

认知语言学赋能文生图提示词优化： 以国际中文教育为例^{*}

郭梦溪¹，崔晓璐²，赵怡静¹

(1. 北京语言大学 信息科学学院；2. 北京语言大学 国际中文学院，北京 100083)

【摘要】研究聚焦于国际中文教育领域文生图技术中提示词的优化问题，构建了一种以认知语言学为指导的方法框架。研究以框架语义学、概念隐喻及意象图式为理论基础，借助向量空间模型测量抽象动词的语义与框架的距离，并将其划分为 3 个认知层级，进而提出域替换、隐喻混合与重构等差异化的文生图提示词策略。在此基础上，设计了包含国际中文教育场景、动作与文化 3 个维度的标准化提示词体系。对照实验显示，认知强化的文生图提示词能够有效增强图像生成的语义准确性及其在教学情境中的抗干扰能力。研究为多模态教学资源生成提供了可操作的新路径，同时拓展了 AIGC 技术在国际中文教育领域的应用潜力。

【关键词】认知语言学；多模态大模型；国际中文教育；图像生成；AIGC

【中图分类号】H195 **【文献标志码】**A **【文章编号】**1674-5639(2026)05-0032-10

DOI: 10.14091/j.cnki.kmxyxb.2026.05.004

一、引言

生成式人工智能的飞速发展，促使文生图技术不断完善，也为国际中文教育的资源开发开辟了新路径。多模态大模型为教学资源搭建提供新思路，但在实际应用中，通用 AIGC 模型与具体教学场景的需求仍存在明显脱节。文生图创作时常面临抽象概念难以呈现、文化语境复杂等问题，核心原因是提示词工程缺少认知层面的理论支撑。当下的文生图提示词多依靠经验总结，缺乏认知语言学的理论指引，使生成图像不符合中文学习者的认知特点。

本研究意在探究认知语言学视角下隐喻理论和相关规律对文生图提示的提升作用，以期在理论上阐明隐喻相关理论不仅是视觉分析的工具，更可以有效提升视觉产品（图像）生成性能，从而深化对人机协同知识生产中认知机制的理解。在实践上，我们则针对国际中文教育中的抽象动词教学，构建了一套从认知理论到技术落地的技术流程。通过对语义与框架距离进行量化，将认知理论融入提示词构建，突破了传统提示词工程仅依赖人工经验的局限，显著提升了 AIGC 资源在复杂教学语境下的语义精准度与文化适配性，为多模态教学资源的智能化生成提供了可计算、可迁移的路径。

二、认知语言学理论对文生图的启示

(一) 相关认知语言学研究

认知语言学认为，语言是认知活动的一部分。语言的产生与理解围绕人类的基础认知能力和心理活动展开。^①人类在认识世界、梳理各类事物时，会依赖多种基础心理模式。其中“框架”是重要的认知方

^{*} **【作者简介】**郭梦溪，女，河南洛阳人，北京语言大学在读博士研究生，研究方向为计算语言学、国际中文智慧教育；崔晓璐，女，山东潍坊人，北京语言大学在读硕士研究生，研究方向为语言资源建设与应用、国际中文智慧教育；赵怡静，女，河南许昌人，北京语言大学在读硕士研究生，研究方向为语言资源建设与应用、国际中文智慧教育。

【基金项目】北京语言大学中央高校基本科研业务费专项资金（18WT03）。

^① Langacker R W. Foundations of Cognitive Grammar: Volume I: Theoretical prerequisites [M]. Stanford: Stanford University Press, 1987; Talmy L. Toward a cognitive semantics: Volume 1: Concept structuring systems [M]. Cambridge, MA: MIT Press, 2000.

式, 包含主体、环境、动作等关键内容, 影响人们认知、解读具体场景和事件。^① 隐喻是人类理解抽象概念的核心方式, 我们往往会凭借具象的经验, 完成对抽象对象映射转化。^② 这一认知规律符合文生图的技术逻辑。例如, “满月寓意团圆”等来源于中文思维的表达, 可以将抽象文化内涵转化为视觉模型可识别的具象的指令, 避免中秋节场景中的文化表达偏差的问题。除框架与隐喻外, 意象图式是另一个核心认知要素。意象图式来自身体与外界的长期、反复互动, 进而形成相对稳定的感官运动认知模式, 带有身体经验属性。^③ 语言与图像的内在联系也建立在这种认知基础上。人类处理视觉与语言信息的过程, 是多重认知表征相互作用、整合加工的结果。^④ 语言信息可激活意象图式与感知模拟机制,^⑤ 引导视觉认知。例如, 人们理解“攀登”时, 会激活其语义和向上空间图式, 这正是文生图语义转图像的认知依据。

