

大语言模型背景下 ChatGPT 与 DeepSeek 翻译质量对比研究

——以历史相关文本为例

潘红

三亚学院

DOI:10.32629/mef.v8i18.16969

[摘要] 本研究以大语言模型ChatGPT与DeepSeek为翻译工具,以中国历史类文本为翻译文本,进行汉译英任务,采用BLEU值与TER值自动评估方法,对比分析ChatGPT-4与DeepSeekV3.1在历史相关文本翻译中的表现。重点考察译文在忠实性、语言风格、创造性及篇章连贯性等方面的差异,旨在为人工智能翻译技术的优化与译后编辑策略提供参考,同时为中国历史相关文本的对外传播探索可行路径。

[关键词] ChatGPT; DeepSeek; 历史相关文本翻译; 翻译质量评估

中图分类号: G633.51 **文献标识码:** A

A Comparative Study on Translation Quality Between ChatGPT and DeepSeek in the Context of Large Language Models——Taking historical texts as an example

Hong Pan

Sanya University

[Abstract] This study employs large language models ChatGPT and DeepSeek as translation tools, using Chinese historical texts as the source material for English translation tasks. By applying automatic evaluation methods such as BLEU and TER scores, it conducts a comparative analysis of the performance of ChatGPT-4 and DeepSeekV3.1 in translating history-related texts. The study focuses on examining differences in translation fidelity, linguistic style, creativity, and textual coherence, aiming to provide references for optimizing AI translation technologies and post-editing strategies. It also explores feasible pathways for the international dissemination of Chinese historical texts.

[Key words] ChatGPT; DeepSeek; history-related text translation; translation quality evaluation

引言

本研究的原著以《资治通鉴》原典为经纬,围绕政治得失与王朝兴衰这一主线展开分析。该书文本有如下特点:第一,夹叙夹议,“论从史出”,原著囊括从“三家分晋”到“大唐盛世终结”等重大历史转折点的考察,包含大量的历史相关信息。第二,有直接引自《资治通鉴》《史记》等历史典籍的原文(文言文或者白话文)。第三,兼有科普性和学术性。依据彼得·纽马克的文本类型说,属于信息型文本。(赵晨2025)原著涉及富含中国文化的专有词汇,例如“天下”“阴阳”等,更包含中国古代国家制度相关词语如“九品中正制”“均田制”和官职名称等内涵丰富的词语。这就对译文提出了相应的挑战:首先,由于原文与译文读者所处的文化背景存在差异,原文省略背景知识,如典籍原文含义和专有词汇意思,将会给英语读者阅读带来障碍,这就需要译者不仅传递字面意思,还要适当补充其背后的文化背景知识、

显化原文中可能存在的隐含关系,帮助译文读者更好地理解原文内容。其次,原文中信息点多且部分内容有重复,翻译时不进行恰当处理会使译文冗长啰嗦。(赵晨2025)译者在翻译时必须深入理解原文,根据译文读者的表达习惯,对原著中的重复性内容进行精简处理。

人工译者用传统的文学翻译方法,通过通读全文、查询背景知识和专有词汇含义,以及对比研究平行文本、查阅词典,依赖译者对文本和历史、文化的理解来完成本书的汉译英工作,面临提升效率的难题。随着人工智能技术的飞速发展,大语言模型(LLM, Large Language Models)已经成为当今最热门的技术之一。OpenAI的ChatGPT、DeepMind的DeepSeek等,通过海量的语料库进行训练,可以学习语言的结构、语义、上下文关系等,从而能够进行语言理解、文本生成。这就为中华典籍外译带来了新的前景。本研究旨在系统评估ChatGPT与DeepSeek在历史相关

文本英译任务中的表现,并通过与人类译员的对比,分析其在翻译准确性、文化专有项处理及语境连贯性等方面的相对优势与局限。重点考察两类大语言模型对中国历史典籍中特有概念(如典章制度、哲学术语及文化负载词)的翻译能力,进而探讨其对中国历史文献对外传播的潜在助力与当前局限。

1 ChatGPT-4与DeepSeekV3.1在《资治通鉴与家国兴衰》(节选)的翻译质量评估

本翻译质量评估文本共获取中文9342个字符,参考英文译文6221个字符。数据收集时间截至2025年11月11日。文中出现历史人物称谓、中华文化特有词汇和国家制度相关词汇,这些构成了翻译中的难点。

