

现代科技知识丛刊

中国的机器翻译

刘涌泉等 编著

ZHONGGUO
DE
JIQI FANYI

知识出版社

XIANDAI KE-JI ZHISHI CONGKAN

现代科技知识丛刊

中国的机器翻译

刘涌泉等著

知识出版社

1984·3·上海

现代科技知识丛刊
中国的机器翻译
ZHONGGUO DE JIQI FANYI

刘涌泉等著
知识出版社出版
(上海古北路850号)

新华书店上海发行所发行 江苏海门印刷厂印刷

开本 787×1092 毫米 1/32 印张 6.25 插页 2 字数 132,000

1984年3月第1版 1984年3月第1次印刷

印数: 1-4,500

书号: 7214·1011 定价: 0.62元

内 容 提 要

本书共收有关机器翻译的文章八篇。前面四篇,介绍机器翻译的概况、数学问题、各种自动句法的分析方法、语文形式的描述等。后面四篇,介绍机器翻译专用软件 ECMT-78 系统,以及用这个系统进行英汉自动翻译的分析、转换和生成等问题,另外关于汉—法/英/日/俄/德多种语言自动翻译的试验报告。本书可供翻译工作者业务上的参考,也是高中以上学生和广大科学爱好者研究、学习机器翻译知识的一本入门向导书。

出版前言

刘涌泉

当今世界在缩小,语言障碍在扩大

先谈一个事实:《西学东渐记》的作者容闳先生(1828~1912)叙述他去美留学时说:“一八四七年一月四日,予等由黄浦首途,船名‘亨特利思’(Huntress),帆船也,……时值东北风大作,解缆扬帆,自黄浦抵圣希利那岛(St. Helena),波平船稳。……舟既过圣希利那岛,折向西北行,遇‘海湾水溜’(Gulf Stream),水急风顺,舟去如矢,未几遂抵纽约。时在一八四七年四月十二日,即予初履美土之第一日也。是行计居舟中凡九十八日。”^①“一八五四年十一月十三日,自纽约首途。……船初行,先以他船拖至桑得荷克(Sandy Hook),迨出口后乃解缆自颺。正值逆风迎面而来,势殊猛烈。风篷不能扯满,则张半帆,旁行斜上,曲折以进。船中载货极少,既欲觅一压舱之重物,亦不可得。以故冲击风浪中,颠簸愈甚。……由桑得荷克以至香港,几无平稳之日。计水程凡一万三千海里,船行历一百五十四日乃达目的。”^②上述描述说明,顺风而下,98天,逆风而上154天,加起来共252天。容先生出国留学八年,路途上竟然用去大半年宝贵时间。然而,现在坐上波音747飞机不到一天就可到达纽约,往返也只需两天。请看,百

① 容闳《西学东渐记》,湖南人民出版社,11~12页。

② 容闳《西学东渐记》,湖南人民出版社,24~25页。

年间的变化有多大！科学技术的发达已把各国人民之间的距离大大缩短，当今的世界已不象唐僧取经和郑和下西洋时那样庞大，它不断地在缩小，远隔万水千山的彼岸犹如近在咫尺。

交通工具越来越完善，人们的交际活动也就越来越频繁。据统计，近年来，每年出席各种国际会议的人就有二百多万。在这种情况下，语言障碍的问题显得越来越尖锐了。如果说，在闭关锁国的年代，有一些从事翻译工作的人（最早的叫“象胥”“象译”，后来叫“译长”“译使”“通事”“通译”“翻译”“译员”……）基本上就能满足需要的话，而今在文化交流、商业贸易蓬勃发展的年代，尽管开办了很多外语院校，培养了一大批外语人材，但仍无济于事。更为严重的是：随着科学技术的迅猛发展，科技文献资料每十年左右就增加一倍。人们都在惊呼“情报爆炸”“情报成灾”，要为消除语言障碍而奋斗。

机器翻译问世

克服语言障碍，是一个老问题。人们曾想通过翻译（自古有之）、通过设计国际辅助语（自莱勃尼茨的万国通用文字算起已经有 300 多年了，提出的方案有 150 种以上）等途径来解决它，但总未见有多大成效。究其原因，恐怕是完全从人身上打主意行不通。出路何在？求助于电子计算机。一般的机器是人的四肢功能的延伸，而电子计算机则是人脑功能的延伸。翻译是一种思维活动。电子计算机的出现给翻译自动化创造了技术条件。

