

机器翻译不仅仅是机器+翻译

—读《机器翻译研究》

刘海涛¹

《机器翻译研究》（以下简称“机译”），冯志伟著，中国对外翻译出版公司，2004年出版，841页。

本书是作者继《自动翻译》（知识出版社，1987，与杨平合著）和《自然语言机器翻译新论》（语文出版社，1995）后，有关机器翻译的又一部著作。和前面两部著作相比，本书最引人注目的是它的篇幅。按照作者冯志伟先生的说法，这是他自己近20部著作中，最放的开的一部。所谓放的开，就是有些问题，在作者过去的著作中由于篇幅和其它方面的限制，无法向读者说透，而在这本书里它们都得到了比较详尽的描述。但详尽，并不意味着艰深。本书的风格仍然是深入浅出、通俗易懂。

本书是“十五”国家重点图书《翻译理论与实务丛书》之《科学翻译研究卷》的一种。这可能意味着机器翻译系统不再被职业翻译视为会抢自己饭碗的坏东西，而是能帮助自己提高工作效率的伙伴了；同时机器翻译研究也有助于人们对于翻译过程本身的认识，其中翻译过程的形式化描述对于翻译的教学和实践都是具有指导意义的。因此，我们可以理解这部一般属于计算语言学中的机器翻译著作为什么会出现在有关“翻译理论”的书系中来。而将其归为“科学翻译”之一种，也说明翻译界对于目前机器的能力是了如指掌的。

本书封底上的介绍文字说“（本书）系统介绍机器翻译的历史和现状、各种类型的机器翻译系统、机器翻译的实现过程。通过大量的实例，有血有肉地描述计算机分析自然语言的过程，使读者对机器翻译有更加清楚而具体的认识。”对于一位科技翻译工作者，这段话的意义何在呢？一个科技翻译工作者为什么要理解一个机器工作过程呢？而且还要对这一过程有清楚和具体的认识呢？我个人认为似乎可以从以下几方面来理解这些问题：

1. 大量实践证明，机器翻译技术可以成为（科技）工作者的得力助手，可以帮助人们翻译的更好、更快；
2. 就目前的人类的技术和认知水平而言，只能是构建一种可称之为“人机智能综合体”的机器翻译系统，这也是我们今天可以见到许多系统被冠之以“人助机译”或“机助人译”的原因；
3. 既然任何“灵智”的机器翻译系统均需要人工参与，那么作为“人机智能体”一部分的人，当然有必要更多地了解他的同伴的工作原理和其它秉性；
4. 翻译工作者对于计算机分析处理语言过程的了解，不但有助于提高自己的工作效率，也有助于研究开发更好的机器翻译系统。

¹ 中国传媒大学应用语言学系教授

因此，翻译工作者了解一些机器翻译的情况，只能是有百利而无一害。

接下来，我简单介绍一下本书的内容，供大家参考。

在题为“机器翻译概述”的第一章里，作者分三节介绍了：什么是机器翻译？人们对机器翻译的一些误解以及机器翻译的简单历史等。因为，本书是一本关于机译的原理的著作，所以 30 多页的篇幅分配的还是比较合理的。通过这一部分的学习，读者基本可以清楚机器翻译的目的和大致过程，对于国内外机译系统的发展也可做到心中有数。科技翻译工作者会欣喜地发现机器翻译最适合的领域就是我们一般所说的科学翻译领域。

按照实现原则的不同，可以把机器翻译系统分为不同的类型。这一方面的内容，是第二章的主题。作者将机译系统分为基于规则的系统、基于语料库的系统，应该说基于规则的系统还是目前实用系统的主流，但基于语料库的系统，其中包括基于统计和基于实例的机译系统，是目前机译研究的主流。为了扬长避短，也出现了把规则方法和语料库方法结合在一起的所谓多引擎机译系统，本书对这种类型的机译系统也做了介绍。口语机器翻译也是目前的一个研究热点，为了让读者对此领域有所了解，作者介绍了 Verbmobil 等几个主要的口语机译系统。这一章还包括了“因特网与机器翻译”和“机器翻译与翻译记忆”两节，严格说来，这两方面的内容似乎和机译系统的类型无关，但鉴于因特网对于人类生活的重要性，目前许多机译系统都希望在因特网上占有一席之地。虽然从实现机译的关键技术看，用于因特网的机译和用于其它领域的系统不会有太大的区别，但其应用也有自己的一些特点，因此有理由将其列为一种特殊的类型。通过第一章的内容，大家可以感觉到在可预见的将来，全自动高质量的机译系统只能是一种理想的。因此，人的参与是必不可少的，而按照人的参与程度，又衍生出来不同的产品，“翻译记忆”就是近年来用的较为广泛的一种技术。正如作者所说的那样“由于目前还没有一种机器翻译产品的效果能让人满意，对于专业翻译来说，翻译记忆技术几乎是唯一的选择。”(65 页)由此可见，翻译记忆对于专业的翻译工作者是非常重要的技术。

