

机器翻译——从实验室走向市场

冯志伟

提要 本文叙述了机器翻译发展从草创期,经过复苏期,到繁荣期的发展过程,这个过程也就是机器翻译从实验室走向市场的过程,特别介绍了基于统计的机器翻译和基于实例的机器翻译。

自然语言处理系统的研究首先是从机器翻译系统的研究开始的。1946 年电子计算机刚一问世,人们在把计算机广泛地应用于数值运算的同时,也想到了利用计算机把一种或几种语言翻译成另外一种语言或另外几种语言。从 50 年代初期到 60 年代中期,机器翻译一直是自然语言处理系统研究的中心课题,当时采用的主要是“词对词”翻译方式,这种不是建立在句法和语义分析的基础上的简单技术,没有得到预期的效果。60 年代中期,人们开始转入对自然语言的句法、语义和语用等基本问题的研究,并尝试着让计算机来理解自然语言。许多学者认为,断定计算机是否理解了自然语言的最直观的方法,就是让人们同计算机对话,看计算机能否对人用自然语言提出的问题作出回答,这样,就出现了“人机对话”系统的研究,并成为人工智能这一新兴学科的重要内容。自然语言处理的理论和方法也就在这些具体的研究中逐渐形成、成熟并完善起来。由于过分地强调自然语言的理解,在很长一段时期内,学术界普遍认为,如果计算机没有理解自然语言,就搞不成机器翻译,使机器翻译的研究走入了“误区”,出现了徘徊不前的局面。今天,由于信息时代的推动,机器翻译开始面对实际的需要,开发实用的产品,终于走出了这个“误区”。机器翻译研究从实验室走向了市场,机器翻译软件产品纷纷到商品市场上参与竞争,各显身手,机器翻译的研究进入了一个新时期。在这篇文章中,我们将简要地回顾机器翻译从实验室走向市场的这个曲折的历程。

在自然语言计算机处理系统的研究中,机器翻译是一个历史悠久的部门。

在 17 世纪,一些有识之士就提出了采用机器词典来克服语言障碍的想法。笛卡儿(Descartes)和莱布尼兹(Leibniz)都试图在统一的数字代码的基础上来编写词典。在 17 世纪中叶,贝克(Cave Beck)、基尔施(Athanasius Kircher)和贝希尔(Johann Becher)等人都出版过这类的词典。由此开展了关于“普遍语言”的运动,一些人试图在逻辑原则和图形符号的基础上,创造出一种无歧义的语言,这样一来,人们就不必再由于误解而产生交际方面的困惑了。维尔金斯(John Wilkins)在《关于真实符号和哲学语言的论文》(Essay towards a Real Character and Philosophical Language, 1668)中提出的国际语(Interlingua)是这方面最著名的成果。

本世纪 30 年代之初,法国科学家阿尔楚尼(G. B. Artsouni)提出了用机器来进行语言翻译的想法。

1933 年,苏联发明家特洛扬斯基(. . .)设计了用机械方法把一种语言翻译为另一种语言的机器,并在同年 9 月 5 日登记了他的发明。但是,由于 30 年代的技术水平还很低,他的翻译机没有制成。

机器翻译系统的研制是从 40 年代末期开始的。可以分为草创期、复苏期、发展期三个时期。

一、草创期(1954 年 - 1970 年):1946 年,美国宾夕法尼亚大学的埃克特(J. P. Eckert)和莫希莱(J. W. 1997 年第 3 期

Mauchly) 设计并制造出了世界上第一台电子计算机 ENIAC。在电子计算机问世的同一年,美国罗克非勒基金会副总裁韦弗(W. Weaver)、英国工程师布斯(A. D. Booth)就提出了利用计算机进行语言自动翻译的想法。1949年,韦弗发表了一份以《翻译》为题的备忘录,正式提出了机器翻译问题。在这份备忘录中,他除了提出各种语言都有许多共同的特征这一论点之外,还有两点值得我们注意:第一,他认为翻译类似于解读密码的过程。第二,他认为原文与译文“说的是同样的事情”,因此,当把语言 A 翻译为语言 B 时,就意味着,从语言 A 出发,经过某一“通用语言”(Universal Language)或“中间语言”(Interlingua),然后转换为语言 B,这种“通用语言”或“中间语言”,可以假定是全人类共同的。可以看出,韦弗没有看到机器翻译在词法分析、句法分析以及语义分析等方面的复杂性。

