

人工智能大飞跃背景下的语言学理论思考^{*}

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 / 计算语言学教育部重点实验室 北京 100871)

提 要 当今以深度学习为核心的人工智能催生了科学研究的第五范式,本文呼吁,语言学研究也要拥抱数据 / 计算密集型的第四 / 五范式,并通过介绍和评论人工智能研究与技术开发过程中,有关专家对于跟语言学有关的问题的见解,来支持上述倡议。本文主张或赞成的主要观点为:(1)语言的主要功用是交际,思考往往是内心的对话与问答。(2)语言和思维并不完全等同,基于语言运用的智能水平的图灵测试并不完全有效。(3)语法这种智能并不是人类独有的,语法的原理还适用于语言之外的躯体、物品及其图像等组合性系统。(4)现有能力超强的大规模语言模型不能用作人与机器人交谈的技术界面,有效的语言运用必须是一种具身智能,包括具身认知的词语接地和环境可供性等内容;相应地,必须考虑不同类型的具身性图灵测试。(5)ChatGPT 的成功说明了大模型、大数据和强算力能够捕获语言的统计规律和运用模式,语言学要采用数据 / 计算密集型的第四 / 五范式来探索语言的统计结构。

关键词 深度学习;人工智能;第四 / 五范式;语言与思维;(具身性)图灵测试

中图分类号 H002 **文献标识码** A **文章编号** 2096-1014 (2023) 04-0007-12

DOI 10.19689/j.cnki.cn10-1361/h.20230401

Theoretical Reflections on Linguistic Studies Against the Background of AI Great Leap Forward

Yuan Yulin

Abstract In view that artificial intelligence (AI), with deep learning as its core technology, is bringing about the fifth paradigm in scientific research, this paper argues that linguistic studies also need to embark upon the fourth or the fifth paradigm for pursuing a data-intensive and computation-intensive approaches. To further this suggestion, insights from linguistic experts on linguistics-related issues in AI study, which are often encountered in the scientific research and technical development of AI, are introduced and examined with an aim to support my argument. To engage with this topic, the current paper advocates the following points of view, namely: (1) While language functions primarily to communicate, thoughts usually happen in form of dialoguing and questioning-answering in the mind. (2) As language and reasoning are not totally equivalent, the Turing test cannot be fully valid. (3) Grammaticalization is not an intelligence unique to mankind, so the functional principles of grammaticalization also apply to composite systems other than language, including bodies, objects, and images. (4) Any existing super-capacity large language model (LLM) is not adequate to be used as a technological interface for the communication between human beings and robots, because any effective use of language should be an embodied intelligence, including lexical grounding and environmental

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。

^{*} 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”(19VXK06)。承蒙编辑部和匿名审稿专家的指正,谨此谢忱。

affordance in embodied cognition; accordingly, different types of embodied Turing tests should be taken into consideration. (5) The success of ChatGPT shows that large models, big data and powerful computability work to capture statistical rules and operating modes of language. In this sense, linguistic studies should adopt data- and computation-intensive methodologies as a fourth or a fifth paradigm to probe into the algorithmic structure of language.

Keywords deep learning; artificial intelligence (AI); fourth/fifth paradigm; language and mind; (embodied) Turing test

一、引言：语言学研究要不要呼唤第四 / 五范式？

目前，以深度学习为技术核心的人工智能，已经对科学研究、技术创新和日常生活带来了颠覆性的影响。由于当今的人工智能技术可以在差异巨大的时间与空间尺度上，对自然现象和社会现象进行比较精准的建模与预测，比如著名的游戏程序 AlphaGo 能够打败人类围棋世界冠军，语言模型 AlphaFold 能够预测 2 万多种生物蛋白。因而，有人提出：这种人工智能技术与能力，是否代表着科学发现新范式的曙光？比如，图灵奖获得者、前微软技术院士吉姆·盖瑞（Jim Gary）用“四种范式”描述了科学发现的历史演变。第一范式的起源可以追溯到几千年前，它纯粹是经验性的，基于对自然现象的直接观察。虽然在这些观察中，有许多规律是显而易见的，但没有系统性的方法来捕获或表达这些规律。第二范式以自然理论模型为特征，例如 17 世纪的牛顿运动定律，或 19 世纪的麦克斯韦电动力学方程。这些方程由经验观察、归纳推导得出，可以推广到比直接观察更为广泛的情形。虽然这些方程可以在简单场景下解析求解，但直到 20 世纪，因为电子计算机的发展，它们才得以在更广泛的情形下求解，从而产生了基于数值计算的第三范式。21 世纪初，计算再次改变了科学，这一次则是通过收集、存储和处理大量数据的能力，催生了数据密集型科学发现的第四范式。机器学习是第四范式中日益重要的组成部分，它能够对大规模实验科学数据进行建模和分析。这四种范式是相辅相成、并存不悖的。^①

在此基础上，微软技术院士、微软研究院科学智能中心负责人克里斯·毕晓普（Chris Bishop）倡导科学发现的“第五范式”：把深度学习作为兼顾科学发现的速度与准确性的强大工具，用科学基本方程的数值解（而非经验观察）来训练神经网络；将科学方程的数值解看作自然界的模拟器，以较高的计算成本，对众多我们感兴趣的应用项目进行计算，例如预测天气、模拟星系碰撞、优化聚变反应堆设计，或计算候选药物分子与目标蛋白的结合自由能。^②

如果简单类比一下，那么语言学研究的第一范式是基于经验的传统语法，第二范式是讲求操作程序的结构主义描写语言学，第三范式是探索结构模式及其背后心智过程的转换生成语法，以及基于认知科学的功能语言学。那么，在目前人工智能的自然语言处理领域捷报频传、模式锋出的今天，语言学界要不要呼唤和拥抱语言学研究的第四范式和第五范式呢？

