

第一章

机器翻译与计算机辅助翻译

一、机器翻译

从古希腊时代起，人们就梦想着用机器翻译语言。直到 1903 年，法国语言学家 Couturat 和法国数学家 Leau 在《通用语言的历史》一书中才第一次提出了“机器翻译”这个概念（梁三云，2004）。法国科学家 Artsouni 于 1933 年发明了“机械脑”，并申请了“翻译机”的专利。同年，苏联发明家 Troyanski 发明了用机械方法把一种语言转换成另一种语言的机器，并登记了自己的发明专利（钱多秀，2011）。

1. 机器翻译的定义

机器翻译（Machine Translation），简称机译或 MT，又称为计算机翻译（Computer Translation）、电脑翻译或自动翻译（Automatic Translation），是利用计算机将一种自然语言（源语言）转换为另一种自然语言（目标语言）的翻译技术（Hutchins，1986）。它涉及计算语言学（Computational Linguistics），是把计算机科学、认知科学、语言学、信息技术等诸多学科有机融合的新领域，是人工智能发达程度的终极体现之一，是自然语言理解、知识表达和机器翻译所共同面临的人工智能难题，不但具有重要的科学研究价值，还具有重要的实用价值。随着经济全球化及互联网的飞速发展，机器翻译技术将起到越来越重要的作用。

2. 机器翻译简史

机器翻译从产生至今经历了跌宕起伏的发展历程。有研究者认为大致可分为四个时期：萌芽开创期、发展受挫期、恢复期和突飞猛进期。

(1) 萌芽开创期

1946年，世界上诞生了第一台电子计算机，它的运算速度在当时看来快得难以置信，这使得翻译人员开始考虑如何利用计算机的高速运算功能进行自然语言的翻译。美国科学家 Weaver 首先意识到翻译也是一种解码过程，可以通过计算机来处理。受他的影响，1954年，美国乔治城大学在国际商业机器公司（IBM）协同下，用 IBM-701 计算机成功完成了英俄机器翻译试验，首次向公众和科学界证实了机器翻译的可行性，同时也拉开了机器翻译研究的序幕。此后，机器翻译研究在美国、苏联、日本、意大利、比利时、英国、德国等国如火如荼地开展起来。

中国紧随其后，早在 1956 年就把机器翻译研究列入了国家科学工作发展规划。次年，中国科学院语言研究所和计算技术研究所合作开展了俄汉机器翻译试验，翻译了 9 种不同类型的较为复杂的句子。

从 20 世纪 50 年代开始到 20 世纪 60 年代前期，机器翻译研究呈不断上升的趋势。美苏两个超级大国出于军事、政治、经济目的，都对机器翻译项目提供了大量的资金支持，而欧洲国家由于地缘政治和经济的需要也对机器翻译研究给予了高度重视，机器翻译火爆一时。这个时期机器翻译虽然只处于萌芽阶段，但已然开始进入了盲目乐观的繁荣期。

(2) 发展受挫期

由于早期的计算机系统相对简单，仅限于对词汇和语法规则的编程，缺乏句法分析功能，这一语义分析上的缺陷使研究者们陷入了困境。为了对当时机器翻译的研究进展做出客观评估，美国科学院于 1964 年成立了语言自动处理咨询委员会，开始了历时两年的全面调查分析与测试。

1966 年 11 月，该委员会公布了关于当时机器翻译技术发展状况的分析与测试结果，结论是：与人工翻译相比，机器翻译较慢，而且不够准确，但其成本却比人工翻译高出一倍，看不到任何前景（钱多秀，2011）。该报告全面否定了机器翻译的可行性，并建议停止对相关项目的资金支持。

这一报告的结论给正在火热发展中的机器翻译研究浇了一大盆冷水，从此该项研究在世界大部分地区陷入僵局，在中国也被搁置了。

(3) 恢复期

步入 20 世纪 70 年代后，随着科学技术的突飞猛进和各国科技情报交流的日趋频繁，传统的人工翻译效率已远远满足不了各国之间的沟通需求。相比之下，计算机在提高翻译效率上具有先天的优势。与此同时，计