机器翻译之可能实现，还在于语言的可译性和翻译工作的可模拟性。语言，从某种意义上说，是一个符号系统。它不

是不可捉摸的。这个系统中的成分(词)以及这些成分的组合(句)都有其物质基础——语言以及反映在书面上的书写符号。由于有物质作基础,并有一定规律可寻,因而就完全有可能设计一套程序让机器进行翻译。再进一步说,语言之中的词虽然不少,但常用的并不多,句子的表现法虽然种类繁多,但是从它们当中也能分析出一定的类型来。如果估计到科技文献的语言特点:绝大多数的词是单义词,句子结构比较简单,词序比较固定,修辞色彩不浓,那么困难程度也就降低了。

机器翻译的原理没有什么奥秘,简单地说,就是让机器模拟人的翻译过程。人在进行翻译之前,必须掌握两种语言的词汇和语法。机器也是这样,它在进行翻译之前,在它的存储器(等于人脑)中已存储(或“记忆”)了语言学家编好的并由数学家加工过的机器词典和机器语法。人进行翻译时所经历的过程,机器也同样遵循:先查词典得到词的意义和一些基本的语法特征(如词类等),如果查到的词不止一个意义,那么就要根据上下文选取所需要的意义,在词汇意义和基本语法特征弄清楚之后,就要进一步明确各个词之间的关系。在这之后,根据译入语的要求组成译文(包括改变词序、翻译原文词的一些形态特征及修辞)。

现实生活要求翻译工作来一个彻底革命,客观上也具备了一定的条件。需要和可能的结合,促成了机器翻译这门边缘学科的诞生。机器翻译是语言学、数学、计算技术、自动化等科学部门共同研究的对象。语言学工作者提供适合于计算机进行加工的词典和语法规则,数学工作者把语言学工作者提供的材料代码化和程序化,即变成机器能懂的语言,计算技术工作者研制便于进行翻译的计算机,自动化工作者解决外

部设备问题，如光电输入装置。机器翻译的实现有赖于这几方面的成就和共同努力。

机器翻译发展史上的马鞍形说明了什么

美、英及苏联注意这个问题比较早，1946年电子计算机一问世，英、美学者就开始讨论用它作翻译的可能性。1952年，美、英学者在麻省理工学院召开了第一次机器翻译会。随后，在不少国家又多次召开过机器翻译专门会议，并在语言学、计算机、信息加工等学科会议上多次讨论过机器翻译专题。

1954年初，美国乔治敦大学在国际商用机器公司的协同下，用IBM-701电子计算机进行了第一次机器翻译试验，把俄文句子译成了英文句子。这次试验所用的机器词典包括250个词，机器语法规则只有六条。

1955(或1956)年，苏联科学院精密仪器与计算技术研究所在БЭСМ计算机上进行了英俄机器翻译试验，词典包括952个词，语法加工系统也比较复杂。

1955(或1956)年，英国柏克培克学院利用Apexc计算机进行了法英机器翻译试验。词典包括250个词。每小时可译1,000个词的资料。这次试验通过电视进行了表演。

1959年，日本东京电工实验室利用yamato翻译专用机进行了英日机器翻译试验。词典包括2,000个词，语法规则是根据初中一、二年级课本编制的。

1959年9月，中国科学院语言研究所和计算技术研究所，在104通用电子计算机上进行了俄汉机器翻译试验。词典包括2,030个词，语法规则系统由29个线路图表组成。

此后，不少国家还在不同型号的计算机上进行过多次试验。

机器翻译在 1966 年以前，发展是比较顺利的，尽管包含一些虚假的繁荣。那时至少有二十来个国家的七十多个研究单位从事机器翻译研究，投资多，成果也多。当时曾发行三种专门期刊，出版专著或论文集几十种。

