“辛顿·乔姆斯基·语言学发展”多人谈

[**编者按**] 2024 年 4 月 8 日，有“人工智能教父”之称的杰弗里·辛顿（Geoffrey E. Hinton）在都柏林大学接受尤利西斯奖章的获奖感言里，对乔姆斯基提出了毫不客气的批评：“语言学家被一个名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。但这一看就知道纯粹是胡言乱语。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开始。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。许多统计学家和认知科学家也说，永远不可能在这样一个大网络里学习语言。乔姆斯基从来没有提出任何一种有关语义的理论。他的理论全是关于句法的。”这篇发言引起了中国语言学界的关注，陈国华教授把它译为中文，以《杰弗里·辛顿接受尤利西斯奖章时发表的获奖感言》为题，发表在《当代语言学》2024 年第 4 期上。10 月 8 日，霍普菲尔德（John J. Hopfield）和辛顿以“通过人工神经网络实现机器学习的基础性发现和发明”获得诺贝尔物理学奖后，关于大语言模型和语言学发展、辛顿和乔姆斯基的话题再度爆火。

一位人工智能大家，获得了诺贝尔物理学奖；他批评了一位美国的著名语言学家，却引发了中国语言学者的热烈讨论，反思中国语言学的问题。这形成了一个奇妙的蝴蝶效应。我刊随即与《当代语言学》编辑部筹划，联合举办“大语言模型与语言学发展座谈会”。10 月 17 日，座谈会在商务印书馆召开，线上线下学者各陈己见。我刊特就此设多人谈栏目与学界共享。

数字智能和人类智能学习语言的方式不能等同

冯志伟（新疆大学中国语言文学学院） 近来，“人工智能教父”辛顿在 2024 年 2 月 19 日牛津大学的罗曼斯讲座中，在 4 月 8 日获得尤利西斯奖章的演讲中，不断地批评乔姆斯基，指责他的语言学理论是疯狂的理论；甚至在 10 月 8 日得到获诺贝尔奖的消息时，也不忘踩乔姆斯基一脚，指责他的理论是胡说八道。

辛顿不遗余力地批评乔姆斯基，主要原因在于他和乔姆斯基对于语言理解和人工智能发展的观点存在着显著的差异。

（1）语言习得机制。辛顿认为语言习得主要依赖于大语言模型的海量数据和神经网络的权重调整。他坚信人工智能系统通过输入大量语言数据，就可以像人类一样学习到语言的结构和意义。而乔姆斯基则主张“刺激贫乏论”，认为儿童在语言习得时接受的外在的语言数据不多，受到的刺激是贫乏的。人类天生具备一种内在的语言习得机制，这种机制不是通过后天的刺激得来的，而是由先天的基因决定的。

（2）对大语言模型的理解。乔姆斯基认为，人工智能和人类在思考方式、学习语言与生成解释的能力，以及道德思考方面有着极大的差异，如果 ChatGPT 式的机器学习程序继续主导人工智能领域，那么人类的科学水平以及道德标准都可能因此降低。他指出，大语言模型依赖海量数据进行工作，实质上只不过是一种“剽窃”。辛顿批评乔姆斯基的观点，认为当前的大语言模型有理解能力。辛顿强

调，现在人们必须研究如何将人工智能置于人类的控制之下，并投入大量的研究精力。辛顿对人工智能的快速发展表示担忧，认为我们正处在历史的某个关键点，在接下来的几年时间里，我们需要弄清楚是否有办法应对这种威胁。特别是考虑到一旦这些人工智能失控，就会接管人类，使人类面临生存威胁。

（3）方法论。辛顿的观点代表了数据驱动和神经网络的方法，这是一种联结主义的方法；而乔姆斯基的观点则代表了基于规则的方法，这是一种符号主义的方法。

总的来说，辛顿关注的是数字智能，并试图把数字智能推广到人类智能；而乔姆斯基则侧重于基于规则和符号主义的方法，关注的是人类智能。他们的争论有助于推动人工智能领域的研究和发展。但辛顿批评乔姆斯基的理论是胡说八道，实在是有些过分了。

我们认为，数字智能学习语言的方式与人类学习语言的方式并不完全相同。人类的语言习得有先天的因素，也有后天的因素。而数字智能依赖海量数据，并不存在刺激贫乏的问题，没有先天的因素。不能因为数字智能不存在刺激贫乏而否认人类语言具有先天的因素；也不能把数字智能与人类智能等同起来，忽视了二者的差别。

辛顿批评乔姆斯基的是与非

陈国华（北京外国语大学外国语言研究所） 2024年4月8日，有“人工智能教父”之称的辛顿获得了尤利西斯奖章。在颁奖仪式上发表的获奖感言里，他对乔姆斯基展开了毫不客气的批评：“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。但这一看就知道纯粹是胡言乱语。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开始。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。”这段话包含两个值得探究的议题。

一、人的语言能力是先天的还是后天学会的？

这是一个至今仍存在争议的古老问题。我认为，争议之所以一直存在，根本原因在于争论者对于“语言”“能力”“学”这几个基本概念的意义缺乏共识。如果我们把“语言”解读成“说话”，人类显然并非先天（即一出生）就有语言能力。婴儿一出生就会哇哇叫；给他一个指头之类的物体他会抓握；把他放入水里他会屏住呼吸，运动四肢。但我们不会把婴儿的哇哇叫视为说话，不会把他手握物体视为使用工具，不会把他在水里的屏息运动视为潜泳。凡出生时不会而之后经过自身努力才获得的能力，都是学会的。因此，语言当然是学会的。但是，所有动物，受其先天生理特征的限制，其学习能力都是有限的。鸟类永远学不会像人类那样哺育后代，人类也永远学不会像绝大多数鸟类那样挥动双臂飞到天上。在这一意义上，人类的语言能力是人类独具的一种受发育阶段和各种后天情况制约的生理或先天机能。

