

Introduction
to Computer-aided Translation
计算机辅助翻译入门

主编 吕奇 杨元刚



WUHAN UNIVERSITY PRESS

武汉大学出版社

Introduction
to Computer-aided Translation
计算机辅助翻译入门

主 编 吕 奇 杨元刚
副主编 刘文高
编 者 吕 奇 杨元刚 刘文高
梅 先 赵丹宏 骆 琪
李 聪 杜 嘉 左媛媛



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

计算机辅助翻译入门/吕奇,杨元刚主编. —武汉:武汉大学出版社,
2015.5

ISBN 978-7-307-15481-0

I. 计… II. ①吕… ②杨… III. 自动翻译系统—研究 IV. TP391.2

中国版本图书馆 CIP 数据核字(2015)第 066370 号

责任编辑:林 莉 责任校对:汪欣怡 版式设计:马 佳

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:cbs22@whu.edu.cn 网址:www.wdp.com.cn)

印刷:崇阳县天人印刷有限责任公司

开本:787×1092 1/16 印张:24.75 字数:579千字 插页:1

版次:2015年5月第1版 2015年5月第1次印刷

ISBN 978-7-307-15481-0 定价:48.00元

版权所有,不得翻印;凡购买我社的图书,如有质量问题,请与当地图书销售部门联系调换。

前 言

一、编写背景

自 2006 年和 2007 年我国分别开设翻译本科专业 (BTI) 和翻译硕士专业 (MTI) 以来, 翻译学科取得了蓬勃发展, 已逐渐成为国内高校新的学科增长点。截至 2014 年 7 月, 全国获准开办翻译硕士专业的高校已达 206 所, 开办翻译本科专业学位的高校达到 152 所。^①与此同时, 国内翻译市场的新变化又给翻译人才培养带来了影响, 为此高校必须顺应翻译行业发展趋势, 结合翻译市场需求, 培养出更多实践型、创新型、应用型翻译人才; 而“计算机辅助翻译”课程建设正是顺应了这一形势要求。全国翻译硕士专业学位教育指导委员会和教育部高等学校翻译本科专业教学协作组在近三年年会上特别强调: 应加强“计算机辅助翻译”课程建设, 重视培养学生掌握现代化翻译工具的能力。教育部高等学校翻译本科专业教学协作组 (2012: 4-8) 更是将“计算机辅助翻译”课程列为“翻译知识与技能”模块, 并将翻译工具能力列为本科翻译专业翻译能力基本要求之一。这些都凸显了“计算机辅助翻译”课程建设的重要性与紧迫性。

然而时至今日, 全球化与本地化的浪潮已席卷世界, 在翻译产业与翻译技术的发展已如火如荼之际, 不少高校师生尚不了解 MT (机器翻译) 与 TM (翻译记忆) 的区别, 对 CAT (计算机辅助翻译) 知之甚少, 甚至以为“译者所需要做的, 就是将包含源语的电子文档导入电脑, 便可静候电脑将翻译好的译文一一呈上” (Samuelsson-Brown, 2010: 83)。他们要么不曾了解、不愿了解甚至不屑了解 CAT; 在他们的认知当中, 所谓的 CAT 能做的不过是产出一些荒诞可笑的译文; 要么是对 CAT 心存敬畏, 仿佛那是充满神秘色彩而触不可及的高科技。类似这种认识上的两重误区, 不仅存在于学生群体, 也存在于一些关在象牙塔里闭门造车的高校学者之中。(吕奇, 2014: 89)

因此, 本书旨在以教材的形式, 为高校翻译本科专业 (BTI) 和翻译硕士专业 (MTI) 的“计算机辅助翻译”课程建设与实践型、创新型、应用型翻译人才培养提供参考, 其适用对象主要为翻译本科专业和翻译硕士专业学生, 也适合于计算机辅助翻译初学者和爱好者自学, 同时为国内高校从事计算机辅助翻译教学的同行提供借鉴。

二、本书特点

作为一本入门级教材, 本书在编写理念、内容安排和行文方式上具有以下三大特点:

① 数据来源: 中国译协网, http://www.tac-online.org.cn/ch/tran/2014-07/23/content_7085283.htm.

