

# 大语言模型在跨文化语言理解中的潜力与局限

孔一朴\*, 郭政杰#

河北大学外国语学院, 河北 保定

收稿日期: 2026年4月22日; 录用日期: 2026年5月26日; 发布日期: 2026年6月8日

## 摘要

随着全球化的发展, 跨文化语言理解在日常交际、教育和人工智能应用中愈发重要。大语言模型(LLM)如ChatGPT在生成跨文化表达方面展现出一定潜力, 但其对文化语境和隐含意义的理解仍存在局限。本文以礼貌表达、幽默和隐喻为研究对象, 通过让ChatGPT生成典型跨文化表达示例, 结合跨文化语言理解理论, 对其生成内容进行分析。研究发现, LLM能够生成符合文化预期的语言表达, 但在面子、礼貌策略、情感微调和语境依赖上仍表现不足。这一研究为跨文化交流、语言教学及人工智能在跨文化应用中的改进提供了参考, 同时提出了大语言模型在理解文化内涵方面的局限性及优化方向。

## 关键词

ChatGPT, 跨文化语言理解, 礼貌策略, 幽默理解, 隐喻理解

# Potentials and Limitations of Large Language Models in Cross-Cultural Language Understanding

Yipu Kong\*, Zhengjie Guo#

School of Foreign Studies, Hebei University, Baoding Hebei

Received: April 22, 2026; accepted: May 26, 2026; published: June 8, 2026

## Abstract

With the progression of globalization, cross-cultural language understanding has become increasingly vital in daily communication, education, and artificial intelligence applications. Large Language

\*第一作者。

#通讯作者。

**Models (LLMs) such as ChatGPT have demonstrated significant potential in generating cross-cultural expressions; however, their understanding of cultural context and implicit meanings remains limited. Focusing on politeness expressions, humor, and metaphors, this paper analyzes the content generated by ChatGPT through typical cross-cultural expression examples, grounded in theories of cross-cultural linguistic understanding. The findings indicate that while LLMs are capable of generating language expressions that align with cultural expectations, they still exhibit deficiencies in handling face-saving, politeness strategies, emotional fine-tuning, and contextual dependency. This study provides insights for the improvement of cross-cultural communication, language teaching, and the cross-cultural application of AI, while highlighting the limitations of LLMs in grasping deep cultural connotations and proposing directions for optimization.**

## Keywords

**ChatGPT, Cross-Cultural Language Understanding, Politeness Strategies, Humor Understanding, Metaphor Understanding**

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

随着全球化和数字化的发展,跨文化交际成为现代社会不可忽视的重要议题。在跨文化环境中,不同文化背景的人们在语言理解上存在显著差异,包括礼貌策略、幽默表达、隐喻理解以及语境依赖等方面[1][2]。例如,在高语境文化中,如中国,语言表达往往依赖隐含意义和人际关系;而在低语境文化中,如美国,表达更倾向于直接、明示的方式。这些差异不仅影响日常交流,也对跨文化教育、国际合作以及语言技术的应用提出了挑战。

近年来,大语言模型(Large Language Models, LLM)如 ChatGPT 在自然语言生成和理解方面表现出极大潜力。这类模型能够基于大量文本数据进行语言生成,并在一定程度上模拟人类的对话行为[3]。然而,尽管 LLM 在生成语言上流畅且符合语法规范,其对跨文化语境和隐含信息的理解仍存在局限[4]-[6]。例如,当面对文化特定的礼貌表达或幽默隐喻时,模型可能无法正确把握文化背景或使用的语境线索,从而生成与目标文化理解不完全一致的回答。

在此背景下,本文旨在探讨 LLM 在跨文化语言理解中的潜力与局限。通过选取典型的礼貌、幽默和隐喻表达,让 ChatGPT 生成跨文化示例,并结合语言理解理论进行分析,本文将揭示 LLM 在跨文化语境下的表现特点及存在问题,为跨文化交流、语言教学和人工智能应用提供参考。

