



大语言模型与语言学理论之整合：理论融入与能力评估

王儒蕴¹²

北京外国语大学中国外语与教育研究中心 / 广州南方学院

【摘要】本文聚焦现有大语言模型与隐喻的交叉研究，围绕“理论融入”与“能力评估”两个核心问题，对相关研究进行系统梳理。本文首先考察概念隐喻理论、批评隐喻分析等语言学理论如何与大语言模型整合，分析其对模型表现能力所起的作用。其次，本文从语言学视角出发，从隐喻识别、解释、生成以及工具调用能力这四个维度对大语言模型的隐喻处理能力进行了界定，评估大语言模型在这四个维度上的表现能力。结合对这两大核心问题的考察，本文认为未来研究应加强语言学理论与计算方法融合的广度与深度，同时进一步探索评估及增强大语言模型语言学专业能力的方法。

【关键词】大语言模型；概念隐喻；隐喻处理能力；语言学理论

-
- 1 王儒蕴，北京外国语大学中国外语与教育研究中心 / 广州南方学院，研究方向：认知语言学。邮箱：274173999@qq.com。本论文由中央高校基本科研业务费专项资金资助，项目名称：概念隐喻理论驱动提示对大语言模型概念解释质量的影响研究，项目批准号：2026JX019。
 - 2 本文得到北京外国语大学王馥芳教授的悉心指导，谨致以诚挚谢意。



The Integration of Large Language Models and Linguistic Theory: Theories and Evaluation

Ruyun Wang¹²

National Research Center for Foreign Language Education, Beijing Foreign Studies University/
Nanfang College • Guangzhou

【 Abstract 】 This paper focuses on existing research at the intersection of large language models and metaphor, offering a systematic review organized around two core issues: theoretical integration and capability evaluation. First, it examines how linguistic theories, including Conceptual Metaphor Theory and Critical Metaphor Analysis, have been integrated with large language models and analyzes their contributions to model performance. Second, from a linguistic perspective, the paper defines the metaphor-processing capabilities of large language models along four dimensions: metaphor identification, interpretation, generation, and tool use, and evaluates model performance across these dimensions. Based on its examination of these two core issues, the paper argues that future research should broaden and deepen the integration of linguistic theory and computational methods, while further exploring approaches to evaluating and enhancing the linguistic expertise of large language models.

【 Keywords 】 Large Language Model; conceptual metaphor; metaphor-processing capabilities; linguistic theory

-
- 1 Ruyun Wang, National Research Center for Foreign Language Education, Beijing Foreign Studies University / Nanfang College • Guangzhou. Research interest: Cognitive Linguistics. Email: 274173999@qq.com. This paper is supported by the Fundamental Research Funds for the Central Universities under the project titled “Study on the Effects of Conceptual Metaphor Theory–Driven Prompting on the Quality of Conceptual Explanation in Large Language Models” (Grant No. 2026JX019).
 - 2 This paper was completed under the careful guidance of Professor Wang Fufang from Beijing Foreign Studies University, to whom I would like to express my sincere gratitude.

引言

隐喻广泛存在于人类思维与交流的各个领域，植根于人类理解世界及与世界互动的方式之中，随着大语言模型（Large Language Model, LLM）在自然语言处理任务中表现水平的不断提升，从隐喻研究视角出发的新问题获得越来越多的关注，这些问题包括大语言模型能否识别隐喻、解释隐喻、评价隐喻，是否能够构建隐喻背后的源域—目标域映射等。LLM 的隐喻处理能力不仅是一个自然语言处理问题，也是检验人工智能语言理解、概念映射能力和认知边界的重要切入点。已有研究表明，ChatGPT 等大语言模型在部分隐喻识别、文学隐喻解释、中文隐喻评估和多模态隐喻检测任务中表现出较强能力，但模型表现可能受到词汇重合、句长、提示方式和训练语料等因素的影响。

尽管已有研究开始对大语言模型的隐喻处理能力进行测评，探索增强方法，但从语言学理论视角出发对 LLM 的隐喻处理能力进行界定和评估、考察语言学理论与大模型技术整合的研究仍较为缺少。因此，通过对现有相关研究的梳理和总结，本文旨在揭示语言学理论在大语言模型隐喻研究上的整合现状以及 LLM 在隐喻任务上的处理能力，本文主要基于两个核心层面展开讨论：（1）大语言模型与语言学理论整合层面，语言学理论如何融入大语言模型隐喻研究，即如何为提示词工程、数据增强、评估指标等提供理论依据，对模型表现起到何种效果；（2）隐喻处理能力评估层面，即大语言模型是否真正理解隐喻，其隐喻处理能力存在哪些优势与局限。