计算机科学、语言学研究的发展，特别是计算机硬件技术的大幅提高，以及人工智能在自然语言处理上的有效应用，从技术层面上推动了机器翻译研究的复苏，各种实用的系统也如雨后春笋般应运而生。许多国家成功研发了不同的翻译系统，其中有些系统运行稳定，至今仍在沿用。加拿大、法国和德国成功研发出至今仍具有相当影响力的 SYSTRAN 系统，后来该系统先后被美国空军和欧共体采用，以处理快速增长的多语文件。随后加拿大蒙特利尔大学和加拿大联邦政府翻译局联合研发的 TAUM-METEO 英法翻译系统投入使用（英语和法语具有很高的语言相似度），着重天气预报资料的双语翻译，每小时翻译 6 万到 30 万词。因为天气预报这一特定领域使用的语言可以被精确限定，此系统每天可翻译 1500 到 2000 篇天气预报资料供电视、报纸使用。METEO 是机器翻译发展史上的里程碑，它的出现标志着机器翻译研究进入了繁荣时期（冯志伟，2004）。

与此同时，中国的机器翻译研究也开始复苏，特别是 80 年代中期以后，发展进一步加快：1987 年中国军事科学院研制的 KY-1 英汉机器翻译系统获得国家科技进步二等奖；中国科学院计算技术研究所研制的 863-IMT 英汉机器翻译系统获得了国家科技进步一等奖（李鲁，2002）。这些成就表明中国在机器翻译技术方面取得了巨大进步。

（4）突飞猛进期

从 20 世纪 40 年代至今，电子计算机硬件经历了电子管时代、晶体管时代、集成电路时代、超大规模集成电路时代，现已发展到了智能化微型电子计算机时代，可以进行思维、学习、记忆、网络通信等工作。

同时，随着互联网、互联网+及大数据的普遍应用，经济全球化进程的高速运行以及国际社会交流的日渐频繁，翻译产业化已经形成并且发展迅速；传统人工翻译的方式到了这一时期，已经无法满足迅猛增长的翻译需求，人们对于机器翻译的渴求空前增长，机器翻译又迎来了一个崭新的发展机遇。中国在这方面取得了前所未有的成就，相继推出了一系列机器翻译软件，例如“译星”“雅信”“灵格斯”“有道”“金山词霸”等等，标志着中国商用机器翻译系统开始迈入实用化阶段。

随着互联网技术、大数据、人工智能技术突飞猛进的发展，各类互联网公司纷纷成立了机器翻译研究组，研发了基于互联网大数据的机器翻译系统，使机器翻译走向千家万户，例如“百度翻译”“谷歌翻译”“有道翻译”“爱词霸在线翻译”等。

为了展示机器翻译近年来的发展，我们做了一个机器翻译准确率的纵向对比小实验。实验基于笔者教学的积累，在 2016 年、2017 年和 2019

年分别使用百度机器翻译翻译了几组相同的句子，从中可以看出其巨大的进步。下面编号 a) 表示的是 2016 年百度机器翻译的结果；编号 b) 表示的是 2017 年百度机器翻译的结果；编号 c) 表示的是 2019 年百度机器翻译的结果。

1. 小刘提交的报告有很大的水分。
 - a) Liu report a lot of water.
 - b) Liu submitted a report of a lot of water.
 - c) Xiao Liu's report is very watery.
2. 植物是靠它的根从土壤中吸收水分。
 - a) Plants are to close its roots to absorb moisture from the soil.
 - b) Plants absorb water from the soil by its roots.
 - c) Plants absorb water from the soil by the roots.

近年来，随着语音识别、计算机真人发音等人工智能技术的日趋成熟，机器翻译技术研究开始迈入口译领域，并向手机应用延伸，出现了不少手机翻译 app。这些手机 app 依托互联网大数据，大都支持语音识别和双语互译，比如“微软翻译”“谷歌翻译”“有道翻译官”等等，都达到了相当高的实用阶段。

3. 机器翻译的质量

或许有些人认为机器翻译偏差大，不能帮助人们解决语言沟通障碍，一个简单的实验就能改变这种看法：先选一种自己完全不懂的语言，比如德语、法语、日语等，然后在互联网上随便访问一个该种语言的网站，随机找出一段话，将其复制粘贴到“百度在线翻译”或“有道在线翻译”里，把目标语言设置成中文，最后看看翻译的结果。一般情况下，我们可能会发现很多句子不太通顺，但是并不影响我们对原文意思的大致理解。在机器翻译的帮助下，我们从完全看不懂原文，到能通过译文对原文内容有大致了解，甚至可以完全理解原文内容，充分说明了机器翻译大有可为。