在编写理念上,本书明确定位为“计算机辅助翻译”课程学习入门级教材,在理论与实践相结合的基础上,意在突出计算机辅助翻译实践性、应用性强的特点,着眼于为读者揭开计算机辅助翻译的神秘面纱,帮助他们“走近”并“走进”计算机辅助翻译的世界——使其通过系统学习本书中的内容,厘清计算机辅助翻译的一些基本概念,消除认识上的一些误区;了解计算机辅助翻译的基本原理,掌握多种国内外主流计算机辅助翻译软件和工具的基本操作,锻炼动手能力;并对翻译项目管理和本地化翻译等相关知识形成初步的认知,以适应相关行业和领域的工作需要。

在内容安排上,本书框架明晰,内容丰富,详略得当。全书共分十章并设置了附录,分为五大模块,可供一个学期使用。

●“模块一”为本书第一章,侧重理论铺垫,主要概述计算机辅助翻译发展简史、基本概念与工作原理。

●“模块二”为本书第二章,侧重广义层面的翻译工具应用介绍。

●“模块三”包括第三章至第九章,为本书核心模块,侧重介绍与讲解数款国内外主流计算机辅助翻译软件的基本功能与操作流程。

●“模块四”为本书第十章,侧重介绍计算机辅助翻译技术在翻译项目管理中的实际运用。

●“模块五”为附录部分,侧重提供国内外计算机辅助翻译网站、论坛、视频教学资源;同时介绍翻译行业服务规范的国家标准。

上述各章均附有思考与练习题,供读者参考使用,结合素材自行操练。需要说明的是,本书第三章至第九章重点介绍的计算机辅助翻译软件,尽量选取其最新版进行讲解(例如SDL Trados Studio 2014、传神 iCAT 和雪人 CAT,国内鲜有同类教材涉及此三款软件最新版的入门指南)。即便如此,在编写期间,有的软件又发布了更新版(例如 Déjà Vu X3 和 memoQ 2014),因而这些内容,有待日后再版时予以更新和修订。

在行文方式上,本书以平实、简洁的语言对计算机辅助翻译基本概念和软件工具操作方法进行述介与讲解。在述介基本概念时,尽量避免过多使用艰涩难懂的术语;如无法避免,则尽量以脚注等形式加以解释。在讲解操作方法时,力求杜绝简单罗列操作说明书上的步骤,而是对流程中每个重要环节随列随释,予以必要的说明、强调和指导,即不仅告诉读者应当怎样做,还要告诉为何要这样做。

除了上述三大特点以外,本书在编写团队人员构成方面也有着创新尝试,即采取“校企合作”模式。编写团队成员除了在高校从事翻译教学和研究工作多年的学者,还邀请了武汉大学珞珈人翻译有限公司总经理刘文高先生、传神(中国)网络科技有限公司项目经理梅先先生和上海予尔信息科技有限公司项目经理赵丹宏先生等在翻译项目管理和计算机辅助翻译实践方面具有丰富实战经验的业内专家参与本书的审校和编写工作。这种在编写团队人员构成上的“校企合作”模式,有效保证了本书在理论和实践上能够有机结合,优势互补,避免“纸上谈兵”的现象。

综上所述,本书所起到的作用,正如为读者打开一扇门,推开一扇窗;而计算机辅助翻译更为奇幻的风景,有待读者进一步去领略、去探索、去发现。

三、编写说明

本书的编写工作得到了以下机构和人士的鼎力相助，作者谨致谢忱：

感谢传神（中国）网络科技有限公司和北京东方雅信软件技术有限公司提供以及上海予尔信息科技有限公司、佛山市雪人计算机有限公司和上海泰彼信息科技有限公司（中国代表处）授权使用的技术手册资料；感谢崔启亮博士、王华树博士、曾立人教授、闫栗丽女士，是他们在作者参加由中国翻译协会、全国翻译硕士专业学位教育指导委员会和教育部高等学校翻译专业教学协作组举办的“首届全国高等院校翻译专业师资翻译与本地化技术、翻译与本地化项目管理培训班”期间，传道授业解惑，将作者带进了计算机辅助翻译的神奇世界；感谢武汉大学出版社叶玲利编辑、林莉编辑和湖北大学外国语学院江晓梅教授，她们为本书的策划、撰写和出版等工作提供了宝贵建议和大力协助，在此表示衷心的感谢。

本书由吕奇和杨元刚共同设计编写框架和体例，刘文高和杨元刚参与部分章节的初稿审校。全书各章节具体编写情况如下：第一章由吕奇编写；第二章由杜嘉和左媛媛共同编写；第三章由吕奇和赵丹宏共同编写；第四章和第七章由骆琪和杨元刚共同编写；第五章和第九章由吕奇和李聪共同编写；第六章由梅先编写；第八章由李聪编写；第十章由刘文高编写；附录部分由吕奇和左媛媛共同整理编写；全书最后由吕奇负责统稿、修订和审校。

