# 基于 MQM 质量评估模型的专业文本机器翻译错误类型实证分析

杨博超，冷冰冰
（上海理工大学 外语学院，上海　200093）

**摘要：**机器译文质量历来是机器翻译的核心问题，作为评价机器译文质量的有效方式，译文错误的归类方法也备受关注。以专业文本的机器翻译错误为研究对象，以 MQM 质量评估模型为分析框架，对目前机器翻译代表引擎 DeepL 产出的译文进行错误归类。实证结果表明，尽管机器翻译能最大限度地实现与原文信息的对等，但在术语不适切、信息不准确、译文不流畅和语法不规范等错误类型上问题突出。该研究可为职业译员培训中的译前编辑和译后编辑提供教学支撑。

**关键词：**机器翻译；MQM 质量评估模型；错误类型
**中图分类号：**H059　　**文献标识码：**A　　**文章编号：**1009－895X(2025)05－0432－07
**DOI:** 10.13256/j.cnki.jusst.sse.240109012

# An Empirical Study of Machine Translation Error Types in Specialized Text Translation: Multidimensional Quality Metrics Perspective

YANG Bochao，LENG Bingbing
（*College of Foreign Languages*, *University of Shanghai for Science and Technology*, *Shanghai 200093*, *China*）

**Abstract:** The quality of machine translation has long been a core issue in machine translation, and categorization of translation error types for quality assessment has attracted much attention. Under the guidance of Multidimensional Quality Metrics (MQM), this paper categorizes translation errors of specialized texts by DeepL, a neural network-based machine translation tool. The findings indicate that, despite optimal equivalence of machine translation between the source and target texts, typical errors are related to term inappropriateness, information ambiguity, text disfluency and nonstandard grammar. The research findings can provide professional translation training with practical pre-editing or post-editing suggestions.

**Keywords**：machine translation；multidimensional quality metrics；error types

机器翻译自 1949 年由 Warren Weaver 提出以来，几易其“芯”，寄托着人们对高质量机器译文的探索与追求。近年来，以深度学习为基础的机器翻译取得诸多成果，各大公司纷纷将其纳入自己的研发战略中，先后推出在线机器翻译引擎，例如目前应用较广的 Google 翻译、DeepL 翻译、

**收稿日期：**2024－01－09
**基金项目：**2022 年度上海理工大学本科重点教改项目
**第一作者：**杨博超，男，硕士研究生。研究方向：机器翻译。E-mail：1028363053@qq.com
**通信作者：**冷冰冰，女，副教授。研究方向：科技翻译理论与实践、职业译者培训。E-mail：iamlengbingbing@163.com

百度翻译等。当下，在线机器翻译系统在英汉、汉英翻译上的整体质量虽然均达到了及格及以上水平，基本能实现交际目的，但离良好及以上水平尚有差距[1－3]，就连为首的 Google 翻译在汉译英时也不乏出现一些“较为隐蔽的、甚至是中级英语学习者都难以轻易识别的错误”[4]。

生成式人工智能工具 ChatGPT 的问世推动了神经网络机器翻译技术的更迭，国内外有关机器翻译的研究层出不穷。但在有关机器翻译译文质量的评估方式、译文错误分类标准的共性上的观点仍较为模糊[2,5－8]。虽然人们可以运用特定尺度评价译文质量，比如使用认可度较高的 SAE J2450 和 LISA QA Model，但前者受限于行业特定性，通用性不足，后者采用“一体适用”的评估方法，缺乏灵活性。此外，BLEU、METEOR 和 TER 等自动评价标准虽能反映机器译文整体质量，但无法体现译文具体存在哪些问题，且弱化了人工评价这一概念[9－11]。

长期以来，行业内缺乏一种能够针对特定客户需求，编译特定文本类型的机器翻译错误，并基于此种错误类型提出译文接受度评估标准的方法[12]。多维翻译质量标准评估模型（multidimensional quality metrics，MQM，简称“MQM 质量评估模型”）应运而生。

