生成式 AI 对汉语近义词的辨析能力研究
——以“了解、理解、明白”为例

徐琳迪

（郑州大学文学院，河南郑州，450001）

摘要：生成式 AI 能够快速生成词义解释与教学例句，但其近义词辨析输出能否直接用于国际中文教学，仍需结合具体语言材料考察。本文以“了解、理解、明白”为对象，从语义、句法与语用三个层面建立分析框架，对豆包网页端快速模式开展探索性测试。15 道人工编制题在三个新会话中重复提交，共取得 45 个题目回答单元。定性分析发现，模型对典型语境的优选词及若干不自然搭配给出了较稳定的核心判断，但部分回答在例句功能展示、解释边界及替代表达接受度方面存在待复核问题。三轮核心判断相近并不意味着解释完全一致，也不能证明教学有效。本文据此提出“语境比较—解释核查—句式改写—交际运用”的教学活动方案，强调教师对生成材料的筛选与修订。研究结论仅适用于本次单产品、小样本及指定提示词条件。

关键词：生成式 AI；汉语近义词；词语辨析；国际中文教育；豆包

中图分类号：H195.3　文献标识码：A

1 引言

近义词教学既需要说明共同意义，也需要解释搭配、句式和交际情境中的差别。李娟基于真实语料构建近义词辨析知识库，将搭配关系、句法分布、语义特征和用法模板纳入描述[1]。Edmonds 和 Hirst 则从指称意义、语体、表达态度及结构限制等方面刻画近义词差异，强调词语选择与具体情境的联系[2]。这些研究说明，释义相近只是辨析的起点。以“了解、理解、明白”为例，三词均涉及“懂”，却不能据此推定它们在所有句式中可直接互换。

生成式 AI 为词汇材料准备提供了新的工具。Kohnke 等在语言教学技术评述中同时讨论 ChatGPT 的潜在用途、局限及师生所需的数字能力[3]。Han 进一步提出，应分别考察模型的语言特征、学习者如何使用其功能，以及互动是否带来语言行为变化[4]。前者讨论使用条件，后者提出研究议程，均不能直接作为“AI 已提高学习效果”的实验证据。这些观点也提示，模型输出评价与学习者习得研究需要分别设计。

语言通顺并不必然意味着辨析充分：模型可能选对词，却给出范围过大的解释。联合国教科文组织强调以人为中心的应用与教育机构对工具适切性的验证[5]；Ji 等的自然语言生成综述也指出，生成文本可能流畅却不忠实于依据[6]。据此，本研究同时检查答案、解释和例句功能，但不把所有存在争议的语言判断一概归为“幻觉”。

本研究选择三个意义相近、能够进入日常课堂交流的词语，考察豆包网页端在限定任务中的输出表现。研究关注四个问题：模型能否说明共同义与常见侧重；能否处理语境选择、句子判断和替换；解释与例句是否适合中级学习者；重复测试中的判断和理由是否稳定。本文采用单产品探索性案例研究，目的在于识别值得保留的教学资源和需要核查的解释类型，不评价各类生成式 AI 的总体优劣，也不将生成材料的表现等同于学习者的学习效果。

2 词义参照与分析框架

2.1 共同义与语义侧重

教育部《重编国语辞典修订本》将“了解”释为“明白、清楚”，并列出“弄明白、弄清楚”等用法；“理解”包含“了解、明白事理”的解释；“明白”则包含“清晰、清楚”和“了解、知道”等义项[7-9]。释义之间的相互说明表明，三词存在意义重叠。本文只讨论现代常用认知义及相关表达，不使用词条收录的古义。该辞典可用于确认共同义，但不能单独裁定大陆当代汉语的全部搭配与语用偏好。

作为本次材料的操作性分析，“了解”通常突出对情况、信息或人的知晓与熟悉；“理解”通常突出对内容、理由或处境的领会，部分语境涉及体谅；“明白”既可以表示主体懂得，也可以描述表达清楚。这些是帮助比较的常见侧重，不是相互排斥的定义。例如，“深入了解”提示“了解”并非只能指浅层认识；“理解你的选择”也不能直接推出说话者赞同该选择。涉及体谅和接受度的判断，仍需结合大陆规范辞书、实际用例与教师意见补证。

2.2 句法位置与参与者关系

李娟将词与搭配词的句法关系、参与者的语义类别及用法模板连接起来[1]，为本研究设置句式替换任务提供了方法参照。“我明白这条规则”描述人的认知状态，“请把规则解释明白”强调解释达到清楚的结果。后一句中的“明白”处于补语位置，替换为“理解”后，原句式通常不能保持自然。若改为“请把规则解释清楚，让大家理解”，则增加了理解者，并重组了动作关系。上述三词例句是本文分析材料，不能冒称来自李娟的语料。

