

机器和机器辅助翻译的目前情况

H.E.Bruderer

本文是作为一种情报来源和技术状态的概括提出来的。许多材料都是以图表形式给出的。

一、机器翻译的原理

语言数据处理是应用语言学的一个分支，是多学科相互交叉的一个领域，还可以进一步分为几个小的领域（图1，图2），而自动翻译是其中心部分（图3）。翻译系统实际上由三部分组成（图4），翻译过程分三个阶段。

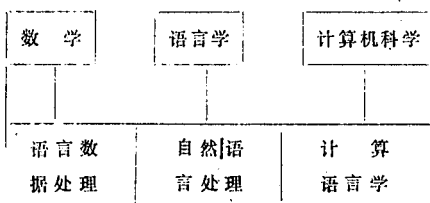


图1 计算语言学组成框图

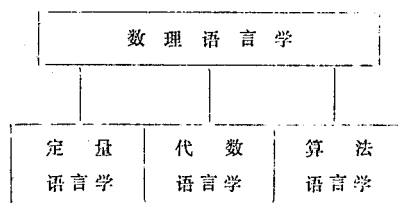


图2 数理语言学结构

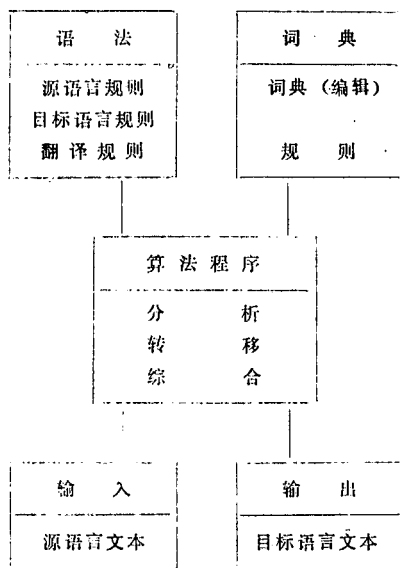


图4 翻译系统的组成

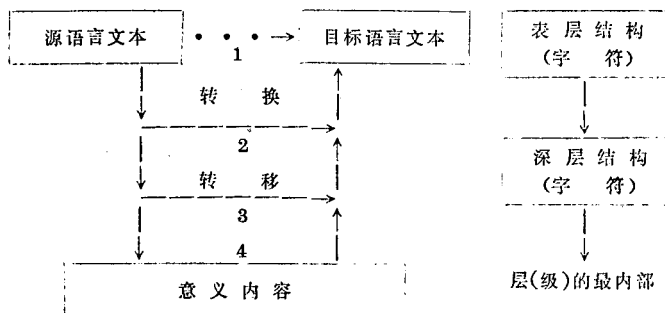


图5 翻译过程的诸阶段

注：方法1—借助字形学分析的词对词翻译。
 方法2—借助字形学和句法分析的翻译。
 方法3—借助字形学、句法和语义分析（除借助文法分析外，还可以借助人工智能）的翻译。
 方法4—借助一种通用的媒介语言（Interlingua）。
 方法1和2今天是可能的；方法3是部分可能的；方法4仍旧远未达到。

二、理论前提

今天已经有了机器和机器辅助翻译的技术基础：计算机有充分的主存储器容量和高的处理速度。但是文本搜集一直是个窄路。翻译的成本通常是难以探知的。但文本翻译的大多数系统是经