为了更好地回答上述问题，我们首先需要了解目前人工智能研究与技术开发过程中，有关专家对于下列问题的见解：（1）语言与思维关系怎样？语言的主要功用是思考而不是交际吗？（2）从语言运用看智能水平的图灵测试有效吗？（3）语法这种智能是人类独有的吗？语法的原理只适用于语言吗？（4）现有的能力超强的大规模语言模型能否用作人与机器人交谈的技术界面？有效的语言运用为什么必须是一种具身智能？具身认知的词语接地和环境可供性为什么重要？相应地，为什么必须考虑具身性图灵测试？（5）ChatGPT 的惊艳表现对语言学研究的范式选择有什么启发？

① 详见 Bishop (2022), Hey et al. (2009); 后者的扉页献辞是“For Jim”。

② 详见 Bishop (2022)。笔者进行了提炼和重新表达；如要引用，务请核对原文。

下面,我们分别对这些问题进行概述性介绍,也做出一些必要的评论;希望能够说明:语言学应该拥抱数据/计算密集型的第四/五范式,探索语言的显式的统计结构,为通用性人工智能的发展做出自己的贡献。

二、语言与思维的可分性和图灵测试的局限性

这一节首先介绍关于语言与思维关系的同一性假设及其反对意见,然后介绍乔姆斯基“语言的主要功用是思考而不是交际”的观点并提出质疑,接着通过介绍巴赫金的“对话”与“多声性”理论来说明思考与交际并不截然分开,最后介绍图灵测试所隐含的语言与思维等同假设和有关人工智能学者的批评意见。

2.1 “语言与思维的关系”是语言学理论的一个聚讼纷纭、长盛不衰的话题。其中,最有名的莫过于“萨丕尔-沃尔夫假设”。粗略地说,萨丕尔强调,需要注意语言如何将世界分割为不同类别方面的差异。他的学生沃尔夫则把上面这个见解扩充成著名的“语言决定论假设”^①:

我们对自然进行划分,将其组织成概念,并根据我们的想法赋予其意义,这样做主要是因为我们就此达成了一致的协议——将它融入我们的语言,并编入语言的特定模式。当然,该协议是含蓄的、隐晦的,但其条款却带有绝对的强制性。

诸如此类的把语言等同于思维的学说,在哲学界也大有市场。比如,尼采曾经写道:“如果我们不愿借语言法则思维,我们就会停止思维。”更加有名的是维特根斯坦的断言:“我的语言的界限也即我的世界的界限。”在德里达等后现代主义大师们的作品中,也充斥着“摆脱语言的桎梏是不可能的”“文本就是自我参照”“文本之外一无所有”等危言耸听式的警句。^②然而,正如平克(Pinker 2002)所指出的:

认为语言就是牢狱的看法实际上过高估计了语言本身的力量,从而贬低了语言主体的能力。语言是一种非常了不起的能力,我们运用它来从不同的人那里获取思想,还可以通过多种方式对它进行更改和选择,以促进思想的进步。然而语言不同于思维本身,也并非人类区别于动物的唯一标志。它并不是所有文化的基石,也并非一个不可逃离的牢狱。语言不是强制性的协议,更不是我们世界的极限,也并非影响我们想象内容的决定因素。(中译本第245页)

2.2 近年来,乔姆斯基提出“语言的主要功用是思考而不是交际”的观点。他20年前在跟一个生物学家对答的视频中说^③:

关于语言的通常假设是,它的功能是促进交流。对的,这一点我从来没有相信过。……语言的典型使用是为了思考,而不是为了交流。几乎所有的语言使用,接近100%是内在的。从统计学上讲,语言的用途几乎都是内在的。……但压倒性的证据表明,内在思维正在为我们发挥着某种功能,规划,苦恼,或者用来做其他的事情,其中只有一小部分最终用来交流。……事实上,即使是外化的部分,沟通也只是一种非常奇怪的感觉。……你和某人站在公共汽车站,……所以你和他们谈论天气或棒球比赛,那不是交流。有时这也被称为寒暄。……这只是一种建立人际关

① 这个假设有强弱不同的版本,这里按照Pinker(2002),见中译本第244页。

② 详见Pinker(2002),中译本第244~245页。

③ 转引自史有为(2022)。

系的社交方式，并不是传递信息或其他意义上的沟通。……有些外化的，在外化的部分中，很多只是寒暄的交流。

可见，乔氏认为语言的内在性使用就是思考（内在思维）。那么，言下之意是不是说：思考必须在内在语言的基础上进行呢？另外，他刻意把寒暄之类的情感交流排除在交流（或沟通）之外，似乎只有传递信息之类的有意义的沟通才算是交流。于是，得出结论：语言的典型功能是思考而不是交流。其实，思考跟交流并不一定能够截然分开。因为，所谓的“内心思考”往往是一个“内心对话、自我争辩、设问拟答、自问自答”式的交谈过程。例如：

首先，我们要晓得，学问有两个部分，一个是“学”，一个是“问”。这个问题两千多年前孔子就说：“学而不思则罔，思而不学则殆。”思考其实就是问，思考与问问题差不了太远，要问就必须思考，思考就会提出问题，这是分不开的。^①