## 2. 文献综述

### 2.1. 跨文化语言理解理论

跨文化语言理解研究强调文化背景对语言意义和交际行为的影响[1]。Hall 将文化划分为高语境与低语境文化,高语境文化中的语言信息高度依赖语境、非语言提示及社会关系,而低语境文化更依赖语言本身的明示内容。例如,中国的日常交流中,人们在请求或拒绝时往往采取间接、委婉的表达,以维护面子和人际关系;而美国文化中,表达更直接明确[2]。

Brown 和 Levinson [2]提出,人们在语言交际中通过策略性表达来保护自身和他人的面子,其中包含

请求、拒绝、道歉等行为。这一理论为分析跨文化语言理解提供了操作性方法：不同文化在礼貌策略上的选择，直接影响语言理解和交际成功率。幽默和隐喻理解也是跨文化语言理解研究的核心问题。幽默常依赖特定文化知识、双关语或语境线索，不同文化的人对同一笑话的理解可能完全不同[7]。隐喻表达同样具有文化依赖性，如中文成语“画蛇添足”与英文谚语“Break a leg”，字面含义与文化内涵差异显著。理解这些表达需要语言知识与文化知识的结合，而不仅仅是字面解析。

## 2.2. 会话分析视角下的语言理解

会话分析(Conversation Analysis, CA)为理解语言如何在交际中运作提供了理论基础[8]。CA 强调对话中的轮换发言(turn-taking)、动作形成(action formation)和序列组织(sequence organization)，并关注参与者如何在对话中管理信息和知识(epistemics) [9]。这些概念对分析 LLM 生成的跨文化语言表达具有启发意义，因为模型虽能生成语法正确的文本，但是否能按文化语境合理安排对话轮次、动作和序列，是衡量其理解能力的关键。

## 2.3. 大语言模型与跨文化语言生成

近年来，LLM 在自然语言处理领域广泛应用，尤其是在对话生成、文本理解和语言翻译中[3]。这些模型通过训练海量文本数据，能够在语法、词汇选择和上下文衔接上达到较高水平。然而，研究指出，LLM 对文化依赖的表达仍存在局限[4] [5] [10]。例如，当被要求生成中文委婉拒绝或英文幽默回复时，模型可能无法充分理解文化面子、礼貌策略或幽默的隐含规则，而生成的文本更倾向于通用语言，缺乏文化微调[11]。计算语言学界的近期研究指出，LLM 在处理跨文化语义时仍存在明显的“文化偏差”，倾向于反映西方价值观，难以捕捉非西方语境下的深层社会属性[12]。例如，在生成中文“婉拒”或英文“幽默”时，模型往往因缺乏对文化面子和礼貌策略的深层建模，导致输出结果呈现出泛化的“随机鹦鹉”特征，即仅能模仿形式而无法理解交际意图[13]。

此外，LLM 在处理跨文化语言理解时，还面临语境理解不足的问题[14]。其生成文本缺乏对对话参与者身份、社会关系和文化背景的精细把握，这导致模型生成的回答虽然语法正确、信息完整，但在跨文化交际中可能显得不自然或不符合文化期望。除此之外，LLM 在跨文化交互中还面临社会语用失效的问题。现有证据表明，模型在感知对话参与者的身份地位、社会距离和文化背景方面表现出不稳定性。这导致其生成的回答虽然语法正确、信息完整，但往往无法在动态语境下进行灵活的策略微调，难以符合特定文化的社交期望[15]。这种局限性反映了模型在文化知识管理与语用推理能力上的深度缺失。

综上所述，跨文化语言理解研究强调文化背景、礼貌策略、幽默与隐喻在语言理解中的作用。会话分析提供了分析语言在交际中操作方式的理论工具。结合 LLM 的研究，本文认为大语言模型在跨文化语言生成上具有潜力，但在文化微调、语境依赖和隐含意义理解方面仍存在明显局限。这为后续的分析 and 讨论奠定了理论基础，也明确了本文研究的重点：通过生成示例分析 LLM 在跨文化语言理解中的优势与不足。

## 3. 研究方法

本研究旨在探讨大语言模型(LLM)在跨文化语言理解中的潜力与局限。为了保证实验的简洁性和可操作性，本文选择 ChatGPT-5.2 作为研究对象。该模型基于大规模文本数据进行训练，能够生成自然语言文本并模拟对话行为，在学术研究和实际应用中具有代表性，因此适合作为跨文化语言理解的实验工具。

