

人机共译交互平台在工程翻译中的运用

——以传神语联网为例

肖凤华 殷白恩

(武汉华夏理工学院 武汉市 430223) (传神语联网网络科技股份有限公司 武汉市 430223)

摘要 本文介绍了从机器辅助翻译 (CAT) 到神经网络机器翻译 (NMT) 的发展, 通过传神语联网的真实案例, 介绍了工程行业翻译的特点以及人机交互平台在工程翻译中的运用, 通过前景精准的产能聚集和适配, 可以大幅节约时间成本, 改变工程行业译员的能力模型, 提升行业商业增值产出价值。

关键词 人机共译 人机交互平台 工程翻译

Abstract This paper introduces the development from computer aided translation (CAT) to neural network machine translation (NMT). With the real cases completed by Transn Company, it unveils the characteristics of engineering industry translation and the application of human-computer interaction platform in engineering translation. Furthermore, it is shown that the capacity gathering and adaptation through accurate projections can greatly save time and enhance the ability of translators in the engineering industry.

Key Words man-machine translation human-computer interactive platform engineering translation

DOI:10.16024/j.cnki.issn1002-0489.2019.03.012

引言

近年来, 机器翻译市场在全球市场分布、用户行业分布、技术成熟度和服务模式等诸多维度都发生了深刻的变化。国内许多大型商业公司、专业翻译公司、科研机构甚至是创业公司也纷纷参与机器翻译行业的竞争, 最近几年陆续推出了 NMT 产品并开始向 B 端甚至 C 端用户提供服务。本文旨在通过真实典型成功案例研究, 揭示人机共译平台在工程翻译项目管理中的优势和工作原理, 并预测其未来的趋势和应用场景。

1 人机共译背景介绍

1.1 CAT 到 NMT 的转变

1947 年计算机诞生以来, 学界就开始尝试利用计算机进行机器辅助翻译 (CAT)^[1], CAT 翻译方法的发展历程体现出人们使用计算机实现自动翻译方法论的不同阶段。第一阶段基于语言规则的方法, 认为只要将人类翻译的理论和手段“灌输”给计算机就可以让其模拟人来实现翻译, 但语言规则无法覆盖复杂多变的语言现象。第二阶段是从实际语言中学习的经验主义方法, 无论是从翻译样例中获取翻

译模版还是统计模型 (SMT), 都属于这个范畴。第三阶段是神经网络翻译 (NMT) 即端到端的整体翻译模式。这三种研究方法的进展体现了机器翻译策略的变革路线图: 从模仿人类翻译规则到语言结构成分的分析 and 重组, 再到“整入整出”的翻译策略^[2]。

Google 发现, 在多个样本的翻译中, 神经网络机器翻译系统将误差降低了 55% - 85% 甚至更多。但是 NMT 仍然存在训练成本太高等问题, Google Brain 的研究人员 Denny Britz, Anna Goldie, Thang Luong, Quoc Le 对 NMT 架构的超参数进行了大规模分析, 并且对建立和扩展 NMT 构架提出了一些新颖观点和实用建议^[3]。

1.2 人机共译

“人机共译”是基于大数据、人工智能和移动互联网, 将机器智慧与人工结合, 平衡机译高效率与人译高质量, 组成“稿件-机译-人译”的产出链的一种智能翻译模式。基于神经网络的机器翻译 (NMT) 代替了传统的统计机器翻译 (SMT), 神经网络机器翻译技术已经使得翻译系统的速度和准确度有了大幅

本文获得湖北省专利局项目资助 (2018 年 45)。项目立项网址 <http://zscqj.hubei.gov.cn/show/34970>

作者电邮: 951294526@qq.com, ybc007@163.com

收稿日期: 2019-05-08 修改日期: 2019-06-25/38

提升。

传神语联网网络科技股份有限公司采用循环神经网络、word embedding 的新一代算法和 GRUs&LSTM 算法以及人机交互翻译平台的开发, 实现翻译从业者高效高质地完成翻译任务, 满足翻译行业的大规模高效工作流程的需要。

