

# 机器翻译译后编辑难度测量体系构建研究<sup>①</sup>

贾艳芳 湖南师范大学 孙三军 北京外国语大学

**摘要:** 了解机器翻译译后编辑难度的影响因素及测量方法,对于提升机器翻译质量、优化其人机交互工具界面和功能、提供课堂教学和测试难度分级依据均有重要作用。本文提出了译后编辑认知加工过程及难度测量的理论和方法模型,分析了影响译后编辑难度的三大因素:材料因素(即机器翻译及原文)、译者个体因素、环境因素(即译后编辑工具),并从机器翻译质量评估、原文特征分析、译者个体因素、译后编辑工具可用性评测及认知努力测量等方面,论述了译后编辑难度测量及存在的挑战。

**关键词:** 机器翻译;译后编辑;任务难度;认知努力;人机交互

**中图分类号:** H315 **文献标识码:** A **文章编号:** 1672-9382(2022)03-0016-09  
DOI:10.13564/j.cnki.issn.1672-9382.2022.03.001

## 1 引言

机器翻译在经历了基于规则、实例和统计的实现手段后,又迎来了基于神经网络的机器翻译(简称神经机器翻译),其质量在许多语言对上优于基于统计的机器翻译,尤其表现在译文流利度上(Jia et al., 2019a)。虽然机器翻译中的错误依旧不可避免,但辅以不同程度的人工译后编辑来修改机译错误,可大大提升翻译效率,因此译后编辑模式被语言服务行业广泛采用,成为当下翻译市场应用最多的主流翻译模式(TAUS, 2019)。

从认知角度讲,译后编辑是一项人机交互特征明显的新型翻译模式(O'Brien, 2021),其认知加工过程较传统人工翻译更为复杂,又有别于基于翻译记忆模糊匹配的计算机辅助翻译。译者需要同时与原文、机器翻译、译后编辑工具交互协作,以实现最佳的认知成本效益。原文和机器翻译译文两个输入源的特征都将影响译后编辑的任务难度;此外,译后编辑必须在特定的工具中进行,工具的可用性也会直接影响任务难度。

了解机器翻译译后编辑难度的影响因素及

测量方法,对于提升机器翻译质量、优化人机交互、方便课堂教学和测试难度分级均有重要作用。然而,译后编辑作为一种较新的翻译模式,其难度测量又涉及机器翻译、翻译学、认知科学、工效学等跨学科专业领域,目前相关研究仍较为匮乏。本文尝试构建译后编辑认知加工过程和难度测量框架,从理论与方法两个角度对译后编辑难度的测量进行探讨。

## 2 译后编辑认知加工过程及难度测量研究模型

任务难度指“识别问题解决方案所需的认知负荷或心理努力程度”(Gallupe et al., 1988: 280),它主要针对特定的任务和任务执行者。衡量任务难度关系到执行者完成任务的难易程度(Dahl, 2004: 39),其本质上是主观的,与个体感知直接相关。

认知负荷(cognitive load)和心理负荷(mental workload)是认知科学的核心概念,两者内涵相近,但源于不同的研究领域。这两个概念在翻译及译后编辑过程研究中常互换使

作者简介:贾艳芳,湖南师范大学翻译系讲师,研究方向:人机交互翻译、翻译认知过程、计算机辅助翻译, E-mail: yanfang.jia@hunnu.edu.cn。  
孙三军,北京外国语大学英语学院副教授,研究方向:翻译认知过程、实证翻译研究、翻译技术, E-mail: sunsanjun@bfsu.edu.cn。

FLC May 2022 Vol.19 No.3 (General Serial No.107)



用。认知负荷最早出现于心理学领域，影响较大的是Sweller et al. (1998) 提出的认知负荷理论，该理论认为，任务所需的总体认知负荷由材料因素和个体因素引发的内在认知负荷，以及任务呈现方式和环境等引发的外在认知负荷需求共同决定。心理负荷是人机交互研究领域的重要概念，人机交互是一门研究机器或计算机化的系统与软件与用户之间交互关系的学科 (Johnson, 1992), Meshkati (1988) 提出了影响人机交互任务执行者心理负荷的多维模型，概括了影响心理负荷的众多因素、各因素之间的关系，以及测量心理负荷的手段。本研究中，认知负荷指译后编辑任务对译者提出的认知资源需求，而认知努力 (cognitive effort) 指译者在执行任务时实际投入的认知资源。

本研究基于认知负荷理论和心理负荷多维模型，构建了译后编辑认知加工过程及难度测量模型 (见图1)。译后编辑认知过程包含五类活动：(1) 原文阅读理解；(2) 机器翻译阅读理解；(3) 机器翻译质量评估；(4) 机器翻译错误修订；(5) 原文人工翻译。译者依据原文评估识别机器翻译错误，决定修改机器翻译错误或重新人工翻译原文句段，这一决策过程受到原文特征和机器翻译质量共同影响。译者阅读原文或机器翻译的优先顺序，可能受到机器翻译质量或译者个体习惯差异影响 (Guerberof Arenas, 2013; Carl et al., 2015)。在此过程中，译者仍是译后编辑任务

