Modern Linguistics 现代语言学, 2023, 11(3), 1197-1205
Published Online March 2023 in Hans. https://www.hanspub.org/journal/ml
https://doi.org/10.12677/ml.2023.113161

Hans汉斯

日汉机器翻译误译分析及改善策略探讨

——以日本首相的致辞文本为例

王杨帆，卜朝晖

广西大学外国语学院，广西 南宁

收稿日期：2023年2月20日；录用日期：2023年3月21日；发布日期：2023年3月31日

摘　要

本文以日本首相的致辞文本为例，研究了百度机器翻译系统在致辞文本翻译上的表现，统计了机器翻译误译类型的分布，分析了机器误译的主要类型及原因，并从译者编辑的角度提出了改进策略。研究结果表明：百度翻译在致辞类文本上的整体表现较差，有超过半数的句子需要进行不同程度的译后编辑。在影响原文理解的机器翻译误译类型中，句段逻辑误译和词组搭配误译分别占32%，其次是专业术语误译，占12%，一般词义误译占9%，多译有12%，最后漏译有3%。针对这些机译问题，提出了译前简化长难句；在计算机翻译技术合成环境中工作；先修改后校对的策略，并通过实例进行验证，以期提高人机合作翻译工作的效率。

关键词

机器翻译，日译汉，致辞文本，改善策略

Analysis of Japanese-Chinese Machine Mistranslation and Discussion of Improvement Strategies

—Using the Text of Japanese Prime Minister's Speech as an Example

Yangfan Wang, Zhaohui Bu

College of Foreign Languages, Guangxi University, Nanning Guangxi

Received: Feb. 20th, 2023; accepted: Mar. 21st, 2023; published: Mar. 31st, 2023

Abstract

This paper takes the speech text of the Prime Minister of Japan as an example and studies the per-

文章引用: 王杨帆, 卜朝晖. 日汉机器翻译误译分析及改善策略探讨[J]. 现代语言学, 2023, 11(3): 1197-1205.
DOI: 10.12677/ml.2023.113161

formance of Baidu's machine translation system on the translation of speech texts, statistics on the distribution of machine translation mistranslation types, analyses the main types of machine mistranslation and their causes. It also proposes improvement strategies from the perspective of translator edits. The research results demonstrate that the overall performance of Baidu Translation on speech texts is weak with more than half of the sentences requiring different degrees of post-translation editing. Among the types of machine translation mistranslation that affect the comprehension of the original text, sentence paragraph logic mistranslation and phrase collocation mistranslation respectively account for 32%, followed by terminology mistranslation at 12%, general word sense mistranslation at 9%, multiple translations at 12%, and finally omission at 3%. In order to address these machine translation problems, the strategies of simplifying long and difficult sentences before translation; working in the synthetic environment of computer translation technology; and revising before proofreading are proposed and verified by examples, with a view to improving the efficiency of human-computer cooperative translation work.

Keywords

Machine Translation, Japanese-Chinese Translation, Speech Text, Improve Strategy

1. 引言

1.1. 研究背景

随着计算机和信息技术的发展，翻译技术也在不断发展。以机器翻译为主的翻译技术受到了业界和学术界的广泛关注。机器翻译(Machine Translation)被国际标准化组织(ISO)规定为“运用计算机体系将文本或语音从一种自然语言自动翻译为另一种语言的过程”。

机器翻译包含基于规则、基于实例、统计和神经网络四种技术类型。2017 年神经网络机器翻译技术的出现，大大提高了译文的流畅度。目前，神经网络机器翻译已成为应用最广泛的机器翻译系统。将机器翻译和人工翻译相结合，衍生以机器翻译、人工译后编辑和翻译项目管理为核心的语言服务正在日益增多[1]。

