ChatGPT 等大型语言模型对语言学理论的挑战与警示*

袁毓林[1,2]

(1. 澳门大学人文学院中国语言文学系,澳门;
2. 北京大学中文系/中国语言学研究中心,北京 100871)

提　要　本文首先简介 ChatGPT 怎样颠覆自然语言处理的工作范式和语言学的有关信条,然后介绍 Piantadosi (2023)怎样通过对比现代大型语言模型的工作机理与生成语言学的研究路径,得出以下这个极具批判性和震撼力的结论:现代机器学习已经颠覆并绕过了乔姆斯基方法的整个理论框架。为了便于大家的阅读与理解,本文围绕着下面三个问题来介绍和评述 Piantadosi (2023):1)现代语言模型能不能看作是一种语言理论?2)现代语言模型及其工作原理能不能禁得住语言学家的质疑?3)现代语言模型及其表现能不能反驳乔姆斯基的语言学主张?最后,文章指出,面对现代大型语言模型在语言运用上的出色表现,我们应该摒弃鸵鸟法、权威法和先验法,采用正视问题、勇于怀疑、不断试错与验证的科学法。

关键词　ChatGPT　现代/大型语言模型　乔姆斯基/生成语言学的主张　科学法

DOI:10.16027/j.cnki.cn31-2043/h.2024.01.003

一、ChatGPT 颠覆了自然语言处理(NLP)的范式和语言学的信条

这一部分介绍本文的写作缘起与若干技术背景。首先简介以 ChatGPT 为代表的现代大型语言模型在语言运用方面的出色表现,特别指出它们的工作机制迥异于语言学理论对于语言的结构方式与运作机制的认识;从而说明语言学研究已经到了必须躬身反省的时候了,并且交代文章接下来将围绕 Piantadosi (2023)进行介绍与评述。

1.1　自 2010 年代以来,随着深度学习(deep learning)算法在图像和语音识别领域的突破性成功,几经沉寂的人工智能研究再次火热。其中,屡屡打败人类冠军的围棋程序 AlphaGo,使得人工智能和机器学习进入大众视野。2022 年末,聊天程序 ChatGPT 的惊艳问世,再次使

* 本课题的研究得到澳门大学讲座教授研究与发展基金(CPG2023-00004-FAH)和启动研究基金(SRG2022-00011-FAH)及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)的资助,谨致谢忱。

人工智能成为出圈的网络热点。从表面上看,ChatGPT 是一款升级版的聊天机器人(ChatBot):你可以用自然语言问它各种问题,包括数学推理和程序代码方面的问题;它都会用十分规范、非常流畅的自然语言来回答你,甚至连贯地维持多个轮次的你问它答。虽然它有时可能会理解偏差和回答出错,但你可以去引导并纠正它。也就是说,它具有在跟人类的交互中不断学习的能力。因为 ChatGPT 不仅能够自然地跟人进行多回合的聊天、回答各种各样的问题,而且能够写邮件、写文案、写摘要、写代码、翻译外语、编写视频脚本等,所以它更像是一个通用的自然语言处理(NLP)平台。它不仅在语言运用方面表现优异、功能强大,而且工作机制非常单纯。正如 ChatGPT 的重要缔造者、OpenAI 的首席科学家 Ilya Sutskever 所说的[①]:

> 我们的工作原理是不断培训神经网络体系,让神经网络去预测下一个单词。基于过去我们收集的文本,ChatGPT 不仅仅是表面上的自我学习,我们希望它能够在当下预测的单词和过去的单词之间达成一定的逻辑上的一致。过去的文本,其实是用于投射到接下来的单词的预测上。

这种用单一的后词预测机制来一体化地解决多种跟自然语言相关的下游任务的工作方式,不仅颠覆了主流的自然语言处理(NLP)范式,而且也颠覆了人们对于人类语言的结构方式的认知,无声地反驳了相关的语言学信条。下面,我们分别对以上所说的两个方面略作说明。第一,因为传统的 NLP 是流水线(pipeline)范式:先做词法处理(比如,分词、命名实体识别等),次做句法处理(比如,自动句法结构和成分关系分析等),再做语义处理(比如,词义消歧、语义角色识别、代词回指确定等);然后,再利用这些特征完成特定的领域任务(比如,智能问答、情感分析、文本摘要等)。在这种范式之下,每个处理模块都是由不同模型来完成的,并需要在不同的标注数据集上训练。而现代大型语言模型(large language model, LLM)出现后,就完全颠覆和取代了流水线模式[②]。第二,相应地,传统语法学关于词法与句法分立、现代语言学关于句法与语义模块化分治的信条,也受到颠覆性的质疑。

1.2 更加严峻的问题是,ChatGPT 似乎非常接近人类理解和运用语言的实际情况。我们普通人在理解、运用自然语言的时候,好像也没有在大脑中将它拆分为很多步不同的任务,逐个任务进行分析,然后再作汇总。这可能不是人类使用自然语言的方式。人们在听到一句话的时候,并不会依次对它的句法结构、语义关系、实体内容与关系、情感倾向等内容逐一分析,然后再拼凑出这句话的意义。更加自然的情况是,人类对语言的理解过程是一个整体性的过程。这种对一句话的整体性理解,可以用自然语言的形式,通过回复的方式整体性地表现出来。就像课堂上老师就课文上的一句话或一段话的内容向学生提问,并且,从学生的回答上看他是否理解了这句话和这段话。这个过程并不是像 ChatGPT 以前的人工智能系统那样,拆分成多个单项的任务,然后逐一输出情感分析的标签、实体信息的片段或是别的某个单个任务的结果,然后再用这些东西拼凑出回复。而 ChatGPT 却能够径直接收自然语言的请求,然后直接用自然语言作出回复,并保证语言的流畅性与逻辑性。其实,这就是人跟人交流的最自然的方式,也是人工智能研究梦寐以求的最理想的“人机交互”方式。正因为如此,所以大家在使用 ChatGPT 时,会由衷地产生一种“很智能”的体验感[③]。

联想到我国传统的语文教育,通常遵循的是“书读百遍,其义自见”的惯例。先生不会去刻意地讲解分析,学童们则“熟读唐诗三百首,不会作诗也会吟”,最终,大家基本上都学会了作文属对。所以,语言学的一些理论、语法学的一些信条,可能到了该反省与检讨的时候了。

1.3　著名语言学家乔姆斯基对于 ChatGPT 之类的大型语言模型有比较尖锐的批评[④]。但是,很快就迎来了加州大学伯克利分校的神经心理学家 Steven T. Piantadosi 博士的反驳文章,即 Piantadosi（2023）。这篇长文通过详细对比基于梯度计算和记忆结构的现代大型语言模型和以乔姆斯基生成语言学为代表的语言学研究路径,得出以下四个极具批判性和震撼力的结论：1）现代语言模型的崛起和成功,实质上削弱了生成语法学提出的关于语言是天生的这种强烈的主张；2）现代机器学习已经颠覆并绕过了乔姆斯基方法的整个理论框架,包括其对特定洞见、原则、结构和过程的核心主张；3）现代语言模型实现了真正的语言理论,包括对句法结构和语义结构的表示；4）生成语法方法在任何领域都不具有竞争力,并且可以说已经逃避了对其核心假设进行实证测试。以下三节将围绕下面三个问题进行介绍和评述：1）现代语言模型能不能看作是一种语言理论？2）现代语言模型及其工作原理能不能禁得住语言学家的质疑？3）现代语言模型及其表现能不能反驳乔姆斯基的语言学主张？

二、现代大型语言模型能否看作是一种语言理论？

这一部分介绍 Piantadosi（2023）怎样把现代大型语言模型看作一种语言理论,并且对标生成语法理论说明这两种理论在实证性方面的差别。然后再对此进行补充说明与评述。

