

由辛顿的获奖感言反思语言学研究的路向 *

袁毓林

(澳门大学人文学院中国语言文学系 澳门 999078)

[摘要] 本文由辛顿在获诺贝尔奖感言中对乔姆斯基语言学的批评,反思语言学研究的路线与方向。首先,扼要引述辛顿在多个场合对乔姆斯基的批评,就此提出两个问题:(1)语言大模型的成功能不能颠覆语言天生理论?(2)生成语法理论有没有把语言学的方向带偏了?然后通过简介 Piantadosi (2023) 和进化神经生物学等有关观点和实验,说明生成语法的研究范式不符合人类朴素的心智。接着,简单比较近 70 年来人工智能和生成语言学的发展历程,说明学派竞争和客观评测是推动学术进步的有效路径。最后,通过分析乔姆斯基对其学术思想及其历程的自述,说明语言天生理论和普遍语法虽然不易证伪,但是更难证明其成立。

[关键词] 辛顿;乔姆斯基;语言大模型;语言天生理论;生成语法;普遍语法

[中图分类号] H0-0 [文献标识码] A [文章编号] 1003-5397 (2025) 01-0086-14

DOI:10.16499/j.cnki.1003-5397.2025.01.005

The Routes and Directions of Linguistic Studies: Reflections on Geoffrey E. Hinton's Criticism of Chomskyan School

YUAN Yulin

Abstract: This paper reflects on Geoffrey E. Hinton's criticism of the Chomskyan linguistic theory following his acceptance of the Nobel Prize, highlighting the routes and directions of linguistic studies. First, a brief overview is provided of Hinton's critiques made on several occasions of the Chomskyan linguistic theory and the following two key questions are posed: (1) Can the success of large language models (LLMs) overthrow the principles of linguistic innatism? (2) Has the theory of generative grammar misled the pursuit of linguistic research? Next, we introduce relevant perspectives from Piantadosi (2023) and findings from evolutionary neurobiology to demonstrate that the generative grammar paradigm does not align with the mentality simplicity of human beings. We then compare the developments in artificial intelligence and generative grammar over the past 70 years, illustrating how competition among different schools and objective evaluating can drive academic advancement. Finally, we analyze Chomsky's personal reflections on his academic thoughts and their evolution, highlighting the challenges of falsifying linguistic innatism and universal grammar, which are often even more difficult to verify.

[收稿日期] 2025-01-03

[作者简介] 袁毓林,澳门大学人文学院中国语言文学系教授,主要研究理论语言学和汉语语言学。

* 本文得到澳门大学讲座教授研究与发展基金(CPG2025-00008-FAH)和启动研究基金(SRG2022-00011-FAH)支持,还承蒙周韧教授和孙晓雪博士指正,谨此致以谢忱。

Keywords: Geoffrey E. Hinton; Chomsky; Large Language Models (LLMs); linguistic innatism; generative grammar; universal grammar

一 辛顿获奖引发的语言学路向的反思

众所周知,2024年的诺贝尔物理学奖得主辛顿(Geoffrey E. Hinton),是深度学习(deep learning)人工智能的奠基人。他称自己1985年建立的使用反向传播的小型神经网络语言模型,是现代大型语言模型(Large Language Models,简称LLMs)的先驱(韩佳琦,2024)。而语言大模型(LLMs)是本轮人工智能大爆发的主要的技术推动力。比如,大家熟知的ChatGPT的基座就是语言大模型GPT-3.5;并且,正是2022年底ChatGPT的爆火,使得人工智能得以走出实验室,进入大众视野和寻常百姓的电脑桌面。

为了“表彰他们通过人工神经网络实现机器学习的奠基性发现和发明”,2024年10月8日,瑞典皇家科学院授予霍普菲尔德(John J. Hopfield)和辛顿诺贝尔物理学奖。一石激起千层浪,许多人不理解“为什么诺贝尔物理学奖居然授予人工智能专家?”或者质疑“他们凭人工智能的成绩怎么能拿物理学奖呢?”。这种争议本来不足为怪,但辛顿获奖居然在语言学界引起了广泛关注和讨论。因为,语言大模型是为人类自然语言建立的一种计算模型,其中涉及设计者对于语言结构和功能的基本假设和独到认识;并且,这种认识不同于主流的语言学理论。长期以来,语言学家乔姆斯基认为语言是人类先天获得的,语言的递归性等复杂的结构特点,远远超出了马尔可夫过程模型的刻画能力,也是机器无法通过统计分析来学会的。但事实是:基于概率统计和马尔可夫过程原理的ChatGPT等语言大模型,在语言生成和理解方面的表现却接近甚至有望超越人类的水平。

可能正是有了这种技术上的突破和工程上的成功,2024年10月8日,辛顿在回应诺奖官网记者Adam Smith采访时,直言不讳(在某些人看来是“大言不惭”)地说:“语言大模型比很多人想象的更接近于真正的理解,乔姆斯基对语言大模型的理解能力的判断是错误的,大模型的成功驳斥了语言学家的理论,……”。^①这种对于影响国际主流语言学近70年的生成语言学的批判,自然会引起语言学界深刻反省:主流语言学的路线与方向有没有搞错?其实,诸如此类的批评,已经不是第一次,也不是第二次,而是好多次了。

二 辛顿对乔姆斯基的多次批评与我们的两个问题

此前,辛顿已经多次在讲演中直截了当地批评乔姆斯基的语言观。比如,在2023年10月的“圆桌论坛:AI时代的智能研究”上,辛顿火药味十足地说:^②

我认为这些大型语言模型已经结束了这种情况,虽然在乔姆斯基的脑海中可能还没有结束,但在几乎所有其他人眼中已经是事实。……值得一提的是,我从本科时期就一直觉得乔姆斯基是完全错误的。

