

基于依存句法的机器译文质量优化方法研究

孔晓宏

上海海事大学外国语学院, 上海

收稿日期: 2025年12月9日; 录用日期: 2025年12月31日; 发布日期: 2026年1月12日

摘要

大语言模型翻译普遍存在受原文结构影响、质量难以保证等问题, 而机器译文质量评估多采用人工评价, 暂无通用的质量评估框架。基于此, 本文提出一种基于依存句法分析的译文质量评估与优化方法。通过构建“人工智能”领域的中英平行新闻语料库, 利用依存句法分析工具提取并分析英文原文与机器译文在平均依存距离、关键词性分布、依存关系类型及从句使用等方面的差异。研究发现, 机器译文存在平均依存距离偏高、形容词修饰语占比过大、名词复合结构占比偏低、较少使用从句等结构性问题。根据研究发现, 本文提出一套句法优化策略, 并对比优化前后的依存结构指标, 验证了该策略能有效提升译文句法结构。

关键词

翻译质量评估, 依存句法, 大语言模型, 平均依存距离, 人工智能

A Study Based on Dependency Syntax for Machine Translation Quality

Xiaohong Kong

College of Foreign Language, Shanghai Maritime University, Shanghai

Received: December 9, 2025; accepted: December 31, 2025; published: January 12, 2026

Abstract

Translations produced by large language models are widely affected by source-text structures, and their quality is difficult to ensure. Meanwhile, current machine translation quality assessment relies largely on human evaluation, and a unified, general evaluation framework is still lacking. In response to these issues, this study proposes a translation quality evaluation and optimization method based on dependency syntactic analysis. By constructing a Chinese-English parallel news corpus in the field of artificial intelligence, dependency parsing tools are employed to extract and

analyze differences between English source texts and machine-translated texts in terms of average dependency distance, distribution of key syntactic elements, types of dependency relations, and the use of subordinate clauses. The results indicate that machine-translated texts exhibit several structural problems, including excessively high average dependency distance, an over proportion of adjectival modifiers, an under representation of nominal compound structures, and limited use of subordinate clauses. Based on these findings, the study proposes a set of syntactic optimization strategies. By comparing dependency-structure indicators before and after optimization, the effectiveness of the proposed strategy in improving the syntactic structure of machine-translated texts is empirically verified.

Keywords

Translation Quality Assessment, Dependency Syntax, Large Language Models, Average Dependency Distance, Artificial Intelligence

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

随着自然语言处理技术的快速发展，机器翻译广泛应用于多个场景。然而，机器译文仍普遍采取直译方式，极易受原文影响，句法结构仍保持原文的结构，导致译文用词不地道，构句不自然。如何构建科学、高效的机器翻译质量评估体系，是提高机器译文质量的首要方式。

当前，机器翻译质量评估方法主要可分为三类：人工评价、自动评价与近年来兴起的交互式评价[1]。人工评价能够综合考量语义、风格与文化适配性，但其成本高昂、效率低下，难以支持大规模系统迭代。自动评价方法以 BLEU、TER 等为代表，侧重于词汇和句法层面的形式匹配，具备高效率与可复现的优势，却难以深入捕捉语义连贯性与逻辑合理性。基于大语言模型的交互式评价虽展现出强大的语义理解与复杂指令执行能力，能够实现错误分类等精细化分析，但其应用仍面临多重挑战，包括高昂的调用成本、复杂的提示工程设计、生成结果的不稳定性以及可解释性不足等问题，限制了其在大规模标准化评估中的推广。

依存句法分析作为一种能够揭示句子内部语法结构关系的语言学工具，为机器翻译质量评估提供了一条技术门槛适中、覆盖范围广泛的可行路径。相较于现有方法，依存句法分析具备显著优势：一方面，已有诸如 Stanford CoreNLP、spaCy 等成熟开源工具支持，技术实现门槛低，资源消耗可控；另一方面，它能够直接解析主谓宾、定状补等句法关系，使评估范围从传统的 n-gram 表层匹配延伸至句法结构合理性的深层分析，有效识别如成分缺失、依存关系错误等传统自动指标难以捕捉的问题。因此，构建基于依存句法的评估体系，有望在评估成本、效率与深度之间取得更优平衡，推动机器翻译质量评估向更精细、可解释的方向发展。

