

开源计算机辅助翻译工具研究

王华树¹, 陈晓婷²

(1. 北京师范大学 外国语言文学学院, 北京 100875; 2. 北京航空航天大学 外国语学院, 北京 100191)

摘要: 相对于商业计算机辅助翻译 (Computer-Aided Translation, 简称 CAT) 工具来说, 开源 CAT 工具有很多显著的优点, 被广泛应用于开源社区。然而, 国内翻译学者对开源 CAT 技术的研究基本上处于空白状态。基于这样的背景, 首先探讨了开源 CAT 技术的概念, 介绍了主要的开源 CAT 工具及其在翻译实践中的应用, 探讨了开源 CAT 工具的优缺点及发展趋势。最后指出开源 CAT 技术在翻译实践和翻译教学中将会发挥越来越重要的作用。

关键词: 开源软件; 计算机辅助翻译; 翻译实践; 翻译教学

中图分类号: H 059 **文献标志码:** A **文章编号:** 1009-895X(2015)03-0211-05

DOI: 10.13256/j.cnki.jusst.sse.2015.03.003

A Study on Open Source Computer-Aided Translation Tools

Wang Huashu¹, Chen Xiaoting²

(1. School of Foreign Languages and Literature, Beijing Normal University, Beijing 100875, China;

2. School of Foreign Languages, Beihang University, Beijing 100191, China)

Abstract: With many significant advantages over commercial profits, open source computer-aided translation (CAT) tools are widely used in open source community. However, Chinese translation scholars have rarely set foot on open source CAT research. Given such a background, the paper first discusses the concept of open source CAT tools, then illustrates the main tools and their applications in translation practice, and analyzes the pros and cons as well as the development trends of open source CAT tools. Finally, the paper holds that the open source CAT technology will play an increasingly important role in the practice and teaching of translation.

Keywords: open source software; computer-aided translation; translation practice; translation teaching

近年来, 随着开源运动的发展, 开源项目覆盖领域广泛, 从专门的应用程序到门户网站, 从电子政务到全面开发开源软件包计划, 开源软件已经渗透到各个不同的行业。开源软件在翻译行业中也日益普及, 翻译领域也出现了众多开源 CAT 工具, 这些开源软件凭借其优势受到企业和用户的青睐, 在国内的应用也越来越广泛。

一、国内开源计算机辅助翻译研究现状

开源即开放源代码 (Open Source), 开源软件 (Open-Source Software, 简称 OSS) 是一种源代码可以任意获取的计算机软件, 这种软件的版权持有人在软件协议的规定之下保留一部分权利并允许用

收稿日期: 2014-09-19

基金项目: 北京市社会科学基金资助项目“现代语言技术体系研究”(14WYB015)

作者简介: 王华树 (1980-), 男, 讲师。研究方向: 应用翻译、翻译技术、翻译管理。E-mail: wanghuashu@vip.qq.com

户学习、修改、增进提高这款软件的质量^[1]。开源的最大特点在于开发人员和使用者可以看到内码、修改程序,软件源代码自由流通,不受知识产权的约束。开源软件最早出现于20世纪70年代,至今经历了数十年的发展历程。在操作系统、数据库、服务器、移动应用等方面都有风靡全球的开源产品,如今已经深入渗透到了人们的社会生活和工作的各个领域,从不同的角度以不同的形式影响着人们的日常生活。近年来,随着 kernel.org, Opensource, Sourceforge, Linux 中国, 开源中国社区等社区的兴起, 开源 CAT 软件的开发和应用呈现出上升趋势。

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^[2], Hutchins^[3], Kay^[4], Kenny^[5], Melby^[6], Quah^[7], Sommers^[8]等, 国内的如冯志伟、徐彬^[12]、张政^[13]、苏明阳^[14]、钱多秀、俞敬松^[15]、王华树^[13,15]。归结起来, 大致可分为狭义和广义两大类。狭义的计算机辅助翻译技术通常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。计算机辅助翻译是以人为主体的翻译活动, 区别于全自动化的机器翻译, 前者可以称为“机助人译”, 后者可以称为“人助机译”。国外的 SDL Trados, Déjà Vu, Wordfast, MemoQ, STAR Transit 等主流的计算机辅助翻译工具, 以及国内的雅信 CAT、传神 TCAT、朗瑞 CAT、雪人 CAT 等工具属于此类技术范畴。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信息技术, 如在线词典、语料库、格式转换、OCR 识别、桌面搜索等。本文主要探讨狭义的开源 CAT 工具。

