

**MACHINE TRANSLATION, PAST, PRESENT AND FUTURE\***

**BY**

**PROFESSOR LOH SHIU-CHANG, PH.D., FBCS, FIP, FRSA,  
THE CHINESE UNIVERSITY OF HONG KONG**

---

Important note

The attached paper "MACHINE TRANSLATION, PAST, PRESENT AND FUTURE" is a direct translation in English by computer of a paper originally written in Chinese. The English translation is reproduced here, without any alterations, from the computer print-out.

---

**EXPERT GROUP MEETING**

**ECONOMIC AND SOCIAL COMMISSION FOR ASIA AND THE PACIFIC**

**BANGKOK**

**8-12 DECEMBER 1975.**

\* This background paper has been prepared by Professor LOH SHIU-CHANG, PH.D., FBCS, FIP, FRSA, Chinese University of Hong Kong. The views expressed in this paper, which is presented in its original form, are those of the author and do not necessarily reflect the views of the ESCAP secretariat or of the United Nations.

### THE NEED FOR MACHINE TRANSLATION

THE RATE AT WHICH MAN HAS CONTINUOUSLY BEEN ACCUMULATING USEFUL KNOWLEDGE ABOUT HIMSELF AND THE ENVIRONMENT HAS ACCELERATED GREATLY, PARTICULARLY DURING THE LAST 100 YEARS. THERE ARE MANY REASONS FOR THIS INCREASE : THERE ARE MORE SCIENTISTS TO-DAY THAN EVER BEFORE, SCIENTIFIC RESEARCH IS RECEIVING SUBSTANTIAL GRANTS AND FINANCIAL SUPPORT FROM GOVERNMENT AND INDUSTRY, THE COLD WAR HAS RESULTED IN A COMPETITIVE ATMOSPHERE IN WHICH NATIONS ARE ENDEAVOURING FOR MILITARY AND TECHNOLOGICAL SUPREMACY, ETC.

BY THE MOST CONSERVATIVE ESTIMATES, EUROPE WAS PUBLISHING 1000 NEW BOOKS PER YEAR BEFORE 1500. BY 1950, FOUR AND A HALF CENTURIES LATER, THE RATE HAS ACCELERATED SO SHARPLY THAT EUROPE WAS PUBLISHING 120,000 NEW BOOKS PER YEAR. BY THE MID-SEVENTIES, THE PUBLICATION OF BOOKS ON A WORLD SCALE, EUROPE INCLUDED, REACHED THE FIGURE OF APPROXIMATELY 1,500 NEW BOOKS PER DAY.

THERE ARE MORE SCIENTIFIC SOCIETIES IN EXISTENCE TODAY THAN EVER BEFORE, AND THESE SOCIETIES ARE HOLDING MORE MEETINGS, CONVENTIONS AND SEMINARS, AND PUBLISH MORE JOURNALS AND PROCEEDINGS THAN EVER BEFORE. THE FIRST TWO SCIENTIFIC JOURNALS APPEARED 300 YEARS AGO. IN 1958, THE NUMBER OF RESPONSIBLE TECHNICAL JOURNALS APPROACHED 100,000: THERE ARE NOW FOUR MILLION JOURNAL ARTICLES, 120,000 TECHNICAL BOOKS AND 100,000 TECHNICAL REPORTS BEING PUBLISHED ANNUALLY. (ON A WORLDWIDE BASIS APPROXIMATELY 60,000,000 PAGES EVERY YEARS) AND THIS RATE OF PUBLICATIONS IS INCREASING EVERY YEAR. TO-DAY, FOR EXAMPLE, THE NUMBER OF SCIENTIFIC AND TECHNOLOGICAL JOURNALS AND ARTICLES IS DOUBLING, LIKE INDUSTRIAL PRODUCTION IN THE INDUSTRIALISED COUNTRIES, ABOUT EVERY FIFTEEN YEARS.

HOWEVER, WE CAN ARGUE THAT EVERY BOOK IS A GAIN FOR THE ADVANCEMENT OF KNOWLEDGE. NEVERTHELESS, IT IS TRUE THAT THE ACCELERATION RATE OF BOOK PUBLICATION IS, IN FACT, PARALLEL TO THE RATE AT WHICH MAN DISCOVERS NEW KNOWLEDGE.

THE RESULT OF ALL THESE PUBLICATION ACTIVITIES IS THAT THE INDIVIDUAL SCIENTIST IS STRUGGLING IN AN OCEAN OF INFORMATION AND IS ENDEAVOURING TO SEEK THE INFORMATION HE NEEDS. FURTHERMORE, ALL JOURNALS AND BOOKS ARE PUBLISHED IN DIFFERENT NATIONAL LANGUAGES, BECAUSE NO ONE COUNTRY CAN MONOPOLIZE THE DEVELOPMENT OF SCIENCE AND TECHNOLOGY. THEN, HOW CAN THE SCIENTIST OF ONE COUNTRY BE FAMILIAR WITH THE SCIENTIFIC AND TECHNOLOGICAL DEVELOPMENTS AND RESEARCH FINDINGS TAKING PLACE IN OTHER COUNTRIES ? SHOULD WE INSIST THAT THE SCIENTIST BE ABLE TO READ ONE, TWO, OR MORE FOREIGN LANGUAGES, SO HE CAN READ THE CURRENT LITERATURE RELATING TO HIS OWN FIELD, TO SEARCH THE LITERATURE ACCUMULATED IN THE PAST, AND TO BE MADE AWARE OF WORK RELATED TO HIS FIELD SO THAT HE CAN PROCEED CONFIDENTLY IN HIS WORK TO AVOID UNINTENTIONAL DUPLICATION.