由于编者水平有限，而本书涉及的领域发展又日新月异，书中难免存在一些疏漏和不妥之处，敬请各位专家、同行和广大读者批评指正。

吕 奇

2015年2月

湖北大学逸夫人文楼

179	第四章 机器翻译	179
187	第六章 CAT 入门	187
187	第一节 CAT 的由来	187
192	第二节 CAT 的启动与设置	192
197	第三节 CAT 主要功能简介	197
199	第四章 WORD 操作介绍 (主翻译流程)	199
202	第五章 EXCEL 操作介绍 (主翻译流程)	202
	第一章 计算机辅助翻译概述	1
	第一节 计算机辅助翻译发展简史	1
	第二节 计算机辅助翻译基本概念	5
	第三节 国内外主流计算机辅助翻译软件简介	12
	第二章 翻译工具应用入门	31
	第一节 翻译工具概述	31
	第二节 文字识别工具的应用	36
	第三节 电子词典与在线翻译工具的应用	51
	第四节 百科全书	64
	第三章 SDL Trados Studio 2014 入门	68
	第一节 软件特色与新增功能介绍	68
	第二节 软件界面与基本设置	71
	第三节 创建翻译记忆库	74
	第四节 创建术语库	86
	第五节 文件翻译工作流程	97
	第六节 审校	114
	第四章 Déjà Vu X2 入门	117
	第一节 Déjà Vu X2 新增功能简介	117
	第二节 创建翻译记忆库	118
	第三节 创建术语库	131
	第四节 创建翻译项目	140
	第五节 项目文件翻译流程	149
	第五章 memoQ 2013 入门	159
	第一节 memoQ 2013 操作界面简介	159
	第二节 创建翻译项目	161
	第三节 项目文件翻译流程	168

第四节 资源控制台	179
第六章 传神 iCAT 入门	187
第一节 iCAT 的安装与登录	187
第二节 iCAT 的启动与设置	192
第三节 iCAT 主要功能按钮介绍	197
第四节 WORD 版操作介绍 (主翻译流程)	199
第五节 EXCEL 版操作介绍 (主翻译流程)	205
第六节 语料管理	210
第七章 雪人 CAT 入门	216
第一节 雪人 CAT 主要功能介绍	216
第二节 雪人 CAT 基本操作指南	218
第三节 雪人 CAT 网络协同翻译平台的使用	245
第八章 雅信机辅笔译教学系统入门	262
第一节 雅信机辅笔译教学系统简介	262
第二节 学生笔译教学系统平台的操作	263
第三节 教师笔译教学系统平台的操作	275
第四节 雅信项目实训子系统的操作	288
第九章 火云术语入门	302
第一节 火云术语简介	302
第二节 登录、注册与找回密码	304
第三节 查词	306
第四节 个人中心	309
第五节 术语管理	312
第六节 云术语库	325
第七节 翻译工具	329
第十章 翻译项目管理入门	338
第一节 待翻译文件的字数统计	338
第二节 文件内容重复率分析	341
第三节 译员的选择	344
第四节 项目包的分发	346
第五节 术语表的创建与分发	351
第六节 项目组的沟通	353
第七节 进度管理	357

第八节 虚拟化解决方案.....	357
第九节 收稿.....	360
第十节 审校和拼写检查.....	361
附录一 CAT 学习网络资源介绍	369
附录二 中华人民共和国国家标准：翻译服务规范.....	373
参考文献.....	381

第一章 计算机辅助翻译概述

计算机辅助翻译, 英文名为 Computer-aided Translation 或 Computer-assisted Translation, 缩写为 CAT。它不同于以往的机器翻译, 不是纯粹依赖于计算机的自动翻译, 而是在人的参与下完成整个翻译过程; 它能够帮助译者优质、高效、轻松地完成翻译工作。本章将简要回顾计算机辅助翻译发展史, 厘清计算机辅助翻译的基本概念, 并对国内外若干主流计算机辅助翻译软件进行介绍。

第一节 计算机辅助翻译发展简史^①

对于计算机辅助翻译发展史, 国内学者如张政(2006)、钱多秀(2011)、陈善伟(2014)等有着不同的划分依据, 在关键概念上也存在计算机翻译、计算机辅助翻译、电脑辅助翻译等差异。此处, 我们借鉴陈善伟(2014)的划分方式, 将计算机辅助翻译^②发展简史划分为以下四个阶段——第一阶段: 萌芽初创期(1967—1983); 第二阶段: 稳步发展期(1984—1992); 第三阶段: 迅速发展期(1993—2002); 第四阶段: 全球发展期(2003至今)。

一、第一阶段: 萌芽初创期

计算机辅助翻译源于机器翻译, 而机器翻译则始于计算机的发明。自1946年第一台计算机 ENIAC 问世以来, 机器翻译在世界各国发展迅速。1947年, 洛克菲勒基金会(Rockefeller Foundation)的瓦伦·韦弗(Warren Weaver)与英国伦敦大学伯克贝克学院(Birkbeck College, University of London)安德鲁·唐纳德·布思(Andrew Donald Booth)是最早提议用新发明的计算机来翻译自然语言的两为学者。(Chan, 2004: 290-291)

随后, 西方一些国家纷纷建立机器翻译研究机构, 相继召开国际会议, 机器翻译蓬勃发展起来。然而, 1966年, 美国语言自动处理咨询委员会(Automatic Language Processing Advisory Committee, 简称 ALPAC)公布的报告指出: 机器翻译是一次失败的尝试, 没有发展前景。受其影响, 机器翻译的研究陷入了萧条。(张政, 2006: 3)。