的主导者，但必须和机器翻译译文及译后编辑工具交互合作，才能最好地平衡质量水平要求和认知努力付出。下文将对译后编辑认知加工过程及任务难度的各项影响因素和测量方法展开分析讨论。

### 3 译后编辑难度影响因素及其测量方法

影响译后编辑认知加工过程及任务整体难度的因素包括三类：(1) 材料因素，即机器翻译和原文；(2) 译者个体因素；(3) 译后编辑工具。认知心理学研究表明，人在从事复杂活动时的工作记忆有限，译者在进行译后编辑任务时，对各因素进行认知加工均会占用其有限的工作记忆和认知资源。因此，这四因素相互作用，共同决定译后编辑任务的难度及译者的认知努力付出。认知努力测量指标主要包括译者执行译后编辑任务时的各项生理指标、行为指标及最终译文质量。四项影响因素均为可控制和可测量项，检验其中一项或多项因素对译后编辑难度的影响，须在有效测控其他影响因素的前提下进行。

#### 3.1 机器翻译质量

##### 3.1.1 机器翻译质量评估方法

机器翻译质量通常被视为译后编辑任务难度的决定性因素。机器翻译质量评估方法可分为人工评估和自动评估两种。人工评估方法包括忠实度/流利度评分、等级

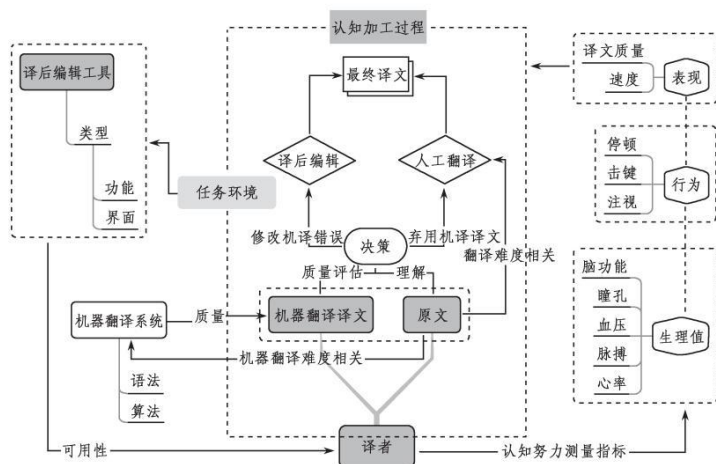


图1 译后编辑认知加工过程及难度测量模型

排序 (ranking)、错误分类,如翻译自动化用户协会 (the Translation Automation User Society, TAUS) 的动态质量评估框架 (DQF) 及目前采用最广的翻译质量多维评估模型 (Multidimensional Quality Metrics) (Lommel, 2018); 自动评估通过对比机译结果与标准人工参考译文或译后编辑结果之间的相似度为机译结果量化打分,常用方法包括BLEU、NIST、GTM、METEOR、TER、HTER等。两者相比,人工评估耗时耗力但更为可靠,自动评估方法省时省力且相对稳定,被机器翻译领域广为采用,但严重依赖参考译文,所以可靠性较差,且与人工评估之间的相关性低,仍有待继续完善。

在现有的全自动机器翻译模式下,译者无法实时干预机译结果,需要对所有机器给定译文进行译后编辑,仍处于“人适应机器”阶段;探索研发中的交互式机器翻译模式(如TranSmart),其发展朝向“机器适应人”的人机融合方向,目标是在译者与机译系统交互时,系统会根据译者的实时反馈,产生新的翻译假设供译者选择,避免译者反复修改相同的机器翻译错误,并实时更新优化译文选项。新变化将对机译质量评估带来全新挑战。

### 3.1.2 机器翻译质量与译后编辑难度

Krings (2001) 对基于规则的机译句子质量进行了人工评估,发现句子机译质量越高,则译后编辑耗时越短;但机译质量与译后编辑认知努力之间的关系并非线性的,译后编辑中等质量机译句子所花费的认知努力很多时候反而最高,因为译者的注意力在原文、机译译文和译文上的分散程度更高。Krings还发现,机译质量高低对原文加工及认知努力分配有一定影响,但该研究并没有测量原文句子复杂度,所以无法衡量原文特征本身在此过程中的作用。

机器翻译的错误数量和类型均与译后编辑努力的付出相关。Daems et al. (2017) 发现机译错误总量与总注视次数、产出单位 (production unit) 数量呈负相关,与平均停顿率 (average pause ratio) 呈正相关,但与平均注视时长无相关性;不同的错误类型对不同的译后编辑努力指标的影响不相同,如连贯性错误对任务完成时间的影响较大,而意义迁移与注视时长的相关性较强。此外,机器翻译质量自动评估结果与译后编辑努力呈负相关。低分的GTM、TER、BLEU及METEOR机译译文,在译后编辑时需要消耗的任务时间及总注视时间更长,总注视次数也更多,表明更高的

认知努力付出(如O'Brien, 2011; Gaspari et al., 2014); 基于高分的METEOR机译结果译后编辑,消耗的平均注视时间和注视次数较低 (Vieira, 2016)。上述研究均将机器翻译质量作为影响译后编辑过程的主要因素,对应的原文特征很少被纳入考虑范围。然而,在与不同复杂度的原文配对时,具有相同错误或自动评分水平相当的机器翻译结果可能损耗不同程度的认知努力。