本次翻译质量评估将ChatGPT-4与DeepSeekV3.1的翻译质量进行对比分析。在使用两种大语言模型翻译时,均使用相同的英语提示词,一段一段地进行汉译英,并且不干预译文走向。为了最大程度的保留两个大语言模型译文的特质,提示词中未添加术语限定的翻译提示语和说明文本来源于固定语言风格的背景信息提示语。(王子云,毛毳2023)

本翻译质量评估数据来源于“试译宝—译文评测工具”中所计算出的BLEU值和TER值,选用原著的人工英译本为参考翻译文本。在机器翻译领域,BLEU值(Bilingual Evaluation Under study)和TER值(Translation Error Rate)是两种常用的翻译质量评估指标。BLEU基于n-gram模型,比较机器译文中的n-gram与参考译文中的n-gram,计算匹配的n-gram数量占机器译文n-gram总数量的比例,用于评估机器译文的充分性和流畅性。(王子云,毛毳2023)BLEU值越高,表明翻译质量越好。TER通过统计机器译文修改为参考译文的后编辑次数,来评估机器译文中词汇和语法的错误率。TER值越低,表明该译文与参考翻译之间的编辑距离越小,与人工译文越接近。表1展示了两个大语言模型的BLEU值和TER值。

表1 ChatGPT-4与DeepSeek-V3.1的译文质量指标单位：%

指标名称	ChatGPT-4	DeepSeek-V3.1
BLEU	55.77	57.9
TER	71.74	71.05

从数据整体看,这两个大语言模型都具备了相当成熟的翻译能力,其分数表明它们生成的译文与高质量参考译文已经具有较高的匹配度,需要人工介入修改的幅度相对可控。BLEU分数均超过55分,表明两款模型在词汇、短语和句法结构的匹配度上表现良好,能够生成基本准确、流畅的译文。TER分数均控制在72以下,说明其译文整体可读性较强,后期编辑的工作量处于合理范围。这表明大语言模型在翻译任务上已经超越了简单的词对词替换,具备了较强的语境理解和语言生成能力。

不过,需要强调的是,BLEU和TER这类自动评估指标主要关注译文与参考译文在表面特征上的匹配度,例如词汇和短语的

匹配(n-gram)以及编辑距离。它们难以完全衡量译文的逻辑流畅性、文化适配性以及语体的得体性等更深层的质量维度。然而,无论是ChatGPT-4还是DeepSeek-V3.1,在官名、中华文化特有词和人物对话的翻译上,准确度都有待提高。下面具体分析。

1.1 人物历史称谓的翻译

第二讲出现了14个人物历史称谓(谥号称谓),包括“XX王”“XX公”和“XX侯”。

ChatGPT-4和DeepSeek-V3.1均将“XX王”简单翻译为“King X of X”,有待修正。把“周幽王”“周平王”“周显王”中的“王”翻译为“King”是合适且准确的,因为他们都是东周君主,在名分上仍是天下的“共主”,其称号“王”在层级上对应于“King”。根据《资治通鉴》记载,公元前325年,秦惠王首开先河,自称为“王”,是为秦国称王之始。因此“秦惠王”译为“King Hui of Qin”是合适的。DeepSeek-V3.1将“秦惠王”译为“King Huiwen of Qin”,可能模糊了其生前使用的实际称号与后世历史评价之间的区别。据《史记》记载,战国时期齐国君主田因齐于公元前323年率先自称“王”,似乎将齐威王译为“King Wei of Qi”是对的。但是原著文中提及“齐威王”的语境,时间却是秦孝公在位时期,即公元前338年之前,此时“齐威王”并未称“王”,因此ChatGPT-4和DeepSeek-V3.1均未能翻译准确。

ChatGPT-4和DeepSeek-V3.1对“XX公”的翻译也有待讨论,它们均无视具体历史时期,使用“Duke”来指代“公”,并不准确。秦襄公是秦国第一位被周天子正式赐予爵位的君主,因此用“Duke Xiang of Qin”来翻译是恰切的。用“Duke”来翻译此处的“公”,旨在准确传达其受封诸侯这一开创性的历史事实,标志着秦国从附庸正式成为诸侯国。但是,“秦穆公”“秦献公”“秦孝公”中的“公”是作为泛化的尊称,翻译为“lord”更合适,因“Lord”是一个含义广泛且中性的词,能准确地传达其作为“一方统治者”或“君主”的实质。

