

上海外语音像出版社“十三五”规划重点教材

Cloud Translation  
Translation Technology Competence  
Professionalization Post Editing  
Language Technology  
Translation Memory Information Retrieval  
Computer-aided Translation

王华树 主编

---

**A Practical Course in  
Translation Technologies**

**翻译技术教程**

---

下册



创于1897

商务印书馆

The Commercial Press



爱未来

上海外语音像出版社

图书在版编目 (CIP) 数据

翻译技术教程. 下册 / 王华树主编. —北京 :  
商务印书馆, 2017

ISBN 978-7-100-13998-4

I. ①翻… II. ①王… III. ①翻译—教材 IV. ①H059

中国版本图书馆 CIP 数据核字 (2017) 第 117817 号

权利保留, 侵权必究。

翻译技术教程. 下册

王华树 主编

---

商务印书馆出版

(北京王府井大街36号 邮政编码 100710)

上海外语音像出版社出版

(上海大连西路550号 邮政编码 200083)

商务印书馆 发行  
上海外语音像出版社

上海新华印刷厂

ISBN 978-7-100-13998-4

2017年8月第1版 开本 787×960 1/16

2017年8月第1次印刷 印张 25 1/4

定价: 49.80 元

# 序 言

2016年5月,中共中央、国务院印发了《国家创新驱动发展战略纲要》,其中强调科技创新是提高社会生产力和综合国力的战略支撑,应加快工业化和信息化深度融合,推进各领域新兴技术跨界创新。这不仅充分表明中央高度重视推动科技创新发展,也给以跨语言、跨文化和跨平台为特点的语言服务科技智库带来蓬勃发展的历史机遇和服务国家的光荣使命。广东外语外贸大学在高校科技智库建设方面做了长远的规划和不懈的努力。近年来不仅集成多学科、多行业、多领域的智慧资源,建设了多个语言服务协同创新发展平台,还通过整合多学科发展战略分析和研判,与世界科技发展前沿和趋势保持一致,为珠三角乃至全国科技创新实现从“跟跑”到“领跑”转变提供智力支持,为“一带一路”国家语言服务战略提供技术保障和支撑服务。2016年11月,广东外语外贸大学牵头成立了世界翻译教育联盟,将中国的教育科技创新与世界的教育科技创新发展联系在一起,也为今后中国翻译教育的国际化发展奠定了良好的合作和共赢基础。

无论是国家语言服务科技智库还是全球翻译教育服务都急需高端的语言技术服务人才。信息技术与通信技术快速变革,技术与商业模式融合发展,技术资源与数据平台泛在共享,全球范围内大规模的技术创新已经形成燎原之势。大数据技术、人工智能、语音识别与机器翻译等技术发展迅速,已经影响到社会各行各业。“技术”已经成为引发社会变革的最具活力因子之一,正在改变着语言服务的面貌。翻译作为语言服务的主体业务,其对象、形式、流程、手段和所处环境等都发生了巨大的变化。语言服务对技术的需求越来越大,通常会涉及计算机辅助翻译、机器翻译、审校、质检、管理、排版等各种各样的技术处理,对人才的技术能力要求越来越高,翻译技术能力已经成为现代译员的必备素质,这对新时代的翻译教育提出了新的挑战。

近观国内高校翻译专业教学,截至2017年1月,215所高校开设了翻译硕士专业(MTI),252所高校开设了翻译本科专业(BTI),可见市场对翻译专业人才的需求量不断增长。在2017年实施的“翻译学位授权点申请基本条件”中要求“笔译方向须配备计算机辅助笔译实验室1个以上。实验室须配置翻译软件和语料库3种以上”,这客观上要求翻译高校应加强翻译技术课程建设,重视培养学生掌握现代化翻译工具的能力。此外,根据《全国翻译硕士专业学位研究生教育与就业调查