基于上述背景，本文聚焦“人工智能”领域的中英新闻文本，首先利用爬虫技术获取原始语料，将中文新闻通过大语言模型译为英文；随后对译文进行清洗与预处理，并开展系统的依存关系分析；最终依据分析结果提出针对性的优化策略，以提升机器翻译的句法合理性与整体质量。

2. 文献综述

目前，人工评价在机器翻译质量评估中仍占据重要地位。Snover [2]等人(2006)提出翻译编辑率(TER)，

用于衡量将机器译文修改至符合特定要求所需的人工编辑量，并进一步发展为面向人工编辑的翻译编辑率(Human-targeted TER)，显著提升了人工评价与评估结果之间的相关性。在翻译错误分析方面，德国人工智能研究中心(DFKI)提出了多维质量指标(MQM)，该框架提供了一套层次化的错误分类体系，并支持定制化扩展，因而具备广泛的适用性。

在自动评估方面，基于字符串匹配的方法长期受到学界关注。Papineni [3]等人(2002)提出的 BLEU 指标基于如下假设：越接近专业人工翻译的文本，其质量越高。该指标通过计算机器译文与参考译文之间 n-gram 的匹配程度来量化翻译质量。Rei [4]等人(2020)进一步提出 COMET，一种专门用于机器翻译评估的跨语言预训练模型。该模型能够深入理解原文、译文与参考译文之间的句法语义关联，并借助回归层输出与人工评分高度一致的评估结果，同时有效判断译文的语义忠实度与语言流畅度。

随着大语言模型的兴起，交互式机器翻译质量评估方法逐渐崭露头角。Yang [5]等人(2023)提出一种基于知识提示的评估方法(KPE)，融合了困惑度、词级与句级相似度等多个维度，并引入思维链(Chain-of-Thought)提示机制，实现了对译文质量的多角度综合评价。

然而，上述评估方法均存在一定局限。人工评估成本高、效率低，且易受主观因素干扰；BLEU、TER 等自动评估方法多局限于词汇与句法层面的形式匹配，难以深入语义理解；交互式评估虽具备多维度、高覆盖与高效率等优势，但仍面临技术门槛高、模型可解释性差等“黑箱”问题。

依存语法[6]描述了句子中词与词之间存在的非对称支配关系，分为支配词与从属词。此类关系可表示为带方向的句法结构树，从而直观刻画词间的依存关系。相较于侧重短语层级的传统评估指标，依存语法通过揭示句中各成分间的依存关联，并构建相应依存树库，有助于更准确地分析译文的流畅程度。此外，该方法聚焦于词语之间的功能联系，能够缓解因语序差异导致的分析偏差，从而提升评估的准确性。

依存距离与依存方向是用于描述词之间依存关系的重要概念。依存距离指在一个句子的依存句法树中，一个支配词与其从属词之间的线性词序距离。距离为 1 表示两个词在句子中直接相邻，这是最紧密的依存关系。距离越大，表明两个在语法上紧密相关的词在句子表层被分隔得越远。平均依存距离指所有依存关系的依存距离的算术平均值。“最小依存距离”原则认为，人类在理解和产出语言时，会下意识地倾向于缩短语法上相关词语之间的距离，以减轻工作记忆的负担。因此：平均依存距离越低，通常意味着句子结构更紧凑，更容易被理解。平均依存距离异常高，可能意味着句子结构扭曲、嵌套复杂、插入成分多，理解起来更费力[7]。依存方向描述的是在一个具体的依存关系中，从属词相对于其支配词的线性位置。依存方向由支配词指向从属词。不同语言通常存在不同的依存方向，如英语中支配词在前，日语中支配词通常在后。通过统计译文句子中不同依存方向的比例，可以评估其语序是否符合目标语言的规范。例如，一篇英译汉的译文中，如果“支配词在后”的比例远低于汉语母语文本的基准，则说明其可能保留了过多英语的语序，显得“翻译腔”过重，不够地道。

