

载《现代语文》，2004年

澄清对机器翻译的一些误解

冯志伟

(教育部语言文字应用研究所)

机器翻译(machine translation)是使用电子计算机把一种语言(源语言, source language)翻译成另外一种语言(目标语言, target language)的一门新学科。这门新学科也同时是一种新技术。它涉及到语言学、计算机科学、数学等许多部门,是非常典型的多边缘的交叉学科。在语言学中,机器翻译是计算语言学的一个研究领域;在计算机科学中,机器翻译是人工智能的一个研究领域,在数学中,机器翻译是数理逻辑和形式化方法的一个研究领域。机器翻译要把这些不同的学科结合起来,综合地进行研究。机器翻译要求不同学科的专家通力合作,取长补短,相得益彰。机器翻译是语言现代化的一个重要方面,目前机器翻译产品虽然不能达到很高的质量,但是,它可以帮助人们及时地获取信息,克服语言的障碍。

然而,不少人对于机器翻译存在着一些误解,我们在这里有必要澄清这些误解。

有人说,“既然计算机可以做翻译,那么,为什么不用计算机把曹雪芹的小说《红楼梦》翻译成英文,或者把莎士比亚的剧作和十四行诗翻译成中文?”

事实上,机器翻译不可能、而且从来也不想翻译这样的世界文学名著。文学翻译要求很高的文学修养,它决不是普通的翻译人员所能达到的;而且,在整个翻译事业中,文学翻译只是很小的一个部分,不能因为机器翻译不能进行文学翻译就说机器翻译一无所能。要求机器翻译系统翻译世界文学名著,正如要求工业机器人表演芭蕾舞剧《天鹅湖》一样,是不应该的,也是没有必要的。众所周知,尽管工业机器人不能表演《天鹅湖》,但是它在其他方面还是非常有用的。

当然,文学作品的翻译是一项充满艰辛却又令人神往的工作,它与任何其他需要人类丰富创造力的领域一样,都需要耗费大量的人力。我们来研究下面摘自十八世纪曹雪芹的小说《红楼梦》中第四十五回末尾处的一段文字,英文翻译来自文学翻译家霍克斯(David Hawkes)的英译本:

“黛玉自在枕上感念宝钗……”

又听见窗外竹梢焦叶之上,雨声淅沥,清寒透幕,不觉又滴下泪来。”

As she lay there alone, Dai-yu's thoughts turned to Bao-chai... Then she listened to the insistent rustle of the rain on the bamboos and plantains outside her window. The coldness penetrated the curtains of her bed. Almost without noticing it she had begun to cry.

黛玉自在枕上感念宝钗...
Dai-yu alone on pillow top think-of-with-gratitude Bao-chai
又听见窗外竹梢焦叶
again listen to window outside bamboo tip plantain leaf of
之上,雨声淅沥,清寒透幕,
On-top rain sound sign drip, clear cold penetrate curtain,
不觉又滴下泪来。
not feeling again fall down tears come.

我们来考虑一下在这种文学翻译中我们需要涉及哪些问题呢？首先是由于曹雪芹常常采用双关语给人物命名而引起的中文人名的翻译问题。霍克斯的选择是对其中的主要人物都采用构成人名的汉字的音译，而对那些仆人的名字则采用意译（比如 Aroma[袭人]、Skybright [晴雯]）。其次中文几乎没有动词时态和语态的变化，因此霍克斯不得不决定将中文“透”翻译为 *penetrated*，而不是 *was penetrating* 或 *had penetrated*。霍克斯还在 *window* 的前面添加了物主代词 *her* 使得 *her window* 比 *the window* 更适合表达那种安静闲适的卧室气氛。为了使不熟悉中国床帷的英文读者能清楚地理解，霍克斯将“幕”翻译为 *curtains of her bed*。最后，短语“竹梢蕉叶”的中文是非常优雅的，这种四字短语是有文化品位的标志，但是如果以词对词的方式翻译为英文，就很糟糕了，因此霍克斯只是简单地将它翻译为 *bamboos and plantains*。这些都反映了霍克斯高超的文学翻译技巧。

显然这类翻译要求我们对源语言和输入文本具备博大精深的理解，同时也需要我们能够老练地、富有诗意地、创造性地支配目标语。因此，将任意文本自动从一种语言生成另外一种语言的高质量的翻译问题是难以自动实现的。

