

# 目 录

---

开 篇 “计算机辅助翻译” 课程的教学与思考 .....	1
一、课程背景 .....	1
二、课程板块及内容 .....	1
三、经验与反思 .....	7
四、结语 .....	8
<b>第一章 从机器翻译到计算机辅助翻译 .....</b>	<b>9</b>
一、背景和需求 .....	9
二、机器翻译的发展历程 .....	11
三、机器翻译的原理 .....	17
四、计算机辅助翻译：萌芽与繁荣 .....	21
五、计算机辅助翻译：主要模块 .....	23
<b>第二章 计算机辅助翻译工具概述 .....</b>	<b>27</b>
一、硬件配置 .....	27
二、基本的软件配置 .....	28
三、电子词典和在线自动翻译工具 .....	31
四、百科全书 .....	36
五、搜索引擎 .....	39
六、狭义的计算机辅助翻译工具 .....	39
<b>第三章 双语语料库的建设与用途 .....</b>	<b>41</b>
一、双语语料库的概念与类型划分 .....	41
二、双语语料库的建设 .....	43
三、双语语料库的应用 .....	48
四、基于语料库的翻译研究 .....	57
五、双语语料库与计算机辅助翻译 .....	60
六、小结 .....	61

<b>第四章 双语语料库对齐与检索使用实例</b> .....	63
一、ParaConc 操作基础 .....	63
二、ParaConc 检索功能及使用 .....	73
三、ParaConc 其他功能与使用简介 .....	87
四、小结 .....	89
<b>第五章 术语、术语库、对齐与翻译记忆</b> .....	93
一、术语与术语学 .....	93
二、术语与计算机辅助翻译 .....	95
三、主流术语库软件 .....	97
四、对齐与翻译记忆 .....	99
五、主流翻译记忆软件简介 .....	105
六、小结 .....	111
<b>第六章 SDL Trados 2007的使用</b> .....	113
一、SDL Translator's Workbench 的使用 .....	113
二、WinAlign 的使用 .....	133
三、SDL MultiTerm 2007 的使用 .....	143
四、其他组件简介 .....	157
五、SDL Trados Studio 2009 的新增功能 .....	158
<b>第七章 雅信辅助翻译教学系统演示</b> .....	163
一、学生笔译练习平台 .....	163
二、专家库建设及检索平台 .....	170
三、雅信翻译项目演练 .....	178
<b>第八章 计算机辅助的译文质量保证</b> .....	191
一、《现代汉语语法信息词典详解》的翻译与校对需求 .....	191
二、计算机辅助的译文校对 .....	192
三、小结 .....	219
<b>第九章 主流机辅翻译软件的比较与评估</b> .....	223
一、国内机辅翻译软件的使用现状 .....	223
二、CAT工具：评估的框架 .....	224

三、评估结果 .....	228
四、给未来CAT工具使用者的建议 .....	233
五、结论 .....	234
<b>第十章 本地化与翻译 .....</b>	<b>237</b>
一、本地化 .....	237
二、本地化与翻译 .....	238
三、主流的本地化工具 .....	242
四、SDL Passolo 使用演示 .....	244
<b>第十一章 桌面排版系统与翻译 .....</b>	<b>259</b>
一、桌面排版系统 .....	259
二、FrameMaker 简介及其在翻译中的应用 .....	263
三、FrameMaker 排版演示 .....	264
四、小结 .....	296
<b>结语 计算机辅助翻译的前景 .....</b>	<b>297</b>
一、使用翻译技术的主要障碍 .....	298
二、翻译技术的前景 .....	299
<b>参考文献 .....</b>	<b>301</b>

# 开篇

## “计算机辅助翻译”课程的教学与思考

这部分内容是对北京航空航天大学外国语学院自 2004 年秋季以来开设“计算机辅助翻译”课程的教学与思考，曾发表在《中国翻译》2009 年第 4 期。由笔者更新后在这里分享出来，供读者参考使用。

