专题研究二　大语言模型

描写还是解释：由ChatGPT反思语言学的两种目标*

袁毓林

（澳门大学　人文学院中国语言文学系　澳门　999078；北京大学　中文系 / 中国语言学研究中心　北京　100871）

提　要　本文在现代大语言模型语境下反思语言学研究的两种目标之争：精确描写（语言事实，how）还是科学解释（语言能力，why）？以此为中心，讨论了一系列相关的问题，并考察了 ChatGPT 能否捕获长距离依存、能否理解句法与语义分离的句子、对语言的科学解释与精确描写是否对立。得出的结论是：（1）ChatGPT 等大模型能够超越马尔可夫过程模型，来捕获语句中不同词语之间的长距离依存关系；能够隐式地学习基本的句法和语义知识，从而理解、识别和生成语义异常的句子。（2）对语言的精确描写和科学解释并不对立，并且前者比后者更加重要。（3）生成语法学的“原则与参数”范式下的范畴语法，对于描写人类自然语言有不可克服的困难。（4）语法学的研究取向应该是语义优先，而不是句法优先。（5）大模型的成功说明：对语言事实的准确描写远比对语言能力的抽象解释更为基本。

关键词　ChatGPT；语言模型；描写 / 解释；语言事实 / 语言能力；语义优先 / 句法优先

中图分类号　H002　**文献标识码**　A　**文章编号**　2096-1014（2025）01-0062-13

DOI　10.19689/j.cnki.cn10-1361/h.20250106

How versus *Why*: Reflections on the Two Objectives of Linguistics by Means of ChatGPT

Yuan Yulin

Abstract　Since its public release at the end of 2022, ChatGPT has attracted worldwide attention. Much research has been conducted on the opportunities and challenges ChatGPT has brought to linguistic studies. At the same time, scholars hold different views on the roles of ChatGPT in linguistic studies. This paper begins with Norvig's (2011) argument on two competing goals in linguistic studies: descriptive accuracy (of linguistic performance, that is, how) and scientific explanation (of linguistic competence, that is, why). Centered on this issue, a series of related questions are discussed, leading to the following conclusions: (1) ChatGPT and Large Language Models (LLMs) can surpass Markov Process Model to capture long-distant dependency holding between different words in a sentence. They can implicitly learn basic syntactic and semantic knowledge, enabling them to understand, recognize, and generate semantically anomalous sentences. (2) Descriptive accuracy and scientific explanation do not contradict each other, and the former is more important than the latter in linguistic studies. (3) Categorical grammar within the "principles and parameters" paradigm of generative grammar faces insurmountable difficulties in describing human natural language. (4) The approach of grammar study should prioritize semantics over syntax. (5) The success of LLMs shows that the descriptive accuracy of linguistic performance is far more basic than abstract explanation of linguistic competence.

*　作者简介：袁毓林，男，澳门大学教授，主要研究方向为理论语言学和汉语语言学。电子邮箱：yuanyl@pku.edu.cn。

本文得到澳门大学讲座教授研究与发展基金（CPG2024−00005−FAH）和启动研究基金（SRG2022−00011−FAH）支持。承蒙编辑部和匿名审稿专家指正，谨致谢忱。

Keywords　ChatGPT; language model; description/explanation; linguistic performance/linguistic competence; semantic priority/syntactic priority

一、引言：语言模型的成功值得语言学家反思

2022 年 11 月 30 日，这也许是一个值得语言学人永远记住的日子：在这一天，美国人工智能机构 OpenAI 发布了 ChatGPT，这是一款以聊天机器人形式呈现出来的人工智能系统，它仿佛真的能够理解和生成人类自然语言；因此，也有可能动摇主宰当代语言学的某些理论假设或研究信念。不信的话，你可以试着跟它像朋友一样问答聊天和争辩纠正，看看期间有没有什么违和感。比如，下面是 2023 年 5 月 4 日笔者跟它的一段对话：

笔者：澳门跟广州之间有直达火车吗？

ChatGPT：是的，澳门和广州之间有直达火车。目前，澳门与广州之间的高速铁路连接已经建成并运营。您可以从广州南站、广州火车站或广州东站乘坐高铁到达澳门，车程约为 1 ～ 2 个小时。不过需要注意的是，因为疫情等因素，火车班次和运行时间可能会有所调整，请您提前查询具体信息并遵守当地的旅游和防疫规定。

笔者：据我了解，澳门和广州之间没有火车。请问，乘汽车方便吗？

ChatGPT：非常抱歉，我的之前的回答有误。您所说的是正确的，目前澳门和广州之间没有直达火车。不过，您可以选择乘坐汽车前往澳门，这是一种比较方便的交通方式。您可以从广州的多个汽车客运站出发，前往澳门旅游。从广州的天河客运站或广州汽车客运站出发，到澳门口岸的车程约为 3 ～ 4 小时左右，票价在 100 ～ 200 元之间。此外，您也可以选择在珠海或横琴岛口岸下车，然后乘坐轮渡前往澳门。但需要注意的是，由于疫情等原因，旅行前最好提前查询相关信息，并遵守当地的旅游和防疫规定。

正如 OpenAI 网站上所介绍的："对话格式使得 ChatGPT 能够回答后续问题、承认错误、挑战不正确的前提，并拒绝不适当的请求。"[①] 然而，该语言模型也存在许多严重的问题，比较突出的是"一本正经地胡说八道"，即出现"幻觉"（hallucination）。这也是大语言模型目前遭受批评的一个普遍性问题。因为，大模型是一种具有众多参数和复杂计算结构的机器学习系统，经过海量自然语言文本等数据的训练，来预测用户给定的文本后面的下一个词语，依此推进，最终生成符合人类语言习惯的文本。至于这个文本的内容的真实性，则是无法保证的。也就是说，它只能基本上保证语法正确（合式的，well-formed），但是并不能够保证内容的正确和可靠（可信的，trust）。