报告》,MTI十年来得到了长足发展,未来应该强化MTI高校的区域特色和专业特色发展,促进多元化课程体系建设(如计算机辅助翻译、机器翻译与译后编辑等),这也要求广大教育工作者要密切关注翻译技术的发展,积极应对技术发展带来的各种挑战。以广东外语外贸大学为例,在2015年,我们率先在硕士层面开设国内第一个本地化专业方向,在翻译本科和硕士阶段开设翻译和本地化技术型课程,同时搭建语言资产和教学案例平台,加强翻译技术实验室建设,优化翻译和本地化教学配套资源,旨在培养懂翻译、懂技术、懂行业且能够适应新时代发展需求的综合性人才。

翻译技术发展迅速,职业翻译工作已离不开翻译技术,很多学者关注教学的同时也关注翻译技术发展。从本质上说,翻译技术是当今翻译行为本质属性的重要构成,翻译技术也由此成为翻译教学与研究领域的重要课题。翻译技术为当今的翻译学引入了许多新的研究内容,例如,计算机辅助翻译过程、译者翻译技术能力、术语管理技术、翻译质量保障技术、机器翻译伦理研究等;也为翻译学带入了一系列新的术语,比如翻译记忆、模糊匹配、本地化翻译、译后编辑、众包翻译等,这些新的内容拓展了翻译学的理论深度与宽度,非常值得深入探讨。

在新时代,我们的翻译教育应该遵循“创新引领、产业驱动和职业导向”的方针,服务于国家发展的宏观战略,并为地方经济社会发展提供智力支撑。为此,我们需要坚持国家科技创新的大政方针,把握语言服务市场的需求变化,确立区域差异化的翻译专业教育体系与评估机制,加强专业化的语言服务人才规划,提升师生对语言服务技术的认知,优化政产学研协同研究机制,深入研究翻译教育和翻译技术的相关问题,努力为国家语言能力建设培养一批具备综合素养的战略性语言服务人才,这是我们广大翻译教育工作者必须要重视的问题。

在这样的背景下,《翻译技术教程》一书问世,可谓恰逢其时。尤为可喜的是,以王华树博士为代表的一批青年学者,既有实战积累,又有理论素养,既继承前人研究的成果,又能紧跟时代步伐,勇于开拓创新。他们的研究成果拓展了翻译技术与教学的思路,为当下渐次铺开的翻译技术教学再添力量。阅毕全书,印象深刻,略述如下:

一、技术引领。该书较为全面地介绍了大数据时代的翻译技术体系构成及最新发展趋势,突出大数据时代翻译技术的热点话题,例如,神经网络机器翻译、机器翻译与译后编辑技术,翻译管理与内容管理技术、云翻译平台技术等,旨在引领翻译专业进行技术创新和教学改革。

二、内涵丰富。该书层次清晰,案例丰富,内容充实,从翻译搜索技术、语料库技术、计算机辅助翻译技术、机器翻译及译后编辑、本地化工程技术、术语管理技术、技术文档写作到翻译管理技术等均有深入讨论;书中还有单独章节专门介绍综

合性的技术与应用,可谓包举无遗,为读者呈现了新时代主流翻译技术的概貌。

三、协同创新。该书一改传统教材编写惯例,编写团队既有学界又有业界人士,且多数是“跨界”专家,包括语言服务企业翻译和本地化技术资深专家、翻译技术沙龙组织者以及具有丰富教学经验的一线翻译技术教师,他们群策群力,各书其长,构建了基于语言服务需求的翻译技术教学框架,体现了现代语言服务生态系统的协同创新特征。

四、突出应用。该书求真务实,兼顾价值理性的同时,凸显翻译技术应用。书中结合国际知名语言服务企业和语言技术研发企业(如SDL、Kilgray、Atril等)的技术实践和解决方案,深入浅出地讲解了多项翻译技术与工具的功能用途及操作要领,为读者呈现了一幅现代翻译技术的工作流程图。

随着全球化的深入推进,市场对语言服务人才的综合素养要求空前提高,反推翻译教学改革。在当前信息化翻译教学如火如荼的局面下,该书的理论和实践价值尤为明显。限于多种因素,书中针对技术内涵和技术发展等内容可能还有待完善,但编者的大胆创新之举已是学界福音,希望该书能够掀开国内翻译技术教学新篇章,引领翻译教学改革新趋势。