THE SO-CALLED " INFORMATION EXPLOSION " HAS ARRIVED. SUITABLE WAYS MUST BE ESTABLISHED TO ASSIST THE SCIENTIST AND RESEARCH WORKER TO READ CURRENT LITERATURE IN THE FIELD RELATING TO HIS WORK, TO SHARE KNOWLEDGE AMONG PEOPLE AND AMONG NATIONS. I BELIEVE THAT MACHINE TRANSLATION OFFERS THE ONLY REALISTIC SOLUTION TO THE PROBLEM WE ARE FACING. A GREAT DEAL OF TIME AND RESOURCES COULD BE SAVED IF DUPLICATION OF RESEARCH EFFORT WERE TO BE ELIMINATED, OR AT LEAST REDUCED TO A MINIMUM BY MUTUAL REPORTING OF PROGRESS AND EXCHANGING IDEAS. THE WORK OF SCIENTISTS SPEAKING DIFFERENT LANGUAGES CAN BE ACCELERATED BY LEARNING WHAT SCIENTISTS IN OTHER COUNTRIES ARE DOING, THEIR ACHIEVEMENTS AND FAILURES, AND THEY CAN PLAN A BETTER COURSE OF ACTION TO SOLVE THEIR OWN PROBLEMS. THE

ONLY MEANS BY WHICH THESE SCIENTISTS CAN COMMUNICATE WITH EACH OTHER IS BY PUBLISHING PAPERS OUTLINING THE METHOD USED AND THE RESULTS OBTAINED, AND THESE PAPERS HAVE TO BE TRANSLATED BEFORE THEY BECOME UNDERSTANDABLE TO THE OTHER SCIENTISTS. AUTOMATIC TRANSLATION IS DEFINITELY NEEDED SINCE HUMAN TRANSLATORS CANNOT POSSIBLY HANDLE THE VOLUME OF WORK OR EVEN HOPE TO ACQUIRE THE SPECIAL VOCABULARY TO MAKE GOOD TRANSLATIONS IN A WIDE RANGE OF TECHNICAL SUBJECTS.

AUTOMATIC TRANSLATION CAN NOT BE PERFECT. WHETHER IT COULD EVEN BE HIGH QUALITY OR NOT IS DEPENDENT ON HOW HIGH THE STANDARDS ARE SET. THE IMMEDIATE GOAL IS NOT TO DESIGN A PERFECT AUTOMATIC TRANSLATION SYSTEM OR TO ACHIEVE HIGH QUALITY MACHINE TRANSLATION, BUT TO DESIGN A MACHINE TRANSLATION SYSTEM THAT IS BETTER AND MORE EFFICIENT THAN THE ONES WE HAVE TO-DAY.

**PAST**

THE SUGGESTION THAT MODERN COMPUTING MACHINES COULD BE USED FOR TRANSLATION ORIGINATED FROM A.D. BOOTH. IN 1946, VARIOUS NEW USES FOR AUTOMATIC DIGITAL COMPUTERS WERE BEING CONSIDERED. THESE APPLICATIONS RANGED FROM THE MORE OBVIOUS APPLICATIONS TO PROBLEMS IN MATHEMATICS AND PHYSICS, TO PHILOSOPHICAL PROBLEMS SUCH AS MECHANIZATION OF HUMAN THOUGHT PROCESSES, THE PLAYING OF GAMES AND THE TRANSLATION OF LANGUAGE. AT THIS EARLY STAGE OF AUTOMATIC TRANSLATION, MOST OF THE RESEARCH WORK DONE IN AMERICA WAS FOCUSED ON GERMAN INTO ENGLISH, WHILE IN RUSSIA THEY FOCUSED ON GERMAN INTO RUSSIAN. THIS WAS PROBABLY BECAUSE GERMANY HAD DEVELOPED VERY ADVANCED TECHNOLOGY DURING WORLD WAR II. HOWEVER, ATTENTION WAS SOON DIRECTED TOWARDS RUSSIAN-ENGLISH TRANSLATION IN AMERICA. WITH REGARDS TO THE AUTOMATIC TRANSLATION OF CHINESE, RUSSIA HAD SPENT A LOT OF TIME AND RESOURCES IN THIS RESPECT BEFORE THE AMERICAN REALIZED ITS IMPORTANCE.

GENERALLY SPEAKING, MACHINE TRANSLATION HAS PASSED THROUGH THREE PHASES OF DEVELOPMENT. THE FIRST, OR HEURISTIC PERIOD, BEGAN WITH BOOTH'S SUGGESTION IN 1946. DURING THIS PERIOD, BOOTH AND RICHENS PERFORMED EXPERIMENTS ON MICRO-GLOSSARY DESIGN WHILST OTHER RESEARCH GROUPS STUDIED MACHINE PRODUCED GRAMMATICAL RULES TO ASSIST IN TRANSLATION.

