

# 大数据背景下的译后编辑项目研究

陆强 罗洪燕 王丽萍

( 中国社 会 出 版 社 北 京 市 100030) ( 中 译 语 通 科 技 股 份 有 限 公 司 北 京 市 100131)

**摘要** 为了提高机器翻译的准确性、规范性和适用性,译后编辑尤为重要。本文基于机器翻译/译后编辑项目的现实要求,研究如何根据翻译准确性、完整性、术语一致性等方面的普遍诉求,制定通用的译后编辑质量规范。同时,根据不同项目的特点,辅以个性化的项目质量规范,作为项目文件,共同指导译后编辑人员开展工作。

**关键词** 译后编辑 语料 神经网络机器翻译

**Abstract** In order to improve the quality of machine translation ( MT ) outputs , post-editing ( PE ) becomes particularly important. Based on the needs of PE projects and such criteria as accuracy , completeness and terminological consistency , this article studies how to develop generic PE guidelines. Taking into account the features of individual PE projects , it also explores the need of project-specific quality checklists which , in conjunction with the generic guidelines , serve as the standards of edition for post-editors.

**Key Words** post-editing corpus Neural machine translation

DOI:10.16024/j.cnki.issn1002-0489.2019.03.011

## 引言

大数据时代,随着信息的几何式增长,翻译的需求急剧增加,以海量语料和神经网络为支撑的机器翻译技术,突破项目的时间和预算瓶颈,为超大型翻译项目提供了可行的解决方案。

“神经网络机器翻译”是一种基于神经网络的机器翻译 ( Machine Translation , MT ) 方法,通过编码-解码系统,将源语言序列进行编码,提取自然语言 ( 源语言 ) 中的信息,通过解码再将这种信息转换到另一种自然语言 ( 目标语言 ) 中,从而完成对语言的翻译。神经网络翻译技术的应用极大地提高了机器翻译的输出质量,推动了机器翻译系统的应用,从而为一些超大型翻译项目的实施真正提出了切实可行的方案。关于机器翻译译后编辑的 ISO 草案标准<sup>[1]</sup>指出,在时间紧迫和预算紧张的情况下,机器翻译系统为项目提供了可行的解决方案,借助机器翻译系统,可以翻译在其他条件下无法翻译的文字材料。换言之,机器翻译并非替代人工翻译,而是为依靠当前的人力或财力无法实现的翻译项目提供了最后的解决方案。

然而,ISO 草案同时也指出了机器翻译输出质量的局限性以及译后编辑 ( Post-editing , PE ) 的必要性。“机器翻译的译后编辑” ( MTPE ) 是指机器翻译和人工编辑相结合,由受过专业训练的译后编辑人员对机器翻译输出进行检查、纠正、改进,以提高译文质量和适用性的过程。作为翻译行业的新兴领域,译后编辑的研究文献相对较少,为实际项目量身定做的指南性文件更加稀缺,而逐渐涌现的数据库等大型翻译需求亟需我们对 MTPE 项目质量标准的制定、质量控制和译后编辑人员的培训开展进一步的研究。

为了更好地服务大型 MTPE 项目客户的现实需要,同时为了进一步研究译后编辑发展趋势,开展译后编辑培训,并为机器翻译系统提供改进意见,中国社 会 出 版 社 与 中 译 语 通 公 司 针对中国政区大典项目专门成立了 MTPE 项目指南编制小组,针对不同领域的中文和英文语料开展双向测试研究。

## 1 MTPE 项目的质量标准和典型错误

初期测试选取了近 30 名译员参加,从业年限从 3-20 年不等,对 CAT 相关技术有不同程度的了解与应用,但对 MTPE 项目较为陌

生。测试前进行了简单的普及性培训,并参考相关文献和已有 MTPE 项目的质量要求,制定了通用的质量标准,包括必须修改项(理解/语序/断句错误、漏译、术语错误/不一致和搭配不当、成分残缺等语法错误)和不建议修改项(如语义的过度增补和背景诠释和语体风格的过度修改,特别是对信息传递基本准确的内容进行的润色和修改)。

根据测试结果,典型错误包括以下四类:

### 1.1 理解/语序/断句错误

例1 ...will have an impact on the city including catalytic effects on urban development and rehabilitation.

