

神经机器翻译质量与译后编辑实证研究

裘白莲^{1,2}, 王明文^{1*}, 罗琪³, 李茂西¹

(1. 江西师范大学计算机信息工程学院, 江西 南昌 330022;

2. 华东交通大学外国语学院, 江西 南昌 330013;

3. 江西师范大学图书馆, 江西 南昌 330022)

摘要: 随着神经机器翻译质量的提高, 机器翻译+译后编辑的工作模式在翻译行业越来越普遍。本文针对神经机器翻译质量对译后编辑过程的影响进行了研究, 考察机器翻译质量、源文特征对译后编辑努力的影响, 同时观察、比较职业译员和学生译员译后编辑过程的异同。实证研究发现机器译文质量影响译后编辑时间努力和技术努力, 而源文句长对时间努力没有影响。职业译员译后编辑速度快于学生译员, 其所做的译后编辑操作少于学生译员。汉英语言对神经机器翻译译后编辑实证研究有助于促进用户角度的机器翻译研究, 深化对译后编辑过程的了解, 同时为译后编辑教学提供有效参考。

关键词: 神经机器翻译; 译后编辑; 译后编辑努力; 过程研究

近年来随着神经机器翻译 (Neural Machine Translation, NMT) 的发展, 机器翻译质量在很多语言对上大有改进。Junczys-Downmunt 等^[1]在 30 多个语言对上将神经机器翻译和统计机器翻译进行对比, 结果表明, 神经机器翻译在 27 个任务上超过了基于短语的统计机器翻译 (PBMT), 这展现了神经机器翻译的强大能力。但整体而言, 神经机器翻译还达不到专业人工翻译水平。只有在一些特定场景, 如垂直细分领域、控制语言、专用机器翻译系统等, 机器翻译质量能达到特定要求, 直接可用, 而在其他场景下一般都需要人工译员对机器译文进行译后编辑 (Machine Translation Post-editing, MTPE)。译后编辑是对机器翻译译文进行修改和加工以使其符合客户质量要求的过程^[2]。大量实证研究表明, 译后编辑与人工翻译或翻译记忆 (translation memory, TM) 辅助翻译相比, 效率更高, 译后编辑可以节省时间, 提高速度^[3-6]。

机器翻译+译后编辑的模式在翻译行业越来越普遍。卡门森斯咨询公司 (Common Sense Advisory, CAS) 2019 年发布的报告显示, 有 51% 的受访企业使用机器翻译, 其

中 81% 部署了 NMT^[7]。王华树、陈涅奥^[8]调查显示, 受访企业中 89.91% 使用机器翻译, 受访企业平均在 30% 左右的翻译项目中采用 MTPE 模式。

神经机器翻译提高了机器翻译的质量标准, 但是以自动评价指标或人工评价来度量的质量提升, 能否带来译后编辑效率的提高, 这仍是个开放的问题, 相关的实证研究还不多。Pereira^[9]研究表明, 统计机器译文和神经机器译文中, 译后编辑人员更偏好后者。Läubli 等^[10]将 TM 辅助翻译与神经机器翻译译后编辑进行对比, 结果表明, 译后编辑可以节省大量时间, 而且其译文质量与 TM 辅助翻译相同或略有提升。Zouhar 等^[11]发现, 性能更好的神经机器翻译系统, 其译后编辑过程中修改更少。国内相关研究仍处于起步阶段, 汉英语言对神经机器翻译译后编辑的实证研究还很少。卢植、孙娟^[12]通过眼动实验, 发现译后编辑比人工翻译可缩短翻译时间, 提高翻译效率, 减少认知努力, 但文中未交待所使用的机器译文来源。王湘玲、王婷婷^[13]将人工翻译与机器翻译译后编辑进行对比实证研究, 发现译后编辑能显著缩短任务时间, 提高翻译速度。其机器译文来自谷歌翻译。王湘玲、赖思、贾艳芳^[14]

收稿日期: 2022-07-24; 录用日期: 2022-

基金项目: 国家自然科学基金 (61876074, 61662031); 教育部人文社科基金 (基于细粒度神经机器翻译错误分析的译后编辑努力研究, 21YJC740040)

*通信作者: mwwang@jxnu.edu.cn

基于眼动追踪和键盘记录数据,对人工翻译与谷歌 NMT 译后编辑进行比较研究,发现译后编辑在处理隐喻表达时,可以减少译者认知负荷,提高译文质量。以上三项研究的被试均为学生译者,没有职业译员参与。

