

新闻语体汉维机器翻译译文质量评测

王文娟, 司马义·阿不都热依木

新疆大学中国语言文学学院, 新疆 乌鲁木齐

收稿日期: 2026年6月19日; 录用日期: 2026年7月24日; 发布日期: 2026年8月6日

摘要

近年来, 机器翻译利用人工智能技术和专业人工翻译经验有机结合, 在跨语言信息传递领域实现了快速发展与广泛应用。为了测评不同翻译软件在新闻语体汉维互译中的译文质量, 帮助新闻领域的译者更有效地利用翻译系统, 从而提升汉维译文质量, 本文选取了必应翻译、izdax翻译、国语助手、DeepSeek-V3翻译、谷歌翻译以及小牛翻译六款机器翻译软件作为评测对象。评测以新闻语体为测试材料, 涵盖了汉语译维吾尔语与维吾尔语译汉语的双向翻译任务。评测方法采用人工打分, 依据忠实度和流畅度两项关键指标分别进行评分, 并通过汇总计算得出各软件的综合平均得分。此外, 研究还结合了笔者实际使用体验的反馈, 并对新闻标题这一特定语体进行了专项测评。基于以上评测结果, 本文从核心功能优化、操作体验提升、付费模式与功能匹配调整三个主要方面提出了改进建议, 旨在为新闻语体译者在MTPE (Machine Translation Post-Editing) 工作流下选择合适的翻译工具提供参考, 同时也为汉维机器翻译系统的针对性优化提供实证依据。

关键词

新闻语体, 汉维机器翻译, 人工测评, 数据分析

Quality Evaluation of News Stylistics Chinese-Uyghur Machine Translation Outputs

Wenjuan Wang, Simayi·Abdureyim

School of Chinese Language and Literature, Xinjiang University, Urumqi Xinjiang

Received: June 19, 2026; accepted: July 24, 2026; published: August 6, 2026

Abstract

In recent years, machine translation, enabled by the coupling of artificial intelligence technology and

professional human translation techniques, has achieved rapid development and wide application in the field of cross-language information transmission. To evaluate the translation quality of different translation software in Chinese-Uyghur mutual translation of journalistic texts and assist translators in the journalism sector in utilizing translation systems more effectively to improve the quality of Chinese-Uyghur translations, this paper selects six machine translation tools, namely Bing Translator, izdax Translator, Guoyu Zhushou, DeepSeek-V3 Translator, Google Translator, and Xiaoniu Translation, as evaluation objects. The evaluation adopts journalistic texts as test materials, covering two-way translation tasks: Chinese to Uyghur and Uyghur to Chinese. Human rating is employed as the evaluation method, with scoring conducted separately based on two key indicators, namely fidelity and fluency, and the overall average score of each software is obtained through aggregated calculation. In addition, this study incorporates feedback from the author's practical usage experience and conducts a special evaluation on the specific genre of news headlines. Based on the above evaluation results, this paper puts forward improvement suggestions from three main aspects: optimization of core functions, improvement of operation experience, and adjustment of the matching between paid modes and functions. The study aims to provide a reference for translators of journalistic texts to select appropriate translation tools under the MTPE workflow, and also offer empirical evidence for the targeted optimization of Chinese-Uyghur machine translation systems.

Keywords

News Stylistics, Chinese-Uyghur Machine Translation, Human Evaluation, Data Analysis

Copyright © 2026 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

随着人工智能和自然语言处理技术的快速进步, 机器翻译在跨语言交流、多民族文化传播中的作用越来越重要。汉维机器翻译连接着汉语和维吾尔语, 对新疆乃至中亚地区的经济文化交流来说, 是不可或缺的桥梁。由于汉语与维吾尔语分属不同语系(维吾尔语属于阿尔泰语系, 汉语属于汉藏语系), 二者在形态变化、句法结构、语法规则等层面存在显著差异[1], 语言类型学隔阂较为突出, 也正因如此, 针对汉维双向机器翻译的系统性研究, 兼具重要的理论研究价值与现实经济社会价值。新闻语体是信息传播的核心载体, 翻译质量好不好, 直接关系到信息能不能准确、及时地传递出去。目前, 现有研究在汉英机器翻译译文质量评测方面已有所涉及[2], 但是针对汉维机器翻译在新闻语体方面的测评研究, 目前尚属空白。

从发展背景来讲, 汉维机器翻译的研究与应用, 是在中国多民族语言政策的推动、边疆地区信息化建设需求的双重作用下开展的。特别是“一带一路”倡议持续推进后, 新疆等地与其他省区、中亚国家的往来越来越频繁, 高质量的汉维翻译成了区域合作里的关键环节。新疆地区作为我国国际贸易的活跃口岸, 机器翻译模型的应用为维吾尔族人民提供了更好的汉语及其他语言的学习工具, 进一步促进不同地区和民族之间的文化交流和贸易往来[3]。不过, 汉维机器翻译也面临着现实难题: 一方面相关数据储备不足, 另一方面汉语和维吾尔语在语法结构、词汇表达、文化背景上差异显著, 这些都让机器处理新闻文本时面临不少挑战。本研究旨在通过科学评测, 识别必应翻译、izdax 翻译、国语助手、DeepSeek-V3 翻译、谷歌翻译、小牛翻译等汉维机器翻译系统在新闻语体中的优缺点, 为技术优化和译后编辑工作提供科学依据, 推动跨语言信息服务均衡发展, 同时为新闻语体翻译使用者提供更好地选择。

