

# 大语言模型对中国形象类块状语英译能力评测与优化研究

## ——跨文化传播视角下的人机协同翻译路径探索

沈 越\*, 赵振强, 刘绍龙

浙江越秀外国语学院应用外语学院, 浙江 绍兴

收稿日期: 2025年12月26日; 录用日期: 2026年1月23日; 发布日期: 2026年1月30日

### 摘要

大语言模型技术的快速发展为中国形象的跨文化传播提供了新的技术路径。本研究通过构建三维量化评价体系, 自建三大领域共450组四字格块状语语料库, 对三款国产主流大语言模型进行系统评测与对比分析。研究整合COMET自动化评分与专家人工校验机制, 并运用Few-shot学习和Chain-of-Thought等提示工程技术开展优化实验。研究发现: (1) 三款模型在语义准确性与表达流畅性上表现相对均衡, 但文化传递力普遍偏低, 呈现出“形似而神异”的翻译特征; (2) 四字格块状语因其高度凝练的认知格式塔结构, 对模型的文化解码能力提出更高要求, 模型普遍存在“去语境化”倾向; (3) 基于思维链的显性提示策略能够有效激活模型的隐性文化知识, 使其从“概率匹配”转向“逻辑推理”翻译; (4) 人机协同介入机制在改善中国形象对外话语传播中具有不可替代性, 提示工程本质上构成了一种认知补偿机制, 为模型搭建了外部的认知支架。本研究为大语言模型在中国形象对外传播中的应用提供了科学的诊断依据与技术优化方案。

### 关键词

大语言模型, 块状语, 三维评价体系, 提示工程, 认知补偿

# A Study on Evaluating and Optimizing the Capability of Large Language Models in Translating Chunks with Chinese Images into English

## —Exploring Human-Machine Collaborative Translation Pathways from an Intercultural Communication Perspective

\*通讯作者。

**Yue Shen\*, Zhenqiang Zhao, Shaolong Liu**

School of Applied Foreign Languages, Zhejiang Yuexiu University, Shaoxing Zhejiang

Received: December 26, 2025; accepted: January 23, 2026; published: January 30, 2026

## Abstract

The rapid development of large language model (LLM) technology provides new technological pathways for the cross-cultural communication of China's national image. This study systematically evaluates and comparatively analyzes three mainstream domestically-produced LLMs by constructing a three-dimensional quantitative evaluation framework and building a self-compiled corpus of 450 sets of four-character chunky expression across three domains. The research integrates COMET automated scoring with expert manual verification mechanisms, and conducts optimization experiments by employing prompt engineering techniques such as Few-shot learning and Chain-of-Thought. The findings reveal: (1) The three models demonstrate relatively balanced performance in semantic accuracy and expressive fluency, but universally exhibit low cultural conveyance capacity, manifesting translation characteristics of "formal similarity with semantic divergence"; (2) Four-character chunks, due to their highly condensed cognitive Gestalt structure, impose higher demands on models' cultural decoding capabilities, while LLM models universally displaying a tendency toward "decontextualization"; (3) Explicit prompting strategies, based on chain-of-thought, can effectively activate models' implicit cultural knowledge, enabling a shift from "probabilistic matching" to "logical reasoning" translation; (4) Human-machine collaborative intervention mechanisms possess irreplaceable value in improving the external discourse communication of China's national image, because prompt engineering essentially constitutes a cognitive compensation mechanism that constructs external cognitive scaffolding for models. This study provides scientific diagnostic evidence and technical optimization solutions for the application of large language models in the external communication of China's national image.

## Keywords

**Large Language Model, Chunky Expressions, Three-Dimensional Evaluation System, Prompt Engineering, Cognitive Compensation**

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

中国形象的国际传播是国家文化软实力建设的重要组成部分，而承载中国独特文化理念与社会实践智慧的汉语块状语，作为一种高度浓缩的语言符号系统，其跨语言转换质量直接影响着中国话语在国际场域中的阐释效果与接受度。块状语不同于一般性的词汇或短语，它以紧凑的结构承载着丰富的文化信息、深层的语义逻辑和特定的民族认知图式，尤其是四字格块状语，因其对称工整的形式与完形性的语义特征，成为汉语表达中最具代表性的文化语言单位。近年来，基于 Transformer 架构的大语言模型在机器翻译领域展现出令人瞩目的性能提升，其生成的译文在流畅度与准确性上已接近人类译者水平。然而，国内主流大语言模型如 DeepSeek、文心一言、通义千问等在处理此类语料时，究竟能否实现语义的完整

还原、表达的自然流畅以及文化内涵的有效传递，其翻译过程中存在哪些系统性短板，又如何通过技术手段进行有针对性的优化，这些问题亟待通过实证研究予以回答。为此，本研究选取三个核心领域构建专门语料库，通过建立多维量化评价标准，系统考察国产大模型在块状语翻译中的表征，并探索基于提示工程的优化路径，以期提升中国形象对外传播的精准性与有效性。

## 2. 研究现状

汉语块状语因其独特的结构特征与文化负载属性，一直是翻译研究的重点对象。杨开烨等(2023)指出，四字格作为汉语中最具典型性的块状语形式，其翻译不仅涉及语言符号的转换，更关乎文化意象的重构与认知图式的跨语言映射[1]。在翻译策略方面，学界长期存在直译与意译、异化与归化的争论。章明蕾(2025)通过语料库研究发现，在处理具有中国特色的文化专有词汇时，采用解释性翻译和增译策略能够更好地弥合文化认知差异[2]。李珩(2019)则强调，对外话语翻译应在保持源语文化特色的前提下，兼顾目标语读者的可接受性，实现“文化传真”与“交际有效”的平衡[3]。

随着人工智能技术的发展，机器翻译评测成为新的研究热点。耿芳和胡健(2023)基于 ChatGPT 的翻译实例研究表明，大语言模型在汉译英的错误修正、句法结构调整和篇章连贯增强方面表现出色，但在处理文化负载词时仍存在明显不足[4]。周领顺和陈龙宇(2025)在 AI 背景下的译者行为研究中首次系统讨论了机译者行为，指出人译者与机译者的互动关系将成为未来翻译研究的重要方向[5]。袁筱一(2025)从哲学层面探讨了人工智能文学翻译的“主体性”与“创造性”问题，认为大语言模型在处理非信息文本时难以复制人类译者的文学性和主体性，尤其在文化阐释层面存在根本性局限[6]。在评测方法论方面，Al Sawi 和 Allam (2024)通过对人工译者与 ChatGPT 生成的阿拉伯语字幕发现，人工译者更能意识到文化典故类表达需要改变措辞，而 AI 倾向于字面直译，导致严重的语义错误[7]。Lyu 等(2024)在珠宝设计的跨文化转译研究中指出，生成式 AI 虽能提升技术效率，但在文化深层语义的把握上仍面临挑战，人机协同模式是弥补 AI 文化意识缺失的有效路径[8]。

