Modern Linguistics 现代语言学, 2026, 14(1), 563-574
Published Online January 2026 in Hans. https://www.hanspub.org/journal/ml
https://doi.org/10.12677/ml.2026.141073

Hans汉斯

在线语料库与AI辅助翻译的“成本－收益”再评估

孙　灵，李升炜

西北师范大学外国语学院，甘肃　兰州

收稿日期：2025年11月27日；录用日期：2026年1月5日；发布日期：2026年1月19日

摘　要

在线文献库极大地提升了翻译研究的资料可及性，但其中存在的影像与转写内容错位、OCR识别误差以及接口与版权限制，使其难以直接作为成熟语料使用；文本工程处理——包括版本甄别、清洗、对齐与归一化——仍是保障学术可靠性的必要前提。与此同时，神经机器翻译与大语言模型显著提高了术语检索、候选生成、对齐核验与风格统一等环节的效率，但也伴随着事实性偏差与体裁漂移的风险。本文提出一条以“权威文本－受控AI－可复核报告”为核心逻辑的写作路径：将在线文献库定位为证据仓库，以联合国官方文件系统(UN ODS)与联合国术语库(UNTERM)为参照基准，借助COMET、BLEURT、MQM等自动化评估工具形成校验共识，并将AI工具限定在可追溯的辅助角色中。研究结论表明，通过权威文本、透明参数与最小可复核单元共同约束工作流程，可以在控制检索、对齐与核验所需人时成本的同时，提升最终结论的可验证性与跨语言可比性。

关键词

在线语料库，语料库译学，神经机器翻译，生成式人工智能，COMET，BLEURT，MQM，证据链，可复核研究

Reassessment of the “Cost-Benefit” of Online Corpora and AI-Assisted Translation

Ling Sun, Shengwei Li

College of Foreign Languages and Literatures, Northwest Normal University, Lanzhou Gansu

Received: November 27, 2025; accepted: January 5, 2026; published: January 19, 2026

Abstract

Online corpora have significantly expanded the accessibility of translation studies, yet challenges

文章引用: 孙灵, 李升炜. 在线语料库与 AI 辅助翻译的“成本-收益”再评估[J]. 现代语言学, 2026, 14(1): 563-574.
DOI: 10.12677/ml.2026.141073

such as image-transcription mismatches, OCR errors, and interface/copyright restrictions prevent them from serving as readily usable linguistic resources. Text engineering—including version authentication, cleaning, alignment, and normalization—remains a prerequisite for academic reliability. Meanwhile, neural machine translation and large language models enhance efficiency in terminology retrieval, candidate generation, alignment verification, and style harmonization, though they introduce risks of factual inaccuracy and genre drift. This paper proposes a writing framework grounded in the "authoritative text-controlled AI-verifiable report" triad: positioning online corpora as evidence repositories anchored by UN ODS and UNTERM, supplemented with evaluation consensus tools like COMET, BLEURT, and MQM, while constraining AI to a traceable assistant role. The conclusion emphasizes that when authoritative texts, transparent parameters, and minimal verifiable documentation collectively govern the workflow, human-hour costs for retrieval, alignment, and verification become controllable, while enhancing the verifiability of conclusions and cross-linguistic comparability.

Keywords

Online Corpus, Corpus-Based Translation Studies, Neural Machine Translation (NMT), Generative Artificial Intelligence (Generative AI), COMET, BlEURT, MQM, Chain of Evidence, Verifiable Research

1. 引言

在线文献库的规模化建设显著降低了跨时段、跨体裁文本的获取门槛，但“可达性”并不自动等同于“可用性”。当原始影像与衍生文本在分页上不对应、OCR 在识别长“s”、连字符及异体字时出现系统性误差，或平台通过视图与字段限制影响批量导出时，未经核验的词频与句法统计极易偏离真实语言使用情况[1] [2]。近年实践表明，ECCO、Google Books 和 HathiTrust 等资源更适合作为证据来源，而非直接用于统计抽样。前两者可用于线索发现与版本追踪，后者则利于影像锚定与交叉验证。而以联合国官方文件系统(UN ODS)为代表的权威资源，因其文号清晰、段落结构明确，天然适合构建“可逆”证据链，并可借助联合国术语库(UNTERM)进行术语回填与标准化处理[3]。这种“来源－版本/文号－段落/页码－链接－访问日期”的链式记录方式，与国内学界倡导的自主知识体系及方法自觉相契合：研究论述与复核路径应在同一文本中实现同构[4]。

神经机器翻译与大语言模型的引入，改变了翻译研究的工作流程，但并未改变科学证据的边界。研究与实践表明，AI 在术语提取、候选生成与对齐核验方面具有显著效率优势，在提供稳定上下文时，也有助于提升篇章连贯性与指称一致性[5]-[7]。然而，译后编辑实践及类 ChatGPT 应用的探索也提醒我们：事实准确性与体裁规范性必须回归权威文本进行锚定，尤其在法规、政策等体裁中，条款编号、日期与专有名词的任何偏差都可能引发累积性后果[8]-[10]。因此，本文主张将 AI 限定为可追溯的辅助工具。其输出在纳入论证前，必须通过“文号－段号”回指联合国官方文件系统(ODS)中的原文，并在尾注中固定链接与访问日期；同时，在正文中需明确标注所使用的模型平台、版本及调用时间等最小披露信息。在效果评测方面，本文以 COMET 与 BLEURT 所评估的语义一致性为主要指标，同时保留 BLEU 与 chrF 分数以维持历时可比性。此外，引入 MQM 错误类型学，将“术语一致性、篇章照应、事实错误”等常见问题转化为可解释的标注类别，以避免单纯依赖分数而导致的评估与现实脱节[3] [11]。