秋为收,冬为藏,更迭之际,得此厚积薄发之作,愈发期待破土之春。

是为序。

赵军峰

广东外语外贸大学高级翻译学院

2017年1月6日



# 前 言

放眼全球,国际语言服务市场朝气蓬勃,语言服务正朝着数字化、专业化、网络化、云端化趋势快速发展。信息资源的全球非对称性已经成为制约区域经济均衡与可持续发展的最主要障碍之一。语言作为重要的信息载体,是实现信息资源共享的关键。因此,破除语言壁垒已成为互联网时代的“刚需”。纵观全局,整个经济社会对语言服务人才的信息素养要求越来越高,新时代的语言服务人才要敢于面对语言服务行业的问题与挑战,充分利用现代信息技术武装自己,才能适应未来语言服务行业更为多元化的需求。

近观国内,随着国家“一带一路”倡议战略构想的提出和实施,中国开始广泛、深度地融入全球经济体系,同时逐步扩大对外文化传播,加强国家软实力建设,现代信息技术和翻译技术就成为“一带一路”国家语言服务建设的助推器,这对翻译专业人才信息技术能力的培养提出了更高的要求。在新时代背景下,深入推进职业化、专业化、国际化人才战略,培养具备现代语言信息技术能力的高级翻译人才,对于满足国家新型语言服务人才的需求,积极推动国家语言服务及语言能力的建设具有举足轻重的作用。

此外,翻译能力的发展在不同时期呈现出不同的阶段性特点。在大数据时代,语言服务企业对于译员计算机辅助翻译工具应用能力的要求越来越高,翻译技术能力已经成为翻译能力必备的一部分。语言服务企业要求翻译人才应该具备基本的计算机操作能力、信息检索能力以及计算机辅助翻译工具应用能力等相关的技能。其中,计算机操作能力是当代译员最基本的技术能力,信息检索能力是现代职业译员高效收集翻译主题相关专业知识的必备技能,计算机辅助翻译工具应用能力是译员必备的更高层次的综合素养和技能。

基于这样的共识,本书根据国际语言服务行业的发展需求,以译员翻译技术能力为基础,结合翻译工具应用和案例分析,以项目案例为驱动,以解决翻译实践中实际问题为导向,着重培养学生的技术思维和技术能力,希望为培养面向全社会的、以应用能力为导向的语言服务行业复合型人才提供借鉴和参考。本书适用对象为外语、翻译专业的本科生、研究生,翻译技术教学及研究人员,以及翻译、审校、项目经理、技术写作等其他语言服务从业者和爱好者。

本书在编写理念、内容编排、使用方法以及作者安排等方面具有以下特点:

在编写理念上,本书兼顾技术理论与技术实践,偏重实践应用,着眼于为读者解

读新时代语言服务技术以及现代译者的技术能力要求;帮助广大读者“走近”并“走进”现代翻译技术——厘清翻译技术的一些基本概念,消除认识上的一些误区;了解语言服务技术及发展趋势,掌握多种现代翻译技术和工具的基本操作;并对语言服务行业和翻译项目管理有宏观上的初步认识,以适应行业和时代发展需要。

在内容编排上,本书兼顾学生和教师,侧重为教师提供丰富的教学案例。本书共有13个章节,每个章节之间的内容相互关联。第一章是翻译技术概述,第二章至第十二章则是翻译技术专题,第十三章是翻译技术的综合应用。授课教师可自由组合,从不同章节中选取合适内容组成一个相对完整的翻译技术应用课程,或者整合为一个课程的不同级别(如初级和高级)。例如,开设“计算机辅助翻译实践”课程,除了讲解计算机辅助翻译工具之外,还可以将翻译与搜索、语料库与翻译、术语管理技术以及翻译管理技术等融入到教学设计之中。开设“本地化翻译”课程,可以将技术写作、计算机辅助翻译技术、术语管理技术、翻译项目管理等内容整合在一起。如此可以实现一书多用,很大程度上解决了翻译技术应用教材短缺的困难。