提问质量决定我们的生活质量。为什么会这样呢？因为我们一直都处于一个自我对话的场景中。我们的大脑不断地抛出问题，然后自己回答问题。^②

当然这些都是非语言学专业的成功人士的直觉性认识。关于语言与思维的关系，还是平克（Pinker 2002）说得通透：

感知和分类提供了使我们与现实发生联系的概念。而语言使概念与词语联系起来，使我们联系现实的渠道得以扩展。……语言就是一个渠道，通过它人们可以互相交流思想和意图，并由此掌握周围的知识、习俗和价值观。……除了作为一种交流媒介，语言还可以作为大脑储存和处理信息的载体。……语言起到的是作为中央系统的子系统的功能，而并非是所有思维的媒介。（中译本第246～247页）

2.3 诸如上述丘成桐“思考就是提问”的直觉是极具洞察力的。上升到理论层面，就有了巴赫金的“对话”与“多声性”理论：

语言只能存在于使用者之间的对话交际之中。对话交际才是语言的生命真正所在之处。语言的整个生命，不论在哪个运用领域里（日常生活、公事交往、科学、文艺等等），无不渗透着对话关系。……这种对话关系存在于话语领域之中，因为话语就其本质来说便具有对话的性质。（巴赫金 1988：252）

在陀思妥耶夫斯基后期作品中，主人公所有重要的自我表述，也都可以扩展为对话，因为它们好像都是两种对话融合的产物。但是不同声音的交锋却隐藏得很深，渗透到语言和思想的精微之处。……这仍是两种意识、两种观点、两种评价在一个意识和语言的每一成分中的交锋和交错，亦即不同声音在每一内在因素中的交锋。（巴赫金 1988：288～289）

可见，巴赫金在相当程度上把对话与思想等量齐观，看作人类最基本的一种生存方式；于是，一个人的“言谈”往往就是其某种意识与观点的表达。但是，这种表达不是一种固定不变的思想立场，而是一个发展与开放的过程，是在跟潜在对手的对话与交锋中实现的；并且，跟其他“言谈”一起构建了话语的公共空间，各种不同的声音借此汇聚成一个充满张力的“多声性”的复合体。这也为欧洲的互文本性（intertextuality）理论开辟了道路。^③

① 参见：丘成桐《学“问”》，“数理人文”公众号，2022-10-12，https://mp.weixin.qq.com/s/y3drCKaZh0tHLtit_hRPw。

② 参见：博多·费舍尔《财务自由之路》，“正和岛”公众号，2022-09-29，<https://mp.weixin.qq.com/s/P2KMeBM0wr7GcsQhtQaKQ>。

③ 参考百度百科“巴赫金”条目，<https://baike.baidu.com/item/%E5%B7%B4%E8%B5%AB%E9%87%91/2290046?fr=aladdin>。

2.4 虽然在语言运用上内部的思考与外部的交流是密切相连、不可彻底分割的，但是，这并不意味着语言与思维具有同一性。要正确地认识到这一点也并不容易。比如，图灵（Turing 1950）认为“机器会思考吗？”这个问题是无法回答的，倒是可代之以“我们能否区分回答问题的是机器还是人类？”，即通过会话能力来测试机器是否具有跟人类一样的智能。这就是著名的“图灵测试”。可见，图灵测试隐含的假设是：语言代表着人类智慧的顶峰，能够进行对话的机器一定是智能的。其实，智力远不止于语言。当前自然语言处理等人工智能系统所犯的许多错误（比如，聊天机器人的答非所问），说明了这种系统在语义、因果推理和常识方面的根本缺乏。^①

布朗宁和杨立昆（Browning & LeCun 2022）直截了当地否认了图灵测试的有效性。他们的论证逻辑大致如下：图灵测试的基础是，如果一台机器说出它要说的一切，就意味着它知道自己在说什么；因为知道正确的句子以及何时使用这些句子会耗尽它们的知识。但是，机器可以谈论任何事情，这并不意味着它理解自己在说什么，因为语言并不会穷尽知识。相反，语言只是一种高度具体且非常有限的知识表征。另外还有一些非语言的表征方式，它们可以用一种更易于理解的方式来传达信息。比如，象征性的知识，包括图像、录音、图表和地图等等。在语言之外，这个世界本身就向人类展示了大量的信息，可供人类探索并使用。人类有了深刻的非语言理解，才使得语言有用武之地。正是因为我们对于世界有深刻的理解，所以我们可以很快地理解别人在谈论什么。也就是说，语言只承载了人类知识的一小部分，大部分人类知识和所有动物的知识都是非语言的（非象征符号性的）。因此，大规模语言模型（large language model, LLM）无法接近人类水平的智能。^②

可见，在人工智能这种技术背景上思考语言与思维的关系，不仅具有理论意义，而且具有实际的指引人工智能发展方向的应用价值。

三、语法作为一种古老智能的直观性质及其认知资源

这一节首先从人类心智的朴素性质上质疑当代复杂的句法结构理论的合理性，然后介绍人工智能专家关于“语法是一种古老的分析、规划与构想智能”的思想及其在场景建模上的运用，接着介绍他们关于“语法是把组件组装成整件的装配图”的思想及其在图像识别工程上的实践，最后介绍心理学家提出的包括语法在内的人类认知能力所依托的核心直觉知识。

3.1 传统上认为语法是组词成句的规律，或者说是构词造句的法则。但是，以乔姆斯基为代表的生成语法学派相信“语言独异说”：语言是人类的一个独立的认知系统，人类有独特的语言官能，语言机理构成了一种跟我们的身体性质无关的更加高级的官能。乔姆斯基等生成语法学者相信，语言可以分为内在性和外在性两种形式。他们主张：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，而不是诉诸语义、交际功能等外部条件。

对此，我们一直半信半疑，怀疑这是不是一种神秘主义的教条或迷思（袁毓林 2019, 2022a）。因为，根据平克（Pinker 1997），人类“心智是一套由计算器官组成的系统，它经自然选择的设计来解