2.1　Piantadosi（2023）指出,当今在计算语言任务上表现最佳的技术,通常使用叫作转换器（transformers）的深度神经网络。这些是文本的模型,在基于互联网文本的巨大数据集上进行训练,以预测即将到来的语言材料[⑤]。他把这些语言模型获得的巨大成就归功于：首先,我们已经能够在大规模的数据集上训练它们。这部分是由于计算的进步（例如,计算任意模型的导数）,部分是由于能够从互联网上获得大量的文本集。一个典型的语言模型可能要在数千亿的词例上进行训练,估计仅能源就需要花费数百万美元。其次,这种模型的架构可以灵活地处理非本地的依存关系（nonlocal dependencies）,以便在预测一个词时可以利用远处的材料。关键的结果是,领先的模型不仅能够生成合语法句子,而且能够生成整个话语、脚本、解释、诗歌等（Piantadosi 2023:2）。即现代语言模型的语言运用水平,总体上已跟人类比较接近。

那么,这些奇迹是如何发生的呢？他归结为现代语言模型的下列两个关键特征：第一,现代语言模型包括一个注意力机制（attentional mechanism）,允许从先前的某个语言材料上预测序列中的下一个词。例如,ChatGPT 在生成《蚂蚁击沉航空母舰》故事时,当它说“其他蚂蚁震惊和好奇于亚历克斯的……”时,它从之前的几十个词中检索出“亚历克斯”这个名字来确定接续什么样的中心语是合适的。这区别于只能利用前面几个词的条件概率的“n 元语法”（n-gram）等早期最流行的模型。第二,现代语言模型整合了语义和句法。这些模型中的词的内部表示被储存在一个向量空间（vector space）中,这些词的位置不仅包括意义的某些方面,还包括决定词如何在序列中出现的属性（比如,句法）。对于上下文和词义如何预测即将出现的材料,有一个相当统一的界面——句法和语义,在模型中没有被分离成不同的组成部分,也没有被分离成不同的预测机制。正因为如此,这些模型找到的网络参数将句法和语义属性融合在一起,两者以非凡的方式相互作用,并跟注意力机制相互作用。这并不意味着模型不能区分句法和语义,或者,比如不考虑语义而反映句法结构;但是,它确实意味着这两者可以相互提供有用的信息。这种模型的一个相关方面是,它们具有数十亿到数万亿参数的巨大记忆容量,这

使它们能够记忆语言的各种特异性。通过这种方式,它们继承了强调构式(constructions)的重要性的语言学家的传统。这种模型还继承了从语料上自底向上学习的传统,继承了显式地连结句法和语义的计算工作的传统(Piantadosi 2023: 5–6)。

关于大型语言模型的注意力机制及其对于词例的表示方式,需要补充一点知识:转换器(transformer)架构从 2018 年开始统治 NLP 领域,推动了 NLP 的飞跃发展。因为预训练的转换器最重要的思想是引入注意力机制(attention mechanism),把多头注意力(multiheaded attention)和自注意力(self-attention)结合起来。而所谓 Attention,指对句子中每个位置的表示(representation,一般是一个稠密向量),是通过其他位置的表示的加权求和而得到的。比如,要理解和翻译 The animal didn't cross the street because it was too tied,必须确定 it 回指上文的 animal 还是 street? 当语言模型编码 it 时,一个注意力头集中在 animal 上,而另一个则集中在 tied 上;结果,在某种意义上相当于用单词 animal 和 tied 的表示来表示 it,从而消解了 it 的回指歧义。同理,在处理 The animal didn't cross the street because it was too noisy 时,注意力机制使得模型相当于用 street 和 noisy 的表示来表示 it,从而消解了 it 的回指歧义[⑥]。

2.2 Piantadosi (2023)指出,对于这种大规模过度参数化的模型是如何工作的,我们应该有一个很好的心理印象,即它们有丰富的潜在空间来推断隐藏的变量和关系。隐性(或潜在)变量一直是计算理论和非正式理论都试图捕捉的语言的关键方面之一。在一个句子的中间,有一个关于句子的潜在结构的隐性变量;在说一个意义上模棱两可的词时,我们心中有一个关于我们打算表达什么意思的隐性变量;在整个话语中,我们心中有一个更大的故事弧线(story arc),它只在多个句子中展开。语言学的各种形式化体系(formalisms)也试图描述这些隐藏的变量。但是,大型语言模型所做的是推断可能的隐藏结构;因为,这种结构允许他们更好地预测即将出现的材料。这使得它们在概念上类似于数学中的嵌入定理(embedding theorems)。这些定理表明,有时一个动态系统的完整的几何结构,可以从其随时间演化的各种状态的低维投影中恢复出来。语料库是句法和思维的低维投影,因此这并非不合情理:一个智能的学习系统,至少可以通过单纯观看文本来恢复这些认知系统的某些方面(Piantadosi 2023: 6–7)。通过详细的分析,可以看到大型语言模型中存在的结构随着模型在训练后生成文本,其内部状态表示了句法结构和语义的潜在方面。训练后的模型内部表示状态和注意模式的结构,可以捕捉树状结构;这种树状结构跟人类标注的解析树有很强的相似性,以至于可以从一个模型的树状结构的程度来预测它的泛化性能。这些模型似乎在涉及追踪正确的潜在状态(比如,功能词和填空式依存关系)的结构上表现良好。事实上,一些模型的内部处理结构似乎自发地形成了一个直观的管道:表示词类,然后是句法解析、语义分析,等等。所有这些都是可能的,因为大型语言模型发展了对关键结构和依存关系的表示,只是这些表示是以语言学不熟悉的参数化的方式进行的。这正如 Baroni (2022)所主张的,各种语言模型应该被当作是不同的语言学理论。特别是一个由各种理论组成的空间被各种模型参数化,跟数据对比来正式地发现哪一种理论是最好的(Piantadosi 2023: 7)。事实上,我们还没有深入地了解这些模型建立的表达是怎样工作的,……它们不同的内部状态是怎样互相关联来达到成功的预测的。我们只能从某些结构比其他结构工作得好一些来获知:正确的注意力机制、预测、语义表示等是重要的。这种状态类似于医疗史,人们弄清楚了哪一种处置方式更加有效(比如,用柠檬对付坏血病),但是并不了解其机制(Piantadosi 2023: 8–9)。那么,现代语言模型是如何整合各种不同的针对语言的计算