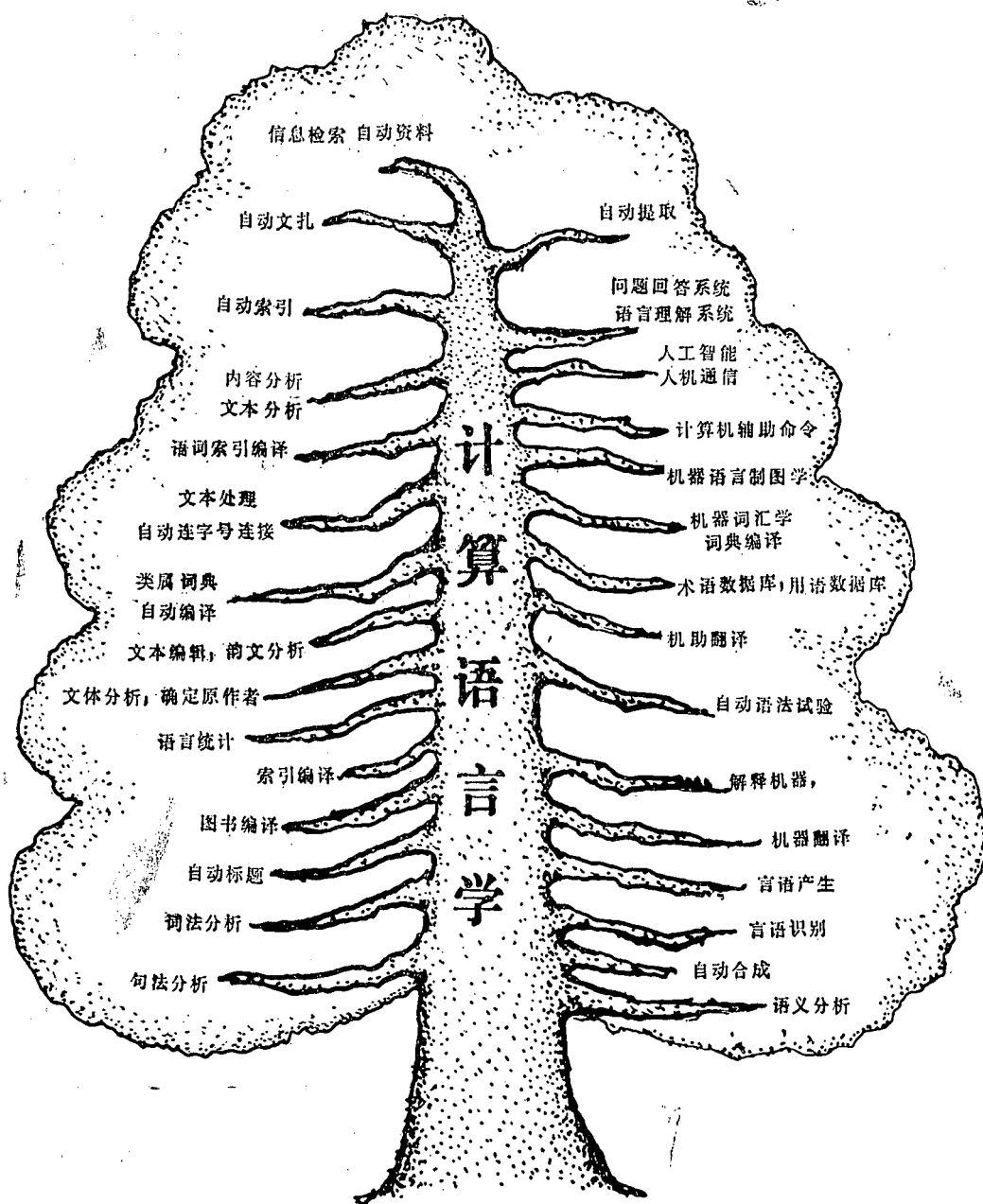


图3 语音数据处理的可能性

济的。然而一直缺乏全自动翻译语言学的前提，至今还没有显现出来能包括为了能充分扩展的语义和语用学因素。困难是由于歧义和代词所指（语义）引起的。

三、实际结果

目前只有少数机器翻译系统在实际上可用：

SYSTRAN (俄—英)，CULT(中—美)，METEO (英—法)，TITUS (德/英/法/西班牙)

和GAT (俄—英)。

在最近的将来会有更多的操作系统加入上述诸系统的行列,其中包括格勒诺勃和蒙特利尔大学, Brigham青年大学和Logos开发公司的系统。

下面的机助翻译系统是能运行的:

LEXIS (联邦语言部(西德)), TEAM (西门子), TERMIUM (蒙特利尔大学), EURODICAUTOM (欧洲共同体), EWF (东德德累斯顿工业大学)。其中联邦语言部有最长的实际经验。

表1 机器翻译的源语言及目标语言

源语言 \ 目标语言	保加利亚文	亚文	中文	德文	英文	世界语	法文	格鲁吉亚文	日文	荷兰文	波斯文	葡萄牙文	俄文	西班牙文	泰文	捷克文	越南文
保加利亚文																	
中文				★													
德文				★				★					★				
英文			★	★			★		★	★	★	★	★	★		★	★
世界语				★													
法文					★						★		★	★			
格鲁吉亚文																	
日文					★		★										
荷兰文					★												
波斯文																	
葡萄牙文																	
俄文	★			★	★		★	★									
西班牙文							★										
泰文					★												
捷克文																	
越南文																	