机器翻译不可避免地存在误差的原因在于，机器翻译运用计算语言学原理，机器自动识别语法，调用存储的词库，自动进行对应翻译，但是由于语法、词法、句法发生变化或者规则不定，语境不同，因此机器翻译的机械过程与语言灵活多变的特质必然产生矛盾，所以出现误差甚至错误在所难免。近年来，随着语料库语言学、智能翻译记忆库和神经网络翻译技术的发展和使用，机器翻译的准确性在不断提高，尤其是在专业领域，机器翻译可以大显身手。

中国数学家、语言学家周海中指出：要提高机译的译文质量，首先要解决的是语言本身问题而不是程序设计问题；单靠若干程序来做机译系统，肯定是无法提高机译的译文质量的。同时，他还指出：在人类尚未明了大脑是如何进行语言的模糊识别和逻辑判断的情况下，机译要想达到“信、达、雅”的程度是不可能的。这一观点道出了制约译文质量的瓶颈所在。

人类对机器翻译的偏见来自对高质量自动化翻译的单向要求，认为所有的翻译都要求高质量。其实，随着翻译对象的多样性不断增加，翻译译文的质量要求也是多元化的。有些翻译并不要求太高的译文质量，只要读者能够了解基本信息即可；而这类稿件可能对于翻译的时效性要求很高，例如电子商务网站的用户评论、网上新闻、学术论文摘要等。这些场景下，机器翻译是最有效的选择。

机器翻译发展至今，影响其发展的最大障碍就是译文的质量。就已有的发展来看，机译的质量离理想目标仍有距离。不论怎样，目前是机译的蓬勃发展时期，这种进步是建立在译界对机译客观认识和理性思考的基础之上的。我们也有理由相信，在计算机专家、语言学家、心理学家、逻辑学家和数学家等各个领域诸多专家的共同努力下，机译的瓶颈问题或将最终得以解决。

二、计算机辅助翻译

1964年，美国语言自动处理咨询委员会在对机器翻译状况进行调研的基础上，建议开发机器辅助工具供译员使用，并继续资助计算机语言方面的基础研究，由此掀开了计算机辅助翻译研究的第一页。

1. 计算机辅助翻译的定义

计算机辅助翻译是以人为主体进行的翻译活动，区别于全自动化的机器翻译。关于计算机辅助翻译的概念，国内外诸多学者均有论述，国外的如Bowker、Hutchins、Kay、Kenny、Melby、Quah、Somers等，国内的如冯志伟、徐彬、张政、苏明阳、钱多秀、俞敬松、王华树等。归纳起来，其概念大致可分为狭义和广义两类。

狭义的计算机辅助翻译是指利用翻译记忆的匹配技术提高翻译效率的翻译技术。它利用计算机模拟人脑记忆功能的机制，将翻译过程中简单、重复性的记忆活动交给计算机来做，将译者从机械性的工作中解放出

来，全力关注翻译本身的问题。这种方式可以称为“机助人译”，国外的 Trados、Déjà Vu、Wordfast、memoQ 等主流的计算机辅助翻译工具，以及国内的雅信 CAT、Transmate、雪人 CAT 等工具皆属于此类技术范畴。

广义的计算机辅助翻译技术则不限于此，可以涵盖译者在翻译过程中可能用到的提高翻译效率的任何信息技术，如在线词典、语料库、格式转换、光学字符识别、桌面搜索等等（王华树，2014）。

2. 计算机辅助翻译简史

计算机辅助翻译（Computer-aided Translation, CAT）技术经历了 50 多年的发展，可以分为四个时期：萌芽开创期、平稳发展期、迅速发展期和全球发展期。

（1）萌芽开创期

CAT 的诞生源于机器翻译的消极表现，其萌芽开创期的起点始于 1966 年美国语言自动处理咨询委员会对机器翻译进行的调查。也就是说，CAT 的萌芽期与机器翻译的受挫期几乎是重合的。从机器翻译的定义和整个发展简史中也不难看出，从狭义上讲，机器翻译发展中后期的 M（MT）与 CAT 的 C 都是指 Computer。也正是在这一时期，产生了影响至今的 CAT 核心概念，即翻译记忆（Translation Memory）（Hutchins, 1998）。1978 年，Melby 将翻译记忆理念运用到了他的 Repetitions Processing 工具之中，为其机器翻译研究小组研发的互动翻译系统搜寻匹配字符串。