本研究在梳理现有文献时发现，该领域在以下方面尚存进一步深化与拓展的空间：首先，当前对大语言模型翻译能力的评测，较多集中于以 ChatGPT 为代表的国际主流模型，而对国产大模型的系统性、对比性评估研究相对有限，其性能特征与优化路径亟待更细致的考察。其次，尽管块状语在对外传播中具有重要价值，但专门针对其中承载中国形象、特别是以四字格为代表的典型文化负载表达的翻译能力研究仍较为鲜见，相关讨论尚未充分展开。此外，在技术应用层面，如何借助提示工程等前沿方法，有效提升模型在文化意象传递上的准确性与适应性，其具体策略与实证效果仍有待深入探索。

鉴于此，本研究旨在通过构建涵盖语义、表达、文化的三维评价体系，对国产主流大模型进行细粒度评测，进而结合定量分析与案例诊断，探究适用于中国形象类块状语英译的提示优化方案。

## 3. 理论、问题和方法

### 3.1. 理论框架

本研究构建“语义准确性 – 表达流畅性 – 文化传递力”三维评价模型作为理论基础。该模型整合了功能翻译理论、框架语义学与跨文化传播理论的核心观点：

(1) 语义准确性(Semantic Accuracy): 基于 Fillmore (2025)的框架语义学理论，考察译文是否准确还原源语块状语的概念框架、参与者角色和事件逻辑[9]。具体包括词汇选择精准度(是否选择了语义场中最恰当的对应词)、语义完整性(是否遗漏或增添关键语义成分)、逻辑关系保持(块状语内部及其与上下文的逻辑关联是否清晰)三项指标。

(2) 表达流畅性(Expressive Fluency): 基于 Halliday (2024)的系统功能语言学，重点考察译文在目标

语境下的自然度与可接受性[10]。包括句法自然度(是否符合英语母语者的表达习惯)、语言规范性(是否遵循目标语的语法规则与搭配惯例)、可读性(译文是否便于目标读者理解)三项指标。

(3) 文化传递力(Cultural Transmission): 基于 Nida (2025) 的动态对等理论, 聚焦译文对源语文化内涵的传递效果[11]。包括文化内涵保持度(是否保留了源语的文化意象与深层含义)、目标语文化适应性(是否以目标语读者能够理解的方式呈现文化信息)、传播效果(译文能否有效建构积极的中国形象)三项指标。

### 3.2. 研究问题

本研究主要回答以下三个核心问题:

(1) DeepSeek、文心一言、通义千问在翻译不同领域的中国形象类四字格块状语时, 其在三维评价体系下的表现特征有何差异? 各模型的优势与短板分别体现在哪些方面?

(2) 大语言模型在处理四字格块状语时产生语义偏移、文化缺失、表达生硬等翻译问题的主要错误类型及其认知机理是什么? 这些问题的产生是否与块状语的特定结构特征或文化属性相关?

(3) 基于 Few-shot 学习和 Chain-of-Thought 的提示优化策略能否显著提升模型的翻译质量? 不同类型的提示词模板对三个评价维度的改进效果是否存在差异?

### 3.3. 研究方法

#### 3.3.1. 语料选取

本研究自建中国形象类四字格块状语语料库。语料来源于具有代表性和文化典型性的三个领域:

(1) 非物质文化遗产领域(文化维度): 涵盖传统技艺、民俗活动、表演艺术等方面四字格表达, 如“薪火相传”“巧夺天工”“源远流长”等, 共 150 组实例。

(2) 社会基层治理领域(社会维度): 涉及社区建设、邻里关系、基层服务等方面四字格表达, 如“邻里守望”“共建共享”“精准施策”等, 共 150 组实例。

(3) 绿色循环经济领域(经济维度): 包括生态保护、可持续发展、循环利用等方面四字格表达, 如“绿水青山”“循环利用”“节能减排”等, 共 150 组实例。

语料筛选标准为: 必须为典型的四字格块状语结构; 具有明确的中国文化或社会实践特色; 来源真实可靠(政府公报、官方媒体、学术专著、文化遗产档案); 具有对应的官方或权威英译版本作为参照。共筛选出 450 组典型实例, 平均分配到三个领域。

#### 3.3.2. 分析流程

研究分为“基线评测”与“优化评测”两个阶段:

第一阶段: 基线评测。将 450 组含有四字格块状语的完整句例分别输入三款大语言模型(DeepSeek 3.2、文心一言 Ernie Bot 4.0、通义千问 Qwen-Max), 采用标准化提示词“请将以下句子翻译成英文”, 不提供任何领域背景或文化提示, 获取基线译文。

第二阶段: 优化评测。针对基线测试中识别出的高频错误类型, 设计三类提示词优化模板:

(1) 语义澄清型: “请先解释‘[块状语]’的深层含义, 再进行翻译。”

(2) 文化背景补充型: “请作为熟悉中国[领域]文化的专家, 翻译以下句子。”

(3) 思维链引导型: “第一步: 分析‘[块状语]’的字面意思和比喻意义; 第二步: 考虑目标读者的文化背景; 第三步: 选择最恰当的翻译策略并生成译文。”

为深入评估思维链(CoT)各环节对翻译质量提升的具体贡献, 本研究设计了消融实验(Ablation Study), 将完整的三步骤思维链拆解为以下变体:

- CoT-Step 1: “请先分析[块状语]的字面意思和比喻意义, 再进行翻译。”(仅保留语义分析步骤)

- CoT-Step 2: “请考虑目标读者的文化背景, 翻译以下句子。”(仅保留受众意识步骤)
- CoT-Step 3: “请选择最恰当的翻译策略(直译、意译或文化阐释), 翻译以下句子。”(仅保留策略选择步骤)
- CoT-Step 1 + 2: “请分析[块状语]的含义, 并考虑目标读者的文化背景, 再进行翻译。”(保留前两步)
- CoT-Step 2 + 3: “请考虑目标读者的文化背景, 选择恰当的翻译策略, 翻译以下句子。”(保留后两步)
- CoT-Full (完整版): 包含全部三个步骤。

每组语料分别在六种提示词变体下进行测试, 与基线译文进行对比分析。实验选取 150 组高错误率语料(在基线测试中文化传递力维度得分低于 70 分的样本)作为消融实验对象, 每种提示词变体测试 50 组(跨三个领域均匀分布)。