现有研究仍主要关注隐喻对词义扩展的主导作用, 杨维、王飞通过“hot”的义项分析揭示隐喻的语义延伸过程; 吴淑琼等推动了语料库技术与认知理论的结合。^⑥ 原型、范畴化等理论也被广泛应用于外语教学, 文旭、徐锦芬等学者通过构建词汇网络、优化语法教学模式, 为外语教学提供了实践支撑。^⑦ 认知语言学与视觉研究的结合集中在多模态隐喻的分析上。研究者大量运用 Lakoff、Johnson 的概念隐喻理论与 Forceville 的多模态隐喻理论, 对广告、电影海报等图文并存的语篇进行实证分析。^⑧ 相关研究共同点在于, 将认知语言学作为一种分析工具, 揭示从源域到目标域的映射机制如何在视觉和语言的交叉中构建意义。^⑨ 由此可见, 概念隐喻和框架语义学已经成为理解视觉叙事和意义构建的重要理论。

国际中文教育场景中, 学习者对动词的理解高度依赖认知结构。我们结合教学实践经验(图示法、互动模拟法等)和相关理论, 认为可以将认知语言视角对复杂语义和语用场景的理解原理与文生图的提示词工程技术进行融合, 将认知结构转化为大模型可执行的提示词指令, 实现认知规律向提示词工程的落地。框架语义学通过激活主体、动作、环境、关系等核心要素, 搭建语义与视觉的结构化映射, 解决图像框架要素缺失问题; 概念隐喻理论明确源域与目标域的映射规则, 为抽象动词视觉化提供符合人类认知的转化路径, 融入中文文化经验, 减少隐喻偏差与文化错位, 强化学习者感知模拟; 意象图式以具身经验为底层逻辑, 约束图像呈现符合物理常理与认知习惯, 避免违背认知规律的生成效果。三者形成“结构搭建、跨域转化、逻辑约束”的认知闭环, 将语言理解规律转化为视觉生成准则。这一机制既优化了文生图技术, 也阐明在国际中文教育中语言激活意象图式, 推动多模态整合, 进而提升抽象词汇的教学效率。

(二) 文生图工作的认知视角

现有研究证实, 文生图可高效生成个性化、场景化视觉教学资源。Zhu 等、何晴等、朱奕瑾等的相关探索, 也为技术落地提供了实践经验。^⑩ 当前相关提示词优化主要有两条路径: 一是经验驱动的人工模板

① Fillmore C J. Frame semantics [A]. *Linguistics in the Morning Calm* [C]. Seoul: Hanshin Publishing Company, 1982: 111-137.

② Lakoff G, Johnson M. *Metaphors We Live By* [M]. Chicago: University of Chicago Press, 1980.

③ Johnson M. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason* [M]. Chicago: University of Chicago Press, 1987.

④ Paivio A. *Mental representations: A dual coding approach* [M]. New York: Oxford University Press, 1986.

⑤ Barsalou L W. Perceptual symbol systems [J]. *Behavioral and Brain Sciences*, 1999, (4): 577-660.

⑥ 杨维, 王飞. 认知语言学视角下“hot”的义项分析 [J]. *现代语言学*, 2024, (4): 307-312; 吴淑琼, 江艳. 认知语言学量化转向的新发展: 认知语料库语言学 [J]. *外国语文研究* (辑刊), 2024, (1): 76-90.

⑦ 文旭, 卢晗. 文化认知与外译传播路径探析 [J]. *外语与翻译*, 2025, (3): 1-9+98; 徐锦芬. 认知语言学在外语教育学中的融合运用: 现状与展望 [J]. *解放军外国语学院学报*, 2023, (6): 79-86.

⑧ 马婧妍. 对比分析中俄公益广告中多模态隐喻的民族文化特点 [J]. *海外文摘*, 2020, (19): 9-11.

⑨ 张定琦. 多模态隐喻在上海公益广告专列中的认知构建 [J]. *现代语言学*, 2023, (11): 5388-5395.

⑩ Zhu Y J, Wang Y, Gao L M, et al. The construction and application of a picture resource system for the international Chinese language education field [A]. *2024 International Conference on Asian Language Processing (IALP)* [C]. IEEE, 2024: 359-365; 何晴, 饶高琦. 国际中文教育生成式图片库的构建与应用 [J]. *国际中文教育* (中英文), 2025, (4): 59-73; 朱奕瑾, 王英艳, 王雨, 等. 基于 AIGC 的中文语法教学视频资源生成式构建 [J]. *语料库语言学*, 2025, (1): 17-33.