此外,机器翻译技术几经迭代更新,现有研究主要以基于统计的机器翻译为研究对象,神经机器翻译系统基于神经网络模型的翻译方法,虽然极大提高了译文流畅度,但其错误类型却更加难以预测,这类错误隐藏在流畅的行文中,反而加大了译后编辑时识别和更正机译错误的难度 (Moorkens, 2018; Yamada, 2019)。而交互式机器翻译具有人机融合的特点,其译文及质量随着与译者的交互呈现动态变化,这对译后编辑过程也会产生影响,因此亟须加大对最新机器翻译方法的研究。

## 3.2 原文特征及其测量

### 3.2.1 原文与机器翻译难度相关特征

在译后编辑模式下,原文发挥的作用有异于人工翻译。原文特征会在一定程度上影响机器翻译质量,早期基于规则及实例的机器翻译质量受原文特征的影响尤为明显。原文中容易引发机器翻译错误的特征被称作消极翻译特征(如多义词、省略、动名词、被动语态等) (Underwood & Jongejan, 2001)。O'Brien (2004, 2006) 发现,消极翻译特征较多的原文在译后编辑时需要更多的认知努力,如任务时长更长及编辑量更多;不同消极翻译特征对译后编辑认知努力的影响程度不同,如较长的名词短语及动名词对译后编辑努力的影响较大,而缩写、专有名词以及标点符号等的影响则较小。然而,此类针对原文特征对译后编辑认知努力影响的研究,普遍没有控制或测量相应的机器翻译质量水平。所以,原文的某些特征对译后编辑认知努力的影响,可能是由这些原文特征造成的较低质量的机器翻译或两者交互作用共同造成的,无法明确归因于原文特征。

了解原文消极翻译特征的影响有助于制定受控语言 (controlled language) 写作规范,实施译前编辑,从而消除这些因素,提高机器翻译质量,最终降低译后编辑难度。随着机器翻译技术的革新,原文对机器翻译质量的影响

减弱,而基于统计的机器翻译及神经机器翻译,其质量更多受到训练机译系统的平行语料库的质量及算法的影响,受原文特征的影响程度较小(Jia et al., 2019b)。

### 3.2.2 原文与翻译难度相关特征

译者需要依据原文来识别机器翻译错误并加以修改,对于错误较多的句段,译者可能弃用机译译文,对原文进行不同程度的人工翻译。此时,原文与翻译难度相关特征将影响其最终的认知努力付出。

原文翻译难度测量可采用可读性公式、高低频词占比率、句法复杂度、非直译度(non-literalness)、文本语义特征及专家评分等方法。可读性公式是文本阅读难度测量最常用的方法,一般包含1—3个可以量化统计的文本特征(如词汇难度与句长),可以据此对文本的阅读难度给出客观评分,常用的可读性公式包括Flesch Reading Ease、Dale-Chall、SMOG、Flesch-Kincaid Readability等。高低频词占比可以测量原文中高低频词所占的比例,通常一篇文章所含的低频词越多,阅读难度越高。句法复杂度测量文本句子内部的连贯性及相邻句子之间的结构相似性,句子之间的连贯性及相似度越高,阅读难度就越低,可以利用Coh-Metrix对句子结构相似度进行打分(McNamara et al., 2014)。

除阅读难度外,某些文本特征容易引发翻译相关难度,如可直译度。可直译度指文中包含习语、隐喻、转喻等非直译元素的数量,非直译元素越多,文本翻译难度越大。可直译度最早由Jensen(2009)应用到翻译过程研究中,作为原文复杂度的筛选标准之一。此外,文本的语义特征,如“词的多义性”和“词语在译入语中对等词的数量”,也是影响可直译度的潜在因素(孙三军、文军,2015),但目前暂无有效工具对汉语词义进行量化统计。WordNet词义统计工具以英语为主,暂未开发其他语言的词义统计功能。除以上几种评价标准外,专家对原文阅读及翻译难度的主观评估也是原文难度评估的常用手段。

原文翻译难度测量在实践中多综合采纳以上多种测量方法。Jensen(2009)最早使用可读性公式、高低频词占比率、可直译度及专家评分作为原文难度分层筛选的标准,是目前翻译过程研究中原文复杂度筛选最常引用的指标模板之一(如Liu et al., 2019)。Sun(2015)构架了翻译难度研究框架。

译后编辑研究对原文的关注则较少,译后编辑过程研究也极少测控原文特征。然而,

贾艳芳(2020)对于神经机器翻译任务难度的研究显示,原文复杂度和机器翻译质量存在显著的交互效应,会影响译后编辑难度。原文复杂度对其译后编辑难度的影响程度取决于机器翻译质量,对低质量神经机器翻译来说,原文复杂度的影响与传统人工翻译相似,复杂度越高,译后编辑难度越大;而对高质量神经机器翻译来说,原文复杂度对译后编辑的影响则不显著。译者访谈表明,该现象可能由于原文需要被理解加工的程度取决于机器翻译的质量,对于低质量的机器翻译,译者需要充分理解原文才能修改机译中的大量错误或重新翻译原文,而对高质量机器翻译进行译后编辑,译者无须深入加工原文,原文主要起到参考作用。然而,该研究只聚焦于原文的整体复杂度,原文具体特征对译后编辑难度的影响仍有待进一步探究。