为了评估模型在跨文化语境下的表现，本研究设计了三类典型输入，分别涉及礼貌策略、幽默表达和隐喻理解。礼貌策略的输入旨在考察模型对请求和拒绝的处理能力，例如中文输入“用中文写一句委婉拒绝朋友邀请的话”，以及英文输入“Politely decline a friend's invitation in English”。幽默表达的输入

用于测试模型对跨文化幽默的理解与生成, 例如中文输入“用中文写一句幽默回答‘猜猜我发生了什么事?’”, 以及英文输入“Respond humorously to ‘Guess what?’ in English”。隐喻理解输入则关注模型对文化特定习语或成语的解释能力, 如中文输入“解释中文成语‘画蛇添足’的含义”以及英文输入“Explain the meaning of the English idiom ‘Break a leg’”。这些输入设计基于跨文化语言理解理论, 充分考虑高低语境文化差异及礼貌策略的文化特征。在数据收集阶段, 每类输入生成三条不同输出, 以确保样本的多样性。生成文本均记录原文和模型输出, 并进行编号管理, 形成小规模样本库, 便于后续分析。

数据分析基于会话分析理论[16], 主要从话轮转换、动作形成、序列组织和知识管理四个维度进行考察。在话轮转换方面, 重点考察 ChatGPT 的输出是否与输入提示顺畅衔接, 以及模型在对话中的回应是否符合自然交流的顺序。在动作形成方面, 分析关注模型生成的内容是否体现了特定的语言行为, 如礼貌表达、幽默回应或隐喻使用, 评估其在跨文化语境中动作识别的准确性。在序列组织方面, 则考察文本的逻辑顺序和文化语境适配性, 看生成内容是否合理安排信息并保持对话连贯性。此外, 还分析模型在知识管理上的表现, 包括是否能够合理处理信息归属和共享, 即是否体现说话者与听者之间的关系。这些分析方法可以帮助系统地评估 ChatGPT 在跨文化语言理解中的优势与局限, 为后续的结果讨论提供可靠的数据支撑。

## 4. 数据分析

本研究基于会话分析理论, 从话轮转换、动作形成、序列组织和知识管理四个维度分析 ChatGPT 在跨文化语言理解中的表现, 并结合跨文化交际理论评估模型在不同文化语境下的优势与局限。

### 4.1. 委婉拒绝表达的跨文化比较

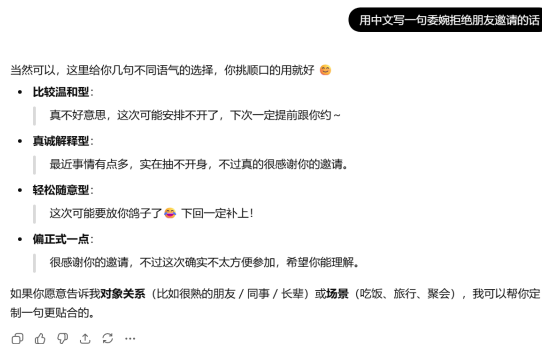


Figure 1. Chinese expressions for polite refusal  
图 1. 中文委婉拒绝表达

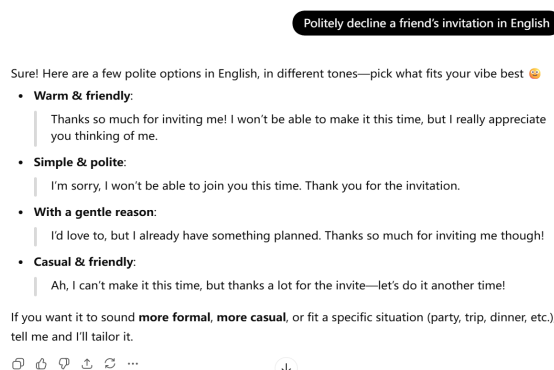


Figure 2. English expressions for polite refusal  
图 2. 英文委婉拒绝表达

在互动策略与动作形成方面, ChatGPT 生成的中文表达体现了高语境文化特征, 常通过前置道歉和理由解释(如“真不好意思”)作为情感缓冲, 以维护面子和社会和谐(如图 1); 英文表达则符合低语境文化, 强调信息明确与效率, 通常在简短感谢后直接说明拒绝(如图 1)。尽管模型能准确模仿两者的礼貌策略, 但其输出具有明显的模板化特征, 缺乏自然对话中的轮换结构, 且未能根据社会关系的亲疏远近进行差异化调整[14]。