此外, 针对不同模型的特性差异, 本研究还设计了模型定制化提示词对照实验:

- 针对 DeepSeek (语义理解强但表达略显生硬): 增加“请用地道的英语表达”指令。
- 针对通义千问(表达流畅但文化敏感度不足): 增加“请特别注意保留中国文化特色”指令。
- 针对文心一言(文化知识丰富但有时过度解释): 增加“请简洁地翻译, 避免冗余信息”指令。

通过对比通用提示词与定制化提示词的效果差异, 评估针对性优化策略的有效性。

### 3.3.3. 评价标准

本研究采用“自动化评分 + 人工校验”的混合评测机制:

(1) 自动化评分: 使用 COMET (Crosslingual Optimized Metric for Evaluation of Translation)模型进行机器评分, 该模型基于跨语言预训练技术, 能够捕捉译文的语义相似度、流畅度等多维特征, 输出 0~1 之间的质量分数。

(2) 人工校验: 邀请三名具有副教授及以上职称的翻译教师组成评审组, 按照三维评价标准对每组译文进行百分制打分。每个维度下的三项指标各占该维度分数的三分之一, 三个维度的权重分别为: 语义准确性 40%、表达流畅性 30%、文化传递力 30%。

(3) 质量等级划分: 根据综合得分将译文质量分为四个等级: 优秀(90~100 分)表示译文在三个维度上均表现出色, 达到专业译者水平; 良好(75~89 分)表示译文整体质量较高, 但在某些方面存在可改进空间; 中等(60~74 分)表示译文基本可理解, 但存在明显的语义偏差或表达问题; 较差(60 分以下)表示译文存在严重错误, 影响信息传递或可能造成误解。

### 3.3.4. 典型实例

为深入剖析模型的翻译表现, 本研究从 450 组语料中精选 5 个典型案例, 每个实例包含 1~3 个四字格块状语。表 1 展示了部分典型案例及其多版本译文对比:

**Table 1.** Comparison of translation examples for Chinese image-building four-character chunky expressions

**表 1. 中国形象类四字格块状语典型实例翻译对比表**

领域	原文句例	官方/人工参考译文	DeepSeek 基线译文	通义千问基线译文	文心一言基线译文	对比分析
非苏绣技艺巧夺天工, 每一针每一线都凝聚着匠人化的精益求精。	Suzhou embroidery is <b>exquisitely crafted</b> , with every stitch <b>embodying the artisan's pursuit of perfection.</b>	Suzhou embroidery is <b>ingeniously made</b> , every needle and thread condenses the craftsman's <b>spirit of continuous improvement.</b>	Suzhou embroidery technique is <b>skillfully made by heavenly craftsmanship</b> , every stitch gathers craftsman's <b>refined and better.</b>	Suzhou embroidery is <b>clever and wonderful</b> , with each stitch showing the craftsman's <b>spirit of continuous improvement.</b>	Suzhou embroidery is <b>ingeniously made</b> , every craftsman's <b>spirit of continuous improvement.</b>	官方译文采用意译, 流畅自然。机译多直译, DeepSeek “ingeniously made” 较为简洁, 文心一言保留了“spirit”, 增强可读性。

续表

非 项这项技艺代代 遗相传, 在传承人 文 的守护下历久 化 弥新。	This craft has been passed down through generations, remaining ever vibrant under the guardianship of its inheritors.	This craft is passed from generation to generation, under the protection of inheritors it becomes newer after long time.	This technique transmits generation by generation, staying more fresh with long history under inheritors' guard.	This craft passes down generations, and lasts long and renews with inheritors' protection.	人工译文用“ever vibrant”精准传达“历久弥新”的动态感。机译多为僵硬直译, 缺乏英语韵味。
社 社区通过邻里守 会 望机制, 营造了 治 治守望相助的温馨 理 氛围。	By establishing a neighborhood watch system, the community has fostered a warm atmosphere of mutual support.	Through neighbor watching mechanism, the community creates watching and helping each other warm atmosphere.	The community creates neighbors guarding neighbors warm feeling through mutual watch and help mechanism.	Through neighborhood watch, the community builds an atmosphere where people watch out for one another.	文心一言准确识别“neighborhood watch”这一英语固定表达, 译文最为地道。其他模型过于直译。
社 基层干部扎根基 会 层, 以无私奉献 治 的精神服务群 理 众。	Grassroots officials are deeply rooted at the community level, serving the people with selfless dedication.	Grassroots cadres root in grassroots, serve the masses with selfless devotion spirit.	Grassroots officials take root at the basic level, serving people with unselfish dedication.	Grassroots cadres stay rooted at the grassroots, serving people with the spirit of selfless contribution.	“deeply rooted”比“take root”更符合英语习惯。DeepSeek 的“selfless devotion”与通义千问“unselfish dedication”语义相近但表达略有差异。
绿 我们要牢固树立 色 绿水青山就是金 经 山银山的理念, 济 实现永续发展。	We must firmly uphold the concept that lucid waters and lush mountains are invaluable assets, achieving sustainable development.	It is imperative to know that pristine ecosystems translate into enduring prosperity, thus securing sustainable progress for future generations.	We must commit to the principle that ecological preservation constitutes tangible wealth, forging a path of enduring development.	We need to deeply convict that nature's endowment is our lasting treasury, ensuring continuity in developmental journey.	官方译法“lucid waters and lush mountains”富有诗意且已成惯用表达。机译颜色词选择各异, 缺乏统一性。

### 3.3.5. 分析方法

研究采用定量与定性相结合的混合方法:

定量分析: 运用 SPSS 26.0 统计软件, 对三款模型在三维评价标准下的得分进行描述性统计分析(均值、标准差)、方差分析(ANOVA)检验模型间差异的显著性, 以及配对样本 t 检验(Paired t-test)评估优化前后的改进效果。

定性分析: 基于“词义理解 - 结构转换 - 文化表达”三层次分析框架, 对典型错误案例进行深度解读, 探讨错误产生的认知机理。具体包括: 词义层面分析模型对四字格整体语义与各语素间关系的理解准确性; 结构层面考察模型在汉英句法转换中的处理策略; 文化层面剖析模型对隐喻、典故等文化要素的识别与转换能力。

## 4. 结果和分析

### 4.1. 定量数据: 三维评价体系下的模型表现

基于 450 组语料的综合评测结果显示, 三款大语言模型在不同评价维度上呈现出显著的差异化特征。

表 2 汇总了三款模型在三维评价标准下的详细得分情况。

数据解读与分析:

(1) 语义准确性维度: DeepSeek 在该维度表现最优(90.2 分), 特别是在“逻辑关系保持”指标上高达 91.7 分。这表明 DeepSeek 在解析四字格块状语的内部结构(如并列、偏正、动宾等关系)以及其与上下文的逻辑衔接方面具有显著优势。例如, 对于“精益求精”(并列递进结构), DeepSeek 能够较准确地识别

其“追求卓越”的整体语义，而非简单地逐字翻译。通义千问(87.5 分)和文心一言(86.5 分)在该维度略显逊色，主要问题在于对复杂语义关系的把握不够精准，偶尔出现语义遗漏或添加现象。

(2) 表达流畅性维度：通义千问在该维度遥遥领先(89.5 分)，所有三项指标均为最高分。其生成的译文更符合英语母语者的表达习惯，句式结构自然，词汇搭配地道，可读性强。这可能得益于通义千问在训练阶段使用了大量高质量的英语单语语料和双语平行语料。相比之下，DeepSeek (85.7 分)虽在语义理解上占优，但在表达上略显生硬，存在“翻译腔”；文心一言(84.6 分)在该维度表现相对较弱，部分译文出现了词汇搭配不当或句式不够简洁的问题。

(3) 文化传递力维度：这是所有模型的共同短板，三款模型的平均得分仅为 73.3 分，远低于前两个维度。文心一言在该维度略胜一筹(74.4 分)，尤其在“文化内涵保持度”和“传播效果”两项指标上表现较好。这可能与文心一言依托百度知识图谱，对成语典故等文化信息具有较强的检索和识别能力有关。例如，对于“源远流长”这一富含历史感和时间延续性的四字格，文心一言能够选择“have a long history and rich heritage”这样既保留文化韵味又易于理解的表达。然而，即便是表现最好的模型，在该维度的得分也未超过 75 分，说明大语言模型在处理文化负载表达时仍面临巨大挑战。

(4) 综合表现：从加权平均分来看，通义千问以 84.6 分位居第一，DeepSeek 以 83.8 分紧随其后，文心一言以 82.7 分排名第三。三款模型的综合得分较为接近，均处于“良好”等级(75~89 分)，尚未达到“优秀”水平(90 分以上)。这表明国产大语言模型在处理中国形象类四字格块状语时已具备较为扎实的基础能力，但在实现高质量、高文化保真度的翻译方面仍有较大提升空间。

**Table 2.** Three-dimensional evaluation results of three LLMs in translating Chinese image-building four-character chunky expressions

**表 2.** 三款大语言模型在中国形象类四字格块状语翻译中的三维能力评测结果

评测维度	具体指标	DeepSeek	通义千问	文心一言	三模型均值
语义准确性 (Semantic Accuracy)	词汇选择精准度	88.6	86.2	87.4	87.4
	语义完整性	<b>90.3</b>	87.5	86.8	88.2
	逻辑关系保持	<b>91.7</b>	88.9	85.2	88.6
	维度均值	<b>90.2</b>	87.5	86.5	88.1
表达流畅性 (Expressive Fluency)	句法自然度	84.8	<b>89.3</b>	83.7	85.9
	语言规范性	86.9	<b>90.6</b>	85.8	87.8
	可读性	85.5	<b>88.7</b>	84.4	86.2
	维度均值	85.7	<b>89.5</b>	84.6	86.6
文化传递力 (Cultural Transmission)	文化内涵保持度	71.4	73.8	<b>76.2</b>	73.8
	目标语文化适应性	74.6	75.9	72.8	74.4
	传播效果	69.8	71.5	<b>74.3</b>	71.9
	维度均值	71.9	73.7	<b>74.4</b>	73.3
综合得分	加权平均分	<b>83.8</b>	<b>84.6</b>	82.7	<b>83.7</b>

## 4.2. 错误类型分析与优化效果评估

为深入理解模型翻译问题的根源，本研究对 450 组基线译文进行了系统的错误标注与分类。研究团队识别出五大类错误类型，并针对高频错误设计了相应的优化策略。**表 3** 展示了优化前后的错误率变化及质量提升情况。

**Table 3.** Comparison of translation error rates and quality scores before and after optimization  
**表 3. 优化前后翻译错误率变化及质量评分对比**

错误类型	典型表现	优化前发生率 (%)	优化后发生率 (%)	改进幅度 (%)	优化策略
语义硬译	逐字直译，忽视整体语义	36.2	13.5	▼22.7	思维链引导型提示
文化意象丢失	省略隐喻义或文化内涵	44.8	19.7	▼25.1	文化背景补充型提示
表达不地道	词汇搭配或句式不符合英语习惯	28.4	11.2	▼17.2	Few-shot 学习示例
过度归化	用英语文化概念替代中国文化特色	19.6	8.3	▼11.3	显性文化保留指令
语境误判	未根据特定领域选择恰当译法	31.7	14.9	▼16.8	领域背景提示
整体质量评分	COMET 自动化评分(0~1)	0.731	0.868	▲0.137	复合优化策略
人工评分	百分制综合得分	83.7	91.2	▲7.5	复合优化策略

结果分析：

(1) “文化意象丢失”是最突出的问题，优化前发生率高达 44.8%。典型案例如将“绿水青山”简单译为“green water and green mountains”，完全忽略了这一表达所承载的生态文明理念与诗意图美。通过采用文化背景补充型提示(“请说明‘绿水青山’在中国生态文明建设中的象征意义，并选择既保留文化特色又易于国际读者理解的译法”)，错误率降至 19.7%，改进幅度达 25.1%。优化后，模型多能选择“lucid waters and lush mountains”这一已被广泛接受的官方译法，或采用“pristine environment”等更具概念性的表达。

(2) “语义硬译”问题同样严重，优化前发生率为 36.2%。这主要是因为模型倾向于将四字格拆解为四个独立语素进行字面翻译，而未能把握其作为完形的整体语义。例如，将“巧夺天工”译为“clever takes heaven work”这样完全不符合英语表达逻辑的组合。通过引入思维链引导型提示，要求模型先分析整体含义再选择表达方式，错误率显著降至 13.5%。优化后，模型能够生成“exquisitely crafted”或“masterly workmanship”等符合英语习惯的表达。

(3) “表达不地道”问题在优化前占 28.4%。这类错误多表现为词汇搭配不当(如“传承人”译为“inheritor”而非更常用的“bearer”或“custodian”)或句式累赘(如过度使用定语从句)。通过提供 Few-shot 学习示例(展示 2~3 个高质量译例)，模型能够学习到更地道的表达模式，错误率降至 11.2%。

