网络奇幻小说《斗罗大陆》中对抗书写的多平台译文质量对比

何雅婷

桂林理工大学，广西桂林，541006；

摘要：本研究依据“忠实性、流畅性、文学性”三原则，构建多维英译质量评估体系，系统比较 ChatGPT、文心一言、Kimi 三种大语言模型与 Google、DeepL、有道三款传统在线翻译平台对网络奇幻小说《斗罗大陆》某“对抗书写”片段的英译表现。结果显示，ChatGPT 在忠实性和文学性两个维度表现突出，能够较全面还原原文氛围和修辞。Kimi 同样展现出高度文学性，尤其在打斗细节再现与词汇多样性方面富有优势。文心一言在流畅性表现良好，但对抗场景的节奏和氛围还原略逊于前两者。Google 译文流畅性最佳，便于理解，但在忠实性及文学性表现有限。DeepL 在忠实度上整体良好，但在细节描写和衔接自然度方面存在不足。有道表现最弱。总体来看，大语言模型在网络玄幻小说中译英任务中整体优于传统反映平台，更能平衡信息忠实传递与文学风格再现。

关键词：网络奇幻小说《斗罗大陆》；大语言模型；传统在线翻译平台；三原则和多维质量评估体系；英译质量对比

DOI：10.64216/3080-1494.25.09.025

引言

随着中华文化对外传播的不断深入，网络玄幻小说凭借独特的修炼等级、宗门体系等文化元素，以及主角成长与团队合作的叙事模式，正日益受到全球读者的关注。其中，《斗罗大陆》以“武魂+魂环”的成长体系和鲜明的人物塑造，成为中国网络玄幻小说海外传播的代表性作品（王明宪，2025）。现有研究主要集中于《斗罗大陆》的跨媒介及海外 IP 运营策略（徐立萍 & 赖雅婷，2024）、与网络游戏的互动机制、文本叙事特征与文化融合等领域。然而，对于该小说英译版本的系统性研究相对不足，尤其是对其中“对抗书写”——即关于力量对决、正邪较量等激烈打斗场景的叙写方式——在传统机器翻译与新兴大语言模型（LLM）等不同翻译模型下的英译质量尚缺乏深入评估。同时，如何在英译过程中兼顾文化特异性与目标语读者的接受度，也有待进一步探讨。实际上，对抗书写不仅推动了情节发展，更集中展现了玄幻文化的美学特征与语言的动态性（王明宪，2025）。基于“忠实性、流畅性、文学性”三大原则，本文构建了多维英译质量评估体系，对《斗罗大陆》一处对抗场景在 ChatGPT、Kimi、文心一言、Google Translate、有道翻译、DeepL 等多种翻译平台下的英译质量进行了质性与量化相结合的对比分析，旨在揭示不同平台在玄幻文本英译中的优势与不足。

1 文献综述

1.1 翻译理论和质量评估标准

翻译理论为译者实践提供了理论基础。如严复提出的“信、达、雅”要求译文忠实原文、通顺易懂、富有文采，但“雅”的含义较为模糊，实际操作中易导致译文过于追求修辞而脱离原意。施莱尔马赫（1813）提出“异化”与“归化”两大准则，前者强调保留原文文化特征，但可能使译文晦涩，后者注重通顺易读，却可能弱化原作风格。黄龙（1981）在“信达雅”基础上进一步明确了“雅”的标准，主张译文应再现原作意境和美学价值，体现节奏与情感深度，实现风格传递与读者共鸣。

翻译质量评估是翻译研究的又一重要方向。如豪斯（1977）提出的译文质量评估模式，通过对原文与译文在语义、语用和语篇层面的细致比对，为评价译文优劣提供了系统依据。近年来，尽管质量评估逐步引入翻译教学，但相关研究仍显薄弱。部分学者将学界与业界的方法结合，引入 MQM（Multidimensional Quality Metrics）等分级量表，将译文错误细分为七个层级，涵盖“准确性”“流畅度”“风格”等维度，实现翻译错误的量化和分级，并强调结合实际情境进行定性分析。此外，流畅性评估常借助福莱施可读性指数、自动可读性指数等工具，客观分析译文的可读性与流畅度（Flesch, 1948；Senter & Smith, 1967；Benjamin, 2021）。Graesser 等（2004）提出的文本分析法及