但是当前的机器翻译系统对一些较简单的翻译任务还是可以胜任的。比如，机器翻译系统的应用常常集中完成下列翻译工作：

- (1)那些粗略翻译就足够的工作；
- (2)那些人工译后编辑（*post-editing*）可用于提高机器翻译输出质量的工作；
- (3)能够产生高质量译文的受限子语言领域的翻译工作。

网络中信息的获取是一种“信息采集”的任务，只要有非常粗略的翻译，读者就会感到满意。我们来看下面摘自 SYSTRAN 机器翻译系统的英汉机器翻译的英文原文和汉语译文：

SYSTRAN Translation from English to Chinese (翻译日期：11-March-2004)

英文原文：

Austrian Airlines Group has announced that it will include Shanghai in its flight service network beginning from April 29th with the operation of three flights a week from Pudong International Airport to Vienna. At the same time the number of flights from Vienna to Beijing will also be increased from four to six this summer. According to a company official, 120,000 Chinese tourists visited Austria last year, and the figure is expected to rise 20 percent annually.

汉语译文：

奥地利航空公司小组宣布，它包括上海在它的飞行服务网络起点从4月29日以三次飞行的操作每星期从浦东国际机场到维也纳。同时飞行的数量从维也纳的向北京并且将被增加从四到六这个夏天。根据公司官员，120,000个中国游人去年访问了奥地利，并且图被预计年年上升百分之20。

这样的机器翻译的译文是大概可以看懂的，从这种粗糙的译文中，我们可以知道奥地利航空公司将增加从上海到维也纳的航班，这个消息对于那些打算到奥地利去旅游的人肯定是会有帮助的。尽管在汉语的译文中，把英语 *figure* 这个多义词错误地翻译为“图”，而正确的翻译应当是“数字”，译文中还有不少的语法错误，但是，译文整体的含义还是清楚的。译文的信息对我们是有用的。

顺便说一下，将机器翻译技术用于文本查询还可以通过更有效的**跨语言信息检索**（cross-language information retrieval）技术来实现，这种技术所关注的是如何采用一种语言的检索词来查询另外一种语言的文本。

粗略翻译还可以作为完整的翻译过程的第一个阶段。机器翻译系统可以先生成翻译草稿，然后由翻译工作者进行译后编辑处理。即使是一个草稿有时也会加快整个翻译处理的进程。严格的说，采用这种方式的系统是**机助人译**（computer-aided human translation，简称为CAHT或CAT）而不是（全自动）机器翻译。机器翻译的这种使用模式特别适合于翻译工作量大又需要快速更新的领域。例如，软件手册的**本地化**（localization）翻译，关于市场变化的财经新闻的翻译，等等。

天气预报是机器翻译在**子语言**（sublanguage）领域的一个实例，在这个领域的机器翻译模型足以产生可以直接使用的原始的机器翻译输出，甚至不用做任何译后编辑。天气预报包括的短语有：Cloudy with a chance of shower today and Thursday（今天和周四多云转阵雨），Low tonight 4（今晚最低4度），high Thursday 10（周四最高10度）以及 Outlook for Friday: Sunny（展望周五：晴），等等。这个领域的翻译内容一般只包括有限个词汇和一些基本的短语类型，歧义很少，而且歧义单词的意义也可以基于局部上下文采用词类和语义特征而得到消除。常用的语义特征有：MONTH（月）、PLACE（地点）、DIRECTION（方向）、TIME POINT（时间点）、TIME DURATION（时间跨度）、DEGREE-OF-POSSIBILITY（可能的程度）等。其他可以作为机器翻译子语言的领域还有：技术设备维护手册、航空旅行查询、约会安排以及饭店推荐等方面的翻译。

有人说，美国的机器翻译系统曾经把英语翻译成俄语，然后又把俄语翻译成英语，最后翻译出来的英语是语无伦次的，结果闹出了大笑话。机器把英语的成语The spirit is willing, but the flesh is weak（心有余而力不足）翻译成俄语之后，这个句子的意思却变成了The whiskey is alright, but the meat is rotten（威士忌没有问题，而肉却烂了）；机器把英语成语的Out of sight, out of mind（眼不见心不烦）翻译成俄语之后，这个句子的意思变成了Invisible idiot（看不见的白痴）。有人说，一个英法机器翻译系统把英语的科技术语 hydraulic ram（水泵）翻译成法语，其意思却是water goat（水山羊）。因此，他们讥讽机器翻译是低能的，是没有什么用处的。