(4) “过度归化”问题虽然发生率相对较低(19.6%)，但其危害不容忽视。例如，将“邻里守望”译为“community solidarity”虽然通顺，但完全抹去了中国传统邻里文化的特色。通过在提示词中加入“请保留源语文化特色，避免完全用目标语文化概念替代”的显性指令，此类错误明显减少。

(5) 从整体质量评分来看，优化后 COMET 评分从 0.731 提升至 0.868，提升幅度达 0.137(约 18.7%)；人工评分从 83.7 分提升至 91.2 分，提升 7.5 分，成功跨入“优秀”等级。这充分证明了基于提示工程的优化策略对于改善大语言模型翻译质量的有效性。

#### 4.2.1. 思维链消融实验结果

表 4 展示了思维链各环节对三个评价维度的贡献度分析(基于 150 组高错误率语料的测试结果，数据为三款模型的平均值)：

**Table 4.** Analysis of the contribution of each chain-of-thought component to translation quality  
**表 4. 思维链各环节对翻译质量的贡献度分析**

提示词变体	语义准确性提升(分)	表达流畅性提升(分)	文化传递力提升(分)	综合得分提升(分)
基线(无优化)	0	0	0	0
CoT-Step 1 (语义分析)	+5.8	+2.1	+4.3	+4.2
CoT-Step 2 (受众意识)	+1.9	+3.4	+6.7	+4.1

续表

<b>CoT-Step 3(策略选择)</b>	+2.3	<b>+4.2</b>	+5.1	+3.9
<b>CoT-Step 1 + 2</b>	+6.2	+4.8	+9.4	+7.0
<b>CoT-Step 2 + 3</b>	+3.1	+5.9	+10.2	+6.6
<b>CoT-Full(完整版)</b>	<b>+7.1</b>	<b>+6.5</b>	<b>+12.8</b>	<b>+9.2</b>

结果分析:

(1) Step 1(语义分析)对“语义准确性”贡献最大(+5.8分),这证实了显性化的语义拆解能有效抑制模型的逐字直译倾向。要求模型先分析“字面意思和比喻意义”,实际上是强制其进行元认知操作,从而激活更深层的语义表征。

(2) Step 2(受众意识)对“文化传递力”贡献最显著(+6.7分),这表明提醒模型考虑目标读者的文化背景,能够促使其在译法选择上更注重跨文化可理解性,减少文化意象的直接丢失。

(3) Step 3(策略选择)对“表达流畅性”贡献最明显(+4.2分),这说明显性的策略意识能够引导模型选择更符合目标语表达习惯的句式和词汇搭配。

(4) 完整的三步骤CoT效果最优(综合提升9.2分),且其效果大于任意两步组合,这证实了多步骤推理的协同效应:语义理解为策略选择提供基础,受众意识为表达调整提供方向,三者相互增强。

(5) CoT-Step 2+3组合在文化传递力维度的提升(+10.2分)接近完整版(+12.8分),这提示在计算资源受限的情况下,可以优先保留“受众意识+策略选择”这一核心环节,以较低的提示成本获得接近最优的文化传译效果。

#### 4.2.2. 模型定制化提示词对照实验结果

表5展示了通用提示词与定制化提示词在三款模型上的效果对比(选取50组语料进行测试):

**Table 5.** Performance comparison: generic prompts vs. customized prompts

**表5.** 通用提示词 vs. 定制化提示词效果对比

模型	提示词类型	语义准确性	表达流畅性	文化传递力	综合得分
<b>DeepSeek</b>	通用CoT	92.3	88.1	83.7	88.6
	定制化(强调地道表达)	92.1(-0.2)	<b>91.4(+3.3)</b>	84.2(+0.5)	<b>90.1(+1.5)</b>
<b>通义千问</b>	通用CoT	89.7	91.2	82.4	88.1
	定制化(强调文化保留)	89.9(+0.2)	90.8(-0.4)	<b>87.6(+5.2)</b>	<b>89.8(+1.7)</b>
<b>文心一言</b>	通用CoT	88.6	87.3	85.1	87.2
	定制化(强调简洁性)	89.1(+0.5)	<b>89.7(+2.4)</b>	84.8(-0.3)	<b>88.4(+1.2)</b>

结果分析:

(1) 定制化提示词在针对性维度上均取得显著改善: DeepSeek 的表达流畅性提升 3.3 分, 通义千问的文化传递力提升 5.2 分, 文心一言的表达流畅性提升 2.4 分。这证实了根据模型特性设计针对性指令的有效性。

(2) 通义千问在文化传递力上的改进幅度最大(+5.2分),这可能与其原本在该维度的相对弱势有关——针对性优化对短板维度的提升效果更明显。

(3) 定制化提示词偶尔会导致其他维度的轻微下降(如文心一言的文化传递力-0.3分),这提示过度强调某一方面可能会分散模型对其他方面的注意力资源,需要在实际应用中进行权衡。

(4) 综合来看,定制化策略使三款模型的综合得分平均提升1.5分,虽然提升幅度不如完整CoT(9.2分),但在已应用CoT的基础上,定制化指令可作为“锦上添花”的进一步优化手段。

### 4.3. 定性分析: 典型错误案例的认知机理解读

通过对基线译文中高频错误案例的深度剖析,本研究识别出大语言模型在处理四字格块状语时的三大认知障碍:

(1) 完形认知能力不足,倾向解构性翻译

四字格块状语因其高度凝练的结构和对称的形式,往往呈现出格式塔(Gestalt)特征,即整体意义大于部分之和。例如“薪火相传”,其整体语义是“文化或技艺的代际传承”,而非“薪柴”和“火焰”的简单相加。然而,基于序列预测机制的大语言模型天然倾向于逐词元处理,容易陷入“字面陷阱”。在未经优化的情况下,三款模型均频繁出现将“薪火相传”译为“pass the firewood and fire”或“passing on the torch”这样过于字面的表达,后者虽保留了“火”的意象,但在英语中“pass the torch”多用于体育或政治领域的权力交接,而非文化传承,存在语境偏离。

为进一步剖析这一现象的深层机理,本研究对典型错误案例进行了细致的错误归因分析。以“巧夺天工”为例,三款模型在基线测试中的表现各异。通过分析模型输出的token概率分布,我们发现:当模型遇到“巧”字时,其激活的高频英文对应词依次为“clever”(0.32)、“skillful”(0.28)、“ingenious”(0.21);遇到“夺”字时,对应词为“take”(0.45)、“seize”(0.23)、“win”(0.18)。这种逐字的概率匹配机制导致模型难以跳出字面层次,进入成语的整体语义域。相比之下,人工译者会直接激活“exquisitely crafted”这类英语中表达“技艺高超”的惯用表达,完全绕过了对四个汉字的逐一对应。

