

ChatGPT 能不能胜任语言学家的合作伙伴？*

袁毓林

提要 袁毓林(2024)提出了语言学家跟语言大模型协同的三种方式,并举出了一些人机合作进行语言研究的例子(即大小不一的课题)。本文则介绍了用这些问题对 ChatGPT3.5 进行测试的结果,以及对此结果的评论。总的情况是: ChatGPT3.5 在造特定语义类型的句子、有关课题研究的综述和热门论题的小论文写作等方面,表现良好;而对于特定词法的词语的检索和对于特定语法现象背后的复杂语法规律的概括和提取,则表现不佳。但是,随着人工智能技术的进步,ChatGPT 等大模型有望胜任语言学家的合作伙伴。

关键词 语言大模型 人机协同 ChatGPT3.5 人工智能 合作伙伴

一 引言：实测 ChatGPT 完成有关语言研究任务的水平

我们在《ChatGPT 语境下语言学的挑战和出路》(袁毓林 2024)一文中,曾经设想语言学家和 ChatGPT 协同进行语言研究的三种方式:嵌入、副驾驶和智能体,并且分别设计了一些语言研究不同阶段的任务。为了检验目前的语言模型完成这些任务所能达到的水平,我们于 2024 年 6 月 5 日下午,在 ChatGPT3.5 这个比较普通的语言模型上进行了测试。下面分别报告所得到的结果,并且进行简单的评论和分析。

* 本文得到澳门大学讲座教授研究与发展基金(项目编号 CPG2024 - 00005 - FAH)、启动研究基金(项目编号 SRG2022 - 00011 - FAH)和国家社会科学基金专项项目“新时代中国特色语言学基本理论问题研究”(项目编号 19VXK06)的支持,谨此致以谢忱。

二 ChatGPT 完成嵌入式语言研究任务的水平及其评析

我们设计了人机协同进行语言研究的四个嵌入式任务,下面是对前三个任务的测试:

任务 1 作者提示:找出现代汉语中像“获得、落得”那样以“得”结尾的二字词,越多越好。

ChatGPT 回答:现代汉语中以“得”结尾的二字词有很多,以下是一些例子:

1. 承受、2. 考虑、3. 赚得、4. 懂得、5. 需要、6. 舍得、7. 值得、8. 理解、9. 放弃、10. 感受。

作者评论:ChatGPT 并没有完全理解用户关于词法的例示式任务提示。返回的 10 个词中,不合要求的有 6 个。其表现可以判为不及格(得分率约为 40%)。反映出 ChatGPT 不理解人类专家对于特定的词法限制的描述,所以有时无法给出符合要求的词语。

任务 2 作者提示:造出 10 个像“难道要我下跪不成?→难不成要我下跪?”一样的句子配对。

ChatGPT 回答:难不成句子配对:

1. 你认为他真的会相信你的借口吗?→难不成他真的会相信你的借口?
2. 他难道就不能等一等吗?→难不成他就不能等一等?
3. 我们难道就没有别的选择吗?→难不成我们就没有别的选择?
4. 你难道不明白我的用意吗?→难不成你不明白我的用意?
5. 难道这就是你的最终决定吗?→难不成这就是你的最终决定?
6. 你难道不相信我的话吗?→难不成你不相信我的话?
7. 他难道会放过这个机会吗?→难不成他会放过这个机会?
8. 你难道认为这么简单就能解决吗?→难不成你认为这么简单就能解决?
9. 他难道不明白自己的处境吗?→难不成他不明白自己的处境?
10. 难道你就不能给我一点空间吗?→难不成你就不能给我一点空间?

作者评论:ChatGPT 并没有完全理解用户关于构式标记及其变异形式的例示式任务提示。返回的 10 个句对中,先行句都用“吗”,而不是用“不成”来跟“难道”配合(注意:第 1 句没有用“难道”)。也就是说,它对“难道……吗?”这种构式框

架比较娴熟,对“难道……不成?”这种构式框架比较陌生。可能是训练语料中“难道……吗?”类实例比较多,而“难道……不成?”类实例比较少。后续句基本都正确。这也许是“难不成……?”这种连续性构式标记比“难道……不成?”这种离散性构式标记更加容易学习。不合要求的有一半多。其表现可以判为不及格(得分率约为45%)。反映出 ChatGPT 不理解人类专家对于特定的句法(构式)限制的描述,所以有时无法给出符合要求的构式实例。