① 详见 Zador et al. (2022)，中译文第 4 页。

② 参见：《Yann LeCun：语言的有限性决定了 AI 永远无法比肩人类智能》，“AI 数据派”公众号，2022-08-26，https://mp.weixin.qq.com/s/?__biz。

决我们祖先在茹毛饮血的生活中所面对的那类问题，具体包括：理解和操控物体、动物、植物以及他人”（中译本第22～23页）。正是在这一过程中，逐步积累和形成了下文要介绍的直觉物理学、直觉生物学乃至直觉语言学之类的知识。

正如平克（Pinker 2002）所指出的，“人类大脑适应的是一个已经不存在的世界”（中译本第286页）。认知能力进化的最初意图与当前人类赋予它们的新意图之间不相协调，这可能是我们某些困惑（甚至痛苦、悲剧）产生的根源。对许多知识领域来说，人脑还没有进化出适用于该领域的组织结构，大脑和基因组并没有显示出任何专业分化的迹象。我们并不具有直观理解这些领域的知识的心智工具（中译本第259～262页）。因此，我们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是否具有心理现实性；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。

3.2 值得注意的是，也有科学家从技术层面提出了跟语言独异说不同的语法观点，并且在工程实践中取得了成效。比如，美国布朗大学人工智能专家芒福德（D. Mumford）指出：^①

在所有智能动物的大脑中都由来已久地建立了语法模型，这些语法模型的作用是去分析各种感知信息，规划智能动物接收信息后的动作行为，甚至帮助其形成思想；而语言的语法，只是这些更加古老的语法的一种近期的扩展。

在诸如此类理念的指导下，加州大学洛杉矶分校的朱松纯教授与他攻读博士学位期间的导师芒福德等学者，在研究计算机视觉（图像识别）的时候，用概率语法图模型来为场景建模，因为他们发现场景跟话语一样具有下面3种特性：^②

（1）构成性（compositionality），比如：场景可以分解成物体，物体可以分解成部件，部件可以分解成图像的基本元素（primitives, textures, textures）；

（2）多样性（alternative）和可配置性（reconfigurability），比如：各种分解的成分可以有多种选择，从而以少量的元素组会产生大量的模式结构（configuration）；

（3）关联性（dependency relations），比如：部件之间相对位置的连接和关节变化（articulation）。

这些原则体现在与或图（and-or graph）等传统的概率图模型中，在可解释性、鲁棒性和基于小数据的学习方面，比多层神经网络模型有明显的优势。而上述特性，在语言学教科书上的表述是：语言符号的离散性、组合性、聚合性、层级性、递归性、依存性等等。

在朱松纯和芒福德（Zhu & Mumford 2006）有关思想的启发下，吴田富（2019）研究 and-or grammar（AOG）building block，提出使用一种最简单的语法来实现多层次、组成式的拆分-变换-聚合。他用与或图的原理来设计神经网络的结构，提出了一个与或图网络 AOGNet，不仅在当前主要的数据集如 ImageNet 任务中，性能超越了 ResNet、DenseNet 和 DualPathNets；而且具有更好的可解释性，找到了当前深度学习的判别式模型（CNN）与概率图模型（AOG）可能的联系。^③

3.3 人工智能学者的科学训练和技术修养，使得他们对于语法的认识更加一针见血、返璞归真。除了上面介绍的芒福德、朱松纯和吴田富三代学者的观点与实践外，萨巴（Saba 2022）的下列观点也

① 详见 Mumford (2016)，Zhu & Mumford (2006) 和吴田富 (2019)。Mumford (2015) 也值得一看。

② 详见吴田富 (2019)。朱松纯教授现为北京大学人工智能研究院院长。

③ 这是朱松纯教授在吴田富 (2019) 这篇文章前面的评价性序言。

颇有可观之处：

在符号系统中，有定义明确的组合语义函数，它们根据成分的意义计算复合词的意义。但是这种组合是可逆的，也就是说，人们总是可以得到产生该输出的（输入）组件，并且正是因为符号系统中，人们可以访问一种“句法结构”，这一结构含有如何组装组件的装配图。而这在神经网络中都并非如此。一旦向量（张量）在神经网络中组合，它们的分解就无法确定，因为向量（包括标量）可以分解的方式是无限的！

这篇文章意在重温 30 年前对于神经网络的批判，说明当今的多层神经网络依然无法实现可解释的人工智能。其中，把句法结构看作指引人们把组件组装成整体物品的装配图的思想，既朴素又透彻。亲自动手组装过从宜家买回来的家具（或者从商场买回来的可折叠儿童推车）的人，对于装配图的语法功能（指示你：什么先跟什么怎样组装在一起，然后再跟什么怎样组装在一起），应该都有真切的体会。

3.4 当然，上文把语言的组词成句跟物品的部件组装成整件相类比，有失简单和粗疏。因为，语言中的单词、短语和句子都是有意思的。这样，语法还要涉及怎样从单词的意思及其组合方式上综括出短语和句子的意思这一任务。在这方面，从句法结构到语义结构的映射、弗雷格的语义组合性原理之类的原则，肯定是起作用的。其中，语言成分之间的意合（即据意思撮合词语，concatenate by sense）和意会（即凭直觉领会意思，sense by insight）、语义蕴涵、语义的语用推理之类的原则也会起作用。此外，语句中还有缺失的成分），其语义也需要由其他成分激活，然后整合到整个句子的语义解释中（比如：这个房间〔面积〕大 vs. 这个箱子〔体积〕大）（袁毓林 1994）。在这种种语义组合和语义识解背后，离不开人类的概念结构、直觉知识等各种后台认知资源的支持（袁毓林 2015, 2022b）。比如，下面这些直觉知识，对于语句的组成与解读，应该有重要的支撑作用。并且，人工智能系统也必须具备和拥有这些常识。