乔姆斯基,一个在麻省理工学院的怪才,一直声称语言完全是天生的,但我们现在知道

情况并不一定如此。回顾乔姆斯基关于语言的理论，他的看法显得有些疯狂；因为语言的核心是传达意义，而乔姆斯基似乎忽略了这一点。就好比你想了解一辆汽车的工作原理，绝大多数人会首先考虑发动机是如何让汽车运转的。但乔姆斯基似乎更关心的是：为什么有三轮车和四轮车，却没有五轮车。这种思维方式类似于他对语言的理论，他专注于为什么某些句法结构不可能出现，而忽视了语言如何表达意义这个基本问题。

2024年2月19日，在牛津大学的演讲《数字智能会取代生物智能吗？》中，他说：^③

乔姆斯基曾说语言是天赋而非学会的，这很荒谬。……一个没有先天知识的大型神经网络仅仅通过观察数据就能实际学习语言的语法和语义，……。

2024年春，在爱尔兰都柏林大学接受尤利西斯奖章的获奖感言中，辛顿(2024a)说：

语言学家被一个叫乔姆斯基的人误导了好几代……。他有一个偏执古怪的理论，即语言不是学会的。他成功地说服很多人相信这一点。这个说法显然是一派胡言。语言显然是学会的。大型神经网络学习语言，不需要任何先天结构，只是从随机权重和大量数据中开始学习。乔姆斯基却仍然在说，但这并非真正的语言，这不算数，这是不对的。

在2024年5月的一个访谈中，辛顿(2024b)说：

乔姆斯基的观点是，除非所有的知识都已经深深植根并且已经成熟，否则你无法学习任何像语言这样复杂的事物。然而，这种观点现在显然是荒谬的。

辛顿这种对乔姆斯基“不依不饶、深揭猛批”的做派，跟乔姆斯基对统计语言模型的“锲而不舍、穷追猛打”的作风堪称匹配^④。不过，现代大型语言模型可以“再造人类语言”（recreate natural language），说明基于统计的办法是可以刻画和理解自然语言的。辛顿说“乔姆斯基误导了几代人”听上去非常刺耳，但是也值得我们语言学界的同仁思考和反省。因此，在2024年10月17日由《当代语言学》和《语言战略研究》联合举办的“大语言模型与语言学发展座谈会”上，我作了题为《这是语言统计技术的胜利，也是语言天生理论的失败》（袁毓林，2024a）的发言；在发言的结尾处，我斗胆提出了两个需要思考和讨论的问题：

- (1) 基于统计的语言大模型的成功，能不能颠覆语言天生理论及普遍语法理论？
- (2) 乔姆斯基生成语法理论及有关主张与研究范式，有没有把语言学的方向带偏了？

当时限于时间，没有展开来讨论。本文则将更深入地介绍前贤和笔者对这两个问题的意见，并且说明基本的理由。

三 语言大模型的成功能颠覆语言天生理论？

关于第一个问题：“基于统计的语言大模型的成功，能不能颠覆语言天生理论及普遍语法理论？”我们首先想到的是Piantadosi（2023），他通过对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。袁毓林（2024b）则以此为基础，围绕着下面三个问题进行比较详细的介绍和评述；现在，我们就下面三个问题，撮要简述如下：

- (1) 现代大型语言模型能不能看作是一种语言理论？

我们的回答是：在语言大模型这里，用于预训练的转换器是以词向量为输入的；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，使得语言大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。但是，如果像 Piantadosi (2023) 那样，认为 GPT-3 之类大型语言模型，是人类语言认知的一种严肃、合理和优良的科学理论，则可能是有问题的。因为诸如此类的语言模型的不同构架，更多的是出于工程方面的考虑。

(2) 现代语言模型及其工作原理能不能禁住语言学家的质疑？

我们的回答是：大型语言模型禁得住乔姆斯基等语言学家的质疑。因为，在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。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处的语言环境中，只需要为数不多的语言刺激（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）、大算力和强算法，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，其表现远远地超出了生成语法学的理论假设。从乔姆斯基关于评价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，它拥有这么多文本语料，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流畅的段落。对此，我们无法否认它“充分地观察”了这么多语料，并且通过庞大的参数和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能完全知道它的机理，就像科学家不知道人脑是怎么涌现出语言能力一样。其实，这对于人类认识和科学历史来说并不奇怪。正如杰出的物理学家理查德·费曼 (Richard Phillips Feynman) 所说的：虽然量子力学是现代人类历史上最伟大的知识成果，但是包括这一领域的奠基人在内，没有人真的理解量子力学。

(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？

我们的回答是：正如 Piantadosi (2023) 所指出的，现代大语言模型的确削弱、反驳了乔姆斯基关于语言学的一些主要主张。比如，乔姆斯基提出这样的观念：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。但是，首先，像 GPT-3 这样的大型语言模型在大量文本数据上进行训练，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和跟他人互动而学到和发展的。第二，大语言模型在执行

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规则的观点。相反,它表明语言是一个学到的和适应的系统,可以通过机器学习算法进行建模并改进。最后,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广泛的话题的语言。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。

值得一提的是,跟乔姆斯基一样,心理语言学家 Steven Pinker 也认为人类大脑中有一个专门的语言学习模块。但是,Pinker (1994)指出:

虽然乔姆斯基因为提出“语言先天假说”而著名于世,但他从来没有对这个观点展开过系统的科学论证,而他的主要论据——“输入贫乏理论”也远非无懈可击。借助大型在线语料库(这是语言学中最为重要的新方法),杰弗里·普勒姆和哲学家芭芭拉·肖尔茨(Barbara Scholz)揭示说,许多据称是儿童从未听过的句式,其实都可以在合理大小的英文样本中找到例子。他们并没有否认输入贫乏的可能性(我认为西蒙案例、尼加拉瓜手语以及第4章中彼得·戈登的“mice-eater”实验都是很好的例证),但他们指出,要确证这一说法,比乔姆斯基和他的追随者所设想的要难得多。^⑤(参见中译本第455页)

也就是说,支持“语言天生假设”的关键论据“儿童语言习得时输入贫乏理论”,本身是需要证明的。目前,它还只是一个有待求证的假设;但是,它却被用来支持另外一个重要且激进的假设。

此外,Beguš 等(2023)报道,为了揭开人工神经网络学习的神秘面纱,看看“神经网络的学习方式是否跟人类相同”。他们设计了一项实验:给人类听一种简单的声音,然后收集人类听到声音后产生的脑电波。同时,把同一种声音输入到神经网络中,分析神经网络所产生的信号。他们对两者进行比较,发现结果竟然出奇相似。他们测试了适用于各种任务的通用神经元组成的网络,发现这些人工神经网络信号跟人脑电波信号非常匹配。这也就说明了这两个系统正在进行类似的活动。比如,说英语的人的脑干对“bah”声音的反应,比说西班牙语的人稍微早一些;而受过英语训练的生成对抗网络(GAN)对相同的声音的反应,也比受过西班牙语训练的模型稍微早一些。人类跟机器的这种反应时间上的差异几乎是相同的,大约是千分之一秒。这也为他们提供了额外的证据,即人类和人工神经网络“很可能是以类似的方式处理语言的”。这项研究的结论,跟乔姆斯基的“人类天生就具有理解语言的能力,这种能力与人脑是硬连接(hard-wired)”的观点不一致;相应地,也不支持乔姆斯基的普遍语法(Universal Grammar)概念,即人脑普遍有一种语言习得机制,使人们能够学会使用语言。^⑥

四 生成语法理论的研究范式是否符合人类朴素的心智?

关于第二个问题:“乔姆斯基生成语法理论及有关主张与研究范式,有没有把语言学的方向带偏了?”我们沿用袁毓林(2023)的观点:

以乔姆斯基为代表的生成语法派相信:人类有独特的语言官能(language faculty),

人类内在性的语言系统是一个不同于其他心理系统的独立的认知系统；因此，必须假设普遍语法这种高度抽象的自主原则系统制约内在性语言，而不是诉诸于语义、交际功能等外部条件。我们则从根本上怀疑形式语法学者设想的由多个 VP-shell 叠床架屋堆砌出来的句法结构具有心理现实性；或者坦率地说，我们朴素简陋的心智还没有合适的工具来处理这种既抽象又繁复的结构。

下面，我们用进化心理学的观点及其两个经典的例子来作进一步的说明。根据 Gazzaniga 等 (2009)，现代人类的大脑是 10 万年前适应更新世 (Pleistocene) 的狩猎和采集的社会环境而进化出来的产物，是为了解决石器时代人类比较简单的生活的。所以，我们在考虑现代人类大脑的功能和能够做什么（包括语言理解和生成）时，应该考虑到我们的大脑只适应解决真实世界的问题。比如，如果你是一个酒吧的保安，你的任务是不让 18 岁以下的人进来喝啤酒。下面四张卡片，一面写着顾客的年龄，另一面写着顾客喝的饮料。那么，你会翻开哪些卡片检查呢？



很显然，你应该翻开 1 号卡片，看看这个顾客是不是 18 岁以下的人；再翻开 4 号卡片，看看这个顾客喝的是不是啤酒；其他则不用理会。因为 2 号卡片是苏打水，谁爱喝就谁喝；3 号卡片是 26 岁的顾客，爱喝啥就喝啥。

接下来是一道稍微抽象一点的逻辑判断题：下面四张卡片，一面是数字，另一面是字母。



请观察上面四张卡片，如果需要你判定下列规则是否有例外的情况，那么你会翻开哪些卡片去验证下面这个规则呢？

如果某张卡片一面是 D，那么它的另一面是 3。

许多人会感觉脑袋发懵，因为这是一个远离日常生活场景的抽象问题，采用的“如果……那么……”这种描述性规则的形式。其实，这个问题的逻辑结构跟上面的问题是一样的，所以答案也是一样的。即翻开 1 号卡片和 4 号卡片来验证。因为：

- (1) 如果 1 号卡片 (字母 D) 的反面不是数字 3，那么上面的规则就有例外情况。这是通过否定后件情况来否定整个蕴涵关系 (Implication) 的成立 (hold)；
- (2) 2 号卡片 (字母 F) 跟上述规则 (的前件和后件都) 无关，不用理会；
- (3) 3 号卡片 (数字 3) 虽然跟上述规则的后件有关，但是也不用理会。因为，这个规则是一种蕴涵关系，肯定了后件并不能肯定前件；并且，假前提反倒是可以推出真结论的，即实质蕴涵怪论 (Material Implication Paradoxes)；
- (4) 如果 4 号卡片 (数字 5) 的反面是字母 D，那么上面的规则就有例外情况。这也是通过否定后件情况和肯定前件情况来否定整个蕴涵关系的成立。

