

ChatGPT 的文学翻译质量评估研究——以《围城》（第一章）英译为例

漆堃 莫秀兰

桂林电子科技大学，广西桂林，541004；

摘要：随着人工智能技术的迅猛发展，以 ChatGPT 为代表的大语言模型在翻译领域展现出巨大优势与广阔应用前景。本文以钱钟书经典作品《围城》为研究对象，探究 ChatGPT 在文学翻译中的表现。研究参考珍妮·凯利 (Jeanne Kelly) 和茅国权 (Nathan K. Mao) 的权威译本，通过计算译文的 BLEU 值、TER 值进行定量分析，并从文化元素传达、语言美感传递的维度展开定性分析，对比 ChatGPT 译文与人工译文在翻译质量上的优势与不足。研究发现，ChatGPT 能够实现语言通顺、语法正确，但在语义精准把握、情感传递、文化意蕴阐释等方面仍存在局限。研究认为，AI 翻译技术尚无法完全取代人工译者，但其高效性与创新性为文学翻译提供了新路径。译者应积极构建人机协同模式，推动中华优秀文学作品的国际传播。

关键词：ChatGPT；文学翻译；《围城》；翻译质量评估

DOI：10.64216/3080-1494.25.12.066

1 引言

1.1 研究背景

随着人工智能技术的不断进步，ChatGPT 等机器翻译依靠其强大的数据处理能力，在翻译领域发挥日益重要的作用，交互式人机对话也为译者开展翻译工作提供了诸多帮助，不仅提升了查阅资料的速度与准确度，更丰富了译者自身的知识储备 (苏晓洋, 2024)。虽然 ChatGPT 凭借其强大的语言生成能力，在多语言翻译任务中展现出高效性与适应性，但其在文学翻译领域的表现仍需深入探讨。文学翻译不仅要求语言转换的准确性，更注重文化内涵、情感风格与审美特质的传递，这对 AI 技术提出了更高挑战。我国大量的优秀文学作品在国际传播中面临知名度低，影响力弱的问题，但 ChatGPT 的出现为破解这一困境提供了技术契机。

1.2 研究对象

《围城》在中文文学界享有极高的声誉，被誉为“新儒林外史”，选择《围城》作为研究对象，是因为该书不仅是理解民国时期中国文化和社会结构的重要文献，也是传播中国文化的理想媒介 (廖钰惠, 2025)。

《围城》的翻译为研究中英文学互鉴、跨语言叙事策略提供了典型案例，其译本对比研究成为翻译学领域的重要课题。其中由 Jeanne Kelly (珍妮·凯利) 和 Nathan K. Mao (茅国权) 合译的首个完整英译本 Fortress

Besieged, 于 1979 年由印第安纳大学出版社 (Indiana University Press) 出版，被纳入“亚洲经典丛书”，至今仍是西方学界研究《围城》的主要版本。

1.3 研究目的与意义

《围城》作为中国文学著名的作品，其英译不仅只是语言转换的过程，更是一次代表着中国文化的艺术品与外界文化的读者的对话。本文通过定量与定性结合的方法，系统分析 ChatGPT 的《围城》英译质量，旨在揭示大语言模型在文学翻译中的潜力与局限，为 AI 技术与人工翻译的协同发展提供参考。本研究为探讨大语言模型在文学翻译中的应用提供了有益的参考，对促进优秀文学作品的传播和跨文化交流具有积极的意义。

2 研究方法

采用定量分析与定性分析相结合的方式。定量分析主要通过有参考自动评估法来评估翻译的准确性和流畅度。而定性分析则对译文进行深入分析和解读，包括词汇选择、句子结构、文化元素传达等方面。通过结合这两种分析方法，可更全面地了解 ChatGPT 在翻译《围城》时的优势和不足。

2.1 定量分析：有参考自动评估法

有参考自动评估指以人工译文为依据，对机器翻译译文与参考译文之间的相似度进行自动计算，二者之间的相似度越接近，机器翻译的译文质量就越高 (王均

松, 2024)。在机器翻译领域, BLEU 值(Bilingual Evaluation Understudy)和 TER 值(Translation Error Rate)是两种常用的用于评估翻译质量的指标。BLUE 值通过 NLTK 工具计算 ChatGPT 译文与参考译文的 n-gram 匹配度, 衡量词汇准确性与流畅度, 计算被测译文和参考译文之间的相似程度, 确定被测译文的质量。TER 值使用 Mozilla 的 TER 工具计算编辑距离, 评估译文与参考译文的差异程度, 衡量机器翻译系统输出与参考翻译之间的编辑距离。

2.2 定性分析: 多维文本对比

定性分析从文化元素、语言美感的角度出发, 对比 ChatGPT 译文与参考译文在跨文化语义传递、文学风格再现及细腻情感表达上的表现。文化元素传达包括译文对习语、典故、专有名词及隐喻等处理, 语言美感指译文对修辞手法巧妙运用及幽默讽刺语言风格传达。

3 研究结果

由于翻译文本字数较多, 计算工具无法对长篇幅文本进行评估, 为了保证数据的准确性, 本研究对《围城》第一章随机采样了十段文字内容, 并分别对其 BLUE 值与 TER 值进行计算, 将其平均值作为最终参考数据: ChatGPT 对《围城》第一章翻译 BLUE 值约为 0.556, TER 值约为 0.487。

3.1 数据分析

BLEU 值衡量机器译文与高质量参考译文的相似度, 范围在 0 到 1 (即 0% 到 100%), 得分越高表示翻译结果与人工翻译越接近。0.556 表示 ChatGPT 译文在词汇选择、语句结构和句子流畅度上, 与权威人工译本有中等程度的匹配。它捕捉到了原文的主要信息和部分风格, 但在精确的词汇选择、文学性传达等方面存在差距。

