

ChatGPT 近义词辨析的效能研究

王海冰

黑龙江大学, 黑龙江 哈尔滨 150000

摘 要 : 选取《HSK 标准教程》中83组近义词作为样本对 ChatGPT 进行近义词辨析提问, ChatGPT 的辨析结果评价: ChatGPT 在辨析实词近义词、不含相同语素近义词方面效能较高, 但对汉语包含人类的情绪和隐喻理解不到位。期望中国式大语言模型的进一步发展, 基于自然语言处理的人工智能助手在汉语学习中实现全面工具化, 在汉语教学中与教师人机互补, 实现国际中文教育高效智能新体系。

关 键 词 : ChatGPT; 近义词; 实验; 效能

Study On The Efficiency Of Discriminating Synonyms of ChatGPT

Wang Haibing

Heilongjiang University, Harbin, Heilongjiang 150000

Abstract : 83 groups of synonyms in HSK Standard Course were selected as samples to analyze and question the synonyms of ChatGPT. The results of analysis of ChatGPT were evaluated as follows: ChatGPT had high efficiency in discriminating synonyms of solid words and synonyms without the same morpheme, but it did not fully understand human emotions and metaphors in Chinese. It is expected that the Chinese-style large language model will be further developed, and the artificial intelligence assistant based on natural language processing will be fully instrumented in Chinese learning and complement with teachers' man-machine in Chinese teaching, so as to realize a new system of efficient and intelligent international Chinese education.

Keywords : ChatGPT; synonyms; experiment; efficiency

引言

近年来, 人工智能技术的快速发展为语言教育领域带来了新的可能性。ChatGPT 为一种基于自然语言处理的人工智能助手, 用于对话和生成文本, 具有快速回答问题、根据上下文进行互动、模拟对话等功能, 广泛应用于语言学习和语言练习。^[1] 因其强大的文本生成能力和交互特性, ChatGPT 成为一个潜在的近义词辨析教学工具。本研究旨在探究 ChatGPT 在国际中文教育中应用于近义词辨析教学的可行性和效果。通过实验, 我们将评估 ChatGPT 在辅助学生学习汉语近义词的效能。研究结果将有助于我们深入了解 ChatGPT 在国际中文教育中的潜在价值, 既为教师提供更多教学工具选择, 也为学生提供更好的学习体验。

一、ChatGPT 近义词辨析功能

ChatGPT 是一种强大的语言模型, 可以理解和生成人类自然语言文本。^[2] ChatGPT 近义词辨析效能是就是评估 ChatGPT 近义词辨析功能的效度, 将本学科在近义词辨析方面的研究成果与 ChatGPT 自身特点相结合, 运用可供参考的辨析角度和方法帮助我们进行效能评估, 通过实验评分的方式为其能力和作用赋值, 用分数高低直接表示效能高低。

根据其效能的高低结果为国际中文教育中运用 ChatGPT 进行近义词辨析的实现程度提供数据参考, 包括教师教学和学生自学。

二、ChatGPT 用于国际中文教育近义词辨析的实验设计

(一) 待考察问题

第一, ChatGPT 是否能够高质量输出汉语近义词辨析的结果?

第二, ChatGPT 对于不同 HSK 等级的近义词、有无相同语素的近义词、实词与虚词近义词的辨析是否存在效能差异?

第三, ChatGPT 输出的答案是否能直接帮助汉语学习者区别近义词细微差别? 甚至是替代教师在教学中使用?

(二) 实验过程

1. 取词

参考“北京语言大学 HSK 动态作文语料库”和“全球汉语中介语语料库”中偏误率较高的近义词,^[3] 从《HSK 标准教程》中随机抽取课文内、课后习题以及词语辨析板块中83组近义词, 皆为成对近义词, 部分近义词有相同语素, 包括实词与虚词, 确保选词在 HSK1—HSK7-9级各级词汇都有涉及。

注: 本研究为全球中文教育主题学术活动计划项目《欧美学习者汉语焦点语感测试与对比判断研究》(项目编号: SH23Y18) 的阶段性研究成果。

作者简介: 王海冰(2000—), 女, 汉族, 黑龙江省大兴安岭人, 学历: 在读硕士研究生, 单位: 黑龙江大学, 研究方向: 汉语国际教育。

2. 操作

将83组近义词逐一输入 ChatGPT 对话框, 统一输入相同形式“你认为**和**有什么区别?”等待 ChatGPT 输出答案, 并将83组近义词辨析答案进行整理方便专家评分。

整理原则:

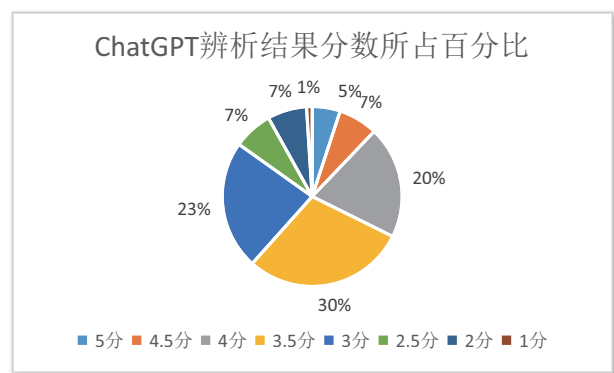
1. 提炼主要观点, 提炼纲目。
2. 去除模棱两可表达, 去除大数据提供的冗余信息。
3. 评分

共5位从教超过10年的国际中文教师对 ChatGPT 在近义词辨析语法、意义、情态和色彩等方面的效能进行评分, 作者也参考《现代汉语正误辞典》《现代汉语八百词》《1700对近义词语用法对比》及教学经验对比 ChatGPT 所给答案进行评分。

将6位专家对83组 ChatGPT 近义词辨析结果的评价得分取平均分, 即为该组近义词辨析的效能评价。对评分存在3分以上差距的近义词辨析结果进行二次赋值, 去掉一个最高分和一个最低分, 再取平均值为 ChatGPT 近义词辨析效能做最终赋值。

(三) 评分结果及说明

1. ChatGPT 全部辨析结果各分数所占百分比



> ChatGPT 近义词辨析效能评分统计图

ChatGPT 对这83组近义词辨析结果中, 总平均分为3.42分。5分占5%, 以其中一个辨析“果断-武断”(HSK7—9级)为例, ChatGPT 能够准确阐述“果断”和“武断”的词义, 明确两者在感情色彩方面的差异, 相对应的例句可以体现词语使用差异, 无法互换使用, 语言内容准确, 逻辑清晰明了。4.5分占7%, 以“容易-简答”(HSK3级)为例, 不同点辨析基本正确, 提供该词主要的用法和搭配对象, 但所提供的例句不准确, 并且并不属于特殊用法。

ChatGPT 辨析近义词的结果中, 4分—3分共占73%, 这一比例证明 ChatGPT 在汉语近义词中可以找到大部分不同点并基本做到一般性知识正确, 错误较少, 逻辑清晰。过于书面语的解析可能也会给汉语非母语的学生造成更难理解的情况, ChatGPT 在语言简化和易懂方面的能力具有不稳定性。2.5分和2分各占7%, 我们发现 ChatGPT 在辨析过程中部分词偶尔会出现主要观点不明确的情况, 找不到关键性不同点; 以其中“从而-因而”(HSK5级, 评分2分)为例, 虽然阐述内容看起来丰富饱满, 但是找不到 ChatGPT 想要表达两者的主要区别是什么, 阅读完整答案还是不清楚“从而”和“因而”有什么不同。1分最少只占1%, 虽不排除本次选词局限性的弊端, 但是也可以证明 ChatGPT 无法用于

近义词辨析的可能性几乎没有。

2. ChatGPT 辨析不同 HSK 等级词平均分数对比

ChatGPT 在处于不同 HSK 等级阶段的近义词辨析平均分数差别不大, 但都保持在3分及以上分数。再次表明, ChatGPT 在辨析近义词方面具有参考价值, 但并不是等级越低的词语 ChatGPT 辨析效能越高, 需要具体词语具体分析。

实验中 HSK1级、HSK4级、HSK7—9级中的词汇辨析分数高于平均分3.42分, 7—9等级的词辨析分数最高, 效能最高。我们认为在高级词汇的近义词辨析中, 符合学生水平(“i+1”)的可用解释词语范围扩大, ChatGPT 被允许更“开放”用词, 在对这些词语进行辨析时, 可以更高效能对其不同点进行描述, 高级阶段的学习者也有能力理解辨析结果。

以 HSK1级词汇“不-没”辨析结果为例, 在辨析结果中明确否定词否定语法成分, 强调“没”用于表示动作在过去或现在未完成, 关键性不同点可以找到, 但例句“我不喜欢这本书。”并不属于“不”与“没”有交叉词义的使用情况, 例句不够典型。若使用“我不去”和“我没去”这样的例句更容易让留学生理解其差异。HSK1级需掌握生词约500个, 都是超高频使用词汇, 对于 ChatGPT 来说, 语料丰富, 辨析难度低。