由于学者的热心倡导,实业界的大力支持,美国的机器翻译研究一时兴盛起来。1954年,美国乔治敦大学在国际商用机器公司(IBM公司)的协同下,用 IBM-701 计算机,进行了世界上第一次机器翻译试验,把几个简单的俄语句子翻译成英语,接着,苏联、英国、日本也进行了机器翻译试验,机器翻译出现热潮。

早期机器翻译系统的研制受到韦弗上述思想的影响,许多研究者都把机器翻译的过程与解读密码的过程相类比,试图通过查询词典的方法来实现词对词的机器翻译,因而译文的可读性很差,难于付诸实用。

1964年,美国科学院成立语言自动处理咨询委员会(Automatic Language Processing Advisory Committee,简称 ALPAC 委员会),调查机器翻译的研究情况,并于 1966 年 11 月公布了一个题为《语言与机器》的报告,简称 ALPAC 报告,对机器翻译采取否定的态度,指出,机器翻译研究遇到了难以克服的“语义障碍”(semantic barrier)。一些学者也认为,没有计算机对于自然语言的理解,机器翻译是没有希望的。ALPAC 报告使机器翻译走入了“误区”。在 ALPAC 报告的影响下,许多国家的机器翻译研究跌入低潮,许多已经建立起来的机器翻译研究单位遇到了行政上和经费上的困难,在世界范围内,机器翻译出现了空前萧条的局面。不过,尽管在萧条时期,法国、日本、加拿大等国,仍然坚持研究,于是,在 70 年代初期,机器翻译又出现了复苏的局面。

二、复苏期(1970年-1976年):在复苏期,研究者们普遍认识到,原语和译语的差异,不仅只表现在词汇的不同上,而且,还表现在句法结构的不同上。为了得到可读性强的译文,必须在自动句法分析上多下功夫。

早在 1957 年,美国学者英格维(V. Yingve)在《句法翻译的框架》(Framework for syntactic translation)一文中就指出,一个好的机器翻译系统,应该分别地对原语和译语都作出恰如其分的描写,这样的描写应该互不影响,相对独立。英格维主张,机器翻译可以分为三个阶段来进行。第一阶段:用代码化的结构标志来表示原语文句的结构;第二阶段:把原语的结构标志转换为译语的结构标志;第三阶段:构成译语的输出文句。第一阶段只涉及原语,不受译语的影响,第三阶段只涉及译语,不受原语的影响,只是在第二阶段才涉及到原语和译语二者。在第一阶段,除了作原语的词法分析之外,还要进行原语的句法分析,才能把原语文句的结构表示为代码化的结构标志。在第二阶段,除了进行原语和译语的词汇转换之外,还要进行原语和译语的结构转换,才能把原语的结构标志变成译语的结构标志。在第三阶段,除了作译语的词法生成之外,还要作译语的句法生成,才能正确地输出译文的文句。

英格维的这些主张被机器翻译系统的开发人员普遍接受,因此,这个时期的机器翻译系统几乎都把句法分析放在第一位,并取得了很大的成绩。

这个时期机器翻译的另一个特点是语法(grammar)与算法(algorithm)分开。

早在 1957 年,英格维就提出了把语法与“机制”(mechanism)分开的思想。英格维所说的“机制”,实质上就是算法。所谓语法与算法分开,就是要把语言分析和程序设计分开,程序设计工作者提出规则描述的方法,而语言学工作者使用这种方法来描述语言的规则。语法和算法分开,是机器翻译技术的一大进步,它非常有利于程序设计工作者与语言工作者的分工合作。

复苏期的机器翻译系统的典型代表是法国格勒诺布尔理科医科大学自动翻译中心的机器翻译系统。这

个自动翻译中心的主任沃古瓦(B. Vauquois)教授明确地提出,一个完整的机器翻译过程可以分为如下六个步骤:(1)原语词法分析,(2)原语句法分析,(3)原语译语词汇转换,(4)原语译语结构转换,(5)译语句法生成,(6)译语词法生成。其中,第一、第二步只与原语有关,第五、第六步只与译语有关,只有第三、第四步牵涉到原语和译语二者。这就是机器翻译中的“独立分析-独立生成-相关转换”的方法。他们用这种方法研制的俄法机器翻译系统,已经接近实用水平。