设计,依赖设计者经验,缺乏统一理论与可迁移性;二是数据模型自动优化,生成效率高,但多以图像视觉效果为目标,忽视语言学习者的认知规律与教学适配性。王东清等、王常圣虽提出了通用提示词优化方法,^①但多面向艺术创作与通用生成场景。国际中文教育领域内,针对抽象词汇教学的认知强化型提示词研究仍属空白,未能结合二语学习者的认知特点开展设计。

AIGC 技术快速发展,使文生图研究从技术实现转向与认知机制深度融合的新范式。Chakrabarty 等^②探索了大模型与扩散模型协同生成视觉隐喻的路径;Su 等^③则基于隐喻理解,研究出视觉隐喻图像高效生成方法。相关识别检测研究为生成策略优化提供了实践经验:He 等^④指出机器识别与生成需精细化分解视觉元素;姚加威^⑤构建了概念隐喻驱动的生成框架及知识增强检测模型;Zhang 等^⑥也证实了多模态识别中认知链推理的有效性。这些研究表明,引入认知理论是提升文生图质量的一个关键因素,认知语言学理论也可以转化为内容生成的工程指导。但目前仍缺乏语言教学方面的研究实践。

这类研究相对不足的原因,在于语言教育领域中知识生产时会面临的认知隔阂,而这种认知隔阂源于生成式人工智能知识的局限性和人类知识的低具象性。陈港指出,人机协同知识生产中,生成式 AI 仅为“弱认知主体”,其认知由人主导,效果高度依赖提示词质量。^⑦大模型的跨模态对齐依靠海量语料,而非对现实世界的具身感知;其知识是数据驱动的概率分布,缺少物理与因果层面的理解。例如大模型对“苹果”等具体名词可视化效果较好,但面对“通过”等抽象动词时,无法依托身体经验形成认知,难以理解从空间穿越到目标达成的隐喻映射。人类对抽象动词的理解根植于情境化、文化性的隐性知识,而大模型仅能进行显性符号组合,缺少具身反馈,图像易出现语义泛化。因此,有必要将认知语言学框架融入提示词优化,使生成图像符合人类认知逻辑,降低认知违和感。

不同词汇的视觉生成难度差异显著,其认知具象程度直接影响视觉意象的激活效果。“辞职”等高具象词语,具有清晰的动作原型与场景特征,隐性认知易外化,视觉呈现难度低;而“把关”等抽象低具象词汇,因缺乏固定物理原型,呈现出千人千面的认知发散特征,认知负荷显著提升。对于低具象性的隐喻词,借助认知语言学可显化隐喻词的隐性特征,通过认知约束让发散语义归为具体表征。此外,语言教学场景中存在文化普适的认知原型,通过提取抽象语义核心义素锁定认知共性,再搭配高辨识度教学符号,可将模糊意象外显化,使生成图像既符合物理常理,又能与教学语境中的抽象语义高度契合。

我们基于认知语言学的系统性的认知对齐与语义映射框架,将抽象语言符号通过认知中介转化为具备教学适用性的视觉表征,以弥合由于大模型缺乏具身性而导致的语义理解偏差。

三、文生图提示词工程中的认知问题

(一) 隐喻映射模糊

提示词未建立源域与目标域的明确隐喻关联,或引入不符合中文认知习惯的西式隐喻,导致抽象概念图像化出现歧义,降低词汇习得效率。例如图 1 的 3 个例子中“思乡”未关联月光、归雁等中式具象

① 王东清,芦飞,张炳会,等. 大语言模型中提示词工程综述 [J]. 计算机系统应用, 2025, (1): 1-10; 王常圣. 面向大模型艺术图像生成的提示词工程研究 [J]. 图学学报, 2024, (6): 1243-1255.

② Chakrabarty T, Saakyan A, Winn O, et al. I spy a metaphor: Large language models and diffusion models co-create visual metaphors [A]. ACL 2023 [C]. Toronto, Canada: ACL, 2023: 7370-7388.

③ Su C, Wang X, Liu S, et al. Efficient visual metaphor image generation based on metaphor understanding [J]. Neural Processing Letters, 2024, (3): 150.

④ HE X, YU L, TIAN S, et al. VIEMF: Multimodal metaphor detection via visual information enhancement with multimodal fusion [J]. Information Processing & Management, 2024, (3): 103652.

⑤ 姚加威. 基于大模型的隐喻图像生成与评估研究 [D]. 广西大学硕士学位论文, 2025.

⑥ Zhang D Y, Lu X Y, Zhuang M L, et al. Multimodal metaphor recognition based on chain-of-cognition prompting [J]. Cognitive Systems Research, 2025, (91): 101356.

⑦ 陈港,孙元涛. 知识“可供性”: 人机协同知识生产如何“双智驱动” [J]. 电化教育研究, 2025, (11): 25-32.