作为计算机辅助翻译中的一个主要概念与功能, 翻译记忆恰好在这一时期出现绝非偶

^① 本部分参考了陈善伟(2014)的相关内容。

^② 陈善伟(2014)在《翻译科技新视野》一书中使用的是“电脑辅助翻译”这一术语, 该术语在香港地区使用更为普遍; 为保证术语的统一规范性, 如无特殊说明, 本书全部使用“计算机辅助翻译”。

然。按照约翰·哈钦斯 (John Hutchins) 的说法, 翻译记忆的概念可以追溯到 20 世纪 60 年代至 80 年代。1978 年, 艾伦·梅尔比 (Alan Melby) 在杨伯翰大学 (Brigham Young University) 翻译研究组研究机器翻译和开发交互式翻译系统 (Interactive Translation System) 时, 已经将翻译记忆的概念融入“重复处理” (Repetition Processing) 工具中, 从中寻找匹配的字符串 (Melby, 1978; Melby & Warner, 1995: 187)。次年, 彼得·阿芬恩 (Peter Arthern) 在欧盟委员会就是否应该采用机器翻译的研讨会上提出“以文本检索翻译” (Translation by Text-retrieval) 的方法。由此可见, 在 20 世纪 70 年代末至 80 年代, 翻译记忆的概念已经确立。虽然哈钦斯认为首先提出翻译记忆理念的是阿芬恩, 但梅尔比与阿芬恩几乎在同一时间提出翻译记忆的概念, 因此可以并称为翻译记忆概念的先驱。(陈善伟, 2014: 3)

翻译记忆的概念最早应用于梅尔比与他在杨伯翰大学的研究伙伴共同开发的商用“自动化语言处理系统” (automated processing systems, ALPS)。该系统可提供之前译过的完全相同的分段 (Hutchins, 1998: 291)。部分学者将这种完全匹配 (full match) 的功能类型归为第一代翻译记忆系统 (Elita & Gavrilu, 2006: 24-26; Gotti et al., 2005; Kavak, 2009)。它的主要缺点在于完全匹配的句子数量极少, 翻译记忆重复使用的机会极少, 翻译记忆库的作用亦很小 (王正, 2011: 141)。

其后, 翻译记忆技术在相当长的一段时间内还只是处于探索阶段, 并未出现真正意义上的商用计算机辅助翻译系统。因此, 从 1967 年到 1983 年这段时期只能算作计算机辅助翻译发展的萌芽初创期。

二、第二阶段: 稳步发展期

1984—1992 年是计算机辅助翻译的稳步发展期。这一时期, 伴随着公司运作, 计算机辅助翻译系统逐步走向商业化, 其发展规模也日趋庞大。

1984 年, 最早的两家计算机辅助翻译公司成立, 它们分别是德国塔多思公司 (Trados GmbH) 和瑞士的 STAR 集团 (STAR Group)。这两家公司对计算机辅助翻译的发展带来了深远的影响。德国塔多思公司由约亨·胡梅尔 (Jochen Hummel) 和希科·克尼森 (Iko Knyphausen) 在斯图加特创立。该公司以软件服务商为开端, 在成立那年, 致力于 IBM 公司的翻译项目, 后来又为协助完成项目而开发计算机辅助翻译软件。因此, Trados 软件的开发与应用可视为计算机辅助翻译稳步发展的起点 (陈善伟, 2014: 4-5)。

计算机辅助翻译系统商业化始于 1988 年, 当时国际商业机器公司日本分公司的住田荣一郎 (Eiichiro Sumita) 和堤丰 (Yutaka Tsutsumi) 发布了 Easy TO Consult (ETOC) 工具, 该工具实质上是一款升级版的电子词典。虽然该系统并未使用“翻译记忆”这个术语, 而是将译文数据库依然称作“词典”, 但它显然已经基本具备了现在“翻译记忆”的基本特征 (Sumita & Tsutsumi, 1988: 2)。

1990 年, 塔多思公司发布首个术语库 MultiTerm, 它是磁盘操作系统 (DOS) 储存与记忆的多术语管理工具。这个工具极具创意, 将所有数据存储在单一及结构自由的数据库, 条目按照用者自定义属性进行分类 (Eurolux Computers, 1992a; <http://www.translationzone.com>; Wassmer, 2011)。

1991年,瑞士STAR集团的Transit 1.0 32位磁盘操作系统版向全世界发行。该版本1987年开始研发,一直只供公司内部使用。Transit 1.0的模块是当前计算机辅助翻译系统的标准功能,具有分隔但又同步的源语与译语窗口及有标记保护的专用翻译编辑器、翻译记忆引擎、术语管理组件及项目管理功能。从系统发展的角度来看,术语管理和项目管理这两个概念始于Transit 1.0。