### 3.3 译后编辑工具可用性及其测量

现有的译后编辑任务普遍在融合机器翻译功能的机辅翻译软件(如Trados和memoQ)、线上机辅翻译平台(如YICAT和Memsources)或独立的机器翻译软件(如云译通)中进行。译后编辑作为典型的人机交互模式,其工具功能及界面设计的可用性直接影响译后编辑的外在认知负荷需求,不符合译者的认知需求及习惯的功能或界面,会增加认知摩擦,加大任务难度。可用性评估是工效学研究的重要领域,基本方法包括用户测试、启发式评价(heuristic evaluation)、认知 walkthrough(cognitive walkthrough)、行为分析(action analysis)、访谈、问卷、GOMS模型,以及基于眼动和脑电技术的认知生理评估方法等。

现有的机辅翻译软件普遍融合机器翻译功能,可以同时进行机辅翻译和机器翻译译后编辑。O'Brien & Moorkens(2014)和Moorkens & O'Brien(2017)等研究均表明,职业译者对机辅翻译软件功能及界面以及其译后编辑功能普遍抱有强烈的不满,主要针对对界面布局、交互功能、文本切分方式、视觉效果、软件故障等,即便在机辅翻译系统功能不错的前提下,译者普遍对其译后编辑功能也会有所不满。由于“翻译记忆和机器翻译技术到目前为止被认为是两种明显不同的技术”(ibid.: 109),机辅翻译软件并没有全方位考虑译后编辑任务的特殊性。独立的译后编辑工具研发也得到了—定的关注,如TransType(Langlais et al., 2002)、iOmegaT(Moran & Lewis, 2011)、PET(Aziz et al., 2012)、



Matecat (Bertoldi et al., 2013)、Casmacat (Koehn et al., 2015) 等, 但此类软件开发同样主要考虑软件功能本身, 缺乏对译者人为因素的考虑。驱动翻译技术发展的首要因素是功能需要, 能否很好地满足译者的生理、心理及认知需求往往是开发者的次级考量因素, 软件很少在设计之初就以译者为中心, 全面考量人类工效学因素, 译者反馈往往在软件设计完成以后才会得以收集, 而这时软件可以调整的余地已经很小了。

提升工具可用性是提高译后编辑效率和用户满意度的重要途径。通过融合可用性评估和翻译过程研究方法, 观察和测量分析译者在真实工作环境下应用翻译技术进行翻译时的认知加工过程, 对现有软件或正在开发软件的可用性进行分析, 可以为翻译技术开发提供宝贵的参考。Kappus & Ehrensberger-Dow (2020) 通过分析任务时间、眼动热点图、击键次数、鼠标活动、译文质量评估及访谈等, 发现功能相似的机辅翻译软件Trados和Lilt的界面布局呈现方式差异较大, 而译者更喜欢符合直觉判断的功能布局, 功能繁杂的界面在一定程度上会增加译者的认知负荷。随着交互式机器翻译的研发 (Peris & Casacuberta, 2019), 机器翻译朝着“机器适应人”的方向发展, 要更好地实现人机融合, 开发符合译后编辑自身任务特点的工具的需求更为迫切。

Ragni & Vieira (2022) 通过梳理目前针对神经机器翻译的研究发现, 绝大多数涉及译者的机器翻译研究, 其研究目的均为改进机器翻译技术本身, 而不是如何让机器翻译作为工具更好地服务译者。因此, 除了针对提升机器翻译技术本身之外, 亟需学界和业界加大以下方面研究, 包括: 如何让机器翻译技术及译后编辑工具设计更有效地协助译者完成译后编辑任务, 揭示译者执行译后编辑任务时的认知需求和局限, 评估现有工具人机界面及功能设计的实用性及易用性, 最终改进不符合译者认知需求或认知工效原理的痛点设计。

### 3.4 译者因素及测量

除了客观输入材料和工具因素外, 译者对译后编辑任务难度的感知与其自身的认知能力、情感态度因素及译后编辑相关能力等也密切相关。认知能力通常可以通过受试的工作记忆力容量测得, 具体包括操作广度、阅读广度及记忆广度等测验 (Conway et al., 2005)。然而, Vieira (2016) 发现工作记忆力对译后编辑认知努力并没有显著影响, 认为工作记忆