可见，只有1号卡片（字母D）和4号卡片（数字5）需要翻开检查，因为它们可以通过否定后件的办法来验证蕴涵关系，从而检查上述规则有无例外。

心理学家发现人们很善于发现违反群体规则等社会契约的行为，从而抓住潜在的欺骗者。但是，人们却无法发现单纯的违背描述性规则（如果……那么……）的现象。在对美国和德国大学生的研究中，75%到90%的人能正确推理与社会交换有关的事情，而当关注点是抽象的或描述性的规则时，只有4%到25%的人能做到。检测潜在的欺骗者是一个很重要的进化问题：那些不能检测出欺骗者的人会失去重要的资源，从而在群体中处于劣势。相反，检测到描述性规则就没有什么进化意义。也就是说，由于进化所带来的优势，人类善于发现条件规则的潜在违背，尤其是当他们处于某种社会情景中时。尽管如此，进化并没有给我们在检测描述性规则的违背时带来优势。^⑦

所以，我们不太相信后期生成语法理论中诸多抽象的句法核查规则或运算方法是在人脑中可以实际运作的；进而，从根本上怀疑后期生成语法那种愈来愈繁复的句法结构和愈来愈抽象的句法操作是人脑所能够处理和应付的。比如，乔姆斯基的“最简方案”假设人类语言共用一套运算系统，只需要有限的几种运算方法就可以构建所有合乎语法的句子。比如，代词的人称和数有实质性的意义区别，可以假设它们是在词库里取值的；而格的特征纯属入句以后语法形式上的要求，没有实质性的意义差别，可以假设是在词库里尚未取值，是进入句法运算系统以后通过“一致”（Agree）这种运算方法以后才取值的。与此不同，助动词的时制特征有实质性的意义区别，可以假设它们是在词库里取值的；而人称和数的特征纯属入句以后语法形式上的要求，没有实质性的意义差别，可以假设是在词库里尚未取值，是进入句法运算系统通过一致这种运算方法以后才取值的。其中，一致这种运算方法的使用条件如下（徐烈炯，2019）：

(5) 使 α 和 β 一致，当且仅当：a. α 与 β 相配；b. β 在 α 域内；c. α 与 β 均待用；d. 没有 γ 介于 α 与 β 之间。

其中，(5a)要求 α 与 β 在取值方面相配，即两个成分 α 与 β 具有相同的形态特征[F]，而且取相同的值；或者其中一个成分 α 的特征[F]已经取值，而另一个成分 β 的特征[F]尚未取值。(5b)是对 α 与 β 所处结构位置的要求， β 在 α 域内指 α 统制 β 。(5c)指 α 与 β 在“LF”（逻辑式）有不可解读的特征。(5d)指 α 与 β 同处于局部范围内最靠近的位置。以上条件都满足时，启动操作。而这些抽象规则是用以解决类似下面句子中，人称代词的格特征和助动词“BE”的人称和数的特征的取值的。例如：

- (6) a. The policemen arrested HE. → b. The policemen arrested him.
- (7) a. The policemen arrested THEY. → b. The policemen arrested them.
- (8) a. He BE arrested by the policemen. → b. He was arrested by the policemen.
- (9) a. They BE arrested by the policemen. → b. They were arrested by the policemen.

直白地说，(6a)中“HE”的单数和第三人称的特征是从词库里带来的，处于动词过去时形式“arrested”宾语位置，使得代词“HE”只能取单数、第三人称和宾格的特征，把这些特征输入“PF”（音系式），就拼写出（spelled out）“him”这种形式。同样，(7a)中“THEY”

的复数和第三人称的特征也是从词库里带来的，处于动词过去时形式“arrested”的宾语位置，使得代词“THEY”只能取复数、第三人称和宾格的特征，把这些特征输入“PF”，就拼写出“them”这种形式。(8a)中“He”的单数和第三人称的特征是从词库里带来的，“BE”的过去时特征也是从词库里带来的；“BE”的未取值的人称和数特征，通过搜索它所统制的范围，探测到受它统制的代词“He”有单数、第三人称的特征，把这些特征拷贝过来，连同自己固有的过去时特征一起输入“PF”（音系式），就拼写出“was”这种形式。而(9a)中“They”的复数和第三人称的特征也是从词库里带来的，“BE”的过去时特征是从词库里带来的；“BE”的未取值的人称和数特征，通过搜索它所统制的范围，探测到受它统制的代词“They”有复数、第三人称的特征，把这些特征拷贝过来，连同自己固有的过去时特征一起输入“PF”，就拼写出“were”这种形式。这种解释绕了一大圈，也没有传统语法的“语法形态的一致关系”来得通透和简洁。

Pinker (1994)说得更加直截了当和简单明了：

在语言学界，乔姆斯基的一个重要特点便是每隔 10 年就对自己的理论进行一番颠覆性改造。目前的“版本”（指“最简方案”）是 5.2 版（当然，这取决于你如何计算），而我在这一章里描述的是 3.2 版的简化版，即“修正的扩充标准理论”（Revised Extended Standard Theory）。不过，对于任何一位正在研读语言学著作的人来说，本章绘制的语法图都不会陌生，因为我所强调的方面都经得起时间的考验，而且可以很容易地转化为其他理论。在我本人的著作中，我一贯支持比乔姆斯基理论更为平实、自然的理论（更少的分枝、更少的语迹、更少的转换），它们的构架清晰可辨，一目了然，例如短语、词语项和结构。琼·布列斯南（Joan Bresnan）的理论（Lexical-Functional Grammar, LFG）就是一个例子，而这种优点可以在雷·杰肯道夫（Ray Jackendoff）和彼得·库里卡弗（Peter Culicover）的新书《简单句法》（*Simple Syntax*, Oxford University Press, 2005）中找到。^⑧（中译本第 457 页）