2. 研究方法

2.1. 语料介绍

本研究的对象为新闻双语文本。汉语新闻文本来自《新疆日报》2025 年 5 月 15~17 日, 第 26972 号今日 4 版。译文新闻文本来自《Jindagan geziti》0-2025519-期, 该文本包含“领导访问, 政治经济, 艺术文化, 科技新闻”等方面的新闻。其中汉文和维文各选取 445 个句子, 汉文字数为 21,525 字, 与汉文对应的维文字数为 19,875 字。

2.2. 研究过程

(1) Tmxmall 在线对齐

将原文与译文导入 Tmxmall 在线对齐平台, 完成新闻语料逐句对应处理, 建立句对句的平行语料库。此举确保语料严格对齐, 便于后续机器翻译的评估与性能测试工作高效开展。

(2) imla 校对软件

首先将对齐后的新闻语料导入 imla 校对工具, 启动系统内置的拼写检查模块对双语文本进行全面扫描。随后根据软件标记的可疑词汇列表, 逐项核对汉语原文与维吾尔译文的拼写规范。对于识别出的拼写错误, 需参照标准词典进行修正, 重点关注专有名词的转写一致性和术语拼写准确性。完成修正后需执行二次校验, 确保所有修正内容符合语言规范并保持原文语义的完整传递, 保证用于机器翻译测评的语料没有任何翻译错误, 拼写错误和标点符号错误等。最后导出校对完成的语料文件, 并在日志中记录拼写修订清单以备核查。

(3) 机器翻译软件

本研究对必应翻译、izdax 翻译、国语助手、DeepSeek-V3 翻译、谷歌翻译及小牛翻译六大机器翻译系统进行了双向翻译质量评估。具体评估方向包括汉译维与维译汉。通过输入统一测试文本, 分别获取各系统在两个翻译方向上的输出结果。所有机器翻译文本均被完整记录, 以备后续进行定量与定性分析。测试过程确保了输入文本的一致性及输出结果的原始性, 旨在客观比较不同系统在汉维互译任务上的实际表现。

(4) 评分表进行评测打分

评分表 rating scale		
五分制档位	忠实度(信息还原准确性)	流畅度(译文可读性)
0 分	完全没有将原文内容翻译出来, 无任何原文信息保留	译文完全不可理解, 无法获取任何有效表意
1 分	仅原文个别单词被孤立翻译, 句子整体语义完全缺失、偏差极大	译文晦涩难懂、不知所云; 只有个别短语、大于单词的语法成分能够勉强理解
2 分	仅原文少数短语、大于单词的语法成分被翻译, 核心语义大量缺失	译文部分内容基本流畅, 仅少量短语、语法成分表意通顺, 整体可读性差; 关键单词孤立译出, 仅对人工编辑有少量用处
3 分	原文六成的概念、概念间逻辑关系被正确翻译; 主谓宾等句子核心主干关系准确	译文大部分内容基本流畅, 大致传递原文整体意思, 仅存在局部出入; 一般需要对照原文修正错误, 不看原文也可大致猜测译文含义
4 分	原文八成的概念、概念间逻辑关系被正确翻译, 核心信息完整, 仅存在少量细节偏差	译文绝大多数内容基本流畅, 主谓宾主干通顺; 无需对照原文即可完整读懂译文; 仅存在少量词形、语序、多义词选择、搭配、得体性问题, 无需参照原文即可轻松修改完善
5 分	原文全部概念、内在逻辑关系精准正确翻译, 无漏译、误译、增译	译文语句自然流畅, 精准完整传递原文全部信息; 语法结构完全正确; 仅存在个别错别字、单复数、表达地道性等微小瑕疵, 仅需极小修改, 甚至无需任何修改

评分法是评价人工译文的一种常见方法, 但同样适用于机器翻译质量评估[4], 且人工评测指标一直以来就是机器翻译质量评估的重要组成部分[5]。基于此, 本次评估依据翻译忠实度(信息还原准确性)、流畅度(译文可读性)两项维度, 采用五分制对平行语料库中的前 200 句(1 万字左右)的翻译句子逐一进行人工评分。人工评测指标一直以来就是机器翻译质量评估的重要组成部分。每句翻译由 4 人进行打分, 评分组由汉语母语者和维吾尔语母语者构成独立的评分, 最终汇总分析整体质量表现。评分标准如下: 5 分为优秀, 4 分为良好, 3 分为合格, 2 分为较差, 1 分为差, 0 分为无效。评分结果将用于全面评估翻译质量的准确性与自然度。

综合打分补充判定规则:

底线原则: 忠实度为核心基础维度, 若忠实度 ≤ 2 分, 流畅度再高, 综合最终得分最高不得超过 2 分;

扣分规则: 忠实度达标, 但译文逻辑混乱、卡顿严重、可读性差, 可在档位基础上扣 0.1~1.0 分;

加分规则: 忠实度满分(5 分)、流畅度 ≥ 4 分, 可直接判定为优质译文, 综合基准分不低于 4.5 分。

3. 评测结果分析

3.1. 汉维双向翻译整体分析

根据本次翻译测评的人工打分平均结果可以发现, 机器翻译在维吾尔语译为汉语方向的表现整体较为理想, 翻译准确性较高; 而在汉语译为维吾尔语方向, 不同工具的翻译结果呈现出明显差距, 整体翻译质量显著低于维译汉方向, 说明汉译维任务对机器翻译而言难度更高, 仍存在较大的改进空间。但是这一现象并非仅由任务本身的技术难度导致, 而是多重因素共同作用的结果:

(1) 语料资源的分布存在天然失衡。维译汉方向上, 维吾尔语使用者为了获取信息, 大量的维吾尔语语料会被翻译成汉语, 形成了规模大、质量高的平行语料库, 为模型训练提供了充足的学习数据; 而汉译维方向上, 汉语是主流语言, 语料基数本身极大, 但主动翻译成维吾尔语的高质量文本相对较少, 公开、高质量的汉译维平行语料库规模远小于维译汉方向, 模型缺乏足够的高质量训练数据。

(2) 语言类型差异带来的翻译复杂度不对称。维吾尔语作为黏着语, 依赖词缀变化表达语法信息, 而汉语作为孤立语, 以语序和虚词为核心语法手段。维译汉时, 模型只需将复杂的词缀信息转化为汉语的语序表达; 而汉译维时, 模型需要精准识别汉语句子中隐含的时态、语态、人称等信息, 并生成符合维吾尔语语法规则的词缀形式, 极易出现形态搭配错误。

(3) 模型优化的场景侧重存在差异。多数翻译工具的核心用户场景以维译汉为主, 厂商在该方向的优化投入、迭代频率更高; 而汉译维方向用户基数较小, 缺乏持续的调优与反馈迭代, 导致模型适配性不足, 不同工具间的表现差距被放大。

(4) 文化与术语翻译的双向不对称。大量维吾尔语文学、民俗、地域特色词汇, 在汉语中都有对应的通用译法, 甚至已经形成了固定翻译规范, 维译汉时直接套用即可, 准确性高; 而汉语中的现代科技、网络热词、政治文化术语等缺乏统一的维吾尔语译法, 汉译维时容易出现生硬直译、表达不符合母语习惯的问题, 直接影响翻译质量与人工评分。

3.2. 单向翻译具体分析

(1) 在汉译维方向

国语助手的平均得分为 4.56 分(见表 1), 在所有平台中表现最佳。该平台在翻译新闻类专有名词方面能力较强, 但在人名、地名翻译中出错率较高。此外, 个别专有名词无法输出, 例如将“漫画”译为乱码“#”。在句式翻译方面也有问题, 如将“一抽一放, 电站与不听话的风光资源形成互补。”译为“bir

tartif, bir qojup berif arqiliq, elektr istansisi bilen «gep aqlimajdikan» menzire bajliqi bir-birini toluqlap” 其中“风光资源”被误译为“menzire bajliqi”, 意为“景观资源”未能准确传达原意。总体来看, 国语助手在政务、新闻等规范文本上的翻译效果优于其它平台, 但专有名词的处理仍有待改进。

izdax 翻译软件的平均得分为 4.5 分(见表 1), 该平台在翻译较长句子时表现精准, 在翻译人名、地名时存在与国语助手相似的问题, 出错率较高, 整体表现略低于国语助手, 但仍处于较好水平。具体错误案例如下: 将“他生前长期致力于推动中乌关系发展”译为“u fiijat waqtida uzaq muddet tyrtke boluf dzuongo-ukraina munasiwitiñ tereqqijatıa kyf fıqirit” 其中“中乌”译为“dzuongo-ukraina”意为“中国与乌克兰”, 通常“中乌关系”多指中国-乌兹别克斯坦关系; 也可指中国与乌克兰、乌拉圭、乌干达等国关系, 但根据上下文可知原文是指“中国与乌拉圭”, 属于国名误译。其中“tyrtke boluf(推动)”一词与“tereqqijat(发展)”为固定搭配。又如, 将“进一步激励广大残疾人自尊、自信、自强、自立”译为“mejiplarni özini fiormetleñ, öziñe ifiniñ, özini kyfıejtiñ, öziñe tajiniñqa jenimu riñbetlendyrıp” 其中“进一步激励广大”被漏译, 导致句子语义不完整。

小牛翻译的平均分为 4.225 分(见表 1), 其在翻译人名与地名时与国语助手相似, 翻译不出的词汇为乱码代替。此外, 个别词汇的翻译过于生硬, 例如将“他的逝世使中国人民失去了一位老朋友、好朋友。”翻译为“u wapat bolup, dzuongo xelqi bir qedinas dosti, jaxfi dostidin ajrilip qaldi.” 其中的“老朋友”直译为“kona dost”, 缺乏自然表达, 影响了译文的整体可读性。

必应翻译的平均得分为 4.185 分(见表 1), 该平台在翻译一般语句时表现尚可, 但在专有名词处理上出错率较高, 例如对“中华”等术语的翻译不够准确。同时, 翻译过程中出现漏译日期等现象, 影响了信息的完整传递。在语义表达方面 also 存在问题, 例如将“努力做残疾人信得过、靠得住的知心人、贴心人。”译为“mejiplar ifenflık, ifenflık sirdaf bolufi kerek”, 其中“信得过”与“靠得住”两个词汇被混译为一词, “知心人”与“贴心人”也被混译为一词, 未能准确区分原文中的语义层次, 导致翻译不准确。

谷歌翻译的平均分为 3.538 分(见表 1), 在各平台中表现较差。该平台在翻译过程中存在大量词汇使用不当的问题。例如: 将“新华社北京 5 月 16 日电”译为“xelq tori ujxurfe qaniliniñ 16-maj bejdıñdin bergen telegrammisi”, 其中“新华社”译为“xelq tori ujxurfe qaniliniñ”意为“人民网维吾尔语频道的”与原词意思完全相悖。又如, 将“在第三十五个全国助残日到来之际作出重要指示强调。”译为“35-nöwetlik mejiplar bajrimi munasiwiti bilen muhim joljoruq berdi we tıkitlidi” 其中“全国助残日”译为“nöwetlik mejiplar bajrimi”, 意为“残疾人节”, 偏离了该节日“扶助、关爱残疾人”的核心语义, 属于概念性误译。