基于上述框架，本文的研究问题相互关联、并行展开：第一，在线文献库在怎样的文本工程与版本甄别条件下，能够从“可获取资源”转化为“可验证证据”；第二，人工智能在何种界限内可带来净效益，而不削弱学术研究的可复核性；第三，如何将评测标准与合规共识转化为可操作的学术体例，使证据链内化为论证的有机组成，而非外部附件。围绕这三个问题，本文在“证据化综述(一)”中探讨影像、转写与版本之间的可逆关系及跨库互证机制；在“证据化综述(二)”中融合语义型自动评测与基于 MQM 的语篇解释，并借助两则联合国文件微型案例(A/RES/75/1 第 1 段；A/RES/71/1 第 41 段)构建“文本事实－评测分数－体裁约束”的闭环验证。方法上，本文遵循小样本、强证据原则，不额外引入人工评估，而是以“最小可复核包”支持研究过程的复查与追溯，具体包括：资源清单与证据定位表(用于来源核验)、文本工程说明(用于过程再现)、参数与评测说明(用于数值可比性)，以及 AI 调用日志(用于生成可追溯记录)。

在学术脉络上，本研究植根于语料库翻译学与翻译质量评估的研究传统，并在此基础上与近年来相关方法论进展保持对话。同时，本文充分关注并吸纳了本土研究在具体语境中所提出的实际问题与理论反思，以使探讨兼具国际视野与在地相关性。本研究的核心目标并非追求样本量的扩张或评估层次的叠加，而是致力于通过构建清晰的证据链与透明的参数体系，将研究中的“成本－收益”量转化为可操作的写作规范，从而在统一的文本框架内，实现研究问题、证据材料与方法流程三者之间的有机结合与相互印证。

2. 研究史与理论支点

自语料库翻译学兴起以来，翻译研究的证据基础经历了从“可读文本”向“可证文本”的范式转变。Baker 以“代表性－可比性”为方法论基石，将语料库方法与“翻译普遍性”命题相结合，确立了以可验证样本支撑理论论证的研究路径[1] [12]。在此基础上，Bowker 和 Pearson 将专门用途语料库的任务导向与术语工作相结合，明确了“面向用途的采样－术语抽取－译前准备”这一操作链，使术语、语境与功能得以系统整合[13]。随后，Zanettin 和 Olohan 分别从“编目－对齐－接口”等技术环节切入，揭示了研究者自建语料库时面临的规模与质量之间的权衡及其再利用困境[14]-[17]。与此同时，McEnery 与 Hardie 则将编码、断句、词形归并及变体统一等环节系统纳入“文本工程”流程，构建了从检索、归一化到可视化的一体化技术框架[2]。上述脉络共同指向一个核心的方法论前提：任何具有可比性的结论都必须基于体裁、版本与对齐单位的明确锚定，而不能以平台规模或检索便利性替代证据本身的可靠性。

与语料库研究范式相并行，翻译技术谱系经历了从统计方法向神经方法的演进。Koehn 通过对统计机器翻译(SMT)与神经机器翻译(NMT)之间延续性与断裂性的谱系梳理，为该领域的纵向比较提供了清晰的时间轴线与术语框架[5] [18]。2010 年代后期，围绕“机器翻译加译后编辑能否达到人工基准”的讨论逐渐转向实证。Toral 和 Way 在文学体裁中检验了神经系统的潜力与局限[19]，而 Castilho 等人通过跨体裁、跨任务的系列研究，揭示了“速度提升－质量风险”之间存在的非线性关系，并强调体裁与任务边界对研究结论外推具有关键制约[20] [21]。这一阶段的实证工作将讨论焦点从单纯的“可达性”重新引向“可用性”：研究者必须在候选生成的效率优势与语篇结构的整体稳定性之间，做出可验证、可复核的权衡。质量评估理论为上述技术路径提供了价值坐标与解释框架。House 提出的质量评估模型将“何为优质译文”的判断置于语用与交际功能之中[8]；Skopos 理论则强调目的的优先性，主张在具体情境中根据目的与规范进行裁定[22]；Toury 的描述翻译学则以“规范－操控”为核心，系统解释译文选择背后的社会文化制约[23] [24]。在文化与话语层面，Venuti 关于归化与异化的讨论，将语域与体裁的社会文化维度重新引入翻译评价的视域[25] [26]。

自动评测方法也经历了相应演进。传统指标如 BLEU 与 chrF 在保持历时可比性方面仍有其价值[27] [28]，但对高性能系统的区分度逐渐有限；Post 为此提出 SacreBLEU，通过统一参数与脚本来修复评测的

可比性[11]。面向语义一致性的新一代指标，如 COMET 与 BLEURT，借助跨语言表征和学习型打分机制，显著提升了与人工评价的相关性[29] [30]。与此同时，MQM 框架将“错误位置、类型与严重程度”转化为可共享的标注体系，并通过与多届 WMT 人工评估实践的对接，赋予自动分数以语篇和体裁层面的解释力[3]。总体而言，翻译质量评测已从早期的“表层形式匹配”转向“语义一致性–错误类型学解释”相结合的范式。这一转向使得人工智能介入翻译所带来的收益与风险，得以在可验证的语言描述和系统比较中被更清晰地刻画与评估。

在平台层面，在线语料的学术演进历程一再印证：平台的规模与检索的便捷性，并不直接等同于证据的可靠性。影像、转写与版本之间的错位，可能削弱基于其开展的统计分析与文献引证的稳定性。因此，在线文献库更宜被视为“证据来源”，而非可直接使用的“就绪语料”。相较于简单依赖“抓取–统计”的研究捷径，当前实践更强调功能化分工：借助转写或派生文本进行线索发现与初步定位，再回归原始影像完成最终引证与定稿；在法规、政策等严肃体裁中，则优先依据“文号–段号/条号”的层级结构组织文本对齐，以保障事实准确性与体裁规范的可验证性。这一“证据链”导向的研究思路，也与国内学界围绕翻译学知识体系建设的讨论相呼应：方法论的选择与学术体例的构建，应共同服务于可追溯的证据、可解释的方法与可复核的结果，形成合力推进研究的透明与可信[4]。