总的来说，我们呼唤更加平实和自然的语言学理论，希望相应的语法学构架清晰可辨、一目了然，其中的句法结构包含更少的分枝、更少的语迹和更少的转换。一句话，希望语言学理论及其有关的描写方式能够符合人类朴素的心智。

下面，简单地比较一下都是有着近 70 年历史的人工智能和生成语言学，以便从中获得一些思考与教益。

五 近 70 年人工智能和生成语言学发展历史的比较

早在 1950 年，图灵（Alan Turing）就发表论文《计算机器与智能》（Computing Machinery and Intelligence）提出了机器思维和图灵测试的概念。但是，直到 1956 年 8 月的达特茅斯会议，即“人工智能夏季研究项目”（Summer Research Project on Artificial Intelligence），人工智能才正式进入科学家们研究的议事日程。同年 9 月，在麻省理工学院（MIT）召开的信息论年会上，乔姆斯基发表论文《语言描写的三种模型》（即 Chomsky, 1956），证明了有限状态句法不能表示某类语言，文中还引用了次年出版的《句法结构》（Chomsky, 1957），引起了语言学、心理学和计算机科学等认知科学界的广泛

关注。如果说 1956 年是实际意义上的人工智能元年，那么也可以说 1956 年是乔姆斯基学派的生成语言学元年。

近 70 年来，人工智能经历了几起几落。1958 年，Frank Rosenblatt 发表《感知器：脑的组织和信息存储的概率模型》，提出了感知器模型；同年 John McCarthy 发明了 LISP（表处理语言），成为人工智能的得力研究工具。1960 年，Newell、Simon 和 Shaw 设计通用问题求解机，用以解决多种类型的数学难题；1965 年，Edward Feigenbaum 等研发专家系统 DENDRAL，成为第一套有效进行工作的专家系统；1966 年，麻省理工学院的 Joseph Weizenbaum 推出聊天机器人 ELIZA。但是，1969 年 Seymour Papert 和 Marvin Minsky 出版《感知机：计算几何学导论》(*Perceptrons: An Introduction to Computational Geometry*)一书，认为单层感知机代表的神经网络的容量是有限的，无法处理现实中的复杂非线性问题，直接导致了神经网络研究将近二十年的低潮期。

1970 年，T. Winograd 开发了 SHRDLU 系统，该系统可以部分理解语言。1972 年，斯坦福大学研发机器人 Shakey，这是首台采用了人工智能学的移动机器人，掀起了早期人工智能研究的小高潮。但是，同一年 Hubert Dreyfus 出版著作《计算机不能做什么》(*What Computers Can't Do*)，在一定程度上打击了人们对人工智能研究的积极性。并且，随着 1974 年应用数学家詹姆斯·莱特希尔 (James Lighthill) 对人工智能的负面报告，导致研发资金大幅削减，人工智能研究进入寒冬。1976 年 Allen Newell 和 Herbert A. Simon 提出物理符号系统假设 (Physical Symbol System Hypothesis)，企图建立人工智能的理论体系。

进入 1980 年代，随着知识表示和知识库、符号主义机器学习等技术的兴起，人工智能研究略有复苏。1982 年，John Hopfield 的《具有集体计算能力的新型神经网络和实际系统》，提出了一种具有联想记忆能力的神经网络，后来被称为“霍普菲尔德网络”；1983 年，T. Sejnowski 和辛顿提出了“隐单元”(hidden units)的概念，并且研制出 Boltzmann 机；1986 年，辛顿等出版《并行分布式处理：认知的微细构造探索》，重新提出了简明有效的误差反向传播算法 (Error Backpropagation Algorithm，即 EBP 算法)，解决了多层感知机 (Multi-layer Perceptron, MLP) 的参数训练问题；1987 年，Marvin Minsky 出版著作《思维社会》，向人们描述了大脑中各种不同层次水平的“智能主体”；1988 年，L. O. Chua 提出了卷积神经网络 (Convolutional Neural Network, CNN)。

进入 1990 年代，并行计算和统计分析产生。1995 年，Vladimir Vapnik 等提出了支持向量机 (Support Vector Machine, SVM) 算法，可以分析数据和识别模式，用于分类和回归分析。1997 年，Sepp Hochreiter 和 Jürgen Schmidhuber 提出长短时记忆算法 (Long Short-term Memory, LSTM)。同年，国际商业机器公司 (IBM) 的国际象棋程序 DeepBlue (深蓝) 战胜当时世界排名第一的加里·卡斯帕罗夫 (Garry Kimovich Kasparov)，重新燃起了人们对人工智能研究的热情。2006 年，辛顿出版著作《深层置信网络的快速算法》，提出了深层置信网络 (Deep Belief Networks, DBN)；同年，辛顿和 LeCun 等提出深度学习 (deep learning) 概念。2011 年，IBM 的 Watson 系统在美国热门问答节目《危险边缘》(Jeopardy)

中,击败了该节目有史以来最强大的两位选手肯·詹宁斯和布拉德·鲁特。同年,苹果公司在 iPhone 产品的演示中,展示了一个名为 Siri 的虚拟助手。2012 年,Google X 实验室采用“神经系统”识别出一只猫。

