

语言和计算机

《语言和计算机》编辑组编

2

LANGUAGE
AND
COMPUTER

中国社会科学出版社

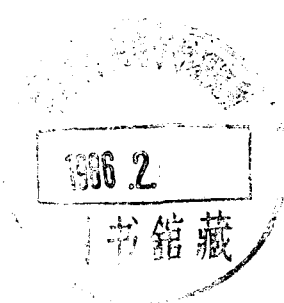
73.87

769

语言和计算机

(2)

《语言和计算机》编辑组编



中国社会科学出版社

8610055

语言和计算机

Yuyan he jisuanji

(2)

*

中国社会科学出版社出版

新华书店 北京发行所发行

八九九二〇部队印刷厂印刷

787×1092毫米 16开本 13 印张 315千字

1985年2月第1版 1985年2月第1次印刷

印数 1—14,500 册

统一书号: 9190·638 定价: 2.30元

编 者 的 话

机器翻译、人工智能(自然语言理解)、情报检索、程序教学、汉字信息处理、语言数据库、言语识别和言语合成,以及其他与语言文字信息处理有关的学科,都属于自然语言的计算机处理,或者叫作计算语言学。

计算语言学的研究,国际上已经取得巨大进展。例如:情报检索早已广泛使用;几十万至数百万词条的语言数据、术语库的建立,已获得成功;机器翻译系统、机助翻译系统和文句处理系统等,也不同程度地投入了运行;言语识别与合成、人工智能等课题正在探索。国内在这些领域的研究除情报检索外,基本上还处于实验研究阶段。

《语言和计算机》,是讨论如何利用计算机处理语言数据和研究语言的学术论文集,是研究自然语言计算机处理的一个新园地。为了促进学术交流,它将继续发表国内的科研成果,介绍国内外关于计算机处理语言方面的研究理论、方法和动态。限于条件,这个论文集每年暂出版一期(第一集已由中国社会科学出版社1982年出版)。我们希望得到读者和作者的积极支持,共同把《语言和计算机》办好,为促进我国计算语言学的研究贡献力量。

目 录

迈向机器翻译的新阶段——中国科技情报学会第一次机器

翻译学术讨论会开幕词	刘涌泉	1
中国科技情报学会第一次机器翻译学术讨论会总结发言	刘涌泉	7
JFY-I 型英汉机器翻译系统概述	刘 伟	10
一种高级的人工智能——机器翻译	刘涌泉 姜一平	30
英语多义词在机器翻译中的处理	王广义 吴 逊	41
机助翻译问题	姜一平	51
机器翻译程序设计中词序调整的实现方法	傅爱平	57
用于机助翻译的英汉自动词典	黄秀铭	64
ECAT机译系统概述	李卫东	74
法汉机器翻译 80TMFC 试验系统简介	乔 毅	94
EH-Q 俄汉机器翻译系统语言分析及其处理的几个问题	秦 璋	108
一种汉语理解系统的设想	刘涌泉 姜一平	115
MT-H78 英汉机器翻译系统概述	黑龙江大学机器翻译研究组	124
HIT-80型俄汉题录机器翻译系统	王 吟 张宝锐 滕芳师 王开铸 王宇颖 俞经善	131
德汉题录机器翻译系统的一种设计方案	胡若莉	137
英汉机器翻译中利用词组切分标志加工句子的新设想	张志庆	152
文本简化式机器翻译	姜一平 王 丽	159
机器翻译和机助翻译	[英] W. J. 哈钦斯著 黄秀铭 冯树仁译 白 水 乔 毅校	161
介绍一个英汉机器翻译规则系统	霍 屏	198
介绍几种有关机器翻译的参考文献	王 丽	198
中国科技情报学会第一次机器翻译学术讨论会论文索引	中国科技情报学会机器翻译专业组供稿	199
我国机器翻译学术会议	广 宇	200
机器翻译科研队伍的生力军	光 益	201