DeepSeek-V3 仅得 2.815 分(见表 1), 在所有平台中表现最差, 翻译结果与原文的契合度最低。该平台在简单句的翻译中尚能顺畅输出, 但在处理略复杂的句子时, 出现大量的人名、地名错译, 词汇漏译, 语法错误, 语序混乱等问题。例如, 将“这已不是艾德莱斯第一次走出国门, 在世界舞台亮相。”译为“bu etrep nusxısiniñ dunja sefiniside namajan bolufidiki tundzi qetim emes”, 其中存在词汇翻译错误和语序混乱的问题。

综上所述, 国语助手的表现最为出色, 其作为面向国内少数民族用户的本地化翻译平台, 具备天然的语料与场景优势, 其训练数据包含大量规范的政务、新闻、民生类汉维双语文本, 同时针对汉译维方向进行了专门的母语者反馈优化与术语库建设, 对维吾尔语的语法规则、本土表达习惯适配度更高, 因此翻译结果更贴合原文语义与母语者阅读习惯。

而 DeepSeek-V3 作为通用大模型, 在汉译维翻译中存在明显短板。一方面, 网络公开的维吾尔语语料本身较少, 高质量汉-维平行语料更是稀缺, 模型缺乏足够的训练样本; 另一方面, 通用大模型的优化重点集中在汉语语言的通用任务上, 未针对维吾尔语的黏着语特性、国内本地化场景进行专项调优, 难以精准处理汉译维中复杂的词缀生成与语法适配问题, 导致翻译结果与原文契合度较低。

(2) 在维译汉方向上

在维译汉翻译任务中, 各平台的评估得分与表现同样存在差异, 具体分析如下。

izdax 翻译软件的平均得分为 4.71 分(见表 1), 在各平台中表现最佳。该平台的翻译结果与原文最为接近, 语句流畅, 意思吻合, 但在部分涉及日期的句子翻译中会出现识别错误的情况。例如, 将“*finxua agentliqi, bejdziñ, 5-ajniñ 16-kyni telegrammisi 7-qetimliq memliketlik özini kyfjetif nemuniñiliri we mejiplerge jardem berif ilkarlirini teqdirlif jiwini 5-ajniñ 16-kyni bejdziñda eñildi*”译为“新华社北京 5 月 16 日电第七次全国自强模范暨助残先进表彰大会 15 日在京召开”, 其中原文中的日期为“16 日”, 但译文错误地译为“15 日”, 属于日期识别与转换错误。

国语助手的平均分为 4.69 分(见表 1), 该平台整体翻译流畅度较好, 但仍存在不足之处。例如, 将“*«tfoñ dölet quruluñidin findzanğa nezer» pytyn taratquluq xewer qiliñ kollekti tiñanñan itkirisig kelip, bu qaltis «elektir torini teñfigytñi»ni zijaret qildi.*”译为“‘大国工程看新疆’全媒体报道团队来到天山深处, 采访这位出色的电网调度。”其中将“*elektir torini teñfigytñi*”意为“电网调节器”译为“电网调度”, 这一概念翻译不准确。

小牛翻译软件的平均分为 4.608 分(见表 1), 该平台翻译基本流畅, 但有部分专有名词翻译错误, 且翻译过于口语化。例如, 将“*bu, etlesniñ tundzi qetim dölet derwazisidin fñiqip, dunja sefnjside amajan qiliniñ emes*”译为“这, 并不是阿特拉斯第一次走出国门, 展现在世界舞台上”, 其中专有名词“艾德莱斯”被误译为“阿特拉斯”, 整句翻译过于口语化, 不符合新闻语体所需的专业性。

必应翻译软件的平均分为 4.258 分(见表 1), 该平台在翻译涉及数字类句子时, 出现内容杂糅的问题。例如, 将“*newette, aptonom rajñe bojiñe dñemij 5956 tyrylyk fier deridzilik xejrij maddij medenijet mirasi wekillik tyri we 6992 neper fier deridzilik wekillik warisi bar*”译为“目前, 全区共有各级各类物质文化遗产代表性项目 6992 名”, 该译文存在核心概念错译、关键数据颠倒、信息要素缺失三大严重问题: 将“物质文化遗产”误译为对立概念“物质文化遗产”, 项目数与传承人数完全互换, 语义残缺、逻辑混乱, 未遵循翻译忠实性原则, 属于严重不合格的误译。又如, 将“*fierqajsi jeza bazarlardin tallanğan xelq iñi senetkarliri bu jerde muqam nomurlirini repetitsije qilidu we orunlajdu*”译为“从各乡镇挑选的民间艺术家在这里排练和表演新疆民间艺术家。”该译文不连贯, 翻译出的文字格式有空格, 存在多处明显翻译失误: 关键专有信息严重漏译, 原文核心内容“木卡姆节目”未译出, 丢失句子的语义重心; 随意增译冗余语义, 原文并未提及“新疆民间艺术家”相关表述, 译文擅自添加额外内容, 违背翻译忠实性; 句式表达生硬破碎, 不符合规范书面语体。

谷歌翻译软件的平均分为 4.25 分(见表 1), 该平台在专有名词的翻译上存在失误, 对句子理解不到位。例如, 将“*koryñyñke qatnañti hem jikinka qatnañti.*”译为“出席会议并参加发布会”。公务活动概念错换, 将“参加会见、出席大会”译为“出席会议、参加发布会”, 歪曲了活动类型。