2 人机交互平台

2.1 人机交互平台的特点

人机交互平台是为系统与用户之间的交互关系提供的用户可见部分的操作环境。翻译的人机交互平台是指译员之间的交流互动协作、译员与翻译团队间进行翻译资产置换和在线协作、翻译项目的管理者和客户相互渗透的平台。平台链接将不同的角色(译员、审校、排版和客户)和不同的机器链接起来共同完成一次项目合作, 同时解决产能聚集、质量保障、效率提升、知识库积累和机器学习等多个任务。

2.2 人机共译交互平台的目标

以传神语联网为例, 通过人机互译平台的搭建, 可以完成以下目标:

1) 翻译管理平台功能

能有效解决翻译团队流程化管理的问题, 提高海量翻译任务的有序管理和调配翻译资源。可以帮助客户有效监控并保证翻译进度与质量。

2) 协同翻译平台功能

可以解决翻译过程中术语统一, 语料复用的问题。解决翻译完成后版式保持, 以及解决翻译过程中出现的低级错误。

3) 术语语料库功能

翻译项目完成后建立行业术语语料库, 能有效解决翻译成果重复利用以及术语语料不断积累的问题。

4) 知识库系统功能

可以解决文档规范化管理问题, 建立多语知识库系统, 将翻译资产最大化应用。

3 人机共译平台在工程翻译中的应用

3.1 工程行业翻译的特点

工程学、工程技术、工程设计和工程施工领域中的翻译可称为工程翻译, 很多科技和外贸翻译的内容也属于工程翻译的范畴^[4]。工程翻译具有用词专业、涉及领域广、项目大等特点, 是翻译行业公认的大型项目领域, 需要

多领域专业人才合作才能完成。工程翻译译员需求量大、专业术语多、单一稿件翻译量巨大、稿件重复几率高、源语言水平参差不齐, 依靠人工翻译完成工程翻译效率低, 终稿风险高。在传神语联网传统笔译项目中, 工程翻译的稿件占翻译总量 40% 以上。

3.2 人机共译平台在工程翻译中的优势

机器翻译速度快、成本低, 但不够准确流利; 人工翻译细致准确, 但是成本高、速度慢。而人机共译交互平台在工程翻译中可以得到很好地应用。

3.3 人机交互平台的工作流程

任务建立——产能匹配(译员产能适配机器产能适配)——协同翻译——成果提交——知识库整理——机器学习——译员评价

基础数据层包括商业大数据、垂直行业语料库、客户的语料库、平行双语语料库以及译员个人的语料库等。在基础数据层的基础上, 通过 CAT、SMT 和 NMT 的机器翻译产能加上译员和翻译公司的产能聚集层到功能聚集层, 也称为基础支撑服务。有了数据支撑、基础支撑和机器翻译支撑后人机交互开始实现。

人机交互层是通过人的能力和机器能力的交互完成项目协作, 知识库共建和机器学习。通过图像识别、文字识别和音视频识别获取信息后, 术语语料库和在线翻译完成机器部分的基础工作。包括术语提取、术语统一、术语管理以及译后的语料回收等。在线翻译通过自动分析文本、智能设置和多种机器辅助完成任务。系统智能纠错引擎检查翻译中出现的低级错误如拼写错误、数字错漏译、单位错漏译、术语漏译、句子段落漏译和标点错漏译等, 节省了校对的时间和成本。翻译积累的术语、语料是人机交互平台中最重要的部分。语料可以审核和修改, 数据在云端存储即时同步更新, 术语语料循环重复利用, 避免了大量的重复翻译。在人机交互层里, 人的能力包括翻译项目管理和创作翻译能力。项目管理控制流程管理、任务管理、译员资源管理等, 可有效控制整个翻译过程。通过译员管理系统, 翻译管理者可看到每个译员的基本情况、历史合作项目、任务饱和度以及专家对译员水平的评价等。整个团队在翻译的过程中, 通过在线协同翻译, 使得项目间翻译成果可以和其他译员共享, 避免了大量的重复劳动。