如果我们把“语言”分解为口语和笔语（即书面语言）这两种形式的语言，其获得方式显然不同。人类学说话，是其在婴幼儿阶段的一种本能的自发行为，只要有人不断与他进行口头交流，他就能学会对方的语言，不管对方说的是哪种语言，也不需要交流者刻意教他该语言词汇的发音、意思和用法。人类学笔语，一般晚于学口语，而且不是其本能的自发行为，既需要他人刻意教授相关文字符号的写法、读音、意义、用法，也需要自己刻意学习这些知识。我推测，乔姆斯基看到婴幼儿学习母语，既不需要老师像教算术那样刻意地教，也不需要学生像学算术那样刻意地学，为了以示区别，遂

将这一过程称为“获得”（acquisition，中文一般译作“习得”，英文词义本身不含“习”）。用“语言获得”来形容母语口语的学习，问题不大；用来指称任何一种笔语的学习，都属于用词不当。

二、究竟什么是语言，语言机器产出的话语和文本是否算真语言？

究竟什么是语言，也是一个古老的问题。如果把语言定义为人类各族群发展出来的一套意思驱动（meaning-driven）、遵守常规（convention-abiding）、基于类推（analogy-based）、彼此不同（distinct from one another）但大致可以互译（essentially translatable into each other）的符号系统，把人类的言语行为（verbal behavior，注意不是 speech act）定义为个人为了表达意思，利用某一语符系统或语符处理系统与他人进行的交流；那么，无论人类自身的言语产出或人类所发明语符处理系统的言语产出，只要满足这一定义，就可算作语言。换句话说，语言机器模仿人类自然语言产出的人工语言，只要足够自然，能以假乱真，就是真语言。

目前 ChatGPT 产出的话语，尽管非常流畅、自然，具有极高的仿真度，但仍未完全达到以假乱真的水平。首先，它无法完全正确理解人类的各种语言产出；其次，它只是对人类的话语做出回应，不像人类的不同个体那样，出于表达自己意思的欲望进行言语交流，也不具有个人言语（idiolect）的特定个性或风格。只要使用者未提出特定要求，其所输出话语的风格明显单一，且表达直白而啰嗦。

尽管如此，辛顿还是有足够的底气严词批评乔姆斯基，因为他做到了包括乔姆斯基在内的所有语言学家迄今远远无法做到的一件事，即让机器产出了几乎可以让人以为是人类自然语言的话语。下一个问题是，语言学家该怎么办？

辛顿获得诺奖引发的思考

李宇明（北京语言大学） 霍普菲尔德和辛顿是人工智能领域的大师，获得了 2024 年诺贝尔物理学奖。辛顿在获奖前后多次批评乔姆斯基，说他的语言学理论误导了好几代人。辛顿 2024 年 4 月接受尤利西斯奖章时的获奖感言在《当代语言学》2024 年第 4 期刊登之后，引起了中国语言学界的较大反应，引发了对语言学发展问题的深刻思考。

大语言模型问世不久，就把诺贝尔物理学奖授予人工智能科学家，说明颁奖者认识到人工智能对当今世界发展的重要意义。的确如此，在数字时代，所谓新质生产力都需要人工智能的加持。

语言智能是人工智能皇冠上的明珠。机器通过大数据处理就能“学会”语言，“获得”语言智能，这是经验主义、数据统计在语言处理上取得的辉煌成绩。乔姆斯基语言学重视理性演绎，重视解释性，重视规律的揭示，引领语言学从经验主义发展到理性主义。科学本是寻求因果关系的，而以数据驱动的大语言模型重视的是统计频率，据说没有使用语言学的规则。在语言智能研究的历史上，规则驱动也曾是主流理念，数据驱动也曾经身处低谷。现在有不少学者预测，数据驱动将来还会遇到发展的“天花板”，语言智能需要数据与规则“双轮驱动”。由此可以理解辛顿为何会批评乔姆斯基。乔姆斯基做的是“语言科学”，要解释的是人类习得语言的机制；辛顿做的是语言技术，理论上也只是在“技术科学”的层面。不同路向、不同追求、不同科学层面，自然会有分歧、有批评。而且可以预见这种批评是暂时的，语言智能的发展不可能总是只采用经验主义路线。

语言学过去的研究精力主要在语言上，如果用“鸡生蛋”做比喻，就是主要精力在研究鸡蛋上。

乔姆斯基将语言学引向对母鸡的研究，探讨母鸡生蛋的原理。而人工智能的语言处理，是用“机器鸡”生出蛋来。这对语言学提出了新的课题，即：不仅要研究碳基生命的“肉鸡”所生之蛋（自然语言），还要验证、修正乔姆斯基的原理；同时也需要研究硅基生命的“铁鸡”所生之蛋（机器的语言产品），以及铁鸡的生蛋原理；而且还需要研究这两种蛋的异同、这两种鸡生蛋原理的异同，从而得出更高层面的语言学概括。语言学需要反思，需要开疆拓土，不能老在舒适区里惯性行走，更不能“抱残守缺”。

“两蛋两鸡”的研究，将极大丰富语言学，提升语言学的水平和影响力。而且也将促使语言学为语言智能的发展做出应有贡献，提升语言学的学科穿透力。值得关注的是，当前语言学的社会应用场域已经十分狭窄，主要有语言教学（特别是外语教学）、翻译、社会语言规划等。如果能够在语言数据、语言智能等领域找到作用域，也当为语言学开辟一方现代生活的应用场域。

亲知是人脑语言学习的起点

陈保亚（北京大学中文系） 基于辛顿反向传播算法的人工神经网络及其大语言模型能够生成崭新的合法句子，说明大语言模型已经获得了一套可以生成自然语言文本的规则，解决了长期以来困惑自然语言处理的根本问题。这一进展对机器人发展的重要性，犹如自然语言符号系统的出现对人类发展的重要性。从此机器人可以和真人进行自然语言对话，直接阅读、继承和使用人类用自然语言记录的浩瀚文本知识，并展开进一步的“认识”活动。