### 一、课程背景

自上世纪 90 年代起，香港中文大学翻译系就开始给本科生和硕士生开设计算机与翻译结合起来的课程，名称不一，如“计算机与翻译”、“计算机在翻译中的应用”、“机器翻译入门”等。鉴于计算机技术给翻译过程带来的巨大便利和翻译行业对从业人员技术水平的要求，该系于 2002 年在亚太地区首次开设“计算机辅助翻译”两年制文学硕士课程，目的是深化学生对语言作为交流基本工具的理解，并使他们掌握与翻译实践有关的各种技术。

2004 年初，笔者从香港中文大学学习归来。经学院安排，从当年起，每年秋季学期给全院英语专业语言学和翻译方向的一年级硕士研究生开设“计算机辅助翻译”（Computer-aided Translation, CAT）课程，每周 2 学时，共 18 周 36 个学时。鉴于现有参考资料和软件界面多为英文，面对的又是英语专业的学生，课程主要工作语言为英文。后来，为配合此课程，学院成立了“翻译科技实验室”，在软硬件设施和课程方面均为全国高校的领先者。本课程开设至今已 6 年，有一些初步的经验和反思，在此总结出来，与大家分享，供有意开设此课程的兄弟院校参考，以帮助修读本课程的学生扩大视野，提高他们实际翻译操作的技能和水平，增加他们毕业后的就业选择面，为他们成为宽口径、复合型的人才做出一点贡献。同时，也希望得到更多的建议，以便于本课程的进一步修订与完善。

### 二、课程板块及内容

本课程从理论讲解到实际操作，主要有如下几个板块：机器翻译和计算机辅助翻译史、计算机辅助翻译的原理、广义与狭义的翻译工具、语料库与计算

机辅助翻译、术语与术语库、对齐与翻译记忆、主流的计算机辅助翻译工具、计算机辅助翻译与全球化和本地化等内容。同时，课程要求修课研究生学会使用主流的语料库分析工具和计算机辅助翻译工具。前面的内容是铺垫，理论性知识较多，后面部分是实践，对动手能力要求较高，简述如下。

## 1. 机器翻译和计算机辅助翻译史

计算机辅助翻译（电脑辅助翻译）是从机器翻译（Machine Translation, MT）或计算机翻译发展而来的，也可以把它称作机器辅助翻译（Machine-aided Translation, MAT）。本课程用计算机辅助翻译来统称 CAT 与 MAT。

CAT 与 MT 既有本质的不同，二者又密不可分。前者可进一步细分为计算机辅助的人工翻译（Computer-aided Human Translation, CAHT）和人工辅助的计算机翻译（Human-aided Computer Translation, HACT），在翻译的过程中，自始至终都有人工因素。后者则是没有人工参与的全自动翻译过程，其最终目标因此也可以叫全自动高质量机器翻译（Fully-automated High Quality Machine Translation, FAHQMT）。

这些基本知识十分重要，能够厘清学生对几种概念的认识，消除误解，帮助他们认识将要接触的是什么、需要他们掌握什么知识和技能，从而对计算机技术在翻译中的应用有合理的预期。因为笔者发现，多年来，大部分人想当然地以为 CAT 就是机器翻译，是全自动的。而现有的自动翻译系统所产出的译文很难让人满意，因此他们对学习 CAT 技术或者一概拒绝，或者怀有很大的疑问。这种区分有助于使大家对 CAT 的期望趋于合理和现实。