本文拟针对神经机器翻译译后编辑回答以下三个问题。第一,神经机器翻译译后编辑效率如何,机器翻译质量对译后编辑努力产生什么影响。第二,源文句长对译后编辑是否产生影响。第三,职业译员和学生译员的译后编辑过程是否有差异,有什么差异。

1 译后编辑努力及影响因素

Krings^[15]将译后编辑努力(post-editing effort)分为三个层面:1) 时间努力;2) 认知努力;3) 技术努力。时间是指译后编辑所需的总时间。认知努力指思考的过程,包括识别机器译文中的错误以及计划如何修改错误。技术努力即实际编辑操作,是指身体动作,在识别出错误和计划好修改方法之后产生译后编辑译文(PE 译文)所需的击键、剪切和粘贴操作。在译后编辑工作流程中,译后编辑努力受到许多因素的影响,除了译后编辑人员个体差异、经验、工具平台,还有机器译文质量、源文特征、错误类型等。

许多研究表明,对高质量机器译文进行译后编辑比人工翻译效率更高。但是,如果机器译文质量低,译后编辑效率反而不如人工翻译^[16]。

机器翻译通常使用自动评价指标来表示机器翻译质量。Tatsumi 等^[17]考察四种自动评价指标(BLEU、HTER、NIST、GTM)与译后编辑时间的相关性,发现四种自动评价指标都与译后编辑时间具有较强的相关性,其中,GTM 的相关性最强。Koehn 和 Germann^[18]对比四种机器翻译系统,发现其中性能最好的系统比最差的系统译后编辑效率提高约 20%。Sanchez-Torrón 和 Koehn^[19]研究发现机器译文 BLEU 值每高 1 个点,每单词译后编辑时间就减少 0.16 秒,减少约 3-4%。这说明机器翻译质量越高,译后编辑效率也越高。该研究还发现 BLEU 值最低的机器翻译系统,其译后编辑译文质量最低,译后编辑

努力最高。Daems 等^[20]研究发现,大部分译后编辑努力指标,包括产品指标和过程指标,都受到机器翻译质量的影响。Stasimioti 等^[21]对比三种机器翻译系统,发现翻译质量最好的系统,即 BLEU 和 METOER 值更高、WER 和 TER 值更低的系统,其所需的译后编辑时间努力(每句所用时间)、技术努力(击键数)和认知努力(眼动数据)均最少。Zouhar 等^[11]研究神经机器翻译质量对译后编辑的影响,结果表明,机器译文质量更高,译后编辑过程中的修改就更少,即机器译文质量与技术努力相关,但是其研究发现机器译文质量与译后编辑时间不直接相关。

翻译过程中,源语文本的特征和难易度会影响译者的行为,在译后编辑过程中同样如此。相关研究发现源文特征是影响译后编辑努力的主要因素之一,源文特征主要包括源文句子长度和句子结构等。Tatsumi^[17]发现源文句子长度和句子结构,以及特定错误类型,导致译后编辑时间长。Tatsumi 和 Roturier^[22]发现非常长和非常短的句子产生更多的译后编辑次数,而不完整句和复杂句、复合句需要更多译后编辑次数。Koponen^[23]发现非常长的句子一般在人工评价中得分低,对译后编辑努力的估计与句长强相关。Vieira^[24]发现介词短语会增加译后编辑努力。

译后编辑是修改机器译文中的错误,因此,机器译文的错误情况,如错误数量、错误频率等,会对译后编辑努力产生影响。除此之外,错误类型也对译后编辑努力产生影响,有些错误类型比其他错误类型更难修改。Koponen 等^[25]表明错误类型影响译后编辑时间,修改一些难度更小的错误类型需要的译后编辑时间更短,如词形错误、同义词替换等。Popović 等^[26]考察五种编辑操作类型(修改词形错误、修改词序错误、增加漏词、删除增词、修改选词错误)与认知努力和译后编辑时间的关系,发现修改错词和词序错误的编辑操作需要的认知努力最多,修改错词所需要的译后编辑时间最多,删除增词对译后编辑认知努力和时间的影响都很小。Zaretskaya 等^[27]发现有些错误类型,如错词、过于直译、命名实体,更难进行译后编

辑，带给译后编辑者更多挑战。

2 实验

2.1 被试

本研究中译后编辑实验的被试分为学生译员和职业译员两组。10 名学生译员选自本地高校翻译专业本科四年级学生，其中包括 8 名女生，2 名男生；年龄最小 21 岁，最大 24 岁，平均年龄 22.2 岁，均以汉语为母语，英语为第二语言；所有学生译员均熟悉计算机操作，完成了计算机辅助翻译课程学习，成绩合格，在课程中接触过译后编辑，但无其他翻译或译后编辑经验。学生均自愿参加译后编辑实验。

