



“十五”国家重点图书

翻译理论与实务丛书

罗进德 主编

TRANSLATION THEORY  
AND PRACTICE SERIES

科学翻译研究卷  
李亚舒 黄忠廉 主编

STUDIES OF  
SCI-TECH TRANSLATION

# 机器翻译研究

冯志伟 著

中国对外翻译出版公司

---

**图书在版编目(CIP)数据**

机器翻译研究/冯志伟著.——北京:中国对外翻译出版公司,  
2004.12

(翻译理论与实务丛书·科学翻译研究卷)

ISBN 7-5001-1274-2

I. 机... II. 冯... III. 机器翻译 研究 IV. H059

中国版本图书馆 CIP 数据核字(2004)第 117798 号

---

---

出版发行/中国对外翻译出版公司

地 址/北京市西城区车公庄大街甲 4 号(物华大厦六层)

电 话/(010)68002481 68002482

邮 编/100044

传 真/(010)68002480

电子邮件/ctpc@public.bta.net.cn

网 址/www.ctpc.com.cn

责任编辑/章婉凝

责任校对/立 序

封面设计/耕者设计工作室

印 刷/北京奥鑫印刷厂

经 销/新华书店北京发行所

规 格/850×1168 毫米 1/32

印 张/27

版 次/2004 年 12 月第一版

印 次/2004 年 12 月第一次

印 数/1-3 000

---

ISBN 7-5001-1274-2/H·385 定价:54.00 元



版权所有 侵权必究

中国对外翻译出版公司

# 为中国第四次翻译高潮贡献精品

## 翻译理论与实务丛书

### 总 序

中国历史上第四次翻译高潮正在神州大地蓬勃推进,方兴未艾,景况壮观。

中国历史上出现过三次翻译高潮:东汉至唐宋的佛经翻译、明末清初的科技翻译和鸦片战争至“五四”的西学翻译。而目前这一次的翻译高潮,无论在规模上、范围上,还是在质量水平和对中国社会发展的贡献上,都是前三次翻译高潮无法比拟的。

这一次翻译高潮的出现,首先是全球信息时代降临的结果。信息爆炸、知识爆炸,同时也就是翻译爆炸。在今日“地球村”,离开翻译谈论知识信息,是不可思议的。同时,这次翻译高潮的出现又是以中国推行改革开放新政,走上社会主义市场经济的现代化强国之路为契机的。历史潮流滚滚向前。在人类高奏和平发展的大乐章中,翻译无疑是其中不可或缺的旋律。在中华民族复兴的大进军里,翻译必然是一支活跃的先头部队。

翻译,无论是作为文化现象、思想运动,还是作为一项职业、一种知识技能,总与所处的时代背景密不可分。翻译的观念、方

法、样式、标准、风格,无不与时俱进。观察一下世界范围和中国国内的翻译实践活动和理论研究,不难得出结论:翻译确实是人类精神文明中最富活力、最敏锐的领域之一。信息时代和市场经济,决定了目前这次翻译高潮最突出的特点。较之前三次翻译高潮,这一次的翻译高潮信息量更庞大,涵盖面更广泛,题材体裁更丰富多样,方式更灵活便捷,技术装备更先进,从业人员更众多,受益者更普遍,理论研究更活跃,人才培养更具规模。如果从经济学观点看,翻译作为信息产业之一支,现在所产生的经济效益也是历史上空前的。如果说前三次翻译高潮,都是外文译入中文为主,那么这一次翻译高潮则正在改变中国在翻译上的“入超”地位,对外介绍中国和外部了解中国的迫切需要,对中译外提出了更高的要求。

在新世纪的门槛上,我们欣喜地看到,翻译作为一门社会迫切需要、大有用处的知识技能,翻译学作为一门内涵深广的人文学科,在最近二十年的新时期有了长足发展。学翻译、教翻译、研究翻译、评论翻译和从事翻译职业的实践,已经成为与对外开放同步前行的社会文化热点之一,持续升温。翻译教学和译学理论研究,其规模之大、阵容之强、水平之高、成果之丰都是前所未有的,正引起国际翻译学术界的瞩目。中国作者的名字和作品在国际权威译学刊物上频频出现,正从一个侧面反映着我们的成就。翻译教学已经从语法为纲的语法复制型翻译模式中解放出来,已经分清了教学翻译和翻译教学这两个本不应混淆却曾长期混淆的基本概念,大学翻译课不再是外语教学的手段,而是在双语知识基础上培养口笔翻译技能的高级课程。翻译理论研究继文艺学、语言学之后,又有了美学、语篇分析、社会符号学、语言文化比较研究等多角度研究翻译的方法或理论工具,为确立翻译学作为一门以翻译为研究对象的,开放的,跨学科的人文科学的地位奠定基础。有关翻译的知识大大地丰富和深化,使越来越多的人认识到:翻译及其