通 讯
与
资 料

LANGUAGE AND COMPUTER

Main Articles

- Towards a New Stage of Machine Translation—Opening Speech at the 1st National Symposium on Machine Translation Sponsored by the China Society for Scientific and Technical Information Liu Yongquan (1)
- Summary Speech at China's 1st National Symposium on Machine Translation Liu Yongquan (7)
- An Outline of JFY- I English-Chinese Machine Translation System Liu Zhuo (10)
- Machine Translation: A High-Level Artificial Intelligence Liu Yongquan and Jiang Yiping (30)
- The Handling of English Polysemantic Words in Machine Translation Wang Guangyi and Wu Xun (41)
- On the Problem of Machine-Aided Translation Jiang Yiping (51)
- The Methodology of Word Order Adjustment in Machine Translation Programming Fu Aiping (57)
- Machine-Aided Translation and Its English-Chinese Automatic Dictionary Huang Xiuming (64)
- An Outline of ECAT Machine Translation Algorithm Li Weidong (74)
- An Outline of 80TMFC French-Chinese Machine Translation Experimental System Qiao Yi (94)
- Some Issues of Language Analysis and Its Treatment Associated with EH-Q Russian-Chinese Machine Translation Algorithm Qin Zhang (108)
- An Approach to a Chinese Comprehension System Liu Yongquan and Jiang Yiping (115)
- An Outline of MT-H78 English-Chinese Machine Translation System Machine Translation Research Group of Heilongjiang University (124)
- An Introduction to HIT-80 Russian-Chinese Machine Translation System for Titles Wang Zhen and Others (131)
- A Scheme of German-Chinese Machine Translation for Titles Hu Ruoli (137)
- A New Method to Process Sentences by Means of Phrase Segmentation Marks in English-Chinese Machine Translation Zhang Zhiqing (152)

迈向机器翻译的新阶段

——中国科技情报学会第一次机器翻译

学术讨论会开幕词

刘涌泉

全国机器翻译第一次学术交流会，现在开始了。多年来，曾几次酝酿开这样一次会，但由于各种原因，都未开成。今天，我们这个会胜利召开了，而且是一个内容相当丰富的会议：论文 32 篇，有谈英汉、俄汉、德汉、法汉机译的，还有谈汉外的，有谈多语之间翻译的，也有谈机助翻译的，有专从语言方面谈的，也有专从程序方面谈的。另外，还有讲如何利用计算机辅助研究的。更可喜的是，还带来了不少实物——机器译文和机器输出的各种统计资料。单就这些来看，可以说我们的机器翻译研究工作已经取得了很大成绩，并已开始走向更加繁荣的新阶段。

谈到机器翻译研究，我们不能不对过去走过的路作一个简单的回顾。

我们起步的时间并不晚。1956 年，机器翻译研究便列入了我国科学工作的发展规划。课题的名称是“机器翻译、自然语言翻译规则的建立和自然语言的数学理论”。1957 年，机器翻译研究工作正式开始进行，到现在，已经二十三个年头了。二十三年来，机器翻译走过的道路是非常曲折的。它的发展大体上可划分为以下三个阶段：

1. 初步发展阶段（1957—1966），1966 年以前，这项工作的发展基本上还是顺利的。我们主要研究了俄汉、英汉机器翻译。俄汉方面我们投入的力量比较大。语言研究所和计算所协作研究制定了一套俄汉机器翻译规则系统，并于 1959 年国庆十周年前夕，在我国 104 大型通用电子计算机上成功地进行了试验。我们是世界上第五个进行机器翻译试验的国家。词典包括 2,030 个词条，语法规则系统由 29 个线路图表组成。这次试验翻译了九个不同类型的较复杂的句子，由于当时没有汉字输出装置，译出的句子是汉字代码。除了上述两个研究单位，北京俄语学院、华南工学院、哈尔滨工业大学等高等院校和中国科技情报研究所也先后开展了机器翻译研究。1960—1961 年期间，一些单位集中在一起，在过去研究成果的基础上又制订了一个新方案（即 1961 型方案）。

英汉方面：语言研究所和北京外国语学院协作，于 1960 年初编制了一套英汉机器翻译规则系统。由于当时人力有限，有些部分编制得比较简单。但是，从这个比较简单的系统中也取得了不少经验。