他们还根据语法与算法分开思想,设计了一套机器翻译软件ARIANE-78,这个软件分为ATEF,ROBRA,TRANSF和SYGMOR四个部分。语言工作者可以利用这个软件来描述自然语言的各种规则。其中,ATEF是一个非确定性的有限状态转换器,用于原语词法分析,它的程序接收原语文句作为输入,并提供出该文句中每个词的形态解释作为输出;ROBRA是一个树形图转换器,它的程序接收词法分析的结果作为输入,借助语法规则对此进行运算,输出能表示文句结构的树形图;ROBRA还可以按同样的方式实现结构转换和句法生成;TRANSF可借助与双语词典实现词汇转换;SYGMOR是一个确定性的树-链转换器,它接收译语句法生成的结果作为输入,并以字符链的形式提供出译文。

通过大量科学实验和实践,研究者们认识到,机器翻译中必须保持原语和译语在语义上的一致,一个好的系统应该把原语的语义准确无误地在译语中表现出来。这样,语义分析在机器翻译中越来越受到重视。

美国斯坦福大学威尔克斯(Y. A. Wilks)提出了“优选语义学”(preference semantics),并在此基础上设计了英法机器翻译系统,这个系统特别强调在原语和译语生成阶段,都要把语义问题放在第一位,英语的输入文句首先被转换成某种一般化的通用的语义表示,然后再由这种语义表示生成法语译文输出。由于这个系统的语义表示方法比较细致,能够解决仅用句法分析方法难于解决的歧义、代词所指等困难问题,译文质量较高。

三、繁荣期(1976年—现在)。繁荣期最重要的特点,是机器翻译研究走向了实用化,机器翻译产品开始进入市场。

机器翻译的繁荣期是以1976年加拿大蒙特利尔大学与加拿大联邦政府翻译局联合开发的实用性机器翻译系统TAUM-METEO正式提供天气预报服务为标志的。这个机器翻译系统投入实用之后,每小时可以翻译6万-30万个词,每天可以翻译1500-2000篇天气预报的资料,并能够通过电视、报纸立即公布。

日本富士通公司开发的ATLAS-I(Automatic Translation System-I)系统是一个建立在大型计算机上的英日机器翻译系统,该系统以句法分析为中心,可进行科学技术文章的翻译,在FACOM M380计算机上,每小时可翻译60000词。日本富士通公司开发的ATLAS-II机器翻译系统也建立在大型计算机上,但其翻译方式与ATLAS-I不同,以语义分析为中心。该系统建立了用于表示概念之间关系和客观世界知识的“世界模型”,在译文生成时,特别注意单词之间的搭配关系和邻接关系,在机器翻译过程中,采用一种叫做“概念构造”的中间语言来作为原语和译语的共同表达。该系统目前用于日英机器翻译。

此外,日本还有一批实用化机器翻译系统。

欧美除TAUM-METEO机器翻译系统之外,也陆续推出了一批实用化的机器翻译系统,如:法国纺织研究所的TITUS-IV系统;美国在乔治敦大学机器翻译系统的基础上进一步开发的机器翻译系统SYSTRAN;美国罗各斯(LOGOS)公司开发的LOGOS-III机器翻译系统;美国国家航空和航天的NASA系统;美国魏德纳(WEIDNER)通讯公司WCC的WEIDNER机器翻译系统;设在华盛顿的泛美卫生组织研制成的PAHO系统;美国卡内基-梅隆大学(CMU)研制了基于知识的机器翻译系统KBMT-85;德国西门子(SIEMENS)公司与美国德克萨斯大学(Texas University)合作研制的METAL系统;德国萨尔大学(Universität des Saarlandes)研制的SUSY(Saarbrücken Automatic Translation System)系统。此外,还有一些大规模的机器翻译系统正在研制之中,例如,EUROTRA计划、Mu系统、ODA计划、DLT系统等。

1978年,欧洲共同体在继续使用和发展SYSTRAN系统的同时,提出了欧共体内七种语言(后来变为九
1997年第3期

种)之间进行任一方向翻译的多语种机器翻译计划 EUROTRA,此计划于 1982 年正式实施,前后延续了十多年,至今尚未达到预期的结果。

日本在提出第五代计算机计划的同时,于 1982 年至 1986 年由政府开展了英日、日英机器翻译 Mu 系统的研制,接着,又由通产省出面,组织与亚洲四个邻国(中国、印度尼西亚、马来西亚、泰国)合作研究日语、汉语、印度尼西亚语、马来语、泰语五种语言互译的多语言机器翻译 ODA 计划,原定于 1987 年至 1992 年完成,后来延长至 1995 年初完成。但是,ODA 计划完成验收时,翻译的效果并不理想。