所谓的开源计算机辅助翻译工具是指开放源代码的翻译记忆系统。在翻译过程中, 系统将已有的原文和译文以句子为单位存储在翻译记忆数据库中, 并自动搜索翻译记忆库中相同或相似的翻译资源供译者参考, 可以帮助译者提高翻译速度, 确保译文术语和风格的一致性。和其他开源工具一样, 开源的翻译记忆系统也具备实现开放源代码、修改程序、软件源代码自由流通、不受知识产权约束的特征。

笔者通过中国知网数据库检索主题词“开源计算机辅助翻译”和“开源”或“计算机辅助翻

译”, 检索时间不限, 检索范围涵盖了全国外语类的学术期刊, 发现探讨开源计算机辅助翻译技术或工具的文章寥寥无几。国内翻译研究者或翻译技术教育者对开源计算机辅助翻译的关注度远远不够, 研究基本上为空白状态。因为计算机辅助翻译涉及到了语言学、翻译学、传播学、信息技术、软件工程等多个学科, 具有明显的跨学科特征, 所以不容易被人理解, 有必要做普及性介绍。

二、主要的开源计算机辅助翻译工具

随着开源运动的发展, 在翻译领域中, 支持多国语言的软件工具不断涌现, 众多开源软件在翻译实践中得到了广泛应用。目前市面大多数 CAT 工具基本上都属于商业系统, 如 SDL Trados, DéjàVu, Wordfast, MemoQ 等。相对商业 CAT 工具来说, 开源 CAT 工具所占比例较少, 主要用于开源社区的翻译项目以及中小型语言服务企业。诸如 Anaphraseus, ForeignDesk, G11NtoolKit, Lokalize, OmegaT, Ooxlate, Open Language Tools, OpenTM2, Open-tran, Pootle, TransFluent, Transolution, Okapi Framework, TinyTM, Virtaal, Wordforge 等工具是开源 CAT 工具的主要代表。从结构上看, 开源 CAT 工具大致可以分为嵌入式和独立式两种。从软件架构上, 大致可以分为 C/S 架构 (Client/Server, 客户机/服务器) 和 B/S 架构 (Browser/Server, 浏览器/服务器)。前者是早期软件系统首选的设计标准, 后者是随着互联网技术的兴起, 对 C/S 架构的一种改进, 也是未来的 CAT 工具发展中占据主导地位的一种设计理念。下面分别介绍几款具有代表性的开源 CAT 工具。

嵌入式的工具, 如 Anaphraseus。它是基于 Open Office.org Basic 的一种宏设置, 可为 OpenOffice.org 用户提供包括文本分割、翻译编辑、模糊匹配、术语识别、词汇表导入导出等辅助功能。它兼容 TMX 标准, 可以导入 OmegaT 翻译记忆库文件, 并且同 Wordfast 的翻译记忆库无缝连接。独立式的工具, 如 OmegaT。作为一个独立的、跨平台 CAT 工具, OmegaT 可辅助译者处理翻译资源, 在翻译项目中管理源文件、翻译记忆和术语。OmegaT 凭借其实用强大的功能, 受到各个领域的欢迎。在商业领域, 微软 Translation Hub 将 OmegaT 视为重要合作伙伴之一, 因为该软件支持众多格式, 能够直接兼容微软的术语集以及多语言种类 TBX 格式的文件, 甚至可以在同一个项目中