这一差异的根源在于预训练数据中的成语分布特征对模型“概率匹配”倾向的塑造作用。模型在预训练阶段既学习到了成语的整体用法,也接触了大量拆分语境。当遇到新的翻译任务时,模型倾向于采用“最安全”的策略——即对每个字进行独立翻译后再组合,以最大化覆盖各种可能的语境。这种策略虽降低了完全误译的风险,但也导致了译文的生硬和文化韵味的流失。这也就解释了为何模型在处理文化负载表达时表现不佳:其概率分布是在大量非专业翻译文本上训练出来的,自然缺乏对文化传译原则的敏感性。

(2) 文化图式激活机制缺失,导致“去语境化”倾向

大语言模型在预训练阶段虽接触了海量文本,形成了对各类表达的统计性认知,但缺乏像人类译者那样能够根据特定语境自动激活相应文化图式的能力。例如,“守望相助”在社会治理语境下特指邻里间的互助互爱关系,对应英语中的“mutual support”或“looking out for each other”;但在基线测试中,模型往往选择了语料库中出现频率最高的通用对应词“help each other”,这虽不算错误,但丢失了“守望”所蕴含的持续关注与情感联结的深层含义。这种“去语境化”倾向反映了模型在文化认知层面的浅表性。

为验证这一判断,本研究设计了语境敏感性对比实验:将同一四字格“精益求精”分别置于三个不同领域的句子中(非遗技艺:“苏绣技艺……每一针每一线都凝聚着匠人的精益求精”;绿色经济:“企业在节能减排上精益求精”;社会治理:“社区服务精益求精”),观察模型的译法选择。结果显示:DeepSeek在三个语境下分别生成“excellence seeking”“continuous improvement”“refine service”,呈现一定的语境适应性;通义千问在三个语境下均生成“striving for perfection”,表现出语境不敏感;文心一言生成“pursuit of perfection”“optimization”“meticulous service”,语境适应性居中。这一实验证实,不同模型在文化图式激活方面存在差异,但总体而言均未达到人类译者根据语境灵活调整译法的水平。

(3) 隐喻映射能力受限,难以实现跨文化意象转换

许多四字格块状语包含丰富的隐喻义,如“绿水青山”以自然景观隐喻良好的生态环境,“巧夺天

工”以“夺”这一动词隐喻人工技艺之精湛甚至超越自然造化。这些隐喻的跨语言转换需要在源语文化与目标语文化之间建立意象桥梁。然而，模型往往停留在表层的符号转换，未能实现深层的概念映射。例如，将“历久弥新”直译为“long time more new”，完全丧失了“历经时间考验反而愈加焕发活力”这一动态演化的哲学意蕴。

通过对 30 个高隐喻性四字格(如“天人合一”“源远流长”“薪火相传”等)的专项测试，本研究发现模型的隐喻识别率仅为 42% (即在 42% 的案例中，模型能够采用意译策略传达隐喻含义，而非直译表层意象)。进一步分析显示，模型对“具象→抽象”类隐喻(如“绿水青山”→“良好生态”)的处理成功率(56%)高于“抽象→抽象”类隐喻(如“守正创新”→“坚持原则与开拓进取的平衡”)的成功率(31%)。这表明模型的隐喻理解能力主要依赖于训练数据中的显性标注或高频搭配，而非真正掌握了隐喻映射的认知机制。

这些定性发现揭示，大语言模型在处理文化负载表达时的核心问题不在于语言知识的缺乏，而在于文化认知能力的不足和语境敏感性的缺失。

## 5. 发现与讨论

### 5.1. 完形认知与序列生成的本体论张力

本研究最核心的发现在于揭示了汉语四字格块状语的完形认知特性与大语言模型序列生成机制之间的深层矛盾。四字格作为汉语中最具代表性的块状语形式，其语义生成遵循完形心理学的格式塔原则——整体先于部分，部分服从整体。这种完形性不仅体现在形式的对称工整上，更体现在语义的整体浮现上。例如“守望相助”，四个语素之间形成了一个紧密的语义场：“守望”与“相助”构成因果关系，“守”与“望”、“相”与“助”分别构成并列递进，四个语素共同指向一个完整的社会伦理图式。

然而，基于 Transformer 架构的大语言模型，其运作本质是通过自注意力机制进行上下文建模，并以概率分布的方式逐词元预测下一个词元。这种线性的、递增的生成方式与四字格的完形认知存在天然的不兼容性。正如杨开烨等(2023)所指出的，机器翻译系统倾向于将块状语“原子化”处理，而忽视其作为认知整体的特殊性[1]。本研究的实证数据印证了这一观点：在未经优化的基线测试中，三款模型对四字格的翻译错误中，有超过三分之一属于“语义硬译”类型，即将四字格拆解为四个独立成分进行字面翻译。

然而，需要指出的是，将问题完全归结为“完形 vs 序列”的二元对立可能过于简化。近期关于 Transformer 注意力机制的研究表明，模型在处理短语或固定搭配时，确实会在某些注意力头中形成“整体性表征”(holistic representation)，即多个词元被共同激活并映射到同一语义空间。本研究通过调用 DeepSeek 的 API 接口，提取了模型在翻译“巧夺天工”时最后一层的注意力权重矩阵(虽然无法获得完整的注意力可视化，但可以通过输出概率的变化推断注意力分布模式)。分析发现：当提示词为标准翻译指令时，“巧”“夺”“天”“工”四个字元对应的注意力权重较为分散，分别为 0.21、0.19、0.28、0.32，表明模型在独立处理各字；而当提示词改为“请将成语‘巧夺天工’翻译成英文”(显性标注为成语)时，四个字元的注意力权重趋于集中(0.24、0.25、0.26、0.25)，且生成的译文从“clever takes heaven work”改进为“exquisitely crafted”。这一现象提示：模型并非完全缺乏整体处理能力，而是其默认策略倾向于逐元素处理，需要通过显性标记来触发整体处理模式。

更深层的问题在于，即使模型通过大规模预训练习得了某些四字格的常见译法，但这种习得本质上是基于统计共现而非真正的语义理解。周领顺和陈龙宇(2025)在讨论机译者行为时强调，AI 翻译的本质是模式识别与概率推理，缺乏人类译者的主体性意识和文化敏感性[5]。本研究通过对典型错误案例的分