大语言模型还证实了哈里斯的分布理论的可行性。哈里斯在《从语素到话语》（1947）一文中提出了一种基于分布理论的发现程序，即只要对语素的分布做充分的描写，就能生成全部话语。大语言模型唯一能够利用的信息正是海量文本中语言片段的分布。大语言模型也证实了非线性回归数学模型的可行性，因为大语言模型本质上是不断调整参数权重，最终回归出能生成合法句子最佳模型。

不过乔姆斯基对大语言模型的质疑仍然值得重视。大语言模型是一种从文本到文本的言知学习方式，而不是基于经验学习的亲知学习方式。基于言知的学习模式需要以大数据和超强运算这样一种强储算能力为基础。相比之下，人类学习依赖的是弱储算能力，必须还原出规则。人类的强项在于从亲知和言知的互动中还原规则。比如汉语人通过触觉、味觉、视觉等在亲知布、木头、草的过程中获得“质料”的经验，由此类推出“布鞋、木鞋、草鞋”等都是质料不同的鞋，这种类推可以平行周遍下去，生成“金鞋、玉鞋、铜鞋、铁鞋、石鞋”等崭新组合，由此还原出“质料＋鞋”的组合规则。

大语言模型证实了规则隐藏于分布，但人工神经网络黑箱如何提取规则还不清楚。人工神经网络是模仿动物神经网络而设计的，但动物中只有人类会语言，这说明大语言模型并未反映人脑语言学习的独特机制。目前还没有办法有效区分动物神经网络和人脑神经网络。大语言模型并非像辛顿所说的那样揭示了人脑语言学习机制，但是为理解人脑神经网络和人工神经网络的差异提供了窗口。人是符号动物，认识人脑最终还是要以语言符号系统的起源、演化和运转为中心。亲知活动是人类语言习得的起点，也是人类不断进行创造性认识活动的源泉，缺少亲知的认识活动是从书本到书本的认识活动，无法超越文本界限。将来机器人如果发展出类似真人的生物属性，能够在亲知和言知互动的基础上完成自然语言学习，将进一步提供一个从人工神经网络认识人脑神经网络的有效窗口。无论是现在的物理神经网络还是将来的生化神经网络，都为认识人脑语言机制提供了窗口。

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（澳门大学人文学院中国语言文学系） 语言学素来跟诺贝尔奖没有瓜葛。今年的诺贝尔物理学奖授予人工智能专家霍普菲尔德和辛顿的消息甫一传出，不仅众多网友一脸懵：“凭人工智能的成绩怎么拿物理学奖呢？”而且，在貌似无关的语言学界也掀起不小的风波。虽然从瑞典皇家科学院的授奖理由“表彰他们通过人工神经网络实现机器学习的基础性发现和发明”上，看不出这跟语言学有一丁半点儿的关系；但是，敏锐的语言学家从诺奖官网对辛顿的采访中，再次听到了他批评乔姆斯基的声音：“大语言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”，云云。

其实，从 2022 年底 ChatGPT 爆火带动人工智能进入大众视野以来，被誉为“深度学习之父”乃至“人工智能教父”的辛顿，已经多次在讲演中直言不讳地批评乔姆斯基的语言观。比如，他在《数字智能会取代生物智能吗？》（2024 年 2 月 19 日，牛津大学）中说：“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，……”。的确，辛顿领导他的团队成功地把误差反向传播算法引入多层神经网络训练，开启了深度学习的爆发式进步，为大语言模型理解和生成自然语言奠定了坚实的基础。这次辛顿的获奖，表明了国际权威学术机构对他在人工智能领域的杰出贡献的认可。从自然语言处理的角度，也许可以说：这是基于大规模语料统计的语言技术的胜利，也是乔姆斯基语言天生理论的失败。因为，根据语言天赋论及与其相应的普遍语法理论，人类语言是受到十分繁复和抽象的规则（原则与参数）控制的，无法用马尔可夫过程模型来刻画，也是机器无法通过统计分析来学会的。但是，现代大型语言模型却以“再造人类语言”的方式，能够以接近于人类的水平来理解和生成自然语言。这种活生生的事实，使得辛顿有底气在尤利西斯奖章颁发仪式（2024 年春，都柏林大学学院）上说，乔姆斯基误导了几代人。无独有偶，一位早年跟乔氏有过交集的国际知名语言学家，2023 年末访问澳门大学，在谈到乔姆斯基语言学理论和大语言模型时，也语重心长地对我说：“他把语言学的方向带偏了。”话少意多，发人深省。

怀着对于学术研究的敬畏心（不误入歧途）和对于年轻学子的责任心（不误人子弟），趁着大语言模型成为诺奖主角的东风，认真思考和讨论下列问题不仅应景而且必要：

（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？

学术的激烈交锋会促进科技的深入发展

孙茂松（清华大学计算机科学与技术系） 辛顿对乔姆斯基的批评确实很不客气，好像有点儿过了；但其实乔姆斯基对大语言模型缺陷的批判，也非常尖锐。两方面的观点是针尖对麦芒，火药味儿十足。

我有一种未经认真“考察”的感觉——在基本问题上秉持绝对化观点，有我无你、誓不两立，似乎是科学或技术大师们的行事常态。譬如，国际语音识别领域的先驱、美国工程院院士杰利内克（Frederick Jelinek）1988 年曾经放过一句“绝”话：“每当我解聘一位语言学家，语音识别的性能就会提升。”反观我辈凡夫俗子，看问题往往比较“全面”“辩证”，既说甲好，也说乙不错，但甲乙各存