意象, 表意模糊; “领跑” 未关联跑道, 生成成功人士形象引发认知偏差; “折腾” 未体现重复动作与痛苦表情, 仅呈现杂乱事物, 难以理解。



图1 隐喻映射模糊图例

(二) 缺失框架要素

提示词仅罗列个别物体、忽视场景整体结构, 导致生成图像信息零散、语义单薄。如图2中“介绍”缺失多人互动场景, “咨询”缺少沟通互动与场景信息, “妥协”未体现从立场对立到缓和的心理过程与情绪, 造成片面理解。只有通过语义框架构建完整场景, 明确空间、状态及实体关系, 才能生成要素全面的图像。



图2 缺失框架图例

(三) 违背认知规律

提示词设计不符合人类认知原理, 如图3举例: 单一图像元素过多超出认知负荷 (“写生”); 主体与背景混淆 (“里”); 核心角色不突出 (“调解”, 易误读为 “谈判”), 影响信息提取与教学有效性。

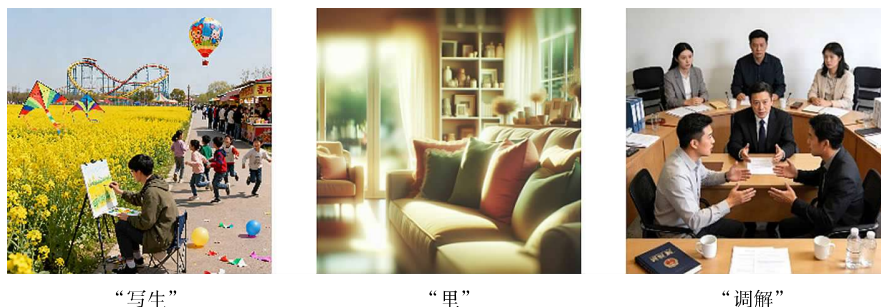


图3 违背认知规律图例

综上, 当前文生图实践在国际中文教育中的认知缺陷集中体现为三点: 一是隐喻映射模糊, 影响词汇情感与语境整合; 二是框架要素缺失, 难以支撑词语准确理解; 三是违背认知规律, 增加学习者解读负担。这些认知层面的问题共同制约了多模态教学资源的构建和使用。

四、认知理论强化的提示词工程

针对国际中文教育中抽象动词可视化的核心难题, 本研究从认知语言学视角构建认知强化的提示词体系。其核心是深度融合认知机制, 通过量化语义与框架距离, 明确抽象动词隐喻转换的认知规律, 设

计差异化策略,构建三层提示词框架,将抽象认知理论与生成策略转化为大模型可精准解读的指令,实现语言认知与视觉生成的系统衔接。本章聚焦认知理论强化提示词设计方案的落地,呈现体系构建逻辑,为后续实验提供理论与工程基础。

(一) 动词语义与框架距离模型

动词抽象化演变本质是从物理源域向抽象目标域的隐喻映射过程。Langacker 提出,该过程的认知难度取决于语义整合效率与认知域固化度^①;施春宏指出,比喻义形成时描述性语义成分基本不变,关涉性语义成分则发生迁移,表现为语义特征的替换与调整^②;王莹认为语义距离与相似度密切相关,并通过定性分析将原义与比喻义的语义距离分为近、较近、远三类^③。因此,本研究构建了语义距离与框架距离两大核心度量指标。

语义距离 (D_{sem}) 可以用来衡量动词原本的物理源域与抽象引申义间的语义差异。我们采用 gte-large-zh 预训练模型分别提取动词物理本义、抽象引申义的高维语义向量,通过余弦距离公式计算得出语义距离数值。数值越高,则源域与目标域的语义距离越大,语义关系越弱,隐喻映射时的认知负荷也就越高。框架距离 (D_{frame}) 的计算有别于传统语义定性分析。由于动词框架特征隐含于上下文搭配规律之中,易混无关变量,因此我们通过构建最小对比句排除无关变量:一组填充具象名词还原动词物理本义如“工人支持了柱子”,另一组采用抽象概念呈现隐喻义如“论据支持了观点”。随后,我们将两类对比句输入 BERT 架构的预训练模型 gte-large-zh,通过计算动词在不同句法环境中隐藏状态向量的余弦距离,判定其框架变化。若数值趋近于 0,说明动词隐喻转换时语义框架变化相对较小;数值越大,则表明语义框架出现重建,如从物理动作向逻辑关系的转变,这会给 AIGC 生成逻辑带来更大挑战。本研究借助预训练语言模型提取向量并计算这两种距离,其核心目的并非构建一个独立的分类算法,而是为了客观衡量动词从物理源域向抽象目标域映射的认知难度,从而为后续制定差异化的提示词策略提供科学的分类依据。