5 名职业译员均从事翻译职业，以汉语为母语，英语为第二语言；均为女性，年龄最小 30 岁，最大 39 岁，平均年龄 33.2 岁，从事翻译职业年限最少的 5 年，最多 14 年，平均从业年限 8 年；其中 4 名职业译员具有全国翻译专业资格（水平）考试（CATTI）二级笔译证书，另 1 名具有上海英语高级口译证书；所有职业译员均了解译后编辑，并在工作中使用过译后编辑。职业译员均有偿参加本次译后编辑实验。

2.2 实验材料

译后编辑实验材料包括两篇英汉翻译文本，均来自 WMT2019 英汉新闻机器翻译测试集。实验使用的机器译文是 KSAI 组（金山 AI）提交的神经机器翻译译文，该小组在 WMT2019 英中机器翻译任务中人工打分排名第一。这代表了当时机器翻译的最高水平，具有很好的翻译质量。KSAI 组基于多种神经机器翻译模型，以 Transformer 作为基线系统，使用了几种数据过滤和回译作为数据清洁和数据增强的方法。其最终模型是经过多模型集成、重排序、后处理的系统组合^[28]。

实验材料为新闻领域材料，难度适中。文本 1 有 21 个句子，468 个英文词。文本 2 有 21 个句子，528 个英文词。每位译员均完成两个文本的译后编辑任务。42 个句子经过 15 位译员译后编辑，共产生 630 个经过译后编辑的句子和相关数据用于分析。研究者向被试说明译后编辑质量要求，提供译后编辑指南。参照 TAUS 译后编辑指南，采用轻度

译后编辑原则，本次译后编辑指南如下：（1）力求译文意思正确、语句通顺；（2）确保没有信息增加或遗漏；（3）尽可能多地使用机器译文；（4）除非影响语义，否则不修改句子结构；（5）单纯的风格问题无需修改。

2.3 实验流程

本研究使用 Translog-II^[29] 击键记录软件收集被试的译后编辑过程数据和最终译文。Translog-II 软件的用户操作界面为左右两栏，左侧呈现源语文本，右侧呈现机器译文。被试可在右侧进行译后编辑，修改机器译文（见图 1）。

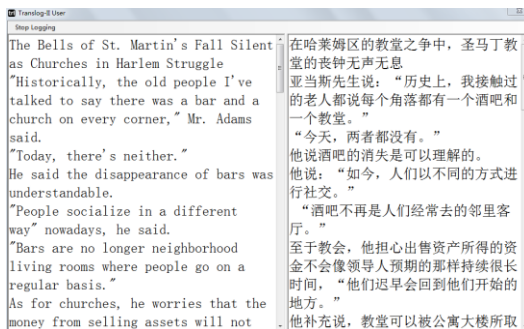


图 1 Translog-II 译后编辑操作界面

为方便数据采集，学生译员译后编辑实验安排集中完成，被试均在语言实验室电脑上完成实验任务，所有电脑上均预先安装了 Translog-II 软件。两个文本的译后编辑任务分为两次进行，中间间隔一周。实验流程相同。实验正式开始之前，向被试介绍研究目的、译后编辑指南，以及软件使用方法等。被试了解实验任务之后签署知情同意书，然后填写前测问卷，用于收集被试性别、年龄等信息。正式实验开始，被试根据译后编辑指南对机器译文进行译后编辑。为最大限度接近真实翻译过程，实验中被试可以查询网络词典等资源。

职业译员译后编辑实验分散进行，均在个人电脑上、自行安排时间完成。研究者通过书面方式告知职业译员此次研究目的、任务要求、译后编辑指南、软件使用方法等。职业译员填写前测问卷，用以收集职业译员年龄、性别、从业年限、资格证书等信息。本研究采用 SPSS25.0 对定量数据进行统计分析。