## 一、MQM 质量评估模型

MQM 质量评估模型是欧盟资助 QTLaunchPad 项目的部分成果，是一种可用以评估译文及原文的动态的、全面的和可定制的评估框架。该模型提供了一个按等级划分，共包含 108 种错误类别的分类系统，并详细定义和区分了各种错误类别。此种层次化的分类结合了主流的质量评估方法（如 TAUS dynamic quality framework、LISA QA Model、SAE J2450 等）、质量检测工具（如 Apsic Xbench、CheckMate、XLIFF:doc 等），以及机器翻译质量多元评估理论[13]。MQM 质量评估模型在顶层结构上将翻译问题划分为 5 个维度，即流利性（fluency）、准确性（accuracy）、真实性（verity）、格式（design）和国际化（internationalization），以及少数无法分类的“其他（other）维度”，最终形成树形结构网。同时，为便于操作与统计，MQM 质量评估模型还提供一个包含父类（3 种）和子类（16 种）共计 19 种错误类别的核心模式（MQM Core，简称“核心模式”），见表 1。此外，MQM 质量评估模型赋予每种错误类型一定的权重比，并按错误的严重程度依次划分为重大错误、严重错误、轻微错误和无错误，进而提供了可用以计算翻译质量的计算公式：TQ（质量得分）= 100 — TP（译文罚分）+ SP（原文罚分）。MQM 质量评估模型“真正做到了以用户为中心，让用户能从不同标准、不同层次、不同粒度上对翻译质量进行评估”[14]。

李颖[15]、王晚秋[16]曾运用 MQM 质量评估模型指导翻译实践，认为该模型在评估标准和参数上具备客观性和完整性。其中：李颖侧重对比其他评估系统以检验 MQM 质量评估模型在指导军事文本英译汉时的实用性和有效性；王晚秋受限于机器翻译在政治文本上的高利用率（海量语料库导致），未能充分体现机器翻译的不足。尽管如此，上述研究对本文运用 MQM 模型在语料选择、错误类型划分和错误规避方法等研究设计上给予了启发。

## 二、研究设计

### （一）问题聚焦

为避免“唯心观（mentalist view）”的译文评估方式[17]，确保译文错误分类的标准化和典型性，本研究从国内外众多的在线神经网络机器翻译系统中选择当下号称“全世界最好的机器翻译”——DeepL 翻译系统作为评价对象[18]，运用核心模式对 DeepL 翻译系统生成的译文进行测评。研究主要回答 3 个问题：（1）以 DeepL 为代表的在线机器翻译引擎在汉译英时的质量如何？（2）典型错误有哪些？（3）该如何避免这些错误？

### （二）语料选择

本次研究使用的是自建小型语料库中的 20 例语料，这些语料均来自 2021 年世界 500 强科技公司的中文官网，这些公司如索尼（Sony）、英特尔（Intel）、道达尔能源（TotalEnergies）、沃博联（Walgreens Boots Alliance）等，语料共计约

**表1 包含19种错误类别的核心模式**

**Tab. 1 MQM Core with 19 error types**

| 错误类别 | 翻译质量维度 | | |
|---|---|---|---|
| 父类（parent） | 准确性（accuracy） | 流利性（fluency） | 真实性（verity） |
| 子类（child） | 术语（terminology） | 语域（register） | 完整性（completeness） |
| | 误译（mistranslation） | 风格（style） | 法律要求（legal requirements） |
| | 漏译（omission） | 一致性（inconsistency） | 地区适用性（locale applicability） |
| | 增译（addition） | 拼写（spelling） | |
| | 未译（untranslated） | 排印（typography） | |
| | | 语法（grammar） | |
| | | 地区差异（locale violation） | |
| | | 可理解度（unintelligible） | |