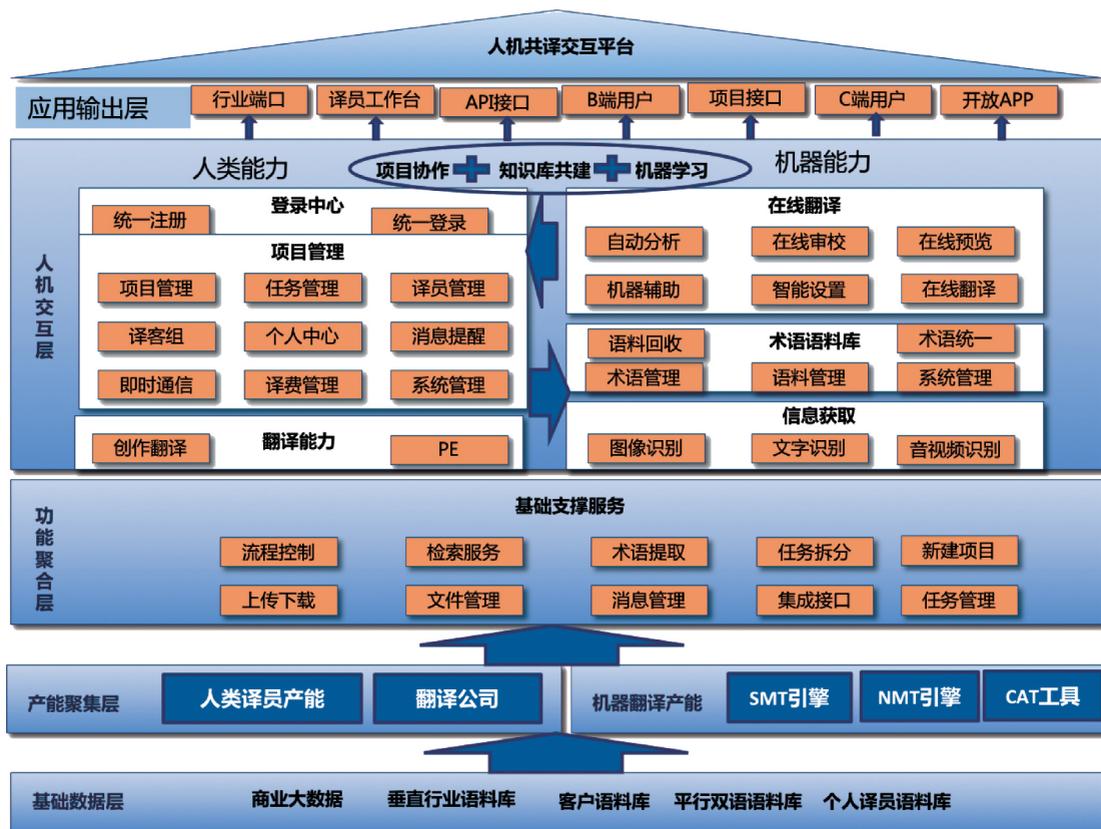


图1 人机共译交互平台

(表格注释: API—application programming interface 应用程序接口; B端用户—business 企业或组织的应用; C端用户—customer 个人用户; 开放APP—公开免费的APP接口; PE—post editing 译后审订)

3.4 人机共译平台在工程行业大规模招投标翻译项目中的应用分析 (传神语联网真实案例)

案例1: 2015年,中国电力工程行业海外并购项目需要在7个自然日内翻译完成多语种资料,14000多页文本,其中英语资料4000多页。内容涉及外方收购对象的公司概况、历年财务状况及重大案例情况。传神语联网公司利用人机共译交互平台,采用人机结合的方式,动用译员100余人,参与项目管理人员10余人,在7个自然日顺利完成项目交付,为客户提供了及时高质量的服务体验,同时,也为传神语联网在大规模工程行业集中需求处理方面积累了宝贵经验。

案例2: 2016年,南美某港口投标工程项目需要在4天内完成4000多页的西班牙语资料翻译,本次项目采用人机共译交互平台,在

取得客户充分信任的基础上,顺利完成了投标任务,同时帮助客户取得了关键信息,节省人工成本60%,大幅度缩短了人工周期。

案例3: 2014-2019年,北京一个工程类客户有多语种翻译长期需求。传神语联网公司采用人机共译交互平台,从项目派发,译员分配,知识库累积和更新,到译员培训,项目管理全流程持续精进,不断提高交付速度和质量。据统计,与该客户平均每年合作量不低于500万人民币,是传神语联网公司推广人机共译平台的典型成功案例。

4 结语

人机共译平台将聚合更多的机器产能和人工智能,两者将在实际翻译项目的运作中更紧密结合。机器产能将更多地利用平台学习人工产能,输出接近人工产能与质量的成果。工程
(下转第59页)

