

英汉机器翻译错误类型例析及其译后编辑策略

——以医学文本为例

方茹丹

陕西科技大学文化与教育学院, 陕西 西安

收稿日期: 2024年10月12日; 录用日期: 2024年11月11日; 发布日期: 2024年11月21日

摘要

随着全球医学领域的交流日益频繁, 准确的医学文本翻译需求逐渐增加。尽管神经机器翻译(NMT)技术的飞速发展在一定程度上提升了翻译效率, 但其在医学文本的翻译中仍存在较多错误, 影响译文的准确性。本文通过对英汉医学文本机器翻译错误的分类及其译后编辑策略进行研究, 重点从词汇、句法和语篇三个层面分析常见的翻译错误, 并提出相应的译后编辑策略, 以期为提高医学文本翻译质量提供参考。

关键词

机器翻译, 错误类型, 译后编辑, 医学文本

Analysis of Error Types in English-Chinese Machine Translation and Post-Editing Strategies

—Taking Medical Texts as an Example

Rudan Fang

School of Culture and Education, Shaanxi University of Science & Technology, Xi'an Shaanxi

Received: Oct. 12th, 2024; accepted: Nov. 11th, 2024; published: Nov. 21st, 2024

Abstract

With the increasing frequency of communication in the global medical field, the demand for accurate medical text translation is gradually increasing. Although the rapid development of neural machine translation (NMT) technology has improved translation efficiency to a certain extent, there are still many errors in the translation of medical texts, which affects the accuracy of the translation. This paper studies the classification of machine translation errors of English-Chinese medical texts

and their post-editing strategies, focusing on analyzing common translation errors from the three levels of vocabulary, syntax and discourse, and proposing corresponding post-editing strategies, in order to provide a reference for improving the quality of medical text translation.

Keywords

Machine Translation, Error Types, Post-Editing, Medical Text

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在全球化进程加速的背景下,医学领域的跨语言交流日益频繁,促使医学文本的翻译需求大幅增加。中英双语在医学研究和临床实践中的重要性,使得高质量的翻译成为保障医学信息准确传递的关键。然而,医学文本因其专业术语繁多、句法复杂以及文化特定性强,给机器翻译带来了巨大挑战。尽管神经机器翻译(NMT)技术的进步在一定程度上提升了译文的流畅度和自然度,但医学文本中的翻译错误依然频繁出现,甚至可能对实际应用造成严重后果。因此,如何有效识别并修正这些错误,成为机器翻译在医学领域应用中亟需解决的问题。

本研究旨在系统分析医学文本英汉机器翻译中的常见错误类型,并探索相应的译后编辑策略,以提升翻译质量和可靠性。本文的分析材料来源于作者在西安雅信诚医学翻译科技有限公司实习期间提供的译稿材料。这些材料由实习公司提供,并经过筛选以确保它们与本研究主题的相关性。研究通过对词汇、句法和篇章三个层面的翻译错误进行详细分类和案例分析,探讨这些错误的产生原因及其在译文中的表现形式,并提出针对性的译后编辑策略,以帮助译者更好地应对医学文本的翻译挑战。通过这项研究,期望为翻译理论的深入研究和实际译后编辑实践提供有益的指导,同时推动机器翻译技术在医学领域的更广泛应用。

2. 文献综述

2.1. 机器翻译与译后编辑

根据 ISO/DIS 17100:2013 标准的定义,机器翻译是“使用计算机系统将文本或语音从一种自然语言自动翻译为另一种语言”(MT, automated translation of text or speech from one natural language to another using a computer system)。译后编辑是“检查和修正机器翻译的输出”(to check and correct MT output) (ISO, 2014) [1]。传统意义上的译后编辑是对机器翻译输出的初始译文进行修改与加工的过程。Allen (2003)将译后编辑描述为“迄今为止与机器翻译最为相关的任务” [2]。Daniel Marcu 提出“译后编辑是指语言专家对机器翻译输出进行编辑以生成人工质量翻译的过程。译后编辑是将机器翻译和高水准专业人工译后编辑人员相结合,以生成达到发布标准的高质量翻译的过程”(SDL, 2013) [3]。DePalma (2013)将译后编辑定义为“通过人工评审,编辑,提高译文质量和适用性的过程。具体而言,机器翻译的译后编辑(MTPE)是通过人工和部分自动化方式增强机器翻译的输出,以满足特定质量目标的过程” [4]。