同时处理多种格式的文档。在教育领域,诸多英美高校高度重视 OmegaT 的教学,英国开设计算机辅助翻译课程的四所高校,即利兹大学 (University of Leeds)、斯旺西大学 (University of Swansea)、伦敦帝国学院 (Imperial College of London) 和诺丁汉大学 (University of Nottingham), 都要求学生需掌握这一世界通用的 CAT 软件^[17]。此外,这一实用工具还受到政府机构的青睐,日本政府资助开发的 Benten 就是使用 OmegaT 代码来处理翻译记忆匹配过程的,南非政府的艺术和文化部门资助开发的 Autshumato translation suite 也是基于 OmegaT 创建的。

Okapi Framework 属于独立式的开源 CAT 工具,不过它是一系列组件和应用程序的集合,包括 Rainbow, CheckMate, Ratel 等几个常用组件,可为文档、软件的本地化和翻译提供多层次的技术支持,广泛应用于各大本地化服务提供商,Microsoft, Google, Pactera, Lionbridge 等公司都在使用 Okapi Framework 工具。

Pootle 是基于 B/S 架构的翻译和管理系统,旨在帮助团队进行翻译协作,在有网络接入的任意地方,译员都能通过这一平台进行多人在线实时协作,操控同一个翻译项目,进行任务分配、信息统计、双语语料回收以及远程翻译数据库共享等工作。开源网页浏览器 Mozilla Firefox, 开源办公软件 OpenOffice. Org 以及由美国麻省理工大学的尼葛洛·庞帝教授提出的“每个孩子一台笔记本电脑”(One Laptop Per Child) 项目^[9]等都纷纷采用 Pootle 进行翻译项目管理。

三、开源计算机辅助翻译工具的优势和局限性分析

计算机辅助翻译技术具有重复利用以前翻译的语料,方便统一术语,保持翻译一致性等强大的功能,所以商业的 CAT 工具被广泛应用于语言服务行业,并占据主要的翻译技术市场。同商业的 CAT 工具相比,开源 CAT 工具也具有明显的优势^[9,18]。

(1) 性价比高。开源 CAT 工具的主要功能与商业 CAT 工具相当,可兼容 TMX 标准,具备模糊匹配、术语管理等主要功能,译者几乎不需花费任何代价,可以节省购买翻译工具的费用。相比之下,商用 CAT 除了采购软件本身的成本,还需要

支付病毒防护成本、技术支持费用、升级开支和与厂商锁定有关的成本,价格十分高昂(如表1所示)。开源 CAT 工具对于个人译者、教育机构和非营利机构等有很大帮助。

(2) 灵活可靠。商用 CAT 经常受制于供应商,有时甚至受到 IP 限制,软件需要经常升级,如果不升级,可能就无法使用。而开源软件即使在较老的机器上仍然能够很好地运行,用户可以自主决定升级的时间。

(3) 定制化强。由于源代码是开放的,用户可以自行获取并修改以便适应自己的需求。这种定制化能力不仅仅是修改软件,更重要的是还添加了用户所需的功能。

(4) 安全性高。Linux 创始人提出著名的“林纳斯定律”:有足够多的人关注,错误将无所遁形(given enough eyeballs, all bugs are shallow)^[10]。意思是说看到和测试代码的人越多,漏洞就越可能被更快地发现和修复。开源的 CAT 软件背后有全球各地的社区进行审查和测试,一有漏洞很快就能修复,而封闭的商用 CAT 软件则难以做到,用户只能从厂商处得知他们在保持着软件的安全和遵守标准的说法。

因此,从某种程度来看,开源的 CAT 工具已经赶上并正在努力超越商用的 CAT 工具。

表1 专有软件和开源软件成本比较

Tab.1 Cost comparison between proprietary software and open source software

专有软件 (Proprietary Software)	开源软件 (Open Source Software)
Microsoft Windows XP Professional: \$262.99	Ubuntu Linux: \$0.00
Microsoft Office XP Professional: \$225.00	Open Office. org: \$0.00
Microsoft Outlook: \$89.99	Mozilla Mail: \$0.00
Quick Books Basic: \$189.99	Gnu Cash: \$0.00
Trados 7 Freelance: \$895.00	Omega T: \$0.00
Total: \$1,662.97	Total: \$220.00