目前，已有部分研究尝试将依存语法应用于机器翻译生成优化与翻译共性分析。例如，Shen [8]等人(2010)借助依存语言模型捕捉词语间的长距离依存关系以优化译文质量；Xu [9]等人(2021)则运用依存语法系统刻画翻译语言的特征规律。然而，利用依存语法优化机器翻译质量评估的研究仍较为欠缺。因此，本文拟基于依存语法，针对新闻类机器译文构建质量评估体系，以弥补现有研究在此方向的不足。

3. 实验设计及结果

本实验旨在通过量化依存句法特征，对大语言模型生成的英译文本进行质量评估与迭代优化。为确保文本的主题相关性与句法分析价值，本研究聚焦于“人工智能”领域。采用网络爬虫技术，分别从《北京周报》和 CNN 新闻采集英文原文新闻与中文原文新闻各 4 篇，构成一个小规模、高质量的双语平行语料库。抓取后对文本进行清洗，去除原始 HTML 标签、广告、脚本及无关元数据。然后将中文原文文本

输入至豆包模型，使用“请将以下科技文本准确、流畅地翻译为英文，保留原文风格。”提示词，生成初始英文译文。随后，采用 Stanford Parser 对所有英文文本进行处理，以获得分词、分句、词性标注及通用依存关系分析结果。平均依存距离是衡量句子结构紧凑性与认知加工负荷的关键指标。本研究首先对每个句子计算其所有依存关系的线性距离绝对值，并计算其均值，得到句子的平均依存距离值。为进行更精细的对比，本研究从以下三个维度提取并量化微观句法特征。1. 关键词性分布：统计英文原文与机器译文中关键词性标签的出现频率。重点关注的词性类别包括：名词(NN)、动词(VB)、形容词(JJ)。主要功能词：介词(IN)、连词(CC)、从属连词。2. 从句结构标记：统计并对比引导名词性从句、定语从句及状语从句的显性标记词(如 *that, which, who, because, if* 等)的使用频率，以评估句法复杂化的方式与习惯。3. 依存关系类型分布：统计关键依存关系类型的分布，以揭示句法构造习惯。重点分析的依存关系包括：形容词性修饰语(*amod*)、名词性主语(*nsubj*)与状语修饰语(*advmod*)。

通过对比关键词性分布，识别实词与功能词使用的偏差；通过对比依存关系频率(如 *amod*、*nsubj* 的使用比例)，深入诊断句式构造的具体问题；通过对比从句标记频率，评估从句使用的习惯与复杂度。根据上述诊断出的具体句法问题，对机器译文进行人工优化。

Table 1. Text character count

表 1. 文本字符数

| | 中文原文 | 英文原文 | 中文译文 |
|-----|------|--------|--------|
| 字符数 | 8037 | 14,333 | 17,245 |

通过 Stanford Parser 对各篇新闻进行依存关系分析，得出英文原文平均依存距离为 3.15，中文译文平均依存距离为 3.46 (见表 1)。同一语言内相比，0.295 的差异非常显著。从具体依存关系来看，中文译文的形容词修饰语(*amod*)占比偏高(8.91%~10.67%)，而英文原文的 *amod* 占比仅为 4.9%~6.37%。