关于把英俄机器翻译翻译的这个传说是毫无根据的，也是不符合事实的。据我所知，这样的传说最早来自60年代美国的《哈泼杂志》。

1962年8月号的《哈泼杂志》(Harper's Magazine)发表了一位叫做古温豪芬(John A. Kouwenhoven)的美国人撰写的题目为《翻译的困扰》(The trouble with translation)的文章，文章中编造了如下的故事：

有几个电子工程师设计了一部自动翻译机，这部机器的词典包含1500个基础英语词汇和相对应的俄语词汇。他们宣称这部机器可以马上进行翻译，而且不会犯人工翻译的错误。第一次试验时，观众要求翻译“Out of sight, out of mind”（眼不见心不烦）这个句子，灯光一阵闪动之后，翻译出来的俄语句子的意思竟然是“看不见的疯子”（Invisible idiot）。他们觉得这样的谚语式的句子比较难于翻译，于是又给机器翻译翻译另一个出自圣经的句子“The spirit is willing, but the flesh is weak”（心有余而力不足），机器翻译出来的俄语句子的意思却是“酒保存得很好，但肉已经腐烂”（The liquor is holding out all right, but the meat has spoiled）。

这样的故事显然是这位古温豪芬先生挖空心思捏造出来的。稍微有一点机器翻译历史常识的人都知道，在60年代，美国只研究过把俄语翻译成英语的机器翻译系统，根本没有研究

过把英语翻译为俄语的机器翻译系统，古温豪芬说机器翻译曾经把英语翻译成俄语，完全是无中生有的无稽之谈。这显然是古温豪芬对于机器翻译的别有用心攻击。古温豪芬的行为已经成为了机器翻译发展史上的一段丑闻。40多年过去了，我们怎么还能相信这样的丑闻呢？

至于有人说某个英法机器翻译系统把习用的术语“hydraulic ram”错误地按字面意义分解而发生翻译的失误，正是机器翻译力图解决的问题。这样的问题可以通过建立机器可读的术语词典来解决。不能因为这样的失误而得出“机器翻译没有用处”的结论。

有人说，“机器翻译的译文质量实在太低劣了，恐怕在实际上派不上什么用处。”

当今机器翻译系统的存在是一个不可否认的事实。尽管有人认为机器翻译派不上什么用处，可是，机器翻译系统却时时刻刻被人们使用着。不论你喜欢它也好，不喜欢它也好，机器翻译系统总是存在着的，也总是在运转着的。这样的机器翻译系统不会因为你不喜欢它而自行消失，也不会因为你喜欢它而马上达到很高的质量。位于加拿大Dorval的加拿大气象中心的METEO机器翻译系统，从1977开始就进行英语和法语的气象预报资料的机器翻译，从来没有中断过。例如，在1990年，METEO机器翻译系统每天能够把45000单词的英文气象资料翻译成法文，并把翻译结果传送到出版、广播和电视部门公之于众。在80年代，内燃机引擎制造厂Perkins Engines由于使用了机器翻译系统WEIDNER来翻译内燃机引擎说明书，从而大大节省了说明书翻译的成本和时间，提高了这个内燃机引擎制造厂的生产力。这些，都有力地说明机器翻译对于科学技术翻译的某些局部的领域的翻译是行之有效的，是大有可为的，机器翻译确实是科学技术翻译的重要手段之一。现在，世界上运行着的机器翻译系统除了METEO之外，还有欧洲联盟正在使用的SYSTRAN系统，此外，LOGOS系统、ALPS系统、ENGSPAN系统、SPANAM系统、METAL系统、GLOBALINK系统以及我国的“华建”系统和“译星”系统也正在运行之中。我们怎么能够说机器翻译系统派不上用处呢？

有人说，“如果机器翻译成功了，那么，翻译人员就只好失业了。”

翻译是一个很重要的产业。翻译的需求与日俱增，翻译人员每天都有干不完的活，根本不可能失业。机器翻译的质量虽然不理想，但是，它可以很快地帮助翻译人员查出文献中的生词，它可以帮助翻译人员核对术语的一致性，它应该是翻译人员的可靠的助手和亲密的朋友。我们应该探索翻译人员与机器翻译合作共事的翻译模式，把人的翻译和机器翻译结合起来，取长补短，相得益彰。