本文以神经网络机器翻译系统的日汉翻译为样本进行研究，分析了机器翻译的误译类型，从而为翻译从业者的编辑工作提供一定的参考，以期提升翻译的整体效率。

1.2. 先行研究

近年来，学者主要致力于英汉机器互译。罗等人[1] [2] [3] [4]从词汇、句法和篇章层面归纳了机器翻译的误译类型。崔等人[5] [6] [7] [8] [9]分析了科技文本的译后编辑误译类型[3]。李奉栖[10]进行了此类在线翻译系统的英汉互译质量对比研究，通过翻译－评分－统计－分析对比的步骤得到了 3 个梯队的机翻质量报告[4]。杨文地[11] [12] [13]以科技文献为例，解析了神经网络机器翻译的以后编辑过程[5]。但日汉机器互译方面的研究较少，特别是机器翻译对致辞类文本的研究就更少了。因此，本文以日本首相的致辞文本为例，深入分析日汉机器翻译的准确度和存在的问题，并对此提出一些供学者参考的意见。

1.3. 研究对象

基于全球中文搜索引擎百度的神经网络翻译系统，对日本前首相安倍在国际律师协会东京年会上的

致辞发言稿 [1] 进行机器日汉互译。致辞发言稿在赖斯文本类型理论中属于信息型文本、表情型文本和感染型文本之间的混合型文本，如图 1 所示。并且，其严谨性与信息型文本有差异，文学性不如表情型文本，呼吁性的感染力在一定程度上接近感染性文本。但是，致辞发言稿同时具备信息型文本的严谨性，表情型文本的文学性，感染性文本的感染性，因此，选用此类文本进行日汉机器互译具有代表性。

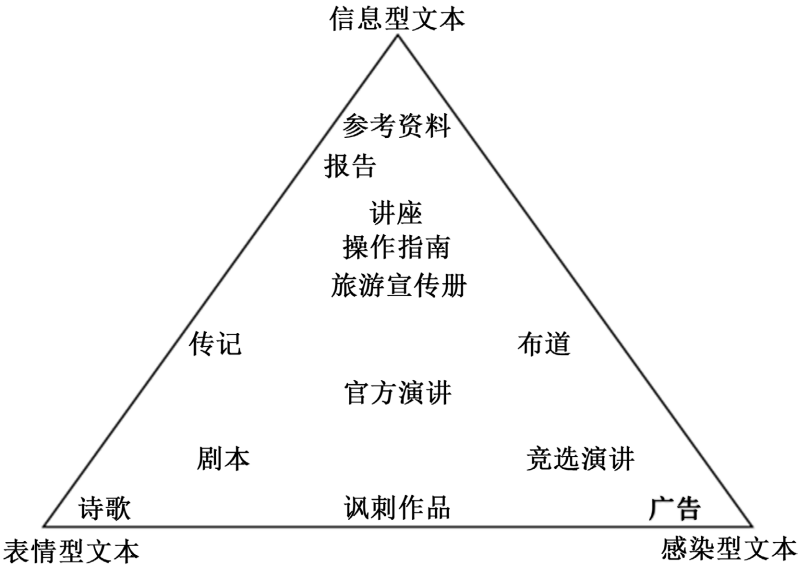

Figure 1. Types of text in Reiss [7]
图 1. 赖斯的文本类型[7]

2. 机器误译分析

2.1. 机器误译整体评价

在本文中误译句子指的是影响读者理解原文的语句，不包含译文风格和标点的误译。作者对机器翻译误译率进行了统计，以百度机器翻译的致辞类文本做出评价，见表 1。在表 1 中，日本前首相安倍在国际律师协会东京年会上的致辞发言稿中句子数有 59，机器互译的误译句数为 32，错误率为 63%。由此可知，神经网络日汉机器翻译的准确度距离理想状态存在一定的差距，翻译工作者需要对原文的二分之一以上的内容进行修正。

Table 1. Number and proportion of sentences mistranslated by Baidu translation
表 1. 百度翻译误译的句数及比例