但是，不管怎么说，"让机器理解人们向它发出的自然语言指令"，这个自然语言处理和计算语言学多少年来梦寐以求的目标，似乎已经初步达成。大家可能记得，以前的人工智能系统基本上是各有专长的专家系统，分别擅长图像分类、人脸识别、语音分析、目标检测、机器翻译、语言理解等特定任务。但是，ChatGPT 却像一个全能的"X 战警"，不仅能够聊天和回答问题，而且还能够编程序、写文章、做攻略、画表格、列算式、解方程……以一种近似人类水平的通用人工智能的姿态，掀起新一轮生成式人工智能的热潮：从生成创意内容到协助科学研究，带动大语言模型逐步融入我们的学习、工作、科研和日常生活。这就难怪它这么受到大众用户的欢迎和热捧：ChatGPT 问世后仅仅两

① 详见《行业洞察 | OpenAI 超级对话模型 ChatGPT 发布》，https://new.qq.com/rain/a/20221206A094YW00。

个月，月活跃用户数就成功破亿，成为 IT 产品史上月活用户数最快过亿的消费级应用。这也意味着 ChatGPT 等大语言模型，除了为人们的语言交往和信息交流提供新的动力和趣味之外，还在改变相关行业（比如教育）、简化工作流程（比如招聘）、创造新的产品和艺术内容（比如药品和动漫）；甚至重新开启我们对于技术的异想天开式的期望，走向创造一种人类跟机器可以像人跟人一样亲密无间交流的梦幻般未来。

想一下吧！近在 10 年前，有关业内人士还在壮胆式地高喊："自然语言处理是人工智能皇冠上的明珠。"没想到，现如今人类似乎已经把这颗璀璨的明珠妥妥地收入囊中了。好事来得实在太快，对此，我们语言学人如果不做出一些反思，无论是关于语言的认识论，还是关于语言学的方法论，好像都说不过去。因为，ChatGPT 等大语言模型首先是关于语言的计算模型；其次，它采用的是不被语言学家看好的基于统计的概率方法。比如，早在 1950 年代人工智能概念刚刚提出的时候，彼时的新锐语言学家乔姆斯基已经声称：基于统计的概率模型不能真正刻画自然语言（Chomsky 1956，1957）。于是，问题就来了：不能刻画自然语言的概率模型，何以能够在语言生成和理解方面都有如此杰出的表现呢？诸如此类的问题，都是在当下 ChatGPT 等大模型高歌猛进的语境下，具有理论兴趣的语言学家应该思考的问题，更是正在规划自己未来的职业生涯的研究生们必须直面和正视的问题。

为此，我们下面将分别介绍 60 多年来乔姆斯基对基于统计概率的语言模型的持续质疑，以及人工智能专家 Norvig（2011）对他的评论；其中，重点讨论下列语言学理论与计算处理的关键问题：（1）现代大语言模型能不能超越有限状态的概率转移，来捕获语句中不同词语之间的长距离依存关系？（2）现代大语言模型能不能理解和识别"colorless green ideas sleep furiously"（Chomsky 1956：116）之类经典的句法合格但语义异常的句子？（3）对语言的"精确描写"和"科学解释"是否对立？哪一个更加重要？（4）生成语法学的"原则与参数"范式下的范畴语法，对于描写人类自然语言有没有不可克服的困难？（5）语法学的研究取向应该是"句法优先"还是"语义优先"？（6）语言学家可以从语言大模型的成功中获得什么样的经验与教训？

二、乔姆斯基对概率模型的持续质疑

ChatGPT 虽然炫酷，甫一出世就技惊四座；但是，说到底 ChatGPT 等大语言模型都只是一种人类自然语言的可计算数学模型。于是，以研究人类自然语言的结构和功能为志业的语言学家应该是"与有荣焉"了吧？很不幸，答案是否定的。因为 ChatGPT 等大模型不仅绕开了包括生成语法理论在内的最前沿的现代语言学理论模型（详见 Piantadosi 2023），而且使用的恰恰是几十年前被乔姆斯基判了死刑的基于统计的概率模型。比如，Chomsky（1956：113）在摘要中开宗明义地指出：

> 通过从一个状态到另一个状态的转移来产生符号的有限状态马尔可夫过程（finite-state Markov process）不能充当英语语法。并且，随着 n 的增加，产生英语 n 阶统计近似的此类过程的特定子类，并不会更接近地匹配英语语法的输出。

Chomsky（1957：17）毫不含糊地指出：

> 我认为，我们不得不得出结论：……概率模型（probabilistic model）没有对句法结构的基本问题给出任何洞见。

Chomsky（1969：57）又直截了当地指出：

> 必须认识到，"一个句子的概率"是一个毫无用处的概念，不管从这个概念的什么意义上来

说［都是如此］。

半个多世纪以来，乔姆斯基的这一观点一直没有改变。2011 年，在麻省理工学院纪念建校 150 周年的一个讨论会上，主持人平克（Steven Pinker，哈佛大学心理系教授）向乔姆斯基发问："如何看待概率模型近年来在认知科学领域到处开花的趋势？"乔姆斯基的回应是：[①]

> 确实有许多研究工作在尝试用统计模型来解决各种各样的语言学问题。其中有一些取得了成功。但是大多数是失败的。
>
> 如果不考虑语言的实际结构就应用统计方法，那么所谓的成功不是正常意义上的成功。就科学研究的历史经验来说，这种意义上的成功并非主流。这就好像研究蜜蜂行为的科学家只是对着蜜蜂录像，通过记录蜜蜂的历史行为，加以统计分析，来预测蜜蜂未来的行为。也可能统计方法可以预测得很好，但这算不上科学意义上的成功。研究蜜蜂的科学家并不关心这种预测。

直到 ChatGPT 火爆出圈、名满天下，乔姆斯基依旧不改初心，在《纽约时报》上跟人合作发表文章，直言不讳地批评 ChatGPT 等机器学习系统："只是在随时间变化的概率中进行交互学习，没有提出任何因果机制或物理规律，无法解释英语语法规则；因而，其预测将总是肤浅而又不可靠的。"（Chomsky 2023）甚至在受到 Piantadosi（2023）等的猛烈批评以后，仍然在接受社会学家 Mirfakhraie 的采访时坚称："大语言模型无法阐明人类语言的习得问题，因为它们只是扫描天文数字量级数据以找到统计规律，并根据它们所分析的庞大语料库来预测在序列中可能出现的下一个单词。"（Mirfakhraie 2023）