以 HSK4级词汇“激动-兴奋”辨析结果为例, 明确指出“激动”可能是正面喜悦也可能是负面愤怒, 而兴奋常是积极愉快的这一关键性差异, 但是没有说明“兴奋”还有名词用法, 缺少词语固定搭配的展示, 比如“激动人心”不能说“兴奋人心”, “兴奋剂”不能说“激动剂”等。

以 HSK7—9级词汇“幸亏-多亏”辨析结果为例, 能准确描述关键性区别, “幸亏”表示庆幸, “多亏”表示感谢, 而且例句“多亏了你的帮助, 我完成了这个项目。”也是不可以换成“幸亏”的, 例句典型且附有讲解。这一段的辨析对同等水平的学习者来说理解容易。

在实验过程中, 发现 ChatGPT 对汉语词汇在等级划分的弊端, ChatGPT 无法准确评估输入词语的难度等级。

3. 不含与含相同语素的近义词辨析结果平均分数对比

ChatGPT 在对这83组近义词辨析中不含相同语素近义词辨析平均分为3.5分(高于总平均分3.42分), 含相同语素近义词辨析平均分为3.35分(低于总平均分3.42分), 不含相同语素近义词辨析分数远远高于含相同语素近义词辨析分数, 很明显 ChatGPT 辨析不含相同语素的近义词效能较高。

以含相同语素“参加-参与”辨析结果为例, 在辨析过程中正确找出两个词语在搭配层面不同点, “参与”搭配活动、过程, “参加”搭配具体的组织或事件等。但例句“他积极参与社区服务项目。”也可以说“他积极参加社区服务项目。”, 例句没有体现“参与”的特殊性, 且“参与”并不一定都表示积极意义, 例如“他也参与了这次的毒品交易。”“参加”表示有计划有目的地加入活动或比赛等, 一般是积极的。语义上的细微差异是 ChatGPT 分析上的薄弱之处。

4. 实词近义词辨析与虚词近义词辨析平均分数对比

实词辨析平均分为3.44分(高于总平均分3.42分), 虚词辨

析平均分为3.27分（低于总平均分3.42分），实词近义词辨析效能大幅度高于虚词近义词辨析效能，但并不是所有实词 ChatGPT 都具有较高辨析效能。

实词近义词辨析以指示代词“这么”和“那么”（评分为1分）为例，首先，我们询问 ChatGPT “这么”和“那么”是什么词性，ChatGPT 告诉我们两者都是副词，表示程度、方式或原因，词性分析是错误的，对“这么”和“那么”地分析带有英语思维倾向，简单认为“这”就是离说话人近，“那”离说话人远。^[4] ChatGPT 对于意义相对虚化和与上下文联系紧密的实词辨析效能低，例如部分代词使用语境和语用功能对其意义都产生巨大影响，代词在语用上的涵盖范围也不同，有些更广泛，有些更具体，这些 ChatGPT 无法高效分析。

由于汉语用虚词来表示语法关系既特殊使用频率又很高，对汉语作为第二语言学习者来说是个难点，要区分它们更难。从生成结果中不难看出，实词近义词辨析中语义差异是 ChatGPT 的强项，语法意义的不同确是其短板，很少在答案中看到充当谓语、充当宾语补语等字眼，而虚词近义词又主要涉及语法关系意义的差异辨析，这也是辨析虚词效能低的原因。ChatGPT 在对“对于”和“关于”的辨析结果中没有体现两者位置关系的差异，都做状语时“关于”一定要放在主语前，而“对于”放在主语前后都可以，另外，“关于”还可以做定语。不了解这些的汉语学习者还是会出现各种偏误。

三、ChatGPT 近义词辨析效能评价

（一）基于人类反馈的强化学习人工智能正面效能

第一，无限陪伴学习者学习和快速可视化。^[5] ChatGPT 不受时间空间限制，可以低成本为学生提供课外随时随地的近义词辨析学习，根据自身想法进行提问式学习，延伸学习的汉语近义词方式和时间，加强对汉语的感知能力。

第二，提供全面一般性知识和沉浸式、交互式环境。ChatGPT 关于语义方面的辨析效能较高，可以给出基本合格的答案，通识性知识的准确率也有一定数据支持与保证。更多的语言输出机会，对 ChatGPT 进行近义词辨析提问即其中之一。与 ChatGPT 进行对话无任何交际负担，不怕说错句，增强学生交际意愿。^[6]