方法的呢？有趣的是，不是通过对它们直接进行编码，而是通过允许它们从内置的架构原则中涌现（emerge）。例如，这些模型知道诸如嵌套句（embedding sentences）和关系小句（relative clauses）；在这个意义上，它们似乎有对层次结构（hierarchy）和递归（recursion）的表征。人们几乎可以肯定这些模型也有约束（constrains）的类似物，可能包括硬约束（如词序）和可违反的、概率性的约束。它们肯定会记住一些构式。所有这些都会在参数中实现，以达到很好地预测文本这个总体目标（Piantadosi 2023：9）。在这里可以发现，Piantadosi（2023）认为文本（从词、句子到段落）似乎有一种不为语言学家所了解的潜在的结构和意义，正好被大模型捕捉到了。对此，我们还需要论证和探索。其实，主要的秘密可能在于转换器一开始接受的由“词嵌入”（word embedding）这种基于神经网络的分布表示方法所产生的词向量。所谓“词嵌入”就是构造词的向量化特征表示，可以非正式地理解为：把词语嵌入到一个数学空间里，即把离散的词语符号转换成连续型数值形式。严格地说，“嵌入”是一个数学概念，表示的是形如 $f(X)\rightarrow Y$ 这样的函数，该函数要满足单射（injective，即每个 Y 值只有一个 X 值跟它对应）和结构保持（structure preserving）的特征。举例来说，如果在某种输入空间中可以定义某种距离度量，而这种距离度量能够线性地变换到输出空间中，那么该距离定义满足了结构保持。“词嵌入”这个术语更多地沿用了原数学概念中结构保持的性质：希望能够把自然语言中的每一个词映射到一个 N 维空间中的一个实值向量，并且在这个 N 维实数空间中，可以形式化地定义词与词之间的相互关系，而这种关系又能够符合我们日常的语言学直觉。比如，假定我们在目标空间中使用欧氏距离或者余弦相似度计算得到的两个词之间的相似关系，跟真实世界中的语言相似性基本一致，那么，我们就得到了一个符合我们的要求的词嵌入[⑦]。于是，通过词嵌入处理，原来作为符号实体的词语抽象成了数学描述，就可以对它进行建模。简而言之，词嵌入就是把词的上下文分布信息嵌入到词的向量表示之中，把一个词语转换成可以反映其意义的向量表示，来让机器读取和处理数据。因为词向量中的数值通过训练得到，并且记录了每个词在训练数据中的上下文信息，所以词向量可以更好地刻画语义信息，从而应用到很多语义匹配（判别同义关系，甚至阅读理解）和生成（甚至自动摘要和写作）任务中。如要比较词与词之间的相似性，可直接通过向量之间的余弦距离的度量来确定。

2.3　Piantadosi（2023）把大型语言模型当作一种科学理论，因为它们是现有的唯一能很好地捕捉人类语言的基本动态（dynamics，即相互作用的方式和机制——引案）的模型。然而，由于是神经网络，它们的状态——至少是初始状态——完全不同于在生成语言学中起主导作用的规则和原则。如上所述，它们用参数来体现一种关于语言的理论，包括对于一个句子和一个话语的潜在状态的表示。在其他科学中，如对于飓风或大流行病的建模，也有完全相同的逻辑：调整参数以形式化，然后比较理论；任何一组假设都会产生各种预测的分布，而假设的调整是为了做出可能的最佳预测。通过这种方式，学习机制在理论空间中配置模型本身，以满足期望的目标函数。对于飓风或大流行病，这是科学所能达到的严格的程度；对于单词序列，每个人似乎都失去了理智（Piantadosi 2023：9–10）。

Piantadosi（2023）指出，乔姆斯基等人 2023 年将这些模型描述为“在一些狭窄的领域”是有用的，但受到“不可消除的缺陷”的阻碍，使它们“跟人类推理和使用语言的方式有深刻的区别”。正如网上迅速指出的那样，他们提出的几个例子（比如，用反事实条件句进行推理，或理解句子 John is too stubborn to talk to（约翰太固执，没法跟他谈）的意义），目前的模型实际上是做得正确的。乔姆斯基等人 2023 年抨击的是这些模型的想象版本[⑧]，而忽略了真正的模型如此恰当地捕捉了句法这一事实。对于这一成功，乔姆斯基和其他人一直声称是不可能的。……

原则上同样强大的模型却表现不同,这就给了我们科学的力量。特别是,我们可以把每个模型或一组建模假定看作是关于心灵如何工作的可能假设。测试一个模型与人类行为的匹配程度,就可以对该模型的假定进行科学测试。例如,该领域就是这样发现注意力机制对模型的表现良好很重要。同样,“消减”(ablation)实验允许研究人员改变网络的一个部分,并利用不同的性能来确定什么原则支持特定的行为(Piantadosi 2023:10)。那么,Piantadosi 把大型语言模型当作一种科学理论,在理论上能不能成立呢?我们认为需要作出一些解释。可以这么来设想:所谓科学理论就是一组关于某种科学问题的假设(hypotheses),而对于大型语言模型来说,它的设计与实现既涉及到对于语言运用的假设(语言学假设),又涉及到怎样在计算上实现特定的语言学假设的假设(计算机假设)。在语言模型这里,用于预训练的转换器是以词向量为输入对象的;而词向量是基于词的分布和“意义即用法”这种语言学假设的;转换器对于输入向量在多头注意力机制的指引下,进行了上下文关联编码和位置编码,使得模型可以从句子的上文有效地预测下文,最终生成合格的句子与文本。其中,每一步计算实现的背后都有相应的语言学假设和计算机假设。就此而言,一种语言模型就是一种整合了有关的语言学假设和计算机假设的科学理论。

2.4 Piantadosi(2023)指出,像所有的科学理论一样,即使我们发现它们在机制或表示方面如何跟人不同,它们仍然是有启发性的。根据乔治·博克斯(George Box)的建议:“所有的模型都是错的,[但是]有些是有用的。”[⑨]我们可以思考这些模型的科学优势、贡献和弱点,而不需要完全接受或否定它们。事实上,通过这种假定测试来帮助我们确定什么是可能的,这些模型已经做出了实质性的科学贡献。比如,在没有内置层级性的情况下,模型能不能发现层级关系?词的预测能不能提供足够的学习信号,来获得大部分的语法?一个计算架构能不能在没有移位(movement)的情况下,掌握特指问句(WH-questions),或者在没有先天的约束原则(binding principles)的情况下使用代词?对于所有这些问题的答案,最近的语言模型显示都是“是”(Piantadosi 2023:10)。这说明:大型语言模型没有借鉴生成语言学的理论,但是取得了令人满意的效果;尽管其工作机理可能异于人类,但仍对我们理解人类智能和语言机制具有极大的启发性。

Piantadosi(2023)指出,这些语言模型实现了好的科学理论的几个核心的目标。首先,它们是足够精确和形式化的,以至于可以用一个真实的计算系统来实现,不同于生成语言学的大多数情况。正是其可实现性使得我们可以看到这些理论的内部一致性和逻辑连贯性。语言模型的向量至少编码了语义的一些方面。至少在一些简单的领域,在一些标注数据的训练下,模型的语义空间可以跟世界对齐。它们学习到的表示可以在一定程度上迁移到其他语言上,这表明它们意指了意义的一些深层次的东西。特别是语言模型预测了一些神经数据。相反,句法的生成理论历来缺乏独立的经验支持,尤其是一直没有跟神经科学发生令人信服的关联。虽然句法的大脑基础对于乔姆斯基(生物)语言学来说是多么重要的,但是句法结构的行为和大脑[神经]证据的缺乏是惊人的。跟基础的短语结构相比,支持诸如移位、语迹/拷贝等非常牵强的理论构件的证据,依然是难以找到的。而且,这些语言模型(作为一种语法理论)是经过经验测试的,它们在许多自然语言处理任务上都达到了当前的最佳水平。而生成句法学的方法不仅在任何领域都无法企及,而且还可以说是逃避对其核心假定进行经验测试。不知道科学史上是否有过这种情况:一个达到如此精确的、已经实现的计算系统,居然被一个不能开发出一种远远无法跟它相比的替代品的领域所鄙薄(Piantadosi 2023:11–13)。这说明:大型语言模型不仅在计算上有效,而且跟有关的神经实验相吻合;并且,批评所谓的生物语言学却缺乏神经心理学的实证性研究,这是对“扶手椅上”的语言学研究的一种警示,读起来令人五味杂陈。

三、现代语言模型能否禁得住乔姆斯基等语言学家的质疑？

Piantadosi（2023）认为，大型语言模型的成功是生成理论的失败，因为它几乎违背了生成理论所推崇的所有原则。事实上，乔姆斯基和他的追随者们长期以来声称必要的原则和先天偏见［例如，约束原则、两叉分支、孤岛限制、空语类原则等］都不需要建立在这些模型中。此外，这些模型的建立没有纳入乔姆斯基的任何重要的方法论主张，比如，确保模型正确考虑语言能力与语言运用，尊重“最小化”或“完美性”，并避免依赖未经分析的数据的统计模式（Piantadosi 2023: 14–15）。下面我们择要介绍和评论。不过，为了照顾语言学同行的阅读兴趣和习惯，我们侧重于介绍和回答大型语言模型能不能禁得住乔姆斯基等语言学家的质疑；希望反过来，从这样一种批判性的角度进一步说明大型语言模型处理语言的基本机制及其成效。