（2）平稳发展期

这一时期，欧洲出现了 CAT 的商业化运营。1984 年，欧洲出现了两家后来具有世界影响力的 CAT 公司，分别是德国的 Trados 和瑞士的 Star Group。

Trados 公司 1984 年成立于德国斯图加特。Trados 这一名称源自三个英语单词，即 Translation、Documentation 和 Software。其中，在 Translation 中取了 TRA，在 Documentation 中取了 DO，在 Software 中取了 S，把这些字母组合起来就是 Trados 了。透过这三个英语单词的含义，我们可以知道，其取名用意恰恰体现了 Trados 软件所要实现的功能和用途。该公司在 20 世纪 80 年代后期开始研发翻译软件，并于 90 年代发布了自己的第一批 Windows 版本软件，1992 年的 MultiTerm 和 1994 年的 Translator's Workbench。

瑞士的 Star Group 公司同样作为语言服务供应商而成立，为整个信息生命周期流程（内容创建、图片、翻译、本地化和印务）提供支持。尤其是该公司除了为世界 500 强公司提供标准化的文档制作和翻译服务外，还承担了美国标准和德国标准的军工项目文档翻译，并直接为瑞士军队项目提供备件目录、培训材料、保养和维修文档及军队专用操作手册的翻译服务。

（3）迅速发展期

在 1993 年以前，计算机辅助翻译市场仅有 Trados Translator's Workbench II、IBM Translation Manager 2 和 STAR Transit 1.0 等系统。之后的十年间，市场上出现了包括 Déjà Vu、Eurolang Optimizer、Wordfisher、ForeignDesk、Trans Suite 2000、Yaxin CAT、Wordfast、OmegaT、MultiTrans、Heartsome 等在内的超过 20 种 CAT 系统。1997 年，微软采用 Trados 进行软件的本土化翻译，Trados 公司从而在 90 年代末期成为桌面翻译记忆软件行业的领头羊。Trados 软件基于翻译记忆库和术语库技术，为快速创建、编辑和审校高质量翻译提供了一套集成的工具。

在这一时期，Déjà Vu 的发展也格外引人注目。Déjà Vu 1.0 于 1993 年发布，是当时首款嵌入 Microsoft Word 界面的 CAT 软件，该软件随后的更新版本 Déjà Vu 1.1 不但各项性能有明显提升，还嵌入了对齐工具（当时的 CAT 对齐工具是非常昂贵的独立产品），为当时的翻译市场设定了新的标准。Déjà Vu 在当时不但因物美价廉而受到青睐，而且在业内许多方面都创造了第一（Chan Sin-wai, 2014）：第一款为 Windows 操作系统而设计的翻译记忆工具，第一款直接嵌于 Microsoft Word 的翻译记忆工具，第一款 32-bit 翻译记忆工具（Déjà Vu 2.0），第一款面向大众的专业翻译软件。

（4）全球发展期

这一时期科技发展的突飞猛进促使 CAT 系统功能获得日新月异的提升。在过去的十几年里，大多数旧的系统已经升级换代或淘汰关闭，消费者的需求灵活多样，CAT 市场的竞争空前激烈。2005 年，Trados 被 SDL 公司收购，翻译软件从此以 SDL Trados 冠名，从嵌入 Microsoft Word 界面的 SDL Trados 2007 到拥有独立翻译项目管理平台的 SDL Trados Studio 2017，实现了质的飞跃。

除了上面提到过的老牌 CAT 公司及其代表性 CAT 软件外，在这一时期还陆续出现了匈牙利 Kilgray 公司的 memoQ、日本 Rozetta 公司的 TraTool、北京东方雅信软件技术有限公司的 Yaxin CAT 2.0 等等。许

多公司的 CAT 软件也在不断地更新换代，比如 SDL Trados 陆续推出了 SDL Trados 2007、2009、2011、2014、2015 和 2017 等升级版本（Chan Sin-Wai, 2014）。