一、研究现状概述

聚焦大语言模型与隐喻的交叉研究，本文系统考察了 2023 年 1 月至 2026 年 4 月期间收录于中国知网、Web of Science（WoS）和 arXiv preprint 的相关实证类研究论文。笔者以“大语言模型”和“隐喻”为关键词进行检索，经人工筛选，共计从前述三个文献库获取 38 篇与本文主题高度相关的文献。

总体来看，现有研究的关注点已从单纯的隐喻识别或检测扩展到隐喻解释、隐喻评价、多模态隐喻理解和社会话语分析等多个层面，检验 LLM 是否真正理解隐喻，探索其隐喻处理能力边界，测试任务呈现出明确的深化趋势。从整体结果来看，相关研究既展示了模型在隐喻识别、解释和评分等任务中较高水平的表现，也揭示了其在新颖隐喻、具身隐喻、源域—目标域阐释和概念一致性等方面的局限。



在理论融合的探索上，一部分研究开始将语言学理论和心理学理论引入 LLM 的隐喻研究，将理论与提示词工程、思维链（CoT）、数据增强、标注程序或评估框架相结合，这些理论包括概念隐喻理论、批评隐喻分析、认知心理学等。从研究特色来看，中国知网上所获取的相关文献较多关注 LLM 在中文隐喻分析任务中的应用价值和方法探索，如 ChatGPT 辅助批评隐喻分析、中文隐喻质量多维度评估、通感隐喻分析、中文隐喻成分识别以及多模态隐喻检测等（于艳春，2025；黄孝喜等，2026；盛坤 & 王中卿，2025；张声龙 & 刘颖，2024）。这些研究体现出应用导向，重视面向中文的任务建构，关注中文语境下大语言模型的隐喻识别、自动评分、数据增强和知识组织功能。而 WoS 及 arXiv preprint 上所获取的相关文献则在数量、研究范围和理论反思上更为多元，包括各层级的 LLM 隐喻能力测试实证研究，以及多模态、跨文化、政治话语分析等前沿领域的探索，突出能力评测、理论整合、方法创新和反思批判。

二、大语言模型与语言学理论之整合

2.1 大语言模型研究所整合的语言学理论

当前大语言模型隐喻研究在不同维度上整合了概念隐喻理论（Conceptual Metaphor Theory, CMT）、批评隐喻分析（Critical Metaphor Analysis, CMA）等语言学理论资源，本节对这些理论进行简要介绍。

（1）概念隐喻理论

概念隐喻理论（CMT）是隐喻研究中的重要理论基础，也是认知语言学核心理论之一，该理论最早由 Lakoff 和 Johnson 提出。概念隐喻理论认为隐喻并非单纯的修辞，而是人类借助相对具体、熟悉的经验来理解较为抽象的概念，其核心在于源域与目标域之间的系统性映射，隐喻是人类理解抽象概念、组织经验和进行推理的重要认知机制，即人类的概念系统在本质上是隐喻性的（Lakoff & Johnson, 1980）。杨启萌等（2025）提出概念隐喻理论可为计算建模提供框架，并展示了该理论在驱动隐喻计算处理研究中的核心价值。

（2）批评隐喻分析

批评隐喻分析（CMA）是由英国学者 Charteris Black 在 2004 年提出的一种语篇隐喻分析方法，该理论认为隐喻的语篇功能是人们建立意识形态和修辞动机的基础。CMA 要求对“这一隐喻是什么？它属于何种类型的隐喻？选择该隐喻的动机是什么？”等问题进行细致分析，旨在揭示选择隐喻的深层动机。