3.1 语言模型能不能分别句法和语义？

Piantadosi（2023）指出，乔姆斯基和其他人长期以来一直强烈地主张，应该将句法作为一个独立的实体来研究；句法不仅跟认知的其他部分，而且跟语言的其他部分都无关。在这种方法中，句法不应该被还原为词与词之间的一般统计数字；但这恰恰是大型语言模型现在提供的东西。现代大型语言模型在基础表示中整合了句法和语义：将单词编码为高维空间中的向量，而没有竭力将诸如词类［等语法范畴］跟语义表示分开，甚至没有在字面（literal word）以外的任何分析层面上进行预测［也就是说，只预测下一个单词］。使这些模型运行良好的部分原因在于，确定如何将语义属性编码到向量中；事实上，是通过编码分布式语义来初始化单词向量的。因此，做预测句法材料的模型不需要假设句法的自主性；而且，这种假设很可能妨碍它［的运作］。我们可以拿经典的 colorless green ideas sleep furiously 来测试，它通常被作为句法功能跟语义分开（而且转移性概率统计用不上）的例子。有趣的是，ChatGPT 不仅可以学习相关的统计数据，而且可以针对用户的问题：“Why is the sentence ‘colorless green ideas sleep furiously’ interesting?”（为什么这种句子有趣？）创造出解释这句话为什么有趣的回答：因为这是一个语法上正确、但是语义上没有意义的句子。甚至，在接到用户的指令 Generate ten other sentences like this（生成 10 个类似的句子）以后，造出了 Purple fluffy clouds dream wildly/Blue glittery unicorns jump excitedly/Black shiny kangaroos hop playfully 等 10 个类似的句子。这个模型成功地得到了 this（“这个”，一个句子）的所指，解决了 like this（“像这样”）中的歧义，明白它在这里指的是句子的结构［即模型理解用户要它创造 10 个跟例句结构相似的句子］。这正是人们本来认为统计模型不应该知道的东西！它在句子中生成了一些并不完全是低频的两词组合（bigrams）。我们可以注意到它的一个弱点：不太容易生成完全无意义的句子，比如 black shiny kangaroos（黑色闪亮的袋鼠）很罕见，但并非不可能。这可能是因为，无意义的语言在训练数据中很罕见。这些结果说明，即使是整合了句法和语义的模型，也能在适当的时候进行句法概括。［即使是］为了显示［独立于语义的］句法行为，句法在基础机制或模型的分析中［也］不需要是自主的（Piantadosi 2023: 15–16）。可见，大型语言模型的词向量及其在注意力机制下的转换等基础表示中，充分地整合了句子的句法和语义信息；并且，在一定的用户指令的提示（prompting）之下，模型能够区分句子的句法构造与语义表达。这说明，对于一种语言理论来说，“句法自治”（the autonomy of syntax）假设是不必要的。

3.2 语言模型能不能利用概率和梯度表示来拟合离散的语法？

Piantadosi（2023）指出，对于现代语言模型来说，概率和信息论是核心。乔姆斯基尽管长期以来一直对概率不屑一顾，但是较新的模型还是使用概率来推断整个生成过程和结构。因为概率预测本质上提供了一个可被使用的错误信号，可以用来调整本身编码了结构和生成过程的参数。在机器学习中也有类似的情况，可能的规则空间被隐含地编码为模型的参数。值得注意的是，出于数字稳定性的考虑，大多数处理概率的模型实际上是用概率的对数工作的。以对数概率工作的模型实际上是在根据描述长度（description length）工作：寻找使数据最可能的参数（最大化概率）与寻找给数据一个简短描述的参数（最小化描述长度或复杂性）是一样的。因此，最佳参数相当于科学理论，在描述长度的确切意义上，它能很好地压缩经验数据。概率远非“完全无用”，它是允许人们实际量化诸如复杂性和最小化的措施（Piantadosi 2023: 16–17）。

这使我们想起 GPT-4 的重要缔造者、OpenAI 的首席科学家 Ilya Sutskever，在前一段时间分别跟英伟达 CEO 黄仁勋（GTC 大会）和前《纽约时报》记者、现 Eye on AI 播客主持人 Craig S. Smith 的两场对话中，反复提到的促使 ChatGPT 成功的两大基础想法：“第一个想法是通过压缩来进行无监督学习”，ChatGPT 实际上压缩了训练数据。从数学意义上讲，通过不断训练这些自回归生成模型实现数据压缩。如果数据压缩得足够好，你就能提取其中存在的所有隐藏信息。训练神经网络预测下一个字符，可以使模型学习到一个可以理解的表示，从而打破无监督学习这种技术瓶颈。即是说，如果有一个神经网络能够预测下一个字符，它就能解决无监督学习问题[10]。“第二个想法是强化学习”，这里不作展开。

Piantadosi（2023）指出，预测是概率性的这一事实是有用的，因为它意味着基础表示是连续的和梯度的。跟生成语言学典型的对离散规则和过程的形式化工作不同，现代语言模型不使用（至少是显式的）规则和原则。它们基于一个连续的计算，允许多种影响因素对即将出现的语言项目产生梯度影响。连续性是重要的，因为它允许这些模型使用梯度方法（其实质是一种微积分的技巧），来计算所有参数的变化方向以最快地减少错误。这并不是说这些模型最终没有离散值，毕竟在英语语料上训练以后，它们鲁棒地（robust）生成了主语居于动词之前的序列。关键的一点是，离散性是连续性模型的特例。这意味着用连续性表示工作的模型得到了两个世界的最佳：在适当的时候拟合了离散的类型，在其他情况下又拟合了梯度的类型。梯度模型超越决定论的规则所取得的成功，启示我们：语言的许多方面是基于梯度计算的。这种成功事实上反映了在数字计算领域“放松”方法的流行性；在该领域，对带有硬约束的优化问题，经常通过一种近似软的、连续的优化问题来解决。这样，跟许多语言学家的直觉相对，即使我们最终想要获得一个硬的、离散式语法，对一个学习者来说，最佳的方式可能还是通过连续的表示来达到（Piantadosi 2023: 17–18）。这启发我们，像优选论那样的音系学理论，可能是一种比较聪明的路子。

3.3 语言模型能不能在无限制的空间里学习成功？

Piantadosi（2023）指出，也许最值得注意的是，尽管现代语言模型关于学习的底层架构相对不受约束，但是它们还是成功了。这是语言的统计学习理论的一个明显的胜利。这些模型能够拟合大量可能的模式，虽然其架构的原则确实制约了它们，使一些模式比其他模式更容易，但所产生的系统是令人难以置信的灵活。尽管缺乏这种约束，该模型还是能够弄清语言的大部分运作方式。