在使用方法上,读者可以通过多元化的学习方式,借助现代化的网络教学平台,实现与编者零距离地学习和交流。本书除了纸质教材之外,后续将陆续推出网络课程,提供涵盖主要知识点的微视频课程,并配备可用的实操软件和丰富的案例资料。配套网络平台设有教材简介、章节作者、章节内容、视频课程、实践练习等栏目。读者可以通过浏览章节简介,学习相应内容和习题。如有疑问,可以在线讨论或者在约定时间内与编者互动。

在作者安排上,本书有别于过去翻译技术教材由高校教师撰写的做法,我们特别邀请了语言服务企业懂翻译、懂技术的资深专家以及活跃在一线的翻译技术沙龙组织者和具有丰富教学经验的翻译技术教师参与编写,力争每位专家编写最擅长的内容。具体分工如下:王华树负责全书的策划、统筹、统稿、审稿及质量管控,并负责编写第一章和第四章内容。张成智、韩林涛编写第二章,管新潮、姜振东编写第三章,刘宁赫参与编写第四章,吴微微编写第五章,周兴华编写第六章,徐彬编写第七章,李梅、邹剑宇编写第八章,崔启亮编写第九章,王少爽、魏勇鹏编写第十章,刘劲松编写第十一章,杨颖波编写第十二章,李艺峰、田惠才编写第十三章。

本书针对时代和市场需求,博取众家之长,以理引路,以例示范,力显业界之鲜活面貌。然而,翻译技术包罗万象,发展日新月异,纵使作者奋发蹈厉,朝乾夕惕,仍不免有所疏漏,恳望方家不吝赐教。

王华树

广东外语外贸大学

外语研究与语言服务协同创新中心

2017年1月1日

# 目 录

<b>第八章 机器翻译及译后编辑</b> .....	1
本章导言 .....	1
8.1 机器翻译概述 .....	1
8.1.1 机器翻译的发展历程 .....	2
8.1.2 机器翻译的工作原理 .....	4
8.2 机器翻译的主要应用 .....	9
8.2.1 机器翻译的适用领域 .....	10
8.2.2 语音翻译系统 .....	12
8.2.3 在线机器翻译系统 .....	13
8.3 机器翻译的质量评测 .....	18
8.3.1 评测分类与内容 .....	18
8.3.2 评测标准与方法 .....	19
8.4 机器翻译的未来展望 .....	21
8.5 译后编辑概述 .....	23
8.5.1 译后编辑及其研究 .....	23
8.5.2 译后编辑的特征和所需技能 .....	24
8.6 译后编辑的市场需求和技术应用 .....	27
8.6.1 译后编辑的市场需求 .....	27
8.6.2 译后编辑的工具 .....	27
8.6.3 译后编辑案例分析 .....	32
8.6.4 译后编辑未来展望 .....	34
思考题 .....	35
<b>第九章 本地化工程技术</b> .....	36
本章导言 .....	36
9.1 本地化与本地化工程简介 .....	36
9.1.1 本地化的概念 .....	36
9.1.2 本地化的对象类型 .....	37

9.1.3	本地化翻译的文本特征 .....	38
9.1.4	本地化工程的概念 .....	41
9.1.5	本地化、本地化翻译、本地化工程之间的关系 .....	42
9.1.6	本地化工程的工作内容 .....	43
9.2	本地化工程技术 .....	45
9.2.1	本地化工程分析技术 .....	45
9.2.2	本地化预处理技术 .....	46
9.2.3	本地化翻译技术 .....	48
9.2.4	本地化后处理技术 .....	50
9.3	Alchemy Catalyst 在软件用户界面本地化翻译案例 .....	51
9.3.1	译前阶段(预处理) .....	52
9.3.2	译中阶段 .....	53
9.3.3	译后阶段(后处理) .....	54
9.4	本地化翻译技术与工具的选用策略 .....	56
9.4.1	基于本地化对象特征 .....	56
9.4.2	基于本地化翻译的质量要求 .....	57
9.4.3	基于组织的本地化能力 .....	57
9.4.4	基于工具功能与市场需求 .....	57
9.5	网站本地化翻译案例分析 .....	58
9.5.1	案例背景 .....	58
9.5.2	需求分析 .....	59
9.5.3	项目实施 .....	59
9.5.4	案例总结 .....	67
9.6	游戏本地化翻译案例分析 .....	68
9.6.1	案例背景 .....	69
9.6.2	需求分析与解决方案 .....	69
9.6.3	项目实施 .....	70
9.6.4	案例总结 .....	81
9.7	手机应用程序本地化翻译案例分析 .....	82
9.7.1	项目概况 .....	82
9.7.2	项目分析 .....	82
9.7.3	项目实施 .....	83
9.7.4	案例总结 .....	91
	思考题 .....	91