鉴于目前大多数实用系统仍然是基于规则的，所以在题为“机器翻译的实现过程”的第三章里，作者比较详细地介绍了基于转换的机译系统的实现过程。为了让读者对于目前机译系统的翻译水平有所了解，作者首先介绍了自己对国内的三个基于规则的英汉机译系统用 1100 个句子进行测试的结果。作者认为“机器翻译已经达到了一定的水平，尽管译文质量还有问题，但是，基本上还是可以读懂的。”(74 页)在进入主题之前，作者又介绍了法国机译专家沃古瓦的“机器翻译金字塔”，通过这个金字塔读者可以直观地了解基于规则的三种主要方法：直接法、转换法和媒介语法。其中用的最多的是基于转换的方法，简称转换法。作者将转换法的实现过程分为以下步骤：特征和特征值的定义，词典编写，形态分析，句法分析，逻辑语义分析，双语转换，目标语言生成。在讲这些步骤时，作者不是只说一些干巴巴的理论，而是以一个英汉系统为例，对无论是词汇还是规则都用具体的例子做了描述。这对于读者理解这些原本比较抽象的东西，很有好处。

在 800 多页的篇幅里，前三章只占了 100 页。由此可见，在篇幅分配上，作者的重心放在了有关机译的原理和实现技术方面，这与书名中所含的“研究”二字是吻合的。

从第 103 页到 466 页，作者较为详细地介绍了形态自动分析、基于转移网络的自动句法分析、基于上下文无关语法的自动句法分析、基于特征结构的自动句法分析、基于依存语法的自动句法分析和基于概率的自动句法分析。除“形态分析”一章外，这一部分的主要内容都是有关自动句法分析的，为什么在一部研究机器翻译的专著中，要用如此之多的篇幅来研究句子的自动分析呢？计算语言学的实践证明，无论你做什么应用，自动句法分析都是最重要的环节，是绝大数机器处理自然语言的基础，因此冯志伟曾在此前的著作中，用“剖析”来翻译英文中的“Parsing”，并用此术语来专指“自动句法分析”。关于剖析，德国计算语言学家 Hans Uszkoreit 有如下名言“*If you can walk, you can dance. If you can talk, you can sing. If you can parse, you can understand.*”由此可以看出，剖析的重要性怎么强调都不过分。可以说，作者在这一部分不但讲了传统的基于转移网络和上下文无关语法的剖析技术和常用算法，也对基于特征结构、基于依存语法和基于概率的剖析技术做了详尽的介绍。按照我自己的了解，这一部分内容是国内目前最详尽的有关剖析技术的著作。

句法分析之所以重要，在于它是语义分析的基础，也就是说，剖析是理解的基础。在第 10 章“语义自动分析”中，作者不但介绍了一阶谓词、语义网络、概念依存、框架等意义的形式化和表示方法，也对知识本体、词网和框架网络做了介绍，这些内容基本上涵盖了目前语义自动处理的主要技术。

我们说过计算语言学的历史可以视为一部与歧义做斗争的历史，机器翻译作为计算语言学中最早的应用和所含领域最多的应用，它所涉及的歧义不仅仅有单语歧义，还有语际转换（双语）歧义。作者单立“词义排歧”一章，可说是明智之举。这一章又分为两节：英语的词汇歧义和几种重要的词义排歧方法。为什么在一本汉语的著作里要讲英语的词汇歧义？而不拿汉语来说事呢？我们考虑可能有这样两个原因使作者做出了这样的选择：因为本书所举的机译系统例子，大多是外译汉，特别是英译汉的系统，所以研究英语的词汇歧义可能更容易说明问题；二是随后介绍的排歧方法许多都是针对英语的自动处理提出来的，所以便于保持前后一致，也有助于读者对这些方法的理解。如果作者也能更多地谈一下结构歧义的问题，这一部分就更丰满了。