人们不应忽视“刺激的贫乏论”长期以来对生成语言学家所起的作用。大型语言模型终结了这种论调以及相关的争论。它们能够生成超出训练集的句子就是经验主义的胜利。生成句法的教科书中,曾经给出“证明”:因为无限、能产的系统不能被学习,所以句法部分必然是天生的。这种关于在一个不受限制的空间里不能学习的证明是站不住脚的。对于任何熟悉生成句法学措辞的人来说,语言的核心结构可以在没有实质性限制的情况下被发现的观念,听起来好像是不可能的。但是,没有限制的学习不仅是可能的,而且已经被很好地理解甚至预测出来。对于学习和推理的形式分析已经显示:学习者可以从可能计算的空间中推导出正确的理论。特别在语言方面,通过只是观察正面的证据,语法的正确的生成系统同样可以从所有的计算空间(可能是最不受限制的空间)中被发现。根据这种观点,大型语言模型的功能多少有点像自动的科学家或自动的语言学家,他也可以在相对不受限制的空间中工作,通过搜索来发现最好且最简单地预测观测数据的理论。生成句法学提出的下列问题值得思考:为什么小孩子不说 The dog is believed's owners to be hungry(那条狗被相信的主人是饥饿的)或者 The dog is believed is hungry(那条狗被相信是饥饿的)。大型语言模型提供的答案是:在由模型发现的用来解释它所看到的语料的理论中,这些句子是不被允许的。天生限制是不需要的 (Piantadosi 2023: 18–19)。根据生成语法的“原则与参数”理论,儿童大脑中具有天生的符合语法理论原则的普遍语法,不同族群的儿童在所暴露的语言环境中,只需要极其贫乏的语言刺激(包括正面的例子和反面的例子),就可以对普遍语法所允许的具有一定的、有限的变异范围的参数进行选择与调整,最终掌握一门特定语言。而大型语言模型却凭借大数据(庞大的语料训练,基本上都是正面的例子)、大算力和巧算法来生成合格的句子和文本。这种在无限制的空间中学会一种语言的卓越表现,远远地超出了生成语法学的理论假设。

3.4 语言模型没有核心表示,能不能发现层次结构?

Piantadosi (2023)指出,大型语言模型不是最小的表示,而是最大的表示。既没有一个核心的表示或结构的小金块(如合并[merge])来引领这些模型取得成功,任何反对派生复杂性的偏见也不可能对模型发挥关键作用,因为一切都只是一个单独的大矩阵的计算。而且,这种计算在结构上并不是极简主义语言学所指的最小或“完美的”。相反,大型语言模型的注意力机制对任意遥远的材料进行调节,也许没有结构上的关联,因为这就是它们在句子之间建立话语模型的方式。一种符合人类记忆无数的语言组块的能力的语法理论,改变了我们应该如何思考推导的图景;如上所述,基于概率的模型为语法中的复杂性概念提供了正式的立足点 (Piantadosi 2023: 19–20)。

Piantadosi (2023)指出,他们在对这些模型的训练中发现了结构——包括层次结构。这些模型当然可以学习基于线性结构而不是层次结构的规则,但数据强烈地引导它们走向层次化、结构化的泛化。这种发现层次结构不是内建层次结构的能力,而是认知心理学家长期强调的能力。例如,通过聚类诱导句法范畴的工作,我们熟悉的助动词倒置(aux-inversion)的例子,是为了说明儿童必须拥有层次化的句法。作为一个简单的实验,我们也可以要求这些模型来形成一些问句。如:

Convert the following into a single question asking if the accordion is in the rain:

“The accordion that is being repaired is out in the rain.”

Is the accordion that is being repaired out in the rain?

甚至早期的语言模型也能通过这种严格的实验,似乎这种模型知道哪一个 is 应该移位似的。当然还可以要求模型生成更多[不同类型]的问句,却不给出问句类型的引导。例略。这些模型只是被

训练去预测文本,但是却可以完成生成各种疑问句的工作。可见,我们没有必要认为疑问句是从陈述句上推导出来的;并且,看起来模型内部也根本不像发生了这种事情。这种模型或许引导我们仔细考虑不同构式之间的关联,比如陈述句及其相应的不同形式的问句。在模型的潜在激活空间中,这些不同形式的构式可能跟目标问句安置在相邻的地方。对于最佳的预测理论来说,构式可能被相互关联起来,但不是按照句法的标准理论预测的那种方式(Piantadosi 2023: 21–22)。

可见,大型语言模型可以从文本的表层序列上发现其背后的层次结构,从而正确地完成基于层次性的造句和理解工作。并且,相关的构式虽然在形式和意义上相互关联,但是未必具有派生或转换关系;也就是说,不同的构式都是独立自主的"形式—意义"配对。

3.5 从语言模型的表现看语言和思想能不能分离?

Piantadosi (2023)指出,在乔姆斯基看来,人类的语言跟人类的思想有着深刻的内在联系。乔姆斯基2002年将语言描述为"一个表达思想的系统",事实上,这个系统主要用于自言自语。有趣的是,他没有借鉴关于内心独白的文献;这些文献显示了个体之间的巨大差异,有些人说自己根本没有使用内部语言。Mahowald & Ivanova 等人2023年在一篇综合评论中认为,大型语言模型在语言能力和思维之间表现出引人注目的分离。这些模型知道这么多的句法,还有语义的各个方面;但是,用适当的逻辑推理任务来绊倒它们并不难。因此,大型语言模型提供了一个原则性的证明:句法可以存在,并可能跟其他更强大的思维和推理形式分开来获得。我们在语言中看到的几乎所有的结构都可以来自于学习一个好的字符串模型,而不是直接对世界进行建模。因此,这些模型显示出一种语言和思想相分离的逻辑可能性。而相当多的神经心理学支持这种观点:语言和思想在人脑中也是相分离的。大量的失语症文献显示,失语症患者经常能够完成需要推理、逻辑、心智理论、数学、音乐、导航等的任务。可见,在生物体内,语言和其他理性思维过程是分离的。这并不是说语言无法跟思想发生联系,我们有时可以用语言自身来解决一些推理和交际问题。一种引人入胜的提议是:语言也许是一种连结其他表示与推理的核心领域的系统(Piantadosi 2023: 22–23)。

在这里,我们要强调的是:语言可以在一定程度上跟思想分离,但是无法彻底脱钩;因为语言的文本是以思维和交际作为内容的,作为思维与交际工具的语言,必然在设计原理与运作机制上受到思维与交际的深刻影响。因此,透过语言文本,我们可以在相当程度上了解我们生活在其中的世界,包括自然界、人类社会,以及人们的内心世界。正如 Ilya Sutskever 在接受《纽约时报》前记者克雷格·史密斯访谈时所说的:"我认为随着我们的生成式模型变得异常优秀,它们将具有我所说的对世界和其许多微妙之处的惊人程度的理解。它是通过文本的角度来看待世界的。它试图通过人类在互联网上所表达的文本空间上的世界投影来更多地了解世界。""每个神经网络通过 Embedding 表示法,即高维向量,来代表单词、句子和概念。我们可以看一下这些高维向量,看看什么与什么相似,以及网络是如何看待这个概念或那个概念的?因此,只需要查看颜色的 Embedding 向量,机器就会知道紫色比红色更接近蓝色,以及红色比紫色更接近橙色。它只是通过文本就能知道所有这些东西。""我认为我们的预训练模型已经知道了它们需要了解的关于基础现实的一切。它们已经具备了有关语言的知识以及有关产生这种语言的世界进程的大量知识。大型生成模型对其数据——在这种情况下是大型语言模型所学习的东西——是对产生这些数据的现实世界过程的压缩表示。"[11]这就解释了为什么 GPT-4 在常识和推理方面具有不俗的表现。