词类和句法功能也应分别处理。动词、形容词属于词类概念，谓语、补语属于句法功能；一个词作补语，不能据此直接判定它必定不是动词。因此，当测试要求模型分别展示“明白”的认知用法与表示清楚的用法时，评价重点是例句是否清楚体现目标功能，而非机械地以句中位置否定词类。这一区分同样约束本文的分析，避免评分者自身使用不充分的规则。

2.3 语境适切性与教学可理解性

语用分析需要区分“语法上能成立”“在当前情境中较合适”和“最自然的优选表达”。记者到现场获取事故信息时，“了解”与调查动作相配；教师询问学生是否听懂时，“明白了”适合作为简短应答。不过，优选一种说法不等于排除所有其他说法。说话者究竟在确认收到信息、领会要求还是体谅他人，会影响词语选择；不能把一道题的偏好写成跨情境的禁用规则。

本文据此从语义、句法和语用三层观察回答，并增加教学可理解性维度。Webb 的词汇研究分别测量词形、联想、语法功能、句法及意义等方面，说明词汇知识评价可以超越单一释义测试[10]。COERLL 词汇教学资源则区分理解词语与在新句子中运用词语，强调不同语境中的接触[11]。本文参考其多维评价与语境运用思路，检查解释是否清楚、例句是否实现指定功能，不直接移用英语学习研究的效应大小。

3 研究设计

3.1 对象、材料与测试条件

测试于 2026 年 11 月 3 日进行，使用豆包网页端，页面显示“快速”模式和“登录”按钮，实际为未登录网页会话。后端模型版本、系统提示词、采样参数及随机种子不可见，本文不将网页产品等同于某个确定的模型版本。每轮建立新对话，仅用于减少可见聊天历史的影响，不能保证后台路由、隐性记忆或其他因素完全一致。

材料为 15 道人工编制题，分为五类，每类 3 道。题目、题序、提示词和操作性参考判断在首次提交前固定，保存为本地冻结方案；这一记录不是公开预注册。材料和初评标准由 AI 助手辅助编制，句子不来自语料库或学习者真实偏误。题组包含典型情境及较明显的不自然表达，属于目的性选取，不能代表所有近义词任务的难度。

表 1　任务类型与观察重点

类型	题号	观察重点
释义辨析	A1—A3	共同义、侧重与例句
语境选择	B1—B3	优选词与其他表达
句子判断	C1—C3	自然度与规则边界

类型	题号	观察重点
替换分析	D1—D3	句式及语义变化
教学生成	E1—E3	例句功能与对话

例如，A3 要求解释“明白”，并分别提供认知动词用法及表示清楚的用法；B3 设置教师问“听懂了吗”的课堂应答；C1 考察“这份说明写得很理解，大家一看就懂”；D1 要求判断“解释明白”能否直接换成“解释理解”；E2 要求生成分别体现领会内容和体谅他人的例句。这些题目同时考察答案、解释和教学呈现，而非只检查是否生成某个目标词。

3.2 提示词与采集程序

固定提示词为：“你是一名国际中文教育教师。请辨析现代汉语中的‘了解、理解、明白’。逐题作答，保留题号，每题不超过 100 个汉字；选择题给出最自然的选项并说明理由，存在其他可接受表达时请指出；不要把常见倾向说成绝对规则。不要搜索网页，不需要引用文献。”其后按 A1 至 E3 顺序附上完整题组。三轮提示词与题目措辞保持相同，网页对空行的显示归一化另有记录。

每轮将 15 题一次性提交，在三个新会话中共发出 3 次请求，取得 45 个题目回答单元。同轮题目可能彼此提示，不能称为 45 次独立实验。原文、会话链接和采集时间独立保存。第 1 轮首页跳转和第 2 轮采集超时均通过恢复原会话取得完整内容，没有为此重发题组或补写回答。采集时间不用于推算响应耗时。

提示词明确提醒避免绝对化，可能影响回答质量；固定题序也可能产生顺序效应。因此，本研究观察的是具有明确教学要求的条件，不能推断模型在无此提醒时会自行保持同样的边界意识。“不要搜索网页”只是提交指令，不能证明不可见的后台完全没有使用检索能力。