THE SECOND PERIOD STARTED AROUND 1950 WHEN ERWING REIFLER ( WASHINGTON UNIVERSITY ) SUGGESTED A LINGUISTIC ANALYSIS AND PROPOSED SOME TRANSLATION PROCEDURES. SUGGESTIONS ON THE USE OF PRE-EDITORS AND POST-EDITORS AND A LOT OF LINGUISTIC TRICKS TO HELP IN SOLVING MACHINE TRANSLATION PROBLEMS

WERE PROPOSED. WITH SUBSTANTIAL GOVERNMENT SUPPORT, THE AMERICAN SCIENTISTS MADE PROGRESS ON A NUMBER OF FRONTS SIMULTANEOUSLY. HOWEVER, THE END OF THE SECOND PERIOD WAS A TIME OF GREAT DISAPPOINTMENT. THERE WERE A LOT OF UNSOLVED PROBLEMS WHICH SEEMED UNLIKELY TO BE SOLVED. MANY GAVE UP HOPE OF ACHIEVING ANYTHING CLOSE TO A PERFECT TRANSLATION AND DIRECTED THEIR EFFORTS TO MACHINE-AIDED TRANSLATION.

IN THE THIRD PERIOD, FROM EARLY 1960'S TO 1966, SOME LIMITED PROGRESS WAS MADE AND THE IMPORTANCE OF LINGUISTIC RESEARCH WAS REALISED. MORE EFFORT WAS DIRECTED TOWARDS RESEARCH INTO SYNTAX AND SEMANTICS IN ORDER TO ESTABLISH MORE PRECISE GRAMMATICAL RULES.

IN 1967, THE AUTOMATIC LANGUAGE PROCESSING ADVISORY COMMITTEE AT THE REQUEST OF DR. HARWORTH, DIRECTOR OF THE NATIONAL SCIENCE FOUNDATION, ANNOUNCED ITS FAMOUS REPORT RECOMMENDING THAT RESEARCH SUPPORT SHOULD BE GIVEN FOR COMPUTATIONAL LINGUISTICS, BUT NOT FOR MACHINE TRANSLATION WHICH, IN THEIR VIEW, WOULD HAVE NO POSSIBILITY OF TANGIBLE SUCCESS IN THE NEAR FUTURE. ALL RESEARCH IN MACHINE TRANSLATION WAS THUS TERMINATED.

PRESENT

THE MACHINE TRANSLATION RESEARCH AT THE CHINESE UNIVERSITY OF HONG KONG BEGAN IN 1969. WE CAREFULLY STUDIED AND REVIEWED RESEARCH REPORTS CONCERNING MACHINE TRANSLATION. THE ACCOMPLISHMENTS AND FAILURES OF PREVIOUS WORKS WERE EXAMINED. A NEW APPROACH TO AUTOMATIC TRANSLATION OF LANGUAGE BY COMPUTER WAS PLANNED.

WE RECOGNISED AT THE OUTSET THAT FULLY AUTOMATIC HIGH QUALITY TRANSLATION (FAHQ) IS NOT FEASIBLE. A GREAT DEAL OF RESEARCH WORK, BOTH IN THEORETICAL LINGUISTICS AND COMPUTER PROGRAMMING HAS TO BE DEVELOPED BEFORE THE PROBLEMS OF SPEED, ACCURACY, READABILITY, QUALITY AND ECONOMIC FACTORS OF AUTOMATIC TRANSLATION CAN BE SOLVED. ACCORDING TO BAR-HILLEL, " HIGH QUALITY HAS TO BE RELATIONIZED WITH REGARD TO USERS AND SITUATIONS. THOSE WHO ARE INTERESTED IN MACHINE TRANSLATION PRIMARILY AS PRACTICAL DEVICE MUST REALISE THAT FULL AUTOMATION OF TRANSLATION PROCESSES IS INCOMPARIBLE WITH HIGH QUALITY. THERE ARE TWO POSSIBLE DIRECTIONS IN WHICH A COMPROMISE COULD BE REACHED: ONE COULD SACRIFICE QUALITY OR ONE COULD REDUCE THE SELF-SUFFICIENCY OF THE MACHINE OUTPUT. " FAHQ MUST THEREFORE BE SACRIFICED FOR THE TIME BEING, AND EFFORT SHOULD BE FOCUSED MORE ON A MACHINE-MAN PARTNERSHIP FOR TRANSLATION.

RESEARCH INTO MACHINE TRANSLATION AT THE CHINESE UNIVERSITY TAKES A DIFFERENT APPROACH THAN THE OTHERS IN THAT THE CHINESE UNIVERSITY OF HONG KONG PLACES A HEAVY EMPHASIS ON PRE-EDITING THE SOURCE TEXT INSTEAD OF POST-EDITING THE TARGET TEXT. IT IS THE ONLY GROUP TAKING THIS APPROACH OF COMPUTER-PRE-EDITOR PARTNERSHIP. ALL THE OTHER GROUPS, WHO REALISED THE FAHQ IS NOT REALLY

ATTAINABLE IN THE NEAR FUTURE AND SO A LESS AMBITIOUS AIM IS DEFINITELY INDICATE, HAD ADOPTED A TENDENCY TO COMPROMISE IN FINDING SOME COMPUTER-POST-EDITOR PARTNERSHIP.