欧洲共同体在 1982 年开始实施 EUROTRA 计划的同时,还支持了多语言机器翻译系统 DLT 的可行性研究。从 1984 年开始,改由荷兰政府和荷兰的一家软件公司 BSO 各出资一半对此系统的研制进行长期的支持,DLT 系统原打算 90 年代中期开始实用化,可是至今尚未得到满意的结果。

传统的机器翻译系统,按其翻译方式来分,可大致分为直译式、转换式和枢轴式三种类型。

1. 直译式:把原理的词或句子直接替换成译语的词或句子,必要时对词序进行适当的调整。这种直译式广泛地应用于早期的机器翻译系统中。目前,仍然有不少系统采用直译式来进行机器翻译。一个极端的例子就是袖珍旅游翻译机,这种翻译机存贮着原语的常用句子及其相应的译语的译文,翻译时直接查出其对应物即可。比较复杂的直译式是句法直译式、语义直译式。在直译时要进行词语的选择和词序的变换。

2. 转换式:在原语和译语之间设定能在一定程度上表现语义关系的中间表达式,根据中间表达式所处平面的不同,又可分为句法转换式和语义转换式。句法转换式一般以树形图作为句法结构的中间表达式,采用这种中间表达式来分析句子,会出现大量的在句法上有歧义的树形图,致使转换规则十分庞杂,规则的抽象程度也比较低。语义转换式一般采用语义网络作为中间表达式。这样,在表层结构上出现歧义的句子,在语义的中间表达式这一平面上不会再有歧义。因此转换规则的数目不多,这是其优点。但其缺点是为了得到这样的中间表达式,需要进行大量的分析和运算,而且由这样的中间表达式去生成译文也是十分困难的。

3. 枢轴式:把语义转换式推到极限,用中间表达式来表示不依赖于任何具体语言的普遍意义,这种普遍意义就是枢轴。在枢轴式中,原语和译语之间不再需要进行什么转换,转换规则的数目减少到零,翻译的过程就是首先把原语的文句变为枢轴中的普遍意义,再由普遍意义生成译文。由于完全取消了转换规则,所以,原语的分析 and 译语的生成都比较复杂。

传统的机器翻译方法都是基于规则的机器翻译方法,由于分析技术的限制,大多数都以句法直译式和句法转换式为其主流。今后的主流如何,尚难断定。

除了传统的机器翻译方法之外,近年来还出现了基于经验的机器翻译方法。所谓“基于经验”,一是指基于统计,一是指基于实例。

早在 1949 年,韦弗在他的备忘录中,就提出了使用统计学的办法来解决机器翻译问题,但是,由于当时尚缺乏高性能的计算机和联机语料,采用基于统计的机器翻译在技术上还不成熟。现在,这种局面已经大大改变了,计算机在速度和容量上都有了大幅度的提高,也有了大量的联机语料可供统计使用,因此,在 90 年代,基于统计的机器翻译又兴盛起来。

基于统计的机器翻译把机器翻译问题看成是一个噪音信道问题,如下所示:

S 噪音信道 T

可以这样来看机器翻译:一种语言 S 由于经过了一个噪音信道而发生了扭曲变形,在信道的另一端呈现为另一种语言 T,翻译问题实际上就是如何根据观察到的语言 T,恢复最为可能的语言 S。语言 S 是信道意义上的输入,在翻译意义上就是目标语言,语言 T 是信道意义上的输出,在翻译意义上就是源语言。从这种观点看来,一种语言中的任何一个句子都有可能是另外一种语言中的某几个句子的译文,只是这些句子的可能性各不相同,机器翻译就是要找出其中可能性最大的句子,也就是对所有可能的目标语言 S 计算出概率最大

的一个作为源语言 T 的译文。由于 S 的数量巨大,可以采用栈式搜索(stack search)的方法。栈式搜索的主要数据结构是表结构,表结构中存放着当前最有希望的对应于 T 的 S,算法不断循环,每次循环扩充一些最有希望的结果,直到表中包含一个得分明显高于其它结果的 S 时结束。栈式搜索不能保证得到最优的结果,它会导致错误的翻译,因而只是一种次优化算法。