3 实验结果和分析

3.1 机器翻译质量与译后编辑努力

本节阐述机器翻译质量对译后编辑努力的影响,包括时间努力和技术努力。由于认知努力是思考过程,无法直接度量,其数据较为复杂,将另文论述。

3.1.1 时间努力

为考察机器翻译质量对译后编辑时间努力的影响,我们使用机器翻译自动评价指标 BLEU 和 TER,考察这两者与译后编辑时间努力的关系。BLEU 和 TER 均反映机器翻译质量,BLEU 是通过机器译文与参考译文的相似度来反映机器翻译质量,BLEU 值越高,表示机器翻译质量越好;而 TER 是通过编辑距离反映机器翻译质量,TER 越低,表示机器翻译质量越好。译后编辑实验中,两个文本共 42 个句子,首先计算每个句子的 BLEU 和 TER。BLEU 使用 mosesdecoder 的 mtevalv13a.pl 计算,TER 使用 terp^[30] 计算。职业译员组和学生组分开讨论,分别计算每个组的平均译后编辑时间。每句单词数除以每句平均译后编辑时间,得到每句平均译后编辑速度,单位为词/秒。两个组每个句子的 BLEU 与平均 PE 速度散点图如图 2。



图 2 BLEU 与平均 PE 速度散点图

图 2 中 x 轴为 BLEU, y 轴为平均 PE 速度。可见,随着 BLEU 值提高,职业译员和学生译员组的平均 PE 速度均提高。即随着机器译文质量提高,PE 速度更快。这种趋势在职业译员组更为明显,可见,机器翻译质量提高对职业译员译后编辑收益更大。

使用 TER 和平均 PE 速度进行分析,得到的结论与 BLEU 与平均 PE 速度的情况相同。

随着 TER 提高,即机器译文质量更差,职业译员和学生译员的平均 PE 速度均降低。通过 BLEU、TER 与 PE 速度的关系,本研究在英汉翻译方向上证实了 Tatsumi 等^[17]、Sanchez-Torron 和 Koehn^[19] 机器翻译质量越高译后编辑效率越高的研究结论。

3.1.2 技术努力

使用 BLEU 表示机器翻译质量,用 HTER 表示译后编辑技术努力。HTER 是每个句子机器译文与 PE 译文之间的编辑距离,可以表明句子经过了多少译后编辑操作,反映译后编辑技术努力。

Snover 等^[31] 提出 TER 计算公式如下:

$$TER = \frac{\# \text{ of edits}}{\text{average \# of reference words}} \quad (1)$$

TER 是将机器译文修改成与参考译文相同所需的最少编辑次数 (# of edits) 除以参考译文的平均长度 (average # of reference words)。HTER 同样通过公式(1)计算得到,将公式中的参考译文换成译后编辑译文即可。HTER 值越低表示机器译文与 PE 译文越相似,值越高表示机器译文与 PE 译文差别越大。

我们使用机器译文与每个译员每个句子的 PE 译文,分职业译员和学生译员组计算每个句子的平均 HTER。图 3 中, x 轴为 BLEU, y 轴为 HTER。可见,随着 BLEU 提高,HTER 降低,即机器译文质量越高,译后编辑人员所进行的译后编辑操作越少。职业译员组和学生译员组均呈现相同趋势。这表明,机器翻译质量影响译后编辑技术努力。



图 3 BLEU 与 HTER 散点图

3.2 源文句长与译后编辑时间努力

本节分析源文句长对译后编辑时间努

力的影响。译后编辑实验任务共 42 个句子，以每句的单词数计算句长。句长情况见表 1，42 个句子句长平均值为 23.61，最小值为 4，最大值为 51，标准差为 12.06。句长分布差异较大。

表 1 句长描述统计

	均值	标准差	最小值	最大值
句长	23.61	12.06	4.00	51.00

考察句长与每句平均译后编辑时间的关系，结果发现源文句长和平均译后编辑时间的 Pearson 相关系数为负，且不显著。这说明并非句长越长，译后编辑时间越长。也就是说，句长与平均译后编辑时间不相关。有趣的是，这个结果与文献[17]、[23]的结论相反。但是与文献[27]的研究结论比较一致，他们发现句长和译后编辑时间仅弱相关。我们的结果表明，在英汉翻译方向，译后编辑时间受机器翻译质量影响，不受源文句长影响。

3.3 职业译员与学生译员比较

3.3.1 译后编辑效率

根据以往相关研究，职业译员和学生译员在改错、翻译任务中的表现是不一样的。Sommers^[32]发现，在改错任务中，专家采取非线性策略，更关注意义和结构，而学生改错集中在句子层面，很少调整语序或增加信息。Hayes 等^[33]发现专家先关注文本的整体结构，而新手关注文本的表面。在翻译任务中，没有经验的译员将翻译任务视为主要是词汇任务，而职业译员更关注连贯和风格问题。学生译员完成翻译任务的时间比职业译员更长^[34]，注视点数和停顿数更多^[35]。

de Almeida 和 O' Brien^[36]发现更有经验的译者在译后编辑任务中速度更快，做出的关键修改和个人偏好修改更多。Moorkens 和 O' Brien^[37]发现职业译员译后编辑效率比新手译员更高，但是对译后编辑的态度更消极。相反，Guerberof^[6]发现有经验译后编辑人员与新手在译后编辑速度上没有显著差异。Daems 等^[20]研究发现学生译员停顿时间远远高于职业译员，学生译员需要更多时间思考怎么修改机器译文错误。