DeepSeek-V3 得分为 4.123 分(见表 1), 该平台在翻译中存在擅自增译信息。例如将“*bu jerde, qajtma su pompisi su turbiniliq generatorlar guruppisidin tñi juqiri syrette ajliniwatatti*”, 原意为“这里, 4 台可逆式水泵水轮发电机组正在高速运转”译为“这里, 可逆式水泵水轮发电机组正以每分钟 500 转的高速运转”, 其中“每分钟 500 转”为原文未提及的细节信息, 属于无依据的增译。此外, 该平台在翻译过程中出现无法识别翻译的情况, 并作出“你好, 这个问题我暂时无法回答, 让我们换个话题再聊聊吧”的回应。DeepSeek-V3 与其他翻译软件存在一定的差距, 但这一差距并非意味着其整体翻译能力较差, 而是由模型功能定位与训练语料的领域适配性不同所致。

总体来说, izdax 作为专注于维吾尔语的专业翻译平台, 长期积累了大量新闻、媒体类规范维汉平行语料, 对新闻语体、专有名词、句式结构的适配性更强, 翻译准确性更高。而 DeepSeek-V3 属于通用大

模型, 训练数据以通用互联网文本为主, 缺乏针对新闻领域的专项语料与优化, 在处理正式、严谨的新闻文本时更容易出现语义偏移或细节缺失, 因此与原文贴合度略低于其他的专业翻译工具。

Table 1. Average scores of manual rating results
表 1. 人工打分结果平均分

	汉译维	维译汉
必应翻译	4.185	4.258
izdax 翻译	4.5	4.71
国语助手	4.56	4.69
DeepSeek-V3	2.815	4.123
谷歌翻译	3.538	4.25
小牛翻译	4.225	4.608

3.3. 综合翻译能力评估

根据“表 2”所示各翻译软件综合翻译平均分, 在整合新闻语料维汉双向 200 句翻译样本、完成加权平均计算后可见: 国语助手综合得分达 4.63 分(见表 2), 位列六款参与测评软件的第一名, 在新闻语料维汉互译场景中综合翻译实力整体领先; izdax 翻译(4.61 分)、小牛翻译(4.42 分)、必应翻译(4.23 分)、谷歌翻译(3.90 分)综合表现依次居中; DeepSeek-V3 综合得分仅为 3.47 分, 为本次测评中的最低分数, 综合翻译水平整体偏弱。

从综合排名可以发现, 专门面向维汉双语翻译的软件普遍得分更高, 国语助手凭借对维语本土语境、新闻语体表达习惯、双向语义精准传递的深度优化, 实现了稳定且高质量的双向翻译输出; 而 DeepSeek-V3 作为通用人工智能模型, 在小语种、少数民族语言的语料积累、语体适配、地道表达还原上存在明显短板, 因此整体综合评分垫底, 翻译效果稳定性与专业适配度不足。

Table 2. Comprehensive translation scores of various translation software
表 2. 各翻译软件综合翻译得分

翻译软件	综合翻译平均分
必应翻译	4.23
izdax 翻译	4.61
国语助手	4.63
DeepSeek-V3	3.47
谷歌翻译	3.9
小牛翻译	4.42

3.4. 新闻小标题翻译表现

如“表 3”所示, 在新闻标题的维汉双向翻译评测中, 六款主流翻译软件的表现存在显著差异。在汉译维方向上, izdax 翻译、DeepSeek-V3 翻译、谷歌翻译及小牛翻译的综合评分普遍偏低; 而在维译汉方向上, 各款软件的翻译效果相对较好。其中, 国语助手翻译在维汉双向翻译中均获得最高分, 表现突出。

进一步分析低分组软件的译文发现, 其在专有名词和语义传达方面存在明显偏差。例如, 在汉译维

中必应翻译将汉语新闻标题“天山深处有个超级充电宝”译为“tjanfanniq iŋkrdʒidin taŋqiri zeretligyŋbar”。该译文中,“天山”这一专有名词采用了音意结合法,译为“tjanfanniq”译法不够准确;而“充电宝”被译为“zeretligyŋ”该词实为“充电器”之义,产生了语义偏差。类似地,谷歌翻译将该标题中的“充电宝”一词译为“ŋon banka”意为“大银行”,与原词意义相差较大。

再以“艾德莱斯”一词为例,这一源自维吾尔族的传统丝绸技艺具有鲜明的文化特色。在维译汉方向上,必应翻译将“etles”误译为“艾斯”,谷歌翻译直接译为“缎子”,而小牛翻译则将其翻译为“亚特兰大”,上述译文均与原词含义产生了显著偏差。

还有“木卡姆乐韵新声”这一标题,在汉译维中,必应翻译软件将“乐韵”这一文化词汇音译为“lejyn”原标题中的“乐韵”指乐曲旋律韵味、音律气韵、风雅气韵,含艺术美感,这样翻译不能准确翻译出其中涵义;izdax 翻译将其翻译为“muqam kyj sadasi”意为“木卡姆曲调”不能准确翻译出标题的深层含义、文学意境与核心词义,存在严重语义删减;国语助手将“乐韵”译为“muzika”意为“押韵”,这里的乐韵不是指诗词的“押韵”,完全曲解词义,这样翻译不能准确翻译出标题的本意、文学意境与文化内涵,属于望文生义、错解词义;谷歌翻译将此句译为“Muqam New Music”翻译为英语,这样翻译不仅偏离指定维吾尔语目标语种,过度简化语义,丢失标题文艺意蕴与文化内涵,翻译准确性与规范性严重不足。在维译汉过程中,多数翻译软件整体译文较为准确,但是必应翻译将“muqam—melodijediki jeŋi sada”错误将其译为“蔼可亲——旋法的新声音”,不仅曲解了木卡姆这一文化专有名词,还出现语义错乱、表意不通的严重误译问题。