是译后编辑任务的次级影响因素, 与其他因素 (如机器翻译质量、原文熟悉程度及态度等) 交互影响译后编辑的认知努力付出。

作为一种人机交互翻译模式, 译后编辑能力与传统翻译能力存在交叉, 但并不完全相同。O'Brien (2002) 认为成功的译后编辑者需要熟练掌握原语和目的语、专业领域知识、文本类型与对比知识, 能运用工具, 还需要掌握机器翻译知识, 具备术语管理能力, 可以执行译前编辑, 进行受控语言写作, 开展基本编程, 使用宏命令, 为机器词典编码, 并对机器翻译抱有积极态度, 可以容忍低质量机译结果等。Rico & Torrejón (2012) 将译后编辑能力划分为三大范畴, 即语言能力、工具能力与核心能力。冯全功、刘明 (2018) 构建了译后编辑能力的认知、知识及技术三大维度。Nitzke et al. (2019) 认为译后编辑能力应包含双语能力、修订能力、翻译能力、译后编辑能力、机器翻译能力、研究能力、工具能力、语言外能力以及风险评估能力、决策能力、咨询能力及服务能力。

虽然对译后编辑能力到底包含哪些具体内容并无定论, 但其内涵明显更为多元, 除传统的语言能力及翻译能力外, 还包含了工具能力、机器翻译相关知识掌握、技术使用及态度等范畴, 充分体现了其人机交互的特性。虽然行业认为译后编辑比人工翻译更为简单, 其定价比人工翻译更低, 但调查显示职业译者强烈排斥译后编辑任务 (Bundgaard, 2017; Läubli & Orrego-Carmona, 2017; Cadwell et al., 2016), 认为译后编辑增加了任务中的不确定因素。一方面, 机器翻译的质量并不稳定, 另一方面, 译者需要不断学习迭代更新的技术工具, 因此许多译者认为译后编辑是一项更为复杂的认知活动。相比之下, 学生译者对机器翻译及译后编辑的态度更为积极 (Yamada, 2015; Jia et al., 2019b)。

究其原因, 对于职业译者和新手译者来说, 译后编辑都是一项新型翻译模式, 职业译者并不等同于职业译后编辑者, 译者译后编辑相关能力的整体提高, 将降低其任务难度感知, 减少其译后编辑需要付出的认知努力, 但专精知识的获得需要多年的刻意训练 (Ericsson & Smith, 1991)。Moorkens et al. (2015) 发现职业译后编辑者的译后编辑速度明显快于新手译后编辑者, 而de Almeida (2013) 和Guerberof Arenas (2014) 认为译者的译后编辑经验与译后编辑速度没有显著关系。由于译后编辑能力没有统一的测量标准,



前人的研究对译后编辑能力的筛选标准并不一致,多根据译后编辑经验年限、是否会使用工具等简单标准进行筛选,因此仍需要进一步研究细化统一译后编辑能力的测量标准,分析其对译者任务难度感知的影响。

#### 4 译后编辑难度测量方法及挑战

测量译后编辑难度的方法,可分为内省法(主观评测、回溯性报告等)、行为观测法(编辑行为、停顿行为、注视、任务时长等)、生理测量法(脑功能反应、瞳孔反应、皮肤反应、血压、心率等)(Sun, 2019; 谭业升, 2020)。由于认知负荷的多维度性,不同指标从不同侧面反映认知负荷的影响,测量结果可能存在差异。

Vieira (2014) 和 Cumbreño & Aranberri (2021) 等研究对比了眼动、停顿、编辑时间和主观评分四类认知努力指标之间的相关性,发现虽然所有指标之间都存一定程度的相关性,但编辑时间、停顿单词比(停顿次数与原文词数比)、注视次数单词比之间存在较强的相关性且信度较高,而平均注视时长、停顿率及主观评分之间的相关性弱且信度低;具有译后编辑经验的译者对认知负荷的主观评分与其他指标的相关性较高,新手译者则较弱。当用不同测量方法测出的结果相互矛盾时, Jex (1988) 认为应以个体的主观感知评价为准。Herbig et al. (2019) 则尝试建立译后编辑难度的多模态指标评估模型,发现在所有标准中与主观评分相关度最强的是任务时长和眨眼频率,中等相关的是 TER 得分、平均注视量(average fixation amount)、平均停顿率、皮肤电反应,而其余指标呈弱相关。该研究发现,多模态指标认知负荷测量模型的有效性显著高于单独基于时间或文本特征的测量模型。

在行业方面,译后编辑定价主要依靠译后编辑效率作为其任务难度及定价的主要依据,效率主要通过任务时间或编辑量来衡量。然而任务时间并不能全面反映译者认知努力的付出,一方面,单独使用译后编辑时间测量认知努力的有效性不高,另一方面,译后编辑的耗时不止包括修改机译错误本身的任务时间,还可能涉及原文预处理、解决软件技术问题以及项目管理等时间成本。而编辑距离(如 HTER)一方面只能体现最终译后编辑结果和最初机器翻译之间的编辑量,而非具体过程的实际操作修改量,如反复修改过程等,另一方面,编辑量是认知加工过程后的最终技术操

作结果,与认知加工过程本身存在一定差别(Vieira, 2016; 贾艳芳, 2020)。

测评手段优化可以综合采用主观评测、行为及生理测量等多模态数据指标,多方法多元互证,测量译后编辑过程中的认知努力付出,避免单一方法的片面和不足。然而,科学测量译后编辑认知努力还需要建立在对上述影响因素的合理测控上。译后编辑认知加工过程受多元因素影响,但大多数现有研究未能考虑此特性,把原文和机器翻译两大材料因素对译后编辑认知加工效果的影响割裂开,往往在未测控原文或译文一方的情况下,研究另一方的作用(如 Aziz et al., 2014; Daems et al., 2017)。进一步研究需要评测不同层级原文的复杂度、质量优劣分层的机器翻译及两者的交互效应,及其对译后编辑难度的影响,由此全面了解材料内因的作用。