在课程的这一部分，笔者主要给学生介绍计算机（辅助）翻译的历史和现状。从上个世纪 30 年代开始，计算机辅助翻译经历了早期的尝试阶段（40—50 年代初）和高度乐观阶段（50—60 年代中期），1966 年美国的语言自动处理咨询委员会（Automatic Language Processing Advisory Committee, ALPAC）的《ALPAC 报告》对机器翻译研究和开发是一次沉重的打击，也是这一领域研究方向的转折点。经过十年左右的沉寂和停滞，70 年代后期，研究和开发有了复兴；80 年代后期到 90 年代早期，是新的转折点，MT 研究开始从纯理论研究转向实际应用，CAT 逐渐为人接受；研发者着手为职业译员提供译员工作台（translator workstation），在受控语言和受限专业系统方面进行开发，并探讨在多语信息系统里如何应用翻译组件。

真正的翻译助手，即上述的译员工作台，出现在上个世纪 90 年代初期，如 Trados 公司的 Translator Workbench、IBM 公司的 Translation Manager/2、瑞

士 STAR Group 的 STAR Transit, 以及 Eurolang Optimizer 等, 综合应用了复杂的文本处理和出版功能、术语管理以及翻译记忆库等各个方面的研究成果。进入 21 世纪后, 在全球化和本地化的浪潮中, 大量的翻译需求赋予与此相关的计算机技术以更大的使用空间。一些主流软件在全球化和本地化行业中得到广泛的应用, 为推动全球范围内的信息交流做出了传统翻译难以企及的贡献。此部分的授课, 可以参考的相关文献包括 William John Hutchins 的个人网站 <http://www.hutchinsweb.me.uk>, 以及 Chan (2008) 和冯志伟 (2004)。

## 2. 计算机(辅助)翻译的原理

早期的翻译系统基本由双语辞典构成, 即一个源语词汇对应一个或多个目标语词汇, 再配备上能产出正确句序目标语的语法规则。但人们很快就意识到, 仅靠词汇和词序是很不够的, 还需要更为系统的句法分析。当时盛行的形式语法似乎能为此寻找一条出路。但随后, 研究者又陷入机器在语义理解上的困境。20 世纪 80 年代, 机器翻译系统的主导策略是通过中间表述 (intermediary representation) 实现间接翻译 (indirect translation), 翻译过程涉及到语义、句法、词法层次的分析, 有时还有非语言知识库的支撑。80 年代末, IBM 研究组公布了他们完全基于统计方法的机器翻译系统 (Candide) 的成果; 其次, 日本一些研究组开始在研发中利用翻译语料, 这种方法后来被称为“基于例句”的翻译。与此前“基于规则”的翻译系统不同, 这两种的特点都是不再利用句法或语义规则来分析文本或选择对等词汇, 而是利用大规模的文本语料, 在翻译模块中则混合了基于规则和基于语料的方法。

在这个时期, 几乎所有的研究活动都致力于在传统的基于规则和“中间语言”模式的基础上进行语言分析和生成方法的探索。这些方法都伴有人工智能类型的知识库, 表现了更具创新性的潮流。译员的需要让位于商业兴趣: 术语管理软件在这个时期出现, 如 Mercury/Termex。ALPNET 在 80 年代上市了一系列翻译工具, 其中还包括“翻译记忆库”(Translation Memory, TM) 的早期版本, 即事先翻译好的双语对齐文本库。

90 年代早期, 机器翻译研究在新兴的基于语料库的方法的推进下向前发展, 主要表现在统计方法的引入(如 IBM 公司的 Candide 计划)以及基于实例的机器翻译等。统计技术解放了先前基于规则的方法。消除歧义问题、代词指代以及更多的惯用法生成等问题, 都因为基于语料库技术的应用而有了解决的希望。与基于规则的方法一样, 统计方法也不能提供更多让人满意的答案, 但是与十年前相比, 现在提高输出译文质量的希望似乎更大了。正如许多观察家指出, 最有前途的方法可能是整合了基于规则和基于语料库的方法。现在许