ACCORDING TO BAR-HILLER AGAIN, " THE ONLY REASONABLE AIM, THEN, FOR SHORT-TERM RESEARCH INTO MI SEEMS TO BE THAT OF FINDING SOME MACHINE-POST-EDITOR PARTNERSHIP THAT WOULD BE COMMERCIALY VIABLE WITH EXISTING HUMAN TRANSLATION AND THEN TO TRY TO IMPROVE THE COMMERCIAL EFFECTIVENESS OF THIS PARTNERSHIP BY IMPROVING THE PROGRAMMING IN ORDER TO DELEGATE TO THE COMPUTER MORE AND MORE OPERATIONS IN THE TOTAL TRANSLATION PROCESS WHICH IT CAN PERFORM MORE EFFECTIVELY THAN THE POST-EDITOR. THESE IMPROVEMENTS WILL, OF COURSE, UTILIZE NOT ONLY DEVELOPMENTS IN HARDWARE, PROGRAMMING AND LINGUISTIC ANALYSIS, BUT ALSO THE EXPERIENCE GAINED BY ANALYSING THE MACHINE OUTPUT ITSELF ". THE RESEARCHERS AT CURK FEEL STRONGLY THAT IF " POST-EDITORS " MENTIONED ABOVE ARE CHANGED TO " PRE-EDITORS ", THEN THE SEMANTIC AND LEXICAL ANALYSIS WILL BE MUCH SIMPLER AND EFFECTIVE. IN THEORY, IF WE CAN PRE-EDIT THE SOURCE LANGUAGE TO MAKE IT AS COMPATIBLE TO THE STRUCTURE OF THE TARGET LANGUAGE AS POSSIBLE AND THEN PROGRAM A SET OF RULES FOR ANALYSIS AND SYNTHESIS, THEN THE TRANSLATED TEXTS OBTAINED WILL BE READABLE AND ACCURATE. TO QUOTE A.G. OPTINGER : " BEFORE MACHINE INSTRUCTIONS FOR TRANSLATION CAN BE FORMULATED, THE SIGNIFICANT PATTERNS OF THE SOURCE LANGUAGE AND TARGET LANGUAGE MUST BE PRECISELY IDENTIFIED AND PUT INTO CORRESPONDENCE .... CORRESPONDING PATTERNS, THEREFORE, MUST BE DEFINED AS CONVEYORS OF EQUIVALENT MEANINGS SINCE, WHATEVER MEANING IS, IT IS GENERALLY AGREED THAT IT MUST BE PRESERVED IN TRANSLATION. "

A FIXED SET OF PRE-EDITING RULES MUST THEREFORE BE FORMULATED TO ENABLE INEXPERIENCED AND EVEN MONO-LINGUAL PEOPLE TO TRANSFORM QUICKLY THE INPUT INTO MACHINE-TRANSLATABLE FORM. WITH THIS ARRANGEMENT, POST-EDITING CAN BE KEPT TO A MINIMUM, IF NOT ALL TOGETHER ELIMINATED. GIVEN TIME AND BETTER PROGRAMMING



TECHNIQUES, THESE PRE-EDITING RULES WILL GRADUALLY BE REDUCED SO THAT THE COMPUTER WILL EVENTUALLY TAKE UP THIS ROUTINE WORK. PRE-EDITING CAN THEREFORE SOLVE MANY OF THE PRESENT LINGUISTIC PROBLEMS THAT ARE OTHERWISE DEPENDENT ON FURTHER RESEARCH IN NATURAL LANGUAGE, COMPUTATIONAL LINGUISTICS AND TRANSFORMATION MATHEMATICS. IN OTHER WORDS, MODELS THAT ARE MUCH MORE COMPREHENSIVE AND SOPHISTICATED THAN THE PRESENT ONES HAVE TO BE DESIGNED. THESE MODELS MAY TAKE YEARS TO PERFECT AND, AT PRESENT, PRE-EDITING IS ABSOLUTELY ESSENTIAL IN ORDER TO ACHIEVE THE GOAL. IN PRESENT STAGE OF OUR DEVELOPMENT, VERY COMPLEX SENTENCES CAN BE TRANSLATED WITH THE AID OF PRE-EDITING. A SENTENCE WHICH HAS A COMPLICATED STRUCTURE CAN BE ANALYSED BY THE EXISTING PROGRAM IF IT IS BROKEN UP INTO SIMPLER SENTENCES WHICH ARE THEN READILY TRANSLATED BY THE COMPUTER. SENTENCES IN CHINESE ARE OFTEN WITHOUT VERBS OR SUBJECTS AND PRE-EDITING CAN ADD THE VERBS OR SUBJECTS SO THAT THESE SENTENCES CAN THUS BE ANALYSED AND TRANSLATED.