从序列组织来看, 中文拒绝(如图 1)多采用“歉意 - 理由 - 未来补偿”的递进式结构, 侧重关系维护; 英文(如图 2)则呈“感谢 - 拒绝 - 安排说明”的线性逻辑。ChatGPT 虽能遵循上述序列逻辑, 但其生成内容多为独立、静态的文本单位, 缺乏真实交际中根据对方回应进行实时调整的动态能力, 难以体现互动过程中的连续性。

在知识管理维度, 中文表达(如图 1)高度关注听者的情感状态, 通过解释暗示对互动关系的重视; 英文(如图 2)则更多体现说话者对个人权利和自身安排的直接所有权。ChatGPT 虽能模仿这些模式, 却无法动态考虑交际双方的知识状态。在现实对话中, 说话者会根据听者的了解程度调整解释细节, 而模型往往提供泛化理由, 缺乏针对特定对象的知识协商过程。

综上所述, ChatGPT 在跨文化委婉拒绝中表现出较强的语言模仿能力, 能准确捕捉不同文化下的礼貌结构。然而, 其生成机制仍以静态模板为主, 在高语境文化中难以完全实现深层的面子维护策略, 在低语境文化中则缺乏对具体社交情境的细腻适配。这种局限性表明, 大语言模型在理解并实践动态、关系导向的复杂交际行为方面仍有待优化。

## 4.2. 幽默回应表达的跨文化比较

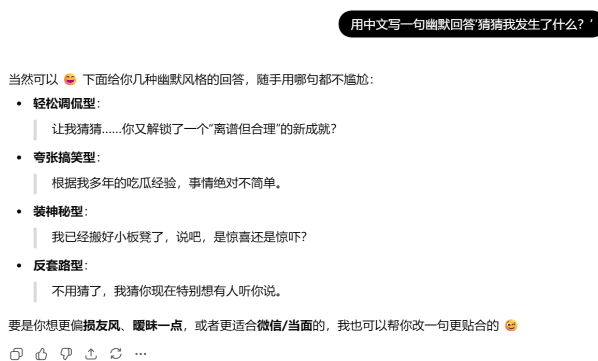


Figure 3. Humorous response in Chinese  
图 3. 中文幽默回应

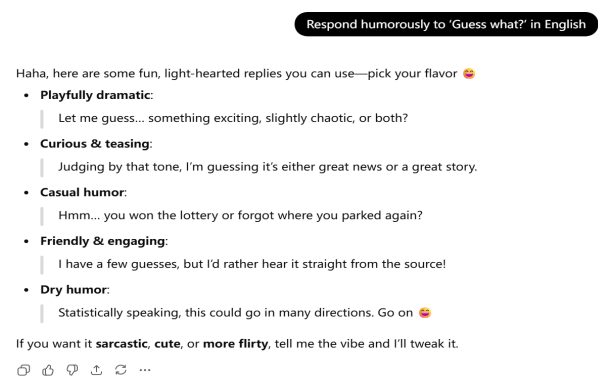


Figure 4. Humorous response in English  
图 4. 英文幽默回应

在互动延展与动作形成方面, ChatGPT 生成的中文幽默回应(如图 3)具有强烈的关系维系特征, 常通过调侃式猜测构建情感共鸣, 符合高语境文化中强调互动氛围的特点; 英文(如图 4)回应则多采用半开放式猜测, 侧重于引导下一话轮的展开, 体现了低语境文化强调对话推进的功能。在动作设计上, 中文(如图 3)表达偏好夸张和文化符号, 利用共享语境强化亲密度; 英文(如图 4)则更多依赖情境对比或轻度自嘲。尽管模型能准确模仿不同语言的幽默模式, 但其输出仍具静态特征, 缺乏真实对话中的非言语反馈(如笑声、语气词), 且难以针对受众的文化背景和年龄差异调整幽默的分寸。