翻译过程是一种分层次的转换过程，一篇文章的层次结构一般为：篇章→句段→句子→子句→短语→词→词素，层次越低，翻译难度就越小。但语言的不同性决定了我们不能总在最低的层次上进行翻译，而语言的模糊性和歧义性又迫使我们把低一层次的成分放到更高一级的层次去考虑。就翻译的理想状况来说，当然最好是能在源语语篇与目标语语篇之间达到等值。因此，语用和篇章也是一个应该考虑的问题。第十一章“语用的自动分析”介绍了“所指判定”、“文本连贯”和“修辞结构理论”等方面的内容。结合形态、句法、语义和语用等层面，对源语进行较全面的自动分析，对于后续的过程是非常有用的。

第十三章“机器翻译”中的转换含有三节：语言的相似性和差异性，从机器翻译看汉语的特点，词汇转换和结构转换。这一部分用大量的例子说明语际转换的困难性和问题，如：词汇选择的差异、形态转换的差异、句法结构的差异、题元关系的差异、中心语的改变、词类范畴的差异等。作者根据自己多年

在机器翻译和计算语言学领域的研究和实践，提出了汉语不同于印欧语言的五个特点：缺乏形态，语法单位界限模糊，词类和句法成分没有明确的一一对应关系，句子成分和语义关系也没有明确的一一对应关系，书面语没有分词连写。这些问题的明确提出，对于进一步提高汉语自动分析的精度是很有益处的。笼统地说，翻译中的语际转换包括词汇转换和结构转换。作者介绍了这两种转换的一般方法。但我们认为，语际转换应该是翻译形式化研究或机器翻译中的一个重点，因为无论是句法还是语义的自动分析都是为语际转换服务的。换言之，应该有一种连贯的翻译形式化理论，把源语分析和语际转换紧密地联系在一起。好在这一方面的内容，读者可以在冯志伟先生的《自然语言机器翻译新论》中读到。我个人认为，国内今后应该加强翻译形式化理论的研究，这种理论不但有助于机器翻译的研究，也会加深人们对翻译过程的理解。

目标语生成是机译过程的最后一个环节。在第十四章里，作者首先介绍了一般的自然语言生成系统的结构，后又在机器翻译中的生成一节向读者介绍了系统功能语法、功能合一语法和意义文本理论，并探讨了这些语言理论对于生成的价值和意义。

词典不但是人进行翻译必不可少的工具，对机器翻译更为重要。考虑到今天语言学研究中的词汇化趋势，越来越多的信息进入词典。第 15 章中的内容有：分析词典中语言信息的描述方法，词典查询结果的排歧，转换词典和生成词典等。由此可见，构造机器使用的词典要比编写面向人类的词典要繁复的多，因为许多对人而言显而易见的东西，必须都要给机器交代清楚。

第 16 章“因特网上的翻译工具”向大家介绍了一些通过网络可以使用的机器翻译系统，使得读者在阅读本书后，马上就可找一些系统来进行试验，这对于读者下一步选用机译系统当然是很有好处了。

至此，全书的主要内容就结束了。但附录所载的“关于机器翻译和计算语言学的对话”，也是很有价值的一篇文献。它不但记录了冯志伟先生个人研究机译和计算语言学的历史，也可视为是一部我国计算语言学和机译研究的简史。书的最后是 12 页的参考文献，相信对于读者进一步学习是有帮助的。但遗憾的是书末没有索引，这不利于读者在一部 800 多页的著作中，快速发现自己需要的东西。

我们认为内容丰富、选材新颖、讲解透彻是本书的主要特点。作者多年来从事机器翻译和计算语言学的研究工作，亲自实现过几个机器翻译系统，这些经验使得本书不仅仅具有理论上的价值，而且也可作为机器翻译系统开发的实践参考书。

机器翻译是多学科交叉、融合的产物。本书在写作中考虑到了跨学科读者的需要，特别是文科背景读者的需要，文体通俗易懂，内容深入浅出，所有这一切使得本书对于从事（科技）翻译、语言学、计算机、自然语言处理等领域的研究者都有重要的参考价值。