在这一时期，几乎所有的 CAT 软件都向着与 Windows 和 Microsoft Office 兼容并保持同步的方向发展，其中 Wordfast 的发展最为明显；向着工作流程控制与 CAT 系统集成方向发展，SDL Trados 的发展最具典型性；向着提供互联网翻译服务方向发展，例如各个 CAT 软件的联网机器翻译模块或线上服务模块；向着适应行业统一格式、实现信息共享与互通方向发展等。

3. 机器翻译与计算机辅助翻译的关系

从 20 世纪 40 年代开始至今，电子计算机硬件的存储量越来越大，运算速度越来越快，操作系统越来越复杂，然而计算机操作却越来越简单，因此普及率也越来越高。在这一背景之下，机器翻译和计算机辅助翻译的软硬件开发也纷至沓来，并且两者在不断地融合，协同应对日趋纷繁复杂的翻译任务，极其显著地提高了翻译的效率，如 SDL Trados Studio 配合 SDL BeGlobal、谷歌云翻译、SDL WordServer、MyMemory Plugin 和 SDL 自动翻译远程服务器等机器翻译模块。相关软件也像当初微软从 DOS 系列到 Windows 系列的跨越一样，CAT 软件设计越来越方便译员的翻译操作，从抽象走向直观，因而普及率也越来越高，从 SDL Trados 2007 以及之前的基于文本操作的若干版本到 SDL Trados Studio 2009-2019 基于项目操作的几个版本的飞跃就是最好的例证。

从计算机、机器翻译和计算机辅助翻译的发展历程来看，计算机技术的发展是机器翻译研究和计算机辅助翻译研究发展的前提和推动力。机器翻译虽先于计算机辅助翻译的产生和发展，但机器翻译的局限性引起了有识之士对于机器、人工智能和人类智慧的思考，从对机器的迷信中恢复理智，在肯定人类智慧不可替代的同时，把视角转向了机器与人类智慧的结合，从而促进了计算机辅助翻译研究的诞生与发展，也反过来促进了机器翻译自身的发展。

与此同时，机器翻译和计算机辅助翻译在实际翻译工作中遇到的诸多问题，也推动了计算机技术针对翻译工作软硬件的开发，促进了计算机相关软硬件技术的提升。一方面，计算机辅助翻译借助于全自动机器翻译，有助于译员提高对原文和译文的理解，从而提高翻译效率和翻译的准确性。另一方面，机器翻译相对于计算机辅助翻译来说，在翻译效率上优势

明显，在翻译质量上有所欠缺，但这一缺陷可以在大数据的环境下得到不断改进。因此，机器翻译和计算机辅助翻译通过互相取长补短来达到翻译质量与效率的共同提高。计算机、机器翻译和计算机辅助翻译既有先后出现的历时性，又有相辅相成的共时性，其发展既承前启后又同步向前，其功能既相对独立又彼此协同。三者最终走到一起是历史使然，更是翻译工作的现实需求和大势所趋。

三、小结

本章从机器翻译和计算机辅助翻译的定义和简史谈起，着重介绍了它们各自发展历程中的四个时期，并简要介绍了它们在中国的发展概况。机器翻译在发展过程中的受挫给计算机辅助翻译发展带来了机遇和空间，后者是对前者的继承，又有着截然不同的理念：机器翻译是“机主人辅”，计算机辅助翻译是“人主机辅”。然而，随着科技的不断进步，以及机器模仿人脑思维过程技术的日趋成熟，“机主人辅”的呼声日渐高涨，而计算机辅助翻译软件也不断更新换代，以 SDL Trados 为代表的计算机辅助翻译系统将机器翻译技术嵌入其中。虽然“人”和“机”在未来的翻译任务中谁会扮演主角，翻译界还在激烈的争论之中，但是两者历史上承前启后、相辅相成，如今互相促进、共同发展的事实是谁也无法否定的。

思考与讨论

1. 什么是机器翻译？最早的机器翻译系统是什么？
2. 简要说明机器翻译的发展受挫期和计算机辅助翻译的萌芽期。
3. 如何看待机器翻译的翻译质量？机器翻译的瓶颈是什么？
4. 什么是计算机辅助翻译？计算机辅助翻译的优势是什么？简要说明计算机辅助翻译的发展史。
5. 简要说明机器翻译与计算机辅助翻译的关系。