TER 值衡量将机器译文编辑成参考译文过程中所需插入、删除、替换、词序调整等步骤的最少编辑操作次数, 占参考译文总词数的比例。TER 值越低表示翻译质量越高, 与参考结果越接近。0.487 表示要将 ChatGPT 译文修改成参考译文, 平均需要改动 ChatGPT 译文中的 48.7% 的单词。这个比例相对较高, ChatGPT 译文与人工译本依然存在较大差异。

3.2 结果分析

ChatGPT 译文基本达到了可读和理解的水平, 能够传达《围城》第一章的主要情节和大部分信息。然而,

与参考译文相比, 它在精确性、流畅度、文学性以及整体文风的优雅度上存在明显的不足。TER 值高于 BLEU 值, 表明 ChatGPT 译文在细节层面需要大量编辑才能达到专业水准。因此, ChatGPT 译文可作为快速获取《围城》第一章内容梗概的工具, 但其翻译质量与人工翻译仍有较大差距。外语读者在欣赏《围城》这类具有较高文学价值的作品, 人工译文的参考是不可替代且不可或缺的。

4 案例分析

本节对 ChatGPT 译文的定性分析将从文化元素传达、语言美感与风格角度展开, 并通过选取文本的实际案例进行分析。

4.1 文化元素传达

原文: “这是七月下旬, 合中国旧历的三伏, 一年最热的时候。”

ChatGPT 译文: “It was the latter part of July, the peak of the ‘three fu’ days in the traditional Chinese calendar—the hottest period of the year.”

人工译文: “It was toward the end of July, equivalent to the ‘san-fu’ period of the lunar calendar—the hottest days of the year.”

分析: 在这里 ChatGPT 将“三伏”直译为“three fu”, 机械拆分“三伏”为“Three Fu”虽字面对应, 但不符合英语表达习惯, 且无法传达“伏”字隐含的“避暑藏伏”文化内涵。而人工译文通过“equivalent to”强调“三伏”与西方读者认知的“hottest days”对应, 同时“san-fu”的拼音保留了文化独特性, 避免直译“three fu”可能引发的理解障碍。此外, 根据中国气象局、国家天文台等官方机构在对外发布的气候报告或文化介绍中, 均采用“Sanfu Days”作为“三伏天”的标准英译。此外联合国教科文组织(UNESCO)也在《中国二十四节气》非遗文档中, 将“三伏”译为“Sanfu”。基于常识, “三伏”的时间概念来自中国传统的农历干支纪日法, 而非公历, 因此需强调“lunar calendar”背景。所以此处的人工译文更加准确的传达文化元素。

4.2 语言美感传递

原文: “法国人的思想是有名的清楚, 他的文章也明白干净, 但是他的做事, 无不混乱、肮脏、喧哗, 但

看这船上的乱糟糟。”

ChatGPT 译文：“The French are famously clear-minded, and their writing is lucid and clean. But when it comes to action, everything becomes chaotic, filthy, and noisy—as evidenced by the mess on this very ship.”

人工译文：“The French are famous for the clarity of their thought and the lucidness of their prose, yet in whatever they do, they never fail to bring chaos, filth, and hubbub, as witness the mess on board the ship.”

分析：

人工译文巧妙利用名词并列的平行结构，用“the clarity of their thought and the lucidness of their prose”

对比后文中的“chaos, filth, and hubbub”，既保留了原文欲扬先抑的节奏，又在用词上更富文学色彩使得讽刺意味更浓；转折处“yet in whatever they do, they never fail to bring...”与“as witness the mess on board the ship”简练有力，既准确传达原意，又自然流畅。而相较之下，ChatGPT 译文用词平淡、结构松散，缺乏抑扬顿挫和幽默讽刺的张力，未能完全体现原文的修辞和语气，因此人工译文在整体表达和风格韵味上更胜一筹。

5 结论

在大语言模型快速发展的背景下，本研究以我国经

典文学作品《围城》(第一章)为研究对象，探讨 ChatGPT 在文学翻译领域的表现。研究发现 ChatGPT 的翻译效率高、译文语句较为通顺、可以基本还原原文的内容情节，但是在语义精准把握、情感传递、文化意蕴阐释等方面仍存在局限。因此人工翻译在当前文学翻译的重要性不可或缺，广大译者也应积极借助大语言模型，构建人机协同模式，推动中华优秀文学作品的国际传播。

参考文献

- [1] Jeanne, K. & Nathan K. M. (1979). *Fortress Besieged*. Bloomington: Indiana University Press.
- [2] 廖钰惠(Liao, YH). (2025). 浅析《围城》汉英翻译中的文化负载词与中国文化的对外传播. 汉字文化, (08), 166-168. doi: 10.14014/j.cnki.cn11-2597/g2.2025.08.050.
- [3] 苏晓洋(Su, XY). (2024). 机器翻译为译者带来的思考与启示——以 Chat GPT 为例. 中阿科技论坛(中英文), (05), 58-62.
- [4] 王均松(Wang, JS), 庄淙茜(Zhuang, ZQ), 魏勇鹏(Wei, YP). (2024). 机器翻译质量评估: 方法、应用及展望. 外国语文, 40(03), 135-144.

作者简介：漆堃，男（2002.02），汉族，安徽六安，本科，研究方向：英汉互译。

基金，项目或课题名称及编号：2025年桂林电子科技大学外国学院研究生创新项目“大语言模型时代下 ChatGPT 翻译质量评估研究——以《围城》为例。