但是，到了 1966 年，机器翻译研究史上出现了一股逆流。从此，机器翻译一段时期内处于低潮，影响了机器翻译的大踏步前进。

这股逆流来自机器翻译的发源地美国。美国曾出现过一阵机器翻译热，还在 60 年代他们就着手开办机器翻译公司，计划发行机器译文。但后来又刮出了一股冷风，1966 年印出了一个黑皮书《语言和机器》。这个黑皮书是由美国语言自动加工咨询委员会提出的。它的主要论点是：(1)机器翻译产品质量不高，人们阅读费劲，而且还有 10~16% 的误差，因而愿意看这种译文的人不多。(2)世界上英文出版的科技文献约占百分之六、七十，美国学者不懂外语已能看懂大部分材料。(3)美国学者如学俄语，经过 200 小时训练，就能看自己专业的书刊。(4)美国的翻译人员很多，翻译费用不贵，完全可以不用机器翻译。(5)机器翻译的质量短期内不会有很大提高，完全自动化高质量的翻译也许是不可能的，起码在不久的将来是不可能的。因此，他们作出结论，认为投入大量资金，短期内又收不到效益，是得不偿失的。黑皮书一出，美国的一些机器翻译工作者就起来反对，说这个报告是一个虚假的报告，所依据的材料很多是不符合事实的，相反的意见也没包括进去。但是，由于资助撤销，不少研究组纷纷下马。后来受到资

助的，只剩下五个单位（研究俄英的两个，研究德英、汉英、越英的各一个）。

黑皮书在国际上也引起了不少反响。有人发表文章抨击它，也有人附和它。有些国家受到了它的影响，但法国、联邦德国、日本等国家并未受影响，相反地，工作比以前做得更多了。没受影响的原因值得研究，我们想，除了这些国家机器翻译研究者的学术观点与美国某些学者不同（实际上也不是与美国学者不同，而是与那些认为“得不偿失”的资本家不同）之外，主要是由需要机器翻译的迫切性不同而决定的。

进入 70 年代以来，不少国家或个人又逐渐认识到有加强机器翻译研究的必要性。例如，苏联曾于 1973 年底、1974 年初专门召开会议研究机器翻译问题，会上检讨了理论脱离实际的倾向，找到了差距（有人讲“落后于西欧和美国八年到十年”），成立了全苏科技文献翻译中心，其中建立了专门的机器翻译研究部，制订了工作计划，决定同时开展四个规则系统（两个英俄、一个法俄、一个德俄）的研制工作。

进入 70 年代之后，机器翻译研究工作又开始走上了一个日趋繁荣的新阶段。这个新阶段的特征已经不再是 50 年代的过分夸张，也不再是 60 年代的大受冷落，而是扎扎实实搞工作，不断赢得新的威望。

目前世界上已有十几个机器翻译和机助翻译系统在初步应用。一般公认，法国、美国和加拿大的工作成果比较多。法国自动翻译研究中心的俄法翻译规则系统已试行应用，其自动词典已达七十万词，语法规则也比较多。南锡大学在语言分析方面进行了不少专题研究，出版的专著就有十几本。1980 年进行了一次英法机器翻译试验。法国纺织研究所文献

中心的 TITUS 系统的特点是具有存储、标引、检索、翻译和排版功能,能进行英、法、德、西四种语言的相互对译。但事先要进行一种看来是可取的加工,即文摘员作文摘时要按规定的十八种“句子格式”进行编写,而输入和翻译的都是这种形式的文摘。另外,法国格里勒布尔大学同欧洲共同体几个国家在合作研制一个 Eurotra 系统,目标是实现多语种的机器翻译。

美国的 Systran 翻译规则系统是目前世界上最出名的一个系统。这个系统也是一个多语种翻译系统,目前它能把俄、法、德等语言译成英语,其中以俄英翻译能力最强。这个系统曾在瑞士、联邦德国、加拿大、卢森堡进行过表演。例如,1975年6月13日在苏黎士大学计算中心表演时,六分钟(中央控制机时间)把三万个俄文词的资料译成了英文(如平均以25个词为一句计算,一分钟可译200句),产品质量对于情报目的是可用的。美国的乔治敦翻译规则系统曾在苏联表演,据说“一次用24小时把400多页的俄文科技书译成了英文,40个辅助人员作技术工作,译文是‘惊人的’,其质量之好(从表达原文意义来说),能使这个专业的工程师很容易核对,他们带着感激的心情于第二天就利用上了这份译文”。美国的 Logos 英越翻译规则系统,特点是技术上比较先进,据说输入是光学扫描器,输出可转到照相排版装置。