本研究借鉴 Xu、Yang^④ 的词语相似度算法,经人工核验,从 HSK 大纲词表中选取了 145 个覆盖不同语义维度的抽象动词作为样本,侧重于抽象动词隐喻类型的多样性。我们对核心抽象动词样本进行全量计算,形成二维散点分布图(图 4)。根据样本在二维坐标系中的分布情况,采用统计学中的 33% 和 66% 分位数作为动态阈值,根据动词从物理源域向抽象目标域迁移过程中的特征,构建了由易到难的 3 个阶段,将二维向量空间划分为 3 个认知区间。而后结合语义距离和框架距离的判定规则,所有动词划分为 3 个等分布的难度区间。这种方法确保了阈值是基于语料数据的客观分布,而非人工主观设定。分类情况如下:

近距离动词:语义距离小于 33% 分位数阈值 ($D_{sem} < 0.3255$),且框架距离也小于 33% 分位数阈值 ($D_{frame} < 0.4595$)。这些词汇位于坐标系左下角,语义与框架距离都较低,核心语义特征在迁移过程中得到了高度保留,容易转换,视觉认知难度最低。“提高、回归、晋升”都属于这类。

远距离动词:语义距离大于等于 66% 分位数阈值 ($D_{sem} \geq 0.4307$) 或框架距离大于等于 66% 分位数阈值 ($D_{frame} \geq 0.5568$),位于坐标系外围包括右侧或上方,动词的物理本义与抽象义项已发生本质性剥离,原有的视觉图式已成为理解干扰项。代表词汇例如“忽悠、挂钩”。

中距离动词:不属于前两类的中间区域都可以归为这一类。这些词汇位于隐喻转换的认知过渡阶段,原有的物理特征不足以覆盖引申出的抽象含义典型动词,源域与目标域呈现半重叠状态。例如“起飞、领跑”。

对比这三类动词的视觉认知难度和 HSK 词汇等级分布,发现二者并不匹配。HSK 大纲中的部分高级

① Langacker R W. Foundations of Cognitive Grammar: Volume I: Theoretical prerequisites [M]. Stanford: Stanford University Press, 1987.

② 施春宏. 比喻义的生成基础及理解策略 [J]. 语文研究, 2003, (4): 19-24.

③ 王莹. 《现代汉语词典》(第 7 版) 动词隐喻义研究 [D]. 北京语言大学硕士学位论文, 2020.

④ Xu H, Yang L. Construction and application of a large-scale Chinese abstractness lexicon based on word similarity [A]. Lu W, Huang S Y, Hong Y X, et al., NLPCC 2022 [C]. Cham: Springer, 2022: 122-130.

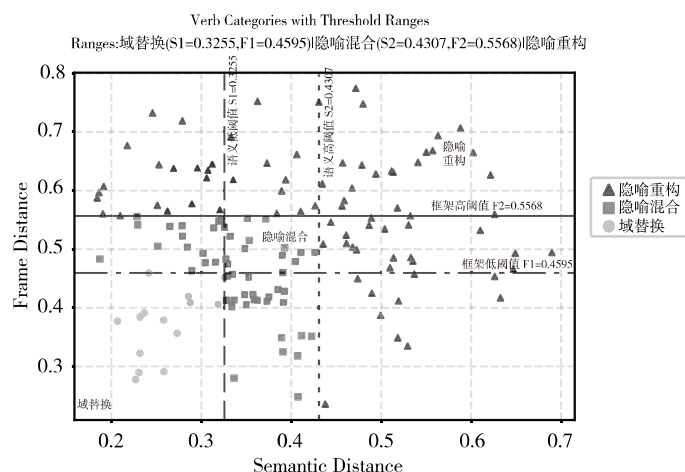


图4 核心抽象动词在语义、框架距离上的分布与策略分类

词, 经计算分类后被归为近距离动词, 认知负荷较低; 而部分需隐喻重构的高难度动词, 却出现在1至2级中。因此词汇的 HSK 等级高低, 不能直接作为判断其视觉化呈现难易的依据。

(二) 基于隐喻的生成策略

按照认知适配的思路, 我们针对三类不同认知难度层级的抽象动词, 设计了不同的隐喻转换方式和视觉呈现逻辑。

1. 域替换策略

根据 Zipf^① 提出的“认知省力原则”, 物理场景的动作逻辑可以运用到抽象语境中。保持动作的基本框架不变, 仅将源域中的核心元素替换为目标域的相关内容, 并可以进行明确的标注。以“进入”的抽象义项“进入新阶段”为例, 提示词无须重构非常复杂的抽象概念, 只需要把物理场景换成“学习者走进标有汉语中高级班的教室”即可。这种处理方式将抽象的目标域特征显性化, 通过人类共通的具身经验约束模型, 降低大模型生成与学习者理解的认知负荷。实践方案如下:

(1) 提取源域具身动作: 通过框架语义学和概念隐喻分析, 找出动词物理本义对应的核心具身动作, 如“隐瞒”的核心具身动作是遮挡或者藏匿。

(2) 筛选目标域相关内容: 结合国际中文教学场景, 选用教材、课堂中的常见教学元素, 筛选与抽象语义直接对应的、无文化壁垒的具象符号。如“隐瞒”的抽象语义是隐瞒信息, 对应物理符号为“不及格试卷”“错误答案”。

(3) 构建替换规则: 保持动作主体、动作逻辑不变, 仅替换动作对象这一核心元素, 形成具身动作框架与相关物理符号相加的生成逻辑。

2. 隐喻混合策略

对这类词汇进行可视化时, 需要提取源域的动作特点和目标域的抽象语义特点, 通过在同一画面中强制使用属于不同域的元素制造认知冲突, 构建融合性的视觉隐喻。以“领跑”为例, 提示词设计为“学习者在标有 HSK 等级的跑道上领先”。这里既保留了跑步的物理动作, 又通过 HSK 等级的跑道标识等显性的视觉符号, 直接建立动作领先与学习领先间的映射关系。实践方案如下:

(1) 分析源域和目标域: 针对源域, 提取可视化的具身动作, 如“假装”对应源域表演的佩戴面具动作; 对于目标域需要提炼抽象语义的核心逻辑, 如“假装”表面行为与真实状态对立。

(2) 视觉融合方案: 确定源域和目标域中元素的视觉融合方式, 构建视觉矛盾体, 帮助大模型与学习者理解表象和本质的复杂认知逻辑。

^① Zipf G K. Human behavior and the principle of least effort: An introduction to human ecology [M]. Cambridge, MA: Addison-Wesley, 1949.

(3) 筛选文化适配符号: 源域符号需选取人类共通的具身经验符号, 如“面具”是跨文化通用的伪装象征, 避免文化误解。

3. 隐喻重构策略

针对远距离动词, 需要采用隐喻重构策略。例如“通过”有达到标准的抽象义项, 物理上的穿过空间难以直接传达如“考试达成目标”这类语义。通过认知分析剥离出动词的隐含义, 寻找具备相同义素且大模型具身感知力更强的新喻体, 重新构建认知场景。以“把关”为例, 若直接描绘关口, 无法体现抽象含义, 因此需要重构动作逻辑和对象, 提示词的设计可以是“教师手持红笔审核标有语法正确的文件”, 背景设定为办公室。通过严格审核这一行为的视觉化, 精准传递了“把关”的核心隐喻义, 将无法直接可视化的逻辑命题, 转为大模型可理解的物理行为。实践方案如下:

(1) 提炼核心义素: 通过认知语言学分析, 分析源域物理义, 并且找出抽象语义的核心义素, 如“把关”的核心义素是“严格审核、筛选合格”。

(2) 筛选新喻体: 选取符合人类共通具身经验、视觉化效果强的新喻体, 且新喻体与核心义素的映射关系清晰, 如“审核文件”对应“把关”的严格筛选义素。

(3) 重构教学场景: 围绕新喻体构建适配国际中文教育的场景, 如教师审核作业、学习者获得证书等, 确保场景与教学目标一致。

(三) 基于隐喻生成策略的提示词工程

认知强化提示词的实质, 是帮助大模型借助认知信息将词汇知识转化为模拟意象。大模型本身缺乏基于现实世界的具身感知和因果理解, 而我们通过融入框架语义、概念隐喻和意象图式, 实质上是在为大模型补充上下文知识空间资源。这种认知维度的介入, 使我们能够更加精准地工程化大模型的跨模态生成能力, 以系统性的语境约束弥补大模型因缺乏具身经验而产生的语义理解偏差。通过槽位填充, 我们手动干预大模型的注意力分配, 明确槽位要素使大模型补全具身细节, 从而弥补大模型缺乏现实世界具身经验的局限。

1. 三层认知映射机制

场景结构层基于框架语义学, 框架语义学不仅是提取要素的工具, 更是一种语境约束。将动词关联的施事、受事及工具等核心要素转化为指令中的[场景]与[主体]槽位。框架决定了场景中必要的参与者角色及其互动方式, 使图像生成趋向语境的结构化表达。

文化语义层基于概念隐喻理论, 负责[隐喻客体]槽位的选择和填充。在这个层面, 我们没有直接描绘词汇的抽象概念, 而是通过映射规则, 将抽象目标域转化为能够体现中国文化内涵的具象符号, 从而达到提升学习者的文化理解能力的目的。

动作逻辑层基于意象图式理论, 对[动作]槽位进行物理性约束。动词展现出来的动态过程能够借助路径图式和力量图式呈现在静态图像中, 通过这种方式, 我们引导学习者关注图像中的动作线索, 进而理解词汇所表达的动态感。