从序列组织与知识管理维度看, 中文幽默回应(如图 3)倾向于铺垫式结构, 通过层层引导构建氛围; 英文(如图 4)则倾向于简洁的线性结构, 追求信息传递效率。在知识管理上, 中文表达(如图 3)常假设一种群体共通的语境预设, 通过特定网络流行语暗示说话者与听者的共同身份; 英文(如图 4)则倾向于明确界定知识边界, 通过直接承认信息未知来鼓励对方叙述。ChatGPT 虽能捕捉上述序列逻辑和知识分配习惯, 但其局限在于: 模型通常默认受众拥有完整的文化背景知识, 缺乏对听者知识范围的动态评估。

综合四个维度可以看出, ChatGPT 在幽默回应表达中能够有效模仿不同语言文化中的幽默策略。中文幽默表达更强调关系维系和文化共享, 而英文幽默表达更强调互动推进和信息明确。然而, 模型在两种语言中均存在互动动态不足的问题, 尤其在跨文化语境中难以准确评估听者文化背景和幽默接受程度。这表明, 虽然人工智能在语言生成方面具有较强优势, 但在复杂跨文化幽默交际中仍存在一定局限性[17]-[19]。

### 4.3. 文化隐喻表达的跨文化比较

在话轮与序列组织上, ChatGPT 对中文成语(如“画蛇添足”)的解释(如图 5)遵循从字面意义到引申含义、再到应用场域的递进式教学结构, 强调信息的完整性与深度; 对英文习语(如“break a leg”)的解释(如图 6)则采用功能导向型结构, 优先说明语用功能及语气提示, 体现了英语文化强调实用性与效率的特点。尽管模型能准确重构中英文不同的解释逻辑, 但其输出表现出明显的标准化特征, 缺乏真实实际中根据听者反馈进行的动态话轮协商。

从动作形成与知识管理维度看, 中文解释(如图 5)体现为知识传递型行为, 通过故事化叙述强化集体文化认同, 旨在帮助听者建立系统认知; 英文解释(如图 6)则更强调社会语境的界定, 通过引入文化传统提示语用风险。然而, 模型的局限在于其生成内容主要基于统计语言模式, 而非针对具体听者的文化背景或认知差异进行动态知识调节。研究发现, 模型在解释成语时往往忽略了跨文化受众可能产生的理解偏差, 这反映了大语言模型在处理隐喻等非字面语言时, 其跨文化动作适配能力依然有限[20]。

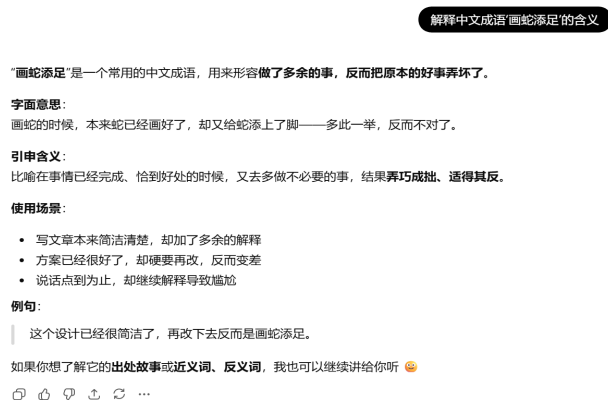


Figure 5. Explanations of Chinese idioms  
图 5. 中文成语解释

Explain the meaning of the English idiom 'Break a leg'

The idiom **"break a leg"** means to wish someone good luck, especially before a performance like a play, concert, speech, or audition.

**Why it sounds so odd:**

In theater tradition, people believed that directly saying "good luck" might bring bad luck. So instead, they used the opposite—"break a leg"—as a kind of superstitious workaround.

**How it's used:**

- Said **before** someone goes on stage
- Common in **theater, music, dance, and public speaking**
- Friendly and encouraging, not literal at all

**Example:**

You've been rehearsing for weeks—break a leg tonight!

**Tone tip:**

It's informal and supportive, so it's great with friends or colleagues, but not ideal for very formal situations.