基于统计的机器翻译进行概率计算时,采用隐马尔可夫模型(Hidden Markov Model,简称 HMM)。隐马尔可夫模型是马尔可夫模型的扩展。马尔可夫模型描述的是一个随机过程,而隐马尔可夫模型中有两个随机过程,一个随机过程描述观察值(例如,具体的单词)和状态(例如,该单词可能标注的词类)之间的概率关系,即观察值是状态的概率函数,另一个随机过程描述状态之间(例如,词类标记与词类标记之间)的转移关系。作为外界的观察者来说,只能看到状态产生的观察值,状态之间的转移是隐藏的,所以叫做隐马尔可夫模型。近年来,利用隐马尔可夫模型在词性标注方面取得了较好的结果,推动了基于统计的机器翻译的研究。

基于实例的机器翻译的思想最早是由日本机器翻译专家长尾真(Makoto Nagao)提出来的。他在 1984 年发表了《采用类比原则进行日-英机器翻译的一个框架》一文,探讨日本人初学英语时翻译句子的基本过程。长尾真认为,初学英语的日本人总是记住一些最基本的英语句子以及一些相对应的日语句子,他们要对比不同的英语句子和相对应的日语句子,并由此推论出句子的结构。参照这个学习过程,在机器翻译中,如果给出一些英语句子的实例以及相对应的日语句子,机器翻译系统来识别和比较这些实例及其译文的相似和相差之处,从而挑选出正确的译文。长尾真指出,人类并不通过做深层的语言学分析来进行翻译,人类的翻译过程是:首先把输入的句子正确地分解为一些短语碎片,接着把这些短语碎片翻译成其他语言的短语碎片,最后再把这些短语碎片构成完整的句子,每个短语碎片的翻译是通过类比的原则来实现的。因此,我们应该在计算机中存储一些实例,并建立由给定的句子找寻类似例句的机制,这是一种由实例引导推理的机器翻译方法。

在基于实例的机器翻译系统中,系统的主要知识源是双语对照的翻译实例库,实例库主要有两个字段,一个保存源语言句子,另一个保存与之对应的译文,每输入一个源语言的句子时,系统把这个句子同实例库中的源语言句子字段进行比较,找出与这个句子最为相似的句子,并模拟与这个句子相应的译文,最后输出译文。

基于实例的机器翻译系统中,翻译知识以实例和义类词典的形式来表示,易于增加或删除,系统的维护简单易行,如果利用了较大的翻译实例库并进行精确的对比,有可能产生高质量译文,而且避免了基于规则的那些传统的机器翻译方法必须进行深层语言学分析的难点。在翻译策略上是很有吸引力的。

要进行基于实例的机器翻译需要研究如下问题:

第一,正确地进行双语自动对齐(alignment):在实例库中要能准确地由源语言例句找到相应的目标语言例句,在基于实例的系统的具体实现中,不仅要求句子一级的对齐,而且还要求词汇甚至短语一级的对齐。

第二,建立有效的实例匹配检索机制:很多研究者认为,基于实例的机器翻译的潜力在于充分利用短语一级的实例碎片,但是,利用的实例碎片越小,碎片的边界越难于确定,歧义情况越多,从而导致翻译质量的下降,为此,要建立一套相似度准则(similarity metric),以便确定两个句子或者短语碎片是否相似。

第三,根据检索到的实例生成与源语言句子相对应的译文:由于基于实例的机器翻译对源语言的分析比较粗,生成译文时往往缺乏必要的信息,为了提高译文生成的质量,可以考虑把基于实例的机器翻译与传统的基于规则的机器翻译方法结合起来,对源语言也进行一定深度的分析。

目前世界上的基于实例的机器翻译系统主要有:日本京都大学长尾真和佐藤的 MB T1 和 MB T2 系统:该系统的翻译过程分为分解(decomposition)、转换(transfer)、合成(composition)三步。在分解阶段,系统根据提交的源语言词汇依存树检索实例库,并利用检索到的实例碎片来表示该源语言句子的依存树,形成源匹配表达式;在转换阶段,系统利用实例库中的对齐信息将源匹配表达式转换成目标匹配表达式;在合成阶段,将目标匹配表达式展开成为目标语言词汇依存树,输出译文。美国卡内基-梅隆大学的多引擎机器翻译系统