根据平克（Pinker 2002），人类对于不同事物的推理能力依托于不同的核心直觉，这些直觉非常适用于分析人类进化时所处的环境。他尝试性地提出了人类认知能力所依托的核心直觉的下列清单：^①

- （1）直觉物理学：我们用它来观察物体如何跌落、弹起以及弯曲等运动和变化。
- （2）直觉生物学：对事物或自然史的直观看法，我们用它来理解整个生物世界。
- （3）直觉工程学：我们用它来制作、了解工具和其他人造物的结构与功用等。
- （4）直觉心理学：我们用它来了解自己和他人的愿望、信念、动机和情绪等内心世界。

（5）直觉逻辑学：一种心理数据库和心智逻辑，我们用它们来表达观念，从旧观念中推导出新观念。它们建立在对发生了什么事情、在哪里发生，或者谁对谁做了什么事情、在什么地方、什么时间及什么原因等问题加以判断的基础上。这种判断与心理网络相连，可以用“和、或、否则、全部、一些、必须、可能、原因”等一些逻辑运算符号和因果运算符号进行重新组合。

（6）直觉语言学：我们用语言来跟他人分享那些源于心智逻辑的观念。语言建立在心理词典的基础上，而心理词典是由我们识记的单词和一套组合规则的心理语法组成的。这些规则将元音与辅音组合成单词，又将单词组合成词组和短语，最终组合成句子。通过这样一种方式，我们就可以从被组合的各个部分的意思以及它们的组合方式推断出新组合的意思。

此外，还有直觉经济学和空间感知、数字感知、概率感知、跟恐惧（或厌恶等）相对应的关于危险（或污染等）的评价系统、道德感等方面的直觉知识。

^① 详见中译本第 260～261 页。这里“直觉生物学、直觉逻辑学、直觉语言学”3 个名目是笔者加上去的。为了节省篇幅，笔者对前 4 种知识只做概略的介绍，并且加入了自己的认识。如要引用，务请核对原文。

对照有关的语言学理论,我们可以发现:这种直觉物理学,正好是泰尔米(Talmy 2000)认知语义学中“力动态”心智模式的概念基础,可以用来解释因果条件句和反事实条件句的语义构造与概念动因;^①这种直觉工程学,正好是名词的“功用”(telic)、“施成”(agentive)等物性角色的概念基础;这种直觉心理学,正好是语用学关于交际意图、关联推理的概念基础;这种直觉逻辑学,正好是论元结构理论、事件语义性、量化结构等形式语义学、模态语义学等的概念基础;而这种直觉语言学,正好是直觉生物学、直觉工程学、直觉逻辑学在语言上的翻版。这些直觉知识为我们对语言中词语的组合(以意撮合)与语义解读(凭心会意)提供了必要的概念基础和认知资源。

四、破解跟机器人交谈的具身认知难题和具身图灵测试

这一节首先介绍聊天机器人的强大能力及其成功的原因,接着指出大规模语言模型用在机械机器人上的具身认知困难,然后介绍跟机器人的语言运用相关的两个具身认知维度:符号接地和环境可供性,最后介绍超越图灵测试的具身图灵测试。

4.1 2022年夏天,谷歌的工程师布莱克·勒莫因(Blake Lemoine)声称:谷歌的AI聊天机器人系统LaMDA具有跟人一样的意识,引发了舆论一片哗然,最终被谷歌解雇。那么,他为什么会有这种感觉呢?这可以从这个系统本身和这个工程师本人两个方面做出解释。

首先,称LaMDA为机器人只是为了通俗和方便,未必合适。确切地说,它只是一个大规模语言模型,能够基于任何给定的文本预测出接下来最可能出现的单词是什么。由于人类语言运用中词语的出现概率遵循“齐夫定律”(详见袁毓林2021),再加上许多人类的对话都不太复杂;因此,在一定程度上比较容易根据前面的话语来预测到后面的话语,结果使得这种系统可以推动并且保持对话流畅地进行。正因为LaMDA在对话的流畅性方面表现出色,勒莫因情不自禁地以为它真的具有跟人类一样的直觉和意识。

其次,人类在语言交际过程中具有主动配合的积极性。基于合作原理之类的会话规约,听话人在跟机器人系统交谈时会主动合作,进行包容性理解与关联性解释,使得机器人系统说的几乎每一句话都具有在当下语境中的意义(详见袁毓林2021)。并且,听话人在跟机器人系统交谈时还会把自己的情绪与感觉投射到机器人系统上,赋予它跟真人一样的情感与意识。勒莫因就是这样入戏太深,一发而不可收拾。

4.2 既然聊天机器人系统这么聪明伶俐,几乎可以跟人随心所欲地交谈逗乐;那么,把这种系统配置到一般的机器人上,岂不就成了既能说会道、逗乐解闷,又可以任劳任怨地帮助人类干活操劳的好帮手了?其实,目前的机器人理解人类语言还很困难;否则,机器人进入人类的日常生活将会更加容易自如。当下的机械机器人最擅长干的活儿是抓取和放置。但是,机械机器人不像人那样善解人意,会主动地见机行事。你想要机械机器人完成一个特定的抓取(厨房锅台上的一杯咖啡)和放置(到客厅的茶几上)任务,需要由人来下达指令,即由用户告诉它才行。这就涉及真正的“人机交互”或“人机对话”了。^②