1961 年以后，各研究单位分别进行总结提高。在这个时期，除整理出一些俄汉、英汉机器翻译的资料外，还公开发表了一些论文和一个普及性小册子《机器翻译浅说》。

这个阶段的后期，语言研究所、计算所和情报所都陆续补充了一些人力，为进一步大规模地开展机器翻译研究创造了有利条件。

2. 停滞阶段(1966—1975): 1966年以来, 直到1975年底, 由于林彪、“四人帮”的干扰破坏, 机器翻译研究工作处于停顿状态。干部流失严重, 好不容易组织起来的几支队伍垮掉了, 这是我国机器翻译事业不可弥补的损失。

这里还需要提到一件事, 在“四人帮”疯狂破坏科学文化事业的日子里, 曾经刮起一股取消机器翻译的歪风。幸亏我们的一些同志顶住了这股歪风, 机器翻译事业才没有被完全砍掉。

3. 恢复发展阶段(1975至今): 邓小平同志主持中央工作期间, 科技事业开始走上正轨。1975年11月, 在中国科技情报研究所设立了一个由情报所、语言所和计算所等单位的工作人员组成的机器翻译联合研究组, 当时只有6人(语言所3, 计算所2, 情报所1)。后来又增加了一些协作人员和工作人员。该研究组的首要任务是研究和实现英汉机器翻译, 并先以冶金题录5,000条为试点, 制定英汉题录翻译规则系统和上机试验。该系统的抽样试验已于1978年5月在计算所111机(64KW)上进行, 一共试验了20条, 达到了预期效果。由于当时尚未配备汉字输出装置, 输出的是汉语拼音。1978年4月, 计算所的同志利用一软件系统进行了英汉、法汉、德汉题录翻译试验。

1978年下半年, 语言所机器翻译研究室正式恢复, 加强了力量。黑龙江大学也正式成立了英汉机器翻译研究组。1979年哈尔滨工业大学成立了以俄汉机器翻译为目标的研究组。

1977年底, 举办了机器翻译训练班, 为各部门培养了一些骨干。1978年下半年, 语言所、情报所、计算所分别招收了一定数量的研究生, 培养水平较高的机器翻译研究人员。为了给研究生提供参考资料, 编辑出版了《机器翻译论文选辑》。

1979年以来, 机器翻译研究工作更加紧张地进行, 上机试验的系统就有三四个。1979年12月在昆明召开的中国科学技术情报学会第一届学术讨论会上, 共提出有关机器翻译的论文14篇, 还展览了语言所协作组、情报所协作组和计算所(三室)三个单位机器输出的译文。有的系统一分钟能翻译4句(穿孔输入时间未计), 大大超过了人的翻译速度。最近, 翻译速度又有所提高。

今年2月, 黑龙江大学制定的英汉机器翻译系统也取得了可喜成果, 打印出几十句合格的机器译文。

哈尔滨工业大学的俄汉题录翻译系统于今年6月试验成功, 并在7月份哈尔滨地区召开的机器翻译学术交流会上作了表演。

目前, 我国主要研究的是英汉机器翻译, 有几个单位在进行研究, 第二个重点是俄汉机译。另外, 德汉、法汉也有部分同志在研究。

除了以全自动为目标的机器翻译研究, 有人也在考虑机助翻译问题, 同时还考虑如何同数据库的建立相结合的问题。

编制一对一或多对一的翻译系统是我们的重点工作, 但制定某种多对多的翻译系统以满足国际交流的需要, 也在酝酿之中。

这次会议共收到论文32篇, 这些论文分别介绍了各单位的研究成果和上机试验的结果。同时, 还收到各单位上机试验的样品, 英文的和俄文的、全文的和题录的。另外, 还收到几份计算机辅助研究的资料——词汇频率统计资料 and 正排、逆排词表。