多商用机器翻译系统正在集成翻译记忆库，而许多翻译记忆库又被机器翻译方法加强。

考虑到授课的主要对象为文科背景居多的英语专业学生，教师可举简单的例子并辅以图表来说明这些原理，给学生以初步的理性认识和感性体验。授课时可参考的资料有冯志伟（1997，2004）、张政（2006，2010）、Bowker（2003）、Chan（2001，2008）、Hutchins & Somers（1992）、L' Homme（2007）、Quah（2006）、Somers（2003）等。

### 3. 广义与狭义的翻译工具

广义的翻译工具指能在翻译过程中提供便利的所有软硬件设施，如文字处理软件、文本格式转换软件、电子辞典、在线辞典和包括计算机、扫描仪、传真机等在内的硬件设备等；狭义的翻译工具是指为提高翻译效率、优化翻译流程而专门设计的 CAT 软件。（徐彬 2007）狭义工具面对的市场较小，用户也较少，目前受到的关注度较低，人们对它的认识也不充分。而多数人在日常工作和生活中都接触过广义的翻译工具，容易把这些广义的工具当作是翻译工具的全部，因而就会产生前面所提到的误解。

本课程主要关注狭义的翻译工具，即 CAT 及其相关知识和产品。在这一部分可引入下文将要接触的几种主流 CAT 工具及其主要功能和技术，如术语库、对齐、翻译记忆、项目管理、质量保证等。这些狭义上的 CAT 工具无论是单独或者集成软件套装，都能够帮助翻译员优质、高效、轻松地完成翻译工作。

授课时，教师可给学生初步介绍各种各样的翻译工具，使学生对它们有初步了解，并有所区分。具体参考资料同上。

### 4. 语料库与 CAT

在完成前面有关历史的讲解后，学生们应该已经了解到，现行 CAT 技术主要是结合语料的建设而研发的，对翻译最终结果至关重要的翻译记忆来说，语料是其成功与否的基础。本课程在这部分讲述与语料库相关的内容，包括定义、分类、建设、维护等。

语料库的分类，尤其是有关平行语料库的内容，对后面将要讲授的术语、对齐和翻译记忆都十分重要。同时，要求学生掌握一种或几种语料分析工具，如 AntConc、MonoConc、ParaConc 和 WordSmith，以及语料分析工具的几个重要组成部分的涵义和用途，如关键词、词频、词表、类型 / 标记比、搭配与

检索、语义韵律等，初步掌握分析英文和对应的中文文本的方法。教师应要求学生在进行语料分析时发现并解决问题，如这些工具在分析中文时的局限性、中文的切分困难及解决办法、文本对应问题等。

就北航的授课实践来看，学生普遍反映，这个部分的内容能让他们感受到计算机技术在分析翻译双语文本时的强大功能，体会到计算机与人工分析时的巨大差异，对与翻译有关的计算机技术兴趣更浓厚了。

此部分的内容可参考文献有 Baker (1993, 1995, 1999), Sinclair (1991), 王克非 (2004), Somers (2003)。此外，国内还有一个很重要的网络资源“语料库语言学在线”可供参考，网址是 <http://www.corpus4u.org>。

## 5. 术语与术语库

对于翻译而言，术语一致性能够保证译文的专业化和高质量。在面对大型的翻译项目时，有一个中心化、电子化、统一化的术语库，对基于网络协同工作的翻译团队来说，其优点不言而喻。在这部分，首先讲授术语的基本知识，如定义、分类、用法、用途等；其次是术语库的创建、编辑扩充和管理维护等知识；然后向学生介绍现有的具代表性的大型术语库，如 EuroTermBank 的概况，以及我们中文语境下在此领域的欠缺，也即未来需要做的工作；最后向学生介绍几个主要的术语库工具，如独立系统 AnyLexic: Terminology Management Software、Sun Gloss、T-Manager 4.5 等，以及配合翻译记忆工具使用的术语工具，如 SDL MultiTerm 2007。同时要求学生具体学会一个工具，初步体会其在翻译团队工作流程中所表现出来的优势。