<b>第十章 翻译术语管理技术</b> .....	92
本章导言 .....	92
10.1 术语管理基础 .....	92
10.1.1 术语的特征 .....	92
10.1.2 术语标准化 .....	94
10.1.3 术语管理的定义 .....	95
10.1.4 译者术语能力 .....	96
10.2 术语管理技术 .....	97
10.2.1 术语提取 .....	97
10.2.2 术语库检索 .....	99
10.2.3 术语管理系统 .....	101
10.3 术语管理流程 .....	103
10.3.1 译前术语管理 .....	103
10.3.2 译中术语管理 .....	104
10.3.3 译后术语管理 .....	105
10.4 术语管理技术应用 .....	106
10.4.1 SDL MultiTerm 2015 的使用 .....	106
10.4.2 memoQ 术语管理模块的使用 .....	132
10.4.3 语帆术语宝的使用 .....	143
10.5 术语管理案例分析 .....	156
10.5.1 案例 1: 语智云帆专利译后编辑项目中的术语管理 .....	157
10.5.2 案例 2: 网站本地化项目中的术语管理实践 .....	162
思考题 .....	176
<b>第十一章 技术文档写作</b> .....	177
本章导言 .....	177
11.1 技术文档写作基础 .....	177
11.1.1 技术文档写作发展的四个阶段 .....	179
11.1.2 技术文档写作的重要意义 .....	181
11.1.3 技术文档写作的概念 .....	183
11.1.4 技术文档写作以用户为中心 .....	188
11.1.5 技术文档写作的结构化 .....	191
11.1.6 技术文档写作的标准化流程 .....	192

11.2 技术文档写作相关法规与标准 .....	199
11.2.1 技术文档写作国内外法规 .....	199
11.2.2 技术文档写作国内外标准 .....	201
11.3 技术文档写作工具 .....	202
11.3.1 内容管理工具 .....	203
11.3.2 文档写作工具 .....	204
11.3.3 页面排版工具 .....	206
11.3.4 制图工具 .....	207
11.3.5 转换工具 .....	207
11.4 技术文档翻译案例 .....	207
11.4.1 术语表述不一致 .....	208
11.4.2 代词指代不清 .....	209
11.4.3 句法提示符以及标点 .....	210
11.4.4 语序 .....	210
思考题 .....	211
<b>第十二章 翻译管理及其技术应用 .....</b>	<b>212</b>
本章导言 .....	212
12.1 翻译管理概述 .....	212
12.1.1 翻译管理的挑战 .....	212
12.1.2 项目管理体系 .....	214
12.1.3 翻译项目管理 .....	216
12.1.4 翻译管理技术及系统 .....	219
12.2 语言服务项目的端到端管理 .....	221
12.2.1 项目资源的准备 .....	222
12.2.2 项目的生产与管理 .....	227
12.2.3 项目的沟通与跟踪 .....	231
12.3 TMS 介绍 .....	234
12.3.1 国外主要 TMS 及示例 .....	234
12.3.2 国内主要互联网翻译平台及示例 .....	246
12.4 小结 .....	257
思考题 .....	258