以上这些足以说明, 机器翻译在我国, 已经显露初步繁荣的景象。当然, 我们知道, 这只是一个好的开端, 今后的任务还极其繁重。

对比一下国际上的发展情况,是很有意思的。世界机器翻译发展经历了一个马鞍形过程,即由1966年以前的高潮,转变为1966年以后的低潮,进入七十年代以后,又逐步向高潮转化。情况是这样:1966年出了一本黑皮书《语言和机器》,这本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%到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,美国学者即使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小反响。有人发表文章抨击它,也有人附和它。有些国家受了它的影响,但法国、西德、日本等国家并没有受影响,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人员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。

进入七十年代以来,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。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。

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、加拿大和美国的工作成果比较多。法国的情况主要由杨平、冯志伟同志介绍,我这里只想谈两点:(1)工作比较扎实:例如,巴黎大学文献加工和语言分析自动化实验室对8,000个动词进行十分详尽的分析。又如南锡大学在搞机译时进行了不少专题研究,出版的专著就有十几本。(2)南锡大学去年进行了一次英法机译试验,速度与我们的差不多。加拿大蒙特利尔大学的Termium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。另一个系统TAUM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。

美国的机器翻译思路较广。有搞机助翻译的,人数不少,最近几年曾开过两次会讨论机助翻译问题。有搞人机对话式的,有Mind系统、ITS系统(An Interactive Translation System)。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。这一派人同人工智能自然语言理解方面联系密切。还有坚持搞一般所说的传统的机译的。这就是世界上最出名的Systran系统。这个系统在世界上许多地方作过表演,一小时能译30万个词(中央控制机时间),一分钟译5,000词,如以25个词为一句计算,一分钟可译200句。据说美国空军一直在使用这个系

统,译后有人进行加工。它能进行俄英、法英、德英翻译,但以俄英最成熟。这个系统的原理,与我们的搞法最接近,机器词典和语法规则系统的考虑也非常相似。

英国目前主要是配合欧洲共同体搞Eurotra,参加协作的有 Manchester 大学、Aston 大学和 Essex 大学。语义型机器翻译系统的设计者 Wilks 教授就在 Essex 大学。

还有香港中文大学的系统也值得一提,他们一直在利用计算机翻译发行我国的《数学通报》,不过人在翻译过程中参与较多,译前有不少加工。

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。今年 4 月在英国剑桥召开的 ALLC (文学与语言计算学会) 第六届学术讨论会上,不少学者讨论它;这个月底在东京召开的计算语言学学术讨论会上,它是几项重要议题之一。

谈到这里,有一个问题需要回答:为什么机器翻译会有再次的兴起呢?这恐怕是人们感兴趣的一个问题。我想主要原因有四:

1. 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,并又蓬勃发展起来,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、西德、日本三个国家所翻译的书就占世界总翻译量的 18% (西德占 7%,法、日各占 5.5%)。又据另一份材料统计:西德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,今年希腊加入共同体又增加了一种。前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,以后需有两万名各懂二、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。”

2. 克服语言障碍的有效措施是求助机器。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,提出的方案已有 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。由于科学技术日新月异,运输工具日益发达,各民族文化交流越来越频繁,语言障碍问题也就越来越严重。解决问题最好的办法还是求助机器。

3. 计算机提供了越来越有利的条件。人们求助机器,机器也不负众望,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 兆字节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(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)。这里顺便提一下:美国最近研制了几种旅游用的翻译机,有半导体收音机大小,装

有三个32K的微型存储元件。能翻译四五种文字，输入利用小键盘，输出利用液晶显示器。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，而且发出声来，是完全可能的。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。黑皮报告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。

4. 机器翻译研究的水平不断提高。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。上面谈到的国内外研究成果足以说明机器翻译的巨大进展，这里就不赘述了。

最后，对今后机器翻译的研究工作谈几点设想，同时对机器翻译的未来作一个粗略的展望。

上面我们谈到，机器翻译已迈进发展的新阶段，这个新阶段的任务是什么呢？

1. 争取早日制定出实用性机译系统。我们的研究工作已进行多年，取得了不少经验和成果，并进行了多次试验。可以毫不夸大地说，我们已经有了一个向实用系统过渡的良好基础。一般认为，实用系统的数量指标应该是：机器词典规模应该是几万或几十万个词条（包括成语等），机器语法几万条或十几万条指令。质量指标应该是能在一专业内进行随机翻译，初期误差不超过5%，争取达到1%。我们所以容许误差，是由于我们的目标是全自动，即不需译前译后加工的翻译。