中国语境中的本土化研究，为上述框架提供了具体化补充与方法论校正。过程研究与语料库建设方面的成果提示了取样、断句与对齐等环节的工程性要求[31] [32]；行业调研则揭示了术语表、风格指南与一致性检查在实际应用中的价值及其落地瓶颈[33]。在系统层面，相关评估与应用研究强调应推进计算机辅助翻译与质量评估体系的对接[10]；而针对生成式人工智能与译后编辑的个案研究进一步提醒，效率提升不应以牺牲事实准确性与体裁规范性为代价[9]。在口译领域，机辅系统的设计与验证表明，“技术辅助”与“译员判断”之间存在可调的协同空间，方法设计应优先保障语篇整体一致性及术语回填的可追溯性[34]。将上述本土化成果与国际评测及理论传统相结合，便形成本文所采用的体例架构：以联合国官方文件系统(UN ODS)与联合国术语库(UNTERM)锚定事实与术语基准；以 COMET 或 BLEURT 等语义一致性指标配合 MQM 错误类型学提供解释维度；最终将“证据仓库–受控 AI–透明报告”整合为一条可操作、可复核的写作路径。

3. 证据化综述(一)：在线库的证据逻辑与文本工程

在线文献库对于翻译研究的根本意义，并不在于其能否便捷地输出可供统计的文本数据，而在于其能否为学术论断提供一个稳定、可追溯且允许交叉验证的实证基础。研究的可靠性并非源自对平台一次性抓取结果的依赖，而是建立在由“文献来源–版本标识–具体位置(页码/段号)–持久链接–获取时间”所构成的完整证据链条之上。这一认识与语料库翻译学内在的学术关切深度契合：自 Baker 的早期研究以来，该领域始终强调，任何基于量化的发现都必须以严谨、透明的文本处理流程为前提，而不能将平台的技术便利性等同于论证的学术有效性[1] [12]。同时，这也与国内学界在构建翻译学知识体系过程中，所倡导的“以可验证证据为基石、以标准化流程为规范”的学术写作体例要求形成了呼应[4]。

以 EEBO-TCP 为例，其基于人工校对的 SGML 和 XML 转写文本，在早期英语的词形识别与版面还原上优于通用 OCR，适用于词形变体检索与历时比较。然而，由于转写文本与原始影像在分页上并非严格对应，且同一作品常存在多个影像版本，因此关键引证仍需回归影像页面进行核实，并在正文或注释中同时注明所用版次与固定访问链接(Text Creation Partnership 2015, 2020)。这种“文本与影像双轨并进”的做法并非冗余，而是将“检索效率”与“引证可信度”进行功能分离：转写文本用于线索发现与初步定位，影像文本则承担最终证据支撑与定稿依据。

类似地，ECCO 与 Google Books 凭借其规模与检索便利，适于追踪版本线索、考察术语首次出现及

表述的跨版本流变，但其受限于版权视图(完整/预览/片段)与 OCR 质量，不宜直接作为统计分析的原始语料(Gale n.d.; Google Books n.d.)。所以更可靠的做法是先借助这类平台完成初步发现，再转向 HathiTrust 或机构馆藏影像以确定具体版本与页码对应。HathiTrust 提供的公共领域影像较为稳定，且平台明确说明 OCR 质量因文献来源与年代而异，这一坦诚的声明反而可被研究者用作影像锚定与跨库互证的依据：当片段预览信息不足时，可借助完整视图精确定位；遇连字符切分错误等问题时，则可退回影像核对，并据此修正检索式或统计脚本(HathiTrust n.d.)。Internet Archive 提供的派生文本(如 EPUB、DjVu)利于快速浏览与定位，但其生成仍高度依赖 OCR，因此量化分析与正式引证仍需以 PDF 影像为准。该平台的方法价值更体现在版本追踪与跨镜像互证，而非作为可直接统计的“洁净文本”(Internet Archive n.d.)。Project Gutenberg 的文本经过多轮志愿者校对，可读性较高，然而由于元数据与版本来源记录不一，重要引证同样需要回溯至影像版本进行复核(Project Gutenberg n.d.)。

联合国官方文件系统(UN ODS)与前述主要收录单语历史文本的资源有所不同，它在跨语言并行性与文档结构层级两方面提供了近乎理想的证据支持环境：以文号为主键统一呈现序言、正文与具体条款，且各语言版本在结构上完全对应，使研究者能够借助“文号–段号/条款号”构建起最小、最精确的证据单元。结合联合国术语库(UNTERM)，术语一致性的判断得以从依赖经验转向可查询的官方条目，从而实现术语标准化与事实核查的系统联动(United Nations ODS n.d.; UNTERM n.d.)。在涉及平行文本研究、AI 辅助译后编辑及术语一致性检查等任务中，这一“回填至权威文本”的路径尤为关键，它将所有候选译文或改写文本重新置于权威原文的约束之下，从而有效避免因追求“语言流畅”而掩盖事实错误或体裁失准的问题。

将上述定位转化为可执行的研究能力，关键在于将文本工程有机嵌入研究设计之中。在采样层面，首先应明确体裁边界(如法规、政策通知、学术著作)与时间跨度，并区分“版本研究”(以版次/印次为线索构建谱系)与“现象研究”(以体裁或主题为线索确保内部同质)的不同抽样逻辑。混合多体裁样本易使语域与篇章结构的系统性差异渗入统计过程，从而削弱结论的解释力。在清洗与归一化层面，针对历史英文文本，需系统处理“长 s”标准化、连字符重建、拼写变体映射(如 colour/color)，以及“i/j”、“v/u”混淆的校正；中文文本则需统一字符编码与标点样式，规范阿拉伯数字与汉字数字的混用，以及全半角符号、日期格式等不一致现象。扫描 PDF 在进入量化分析前，应进行 OCR 质量评估，对置信度较低或版式复杂的区域安排人工核对，并编制“可疑区段清单”作为附录，以便后续核查与验证。

在断句与对齐层面，应避免仅依据句点进行机械切分。中文法规与政策文本通常以“条–款–项”的层级结构组织语义，宜采取“以条款/段落为主、以句子为辅”的分层对齐策略，先确立条款或段落之间的宏观对应，再在各条款内部进行句子级切分与映射。英文文本中的长从句与并列结构常使简单句点切分破坏修饰关系与核心谓语结构，应结合“that/which”等关联词、非限定性从句以及体裁标记进行二次切分。对于平行文本，应优先确保结构同构：例如，当中文使用“(a) (b) (c)”等标号分项列举，而英文采用分号串联表达时，应以“并列项”作为对齐单元，而非简单追求逐句对应。