CULT ( CHINESE UNIVERSITY LANGUAGE TRANSLATOR ) WAS DEVELOPED BASED ON THE PRINCIPLE MENTIONED ABOVE AND HAS BEEN RIGOROUSLY EXAMINED AND TESTED. THE TRANSLATION ALGORITHM IS GIVEN IN FIG. 1. SINCE THE BEGINNING OF THIS YEAR, THE CULT SYSTEM HAS BEEN USED ON A REGULAR BASIS TO TRANSLATED TWO CHINESE SCIENTIFIC JOURNALS, ACTA MATHEMATICA SINICA AND ACTA PHYSICA SINICA, WHICH ARE PUBLISHED BY THE PEKING ACADEMY OF SCIENCE. THIS ACCOMPLISHMENT INDICATE THE CORRECTNESS OF OUR APPROACH AND THE POTENTIAL CAPABILITY OF CULT.

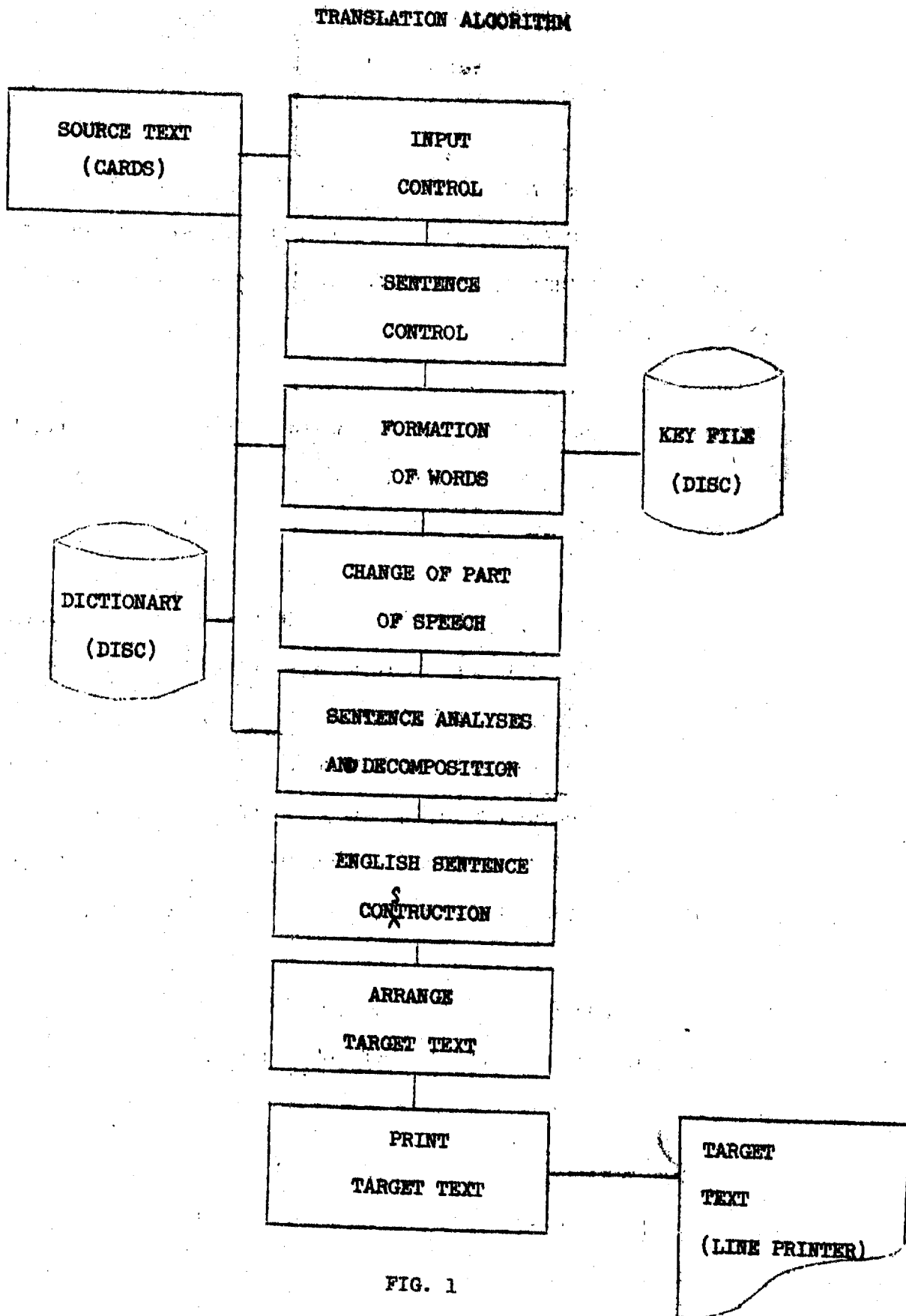


FIG. 1

**FUTURE**

IN THE NEXT FIVE TO TEN YEARS, DEVICES MIGHT BE AVAILABLE AT A REASONABLE COST FOR INPUTTING AND OUTPUTTING CHINESE OR OTHER NON-ALPHEBETIC CHARACTERS. THIS IS, AT PRESENT, THE PRINCIPAL PROBLEM FACING THE PROCESSING OF LANGUAGES OTHER THAN THOSE USING ALPHABETS.