授课时的参考书籍包括冯志伟 (1997) 和 Wright & Gerhard (2001)，同时也有网络资源，如维基百科相关条目。

## 6. 对齐与翻译记忆

翻译记忆是 CAT 的核心技术，指的是一个储存有翻译句子或句段及其源语言对应句子或句段的数据库。这些句段对被称为翻译单元。翻译记忆使用数据库技术在翻译时自动提供以前的翻译译文作参考，而且不仅是提供完全匹配译文，还可以通过模糊匹配查找在语言上与待译原文句段相关的数据库词条，并将有关匹配粘贴到译文中。在翻译过程中，新增的译文和更新后的译文被自动添加到数据库，所以翻译记忆会动态地增大。

而对齐是把原文的句子和译文的句子或句段一一匹配对应起来，其结果对

建设翻译记忆作用巨大。通过将已有的原文与译文的自动对齐，可以将对齐结果导入翻译记忆，供未来的翻译过程使用。这个工具在对某一个专业领域（如金融、科技、医学、政治文献）已有的翻译结果进行处理时，会有较高的匹配率，较少需要人工后期加工。授课时，教师可提醒学生一些需要注意的事项，如中英文文本对齐时可能会出现乱码及处理，并对可能花费的大量时间有思想准备，将对齐结果导入翻译记忆库时需要注意的文本格式和设置等。

关于项目管理功能，可以在了解和掌握了其他几个主要功能的基础上，有所尝试，以了解大型翻译项目的处理流程。前面提到的许多参考资料对此都有涉及，可供参考。

## 7. 主流的 CAT 工具

这个部分将介绍几种主流 CAT 软件的使用，如 Déjà Vu X、SDL Trados、TransWhiz、Wordfast 和雅信 CAT。要求学生学会使用几种软件的主要功能，如前面提到的术语库、对齐、翻译记忆、项目管理等，并通过使用和比较，发现并比较各软件的特色和不足。然后根据现有对各种翻译公司或自由译员的调查（如李燕、娄平 2007），对学生们选择使用何种 CAT 工具提出建议并给出理由。

这个部分需要学生加强动手能力，学习操作的过程比较繁琐，通常需要反复数次才能熟练掌握，应该给出充分的时间，教师也需要提供及时的指导。

## 8. CAT 与全球化和本地化

CAT 在全球化和本地化的过程中使用非常广泛，是推动全球化和本地化必不可少的工具。这个部分向学生介绍 CAT 在全球化和本地化中所起的作用、全球化和本地化对 CAT 工具的要求、使用 CAT 工具的跨国公司和机构（如欧盟和 IBM），以及如何使翻译人员更好地适应本地化和全球化对他们在技术上的要求，进一步提高自己，不仅要掌握双文化、双语言，还要掌握一定的计算机技术。另一个内容是要求学生了解和掌握本地化的主流软件，如 SDL Passolo，并尝试使用本地化工具将某个英语界面的小软件转化成汉语界面（即汉化），或将某个汉语界面的小软件转化成英语界面。如果有可能，还可以引入一些桌面排版系统（Desktop Publishing, DTP）的知识和工具，如 Framemaker 系列。桌面排版是最近几年来新兴的行业，对译文的最终排版和图形图像处理非常重要。译员如能综合掌握这些技能，也能提高翻译的效率和质量。

### 三、经验与反思

相对于传统的人工翻译，CAT 有着无可比拟的优越性，能显著提高译员的翻译质量和数量。人工翻译的日工作极限为 5,000 字左右，而有了 CAT 工具的帮助，译员可以将工作量提高到日均 10,000 字以上。在专业领域，如科技、金融、医学、新闻、法律等，CAT 能够保证译文内容的准确性、术语使用的一致性和译文产出的经济性。