上述译例表明,主流翻译软件在处理新闻标题中的专有名词与文化负载词时,仍存在显著的语义偏差与译名规范化缺陷。相比之下,国语助手的高分表现主要源于两方面优势:一是内置新疆地域词库,能够准确处理“艾德莱斯”“木卡姆”等特色词汇;二是通过对抗训练技术,有效区分新闻标题中的政治表述与口语化表达。

上述评测结果揭示,针对特定语言翻译质量的提升,需融合语言学家标注的语法规则与深度学习技术,尤其在处理新闻标题这类高度凝练的文本时更需注重文化语境的重构。

Table 3. Average scores of the scoring results for news subheadings
表 3. 新闻小标题打分结果平均分

	汉译维	维译汉
必应翻译	4.28	3.69
izdax 翻译	3.78	4.69
国语助手	4.72	4.94
DeepSeek-V3	2.88	4.59
谷歌翻译	3.38	4.38
小牛翻译	3.88	4.75

4. 译后编辑要点与 MTPE workflow 适用性分析

4.1. 译后编辑要点总结

综合上述各平台在汉译维与维译汉方向的表现,译后编辑者在进行译文修正时,需重点关注以下八类问题:

- (1) 专有名词的准确性修正。多数平台在人名、地名、机构名等专有名词翻译上均存在较高出错率,

具体表现为:无法译出(以乱码“#”替代)、音译错误、国名混淆(如将“乌拉圭”译为“乌克兰”)、政治术语误译(如将“国家主席”译为“总统”)。译后编辑者需逐一核对原文中的专有名词,参照规范译名表进行统一校正。

(2) 数字与日期的核查。多个平台在翻译涉及数字、日期的句子时出现识别错误或漏译,如将“16日”错译为“15日”,或将项目数与传承人数颠倒。译后编辑者需仔细比对原文中的数字信息与日期信息,确保数值准确无误。

(3) 语义层次的区分与忠实表达。部分平台存在将多个近义词混译为一词的现象(如将“信得过、靠得住”混译,“知心人、贴心人”混译),导致原文中丰富的语义层次被压缩或丢失。译后编辑者应根据原文语义,对译文进行细化区分,恢复应有的表达精度。

(4) 语体风格的统一。小牛翻译等平台在处理新闻类文本时出现口语化过强的问题,不符合新闻语体的专业性要求。译后编辑者需根据文本类型调整译文风格,将口语化表达转换为规范的书面语体。

(5) 语法与语序的调整。DeepSeek-V3等通用模型在处理复杂句子时易出现语法错误、语序混乱、成分残缺等问题。译后编辑者需依据维吾尔语或汉语的语法规则,重新梳理句子结构,调整语序,补全缺失成分,使译文通顺自然。

(6) 漏译内容的补充。多个平台存在不同程度的漏译现象,包括词汇漏译(如“进一步激励广大”)、专有信息漏译(如“木卡姆节目”)、日期漏译等。译后编辑者需对照原文,逐句检查是否有内容遗漏,并进行补充翻译。

(7) 无法识别内容的处理。DeepSeek-V3在某些情况下无法识别翻译请求,直接返回“你好,这个问题我暂时无法回答”等无效回应。译后编辑者需对此类未完成翻译的内容进行人工补译。

(8) 概念性误译的纠正。谷歌翻译等平台存在概念层面的误译,如将“全国助残日”译为“残疾人节”,将“非物质文化遗产”误译为“物质文化遗产”。此类错误属于核心语义的偏离,译后编辑者需从概念定义出发进行重新翻译,而非简单修改用词。

通过上述针对性的译后编辑操作,可显著提升机器翻译产出的准确性与可读性,使译文更贴近原文语义与目标语言的表达习惯。

4.2. 翻译软件使用感受

在本次评估中,笔者分别使用了必应翻译、izdax 翻译、国语助手、DeepSeek-V3 翻译、谷歌翻译以及小牛翻译等六种机器翻译系统。除译文质量本身外,系统的操作体验、运行稳定性、付费模式等因素同样直接影响译者在 MTPE (机器翻译译后编辑) workflow 中的整体效率。本节从以下三个维度对各系统的 MTPE 适用性进行评估:

(1) 必应翻译的网页界面设计简洁明了,操作流畅,并具备单次 5000 字的大文本处理能力,有利于减少长文本翻译时的分段操作成本。在维译汉任务中,部分句子存在结构不对称问题,可通过简单人工调整修正;汉译维方向对齐不整齐的现象更为常见,需借助句子对齐工具辅助调整。总体而言,必应翻译在大文本快速获取初译结果方面优势明显,但在双向翻译的格式一致性上仍有改进空间,译后编辑时需额外投入格式整理时间。

(2) izdax 翻译作为一款微信小程序内的翻译工具,实行付费使用模式。建议用户采用单句方式进行翻译,批量翻译时容易出现格式对仗不齐问题,换行导致的单词被分割现象在复制到其他应用时无法自动合并,显著增加译后编辑的格式修复成本。部分专有名词(如“玛纳斯”)存在漏译。但其优势在于处理连续大量句子的翻译任务时运行流畅,未出现卡顿,对需要稳定连续输出的 MTPE 场景具有一定价值,但付费门槛与格式问题制约了其综合效率。