仁者见仁,智者见智。这些专家学者的核心要义即是译后编辑在完善机器翻译译文过程中起着至关重要的作用。因此,无论是译者个体还是翻译团队在今后的翻译过程中都应该重视译后编辑的有效使用。

2.2. 译后编辑的研究现状

医学文本翻译作为科技翻译中的一种,其特点就是准确、清晰、精炼、严密[5]。因此,对医学领域的翻译,不仅需要语法以及长难句的分析,就算结合常识推理,也不一定能够翻译出准确的译文,所以需要译者在机器翻译后进行修改润色,使句子通俗易懂。但同时,机器翻译无法结合医学翻译背景进行翻译,导致翻译内容逻辑错误、专业术语使用不准确,特别是在多义词的选择和词性的判断时,如果没有考虑语境和背景,很容易发生误译或错译。

而随着技术的不断升级,使得译后编辑在各技术时间段的基本概念存在细微差别,但核心内容没有改变,即译后编辑的实现需要机器翻译、计算机辅助翻译、翻译辅助工具等软件的支持才能生成高质量的译文。人机交互的翻译模式将成为今后翻译的主流模式,译后编辑(者)作为译后编辑过程中的重要桥梁,其能力培养工作也迫在眉睫[6]-[9]。其次是译后编辑错误类型分析以及修改编辑策略。通过案例论述机器翻译的主要错误类型,总结对应的修改策略,为译文修改指明方向,完善译后编辑工作[10]-[13]。

虽然译后编辑研究话题广泛,但当前译后编辑的主要任务还是完善借助翻译工具所生成的译文质量,提升译后编辑的专业性、可操作性,因此分析基于翻译辅助工具的译文错误类型则十分重要。尽管国内学者在该领域的研究有了一定的发展,但将译后编辑应用于医学领域的研究相对较少。因此,研究译后编辑在医学文本翻译中的应用十分有必要。此外,大多数学者都聚焦机器翻译错误,很少有学者探讨译后编辑策略。因此,本文对有道翻译的译文进行修改、加工,使译文更加简洁、连贯。

3. 机器翻译中的错误类型分析

机器翻译在医学文本翻译中常见的错误可以分为词汇、句法和语篇三个层面。下面将详细细化每一个层面的错误类型,并通过具体的例子进行说明,探讨这些错误背后的成因以及其译后编辑策略。

3.1. 词汇层面的翻译错误

词汇层面的错误是医学文本机器翻译中最常见的错误类型之一,主要表现为术语误译、多义词处理不当和词汇一致性问题。这类错误直接影响译文的准确性,尤其是在医学领域,任何词汇错误都可能导致严重后果。以下细分为几种具体类型:

3.1.1. 术语误译

医学文本中充斥着大量专业术语,这些术语在不同语言之间可能没有直接对应的翻译。因此,机器翻译往往在处理这些术语时出现问题,如选择了不合适的目标语言词汇或是错误理解了术语的语境。

例: Patients with myocardial infarction are at high risk of death.

机器翻译: 患有心脏病发作的患者死亡风险高。

译后编辑: 患有心肌梗死的患者死亡风险高。

分析: 机器翻译系统无法区分“心肌梗死”和“心脏病发作”之间的细微差别,造成术语的误译。机器翻译将“myocardial infarction”(心肌梗死)错误地译为“心脏病”。尽管两者相关,但“心脏病”是一个更广泛的概念,不能准确表达原文中的具体病症。译者需查阅专业医学词典,确保术语翻译的精准性,避免误导读者。

例: The patient was diagnosed with angina.