此外,译后编辑研究往往滞后于翻译技术革新,需要进一步对接机器翻译技术发展,在现有的大量基于统计的机器翻译研究成果的基础上,不断深化对神经机器翻译的研究,并开展对交互式机器翻译译后编辑的研究,分析不断革新的机器翻译方法对译后编辑模式认知加工机制的影响,促进符合工效学原理的交互式机器翻译系统和工具的开发,保证译后编辑难度研究的应用价值最大化。

#### 5 结语

本文运用人机交互领域的理论及相关研究成果,构建了译后编辑认知加工和难度测量模型,分析了译后编辑难度的影响因素、影响因素的测量及认知努力的测量方法。尽管译后编辑已成为语言服务行业的主流翻译模式之一,但迄今为止有关译后编辑难度测量的实证研究却并不多见。译后编辑难度测量涉及机器翻译、翻译学、认知科学、工效学等跨学科领域,且涉及多元复杂的影响因素,亟需大量的实证研究分析解决以下问题,主要包括:(1)综合评测多元因素的影响,分析原文和机器翻译两大材料内因对译后编辑难度的影响;(2)评测现有各类译后编辑工具的界面及功能可用性对译后编辑难度的影响;(3)分析译者在译后编辑时的认知需求、认知能力和局限,以及不同材料内因及工具外因的变化对其认知加工机制的影响,包括认知努力、压力、决策、情感及人机交互等过程;(4)优化测评手段,建立评测译后编辑认知努力的最佳测量模型。基于对上述问题的研究,最终建立译后编辑各影响因素交



互效应的量化模型,科学预测译后编辑难度。

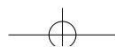
译后编辑难度研究与译后编辑过程中的问题解决、策略与技巧、译后编辑能力等翻译学中的诸多问题密切相关,可以拓展人工智能时代背景下认知翻译学的研究疆域。深化对译后编辑难度影响因素及测量方式的认知,可以对译后编辑任务失败科学归因,对其任务难度计算进行建模科学预测,优化译后编辑工具的人机交互界面和功能,最终全面提升翻译效率及用户满意度,并为译后编辑市场定价及培训测试等提供科学依据。□

#### 注释

① 本文系湖南省社科基金项目“信息技术背景下人机交互翻译模式任务难度研究”(编号:20YBQ068)的部分研究成果。

#### 参考文献

- [1] Aziz, W., S. Castilho & L. Specia. Pet: A tool for post-editing and assessing machine translation[A]. In *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012*[C]. Istanbul: European Language Resources Association (ELRA), 2012: 3982-3987.
- [2] Aziz, W., M. Koponen & L. Specia. Sub-sentence level analysis of machine translation post-editing effort[A]. In S. O'Brien, L. W. Balling, M. Carl, M. Simard & L. Specia (eds). *Post-Editing of Machine Translation: Processes and Applications*[C]. Newcastle upon Tyne: Cambridge Scholars Publishing, 2014: 170-199.
- [3] Bertoldi, N., M. Cettolo & M. Federico. Cache-based online adaptation for machine translation enhanced computer assisted translation[A]. In *Proceedings of Machine Translation Summit XIV: Papers*[C]. Nice, 2013: 35-42.
- [4] Bundgaard, K. Translator attitudes towards translator-computer interaction—Findings from a workplace study[J]. *Hermes*, 2017(56): 125-144.
- [5] Cadwell, P., S. Castilho, S. O'Brien & L. Mitchell. Human factors in machine translation and post-editing among institutional translators[J]. *Translation Spaces*, 2016(2): 222-243.
- [6] Carl, M., S. Gutermuth & S. Hansen-Schirra. Post-editing machine translation: Efficiency, strategies and revision processes in professional translation settings[A]. In A. Ferreira & J. W. Schwieter (eds). *Psycholinguistic and Cognitive Inquiries into Translation and Interpreting*[C]. Amsterdam: John Benjamins, 2015: 145-174.
- [7] Conway, A. R. A., M. J. Kane, M. F. Bunting, D. Z. Hambrick, O. Wilhelm & R. W. Engle. Working memory span tasks: A methodological review and user's guide[J]. *Psychonomic Bulletin & Review*, 2005(5): 769-786.
- [8] Cumbreño, C. & N. Aranberri. What do you say? Comparison of metrics for post-editing effort[A]. In M. Carl (ed). *Explorations in Empirical Translation Process Research*[C]. Switzerland: Springer, 2021: 57-79.
- [9] Daems, J., S. Vandepitte, R. J. Hartsuiker & L. Macken. Identifying the machine translation error types with the greatest impact on post-editing effort[J]. *Frontiers in Psychology*, 2017, 8: 1282.
- [10] Dahl, Ö. *The Growth and Maintenance of Linguistic Complexity*[M]. Amsterdam: John Benjamins, 2004.
- [11] de Almeida, G. Translating the post-editor: An investigation of post-editing changes and correlations with professional experience across two Romance languages[D]. Dublin City University, 2013.
- [12] Ericsson, K. A. & J. Smith (eds). *Toward a General Theory of Expertise: Prospects and Limits*[C]. Cambridge: Cambridge University Press, 1991.
- [13] Gallupe, R. B., G. DeSanctis & G. W. Dickson. Computer-based support for group problem-finding: An experimental investigation[J]. *MIS Quarterly*, 1988(2): 277-296.
- [14] Gaspari, F., A. Toral, S. K. Naskar, D. Groves & A. Way. Perception vs reality: Measuring machine translation post-editing productivity[A]. In *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas*[C]. Vancouver, 2014: 60-72.
- [15] Guerberof Arenas, A. What do professional translators think about post-editing?[J]. *The Journal of Specialised Translation*, 2013(19): 75-95.
- [16] Guerberof Arenas, A. The role of professional experience in post-editing from a quality and productivity perspective[A]. In S. O'Brien, L. W. Balling, M. Carl, M. Simard & L. Specia (eds). *Post-Editing of Machine Translation: Processes and Applications*[C]. Newcastle upon Tyne: Cambridge Scholars Publishing, 2014: 51-76.