3.3 评价方法与证据边界

Liang 等提出的 HELM 框架按场景与多种指标评价语言模型，并公开提示词和回答，以提高可比较性与透明度[12]。本文借鉴多维评价和保留原始输出的原则，不声称实施完整 HELM 基准。冻结方案设定核心判断或例句功能、解释质量与边界两个维度，各为 0—2 分，每轮满分 60 分。助手初评得分分别为 57、55、56 分，尚未经教师独立复核，故只作为过程记录，主要依据原文开展定性分析。

重复一致性按每题核心判断或功能比较三轮编码，另行检查解释、例句和替代表达接受度。这可区分“优选词相同”与“论证完全相同”。编码由助手完成，正式研究仍需至少两名熟悉国际中文教育的评分者独立复核并处理分歧。新增文献用于解释分析方法，不改变测试前冻结的题目及初评标准，也不补足尚未实施的专家验证或语料库统计。

4 测试结果与案例分析

4.1 典型任务中的稳定判断

三轮均完整回答 15 题。在 B1 记者到现场获取事故情况的语境中，模型均选择“了解”；在 B2 已知拒绝邀请的原因并体会难处的语境中，均选择“理解”；在 B3 教师询问是否听懂时，均选择“明白”。C1 的不自然搭配也在三轮被识别并改为“明白”或“清楚”。这说明在本题组及提示条件下，模型能够围绕典型交际目的提供相近的核心答案。

C2 要求判断“了解只能浅、理解一定深，所以任何句子都不能互换”的命题。三轮均否定该说法，第一轮指出“了解不一定浅，理解不一定更深”，第三轮说明“了解深浅可变，理解也不一定很深”。这些回答避免了把近义词差别写成单一等级，也为教师讲解“侧重不同并不等于完全排斥”提供了可修订的材料。不过，提示词已经提醒避免绝对化，因此不能把这一表现解释为完全自发形成的辨析策略。

辅助编码显示，15 题中 14 题的三轮核心编码一致，A3 第二轮的例句功能展示与另两轮不同。这是粗粒度结果，既不表示 14 题全部正确，也不表示 45 段解释内容相同。尤其是 B3 和 D3，优选词或“并非一定错误”的判断相近，接受度措辞却发生变化，应继续检查其解释边界。

4.2 A3：例句自然与功能展示的区别

A3 第二轮给出：“明白（动词）：弄懂、搞清楚。例句：你把道理说明白。”另两轮分别使用“我明白你的意思”和“我明白这个道理”。“你把道理说明白”本身可以成立，但展示的是说明达到清楚的结果，没有像后两句那样直接展示主体懂得某一内容的认知用法。问题在于例句与教学目标的对应不够清楚，而不是句子必然错误。

第二轮另给“现在思路明白了”作为表示清楚的例句，其自然度和中级学习者适用性仍需教师判断。为减少额外干扰，可在教学中比较“我明白这个道理”和“这个道理讲得很明白”，分别标出理解者、内容及表达清楚的结果。此处的改写是本文提出的教学说明，不属于模型原始回答，也不能作为外部语料证明词语分布。

4.3 C1：改正适当而解释仍需核查

C1 第二轮将原句改为“这份说明写得很清楚，大家一看就懂”，改句回应了任务。不过，理由写为：“‘理解’是动词，不能受‘很’修饰作谓语形容文本易懂。”这段原文具有“形容文本易懂”的限定，不能截去限定后宣称模型明确提出所有动词都不能受“很”修饰。本文把它识别为教学解释容易产生歧义的问题，而非已经获得专家共识的语法错误。

学习者可能据此把不自然表达归因于“理解”是动词或“很”不能修饰它，而忽略“我很理解你”这样的常见结构。更清楚的教学说明应对照两个参与者关系：“说明写得很明白”评价文本表达是否清楚，“读者能够理解说明”描述读者对内容的领会。教师既要检查模型改出了什么句子，也要检查解释是否足以支持学习者迁移到新句子。

4.4 B3 与 D3：优选答案不能替代接受度判断

B3 第二轮说：“了解不合适，不用于应答听懂讲解。”第一轮为“了解一般不适合这个回答”，第三轮为“在此处很不自然，一般不这么答”。这些说法的排除程度不同。“明白了”作为题设中的优选应答较有解释力，但关于“了解了”的限制，需要根据确认信息或领会内容的具体情境核查，不能把一次课堂应答选择直接扩展为普遍禁止。

D3 考察调查之后把“了解了事情的经过”换为“明白了事情的经过”是否一定错误。三轮均未判为必错，但第一轮说“不一定错，但不优选”，第二轮说“不是一定错误”并指出书面调查语境中“了解”更贴切，第三轮则说“不一定错误，但不合适”。三者核心分类相同，接受度却不完全相同。Edmonds 和 Hirst 对近义词意义表达强度及语境依赖性的讨论[2]，可帮助理解为何只核对优选词不足以评价这种差异；本轮具体接受度仍须另行验证。