3.6 语言模型能不能回答“语言为什么是这样而不是那样”的问题？

Piantadosi（2023）指出，乔姆斯基坚称大型语言模型毫无成就，因为不能解释“为什么是这样？为什么不是那样？”[12]其实，这些模型能否解释为什么人类语言具有现在这种样子，是一个有趣的问题。这个问题看来依赖于语言系统是否从语言之前[的某种事物或状态]进化而来，或者[跟人类]同时产生的。如果语言为了一般的序列预测而征用神经系统，那么这种情形是可能的：在我们拥有语言之前，我们拥有类似这些模型的某种构架；所以，语言的形式由先前存在的计算构架来解释。反之，如果两者是共同进化的，那么语言可能就不能用加工机制来解释。考虑到这种不确定性，我们承认有些“为什么”问题大型语言模型也许是不能回答的；但是，这并不意味着它们没有科学价值。同样，牛顿定律没有回答为什么是这些成为定律，而不是相对的其他任何东西；但是，它们依然包含了深刻的科学洞察力。应该认识到，任何人对于“为什么”的回答，都是基于假定。但是，反观乔姆斯基自己的理论，也没有允许对它们提出特别深刻的“为什么”问题。在许多时候，他只是简单地说，答案是遗传学或简单性或完美性，而没有为这些论断提供任何独立的、正当的证明（Piantadosi 2023: 23–24）。可见，在科学探索方面，回答 What（是什么）问题虽然不如回答 Why（为什么）问题深刻，但是依然十分重要和具有实用价值。从乔姆斯基评价关于语言理论的三个充分性的角度看，ChatGPT 作为一种运作良好的语言模型，已经达到了观察的充分性和描写的充分性，只是在解释的充分性方面问题较多。试想一下，它被“喂”了这么多文本语料，又经过反复的预训练和微调，最终能够生成合语法的句子和表述流畅的段落；对此，我们无法否认它“充分地观察”了这些语料，并且通过庞大的参数和“过分”地参数化来拟合人类的语言运用，从而达到了描写的充分性。至于解释的充分性，的确连开发人员都不完全知道它的机理，就像科学家不知道人脑是怎么涌现出语言能力的一样。

四、现代语言模型能否颠覆乔姆斯基的语言研究路径？

这一部分先介绍 Piantadosi（2023）的结论性观点，然后对此进行补充说明与评述。

4.1 Piantadosi（2023）详细地解释了现代大型语言模型如何削弱、反驳和颠覆了乔姆斯基关于语言学的一些主要主张。在文章结束时指出：乔姆斯基是一位杰出的语言学家和哲学家，他在语言学和语言理论领域做出了重大的贡献。他提出这样的观念：语言是一种天生的、由生物学决定的能力，这种能力是人类独有的；所有人类都拥有一种普遍语法或一套天生的语言规则，它使我们能够理解和产出语言。然而，像 GPT-3 这样的大型语言模型的开发，已经挑战了乔姆斯基关于语言学和语言本质的一些主要的主张。表现为：首先，语言模型可以在大量文本数据上进行训练，并且可以在没有任何显式的语法或句法指令的情况下生成类似人类的语言，这表明语言可能不像乔姆斯基声称的那样是由生物学决定的。相反，它表明语言是可以通过接触语言、跟他人互动而习得和发展。第二，大型语言模型在执行各种语言任务（诸如翻译、摘要和问答等）方面的成功，已经挑战了乔姆斯基关于语言基于一套天生的规则的观点。相反，它表明语言是一个习得的和适应的系统，可以通过机器学习算法进行建模并改进。第三，语言模型可以在以前从未见过有关话题的情况下，产生连贯的关于这些话题的广泛的语言。这表明语言可能不像乔姆斯基所说的那样是基于规则的。相反，它可能是更具概率性的和上下文依存的，依赖于从训练它的文本

数据中学习到的模式和关联。总之,尽管乔姆斯基对语言学领域的贡献是重大的,但大型语言模型的开发对他的一些主要的主张提出了挑战,并且为探索语言的本质以及它跟机器学习和人工智能的关系开辟了新的途径(Piantadosi 2023:31)。

我们认为:第一,对于"语言是不是一种天生的、由生物学决定的能力?这种能力是不是人类独有的?"等问题,需要反复掂量,多方求证。比如,鸟类会飞翔,这可以说是它们的一种天生的、由生物学决定的能力;但是,苍蝇、蚊子、蜻蜓、蝴蝶等昆虫也会飞,甚至蝙蝠等哺乳动物也会飞,飞翔这种能力就不是鸟类独有的,而是一种生物适应性进化的产物。并且,由于人类制造的飞机也会飞翔,因而飞翔这种能力甚至不只是动物才具有的。因此,在没有确切地排除其他动物的交际系统没有语言的情况下,也不能贸然肯定语言能力是人类独有的。更何况,现在大型语言模型在语言运用方面表现惊艳,远远超出人们的想象。第二,大型语言模型能够比较完美地执行翻译、摘要和问答等各种语言任务,能不能认为它们具有语言能力呢?这种情形,就像问"潜水艇会不会游泳?"一样,不可能得到简单的答案:一方面,根据"游泳"的词典释义"人或动物在水里游动"[13],那么潜水艇显然不会游泳;另一方面,潜水艇确实在水里游动。这样,你既不能说它会游泳,也不能说它不会游泳。第三,大型语言模型不是利用离散性的语法规则,而是依赖于条件概率和从文本数据中学习到的模式和关联,也可以像人类一样创造性地生成语言。但是,这并不能证明人类语言不是基于离散性的语法规则的。就像莱特兄弟发明了由发动机推动螺旋桨旋转来带动滑翔机飞行的飞机,但是这并不能否定鸟类是靠扇动翅膀来产生巨大的下压抵抗力以推动鸟体快速向前飞行的。

4.2 Piantadosi(2023)指出,在现代科学史上,许多计算科学家已经注意到涌现(emergence)现象,其中,系统的行为看上去有点儿不同于从其底层规则可以预期到的样子。股票市场是不可预测的,尽管个体交易者可能遵循着简单的规则(利润最大化)。市场的涨跌是几百万不同决定的涌现结果。高层次的现象可能难以直观地把握,哪怕拥有了充分的关于交易者的策略和局部目标的知识。复杂系统领域已经记录下来的涌现现象几乎到处都有,从社会动态到神经系统、类晶体、蜂群决策。通常,研究复杂系统的唯一方法是通过模拟。我们一般不能直观地知道一组底层规则的结果,但是计算工具允许我们模拟和看到发生了什么。关键的是,模拟测试了模型中的底层假定和原则:如果我们模拟了交易者,但是不看股票市场的高层次的统计数据;那么,我们一定会错过一些关键的原则;如果我们为蜜蜂们建立个体决策的模型,但是不看涌现出来的蜂群关于去哪儿觅食或在何时成群地飞走的决策,那么,我们一定会错失一些原则。我们不能直接对原则进行测试,因为这种系统太复杂了。我们只能通过这样的方式来获得原则:看模拟是否重现了我们感兴趣的系统同样的高层次属性。事实上,对于大型语言模型表现的吃惊,表明了我们对于语言学习系统缺少好的直觉。……一种有效的研究语言的方案或许已经被仔细地考虑,甚至可能已经发展了这些种类的语言模型,并且寻求将类似极简主义的原则跟支配神经网络的原则进行比较(Piantadosi 2023:26–27)。语言学界普遍认同语言是一个复杂系统,也愿意探索语言是怎样在人类适应性进化的过程中不断涌现出新的结构与功能以及系统性行为特征。但是,语言的适应性进化是一个长时段的过程,不容易在短期内进行观察和对比研究。因此,利用大型语言模型、结合神经心理学实验,可能是一个有前途的研究方向。

4.3 Piantadosi(2023)指出,我们必须坦诚地对待当前捕捉句法的语言模型的水平。在所有的语言学理论中,没有任何东西能跟大型语言模型在句法和语义方面的力量相提并论,更不用说话语连贯性、风格、语用学、翻译、元语言意识、非语言任务等等。它们在所有方面都是游戏规