MT: 对城市产生影响,包括对城市的催化作用发展和复兴

错误分析 “催化作用” “发展和复兴” 语序颠倒

PE: 对城市产生影响,包括对城市发展和复兴的催化作用

例2 ...是 21 世纪海上丝绸之路联通大洋洲、欧洲和非洲等地区的关键节点,在全球战略上居重要地位。

MT: It is the 21st century maritime Silk Road connecting Oceania. Key nodes in regions such as Europe and Africa play an important role in global strategy.

错误分析 “关键节点” 前后断句错误

PE: It is an important country along the 21st-century Maritime Silk Road which connects Asia with Oceania, Europe and Africa, and plays an important role in the global strategy.

### 1.2 漏译

例1 It is therefore important to establish at the onset leadership within the department.

MT: 因此,重要的是在部门内建立领导。

错误分析: 机译漏译“at the onset”(一开始)

PE: 因此,重要的是一开始就在部门内建立领导地位。

例2 ... each theme has a specific deliverable for transport. . .

MT: 每个主题都有具体的可交付成果

错误分析: 漏译“for transport”

PE: 针对交通领域,每个主题都有具体的可交付成果

### 1.3 术语错误/不一致

例1 经济新常态

MT: New economic normality

错误分析: 术语“新常态”翻译错误

PE: economic “new normal”

例2 吐鲁番……吐鲁番……吐鲁番

MT: Turpan. . . Turfan. . . Tulufan

错误分析: 在同一篇章中,地名翻译不一致

PE: (地名统一译为) Turpan

### 1.4 语法错误(搭配不当、成分残缺等)

例1 (XX 省) 耕地面积 21728.8 万亩,是国家重要的商品粮生产基地。

MT: The total area of cultivated land 217, 288, 000 mu, is an important commodity grain production base of the State.

错误分析: 原文的语法不规范地复制到机器翻译输出,译后编辑中应对其错误予以相应修改。

PE: The total area of cultivated land is 217, 288, 000 mu, making the province an important commodity grain producer of the country.

例2 印尼群岛东西达 5 300 公里,南北约 2 100 公里,其中 6 000 个岛屿有人居住。

MT: Indonesia Islands to 5, 300 km east-west, about 2, 100 km north-south, of which 6, 000 islands are inhabited.

错误分析: 句子缺少谓语成分

PE: The Indonesian archipelago extends 5, 300 km east-west, about 2, 100 km north-south, and 6, 000 of the islands are inhabited.

## 2 其它发现

经过对机器翻译输出的质量研究,发现机器翻译质量与源语文件类型有一定关系,但是影响因素更多的是用于机翻训练的语料的相关性和丰富程度以及原文的规范程度。因此,即便在同一类型文件中,机器输出的质量也经常存在差异。如:

例1 包容性增长又称包容性发展或共享式增长,指在保持较快经济增长的同时,更多关注社会领域发展,关注弱势群体,让更多的人享受经济全球化的成果。

MT Output:

Inclusive growth, also known as inclusive development or shared growth, refers to maintaining

faster economic growth while paying more attention to social development, focusing on vulnerable groups, so that more people enjoy the fruits of economic globalization.

例2 实施国际产能合作是顺应经济全球化潮流,秉持开放的合作精神,致力于维护全球自由贸易体系和开放型世界经济,深入参与世界分工与合作、促进经济要素有序自由流动、资源高效配置和市场深度融合,保持经济社会持续健康稳定发展的现实选择。

MT Output:

The implementation of international capacity cooperation is in compliance with the trend of economic globalization, uphold the open spirit of cooperation, is committed to safeguarding the global free trade system and open world economy, in-depth participation in the World Division of Labor and Cooperation, promote the orderly and free flow of economic elements, efficient allocation of resources and market depth integration, to maintain the economic and social sustainable and healthy development of the realistic choice.