机器翻译: 患者被诊断为咽喉炎。

译后编辑：患者被诊断为心绞痛。

分析：“Angina”在医学中通常指“心绞痛”，但在机器翻译中可能被错误地译为“咽喉炎”（由于“angina”在非医学语境下也可以表示“喉痛”）。这种错误源于机器翻译未能正确识别术语的专业语境。

3.1.2. 多义词处理不当

多义词是指那些在不同语境下具有不同含义的词汇。在医学文本中，某些术语在普通语境中有其他含义，因此机器翻译可能会选择错误的翻译。机器翻译在处理同义词时，未能准确匹配最合适的词汇，导致不符合语境的词义选择。例如，“administration”在医疗场景中常指“给药”或“药物管理”，但机器翻译可能会选择为“管理”。

例：The administration of insulin must be carefully monitored.

机器翻译：胰岛素的管理必须仔细监控。

译后编辑：胰岛素的给药必须仔细监控。

分析：机器翻译系统未能识别“administration”在医学场景中的特定含义，选择了通用含义。译者需根据语境判断同义词的适用性。

例：The examination revealed a mass in the patient's lung.

机器翻译：检查发现患者肺部有一块肿块。

译后编辑：检查发现患者肺部有一块肿瘤。

分析：在此例中，“mass”在医学语境中指“肿块”或“肿瘤”，但可能被错误地译为“块”或其他不准确的词汇。这种错误反映了机器翻译在多义词处理上的不足。

3.1.3. 词汇一致性问题

机器翻译在处理重复出现的术语时，可能会采用不同的翻译方式，导致词汇不一致，影响专业文本的准确性和连贯性。例如，“antibiotic”在一篇文章中可能被翻译为“抗菌素”和“抗生素”两种不同的表达方式。

例：Antibiotics are used to treat bacterial infections. The overuse of antibiotics can lead to resistance.

机器翻译：抗菌素用于治疗细菌感染。抗生素的过度使用可能导致抗药性。

译后编辑：抗生素用于治疗细菌感染。抗生素的过度使用可能导致抗药性。

分析：机器翻译系统未能保证术语的一致性，造成同一术语不同译法。译者应统一术语翻译，避免混淆。

3.2. 句法层面的翻译错误

句法层面的错误主要出现在长句处理不当、被动句误译和从句结构紊乱等方面。机器翻译系统在处理复杂句子时，往往难以保持句法结构的完整和正确性。

3.2.1. 长句处理不当

医学文本中的长句较为常见，机器翻译系统在翻译长句时容易打乱句子结构，导致译文不连贯或信息丢失。例如，长句中包含多个从句，机器翻译系统可能将其拆解为多个短句，或者丢失主句与从句之

间的逻辑关系。

例：The patient, who was diagnosed with diabetes and hypertension, was prescribed insulin and a low-sodium diet to manage his conditions.

机器翻译：患者，被诊断为糖尿病和高血压，使用胰岛素和低钠饮食来管理他的病情。

译后编辑：该患者被诊断患有糖尿病和高血压，医生为其开了胰岛素并建议采用低钠饮食来管理病情。

分析：机器翻译将句子分割为多个短句，导致信息表达不清，且主从句关系混乱。译者应根据汉语的表达习惯对长句进行重组，使信息更加连贯。

3.2.2. 被动句误译

英语和汉语在使用被动语态的频率和方式上存在显著差异，若未能正确转换，可能导致译文在语法上不符合目标语言的规则或风格。而被动语态在英语医学文本中经常使用，但汉语较少使用被动句，因此机器翻译系统常常直接将英语的被动句结构照搬到中文中，导致译文不自然。例如，“The treatment was administered by the doctor”可能被直接翻译为“治疗是由医生给予的”。

例：The medication was prescribed by the physician.

机器翻译：药物是由医生开的。

译后编辑：医生开了药物。

分析：机器翻译未能处理好英语被动句在汉语中的转换，译文显得生硬。译者应根据汉语的表达习惯，将被动句转化为主动句。

3.2.3. 从句结构紊乱

英语中经常使用从句来表达复杂信息，而汉语则倾向于用较为简单的句式或分句来表达。这种句法差异可能导致机器翻译在处理从句时产生问题，如从句被忽略、错误简化或未能正确分离。所以英语中的从句结构复杂，包含多层嵌套从句时，机器翻译系统难以正确处理，常出现从句与主句脱节或逻辑关系混乱的问题。

例：The patient, who had been suffering from chronic fatigue, was found to have a vitamin D deficiency, which explained his symptoms.