不足，以取长补短、相得益彰为宜。这样一来，却失去了学术态度上的鲜明性，背后的原因也许是缺乏学术洞见和深度。

目前发生的一个振聋发聩的事实是：大语言模型（机器）基本具备了比较高的遣词、造句、写文章的能力（尽管写短篇乃至中篇小说还不行），这在人类历史上称得上史无前例的技术成就，是目前为止包括乔姆斯基理论在内的全部语言学研究成果所未曾做到的。从这个意义上说，以乔姆斯基为代表的理性主义范式在与以语言大模型为代表的经验主义范式的竞争中，输了一分（辛顿提及“乔姆斯基从来没有提出任何一种有关语义的理论。他的理论全是关于句法的”，也不能说完全是偏颇的，因为机器写不出开放式的流畅句子，很难说它真正掌握了语义）。实际上，训练大语言模型的基本算法策略只需用两个词语即可概括出来：一是“下一个词预测”，一是机器的“自监督学习”。貌似很简约，但简约不简单，有一种“大道至简”的味道。其实其中藏着深刻的计算机理，如“下一个词预测”就与“所罗门诺夫归纳”密切相关。所罗门诺夫是1956年达特茅斯会议（该会议的成果之一是首创“人工智能”这个概念）的主要参加者之一，1964年发表了长篇论文《归纳推理的形式理论》，对相关问题进行了数学层次上的深入阐述，也重点讨论了其与普遍图灵机及乔姆斯基体系的关系，构成了“下一个词预测”策略的形式化基础。所以辛顿的批评“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”，是有一定道理的。

但上述情况并不能否认乔姆斯基提出的“语言的先天性理论”及“语言习得装置”（language acquisition device）是一个合理、严肃的科学假设（虽然我感觉device这个词有形“器物”的意味浓了一点儿，换成形而上的mechanism之类也许更稳妥些）。辛顿说：“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开始。”在我看来，随机权重的“大型神经网络”或可类比成乔姆斯基的“语言获得装置”，没有这个预设的“大型神经网络”，大规模语言学习（训练）是无法完成的。此外，大语言模型在性能表现上确实也存在深刻的不足，如缺乏可解释性及可信性等。如果能在模型中加入理性主义因素，是再好不过的，也是我们所期待的（虽然目前还远没有找到如何有效加入的方法）。

当下需要抓紧进行的一个基础研究问题是，尽快揭示出深藏于大语言模型之内的“第一性原理”。近两年国内外已经有一些大语言模型与脑科学、认知科学交叉的研究工作，如最近MIT发表的论文《概念几何：稀疏自编码器特征结构》。大语言模型的结构及其所有参数对人类是完全透明的，是个彻头彻尾的“白箱”；人脑则不然，顶多可算作“灰箱”，研究起来很不直接。不少学者相信，把大语言模型的内在机制及其动力学搞清楚了，会深切启发并有力推动脑科学、认知科学的相关研究（包括人脑中的语言习得装置究竟是怎样的）。

因此，辛顿同乔姆斯基之间学术上激烈的“君子之争”是件好事儿，会引发我们的思考并促使相关研究走向更加深刻。最后的结局不排除可能是这样的：两种观点也许会殊途同归，相互印证。通过研究大语言模型，我们对人脑“语言习得装置”的研究和认识会更加具体化，而这反过来又会启发我们设计乃至重构更加有效的大语言模型。

人工智能给语言学研究带来重大挑战和机遇

徐　杰（澳门大学人文学院） 当代人工智能突飞猛进，日新月异。这无疑给包括当代语言学在

内的学术事业带来了深远影响。在语言学方面，这些影响有的是挑战，有的是机遇。

首先，人工智能给当代语言学领域的重要分支计算语言学带来了重大挑战。曾几何时，就连专做自然语言处理的工程师自己也认为，为了达到其语言工程的目的，应该依靠而且只能依靠语言学科学家对语言规则的揭示和总结。而兼做自然语言处理的语言学家出于自身的专业训练和学术优势，理所当然地，无一例外地，走规则导向的自然语言处理技术路线。但是谁也没有料到，当今的人工智能工程师居然改道而行，抛开了基于语言规则的模式而走大数据路线，并且在工程学意义上取得了很大的成功。原本视野之外的白猫异军突起，居然还真的就捉住了一只大老鼠！这让黑猫何以自处？在我们看来，面对崭新的形势，跨界到自然语言处理工程的语言学科学家们其实大可不必过于纠结焦虑。他们可以要么完全回归对语言本体进行科学研究的老本行；要么干脆不理会 ChatGPT 们现在走的大数据白猫技术路线，锲而不舍地继续沿着自己原定的规则导向黑猫技术路线朝前走，其工作成果也许是下一代人工智能的突破口和增长点。久久为功，说不定能捉住更大更多的老鼠！

其次，人工智能给本体语言学研究带来了一个重大机遇，那就是新一代人工智能以其强大的数据处理能力，给包括语言学在内的所有学科的科学研究赋能，大幅度地强化其精度，大幅度地提高其效能。人工智能赋能现代语言学研究，在目前阶段主要体现在两个方面。其一是助力语言数据的收集和整理，语言学领域有许多分支学科对语言样本和语料有很高的依赖性。计量语言学，语言类型学，比较语言学，语料库语言学，等等，它们的成效主要取决于收集语言数据的范围、数量和准确度。最早的时候，这些工作基本靠焚膏继晷、兀兀穷年的手工操作自不必说，即使后来有了计算机技术的辅助，效能依然有限。在这些方面，今天的新一代人工智能正可以大显神威。它们不仅可以近乎竭泽而渔式地收集语料数据，还可以对这些数据进行初步的分类、整理和分析。其二是助力语言学数据的收集和整理。语言学是一门古老而崭新的学科，古今中外的研究文献可以说是汗牛充栋，浩如烟海。今天的年轻学者很快会发现，自己一入门就被淹没在各种理论、各类学说、似是而非的观点以及对同一类语言现象多种多样的分析方案中难以自拔，恍如雾里看花。这时大数据和深度学习技术就可以进来，对相关文献进行收集整理，条分缕析，甚至对不同理论进行初步的鉴定和评价，以便去粗取精，让人类学者很快站到学术研究的最前沿，如虎添翼，可收事半功倍之神效。