尽管辛顿在 20 世纪 70 年代攻读博士学位时就开始探索神经网络,但直到 2012 年,他和两名研究生在 ImageNet 竞赛中取得了最好的成绩,才使人工智能行业看到了神经网络的发展前景。辛顿对神经网络和深度学习的研究,使得人工神经网络系统通过学习处理大量的数据,能够作出较为准确的预测,这为自然语言处理和语音识别等人工智能工程奠定了基础。2014 年,亚马逊发布了其自主研发的虚拟助手 Alexa。苹果的 Siri 和亚马逊的 Alexa 都具有自然语言处理功能,能够理解语音问题并给出回答。2016 年 DeepMind 实验室(后被 Google 收购)研发的 AlphaGo 围棋程序,击败了当时世界上最优秀的围棋选手之一李世石。2017 年 Google 推出转换器(Transformer)架构,在多层人工神经网络中引入多头注意力机制。2018 年,OpenAI 推出基于转换器的生成式预训练模型 GPT-1,并发表论文《通过生成式预训练提高语言理解能力》,介绍了该模型以及基于转换器的生成式预训练模型(GPT)的总体概念;随后几年,它陆续推出训练参数更多的 GPT-2(15 亿个)、GPT-3(1750 亿个)。直到 2022 年底推出 ChatGPT,引起轰动;2023 年它推出 GPT-4,再次提升了其产品的能力,可以生成更加细致和富有创意的回答,并能参与越来越多的活动(如,通过律师资格考试)。随后,微软将 ChatGPT 整合到其搜索引擎必应(Bing)中,Google 也发布了其 GPT 聊天机器人 Bard,开始了生成式人工智能大发展阶段。^⑨

乔姆斯基的生成语法理论,大概经历了不断自我修正和框架变化的 6 个版本。第一版是以 Chomsky (1957) 为代表的“古典理论”(Classic Theory),创立了转换生成语法,提出语法是一种带有精确和有限的规则的数学系统,可以生成无限的句子。第二版是以 Chomsky (1965) 为代表的“标准理论”(Standard Theory),阐述了语言的心理现实性和语言研究的哲学,比如:语言学研究的目标(观察、描写和解释的充分性);展示了生成语法的一些概念和技术细节,比如:语言能力和语言运用、深层结构和表层结构、语言获得和普遍语法。值得注意的是,标准理论抛弃了古典理论中“核心句”的概念,代之以深层结构;还在整个语法模型中增加了语义部分,使用递归性短语结构规则,并且在词项中引入句法特征来实现次范畴化(subcategorization)。第三版是以 Chomsky (1972) 为代表的“扩充式标准理论”(Extended Standard Theory),从被动转换和量词辖域等方面说明句子的表层结构也影响其语义解释;于是,在这个新版本的语法理论模型中,语义解释是由深层结构和表层结构共同决定的。第四版是以 Chomsky (1977) 为代表的“修正的扩充式标准理论”(Revised Extended Standard Theory, REST),通过在表层结构引入“语迹”(trace)的概念来反映移位成分在深层结构中的位置,从而让表层结构充分反映转换前的深层结构中的有关信息;最终,使得这种经过充实的表层结构直接决定句子的部分语义,接下来再通过逻辑式与其他思维结构的配合来决定整个句子的语义解释。第五版是以 Chomsky (1981) 为代表的“管辖与约束理论”(Government and Binding Theory, GB),假设普遍语

法(UG)是每个人类儿童与生俱来的语言普遍性的基本集合,其中包含许多适用于所有语言的固定“原则”;还嵌入了灵活的“参数”,这些参数必须通过经验来确定。随着人类儿童获得语言经验,他们的大脑利用其掌握的有限语言证据(贫乏的刺激输入)来确定普遍语法的参数,并以非归纳的方式产生儿童第一语言的核心语法。其中,提出了比较可操作的X加杠理论、θ理论、格理论、管辖理论、约束理论、控制理论、界限理论,把这些原则搭配起来,可以解释代词回指等复杂的句法语义问题(徐烈炯,2019)。这期间产生了一大批跨语言的研究成果,堪称是生成语法的黄金时期。

第六版是以Chomsky(1995)为代表的“最简方案”。他在最简方案中指出:人类大脑中存在着语言官能,它使得人类能使用和理解语言。语言官能的一个组成成分是一种生成程序,它生成了语义的、音系的结构描述。这些结构描述就是语言的表达式(expressions),它们是属于语言能力系统中的。由于语言是内包在(embedded)语言运用系统中的,因而可以把表达式看作是对运用系统的一套指令(a complex of instructions)。运用系统似乎可以分为“发音—听觉”(articulatory-perceptual)和“概念—意向”(conceptual-intentional)两个子系统,因此语言表达式应包含对这两个子系统分别作出的指令。对发音—听觉系统的指令就是音系式,对概念—意向系统的指令就是逻辑式。音系式和逻辑式代表语言和语言运用系统的接口层面。出于经济原则的考虑,一个最简的语言学理论方案应该只由概念上必不可少的因素组成。接口层面就是这样的因素,而属于非接口层面的D-结构和S-结构就应该废除。^⑩这个时期差不多是生成语法的高潮期。

通过比较近70年来人工智能研究和生成语法学的发展历程,可以发现:前者道路多歧,既有理论方向的分歧、又有技术路线的差异;但是,最终以相对客观的评测为基准,来裁定高下优劣,逐步走向通用人工智能(AGI)的道路。后者虽然理论框架变化繁多、模型版本更迭不断,但是学派的主心骨稳定,难免存在学派内部不容易出现真正的革新者这种局面。中间除了1960年代乔姆斯基的一些同事和学生(Paul Postal,George Lakoff,James McCawley,John Robert Rose,等等)对标准理论提出异议,逐步形成“生成语义学”(generative semantics),但流行了不到10年外(徐烈炯,2019),其余时间基本上是“一个学派、一个宗师、一个范式、一个框架”的。