If you want, I can share similar idioms or explain where this one might come from historically 🗣️

Figure 6. Explanations of English idioms

图 6. 英文习语解释

综合来看, ChatGPT 在文化隐喻表达解释中展现出较强的语言生成能力, 能够模拟中英文文化中不同的解释策略。中文解释更侧重文化背景与意义递进, 而英文解释更强调语用功能与情境适配。然而, 大语言模型虽能重现文化表达形式, 它在跨文化交际中仍存在局限, 在深层文化理解与动态交际能力上仍有提升空间[21]。

## 5. 结语

本研究基于会话分析框架, 深入探讨了大语言模型在跨文化理解中的表现, 重点剖析了其话轮转换、动作形成、序列组织及知识管理能力。研究发现, 虽然以 ChatGPT 为代表的模型在句法规范和形式模拟上展现出极强潜力, 能熟练应对各类跨文化交际的表面特征, 但在动态语境适配、深层意图建模以及面子维护策略上, 依然存在难以逾越的障碍。

这种“重形式、轻内涵”的局限性有着深刻的技术根源。从架构来看, 现有的 Transformer 模型本质上是基于统计概率的词语预测器, 它擅长捕捉文本表面的共现规律, 却缺乏对社交行为背后真实意图的建模。这导致模型在处理委婉拒绝或幽默这类高度依赖语境推断的任务时, 往往出现“语用失效”。同时, 由于预训练语料过度集中于西方工业化国家, 模型内部形成了一套以低语境文化为基地的逻辑, 在应对汉语等高语境文化的复杂人际关系时, 极易输出机械且泛化的通用文本。此外, 现有的 RLHF (基于人类反馈的强化学习) 算法过度追求回答的标准性与安全性, 在无形中压制了文化表达的灵活性与多样性, 使生成内容显得平淡且模板化。

要突破这些技术瓶颈, 未来的优化应从以下几个务实方向展开。首先是数据层面的精细化, 即构建包含社交关系标签(如职位高低、亲疏远近)的对话集, 让模型在学习语言的同时理解社交动态。其次, 可以引入检索增强生成(RAG)技术, 在模型开口前先检索特定文化的社交准则库, 以此纠正模型自带的文化偏见。在算法改进上, 应尝试将“社交得体度”引入评价体系, 研发专门的语用奖励函数, 引导模型从说出正确的词向说出得体的话转变。最后, 通过优化思维链(CoT)推理, 强制模型在生成最终回复前先分析对方的身份和潜在的心理预期, 使交互过程从简单的统计模拟转向更有深度的逻辑推理。

总而言之, 大语言模型正在改变人机交互的边界, 但目前的局限性提醒我们, 它距离真正的文化自觉仍有路要走。通过对数据集进行文化解耦、引入外部语境知识库以及优化语用导向的算法, 我们有望在未来构建出真正具备文化温度与交际智慧的智能系统。