当源语言与目标语言的结构差异较大时，译文往往会受到源语言结构的影响。翻译过程中并不总是最大化目标语言的习惯表达，而可能因为源语言的干扰(*source-language interference/shining-through*)而使译文句法结构偏向源语[10]。Cheung [11] (2017)指出，汉语名词短语内部的信息更倾向于集中在支配词前，而英语通常通过介词短语、关系从句或名词化结构表达同类语义。以本研究语料之一为例，中文原文中大量出现“名词 + 名词(*nn*)”偏正结构，如“文化偏见”“全球刻板印象”“跨文化传播”“结构性不公”等，这类高密度名词性前置修饰在汉语中极为常见，且符合汉语“修饰突出型”的结构特征。而在机器译文中，这些短语则被译为“形容词 + 名词”和“名词 + 介词短语”结构，如 *cultural biases, stereotypes from around the world, cross-cultural dissemination of stereotypes, structural inequities*。

此外，本研究语料中英语原文展现出更高的名词复合修饰密度(*nn* 比例 9.67%~16.57%)，如 *AI Litigation Task Force, state AI laws, federal policy framework* 等名词短语结构多次出现。而在译文中，*nn* 占比显著偏低(6.06%~7.85%)。汉语修饰语的高密度前置不仅是句法特征，也是一种信息组织策略，在跨语言转换时容易被直译为形容词修饰，从而改变目标语的修饰比例[11]。本研究语料中，其中一篇英文原文新闻出现大量高密度名词短语，如 *Air traffic controller staffing, Federal Aviation Administration's Air Traffic Controller Workforce Plan* 等。而中文原文中也出现大量的 *nn* 结构，如“偏见行李”“文化漂移”等四字词，以及“性别歧视”“文化偏见”“语言不平等”等结构。英语作为“层级包装型”语言，更倾向于通过名词复合和从句实现信息集成。大语言模型的译文呈现出“去名词化”和“修饰语显性化”的双重特征，既是源语言结构的干扰体现，也是模型在确保语义忠实度前提下的一种结构简化策略。由于这些修饰语在源文中承载了大量语义信息，大语言模型进行逐层转换时更容易形成紧密的定语链条，使译文平

均依存距离增加、*amod* 占比提高,同时 *nn* 占比下降。这既反映了源语结构的投射效应,也反映了译文结构在信息分布上的重构——由原文的高密度名词结构转向修饰语显性化和介词化表达。

在从句结构的使用上,本研究识别出英语原文与机器译文之间存在显著差异。本研究所选取的一篇英语原文新闻大量使用多层次从句来显现事件间的逻辑关系。例如,原文中出现大量原因从句(*because it demands almost near perfect performance*)、让步从句(*even if the technology were to be used...*)、非限定性定语从句(*controllers, who require rigorous training...*)、补语从句(*experts highlighted that...*)等结构。此外,多个句子跨越多重子句,例如:“*While AI does have some practical uses..., experts highlighted a range of reasons why the technology won't be a substitute...*”。这类结构是英语原文新闻文本高度层级化(*hypotactic*)的句法特征。

相比之下,本研究语料中的机器译文呈现出更低的从句密度。例如,“*Artificial intelligence (AI) has become an inseparable ‘companion’ in our lives.*”“*According to a report..., an international study points out that...*”“*This inevitably leads to deep thought...*”等。这类结构以短句、并列句或线性铺陈为主,极少出现英语原文文本常用的复杂从句嵌套。即便出现 *that* 引导的从句,也多为“结构对应式”翻译,而不是通过英语常见的关系从句或多层补语从句扩展语义层级。例如:“*research shows that...*”或“*the study points out that...*”均是对中文“指出”“显示”等动词的线性转换,而非重构为英语典型的层级句式。

导致这种差异的根源在于中文的句法组织方式。中文新闻文本广泛依赖松散并列结构,这些结构在语义上高度压缩,但在句法上不依赖从句进行层级表达。比如,中文原文“研究显示,模型表现不佳。它们还产生负面刻板印象。”,大语言模型倾向于用简单的从句嵌套“*Research shows that the models perform poorly, and that they also generate negative stereotypes.*”。