2. 为了达到上述的目的，必须在积极进行实验研究的同时，加强专题研究。从语言研究方面说，任务还很重。语言工作是机器翻译的基础。没有一个坚实的基础，等于在沙滩上建楼阁，机器翻译搞不好。语言规则必须在大量材料的基础上总结才能可靠。因此，对于一些重点问题，开展专题研究是完全必要的。

程序设计，尤其是机器翻译的程序设计，绝不是一种简单的技术性工作，而是包括大量研究课题在内的工作。软件技术近年来发展很快，如何把这方面的新成就用于机器翻译就是一个重要课题。机器翻译规则系统效用大小，不仅取决于语言研究工作，而且在相当大的程度上还取决于程序设计。

3. 为了加速机器翻译研究，有必要加强计算机辅助研究的工作。利用计算机加工文字材料，编制正排逆排词表、索引和逐词索引、根据不同要求得出各种统计数据和进行各种分析工作，不仅便于编制机器词典和进行专题研究，而且还能更广泛地用于文字处理的其他方面。

4. 为了使机器翻译事业更好地发展，除了集中精力搞好英汉机器翻译研究外，加强其他语种的研究也是必要的（目前已开展工作的有俄汉、德汉、法汉，但还缺少日汉）。看来，多对多翻译问题、汉外机器翻译研究问题、机助翻译问题，也应提到日程上来了。

5. 为了使机器翻译早日应用，进一步加强领导，合理规划，补充研究力量，改善设备条件，把组织协调工作切实做好，也是急需解决的问题。

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。我自己是坚信这一点的。但是，哪个国家先实现这个目的，一个国家何时实现这个目的，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）。翻译机器不是靠引进所能解决的，必须靠自力更生。当然，学习外国先进经验，还是非常必要的。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，促使机器翻译在我国早日实现。

中国科技情报学会第一次机器 翻译学术讨论会 总结发言

刘 涌 泉

同志们：

第一次全国机器翻译学术讨论会经过五天紧张的学术活动就要结束了。由于全体同志的共同努力，我们这个会确实开成了一个内容非常丰富、学术成果累累、开脑筋、鼓干劲的会议。所谓内容丰富，我指的是除了宣读或散发了三十二篇学术论文外，还听取了两个介绍国外情况的学术报告。论文内容涉及的面很广，大多数是谈机译系统编制方法的，也有少数是谈如何利用计算机进行统计研究和建立数据库以辅助机器翻译研究，以及其他一些更广泛的问题的。我所指的内容丰富，还表现在除了学术报告，还有成果展览和会后的机器翻译表演（到两个单位参观三个系统的表演）。所谓成果累累，是相对而言的。这既是一个象征性的说法，也不完全是象征性的说法。从会议正式的代表人数说，不到四十人，提出论文和报告三十四篇；从时间说，我们恢复机器翻译研究才几年，今天不仅拿出这些成品，而且还能作机器翻译表演，这样的表演在国内还是第一次，世界上也只有极个别的国家能做到这一点。所以，这不能不说是成果累累。所谓开脑筋，这次谈了几种类型的系统，同时还比较详细地介绍了法国机译系统的专用语言，大家都有不少收获，同志们思路很开阔，提出了不少建议，有方向性的，也有比较具体的。这是一方面。另一方面，通过讨论，大家更加明确了一些问题：

