

●有了這套實用的語言系統
連電腦都成為翻譯大師

●總有一天您也可以像李麥克一樣
用自然語言和您的電腦伙伴交談



●用OCR掃描輸入英日原文
便可像影印機把中文輸出
有照片為證不再是癡人夢話



●國內英中翻譯已近成熟
日中系統急起直追

電腦自動翻譯

來龍
去脈

作者／長尾 真
譯者／楊德 輝

國內唯一介紹有關電腦翻譯原理的書

電腦自動翻譯 來龍去脈

原作者／長尾 真

譯 者／楊德輝

地址／台北市臨沂街69巷17弄4號3樓

電話／321-6092

郵撥／1009944-4 戶名／楊德輝

印製／漢陞印刷有限公司

地址／台北市潮州街91之9號5樓

總經銷／吳氏圖書有限公司

電話／303-4150

郵撥／0798349-5吳氏圖書有限公司

郵政信箱／台北郵政信箱30-272號

初版／民國79年 5月

定價／160元

○如有缺頁・破損・倒裝，請寄回總經銷調換

○作者自行出版・依法免辦出版登記

○劃撥一律9折

推薦序

這是一本關於機械翻譯的入門書，但不是一本普通的入門書，而是由當今「機械翻譯」的權威專家日本京都大學長尾真教授所著，不但適合一般讀者閱讀，即使搞「機械翻譯」這一行的也值得仔細讀幾遍。

「機械翻譯」已有四十年左右的歷史，歐美、日本等國歷年投入的研究經費及人力不知有多少。可是，就個人目前所見的系統而言，仍難達到「實用化」。根據過去三年在電子工業研究所從事「機械翻譯」研發工作的經驗，個人認為主要原因有三：

一、人類自然語言的現象非常複雜，要讓電腦能適度地了解自然語言，而進一步做翻譯工作，決不是幾百條文法分析規則就能辦得到。

二、無論是翻譯過程的那個階段都必須藉助好的詞典才能使翻譯品質提升，而一本詞典至少要五萬個詞才夠派上用場，每個詞必須記載許多關於這個語法和語意的資料。

三、要讓電腦了解並翻譯自然語言還必須將現實世界的知識及翻譯技巧以一種能讓電腦使用的表現方式寫成規則，而這類的知識又是一個無底洞。

話雖如此，「機械翻譯」實用化的夢想並非絕然無望。近幾年來，日本各大電子公司

紛紛推出自己的機械翻譯系統，有名的廠牌就不只十種，可見前途還是有的。

近一、二年國內市面上有出現一種手掌大小的「翻譯機」，主要功能不外乎：一、輸入英文查中文解釋，二、輸入中文查英文，三、固定句子的翻譯。其實，這種產品應該叫「詞典機」比較妥當。真正翻譯機必須做到翻譯任何文章，而且譯文還可以讓人看懂。這一點希望讀者不要混淆。

讀者看完這本書或許會問：既然「機械翻譯」還有一些困難尚未解決，為何有那麼多日本企業投入研發工作，為何長尾真教授從一九六三年起一直從事「機械翻譯」研究至今？或許今日日本已儼然成為「科技大國」，而我們卻還喊著「引進國外科技」——也就是七十年前的「賽先生」——的原因之一吧。

楊德輝先生從事工業日文的翻譯及教學工作多年，並著有多種日文參考書。於翻譯本書時曾與本人及同事探討書的內容，這種經過消化吸收後才翻譯的譯文，更適合一般讀者閱讀。

李炳煌

民國七十八年十二月四日

於工業研究院電子工業研究所

目錄

作者原序·····	1
-----------	---

1 機械翻譯的啟蒙期·····	7
-----------------	---

史潑托克衛星的震撼／日本的機械翻譯發展概況／句子分析的困難性／ALPAC報告書的打擊／反省的時代／專門術語的作成／T A U M M E T E O系統成功鼓舞士氣／再度回頭做基礎研究