我们首先比较职业译员与学生译员的译后编辑效率，效率体现为速度。译后编辑速度通过将源文词数除以译后编辑时间（秒）得到，单位为词/秒。在数据统计之前，按照 Tatsumi^[38]等类似研究的做法，首先删除异常值，以免影响分析结果。异常值是指译员处理某个句子的时间异常长，此时译员的关注点可能不在译后编辑任务上。表 2 和表 3 表明职业译员和学生译员句子级平均译后编辑速度，TR 表示职业译员，ST 表示学生译员。

从表 2 可见，PE 速度最快的职业译员是 TR2，速度为 0.716 词/秒，大致相当于 2500 词/小时。PE 速度最慢的是 TR4，速度为 0.274 词/秒，大致相当于 1000 词/小时。也就是说，在英汉翻译方向，职业译员 NMT 译后编辑速度大致为 1000-2500 词/小时。

从表 3 可见，学生译员中 PE 速度最快的是译员 ST9，速度为 0.342 词/秒，相当于 1200 词/小时。PE 速度最慢的是译员 ST4，速度为 0.229 词/秒，大致相当于 800 词/小时，即学生译员 NMT 译后编辑速度大致为 800-1200 词/小时。与职业译员相比，学生译员 PE 速度更慢。此外，学生译员标准差比职业译员更小，表明学生译员个人差异和学生译员之间的差异更小，而职业译员的差异大得多，即个体差异在职业译员上体现得更加明显。

表 2 职业译员 PE 速度（词/秒）

	均值	标准差	最小值	最大值
TR1	0.528	0.425	0.09	2.00
TR2	0.716	1.032	0.02	4.20
TR3	0.464	0.609	0.05	4.00
TR4	0.274	0.306	0.02	1.62
TR5	0.473	0.264	0.12	1.25
均值	0.491	0.527		

表 3 学生译员 PE 速度（词/秒）

	均值	标准差	最小值	最大值
ST1	0.304	0.314	0.03	1.70
ST2	0.282	0.469	0.06	3.11
ST3	0.252	0.274	0.02	1.26
ST4	0.229	0.201	0.05	1.16

ST5	0.240	0.232	0.02	1.04
ST6	0.232	0.154	0.08	0.68
ST7	0.315	0.307	0.02	1.20
ST8	0.334	0.278	0.04	1.18
ST9	0.342	0.316	0.05	1.63
ST10	0.280	0.343	0.05	1.70
均值	0.281	0.289		

从表 2、3 可知，职业译员平均速度为 0.491 词/秒，学生译员平均速度为 0.281 词/秒，职业译员比学生译员平均 PE 速度快 0.21 词/秒。这与文献[37]在英德译后编辑实验中的研究结果一致。他们发现职业译员平均速度是 0.387 词/秒，学生译员 0.126 词/秒，职业译员平均 PE 速度比学生译员快 0.261 词/秒。我们的职业译员与学生译员的速度差与他们的实验结果非常接近。他们用于译后编辑的机器译文是 SMT 产出，我们是 NMT 产出，我们实验中职业译员和学生译员的 PE 速度均比他们的更快。文献[4]在英语到法语、德语和西班牙语 3 个翻译方向 SMT 译后编辑实验中，职业译员 PE 速度为 800-1800 词/小时。与以上两个研究相比，我们的 PE 速度都更快，如果不考虑语言对的差异，这说明 NMT 译后编辑的速度快于 SMT。

3.3.2 编辑距离

编辑距离是指两个字符串之间，由一个转成另一个所需要的最少编辑操作次数。本文使用 HTER 表示译后编辑译文与机器译文之间的编辑距离。

本次实验中，职业译员组篇章级 HTER 平均值为 18.05。该组进行译后编辑最多的句子 HTER 值 92.86。该句机器译文及职业译员译后编辑译文见例 1。ST 表示源文句子，MT 表示机器译文，PE 表示译后编辑译文。该例中，职业译员对机器译文做了大量修改，主要体现为将机器译文非正式的口语风格改成了较正式的书面语风格，如将“为什么”改为“因此”，“对……来说”改为“对……而言”，“今天”改为“如今”，“是很重要的”改为“十分重要”。

例 1 职业译员译后编辑例子

ST: This is why putting the investment into language teaching now is important for today's children.