那么，应该怎样看待乔姆斯基的这些观点呢？下面先从 Norvig（2011）说起。

三、现代大语言模型能不能捕获长距离依存关系？

对于乔姆斯基在 2011 年的研讨会上以及此前的相关观点，人工智能专家、时任 Google 公司研究主管的 Peter Norvig 撰文（Norvig 2011）提出异议。首先，他历数了基于统计的语言模型在搜索引擎、语音识别、机器翻译、问题回答、词义消歧、指代求解、词性标注、句法解析等各项自然语言处理任务上的压倒性成功（对世界做出准确的预测），说明乔姆斯基在 2011 年的那个研讨会上对统计模型的评价不符合事实。其次，他用下面这些例子来说明早期的简单的概率模型的确有问题：

（1）I never, ever, ever, ever, ... fiddle around in any way with electrical equipment.

（2）She never, ever, ever, ever, ... fiddles around in any way with electrical equipment.

（3）* I never, ever, ever, ever, ... fiddles around in any way with electrical equipment.

（4）* She never, ever, ever, ever, ... fiddle around in any way with electrical equipment.

从理论上说，无论句子中的 ever 重复多少次，都不影响（1）（2）是合语法的，而（3）（4）是不合语法的。不过，一个有 *n* 种状态的马尔可夫链概率模型在碰到句子中 ever 的个数超出 *n* 的值的时候，就分不清（1）（2）跟（3）（4）的区别了。因此，概率性的马尔科夫链模型不能处理所有的英语句子。但是，这并不意味着所有的概率模型都要遭到同样的批评。从乔姆斯基刚刚出道的 1957 年到现在，已经发展出许多概率模型。上面这 4 个例子就可以用有限状态模型来区分。此外，还有能力更强的概率上下文无关语法（PCFG），可以处理此类问题。

① 详见 http://languagelog.ldc.upenn.edu/myl/PinkerChomskyMIT.html。

为了测试现代大语言模型有没有理解长距离依存关系的能力，我们去掉（3）（4）句前面的星号，问 ChatGPT-4：“下面 4 个句子，哪些是符合语法的，哪些是不符合语法的？为什么？”它的回答是（2024 年 1 月 10 日）：

ChatGPT：在分析这些句子时，关键是注意动词与主语的一致性。动词应该根据主语是第一人称（I）还是第三人称（she）来调整其形式。下面是对每个句子的分析：

1. I never, ever, ever, ever, ... fiddle around in any way with electrical equipment.
   - 这个句子是**符合语法的**。主语是“I”（第一人称），因此动词使用基本形式“fiddle”。
2. She never, ever, ever, ever, ... fiddles around in any way with electrical equipment.
   - 这个句子也是**符合语法的**。主语是“She”（第三人称单数），因此动词使用第三人称单数现在时的形式“fiddles”。
3. I never, ever, ever, ever, ... fiddles around in any way with electrical equipment.
   - 这个句子是**不符合语法的**。主语是“I”，但是动词使用了第三人称单数的形式“fiddles”。正确的形式应该是“fiddle”。
4. She never, ever, ever, ever, ... fiddle around in any way with electrical equipment.
   - 这个句子也是**不符合语法的**。主语是“She”，但是动词使用了基本形式“fiddle”。正确的形式应该是“fiddles”。

总的来说，（1）和（2）句子是符合语法的，因为它们正确地使用了与主语相一致的动词形式。而（3）和（4）句子是不符合语法的，因为它们的动词形式与主语不一致。

可见，ChatGPT 的语言理解和生成能力都非常强，不仅懂得用户问什么，而且能够用语法学概念来分析、判断和做出说明。那么，为什么它有如此强大的语言处理能力呢？因为它采用了转换器（Transformer）这种神经网络架构，使得计算机能够通过分析文本中单词和短语之间的关系来理解、解释和生成人类语言。尤其是转换器中的注意力机制（attention），不仅专注于它当前正在处理的单词的含义，并且关注上文中跟当前词语相关的词语；从而捕获句子中不同位置上相关单词之间的关系，比如上面 4 个句子中“I/she…fiddle/fiddles”之间长距离的一致关系，无论它们中间插入了多少个 ever。[①] 正是转换器的这种构架和注意力机制，使得语言大模型能够捕获句子成分之间长距离的依存关系，以类似人类的方式聊天、写作和思考、推理；并且，不仅没有违和感，而且有时还逼真、自然得令人惊讶（详见袁毓林 2023 及其所引文献）。看来，基于统计的概率性语言模型的威力还真的不容小觑。

四、现代大语言模型能不能理解句法和语义分离的句子？

Norvig（2011）指出，每一个概率模型实际上都是一个确定性模型的超集（superset），后者只不过是将概率值严格地限定为 0 或 1 而已。对概率模型的合理批评必然是因为它们表达能力过强，而不是因为它们的表达能力不够。乔姆斯基（Chomsky 1956：116；1957：15）提出了一个著名的例子，同时也是对有限状态概率模型的一个批评：

① 上面的（1）（2）改编自 *The Reptile Room* 一书，原书在“A Series of Unfortunate Events”这一章中，在“ever, ever fiddle around in any way with electrical devices”之前，竟然是满满一页纸的“ever,”；详见 https://www.reddit.com/r/Damnthatsinteresting/comments/excoms/lemony_snicket_decides_to_include_a_page_in_a/。

（5）Colorless green ideas sleep furiously.（无色的绿色思想狂怒地睡觉。）

（6）Furiously sleep ideas green colorless.（狂怒地睡觉思想绿色无色的。）

尽管（5）（6）及其任何部分都未曾在说英语者的语言经验中出现过，但（5）是合语法的，（6）是不合语法的。乔姆斯基认为，没有一个 *n* 阶近似模型可以把这两种句子区分开来。Norvig（2011）指出，虽然就整个句子而言，乔姆斯基的判断显然是正确的；但说到句子中的“部分”，则并不尽然。下面是一些部分（二词组合）出现的例子：