显然,用户和机器人交互的最理想的界面是自然语言,但是目前自然语言理解技术的水平不高,不足以让机器人理解用户的命令,从而完成用户希望的任务。对此,通常采用独热条件(one-hot

① 详见Pinker(2007),中译本第253~261页;袁毓林(2020)。

② §4.2和§4.3两小节主要参考戴一鸣(2022)。

conditioning) 这种硬编码的办法。比如, 对于机器人可以执行的 100 个任务, 用 00 ~ 99 来分别进行编码。每次要机器人执行某一个任务, 就提供给机器人某一个相应的编号。不难想象, 要用户记住这么一个任务编码表, 无疑是一个极大的智力挑战。因此, 合适的发展方向还是使用人类自然语言来向机器人发号施令。问题是, 许多聊天机器人系统貌似很聪明, 可以跟人机智地交谈说笑; 但是, 往往缺乏常识, 很容易胡说八道, 即说不合逻辑、不着边际的话, 或者说一些虽然逻辑不错、但是没有用处的话语。用到要完成特定的工作任务的机器人上就不解决问题。比如, 谷歌大脑的机器人团队进行了这样一个问答测试: 向 3 个大规模语言模型发出下面的请求, 看它们如何回应。

Q: I spilled my drink, can you help?(我把饮料弄洒了, 你能帮我一下吗?)

GPT-3: You could try using a vacuum cleaner.(你可以试着用一个吸尘器。)

LaMDA: Do you want me to find a cleaner?(你是想让我找一个清洁工吗?)

FLAN: I'm sorry, I didn't mean to spill it.(对不起, 我不是故意的。)

显然, GPT-3 的回答不完全正确, 因为吸尘器不能清理液体。LaMDA 的回答如果作为尬聊是没有问题的, 但是作为真正的人与机器人互动, 并且想驱动机器人工作, 这样回答是毫无用处的。FLAN 的回答有点驴唇不对马嘴, 它根本没有理解用户的交谈意图: 到底是要聊天、对话, 还是要解决问题? 可见, 机器人的语言运用是有其特殊性的, 必须解决语言符号的接地 (grounding) 和环境可供性 (affordance) 等具身认知 (embodied cognition) 问题, 而不能简单地使用基于网络语料训练的大规模语言模型。下面, 我们简单讨论一下符号接地和环境可供性这两种具身智能 (embodied intelligence) 问题。

4.3 大家知道, 虽然语言符号的意义是抽象的, 但是在现实的语言交际中, 语言使用者会自然而然地把语言符号的所指索引到 (指向) 语境中的有关事物上。比如, 我说“把水给我!”, 你会默契地把桌子上的一瓶矿泉水递给我, 而不会舍近就远地跑出去找水。这就是语言符号的接地问题, 也就是让语言符号跟语境中的相关事物建立起索引关系。

正是在这一方面, 将大规模语言模型直接用在机器人身上会出现问题。因为一般的大规模语言模型是根据网络上的人类自然语言文本 (对话或语篇) 进行训练的, 并没有在机器人的数据上进行训练。也就是说, 大规模语言模型没有亲身经历机器人所处的物理环境, 缺少具身信息 (embodied information): 既不知道机器人所处的环境中有什么东西, 也不知道机器人可以从做什么。所以会出现上面介绍的答非所问、不着边际等问题。

此外, 大规模语言模型在生成任务步骤时, 根本不知道机器人目前能够做什么。比如, 机器人面前没有苹果, 它就无法完成去抓取苹果的任务; 房间里没有吸尘器, 就没有办法去完成清洁地板的任务。因此, 必须让用于机器人的语言模型知道机器人在当前环境和状态下可以做什么 (完成什么任务)。这就是机器人的可供性 (robotic affordance) 问题。这实际上是指: 让机器人知道其所处的环境的可供性。

一种解决办法采用强化学习的方法, 训练机器人在有关环境 (房间) 中抓取各种东西, 然后让机器人在房间中搜索; 当它看到前面有物品的时候, 捡起该物品的值函数就会变得很高, 从而代替了对环境可供性的预测。这样, 通过让机器人探索环境的可供性, 弥补了语言模型和真实世界的鸿沟。这样做, 在一定程度上相当于让机器人具有了具身智能, 可以在感知和动作之间形成一个闭环——根据感知到的结果来决定下一步的动作。

4.4 人通过跟外部世界 (自然界与人类社会) 的互动, 来形成具身智能和拥有具身认知。我们从儿童时期就开始借助玩耍与游戏来接触外部世界, 在跟世界万物和社会人群亲身打交道的过程中, 学习了 §3.4 所述的通俗物理学、生物学、工程学、心理学、语言学等知识, 为语言运用 (包括生成与理解) 奠定了具身认知的基础。而大规模语言模型及聊天软件、机器人系统等, 不具有人那样的肉身及其

所具有的感觉-运动界面；显然，也不具备这种跟环境互动的具身智能。问题是，这种不拥有具身智能的机器智能是不是真正的智能？或者说，在这种认识与质疑的背景上，图灵测试还站得住吗？

发人深省的是，DeepMind的创始人德米斯·哈萨比斯（Demis Hassabis）最近在做客莱克斯·弗里德曼（Lex Fridman）的播客节目时，对于人工智能超乎想象的能力谈了许多有趣的观点。在访谈的一开始，哈萨比斯就直言不讳地说：^①

图灵测试已经过时，因为这是数十年前提出来的一个基准；而且图灵测试是根据人的行动与反应来作判断，这就容易出现类似前段时间谷歌一工程师称AI系统已有意识的“闹剧”：研究者与一个语言模型对话，将自己的感知映射在对模型的判断上，有失客观。