MT: 这就是为什么现在对语言教学的投入对今天的孩子来说是很重要的。

PE: 因此，投入开展语言教学对如今的孩子而言十分重要。

学生译员组篇章级 HTER 平均值为 27.57。与职业译员组相比，学生组的译后编辑技术努力更高，即该组所做的译后编辑操作更多。其中进行译后编辑最多的句子 HTER 值为 146.15。该句机器译文及学生译员译后编辑译文见例 2。该例中，机器译文“放在比赛的前面”系词义错误，学生译员正确地将之修改为“增强……竞争力”，同时修改了机器译文中不流畅的表达，如将“来自外国的人”改为“外国人”，“多语言的人”改为“多语种学习者”等，还调整了句子结构。这些修改导致该句 HTER 高，译后编辑操作多，技术努力高。

例 2 学生译员译后编辑例子

ST: Being able to communicate directly with people from foreign countries will automatically put a multilingual person ahead of the competition.

MT: 能够与来自外国的人直接交流，将自动把一个多语言的人放在比赛的前面。

PE: 与外国人交流的能力可以增强多语种学习者的竞争力。

在编辑距离上，我们的结果与文献[37]观察到的情况相反。他们职业译员组篇章级 HTER 为 30.31，而新手译员篇章级 HTER 为 27.15，即新手译员进行的译后编辑操作更少。可能的原因在于，本次实验中，由于学生译员缺乏译后编辑经验，对轻度译后编辑原则的理解不够到位，或者对机器翻译了解不够、信心不足，倾向于做更多的修改。相反，职业译员对机器翻译和轻度译后编辑原则的了解更深入，能够进行更精准、更少量

的修改，同时保证 PE 译文质量。

例 3 给出了部分职业译员 PE 译文和学生 PE 译文的比较。该例中，机器译文中有三处错词，分别为“辆”、“车”、“铃铛”。职业译员点对点精准地修改错词，而学生译员除了修改错词，还对句子结构或语序进行了部分调整，因此对机器译文进行的修改更多，HTER 自然更高。

例 3 部分 PE 译文比较

ST: The Rev. David Johnson, Father Johnson's son and successor at St. Martin's, proudly called the carillon "the poor people's bells."

MT: 约翰逊神父的儿子、圣马丁教堂的继承人大卫·约翰逊牧师自豪地称这辆车为“穷人的铃铛”。

TR1 PE: 约翰逊神父的儿子、圣马丁教堂的继承人大卫·约翰逊牧师自豪地称这个钟琴为“穷人的钟”。

TR2 PE: 约翰逊神父的儿子、圣马丁教堂的继承人大卫·约翰逊牧师自豪地称之为“穷人的大钟”。

ST1 PE: 大卫·约翰逊牧师是约翰逊神父的儿子，也是圣马丁教堂的继承人，他自豪地称这架钟琴为“穷人的铃铛”。

ST2 PE: 约翰逊神父的儿子，同时也是圣马丁教堂的继承人大卫·约翰逊牧师，就自豪地称这首钟乐为“穷人之乐”。

4 结论

机器翻译与人工译后编辑相结合的工作模式越来越普遍，受到的关注也越来越多。机器翻译译后编辑比传统人工翻译的主要优势在于其潜在的效率提升，提高翻译速度，节省翻译成本。机器翻译质量影响译后编辑，在译后编辑任务中，机器翻译质量好坏最终需要译后编辑人员为之买单，体现在译后编辑效率和努力上，因此需要加强用户角度的研究。

本文通过神经机器翻译译后编辑实验，考察机器翻译质量和原文特征对译后编辑

努力的影响，同时比较职业译员和学生译员在译后编辑过程中的异同。研究发现，机器翻译质量影响译后编辑时间努力和技术努力，这对职业译员和学生译员同样适用。原文句长不影响时间努力。职业译员译后编辑效率高于学生译员，对机器译文进行的译后编辑操作少于学生译员。从用户角度对神经机器翻译译后编辑进行的实证研究，有助于加强对译后编辑过程的了解，推动机器翻译朝更加适合译后编辑人员需求的方向发展。

参考文献:

- [1] JUNCZYS-DOWMUNT M, DWOJAK T, HOANG H. Is neural machine translation ready for deployment? A case study on 30 translation directions [C] // Proceedings of the Ninth International Workshop on Spoken Language Translation. Seattle: IWSLT, 2016.
- [2] ISO 18587. Translation Services - Post-editing of machine translation output - Requirements. Geneva: International Organization for Standardization, 2017.
- [3] GUERBEROF A. Productivity and quality in MT post-editing [C] // MT Summit XII-Workshop: Beyond Translation Memories: New Tools for Translators MT. Beijing: MT Summit, 2009: 137-144.
- [4] PLITT M, MASSELOT F. A productivity test of statistical machine translation post-editing in a typical localization context [J]. The Prague Bulletin of Mathematical Linguistics, 2010, 93(1): 7-16.
- [5] DE SOUSA S, AZIZ W, SPECIA L. Assessing the post-editing effort for automatic and semi-automatic translations of DVD subtitles [C] // Proceedings of Recent Advances in Natural Language Processing. Hissar: RANLP, 2011: 97-103.
- [6] GUERBEROF A. Correlations between productivity and quality when post-editing in a professional context [J]. Machine Translation, 2014, 28: 165-186.
- [7] PIELMEIER H & LOMMEL A. Machine translation use at LSPs: data on how Language Service Providers use MT [R]. Cambridge: CSA Research, 2019.
- [8] 王华树,陈涅奥.中国语言服务企业机器翻译与译后编辑应用调查研究[J].北京第二外国语学院学报, 2021,283(5):23-37.
- [9] PEREIRA A. Determining translators' perception, productivity and post-editing effort when using SMT and NMT system [C] // Proceedings of the 21st Annual Conference of the European Association for Machine Translation. Alacant: EAMT. 2018:327-338.
- [10] LAUBLI S, AMRHEIN C, DUGGELIN P, et al. Post-editing productivity with neural machine translation: an empirical assessment

