藏在主要記憶部中的運算資料應包括指示電腦工作的指令(Instruction)和所需的資料(Data)。這些資料的形式有三，一種為數字形式、二為文字形式、三為前二者的混合形式。無論這些資料原來採用什麼形式，它們在記憶部中都是用兩進位表象來代表的(Binary Representation)。

2. 中央運算部 (Central Processing Unit)

算術邏輯單位 (Arithmetic and Logic Unit) 是電腦進行簡單之算術或邏輯運算的地方。這個單位可以在控制單位的控制之下進行加、減、乘、除等算術運算，以及分析、比較等邏輯運算工作。

控制單位 (Control Unit) 是電腦的神經中樞。它由記憶部以一定的規則依次或逐步地取出指令加以分析，然後按指令所指示的內容去控制電腦其它部份的操作，例如加一個數，讀入一張資料卡或印一行字等。

3. 輸入部 (Input)

輸入部是電腦輸入資料的地方。運算所需的資料首先要變成電腦能閱讀的形式，例如打孔卡、打孔紙帶及磁帶等等，然後經過輸入器 (Input Device) 在控制單位控制之下貯藏在記憶部中。

輸入器包括讀卡機 (Card Reader)、讀紙帶機 (Tape Reader)、電傳打字機 (Teletypewriter)、磁帶機 (Magnetic Tape)、磁碟機 (Disc Unit) 等。

4. 輸出部 (Output)

輸出部是電腦將輸入的資料經過運算之後所得到的結果予以輸出的地方，它是電腦內部與外界通訊的媒介物。

如果輸出的資料不準備繼續作為輸入資料應用的話，則通常可用行印機 (Line Printer) 印出。行印機和普通的印刷機頗相似，它只是在電腦的控制之下運行印製。行印機每行可印的字數，一般都在 96 到 160 個字母之間。

有時通過電腦所運算的結果往往需要歸入檔案以作為永久性的參考資料，或者這些結果仍要被用作再次運算的輸入資料。紙卡、紙帶、磁帶及磁碟都是保存資料的最佳媒介體，它們不但能用作輸出媒介體，而且還能作為輸入媒介體。一般如要將輸出的資料保存

在輸出媒介體上必須要通過卡片打孔機(Card Punch)、紙帶打孔機(Paper Tape Punch)、磁帶機或磁碟機等輸出單位，並在特定設計的程式計劃控制下完成儲存資料的工作。上述那些媒介體往往被稱為電腦的外圍記憶單位或輔助儲藏體(Auxiliary Storage)。所謂「輔助」之意在於輔助及彌補電腦內記憶部容量之不足而已，其主要缺點是由於它的儲存速度不及記憶部那麼快。其中尤以卡片打孔機及紙帶打孔機的運轉速度較慢以致影響儲存資料的效率。

磁碟(Magnetic Disc)是一種類似唱片那樣的圓盤形碟式儲存體，每一圓碟的表面塗上磁性物質，可分為很多條同心圓形的軌跡(Track)，每一條軌跡又被分為許多部份(Section)，每一條圓形軌跡及弧形部份都編有號碼以便於尋找取存資料的位址。在磁碟高速的運轉之下通過讀寫磁頭(Read-Write Head)的運動才能取存所需的資料。一般電腦所用的磁碟是由一片或多至數片的圓碟疊合而成，故其儲存量要比電腦的記憶部為多，由於它的取存資料的方式是採用任意取存式(Random Access)，且能節省大量取存資料的時間。故在一般運算過程中磁碟常被用作輔助儲存體。

三 電腦的程式設計(Computer Programming)及程式計劃(Program)

電腦工程師在設計一副電腦之時，只賦予這副電腦可以根據一定的運算法則去做些簡單的算術運算如加、減、乘、除及邏輯運算。如要電腦工作，首先要給它一套指令去指示它怎樣做，這些指令的寫法也要根據一定的規則。

電腦程式師(Computer Programmer)根據特定的法則編撰出一套完整的指使電腦進行一系列運算的指令，這一項工作過程可稱之為電腦的程式設計，而這一套完整的指令被稱為電腦的程式計劃(Computer Program)。電腦的程式計劃語(Computer Programming Language)是由一系列特定的規則及專用的符號所組成的語言。