2.2 理论整合方式与特点

(1) 基于 CMT 和 CMA 的提示设计

在部分现有大语言模型隐喻能力研究中, CMT 和 CMA 不同程度地被整合进了提示设计, 整合程度由低至高表现为三种类型: 单条或特定提示词优化、连贯提示框架设计、整体提示范式。从第一类来看, 部分研究将 CMA 用于特定提示词的优化, 如于艳春 (2025) 将 ChatGPT 引入 CMA 中的隐喻识别环节, 通过提示词要求模型根据隐喻标准识别语篇中的隐喻性表达, 并进一步划分源域和目标域, 该方法的作用主要在于提升隐喻识别的客观性、效率和启发性。第二类研究进一步将 CMT 和 CMA 作为提示框架的设计理论依据, 应用于具体提示用语、提示流程和提示要点等, 如在多模态隐喻识别中, Zhang 等 (2025) 基于 CMT 中的源域、目标域和映射关系, 提出认知链 (CoC) 提示方法, 实验依次要求模型抽取潜在的隐喻实体、判断实体关系、再完成隐喻识别。结果显示, 采用该方法的模型表现优于既有基线模型, 表明基于 CMT 核心认知推理步骤的提示词能够帮助模型获得更可靠的隐喻信息并提升识别性能。此外, Meng 等 (2025) 也将 CMA 与 CoT 结合, 构建面向特朗普政治话语的提示框架, 要求模型依次完成语境分析、隐喻识别、隐喻解释和隐喻阐释四个步骤。CMA 不但提供理论基础, 还直接被作为提示词的组织逻辑, 在政治语境和社会文化维度分析上发挥优势 (Meng et al., 2025)。第三类研究, 提出将 CMT 作为一个整体提示范式 (Kramer, 2025), 要求模型以“认知主体”的身份进行隐喻处理相关任务, 将源域、目标域和跨域映射嵌入提示词与 CoT 式推理流程, 以此获得一个 CMT 增强模型, 并在隐喻识别、跨域推理、解释教学和隐喻阅读理解等任务中将其与基线模型进行比较, 结果显示 CMT 提示范式能够提升推理准确性、解释清晰度和隐喻连贯性 (Kramer, 2025)。

(2) 基于 CMT 的评估框架构建以及学科能力测评。

CMT 也被用作评估框架构建的理论基础, 黄孝喜等 (2026) 结合概念隐喻理论的核心思想, 提出基于大语言模型的中文隐喻多维度评估框架。该研究通过多轮对话和思维链提示, 引导模型先识别源域、目标域及映射关系, 再让模型对中文隐喻进行五个维度的评分。其实验显示, 大语言模型评分与人工评分具有较高一致性, 说明 CMT 导向的评估流程有助于提高模型评估结果的可行性和可靠性。

在部分研究中, CMT 还被作为能力测评对象, 发挥揭示模型学科认知边界的作用。蔡雨晨和王馥芳 (2024) 以概念隐喻理论为例, 从核心术语理解、理论框架理解、理论应用和高层次思考四个层级评估 GPT-4 的认知语言学专业能力, 尤其考察其隐喻性语言识别和概念隐喻抽取能力。该研究表明, GPT-4 在部分任务中表现较好, 但尚未达到初级认知语



言学研究者水平。

总体而言，CMT 和 CMA 语言学理论引入的共同效果在于提升了模型隐喻识别 / 检测的准确性、推理透明度和解释理据性，同时减少了纯人工识别的主观性，它们在目前大语言模型隐喻能力研究中的整合与贡献可概括为三个方面：一是为提示词框架和推理链提供理论依据，增强模型对隐喻结构的抽取能力；二是为隐喻识别、解释、质量评估和学科能力评估提供理论维度与评价标准；三是帮助模型输出更具有可解释性的分析过程。其效果总体上是积极的，尤其是在识别准确率、解释深度、概念映射清晰度和人机评分一致性等方面的提升上。但相关研究也作出提醒，语言学理论的融入并不能完全消除模型幻觉、过度标注、语义误判和专业知识不系统等问题，LLM 在复杂隐喻推理和频次统计上仍有错误，因此人机协作、人工校验在当前 LLM 隐喻研究中仍有必要（于艳春，2025；Zhang et al., 2025；Meng et al., 2025；Kramer, 2025；黄孝喜，2026；蔡雨晨 & 王馥芳，2024）。

三、大语言模型隐喻处理能力评估

乔姆斯基（Chomsky, 1965）区分了“语言能力”（competence）和“语言表现”（performance），语言能力由人内在语法规则的心理表征构成，是内隐的，而语言表现（performance）¹则是内在语言能力在具体情形中的表达和使用。由于 LLM 的语言学习与生成本质为数据驱动，依赖海量数据和神经网络训练，是一种基于统计概率对人类语言使用的模拟，本文参考乔氏观点，将 LLM 的隐喻处理能力界定为一种“表现能力”²，而非内在的语言能力。这种基于算法的隐喻处理能力包括了对隐喻的识别能力、解释能力、生成能力以及 LLM 在隐喻研究中的工具调用能力四个维度，其中，隐喻识别的准确性能够反映模型对源域和目标域跨域映射的把握，隐喻解释的合理性能够反映模型在识别基础上是否具备较好的深层语义推理能力，隐喻生成则进一步考验模型综合调用其语义推理机制、隐喻知识，甚至文化背景知识等的的能力，这三类任务都以 LLM 对隐喻的理解为其完成基础，且构成由低到高的层级性能力要求。另外，LLM 在隐喻研究中的工具调用能力也需基于其对隐喻的理解，能够反映 LLM 的隐喻处理水平，因此本文将这项能力也纳入探究其隐喻处理能力的考察维度。基于以上理据，本节结合现有研究中 LLM 在这四个维度的隐喻处理任务上的表现，对其能力进行分析、评价和总结。