理论研究不应是应用语言学的一个分支,而应享有独立的学科地位。翻译教学和理论研究领域这些可喜的变化,无疑对造就高水平的翻译人才、提高我国数以十万计的翻译从业者的业务水平和工作质量发挥着积极的作用。

理论与实践互动,是翻译事业健康发展的必要条件。在这方面有两个良好趋势值得注意。一是随着新一代翻译工作者和翻译教师、翻译研究家的迅速成长和翻译学科的壮大(以翻译学和翻译理论与实践为方向的硕士研究生教学点已有近百个,博士点约有十个),以及众多翻译学术刊物的出现,学术论文的发表,学术专著的出版,特别是由于译学研究成果的可信和有用,“翻译无理论”和“翻译理论无用”的声音,除了极个别场合外已经不大听得到了。这当然是个积极的变化。二是译学著述不再是文学翻译家和大学翻译教师的专利,各行各业从事翻译工作的有心人,他们挟着丰厚的实践经验,迈步跨入这个领域,著书立说,带来一阵阵清新气息。这方面以科技翻译界贡献最为突出,其他如外交、贸易、金融、工商企业、编辑出版、对外宣传等各界,也都有作者涌现出来。翻译实务的经验以此得到总结,得以条理化、系统化,为后来人指点门径,同时也为我们的译学建设作了贡献。基础研究与应用研究相互为用,有血有肉,这一趋势无疑是健康的,值得欢迎的。

这套“翻译理论与实务丛书”正是在上述喜人形势之下应运而生,编辑出版的。

中国对外翻译出版公司,作为国内惟一以翻译为特色的国家级出版机构,成立二十多年来已经陆续编辑出版数十种翻译理论与技巧类图书,其中若干种已经成为翻译研究或教学领域脍炙人口的必读书,在国内外翻译界赢得了声誉,奠定了公司在译学书籍

出版方面的领先地位。尽心竭力继续为翻译界广大专家学者、教师、从业人员、学生和人数众多的翻译爱好者提供高质量的专业读物,是我们义不容辞的责任。

现在,在已有的成绩和经验的基础上推出这套“翻译理论与实务丛书”,我们有点想法:一、加强自主策划,提高原创性,体现信息时代和市场经济环境下翻译的新任务和新特点;二、加强实用性和针对性,不仅要反映翻译教学和研究的新趋势、新成果,更要关注科技、经济、法律、贸易、金融、旅游、传媒等非文学领域的翻译实务;三、笔译与口译兼顾,英译中与中译英兼顾,着重于中英互译和中译英;四、有选择地引进境外著作,港台海外华人翻译家和翻译理论家不乏佳作,立足中华语言文化,有其独特魅力,是我们的首选。

这些想法能不能实现,能不能把我们贡献精品的愿望变成现实,全靠翻译界广大学者专家、学校师生以及广大读者的支持与合作,没有这种支持与合作,我们的想法再好也只是空话而已。我们所期待的支持与合作,既包括惠赐书稿、推荐选题、介绍作者,也包括对我们已出的书发表评论,无论是表扬还是批评,都将鼓励我们把工作做得更好,是我们无任欢迎的。

丛书主编 罗进德 谨识

## 本卷主编寄语

在总序中，主编罗进德先生指出：“译学著述不再是文学翻译家和大学翻译教师的专利，各行各业从事翻译工作的有心人……著书立说，带来一阵阵清新气息。这方面以科技翻译界贡献最为突出。”“科学翻译研究卷”即是对这一席话的验证与体现。