机器翻译：患者，谁患有慢性疲劳，被发现患有维生素 D 缺乏，这解释了他的症状。

译后编辑：该患者长期疲劳，被发现是由于维生素 D 缺乏，这解释了他的症状。

分析：机器翻译系统在处理从句时，句法关系混乱，且从句指代不清。译者应理顺主从句的逻辑关系，并调整从句表达方式。

例：The doctor, who had been treating the patient for years, suggested a new treatment plan.

机器翻译：一直在治疗这位病人的医生，建议了一个新的治疗方案。

译后编辑：这位医生一直在治疗该病人，建议了一种新的治疗方案。

分析：机器翻译未能正确处理定语从句“who had been treating the patient for years”，导致译文结构不自然，信息表达不清晰。正确译文应将定语从句独立出来。

3.3. 语篇层面的翻译错误

语篇层面的错误主要涉及文本的连贯性、逻辑衔接不当以及上下文指代混乱等问题。由于机器翻译系统一般逐句处理文本，难以处理整个语篇中的信息连贯性和逻辑关系。

3.3.1. 逻辑衔接不当

医学文本通常包含复杂的逻辑关系，如因果关系、对比关系等。机器翻译系统在处理这些关系时，容易错用连接词或忽略逻辑关系，导致译文逻辑不清。例如，“therefore”和“however”的错误使用常见于机器翻译中。

例：The patient did not respond to the first treatment. Therefore, a new treatment plan was introduced.

机器翻译：患者没有对第一次治疗产生反应。然而，采用了新的治疗计划。

译后编辑：患者没有对第一次治疗产生反应。因此，采用了新的治疗计划。

分析：机器翻译将因果关系误译为对比关系，使用了不当的连接词。译者应根据上下文调整连接词，确保逻辑关系准确。

3.3.2. 上下文指代不明

在长篇医学文本中，指代词(如“it”、“this”等)在机器翻译中经常被误译，导致上下文衔接不明或指代对象不清。例如，指代词在医学文本中的具体含义可能随上下文变化，机器翻译系统难以准确判断其指代对象。

例：The patient's condition worsened after the surgery, which led to additional complications.

机器翻译：患者的情况在手术后恶化，这导致了额外的并发症。

译后编辑：患者的病情在手术后恶化，进而引发了额外的并发症。

分析：机器翻译未能准确指代“this”所指的“患者情况的恶化”。译者应仔细检查指代词的前后文，确保指代清晰。

3.3.3. 信息丢失

机器翻译在处理语篇时，有时会因为复杂句式或篇章长度过长而丢失部分信息，导致译文不完整。例如，机器翻译可能会忽略句子中的修饰部分或从句，影响信息的完整性。

例：After several tests, which included blood work and MRI scans, the doctor concluded that the patient needed immediate surgery.

机器翻译：经过几次检查，医生得出结论，患者需要立即手术。

译后编辑：经过几次检查，包括血液检查和MRI扫描，医生得出结论，患者需要立即手术。

分析：机器翻译忽略了句中对“检查”的详细描述，导致信息丢失。译者应补充完整的从句信息，确保信息的全面性。

4. 译后编辑策略

通过分析词汇、句法和语篇三个层面的机器翻译错误，可以看出，机器翻译在医学文本中的表现仍有许多局限性，尤其在处理专业术语、复杂句法结构和上下文逻辑关系时容易出错。译后编辑策略在确

保译文的准确性和连贯性方面发挥了关键作用，未来应结合更多专业领域的翻译需求，优化机器翻译算法，提升其在医学文本翻译中的表现。

针对上述三个层面的错误，译后编辑策略应包括以下几个方面：

1) 词汇层面：应通过对专业术语进行一致性检查，确保所有术语的翻译符合医学领域的使用标准。同时，可以借助术语库和翻译记忆库来减少术语误译的发生，加强专业术语库的建设和使用，提高术语翻译的准确性。