MORE LINGUISTIC RESEARCH APPLICABLE TO MACHINE TRANSLATION MAY YIELD USEFUL RESULTS SO THAT MORE COMPREHENSIVE GRAMMATICAL RULES AND SENTENCE STRUCTURES BETWEEN LANGUAGES CAN <sup>BE</sup> FORMULATED.

AT THE CHINESE UNIVERSITY, WE ARE AT THE PRESENT WORKING ON A LANGUAGE TRANSLATOR WHICH WILL ENABLE US TO TRANSLATE CHINESE INTO ENGLISH AS WELL AS TO TRANSLATE ENGLISH INTO CHINESE. A SIMULATION STUDY HAS BEEN DONE WHICH INDICATES THE POSSIBILITY OF SUCCESS. WE ARE HOPING THAT, IN A FEW MONTHS, A MODEL MAY BE DESIGNED TO DEMONSTRATE THE SYSTEM. IF IT IS SUCCESSFUL, WE ARE OF THE OPINION THAT A UNIVERSAL TRANSLATOR MAY BE DESIGNED SO THAT ONLY ONE TRANSLATOR WILL BE REQUIRED TO DO TRANSLATION BETWEEN LANGUAGES.

ASK NOT WHAT THE COMPUTER CAN DO FOR YOU. ASK WHAT YOU, WITH THE AID OF THE COMPUTER, CAN DO FOR THE BETTERMENT OF THE MANKIND.

機器翻譯：過去，現在及將來

樂秀章教授  
香港中文大學

專家小組會議  
亞洲太平洋經濟和社會委員會  
泰國，曼谷

一九七五年十二月八日至十二日

## 機器翻譯的需要

人們不斷地累積關於本身及環境的有用知識的速率已大大地加速了，特別在最近100年來。這種增加的原因有很多：現今比以前有更多的科學家，科學研究從政府及工業方面獲得大量的資助及財政支持，冷酷的戰爭導致一個各國都致力於軍事和工藝技術霸權的競爭氣紛，等等。

由最保守的估計，在一五〇〇年之前歐洲每年出版1000本新書。到一九五〇年，四個半世紀後，出版率峭急地加速，歐洲每年出版120,000本新書。到七十年代中期，書的出版率在全世界，歐洲包括在內，達到每天接近於1,500本新書的數目。

現今比以前有更多的科學學會存在。而這些學會比以前舉行更多的會議，年會和研究會，且出版了更多的刊物和議事錄。最先的兩本科學刊物出現於300年前。在一九五八年，有代表性的學術性刊物的數目接近100,000本；現在每一年有四百萬篇刊物文章以及120,000學術性書籍和100,000學術性報告出版（就整個世界來說，每年接近於60,000,000頁）。且這出版率每年都在增加。現今，比如，好像工業國家中的工業生產一樣，每十五年，科技刊物及文章都倍增。

然而，我們可以爭辯，每一本書都是知識的進步的一個收穫。但無論如何，書出版的加速率，事實上，是與人們發現知識的速率相平行是真實的。

這些出版活動的結果是使到各個科學家都在一個資料的海洋中掙扎及致力於找尋他所需的資料。進一步，所有刊物及書籍都是用不同的國家語言出版，因為沒有一個國家能夠獨斷科技的發展。則一個國家的科學家怎樣能夠認識發生在其他國家中的科技發展及研究發現呢？我們應不應該堅持要科學家有能力讀一種，二種或更多的外國語言，以便他可以閱讀當前有關他自己科目的文獻，去探索在以前累積了的文獻，且知道有關自己科目的工作，因而使他有信心地進行他的工作以避免無意的重複。

所謂“資料爆炸”已到來了，適當的方法必須設立來幫助科學家及研究工作者閱讀當前與他的工作有關的科目的文獻，來分享人與人之間，國與國之間的知識。我相信機器翻譯提供我們面對的問題的唯一切合實際的解答。如果由互相報告進展及交換意見來消除研究工作的重複性，或者最少減至最低限度，則大量的時間及人力物力就可以節省。由學習其他國家中的科學家所做的東西，他們的成就及失敗，聽不同語言的科學家的工作就能夠加速，且他們可以計劃一個更好的工作方向來解決他們自己的問題。這些科學家們能夠互相聯繫的隨

一辦法是由出版文章來簡述所用的方法和所得的結果，且在它們可成爲其他科學家可明白之前，這些文章必須要翻譯。自動化翻譯是肯定地需要的，由於人翻譯員不可能處理所有的工作或者甚至希望在廣大的學術學科中獲得要達到好的翻譯的特別字匯。

自動化翻譯是不可能爲完美的。至於它甚至是否能夠爲高質是決定於怎樣去定標準。最迫切的目標不是設計一個完美自動化翻譯系統或者獲得高質機器翻譯，而是設計一個比現今我們有的更好和更有效的機器翻譯系統。