原文句数	误译句数	误译占比
59	37	63%

2.2. 机器误译类型分布

笔者对上文中的误译句子进行分析，将误译的句子进行统计分类，结构如表 2 所示。在表 2 中，相同的词语在前后文重复误译时，视为一处误译。

本文将机器翻译的问题分为错译、多译、漏译和译文风格四种类型。其中，错译根据语言单位的不

[1] 日本首相的致辞文本来自日本首相官邸网站 https://www.kantei.go.jp/cn/96_abe/statement/201410/1019iba_speech.html。

同，分为词汇意义误译、词组搭配误译和句子逻辑误译三种类型。词汇意义误译当中，较为明显的有专业术语误译，另外出现有一般词义的误译。

Table 2. Machine mistranslation classification and statistics
表 2. 机器误译的分类及统计

错译				多译	漏译	译文风格问题	合计
专业术语误译	一般词义误译	词组搭配误译	句子逻辑误译				
4	3	11	11	4	1	6	40

从分析结果可知，错译、多译和漏译的机器翻译误译问题将影响读者的正常阅读和对词汇的理解；译文风格问题对于读者理解致辞类文本的影响不大。由于 MTPE 翻译模式的最大目的之一是提升翻译效率，因此译文风格问题暂不纳入译者编辑的考虑范围之内。

将机器误译类型进行可视化分析，得到如图 2 所示的结果。

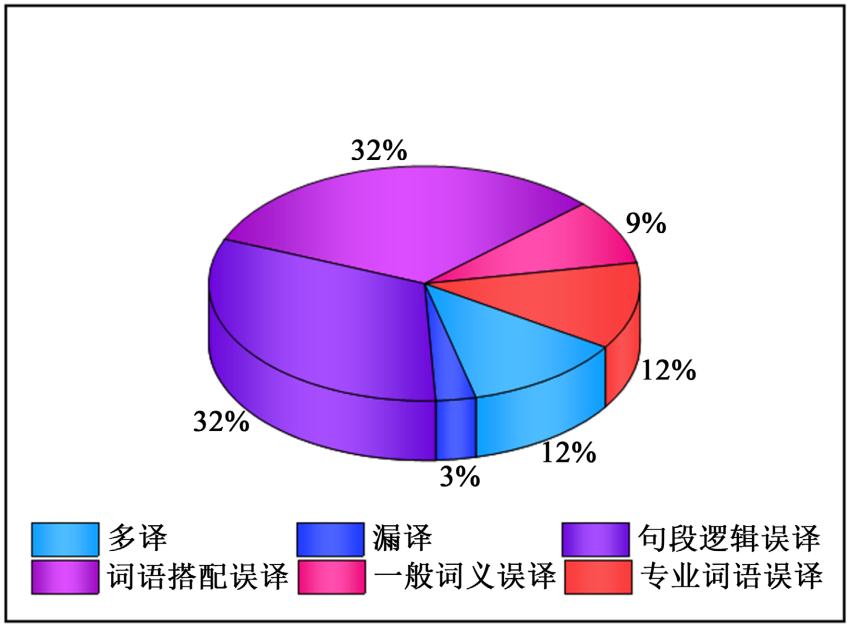

Figure 2. The distribution of machine mistranslation types affecting the comprehension of the original text
图 2. 影响原文理解的机器误译类型分布

由图 2 可知，在影响原文理解的机器误译的类型中，词组搭配误译和句段逻辑误译分别占误译总量的 32%，该类型为典型的两种机器误译问题。

其次，专业术语误译占 12%，一般词义误译占 9%，两者相加为词汇误译，占整体的 21%。由此可知，机器翻译能正确翻译日本前首相安倍在国际律师协会东京年会上的致辞发言稿中大部分词汇，但是对于部分专业术语及部分一般词汇仍难做到准确翻译。从图 2 中可知，多译问题的占比为 12%，与专业术语误译问题相当，较为突出。漏译问题所占比例较小，仅占 3%。