在版本锁定与去重层面，跨库采集往往会产生多份页式不一的数据镜像。建议以“书目–版次–页码–行/段”作为复合主键制定去重规则，优先保留影像质量高、元数据完整的版本，并在研究日志中明确记录判断依据与选择理由。若同一作品存在作者修订或出版方重排等不同版本，应在文中明确说明纳入统计的具体版本，并对可能受版本差异影响的指标加以标注和说明。在证据标注层面，正文中应以“来源–页码/段号”的形式即时标识引证位置，尾注则需补充机构信息、版本/文号、固定链接及访问日期，确保读者能够逆向追溯至原始文本。平行文本在首次引用时，应同时提供对应文号与段号的双语链接(如 ODS 英文页与中文页)；术语首次出现时，宜同步标注其在 UNTERM 中的对应条目，从而形成“术语–文本出处–术语库条目”三者互证的闭环结构。针对历史文本中的关键引文，建议同时提供影像页的永

久链接与转写文本的段落定位，以便于后续对照印刷差异与识别 OCR 或转写过程中可能产生的误差。

跨库互证与误差分析是统计分析前后关键的质控环节。在统计开始前，可借助 Google Books、HathiTrust 及 Internet Archive 等多源平台相互校验，以确认文本完整性与分页一致性；统计分析后，则应对贡献度较高的词项及关键共现结构进行随机抽样，回溯至影像页面核查因 OCR 识别、连字符处理或历史拼写变异所导致的误差比例，并在方法部分说明此类误差对统计指标的潜在影响。即使是 EEBO-TCP 这类经过高质量人工转写的语料，仍建议保留一定比例的影像复核机制，以避免将转写过程中的规范处理误判为原始文本的语言事实(Text Creation Partnership 2015, 2020; Google Books n.d.; HathiTrust n.d.; Internet Archive n.d.)。

文本工程的成果输出不应止于一份“处理后的文本”，而应形成一套完整的方法文档，至少包括：证据台账(逐条记录来源、版本、页码/段落、链接及访问日期)、对齐对照表(呈现条款或句子层级的映射关系及例外情况)、清洗规则说明(列明所用正则表达式、拼写变体映射表、术语白名单等)、误差分析报告(说明抽检比例、主要错误类型及处置方式)以及研究日志(记载版本选择依据、去重判断标准与特殊情形处理)。这些材料共同构成本文在方法部分所设立的最小可复核包，旨在使第三方研究者无需依赖特定平台权限，仅凭公开可得的链接与明确记录的规则即可完整复现研究过程。这一做法也呼应了国内学界倡导的“以可复核性支撑学科话语体系构建”的方法自觉[4]。

在此框架下，UN ODS 与 UNTERM 的耦合尤为体现“证据仓库”的价值：ODS 通过文号与段落结构确保事实可追溯，UNTERM 则提供可供检索的规范术语条目。当 AI 辅助译后编辑或一致性检查时，任何候选译文均可“回填”至 ODS 原文，并依据 UNTERM 的标准进行校正。例如，若将“fundamental freedoms”错误合并至“basic rights”，即可依据 ODS 原文与 UNTERM 条目判定为语义范围缩减或术语使用不当，并在研究方法部分说明该判断如何纳入误差分析及对结论的限定(United Nations ODS n.d.; UNTERM n.d.)。

需要强调的是，将在线文献库定位为“证据仓库”，并非否认其在语料构建中的作用，而是倡导更清晰的功能分工：原始影像提供证据的可靠性，转写或派生文本提供检索与计算效率，二者通过版本锁定与跨库互证相结合，方能在统计分析与文献引证之间建立稳健的桥梁。EEBO-TCP 的高质量转写不能替代影像核验；ECCO 与 Google Books 的规模优势不能掩盖其 OCR 与元数据偏差；HathiTrust 和 Internet Archive 的影像稳定性不意味着可直接用于统计；Project Gutenberg 的志愿者校对亦不等同于版本同一性得到保证。唯有在文本工程层面系统完成“分层定位–跨库互证–权威回填”，在线文献库才能真正从“文本提供者”转变为“证据生成器”(Text Creation Partnership 2015, 2020; Gale n.d.; Google Books n.d.; HathiTrust n.d.; Internet Archive n.d.; Project Gutenberg n.d.)。

因此，在线文献库的证据运用逻辑可归纳为以下三个相辅相成的原则：第一，影像优先、文本助检，确保所有引证与统计最终可回溯至原始影像；第二，结构同构、层级对齐，维护法规、政策等刚性体裁的篇章结构一致性；第三，跨库互证、可逆披露，系统控制 OCR 及元数据偏差，并落实研究过程的可复核性。在此基础上，结合 UN ODS 与 UNTERM 所提供的术语–事实权威回填机制，在线文献库在人工智能时代不仅不会因“自动化”而削弱其学术价值，反而能够为后续关于翻译质量评测以及 AI 受控使用的论述，提供更为扎实、可信的证据基础(United Nations ODS n.d.; UNTERM n.d.)。

4. 证据化综述(二)：AI 的受控使用、评测共识与语篇约束

自神经机器翻译与大语言模型进入翻译领域以来，工作分工的变化主要表现为环节的“前移”而非“替代”：术语提取、候选生成、对齐检验、风格统一与文档级一致性预警等高耗时任务得以压缩，使文本工程与证据构建能够更早展开。然而，效率的提升并不自动转化为证据效力的增强。若缺乏来源锚定

与体裁约束，流畅的生成结果反而容易掩盖术语偏移、指称错位、条文误引等实质错误，进而增加后续复核的成本。因此，AI 在研究写作中的定位，必须由证据链与语篇规范共同界定：来源可追溯、体裁可验证、参数可公开，三者缺一不可[4] [8]。

在术语处理层面，受控使用首先要求将“可接受译名集合”明确化为术语白名单，并将其与文档级一致性要求相绑定。神经系统在封闭语域内虽具备较强的模板复现能力，但在跨域迁移或需要高频跨段落复现时，往往容易出现术语语义范围的窄化或泛化。以 UNTERM 为基准，明确区分“首选－允许－禁用”三类表达方式，并在文中追踪术语“首次出现－跨段复现”的轨迹，能够将术语判断从直觉争论转化为可验证的依据，同时生成“覆盖率”与“一致性”两项量化指标，作为自动评测的解释维度[29] [30]。这种基于规范的本体化处理方法，使术语判断回归到规范层面，避免因模型的“顺滑化”倾向而导致术语信息密度的降低。