(Multi-engine Machine Translation) PANGLOSS 系统:这个系统的主要引擎是基于知识的机器翻译系统,基于实例的机器翻译系统只是它的一个引擎,为整个多引擎机器系统提供候选结果。日本口语翻译通信研究实验室 ATR 的 ETOC 和 EBMT 系统:ETOC 系统能够检索出与给定的源语言句子相似的实例,EBMT 系统能够利用实例库来消解歧义,这两个基于实例的机器翻译系统还不完整。

我国是继美、苏、英之后,世界上第四个开展机器翻译研究的国家。当今在机器翻译方面居于先进水平的日本,是在 1958 年才开始进行机器翻译的,起步比我国晚。与国外机器翻译的发展情况相比较,我国机器翻译除了有草创期、复苏期和繁荣期之外,由于“文化革命”的影响,还有一个停滞期,而且,由于在理论上和方法上以及设备上的底子都很薄,每一个时期又都比国外机器翻译的同样时期稍微滞后。

(1) 草创期(1956 年—1966 年)。在这个时期,我国学者对机器翻译进行了初步的探索和试验。1956 年,国家便把机器翻译研究列入了我国科学工作的发展规划,成为其中的一个课题,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,中国科学院语言研究所与计算技术研究所合作,开展俄汉机器翻译的研究。1959 年,他们在我国制造的 104 大型通用电子计算机上,进行了俄汉机器翻译试验,翻译了 9 个不同类型的、较为复杂的句子。在这个时期,北京外国语学院、北京俄语学院、广州华南工学院、哈尔滨工业大学也分别成立了机器翻译研究组,开展俄汉或英汉机器翻译的试验。草创期的计算机技术还非常落后,工作时没有可观察翻译过程的屏幕,而靠穿孔纸带来与计算机交换信息,翻译的结果用穿孔纸带上的小孔表示,没有汉字输出设备。前辈们在这样艰苦的条件下进行俄汉机器翻译研究,是难能可贵的。

(2) 停滞期(1966 年—1975 年)。在这个时期,除了极少数的机器翻译研究者在极端恶劣的条件下继续进行理论探索之外,没有进行任何的机器翻译研究和试验。

(3) 复苏期(1975 年—1987 年)。在这个时期,我国机器翻译研究重振旗鼓。1975 年 11 月,在中国科学技术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协作研究组,以冶金题录 5000 条为试验材料,制定英汉机器翻译方案并上机试验。1978 年 5 月,在计算所 111 机上进行抽产试验,抽样 20 条,达到了预期的效果。在这个时期,我国学者还开展了法汉、德汉、日汉以及汉-法/英/日/俄/德多语言机器翻译试验,取得了一定的成效。

(4) 繁荣期(1987 年—现在)。这个时期是以“译星 1 号”机器翻译系统的问世为标志的。继“译星 1 号”之后,“高立英汉机器翻译系统”“863 智能英汉机器翻译系统”等一系列的实用化商品化的机器翻译系统如雨后春笋般地推向市场,接着,天津的“通译”系统和西安的“朗威”系统也开始上市,并且拥有了相当数量的用户。我国的机器翻译终于走出了实验室,迈向了实用化和商品化的阶段。在这个时期,我国也开展了基于统计和基于实例的机器翻译研究以及在互联网(INTERNET)上的英汉机器翻译研究,并取得了初步的成果。

机器翻译的研究在我国有着广阔的市场和美好的应用前景。天津的“通译”计算机软件研究所和北京连邦软件专卖店曾联合组织过一次“机译软件推广月”活动,专门作了一次机器翻译软件的市场调查,全国 40 多家连邦软件专卖店在一个月內共接到用户咨询电话两万人次,销售机器翻译软件 500 多套。由此可见,我国机器翻译软件市场的前景是非常乐观的,机器翻译研究有着巨大的潜力经济效益。

参考文献

- [1] 冯志伟《自然语言机器翻译新论》,语文出版社,1995。
 - [2] 冯志伟《自动翻译》,上海知识出版社,1987。
 - [3] M. Tomita, Generalized LR Parsing, Kluwer Academic Publishers, Norwell, Massachusetts, USA, 1991.
 - [4] B. Merialdo, tagging English text with a probabilistic model, Computational Linguistics, 20(2), 155-171, 1994.
 - [5] J. Hutchins, latest Development in MT Technology: Beginning A New Era in MT Research, Proc. of MT Summit IV, 1993.
- (冯志伟 国家语委语言文字应用研究所, 邮编:100010)