1. 关于应用与研究：我们国家正在搞四个现代化，各条战线极需参考外国文献，但我们懂外语的人太少，按人口平均计算，我国大概是外语力量最少的国家。说我国是世界上最需要机器翻译的国家，恐怕一点儿不过分。因此，搞机器翻译，必须要充分考虑这一现实，时刻不要忘记早日付诸应用。为了使机器翻译能达到应用，又必须加强研究工作。语言研究是机器翻译的基础，为了打好基础，我们还有大量的对比研究工作要做。如何把软件技术的成果应用到机器翻译中来，也是一个重要的研究课题。同时，大家还认识到，只有认真进行研究，反复实验，不断提高，才能使我们的译品达到试用阶段，并进而提高到真正的实用水平。也只有这样，才能防止“三品（样品、展品、奖品）”现象产生。

2. 关于形式结构分析和语义分析：大家认识到，机器翻译发展到目前的水平，除了依靠形式结构分析，还应重视语义分析，但是机器翻译不同于人工智能，语义分析应用到什么程度是需要进一步探讨的。目前，不同的方案，不同的语种可以根据自己的具体情况来考虑这个问题。

3. 关于质量标准问题：我们主张全自动，但不坚持高质量，因为高质量在一个相当长的时期是不可能的。机器翻译史上曾以100%和95%来区分是否高质量。我们主张的是后者。

(1)全自动:指的是不要译前、译后加工。其实在输入时把句号同小数点分开这样的译前加工也是有的,但我们不采用有些人主张的那种译前加工。因为要做那样的加工,需要懂原文和专业,而且很费时间,有这个时间,大概不用机器也译出来了。为什么不考虑译后加工呢?就目前情况看,如果有快速打印机,我们有的系统一分钟就能译三四十句,将来设备改善,也可能象美国的系统一样,一分钟译二百多句。让人校对这样大量的输出材料是很难的。有些同志可能没看到过机译文,机器如翻译乱了,校对起来很不容易。另外,汉字都堆在一起,中间无空白,比外文更难校。(2)质量标准:开幕词中提到了一个百分比,同志们有不少反应,昨天在小组会上我补充谈了一下,因为同志们都关心这个问题,这里再谈谈:95%是指正确能懂,不误事。这样的译文将带有浓厚的机器味,这是肯定的,机器味能逐渐减少,这也是肯定无疑的。但是要完全去掉,恐怕很难。再有,从阅读时间上说,比看人译的材料多用一两倍时间,只要能分析出正确意思,就算合格。某些原文词没译出,不误事,也算合格。另外,我是把这样一个标准当作真正实用的商品来要求的,在真正商品登记注册以前,可以有各种试用品(质量为70%的、80%的、90%的)。如果真想很快试用,我认为把范围缩小一点,投入一定人力,现有的系统明年就能提供质量为60%以上的题录翻译试用品。不过我总认为,错误太多不好说是真正实用的商品,十句中四句有错,而且错的和对的混在一起,还有的句子错的跟对的比较起来差不多(因为对的有时也带有很浓的机器味),甚至有时错的比对的更通顺一些。总起来说,作为真正实用的商品标准应高一些,而作为试用品可定得低一些。对全文要求可放低些,对无上下文参考的题录要高些。因此,补充一些阶段性的指标(90%、80%、70%的试用品)是完全必要的,这有利于总结中间成果,促进科研发展。

所谓鼓了干劲,是指大家看到这些丰硕成果,信心增强,确实感到机译已不是纸上谈兵,它已有一定基础,在这个基础上进一步努力,就可使机器翻译实现,而且这个实现的日期,不是在看不见的未来,而是就在八十年代。当然,我们现在还不敢说究竟在哪一年,这要看我们的努力,也要看其他方面的条件。

谈到条件,最主要的是加强领导。这次会上不少同志呼吁应该迅速改组原先的机器翻译领导小组,适当补充一些专业人员,并使这个组织真正发挥集体领导的作用,在制定规划、鉴定方案、审批经费、协调全国工作等方面切实起到监督领导作用。

与加强领导有关的还有建立专门机构问题,看来把有关自然语言计算机处理的学科组织起来统一领导,对机器翻译等专业的发展也是有很大促进作用的。

加强规划和协调工作,在目前百花齐放的情况下,分清轻重缓急,合理规划和协调各部门的工作,更是必要的。有同志建议,目前的安排重点应该是英汉,其次是俄汉,再其次是其他方面。看来是可以考虑的。