4.4万字，内容以科技报告、技术介绍、新闻报道等在特定领域中使用的专业知识为主。此外，上述中文语料均可在其英文官网中溯源到对应的英文原文，可作为后续在比对中文原文与机器译文时的参照版本。选择此类文本的主要原因在于其英文原文地道，在线机器翻译引擎可直接利用的英译文少。同时，专业文本翻译需求快速增长，对其深入探讨有助于提升译前和译后编辑人员的工作效率，帮助快速识别错误类型、优化译文质量，并为机器翻译的算法改进和模型训练指明方向。

### （三）统计步骤

（1）将选定的20例语料分别导入DeepL翻译系统中，并将平台生成的译文标记存档。

（2）在Translate5[①]中创建项目，并将（1）中的机器译文依次导入此平台。

（3）参照核心模式对不同错误的定义，在充分掌握全部错误类型及定义后，人工识别并标记译文错误。若标记过程中出现争议点，则由多人协商一致决定。由于核心模式并未包含所有错误类型，若出现表1未列的错误类型时，为显化统计结果，将其归纳至对应的父类或其他维度中（归类参照完整版MQM质量评估模型[②]）。同时，为突出机器译文错误的类型化（而非错误的重复性与严重性），对同一句子中多次出现的同种错误只标注统计一次，不同错误则分类标注统计。

（4）错误类型统计与对比分析。由于Translate5中尚未搭载用于自动统计错误重复率及错误严重性权重比的Tabular Scorecard，最终统计结果需由人工汇总至Excel表格中。统计结果由人工识别标记的错误和平台自身识别的错误两部分组成，其中，“单个错误百分比”等于“单个错误重复频率”除以“错误总频率”。

## 三、错误类型统计与对比分析

统计结果显示，DeepL在线翻译平台在专业文本汉译英上存在不少问题，特别集中在准确性和流利性两方面，呈现出类型化和重复化特征，见表2。DeepL翻译引擎虽能较大程度地实现译文与原文的对等，但受限于自身模型缺陷，在处理结构复杂、语义模糊，且含多个限定成分修饰的句子时，也会显得力不从心，主要表现在因过度顺应原文而导致的术语不适切、信息误译等准确性维度上，错误率总计31.62%。诚然，机器翻译的双语转换规则和匹配性原则均以语句为基本单位，较少顾及段落层面和篇章层面等语境因素，因而无法采用对偶式翻译、提要式翻译和删译等方法[19]，译文的流利性大打折扣。比如：此次对比语料中，机器译文常忽略或过度对应前后并列成分、前后副标题等在行文结构上的一致性，导

① Translate5是QTLaunchPad项目的一部分，为开源的翻译编辑平台，参见https://translate5.net。

② MQM质量评估模型（含MQM Core）：http://qt21.eu/mqm-definition.（MQM 1.0）。由于MQM质量评估模型尚未证明错误严重性乘积计算公式的有效性，因此，本研究不额外对错误的严重性进行赋分处理。

**表 2　DeepL 机器译文错误类型分布统计表**

**Tab. 2　Statistics of error types by DeepL**

| 错误类型 | 错误频率 | 百分比/% |
|---|---|---|
| （总）准确性 | （113）22 | （44.66）8.70 |
| 术语 | 34 | 13.44 |
| 误译 | 46 | 18.18 |
| 漏译 | 8 | 3.16 |
| 增译 | 3 | 1.19 |
| 未译 | — | — |
| （总）流利性 | （140）10 | （55.34）3.95 |
| 语域 | — | — |
| 风格 | 18 | 7.11 |
| 一致性 | 39 | 15.42 |
| 拼写 | 5 | 1.98 |
| 排印 | 18 | 7.11 |
| 语法 | 50 | 19.76 |
| 地区差异 | — | — |
| 可理解度 | — | — |
| （总）真实性 | （0）0 | （0）0 |
| 完整性 | — | — |
| 法律要求 | — | — |
| 地区适用性 | — | — |
| 总计 | 253 | 100 |