2. 提示词生成的标准化公式

为确保策略在大模型中的高效执行, 我们整理上述理论要点形成了标准化提示词模板: [场景] + [主体] + [动作] + [隐喻客体] + [约束]。各槽位的填充规则如下:

(1) 场景: 根据目标动词的认知层级和生成策略, 确定场景类型, 构建场景结构层, 从而填入相应的空间描述。每个场景需明确贴合国际中文教学场景, 并选取无文化壁垒、视觉辨识度高的环境元素, 如课桌、黑板、证书等, 同时控制无关元素的数量。

(2) 主体: 优先选用与国际中文教育相关的要素, 并根据词汇情感色彩匹配相应的面部表情来辅助语义传达。

(3) 动作: 根据抽象动词的核心物理动作或互动行为, 结合生成策略与动词认知本质, 提取具身动作。明确动作的细节, 避免模糊表述。

(4) 隐喻客体: 选择能够承载抽象义的关键物体或物理符号。在使用文化符号时, 需要明确文化符

号与抽象语义之间的关联; 并且文化符号占比不应该超过场景元素的三分之一, 避免符号的滥用和因元素过多对核心语义造成干扰。

(5) 约束: 用于规定艺术风格、观察视角和负向提示词。风格上可以采用教育插画或现实风格; 视角的确定需要根据与词汇相关的教学点; 负向提示词则是从像素层面过滤不合理的视觉元素。

五、认知强化的文生图实验

本章通过实验和案例分析, 系统验证前文构建的优化提示词路径, 考察其有效性及其与教学需求的适配程度。实验以核心抽象动词为对象, 设计了两组对照, 并采用大模型盲选评估和学习者视角下的近义词辨析, 验证认知强化策略的整体效果, 并进一步探索在复杂教学场景中的应用。结果显示, 认知强化提示词能够明显提升教学图像的语义精准度和教学适用性。

(一) 实验设置与生成配置

本研究人工复核前述样本是否具备视觉映射路径, 最终形成了由 140 个核心抽象动词组成的评测实验数据集。实验设置了 3 组不同数据, 来评估认知强化提示词策略的实际效果, 针对 140 个目标动词累计生成 420 张图像。其中, 直译组 BaselineA 模拟最原始的用户行为, 输入动词本身; 释义组 BaselineB 引入词典定义, 通过增加文本信息来优化生成效果; 而认知强化组 Ours 则是根据本研究构建的标准化体系进行图像生成。

在提示词上, 本研究采用 GPT-4o 进行逻辑推理, 将目标动词转化为符合三层认知机制的结构化提示词, 并统一调用即梦模型进行图像生成。从而确保了所有样本在画风与渲染风格质量上的一致性, 排除不同生成模型对实验结果的影响。

(二) 多模态大模型的语义选择实验

本研究采用大模型盲选的评估方式, 以减少人工评分的主观偏差。具体而言, 由豆包模型在认知强化组 Ours、直译组 BaselineA 和释义组 BaselineB 生成的三组图像中进行客观选择。大模型需从 3 个核心维度筛选最适合国际中文教学的图像并说明理由: 一是语义准确度, 判断图像是否精准传递动词的抽象隐喻; 二是视觉清晰度; 三是教学适用性与词汇的整体匹配度。

在 140 个核心动词的测试中, 我们提出的认知强化策略 (Ours) 有 137 次被选为最优选项, 胜率高达 97.86%; 相比之下, 释义组胜率仅 1.43%, 直译组更是低至 0.71%。这一数据分布反映了当前大模型在抽象语义理解上的不足: 仅依靠直译或词典定义作为原始提示词, 难以实现动词从抽象语义到具体图像的认知转换。在国际中文教育中的抽象概念生成任务中, 引入认知强化可以驱动大模型实现精准教学产出。

(三) 学习者视角的抗干扰效能实验

为验证图像脱离文字语境后的独立表意能力, 我们设置了图像与词汇的匹配测试, 模拟真实的生词学习场景。实验设计加入了高难度干扰项, 例如让大模型识别“上调”这个词时, 要能把它和“提高”“提升”这类常用近义词区分开。这说明图像不仅要传达大致的动作方向, 更需通过特定的视觉隐喻例如“精细调节参数的手势”来表达词汇的特异性。

实验结果显示, 认知强化组、直译组和释义组在该任务中的准确率分别为 54.3%、46.8% 和 45.3%。认知强化组整体表现高于另外两组, 说明使用认知强化策略生成的图像能有效提升课堂近义词辨析的教学效率, 有助于学习者识别近义词之间较难直接观察的语义差异。相比单纯的文字释义, 视觉信息能够提供更直观的理解线索。