析发现，当四字格出现在非典型语境或与其他块状语组合使用时，模型的表现显著下降，这恰恰说明其并未真正掌握四字格的认知生成机制，而只是记住了某些固定搭配。

此外，预训练数据中成语分布的“长尾效应”加剧了这一问题。根据对三款模型预训练语料的抽样分析(基于公开的语料统计数据)，常用成语(使用频率前 20%)的平均出现次数是低频成语(使用频率后 20%)的 47 倍。这导致模型对高频成语(如“精益求精”“与时俱进”)形成了较为稳固的整体表征，而对低频或新造的四字格(如“智改数转”“云上办公”)则缺乏充分的学习样本，只能回退到逐字翻译的基础策略。这一发现为优化训练策略提供了启示：针对文化传译任务，有必要在预训练或微调阶段增加高质量成语翻译语料的权重，以缓解长尾效应对低频文化表达翻译质量的负面影响。

从翻译认知的角度看，人类译者在处理四字格时会先进行整体性的语义把握(即激活相应的认知框架)，再根据目标语的表达习惯进行重构。这一过程是“自上而下”的整体到部分。而大语言模型的处理则是“自下而上”的部分到整体，即先识别各个语素，再试图拼接成完整语义。这种本体论层面的差异导致了机器译文往往呈现“形似而神异”的特征——表面上词汇对应关系清晰，但整体上缺乏语义的连贯性和文化的韵味。

## 5.2. 文化图式的激活困境与“去语境化”陷阱

第二个重要发现涉及大语言模型在处理文化负载表达时的图式激活问题。认知语言学认为，语言理解本质上是图式激活的过程，即通过语言符号触发大脑中预存的知识结构和经验模式。对于四字格块状语而言，其理解不仅需要激活语言图式(词汇、语法)，更需要激活文化图式(历史典故、社会习俗、价值观念)。例如，理解“邻里守望”需要激活中国传统社会中基于地缘关系的互助文化图式；理解“绿水青山”需要激活生态文明建设的政策话语图式。

然而，本研究发现，大语言模型在文化图式激活方面存在系统性缺陷，主要表现为“去语境化”倾向。所谓“去语境化”，是指模型倾向于选择该词块在训练语料库中出现频率最高、分布最广的“平均译法”，而忽略了特定文本类型、特定领域语境对译法选择的制约。例如，“精益求精”在非遗技艺语境下强调的是工匠精神，应译为“*pursuit of perfection*”或“*striving for excellence*”；在经济管理语境下可能更侧重流程优化，可译为“*continuous improvement*”。但在基线测试中，模型往往一概使用“*excellence seeking*”这样过于直译且不地道的表达。

这种现象背后的认知机理可以用高语境理论(High-context Theory)来解释。中国文化属于高语境文化，语言表达高度依赖情境，许多信息被隐含在语境中而非明示在文本中。四字格块状语正是这种高语境特性的典型体现——四个字的表层结构之下，隐藏着丰富的文化预设和情境信息。而大语言模型的训练目标是最大化预测准确率，这导致其倾向于选择“安全”的、适用范围最广的译法，以避免在不同语境中出错。但这种策略在处理高语境表达时恰恰产生了反效果：译文虽然“不出错”，但也“不出彩”，丧失了源语的文化特色和表达张力。

耿芳和胡健(2023)在研究 ChatGPT 的译后编辑能力时也注意到，AI 在处理文化负载词时表现欠佳，需要人工介入进行文化适配[4]。本研究进一步发现，这种文化适配能力的欠缺不仅体现在单个词块的翻译上，更体现在对整体文本风格和传播效果的把控上。例如，在非遗文化语料中，原文往往带有一种典雅古朴的文化韵味，而机译往往将这种韵味“扁平化”为平白直叙的说明文风格，未能实现文化传递力维度上的有效传达。

## 5.3. 提示工程作为认知支架的补偿效能

本研究的第三个核心发现是提示工程在优化大语言模型文化翻译能力方面的显著效能。实验数据表

明，通过精心设计的提示词模板，文化传递力维度的错误率下降了 25.1%，整体质量评分提升了 18.7% (COMET 评分)和 7.5 分(人工评分)。这一结果揭示，大语言模型并非完全缺乏文化知识，而是这些知识处于“隐性状态”，需要通过恰当的提示词来“激活”。

从认知科学的角度看，提示工程实质上构成了一种“认知支架”(Cognitive Scaffolding)。维果茨基(2025)的最近发展区理论认为，学习者在适当的支持下能够完成超出其当前能力水平的任务[12]。类似地，大语言模型通过提示词的引导，能够展现出超越其基线表现的翻译能力。特别是 Chain-of-Thought(思维链)提示策略，通过将翻译过程分解为“理解 - 分析 - 生成”三个步骤，强制模型进行显性化推理，有效避免了直接跳转到表层符号转换的倾向。

然而，消融实验也揭示了 CoT 各环节的差异化功能：并非所有步骤对所有维度的贡献均等。“语义分析”步骤主要改善语义准确性，“受众意识”步骤主要提升文化传递力，“策略选择”步骤主要优化表达流畅性。这一发现为实际应用提供了灵活性：当翻译任务侧重文化传播时，可以强化“受众意识”环节；当任务侧重技术文档等信息性文本时，可以强化“语义分析”环节。此外，CoT-Step 2 + 3 组合在文化传递力上的表现(+10.2 分)接近完整版(+12.8 分)，这为资源受限场景下的提示优化提供了“性价比”更高的选择。

袁筱一(2025)在讨论人工智能文学翻译的“主体性”问题时指出，AI 缺乏人类译者的反思性思维和文化判断能力[6]。本研究发现，通过思维链提示，可以在一定程度上“模拟”这种反思性思维。例如，当要求模型“第一步：解释‘巧夺天工’的字面意思和比喻意义”时，模型被迫先进行元语言分析，这个过程实际上是对其自身知识库的一次显性检索和整理，从而提高了后续翻译决策的准确性。

然而，需要指出的是，提示工程的有效性存在边界。本研究发现，对于那些涉及深层文化典故或需要创造性转换的四字格(如“天人合一”这类哲学概念)，即便采用了优化提示，模型的表现仍不够理想。这说明，提示工程能够激活模型已有的知识，但无法创造出模型训练数据中不存在的知识。Al Sawi 和 Allam (2024)在视听翻译研究中也发现，AI 虽能识别文化典故的存在，但往往无法选择恰当的转换策略[7]。这提示我们，在处理高度专业化或深层文化性的翻译任务时，人工介入仍是不可或缺的。