在句法与信息结构层面，神经系统呈现出另一组内在张力。其在处理并列结构、非限定定语从句与插入语等局部重组任务时表现日趋稳定，若提供充分上下文，句际衔接亦可得到改善[35]。然而，当输入文本被强行以句子为单位切分、跨段落的指称链拉长，或先行成分被省略时，修饰语的归属关系与主位－述位的信息推进模式，往往被系统的“顺滑化”策略所掩盖。对于法规、政策等体裁，条款层级具有刚性约束力，任何在句子层面进行的重写若破坏了这一层级，都可能导致篇章结构的不可逆损伤。因此，受控使用要求以条款或段落作为主要的对齐与复核单位，确保句子级生成严格服从整体结构同构。同时，应建立以“机构－国家－文书－条款”为骨架的指称链标注体系，跨段落检验指称的延续性与唯一性[19]-[21]。

在事实性与引证层面，神经模型尤其容易在条文编号、日期及专有名词等细节上“自信地”产生错误，当面对语义相近但实质不同的条文时，系统往往倾向于基于语义邻近性生成内容，而忽略事实核查。为此，应在正文中通过标注“文号－段号”建立明确的回指路径，使读者能够逆向追溯至权威原文；同时在尾注中固定 ODS 链接与访问日期，将每一次事实判断与一次具体的、可复核的访问记录相绑定，从而在论证层面阻断错误传播[11]。例如，在 A/RES/75/1 第 1 段中，日期、主语与纪念对象在中英文版本之间严格对应；任何将“heads of state and government”简化为“leaders”的改写，均属于语类降格，会削弱政治实体的专指性与条文整体的严谨性。又如 A/RES/71/1 第 41 段，“安全、尊严、人权和基本自由/safety, dignity and human rights and fundamental freedoms”的并列结构不仅是语义上的权利枚举，也体现了该类文本的体裁惯例；若将“fundamental freedoms”并入“basic rights”，不仅压缩了术语的语义边界，也重构了权利的表达结构。此类判断必须通过 UNTERM 条目与 ODS 原文进行双重回填，从而在术语、事实与语篇三个层面上形成一致且可验证的约束。

自动评测为受控 AI 的使用提供了量化参考，但任何分数都必须置于具体的语篇语境中予以解读。传统指标如 BLEU 与 chrF 在保持历时可比性方面仍有其价值[27] [28]，但对区分高质量系统间的细微差异能力有限；而基于语义的评测方法如 COMET 和 BLEURT，凭借其跨语言表征能力，能更敏锐地识别“语言通顺但含义有偏”的现象，与人工评价的一致性也更高[29] [30]。然而，若脱离体裁特征、对齐单位与证据路径的说明，分数的升降容易与文本实际内容脱节。针对法规性文本，应首先检查条款编号是否完整、“权利－义务－例外”的三分结构是否得以保持，以及文内引用与回指是否可追溯；而对于新闻通稿或政策解读类文本，则需评估信息核心重组是否忠实、语域风格是否匹配。因此，有效的评测应结合句级语义指标与文档级一致性特征，并借助 MQM 框架中的“术语－篇章－事实性”等错误类型标签，对主要问题进行可描述、可归因的分析[3]。只有通过这种多维度的评估方式，才能将抽象的分数对应到具体的文本现象，避免“指标与现实”之间的错位。

从方法论而言，受控 AI 的运用并非走向“机器”与“人工”的对立，而是通过将技术流程嵌入研究环节，实现分工的重构与优化。在候选生成阶段，需记录所用平台、模型、版本及调用时间，并保留代表性样例；术语回填阶段，则以预设白名单比对生成结果，标注冲突条目与未覆盖项；事实核验阶段，严格依据“文号–段号”回指至 ODS 页面进行对齐校验，必要时可并列中英文页面以兼顾跨语言读者的核查需求；篇章一致化阶段，则先复核指称链、编号体系及跨段复现情况，再施行最小必要的风格调和。每一环节均需保留操作痕迹，从而形成“原句–AI 候选–人工裁定–证据锚点”的完整可逆路径。与此配套，应建立“风险–缓释”对照机制：针对术语漂移，设置白名单与复现频次阈值；针对编号错引，施加文号–段号的硬性约束；针对日期错位，采用规范化解析与双向校对；针对指称错配，实行先行项标注与跨段一致性检查；针对过度顺滑，则制定并遵循体裁特有的名词化或被动语态保持规则。由此，效率提升的速度红利方能转化为可解释、可复现的工程红利，使研究者的精力聚焦于“专业裁定”而非“低效搜寻”[5] [21]。

在中文场景下，语篇层面的约束尤为需要凸显其必要性。法规、政策与规范性文本中的层级编号本身承载着特定的语义功能，句点在此类文本中并非天然的意义切分界限；采用以条款为主、句子为辅的对齐策略，有助于在译文评测与后续修订中保持原文的篇章结构与逻辑骨架。中文常使用零照应与成分省略，这往往要求回推话题来源，而模型在自动补全时容易引入推测性的先行项，与法律、政策文本所要求的严谨性相冲突。此外，数字与日期的格式差异、中文数字与阿拉伯数字的混用、量词系统的选择以及名词化表达的节奏特点，均需在文本清洗与归一化阶段先行统一，并在生成过程中受明确的规则约束。唯有在此基础上，模型的输出才能在形式上与源文本保持可比，在事实上与权威文本达成一致[8] [25] [26]。