第十三章 翻译辅助综合技术与应用 .....	259
本章导言 .....	259
13.1 其他翻译辅助技术概览 .....	259
13.1.1 概述 .....	259
13.1.2 译前文件处理 .....	260
13.1.3 译中翻译辅助 .....	264
13.1.4 译后文件辅助 .....	269
13.2 文件处理之格式转换 .....	278
13.2.1 CAT 处理的文件综述 .....	278
13.2.2 CAT 各种工具间的互相配合 .....	289
13.2.3 文件转换实例分析 .....	297
13.3 文件预处理之 OCR .....	302
13.3.1 OCR 技术简介 .....	302
13.3.2 OCR 软件对比说明 .....	303
13.3.3 ABBYY FineReader 使用演示 .....	304
13.3.4 OCR 后处理-译前处理 .....	313
13.3.5 OCR 排版标准模板 .....	321
13.4 翻译辅助技术之正则表达式 .....	322
13.4.1 正则表达式 .....	322
13.4.2 CAT 中的正则表达式应用——Trados 为例 .....	335
13.5 其他计算机辅助翻译技术和平台 .....	345
13.5.1 CAT 扩展程序 .....	346
13.5.2 TransTools 使用与操作 .....	349
13.5.3 国内新兴技术平台 .....	359
思考题 .....	387
附录 参考文献与拓展阅读 .....	388
致谢 .....	397

# 第八章 机器翻译及译后编辑

## ✿ 本章导言

用机器替代人工进行翻译的设想由来已久。20世纪计算机的出现终于催生了机器翻译梦想的实现。经过半个多世纪的发展,今天的机器翻译发展成绩斐然,为人类跨语言交流提供了极大可能。尽管如此,机器翻译大多情况下仍需经过最后一道工序译后编辑才能达到人们不同程度的翻译需求。本章将从追溯机器翻译的发展历史开始,介绍机器翻译的工作原理及质量评测标准。我们还将展示机器翻译的应用和主流在线机器翻译系统。之后,本章描述了译后编辑的基本概念、特征及标准,讨论了译后编辑的应用场景并进行译后编辑案例分析。在本章对机器翻译和译后编辑的未来展望中读者们将会发现机器翻译和译后编辑的巨大潜力和 market 价值。

## 8.1 机器翻译概述

机器翻译亦称自动翻译,称其自动是因为将文本或语音形式的自然语言从源语翻译成目标语的过程不是由人工而是由计算机加工处理后完成的。机器翻译是一门涉及语言学、计算机科学以及数学等跨学科的研究领域。从语言学角度来看,机器翻译为计算语言学的一个分支。从计算机科学角度来看,它属于人工智能的范畴。而就数学而言,它涵盖在数理逻辑之下(冯志伟,2005:1)。一般来说,自然语言的机器处理可以分为从规则到语料库两个阶段(吴军,2014:16)。自2011年底谷歌发布了一项基于深度学习的“谷歌大脑”新技术以来,人工神经网络在机器翻译中的应用研究飞速发展。2014年前后基于深度神经网络的机器翻译技术研究开始兴起,2016年9月谷歌宣布由谷歌翻译移动端及互联网App进行的汉英翻

译均由谷歌神经机器翻译完成<sup>①</sup>。标志着机器翻译从此进入了一个新的时代。

### 8.1.1 机器翻译的发展历程

机器翻译与计算机技术的发展与应用密不可分。计算机科学之父阿兰·图灵(Alan Turing)于1950年最早提出了机器智能的设想。其实,他并没有提机器翻译的研究方法,而是提出了一种让人与机器进行交流来验证机器是否有智能的方法,其基本思想是如果人无法判断自己交流的对象是人还是机器,就说明这个机器有智能了。这种方法被后人称为图灵测试(吴军,2014:16)。图灵测试的核心不是计算机能否和人对话,而是计算机能否在智力行为上表现得和人无法区分。事实上,机器翻译的历史可以追溯到许多年前(Hutchins, 1986:3—15),下面简要介绍一下机器翻译的历史和发展。