### 過 去

近代計算機器能用于翻譯的建議起源於A·D·BOOTH。在一九四六年，自動數字計算機的各种新的用途都被考慮到。這些應用從較明顯的數學和物理學上的問題的應用到哲學上的問題如人類思維過程的機械化，玩遊戲以及語言翻譯。在這個自動化翻譯的最初期，在美國絕大多數的研究工作都集中于德譯英，同時在蘇聯，他們集中于德譯俄。這大概是因爲在第二次世界大戰時德國已發展了高度進步的科學技術。然而，在美國注意力不久就轉向俄譯英的翻譯。對於中文的自動化翻譯，蘇聯早在美國人認識到其重要性之前在這一方面花費了很多的時間，人力和物力。

一般地說，機器翻譯經過了三個發展階段。第一個或摸索時期以一九四六年BOOTH'S的建議開始。在這時期，BOOTH和RICHENS作了專門字典設計的試驗，而同時其他的研究組研究了機械化文法規則來幫助翻譯。

第二個時期始于一九五〇年左右，當ERWING REIFLER（華盛頓大學）提出了一語言分析和提議了一些翻譯程序。提議了用預編工作和後編工作及很多語言學的技巧來幫助解決機器翻譯問題的建議。以大量的政府支持，美國的科學家在多方面同時地有了進展。然而，第二個時期的末期却是一個使人很失望的時候。有很多未解決的問題看來無希望解決。許多人放棄了爭取得到一些近似完美翻譯的希望且轉移他們的努力于機器幫助翻譯。

在第三個時期，從一九六〇年初至一九六六年，得到了一些很有限的進展和認識了語言學研究的重要性。較多的努力轉向了造句和語義的研究爲了要建立較準確的文法規則。

在一九六七年，在國家科學基金會主任Dr·HARWORTH的請求下自動化語言處理諮詢委員會發表了它的著名報告書。推薦研究的支持應給予計算語

音學，而不是給予在他們看來在不久將來沒有確實成功的可能性的機器翻譯。所有機器翻譯研究就這樣停止了。

### 現 在

在香港中文大學，機器翻譯研究始于一九六九年。我們小心地研究和覆查了關於機器翻譯的研究報告。審查了以前工作的成就及失敗。計劃了由電腦自動翻譯語言的一種新的探討。

在最初我們就認識到全自動高質的翻譯(FAHQ T)是不易辦到的，在可以解決自動翻譯的速度，準確性，通順性，質和經濟因素的問題之前許多在理論語言學和電腦程序設計兩方面的研究工作一定要展開。根據 BAR-HILLEL，"高質一定是與使用者及情形有關聯。那些有興趣於機器翻譯主要地作為一種實用的器械的必須了解到翻譯過程的全自動化是與高質相矛盾的。有二個可能的方向可達到一個妥協；可犧牲質方面或可減低機器輸出的自足。" FAHQ T 因此必須暫時被犧牲，且努力應較集中于翻譯的一個機器一人合作。

在中文大學，機器翻譯的研究採取一個與其他不同的探討。香港中文大學着重於預編原文代替後編譯文。這是唯一採用這種電腦—預編工作人員合作的探討的一組。所有了解到 FAHQ T 不能在不久的將來真正可達到因而肯定地表示了較少野心目的的其他組，已採用了一種找尋電腦—後編工作人員合作的傾向。

再根據 BAR-HILLEL，"唯一合理的目的，則，對於短期的機器翻譯研究看來是找尋一些商業上能與已存在的人翻譯共生存的機器—後編工作人員的合作。然後嘗試用改進程序設計以使到在整個翻譯的過程中委派給電腦更多的工作，這些它能夠比後編工作人員做得更有效來改進這種合作的商業效應。這些改進當然將利用不只在硬件，程序設計以及語言學分析上的發展，而且由分析機器輸出本身所得的經驗。"香港中文大學的研究工作者強烈地感到，如果以上所提到的"後編工作人員"變為"預編工作人員"，則語義和詞彙分析將會更簡單和有效。在理論上如果我們可以預編原文使它盡可能與譯文的結構相合，然後為分析和語辭的合成定下一組規則，則所得的譯文將會是通順的及準確的。引用 A.G. OETTINGER：'在翻譯的機器指令可被有系統地組成之前，來源語言及目標語言的重要模式必須確切地鑑定和統一起來...互相對應的模式。所以，必須定義為同等意思的轉送者。因為，不論意思是什麼，一般認為，在翻譯中這是必須要維持的。'

一組固定的預編規則因此必須有系統地組成使沒經驗和甚至單語言的人能很快地變換原文成機器可譯的形式。用這種整理，後編工作可以保持到最少限度，如果沒有全部消除的話。給予時間和更好的程序設計技巧，這些預編工作規則會漸漸地減少因而電腦會最終地取代這種常規工作。預編工作因此可以解決很多目前的語言學問題，這些否則要依賴於在自然語言，計算語言學和變換數學的更進一步的研究。換句話說，比目前一個更豐富的和更完整的模型一定要設計出來。這些模型可能需要多年才能完美。而在目前，爲了要達到這個目標預編工作是緊要的。在我們發展的目前階段，以預編工作的幫助，非常複雜的句子也可以被翻譯。一句有複雜結構的句子可以由現存的程序設計來分析如果它折成容易由電腦翻譯的簡單句子。中文句子是時常沒有動詞和主語的，而預編工作可以加動詞或主語因而這些句子就可以被分析和翻譯了。