当今世界，谁在经济上是强手，谁就是未来的主宰。政治实力取决于军事实力，军事实力取决于经济实力，经济实力成了左右世界政治格局的实质力量，而经济实力的支柱是科技实力。经济和科技的发展需要科学翻译，对发展中国家，尤其如此。

什么是科学翻译？在《科学翻译学》一书之首，我们开宗明义，指出：

科学翻译，概略地说，是以传达科学信息为主的翻译活动。

科学翻译，具体地说，是译者用译语表达原语科学信息以求信息量相似的思维活动和语际活动。

1977年联邦德国学者沃·威尔斯在《翻译学：问题与方法》一书中指出：在书斋中苦心翻译文学作品者只占少数。1996年9月22日张研在《中国青年报》上撰文《文学翻译：追根溯源看危机》，从源远流长的中国译史看现状，留念往日的辉煌，指出中国文学翻译面临着新的挑战，文学翻译力量后继乏人。改革开放以来，《人民日报》、《光明日报》、《中华读书报》、《科学时报》、《文汇读书周报》等都在关注科学翻译与文学翻译的发展，期盼重振翻译大国的雄风。

科学翻译事业的蓬勃发展正是适时地融入了这种重振雄风的乐章。在从业人数上科学翻译者是压倒的大多数，外语毕业生从译者绝大多数从事的是科学翻译，文学翻译只是偶尔练笔的一种方式，靠它生存者少。谁也无法否定科技在当代社会的巨大作用。科学技术是第一生产力，科学翻译是其有机组成部分。我国的思想解放、科技进步、经贸发展、文化交流，哪一项离得开科学翻译？林纾之后，世界文学名著的翻译越来越少，计算机诞生后，信息传播速度越来越快，大量的科学信息需要传播，科学翻译更是显出强劲的势头，翻译与信息科学的嫁接最适合科学翻译，机器翻译在 1990 年代的再度兴起，昭示着科学翻译的无量前程。科学促进了翻译事业，翻译繁荣了科学发展。

《科学翻译学》是“科学翻译研究卷”首推之作，具有综论和指导性质；其他几种或与之相关，或从中独立出来，单列论述，从科学文体、科学普及、翻译批评、翻译教学、词典译编、机器翻译、科学口译、翻译方法论等角度作出具体的研究。“科学翻译研究卷”构成了科学翻译理论研究体系与实践研究体系，从一般理论研究到应用理论研究，互为见证，互为参照，自成体系。可见科学翻译学和“科学翻译研究卷”的研究宗旨有别于文学翻译及其研究，突出了科学翻译的作用、特征及其学术、现实意义。

“科学翻译研究卷”吸引了译界老将新秀共襄科学译事。我们的初衷是成熟一本，推出一本。本卷的各位作者或者学有所成就，或者素有研究，都是理论与实践某个方面的专门家，各本专著都是他们的多年思考与研究所得，有的还是国家社会科学基金项目、教育部人文社科重点研究基地重大项目的成果等。

本卷仍坚持罗进德先生倡导的“理论与实务相结合”的原则，按照求严求新的写作标准，不堆砌玄妙词语，即使有新观念新概念，也力求界定清楚，表述上深入浅出，旨在普及与提高。



---

用实在之笔，道思考所得，同时追求学术性和系统性，力求出思想，有观点，立新意。

李亚舒 黄忠廉

2003 年盛夏

## 前 言

我曾经写过两本关于机器翻译的书，一本是《自动翻译》，于1987年由上海知识出版社出版，一本是《自然语言机器翻译新论》，于1995年由语文出版社出版。在这两本书中，我论述了“独立分析—独立生成—相关转换”的原则以及“语言规则和计算机程序分开”的原则，介绍了机器翻译的基本理论。这两本书的出版，对于我国的机器翻译研究，起了一定的促进作用。

但是，这两本书主要介绍理论，对于机器翻译的方法论述不多，实例较少，内容比较抽象，理解起来比较困难。因此，很多朋友曾经建议我再写一本关于机器翻译的书，着重讲述机器翻译的方法，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，使读者对于机器翻译得到更加具体的认识。朋友们的建议当然很有道理，可是，我因为忙于其他工作，总是抽不出时间动笔，再写新著仅只是我的一个愿望，难以实现。