从机器日汉互译结果来看，错译是日汉机器翻译的主要问题，主要表现在句段逻辑、词组搭配、词汇的误译问题，占比高达 85%。其次，多译问题占比 12%。最后，漏译问题占比 3%。下文将对各个类型的机器翻译误译问题进行实例分析，说明其所指意义。

2.3. 机器误译类型的实例分析

2.3.1. 专业术语误译

在此前的研究中，专业术语误译被认为是机译中最严重的问题，在本次的研究结果中，专业性术语误译比一般词义误译多出3%。但是，两者的比例差距较小，因此，在本文中将专业性术语包含于词义误译中，作为词义误译中的特殊现象。在本研究的测试文本中，专业术语在文中出现频率高，属于核心概念。百度机器翻译对专业术语的误译将影响对读者原文主旨的把握。此外，前后文专业术语译文不一将导致读者产生疑惑和混乱。

例如专业术语“国際法曹協会(IBA)”在原文共出现5次，百度翻译机器译文分别翻译为“国际法律协会”2次、“国际法曹协会”3次。《日本国语大辞典》对“法曹”的解释为“法務をつかさどる者。法律の事務に従事する人。特に、裁判官、検察官、弁護士などをさす。法律家。(掌管法务、从事法律事务的人。特指法官、检察官、律师等。)”指的是“司法工作者”，即某类人的总称，人工译文为“国际律师协会”。

对于百度机器翻译给出的译文，读者虽然可以大致猜测出其为法律领域的术语，但对于其具体所指是“律师”或“法律”仍将产生误解，对于机器误译的“法曹”将产生疑问，同时同一术语的多种译法将给读者造成混乱，影响阅读的流畅度。

2.3.2. 一般词义误译

机器翻译可以快速处理原文得到译文，在某种程度上将减少译员的认知努力，从而提高效率。但是机译译文对词汇的翻译有时并不正确。有时译员无需对照原文，仅通过判断机译译文语句是否通顺、意义是否连贯就能判断误译之处并做出修改，如原文中的“子の連れ去り問題”机器误译为“带走孩子的问题”，显然“带走孩子”不能成为“问题”，因此译员能够判断此处有误，根据上下文将“带走”修改为“掠拐”。

有时译者仅通读机器译文无法发现机器误译，对照原文后才能发现词义误译。如原文中“天下に道のある時”百度翻译的“天下有路的时候”单看没有语法错误，对照原文才能发现这是引用汉语古语，这里的“道”从“走的路”抽象为“道义选择”，因此应修改为“天下有道的时候”。

对于一般词义的误译，母语译者有时仅通过预感、语用搭配就能发现问题、作出修正；有时词义的误译则比较隐蔽，译者需要照原文才能发现。因此在译后编辑时可以先通读译文，修改明显的词义问题，这一步骤将减少后续编辑所花费的时间。

2.3.3. 词组搭配误译

词组搭配误译是由于汉语和日语语言使用习惯不同，机器翻译采用直译后产生的不自然表达。如原文中“考え方に親しんだ”机器翻译为“对(某)想法很亲近”，很明显汉语中的“亲近”和有生命的人物搭配，对于无生命的事物或抽象事物，用“熟悉”搭配。因此修改为“对(某)想法很熟悉”，更符合汉语搭配习惯。类似的还有“符合利益”机译为“是利益”、“完善制度”机译为“整备制度”，机器译文的搭配在汉语中无法成立，并且带有浓重的翻译腔，影响了译文流畅度和译文风格。

再如原文中“オールジャパンの態勢”机器译为“以全日本的态势”，作为母语译员能感受到日语“態勢”和汉语“态势”的微妙语感。在CCL汉语语料库中检索“态势”，发现前面一般跟动词(如“发展的态势”)、形容词(如“有利的态势”)，而不跟地点名词。此处替换为“举……之力”搭配更为恰当。