5. 明晰 AI 辅助的局限和边界

必须清醒认识到，模型能力存在内在边界，某些场景下不宜让 AI 承担决定性角色。跨体裁迁移引发的语域漂移、长文档中指称链的断裂、相近条文间的编号错乱等问题短期内仍难以彻底解决。尽管检索增强生成和文档级上下文处理能在一定程度上改善文本的连贯性与事实核查，但在法规、政策等高风险的文本类型中，只有与权威文本回填机制紧密结合，AI 的辅助才具备学术可信度。将术语白名单、指称链标注和事实核验规则，前置为提示词模板或解码约束，以降低后期修订成本；同时推动 MQM 错误标签与自动评测指标的联动，生成具备解释力的质量报告，将抽象的分数变化具体映射为术语、篇章和事实性等维度的提升或退化[3] [11]。但无论技术如何演进，受控 AI 的底层逻辑应始终如一：判断标准源于文本事实，评价尺度基于可靠评测，应用边界取决于体裁规范；所有由模型生成的“建议”，都必须经由证据链牵引回归至权威文本完成最终裁定[4]。

因此，在使用 AI 时，我们也必须建立一种兼具动态性与前瞻性的视角：其能力边界并非固定不变，而是正随着大语言模型等技术的迅猛发展持续迁移与拓展。当模型在事实核查的准确性、长文本的深度理解以及复杂指令的精准遵循等核心能力上取得实质性突破时，AI 的角色将发生根本性转变——从一个需要严格监控的被动工具，演变为一个具备初步认知与执行自主性的“协作者”甚至“伙伴”。这种演进迫使我们对既有的“人机边界”划分与“控制”策略进行深刻的反思和系统性重构。传统的、基于“人类处理复杂抽象任务、机器处理重复规则任务”的静态分工模式将逐渐失效，边界将向更高层次、更具创造性与责任性的领域上移。

人类的角色重心，也必须从弥补机器的“不足”，如纠正事实错误和理解长文本，转向驾驭机器的“能力”，即进行最终的价值判断、伦理权衡、创造性决策以及对后果承担绝对责任。这意味着，即使 AI 能够提供逻辑严密、证据详备的分析和建议，人类仍需牢牢掌握设定目标、审查伦理合规性并在关键

节点做出裁定的终极权威。与此同时，控制策略的范式也需要同步升级，从主要依赖结果端的外部校验与过滤，转向深度融合于系统设计、运行与交互全过程的协同治理。这包括在技术层面嵌入“可解释性开关”、在关键决策流程预设“动态人工介入节点”，并建立完整的输出追溯机制，使 AI 的推理链条与决策依据透明化、可审计。在伦理与法律层面，则需明确“技术赋能”与“责任赋权”的不可逾越的界限，任何情况下都不应将关乎重大利益的最终决策权单独让渡给 AI 系统。

综上所述，AI 在翻译研究中的价值并非追求以“替代”为名的捷径，而是构建一种以证据与规范为前提的深度协作。以联合国文件系统(UN ODS)与术语库(UNTERM)为来源锚点，可使术语选择与事实判断始终具备可溯性；依靠 COMET、BLEURT 等语义指标与 MQM 错误类型学形成的评测共识，则让自动评分结果获得可解释性；而依托“文号－段号－固定链接－访问日期”所建立的可逆路径，确保了每一处论述皆可被复核与验证。将 AI 稳固定位于“可追溯的助手”，而非自主决策的主体，方能在效率提升与学术可靠之间，建立一种可持续、可检验的平衡[3] [5] [8] [12] [29] [30]。

6. 方法、伦理与最小可复核包：从操作步骤到统计控制

要使“证据链－评测－语篇”在一项研究中形成可检验的统一体，方法不应仅是文末附上的“技术说明”，而应作为贯穿全文的写作逻辑先行确立。所有进入统计与论证的材料，都必须受到一条可回溯的证据路径约束；所有用于比较与评价的量化分数，都必须在具体体裁和语篇结构中获得解释；所有调用自动化工具的过程，都应以透明的参数和明确的边界标示其适用范围。这样的体例设计并非对已有惯例的简单拼贴，而是将“概念－术语－话语”的学科表达，与“流程－参数－留痕”的技术治理有机结合，使本土学术叙事与新技术条件相互协同[4]。在样本组织原则上，法规与政策文本中的“条－款－项”并非仅是版式设计，而是承载论证逻辑的篇章单元。因此，条款或段落应作为对齐与分析的首要层级，句子切分仅在此框架内进行。中文材料尤需避免仅凭标点进行断句，因为零形指代与编号嵌套常将关键信息“嵌入”在层级结构内部；平行文本的对齐同样应遵循结构同构优先原则——并列项目、编号体系及跨段回指关系，比逐句机械对应更能保障事实准确与逻辑完整。

将层级优先原则写入研究方法，实质是把“何以为证”的判断标准嵌入文本工程的每个环节，使得后续所有统计分析与文献引证，都能在同一框架下接受复核[4]。文本工程的要点不在“多做几步”，而在“把关键环节变成可检约束”。跨时段英文文本需要先行完成“长 s”归一、连字符重组与拼写变体映射；中文文本则要统一全/半角、数字书写与日期格式。看似“预处理”的操作，实则是在为后续的术语一致与搭配比较清除系统噪音。更重要的是，专名与关键术语不应只以比例指标出现，而要与权威条目相挂钩：当研究对象涉及国际组织文本时，以 UNTERM 为白名单来约束“首选－允许－禁用”的表达，把术语一致从“通顺”提升为“可证”。这既回应了学科层面对“术语－概念－话语”三者协同的要求，也为后续的自动评测提供可解释支点[3] [4]。

清洗、去重与误差分析的作用，在于为“分数为何波动”提供来自文本层面的解释。跨库采集常导致同一文献的多个镜像或不同版本并存，若不依据版本或文号进行锁定与去重，统计分析便可能无意中重复计算同一内容。扫描 PDF 生成的派生文本应在入库前进行 OCR 质量抽检，对低置信度段落回溯至原始影像核对，并将典型处理案例记录于附录。由此建立的误差剖面，并非仅为“补充材料”，而是将评测中的异常波动(如文档级 COMET 或句级 BLEURT 的局部下降)关联到具体的文本特征－例如排版差异、连字符处理或历史拼写变体，避免将一切表现差异笼统归因于“模型失效”。这种从文本实际出发的可解释评估路径，与立足于本土话语体系组织分析与论证的学术取向相一致：数据差异须在语篇结构与体裁规范中得到具体说明，而非停留于抽象的数字层面[4] [29] [30]。