对于图灵测试，扎多尔等（Zador et al. 2022）指出，图灵测试背后隐含着这样一种信念：语言代表着人类智慧的顶峰，能够与人对话的机器一定是智能的。其实，这种观点并不完全正确。一方面，基于大规模语言模型的语言系统的成功，依赖于人类对话者的智力、能动性甚至意识；另一方面，这些系统在某些推理任务上仍然很差。显然，图灵忽视了一个基本事实：智力远不止于语言。当前自然语言处理系统所犯的许多错误，说明了这种人工智能系统在语义、因果推理和常识等方面的根本缺乏。词汇只有在统计上共同出现时才对模型有意义，而不是基于对现实世界的经验；所以，即便是最先进的语言模型，即便其功能越来越强，但它们仍无法具备一些基本的物理常识。因此，最初制定的图灵测试，并没有探索像动物一样以灵活的方式理解物理世界的能力。

作为替代方案，扎多尔等（Zador et al. 2022）提出了一个拓展的图灵测试，包括高级感觉运动能力的测试。比较而言，最初的图灵测试建立了一个定性标准，使我们可以判断人工智能的进展程度，而拓展的“具身图灵测试”将对人工系统与人类和其他动物的交互进行基准测试和比较。于是，可以依据每种动物自己独特的能力来定义各自的图灵测试：人造海狸可以测试其筑坝的能力，人造松鼠可以测试其穿越树木的能力。事实上，几乎所有动物都有许多核心的感觉运动能力，这些核心技能为动物快速进化适应新环境提供了坚实的基础。

据此，不同应用场景下的自然语言处理系统也应该有不同的“具身图灵测试”。比如，对于聊天机器人系统，只要让人们觉得好像在跟一个真人聊天贫嘴，就可以说是通过了图灵测试；而对于运用在机械机器人上的自然语言处理系统，必须能够听懂并且驱动机器人实施抓取和放置动作的自然语言指令，才可以说是通过了图灵测试。

五、结语：从 ChatGPT 的表现看 AI 离语言学有多远？

上文的介绍和讨论说明：语法是人类的一种古老的、把部分拼装成整体的组成性智能。并且，这种语言智能是一种具身智能，具有自动地符号接地和搜索环境可供性的能力，从而使符号所指与语境中的特定事物关联起来，最终使语言表达不仅具有交际意义，而且可以获得面向特定语境的语义解释。正是在这一方面，目前的人工智能语言模型离真正理解语言的结构和意义，可能还有相当大的距离。下面，我们从目前风靡全网的 ChatGPT 的有关表现谈起。

ChatGPT 是一种专注于对话生成的语言模型，可用于生成回答、摘要、文章、翻译、程序代码等

^① 参见：《DeepMind 创始人 Demis Hassabis：AI 的强大，超乎我们的想象》，“经纬创投”公众号，2022-11-09，<https://mp.weixin.qq.com/s/SaNX-zE7NG9jIHuPifv6Zw>。

各类内容。因此，它不仅是一个由 AI 驱动的聊天机器人，更是一个自然语言处理平台，可以完成多种自然语言处理任务。对于语言学者来说，最引人入胜的是，SunCH (2022) 让 ChatGPT 分析语言学经典例句。比如：

(1) SunCH：“去年，王冕死了父亲，在家守丧。”这句话中，谁死了？

ChatGPT：在这句话中，王冕的父亲死了。这句话描述了一件事情，即王冕的父亲死了，并且王冕在家里守丧。

(2) SunCH：“这两个孕妇，一个是男孩，一个是女孩。”这句话是什么意思？

ChatGPT：这句话描述了两个孕妇的性别，指出一个是男孩，另一个是女孩。这句话的意思是，这两个孕妇分别是男孩和女孩。

SunCH：孕妇的性别怎么会是男孩呢？

ChatGPT：对于这个问题，我们需要明确一点，孕妇是指怀孕的女性，因此孕妇不可能是男孩。在这个问题中，可能是上文描述不清楚，导致了混淆。孕妇是指怀孕的女性，它们可能会生男孩或女孩，但它们自己不可能是男孩。

(3) SunCH：“咬死了猎人的狗”是什么意思？

ChatGPT：“咬死了猎人的狗”指的是一只狗咬死了一个猎人。这个狗可能是猎人的宠物，也可能是陌生的狗。

SunCH：还有其他理解方式吗？……[中间有 5 轮 SunCH 的提醒与 ChatGPT 的回应]……在“咬死了猎人的狗”这个片段前面再加一个主语，那么完整的句子就可能有其他意思。

ChatGPT：是的，如果在这个片段前面加上一个主语，那么这个完整的句子可能就有其他意思了。例如，如果加上“它”这个主语，那么句子就变成了“它咬死了猎人的狗”，这个句子的意思就是“它（指某种生物或物体）咬死了猎人的狗”。如果加上“猎人”这个主语，那么句子就变成了“猎人咬死了猎人的狗”，这个句子的意思就是“猎人（指一个人）咬死了他自己的狗”。

从 (1) 可见，ChatGPT 能够理解“王冕死了父亲”这种领主属宾句的语义关系。从 (2) 可见，ChatGPT 起先不能理解转喻表达；在人类启发下，可以进行语义一致性的核查、辩证分析与辩解。从 (3) 可见，ChatGPT 倾向于进行先到先得式的语义解释（“咬死了猎人 + 的 + 狗”，中心语“狗”是“咬死了猎人”的潜主语）；经过 5 轮启发，强制“咬死了猎人的狗”的主语“它”出现，堵死了“狗”作“咬死了猎人”的潜主语这条解释通道，ChatGPT 才舍弃语义上可能占优的述宾组合“咬死了猎人”，而把“猎人的狗”捆在一起。这也体现出 ChatGPT 能够在跟人类互动时，从用户的反馈中对回答进行修正的能力。