众多开源 CAT 系统的出现,可以满足不同用户的需求,根据市场需要还将出现更多新的工具。许多开源社区的用户在使用过程中,也看到了开源 CAT 系统的不足之处。归结起来,主要有以下几个方面。

(1) 大多数开源 CAT 工具采取独立式架构,

许多工具采用了 PO、XLIFF、TMX 等标准的翻译数据格式,但是具有开放的应用程序接口(Application Programming Interface)系统仍在少数,不利于数据交换或远程服务调用。

(2) 开源 CAT 工具之间的集成性和互操作性相对较低,缺少资金维护,用户体验不佳。不同的开源社区根据特定的需求会开发出不同的工具集,会造成多数功能重复。

(3) 多数开源 CAT 工具主要提供翻译编辑功能,具备流程管理、跟踪资产和项目状态、角色管理和任务以及进度控制的开源工具非常少。

凡事有利有弊,我们可以看到开源 CAT 工具显著的优点,已经被广泛应用于翻译和本地化实践之中,甚至是一些大型的商业项目。总体看来,其利大于弊,用户需要有合理的心理预期,不能处处与商业 CAT 工具(如 SDL Trados)一比高下。在翻译实践中,用户可以根据项目的具体需求,结合开源 CAT 工具的可用功能,进行合理取舍或定制开发。

四、开源计算机辅助翻译技术发展趋势

在互联网技术的推波助澜之下,SaaS (Software-as-a-Service,软件即服务)模式的普及对开源技术的普及起到极大的推动作用,开源 CAT 工具逐步从单机版走向平台化协作。诸如 Transifex、Transwiki、Weblate 和 Zanata 等多功能翻译平台,已经在国际语言服务项目中大显身手。

Transifex 是一个 B/S 架构的集成式翻译平台,通过与上游项目的版本控制系统 (Version Control System) 紧密集成,使译员的翻译成果能够方便、快速地复用到整个翻译项目中,充分共享语言资源。Fedora 系统的本地化便采用了 Transifex 平台。此外,Community Answers (类似 Yahoo Answers 的 Web 2.0 QnA 工具)和文件管理工具 Dropfiles 也运用了 Transifex 进行本地化。

Transwiki 是中国万维网联盟 (W3CHINA.ORG) 组织的开放翻译计划 (Open Translation Project,简称 OTP)。该系统为译员提供了一个易用的界面,利用 WIKI 协作特性开展 W3C 及其他重要标准的中文翻译工作。TED (Technology, Entertainment, Design,简称 TED) 就是此开放翻

译计划的受益者,世界各地的翻译志愿者通过该平台为 TED 演讲翻译字幕。

Weblate 是一款高度 Git 集成的翻译管理系统,能够实现子项目间翻译同步,自定义质量检查,自动链接到源文件及云支持等功能。在翻译过程中,Weblate 支持离线翻译,译员可以从该平台下载文章,进行离线翻译,译完后再上传合并。译文提交后,译员不仅可以自定义检查翻译质量,还能够通过多种方式复查他人的译文。语音控制软件 Voice Control 便是在 Weblate 平台进行本地化翻译的。

Zanata 是一个基于 Web 的系统,这个开源的翻译平台旨在管理本地化项目。除了翻译记忆的功能之外,Zanata 还提供一个在线翻译编辑器和与 REST APIs 和命令行工具集成的工作流。译员可以利用 Zanata 编辑器翻译同一项目,编辑器上的译文保存之后,所有译员便可立即共享查看。项目中的每种语言都有对应的聊天室,方便译员实时沟通。Zanata 将集中的本地化资源库和翻译工具综合一体,能够处理整个翻译工作流程,译员无需费时处理格式,可以完全专注于翻译本身。

此外,云计算 (Cloud Computing) 在多个行业迅速得到了应用,对翻译行业的组织形式、流程、市场产生了新的影响^[11]。随着社会网络的普及,无处不在的云计算服务将深入发展。开源 CAT 系统与开放的云平台的整合,能够为用户提供海量的翻译数据支持,快速响应大型翻译项目的需求,增强翻译业务的处理能力,必将大幅度提升翻译行业的生产效率。