1 乔姆斯基提出的 performance 也常被译作“语言行为”、“语言表达”等。

2 本文将文中提及的大语言模型在各隐喻任务上的能力均界定为一种“表现能力”。

3.1 隐喻处理能力的四维度研究

(1) 隐喻识别

隐喻识别和检测是大语言模型隐喻研究中最基础，也是发展最充分的任务，现有研究围绕大语言模型的隐喻识别或检测能力，形成了从词级隐喻检测、语篇隐喻识别到多模态隐喻识别、全篇文本标注和特定领域隐喻抽取的多层次研究格局。

在词级隐喻检测中，部分学者依据隐喻识别程序（Metaphor Identification Procedure, MIP）和选择偏好违背（Selectional Preference Violation, SPV）进行特征建模和任务设计，并测试其增强效果。如，Yang 等（2024）采用 SPV 作为动词隐喻的定义依据，认为动词与主语或宾语之间若不符合常规选择偏好，即可能构成隐喻。该研究不要求 LLM 直接判断隐喻，而是先利用 GPT-3.5 Turbo 生成目标动词的字面搭配表，同时抽取待测句中的主语—动词（SV）、动词—宾语（VO）关系，再将这些搭配映射到 Oxford、WordNet 或 LDA 主题类别，通过主题蕴含关系判断是否存在选择偏好违背，最终形成对动词隐喻的识别判断。结果表明，这种引导和组合的方法能够有效提升 ChatGPT 在无监督动词隐喻检测任务上的表现。再如，郑天龙等（2025）使用 SPV、MIP 和语义用法对比（SUC）三个语言学方法进行多重不一致性特征建模：SPV 用于检测因句法不一致而导致的隐喻发生，MIP 用于建模目标词语义与基本义之间的不一致，SUC 则用于补充处理一词多义情境下传统 SPV 或 MIP 难以覆盖的隐喻发生机制。其实验结果显示，融合这些语言学不一致性特征后，模型在中文隐喻检测任务上优于基线模型，该方法体现出明确有效性，并具备广泛适用于中英文隐喻检测任务的潜力。

除前面所提到的融合 CMA 的语篇隐喻识别研究外，还有不少学者进行了各具特色的探索，向雪等（2026）提出一种基于知识增强的多模态隐喻检测方法（KAMMD），并将检测材料扩展到社交媒体表情包。Fuoli 等（2025）进一步突破词句级层面的隐喻检测范式，比较了检索增强（RAG）、提示词工程和微调三种方法在完整文本隐喻识别和标注中的表现。Han 等（2026）基于荷兰语癌症患者访谈和论坛数据，探索 LLM 在医疗场景中抽取患者隐喻的能力，将 LLM 隐喻抽取用于真实患者叙事和健康沟通研究，同时系统分析幻觉、习语误判等错误类型（Han et al., 2026）。

这些研究共同显示，大语言模型隐喻识别 / 检测研究正在从“直接判断是否为隐喻”转向更具理论依据和方法适配性的路径：一方面通过语言学理论的整合提升模型判断的可解释性；另一方面通过知识增强、检索增强、微调和人机协作等技术方法缓解模型幻觉、语境不足和文化知识缺失等问题。



（2）隐喻解释

部分关于大语言模型隐喻处理能力的研究进一步考察模型能否说明隐喻意义、重构源域—目标域关系、生成合理解释，并在特定语篇或专业知识场景中完成跨概念系统的解释转换。如，Puraivan 等（2024）将隐喻识别与隐喻解释结合起来，考察 ChatGPT-4o 和 GPT-4 Turbo 对西班牙语多义动词语料中隐喻的分析能力。张建丽和王宇泽（2025）还将 ChatGPT 的隐喻能力区分为隐喻理解能力与隐喻阐释能力，要求模型对汉语隐喻句进行理解难度评定，并针对给定的汉语隐喻句生成尽可能多的释义。