这两个译例均出自社科类文本,但机器翻译输出的效果差异极大。例1的语法、断句正确,所需的译后编辑工作量较少;相比之下,例2对长句的判断发生错误,兼之体例和术语问题,所需的译后编辑工作量较大。

此外,通过译后编辑的时间统计发现,未经培训的译后编辑人员并不一定都能提高效率。其中约一半以上的译员相较日常的翻译速度与效率,有较明显的提升,6-7人无明显提升,有4-5人甚至有速度的下降。

按照事先设置的必改项对全部测试样本进行质量评定发现,时间和质量并无明显的正相关性,影响修改质量的相关因素更多的是译员本身的资历以及对相关领域的熟悉程度。

按照事先设置的不建议修改项对全部测试样本进行质量评定发现,从业年限与过度修改现象不存在明显的相关性,但相较从业年限处于中等时段(5-10年)的译员,从业年限较长或较短的译者更容易因为语言风格等因素,对机器翻译的输出进行大规模修改。

### 3 进一步的启示和结论

#### 3.1 机器翻译输出的质量

ISO 在关于机器翻译译后编辑的标准草案

中指出“翻译服务商需明确相关流程,在必要情况下与客户协商,源语言的内容是否适合机器翻译和后续的译后编辑,机器翻译+译后编辑的效率取决于机器翻译系统、语言和原文内容涉及的领域/风格。”在 MTPE 项目测试中笔者发现,机器翻译输出的质量和原文的规范程度以及对机器翻译引擎进行训练的底层语料具有较大的关联性,同一类型甚至同一篇章中不同段落的机器翻译效果和译后编辑难度都存在差异,有时很难按照专业标准对机器翻译输出效果进行一刀切的预判。因此,对于具体的 MTPE 项目而言,有必要在项目启动前,评估项目客户质量预期与机器翻译输出效果的差异,从而确保译后编辑的效果和效率。

#### 3.2 译后编辑的质量标准

在确定译后编辑项目的质量标准时,需要明确对质量的预期以及译后编辑的程度。Sharon O'Brien<sup>[2]</sup>针对译后编辑程度的不同提出了两组术语: Fast/Rapid/Gist/Light Post-Editing 以及 Conventional/Full Post-editing。对于后者,TAUS<sup>[3]</sup>提出了“出版级”(publishable quality)或“近似高质量人工翻译加审校”的质量标准,而对于前者,则提出了“可接受”(good enough)或“适用”(fit for purpose)的质量标准。TAUS 将“good enough”的质量标准进一步细化为: 1) 确保语义准确; 2) 无信息误删或误增; 3) 对具有攻击性、文化上不可接受的、不妥当的内容进行编辑; 4) 尽可能多利用 MT 输出的译文; 5) 遵循拼写、标点、基本规范; 6) 无需对文风进行修改; 7) 无需因追求表述流畅而对句式进行调整。

项目组参照上述原则初步制定了 MTPE 项目的通用质量标准。但在实际业务中,鉴于对机器翻译输出质量的评价和客户对成品质量预期的复杂性,通用标准无法完全满足项目的需求,还需要根据不同项目的性质与客户进行反复沟通,确定相应的质量标准,确定项目质量规范,发给译后编辑人员参考。

#### 3.3 译后编辑人员的技能要求

在 MTPE 项目测试中,我们发现,合格的翻译员不一定是合格的译后编辑人员,然而合格的译后编辑首先必须是合格的翻译,译后编辑人员和翻译相比,并不意味着专业性和能力有所退化,相反,所需要的技能更多且更广

(下转第 26 页)

and is not a partnership.