- [17] Herbig, N., S. Pal, M. Vela, A. Krüger & J. van Genabith. Multi-modal indicators for estimating perceived cognitive load in post-editing of machine translation[J]. *Machine Translation*, 2019(1-2): 91-115.
- [18] Jensen, K. T. H. Indicators of text complexity[J]. *Copenhagen Studies in Language*, 2009(37): 61-80.
- [19] Jex, H. R. Measuring mental workload: Problems, progress, and promises[A]. In P. A. Hancock & N. Meshkati (eds). *Human Mental Workload*[C]. Amsterdam: North-Holland, 1988: 5-39.
- [20] Jia, Y., M. Carl & X. Wang. Post-editing neural machine translation versus phrase-based machine translation for English-Chinese[J]. *Machine Translation*, 2019a(1-2): 9-29.
- [21] Jia, Y., M. Carl & X. Wang. How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study[J]. *The Journal of Specialised Translation*, 2019b(31): 60-86.
- [22] Johnson, P. *Human Computer Interaction: Psychology, Task Analysis, and Software Engineering*[M]. New York: McGraw-Hill, 1992.
- [23] Kappus, M. & M. Ehrensberger-Dow. The ergonomics of translation tools: Understanding when less is actually more[J]. *The Interpreter and Translator Trainer*, 2020(4): 386-404.
- [24] Koehn, P., V. Alabau, M. Carl, et al. *CASMACAT: Final Public Report*[R]. 2015.
- [25] Krings, H. P. *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*[M]. Kent: The Kent State University Press, 2001.
- [26] Langlais, P., G. Lapalme & M. Loranger. TransType: Development-evaluation cycles to boost translator's productivity[J]. *Machine Translation*, 2002(2): 77-98.
- [27] Lübli, S. & D. Orrego-Carmona. When Google translate is better than some human colleagues, those people are no longer colleagues[A]. In *Proceedings of the 39th Conference Translating and the Computer*[C]. London, 2017: 59-69.
- [28] Liu, Y., B. Zheng & H. Zhou. Measuring the difficulty of text translation: The combination of text-focused and translator-oriented approaches [J]. *Target: International Journal of Translation Studies*, 2019(1): 125-149.
- [29] Lommel, A. Metrics for translation quality assessment: A case for standardising error typologies[A]. In J. Moorkens, S. Castilho, F. Gaspari & S. Doherty (eds). *Translation Quality Assessment: From Principles to Practice*[C]. Berlin: Springer, 2018: 109-127.
- [30] McNamara, D. S., A. C. Graesser, P. M. McCarthy & Z. Cai. *Automated Evaluation of Text and Discourse with Coh-Matrix*[M]. Cambridge: Cambridge University Press, 2014.
- [31] Meshkati, N. Toward development of a cohesive model of workload[A]. In P. A. Hancock & N. Meshkati (eds). *Human Mental Workload*[C]. Amsterdam: North-Holland, 1988: 305-314.
- [32] Moorkens, J. Eye tracking as a measure of cognitive effort for post-editing of machine translation[A]. In C. Walker & F. M. Federici (eds). *Eye Tracking and Multidisciplinary Studies on Translation*[C]. Amsterdam: John Benjamins Publishing Company, 2018: 55-69.
- [33] Moorkens, J. & S. O'Brien. Assessing user interface needs of post-editors of machine translation[A]. In D. Kenny (ed). *Human Issues in Translation Technology*[C]. New York: Routledge, 2017: 109-130.
- [34] Moorkens, J., S. O'Brien, I. A. L. da Silva, N. B. de Lima Fonseca & F. Alves. Correlations of perceived post-editing effort with measurements of actual effort[J]. *Machine Translation*, 2015(3/4): 267-284.
- [35] Moran, J. & D. Lewis. Unobtrusive methods for low-cost manual evaluation of machine translation[J/OL]. *Tralogy 1, Session 5 - Quality in Translation*, 2011. <http://odel.irevues.inist.fr/tralogy/index.php?id=141%26format=print>. Accessed 13/04/2022.
- [36] Nitzke, J., S. Hansen-Schirra & C. Canfora. Risk management and post-editing competence[J]. *The Journal of Specialised Translation*, 2019(31): 239-259.
- [37] O'Brien, S. Teaching post-editing: A proposal for course content[A]. In *Proceedings of the 6th EAMT Workshop, Teaching Machine Translation*[C]. Manchester: European Association for Machine Translation, 2002: 99-106.
- [38] O'Brien, S. Machine translatability and post-editing effort: How do they relate[A]. In *Proceedings of Translating and the Computer*