致前后译文可读性差，错误率总计 15.42%；用于连贯句段所使用的连词、分词、代词等功能词语法错误也十分突出，错误率总计 19.76%。

MQM 质量评估模型对“真实性”维度的解释为：在源语言环境中为真的事实陈述在译文语言环境中为假，涵盖流程制度完整性、法律要求合规性、地区适用性等。而本次研究尚未发现上述问题。因此，准确性维度和流利性维度是本研究的研究目标。下文将对高频错误予以举例说明。

### （一）准确性维度

MQM 质量评估模型对“准确性”的定义为：除规定的差异外，目标文本中不能准确传递源文本内容的错误，涵盖术语（错误）、误译、漏译、增译、未译等。需特别说明的是，MQM 质量评估模型中对“术语”这一错误类型的定义为特定领域的术语被翻译成不同于该领域预期的术语或惯用的术语，并且与“误译”处于同一级别，因此为显化错误类型，此次统计过程中将术语误译归纳在“术语”类别中。

#### 1. 术语的适切性

机器翻译基于庞大的语料库，能将大部分术语翻译准确，但由于术语在技术交流中形式灵活而存在翻译错误，出现非领域内或非惯用的术语。

**例 1**　碳捕集领域的颠覆性创新，助力中国双碳目标。

DeepL: A disruptive innovation in carbon capture to help *China's dual carbon target*.

Reference: Innovation in carbon capture to help *China's carbon peaking and carbon neutrality goals*.

例 1 中的“双碳”是我国在第七十五届联合国大会上提出的两项目标，为“碳达峰”与“碳中和”的简称。由于原文术语的简化表达，机器翻译无法译出该术语的内涵信息，西方读者也就无法领略该术语背后承载的科技内容。因此，为规避此类情况，译者可直接在译后阶段对译文进行二次加工，直译出隐含义，或在译前编辑阶段对原文进行预编辑，补全核心信息。因此，“双碳”目标最终可译为“carbon peaking and carbon neutrality”或“double carbon targets, carbon peaking and carbon neutrality”。

此外，机器翻译受限于“逻辑推理能力与语境主观识别能力”[20]，难以做到按语境匹配得体的术语表达。如例 2 中“抗腐蚀性”应与其他类别“机械特性”“电特性”“热特性”等译为同一个词性，以表达相似的信息内容。但是因为机器翻译无法主观识别该语境对并列词组的词性同一性要求，译成了“corrosion resistance”，因而造成术语变体的使用错误。

**例 2**　几乎所有固体材料的重要特性都可以被划分为六个不同的类别：机械、电、热、磁、光学和耐腐蚀性。

DeepL: The important properties of almost all solid materials can be classified into six different categories: mechanical, electrical, thermal, magnetic, optical and *corrosion resistance*.

Reference: The important properties of almost all

solid materials can be classified into six different categories: mechanical, electrical, thermal, magnetic, optical and *corrosion resistant/deteriorative*.

可见，译者除了善用搜索引擎，增强信息获取能力外，还在机译过程中提供一定容量的术语库辅助以提高机器译文的质量也非常重要。比如，可在译前编辑阶段为术语单独设定对应的翻译规则，或在译后编辑阶段对机译术语进行二次加工，阐释术语内涵或选用得体的术语变体来提高专业文本的可读性。

**2. 信息的准确性**

机器翻译引擎受限于训练模型的缺陷极易造成误译，使得译文与原文在词汇、句法、篇章等方面信息不对等。在本研究使用的语料中，此类错误占比约为18.18%。如：例3中“妨碍血液循环”的诱因实则由“脱水”引起，而非“含酒精的饮料”；例4中患者“需求亦不同”的原因实则是患者自身“病情或情况不同”。在这两个译例中，因为汉语逻辑不外显，机器翻译引擎不具备对语义逻辑的推理能力，因而句段中各分句关系的错误切分导致了译文信息的错误传递。

**例3** 最易得此病的人是在飞行中大量饮酒的人，因为含酒精的饮料会引起脱水，也会妨碍血液循环。

DeepL: The people most at risk are those who drink a lot of alcohol on flights, as alcoholic beverages can cause dehydration *and also impede blood circulation*.