五、结束语

互联网的精神实质是自由、平等和分享,开源翻译技术正是这种精神的体现。近年来,企业用户、语言服务供应商以及广大翻译从业者共同见证了一系列开源翻译技术的发展。开源的机器翻译(如 Moses、Apertium)、字幕翻译(如 Gaupol、Aegisub)、移动应用翻译(如 TransFluent)、翻译管理系统(如 Project Open、GlobalSight)等不断涌入国际翻译社区。开源 CAT 软件的兴起和发展将是持续的发展过程,将对商业 CAT 软件产生重要影响,并且与之长期共存。IBM 在 2010 年 6 月与 LISA、Welocalize、Cisco 和 Linux Solution Group 等机构将其原来专有的 CAT 工具 IBM TM/2 开源,并成立了一个专门的指导小组,此小组负责指导从

TM/2 所有权解决方案到开源产品的过渡。老牌的商业 CAT 工具 Heartsome 也于 2014 年 8 月加入了开源 CAT 工具的行列。它们的加盟将大大推动交互性合作的计算机辅助翻译技术的广阔发展,软件操作灵活性大幅度提高,为个人译者和广大的语言服务商提供了更多的技术选择,对开源翻译技术的发展具有积极意义。随着翻译实践的深入,有关开源 CAT 软件的研究和开发将会更加深入,更加多样化。这对于优化开源 CAT 的功能设计,促进商业翻译和本地化技术开发,提高开源社区的翻译效率,降低翻译生产成本,改善高校翻译技术教学现状,都具有举足轻重的作用。

参考文献:

- [1] Laurent A M S. Understanding Open Source and Free Software Licensing [M]. California: O' Reilly Media, 2004:4.
- [2] Bowker L. Computer-Aided Translation Technology: A Practical Introduction [M]. Ottawa: University of Ottawa Press, 2002.
- [3] Hutchins W J. Machine Translation: Past, Present, Future [M]. Chichester: Ellis Horwood Limited, 1986.
- [4] Kay M. The Proper Place of Men and Machines in Language Translation [J]. Machine Translation, 1997, 12 (1/2): 3-23.
- [5] Kenny D. CAT Tools in an Academic Environment: What are They Good for? [J]. Target, 1999, 11(1): 65-82.
- [6] Melby A K. Computer-Assisted Translation Systems: The Standard Design and a Multi-Level Design [C] // Association for Computational Linguistics: Proceedings of the First Conference on Applied Natural Language Proceedings, 1983.
- [7] Quah C K. Translation and Technology [M]. Hampshire and New York: Palgrave Macmillan, 2006.
- [8] Somers H. Computers and Translation: A Translator's Guide [M]. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2003.
- [9] Seeburg D. Free and Open Source Software for Translators [C]. New York: ATA 48th Annual Conference, 2007.
- [10] Raymond E S. The Cathedral and the Bazaar [M]. California: O'Reilly Media, 2001:1.
- [11] Muzii L. Cloud translation [R]. Ekaterinburg: Translation Russia Forum, 2011.
- [12] 徐彬. 计算机技术在翻译实践中的应用及其影响 [D]. 济南: 山东大学, 2004.
- [13] 张政. 计算机翻译研究 [M]. 北京: 清华大学出版社, 2006.
- [14] 苏明阳. 翻译记忆系统的现状及其启示 [J]. 外语研究, 2007(5): 70-74.
- [15] 俞敬松, 王华树. 计算机辅助翻译硕士专业教学探讨 [J]. 中国翻译, 2010(3): 38-42.
- [16] 王华树. 信息化时代背景下的翻译技术教学实践 [J]. 中国翻译, 2012(3): 57-62.
- [17] 张霄军. 英国高校的“翻译技术”教学及其启示 [J]. 外语研究, 2010(6): 76-79.
- [18] McKay C. Free and open source software for translators [EB/OL]. (2014-08-16) [2014-09-01]. <http://www.translatewrite.com/foss/atapaper05.pdf>.

(编辑: 朱渭波)