两个大语言模型在处理“XX侯”的翻译时也过于简单化,一概翻译为“Marquis”。“Marquis”源自欧洲封建制度,其内涵(如与采邑、封臣制的特定关系)与战国时期诸侯国的独立政权性质有显著区别。

1.2 中国古代制度名称的翻译

表2 中国古代制度翻译

制度名称	参考译文	ChatGPT-4	DeepSeek-V3.1
保甲	Neighborhood-Based-Security System	the bao-jia (mutual responsibility) system	the "shi-wu" mutual surveillance system
井田制	the Well-Field System	the well-field system	the Well-Field system
察举	observation	chaju	recommendation
征辟	recommendation	zhengpi	direct summons
九品中正制	the Nine-Rank System	the Nine-Rank system	the Nine-Rank System
科举	the Imperial Examination	the imperial examination	the civil service examinations

第二讲中包含了诸多中国古代制度名称,包含基层管理制度、经济与土地制度和选官与人才选拔制度。

大语言模型对“井田制”“九品中正制”的翻译均正确,但是对其他制度的翻译过度依赖音译(如察举、征辟)或选择不够精准的词汇(如保甲),未能充分传达制度的本质。对“保甲”的翻译,ChatGPT-4采用了音译加解释的方式,“mutual responsibility”的表述虽然部分传达了“联保互卫”的核心思想,但未能完整涵盖该制度作为基层户籍管理与治安组织的行政属性。DeepSeek-V3.1的译法则有较大偏差,“shi-wu”并非“保甲”的标准音译,且“mutual surveillance”过于强调“互相监视”这一具体手段。相比之下,参考译文“Neighborhood-Based-Security System”更准确地概括了该制度以邻里为基础、以安全防控为核心功能的特性,是更可取的译法。

察举与征辟是两组关联密切的制度。察举制的核心是地方官考察并举荐人才。ChatGPT-4的音译“chaju”未能传递任何意义,而DeepSeek-V3.1的“recommendation”虽涉及“举荐”,但未能与“征辟”中的“辟”(亦含征召、推荐之意)很好区分,且忽略了“察”的基础作用。ChatGPT-4对“征辟”的音译“zhengpi”同样无意义。

DeepSeek-V3.1对“科举”的翻译“civil service examinations”(文官考试)从功能角度描述,突出了其为选拔政府官员而设的目的,或是另一个可行的译法。总之,翻译好中国古代制度名称,需要兼顾历史准确性、文化独特性和译文可读性。在动笔翻译前,必须透彻理解该制度的本质、功能和历史语境。对于已在国际汉学界广泛接受的译名,如“科举制”译为“imperial examination system”,应优先采用以保证学术交流的顺畅。

2 结论

综合来看,大语言模型已经重塑了翻译行业的格局。与人类译员相比,无论是ChatGPT-4还是DeepSeek-V3.1,优势和局限都很明显。在历史类文本翻译中,大语言模型在处理大规模文本、进行基础信息提取和初步翻译方面展现出显著的效率优势。然而,技术并未能解决所有问题。在缺乏足够高质量专业数据训练的领域,大语言模型知识储备明显不足。尤其是在处理富含文化、历史背景的复杂文本时,其深层理解力仍无法完全替代人类译员的专业判断。历史类文本外译方面,人机协同将更有优势。利用大模型处理海量信息、提供初稿和提升效率,而人类译者则专注于质量把控、文化转换和风格雕琢等更高层次的任务。

本论文是2025年度三亚学院研究生教育教学改革研究项目“大语言模型技术驱动下翻译课程思政创新路径研究”(项目编号:USY25JG01)研究成果。

[参考文献]

[1] 鹿雅心,王晓燕.大语言模型背景下ChatGPT翻译质量研究——以陕西本土文学作品《人生》(节选)为例[J].现代外语,2023,(22):67-70.

[2] 赵晨.历史类文本中语篇信息的流畅性优化[D].大连外国语学院,2025.

[3] 王子云,毛毳.ChatGPT译文质量的评估与提升——以陶瓷类文本汉英翻译为例[J].山东陶瓷,2023,46(04):20-27.

[4] Li,Feng.“‘Feudalism’ and Western Zhou China:A Criticism.” Harvard Journal of Asiatic Studies,vol.63,no.1,2003,pp.115-44.

作者简介:

潘红(1985—),女,汉族,安徽芜湖人,硕士研究生,讲师,研究方向为英语专业、翻译理论与实践。