Coh-Metrix 工具，也被广泛用于分析词汇多样性、词汇高级性、内容词比例、句法依存深度、复杂从属结构和平均句长等指标，为英译文本的质量评估提供了多维视角。

1.2 大语言模型辅助下的多体裁翻译

随着科技的发展，翻译研究的视角已从聚焦人类译者及其译作逐步转向传统

机器翻译与大语言模型。传统机器翻译受限于语境理解、表达灵活性和隐喻文化处理等方面的不足。在跨句照应与词义消歧等语境推理任务中，翻译因缺乏足够上下文而产生译文逻辑断裂或意义漂移，对于变调丰富、修辞密集的文学文本，机器翻译往往无法有效再现原作的节奏感与修辞张力，使译文在语用层面显得贫乏而僵硬（Hu & Li, 2023）。再者，传统机器翻译面对含隐喻与文化意象的政治语篇时，常局限于逐词直译，无法准确再现其深层意识形态与文化指涉（文旭 & 田亚灵, 2024）。而大语言模型则能够在忠实性、流畅度和风格等方面实现明显提升。目前，学界对大语言模型辅助翻译的关注点不仅涵盖了译者身份和人机协同素养等议题。就职业身份而言，ChatGPT 等 LLM 促使译者从“纯语言专家”转型为“技术驱动的后编辑师”；人机协同层面，Wang&Luo（2025）提出“智能翻译素养”概念，强调译者需具备人机协商、资源管理等多元能力，以实现与大语言模型的高效互动从而达到“人智融合”。同时，还包括与传统机器翻译和人工译文在不同体裁下的质量对比。例如，在中国古典诗歌英译中，ChatGPT 在忠实性、流畅度与风格保持方面优于 Google 和 DeepL，但在深层文化意蕴与典故翻译上偶有直译或误译现象（Gao et al., 2024）。法律文本翻译方面，阿译英方向，人工在术语准确性和法律文体规范性上表现更优，能够准确传递法律意图与权威风格，而 AI 系统则易出现术语直译和表达口语化等问题；同样地，在英译阿方向，人类译者在内容还原和语言规范性方面仍具优势，AI 易出现术语误用、语序混乱等现象（Moneus & Sahari, 2024）。在现代小说领域，文心一言和讯飞星火在翻译准确性和文学性方面表现尤为突出，尤其得益于庞大且高质量的训练语料，其综合表现已超越传统及主流大语言模型，包括 ChatGPT，这进一步印证了语料类型与质量对翻译模型性能的关键作用（赵衍等，2024）。

然而现有研究多集中于诗歌、散文、戏剧等传统文学体裁的英译，对网络奇幻小说中对抗书写等新型叙事文本的系统性分析相对不足。此外，相关研究通常侧重于探讨技术进步对翻译效率的提升，而对译文本在目标语语境中的接受度、读者互动与反馈等方面关注不足。鉴于此，本文从网络奇幻小说中“对抗书写”这一独特叙事单元出发，结合“忠实性、流畅性、文学性”三大原则，基于自建的多维质量评估体系，对《斗罗大陆》在不同翻译平台下的英译版本进行系统性对比分析。探讨多翻平台在网络奇幻小说对抗书写英译中的整体表现，以及在忠实性、流畅性与文学性三个维度上各自的优势与不足。

2 实验设计

2.1 英译原则和评估方法

本研究以“忠实性、流畅性、文学性”为原则，构建整体英译评估框架。忠实性方面，采用 MQM 评估模型，通过量化漏译、增译和错译等现象，客观分析译文的语义保留与完整性。流畅性方面，利用福莱施阅读容易度、自动可读性指数和平均阅读水平共识等测评工具（Flesch, 1948;Senter & Smith, 1967;Benjamin, 2021）对译文进行量化分析，并结合 20 名英语母语者的问卷反馈，交叉验证可读性结果。文学性方面，参考朱莉安·豪斯的文学性评价指标及 Graesser 等（2004）文本分析方法，从词汇和句法复杂度、打斗动作描述、明喻修辞及目标语读者的文化接受度等多方面综合评估译文质量。