所以，尽管机器翻译的译文质量至今还不完美，但是，机器翻译是有用的，机器翻译系统的存在已经是一个不可争议的事实。我们应该解除误解，正视事实，用正确的态度来对待机器翻译。

现在我们已经进入了信息化的时代，语言是信息的最主要的负荷者，如何有效地使用现代化手段来突破人们之间的语言障碍，成为了全人类面临的共同问题。机器翻译采用电子计算机来进行不同语言之间的自动翻译，是解决这个问题的有力手段之一。

不论在社会政治上、商业上、科学技术上来说，机器翻译都无疑是一个重要的学科。

从社会政治上来说，语言对于人类交流思想的重要作用是无庸置疑的。为了克服语言的障碍，曾经有人提出使用人类通用语言（lingua franca）来替代各种不同语言的想法。但是，这样的想法显然是很难实现的，就是有了这样的通用语言，它也代替不了各种不同的语言。因为语言是民族文化的象征，放弃民族语言就意味着放弃民族的文化，如果全人类都将一种通用的语言，各具特色的、丰富多彩的民族文化也就黯然失色了。这显然不是一件好事。尽管英语在英国和美国占绝对的统治地位，在英国的威尔士，人们还在讲威尔士语，在美国的

某些地区，还有很多人在讲西班牙语。至于像加拿大和瑞士这样的双语和多语国家，像欧洲联盟和联合国这样的组织，多语言的使用（multilingualism）已经成为了日常生活中的基本原则和普遍现象。而多语言的使用，不同语言之间的翻译就非常重要了。翻译的需求会变得越来越迫切和尖锐。

我们以欧洲联盟（以下简称欧盟）的多种语言翻译为例来说明这个问题。欧盟作为一个共同体，成功地建立了统一大市场，促进了各成员国的经济大发展；启动了欧洲统一货币，形成了横跨欧洲大陆的“泛欧元区”，足可以同美元抗衡；基本实现了共同边界，建立了共同外交和安全政策；现在正在讨论制定欧洲宪法，进一步加速欧洲经济、政治一体化的进程。但是语言问题一直困扰着这个国际组织，而且至今还没有应对的良策。按照欧盟的前身欧洲经济共同体 1957 年首脑会议通过的关于语言多样化的决议，各成员国都有平等使用本国语言的权利，各成员国的官方语言即成为这个组织的工作语言。当时只有 6 个创始国，开会时只使用 4 种语言，语言问题还不十分突出。1995 年欧盟扩大到 15 个国家，官方语言增加到 11 种（比利时、奥地利没有自己的独特的语言，爱尔兰官方语言为英语、爱尔兰语，卢森堡因其官方语言为德语、法语和卢森堡语，没有要求使用卢森堡语），语言问题开始突显出来。按照规定，在召开欧盟的“三驾马车”欧洲议会、欧盟首脑会议、欧盟委员会等正式会议时，这 11 种语言就是会议的工作语言。也就是说，欧盟在召开正式大会时应当提供 11 种语言的翻译服务，大会文件要译成 11 种语言，由此增加的翻译工作量之大，财政支出之高是可以想见的。据统计，欧盟每个正式会晤至少需要 33 个翻译人员，才能完成 11 种语言间高达 110 种可能的互译。据媒体报道，每年欧盟大小会议共 1 万多个，为此要付出 15 万个翻译人/日，全年翻译的文件达 113 万页以上。每次为期 5 天的欧洲议会起码需要 450 名翻译人员。目前欧盟雇用的常年翻译有 460 人，临时翻译 1500 多人。即便如此，仍不能满足所有成员国的所有语言要求，这支庞大的翻译队伍如今越来越难以应付巨大的工作压力，已陷入捉襟见肘和穷于应付的困境。据报道，2003 年欧盟的文件翻译总量有 148 万页之多。不用说，翻译方面的支出也是巨大的，据媒体披露，整个欧盟目前一年用在翻译方面的支出高达 5.5 亿欧元，每年用于翻译的支出约占全部行政管理预算的一半左右。语言的困扰随着欧盟成员国的增加而愈加尖锐：一方面，所有成员国的代表当然有权使用本国的语言表达各自的意见，另一方面，由此增加的翻译工作量和财政支出却并非是所有成员国都乐于承担的。从 2004 年 5 月 1 日开始，欧盟成员国将扩大到 25 个，这就意味着欧盟正式会议的工作语言将至少有 20 种，一场会议的同声传译至少需要 60 名翻译人员。欧盟每天都有七八百名翻译人员在进行同声传译工作。现在欧盟就因同声传译跟不上，笔译错误层出不穷而叫苦不迭，一旦 20 种官方语言同时互译，其后果将更加混乱，而用于翻译方面的财政支出将大幅度提高，据报道，2004 年由于 10 个新成员国的加入，文件翻译总量将达到 206.5 万页，2005 年更是将高达 237 万页，其支出将增至 8.08 亿欧元，对欧盟各成员国无疑会增加更大的负担，在语言方面带来更多的矛盾与困扰。以至于在 2004 年 5 月，欧盟不得不通过一项新的议案，规定所有欧盟成员国在欧盟会议上的文件都不得超过 15 页 A4 号纸，以减轻翻译人员的工作量。在我们这个多语言的世界里，翻译的重要性是不言而喻的，而随着人类交际活动的发展，翻译的数量越来越多，翻译的需求将越来越迫切，人工的翻译已经远远满足不了社会的需求，在这种情况下，机器翻译必然会作为人工翻译的一个重要的补充而发展起来。