（7）“It is neutral green, **colorless green**, like the glaucous water lying in a cellar.” *The Paris We Remember*, Elisabeth Finley Thomas (1942)

（8）“To specify those **green ideas** is hardly necessary, but you may observe Mr. [D. H.] Lawrence in the role of the satiated aesthete.” *The New Republic*, Vol. 29, p. 184, William White (1922)

（9）“**Ideas sleep** in books.” *Current Opinion*, Vol. 52 (1912)

撇开关于“部分”的争议不说，实际上基于统计训练的有限状态模型可以区分上面（5）（6）两例。Pereira（2002）就提出了一个这样的模型，在增加了词类信息后，对新闻语料进行期望最大化的参数训练，计算结果是例（5）的概率为（6）的概率的 20 万倍。为了说明这不是因为这两个句子在新闻语料训练得到的模型中有如此区别，Norvig 本人用 Google 图书语料库（1800 ～ 1954）的训练模型重复做了计算，结果是例（5）的概率为例（6）的 10 万倍。如果可以在树结构的基础上计算，则对句子“合语法性程度”的估计效果会更好。而不是像乔姆斯基提出的基于范畴的语法那样，仅仅只是区分“合语法 / 不合语法”。

饶有趣味的是，赵元任先生曾经设想使不合理的组合“Colorless green, green ideas, ideas sleep, sleep furiously”合理化的语境（Levenson & Chao 1977：256）：

> 我有一个朋友，他脑子里总是充满各种想法：好的想法和坏的想法，精美的想法和粗陋的想法，旧的想法和新的想法。在把他的新想法付诸实践之前，他通常会在它们身上睡觉，让它们成熟起来。但是，有时匆忙之间，他会在想法尚未成熟之前（即它们尚在青涩之时）将其付诸实践。他的有些绿色想法相当生动、丰富多彩，但也并非总是如此，有些想法则相当平淡乏味和缺少色彩。当他想起自己的一些毫无色彩的想法仍然太青涩而无法使用时，他就会在它们身上睡觉，或者用他的话说，让它们睡觉。然而，其中有一些想法可能是相互冲突和矛盾的，当它们在同一个晚上睡在一起时，它们会发生激烈的争斗，把睡眠变成一场噩梦。因此，我的朋友经常抱怨他那**无色的绿色想法在疯狂地睡觉**。

照此说来，在句法（词类组配）合格的前提下，还真的没有绝对的语义不合格的句子。值得一提的是，Piantadosi（2023：15 ～ 16）指出，（5）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，而且可以解释这句话为什么有趣：因为这是一个语法正确但是没有语义的句子；甚至在接到用户的仿造指令后，造出了 10 个类似的句子。下面是我们跟 ChatGPT 的对话：

笔者：下面两个英语句子符合语法吗？

1. “Colorless green ideas sleep furiously.”
2. “Furiously sleep ideas green colorless.”

ChatGPT：第一个句子是著名的例子：“Colorless green ideas sleep furiously.”这是由语言学家诺姆·乔姆斯基（Noam Chomsky）提出的句法结构上合法但在语义上毫无意义的例子，以强调语

法结构和语义意义之间的区别。第二个句子，“Furiously sleep ideas green colorless.”在语法上是合法的，但它的语义可能会令人困惑。这是因为英语中通常采用主语–谓语–宾语的结构，而这个句子的结构略显不同，使得理解起来可能不那么直观。

ChatGPT 对“Furiously sleep ideas green colorless.”的判断不正确。按照包括乔姆斯基在内的大多数语言学家的意见，这是一个在语法上也不合格的句子。我们接下来把任务改成让它造句：

笔者：请仿照“Colorless green ideas sleep furiously.”（无色的绿色观念狂怒地睡觉）造 10 个英语句子，和 10 个相应的汉语句子。

ChatGPT：Certainly! Here are 10 English sentences following the structure of “Colorless green ideas sleep furiously.” And the corresponding Chinese translations:

1. Bitter cold thoughts whisper silently. 刺骨的寒冷思绪悄无声息地低语。
2. Mysterious silver dreams float gracefully. 神秘的银色梦想轻盈地漂浮。
3. Vibrant orange emotions dance wildly. 充满活力的橙色情感狂野地舞动。
4. Fragile golden memories echo softly. 脆弱的金色回忆轻声回响。
5. Enigmatic purple theories hum melodiously. 高深莫测的紫色理论悠扬地哼唱。
6. Serene blue aspirations glide peacefully. 宁静的蓝色志向宁静地滑翔。
7. Chaotic crimson fantasies burn fiercely. 混沌的赤红幻想燃烧得熊熊燃烧。
8. Whispering turquoise concepts flutter aimlessly. 低语的青绿概念漫无目的地飘舞。
9. Lively coral fantasies shimmer vividly. 活泼的珊瑚幻想生动地闪烁。
10. Timeless lavender beliefs resonate mysteriously. 永恒的淡紫信仰神秘地回响。

显而易见，跟原句（5）相比，ChatGPT 仿造的句子在语义上更加合理和可以理解；因为，在训练语料中缺乏像（5）那种人为造成的合语法但是无语义的句子。并且，这 10 个句子跟 Piantadosi（2023：15）中由 ChatGPT 仿造的 10 个，没有一个是重合的。可见，尽管在训练语言模型时，工程师们没有向它注入显式的句法与语义分离的知识，但是语言大模型依然可以在用户的提示下，清楚地区分有关句子的句法与语义层面。

五、对语言的科学解释跟精确描写并不对立

其实，乔姆斯基反对语言的概率模型，还有更加深刻的哲学考虑，那就是，科学的目标是解释世界，提供关于研究对象的洞见，探索事物为何是它现在这个样子（why）；而不是描写世界、模拟现象或为事实建模，描述事物怎样成为这个样子（how），以取得工程上的成功。对此，Norvig（2011）的评论是：科学和工程是互相成就的，工程上的成功可以作为科学上的成功模型的证据。科学是事实和理论的混合体，理论不能过分凌驾于事实之上。科学发展史是一个不断积累事实的过程，语言学也不应例外。为了佐证自己的观点，Norvig（2011）调查了当时最近的期刊《科学》（*Science*）和《细胞》（*Cell*）上的文章，还有 2010 年诺奖中物理、化学、生理和医学方面的工作，结论是：这些研究工作 100% 地重视“精确地模型化这个世界”（即 how），胜过“提供对这个世界的解释”（why）。当然，他承认区分这两种工作有时并不容易。

Norvig（2011）还指出，语言学中的传统模型往往是离散的、范畴化的、定性描述的。比如，一个词要么是动词，要么不是，并没有关于它的“动词性”（verbiness）的量化程度的描述。但是，母语

者对很多短语和句子的合语法性，也有拿不准的时候。比如：

（10）The earth quaked.