我院翻译科技实验室现有 10 台硬件配置较高的计算机，安装有主流 CAT 软件，如 SDL Trados 2009 和雅信 CAT 4.0，均为定期升级后的最新版本，修课学生可以选择学习使用。学生们普遍反映学习本课程很有收获。将来如果条件许可，还可安装更多其他主流软件，如 Wordfast 和 Déjà Vu X。

经过几年的教学实践和软硬件建设，笔者有一些体会，可以总结为以下七点：

第一点是教师的知识储备和实际操作能力。2004 年开始讲授“计算机辅助翻译”课程时，内地高校几乎还没有同类课程，也没有可供参考和借鉴的经验，需要教师大量阅读有关文献，并学习掌握多种软件，努力打造跨学科的知识结构和操作技能。这对文科教师来说，有一定难度，需要精力和时间的大量投入。现在的情况有所好转，因为越来越多的人意识到了这些工具对优化翻译过程的巨大作用，而且也出现了不少可供参考的指南和学习交流网站。

第二点是有效地利用网络免费资源。随着网络建设的快速发展，互联网已经成为一个海量知识库，教师和学生都可以从中受益。维基百科、百度百科、中国知网以及各种在线的资料库和网站、在线辞典等，可以给初学者提供很多基础知识，如在维基百科上查询 computer-aided translation，就会得到有关 CAT、MT、CAT 工具等信息；查询 terminology，就会得到有关术语、术语库、术语工具等信息。

第三点是因材施教和教学互动。五年来的教学过程表明，根据学生的情况适当调整教学内容，能促进学生对知识的吸收和理解；而学生有时会发现和提出一些教师没有预料到的问题，促使教师进一步思考，对教师知识和技能的提高很有帮助，这也充分印证了教学相长的道理。

第四点是授课时文本的选择。为了更好地说明问题，应该选择内容和格式都比较规范、定期更新的文本，比如政治文献、科技文献、财经报告等。笔者授课时采用了某银行年报的中英文版，这种文本的特点是词汇量有限、词汇使用重复率高、句式简单、表达准确、年度重复率高，对讲授课料库、术语、对

齐、翻译记忆等部分的内容非常具有说服力。政治文献，如十六大报告和十七大报告的中英文版，对讲授和掌握上述知识和技能也非常有帮助。

第五点是学生之间的合作。学生在修读本课程前，基本对 CAT 没有接触，没有多少计算机辅助翻译的基础。在短暂的时间内，单凭个人去了解 and 掌握有关知识和实际操作，确实有困难，而团队分工协作会促使他们互相学习和帮助，进步的程度更大，速度更快。

第六点是与校外教学和科研机构的交流。几年来，与香港中文大学翻译系和北京大学计算语言学所有关专家学者的交流，对笔者扩大视野和了解新动向，很有帮助。

最后一点是保持与供应商和翻译行业的联系。笔者在讲授此课程之前，就与主流产品的供应商有接触交流；在讲授过程中，更得到他们的大力帮助，使本课程得以顺利进行。同时，通过与一些大型翻译公司的交流，能了解行业对从业人员技术水平的要求，对教学有一定的导向性作用。

## 四、结语

从 2007 年起，为适应经济全球化及提高国家国际竞争力的需要，培养适应国家经济、文化、社会建设需要的高层次、应用型、专业性口笔译人才，我国先后有近 200 所高校开设了翻译硕士专业学位（MTI）教育，有 30 多所高校开设了翻译专业本科课程（截止到 2010 年）。在硕士课程阶段，“计算机辅助翻译”被列为选修课，而在翻译专业本科教学计划里，它更是一门必修课，这都说明了 CAT 的重要性和实用性。（吕立松，穆雷 2007）随着越来越多的人对 CAT 的深入了解，以及更多的高校开设翻译本科和翻译硕士专业学位教育，“计算机辅助翻译”这门课程必将在更大的范围内开设。我们愿意与各方交流，共同进步，把这门培养学生跨学科知识和技能以适应时代需求的课程建设得更好。