Reference: The most likely victims are those who drink a lot on their flights, as alcoholic beverages would cause dehydration *which also interferes with blood circulation*.

**例4** 每个患者病情或情况不同，需求亦不同。

DeepL: Each patient’s condition or situation is different and *so are their needs*.

Reference: Each patient is an individual with different needs, *depending on their special illness or condition*.

此种错误类型需要译者在译前阶段揣摩原文内涵，增补逻辑关系词，明确分句间的具体切分，进而提升机器译文的质量，减少译后编辑工作量。

**（二）流利性维度**

MQM质量评估模型所定义的“流利性”与文本的形式或内容密切相关，涵盖内容、惯例和可理解度三大类。其中：内容主要包括语域、风格指南、一致性及是否有歧义等；惯例包括拼写、排印、语法、地区差异等。

**1. 译文的流畅性**

机器翻译系统虽能最大限度地调换各分句位置以凸显隐性逻辑，避免因“顺句驱动”造成“死译”，但中文的“流水句”结构使得并列小句层层铺排，令机器译文的流畅性大打折扣[21]。在本研究语料中，此类错误主要表现在译文用词重复、平行结构使用错误等，错误占比为19.37%。

省略重复成分是汉英翻译中规避中式英语、增强译文可读性的常用手段。由于机器翻译引擎尚无法对相同词块灵活传译，导致英译文重复的问题非常突出。如例5中“几千瓦热实验”被译为“...kW thermal experiment”，这一表达在DeepL机器译文中重复了3次（对照参考译文的处理），这并不符合英语语言“崇尚简洁”的本质要求。译文重复仍然是当前机器翻译引擎难以避免的问题。

**例5** 比如在法国国家石油与新能源研究院（IFPEN）进行的10千瓦热实验，在清华大学（THU）进行的30千瓦热实验，以及挪威科技工业研究院（SINTEF）的150千瓦热实验。

DeepL: Examples include the *10 kW thermal experiment* at the Institut National de la Pétrole et des Energies Nouvelles (IFPEN), the *30 kW thermal experiment* at Tsinghua University (THU) and the *150 kW thermal experiment* at the Norwegian Institute of Science, Technology and Industry (SINTEF).

Reference: Such as a *10 kW thermal experiment* at IFPEN, a *30 kWth* at THU, and a *150 kWth* at SINTEF.

值得注意的是，机器翻译中另一类重复问题是句子某一部分结构的重复翻译，如例6。崔启亮等基于专业文本的英汉机器翻译也提出了同样的问题，并归因于机器翻译语料训练不足[22]。在本次对比语料中，此类错误占了3.95%，可见，这类错误是英汉互译的通病。

**例6** 挤压铸件的微观结构不像传统压铸件看

起来那样细。

DeepL: The microstructure of extruded castings is not as fine as *it appears to be in conventional die casting*.

Reference: The microstructure of the squeeze casting is not as fine as *that observed in conventional die casting*.

总之，目前以 DeepL 为代表的机器翻译引擎存在的一个重要译文错误类型是不善于使用除代词以外的词汇来衔接语篇，提升可读性，这就需要“采用上义词、同义词、近义词、总称词等重复以外的词汇衔接来避免此类问题，达到求雅换词的效果”[18]。

**2. 语法的规范性**

由于中英文语言结构的不同，当中文各分句间有明显连接词时，机器翻译可最大限度地还原其内部逻辑（尽管偶尔也会出现连接词的混用或误用等规范性问题）；但若原文无明显连接词，则译文出错率倍增。罗季美等也曾强调机器翻译研究者应侧重“对连接词，尤其是对高频率连接词的形式化研究”[23]。本次语料统计中，连接词错误主要包括分词错误和连词错误，合计占比约为 16.97%。