2 重新受到肯定的機械翻譯·····	33
--------------------	----

再出發／論文標題的翻譯系統／日本的現況／外國的動態

3	機械翻譯的立足點.....	47
---	---------------	----

翻譯的真諦／翻譯原文的種類／事實・說話者・環境／機械翻譯的文章

4	機械翻譯的作業程序.....	63
---	----------------	----

要素合成原理／句子的分析與翻譯／句子結構分析的問題點
／利用語意來解決曖昧性／語意在句法分析中扮演的角色

5	句子構造分析的方法.....	89
---	----------------	----

詞組結構文法／格變文法／句子的分析步驟／句中句

6	譯詞的選擇和譯句的生成.....	113
---	------------------	-----

中間表現／譯詞的選詞／譯句的生成／日語和英語——文體

的差異

7	機械翻譯的系統.....	131
---	--------------	-----

實際的步驟／軟體系統／辭典／前修改與後修改

8	機械翻譯系統未來的展望.....	147
---	------------------	-----

朝更高級的翻譯發展／語言的理解／實用系統的未來發展——

——(1) 大量高速翻譯 (2) 少量個人翻譯 (3) 應用系統

／必須經過長期基礎研究的努力

作者原序

在二十年前，如果有人談論到「從工程的角度來處理語言」，保證會被人譏為癡人說夢話，絕對行不通的幻想。曾幾何時，隨著現代資訊社會，資訊化時代這些名詞都漸成事實，甚至成為人們每天生活工作的必備工具，這時人們才漸漸認清到，其實支撐現代尖端資訊科技時代整個系統的最大幕後功臣，原來就是語言的工程技術本身。然而，語言工程方面的研究，將會面臨其他研究對象所從未有過的許多不同性質的困難。因為傳統的科學，工程之研究對象都離不開物理現象，其理論基礎不外乎是數學、物理、化學而已。

但這些基礎科學卻沒有一樣能夠在語言的研究上派上用場。語言的造句方法或句子構造的最基本部份，或許勉強強強可以利用所謂的數學基礎理論，自動化裝置理論，形式語言理論，符號邏輯學等理論來加以說明。但這也僅僅能夠對整個語言的基本部份做極少部份的說明有幫助而已，但對於語言所擁有的豐富內容（這或許才是語言的真正本質吧）目前都沒有任何自然科學的理論或方法可以解釋得清楚，實在悲哀。

語言的工程系統研究是個極高難度的題目。為何這麼說呢？原因是，語言是人類頭腦活動的外在表現，要把語言的原理、結構搞清楚，簡直就等於和在研究頭腦的活動本身是一體

兩面的事。正因為認識到這點，最近對於頭腦活動的原理構造之研究相當蓬勃，除了由醫學、神經生理學這些領域來做腦部研究外，心理學、認知科學、資訊科學這些原本以為不太相關的領域也積極地參與研究。但恐怕還得花數十年，人類才能解開自己頭腦內部活動的奧秘吧。

但，難道說，在这一切都未明朗化之前，語言的工程學系統就沒有成功的希望了嗎？答案是否定的。你看，目前人類對鳥類如何會飛的原理都還沒完全搞懂，可是超音噴射飛機、空中巴士還不是照樣已在天空飛來飛去，當仁不讓了。可見語言的基本架構原理，雖然還是含糊不清，可是我們照樣可以在部份已經懂的範圍，逐步地把語言製作成工程系統，應該沒有人反對吧。

在這個共識下，首先誕生的就是利用機器（電腦）將某種語言翻譯成另一種語言的機械翻譯系統，或人稱「自動翻譯系統」這項產品。機械翻譯系統的歷史雖然可以遠溯到電腦才剛發明不久的一九四〇年代，但它的生涯卻極坎坷，經過漫長歲月的研究，到最近才好不容易有點頭緒，有產品正式上市。另一方面結合聲音辨識、合成的技術，發明一種自動翻譯電話系統的構想，目前還正開始進行基礎的檢討。

語言在工程方面應用系統的第二個方向，就是開發一種人機之間的交談系統，透過語言在人類和電腦資訊系統之間溝通交談，好像霹靂遊俠李麥克一樣，把人類的意志傳遞給機械，請它按照我們的意思來做事，而且還會像朋友般地回答有意義的話給我們。要實現這個夢

想，勢必要讓無生命的機械也能夠像人類聽得懂一句話似地「聽懂人類的語言」，這也是目前研究人員迫切想破解其間奧妙原理的研究主題，雖然難，但卻充滿了樂趣的大題目。

除此之外，還有許多語言的應用系統值得一試。無論如何，在未來的資訊社會中，凡是需要做人機溝通的所有設備上，「語言」勢必非參一腳不可。從這層意義來看，建立一套工程系統來處理語言的技巧，將是們今後重要的研究主題。

與其說語言可以用一般性法則來說明，因為其例外現象很多，倒不如說「龐大的例外集合加起來構成語言」更貼切些。這也就是為什麼每本辭典都這麼厚的最主要原因。站在做學問的立場，當然必須由辭典或語言龐大的內容中尋找一般性的規則，但如果你是站在要設計出一套具體實用系統的立場，那就不是光靠原理結構就可以滿足的。這也就是為什麼我們既不能把所有語言現象全都當成例外來一一例舉處理，而需花費精力辛苦地另想辦法去建立介於兩者之間（既有規則，又有例外）某種模式的體系。

在學習英文的課本中，經常會舉出一個實例，告訴日本人不可以把「あなたのオフィスに明日行きます（我明天要去你的辦公室）」用「I will go to your office tomorrow」這句英文來表達，而表須說成「I will come to...（我將到...）」。許多這類微妙差異現象，目前的機械仍無能為力。對機械翻譯來說，一方面要儘量涵蓋許多例外的現象，另一方面又要顧及一般通論規則來設計語言系統的架構，因此在這兩難下，目前階段如果說無法處理語言使用上較深奧高級的部份，也是情有可原的。