2.3.4. 句段逻辑误译

机器翻译长难句时容易出现逻辑问题。《百科事典》将超过“37~47”字的句子界定为长句。长句因

结构复杂、修饰成分多，对于译者来说同样是翻译的难点。而日语的语言特点决定了日语中常有定语修饰的现象，常有长句出现。此外，日语作为高度依赖语境的语言，常用授受、被动、使役等方式内隐式地传达意义，容易使机器产生逻辑误译。例如，(1)

原文：“「法の支配」の本質は、権力は絶対ではなく、権力の上に、権力が奉仕すべき、また、権力が縛られる道徳的実在がある、と言うことです。”

机译译文：“法的支配”的本质是，权力不是绝对的，权力应该服务于权力之上，还有权力被束缚的道德实际存在。

人工译文：“法治”的本质可谓，权力并非绝对，权力应服务于其上的客观存在以及可束缚权力的道德。

上文的例子中，中心词“道德”之前存在两个连续的修饰语“権力が奉仕すべき(权利应服务的)”和“権力が縛られる(束缚权利的)”。其中对于第二个修饰语，机器翻译按照直译的方式译为“权利被束缚的”，混淆了主谓语的逻辑表达顺序，造成了理解的困难。

2.3.5. 多译

多译是指原文中无需翻译出的内容被机器直接翻译，造成译文冗余的情况。例如，(2)

原文：それは、東アジアでは、例えば「仁」とか「慈悲」と呼ばれています。

机译译文：在东亚，例如被称为“仁”或“慈悲”。

人工译文：在东亚，被称为“仁”或是“慈悲”。

这句话中的“例えば”是语气词，只表示语气的强弱，不影响句意。日语还有很多类似的表达如“くらい(大约)”“ほど(程度)”、“と思われる”、“と言われる”等，为模糊和缓和语气而采用的暧昧表达，在翻译时可以适当省略。机器则很难判断日语中的语气词，造成多译的问题。再如机译译文“俄罗斯的文豪托尔斯泰可能都说了同样的话。”当中的“可能”也属于这一类型的多译。因为某些日语句子中的“可能”只是一种缓和语气的习惯性表达，并无实际的推测意义。

2.3.6. 漏译

漏译是指翻译时跳过原文中的某些词句造成译文信息缺失的情况。漏译虽然在机器翻译中不常出现，但是如果出现漏译，将在一定程度上影响读者对原文的理解。同时，译员需要仔细核查原文才能发现机译译文的错漏之处，这在很大程度上增加了译后编译的工作量。例如，(3)

原文：……若き侍たちが、やがて大きく歴史を動かし、日本を近代化に導くことになりました。

机器译文：年轻武士们，不久就开始推动历史，引导日本走向近代化。

人工译文：年轻武士们，此后成为推动历史的巨大力量，引领日本走向现代化。

原文中“大きく歴史を動かし”是“极大地推动历史”之意，机器译文中省去了“大きく(大大地，极大地)”的翻译，造成了漏译。但是这一程度副词的翻译对译文的理解的影响并不大，对于译文的理解造成了很小的偏差。机器翻译的漏译一般不会出现在文段中心词、专业术语及核心意义的翻译上，漏译的发生可能与输入机译系统的文段过长有关。

2.3.7. 译文风格

本研究中的文本为日本首相的致辞，带有口语特点，短语较多。因场合需要使用了日语敬体和敬语，语体庄重。因出现于律师年会上，因此文本较多地涉及法律专业术语，这对于非法律背景的读者来说，可能有一定的晦涩感。总体来看，原文风格应是措辞严谨、语气庄重、文雅、略为晦涩的。

机器译文基本采用直译，语言直白，现代感更强，不能体现出原文的庄重和文雅；对比之下，人工译文语言精炼含蓄，增加了古雅之感，更贴近原文风格。这一点在介词的翻译上最为明显：机器翻译更