中国科学院科技翻译工作者协会副会长李亚舒教授和华中师范大学黄忠廉教授正在编写一套关于科学翻译的书，去年秋天，他们对我说，机器翻译是科学翻译中很重要的内容，他们要求我为这套书写一本关于机器翻译的著作，他们的这个要求正好为实现我再写新著的愿望提供了机会，于是，我顺水推舟，愉快地接受了他们的要求，开始写这本《机器翻译研究》的新著。

机器翻译与科学翻译有着非常密切的联系。早期的机器翻译研究主要就是为翻译科学技术文献服务的。当今科学技术的发展日新月异，科学技术文献的数量与日俱增，使我们目不暇接，手

工的翻译虽然译文质量很高，但是速度太慢，难于满足科学技术文献日益增长的需要，机器翻译是解决这个问题的重要手段。我在1981年至1985年期间，曾在中国科学技术信息研究所计算中心担任机器翻译研究组的组长，我们这个研究组的主要任务，就是使用计算机进行冶金方面的科学技术文献的翻译，建立英汉冶金文献的机器翻译系统。机器翻译确实应该成为科学翻译的一个组成部分。我希望科学翻译界的朋友能够认可机器翻译，欢迎机器翻译，使用机器翻译。

接受了写作《机器翻译研究》的任务后，我就到韩国科学技术院电子工程与计算机科学系担任访问教授，用英语讲课，备课负担很重，因此，我只好见缝插针，利用业余时间边写边改，历时一年，现在终于写成了这本书。

本书在写作过程中，曾参考过国内外时贤著作多种，特别是参考了我的美国朋友汝拉夫斯基（D. Jurafsky）和马丁（J. Martin）2000年出版的《语音和语言处理》（Speech and Language Processing）一书，并得到他们热情的指点，在本书出版之际，特向他们表示衷心的感谢。

冯志伟

2004年9月20日

# 目 录

|                              |    |
|------------------------------|----|
| 丛书总序 .....                   | i  |
| 本卷主编寄语 .....                 | v  |
| 前 言 .....                    | ix |
| <br>                         |    |
| 第一章 机器翻译概述 .....             | 1  |
| 1.1 什么是机器翻译 .....            | 1  |
| 1.2 对于机器翻译的一些误解 .....        | 5  |
| 1.3 机器翻译发展的曲折道路 .....        | 12 |
| <br>                         |    |
| 第二章 各种类型的机器翻译系统 .....        | 35 |
| 2.1 基于规则的机器翻译系统 .....        | 35 |
| 2.1.1 国外的系统 .....            | 35 |
| 2.1.2 国内的系统 .....            | 38 |
| 2.1.3 基于规则的机器翻译系统的类型 .....   | 40 |
| 2.2 基于语料库的机器翻译系统 .....       | 44 |
| 2.2.1 基于统计的机器翻译系统 .....      | 44 |
| 2.2.2 基于实例的机器翻译系统 .....      | 50 |
| 2.3 多引擎机器翻译系统 .....          | 52 |
| 2.4 口语机器翻译系统 .....           | 56 |
| 2.4.1 AT&T 公司的语音机器翻译系统 ..... | 56 |
| 2.4.2 Verbmobil 系统 .....     | 57 |
| 2.4.3 其他的口语机器翻译系统 .....      | 60 |

---

|                            |            |
|----------------------------|------------|
| 2.5 因特网与机器翻译·····          | 61         |
| 2.6 机器翻译与翻译记忆·····         | 64         |
| <br>                       |            |
| <b>第三章 机器翻译的实现过程</b> ····· | <b>67</b>  |
| 3.1 机器翻译结果比较·····          | 67         |
| 3.2 机器翻译金字塔·····           | 74         |
| 3.3 基于转换的机器翻译的实现过程·····    | 76         |
| 3.3.1 特征和特征值的定义·····       | 76         |
| 3.3.2 词典编写·····            | 80         |
| 3.3.3 形态分析·····            | 81         |
| 3.3.4 句法分析·····            | 85         |
| 3.3.5 逻辑语义分析·····          | 89         |
| 3.3.6 双语转换·····            | 91         |
| 3.3.7 目标语言生成·····          | 93         |
| 3.3.8 基于转换的英汉机器翻译实例·····   | 94         |
| 3.4 直接翻译系统的实现过程·····       | 99         |
| <br>                       |            |
| <b>第四章 形态自动分析</b> ·····    | <b>103</b> |
| 4.1 有限状态转移网络 ·····         | 103        |
| 4.1.1 有限状态转移网络的基本原理 ·····  | 103        |
| 4.1.2 有限状态转换机 ·····        | 113        |
| 4.1.3 有限状态机器的局限性 ·····     | 116        |
| 4.2 不同类型语言的形态自动分析 ·····    | 119        |
| 4.2.1 不同类型语言形态分析的特点 ·····  | 119        |
| 4.2.2 形态自动分析规则的编制方法 ·····  | 123        |
| 4.2.3 形态分析的理论依据 ·····      | 128        |
| 4.3 词的形式化描述与分析 ·····       | 130        |
| 4.3.1 词和词形式 ·····          | 130        |