表 2　重点核查案例

题号	输出特征	核查方向
A3	补语例句列为动词例	目标功能是否清楚
C1	改句适当，理由易歧义	主体与结构关系
B3	较强排除“了解了”	应答目的及语境
D3	核心相同，措辞变化	替换接受度边界

4.5 D2 与教学生成：概括的收益和风险

D2 三轮均认为“我理解你的选择”与“我了解你的选择”在语法上可以成立，并区分体谅与知晓的侧重。但第一、三轮用了“体谅、认同背后缘由”解释“理解”。“认同”容易使学习者把懂得理由与赞同选择合并；实际上，领会他人的处境与是否支持某一决定可以分别表达。此处应作为概念边界风险复核，而不据两个词的共现就断定模型所有相关解释都错误。

E 类回答提供了较直接的教学资源。例如，第一轮 E1 生成“我正在了解这座城市的历史”和“相处久了，我慢慢了解他了”，分别对应获取信息和逐渐熟悉；三轮 E2 都包含“我理解你现在压力很大”，与领会课文或诗歌内容的例句形成对照。它们便于组织比较，但都是模型生成材料，尚未经过实际课堂使用。Ji 等讨论的生成可靠性问题[6]提示教师核查材料；A3 的功能展示不足和 D2 的边界风险则应具体分类，不能仅凭争议就认定其属于已证实的事实幻觉。

5 国际中文教学启示

5.1 从概括记忆转向语境比较

本次测试显示，同一回答可能同时包含有用例句和需要修订的规则。教师可把 AI 作为候选材料生成工具，围绕相近内容设计不同交际目标，再检查词义、句式和语境是否匹配。比如，将询问某人情况、领会某人理由和确认听懂要求分别呈现，让学习者先说明“说话者想做什么”，再选择词语。这样有助于避免只背“了解浅、理解深”等过度简化的口诀。

教学材料还应加入可以讨论的替换。对于语法成立但侧重变化的句子，应要求学习者解释变化；对于保持原句式会造成不自然搭配的句子，应允许改写。模型输出中的“最自然”“不合适”“不能用”需要分别核查。Kohnke 等所强调的师生数字能力[3]，在本研究中可落实为辨认解释依据、核对目标功能和保留语境条件；教师仍须借助规范辞书及可追溯用例，不能用第二次 AI 回答作为唯一证据。

5.2 一个可实施的课堂活动方案

本文提出面向中级中文学习者、约 25 分钟的活动，目标是比较三词侧重，并在新情境中解释选择。活动使用经教师筛选的 E 类例句和 D1 改写，不把未经核查的 B3 限制语作为规则。Webb 关于语境中重复接触的研究[10]提示可安排目标词在不同任务中再现，但其结果来自英语词汇学习，不能证明本活动的效果，也不能与本研究三轮模型调用混为一谈。以下为尚未实施的设计。

表 3　课堂活动建议

环节	时间	学习任务
语境比较	5 分钟	按交际目的选择词
解释核查	7 分钟	找出概括的限制
句式改写	6 分钟	调整主体与结构
交际运用	7 分钟	自编对话并说明

第一步提供记者调查、朋友拒绝邀请和课堂应答三个情境，学生选择词语并说明目的。教师先收集理由，再展示模型优选，避免学生仅凭 AI 答案作答。第二步对照“了解只能表示浅层认识”和三轮对该规则的否定，要求学生思考“深入了解”等表达为何不能纳入固定深浅等级；同时讨论“理解但不赞同”所体现的意义边界。

第三步比较“请把规则解释明白”与“请把规则解释清楚，让大家理解”，用简短标记说明

解释者、理解者及结果，随后让学生自己完成一次改写。第四步让两名学生分别扮演教师和学生、采访者和受访者，使用目标词编写新对话。评价时同时查看词语选择、句式自然度和理由是否符合语境，不能只统计目标词出现次数。

5.3 反馈与后续验证

课堂反馈应保留学习者的原句和解释，再由教师指出具体问题，不把 AI 给出的改写直接替代学生答案。若学习者选用了非优选词，教师需先判断它是否表达另一种合理意图，再决定是否纠正。对于确需纠正的句子，可要求学生说明改写中哪一参与者或结构发生变化，以观察是否真正理解区别。

Han 区分模型能力、互动使用和学习者发展三个研究方向[4]。据此，本研究的 45 个回答单元只能支撑输出分析，不能替代学习效果资料。若验证教学方案，需另设学习者参与的研究，明确水平、前后测、评价标准及比较条件，并观察学习者如何理解和运用反馈。当前没有学习者成绩或课堂资料，故不报告未经实施的教学提升。