试译: 既包括自然人也包括 (a) 由两个或三个并非全部是法人团体的机构构成的合伙; (b) 非法人团体或者不完全由法人团体构成且不具备合伙资格的机构。

释义 (a) (b) 中的 person 并非指自然人而是 corporation that is recognized by law as the subject of rights and duties (《韦氏法律词典》<sup>[9]</sup> p. 358)。显然, 从概念的逻辑关系分析, (例 3) 中 individual 的所指应该是 (a) (b) 所提的机构, 而不是自然人, 所以不能译为“个人”。

## 2.2 《美国著作权法》地域适用范围存在误译

著作权是知识产权的一种, 而知识产权的一个重要法律特征是其具有地域性。在《美国著作权法》中, “美国” (the United States) 是一个术语, 具有非常严格的定义:

The “United States”, when used in a geographical sense, comprises the several States, the District of Columbia and the Commonwealth of Puerto Rico, and the organized territories under the jurisdiction of the United States Government.

地理意义上的美国, 是指数个州、哥伦比亚特区、波多黎各自由邦和由联邦政府管辖的领地。

译者将 “the several States” 译为 “数个州” 是一种误译, 此处的 several 并非指数目, 《英美法律术语词典》对其的解释是: applied

or regarded separately. Often contrasted with “joint”. 各自的; 分别的<sup>[10]</sup>。 “the several States” 指的是美国各州, several 体现了各州之间互不隶属的关系, 因此, 我们可以将其译为 “50 个州” 或者 “各州”, 而不是 “数个州”。

## 3 结语

术语译者有时也可视作新术语的创制者, 翻译时, 译者不能率尔操觚, 须综合考虑法律英语术语翻译所涉及的诸项原则, 尤其是术语定名的原则, 权衡利弊, 谨慎翻译。总之, 术语翻译绝非易事, 译者必须具备一定的术语学知识和术语意识。

## 4 参考文献

- 1 杜颖, 张启晨. 美国著作权法. 北京: 知识产权出版社, 2013: 3, 7, 10, 12, 18
- 2 美国版权局官网 Reliance Party 定义. <https://www.copyright.gov/docs/reliance.html>
- 3 格里尼奥夫. 术语学. 北京: 商务印书馆, 2011: 103
- 4 李克兴. 论法律文本的静态对等翻译. 外语教学与研究, 2010 (1): 59-65
- 5 冯志伟. 现代术语学引论. 北京: 商务印书馆, 2011: 36
- 6 张雨泽. 版权法导论. 北京: 知识产权出版社, 2009: 17
- 7 牛津高阶英汉双解词典. Hornby. 商务印书馆, 1997
- 8 朗文法律词典. 北京: 法律出版社, 2007: 299
- 9 韦氏法律词典. 北京: 中国法制出版社, 2014
- 10 张法连. 英美法律术语词典. 上海外语教育出版社, 2014: 389

(上接第 37 页)

泛。崔启亮<sup>[4]</sup>认为, “创建高质量译后编辑内容需要具有多种技能: 积极的态度, 理解和认同机器翻译输出译文的翻译方式, 了解机器翻译的基本原理, 了解机器翻译的错误类型, 熟悉所涉及的项目、领域和术语以及翻译环境。具有需要快速阅读、识别影响读者理解的重大机器翻译错误、修改与原作者意图相反的译文的能力。”在大数据时代已经到来的今天, 尽管围绕新技术还有各种争议, MTPE 作为一种为客户提供的最后解决方案, 已悄然成为现实。如何提高自身翻译与技术素养, 掌握更多技能, 在时代变化中立于不败之地, 是值得翻

译从业人员深思的问题。

## 4 参考文献

- 1 ISO/TC 37. ISO 18587 Translation services — Post-editing of machine translation output — Requirements. 2014
- 2 Sharon O’ Brien, Introduction to Post-Editing: Who, What, How, and Where to Next? 2010. <https://amta2010.amtaweb.org/AMTA/papers/6-01-ObrienPostEdit.pdf>
- 3 TAUS. MT Post-editing Guidelines. 2016. <https://www.taus.net/think-tank/articles/postedit-articles/taus-post-editing-guidelines>
- 4 崔启亮. 论机器翻译的译后编辑. 中国翻译, 2014 (6)