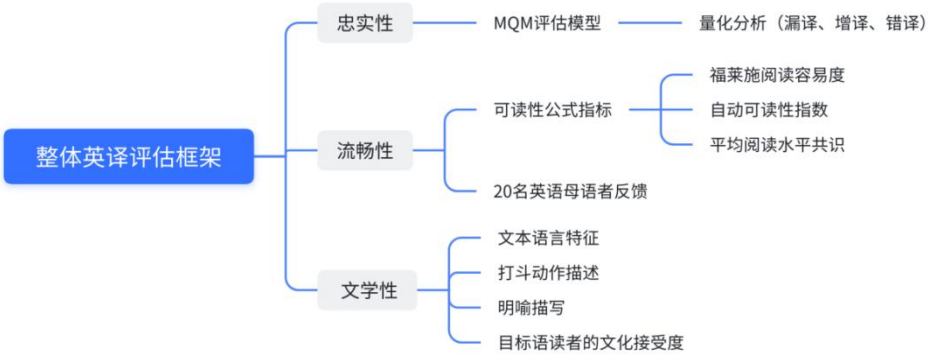

图 1 整体英译评估框架

3.2 翻译工具和测试源文的选取

本研究选取了 ChatGPT-4.0,文心一言和 kimi 等主流大语言模型，以及 Google 翻译、有道翻译和 DeepL 三个传统在线翻译平台作为英译工具。以《斗罗大陆》为测试源文，该作品为唐家三少于 2008 年创作的网络玄幻小说，讲述天赋异禀的少年唐三在斗罗大陆中修炼武魂、结识伙伴、经历冒险并最终成长为强者的故事。已被改编为动画、漫画等多种形式，成为中国网络文学的代表作之一。本研究选取了《斗罗大陆》第九章第四节作为测试文本，共计 4790 字。该选段讲述主角唐三（诺丁学院学生）与师父在猎魂森林外围遭遇剧毒魂兽曼陀罗蛇。师父与自己的魂兽“罗三炮”奋力抵抗，在危急时刻，唐三巧用暗器重创曼陀罗蛇，并最终以短剑将其击杀，从而获得了自己的第一个魂环。上述选段情节紧张，涵盖“武魂”“魂环”等专有名词和复杂战斗过程，常用长句、插叙和明喻，动作与心理活动描写交替，对各平台英译的忠实性、流畅性和文学性提出了全面考验。

3.3 英译实践步骤

本研究针对三种大语言模型，统一设计如下指令：“你好，我是一名翻译专业研究生，现需将如下片段翻译为英文。请你以翻译专家的身份进行英译，要求：忠实反映原文内容，不得删减或增译；译文自然流畅，符合英语母语习惯；最大程度保留原文修辞、意象及文化特色，使英文读者能感受到原文的文学风格。对于专有名词可适当解释或音译。请仅输出英文译文，无需注释或说明。”对于三大传统翻译平台，则直接将小说片段输入获取译文即可。

4 结果分析

4.1 忠实性

本研究从漏译、增译、错译三个维度对六大翻译平台的英译进行了对比分析。

4.1.1 漏译

漏译指译文未完整传达原文语义，尤其是遗漏关键术语、专有名词或设定等内容。本研究依据漏译字数占原文总字数的比例，将译文分为四级：A 级（≤0.5%，0－23 字）、B 级（0.5%~1%，24－47 字）、C 级（1%~2%，48－94 字）、D 级（＞2%，95 字以上）。统计结果显示（见表 2），所有平台英译时漏译字数均较少，均为 A 级，语义遗漏集中在修饰语、专有名词、谓语或主语等类型。例如，ChatGPT 漏译了修饰语“隐约间”，Kimi 漏译了专有名词“紫极魔瞳”，Google 遗漏了谓语“电射而出”，有道漏译了主语描述“沉声”。