(3) 国语助手翻译软件在电脑版应用中常出现无法正常运行的问题, 但可通过微信电脑版的小程序实现翻译功能, 前提是用户需开通会员。在处理大批量汉译维翻译时支持文件上传翻译, 但部分句子排列出现错行, 需手动调整。然而, 在维译汉文件翻译过程中软件无法自动识别文件内容, 建议采用单句翻译方式进行, 大幅降低了批量处理效率。在大量翻译场景下, 文本格式难以保持一致, 且部分人名无法识别, 部分人名以“#”号代替。持续进行大量翻译任务时, 软件易出现卡顿现象, 翻译引擎运行速度明显减慢, 对 MTPE 工作流的连续性和时效性构成制约。

(4) DeepSeek-V3 翻译网页版需以指令方式驱动, 例如: 将词句翻译成维吾尔语或将此句翻译成汉语。无指令则无法识别。虽然在持续发布三四个指令后同一对话框可免去重复指令, 但翻译速度较慢, 持续大量翻译后机器严重卡顿, 频繁显示“无法识别此问题”, 部分句子多次翻译仍无法完成。约翻译 150 句后需开启新对话框并重新进行指令学习方可继续使用。上述操作门槛与不稳定性在 MTPE 工作流中显著增加了非译后编辑本身的操作成本, 对新闻生产所需的时效性与连续性构成较大挑战。

(5) 谷歌翻译与小牛翻译这两款软件在界面响应与操作流畅度方面表现较为稳定, 未出现明显的运行卡顿或功能失效问题。其中, 小牛翻译在译文上存在口语化过强的问题; 谷歌翻译存在一些概念层面的误译问题。从 MTPE workflow 效率来看, 两者的操作稳定性为译后编辑提供了相对顺畅的前置条件, 但由于上述译文质量问题的存在, 仍需要投入相应的修正时间。

综上, 各系统在 MTPE 工作流中的适用性差异明显。必应翻译适合大文本快速初译; izdax 运行稳定但受制于付费与格式问题; 国语助手文件上传便利, 但稳定性与维译汉识别欠佳; DeepSeek-V3 指令依赖与卡顿问题突出, 与新闻生产场景匹配度低; 谷歌翻译与小牛翻译操作稳定, 译后编辑精力可集中于译文质量本身。译者应综合译文质量与操作效率, 按需选配工具。

5. 改进建议

5.1. 译后编辑建议总结

(1) 翻译准确性与完整性

建立汉维专有名词词库, 重点收录政治、文化类高频词汇及常见人名, 避免出现字母、#号替代的情况, 确保特殊词汇精准识别。优化汉维互译的句子结构对齐算法, 针对维译汉的结构不对称、汉译维的格式不齐问题, 自动修正语法逻辑与排版, 减少人工调整成本。

(2) 文本处理能力

统一提升大文本翻译支持, 将必应翻译五千字上限作为基础标准, 其他系统(如 izdax、DeepSeek-V3)需突破单次翻译字数限制, 同时保障批量文本的格式稳定性。可新增格式自动修复功能, 解决 izdax 换行分割单词、国语助手错行等问题, 确保翻译后文本复制到其他应用时无需二次调整。

5.2. 操作体验升级

(1) 指令与交互简化

DeepSeek-V3 需取消“强制指令”要求, 默认识别用户输入的翻译需求(如自动判断汉维互译方向), 仅在复杂场景下提示补充指令; 同时优化对话记忆功能, 避免频繁开启新对话框重新学习指令。所有系统需优化界面引导, 如在批量翻译入口标注“建议单次文本量”“格式注意事项”, 帮助用户快速掌握正确使用方式。

(2) 性能稳定性提升

针对国语助手、DeepSeek-V3 的卡顿问题, 优化翻译引擎负载能力, 确保持续处理 150 句以上文本时, 运行速度与识别准确率不下降。可新增“翻译进度实时显示”功能, 避免用户因等待时间过长重复

操作, 提升大文本翻译时的体验确定性。

5.3. 付费与功能适配调整

(1) 付费模式与功能匹配

izdax 需调整付费权益, 如免费用户开放“单句翻译 + 基础格式修复”, 付费用户解锁“批量翻译 + 高级对齐工具”, 同时明确标注不同付费档位的文本处理上限, 避免用户付费后仍面临格式问题。

(2) 多端适配优化

国语助手需修复电脑版应用运行故障, 确保电脑端与微信小程序功能一致(如均支持文件上传、批量翻译), 避免用户因设备限制被迫切换使用场景。所有小程序端系统(如 izdax、国语助手)需优化文本输入框体验, 支持粘贴文本后自动去除多余格式(如空格、换行), 减少预处理操作。

6. 结论

本研究发现翻译质量呈现较为明显的“方向差异性”。在本次评测的语料范围内, 维译汉整体表现优于汉译维, 各系统平均得分普遍高于汉译维方向。其中, izdax 翻译在维译汉任务中的得分相对突出(4.71分)(见表 1), 国语助手在汉译维方向得分相对较高(4.56分)(见表 1); 而 DeepSeek-V3 在双向翻译中得分均偏低(汉译维 2.815 分、维译汉 4.123 分)(见表 1)。这一结果在一定程度上反映出, 当前汉维机器翻译技术对“汉语到黏着语”转换的适配性可能弱于“维吾尔语到汉语”方向, 但上述差异均基于本次评测数据得出, 未经过统计显著性检验, 其是否具有普遍性尚需进一步验证。