第三，无条件辅助教师教学。低时间成本且随时随地为教师提供大量近义词辨析例句、训练方式、习题等。

（二）技术局限导致近义词辨析效能受限

ChatGPT 是基于网络大数据的人工智能帮手，但对知识真伪的甄别、相关知识的整合能力、信息和语言的处理能力有待提高。对语言知识去伪存真、去粗取精、化繁为简的能力也有所欠缺。^[7]

第一，缺乏对词的兼类和语法上的多功能性的归纳，缺乏对搭配规律的总结。比如“精力”（名词）和“精神”（兼类：名词和形容词），ChatGPT 生成文本中没有体现出“精神”的兼类和语法上的多种功能，对两个词的辨析缺乏全面的解释力。

第二，生成答案不符合学习者理解水平，缺乏针对性辨析。因为技术限制，ChatGPT 无法根据问题判断提问者的语言水平，以致在生成文本时出现超纲词和超越等级词，初、中级的汉语学

习者使用 ChatGPT 帮助词语辨析效能会受到影响。

第三，例句有时不够准确鲜活，甚至不符合汉语表达习惯。^[8] “扔”和“撇”的区别中，ChatGPT 给出例句为“他扔了一块石头进了湖里。”无特殊语用要求的情况下，“他往湖里扔了一块石头。”或者“他扔了一块石头进湖里。”更符合汉语表达习惯。

第四，有时生成错误答案。图“扔”和“撇”的区别中，ChatGPT 输出“撇”的解释就是错误的。汉语二语学习者很难识别 ChatGPT 生成文本中的错误内容，也无法知晓产出语句是否是真实意义的表达。

总之，ChatGPT 只能按照用户要求输出答案，学习者被动接受知识。无法像教师一样，启发学习者主动思考发现近义词的不同点，锻炼汉语思维能力。^[9]

四、结论与启示

总体来说，ChatGPT 在对外汉语近义词辨析中，能够迅速且比较全面地展示出近义词的不同点。^[10] 根据我们的实验结果，得出以下结论：1.ChatGPT 完全可以作为在对外汉语教学中有关近义词辨析部分的教学工具，但不能代替对外汉语教师的位置。2. 我们不能简单认定 ChatGPT 对 HSK 等级低、简单的词语辨析效能高，任何词语辨析都不能无条件依赖 ChatGPT，只能是依据其输出结果考虑参考程度。3.ChatGPT 在不含相同语素的近义词辨析下表现出较高的效能，当涉及含有相同语素的近义词辨析时，其辨析效能较低。4.ChatGPT 在辨析语义明确的实词方面表现出较高的效能，在处理语义虚化的实词和虚词时，容易出现较模糊或不准确的辨析结果。

ChatGPT 作为一个正在不断升级的大语言模型，也在持续学习更新数据库，辨析能力都在稳定提高。同时我们也欣喜地看到，国内讯飞等公司出品的基于自然语言处理的人工智能助手也日渐成熟。中国式语言智能技术的发展加速了汉语数字化处理研究，目标是建立起符合汉语使用习惯的语言模型，为全世界汉语爱好者和学习者提供更加合理的汉语智能服务产品。

参考文献

- [1] 刘利，周小兵，高雪松等。“ChatGPT 来了：国际中文教育的新机遇与新挑战”大家谈（上）[J]. 语言教学与研究，2023(03):1-14.
- [2] 李绍林. 对外汉语教学词义辨析的对象和原则 [J]. 世界汉语教学，2010,24(03):406-414.
- [3] 吴红波. 论《HSK 标准教程》中近义词辨析的处理 [J]. 重庆科技学院学报（社会科学版），2018(6):75-77,80.
- [4] 孟祥英. 谈对外汉语教学中的近义词辨析 [J]. 天津师大学报（社会科学版），1997(03):72-76.
- [5] 马丽娟.《HSK 标准教程（4-6）》中近义词的处理考察研究 [D]. 广西师范大学，2023.
- [6] 郭志良. 对外汉语教学中词义辨析的几个问题 [J]. 世界汉语教学，1988(01):19-24.
- [7] 张博. 同义词、近义词、易混淆词：从汉语到中介语的视角转移 [J]. 世界汉语教学，2007,(03):98-107+3.
- [8] 代璐. 国际中文教育高级综合汉语教材中的近义词探究 [D]. 重庆交通大学，2023.
- [9] 马丽娟.《HSK 标准教程（4-6）》中近义词的处理考察研究 [D]. 广西师范大学，2023.
- [10] 李立冬. 对外汉语教学中的近义词辨析 [D]. 河北大学，2011.