表2 机助翻译的源语言及目标语言

机构 \ 语言	蒙特利尔大学 (加拿大)	纽约国际商业机器公司 (IBM)	联邦语言部 (Huerth)	慕尼黑西门子公司	德累斯顿工业大学 (东德)	欧洲共同体 (卢森堡)
丹麦文						★
德文		★	★	★	★	★
英文	★	★	★	★	★	★
法文			★	★	★	★
意大利文	★		★	★		★
荷兰文				★		★
葡萄牙文			★	★		
俄文			★	★		
西班牙文				★		

四、实验和操作系统概况

(一) 机器翻译系统

1. 在美洲有加拿大的Lokehead大学, 蒙特利尔大学, 美国的原子能委员会、美空军、国家航空与宇宙航行局, Brigham青年大学, 麻省理工学院、贝克莱加州大学、德克萨斯大学, 耶鲁大学, Latsec公司/世界翻译中心公司, Smart通信公司, xyzyx公司等。

2. 在亚洲有香港的中文大学; 日本的京都大学, 九州大学, 东京电子技术综合研究所; 黎巴嫩的贝鲁特国际语言中心; 马来西亚的檳城科学大学。

3. 在欧洲有比利时的安特卫普大学; 保加利亚的保加利亚科学院; 德意志联邦共和国的海得尔堡/Constance大学, 科隆大学, 鲁尔大学, 萨尔大学, 纺织品资料和情报中心; 法国的格雷诺伯科学和医学大学, 巴黎纺织品研究所; 英国的加的夫综合大学, 艾色克斯大学, 伦敦珍珠保险有限公司/自然语言翻译专家小组; 意大利的欧洲原子能联营; 苏联的列宁格勒大学, 语言统计小组, 莫斯科专利情报中心研究所, 莫斯科核子能情报中心, 莫斯科电工研究所资料中心, 莫斯科应用数学研究所, 莫斯科国立外语师范学院, 莫斯科全苏科学技术文献资料翻译中心, 第比利斯电子学、自动化和遥控机械研究所; 捷克斯洛伐克的布拉格卡罗维大学。

(二) 机器辅助翻译系统

1. 在美洲有加拿大的蒙特利尔大学/渥太华国务秘书部; 美国的纽约国际商业机器公司;

2. 在欧洲有德意志联邦共和国的联邦语言部, DEMAG公司, 慕尼黑西门子公司; 德意志民主共和国的德累斯顿工业大学; 卢森堡的欧洲共同体; 荷兰的荷兰外交部。

五、两个翻译系统

(一) 萨尔大学机器翻译系统

这是一个自动翻译系统, 1967年开始研究, 1974年进行翻译工作(实际上1939—1973年间已进行了早期的翻译工作)。翻译的语言对是俄—德, 英—德, 世界语—德, 翻译的方向是不可逆的。在译前和译后不需人编辑。可翻译教学和语言出版物; 通俗科学课本(几百万字)。对于相当简单的句子结构翻译是可能的。每小时可翻译15000—20000词。

该系统使用德律风根TR440计算机, 操作系统的型号是MV17, 主存储器的容量是52K字(翻译程序被分成五个连续运行的主程序, 否则需要200K)。存储媒介是磁盘、磁带。程序语言是FORTRAN, 德律风根汇编语言(TAS)。用穿孔卡片, 光屏进行数据采集, 使用打印机, 萤光屏输出结果。

萨尔大学是1974年成立的“来布尼兹”自动翻译国际研究小组的成员。该大学的翻译系统于1976年9月24日在萨尔布吕肯成功地做了表演。

(二) 西门子公司(西德慕尼黑)的机助翻译系统

这是一个机器辅助翻译系统, 1967年开始研究, 1970年实际开始使用, 1976年10月公布于众。其源语言和目标语言是德、英、法、意大利、葡萄牙、俄、西班牙, 翻译的方向视要求而定。目前可扩展到九种语言。可翻译各类技术文献、标准、手册、说明书等等。翻译程序由机器辅助进行, 人干预反馈到基本形。由萤光屏和(或)电传机直接回答询问结果。只懂一种语言的研究者可使用该系统。通过补充程序和课文索引可进行统计语言学的研究和机助语言教学。

该系统使用西门子4004/35或更大的计算机, Unidata系到7000计算机, 操作系统的型号是BS1000, 内存容量至少65K。存储媒介是磁盘、磁带。使用汇编程序。使用6道TTS穿孔带和