多地使用介词“的”，增加了文章的松散程度；人工翻译则更多地使用了“其”、“之”等半文言说法，使文章显得更具古韵。从实词的翻译来看，同一词语，机器翻译往往采用直译的二字词语，更接近日常口语说话；人工翻译则更多地使用四字词语或成语，使文章显得的更为庄重，更加适宜致辞场合。从文章引用语的翻译来看，机器翻译将引用语译为白话文，更利于读者理解；人工翻译采用了引用语的原始出处，增加了文章的历史厚重感。

根据以上分析可知，百度机器翻译难以译出严肃的致辞文本的风格，目前仍停留在表层意义的翻译上。

3. 机译编辑策略

通过上文对机器翻译误译类型的总结和分析，得到了 6 种不同的误译类型。明确了各误译问题的占比。接下来针对各误译问题提出相应的编辑策略。

3.1. 译前简化长难句

首先针对最为显著的句段逻辑误译问题，提出译前编辑策略。因其发生的主要原因是，长句的结构过于复杂，从句过多，句子成份难以辨析。针对此问题，提出机译前通过切分重组、调整语序等方式简化长难句的策略。化简长难句不仅能使机器翻译达到更好的效果，也能帮助译者理清句子成分之间的关系，从而缩短译后编辑的时间。下文将举例说明，下文中原文用 ST (Source Text)表示，原文的简化版本用 ST’表示，机器翻译的译文用 MTTT (Machine Translation Target Text)和 MTTT’表示。

ST：「法の支配」の本質は、権力は絶対ではなく、権力の上に、権力が奉仕すべき、また、権力が縛られる道徳的実在がある、と言うことです。

ST’：「法の支配」の本質は、権力は絶対ではなく、権力の上に道徳的実在がある、と言うことです。権力は道徳的実在を奉仕すべきであり、道徳的実在は権力を縛る。

提取原文 ST 的主干作为第一句话，提取中心语“道徳的実在”前的长修饰语，作为另外一句话。因为日语句子的主干一般是“……は……です· だ· である”或者“……は……を V”的形式，根据标志词可以迅速找到主干，所以提取并不困难。剩下的部分同样修改为日语句子的基本形式。以上分析可知，日语长难句的简化工作比直接翻译或直接修改机器译文都更加简单。

长难句简化处理后，对机器翻译结果进行展示和分析。

MTTT：“法的支配”的本质是，权力不是绝对的，权力应该服务于权力之上，还有权力被束缚的道德客观存在。

MTTT’：“法的支配”的本质是，权力不是绝对的，权力上有道德上的实际存在。权力应该服务于道德的实际存在，道德的客观存在束缚着权力。

人工译文：“法治”的本质可谓，权力并非绝对，权力应服务于权力之上的客观存在以及可束缚权力的道德。

如上所示，MTTT 中的分句“还有权利被束缚的道德实际”存在语病，而简化后的机器译文 MTTT‘中则不存在，且 MTTT’译文流畅无误。这表明，相较于多修饰语和被动语态的日语长难句来说，百度机器翻译在多个日语短句的翻译上表现更好。由此可知，机器翻译译前对长难句进行切分，可以降低译者翻译和审校的难度，提高机器翻译的正确率，从而增加翻译工作的效率。

3.2. 在计算机翻译技术合成环境中工作

针对于专业术语误译和词组搭配误译的问题，在计算机翻译技术合成环境中工作的应对策略。

其中，专业术语始终是翻译实践中的重难点，在使用机器翻译之前，有必要浏览原文并调查专业术语的译法，在计算机辅助软件中提前创建术语库，登录术语资源。在计算机辅助软件内部使用机器翻译，并得到切分好的可直接编辑的句对，节省了一般从机器翻译引擎复制提取译文的时间，在工作量大的情况下效果更为显著。译后编辑完成后还可利用计算机辅助翻译软件中的核查功能，自动对比原文与译文词数或句长，帮助译员及时发现漏译的情况。