四 電腦是否萬能

1. 電腦的特性

高速度—電腦運算之速度是驚人的，現代的電腦每秒鐘可進行算術運算例如加法達近一億次，而整個電腦系統的工作速度正在不斷的改進中。

準確性—如果電腦程式寫得正確無誤的話，電腦的運算一定準確，不會舉一反三。

可靠性—貯藏在電腦系統的資料不會失去。一個好的電腦系統通常有很安全的保護措施。

服從性—電腦對指令的執行是絕對服從的，決無人類的反叛性。正由於如此，電腦也沒有人類所具有的幻想力和創造力。因此運算的結果正確與否將取決於〔人〕。

2. 電腦的功用

根據上述的提示所知，電腦不是萬能的，電腦本身只會做簡單的運算工作。由於電腦的特性，任何複雜的工作例如科學研究中的數學問題，工商業所需的資料處理問題以及自動控制生產問題等等，只要它們可以簡化為一系列的簡單運算，都可以通過電腦獲得迅速而準確地解決。

在用電腦去解決某一類難題之前，我們首先要對這類難題的特性進行研究，找出它們的規律性從而得出解決這類難題的運算數法。再將這個運算數法寫成程式計劃。

那些沒有一定規律式的問題或重複性不多的問題，或者是有一定的規律，但例外的情形却為數不少的問題，是很難用電腦來解決的。但却可用電腦來幫助解決。例如我們可以用電腦來迅速地取存資料，預測市場遠景以作出公司的經營方案等。

電腦按照程式計劃所得到的結果在形式上是千篇一律，沒有創造的能力，也沒有感情可言。因此我們就無法用電腦寫文學作品，也無法利用它來表達一個人的感情。我們可以說電腦是一個絕對服從命令及高速反應的「大笨蛋」。

第三章 電腦翻譯簡史

一 電腦翻譯簡史

電腦翻譯至今已有廿多年的歷史。爲了使讀者對電腦翻譯發展史有一較系統的瞭解，現將致力於電腦翻譯的各國科學家的姓名、所研究的課題及其所屬研究機構之名稱及所在地按年份之先後次序，排列於表(-)中。

電腦翻譯的發展約可分爲五個階段：

-1945 美國科學家雲紐曼(Von Neumann)設計了第一架可儲程式計劃之電腦，而使電腦的應用範圍進一步推廣至工商業方面。在這以前，科學家一直都在研究設計一種能自動翻譯人類語言之機器。

1947-1952 隨着電腦之發明及其愈來愈廣泛之應用，各國科學家提議用電腦翻譯語言，並且開始進行簡單而有限度的試驗與研究工作。

1953-1960 電腦翻譯的初步嘗試獲得了成功，使科學家充滿了信心去展開「高速優質全自動翻譯系統」之研究。

這段時期可稱爲電腦翻譯之全盛時期。該項研究得到了各國政府之大力支持。尤其在英、美及蘇聯各國，電腦翻譯研究中心有如雨後春筍般地相繼建立起來。到一九六二年爲止，其數目已達到四十八個之多。由該項研究之發展速度及受重視的程度，可見電腦翻譯在政治、經濟及科技上所具有的意義是何等的深遠和巨大。

1960-1967 「高速優質全自動翻譯系統」之研究終於遭到了失敗的命運。科學家開始探求及分析是次研究之失敗原因，是在於：對人類語言的特性瞭解不夠，即人類語言在語法結構上之不嚴謹，以及在語義上之多義性及曖昧性；其次

是對人類語言缺乏系統的分析與研究。因此科學家認為：
 [電腦翻譯的研究能否獲得成功，其關鍵在於建立一門統一的較科學化的語言學]。於是科學家開始分道揚鑣，分別進行電腦輔助人翻譯，句法分析及語義分析等等之研究。

1967—目前 科學家繼續向語言學之荒漠地進軍。計算語言學、語法結構學及語義學等科學化的語言學說也應時而建立起來。其中最突出的是以[電腦輔助人翻譯系統]取代了[高速校質全自動翻譯系統]，前者是在電腦輔助之下，使科技性論文之翻譯工業化，成為可能。我們希望在不久的將來，[科學文獻的翻譯]將不再被視為翻譯家個人的藝術結晶品。

表(一) 電腦翻譯年誌表

年 份	主持人	最初之主要研究課題	機構或國家
1933	托揚斯基 (P.P. Troyansky)	以翻譯機之設計圖案而發 作者證書	蘇 聯
1945	雲紐曼 (Von Neumann)	設計第一架商用電腦	美國賓夕凡尼亞大學 (Univ. of Pennsylvania)
1947	安德魯·布夫 (Andrew D. Booth)	電腦翻譯首倡者	倫敦波比學院 (Bribeck College)
1949	大衛·哈士 (David Hays)	支持各項電腦翻譯之 研究	美國蘭特公司 (The Rand Corporation)