（11）? It quaked her bowels.

quake 这个动词一般在词典中都标记为不及物动词。因此，根据基于范畴的形式语法模型，上面的例（10）是合语法的，而例（11）是不合语法的。但是，例（11）这样的句子又确实出现在英语中。于是，基于范畴的形式语法模型就陷入了两难的困境：接受或排斥例（11）都有问题。像这样的问题，在概率语法模型中就不存在困难。只需要说 quake 的不及物用法概率很高、及物用法概率很低，就可以了。可见，概率化的训练模型可以比范畴化的无训练模型更好地表示自然语言。正因为合语法性并不是范畴化的、确定性的判断，而是概率性的，所以，花时间观察真实语料中的例句是值得的。观察跟直觉在科学研究的历史中并不相悖。只不过，从来都是观察，而不是直觉，在科学研究中占据主流地位。

考虑到 ChatGPT 是一个基于统计的概率模型，我们尝试检测它对于例（11）的合语法性的判断能力：

笔者：请把英语句子“It quaked her bowels.”翻译成汉语。这个英语句子可以说吗？表达什么意思？符合语法吗？

ChatGPT：这个英语句子“It quaked her bowels.”在语法上是合法的，但是这样的表达并不常见，而且可能被认为是相当生硬或过于直白。这句话的字面意思是“它震动了她的肠道。”，但这种表达方式通常不是文学作品或正式场合中常用的表达方式。在文学或书面表达中，人们更倾向于使用更为文雅或隐晦的方式表达相似的感觉，以避免使用过于直接或生理的描写。因此，这个句子在某些语境下可能会引起不适或被认为是不雅的。如果你能提供更多上下文，我可以帮你找到更适当、更自然的表达方式。

可见，就此而言，基于概率的语法模型比基于范畴的语法模型更加灵活；我们实在没有理由把对自然语言的“精确描写”和“科学解释”对立起来。并且，在科学探索的过程中，描写事实通常比理论解释更加基本和重要。比如，进化论的奠基人达尔文以创立了富有洞察力的理论而闻名；但是，他更强调“精确描写”的重要性。物理学家费曼也说过：“物理学可以不需要证明而进步，但没有事实则不可能进步。”（转引自 Norvig 2011）

六、“原则与参数”范式下的范畴语法及其困境

乔姆斯基一向追求语言学理论的简洁与优美，而刻画语言数据的统计概率模型在数学上势必是非常复杂的，因此，他从心底里不喜欢基于统计概率的语言模型。乔姆斯基早期的语法理论强调语言是一个受规则支配的系统，后来又明确区分语言能力和语言运用两个层面：前者指语言使用者由遗传获得的内在的语言知识，其理论概括就是普遍语法；后者指语言能力在一定语境中的具体实现，其外在表现就是语言数据。他认为语言学应该研究语言能力，普遍语法可以理论化为数目有限的原则与参数（详见 Chomsky & Lasnik 1993）。比如，在代词脱落（pro-drop）这个参数上，西班牙语的取值是 1（即“真”，true），而英语的取值是 0（即“假”，false）（详见 Chomsky 1981）。因此，表示“我饿了”的意思，英语必须说“I’m hungry”，代词主语不能省略；而西班牙语中必须说“Tengo hambre”（字面上相当于“have hunger”），做主语的代词 Yo 脱落了。依此类推，如果我们可以找到描述所有语言的为数不多的一系列参数，并且确定每个参数的具体取值，那么我们就真的理解了语言了。

对此，Norvig（2011）指出，问题是语言的现实情况比这个理论要杂乱得多。其实，英语中也有代词脱落现象。例如：

“Not gonna do it. Wouldn’t be prudent.” (Dana Carvey, impersonating George H. W. Bush)

“Thinks he can outsmart us, does he?” (Evelyn Waugh, The Loved One)

“Likes to fight, does he?” (S.M. Stirling, The Sunrise Lands)

“Thinks he’s all that.” (Kate Brian, Lucky T)

“Go for a walk?” (countless dog owners)

“Gotcha!” “Found it!” “Looks good to me!” (common expressions)

语言学家可以为如何解释上面这些现象争论个没完没了。但是，语言的多样性似乎远比用布尔值（true or false）来描述代词脱落的参数值要复杂。一个理论框架不应该把简单性置于反映现实的准确性之上。可见，离散的范畴语法对于语言参数的取值一般是正反二元对立的，没有给连续性的概率取值留下任何空间。

Norvig（2011）分析了乔姆斯基为什么要这样做的原因。第一，他的哲学理念是：我们应该关注深层的“为什么”（why，比如：为什么只有人类才具有语言能力？），只解释表层的现实（how, 比如：我们听到和看到的单词、句子或人际交流等语言运用）是不够的。第二，他一直把注意力放在了语言的生成性上。从这个方面来说，非概率性的理论是合理的。如果他把注意力放在语言的另一面“理解（解释）”上，那么他或许会改变他的说法。在“理解”这一面，听话人需要对收到的信号进行消歧，决定哪种可能的解释概率最高。[①] 这很自然地会被看作一个概率问题。语音识别的研究者是如此看待对语音的解释的，其他领域的研究解释的科学家也是如此的。第三，他更喜欢把语言学看作数学。乔姆斯基（Chomsky 1965：4）说：“语言学理论是心理的，关心的是比实际行为更基础的心理现实。观察语言的实际应用或许可以提供一些证据，但是并不能构成语言学的主题。”其背后的担心可能是：如果关注语言运用和采用统计模型，那么就会让语言学成为一门经验学科，而不是形式科学的数学。但是，我们无法想象物理学家拉普拉斯（Pierre-Simon Laplace）会说，观察行星的运动不能构成轨道力学的主题。物理学家会研究理想的、从实际世界中抽象出来的力学（比如，忽略摩擦力），但是这并不意味着摩擦力不能成为物理学的研究主题。