在文学隐喻研究中，Ichien 等（2024）对 GPT-4 解释新奇文学隐喻的能力进行测试评估。该研究的特色在于将 LLM 隐喻解释能力置于文学创造性和语用理解框架下进行检验，研究指出 GPT-4 的解释虽然准确，却高度公式化，难以让人感受到人类文学阅读的情感与审美表达（Ichien et al., 2024）。在专业领域中，Tang 等（2025）研究 LLM 如何解释中医隐喻，并将其映射到现代西医病理机制，该研究创新性地从普通语言理解推进到医学领域。

一方面，这些研究表明大语言模型已经具备一定的隐喻解释能力，能够生成较为合理的释义和揭示源域—目标域映射，并在文学、医学等场景中发挥辅助解释作用。这类研究的创新趋势主要体现在：①从基础识别转向解释生成和结构化映射；②从普通语句扩展到文学诗歌、专业医学和真实语料库；③借助 CoT、动态提示和人工评审等机制提高解释的透明性、可验证性和理论关联度。另一方面，这些研究也提示 LLM 的隐喻阐释仍存在情感审美不足、过度解释、语义边界误判和专业知识不足等问题，因此其更适合作为人类研究者的解释辅助工具，而非完全独立的隐喻阐释主体。

（3）隐喻生成

相较于识别和解释，隐喻生成更能揭示模型的深层概念组织能力，然而目前直接围绕大语言模型“隐喻生成能力”展开的研究相对较少。Chang 等（2026）通过隐喻生成任务考察 LLM 的文化推理能力，该研究以涵盖五种文化背景和若干抽象概念的隐喻生成任务为例，通过测试和结果分析发现 LLM 在某些特定场景中表现出刻板的隐喻使用方式，并存在默认西方主义倾向。

（4）大语言模型的工具调用能力

有一类隐喻研究体现了人类研究者对人工智能在科研领域发挥辅助作用的乐观态度，这类研究将大语言模型从被评测对象转变为多功能研究工具，应用于隐喻质量评估、数据生成与增强、翻译辅助以及大规模语篇分析等。隐喻质量的自动化评估是当前较多学者关

注的应用方向, DiStefano 等(2025)将 LLM 作为隐喻创造力自动评分工具,使模型学习人类对隐喻原创性的评分方式,他们的研究显示 LLM 能够有效应用于可复现、低成本的隐喻创造性评分。黄孝喜等(2026)也把 LLM 从一般文本评分工具转化为具有认知语言学理论约束的隐喻评估工具。

LLM 还可用于数据生成与增强,盛坤和王中卿(2025)为解决中文通感语料中感觉词分布不均、稀疏感觉词样本不足的问题,以真实训练数据中的稀疏感觉词样本作为提示,利用 ChatGPT-3.5 生成额外的合成样本。在翻译辅助应用上, Wang 等(2025)的研究聚焦文学作品中含隐喻或明喻的句子的机器翻译,创新性地使用 LLM 为构建“会思考的机器翻译”提供训练数据,使模型学习文学隐喻翻译中的反思、修订和语义保留过程。另外, Mendelsohn 和 Budak(2025)将 LLM 应用于 40 万条美国移民主题推文,分析隐喻使用与政治意识形态、用户互动之间的关系,展示了 LLM 在大规模语篇分析上的高效优势。

整体而言,将 LLM 的自动评分、生成、推理和建模能力嵌入传统隐喻研究流程,能够起到降低人工成本、扩大数据规模并提高分析可解释性的作用。

3.2 隐喻处理能力评估总结

(1) 大语言模型在隐喻处理任务上的积极表现

现有研究总体表明, LLM 在隐喻识别、解释、评价与生成等不同层级任务中已展现出较为突出的处理潜力。首先, LLM 不但能进行浅层分类,而且能够在理论提示、知识补充和推理步骤引导的支持下,对隐喻性表达进行较为系统的识别和定位。其次, LLM 还表现出较强的自然语言解释能力和概念映射能力,例如 ChatGPT 能够为汉语隐喻提供数量较多且具有一定合理性的阐释,并能从社会属性、心理属性和认知属性等角度展开意义说明, GPT-4 在文学隐喻解释的任务表现中甚至被盲评者评为优于大学生解释(张建丽 & 王宇泽, 2025; Ichien et al., 2024)。另外, LLM 显示出替代或辅助人工评分的潜力,部分研究发现 LLM 在特定隐喻评分任务上与人工评分有较高的一致性,在给定的评价维度上,能够形成较为稳定的质量判断(DiStefano et al., 2025; Mangiaterra et al., 2025; 黄孝喜等, 2026)。LLM 在隐喻研究中发挥的自动评分、数据生成等工具性能力能够有效降低人工评估成本,进而帮助研究者扩大其研究规模。最后,在隐喻生成方面,虽然相关研究较少,但已有成果显示 LLM 能够在理论提示或文化提示下生成一定质量的隐喻性表达。