2) 句法层面：译者应根据目标语言的语法规则，对长句和复杂句进行重新结构化，以确保句子主次信息表达清晰，并注意修正错误的句法结构。

3) 语篇层面：译者应特别关注上下文的一致性和逻辑性，确保译文在语篇层面的连贯与流畅，必要时调整指代词和连接词，以增强译文的整体衔接性。

5. 结论

本文深入探讨了英汉医学文本机器翻译中的错误分类及其译后编辑策略。作者通过细致的分析，从词汇、句法和语篇三个核心层面，系统地揭示了机器翻译在医学文本中常见的错误类型。在词汇层面，错误主要包括术语误译、多义词处理不当和词汇一致性问题等；句法层面则涉及长句处理不当、被动句误译和从句结构紊乱等问题；语篇层面则主要关注文本的连贯性、逻辑衔接不当以及上下文指代混乱方面的不足。针对这些错误，作者提出了相应的译后编辑策略。在词汇层面，建议加强专业术语库的建设和使用，提高术语翻译的准确性；在句法层面，强调对机器翻译输出的句子结构进行细致检查，确保句子主次信息表达清晰；在语篇层面，则注重提升文本的连贯性和一致性，使翻译结果更加符合医学文本的专业性和严谨性。

本文首次从词汇、句法和语篇三个层面，对英汉医学文本机器翻译中的错误进行了全面而系统的分析，为后续的译后编辑工作提供了清晰的理论框架。针对分析出的错误类型，作者提出了具体且针对性的译后编辑策略，这些策略不仅具有理论价值，更具有较强的实践指导意义。本文聚焦于医学文本这一特定领域，从专业角度出发，深入剖析了机器翻译在医学翻译中的局限性，为提升医学文本翻译质量提供了有力支持。

虽然本文选取了一定数量的医学文本作为研究样本，但样本数量相对有限，可能无法全面反映机器翻译在医学文本中的所有错误类型；在错误分类和译后编辑策略的制定过程中，不可避免地存在一定的主观性，这可能对研究结果的客观性和准确性产生一定影响；本文虽然提出了相应的译后编辑策略，但并未对这些策略在实际翻译工作中的效果进行充分的实践验证，但所提出的策略仍具有较强的理论价值和实践指导意义。

希望未来研究可以引入量化分析方法，对错误类型进行更精确的统计和分类，以提高研究结果的客观性和准确性；将提出的译后编辑策略应用于实际翻译工作中，通过实践验证其效果，并根据反馈进行策略的优化和调整；加强与其他学科的合作与交流，如计算机科学、语言学等，共同探索提升机器翻译在医学文本中准确性和流畅性的新方法和技术。

参考文献

- [1] ISO TC 37. ISO 18587 (2014) Translation Services—Post-Editing of Machine Translation Output—Requirements.
- [2] Allen, J. (2003) Post-Editing. In: Somers, H. Ed., *Computers and Translation: A Translator's Guide*. John Benjamins Publishing.
- [3] SDL. 塑造内容翻译的未来——机器翻译和译后编辑简介[EB/OL]. 2013-03-29. <https://dblp.uni-trier.de/pid/05/5326.html>, 2014-02-21.

- [4] DePalma, D.A. (2013) Post-Edited Machine Translation Define. <https://dblp.uni-trier.de/pid/40/3149.html>
- [5] 林娜, 唐跃勤, 黎斌. 谈受限汉语在机器翻译中的运用: 以中文简历英译为例[J]. 西南交通大学学报: 社会科学版, 2012, 13(5): 49-53.
- [6] 崔启亮. 论机器翻译的译后编辑[J]. 中国翻译, 2014, 35(6): 68-73.
- [7] 魏长宏, 张春柏. 机器翻译的译后编辑[J]. 中国科技翻译, 2007, 20(3): 22-24, 9.
- [8] 冯全功, 张慧玉. 全球语言服务行业背景下译后编辑者培养研究[J]. 外语界, 2015(1): 65-72.
- [9] 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.
- [10] 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.
- [11] 李梅, 朱锡明. 英汉机译错误分类及数据统计分析[J]. 上海理工大学学报: 社会科学版, 2013, 35(3): 201-207.
- [12] 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.
- [13] 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.