#### (1) 从萌芽到低谷(1946—1966)

机器翻译的历史可以追溯到17世纪。被称为西方哲学之父的法国学者笛卡尔(Descartes)提出了通用语言(universal language)的设想<sup>②</sup>。之后人们开始尝试在逻辑语言和统一的图形符号基础上创造一种没有交流障碍的语言。而最早使用“机器/机械翻译(ein mechanisches Uebersetzen)”这个术语的是德国人利格(Rieger),他首次提出利用机械原理以数字语法加词典辅助的方法将源语翻译为目标语(冯志伟,2005:12)。20世纪初期,专家们一直在摸索用机械辅助的方式处理语言之间的语码转换,但均未获得实质性进展,直到1946年第一台电子计算机ENIAC问世才给机器翻译的研究带来了生机。计算机惊人的运算速度启发了人们重新思考利用机器进行翻译的可行性。

1949年,纽约洛克菲勒基金会总裁韦弗(Warren Weaver)发表了一篇后来影响世界的文章(详见下面的8.1.2节),在该文中他阐述了利用计算机进行翻译的观点。此后,他的观点迅速扩散,使机器翻译已然成为计算机的第一个非数字领域的应用。1952年召开了第一届机器翻译大会。紧接着,1954年便出现机器翻译系统的样机。媒体对此事大肆渲染,竞相报道。人们误以为机器翻译时代就要到来,以

---

<sup>①</sup> 2016年9月28日谷歌于Research Blog介绍其神经机器翻译系统(GNMT),宣布将GNMT投入到了非常困难的汉语-英语语言对的翻译生产中,引起了业内的极大关注。有关GNMT详见谷歌于ArXiv.org上发表的论文“Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation”。读者亦可参见网易科技之“机器之心”了解相应中文版介绍:<http://tech.163.com/16/0928/13/C228USK900097U80.html>。

<sup>②</sup> 笛卡尔假设通用语言的词汇由原始词素组成。根据语法规则,这些元素的系统组合将产生人类语言所需的计算结构的无限组合。故笛卡尔可谓是乔姆斯基定义的计算语言学的奠基人。详见Chomsky(2000)及Hatfield(2014)。

为一按电脑键盘便可随时随地轻松地进行翻译。于是,有人开始担心机器翻译将导致大量人工译员失业。虽然这只是虚惊一场,但很快第一个机器翻译系统就问世了,不过那时的翻译质量还非常初级。

此时,远在大洋彼岸的中国也开始了机器翻译的研究(冯志伟,2005:21)。1956年机器翻译便赫然列入了我国科学工作发展规划中。1957年,中国科学院语言研究所与计算技术研究所合作进行汉俄机器翻译的研究。1959年,他们的研究初见成效。测试显示:他们研制的机器翻译系统成功地翻译了9个不同句型的句子。同年,IBM公司为美国空军的国外技术部安装了第一个机器翻译系统。之后,1963年和1964年走在机器翻译研究前沿的乔治敦大学分别为欧洲原子能共同体和美国原子能机构建立了机器翻译系统。

此时的机器翻译尚处于萌芽状态,由于机器译文的低质量与人们对它的高期待相去甚远,人们对机器翻译的结果感到深深的失望。在此背景下,1964年美国政府成立了“自动语言处理咨询委员会”(Automatic Language Processing Advisory Committee),对机器翻译的可行性进行了评估。1966年,他们发布了一份题为“语言与机器”的ALPAC报告,几乎全面否定了机器翻译。报告称机器翻译生成的译文质量非常低劣,不值得政府继续为其提供赞助并建议美国终止机器翻译的研究。这份致命的报告导致大多美国的机器翻译研究项目下马了。不久,苏联也决定不再进行机器翻译的研发了,因为该国计算机的运算能力远没有美国的强大(Hutchins, 2005:6,冯志伟,2005:16)。而在中国,从1966年开始,机器翻译研究被迫中断,由此机器翻译陷入停滞状态。

## (2) 从崛起到兴盛(1970—2000)

令人欣慰的是机器翻译研究并没有从此消失,它仍在逆境中顽强地继续着。成立于1968年的SYSTRAN公司是世界上最早的机器翻译公司之一。SYSTRAN机器翻译系统是ALPAC报告发布后政府大幅削减对机器翻译资助后存活的少数几个机译系统。冷战期间,该公司用新开发的机器翻译Systran系统为美国空军进行了大量的俄英翻译<sup>③</sup>。1976年迎来了机器翻译的一个转折点。是年,加拿大气象局安装了METEO机器翻译系统,首次成功地为大众提供天气预报翻译服务。同一年,欧盟委员会决定购买Systran系统。从那以后,该系统不断更新并持续为欧盟委员会内部提供翻译服务。该委员会甚至决定投资自主研发一个比该系统更好