六 不易证伪但更难以证成的生成语言学

最后,我们不得不承认:批评乔姆斯基是不明智的,因为他对一些语言学的关键概念和术语的理解和使用是别出心裁的,而且是不断变化的。比如,“普遍语法”(UG)这个概念,在传统语法学中指所有语言都有的共性部分,Chomsky(1965)中曾引用James Beattie(1788)中的一段话“Those things,that all languages have in common, or that are necessary to every language, are treated of in a science,which some have called Universal or Philosophical grammar”,但是,乔姆斯基却旧瓶装新酒,用这个早已为人们“广为接受”的名称,来给他自己的假设(内部语言、语言官能)命名。这种抽换概念含义的策略增加了他的有关理论假设的“迷惑程度”。也正是由于诸如此类的概念含义的飘忽不

定，其有关论断也是不易证伪、但更加难以证明其成立的。为了保险，我们看 Milak 和 Tankosić (2024) 中乔姆斯基自己最近的表述：

50 年代初，……我搬到了哈佛大学，在那里我遇到了莫里斯 - 哈勒 (Morris Halle) 和 埃里克 - 伦伯格 (Eric Heinz Lenneberg)，……。我们三人互相交流，渐渐地对语言的一些假设开始成形。第一个假设是，无论语言是什么，它都是人类大脑的一种生物属性。第二点是，语言的使用具有某种创造能力，这种能力使语言具有创新性和无限性，但又适合环境，尽管不是由环境造成的。

所有语言都有一套共同的属性，即普遍语法 (UG)，……而 UG 的另一种含义就是试图找出它们的相同之处。然而，在生成语法的早期框架内，UG 有了更为专业的含义。它变成了对先天的、由基因决定的语言能力基础的研究。对我来说，这种基础存在的证据是显而易见的：人类会说话，而且他们几乎是反射性地从非常零散的数据中拾取这些语言。孩子们并不试图学习语言，也没有人教他们语言，但语言还是很快就产生了。那么，一定是有某种内在系统使儿童能够从环境中挑选出这些随机噪音，并迅速从中构建出一种语言，这反过来又意味着一定有其先天基础。那么，UG 就是关于这种先天基础的理论，而 (生成) 语法是一个早期的、有点含糊不清的术语，用来描述一系列生成程序，这些程序相当于我们所说的“语言”。……，语言是什么？——语言是由你决定的。一般来说，这种性质的问题是没有答案的。……我在工作中提出的专业术语是“内在语言” (I-language)，以区别于一般概念，但就像“语法” (Grammar) 一样，它指的是同一个系统，是内在的、个体的、内向的。也就是说，它关注的是程序而非输出，而输出实际上只是附属现象。

现在已经证明，语言的规则和原则并不是在你听到的字串上运行的，而是在你的头脑构建的结构上运行的，这种结构是普遍的，它体现在语言理解和使用的各个方面。

既然生成语言学的研究对象是看不见、摸不着的“内在语言”和“普遍语法”，那么这种学问的有关论断就不易捉摸和难以证伪了，其结果是，对于它们的所有质疑和批评无异于堂吉诃德大战风车——一切都是徒劳无功的。但是，我们还是想冒着“误读乔姆斯基”的风险，进行逻辑推论方面的讨论。根据乔姆斯基的上述说法：语言的规则和原则是运行在人脑中构建的抽象的普遍结构 (普遍语法) 上的，语言 (能力或普遍语法) 是人脑中产生无穷数目的 (外部语言) 表达的最简单的计算程序。也就是说，普遍语法这种抽象结构 (p) 是形成语言能力并产出外部语言 (q) 的必要条件 (necessary condition)，即：无 p 一定无 q ，有 p 不一定有 q ；且无 q 不一定无 p ，有 q 一定有 p ），甚至是充分必要条件 (sufficient necessary condition)，即：无 p 一定无 q ，有 p 一定有 q ；且无 q 一定无 p ，有 q 一定有 p ）。果真如此，那么可以推论：一个物种 (包括人工系统) 如果没有这种抽象的普遍结构 (无 p)，那么就不可能具有语言能力并且产出外部语言 (无 q)。比如，因为人类以外的自然物种没有普遍语法这种抽象结构 (无 p)，所以我们可以推出它们没有语言 (无 q)；语言大模型等人工系统内部也没有普遍语法这种抽象结构 (无 p)，所以它们应该没有语言 (无 q)。但是，实际上，大型语言模型却能够产出语言 (以类似于人类的水平来理解和生成语言，即有 q)。显然，不管是充分条件还是充分必要条件，“有 q ”都能推出“有 p ”。结果，

事实(无 p)跟推论(有 p)相矛盾。在这种情况下,我们虽然不知道能不能证伪(falsify)普遍语法理论,^⑩但是,我们知道一定不能证成(verify)普遍语法理论。

[附注]