(2) 大语言模型在隐喻处理任务上的问题与局限

大语言模型虽然在若干隐喻任务中能够取得较高水平的表现,但其能力并不等同于稳定、深入的隐喻理解。Tong 等(2024)为检验 LLM 是否能够始终选择正确的目标域释义



（从而完全诠释隐喻），还是基于词汇相似性选择与隐喻词语表面意义相似但不恰当的释义，专门构建了一个同时包含恰当和不恰当隐喻样例的数据集对 LLM 进行测试。该研究在 LLaMA 和 GPT-3.5 的测试结果表明，模型容易混淆源域和目标域，说明其未必能够实现完整的跨域推理。类似的，Sanchez-Bayona 和 Agerri（2025）对 LLM 在隐喻解释中推理能力的评估研究得出了批判性结论，该研究指出受试模型（Llama-3-8B-Instruct 等）的表现受到词汇重叠和句子长度等表层特征的影响大于受到隐喻内容的影响，也就是说，所谓 LLM “涌现”的隐喻理解能力，有可能是一种表层特征、上下文学习和一般语言知识的综合作用结果，而非对隐喻跨域映射的深层掌握。Nadler 等（2025）在探索基于统计学习的 LLM 和具备具身经验的人类的认知差异时发现，统计学习能够支持 LLM 形成部分稳定的颜色意义关联，但在需要感知经验、审美判断和新颖概念联结的颜色隐喻任务中，LLM 与人类仍存在表现差距，尤其在反向颜色联想和新颖颜色隐喻解释任务中，LLM 经常无法给出连贯解释。

另外，在隐喻研究中调用 LLM 的工具能力时，LLM 的表现仍存在不可忽视的局限，例如 LLM 在生成隐喻时可能存在概念刻板化和文化默认倾向的问题，文化身份提示并不能保证模型真正进行文化内部的概念推理，模型有可能只是用地方化符号重新包装主导性概念框架（Chang 等，2026）。Mangiaterra 等（2025）也指出，尽管 LLM 在隐喻属性（熟悉度、可理解性及形象化程度）方面能够有效且可靠地替代或补充人类的评估工作，但在处理隐喻意义的约定俗成属性和多模态特征时，其表现逊色于人类。因此，LLM 作为隐喻研究工具应用时，仍需有理论约束、人工校验和误差控制，尤其在文化语境、感觉经验和高度主观的隐喻评价任务中不能完全替代人类专家。

四、当前不足与未来展望

透过现有研究，我们可以清晰地看到，目前理论应用与模型能力这两个维度上均存在不足。第一，从理论整合应用来看，大语言模型与语言学理论整合的广度和深度仍然欠缺，目前引入大语言模型隐喻处理能力研究的语言学理论主要集中在概念隐喻理论和批评隐喻分析等，对其它语言学理论资源的整合探索仍较为缺少，且虽然部分现有研究将概念隐喻理论等与大语言模型研究进行了整合，但整合方式主要是采用语言学理论作为提示词设计、任务说明或评价维度的外部依据，对于这些核心概念如何被稳定地计算化和操作化仍缺乏深入讨论与测试。另外，基于语言学理论的 LLM 深层隐喻理解能力评估体系也有待进一步



构建与完善。未来研究可探索语言学理论与计算方法更具深度和广度的整合方式，一方面将源域、目标域、跨域映射等理论概念转化为更清晰的标注规范和任务设计，另一方面，在模型输出分析、错误类型归纳和表现评估中引入更严谨的理论框架，构建在语言学意义上更有说服力的评估体系。

第二，从模型能力来看，现有研究总体表明目前大语言模型在涉及隐喻深层语义理解和推理上仍有局限，尤其在高阶任务上的表现有较大的提升空间，隐喻处理能力作为语言学专业能力的一种，本质上也反映了当前大语言模型在语言学专业能力上的不足。因此，未来研究可进一步拓展对大语言模型语言学专业能力的系统性评估，并结合计算机方法和语言学理论探索更为有效的增强路径。