如果译员能够在 CAT 软件中进行译后编辑，利用 CAT 术语库功能，一次编辑替换重复误译，编辑量则大大降低。前文统计了需要做译后编辑的句子数量，如果能够利用计算机翻译技术合成环境，译前完成术语的统一，则可以大大降低译后编辑的工作量。通过自动替换重复误译后的译后编辑量如表 3 所示。从表 3 可知，译后编辑工作量从 63%下降至 49%，下降了 14%。这在一定程度上减轻了译后编辑的工作量。

Table 3. Numbers of mistranslations in the translation technology synthesis environment
表 3. 翻译技术合成环境中的误译数量

原文句数	误译句数	误译占比
59	29	49%

3.3. 先修改后校对

译后编辑(Post Editing)，ISO 规定为“检查和修正机器翻译的输出”，并将译后编辑分快速译后编辑(Light post-editing)和完全译后编辑(Full post-editing)。快速译后编辑得到的是可读性基本得到保证的译文，完全译后编辑要求译文与原文风格一致，读起来就像用目标语写成的文章。

根据上文分析，词组搭配误译是由于中日表达习惯不同造成的，对于汉语为母语的译员来说，可以不看原文，补充出正确的搭配。对于多译问题，译员也可以不看原文直接删去多余成分。译员不看原文进行编辑，可以节省部分时间可精力。编辑至译文通顺后，可以对照原文，进行最后一步的审校，此时左右对照可以加快审阅速度，最后再完成细节编辑。此方法属于译后编辑的范畴。

4. 结论

本文研究了百度翻译在日本首相致辞文本汉译上的表现，通过统计误译句子的错误率，得出百度翻译日译汉在致辞文本上表现不容乐观的结论，有超过二分之一的机器输出需要译后编辑。然后，归纳出机器翻译误译类型，并作了分布统计，得出句段逻辑和词组搭配是最主要的机译问题。其次，是专业术语等词语误译以及多译问题。针对这些机器误译问题，提出了译前切分复杂长句、在计算机翻译技术合成环境工作、先修改后校对的译后编辑策略。同时，本文补充了日汉机器翻译对于不同类型文本表现的研究，希望能够对日汉机器翻译系统开发者和使用者提供一定的参考。

参考文献

[1] 罗季美. 机器翻译句法错误分析[J]. 同济大学学报(社会科学版), 2014, 25(1): 111-118+124.
[2] 里米·芒迪. 翻译学导论: 理论与实践[M]. 李德凤, 译. 北京: 商务印书馆, 2007.
[3] 孙逸群, 周敏康. 国际新闻机辅翻译的编辑剖析[J]. 中国科技翻译, 2022, 35(2): 20-23.
[4] 张法连. 法律翻译中的机器翻译技术刍议[J]. 外语电化教学, 2020(1): 53-58.
[5] 崔启亮. 论机器翻译的译后编辑[J]. 中国翻译, 2014, 35(6): 68-73.
[6] 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.

[7] 李梅. 机器翻译译后编辑过程中原文对译员影响研究[J]. 外语教学, 2021, 42(4): 93-99.

[8] 车彤. 汉译日机器翻译质量评估及译后编辑策略研究[D]: [硕士学位论文]. 北京: 北京外国语大学, 2021.

[9] 冯全功, 李嘉伟. 新闻翻译的译后编辑模式研究[J]. 外语电化教学, 2016(6): 74-79.

[10] 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.

[11] 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, 161(6): 54-59.

[12] 陈胜, 田传茂. 在线翻译平台汉英翻译的问题及译后编辑——以石油地质文献为例[J]. 中国科技翻译, 2021, 34(1): 31-34+49.

[13] 杨玉婉. 神经机器翻译的译后编辑——以《潜艇水动力学》英汉互译为例[J]. 中国科技翻译, 2020, 33(4): 21-23+42.