参数披露与随机性控制，是将“可比性”内化于研究体例的关键。凡使用自动评测，须明确说明文

本切分方式、分词规则、大小写处理、所用脚本及版本号；解码策略(如温度、束宽)与随机种子亦应固定并完整记录。在小样本比较中，建议辅以 bootstrap 或置换检验给出置信区间；进行跨文档比较时，则应同时汇报宏观(文档级)与微观(句级)的聚合结果，以揭示体裁与文本长度对分数分布的影响。更重要的是，需将自动评分与可解释的标注维度相结合：可依据 MQM 框架，对术语、篇章、事实性三类错误进行抽样标注，并与术语白名单覆盖率、指称链稳定性、编号完整率等文档级特征并列分析。借此，每一次分数的变化都能在文本层面被定位、归因与复述。由此，评测便从单纯“分数高低”的外在比较，转向对“何处出错、错误何类、程度何如”的内在解释[3] [11]。

伦理与合规问题并非“只要不涉及个人数据即可豁免”，而在于明确揭示模型在研究中的参与程度及其边界：凡使用人工智能的环节，均应在正文或注释中提供最小必要的信息披露，包括平台与模型名称、版本与调用时间、核心参数及运行环境等。对于法规、合同、公告、国际组织对外文本等高风险体裁，必须严格限定 AI 的使用范围——仅可用于术语初步筛选、候选译文比对与一致性检查等辅助环节，不得直接输出作为定稿。若采用检索增强生成或结构化提示模板，亦须说明所用语料来源、召回阈值及结果过滤规则。这种“将技术纳入体例规训之中”的做法，正体现了在新技术条件下推动知识体系建设、实现研究范式协同与方法融合的必然要求[4]。

在这一框架下，“最小可复核包”(MRP)不再只是文末的补充材料，而是伴随全文呈现的一种可验证的能力交付。它具体包含以下组成部分：资源清单将机构、题名、版本/文号、固定链接与访问日期串联为一条可回溯的路径；文本工程说明以可操作的语句阐述采样、清洗、对齐与归一化的关键规则，并提供示例以说明如何复现；参数与评测说明记录所用指标、脚本版本、随机种子与解码策略，并附上区间估计及对异常结果的解释；证据定位表以“文号 - 段号 - 页码/段落编号”为索引，贯通正文引证与尾注链接；AI 调用日志则通过“原句 - 候选 - 裁定 - 证据”四列式样例，展示如何将生成性建议回填至权威文本。MRP 的根本意义，在于将“可信、可比、可复核”转化为写作结构的内在属性：读者无需依赖作者的主观承诺，即可沿既定的、透明的路径完整复现研究过程。这种将研究方法与伦理要求前置为体例约束、将证据链与参数设置固化为文本结构的做法，使得在线文献库能够稳定作为“证据仓库”，AI 被严格限定为“可追溯的助手”，而研究者与读者则在同一种可检验的规则与路径中展开对话[4]。

7. 结语

本研究将在线文献库重新定位为“可验证的证据仓库”，以结构化证据链为基础前置文本工程，使研究在源头具备可复核性。AI 被严格限定为“可追溯的助手”，其输出必须回填至权威文本并全程留痕，从而将效率提升转化为可检验的工程价值。评测体系融合句级语义指标与文档级一致性特征，借助错误类型学将分数变化解释为术语、篇章与事实性的具体差异。最终，通过“最小可复核包”与权威来源锚点，本研究构建了一种“可透明、可重现”的写作体例，在控制成本的同时，为后续跨语、历时与检索增强研究提供了可衔接的路径与规范。

参考文献

[1] Baker, M. (1995) Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target. International Journal of Translation Studies*, **7**, 223-243. https://doi.org/10.1075/target.7.2.03bak

[2] McEnery, T. and Hardie, A. (2012) Corpus Linguistics: Method, Theory and Practice. Cambridge University Press. https://doi.org/10.1017/cbo9780511981395

[3] Lommel, A., Uszkoreit, H. and Burchardt, A. (2014) Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. *Tradumàtica tecnologies de la traducció*, **12**, 455-463. https://doi.org/10.5565/rev/tradumatica.77

[4] 李晓倩. 中国翻译学知识体系的构建: 主要议题与未来发展[J]. 中国翻译, 2025, 46(3): 27-33.

[5] Koehn, P. (2020) Neural Machine Translation. Cambridge University Press. https://doi.org/10.1017/9781108608480

[6] 王金铨. 计算机辅助翻译评价系统中的翻译质量评估[J]. 上海翻译, 2023(6): 52-57.

[7] 王巍巍. 中国语言服务行业应用人工智能辅助机器翻译工具的现状调研[J]. 外语电化教学, 2025(2): 25-30, 100.

[8] House, J. (2015) Translation Quality Assessment: Past and Present. Routledge.

[9] 耿芳, 胡健. 人工智能辅助译后编辑新方向——基于 ChatGPT 的翻译实例研究[J]. 中国外语, 2023, 20(3): 41-47.

[10] 周兴华, 王传英. 人工智能技术在计算机辅助翻译软件中的应用与评价[J]. 中国翻译, 2020, 41(5): 121-129.

[11] Post, M. (2018) A Call for Clarity in Reporting BLEU Scores. *Proceedings of the Third Conference on Machine Translation*: *Research Papers*, Brussels, 31 October-1 November 2018, 186-191. https://doi.org/10.18653/v1/w18-6319

[12] Baker, M. (1993) Corpus Linguistics and Translation Studies—Implications and Applications. In: Baker, M., Francis, G. and Tognini-Bonelli, E., Eds., *Text and Technology*, John Benjamins Publishing Company, 233-250. https://doi.org/10.1075/z.64.15bak

[13] Bowker, L. and Pearson, J. (2002) Working with Specialized Language: A Practical Guide to Using Corpora. Routledge. https://doi.org/10.4324/9780203469255

[14] Olohan, M. (2004) Introducing Corpora in Translation Studies. Routledge. https://doi.org/10.4324/9780203640005

[15] Olohan, M. (2016) Scientific and Technical Translation. Routledge. https://doi.org/10.4324/9781315679600

[16] Zanettin, F. (2012) Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies. Routledge.