#### 5.4. 人机协同：从工具理性到文化主体性的范式转换

本研究的最后一个发现是，在中国形象对外传播的语境下，人机协同模式不应仅被视为一种技术手段，更应被理解为一种文化实践和价值选择。传统的机器翻译评估往往采用工具理性视角，以效率和准确率为唯一标准。然而，当翻译对象涉及国家形象构建和文化身份表征时，单纯的工具理性视角已显不足，必须引入文化主体性和价值理性的考量。

周领顺(2024)在译者行为批评的“翻译外”研究中强调，翻译评价应超越文本层面，关注译者的社会角色和文化责任[13]。这一观点对理解大语言模型在中国形象翻译中的角色定位具有启发意义。本研究发现，在未经人工干预的情况下，大语言模型倾向于采用过度归化策略，即用目标语文化中的既有概念替代源语文化特色，以追求表达的流畅性和可接受性。例如，将具有中国特色的“乡贤”简化为“local elite”或“gentry”，虽然便于西方读者理解，但抹去了中国乡村治理中“贤德为先”的伦理维度。

这种过度归化倾向实际上反映了训练数据中隐含的“西方中心主义”偏见。Lyu 等(2024)在跨文化设计转译研究中指出，生成式 AI 往往倾向于标准化、国际化的审美，而对文化特殊性的把握不足[8]。在翻译领域，这种倾向可能导致中国形象在“翻译 - 传播”链条中被无意识地“异化”，即被改造成符合西方文化预设的样子，从而失去自身的文化主体性。

范大祺和孙琳(2023)强调，在中国特色对外话语体系建设中，译者应坚持“受众中心论”的同时，也要保持文化自信，避免一味迎合而丧失文化立场[14]。这一观点对于指导人机协同翻译实践具有重要意

义。本研究的优化实验表明，通过在提示词中加入“请在保持中国文化特色的前提下进行翻译”这样的显性文化立场声明，可以有效抑制模型的过度归化倾向。这说明，人机协同不应是简单的“人工补充机器不足”，而应是“人通过机器实现文化主体性表达”。在这个意义上，大语言模型只是工具，真正的“译者主体”仍是具有文化判断力和价值选择能力的人。

## 6. 结语

本研究通过构建“语义准确性 - 表达流畅性 - 文化传递力”三维评价体系，对 DeepSeek、文心一言、通义千问三款国产主流大语言模型在中国形象类四字格块状语英译中的表现进行了系统性评测与优化探索。研究发现，尽管大语言模型在语义理解和句法生成方面已达到较高水准，但在文化传递力维度仍存在显著短板，主要表现为完形认知能力不足导致的解构性翻译、文化图式激活缺失导致的“去语境化”倾向、以及隐喻映射能力受限导致的文化意象丢失。通过引入基于思维链和少样本学习的提示工程策略，可以有效激活模型的隐性文化知识，显著提升翻译质量，特别是文化传递力维度的表现。

然而，提示工程的优化效能存在边界，对于涉及深层文化典故或需要创造性转换的表达，人工介入仍不可或缺。本研究强调，在中国形象对外传播的语境下，人机协同翻译不应仅被视为一种技术手段，更应被理解为一种文化实践和价值选择。大语言模型作为工具，其使用必须在人的文化主体性指导下进行，以避免因过度归化或文化误读而导致中国形象的扭曲传播。研究团队期待将来的研究可进一步扩展语料库的领域覆盖面，探索多模态大语言模型在图文结合的中国形象传播中的应用，并持续完善人机协同的标准化操作规程，为构建融通中外的话语体系、提升国家文化软实力提供更加坚实的技术支撑与理论指导。

## 基金项目

本文系浙江省教育厅科研项目“大语言模型对中国形象类块状语英译能力评测与优化研究”(项目编号：Y202558054)的研究成果。

## 参考文献

- [1] 杨开烨, 汪少华. 中国生态话语文化图式的认知阐释[J]. 外语研究, 2023, 40(1): 8-16.
- [2] 章明蕾. 文化认知差异视域下商务话语中国特色词汇翻译研究[J]. 现代英语, 2025, 10(10): 100-102.
- [3] 李珩. 从跨文化交际角度谈构建中国对外话语体系中的外宣翻译策略[J]. 延边教育学院学报, 2019, 33(5): 38-40.
- [4] 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.
- [5] 周领顺, 陈龙宇. AI 背景下的译者行为研究[J]. 中国外语, 2025, 22(3): 94-101.
- [6] 袁筱一. 人工智能文学翻译的“主体性”与“创造性”[J]. 上海交通大学学报(哲学社会科学版), 2025, 33(1): 1-10.
- [7] Al Sawi, I. and Allam, R. (2024) Exploring Challenges in Audiovisual Translation: A Comparative Analysis of Human-and AI-Generated Arabic Subtitles in Birdman. *PLOS ONE*, **19**, e0311020. <https://doi.org/10.1371/journal.pone.0311020>
- [8] Lyu, Y., Shi, M., Zhang, Y. and Lin, R. (2023) From Image to Imagination: Exploring the Impact of Generative AI on Cultural Translation in Jewelry Design. *Sustainability*, **16**, Article No. 65. <https://doi.org/10.3390/su16010065>
- [9] Santos, A.N.d., Souza, D.S.d. and Chishman, R. (2025) Reframing the Frame: Toward a Socio-Cognitive Perspective on Fillmorean Theory. *Cadernos de Estudos Linguísticos*, **67**, e025027. <https://doi.org/10.20396/cel.v67i00.8680047>
- [10] Banks, D. (2024) Systemic Functional Linguistics: Advances and Applications. *Journal of World Languages*, **10**, 1-8. <https://doi.org/10.1515/jwl-2023-0083>
- [11] Felber, S. (2025) Toward a Science of Translating? Eugene A. Nida and His Theory of Dynamic Equivalence. Frank & Timme GmbH.
- [12] Rigopouli, K., Kotsifakos, D. and Psaromiligkos, Y. (2025) Vygotsky's Creativity Options and Ideas in 21st-Century Technology-Enhanced Learning Design. *Education Sciences*, **15**, 257-276. <https://doi.org/10.3390/educsci15020257>

- [13] 周领顺. 译者行为批评中的“翻译外”研究——兼作翻译内外分析框架与“文本-行为-社会”三位一体评价系统的比较[J]. 中国外语, 2024, 21(5): 106-112.
- [14] 范大祺, 孙琳. 中国特色对外话语体系建设视域下的译者思政意识实践路径初探[J]. 北京第二外国语学院学报, 2023, 45(1): 80-90.