- of speed and quality in the banking and finance domain [C]// Proceedings of MT summit XVII. Dublin: MT-Summit XVII. 2019: 267-273.
- [11] ZOUHAR V, TAMCHYNA A, POPEL M, et al. Neural machine translation quality and post-editing performance [C]// Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Punta Cana: EMNLP. 2021: 10204-10214.
- [12] 卢植,孙娟. 人工翻译和译后编辑中认知加工的眼动实验研究[J]. 外语教学与研究, 2018,50(5):760-769.
- [13] 王湘玲,李婷婷. 人工翻译与机器翻译译后编辑对比实证研究[J]. 外国语言与文化, 2019,3(4):83-93.
- [14] 王湘玲,赖思,贾艳芳. 人工翻译与神经网络机器翻译译后编辑比较研究[J]. 外语教学理论与实践, 2021,(4):115-126.
- [15] KRINGS H P. Repairing texts: empirical investigations of machine translation post-editing process. Kent: Kent State University Press, 2001.
- [16] FONTES L. Evaluating machine translation: Preliminary findings from the first DGT-wide translators' survey [J]. Languages and Translation, 2013,(6):10-11.
- [17] TATSUMI M. Correlation between automatic evaluation metric scores, post-editing speed, and some other factors [C]// Proceedings of The Twelfth Machine Translation Summit. Ontario: MT-Summit XII, 2009:332-339.
- [18] KOEHN P, GERMANN U. The impact of machine translation quality on human post-editing [C]// Proceedings of Workshop on Humans and Computer-assisted Translation. Gothenburg: HaCaT. 2014:38-46.
- [19] SANCHEZ-TORRON M, KOEHN P. Machine translation quality and post-editor productivity [C]// Proceedings of AMTA. Austin: AMTA. 2016:16-27.
- [20] DAEMS J, VANDEPITTE S, HARTSUIKER R J, et al. Identifying the machine translation error types with the greatest impact on post-editing effort [J]. Frontiers in Psychology, 2017,(8):1282-1296.
- [21] STASIMIOTI M, SOSONI V, MOURATIDIS D, et al. Machine translation quality: A comparative evaluation of SMT, NMT and tailored-NMT outputs [C]// 22nd Annual Conference of the European Association for Machine Translation. Lisboa: EAMT. 2020.
- [22] TATSUMI M, ROTURIER J. Source text characteristics and technical and temporal post-editing effort: What is their relationship? [C]// Proceedings of the Second Joint EM +/CNGL Workshop "Bringing MT to the User: Research on Integrating MT in the Translation Industry". Denver: JEC. 2010: 43-51.
- [23] KOPONEN M. Comparing human perceptions of post-editing effort with post-editing operations [C]// Proceedings of the 7th Workshop on Statistical Machine Translation. Montreal: WMT. 2012:181-190.
- [24] VIEIRA L N, ALONSO E, BYWOOD L. Introduction: post-editing in practice – process, product and networks [J]. The Journal of Specialised Translation. 2019, (31):2-13.
- [25] KOPONEN M, AZIZ W, RAMOS L, et al. Post-editing time as a measure of cognitive effort [C]// Workshop on Post-Editing Technology and Practice. San Diego:AMTA,2012.
- [26] POPOVIC M, LOMMEL A, BURCHARDT A, et al. Relations between different types of post-editing operations, cognitive effort and temporal effort [C]// Proceedings of the 17th Annual Conference of the European Association for Machine Translation. Dubrovnik:EAMT,2014: 191-198.
- [27] ZARETSKAYA A, VELA M, PASTOR G C, et al. Measuring post-editing time and effort for different types of machine translation errors [J]. New Voices in Translation Studies, 2016,15(15):63-92.
- [28] BARRAULT L, BOJAR O, COSTA-JUSSA M R, et al. Findings of the 2019 Conference on Machine Translation (WMT19) [C]// Proceedings of the Fourth Conference on Machine Translation . Florence: WMT, 2019: 1-61.
- [29] CARL M. Translog-II: a program for recording user activity data for empirical reading and writing research [C]// Language Resources and Evaluation. ELRA, 2012: 4108-4112.
- [30] SNOVER M, MADANANI N, DORR B, et al. TER-Plus: paraphrase, semantic, and alignment enhancements to Translation Edit Rate [J]. Machine Translation, 2009,(23):117-127.
- [31] SNOVER M, DORR B, SCHWARTZ R, et al. A Study of Translation Edit Rate with Targeted Human Annotation [C]// Proceedings of the 7th Conference of the Association for Machine Translation in the Americas. Cambridge: AMTA, 2006: 223-231.
- [32] SOMMERS N. Revision strategies of student writers and experienced adult writer [J]. Coll. Compos. Commun. 31:378-388.
- [33] HAYES J, FLOWER L, SCRIVER K, et al. Cognitive processing in revision [J]. Advances in Applied Psycholinguistics: Reading, Writing, and Language Processes.1987, (2): 176-240.
- [34] TIRKKONEN-CONDIT S. Professional vs. non-professional translation: a think-aloud protocol study [C]// Learning, Keeping and Using Language: Selected Papers from the Eighth World Congress of Applied Linguistics. Amsterdam: John Benjamins, 1991:381-394.
- [35] DRAGSTED B. Coordination of reading and writing processes in translation: an eye on uncharted territory [J]. Starch Starke, 2010.
- [36] DE Almeida G, O'BRIEN S. Analysing

- post-editing performance: correlations with years of translation experience [C]// Proceedings of the 14th Annual Conference of the European Association for Machine Translation. Saint Raphael: EAMT, 2010.
- [37] MOORKENS J, O'BRIEN S. Post-editing evaluation: trade-offs between novice and professional participants [C]// Proceedings of the 18 Annual Conference of the European Association for Machine Translation. Antalya: EAMT, 2015.
- [38] TATSUMI M. Post-editing machine translated text in a commercial setting: observation and statistical analysis [D]. Dublin City University, 2010.

An Empirical Study on Neural Machine Translation Quality and Post-editing Process

QIU Bailian^{1,2}, WANG Mingwen^{1*}, LUO Qi³, LI Maoxi¹

(1. School of Computer Information Engineering, Jiangxi Normal University, Nanchang 330022, China;

2. School of Foreign Languages, East China Jiaotong University, Nanchang 330013, China;

3. Library, Jiangxi Normal University, Nanchang 330022, China)

Abstract: With the quality improvement of neural machine translation, machine translation post-editing (MTPE) has become more and more popular in translation industry. This paper provides a process research on neural machine translation post-editing, with the aim of exploring the impact of machine translation quality and source text characteristics on post-editing effort, and at the same time comparing the post-editing processes of professional translators and students. We find that machine translation quality influences temporal effort and technical effort, while source sentence length has no impact on temporal effort. Professional translators have faster speed and less post-editing operations than students. The empirical study on Chinese English language pair neural machine translation post-editing is helpful for a user-oriented machine translation research and a deeper understanding of post-editing process, and provides useful reference for post-editor training.

Keywords: neural machine translation; post-editing; post-editing effort; process research