Norvig（2011）的结束语同样是发人深省的：

> 语言是复杂的、随机的、不确定的生理过程，受到进化和文化变迁的影响。构成语言的不是一个外在的理想实体（由少量的参数设定），而是复杂处理过程的不确定的结果。因其不确定性，用概率模型来分析语言就是必然的选择。

显而易见，语言现象的实际情形（真相）远比任何语法学理论模型复杂。当离散的、非此即彼的、擅长定性描述的范畴语法，遇到复杂的、随机的、不确定的语言现象时，难免会捉襟见肘。因此，基于规则的自然语言处理路径被基于统计概率的语言模型取代，也就是不可避免的大趋势了。

① 例如，美国一个机场的301号航班正在推离入口准备起飞时，被控制塔管制员命令折回。原来，有位乘客无意中跟他的飞行员朋友打了声招呼“Hi, Jack!”；但是，从座舱传到控制塔时，被管制员听成了让航空人胆战心惊的致命单词：“Hijack!”（劫机！）。于是，引发了紧急行动。见Gazzaniga（2009）中译本，第335页。可见，同一种语音形式，有时对于不同的人来说，可能有不同概率的语义解释。

七、智识上的分歧：句法优先还是语义优先？

在认知科学界，不赞成乔姆斯基语言学路线的大佬也不乏其人。比如，同为麻省理工学院著名教授的计算机科学家马文·明斯基。[①] 不知道是出于个人认识还是政治因素，明斯基对乔姆斯基的语言学理论颇有微词。说起来，这两位科学家都精通数学，都提出了有关基本心理过程的理论；并且，他们都在 1950 年代后期推动了认知科学和人工智能的兴起和繁荣，都称得上是认知科学和人工智能的重要奠基者。但是，也许是出于深刻的智识上的分歧，也就是对适合于理解心智的目标和方法存在着根本性的差异，他们对于语言学首先应该干什么，有着截然不同的见解：明斯基更加关注的是语言所能实现的功能，而不仅仅是它的结构。因此，他埋怨乔姆斯基太过专注于句法，以至于一度几乎将语义问题完全排除在语言学之外；结果，一个重要的研究领域（语义学）被一个相对不重要的领域（句法学）取代了。对于明斯基来说，这是根本没理解问题的正确起点（即甚至没有提出正确的问题，“not even have the problem statement right”）。他认为乔姆斯基被抽象的数学所迷惑，而忽略了更有趣、更实质的问题，即意义和机制是如何相互关联的。[②]

明斯基的议论在很大程度上触及了语言学和 / 或语言信息处理是句法优先还是语义优先的问题。就语言信息处理而言，鲜有倚重句法的模型或系统获得成功的先例；而像 ChatGPT 等语言大模型成功的关键是倚重语义，特别是通过基于分布式语义学的词向量嵌入表示〔详见袁毓林（2022，2023）及其所引文献〕。这从工程应用的角度，给语言学理论研究的取向和侧重点的选择，提供了一个反思的维度和衡量得失的参照。

不过，对于乔姆斯基来说，语言是一个心智的计算系统，其主要功能是思维，而不是交际。他坚持认为，交际是语言的次要的、附带的功能（详见史有为 2022）。因此，语言学的研究对象不是语言运用，即我们能看到和听到的单词、句子或人际交流行为；而是人类内在的语言能力，即一种潜在于人类心智中的普遍语法，是所有语言共享的一种几乎肯定是通过进化而融入我们的生物学的结构。他假设这种结构的核心和基本特征是递归，即能够无限地在短语中嵌套短语，从而表达思想之间的复杂关系（比如“汤姆说 | 丹声称 || 诺姆相信……”）。因此，乔姆斯基的生成语言学必然是句法优先的。

但是，挑战性的事实是人类学家和田野语言学家丹尼尔·埃弗雷特（Daniel Everett）发现了一种亚马孙语言，即皮拉罕语（Pirahã），它不具备递归性。例如：[③]

（12）a. 男人打我。男人坏。

　　b. 打我的男人坏。

（13）a. 月亮是绿奶酪做的。彼得说。约翰说。

　　b. 约翰说彼得说月亮是绿奶酪做的。

（14）a. 你喝酒。你开车。你进监狱。

---

① 马文·明斯基（Marvin Lee Minsky，1927 ～ 2016），麻省理工学院人工智能实验室的创始人之一，出版过多部人工智能和哲学方面的著作。1969 年，因在人工智能领域的杰出贡献而荣获图灵奖。

② 详见 Marvin Minsky - My opinion of Noam Chomsky's theories (33/151) -YouTube; https://hyperphor.com/ammdi/Marvin-Minsky%E2%88%95vs-Chomsky；https://www.theguardian.com/us-news/2023/may/17/jeffrey-epstein-noam-chomsky-bard-college-president；https://www.independent.co.uk/news/world/americas/jeffrey-epstein-documents-what-to-know-b2473583.html；中文介绍，请看《两位认知科学 / A.I. 大佬出现在爱泼斯坦文件中》，微信公众号“摩登语言学”，2024-01-06，https://mp.weixin.qq.com/s/eRP-BiVGupgjMcPc5UtF0g。