表 1 各平台漏译情况

翻译平台	评级	漏译字数	漏译类型
DeepL	A	5	主语
Google	A	8	修饰语，谓语
文心一言	A	8	谓语，专有名词
ChatGPT	A	10	修饰语，专有名词
有道传统翻译	A	12	主语，谓语
Kimi	A	16	主语，修饰语

4.1.2 增译

增译是指译文中出现原文未有的内容。本研究根据增译字数占原文总字数的比例，将译文分为四级：A 级（≤0.5%，0－23 字）、B 级（0.5%~1%，24－47 字）、C 级（1%~2%，48－94 字）、D 级（＞2%，95 字以上）。统计结果显示（见表 3），所有翻译平台的增译字数极少，均为 A 级。仅 Google 和有道传统翻译出现了增译现象，主要表现为在“don’t know”或“I just don’t know”前添加了主语“I”。此类表达并非原文所有，将原本客观的描述转变为主观表态，改变了原文的语气和逻辑，容易引起误解。

例如：Google：“不知是否因为被罗三炮的屁和雄黄粉激怒了……”

译文：I don’t know if it was angered by Luo Sanpao’s fart and realgar powder.

有道：“……就是不知道这种有魂力的蛇是怎样的攻击方式了。”

译文：I just don’t know what kind of attack this soulful snake has.

表 2 各平台增译情况

翻译平台	评级	数量	增译类型
ChatGPT	A	0	无
Kimi	A	0	无
文心一言	A	0	无
DeepL	A	0	无
有道传统翻译	A	3	主语
Google	A	6	主语

4.1.3 错译

错译包括语法翻译错误、专有名词或其他文化负载词的翻译错误，以及这些词语前后翻译不一致的情况。

本研究根据错误类型设定权重:语法翻译错误每处2分，专有名词或文化负载词错误每处1分。累计得分用于划分译文质量：0－12分为A级，13－24分为B级，25－40分为C级，41分及以上为D级（错译严重。分析结果如图4所示:ChatGPT、文心一言与Google错译较少，均为A级。DeepL和Kimi在专有名词与语法方面出现部分错译，评为B级，如Kimi将“曼陀罗蛇”误译为“Mandrake Snake”。而有道翻译错译最多，包括5处语法、18处专有名词及11处其他文化负载词错误，被评为C级。如有道误将“唐三”译为“Tang SAN”或“Mistress”、武功招式“玄天功”误译为“Xuan day work”，“百足之虫死而不僵”误译为“Knowing that a snake, even after death, retained its venomous bite.”。

表2 各平台错译情况

翻译平台	语法错误（类型/次数）	漏译字数	漏译类型
ChatGPT	2（标点1，句子结构1）	5	主语
文心一言	0	8	修饰语，谓语
Google	时态不一致1	8	谓语，专有名词
DeepL	时态不一致2	10	修饰语，专有名词
Kimi	0	12	主语，谓语
有道传统翻译		16	主语，修饰语

4.2 流畅性

本研究采用三项可读性公式指标对六个平台的译文流畅性进行对比分析：1)福莱施阅读容易度（Flesch Reading Ease，分数越高表示文本更简洁流畅）；2)自动可读性指数（Automated Readability Index，分数越低译文越易于理解）；3)平均阅读水平共识（Average Reading Level Consensus，分数越低整体阅读难度越低）。统计结果见表5。Google在三项指标上均排名第一（福莱施阅读容易度73.0，自动可读性指数7.95，平均阅读水平共识7.83），整体流畅性最优。文心一言仅次于Google,在福莱施阅读容易度和平均阅读水平共识上分别位列第三，自动可读性指数排名第二，表现为大语言模型中最佳。ChatGPT和Kimi在三个指标上的表现均相对较弱。DeepL得分最低，三项指标均为倒数第一，流畅性最差。同时，另外，20名英语母语者对译文流畅性主观评分整体趋势与上述可读性指标一致。

表4 可读性公式指标分数统计

	福莱施阅读容易度	自动可读性指数	平均阅读水平共识
ChatGPT	65.0	10.28	9.46
文心一言	69.0	8.85	8.58
Kimi	65.0	10.6	9.83
Google	73.0	7.95	7.83
DeepL	61.0	13.07	10.26
有道传统翻译	70.0	10.47	8.56