<sup>③</sup> SYSTRAN系统为发源于基于规则的机器翻译技术。其于2010年发布的SYSTRAN服务器7使用了基于规则/统计的混合式系统技术,是市场上出现的第一款混合式系统。SYSTRAN为Yahoo! BabelFish等公司提供技术支持。2007年前Google的语言工具也由SYSTRAN提供技术支持。详见Systran: Past and Present-[https://lilab.unibas.ch/staff/tenhacken/Applied-CL/3\\_Systran/3\\_Systran.html#history](https://lilab.unibas.ch/staff/tenhacken/Applied-CL/3_Systran/3_Systran.html#history) 以及 <https://en.wikipedia.org/wiki/SYSTRAN>。

的机器翻译系统,遗憾的是几年后项目无疾而终。在中国,经历了十年动乱后,机器翻译研究开始复苏了。1978年中国科学技术情报研究所从他们在5 000条冶金题录中提取20条进行测试,取得了英汉机器翻译的预期效果。

20世纪70年代陆续开发的一些机器翻译系统也渐渐被一些大公司接受。随着个人电脑迅速走进人们的生活和工作,1981年出现了第一个机器翻译软件,机器翻译的应用渐渐扩大开来。80年代终于迎来了机器翻译研究的复苏,日本公司开始生产商用系统。1987年,我国的“科译1号”机器翻译系统问世。这是我国第一款商用机器翻译系统,被列为我国1988年计算机界十件大事之一。后来走入大众视野的“译星”机器翻译系统便是在此基础上发展而成。90年代翻译工作站进入了市场,而进入本世纪随着互联网的迅猛发展,机器翻译成了互联网上的在线翻译(Hutchins, 2005:7),详见下文。

### 8.1.2 机器翻译的工作原理

机器翻译技术主要为基于规则和基于语料库两大类。20世纪70年代前研发的机器翻译系统多基于规则,80年代及其后多基于语料库。上述基于规则和基于语料库的机器翻译技术各有利弊,但它们在一定程度上互为补充。于是出现了综合基于规则和基于语料库的多引擎机器翻译技术。而21世纪逐渐走入人们视野的神经网络机器翻译技术则刷新了对以前机器翻译技术的认识。这里介绍基于规则机器翻译、基于语料库机器翻译和综合两者的多引擎机器翻译技术的工作原理<sup>④</sup>以及基于这些原理而开发的不同机器翻译系统。神经网络机器翻译技术见下面的8.4节。

上文中提到韦弗于1949年发表的一篇影响巨大的文章,让我们来了解一下韦弗当年的前瞻性思考以理解机器翻译的工作原理。从今天已广为接受的概念来说,他的思想可以从语言逻辑、上下文、解码等角度来诠释<sup>⑤</sup>。

#### (1) 基于规则机器翻译

从语言逻辑角度,韦弗认为文本的表达是有逻辑性的,这种逻辑性在语言形式上体现出来。所以应该对句子进行结构化的句法和语义分析,因为这是逻辑推演的基础。鉴于语义表达具有天马行空的非逻辑性,他指出上下文在解决词汇歧义中的重

<sup>④</sup> 关于机器翻译的不同系统参见冯志伟(2005:35—66)。

<sup>⑤</sup> 这里对韦弗思想的概述主要基于孙茂松、周建设(2016)。孙茂松、周建设(2016)对韦弗的重要观点阐述分为意义与上下文、语言与逻辑、密码学的角度以及多语言之间的翻译问题等四个方面,本章作者根据其与本文的关系有所变化。本文中未涉及的针对多语言之间的翻译问题,韦弗认为应研究人类通信的共同基础——普遍语言(亦称为语言的逻辑结构)。著名语言学家乔姆斯基提倡的普遍语法理论以及机器翻译研究中的“中间语言”思路皆与之相关。