最近几年,美国搞机助翻译的人不少,曾开过两次会讨论机助翻译问题。人机对话式的翻译系统有 ITS 系统 (An Interactive Translation System)、有 MIND 系统。这种系统是随着计算机的发展在译间加工的基础上产生的。据说,这种系统能保证翻译质量,且速度也能比人快几倍。

加拿大的 TAUM 系统是一个全自动的英法翻译系统,目前在准备为加拿大空军翻译美国航天学手册。这个系统的简化型 TAUM METEO 从 1976 年起每天翻译 1,500~2,000 条天气预报。尽管范围很窄,有时还出现 20% 的误差,但它已经是一个正式使用的工具。蒙特利尔大学的 TERMUM 机助翻译系统,实际上是加拿大翻译公司的总术语库,词条 400 万,是世界上最早最大的术语数据库,现在可以经常向翻译工作者服务。

近年来,机器翻译的兴旺发达,还表现在,它又成了许多国际会议的议题。例如,国际性学术组织——文学与语言计算学会 (Association for Literary and Linguistic Computing, 简称 ALLC) 第六届和第七届学术讨论会(前者 1980 年英国剑桥、后者 1982 年比萨)、第八届(1980 年东京)和第九届(1982 年布拉格)国际计算语言学会议,都有不少机器翻译的论文发表。

以上所说的就是机器翻译发展史上所经历的一个马鞍形过程,即由 1966 年以前的高潮,转变为 1966 年以后的低潮,进入 70 年代以后,又逐步向高潮转化。人们不禁要问:为什么机器翻译会有再次的兴起呢?我想,主要原因有三:

一、社会需要是巨大的推动力

机器翻译经受严重打击,能够保存下来,并又蓬勃发展,主要原因是人们迫切需要它。根据联合国《Index Translationum 翻译索引》(25)统计,1972 年法国、联邦德国、日本三个国家所翻译的书就占世界总翻译量的 18% (联邦德国占 7%, 法、日各占 5.5%)。又据另一份材料统计:联邦德国每八本书中就有一本是译文书。加拿大议会使用英、法两种语言,每年

要投以巨资搞文件翻译。欧洲经济共同体的情况更为严重：共同体的重要文件原来使用七种文字（英、法、德、意、荷、丹麦、佛来芒）发表，希腊加入共同体后又增加了一种。前几年雇请一百多名翻译专家，在卢森堡搞了一个翻译中心，只完成工作量的十分之一。据有人估计，以后需有上万名各懂二、三种语言的专家才能应付浩繁的翻译任务。为此，共同体曾制定《改进欧洲语言间信息传递的行动计划》，其中宣布，共同体委员会准备采用一个机器翻译实用系统，并支持机器翻译研究计划。

我国是科学技术落后的国家，在实现“四化”的过程中急需参考国外科学技术，然而，由于“四人帮”对教育制度的破坏，我国科技干部的外文水平很低，在这种情况下，我国对机器翻译的需要更加迫切。如果别的国家对机器翻译质量要求高，我们则可降低。这由需要的迫切程度决定。记得我们在辩论机器翻译问题时吕叔湘先生曾用比喻十分恰当地说明了这个问题，他说：“口渴想吃冰棍的人，能买到一角一根的，当然很满意，买不到，五分的也可以，五分的也没有，三分的也凑合。”

二、计算机提供了越来越有利的条件

人们求助计算机克服语言障碍，计算机也不负众望。运算速度高达每秒一亿五千万次的计算机已经问世（1959年我们进行俄汉机器翻译试验时所用的计算机是当时运算速度比较高的，每秒一万次。如果说当时机器翻译速度与人相等，那末，现在就比人大约快一万五千倍了）；大量存储的问题也已解决，内存2,048兆字即或更大的计算机已不止一种；各种类型的高速输出装置也已齐备，所差的只是缺乏理想的输入装置（据报道，一种字体的光学自动阅读器已研制成功，目前各