即使像人那麼有智慧，對這類語言細膩部份，也都還經常會犯錯。儘管如此，這些人批評起來可不留情，有人說「機械翻譯系統會譯出古里古怪的文章，所以毫無用處」，另外也有人說「像翻譯這麼高級，只有人才會做的工作，交給機械，它當然不會」，人實在有趣，待「機」嚴，律已寬。至於像核融合的研究，即使過了二十一世紀中期都還無法確定是否能夠成功的研究，他們卻一點都不抱持否定態度，難道他們一直心裡面認定語言這種只有人類才應該會的天賦，現在居然冷冰冰的機械也敢來搶飯碗，而憤怒不講理地排斥嗎？

不管怎樣，語言在自然科學及工程方面的研究才剛剛起步，研究者的數目和經費比起像核融合這類龐大的科學計劃，微少得不成比例。因此，儘管今天它還沒有堪令人滿意的產品問世，但卻不能小看它，認定它永遠沒有揚眉吐氣的一天。只要經年累月地展開正式的研究，我們有絕對的信心，可以製造出相當水準的產品來。

針對「機械翻譯」這個題目而編寫的書籍，即使把全世界加起來，恐怕也是鳳毛麟角，數目極少。機械翻譯是屬於包羅種種學問要素的綜合技術，絕非根據一種明確的原理就可以說明清楚的，因此要編寫一本淺顯易懂的書十分困難。這恐怕也是為什麼這種書市面上根本買不到的一個理由。在這萬難之中，雖然這本書已儘量寫得能讓一般人士都看得懂，儘量避開技術上太過複雜的部分，設法努力說明清楚機械翻譯的過程是如何進行地，但還是很難拿捏得恰到好處。

另一方面，我們也介紹了機械翻譯系統開發的狀況，運用系統時應有的觀念，但要把豐

富的內容塞進有限的篇幅中，對於機械翻譯系統如此龐大的內容，而且又是由許多研究者耗費多年努力才一點一滴建立起來的系統，想要寫得讓人很輕鬆地就能掌握全貌，實在需要高度的安排規劃。如果想要更詳細更正確地瞭解機械翻譯系統，必須另行參考更專門的書。本書中用來做實例講解的機械翻譯範例結果，許多都是由日本目前正在推展的機械翻譯計劃成果摘錄出來的。本書如果能夠提供些微的助益給凡是對語言工程應用有興趣的同好人士，就是我們最大的榮幸。

當本書初稿完成的一九八五年秋天，一向執世界機械翻譯牛耳的法國 Grenoble 大學教授 B・包可華先生不幸以五十六歲英年早逝。長期擔任計算機語言學國際委員會委員長的他，過去招聘許多日本機械翻譯的研究者到 Grenoble 大學參加研究計劃。對我而言，失去一位友好的競爭對手，深感遺憾。在此謹以此書奉上 B・包可華教授靈前悼念故友。

對於鼓勵我撰寫本書的岩波書局小林茂樹先生，以及編輯出版本書費心盡力的片山宏海先生，謹申謝意。

一九八六年一月 著者

長尾 真

1

機械翻譯的啟蒙期

史潑托克衛星的震撼

日本的機械翻譯發展概況

句子分析的困難性

ALPAC報告書的致命一擊

反省的時代

專門術語的作成

TAUMMETEO系統成功鼓舞士氣

再度回頭做基礎研究

史潑托克衛星的震憾

在我身邊有一份報告書。這是日本機械翻譯研究開拓者和田弘老師，數年前在整理他書房的書籍時，送給作者的資料。和田先生擔任通產省工業技術院電氣試驗所（目前的電子技術綜合研究所）電子部長，為日本的電腦及其應用的發展貢獻畢生心力，他同時也是一九五〇年代後半開始推展的日本機械翻譯研究計劃創始人之一。

這本報告書，是由美國眾議院科學太空委員會機械翻譯研究調查特別委員會向眾議院主席提出，於一九六〇年六月二十八日眾院第八十六議會所印發的審議資料。我大致翻閱一下，裡頭內容提到了由W·韋伯在一九四六年首度提出機械翻譯的可行性，以及從那時起至一九六〇年為止，這段期間美國發展機械翻譯研究的歷史，其研究資金的來源，美國在一九六〇年當時研究的狀況，還有包括日本在內的其他國家研究概況，都有詳細的分析和說明。最後更進一步地向議會提出範圍涉及十三個項目的結論和建議。

作者踏入機械翻譯的世界，是在研究所畢業後的一九六三年那年。在由和田先生手中獲得這篇報告書之前，我甚至完全不知道有這些典故，雖然在當時已經有兩三本專門在談機械翻譯的書出版了。透過這些資料和書籍，我從一九五〇年代機械翻譯的歷史學到許多事情，而由一九六〇年代迄今，一直從事這方面的研究。事隔多年，今天再度重新閱讀這篇報告書