以“上调”为例, 直译组和释义组往往生成“箭头向上”这种通用的图像 (图 5 左)。这类图像只能呈现“向上变化”这一宽泛概念, 难以将其与“提升”等近义表达区分开。认知强化组则构建了特定隐喻场景 (图 5 右), 能更好地呈现“上调”所包含的调节、变动等特征。

需要注意的是, 认知强化组虽取得三组中的最高准确率, 但结果也没有超过 60%。这一表现说明, 仅依靠静态图像仍难以完整呈现国际中文教学中近义词之间的细微差异。此类差异不仅涉及词义, 还与具体语境、搭配习惯和语体特征有关, 其中部分信息无法通过单幅静态图像充分表达。因此, 文生图技术在国际中文教育中应定位于认知辅助工具, 而不宜独立承担语义定义功能, 应与文字释义、语境练习



图 5 “上调”图例

和课堂互动配合使用。实际教学可将图像、例句与语境结合起来,先利用认知强化图像突出近义词之间的概念差异,再通过不同的例句呈现其搭配特点和使用条件,最后借助课堂练习帮助学习者完成从理解到运用的转换。

(四) 案例解析与对比

为了具体说明不同生成策略,本研究从 3 个认知层级中各选取一个典型动词,对照分析对照组与认知强化组的生成图像,考察认知识别对视觉转译的影响。

1. 域替换策略案例:“隐瞒”对比分析

“隐瞒”物理源域为遮挡和藏匿,对照组(图 6 左)为男子用信封挡脸的画面,偏向躲避行为,与“隐瞒信息”的教学语义存在一定偏差。认知强化组(图 6 右)保留“置于身后”的动作框架,通过显化不及格试卷这一教学符号,将物理源域的藏匿动作导向国际中文教育的目标域,既保留遮挡动作,又通过试卷符号将抽象隐喻显化。



图 6 “隐瞒”图例

2. 隐喻混合策略案例:“假装”对比分析

“假装”核心义是表面与真实的对立,对照组(图 7 左)多生成夸张笑容画面,没有表现出伪装的含义。认知强化组(图 7 右)采用隐喻混合策略,将假笑面具与真实泪滴置于同一场景,制造了视觉冲突。画面中人手持表情欢乐的面具,面具之后却正在流泪,表面笑容与真实悲伤形成明显反差,使“假装”所包含的表里不一得以呈现。



图 7 “假装”图例

3. 隐喻重构策略案例: “通过” 对比分析

“通过”的物理义指空间穿越, 抽象义指达标、获得批准, 二者对应的视觉表征不存在重合。对照组(图8左)生成人物站在走廊的画面, 没有体现抽象动态语义。认知强化组(图8右)不再采用物理位移的可视化思路, 转而对达标的底层义素进行认知重建, 生成学生展示 HSK 证书的画面, 以此解决远距离词汇生成难题。



对照组 “通过”

认知强化组 “通过”

图8 “通过” 图例

六、结 论

本文围绕认知语言学与文生图技术的结合方式及其量化实现展开分析, 将框架语义学、概念隐喻和意象图式等定性理论, 转化为三层认知映射的可操作体系, 并进一步形成可执行的分析与生成路径。该体系不仅拓展了认知语言学在 AIGC 场景的应用方式, 也为国际中文智慧教育中的多模态资源建设提供了可借鉴的方法框架。

目前, 本研究形成的相关技术成果已应用于国际中文智慧教学系统的建设与运营过程, 用于抽象词汇的可视化生成, 从而将理论研究成果转化为具体的智慧教学实践。在未来的研究中, 可以进一步深化面向大模型的上下文工程, 并且将“认知强化提示词”从图像拓展到视频任务, 探索动态画面精准呈现动词乃至事件语义的可行方案。

Optimizing Text-to-Image Prompts Based on Cognitive Linguistics: A Case Study in International Chinese Education

GUO Mengxi¹, Cui Xiaolu², ZHAO Yijing¹

(1. School of Information Science, Beijing Language and Culture University;

2. International Chinese College, Beijing Language and Culture University, Beijing 100083, China)

Abstract: This study focuses on the optimization of prompts in text-to-image technology within the field of international Chinese education, constructing a methodological framework guided by cognitive linguistics. By integrating frame semantics, conceptual metaphor theory and image schema theory, the study employs a vector space model to measure the semantic distance between abstract verbs and frames, dividing them into three cognitive levels. Furthermore, it proposes differentiated text-to-image prompt strategies such as domain substitution, metaphor combination and reconstruction. Controlled experiments show that cognitively enhanced text-to-image prompts can effectively improve the semantic accuracy of generated images and their anti-interference ability in teaching contexts. This research provides an operable new path for the generation of multimodal teaching resources and expands the application potential of AIGC technology in international Chinese education.

Key words: cognitive linguistics; multimodal large models; international Chinese education; image generation; AIGC

(责任编辑: 杨 恬)