- ① 这个电话采访的中文翻译有多种版本,文字方面出入不小,但是,基本意思比较相近。
- ② 详见 ML 编辑部(2024),视频见 <https://www.youtube.com/watch?v=urBFz6-gHGY>。
- ③ 详见“卫 sir 说”微信公众号刊发的《神经网络之父 Hinton 最新公开演讲: 数字智能会取代生物智能吗?》,2024-03-08,<https://mp.weixin.qq.com/s/u72sPc0PxwIaQBFK-TCKJw>。
- ④ 关于乔姆斯基对于统计语言模型的持续质疑,详见袁毓林(2025)。
- ⑤ 本文参考的 Pinker (1994) 中译本为欧阳明亮翻译的《语言本能: 人类语言进化的奥秘》,浙江人民出版社 2015 年出版。
- ⑥ 参考了 Beguš (2023) 的中文介绍,详见“脑机接口社区”微信公众号刊发的《AI 学语言与人脑极为相似! 新研究证明: 语言并非人类与生特有的能力, 机器也能学》,2024-07-20,<https://mp.weixin.qq.com/s/qT85kzQRrMUIxoIeYAqo-Q>。
- ⑦ 参见 Gazzaniga 等(2009)中译本《认知神经科学——关于心智的生物学》第 566~568、578 页,周晓林、高定国等译,中国轻工业出版社 2011 版。
- ⑧ 详见 Pinker (1994) 中译本第 457 页,我补充了一些人物的英语姓名和著作或理论的名称。
- ⑨ 详见“未来智能实验室”微信公众号刊发的《人工智能发展时间轴》,2023-02-04,https://mp.weixin.qq.com/s/q74RdM4iSB9tXO_iBW9OXA;“智能数码之家”微信公众号刊发的《探秘 AI 历史: 人工智能发展时间轴的重大事件》,2024-11-11,<http://www.aividigi.com/newsinfo/7580961.html>。
- ⑩ 详见 Chomsky (1995) 第 167~172 页,中文介绍请看程工(1994)第 2 页。
- ⑪ 因为乔姆斯基可以通过认定语言大模型生成的语言“并非真正的语言”来逃避上面这种证伪。

[参考文献]

- [1] 程工. Chomsky 新论: 语言学理论最简方案 [J]. 国外语言学, 1994, (3).
- [2] 韩佳琦. AI 教父 Hinton 46 分钟对谈: AI 可复制人类心智, 或加剧全球贫富差距 [Z]. 智东西, 2024-06-17,<https://mp.weixin.qq.com/s/nvhKpEkY9tyEOlhVS89Fdg>.
- [3] ML 编辑部. 杰弗里·辛顿: 乔姆斯基是个疯子, 形式学派大错特错 [Z]. 摩登语言学, 2024-10-13,<https://mp.weixin.qq.com/s/MGiBl62gfYa4OxHhgII3aw>.
- [4] 辛顿. 杰弗里·辛顿接受尤利西斯奖章时发表的获奖感言 [J]. 陈国华译, 当代语言学, 2024a, (4).
- [5] 辛顿. 神经网络教父 Hinton 最新万字精彩访谈: 直觉, AI 创新的洞见和思考, 未来 [Z]. 汉语堂, 2024b-05-30,https://mp.weixin.qq.com/s/gxu-l_M5tvsofSUqWnYhQw.
- [6] 徐烈炯. 生成语法理论: 标准理论到最简方案 [M]. 上海: 上海外语教育出版社, 2019.
- [7] 袁毓林. 人工智能大飞跃背景下的语言学理论思考 [J]. 语言战略研究, 2023, (4).
- [8] 袁毓林. 这是语言统计技术的胜利, 也是语言天生理论的失败 [J]. 语言战略研究, 2024a, (6).
- [9] 袁毓林. ChatGPT 等大型语言模型对语言学理论的挑战与警示 [J]. 当代修辞学, 2024b, (1).
- [10] 袁毓林. 描写还是解释: 由 ChatGPT 反思语言学的两种目标 [J]. 语言战略研究, 2025, (1).
- [11] Beattie, J. *The Theory of Language* [M]. Menston: Scolar Press, 1788.

- [12] Beguš, G., Zhou, A. & Zhao, T. *Encoding of speech in convolutional layers and the brain stem based on language experience* [DB/OL]. (2023-04-20) [2024-12-10].<https://www.nature.com/articles/s41598-023-33384-9>.
- [13] Chomsky, N. Three Models for the Description of Language[J]. *IRE Transactions on Information Theory*, 1956, (3).
- [14] Chomsky, N. *Syntactic Structures* [M]. The Hague/Paris: Mouton, 1957.
- [15] Chomsky, N. *Aspects of the Theory of Syntax* [M]. Cambridge, Massachusetts: MIT Press, 1965.
- [16] Chomsky, N. Deep Structure, Surface Structure, and Semantic Interpretation[A]. *Chomsky Studies on Semantics in Generative Grammar* [C]. The Hague: Mouton, 1972.
- [17] Chomsky, N. *Essays on Form and Interpretation* [M]. New York: North-Holland, 1977.
- [18] Chomsky, N. *Lectures on Government and Binding* [M]. Dordrecht: Foris Publications, 1981.
- [19] Chomsky, N. *The Minimalist Program*. Cambridge [M]. Massachusetts: MIT Press, 1995.
- [20] Gazzaniga, M., Ivry, R. & Mangun, G. *Cognitive Neuroscience: The Biology of the Mind (3nd Edition)* [M]. New York: W. W. Norton & Company, 2009.
- [21] Milak, E. & Tankosić, A. *Becoming a Linguist: Advice from Key Thinkers in Language Studies* [M]. London: Routledge, 2024.
- [22] Piantadosi, S. T. *Modern language models refute Chomsky's approach to language* [DB/OL]. [2023-03-15]. <https://lingbuzz.net/lingbuzz/007180>.
- [23] Pinker, S. *The Language Instinct: How the Mind Creates Language* [M]. New York: W. Morrow and Co., 1994.

(责任编辑 常文斐)