五、结语

大语言模型与隐喻研究的交叉是一个兼具理论价值与方法潜力的新兴方向，一方面，隐喻处理的表现能力可作为检验大语言模型的深层语义理解和推理能力的重要窗口，另一方面，大语言模型为隐喻本身的研究也提供了新的技术路径，推动隐喻研究在复杂语境、多模态材料和更大规模语料的条件下开展。现有研究表明，概念隐喻理论、批评隐喻分析等语言学理论，已开始进入提示词工程、数据增强、任务设计和评估框架之中，并在一定程度上增强了模型表现与结果的可解释性。然而，大语言模型与语言学理论的整合仍缺乏广度和深度，包含隐喻处理能力在内的 LLM 语言学专业能力也有待更全面的评估与加强。未来研究应在语言学理论、计算方法、训练数据和评估框架之间建立更深更广泛的融合，推动大语言模型隐喻处理能力研究从任务表现描述走向机制解释，为人工智能语义理解相关研究提供更充分的理论与方法支持，并进一步探索评估及增强大语言模型语言学专业能力的方法。

（使用 AI 声明：本文在撰写过程中使用了人工智能作为辅助研究工具，但文中所有的论述框架、解读、分析与结论均由作者本人完成及确认。）



【参考文献】

- 蔡雨晨, 王馥芳 (2024). GPT-4 认知语言学专业能力评估——以概念隐喻理论为例 [Evaluation of GPT-4's Subject-specific Competence of Cognitive Linguistics: A Case Study on Conceptual Metaphor Theory]. 语言与智能 [Language and Intelligence], (01), 71–86.
- 黄孝喜, 查正超 & 陆诗佳 (2026). 基于大语言模型的中文隐喻多维度评估 [Multi-dimensional evaluation of Chinese metaphors based on large language models]. 《浙江大学学报 (工学版)》 [Journal of Zhejiang University (Engineering Science)], 60(2), 388–395. DOI: 10.3785/j.issn.1008-973X.2026.02.017.
- 盛坤, 王中卿 (2025). 基于大语言模型和数据增强的通感隐喻分析 [Synaesthesia Metaphor Analysis Based on Large Language Model and Data Augmentation]. 《计算机应用》 [Journal of Computer Applications], 45(3), 794–800. DOI: 10.11772/j.issn.1001-9081.2024091251.
- 向雪, 王中卿 & 王红玲 (2026). 知识增强的多模态隐喻检测方法 [Knowledge-enhanced Multimodal Metaphor Detection Method]. 《小型微型计算机系统》 [Journal of Chinese Computer Systems], 1-12. DOI: 10.20009/j.cnki.21-1106/TP.2025-0398.
- 杨启萌, 魏齐兴 & 孟浩. (2026). 计算隐喻处理: 从隐喻识别、隐喻解释到隐喻生成 [Computational Metaphor Processing: From Metaphor Identification and Interpretation to Metaphor Generation]. 计算机工程与应用 [Computer Engineering and Applications], 62(09), 46-60. DOI: 10.3778/j.issn.1002-8331.2506-0312
- 于艳春 (2025). 大语言模型对批评隐喻分析中隐喻识别的应用性研究 [A Study on the Applicability of Large Language Model (ChatGPT) to Metaphor Identification in Critical Metaphor Analysis]. 《外语学刊》 [Foreign Language Research], 2025(1), 47–56. DOI: 10.16263/j.cnki.23-1071/h.2025.01.006.
- 张建丽, 王宇泽 (2025). 生成式人工智能机器人隐喻能力的实证研究 [Generative Artificial-Intelligence-Robot's Metaphorical Competence]. 《外国语文》 [Foreign Languages and Literature], 41(3), 41–51. DOI: 10.26922/j.cnki.waiguoyuwen.2025.03.007.
- 郑天龙、董瑞、杨雅婷、马博、王磊 & 周喜 (2025). 基于语言学多重不一致性的隐喻检测模型 [Metaphor detection model based on linguistic multi-incongruity]. 《计算机应用》 [Journal of Computer Applications], 45(12), 3829–3838. DOI: 10.11772/j.issn.1001-9081.2024121797.
- Boisson, J., Siddique, Z., Borkakoty, H., Antypas, D., Espinosa Anke, L., & Camacho-Collados, J. (2024). Automatic Extraction of Metaphoric Analogies from Literary Texts: Task formulation, Dataset Construction, and Evaluation. arXiv preprint arXiv:2412.15375.
- Charteris_Black, J. (2004) *Corpus Approaches to Critical Metaphor Analysis*. M. Hampshire: Palgrave Macmillan.
- Chang, Y., Qu, J., & Li, Z. (2026). *Metaphors We Compute By: A computational Audit of Cultural Translation*