介词结构组合也出现显著差异。已有研究表明,英语原文在信息表达中往往呈现出多样化的介词选择,并根据语义细节与修辞目的灵活搭配,这在科技新闻中尤为明显。例如,在本次研究语料中可以看到英文原文新闻使用不同介词如 *around*、*into*、*on*、*against*、*through*、*before* 等,根据情境传递细微的语义差别和事件关系(如“*regulations around artificial intelligence*”,“*extending into more areas of life*”,“*comments on Truth Social*”,“*protecting against online censorship*”)。机器译文则多次使用距离长、固定化的介词结构(如“*of...in...*”,“*of...with...*”等)来承载多个关系,导致句子结构更线性。例如其中一篇机器译文中大量出现“*biases in...*”,“*limitations of...*”,“*rooted in...*”,“*problems of...in...*”等结构。这些结构多为训练语料中的高频组合,在处理复杂关系时会将语义压缩为几个相对稳定的介词短语,减少了多样化介词的使用。

4. 基于依存句法分析的优化策略

根据以上分析,对其中一篇机器译文进行修改优化。以中文原文“研究人员使用 16 种语言设计交互式提示,并测试了数种主流语言模型对这些偏见的反应”的翻译为例,初始机器译文为:

“*The researchers designed interactive prompts in 16 languages and tested the responses of several mainstream language models to these biases.*”该译文在依存关系上呈现出典型的“翻译腔”特征:首先, *the responses of ... to...* 结构构成了冗长的“名词 + *of* + 名词 + *to* + 名词”介词结构,平均依存距离较长;其次,核心语义被包装在静态名词 *responses* 中,缺乏动态性;最后,整个句子虽为并列结构,但未使用任何从句, *ccomp* 比例为 0%。

根据本研究所建立的优化框架,将该句修改为:“*Researchers devised interactive prompts in 16 languages and tested several mainstream language models to see how they responded to these biases.*”

对句子进行优化后,利用相同方式得出了优化前后的依存关系对比(见表 2)。

Table 2. Dependency table
表 2. 依存关系表

| 依存关系类型 | 优化前计数 | 优化前比例 | 优化后计数 | 优化后比例 |
|---------|-------|--------|-------|--------|
| det | 3 | 18.75% | 1 | 5.56% |
| dobj | 2 | 12.50% | 2 | 11.11% |
| amod | 2 | 12.50% | 2 | 11.11% |
| nn | 2 | 12.50% | 2 | 11.11% |
| nsubj | 1 | 6.25% | 2 | 11.11% |
| prep_of | 1 | 6.25% | 0 | 0% |
| prep_to | 1 | 6.25% | 1 | 5.56% |
| ccomp | 0 | 0% | 1 | 5.56% |
| aux | 0 | 0% | 1 | 5.56% |
| vmod | 0 | 0% | 1 | 5.56% |
| advmod | 0 | 0% | 1 | 5.56% |

可以看出, 优化前平均依存距离为 2.88, 优化后降至 2.56, 降幅达 11.1%。对该机器译文采取了如下优化策略: 优化长距离介词短语: 原有的长距离 `prep_to` (距离 10) 被替换为更紧凑的从句结构; 使用从句显化句子关系: 优化前 `ccomp` 比例为 0%, 优化后提升至 5.56%, 丰富了句式多样性, 避免依存关系的过度集中; `det` 比例从 18.75% 降至 5.56%, 减少了限定词的机械使用。

机器译文倾向于在译文中保留汉语的名词短语结构, 导致依存距离拉长、介词结构笨重; 而优化后的译文根据英语的“动词优势”和“形合特征”, 重构了信息表达方式。

在中文原文中, “的”字被广泛用于组合名词与形容词, 且中文习惯用较多四字词语, 导致机器翻译时受到源语结构影响, 出现较多的 `amod` 依存关系。比如, 原文中“全球的刻板印象”这一短语原意本指针对全球的刻板印象, 而因受到原文“的”字影响, 机器翻译倾向于翻译成“*global stereotypes*”。基于此, 应当将其修改为“*stereotypes across the globe*”, 这样更准确传达原文意思, 并避免了过度使用 `amod` 结构。