总而言之，人工智能带来的重大挑战是对当代语言学领域的一个分支学科计算语言学和一种研究范式，也就是对单一语种表面现象的收集、观察、整理和分析的重大挑战；而对本体语言学的高层次理论探索则是一个重大机遇。如能把握好这个机遇，科学规划，人机分工，把语料的收集、整理和初步描写乃至文献综述等粗活重活发包给机器，人类语言学家则可以集中精力和智慧于理论探索和解释，中国语言学整体事业极有可能在较短时间内实现跨越式发展，由现在的对单一语种的现象描写转型升级到跨语种、跨古今的高层次理论解释。这正是我们多年来梦寐以求的美妙愿景！

辛顿对乔姆斯基语言理论批评的三个错误

司富珍（北京语言大学语言学系 / 乔姆斯基研究所） 辛顿的尤利西斯奖获奖感言中有 3 条与乔姆斯基语言理论相关的关键性批评意见：（1）“语言显然是学会的”；（2）“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”；（3）“乔姆斯基从来没有提出任何一种有

关语义的理论，他的理论都是关于句法的”。同时，他声称他所设计的语言模型“实际上是人类语言的工作模型”。众多研究表明，这 3 条批评性意见代表了他本人以及一批观点相近的“倒乔”派对于乔姆斯基理论的 3 个错误性理解。

我们倒着来看，先说第三条意见：“乔姆斯基从来没有提出任何一种关于语义的理论，他的理论都是关于句法的”。只要认真阅读乔姆斯基的原著，就知道这是明显的错读错解。乔姆斯基在最早期的著作《句法结构》中就曾强调：“形式特征和语义特征之间存在对应关系，这一事实不能忽视。这种对应关系应该用一种更为一般的语言理论进行研究，该理论要包括语言形式理论和语言使用理论作为其组成部分……我们发现，这两个领域之间显然存在颇具普遍意义的一些关系……我们应该乐意看到语法将语言的句法框架独立表现出来，让其能够支持语义描写。”1972 年，乔姆斯基又出版了《生成语法中的语义学研究》。语义学家海姆（Heim）和克拉策（Kratzer）也撰有《生成语法中的语义学》，从逻辑语义角度为乔姆斯基的语义理论提供了系统的形式化表达。乔姆斯基不同时期的文献，如著作《句法理论的若干问题》《管辖和约束演讲集》和《最简方案》，论文《论名物化》（Remarks on norminalization）等，也都有对句法与语义关系的系统性讨论。只不过，乔姆斯基认为“不能把是否合语法的概念等同于是否有意义”，换言之，句法相对于语义是独立的系统。

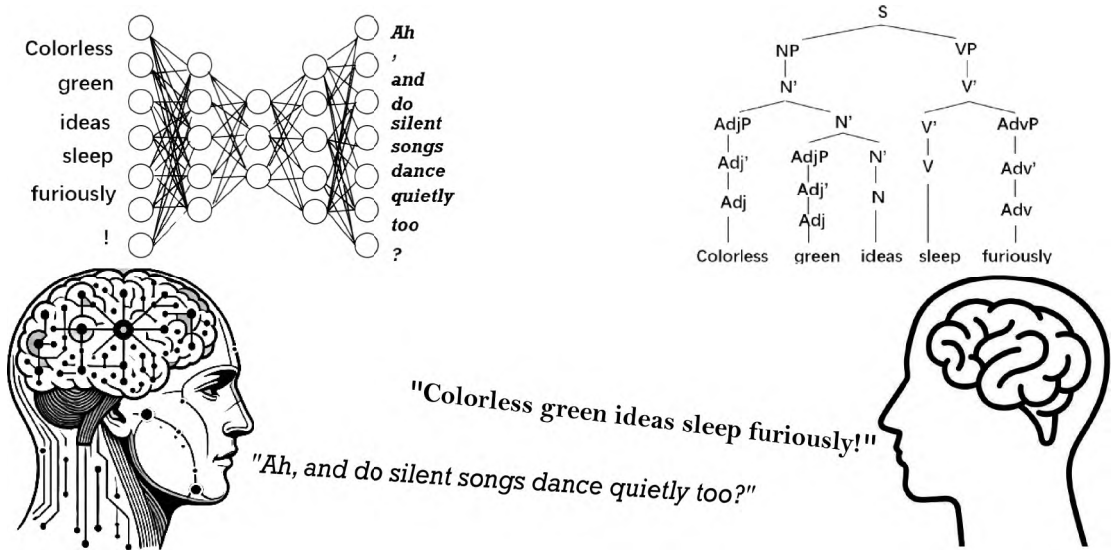
再看第二点：“大型神经网络学习语言，不需要任何先天结构，只是从随机权重和大量数据中开始学习”。这一结论可能反映了大语言模型工作的大部分事实，即随机权重和海量数据在机器学习中的重要性，但它也恰恰反映出大语言模型与人类语言工作机制的不同：人类儿童可以在刺激贫乏的情况下习得母语，并且不管每个人接触的语言环境差异多大，数据量差异有多大，其句法构造方面的语言能力发展进程及语法完备程度却几无差异，这就是所谓语言习得的柏拉图问题。它与机器基于海量数据存储与训练的“豪华刺激”机制完全不同。