1992年,塔多思公司发布了名为Trados的第一套商用计算机辅助翻译系统,标志着商用计算机辅助翻译系统的开端(陈善伟,2014:6)。

这一年也可以看作是计算机辅助翻译区域扩展的开端,各国生产的翻译软件进步神速:

德国塔多思公司发布了计算机辅助翻译系统Translator's Workbench I及DOS版的Translator's Workbench II。此外,塔多思公司开始在全球建立分公司,扩大市场(Brace,1994;Eurolux Computers,1992;http://www.translationzone.com;Hutchins,1988:287-307)。美国国际商业机器公司发布Translation Manager/2(TM/2),配备已整合在演示管理界面(presentation manager interface)下的多种翻译辅助工具的操作系统Operating System/2(OS/2)。此系统可能是第一个加入机器翻译系统的混合计算机辅助翻译系统(Brace,1993;Wassmer,2011)。此外,斯韦特兰娜·索科洛娃(Svetlana Sokolova)和亚历山大·谢列布里亚科夫(Alexander Serebryakov)这两位俄罗斯专家于1991年在圣彼得堡创建Promt公司。该公司除了研发机器翻译技术外,还向用户提供全方位的翻译解决方案。同时,英国也成立了专门从事翻译软件生产的公司。由马克·兰开斯特(Mark Lancaster)成立的SDL国际有限公司,提供软件全球化服务(陈善伟,2014:7)。

三、第三阶段:迅速发展期

1993—2002年是计算机辅助翻译的迅速发展期。之所以将其划分为迅速发展期,主要是以下列三个方面为特征:

一是商用计算机辅助系统越来越多,其内置功能越来越强大。1993年以前,市场上只有三个计算机辅助翻译系统,即塔多思Translator's Workbench II、国际商业机器公司的TM/2及STAR集团的Transit 1.0。1993年到2003年这10年间,约有20个计算机辅助翻译系统经开发后在市场销售,包括被人熟知的Déjà Vu、Eurolang Optimized、Wordfisher、SDLX、ForeignDesk、TransSuite 2000、雅信CAT、Wordfast、Across、OmegaT、MultiTrans、华建、Heartsome^①及译经。在这个时期,商用计算机辅助翻译系统增长速度是以前的6倍。同时,这一阶段开发的基本组件功能更多,有更多组件逐渐嵌入计算机辅助翻译系统中。在所有新开发的功能中,对齐、机器翻译及项目管理的工具最为突出。例如塔多思的Translator's Workbench II中嵌入了T Align对齐功能,后来被称为WinAlign。其他系统例如Déjà Vu、SDLX、Wordfisher及MultiTrans也相继采取这一行动。机器翻译也嵌入计算机辅助翻译系统之中,并用于处理翻译记忆库中无法找到的字段(陈善伟,2014:8-13)。值

^① 瀚特盛(Heartsome)公司,因财务状况等方面的原因,已于2014年7月31日停止运营,并终止任何技术支持。

值得一提的是，随着市场领导地位越来越巩固，这一时期的塔多思，已然成为业界先锋和市场领导者，拥有包括 Trados Translator's Workbench (Windows 和 DOS 版本)、MultiTerm Pro、MultiTerm Lite 及 Multi Term Dictionary 等一系列翻译软件。

二是窗口操作系统 (windows operating system) 越来越居于主导地位。1993 年之前建立的计算机辅助翻译系统是在 DOS 或 OS/2 系统上进行的。1993 年这些系统的 Windows 版纷纷推出，逐渐成为主流。例如，国际商业机器公司和塔多思公司分别在年中发布 Windows 版的 TM/2 和 Translator's Workbench，当年 6 月在西班牙发布的 ATRIL Déjà Vu 1.0 初版，其他在 Windows 运行的新发布的系统包括 SDLX、ForeignDesk、Trans Suite 2000、雅信 CAT、Across、MultiTrans、华建及译经 (陈善伟，2014: 13)。

三是支持的文件格式越来越多，支持翻译的语言数目越来越大。这阶段的计算机辅助翻译系统可以直接或通过过滤器 (filter) 处理更多的文件格式，包括 Adobe InDesign、FrameMaker、HTML、Microsoft PowerPoint、Excel、Word、QuarkXPress 甚至是 PDF 等。相比于 1992 年的 Translator Workbench Editor (只支持 5 种欧洲语言，即德语、英语、法语、意大利语和西班牙语)，这一时期由国际商业机器公司研发的 TM/2 能够支持 19 种语言，包括汉语、韩语和其他 OS/2 兼容的编码字符集。这在很大程度上是由于 1994 年统一码 (Unicode 3.0) 的贡献。统一码提供了任何语言在所有现代软件的文本数据处理、存储和交换的依据，从而让计算机辅助翻译系统开发商逐步解决语言处理上的障碍 (陈善伟，2014: 14)。