专有名词与格式问题是制约各系统译文质量的核心瓶颈之一。所有被测系统在处理新闻语体中的政治术语、人名、日期时均不同程度地存在偏差, 具体表现为漏译、以“#”号遮挡(如国语助手)或直接翻译缺失(如必应翻译)等现象。此外, 在批量翻译场景下, 格式错乱问题频发, 如 izdax 翻译的单词换行分割异常、国语助手的句子错行等, 此类问题需投入大量人工进行译后编辑调整, 方能满足新闻文本的规范性要求。

各系统功能与新闻场景的适配度存在明显差异。必应翻译凭借简洁界面与支持 5000 字大文本处理能力, 在新闻“快译”需求场景中相对更具操作优势; izdax 翻译虽批量翻译流程较为流畅, 但存在付费门槛与格式处理问题; 国语助手的文件上传功能在实际使用中具有一定便利性, 但受限于电脑端故障及维译汉文件识别失效等问题; DeepSeek-V3 的“强制指令”设计及高频卡顿问题, 与新闻生产的时效性与连续性要求匹配度较低, 实用性有待提升。

综合评测结果来看, 现有汉维机器翻译系统在一定程度上能够满足新闻语体的基础信息传递需求, 但距离“零调整”的实用标准仍有明显差距。要实现这一目标, 需优先在以下三个方面着力突破: 一是构建新闻专用汉维术语库, 提升专有名词翻译的准确率和一致性; 二是优化大文本格式自动修复算法, 降低人工校对成本; 三是根据翻译方向差异调整模型训练重心, 尤其应加强汉译维方向的语法对齐能力。

研究局限与展望: 本研究为小样本探索性评测, 测试语料规模有限(约 200 句, 1 万字), 且语料来源集中于单一类型新闻媒体, 可能不足以全面反映各系统在更广泛新闻子语体(如时政、经济、社会、国际新闻)中的表现差异。在评分环节, 每句译文由 4 名评分者独立打分, 评分组由汉语母语者和维吾尔语母语者共同构成, 这在一定程度上保证了评分的语言敏感性, 但本研究未进行评分者间一致性检验(如 Kappa 系数); 各系统间的得分差异亦未经统计显著性检验(如配对样本 T 检验或非参数检验)。因此, 本研究所呈现的系统排名与优劣判断应视为基于当前数据的一种趋势性观察, 而非确定性结论, 读者在参考时需审慎对待。未来研究可通过扩大语料规模、引入多元新闻来源、增加统计显著性检验以及开展纵向跟踪评测, 以进一步增强结论的稳健性和外部效度。

本次评测为新闻译者在机器翻译译后编辑(MTPE) workflows 中的工具选择提供了一定参考。在本次评测中, 维译汉任务中 izdax 翻译表现相对较好, 汉译维任务中国语助手得分相对占优, 大文本处理场景中必应翻译较为适用, 新闻标题翻译可优先尝试国语助手。但具体选用何种系统, 仍需译者结合具体任务需求、文本类型及对译文质量的预期做出综合判断。本研究的更大意义在于, 通过系统性的评测揭示了汉维机器翻译在新闻语体处理中仍存在的结构化改进空间, 为后续技术优化和评测体系的进一步完善提供了实证基础, 期望能推动汉维机器翻译从“可用”向“好用”持续升级, 更好地支撑跨语言新闻传播与多民族文化交流。

致谢

本论文得以顺利完成, 离不开众多师长、同学、朋友和家人的支持与帮助, 在此谨致以诚挚的谢意。

首先, 衷心感谢我的导师。从选题确定到研究框架的搭建, 从机器测评到论文的反复修改, 导师始终给予我耐心细致的指导。导师严谨的治学态度、深厚的学术功底和敏锐的问题意识, 使我受益匪浅。在论文写作过程中, 导师不仅关注研究的学术价值, 也多次提醒我要关注成果的实际应用意义, 这些教诲将使我终身受用。

感谢我在新疆求学期间结识的各位老师和同学。在田野调查和语料收集过程中, 新疆当地的多位维吾尔族朋友和同学给予了热情的帮助, 为我理解维吾尔语新闻文本的特点提供了宝贵的母语者视角。没有他们的协助, 本研究的语料建设和译文评分工将难以顺利开展。

感谢我的同窗好友们, 在论文写作的日日夜夜中, 我们相互鼓励、彼此扶持。与你们的讨论和交流时常给我带来新的启发, 也让这段艰辛的写作历程充满温暖。

最后, 我要特别感谢我的家人。你们始终是最坚实的后盾, 在我求学路上给予我无条件的理解、支持与包容。每当我感到迷茫和疲惫时, 你们的关怀与鼓励总能让我重新振作。

本研究的完成并非终点, 而是新的起点。汉维机器翻译的评测与优化仍有大量值得深入探索的问题, 我愿在今后的学习与工作中继续关注这一领域, 为促进跨语言信息服务和多民族文化交流贡献一份微薄之力。

参考文献

- [1] 董兴华, 周俊林, 郭树盛, 等. 基于短语的汉维/维汉统计机器翻译[J]. 计算机工程, 2011, 37(9): 16-18+21.
- [2] 范新瑜. 海洋科技文本机器翻译译文质量评测[J]. 武汉冶金管理干部学院学报, 2023, 33(4): 31-34.
- [3] 哈里旦木·阿布都克里木, 侯钰涛, 姚登峰, 等. 维吾尔语机器翻译研究综述[J]. 计算机工程, 2024, 50(1): 1-16.
- [4] 王均松, 庄淙茜, 魏勇鹏. 机器翻译质量评估: 方法、应用及展望[J]. 外国语文, 2024, 40(3): 135-144.
- [5] 阳琼. 在线翻译质量文类差异: 基于人工测评的比较分析[J]. 东方翻译, 2021(4): 54-59+66.