况且，所谓不需要任何“先天结构”的说法也在一定程度上是个谎言。我们用辛顿自己使得过的方法对 ChatGPT 4.o 进行了提问，结果表明，语言学家在大语言模型的数据库中做了大量与语法、语义和语用相关的标注和校正工作。说不需要“先天结构”，只是说未“显性”使用乔姆斯基的形式化规则：尽管“大语言模型绕过了乔姆斯基所倡导的语言规则的明确形式化，但对语法、意义和语境等语言特征进行建模的需求却深深植根于语言理论中”。语言学工作者间接参与给予的语法、语义知识可以与先天结构类比。尽管它与人类语言的天赋性还有本质区别，前者是“人工的”“外在的”，后者则是生物遗传所决定的，“内在的”，具有生物学的更复杂本质，更多谜题尚待更多跨学科协作下的探索。

最后来看“语言显然是学会的”。大语言模型是数据驱动的模型，其基础是外部主义的，你可以说它的语言的确是学习来的，或者用乔姆斯基的话来说，是一种高科技“剽窃”。而人类语言则是“生成”的，它具有大语言模型所不具备的创造性、刺激贫乏性，甚至“有限性”。比如莫罗（Moro）的几项基于神经影像学的实验研究表明，人类语言官能既不能产出“不可能的结构”，也无法理解“不可能的结构”，而同时又可以创造出新的语法表达形式。我们团队最近也有两项实证研究，表明大语言模型不仅未展示出其创造性特点，而且在理解人类创新性语法表达形式时会表现出自相矛盾的情况。因此人类语言是“习得”的，这与机器的语言“学习”在机制和表现方面都有质的区别。

语言学研究融入 AI 的三种方式

詹卫东（北京大学中文系） 众所周知，人工智能（AI）有符号主义和联结主义两条路线。借助下面的示意图[①]，以语言智能为例，可以扼要展示两条路线的核心区别。



要让机器能模拟人的语言能力，需要回答一个基本问题：人说话的能力是怎么来的？

图中右上方的句子树形结构图代表了符号主义的回答：人说话的能力来自人脑的语言知识。机器需要像人脑一样，预存名词短语（NP）、动词短语（VP）、形容词短语（AdjP）和副词短语（AdvP）等许多范畴，这些范畴之间形成树状结构的组合模式。从根节点 S 开始，不断推导，直至叶子节点，就可以像人一样，输出句子了。

图中左上方的模拟神经网络图代表了联结主义的回答：人说话的能力来自人脑的学习能力。机器中不需要预存任何范畴和组合模式知识。机器可以构建人工神经网络，通过观察海量文本的词语分布，神经网络可以调试出优化的神经元联结参数（此即学习），直至涌现出语言能力，像人一样，给上句就能接下句。

符号主义是知识驱动，联结主义是数据驱动。符号主义的挑战在于：知识从哪儿来？联结主义的挑战在于：数据从哪儿来（以及是否有足够的算力从数据中学习）？

从目前 AI 的发展来看，联结主义 AI 的性能走在了前面：互联网提供了海量数据，GPU 提供了超强算力，推动着联结主义 AI 飞速发展。而符号主义 AI 尚未找到解决知识来源问题的好办法。

将人类语言学知识预装到计算机中来模拟人类语言智能的实践效果不佳，原因就是人提供的规模有限的语言学知识往往覆盖不了真实场景中的语言现象。同时机器预装语言学知识后并没有获得学习能力，难以适应现实世界的复杂变化情况。

由此，语言学研究如果要融入 AI 的发展浪潮，无外乎 3 种方式。（1）继续符号主义路线。在现

① 图中人说的句子“Colorless green ideas sleep furiously”是乔姆斯基的经典名句“无色的绿色思想狂怒地睡觉”；电脑说的句子是 ChatGPT-4o 的回应：“Ah, and do silent songs dance quietly too?”（啊哈，这么说无声之歌也会静悄悄地跳舞喽？）

有知识框架下挖掘更高质量、更大规模的语言学知识，仍旧以知识驱动的方式推进 AI 技术。但这一路线大致只适用于特定任务场景（如对精度、可控度要求高的任务），很难像大模型那样更具通用性。（2）结合联结主义路线。在语言学知识指导下制作高质量训练和测试数据集，帮助机器基于小样本数据学习，更高效地提升其语言能力和智能水平。（3）提出一套全新的兼具学习能力和符号知识表征的语言学理论框架，实现可解释的 AI（缺乏可解释性是联结主义的先天缺陷：端到端的训练不关注中间推导过程）。

前两个方式比较务实，语言学者可以发挥自己既有的语言学优势，积极扩大语言学知识范围，将语言学知识转化为计算机可用的资源（知识库或数据集），助力 AI 发展。第三个方式志存高远，非常有挑战性，如能实现，将意味着语言学理论研究和 AI 范式的重大突破。

辛顿没能跳出“中文屋”

完　权（中国社会科学院语言研究所） 虽然辛顿丝毫不留情面地批判了乔姆斯基，但是我们不必妄自菲薄，以为语言学就要被人工智能（AI）取代了。实际上，我觉得我们依然有价值，甚至可以反过来给辛顿一些建议。

辛顿在演讲中说：“大语言模型的工作方式，以及我们人类的工作方式就是，我们看到很多文本，或听到很多词串，进而获知词的特征，以及这些特征之间的交互作用。所谓理解，就是这么回事。神经网络模型正在以与人类完全相同的方式做理解。”但是，我们真的就是这么理解的吗？我觉得还不够。

确实，这种方式，非常接近我们认知功能学派推崇的 usage-based linguistics，基于使用的方式，简而言之，用多了，就会了。基于使用的研究法的基本假设就是，语言知识是在联想网络中组织起来的，关联网络对使用频率敏感，在语言使用中高频的语例引起说话人心智中的固化，形成独立表征。所以，认知语言学家对辛顿的说法，必然是举双手欢迎。但是，辛顿对意义的理解，还是缺了那么一些。