四、第四阶段：全球发展期

2003 年以后，计算机辅助翻译在全球众多国家都取得了长足的发展，翻译技术继续走向全球化。北美 (如加拿大、美国等国)、欧洲 (如英国、法国、德国、匈牙利、卢森堡、瑞士、捷克、俄罗斯等国)、亚洲 (如中国、日本等国) 均研发或升级了最新的计算机辅助翻译系统，其功能也日趋强大。陈善伟 (2014: 28-29) 将这段时期呈现出的主要趋势归纳为以下四点：

一是注重 Windows 和 Microsoft Office 的系统兼容性。目前市场上的 67 种商用系统，只有一种不可以在 Windows 操作系统上运行。许多计算机辅助翻译系统为求兼容，都力求紧跟 Windows 和 Microsoft Office 的前进步伐。

二是将工作流程控制嵌入计算机辅助翻译系统中。除了再用或循环使用重复性文本和文本术语的翻译外，在此期间开发的系统都有新增功能，例如项目管理、拼写检查、质量保证和内容控制。以 SDL Trados Studio 2011 为例，该版本可为大多数语言进行拼写检查，还配备了 PerfectMatch 2.0，可对源文件进行追踪修订。市场上大多数系统都可以执行语境匹配功能，即翻译文件和翻译记忆库中的上下文语段是完全相同的匹配。

三是利用网络或在线系统。由于新的信息技术快速发展，这一阶段大多数计算机辅助翻译系统以服务器、网络甚至是云端为基础，有很大的数据存储量。到 2012 年底，市场上可供个人或企业使用的云端计算机辅助翻译系统共有 15 种。如 Lingotek Collaborative Translation Platform，SDL World Server 及 XTM Cloud (Muegge, 2012: 17-21)。

四是采用计算机辅助翻译系统新格式。由于不同计算机辅助翻译系统有不同格式，系

统与系统之间的数据交换一直是一个棘手的问题，如 Déjà Vu X 的格式是 dvmdb，而 SDL Trados Translator's Workbench 8.0 的格式是 tmw。由于这些特定格式的程序不能相互识别，行业内便不能实现数据共享。过去，国际本地化行业标准协会在数据交换标准的发展和推广中曾起过显著作用，如断句规则交换标准，翻译记忆交换格式、术语库交换标准及 XML 本地化交换文件格式。可以预测，为进一步方便数据交换，未来系统的发展方向之一是格式要符合行业标准。

所有数据表明，在过去几十年间，翻译技术发展日新月异，且必将在未来数十年继续保持其发展势头（钱多秀，2011）。

第二节 计算机辅助翻译基本概念

一、计算机辅助翻译 = 机器翻译？

不少计算机辅助翻译初学者，时常混淆计算机辅助翻译和机器翻译两个概念之间的差别，事实上二者在核心技术、工作原理和人工参与程度等方面均有本质不同。钱多秀（2011：2）曾指出，多年来大部分人想当然地以为计算机辅助翻译就是机器翻译，是全自动的。因此我们只有厘清计算机辅助翻译与机器翻译的几个关键概念，消除误解，才能为下一步的学习作好铺垫。这种区分有助于初学者对 CAT 有更为准确的定位和认知。下面我们首先初步了解一下机器翻译的基本概念与工作原理。