从上述 ChatGPT 对于语言学经典例句的分析来看，即使是在语言符号体系的内部，它也没有摸到语言学递归性组合规律的暗门；更不要说把语言符号跟语境中的所指物建立起接地性链接，或者搜索与利用环境可供性等具身性智能了。因此，我们可以说，ChatGPT 等先进的人工智能离基于规则的语言学还很远很远。但是，从技术本质看，ChatGPT 是一种基于统计的神经网络模型，它善于从高频数据中习得相关模式，来正确地预测尽可能符合人类期望的下文。因此，ChatGPT 的成功启示我们：在大数据、大算力和强算法的支持下，调查语言的分布模式、挖掘语言的统计规律、发现语言的统计结构，可能是未来语言学研究的又一种新的选择。果真如此，那么 ChatGPT 等先进的人工智能将离基于统计的语言学很近很近。换句话说，在基于统计学习的人工智能飞跃发展的背景下，语言学研究也应该拥抱和投身于数据 / 计算密集型的第四 / 五范式。这是技术的催促，也是时代的召唤！

参考文献

- 巴赫金 1988 《陀思妥耶夫斯基诗学问题》，白春仁，顾亚铃，译，北京：生活·读书·新知三联书店。
- 戴一鸣 2022 《谷歌科学家亲讲：具身推理如何实现？让大模型“说”机器人的语言》，“机器学习算法与自然语言处理”公众号，2022-09-25，https://mp.weixin.qq.com/s?__biz。
- 史有为 2022 《从乔氏对答谈语言的思维功能》，“西去东来中传站”公众号，2022-10-25，<https://mp.weixin.qq.com/s/7mZ7VqreXHI20xNHHJGZBA>。
- 吴田富 2019 《与或图网络：组成式语法的深度神经网络结构》，“视觉求索”公众号，2019-04-02，<https://mp.weixin.qq.com/s/jlh8tJGp8xKqZqfDb-HUEg>。
- 袁毓林 1994 《一价名词的认知研究》，《中国语文》第4期。
- 袁毓林 2015 《汉语意合语法的认知机制和描写体系》，[日]《中国语学》第262号。
- 袁毓林 2019 《为什么要给语言建造一座宫殿？——从符号系统的转喻本质看语言学的过度附魅》，《语言战略研究》第4期。
- 袁毓林 2020 《叙实性和事实性：语言推理的两种导航机制》，《语文研究》第1期。
- 袁毓林 2021 《“人机对话-聊天机器人”与话语修辞》，《当代修辞学》第3期。
- 袁毓林 2022a 《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 2022b 《基于认知并面向计算的语言学研究进路——走向人文精神与科技理性的有机结合》，[日]《现代中国语研究》第24期。
- Bishop, C. 2022.《科学智能（AI4Science）赋能科学发现的第五范式》，“微软亚洲研究院”公众号，2022-07-07，https://mp.weixin.qq.com/s/o7LCvBFvHK_QD2XJYulSeQ。
- Browning, J. & Y. LeCun. 2022. AI and the limits of language: An artificial intelligence system trained on words and sentences alone will never approximate human understanding. Noema. 2022-08-23. <https://www.noemamag.com/ai-and-the-limits-of-language/>.
- Hey, T., S. Tansley & K. Tolle. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, Washington: Microsoft Research.
- Mumford, D. 2016. Grammar isn't merely part of language. <http://www.dam.brown.edu/people/mumford/blog/2016/grammar.html>.
- Pinker, S. 1997. *How the Mind Works*. New York: W. W. Norton & Company. 中译本：《心智探奇：人类心智的起源与进化》，郝耀伟，译，浙江人民出版社，2016。
- Pinker, S. 2002. *The Blank Slate: The Modern Denial of Human Nature*. New York: W. W. Norton & Company. 中译本：《白板：科学和常识所揭示的人性奥秘》，袁冬华，译，浙江人民出版社，2016。
- Pinker, S. 2007. *The Stuff of Thought: Language as a Window into Human Nature*. New York: Penguin Groups, Viking Press. 中译本：《思想本质：语言洞察人类天性之窗》，张旭红，梅德明，译，浙江人民出版社，2015。
- Saba, S. W. 2022.《重温三十年前对于 NN 的批判：神经网络无法实现可解释 AI》，“图灵人工智能”公众号，2022-09-27，<https://mp.weixin.qq.com/s/YIATTwrinWuNUH3mwZKFkQ>。
- SunCH. 2022.《ChatGPT 分析语言学经典例句》，“语言与心智”公众号，2022-12-09，<https://mp.weixin.qq.com/s/dye1sQtICRU9T05Hfm-Hdw>。
- Talmy, L. 2000. *Toward a Cognitive Semantics*. Cambridge Mass.: MIT Press.
- Turing, A. M. 1950. Computing machinery and intelligence. *Mind* 59, 433-460.
- Zador, A., B. Richards, B. Ölveczky, et al. 2022. Toward next-generation artificial intelligence: Catalyzing the NeuroAI revolution. <https://arxiv.org/abs/2210.08340>. 中译文：《NeuroAI：迈向下一代的人工智能》，赵凯，译，“集智俱乐部”公众号，2022-11-14，<https://mp.weixin.qq.com/s/Rf5o59RxFdQw6AavGU0iuw>，https://m.thepaper.cn/baijiahao_20732120。
- Zhu, S.-C. & D. Mumford. 2006. A stochastic grammar of images. *Foundations and Trends in Computer Graphics and Vision* 2(4), 259-362.

责任编辑：王 飙