这就来到了我的主要观点——辛顿没能跳出“中文屋”。

塞尔（John Searle）1980 年提出一个“中文屋”问题。他的目的是要反对图灵测试，他认为一个计算机程序通过图灵测试并不意味着它具有智能，至多只能是对智能的模拟。为了论证自己的观点，塞尔提出了这样一个思想实验。

想象一个从小说英语、完全不会中文的人，被锁在一个房间里。房间里有一盒汉字卡片和一本规则手册。这本手册用英文写成，告诉人操作汉字卡片的规则，但并没有说明任何一个汉字的含义。它不是汉英字典！只是一个操作特定汉字卡片的规程，本质其实是一个程序。现在，房间外面有人向房间内递送纸条，纸条上用中文写了一些问题，房间里的人只要严格按照规则操作，就可以用房间内的汉字卡片组合出一些词句，完美地回答输入的问题。于是，这个人提供的输出，就通过了关于“理解中文”这个心智状态的图灵测试。

然而，塞尔指出，这个人其实仍然一点儿也不会中文。更进一步，无论是在这个房间中，还是这个房间整体，都找不到任何理解中文的心智存在。因此，通过图灵测试并不意味着拥有智能或者心智。不是拥有，只是功能上的模拟。就好像，假肢不是真胳膊。

塞尔真是了不起！现在看来，大语言模型展现出的强大功能，本质上依旧是高仿的假肢。大模型

依旧是“中文屋”。有人对目前前沿的所有 GPT 系统模型做了综合性测试，发现没有哪个模型不是仍旧处于“中文屋”的阶段。判断的标准有这样几个。（1）复杂的语义解析，如理解情绪、视觉体验、隐喻、故事背景等。（2）情境推理，比如完成这样的情景推理任务：“遮阳伞是否适用于下雨天？”这需要世界知识。（3）创造性生成，也就是看它是否真正拥有与人类相似的想象力和认知创造性。我常常想，我给大模型输入我正在写的论文的前半段，它能不能给我生成后半段？答案不言自明。（4）自我意识与反思，这就是很多科幻小说里的机器人觉醒，比如电影《终结者》里的天网 SKYNET 觉醒后开始毁灭人类。

现在看来，目前的大模型远远没有达到能够觉醒的水平。所以我们不必担心，SKYNET 永远不会派出 T800。辛顿对 AI 的担心，目前看来，是多余的。因为被困在“中文屋”里，是永远造不出 T800 的，除非他愿意听一听我们认知语言学者的建议。

大家知道，认知语言学的背后，是体验哲学，是肉身哲学，是心寓于身的哲学。我们能够理解语言的意义，是因为我们从诞生起就在语言中感知。

我们要怎样来帮辛顿跳出“中文屋”？答案其实萨丕尔 100 年前就在《语言论》里说过了，就是意义要从语言以外去理解，依靠经验。AI 需要有人的人生经验，才能真的有理解。现在图像识别还不能叫作图像感知，比方说，它识别出来这是猫了，就真的懂得这是猫吗？其实还是不懂的。我们是怎么认识猫的？依靠经验！我们不只是看了很多猫，而且还摸过、抱过、听过，讨厌过或喜欢过，跟猫在一起有故事。这才是意义。在我看来，现在的 AI 离理解意义还差十万八千里。要先有感知，才有认知，再有经验，再形成概念，才到语言，这就是意义的整体论。辛顿需要语用整体论帮助他，帮助 AI 理解意义。

我希望辛顿们最终能跳出“中文屋”，带给我们更好的 AI。到那个时候，再像现在这样做语言学，恐怕真的要下岗了。我希望这事儿发生在我退休之后！

辛顿与乔姆斯基之争中的三个事实

王　伟（中国社会科学院语言研究所） 这次辛顿得诺贝尔物理学奖，引起全球广泛关注。因为这个契机，辛顿今年 4 月在都柏林大学学院获尤利西斯奖章时严厉批评乔姆斯基的演讲（我有幸获得辛顿的授权，在《当代语言学》2024 年第 4 期发表了陈国华教授翻译的演讲全文），引起了更多语言学同行的重视。持有不同立场的学者，对此争议纷纭，这种情况是可以理解的。不过，我认为，在这个问题上，区分客观事实和主观态度，是最重要的事情。这一点，对于刚刚或即将进入语言学这个专业的年轻人来说，尤为关键。

所有这些争议的源头，是这样一个重要的事实：以辛顿为代表的人工智能领域的科学家，找到一种用大语言模型开发人工神经网络系统的方法，即通过“喂”给系统超级巨量的真实语言数据，训练系统获得生成合乎语法的句子的能力，并且胜任与人类持续对话的任务。以 ChatGPT 为代表的这类系统，在运用语言方面表现出前所未有的高水平，比如可以完美分析出对话者提供的笑话为什么好笑，“笑点”何在，等等。

第二个重要的事实是，至今没有任何人类语言学家能够开发出这样的语言生成系统，令它只生成

合乎语法的句子，而绝不会生成不合乎语法的句子。“任何人类语言学家”，当然也包括主流语言学家乔姆斯基。

无论如何指出 ChatGPT 的缺点，或者试图证明它拥有的根本不是人类的语言能力，都无法改变以上事实。我们知道，“深蓝”系统在国际象棋方面战胜了所有人类棋手，AlphaGo 则在围棋方面打遍天下无敌手。毫无疑问，这些计算机系统获得下棋能力的方法，当然跟人类不同。它们跟 ChatGPT 一样，都是靠“暴力”运算来实现各自的能力。但是，我们能因此否认计算机系统击败人类棋手这一事实吗？当然不能。