与“主子”相对。

3 《中国丛报》中国文化特色词译介影响

《中国丛报》译介了大量的中国文化特色词，这些词也是中国文化的关键词，是中国文化的内核，它们为西方读者了解中国文化提纲挈领，提供了快速通道。《中国丛报》的文化词英译为后来的译者提供了参考或“模板”，也为当下中国文化特色词英译提供了启示。可谓意义重大，影响深远。

3.1 将中国文化介绍到西方世界

《中国丛报》将中国的政治制度、社会生活、经济贸易、时政新闻等各个方面的内容介绍到西方世界，让西方世界了解中国。毫无疑问，这也促进了中西文化的碰撞、交流与融合。通过对中国经典的译介，将中国文学传播到西方，在西方社会掀起了汉学热，客观上起到“开文学之路，除两地之坑堑”的作用^[3]；将中国思想介绍到西方，促进了中西思想的对比讨论，并且在一定程度上影响了西方思想的发展。

3.2 为西方读者建构中国形象

不可否认，在编译过程中，译者实际上参与了中国形象的建构过程。《中国丛报》中国文化特色词的英译为英语读者建构了中国形象。这些形象中既有正面的，也有负面的，而且就整体而言，负面形象占据了主导地位。《中国丛报》是当时一份向西方全面介绍中国的报纸，通过它的译介，向西方展示了真实的中国，消除了西方人对中国的许多误解，同时通过中西对比，凸显了中国的落后和愚昧，彰显了西方的先进和文明，体现了殖民语境下翻译的典型特点^[4]。

3.3 为中国文化特色词英译提供参考

《中国丛报》所译中国文化特色词为这些词汇的翻译提供了参考，其中许多英文词汇后来成为这些汉语词汇的英译范本，为中西方作

者和译者广泛使用，为读者广泛接受。事实上，我们当下所使用的中国文化特色词的英译，许多都出自《中国丛报》的翻译。

3.4 为中国文化特色词翻译提供方法启示

如上所述，《中国丛报》中国文化特色词的翻译常用义译和直译。义译传递原文的真实意义，却无法揭示原文的表达方式。直译往往传递原文的字面意义，利于读者了解汉语原文的表达内容，不失为一种应对中西文化差异之困惑的方法^[5]，然而有可能遮蔽原文的真正意义或隐含意义，甚至可能导致无法理解。时至今日，对于中国文化特色词，义译和直译仍然是我们常用的翻译方法。《中国丛报》中国文化特色词的翻译方法具有开创和启蒙意义。

4 结语

《中国丛报》中国文化特色词译介无论对思想界、新闻界还是翻译界都具有积极的意义。透过《中国丛报》这一媒介和管道，西方世界开始关注中国、思考中国，这一过程无疑对近代中国“开眼看世界”的开明绅士寻求救国之道提供了新的视角和启示。《中国丛报》中国文化特色词译介促进了中西文化的交流与融合，也为译界提供了方法论上的思考，许多出自《中国丛报》的中国文化特色词为后人广泛使用，对翻译理论和译学发展也做出了积极的贡献。

5 参考文献

- 1 王化文. 《中国丛报》主要作者群研究. 商品与质量, 2011 (4): 99-100
- 2 顾钧. 中国的第一份英文刊物. 博览群书, 2011 (10): 94-97
- 3 姚莹. 康輶纪行. 中华书局, 1845/2014: 20
- 4 张建英. 《聊斋志异》在《中国丛报》的译介. 东方翻译, 2016 (6): 26-32
- 5 王海, 王乐. 《京报》英译活动的跨文化传播策略与技巧——以《中国丛报》文本为例. 国际新闻界, 2014 (10): 62-81

(上接第40页)

行业大规模的翻译需求对平台的依存可能成为人机共译的突破口，真正在翻译行业中形成人机共译大幅节约时间和人力成本的模式。

5 参考文献

- 1 <https://zhuanlan.zhihu.com/p/24501422>

- 2 秦颖. 基于神经网络的机器翻译质量评析对翻译教学的影响. 外语电化教学, 2018 (4)
- 3 <https://www.leiphone.com/news/201703/GNHHxktmFO-Mw3Qnk.html>
- 4 陆伟诚. 怎样做好工程翻译工作. 有色冶金设计与研究, 1995 (9)