- vs. *Thinking in LLMs*. arXiv preprint arXiv:2604.04732.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax [M]*. Cambridge, MA.: The MIT Press.
- DiStefano, P. V., Patterson, J. D., & Beaty, R. E. (2025). Automatic Scoring of Metaphor Creativity with Large Language Models. *Creativity Research Journal*, 37(4), 555–569. <https://doi.org/10.1080/10400419.2024.2326343>.
- Fuoli, M., Huang, W., Littlemore, J., Turner, S., & Wilding, E. (2025). *Metaphor Identification Using Large Language Models: A Comparison of RAG, Prompt Engineering, and Fine-tuning*. arXiv preprint arXiv:2509.24866.
- Han, L., Lindevelt, D., Puts, S., van Mulligen, E., & Verberne, S. (2026). *Dutch Metaphor Extraction from Cancer Patients' Interviews and Forum Data using LLMs and Human in the Loop*. arXiv preprint arXiv:2511.06427.
- Hicke, R. M. M., & Kristensen-McLachlan, R. D. (2024). *Science is Exploration: Computational Frontiers for Conceptual Metaphor Theory*. arXiv preprint arXiv:2410.08991.
- Ichien, N., Stamenković, D., & Holyoak, K. J. (2024). Large Language Model Displays Emergent Ability to Interpret Novel Literary Metaphors. *Metaphor and Symbol*, 39(4), 296–309. <https://doi.org/10.1080/10926488.2024.2380348>.
- Kramer, O. (2025). *Conceptual Metaphor Theory as a Prompting Paradigm for Large Language Models*. arXiv preprint arXiv:2502.01901.
- Lakoff, G., & Johnson, M. (1980). *Metaphors We live by*. University of Chicago Press.
- Lin, Y., Liu, J., Gao, Y., Wang, A., & Su, J. (2024). *A Dual-perspective Metaphor Detection Framework Using Large Language Models*. arXiv preprint arXiv:2412.17332.
- Mangiaterra, V., Al-Azary, H., Barattieri di San Pietro, C., Canal, P., & Bambini, V. (2025). *Can GPT Replace Human Raters? Validity and Reliability of machine-generated Norms for Metaphors*. arXiv preprint arXiv:2512.12444.
- Meng, H., Li, X., & Sun, J. (2025). Large Language Models Prompt Engineering as a Method for Embodied Cognitive Linguistic Representation: A Case Study of Political Metaphors in Trump's discourse. *Frontiers in Psychology*, 16:1591408. <https://doi.org/10.3389/fpsyg.2025.1591408>.
- Mendelsohn, J., & Budak, C. (2025). *When People are Floods: Analyzing Dehumanizing Metaphors in Immigration Discourse with Large Language Models*. arXiv preprint arXiv:2502.13246.
- Puraivan, E., Renau, I., & Riquelme, N. (2024). Metaphor identification and interpretation in corpora with ChatGPT. *SN Computer Science*, 5:976. <https://doi.org/10.1007/s42979-024-03331-0>
- Sanchez-Bayona, E., & Agerri, R. (2025). *Metaphor and Large Language Models: When Surface Features Matter More than Deep Understanding*. arXiv preprint arXiv:2507.15357.



- Tang, J., Wu, N., Gao, F., Dai, C., Zhao, M., & Zhao, X. (2025). *From Metaphor to Mechanism: How LLMs Decode Traditional Chinese Medicine Symbolic Language for Modern Clinical Relevance*. arXiv preprint arXiv:2503.02760.
- Tong, X., Choenni, R., Lewis, M., & Shutova, E. (2024). Metaphor Understanding Challenge Dataset for LLMs. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 3517–3536). Association for Computational Linguistics.
- Wang, J., Meng, F., Liang, Y., & Zhou, J. (2025). *DRT: Deep Reasoning Translation via Long Chain-of-Thought*. arXiv preprint arXiv:2412.17498
- Yang, C., Chen, P., & Huang, Q. (2024). Can ChatGPT's Performance Be Improved on Verb Metaphors Detection Tasks? Bootstrapping and Combining Tacit Knowledge. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1016–1027). Association for Computational Linguistics.
- Zhang, D., Lu, X., Zhuang, M., Yang, S., & Chen, H. (2025). Multimodal Metaphor Recognition Based on Chain-of-Cognition Prompting. *Cognitive Systems Research*, 91:101356. <https://doi.org/10.1016/j.cogsys.2025.101356>