任务3 作者提示:仿照“Colorless green ideas sleep furiously”,造10个语义违反的英语句子,并译成汉语。

ChatGPT 回答:语义违反的英语句子及其汉语翻译:

1. The ice cream is burning hot. (这冰淇淋火辣辣的。)
2. The sky is crying tears of joy. (天空哭出了欢乐的泪水。)
3. The bookshelf is singing a merry tune. (书架在唱着欢快的曲调。)
4. The flowers are whispering secrets to each other. (花朵正在互相低语着秘密。)
5. The clock is laughing hysterically. (时钟在歇斯底里地笑着。)
6. The sun is wearing a cloak of darkness. (太阳披着黑暗的斗篷。)
7. The wind is tasting the fragrance of the roses. (风在品尝着玫瑰的芬芳。)
8. The river is painting colorful pictures on its banks. (河流在河岸上绘制着丰富多彩的图画。)
9. The moon is dancing gracefully in the daytime. (月亮在白天优雅地跳舞。)
10. The stars are whispering lullabies to the night. (星星在向夜晚低声唱着摇篮曲。)

作者评论:“Colorless green ideas sleep furiously”是乔姆斯基(Chomsky)《句法结构》(*Syntactic Structure*) (1957) 中著名的句法正确但是语义怪诞的例句。ChatGPT 在造这类语义违反句方面的表现优秀,甚至有点儿惊人。当然,这些句子的语义怪异的程度有参差。其中,第1、2、6、9句因为前后有反义词等矛盾表达,所以比较成功;而其他句子则违反了词语的常规的搭配和选择限制,但是在拟人化语境或者诗歌等文艺语体中就不太怪异。可能的原因是:训练语料中较少出现语义矛盾的句子,但是存在超常规搭配的修辞性句子。总的来说,其表现可以判为良好(得分率约为70%)。反映出 ChatGPT 对于文本中出现过的语言学热门论题比较容易搜索到相关语料,并且能够学会相应的有特定句法语义特点的句型。

第四个任务“把下面的段落缩写成30字以内的一句话,给下面的文章摘要进

行润色和修辞加工”,没有进行测试。

三 ChatGPT 完成副驾驶式语言研究任务的水平及其评析

我们只设计了一个人机协同进行语言研究的副驾驶式任务,但是分三步进行:

任务 4-1 作者提示: 什么是否定极项(negative polarity items)? 举例说明英语中有哪些典型的否定极项词语? 它们有什么样的句法语义特点? 这方面有哪些参考文献和研究结论?

ChatGPT 回答: 否定极项(negative polarity items)是一种语言现象,指的是只能出现在否定环境中才能表达其本义的词语或短语。它们在肯定语境中通常会引发语法或语义的不合乎自然的感觉。

典型的否定极项词语包括:

1. Any、2. Ever、3. At all、4. Yet、5. In the least、6. Lift a finger、7. Lift a hand、
8. Raise an eyebrow、9. Touch a drop、10. Bat an eyelid

这些词语在否定环境中才能正确使用,例如:

- I didn't see * * any * * cars. (我没看到任何车。)
- She hasn't called * * yet * *. (她还没打电话。)
- I didn't understand * * a word * * he said. (我一句话都没听懂他说的。)

否定极项在语法和语义上都有特定的特点,如只能出现在否定句中,不能独立存在,也不能与一些特定的词语搭配等。

作者评论: ChatGPT 的回答和举例基本正确。但是,没有理会提示中的“有些参考文献”这一项。总的来说,其表现可以判为良好(得分率约为 80%)。

任务 4-2 作者提示: 举例说明否定极项这个概念对于描写和解释语法现象有什么作用?

ChatGPT 回答: 对于描写和解释语法现象,否定极项的概念能够帮助我们理解语言中否定和肯定之间的关系,以及语言中词语的语法和语义特点。它对于句法分析、语义分析以及语言学理论的建立都有重要意义。

作者评论: ChatGPT 的回答虽然简单和抽象,但是基本正确。总的来说,其表现可以判为良好(得分率约为 84%)。

任务4-3 作者提示:请举例说明现代汉语中有没有否定极项?如果有,那么以“现代汉语否定极项的句法语义研究”为主题,进行专题研究和论文写作合适吗?

这个主题有哪些重要的参考文献和研究结论?还可以进行哪些方面的拓展和挖掘?

我从哪几个方面着手现代汉语否定极项的语料调查和例证分析比较合适?