则的改变者。那些怀疑语言模型能否作为习得模型发挥作用的人应该看到梯度表示、架构假定和隐性或涌现原则作为语法理论的成功。这些模型打开了可信的语言学理论的空间,使我们能够测试传统意义上影响语言学家的原则之外的原则,使我们最终能够发展出令人信服的结构和统计学相互作用的理论,而且似乎解决了生成式句法学家的许多问题,但没有使用他们的任何理论工具和理论构件。大型语言模型改写了语言研究的方法论哲学(Piantadosi 2023:29–30)。的确,以前语言学家过分相信自然语言不能用马尔可夫过程模型来刻画,过分相信离散性的语法规则对于描写语言结构的重要性。现在,基于概率统计的大型语言模型在语言生成方面的成功,足以促使我们反思真实交际中的语言(或者说是言语)的构造与生成机理。

4.4 Piantadosi (2023)赞成 Pater (2019)所阐述的,应寻求将语言学与现代机器学习(包括神经网络)相结合的方法。我们应该培养一种多元化的语言学,以尽可能少的先入之见来处理语言问题——也许甚至可以从根本上重新认识和构想语言的作用和它的模样。也许,乔姆斯基理论所关注的许多"句法的"现象,实际上是关于其他东西的,比如,语用学或记住的构式。也许,语言的普遍性——如果有的话——来自使用的各个方面,比如,交际和认知的压力,或者其他文化因素。也许,语言学可以向认知科学的方法学习。也许,语法理论应该尊重人类对序列材料无与伦比的记忆能力。也许,我们应该让语言学专业的学生学习信息论、概率论、神经网络、机器学习、人类学、数字方法、模型比较、科尔莫戈罗夫复杂性、认知心理学、语言处理、多代理系统等等(Piantadosi 2023:29)。如果说语言学是一门研究语言的结构与功能的经验性科学,那么在大型语言模型完全不理会号称语言学最前沿的生成语法学的情况下,居然在生成和理解语言方面取得如此惊人的成功,这令人不得不思考理论语言学在许多方面有点儿不对劲;寻求语言研究跟人工智能的大型语言模型建构、神经心理学的行为与影像实验的结合,是一个不得不考虑的方向。如果是这样,那么语言学的知识结构、教育方式与课程体系,也必将作出重大调整。这样,语言学才可能更加像是认知科学,而不是笛卡尔式的"扶手椅上"的哲学。

五、结语:人工智能大型语言模型作为语言学理论的参照系

首先说明 ChatGPT 等大型语言模型为什么具有信息压缩功能,以及由词的稠密向量表示而引起的对于语义学理论的重新思考;然后说明面对现代大型语言模型在语言运用上的出色表现,我们应该摒弃鸵鸟法、权威法和先验法,采用正视问题、勇于怀疑、不断试错与验证的科学法。

5.1 值得一提的是,在上文 3.2 中,我们介绍了 Piantadosi 和 Sutskever 关于 GPT 具有压缩信息的功能的观点。那么,何以能够如此呢?我们认为关键在于 GPT 所采用的计算构架:转换器输入的是词向量,随后又在多头注意力机制的作用下,经过在大样本语料上训练和微调,使得每一个词的稠密向量中,不仅包含了它跟其他相关词语的共现、语序与选择限制等句法语义方面的语言学信息,而且,还通过词语组合与语篇组织,反映了有关的世界知识、百科知识和专业知识。比如,上文2.1 中介绍的 the animal 与 tied, the street 与 noisy 之间的属性描述关系, 2.3 中介绍的一个 too stubborn 的人,你 talk to him 的必然结果(是没法跟他说理)这种常识, 3.1 中介绍的 colorless, green, idea 与 sleep, furiously 之间的低频共现关系,及其背后所蕴含的不合情理的属性描述与陈述关系等常识。基于同样的道理,在科学文本中 AI 与 learn, predicate, make, understand, do 等有关词语的共现与组合反映了关于 AI 的功能方面的科学知识。其中,一个词的意义在文本的分布描述中,得到了全方位的知识表示。这类似马

克思1845年春天在《关于费尔巴哈的提纲》中所写下的那句传世名言“人的本质不是单个人所固有的抽象物,在其现实性上,它是一切社会关系的总和”所揭示的节点与网络的关系[14]。类似地,词的意义远远超出一个词本身的指称,而是其出现的一切场合及其网络关系的总和。

事实上,关于意义的语义学理论,语言学和计算机科学领域的主流理论是指称语义学(denotational semantics),即一个单词、短语或句子的语义就是它所指代的客观世界的对象。与之形成鲜明对比的是,深度学习NLP遵循的是分布式语义学(distributional semantics),也就是认为单词的语义可由其出现的语境所决定。用斯坦福大学人工智能大佬、著名NLP学者Chris Manning的话来说:“Meaning arises from understanding the network of connections between a linguistic form and other things, whether they be objects in the world or other linguistic forms.”(意义来源于对语言形式跟其他事物之间连接网络的理解,无论它们是世界上的其他物体还是语言形式。)[15]

5.2 语言学理论往往是一种“大胆的假设”,比如,生成语言学认为语言能力是天生的(innate)。但是,以研究皮拉哈语(Piraha)著称的美国语言学家丹尼尔·埃弗雷特(Daniel Everett)在接受《德黑兰时报》(Tehran Times)的采访时直言不讳地指出:ChatGPT已经证明,一种语言是如何在没有任何硬性语法原则的情况下被学会的;ChatGPT以最严酷的方式驳斥了乔姆斯基关于学习语言需要先天语言原则的说法[16]。再比如,从结构主义描写语言学一直到生成语言学都相信,句法结构的递归性(recursion)是一种语言普遍现象。但是,丹尼尔·埃弗雷特在这个访谈中指出:皮拉哈语是世界上近8000种语言中的一种。在最近的工作中,Geoffrey K. Pullum关于跨语言递归的比较研究表明,其他几种语言与皮拉哈语一样,似乎缺乏句法递归。实际情况到底如何,就需要“小心地求证”了。

面对着人工智能大型语言模型对有关语言学理论的挑战,怎么办?我们不妨重温美国实用主义哲学家皮尔斯讨论过的、人类思想探究与信念确定的历史上出现过的四种方法:1)固执法(the method of tenacity),就是认死理,谁劝都不听;不管别人怎么讲,外界怎么看,我的想法或做法就是这样。就像鸵鸟被追急了,把头埋在沙子里一样。用皮尔斯的话说就是,“它把危险藏起来,然后就没有危险;因为只要它看不见,就没有了。”2)权威法(the method of authority),就是仰仗具有一言九鼎的绝对权威的统治者等,不去怀疑和思考。3)先验法(the method of a priori),就是西方传统哲学的“形而上学的方法”。其核心是先建立一些理念,这些理念有着放之四海皆准的真理特质,慢慢自我实现。数学、几何学也用“先天方法”,先设立公理、公设,然后通过演算将其包含的所有定理和结果全部推出来。它的好处是理想化,追求全真全善全美。从一开始就试图设计出一个完美的蓝图,将所有的可能性都囊括进去。这个蓝图指向“大全”,无所不知,无所不晓,导向一种独断的真理。也就是去找那个绝对不怀疑的阿基米德点,建立起一所无比宏伟、无比辉煌的科学知识大厦。皮尔斯说,这种方法看似理性,但实际上忽视了人的理性的最本质的特征,违背了苏格拉底给我们的关于人类“理性”的教诲:人是有限的,我们的本性是“自知己无知”。4)科学法(the method of science),这种方法的本质就是去“祛疑”或叫“满足我们的怀疑”(to satisfy our doubts)。我们的认知和思想过程是先有信念,然后有怀疑,有怀疑就要释疑。科学探索和再探索的过程说明,信念的确立、知识的获得既不全是依靠超越于人的外在理念或实在,也不全是由人的内在理性所决定。它既源于人又超出于人。上面讨论的前三种方法都和人的主观意愿有关,如个人的意念和意志,统治者的意念和意志,或者说某种绝对的理念。所有这些,都是由人所确定的东西,而苏格拉底说人的认知是有限的。科学就是要超出我们个人乃至人类的主观臆想和意志。所以,如果我们要真正确立真假、对错的信念,遵循科学的方法是唯一正确的做法[17]。