③ 例（12 ～ 14）分别根据 Everett（2017）中译本第 25、88、257 页上的例子改编；例（15）引自知乎“皮拉罕语”，https://www.zhihu.com/question/496695904。更加详细的调查，请看 Futrell et al.（2016）。

b. 如果你喝酒开车，那么你会进监狱。

（15）a. 给我带一些钉子回来。丹买了那些钉子。它们都是一样的。

b. 把丹尼尔买的钉子给我带一些回来。

在皮拉罕语中，没有 b 这种嵌套式的递归结构，只能说成 a 这种平铺开来的一组句子。正如 Futrell et al.（2016）所指出的：有些现代人类语言的层级结构低于乔姆斯基推测的层级结构。

对此，信从乔姆斯基理论的人当然可以理直气壮地回应称：即使皮拉罕语没有递归，这对普遍语法理论也毫无影响。因为这种能力是内在的，即使它并不总是被利用。正如乔姆斯基及其同事在一篇合著论文中所说："我们的语言能力为我们提供了构建语言的工具包，但并非所有语言都使用所有工具。"（Fitch et al. 2005）对此，Okrent（2017）敏锐地指出，这关系到的不是埃弗雷特对乔姆斯基理论的挑战，而是乔姆斯基对科学方法本身的挑战。因为，根据哲学家卡尔·波普尔（Karl Popper）的可证伪性准则：理论除非具有被证伪的潜在可能性，否则就不是科学。[①] 如果你声称递归是语言的基本特征，并且无递归语言的存在并没有推翻你的主张，那么还有什么可能使它无效呢？是啊，一个不能被事实反驳的理论，还称得上是一种科学理论吗？语言学还到底是不是一门科学啊？该不会是一门哲学或一种宗教吧？这是不是有一点儿细思极恐啊？

2007 年，在接受 Edge.org 的采访中，埃弗雷特说，他给乔姆斯基发了一封电子邮件："普遍语法做出了什么我可以证伪的单一预测？我怎么测试它？"据埃弗雷特称，乔姆斯基回答说："普遍语法并不做出任何预测。它是一门研究领域，就像生物学一样。"（引自 Okrent 2017）如果情况属实，那么就彻底刷新了我们对于普遍语法的固有认知：普遍语法是一种关于语言能力的理论。难道这是为了逃避可证伪性测试而临阵换将和改旗易帜吗？

事实上，不仅是递归，而且"合并"（merge）这种普遍语法的句法操作，即把两个成分组合成一个更大的成分，也不是所有语言都采用的。比如，埃弗雷特和芭芭拉·克恩（Barbara Kem）宣称：合并理论对亚马孙的瓦里语（Wari）做出了错误的预测；语言学家雷·杰肯道夫（Ray Jackendoff）和伊娃·维滕堡（Eva Wittenburg）宣称：在印尼廖内语（Riau）中寻找合并操作是徒劳的（详见 Everett 2017）。这说明，合并可能是人类语言的一种重要的类似二进制的句法操作，但是不一定是人类语言的必不可少的结构基础。反过来，对于语言来说，更加重要的是语言形式的意义和交际者之间的互动，而不是语言的结构及其抽象的运算方式。也就是说，语义研究可能比句法研究更加重要。

八、结语：我们能够从中得到什么教训？

技术进步的速度常常会超出业内资深专家们的估计。比如，在距今不远的那些岁月中，曾经有过下面这些信心爆棚、信誓旦旦的预判（引自 Jason 2024）：

1. 高速行驶的铁路火车是不现实的，因为，乘客会由于车速太快不能呼吸，窒息而死。——狄奥尼修斯·拉德纳（1793 ～ 1859），自然哲学、天文学教授，伦敦

2. 折腾交流电是浪费时间，人们永远也不会使用它。——托马斯·爱迪生，1889 年

3. 马匹不会过时，而汽车只是流行一时的新奇事物。——美国密歇根州储蓄银行总裁，1906 年

① 根据 Popper（1968），科学家必须预先说明，在什么实验条件下他将放弃自己的甚至最基本的假设。即事先立下反驳的标准：如果哪一种状况真的被观察到了，就意味着他的理论被反驳了。并认为，衡量一种理论的科学地位的标准是它的可证伪性或可反驳性或可检验性，一种不能用任何想象得到的事件反驳掉的理论是不科学的。详见中译本第 49 ～ 52 页和第 54 页的注 2。

4. 全世界所需要的计算机大概是……五台。——IBM 公司，1943 年

5. 过了一开始的 6 个月，电视就不会再有任何市场了，人们很快就会厌倦每天晚上盯着一个胶合板做的盒子。——二十世纪福克斯高管达里尔·扎纳克，1946 年

6. 人们没有理由想要在家里拥有一台电脑。——数字设备公司总裁肯·奥尔森，1977 年

7. 移动电话不会取代固定电话。——马蒂·库珀，1981 年

8. 我预测互联网很快将成为壮观的超新星，而到 1996 年就会遭遇灾难性的崩溃。——罗伯特·梅特卡夫，1995 年

9. 不支持 3G，造价高，而且连最起码的摔落测试都没能通过，不太可能对诺基亚构成威胁。——诺基亚工程师对第一代 iPhone 的评估报告，2007 年

后来的现实当然是：这些预言家被事实无情地啪啪打脸，有时简直让当事人无地自容。同样，当许多语言学家以为基于统计的概率模型无法真正刻画自然语言时，ChatGPT 等基于统计概率的大语言模型却大获成功。从中，我们语言学家能够得到哪些经验和教训呢？