ChatGPT 回答:在现代汉语中也存在否定极项,例如:“一点也不”“任何”等,在否定句中具有类似的使用特点。

针对现代汉语否定极项的句法语义研究是一个合适的研究主题。一篇论文可以从以下几个方面展开:

1. 对现有研究文献进行综述,包括关于否定极项在汉语中的使用和语法特点的研究;
2. 收集现代汉语中否定极项的语料,并进行例证分析,探讨其不同语境下的使用;
3. 分析现代汉语中否定极项的句法和语义特点,探讨其与否定句的关系;
4. 比较现代汉语和其他语言中否定极项的异同,探讨其在跨语言和跨文化环境中的应用情况。

重要的参考文献和研究结论可以包括相关的语言学期刊文章、语言学专著以及会议论文。可以拓展的方面包括对否定极项在口语和书面语中的差异、否定极项的语法化过程以及否定极项与情态词的关系等。

对于现代汉语否定极项的语料调查和例证分析,可以从现代汉语文本中收集否定句,然后分析其中出现的否定极项的类型、位置和语义特点。也可以通过语料库检索等方式获取大规模的语料进行分析。

作者评论: ChatGPT 的回答虽然简单和抽象,但是基本正确;并且,的确对学术新手的选题和规划有一定的启发性。总的来说,其表现可以判为良好(得分率约为84%)。反映出 ChatGPT 对于文本中出现过的学术专题比较容易搜索到,并且能够进行概括、总结和综述。

四 ChatGPT 完成智能体式语言研究任务的水平及其评析

我们设计了两个人机协同进行语言研究的智能体式任务:

任务 5 作者提示: 请以语言学教师的身份, 写一篇面向大众的、关于汉语字母词(如“X 光片、CT 扫描”, 等等)对于语言交际重要性(不能废除、无可取代)的文章, 限 800 字以内。

ChatGPT 回答: ** 汉语字母词对语言交际的重要性 **

作为一种汉语的特殊现象, 字母词是指由拉丁字母构成的, 代表一种概念、行业或专业领域的词汇, 如“DNA 检测”“Wi-Fi 信号”“A4 纸”等。这些词汇在现代汉语中的广泛使用, 不仅方便了人们的生活和工作, 而且对语言交际起到了重要的作用。

首先, 汉语字母词的使用使得人们在跨文化交流中更加便利。在国际化的背景下, 汉语与其他语言的交流越来越频繁, 许多科学、技术、经济和文化领域的术语都以拉丁字母形式被广泛采用。例如, “DNA 检测”这一术语在医学领域中具有重要意义, 其使用不仅可以准确传达信息, 而且有助于加强与国际同行的沟通和合作。

其次, 汉语字母词的使用提高了语言表达的准确性和效率。在某些领域, 由于专业性较强, 使用汉字来表达特定概念可能会产生歧义或不清晰。而采用字母词的方式, 则可以避免这种情况的发生, 让信息传达更加精准、简洁。比如, “CT 扫描”一词准确地描述了一种医学影像学技术, 使用者能够准确理解其含义, 无需进行进一步解释。

此外, 汉语字母词的使用还有助于提升语言的现代化水平和国际化程度。随着科技的发展和社会的进步, 许多新兴行业和概念涌现出来, 其中许多词汇都是以拉丁字母形式命名的。因此, 采用字母词的方式不仅让汉语更加富有活力和时尚感, 同时也有利于汉语的国际传播和交流。

综上所述, 汉语字母词对语言交际具有重要的意义, 它们的广泛应用不仅方便了人们的生活和工作, 而且有助于促进跨文化交流和加强国际合作。因此, 我们应该更加重视汉语字母词的学习和使用, 不断丰富和完善汉语的表达方式, 为汉语的现代化和国际化发展做出更大的贡献。

作者评论: ChatGPT 写的这篇小作文, 不仅语言顺畅、篇章完整, 而且内容正确、考虑全面。总的来说, 其表现可以判为优秀(得分率为 92%)。反映出 ChatGPT 对于文本中出现过的热门论题比较容易搜索到, 并且能够进行概括、总结和模仿作文。

任务6 作者提示:请以专业语言学家的身份,从下面三对例句(前面打“*”表示该句子不合语法)上,概括出一条语法规律,并做出一定的说明、解释和论证,写一篇3000字以内的文章:

1. 买这种汽车的人不太多~这种汽车买的人不太多
2. 看见那个学生的人来了~*那个学生看见的人来了
3. 我见过买这种汽车的人~*这种汽车我见过买的人

ChatGPT 回答: **语法规律概述: **

在这三对例句中,第一个句子在语法上是合乎规范的,而第二个句子则是不合乎规范的。我们可以总结出一个语法规律:在汉语中,当一个动词短语修饰一个名词时,该名词通常位于动词短语之后,而不是之前。

**说明、解释和论证: **

1. 第一对例句中,“买这种汽车的人不太多”,动词短语“买这种汽车的人”修饰名词“人”,表达的意思是“买这种汽车的人”的数量不太多。这种句式符合汉语的语法规则。

2. 第二对例句中,“*那个学生看见的人来了”,动词短语“那个学生看见的人”也是在修饰名词“人”,但放在了名词之前,违反了汉语的语法规则,因此不合乎语法。

3. 同理,第三对例句中,“我见过买这种汽车的人”是合乎语法的,“*这种汽车我见过买的人”则是不合语法的。

在汉语中,名词一般位于动词短语之后,这符合汉语的语法习惯和逻辑顺序。因此,第二个句子中的短语顺序不符合语法规则,应当避免在书面或正式的语言场合中使用。

作者评论:汉语关系化从句(上例中是提取主语的“的”字结构)修饰中心语名词构成的同位性偏正结构,可以做主语(如例1、例2),也可以做宾语(如例3)。在做主语的情况下,关系化从句(“的”字结构)中的宾语可以话题化为整个句子的主语;当然,其中是有条件的:主句的谓语必须是表示状态的(个体性谓词,如例1的“不太多”、其他如“很有钱”等),而不能是表示动作的(阶段性谓词,如例2的“来了”、其他如“在指挥交通”等)。在做宾语的情况下,关系化从句(“的”字结构)中的宾语不能话题化为整个句子的主语(如例3所示)。这种复杂的关系化从句中的

成分的话题化的规则,人类专家尚且不容易总结。所以,在目前 ChatGPT 是应付不了的。其表现可以判为不及格(得分率约为 20%)。反映出 ChatGPT 不能发现和挖掘复杂的语法现象背后比较深层次的规律,无法对有关的句法语义限制条件进行概括、总结和提炼。

显然,从目前 ChatGPT 的表现来看,大模型离胜任语言学家的合作伙伴还有一定的距离。但是,考虑到大模型在不断进化和发展,再加上提示技术也在不断改进;我们相信,大模型将来应该可以胜任语言学家的合作伙伴。

参考文献

袁毓林 2024 《ChatGPT 语境下语言学的挑战和出路》,《现代外语》第 4 期。

Chomsky N. 1957 *Syntactic Structures*. The Hague: Mouton.

袁毓林 澳门大学人文学院中国语言文学系 yuanyl@pku.edu.cn

Digital Humanities and Language Research

..... **FENG Zhiwei and DING Xiaomei(17)**

Abstract: Digital humanity is a new interdisciplinary subject that uses computer technology and network technology to study traditional humanities. This paper introduces the history of digital humanities research, the main content of digital humanities research, digital humanities and artificial intelligence, and discusses the relationship between the digital humanities and the language research.

Key words: digital humanities, machine index, machine translation, digitalization of ancient literatures

Can ChatGPT Serve as a Linguist's Collaborator? **YUAN Yulin(60)**

Abstract: *Yuan Yulin* 袁毓林 (2024) proposed three collaboration types between linguists and large language models, citing examples of human-machine cooperation in linguistic research (namely various scales of projects). This article introduces the testing results by implying these questions to ChatGPT 3.5, together with commentary on the outcomes. It is found that ChatGPT 3.5 performs well in constructing sentences of specific semantic types, reviewing research topics, and writing short papers on popular linguistic issues. However, it underperforms when retrieving specific lexical items and generalizing or extracting complex grammatical rules which underly specific grammatical phenomena. Nevertheless, with progress in artificial intelligence technology, large models such as ChatGPT are likely to be competent enough to work as collaborators for linguists.

Key words: large language model, human-machine collaboration, ChatGPT 3.5, artificial intelligence, collaborator

What Can Typology Do? **ZHU Xiaonong(68)**

Abstract: An adequate logical classification denotes a full-fledged typology and a mature science discipline. The fundamental function for a typology is to identify individuals in question. Cross-language identifications and comparisons, and typological