4.3 文学性

本研究从文本语言特征、打斗动作描述、明喻描写、目标语读者的文化接受度四方面对比分析六个翻译平台译文的文学性。

4.3.1 文本语言特征

本研究基于Coh-Metrix,分析了六个翻译平台在词汇与句法复杂度上的表现。词汇层面考察词汇多样性、词汇高级性和内容词占比；句法层面包括句法依存深度、复杂从属结构占比及平均句长（详见表6）。整体来看，AI大语言模型优于传统在线翻译平台。ChatGPT译文在词汇多样性（405.04，第二）、词汇高级性（418.97，第二）、句法依存深度（8.89，第一）表现突出；Kimi译文在词汇多样性（405.44，第一）、内容词占比（436.17，第一）、句法依存深度（8.30，第二）具优势。传统翻译系统如Google、有道整体成绩较低，如Google译文词汇多样性（397.89）最低，有道译文句法依存深度（7.87）和复杂从属结构（148）均排名倒数第二。

表 5 各平台译文词汇多样性和句法复杂度指标分数统计

	词汇多样性	词汇高级性	内容词占比	句法依存深度	复杂从属结构占比	平均句长
ChatGPT	405.04	418.97	435.58	8.89	178	3.30
文心一言	402.22	418.53	433.80	7.54	198	3.00
DeepL	400.61	419.91	434.49	15.23	142	2.15
Kimi	405.44	414.17	436.17	8.30	173	2.84
Google	397.89	416.81	432.12	8.23	211	3.64
有道传统翻译	402.18	418.86	437.75	7.87	148	3.00

4.3.2 打斗动作描述

ChatGPT 与 Kimi 在动作画面的还原、节奏把控和角色个性展现上表现突出，明显优于其他平台。以罗三炮攻击场面为例，ChatGPT 译为“powerful thunderous flatulence”和“area-effect”，不仅赋予幽默感，还准确传达了招式的广泛影响，增强了场面层次感与紧张气氛。Kimi 在曼陀罗蛇追击描写中，译文“Each contact with the ground was like a springboard, propelling it forward with whip-like agility.”，通过“springboard”和“propelling”精准捕捉蛇的弹射动作，“whip-like agility”细致展现其速度与敏捷，短句结构则强化了追击的连贯与紧迫。相比之下，Google 虽语言流畅，但缺乏动作节奏和人物特色；文心一言描写层次单一，难以还原激烈氛围；DeepL 逻辑衔接不畅，影响场面连贯性；有道对动作和角色细节刻画较为简略。

4.3.3 明喻修辞

《斗罗大陆》中大量明喻描写，使动作和场景更具体生动，增强读者共鸣。ChatGPT 译为“whip-like body made contact with the ground, it would immediately spring like a springboard”，明确对应“身体”为本体，“鞭子”“跳板”为喻体，通过“like”自然建立二者关系，既展现蛇的柔韧，也突出其弹跃，画面感和表现力都很强。相比之下，Google 的译文“its whip-like body would immediately spring up like a spring”虽然保留了比喻，但“spring”多重用法导致喻体重复、关系不清，表达较为生硬，形象性和流畅性不足。

4.3.4 目标语读者的文化接受度

本研究邀请 20 名英语母语者对各平台译文的文化接受度进行评分。结果

显示，ChatGPT 和 Kimi 的译文更符合英语读者对玄幻文学的审美和理解习惯，文化接受度较高。Google 和文心一言整体表现良好，但部分译文被目标语读者认为表达平淡或缺乏文学感染力。DeepL 和有道的译文因表达生硬、细节缺失，接受度相对较低。