分词因具有比从句更简洁的特征而在专业文本中被广泛运用[23]。但若多个分词连用（如例 7 的 DeepL 译文），则句间松散，小句信息层次不清，有堆叠之感，违背英语语言的表达习惯。此类错误通常由于受到汉语流水句结构的影响。对应的预防措施包括译前编辑原文，增加相应的汉语关联词，合并小句或者断句重组，这样可以降低机器译文错误率，提高译文产出效率。

**例 7**　ABB 正在 100 多辆 Re460 型机车上安装节能型牵引变流器，使其延长20年的使用寿命，并提供三个车引变电站，以帮助提高功路和增加电网容量。

DeepL: ABB is installing energy-efficient traction converters on more than 100 Re460 locomotives, *extending their service life by 20 years*, and *providing three traction substations to help boost power and increase grid capacity*.

Reference: ABB is installing energy-efficient traction converters in more than 100 Re460 locomotives to extend their life by 20 years, *and three traction substations are provided to help boost power and increase capacity network*.

连词作为连接词与词、短语与短语、句与句的虚词，可大致分为并列连词和从属连词。在机器翻译中，并列连词错误最高频。本研究数据中，DeepL 犯错频次较高的连词主要有“和”“与”“同”“以及”等。如例 8 DeepL 中“as well as”的使用是直译了“以及”。但是从英语语法上说，当提及两个以上的并列成分时，第一、二个并列成分之间需要使用顿号或逗号隔开，而不宜使用连词，而且“as well as”这个连词强调的是前面的并列成分，这显然扭曲了原文的含义。例 8 机器译文犯错的原因与汉语原文表达的不严谨有直接关系，建议译者一方面在译前可对原文进行优化，另一方面还须扎实掌握功能词的用法，以便在译后编辑阶段提高编辑效率，产出高质量译文。

**例 8**　博世的业务范围非常广泛，包括动力总成解决方案，以及底盘系统控制，电气驱动，汽车转向部门和成立于 2021 年初的跨域计算解决方案部门。

DeepL: Bosch has a broad range of activities, including powertrain solutions, *as well as* the Chassis Systems Control, Electrical Drives, Automotive Steering divisions and the Cross-Domain Computing Solutions division, which was established in early 2021.

Reference: Bosch is involved in many diverse fields, including Powertrain Solutions, Chassis Systems Control and Automotive Steering divisions, and the Cross-Domain Computing Solutions division set up at the start of 2021.

## 四、结语

本文基于自建的专业文本小型语料库，对搭载神经网络的 DeepL 翻译系统在汉译英时的表现进行了实证研究。结果表明，DeepL 翻译系统总体上能满足信息类文本的基本翻译需求，但在处理复杂句子时仍捉襟见肘，翻译错误主要集中在“准确性”和“流利性”两个维度，典型错误表现为术语不适切、信息不准确、译文不流畅和语

法不规范4个方面。

诚然，语言服务行业随着人工智能的普及不断地催生新的翻译技术，但实际上“没有任何一个机器翻译系统能在兼顾源语文本概念意义和语篇意义的同时，再现人际交流意义”[19]。因此，掌握以“机器翻译+译后编辑”模式为代表的翻译技术，理解机器翻译工作原理，能帮助译者快速甄别机器翻译中常见的错误，而通过“机器翻译+MQM质量评估模型”模式的有机结合，译者更是能基于此模型中对各项翻译错误的定义和分类，将其作为辅助性文档，定点加工译文，进行精细化作业。如：通过译后编辑显化术语概念，使用得体的术语来增强术语表达的适切性；消除内容歧义，提升原文的准确度；给原文适度“减肥”，删去重复用词或冗余信息，增强可读性；通过译后编辑规范关联词，紧凑句段结构；等等。此外，MQM质量评估模型也可帮助翻译学习者乃至翻译技术开发人员明确机器翻译模型的典型缺陷，主动绕障，提高译文输出精度，从而让机器翻译更好地辅助人工翻译。

**参考文献：**


[1] 李奉栖. 基于神经网络的在线机器翻译系统英汉互译质量对比研究 [J]. 上海翻译, 2021(4): 46－52.