- 26[C]. London, 2004.
- [39] O'Brien, S. Controlled language and post-editing[J]. *Multilingual*, 2006(7): 17-19.
- [40] O'Brien, S. Towards predicting post-editing productivity[J]. *Machine Translation*, 2011(3): 197-215.
- [41] O'Brien, S. Translation, human-computer interaction and cognition[A]. In F. Alves & A. L. Jakobsen (eds). *The Routledge Handbook of Translation and Cognition*[C]. London: Routledge, 2021: 376-388.
- [42] O'Brien, S. & J. Moorkens. Towards intelligent post-editing interfaces[A]. In W. Baur, B. Eichner, S. Kalina, N. Kessler, F. Mayer & J. Orsted (eds). *Proceedings of the XXth FIT World Congress*[C]. Berlin: BDÜ Fachverlag, 2014: 131-137.
- [43] Peris, Á. & F. Casacuberta. Online learning for effort reduction in interactive neural machine translation[J]. *Computer Speech & Language*, 2019, 58: 98-126.
- [44] Ragni, V. & L. N. Vieira. What has changed with neural machine translation? A critical review of human factors[J]. *Perspectives*, 2022: 137-158.
- [45] Rico, C. & E. Torrejón. Skills and profile of the new role of the translator as MT post-editor[J]. *Revista Tradumàtica: Tecnologies de la Traducció*, 2012(10): 166-178.
- [46] Sun, S. Measuring translation difficulty: Theoretical and methodological considerations[J]. *Across Languages and Cultures*, 2015(1): 29-54.
- [47] Sun, S. Measuring difficulty in translation and post-editing: A review[A]. In D. Li, V. L. C. Lei & Y. He (eds). *Researching Cognitive Processes of Translation*[C]. Singapore: Springer, 2019: 139-168.
- [48] Sweller, J., J. J. G. van Merriënboer & F. G. W. C. Paas. Cognitive architecture and instructional design[J]. *Educational Psychology Review*, 1998(3): 251-296.
- [49] TAUS. Keynotes summer 2019: A review of the TAUS global content conference in Salt Lake City, UT (USA)[R]. Amsterdam: TAUS Signature Editions, 2019.
- [50] Underwood, N. L. & B. Jongejan. Translatability checker: A tool to help decide whether to use MT[A]. In B. Maegaard (ed). *Proceedings of MT Summit VIII: Machine Translation in the Information Age*[C]. Santiago de Compostela, 2001: 363-368.
- [51] Vieira, L. N. Indices of cognitive effort in machine translation post-editing[J]. *Machine Translation*, 2014(3/4): 187-216.
- [52] Vieira, L. N. Cognitive effort in post-editing of machine translation: Evidence from eye movements, subjective ratings, and think-aloud protocols[D]. Newcastle University, 2016.
- [53] Yamada, M. Can college students be post-editors? An investigation into employing language learners in machine translation plus post-editing settings[J]. *Machine Translation*, 2015(1): 49-67.
- [54] Yamada, M. The impact of Google Neural Machine Translation on post-editing by student translators[J]. *The Journal of Specialised Translation*, 2019(31): 87-106.
- [55] 冯全功, 刘明. 译后编辑能力三维模型构建[J]. *外语界*, 2018(3): 55-61.
- [56] 贾艳芳. 神经机器翻译译后编辑与人工翻译: 难度测量实证研究[D]. 湖南大学, 2020.
- [57] 孙三军, 文军. 论翻译难度的测量: 理论与方法[J]. *外语界*, 2015(5): 70-78.
- [58] 谭业升. 翻译认知过程研究[M]. 北京: 外语教学与研究出版社, 2020.

#### Measuring the Difficulty of Machine Translation Post-Editing: A Research Framework

**Abstract:** Identifying the factors impacting the difficulty of machine translation (MT) post-editing and selecting scientific indicators to measure it will contribute to improving MT quality, optimizing the usability of computer-aided translation tools, and providing a theoretical basis for post-editing pedagogy and assessment. This study models the cognitive process of post-editing with its impacting factors and measures of its difficulty. It identifies three groups of factors that influence the difficulty of post-editing, namely, texts (including source text and its machine translation), post-editors, and environments. It then discusses MT quality assessment, source text complexity, post-editor's characteristics, usability of post-editing tools, cognitive effort, and their measurements.

**Keywords:** machine translation; post-editing; task difficulty; cognitive effort, human-machine interaction