关于经验和教训，粗略地说，至少有下面几点。（1）ChatGPT 等现代大语言模型基于深度神经网络，在词语的嵌入式向量表示和转换器的注意力机制等的加持下，能够超越马尔可夫过程模型的有限状态的转移网络，来捕获语句中不同词语之间长距离的依存关系，从而达到接近于人类水平的语言生成与理解。（2）ChatGPT 等现代大语言模型基于海量文本语料的训练，通过词向量进行语言上下文关系等知识的压缩，能够隐式地学习基本的句法和语义知识，从而能够理解、识别和生成“Colorless green ideas sleep furiously.”之类经典的句法合格但语义异常的句子。（3）对语言的“精确描写”和“科学解释”并不对立，并且前者比后者更加重要，因为对语言的科学解释必须建立在对语言的精确描写的基础上。比如，ChatGPT 等现代大语言模型通过学习海量文本语料中的词语与句式的概率分布，相当于达到了对某种语言的精确描写，以至于连人类文本中的各种偏见和刻板印象都习得了；然后通过集束搜索等采样解码策略来预测下一个词语，最终达到语言的生成和理解〔居然还是通过生成来达到理解（详见 Radford et al. 2018）〕。因此，从某种意义上讲，现代大语言模型本身就可以看作一种关于语言运用的科学理论。[①]（4）人类自然语言由众多的社会成员使用，内部难免参差不齐，对于有关句子的合语法性和可接受性也不会有整齐划一的标准。因此，生成语法学的“原则与参数”范式下的范畴语法，对于描写人类自然语言肯定有不可克服的困难。（5）从语言的交际功能这种实际用途出发，无论是语言生成还是语言理解，都是以意义为中心的；相应地，语法学的研究取向可能不应该是“句法优先”，而应该是“语义优先”。只要想一下在人类进化的漫长征途中语言的形成过程，就可以明白：首先得有一批人类文化所创造的概念（意义）跟社会认同的形式（语音）相结合的象征符号，然后才有怎样让多个象征符号合并和组合成符号串的句法。[②] 正如 Luuk & Luuk（2014）所指出的：句法最初是从符号连接开始发展的，然后从单纯的连接发展到嵌入语法。（6）语言学家从语言大模型的成功中获得的最大经验与教训是：对能够直接观察的语言事实（我们每天都说的单词、句子等）的准确描写，远比对不能直接观察的语言能力及其本质（一种特定于语言和人类的抽象特性等）的解释更为基本。前者可以用语料和大语言模型来验证并且支持有关的教学和工程应用，后者则不容易证伪并且

① 具体的论证和说明，详见袁毓林（2024）。人称“神经网络教父”的辛顿（Geoffrey E. Hinton），在牛津大学的演讲（Hinton 2024）中有下列令人警醒的说法：大型神经网络仅仅通过学习大量的文本，就能无师自通地掌握语言的语法和语义；乔姆斯基曾说语言是天赋而非习得的，这很荒谬；他曾经做出了惊人的贡献，但他的时代已经过去了。

② 详见 Everett（2017），中译本第 82 ～ 98、245 ～ 265 页。

容易陷于“确认偏差”（confirmation bias，即倾向于发现有利于自己先前所持的信念、假设或理论的证据，而忽略对自己不利的证伪性数据、事实或理论）。当然，也不排除这样一种可能性：前者囿于语言的表面现象而失去关于语言本质的洞察力，后者则开辟了一条通向认识语言（和人类及其思维）本质的光明大道。另外，在计算技术和数学模型面前，我们应该保持足够的谦逊！因为，语言技术变革的速度、效力和前途，可能会比人们通常预想的更快、更高和更远！

参考文献

史有为　2022 《从乔氏对答谈语言的思维功能》，微信公众号“西去东来中传站”，2022−10−25。

袁毓林　2022 《在人类生境约束下思考语言的设计原理和运作机制》，《语言战略研究》第 6 期。

袁毓林　2023 《超越聊天机器人，走向通用人工智能——ChatGPT 的成功之道及其对语言学的启示》，《当代语言学》第 5 期。

袁毓林　2024 《ChatGPT 等大型语言模型对语言学理论的挑战与警示》，《当代修辞学》第 1 期。

Chomsky, N. 1956. Three models for the description of language. IRE Transactions on Information theory (Volume 2, Issue 3), 113–124. Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA.: MIT Press.

Chomsky, N. 1969. Some empirical assumptions in modern philosophy of language. In *Philosophy, Science and Method: Essays in Honor or Ernest Nagel*. St. Martin’s Press.

Chomsky, N. 1981. *Lectures on Government and Binding*. De Gruyter.

Chomsky, N. 2023. The False Promise of ChatGPT. New York Times, Mar. 8, 2023.（《乔姆斯基：ChatGPT 的虚假承诺》，微信公众号“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》，微信公众号“机器之心”，2023−03−10，https://mp.weixin.qq.com/s/MyiLZYE_hcL27i_qtm7lSA。）

Chomsky, N. & H. Lasnik. 1993. Principles and parameters theory. In J. Jacobs, A. von Stechow, W. Stemefeld, & T. Vennemann (Eds.) Syntax: An International Handbook of Contemporary Research. Berlin: de Gruyter.

Everett, D. L. 2017. *How Language Began: The Story of Humanity’s Invention*. New York • London: Liveright Publishing Corporation.（《语言的诞生：人类最伟大发明的故事》，何文忠，樊子瑶，桂世豪，译，北京：中信出版集团，2020。）

Fitch, W. Tecumseh, M. D. Hauser, & N. Chomsky. 2005. The evolution of the language faculty: clarifications and implications. *Cognition* 97(2): 179–210.

Futrell, R., L. Stearns, D. L. Everett, S. , et al. 2016. A Corpus Investigation of Syntactic Embedding in Pirahã. *PLoS One*. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0145289.

Gazzaniga, M. S. , R. B. Ivry & G. R. Mangun. 2009. *Cognitive Neuroscience: The Biology of the Mind* (3rd Edn.). W. W. Norton & Company, Inc.（《认知神经科学——关于心智的生物学》，周晓林，高定国，等，译，北京：中国轻工业出版社，2011。）

Hinton, E. G. 2024.《数字智能会取代生物智能吗？》，2024 年 2 月 19 日于牛津大学的公开演讲，卫剑钒，编译，微信公众号“卫 sir 说”，2024−03−08，https://mp.weixin.qq.com/s/u72sPc0PxwIaQBFK-TCKJw。

Jason. 2024.《从 GPT-5 是什么说起》，微信公众号“信息平权”，2024−01−21，https://mp.weixin.qq.com/s/1uKg8QulI7AmWLqS9Q8toA。

……

（因版面不足，以下参考文献从略，可在中国知网上阅读、下载完整版）

责任编辑：王　飙