可见,我们不能像鸵鸟那样不去正视大型语言模型的语言运用能力及其理论蕴涵,也不能迷信权

威和不加怀疑地相信其断言“ChatGPT 的虚假承诺”,也不能迷信那种先验的“笛卡尔式的语言学”及其自封的“哥白尼方法”;而是要不断思考、不断提问、不断怀疑、不断探索、不断试错、不断纠错,逐步逼近真相与真理,建立新的信念与理论。其中,实验与验证,包括复现(reproduce)与逆向工程(reverse engineering),是最为重要的环节。正如杰出的物理学家理查德·费曼(Richard Phillips Feynman,1918—1988)所说的:“你不能理解一件事情,除非你能够构建它。”而现在,那个语言运用能力超强的ChatGPT,就像一头大象一样在我们身边走来走去,我们还能够视若无睹吗?事实上,不管我们承认还是不承认、愿意还是不愿意,人工智能的大型语言模型已经成为检验语言学理论的一个重要的参照系。

注　　释

① 详见李海丹、周小燕(2023)。

② 参考 Nick (2023)。

③ 参考 Nick (2023)。

④ 详见 Chomsky (2023)及其中文介绍。

⑤ 通常是一种对单纯词、小于词但大于字符(subwords and supercharacters)的语言单位(比如,构成词的前缀、后缀等广泛出现的语素形式)的编码,笼统地称为“词例”(tokens)。之所以要这样,是因为要减少词汇的数目,便于表示和计算。

⑥ 参考丁小虎(2023)等,但是加入了我们的理解与补充;如果要引用,务请核对原文。

⑦ 关于“词嵌入”的数学背景知识,详见李沐等(2018),第 143—144 页。

⑧ 关于乔姆斯基的这个评论,详见 Chomsky (2023)。

⑨ 我们把为了便于读者理解而补入的词句,放在方括号中。下同。

⑩ 参考王庆法(2023)、元宇宙三十人论坛(2023)和 OneFlow (2023),我们根据自己的理解进行了补充和阐释;如果要引用,务请核对原文。

⑪ 详见王庆法(2023)。

⑫ 关于乔姆斯基的这个评论,详见 Chomsky (2023)。

⑬ 释义引自《现代汉语词典》(第 7 版),第 1587 页。

⑭ 详见《马克思恩格斯文集》第一卷,人民出版社 2009 年版,第 501 页。

⑮ 详见丁小虎(2023)。

⑯ 详见 Mazhari (2023)。

⑰ 详见 Peirce (1877, 1878)和王庆节 (2022)。

参考文献

丁小虎　2023　人类生产力的解放?揭晓从大模型到 AIGC 的新魔法,阿里开发者, 2023. 04. 14;https://mp. weixin. qq. com/s/i6niDdbptYrQQEBYB0jdXA.

李海丹、周小燕　2023　黄仁勋对话 Ilya Sutskever:ChatGPT 有哪些“激动人心的时刻”?,学术头条, 2023. 03. 26;https://mp. weixin. qq. com/s/r3C_fVEHNeBbpK3FeRpzRw.

李　沐等　2018　《机器翻译》,高等教育出版社。

王庆法　2023　OpenAI 首席科学家透露 GPT4 技术原理,清熙, 2023. 03. 17;https://mp. weixin. qq. com/s/H8vNSn-0Ho2Ho4I0n7YDfQ.

王庆节　2022　皮尔斯与美国实用主义哲学的开端,《学术月刊》第 3 期。

元宇宙三十人论坛　2023　黄仁勋对话 OpenAI 创始人:AI 的今天与未来,元宇宙三十人论坛, 2023. 02. 24;https://mp. weixin. qq. com/s/MHPPxbz8wtqK5UntVpfh5Q.

Baroni, Marco 2022 On the proper role of linguistically oriented deep net analysis in linguistic theorising. In *Algebraic Structures in Natural Language*, 1-16. Boca Raton: CRC Press.

Chomsky, Noam 2023 The false promise of ChatGPT. *New York Times*, Mar. 8, 2023. 乔姆斯基:ChatGPT 的虚假承诺,语言治理, 2023. 03. 10; https://mp. weixin. qq. com/s/e9KDOZ3vwd10PFvH6hbtmg; 终于,乔姆斯基出手了:追捧 ChatGPT 是浪费资源,机器之心,2023. 03. 10; https://mp. weixin. qq. com/s/MyiLZYE_hcL27i_qtm7lSA.

Mahowald & Ivanova, Idan A. Blank, Nancy Kanwisher, Joshua B. Tenenbaum & Evelina Fedorenko 2023 Dissociating language and thought in large language models: A cognitive perspective. *arXiv preprint arXiv:* 2301. 06627.

Mazhari 2023 乔姆斯基死敌 Daniel Everett 说 ChatGPT 挑战了"语言先天原则", 摩登语言学, 2023. 04. 08; https://urldefense. proofpoint. com/v2/url?u.

Nick 2023 ChatGPT 的前世今生:"文本生成 + 指令"新范式开启,传统 NLP 技术不会消亡!,算法邦, 2023. 02. 28; https://mp. weixin. qq. com/s/sYeAMrQR7M90OYo6ihT7tw.

OneFlow 2023 GPT-4 创造者:第二次改变 AI 浪潮的方向,贾川、杨婷、徐佳渝译,OneFlow 社区, 2023. 03. 27; https://mp. weixin. qq. com/s/rZBEDlxFVsVXoL5YUVU3XQ.

Pater, Joe 2019 Generative linguistics and neural networks at 60: Foundation, friction, and fusion. *Language* 95 (1): e41-e74.

Piantadosi, T. Steven 2023 Modern language models refute Chomsky's approach to language, https://lingbuzz. net/007180. 乔姆斯基的语言学研究路径是否被计算科学颠覆了?,语言治理, 2023. 03. 15; https://mp. weixin. qq. com/s/sYeAMrQR7M90OYo6ihT7tw.

Peirce, S. Charles 1877 The fixation of belief, *Popular Science Monthly* 12 (November, 1877): 1–15.

Peirce, S. Charles 1878 How to make our ideas clear, *Popular Science Monthly* 12 (January, 1878): 286–302.

Challenges and Warnings from Large Language Models like ChatGPT on Linguistic Theories

Yuan Yulin

Abstract: This article first introduces how ChatGPT discards the working paradigm of natural language processing (NLP) and challenges linguistic beliefs. It then discusses the critical and groundbreaking conclusion which was made by comparing the mechanisms of modern large language models with the research path of generative linguistics in Piantadosi (2023): modern machine learning has subverted and bypassed the entire theoretical framework of Chomsky's approach. To facilitate reading and understanding, this article focuses on the following three questions to introduce and evaluate Piantadosi's (2023) work: 1) Can modern language models be considered genuine theories of language? 2) Can modern language models and their principles be attested under linguistic scrutiny? 3) Can modern language models and their performance refute Chomsky's linguistic claims? Finally, the article suggests that in the face of the outstanding performance of modern large language models in language application, we should abandon the methods of tenacity, authority, and a priori, and instead adopt a method of science that involves confronting problems, embracing skepticism, and continuously experimenting and validating.

Keywords: ChatGPT, modern/large language models, Chomskyan/Generative Linguistic Claims, the method of science