---

|            |                                      |            |
|------------|--------------------------------------|------------|
| 4.3.2      | 语素和语素变体 .....                        | 135        |
| 4.3.3      | 词形式的形态分析步骤 .....                     | 137        |
| 4.3.4      | 词形式的自动识别方法 .....                     | 137        |
| 4.3.5      | 语素变体规则(Allo-rules) .....             | 141        |
| 4.3.6      | 组合规则(Combi-rules) .....              | 146        |
| 4.3.7      | 毗连模式(Concatenation patterns)示例 ..... | 147        |
| 4.4        | 词类标记集和词类标注 .....                     | 149        |
| 4.4.1      | 英语中使用的主要的词类标记集 .....                 | 149        |
| 4.4.2      | 词类标注 .....                           | 152        |
| 4.5        | 英语分析中的兼类词消歧策略 .....                  | 155        |
| 4.5.1      | 英语中词的兼类现象 .....                      | 157        |
| 4.5.2      | 英语中兼类词的消歧策略 .....                    | 166        |
| 4.5.3      | E-to-J 机器翻译系统的同形兼类词的<br>消歧规则概览 ..... | 167        |
| <b>第五章</b> | <b>基于转移网络的自动句法分析</b> .....           | <b>181</b> |
| 5.1        | 递归转移网络 .....                         | 181        |
| 5.1.1      | 自然语言的递归特性 .....                      | 181        |
| 5.1.2      | 从有限状态网络(FSTN)到递归<br>转移网络(RTN) .....  | 182        |
| 5.1.3      | 用递归转移网络(RTN)进行句子识别 .....             | 186        |
| 5.1.4      | 用递归转移网络(RTN)进行随机生成 .....             | 199        |
| 5.1.5      | 下推转录机 .....                          | 200        |
| 5.2        | 扩充转移网络 .....                         | 202        |
| 5.2.1      | RTN 的局限性 .....                       | 202        |
| 5.2.2      | 扩充转移网络在机器翻译中的应用 .....                | 203        |

---

|                                   |     |
|-----------------------------------|-----|
| 第六章 基于上下文无关语法的自动句法分析              | 208 |
| 6.1 短语结构语法                        | 208 |
| 6.1.1 短语结构语法的代数定义                 | 210 |
| 6.1.2 形式语法的乔姆斯基分类                 | 211 |
| 6.1.3 形式语法的生成能力                   | 212 |
| 6.2 英语的上下文无关语法                    | 219 |
| 6.2.1 上下文无关规则和树形图                 | 219 |
| 6.2.2 英语的句子结构                     | 222 |
| 6.2.3 英语的名词短语                     | 224 |
| 6.2.4 联合结构                        | 225 |
| 6.2.5 一致关系                        | 226 |
| 6.2.6 英语的动词短语和动词的再分类              | 227 |
| 6.2.7 助动词                         | 229 |
| 6.2.8 语法的等价与乔姆斯基范式                | 230 |
| 6.2.9 有限状态语法与上下文无关语法的关系           | 231 |
| 6.3 自底向上剖析                        | 234 |
| 6.3.1 剖析的定义                       | 234 |
| 6.3.2 剖析就是搜索                      | 235 |
| 6.3.3 自底向上剖析的过程                   | 236 |
| 6.4 自顶向下剖析                        | 240 |
| 6.4.1 自顶向下剖析的过程                   | 240 |
| 6.4.2 自顶向下剖析与自底向上剖析的比较            | 244 |
| 6.5 左角分析法                         | 244 |
| 6.5.1 左角的概念                       | 244 |
| 6.5.2 自底向上过滤(Bottom-up filtering) | 246 |
| 6.6 自顶向下剖析的问题                     | 250 |
| 6.6.1 左递归(Left-recursion)         | 250 |
| 6.6.2 结构歧义(Structure ambiguity)   | 251 |