1. 机器翻译的基本概念

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（冯志伟，2007）。

Hutchins & Somers（1992：148）将翻译过程中人与机器的参与程度和不同角色总结为图 1.1。

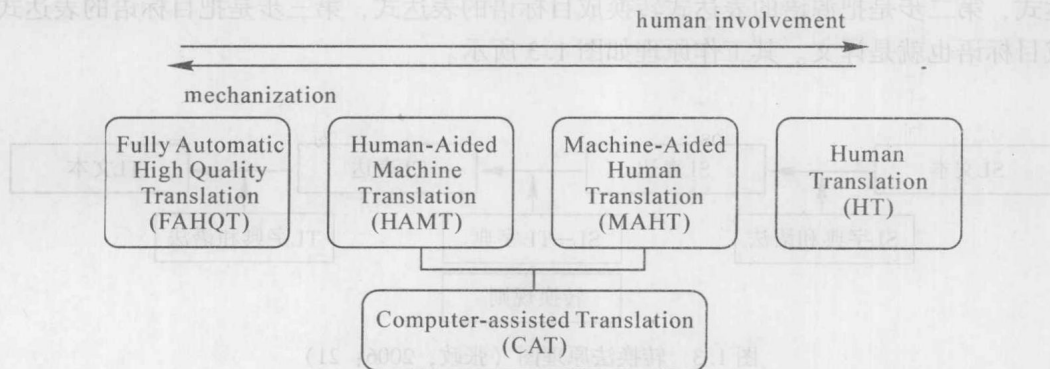


图 1.1 机器翻译与人工翻译参与程度关系图

图 1.1 最左侧一级为全自动高质量机器翻译 (fully automatic high quality machine translation, 简称为 FAHQT), 此时机器参与程度最大, 人工参与程度最小 (无限接近于零), 是最接近真正意义上的机器翻译形式, 也是最为理想的一种机器翻译形式, 但目前在现实中难以实现; 最右侧一级为人工翻译 (human translation, 简称为 HT), 此时人工参与程度最大, 机器参与程度最小 (无限接近于零)。图示中居中的两级, 称作机器辅助翻译 (machine assisted translation, 简称为 MAT), 是利用人机交互来共同完成翻译任务的一种方式, 也是最接近真正意义上的计算机辅助翻译形式。MAT 又可以进一步划分为人辅机译 (human-aided machine translation, 简称为 HAMT) 和机辅人译 (machine-aided human translation, 简称 MAHT)。二者的区别在于, 人辅机译以机器为主, 人在关键部分或计算机处理较难的部分给机器以辅助, 如消歧; 而机辅人译以人为主, 机器帮助人进行各种简单的翻译工作, 如查词典、查例句等。

2. 机器翻译的工作原理^①

根据核心技术和算法不同, 机器翻译可分为四种基本类型:

(1) 基于规则的机器翻译

基于规则是指机器翻译系统建立在语言规则的基础上, 包括直接法、转换法和中间语法法。

直接法是指把源语 (source language, 简称为 SL) 的单词和句子直接替换成目标语 (target language, 简称为 TL) 的单词和句子, 在必要的时候对语序进行适当调整 (Hutchins & Somers, 1992: 72)。其工作原理如图 1.2 所示。

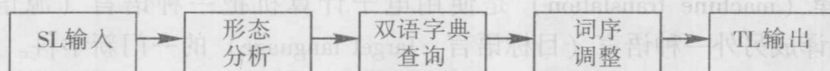


图 1.2 直接法原理图 (张政, 2006: 19)

转换法是指利用中间表达式在源语和目标语之间过渡, 又可进一步分为句法转换和语义转换 (Hutchins, 1986: 54)。转换法分三个步骤进行: 第一步是把源语转换成源语的表达式, 第二步是把源语的表达式转换成目标语的表达式, 第三步是把目标语的表达式转换成目标语也就是译文。其工作原理如图 1.3 所示。



图 1.3 转换法原理图 (张政, 2006: 21)

^① 本部分参考了钱多秀 (2011) 的相关内容。

中间语言法是指把源语转换成一种无歧义的、对任何语言都通用的中间语言 (interlingua), 再用目标语的词汇和句法结构表达中间语言的意义 (Hutchins, 1986: 55)。其工作原理如图 1.4 所示。

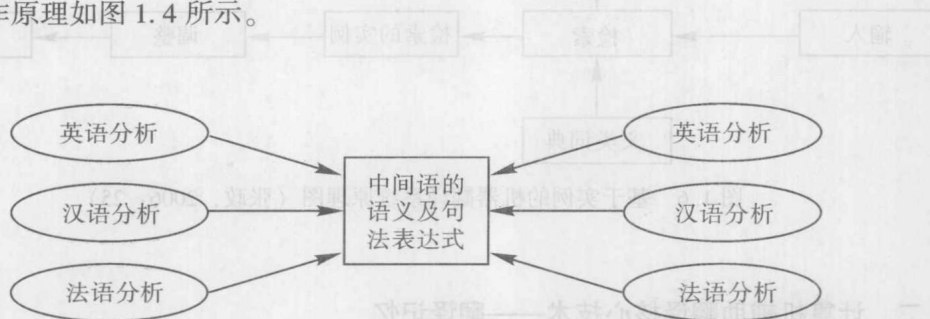


图 1.4 中间语言法原理图 (张政, 2006: 21)

(2) 基于统计的机器翻译

基于统计的机器翻译方法把机器翻译看成是一个信息传输的过程, 用一种信道模型对机器翻译进行解释。源语言句子到目标语言句子的翻译是一个概率问题, 任何一个目标语言句子都有可能是任何一个源语言句子的译文, 只是概率不同, 机器翻译的任务就是找到概率最大的句子。具体方法是将翻译看做对原文通过模型转换为译文的解码过程。因此统计机器翻译又可以分为以下几个问题: 模型问题、训练问题、解码问题。所谓模型问题, 就是为机器翻译建立概率模型, 也就是要定义源语言句子到目标语言句子的翻译概率的计算方法。而训练问题, 是要利用语料库来得到这个模型的所有参数。所谓解码问题, 则是在已知模型和参数的基础上, 对于任何一个输入的源语言句子, 去查找概率最大的译文。基于统计的机器翻译工作原理如图 1.5 所示。