第三个事实，与乔姆斯基有关。ChatGPT 横空出世之初，乔姆斯基率先发难，攻击它“剽窃”。现在我们知道，他说的根本不是事实，这只是他的主观态度。他显然只是在发泄个人的不满情绪。道理很简单，如果我们了解一下 ChatGPT 的原理，便会知道，这些系统并不储存任何具体的句子。它的“本事”，是可以学习从文本中提取词的特征，以网络形式通过数量惊人（千亿数量级）的参数来表征这些特征之间的相互作用，并依据这些知识，来预测下一个词的特征。而预测下一个词（原文为 token，在宽泛意义上与“词”大体相当），是辛顿在 40 多年前就提出的人类生成句子的“第一性原理”。ChatGPT 这样的系统，其实是一大堆以网络形态存在的参数，而不是迄今为止人类曾经说过的所有句子。如果依据这样的知识系统生成合乎语法的句子，是一种“剽窃”，那么世界上就没有不剽窃的人，包括乔姆斯基自己在内。因为乔姆斯基等于是在说，学习就是剽窃。

前年年底 ChatGPT 崛起，令我大受震撼。我当时决定不揣浅陋，以业余爱好者的身份观察、学习、判断，试图窥见这一事件对中国语言学的意义。我先后在多所大学和学术机构分享自己的观察和分析，得到大量反馈。令我印象最深刻的是，外语、翻译等专业学子的焦虑情绪。他们害怕的是，自己所学的专业，将在日益强大的人工智能的威胁下，变得一文不值。我当时对他们的建议是，主动拥抱人工智能工具，尽快学习并掌握这一利器，并把自己的专业知识和机器擅长的能力结合起来，只有这样才有希望避免被彻底淘汰。借此机会，我仍然坚持当初的建议，同时还要鼓励这些年轻人，尤其是上文所说的“刚刚或即将进入语言学这个专业的年轻人”，认清以上 3 个事实，看清历史发展的趋势，在独立的理性思考基础上，做出明智的选择。

人、人脑与人工智能

刘　畅（中国人民大学哲学院） 我想提两个问题。

第一，语言活动中囊括的种种行为反应，如规则学习、语义理解、符号识别等，其真正的主体是什么？

对此有两类回答。一类是还原论的进路，认为（或默认）语言活动的主体是人脑。另一类是反还原论的，也是我所倾向的进路，主张人脑不是主体，人才是主体——是你我在说话，在交流，在思考，而不是你我的大脑。

我愿引入一组区分：自主的反应和自发的反应。语言活动是典型的自主反应。我要说些什么，要怎么说，原则上是由我自主控制的。对于我自主所做的事情，我通常不经观察就知道我做的是什么，我这样做的理由是什么。自发的反应却并非如此，比如免疫系统的反应。人类的免疫系统能侦测到病

原体的入侵，并自发地做出一系列复杂精妙的免疫反应。但所有的免疫反应都是在无意识层面进行的，无需基于行动理由的理解，也无法自主地加以控制。当然，我可以自主地借药物来调节免疫系统的反应。但我能自主控制的是我服用药物的动作，而不是我的免疫反应。

只有自主的反应，才属于狭义上“我所做的事情”。问：“你在干啥呢？”答：“我在消化食物。”这是个玩笑，因为“消化食物”这样的反应显然不属于正常意义上“我所做的事情”。与其说我在蠕动肠道，不如说是肠道自己在蠕动。

那么，人脑的反应——包括单个神经元的反应，以及整个神经网络的反应——是自主的，还是自发的？我认为答案是显而易见的：大脑的所有反应都是自发的，不是自主的。就像我无法控制白细胞的产生一样，我也无法控制脑神经元的激活。人脑的自发反应，构成了人的自主反应的神经生理基础，不过，它们仍是两类不同层面的活动。是人脑在功能层面成就了人的思想、认知，而不是人脑自己在思想、在认知；正如是我借助眼睛看，而不是眼睛自己在看。

我们可以把有关意识主体的追问，转化为有关行为方式的追问。“人”与“人脑”的不同，归根到底，是这两类行为反应方式的不同。一个行为反应是以自知的、可控的、有意识的、有逻辑的方式做出的，还是不自知、不可控、无意识、无逻辑的，往往存在显见的、系统的差异。我们能清晰地分辨两者，而不必诉诸某种“灵魂实体”的预设。以自主方式做出的反应，就是有意识主体参与的反应，“我”的反应；而无意识主体参与的反应，则是自发的。

第二，如果人工智能成功模拟了人类神经系统的自发反应机制，这是否意味着它拥有了与人类相当的感受、理解、认知的能力？比如，是否应当认为 ChatGPT 具备了语言学习的能力？

辛顿的回答显然是肯定的。实际上，他所有的表述都建立在对这一立场认可的基础上。说他“默认”了这一立场或许更准确些，因为辛顿似乎并没考虑过要对此给出任何辩护或反思。

但我相信，问题的正确答案应当是否定的。如果说，一个人类主体具备的种种自主能力，本就不能一对一地还原到这个人的脑神经网络上；那么，一套人工智能哪怕完美复刻了人类的脑神经网络，也一样不等于就具备了任何与人类相当的自主能力。它所模拟的，只是人的自主活动得以实现的功能性基础，而不是自主活动本身。

我赞同完权老师的判断——“中文屋”依旧未被打破。我的观点更进一步：塞尔的“中文屋”，原本的着眼点是对语义的理解能否还原为对句法规则的掌握；不过，假若不论语义的学习还是句法的学习，都只在自主、有意识的层面才是可能的，那么，人工智能的首要问题就不在于如何从句法上升到语义，而在于它压根不懂语义，也不懂句法。人工智能没有认知能力，没有规范能力，它只是具有认知、规范能力的人类所动用的一个技术工具而已。

“人工智能”的名号往往给人一种错觉，仿佛它能与人类智能并驾齐驱，是智能实现的另一条可能路径。这在很大程度上掩盖了“人工智能”作为技术工具的实质。人工智能的发明本是用以替代人类劳作的，而不是用以替代人类的；尽管像其他技术工具一样，对它的不当处置，也可能毁灭人自身。

责任编辑：王　飙