[17] Zanettin, F. and Rundle, C. (2022) The Routledge Handbook of Translation and Methodology. Routledge. https://doi.org/10.4324/9781315158945

[18] Koehn, P. (2010) Statistical Machine Translation. Cambridge University Press. https://doi.org/10.1017/cbo9780511815829

[19] Toral, A. and Way, A. (2018) What Level of Quality Can Neural Machine Translation Attain on Literary Text? In: Moorkens, J., Castilho, S., Gaspari, F. and Doherty, S., Eds., *Translation Quality Assessment*, Springer International Publishing, 263-287. https://doi.org/10.1007/978-3-319-91241-7_12

[20] Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J. and Way, A. (2017) Is Neural Machine Translation the New State of the Art? *The Prague Bulletin of Mathematical Linguistics*, **108**, 109-120. https://doi.org/10.1515/pralin-2017-0013

[21] Castilho, S., Moorkens, J., Way, A. and Gaspari, F. (2020) Machine Translation and Post-Editing in Practice. Springer.

[22] Reiß, K. and Vermeer, H.J. (1984) Grundlegung Einer Allgemeinen Translationstheorie. Niemeyer. https://doi.org/10.1515/9783111351919

[23] Toury, G. (1995) Descriptive Translation Studies—And Beyond. John Benjamins Publishing Company. https://doi.org/10.1075/btl.4

[24] Toury, G. (2012) Descriptive Translation Studies—And Beyond. 2nd Edition, John Benjamins Publishing Company. https://doi.org/10.1075/btl.100

[25] Venuti, L. (1995) The Translator's Invisibility: A History of Translation. Routledge.

[26] Venuti, L. (2017) The Translator's Invisibility. 2nd Edition, Routledge.

[27] Papineni, K., Roukos, S., Ward, T. and Zhu, W. (2001) BLEU: A Method for Automatic Evaluation of Machine Translation. *Proceedings of the* 40*th Annual Meeting on Association for Computational Linguistics—ACL*'02, Philadelphia, 7-12 July 2002, 311-318. https://doi.org/10.3115/1073083.1073135

[28] Popović, M. (2015) chrF: Character N-Gram F-Score for Automatic MT Evaluation. *Proceedings of the Tenth Workshop on Statistical Machine Translation*, Lisbon, 17-18 September 2015, 392-395. https://doi.org/10.18653/v1/w15-3049

[29] Rei, R., Stewart, C., Farinha, A.C. and Lavie, A. (2020) COMET: A Neural Framework for MT Evaluation. *Proceedings of the* 2020 *Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), 16-20 November 2020, 2685-2702. https://doi.org/10.18653/v1/2020.emnlp-main.213

[30] Sellam, T., Das, D. and Parikh, A.P. (2020) BLEURT: Learning Robust Metrics for Text Generation. *Proceedings of ACL* 2020, 5-10 July 2020, 7881-7892. https://aclanthology.org/2020.acl-main.704/

[31] 侯林平. 语料库辅助的翻译认知过程研究模式: 特征与趋势[J]. 外语研究, 2019, 36(6): 69-75.

[32] 刘晓东. 认知导向的翻译语料库研制与评析[J]. 外语学刊, 2023(4): 52-60.

[33] 王华树, 刘世界. 中国语言服务企业机器翻译与译后编辑应用调查研究[J]. 北京第二外国语学院学报, 2021, 43(5): 23-37.

[34] 刘济超, Ömer Sahin Ganiyusufoglu，许文胜. 计算机辅助同声传译系统的设计、开发与验证[J]. 外语教学与研究, 2025, 57(3): 463-475.

[35] Läubli, S., Sennrich, R. and Volk, M. (2018) A Case for Document-Level Evaluation in Machine Translation. *Proceedings of the Third Conference on Machine Translation* (*WMT* 2018), Brussels, 31 October-1 November, 1134-1144.

数据库与在线资源

1. Text Creation Partnership (2025) EEBO-TCP Early English Books Online (Phase I/II Open Access Statement). https://textcreationpartnership.org/tcp-texts/eebo-tcp-early-english-books-online
2. Text Creation Partnership (2015) Licensing and Access. https://textcreationpartnership.org/about-the-tcp/about-partner-libraries/licensing-and-access/
3. Gale (ECCO) (2025) Eighteenth Century Collections Online (ECCO). https://www.gale.com/primary-sources/eighteenth-century-collections-online
4. Google Books (2025) Common Questions (Full/Preview/Snippet). https://books.google.com/googlebooks/common.html
5. HathiTrust (2015) Quality in HathiTrust. https://www.hathitrust.org/blogs/quality-in-hathitrust
6. HathiTrust (2016) Commitment to Quality. https://www.hathitrust.org/the-collection/preservation/commitment-to-quality
7. Internet Archive (2025) Books and Texts—Tips & Troubleshooting. https://help.archive.org/help/books-and-texts-tips-troubleshootin
8. Project Gutenberg/Distributed Proofreaders (2025) Distributed Proofreaders (DP) Workflow. https://www.pgdp.net
9. United Nations (2025) Official Document System (ODS). https://documents.un.org/
10. United Nations General Assembly (2020) A/RES/75/1: Declaration on the Commemoration of the Seventy-Fifth Anniversary of the United Nations (English). https://docs.un.org/en/A/RES/75/1
11. 联合国大会. A/RES/75/1: 联合国成立七十五周年纪念宣言(中文) [EB/OL]. https://digitallibrary.un.org, 2025-09-13.
12. UNTERM (United Nations Terminology Database) (2025) About; Main Portal. https://unterm.un.org/unterm2/en
13. United Nations General Assembly (2016) A/RES/71/1: New York Declaration for Refugees and Migrants. https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_71_1.pdf
14. 联合国大会. A/RES/71/1 (中文 ODS 页面: 联合国中文网站镜像) [EB/OL]. https://docs.un.org/zh/A/RES/71/1
15. 联合国. 关于难民和移民的纽约宣言[EB/OL]. https://www.un.org/zh/documents/treaty/A-RES-71-1, 2025-09-13.