国正在大力研制多种字体的阅读器，估计不久的将来定会突破)。美国、日本最近研制了几种旅游用的袖珍翻译机，有的能翻译四、五种文字，输入利用小键盘，输出利用荧光屏显示。严格说来，这不是机器翻译，而只是交际单位(整个句子)的对号转换。但是加以改进，作为旅游用，甚至不仅显示文字，而且发出比较清晰的声音来，完全是可能的。我谈这一点不是说明别的，而只是想指出近年来计算机的飞跃发展。黑皮书的作者们如果估计到这一点，我想，他们也不会作出那样的结论。

三、机器翻译研究的水平已有显著提高

机器翻译的发展，同任何其他事物一样，也是由简单到复杂、由初级到高级循序渐进的。有人模仿计算机的分类，把机器翻译系统分作三代：词对词翻译是第一代，具有语法分析(尤其是句法分析)能力的是第二代，具有语义分析能力的是第三代。这个发展过程，正好反映了对翻译问题的复杂性逐步加深认识的过程，同时也反映了机器翻译研究不断提高和机器翻译系统不断完善的过程。

机器翻译是加速实现四个现代化的 的一个重要工具

目前，全国各条战线都在响应党中央的号召，为实现四个现代化而努力奋斗。在这万马奔腾的形势下，科学技术现代化问题更显得突出。正象1977年第7期《红旗》杂志评论员指出的，“实现农业、工业、国防和科学技术现代化，关键在于实现科学技术现代化”。无论从历史上说，还是从现实来说，的确都是如此。有的科技情报刊物就曾报道说，“根据不同国

家的估价，目前 60~80% 劳动生产率的增长是由于利用最新的科技成就而获得的”。

科技情报工作是科学技术的重要组成部分，是科技战线上的一支尖兵，它要为领导部门制定规划和进行生产部署当好参谋，为科研生产部门掌握科技发展动向和赶超世界先进水平当好情报员。科学技术现代化要走在其他三化的前面，作为科技大军中的情报尖刀连，更要走在前面，尽快自动化现代化，这是理所当然的事。钱学森同志把它列为几项业务工作之首，道理也就在此。

科技情报工作自动化包括存储、检索、翻译、复印缩微、电传网络等几个重要环节。所谓自动化现代化，最终目的就是利用电子计算机等机器设备把这些环节联结起来，构成一条自动化流水线，使科技人员在任何地方都能通过情报网查询中文或外文资料、要求把外文资料译成中文、复印成册或缩微保存。

从这条自动化流水线上我们可以看出机器翻译这一环节的重要性。在某种意义上讲，翻译工作没有实现自动化，其他环节即使实现了自动化，情报工作自动化的问题也只能说是解决了一半。科技尖兵任务完成不好，将直接影响科学技术现代化以及其他三化的进程。

最后应该指出，根据目前发展的情况来看，实用性机器翻译系统的出现，不是在遥远的未来，而是指日可待了。当然初期的机译品质量较差，不过用于情报目的完全合格。随着研究的不断深入，质量定会逐步提高。另外，人们不仅可以得到文字翻译机，而且还会得到口语翻译机。根据学者们的一般估计，机译品将在 1980~1990 年期间流行于世，这是完全可

能的。但是，哪个国家先实现这个目标，一个国家何时实现这个目标，这就要看各个国家的努力了（这里面包括人力的配备，技术条件的改善以及研究人员的聪明才智和辛勤劳动）。我们国家有巨大的有利因素，这就是我们的社会主义制度，但也有许多不利因素，主要是科学技术落后。我们应该加倍努力，积极创造条件，促进机器翻译在我国早日实现，使它成为加速实现我国四个现代化的一个重要工具。

目 录

出版前言·····	刘涌泉(1)
中国的机器翻译·····	刘涌泉(1)
机器翻译与介词研究·····	刘涌泉 姜一平(15)
逻辑语义及其在机译中的应用·····	董振东(25)
外汉机器翻译中的中介成分体系·····	刘涌泉(46)
JFY-Ⅱ型英汉机器翻译系统概述·····	刘 倬(58)
ECMT-78 英汉机器翻译系统简介·····	
	刘涌泉 吴 逊 王广义 傅爱平(81)
汉一法/英/日/俄/德多语言自动翻译试验·····	
	冯志伟(103)