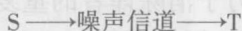


图 1.5 基于统计的机器翻译工作原理图 (冯志伟, 2004: 45)

(3) 基于实例的机器翻译

与基于统计的机器翻译方法相同, 基于实例的机器翻译方法也是一种基于语料库的方法, 其基本思想由日本著名的机器翻译专家长尾真提出。他研究了英语初学者的基本模式, 发现初学英语的人总是先记住最基本的英语句子和对应的日语句子, 而后做替换练习。参照这个学习过程, 他提出了基于实例的机器翻译思想, 即不经过深层分析, 仅仅通过已有的经验知识, 通过类比原理进行翻译。其翻译过程是首先将源语言正确分解为句子, 再分解为短语碎片, 接着通过类比的方法把这些短语碎片译成目标语言短语, 最后把这些短语合并成长句。对于实例方法的系统而言, 其主要知识源就是双语对照的实例库, 不需要什么字典、语法规则库之类的东西, 核心的问题就是通过最大限度的统计, 得出双语对照实例库。基于实例的机器翻译系统原理图如图 1.6 所示。

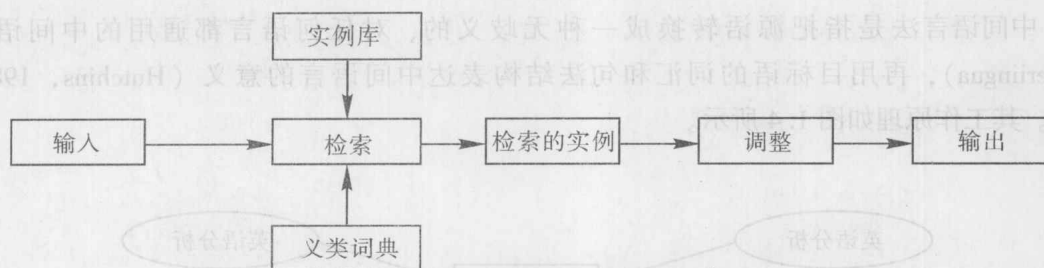


图 1.6 基于实例的机器翻译系统原理图 (张政, 2006: 25)

二、计算机辅助翻译核心技术——翻译记忆

1. 翻译记忆的基本概念

所谓翻译记忆 (亦称翻译内存、翻译记忆库, 英文名为 translation memory, 缩写为 TM), 是指将已经翻译并匹配好的源语和目标语字段储存在数据库中以备将来重新利用。这些已翻译好的字段仍然是人工翻译的结果。它的特点是能够将翻译流程中涉及纯粹记忆的活动, 比如术语的匹配和自动搜索提示、高度相似句子的记忆和复现交给计算机来做, 免除翻译人员反复查找名词之苦, 使其能全力对付语义的转换和传递 (Somers, 2003: 31)。

2. 翻译记忆的工作原理

由于专业翻译领域所涉及的翻译资料数量巨大, 而范围相对狭窄, 集中于某个或某几个专业, 如政治、经济、军事、航天、计算机、通信等专业都有自己的专业翻译公司或部门, 这就必然带来翻译资料的不同程度的重复。据统计, 在不同行业和部门, 这种资料的重复率达到 20%~70%。这就意味着译者至少有 20% 以上的工作是无谓的重复劳动。翻译记忆技术就是从这里着手, 首先致力于消除译者的重复劳动, 从而提高工作效率。

翻译记忆 (TM) 的工作原理是: 译者利用已有的原文和译文, 建立起一个或多个翻译记忆库 (translation memory), 在翻译过程中, 系统将自动搜索翻译记忆库中相同或相似的翻译资源 (如句子、段落), 给出参考译文, 使译者避免无谓的重复劳动, 只需专注于新内容的翻译。翻译记忆库同时在后台不断学习和自动储存新的译文, 变得越来越“聪明”, 效率越来越高。如果用通俗的语言来概括翻译记忆的作用, 那就是: 有了翻译记忆, 译者永远不必对同一句话翻译两遍。

以图 1.7 为例, 译者将双语对齐句段 “XYZ is a software company” 和 “XYZ 是一个翻译公司” 存储到翻译记忆库中, 后又将 “XYZ provides Translation Memory technology” 和 “XYZ 提供翻译记忆技术” 存储到同一个翻译记忆库中。以此类推, 该翻译记忆库会通过不断积累, 存储越来越多的双语对齐句段。译者每次翻译新的文档, 程序会先分析这段文字, 试着在记忆库里找寻当前翻译句段是否与过去曾经翻译过的句段相符。如果找到相符的句段, 则会呈现出来给译者检阅编辑。例如, 当译者翻译的当前句段为 “XYZ is a software company” 时, 会提示译者该句段与记忆库中存储句段 100% 匹配; 当译者翻译的