(或) OCR—B板(字符阅读器)。

在进行与课文有关的询问时,翻译人员在他所不理解的课文下面画线,然后将其作为问题清单通过穿孔卡片或带输入给系统,系统通过一个高速打印机提供一个经选择的与课文有关的按字母顺序排好的表,作为回答。此外,还有为词典编辑的询问。

该系统的主要用户是西门子公司翻译部,荷兰外交部,专业翻译者合作组织及一些出版商。一星期大约有500次询问,每次询问小于500毫秒,每个问题需0.15—0.08马克。

这是一个为解答术语和词典编辑的程序系统,特别是为机器翻译作准备,包含在其数据库中的文件亦可通过其接口程序引进到GOLEM信息系统。借助于面向文本的技术词汇表可望将该系统的生产率提高60%,而且,在将大量翻译工作在几个翻译者间进行分配的情况下,这种表能保证术语一致,从而提高了翻译质量,询问是以自然语言进行的,目前正在研究一种自动询问方法,即自动识别和分配机器可读文本中的单词和多词术语。

从1975年开始,在语言数据处理、机器翻译和术语数据库方面的重要会议如下:

- | | | | | |
|-----------|--------------------------------------|-----------|-------------------|-----------|
| —1975年2月 | 词典编辑计算机辅助技术国际专题讨论会
(东德、德累斯顿) | —1976年4月 | 第三次欧洲控制论和系统研究会议 | 维也纳 |
| —1975年3月 | 来布尼兹小组会议(自动翻译国际研究小组)
Lugano | —1976年5月 | 语言学与信息科学专题讨论会 | 斯德哥尔摩 |
| —1975年3月 | 计算语义学研究班 | —1976年5月 | 来布尼兹小组会议 | 布鲁塞尔 |
| —1975年4月 | 在人文科学中的计算第二次国际会议
美国洛杉矶 | —1976年6月 | 自动翻译讨论会 | 卢森堡 |
| —1975年4月 | 术语学国际合作第一次讨论会 | —1976年6月 | 国际术语学讨论会 | 巴黎 |
| —1975年5月 | 数学模型和计算机在语言学中的应用第一次国际会议
(保加利亚瓦尔纳) | —1976年6月 | 第六次国际计算语言学会议 | 渥太华 |
| —1975年6月 | SYSTRAN系统在波恩、卢森堡、苏黎世的表演 | —1976年9月 | 自动词典编辑、分析和翻译国际讨论会 | 萨尔布吕肯 |
| —1975年6月 | 来布尼兹小组会议,波恩 | —1976年10月 | 自然语言处理专题讨论会 | 阿姆斯特丹 |
| —1975年8月 | 第四次国际应用语言学会议 | —1977年3月 | 瑞士语言数据处理会议 | 苏黎世 |
| —1975年10月 | 来布尼兹小组会议 | —1977年4月 | SYSTRAN讨论 | Compiègne |
| —1975年11月 | 机器翻译国际讨论会 | —1977年春 | 术语国际合作第二次讨论会 | 维也纳 |
| —1976年2月 | SYSTRAN系统讨论 | —1977年5月 | 第三次欧洲信息系统和网络会议 | 卢森堡 |
| —1976年3月 | 外国广播情报机器翻译服务讨论会 | —1977年6月 | 第八届世界翻译会议 | 蒙特利尔 |
| —1976年3月 | SYSTRAN专题讨论会 | —1977年8月 | 在人文科学中的计算第三次国际会议 | 比利时霍铁卢 |
| | | —1977年8月 | 第十二届国际语言学会议 | 维也纳 |
| | | —1978年8月 | 第五届国际应用语言学会议 | 蒙特利尔 |
| | | —1978年 | 第二次国际机器翻译讨论会 | 莫斯科 |

注:上面这个表自然是远不够详尽的,例如没有提到计算语言学暑期学校,文学和语言学计算协会年会,计算语言学学会,英国计算机学会的自然语言翻译专家小组,历届人工智能国际会议等等。

不幸的是,对机器翻译的研究在过去几年取得的进展很少。今天世界上还没有一个单一的翻译系统能够做到把一种自然语言的科学文本全自动、高质量、完美准确地翻译成另一种自然语言文本。然而这对一些国际组织,例如,联合国、欧洲共同体却是很重要很迫切的事,因为语言障碍将会变得很突出,为保证语言翻译的准确、迅速,许多国家和许多国际机构正在密切注意研究结果。此外具有各种术语的数据库的经验也将是有较高教益的。

徐志敏 编译自《Third European Congress on Information Systems and Networks
Overcoming the Language Barrier》 Luxembourg, 3—6 May 1977