从商业上来说，在这个多语言的世界里，翻译对于推销商品的重要性是众人皆知的。如果中国的工业产品在美国市场上销售，美国人当然希望这种产品的说明书是用英文写的，而不是用中文写的。如果德国的药品在中国市场上销售，中国的顾客当然也希望这种产品的说明书是用中文写的，而不是用德文写的。翻译是一种高智能的劳动，它要求熟练的翻译技巧，它需要丰富的语言知识和专业知识，因此，翻译的开销是很高的。如果采用机器翻译来减轻翻译人员的负担，提高翻译工作的效益，它的经济价值也是很高的。根据国际权威机构对于

世界翻译市场的调查显示，翻译市场的规模在 2005 年将达到 227 亿美元，而中国的翻译市场将达到 200 亿人民币。随着因特网应用范围的扩大和国际电子商务市场的日渐成熟，到 2007 年，只是网页的翻译业务将达到 17 亿美元的规模。目前，我国翻译能力严重不足，我国翻译市场的规模尽管已经超过了 100 亿人民币，但是现有的国内翻译公司只能消化 10% 左右，由于无法消化大量从国际上传来的信息流，我们的信息不灵，就有可能使我们在国际竞争中失去大量的商业机会。传统的人工翻译已经难以满足实际的需要，在这种情况下，机器翻译将可能改变我国这种翻译能力严重不足的局面。

从科学技术上来说，机器翻译是计算机科学、数学和语言学的一个应用领域，机器翻译的思想，对于这些学科的发展会产生重要的影响。例如，著名的逻辑程序语言 Prolog 的基本原理，就是他的设计人柯尔迈罗埃（A. Colmerauer）在研究用于机器翻译的“Q-系统”

（Q-system）的基础上奠定的。机器翻译的理论和技术的，成为了推动这些学科发展的取之不尽的源泉。

总而言之，机器翻译使用计算机来进行不同自然语言之间的翻译，它是自然语言计算机处理的一个历史悠久的部门，是横跨语言学、数学、计算机科学的综合性学科，也是信息时代语言应用的一个重要领域。随着计算机网络的迅速普及和推广，随着信息高速公路的发展，网络上不同语言之间交际越来越普遍，语言的障碍也显得越来越严重，机器翻译是克服信息时代的语言障碍的不可缺少的手段，它在现代信息社会中的巨大作用将会越来越明显。由于自然语言的极端复杂性，机器翻译是当代科学技术的十大难题之一。这是一个迅速发展中的新学科，我们应该支持它，爱护它，大力促进它的发展。目前机器翻译的质量仍然还不能令人满意，要实现高质量的机器翻译，我们确实还要进行艰苦的探索。不过，无论如何，机器翻译已经从人们的梦想逐步变成活生生的现实，这是令我们感到振奋的。机器翻译随着计算机的诞生而诞生，它也将随着计算机的发展而发展，只要有计算机存在，机器翻译的研究就会存在。机器翻译永远是一个与计算机共生共存的研究领域。