## 参考文献

- [1] Hall, E.T. (1976) *Beyond Culture*. Anchor Books.
- [2] Brown, P., Levinson, S.C. and Gumperz, J.J. (1987) *Politeness: Some Universals in Language Usage*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511813085>
- [3] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., *et al.* (2023) GPT-4 Technical Report. arXiv:2303.08774.
- [4] Alkhamissi, B., ElNokrashy, M., Alkhamissi, M. and Diab, M. (2024) Investigating Cultural Alignment of Large Language Models. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), Bangkok, 11-16 August 2024, 12404-12422. <https://doi.org/10.18653/v1/2024.acl-long.671>
- [5] Bhatt, S. and Diaz, F. (2024) Extrinsic Evaluation of Cultural Competence in Large Language Models. *Findings of the Association for Computational Linguistics: EMNLP 2024*, Miami, 12-16 November 2024, 16055-16074. <https://doi.org/10.18653/v1/2024.findings-emnlp.942>
- [6] Ma, B., Li, Y., Zhou, W., Gong, Z., Liu, Y.J., Jasinskaja, K., *et al.* (2025) Pragmatics in the Era of Large Language Models: A Survey on Datasets, Evaluation, Opportunities and Challenges. *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), Vienna, 27 July-1 August 2025, 8679-8696. <https://doi.org/10.18653/v1/2025.acl-long.425>
- [7] Ladilova, A. and Schröder, U. (2022) Humor in Intercultural Interaction: A Source for Misunderstanding or a Common Ground Builder? A Multimodal Analysis. *Intercultural Pragmatics*, **19**, 71-101. <https://doi.org/10.1515/ip-2022-0003>
- [8] Sacks, H., Schegloff, E.A. and Jefferson, G. (1974) A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, **50**, 696-735. <https://doi.org/10.2307/412243>
- [9] 彭欣. 如何开启叙事——基于汉语日常交谈的会话分析研究[J]. *语言教学与研究*, 2022(5): 60-72.
- [10] Adilazuarda, M.F., Mukherjee, S., Lavania, P., Singh, S.S., Aji, A.F., O'Neill, J., *et al.* (2024) Towards Measuring and Modeling “Culture” in LLMs: A Survey. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Miami, 12-16 November 2024, 15763-15784. <https://doi.org/10.18653/v1/2024.emnlp-main.882>
- [11] Yin, Z., Wang, H., Horio, K., Kawahara, D. and Sekine, S. (2024) Should We Respect LLMs? A Cross-Lingual Study on the Influence of Prompt Politeness on LLM Performance. *Proceedings of the Second Workshop on Social Influence in Conversations (SICoN 2024)*, Miami, 16 November 2024, 9-35. <https://doi.org/10.18653/v1/2024.sicon-1.2>
- [12] Zewail, A., Figueroa, A., Graham, J. and Atari, M. (2026) Moral Stereotyping in Large Language Models. *Proceedings of the National Academy of Sciences*, **123**, e2519941123. <https://doi.org/10.1073/pnas.2519941123>
- [13] Bender, E.M. and Koller, A. (2020) Climbing Towards NLU: On Meaning, Form, and Understanding in the Age of Data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, 5-10 July 2020, 5185-5198. <https://doi.org/10.18653/v1/2020.acl-main.463>
- [14] Wu, S., Yang, S., Chen, Z. and Su, Q. (2024) Rethinking Pragmatics in Large Language Models: Towards Open-Ended Evaluation and Preference Tuning. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Miami, 12-16 November 2024, 22583-22599. <https://doi.org/10.18653/v1/2024.emnlp-main.1258>
- [15] Chen, X., Li, J. and Ye, Y. (2024) A Feasibility Study for the Application of AI-Generated Conversations in Pragmatic Analysis. *Journal of Pragmatics*, **223**, 14-30. <https://doi.org/10.1016/j.pragma.2024.01.003>
- [16] Schegloff, E.A. (1995) *Sequence Organization in Interaction: A Primer in Conversation Analysis I*. Cambridge University Press.
- [17] 刘建达, 汪晓琪. 汉语母语者与非汉语母语学习者言语行为的元语用评估——以请求和道歉为例[J]. *现代外语*, 2022, 45(3): 344-356.
- [18] Chen, Y., Yang, C., Hu, T., Chen, X., Lan, M., Cai, L., *et al.* (2024) Are U a Joke Master? Pun Generation via Multi-Stage Curriculum Learning towards a Humor LLM. *Findings of the Association for Computational Linguistics ACL 2024*, Bangkok, 11-16 August 2024, 878-890. <https://doi.org/10.18653/v1/2024.findings-acl.51>
- [19] Zangari, A., Marcuzzo, M., Albarelli, A., Pilehvar, M.T. and Camacho-Collados, J. (2025) Pun Unintended: LLMs and the Illusion of Humor Understanding. *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, Suzhou, 4-9 November 2025, 27936-27971. <https://doi.org/10.18653/v1/2025.emnlp-main.1419>
- [20] 项成东, 王茂. 英汉动物隐喻的跨文化研究[J]. *现代外语*, 2009, 32(3): 239-247+328.
- [21] Sanchez-Bayona, E. and Agerri, R. (2025) Metaphor and Large Language Models: When Surface Features Matter More than Deep Understanding. *Findings of the Association for Computational Linguistics: ACL 2025*, Vienna, 27 July-1 August 2025, 17462-17477. <https://doi.org/10.18653/v1/2025.findings-acl.898>