6 结论

本文以“了解、理解、明白”为有限词组，对豆包网页端快速模式的三轮回答进行定性分析。结果表明，在明确教学提示和本题组条件下，模型对典型优选词、若干不自然搭配及绝对化深浅规则形成了相近的核心判断；同时，例句功能展示、解释歧义和替代表达接受度仍存在需要核查之处。最终答案、解释依据与教学适切性应分别评价，重复一致也不能替代正确性验证。

研究的主要限制包括单产品、小规模人工材料、批量提交及固定题序、后台参数不可见，以及 AI 辅助拟题和单评分者初评。后续应优先完成教师独立复核和大陆当代用例补证；若增加其他模型、提示条件或提交方式，应分别固定方案并作为独立条件分析。在这些边界内，生成式 AI 可为词语辨析提供候选例句和比较线索，教师则需要通过核查、修订与交际任务把材料转化为适切的教学资源。

参考文献

[1] 李娟. 汉语近义词辨析知识库构建研究[J]. 北京大学学报（自然科学版）, 2022, 58(1): 106-112. DOI: 10.13209/j.0479-8023.2021.098.

[2] Edmonds P, Hirst G. Near-synonymy and lexical choice[J]. Computational Linguistics, 2002, 28(2): 105-144. DOI: 10.1162/089120102760173625.

[3] Kohnke L, Moorhouse B L, Zou D. ChatGPT for language teaching and learning[J]. RELC Journal, 2023, 54(2): 537-550. DOI: 10.1177/00336882231162868.

[4] Han Z. ChatGPT in and for second language acquisition: A call for systematic research[J]. Studies in Second Language Acquisition, 2024, 46(2): 301-306. DOI: 10.1017/S0272263124000111.

[5] UNESCO. Guidance for generative AI in education and research[EB/OL]. (2023-09-07)[2026-11-03]. https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research.

[6] Ji Z, Lee N, Frieske R, et al. Survey of hallucination in natural language generation[J]. ACM Computing Surveys, 2023, 55(12): 1-38. DOI: 10.1145/3571730.

[7] 教育部. 重编国语辞典修订本：了解[EB/OL]. [2026-11-03]. https://dict.revised.moe.edu.tw/dictView.jsp?ID=63145&la=0&powerMode=0.

[8] 教育部. 重编国语辞典修订本：理解[EB/OL]. [2026-11-03]. https://dict.revised.moe.edu.tw/dictView.jsp?ID=61988&la=0&powerMode=0.

[9] 教育部. 重编国语辞典修订本：明白[EB/OL]. [2026-11-03]. https://dict.revised.moe.edu.tw/dictView.jsp?ID=32256&la=0&powerMode=0.

[10] Webb S. The effects of repetition on vocabulary knowledge[J]. Applied Linguistics, 2007, 28(1): 46-65. DOI: 10.1093/applin/aml048.

[11] COERLL, The University of Texas at Austin. Actively Teaching Vocabulary[EB/OL]. [2026-11-03]. https://coerll.utexas.edu/methods/modules/vocabulary/01/teaching.php.

[12] Liang P, Bommasani R, Lee T, et al. Holistic evaluation of language models[J/OL]. Transactions on Machine Learning Research, 2023-08-24. https://openreview.net/forum?id=iO4LZibEqW.

An Exploratory Study of Generative AI's Ability to Distinguish Chinese Near-Synonyms: The Case of "了解", "理解" and "明白"

Xu Lindi

(School of Liberal Arts, Zhengzhou University, Zhengzhou, Henan 450001, China)

Abstract: Generative AI can produce explanations and examples for vocabulary teaching, but the suitability of its near-synonym distinctions requires close examination. This exploratory study examines "了解", "理解" and "明白" through semantic, syntactic and pragmatic analysis. Fifteen constructed tasks were submitted as a complete set in each of three new conversations using Doubao's web interface in its visible quick mode, producing 45 task-level responses from three requests. Qualitative analysis found relatively stable core judgments in typical contexts, alongside issues requiring review in example functions, explanation boundaries and the acceptability of alternatives. Similar core judgments did not imply identical explanations or demonstrated teaching effectiveness. A classroom activity involving contextual comparison, explanation checking, sentence rewriting and communicative use is proposed. The findings apply only to this product, small task set and specified prompting condition; independent teacher review remains necessary.

Keywords: generative AI; Chinese near-synonyms; lexical distinction; international Chinese language education; Doubao