CULT（中文大學語言翻譯器）是根據于以上所述的原理發展起來的，而且已嚴格地審查及試驗過。翻譯程序在圖1給出。自年初以來，CULT系統已正規地用于翻譯二本中文科學刊物，數學學報及物理學報，它們是由北京科學院出版的。這些成就表示出，我們探討的正確性和CULT的潛在能力。

#### 將 來

在未來的五年至十年中，中文和其他非字母的輸入和輸出器或可會以合理的價錢供應。這是目前面對除字母外的語言處理的主要問題。

更多有用于機器翻譯的語言學研究可能產生有用的結果，因而在語言之間更豐富的文法規則和句子結構可以有系統地組成。

在中文大學，我們目前正在從事于設計一不但能使我們翻譯中文至英文而且翻譯英文至中文語言翻譯器。一個模擬研究已經進行且顯示出成功的可能性。我們希望，在幾個月時間內可以設計出一個模型來示範這個系統。如果成功的話，就我們看來，一個萬用的翻譯器可被設計出來，因而只需一個翻譯器來進行語言間的翻譯。

"不要問電腦可以做什麼，問你自己，用電腦的幫助可以做什麼，爲了人類的改進。"



翻 譯 程 序

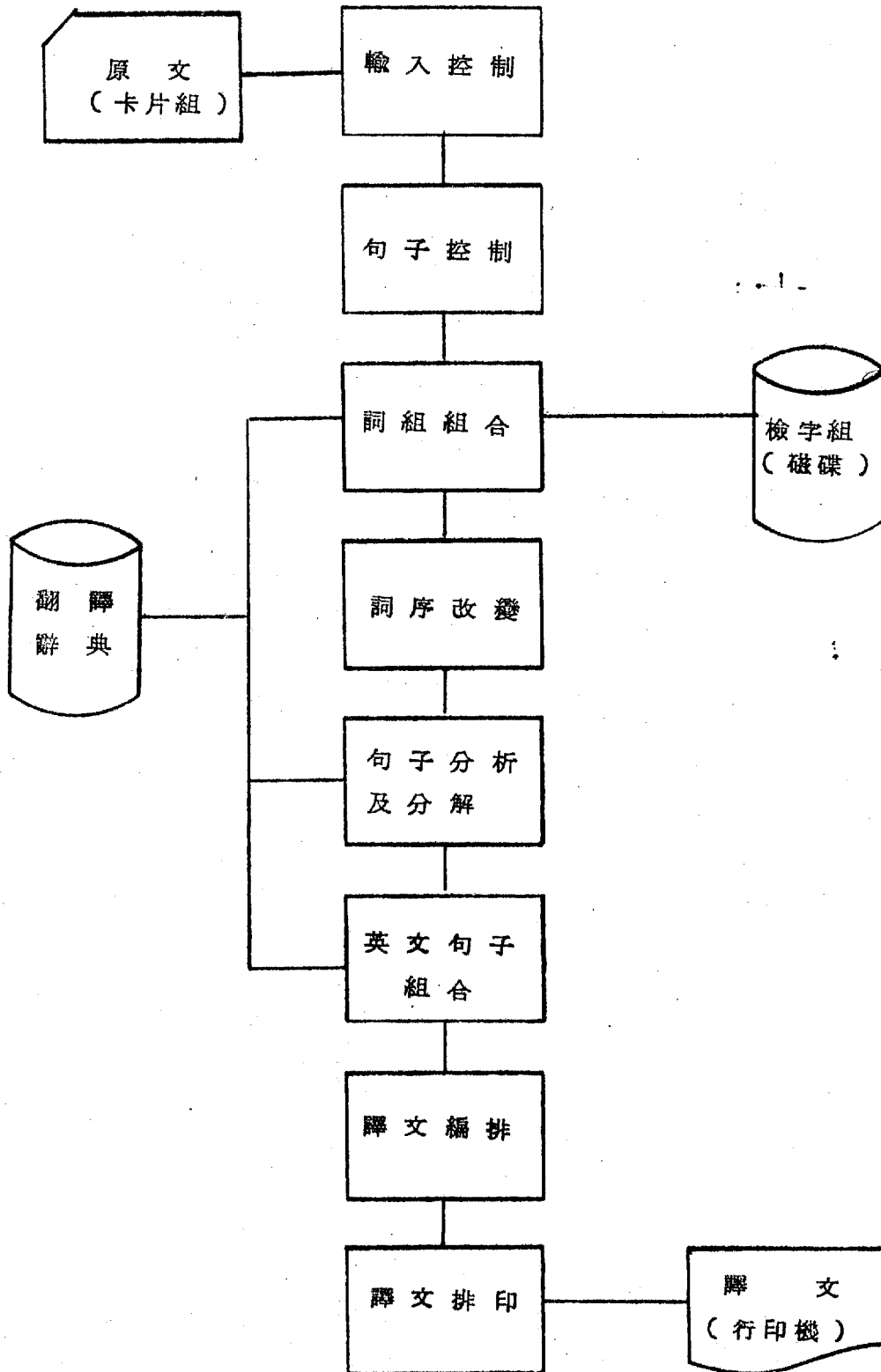


圖 1