这次会议当中,同志们提了不少建议,现在把有关的情况向同志们讲一下。

1. 关于成立机器翻译研究会,在前天的座谈会上情报学会的副理事长张承秉同志和秘书长李新同志已同意我们成立研究会,这对我们的机器翻译工作是一个很大的支持,为今后的国内外学术交流创造了更加有利的条件。

2. 关于建立语料库、术语数据库,这是一个好的建议。语料库现在已有两个单位在搞,语言所已将《Understanding Science》(《了解科学》)送入计算机并进行了统计调查,今后还将继续输入材料。交大也在建立这样一个语料库。今后两家准备合作进行。术语数据库的问题

尚需进一步磋商解决，落实承担单位。

3. 关于加强学术交流和创办刊物问题，语言所有一个用计算机编辑机器翻译论文索引的课题，准备工作已做得差不多了。这个课题完成后，预计可以按季度提供外文新书刊的情报。机译研究会办刊物现在似乎条件还不成熟，准备过一段时间再考虑。目前《情报科学》有机器翻译专栏，语言所计划出一个不定期刊《语言与计算机》。另外，还有一些情报期刊可投稿。

4. 关于学会活动，这里只谈全国性的。全国性学术交流会原则上每年举行一次，但也准备考虑组织小规模的主题讨论会。

最后，我想重复一下大家的意见：机器翻译是一门边缘学科，是一项集体性工作。机器翻译这个研究课题，不同于哥德巴赫猜想那样的课题，是不能单靠某个人或某一个学科所能完成的。因此，加强协作，加强学术交流是非常必要的。学术上有不同意见是正常的，但这不应影响我们共同攻坚，相反地更应促进我们去寻求真理。同志们，让我们共同努力，为科技情报工作自动化的实现，为祖国社会主义四个现代化的实现贡献自己的一切。