[2] 文旭, 田亚灵. ChatGPT应用于中国特色话语翻译的有效性研究 [J]. 上海翻译, 2024(2): 27－34; 94－95.

[3] 戴光荣, 刘思圻. 神经网络机器翻译: 进展与挑战 [J]. 外语教学, 2023, 44(1): 82－89.

[4] 包凯. 谷歌翻译汉译英错误类型及纠错方法初探 [J]. 中国科技翻译, 2017, 30(4): 20－23.

[5] 项睿. 在线翻译软件灰色综合评价 [J]. 信息技术, 2016(9): 77－80.

[6] 孙逸群, 周敏康. 机器翻译质量综合评价方法研究 [J]. 中国科技翻译, 2017, 30(2): 20－24.

[7] 刘世界. 涉海翻译中的机器翻译应用效能: 基于BLEU、chrF++和BERTScore指标的综合评估 [J]. 中国海洋大学学报 (社会科学版), 2024(2): 21－31.

[8] 杨文地, 范梓锐. 科技语篇机器翻译的译后编辑例析 [J]. 上海翻译, 2021(6): 54－59.

[9] PAPINENI K, ROUKOS S, WARD T, et al. Bleu: a method for automatic evaluation of machine translation[C]//Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia: Association for Computational Linguistics, 2002: 311－318.

[10] LAVIE A, AGARWAL A. METEOR: an automatic metric for MT evaluation with high levels of correlation with human judgments[C]//Proceedings of the Second Workshop on Statistical Machine Translation. Prague: Association for Computational Linguistics, 2007: 228－231.

[11] 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: Technical Papers. Cambridge: Association for Machine Translation in the Americas, 2006: 223－231.

[12] MARHEINECKE K. Can quality metrics become the drivers of machine translation uptake? an industry perspective[C]//Proceedings of the LREC 2016 Workshop “Translation Evaluation: From Fragmented Tools and Data Sets to an Integrated Ecosystem”. Slovenia: European Language Resources Association, 2016: 71－75.

[13] BURCHARDT A. Multidimensional quality metrics: a flexible system for assessing translation quality[C]//Proceedings of Translating and the Computer 35. London: Aslib, 2013: 1－7.

[14] 田朋. 翻译多维质量标准MQM模型介评与启示 [J]. 东方翻译, 2020(3): 23－30.

[15] 李颖. 多维度质量指标体系下军事纪实文本译文的翻译质量评估 [D]. 济南: 山东师范大学, 2020.

[16] 王晚秋. 基于多维度质量指标框架的DeepL翻译器政治文本翻译质量评估 [D]. 北京: 北京邮电大学, 2021.

[17] HOUSE J. Translation Quality Assessment: Past and Present[M]. New York: Routledge: 2015: 241－264.

[18] Why DeepL[EB/OL]. [2024-04-03]. https://www.deepl.com/zh/whydeepl.

[19] 胡开宝, 李翼. 机器翻译特征及其与人工翻译关系的研究 [J]. 中国翻译, 2016, 37(5): 10－14.

[20] 翁义明, 王金平. 文学语篇机器翻译的特征与局限——汉语流水句人机英译对比研究 [J]. 当代外语研究, 2020(6): 128－137.

[21] 李奉栖. 翻译硕士专业 (MTI) 学生汉英翻译中的语篇衔接问题研究 [J]. 外语教学理论与实践, 2021(1): 138－152.

[22] 崔启亮, 李闻. 译后编辑错误类型研究——基于科技文本英汉机器翻译 [J]. 中国科技翻译, 2015, 28(4): 19－22.

[23] 罗季美, 李梅. 机器翻译译文错误分析 [J]. 中国翻译, 2012, 33(5): 84－89.


（责编：朱渭波）