5 结论

在“忠实性、流畅性、文学性”三原则的指导下，本研究采用多维评估方法，综合分析了 ChatGPT、文心一言和 Kimi 三款大语言模型，以及 Google、DeepL、有道三款传统在线翻译平台，对网络玄幻小说《斗罗大陆》中某对抗场景的英译质量。研究结果显示，ChatGPT 在忠实性和文学性指标上表现尤为突出，译文能够详细还原打斗场面、准确传达原著氛围，并恰当地再现了修辞和文学色彩。Kimi 同样展现出较强的文学感染力，尤其在词汇多样性和打斗细节还原方面具备一定优势。文心一言整体表现良好，流畅性及部分文学性维度排名靠前，但在打斗节奏与氛围还原方面略逊于 ChatGPT 和 Kimi。在传统翻译平台中，Google 的译文流畅性最佳，文本可读性较高，但在忠实性和文学性方面不及大语言模型。DeepL 则在忠实度上表现良好，但流畅性及打斗细节描写一般，逻辑衔接略显生硬。有道的整体表现最为薄弱，在忠实性、流畅性与文学性三项指标上均位列末位。总体来看，大语言模型，尤其是 ChatGPT 和 Kimi，在中文玄幻文学英译任务中整体表现优于传统翻译平台，不仅译文忠实且完整，还能高度还原原著的风格与文学审美。

这一结论与 Gao 等人（2024）的研究结果相一致，即 ChatGPT 翻译中国古典诗歌时，忠实性、流畅性、文化感知和创新性等方面均显著优于 Google Translate 等传统翻译平台。ChatGPT 优越表现可能得益于大语言模型依托的丰富多元语料、跨文化表达能力及对复杂语境的深度理解，充分彰显了 AI 技术在文学翻译领域的巨大潜力与优势。

值得注意的是，本研究的结论与赵衍等人（2024）

的观点存在差异。赵衍等人认为，在现代小说汉译英任务中，文心一言和讯飞星火两款中文大语言模型不仅优于国内外传统机器翻译平台，也优于 ChatGPT。造成这一分歧的主要原因，可能有以下两点：首先，研究对象的文本体裁不同。本研究聚焦于虚构性强、修辞丰富的网络玄幻小说对抗场景，更注重修辞和氛围的还原。因此，ChatGPT 凭借广泛的跨文化语料和文体迁移能力，展现出更强的文学再现力。相比之下，赵衍等人关注的是贴近现实生活的现代小说，中文大模型在现代话语和本土表达方面语料更充足，因此在现代小说翻译上表现更佳。其次，两个研究所采用的评价体系和模型版本也存在差异，这同样会影响比较结果。

综上所述，不同文本体裁、模型训练语料以及评价标准的差异，都是影响大语言模型翻译表现和研究结论差异的重要因素。

本研究以网络玄幻小说《斗罗大陆》中的一处对抗场景为例，考察了大语言模型与传统翻译平台在汉译英任务中的表现。虽然所选片段囊括了多样的中文句式和词汇，但有限的样本量可能影响结果的代表性和普适性。此外，研究仅聚焦于特定文学体裁和场景。未来可进一步扩大样本范围，涵盖更多文学体裁，并深入比较各平台在译文情感传递等方面的表现，以获得更加全面和深入的研究结论。

参考文献

[1]文旭&田亚灵.(2024).ChatGPT 应用于中国特色话语翻译的有效性研究.上海翻译,(02),27-34+94-95.

[2]王明宪.(2025).网络玄幻小说：中国本土的奇幻创意故事.文学评论,(2),45－58.

[3]王少爽&骆潇洋.(2025).GenAI 时代的智能翻译素养:现实基础、学术理据与概念框架.外语教学,46(01),59-65.

[4]徐立萍,赖雅婷.(2024).网络文学 IP 开发：从“网文”到“网文+”.出版广角,(16),69－74.

[5]赵衍,张慧,杨祎辰.(2024).大语言模型在文本翻译中的质量比较研究：以《繁花》翻译为例.外语电化教学,(4),60－66+109.

作者简介：何雅婷，桂林理工大学外国语学院翻译专业研究生。

指导老师崔丽莉，桂林理工大学外国语学院教师。

本文为桂林理工大学校级教改课题“人机共轭——学科融合——服务东盟：广西高校外语专业人才语言服务能力培养路径研究（2025B20）”和校级数智赋能课程建设“智能思辨写作”的成果之一。