JFY-Ⅱ型英汉机器翻译系统概述

刘 倬

一、 前言

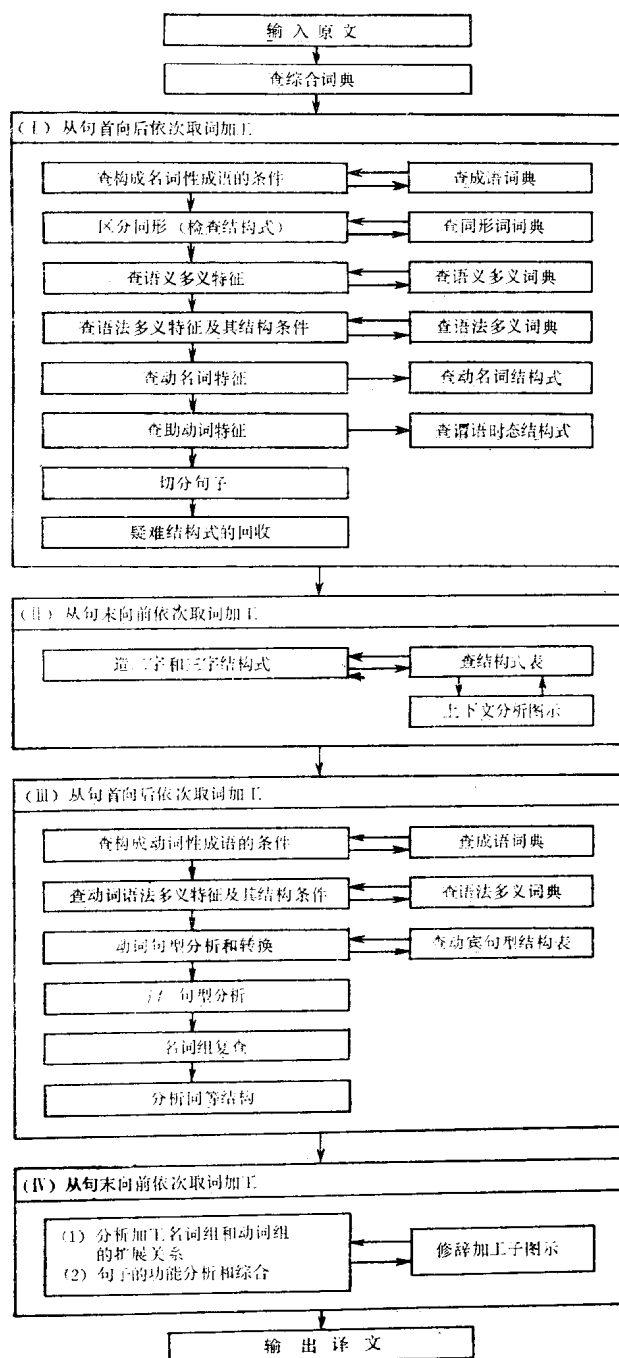
英汉机器翻译的研究和试验，我们是从1975年12月开始的。当时，在国家科委的直接关怀和支持下，由中国科技情报所主办，成立了一个机器翻译研究协作组。参加这个协作组的还有中国社会科学院语言研究所、中国科学院计算技术研究所、北京化工研究院、冶金部情报所、上海科技情报所和林业科学院情报所等单位的同志。在几年的协作中，以5,000条冶金题录和200条冶金文献的文句为基础材料，由语言所提出规则系统的总体设计和程序中各基本加工过程的算法，大家一起群策群力，共编制了两套翻译规则系统，进行了三次上机试验。这两套翻译系统是英汉题录翻译系统和英汉全文翻译系统。三次试验中，英汉题录翻译系统第一次试验是在计算所国产111机上进行的（1976—1978年），编写程序使用的是111机汇编语言；其他两次，一次题录翻译试验，一次全文翻译试验，都是在日本TK-70机上进行的（1978—1980年），编写程序使用的是BOL语言。第一次试验只打出了20条题录抽样的译文，后两次试验打出了5,000条题录和200句文句的译文，结果基本上是令人满意的^①。由于全文翻译系统是题录翻译系统的进一步扩充和完善，它包括了题录翻译系统的全部设计思想。所以下面仅就全文翻译系统的设计作一个简要的介绍，不妥之处希望大家提出批评和指正。

二、 JFY-Ⅰ型系统

我们定名为“JFY-Ⅰ型系统”的英汉全文机器翻译系统，是一套以实际应用为设计思想的试用型系统。从我国的实际出发，这套系统作翻译时不需要译前编辑和译后编辑。换句话说，英文题录或句子经穿孔后就可以直接输入计算机，进行翻译，得出汉语译文。这套翻译系统由两大部分——自动词典和自动语法规则系统组成。其中，自动词典包括一部综合词典、一部成语词典、一部语义多义词典、五部语法多义词典和一部汉语词典。自动语法规则系统包括四线加工：第一线的任务是区分同形，切分句子，分析确定谓语以及查询某些词的强支配关系。第二线是根据给定的结构式表构成名词性词组。第三线分析动词性词组，同时按给出的英汉语对比句型确定动宾关系的类型和译法。另外，在这一线中还要分析解决名词的同等关系。第四线的任务比较复杂：其一，要通过介词分析，解决名词性词组和动词性词组的扩展关系；其二，解决前几线加工中的遗留问题，如某些副词的分析，某些标点的特殊用法的分析，等等；其三，根据给定的英汉语中介成分（即一种特定的调位信息）调整词序；其四，以独立的语段或句子为基本加工单位，进行译文的修辞加工；最后是查汉语词

^① 参阅附件《英汉全文翻译的试样》。

典，输出打印汉语译文。JFY-Ⅱ型翻译系统的工作流程可图示如下：



(一)自动词典及其程序设计

在机器翻译中，自动词典实际上是一个为计算机作翻译服务的自然语言词汇的数据库。如上所述，JFY-Ⅱ型机译系统的自动词典是由一部综合词典和几部分词典组成的。这些词典虽然任务和内容各不相同，但都是按一维符号表的形式存放，它们都是用地址指示字（链字）彼此联接，构成一个独立而完整的词典系统，统称“自动词典”。其中，词典与词典之