本研究提出的句法优化策略不仅具有操作层面的指导意义, 也在一定程度上给予了翻译共性理论——归化(*domestication*)和异化(*foreignization*)——可量化的标准。传统的翻译共性研究多基于词汇、搭配或语篇特征, 而依存句法结构提供了一种更底层、更连贯的量化框架。它不仅能描述“发生了什么变化”(如从句变少了), 更能揭示“变化是如何在句法关系中实现的”(如通过减少 `ccomp` 依存、增加 `conj` 并列关系等), 提高了理论的可解释性。

5. 结语

本研究围绕“大语言模型译文质量评估与优化”这一议题, 构建了基于依存句法的译文质量评估框架, 并将其应用于人工智能领域的双语新闻平行语料。在对比英文原文与机器译文的依存结构特征时, 我们发现译文存在平均依存距离显著偏高、形容词性修饰语(`amod`)比例过多、名词复合结构(`nn`)比例不足等问题, 这些结构性偏差与汉语源语言特征(如高频“的”字结构、四字词语的紧凑修饰模式)密切相关。同时也为翻译共性理论中的“归化”与“异化”倾向提供了可量化的标准。与此同时, 译文的从句密度和介词多样性均低于英文原文, 呈现出“翻译腔”与信息线性铺陈的双重特征。针对上述问题, 本文提

出了多项优化策略,包括将部分 amod 结构转换为介词短语或名词复合结构、增加非限制性定语从句和原因/让步从句、缩短长链介词结构以降低平均依存距离等,这些策略本质上是在引导译文从反映源语结构的“异化”倾向,转向更符合目标语认知习惯与句法规范的“归化”表达。实验验证表明,这些策略显著改善了译文的句法紧凑性和信息层级化表达,使译文更接近英文母语新闻的篇章特征,也为基于依存句法的翻译共性研究与译文自动优化提供了可计算、可干预的实践路径。

参考文献

- [1] 韩林涛,陈重宇. 机器翻译质量评估:发展历程与新路径[J]. 语言服务研究, 2025, 5(1): 53-71.
- [2] Snover, et al. (2006) A Study of Translation Edit Rate with Targeted Human Annotation. In: *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas*, Cambridge, 8-12 August 2006, 223-231.
- [3] Papineni, K., Roukos, S., Ward, T. and Zhu, W. (2001) BLEU: A Method for Automatic Evaluation of Machine Translation. In: *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics-ACL '02*, Philadelphia, 6-12 July 2002, 311-318. <https://doi.org/10.3115/1073083.1073135>
- [4] 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)*, November 2020, 2685-2702. <https://doi.org/10.18653/v1/2020.emnlp-main.213>
- [5] Yang, Z., Li, J., Wang, M. and Zhou, M. (2023) Knowledge-Prompted Evaluation for Machine Translation with Large Language Models. *Transactions of the Association for Computational Linguistics*, **11**, 1572-1590.
- [6] Tesnière, L. (1959) *Éléments de syntaxe structurale*. Klincksieck.
- [7] 刘海涛. 依存语法的理论与实践[M]. 北京: 科学出版社, 2009.
- [8] Shen, L., Xu, J. and Weischedel, R. (2010) A New String-to-Dependency Machine Translation Algorithm with a Target Dependency Language Model. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, 11-16 July 2010, 577-585.
- [9] Xu, J., Zhang, M. and Li, C. (2021) Exploring Universal Dependencies for Translation Universals. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, 19-23 April 2021, 3123-3133.
- [10] Xiao, R. (2010) How Different Is Translated Chinese from Native Chinese? A Corpus-Based Study of Characteristic Features. *International Journal of Corpus Linguistics*, **15**, 5-35. <https://doi.org/10.1075/ijcl.15.1.01xia>
- [11] Cheung, K.H. (2017) *A Corpus-Based Study of the Translation of Chinese Noun Phrases into English*. The Hong Kong Polytechnic University.