---

|                                                               |            |
|---------------------------------------------------------------|------------|
| 6.6.3 子树无效的重复分析<br>(Inefficient re-parsing of sub-tree) ..... | 265        |
| 6.7 Earley 算法 .....                                           | 268        |
| 6.7.1 线图和点规则 (Chart and dotted rule) .....                    | 269        |
| 6.7.2 Earley 算法的三种操作 .....                                    | 270        |
| 6.7.3 Earley 算法示例 .....                                       | 272        |
| 6.7.4 从线图中检索剖析树 .....                                         | 275        |
| 6.8 花园幽径句的自动分析 .....                                          | 287        |
| 6.8.1 英语中的花园幽径句 .....                                         | 287        |
| 6.8.2 汉语中的花园幽径句 .....                                         | 290        |
| 6.8.3 花园幽径度 .....                                             | 297        |
| 6.8.4 花园幽径句的三个特性 .....                                        | 299        |
| 6.8.5 花园幽径句的自动分析算法 .....                                      | 303        |
| 6.8.6 准花园幽径句 .....                                            | 313        |
| 6.8.7 语义花园幽径句 .....                                           | 314        |
| 6.9 CYK 算法 .....                                              | 316        |
| 6.10 富田算法 .....                                               | 324        |
| 6.11 浅层句法分析 .....                                             | 329        |
| 6.11.1 基于有限状态自动机的浅层分析 .....                                   | 330        |
| 6.11.2 基于隐马尔可夫模型的浅层句法分析 .....                                 | 337        |
| 6.11.3 基于中心词依存概率的浅层句法分析 .....                                 | 338        |
| 6.11.4 模仿词性标注的浅层句法分析 .....                                    | 339        |
| <b>第七章 基于特征结构的自动句法分析 .....</b>                                | <b>343</b> |
| 7.1 多值标记函数与复杂特征 .....                                         | 343        |
| 7.2 特征结构与合一运算 .....                                           | 357        |
| 7.2.1 “属性—值”矩阵 .....                                          | 357        |
| 7.2.2 特征路径和重入结构 .....                                         | 360        |



|                                     |            |
|-------------------------------------|------------|
| 7.2.3 特征结构的合一 .....                 | 362        |
| 7.2.4 “包孕于”关系 .....                 | 370        |
| 7.2.5 “合一”的形式定义 .....               | 371        |
| 7.3 特征结构与语法 .....                   | 373        |
| 7.3.1 使用特征结构来增强上下文无关<br>语法的规则 ..... | 373        |
| 7.3.2 一致关系 .....                    | 374        |
| 7.3.3 中心语特征 .....                   | 376        |
| 7.3.4 次范畴化 .....                    | 378        |
| 7.3.4.1 原子特征 SUBCAT 的描述方法 .....     | 378        |
| 7.3.4.2 次范畴化框架 .....                | 380        |
| 7.3.5 长距离依存关系 .....                 | 382        |
| 7.4 合一运算的实现 .....                   | 384        |
| 7.4.1 合一的数据结构 .....                 | 384        |
| 7.4.2 合一运算的操作过程 .....               | 384        |
| 7.4.3 合一运算实现的其他示例 .....             | 388        |
| 7.4.4 带合一约束的自动句法分析 .....            | 402        |
| 7.5 类型与继承关系 .....                   | 406        |
| 7.5.1 简单类型 .....                    | 407        |
| 7.5.2 复杂类型 .....                    | 408        |
| <b>第八章 基于依存语法的自动句法分析</b> .....      | <b>412</b> |
| 8.1 从依存语法到配价语法 .....                | 412        |
| 8.2 依存语法与短语结构语法的等价性 .....           | 419        |
| 8.3 依存语法与从属树的一些形式特性 .....           | 429        |
| 8.4 依存语法在机器翻译系统中的应用 .....           | 434        |