从范式嬗变看语言学与人工智能的融合路径*

詹卫东

（北京大学　中文系　北京　100871）

提　要　从规范语言学、比较语言学、结构语言学到生成语言学，语言学跨越千年的范式变迁，终极目标是为人脑内部语言建模。语言学者虽已积累大量微观语言学成果，却难以由量变引起质变，距建构整体语言模型仍有巨大鸿沟。自 1950 年代至今，人工智能技术范式从符号主义、经验主义发展到联结主义，生成式大语言模型已经可以依靠强大算力和海量数据，对人类语言进行全量建模。但在挖掘表层分布没有充分表达的深层认知规律方面，还无法达到母语者水平。语言学者可以基于自身的语言学洞察力，充分挖掘微观语言现象中蕴含的深层语义问题，把语言知识转化为高质量语言数据，帮助提高人工智能的语言能力。同时利用人工智能技术，在观察充分、描写充分、解释充分的基础上，进一步实现“生成充分”。如何由理论研究成果驱动，由纯手工到半自动再到全自动地生成能与人工智能直接交互的语言数据，是人工智能时代语言学者要认真思考的问题。

关键词　生成式人工智能；形式文法；大语言模型；深度学习

中图分类号　H002　**文献标识码**　A　**文章编号**　2096-1014（2026）03-0041-12

DOI　10.19689/j.cnki.cn10-1361/h.20260303

Rethinking the Integration of Linguistics and Artificial Intelligence Through Paradigm Shifts

Zhan Weidong

Abstract　Across its long intellectual history, from prescriptivism and descriptivism to structuralism and generative grammar, linguistics has the ultimate goal of modelling the internal language faculty of the human mind. Despite the accumulation of extensive body of micro-level findings, a substantial gap persists between such fragmented insights and the construction of a holistic model of human language as a cognitive system. Since the 1950s, the dominant paradigm trajectory in artificial intelligence has evolved from symbolism through empiricism to connectionism. Although contemporary large language models (LLMs), powered by leveraging massive computation and vast datasets, can model human language at unprecedented computational power, they remain limited in capturing the deep cognitive regularities that are not fully encoded in surface distributions and therefore still fall short of native-speaker competence in extracting and internalizing deeper cognitive and semantic regularities that are weakly expressed in surface distributions alone. In this context, it opens a critical space for collaboration. Linguists, drawing on their theoretical sensitivity to subtle semantic, pragmatic, and structural phenomena, can contribute by leveraging domain-specific insight to uncover the deep semantic issues embedded in micro-level linguistic phenomena and to transform linguistic knowledge into high-quality, structured data for improving AI’s linguistic capabilities.

*　作者简介：詹卫东，男，北京大学教授，主要研究方向为计算语言学、语言知识工程、中文信息处理。电子邮箱：zwd@pku.edu.cn。

教育部人文社会科学重点研究基地重大项目“面向机器语言能力评测的综合型语言知识库研究”（22JJD740004）。本研究同时得到多媒体信息处理全国重点实验室开放课题基金（SKLMIP−KF−2025−01）的支持，特此致谢！

At the same time, AI offers linguistics the possibility of moving beyond observational, descriptive, and explanatory adequacy toward generative adequacy. Consequently, the automated transformation of theoretical research into interactive linguistic data—evolving from manual to fully autonomous processes—constitutes a central challenge for linguists in the AI era, and addressing this challenge is essential if linguistics is to play a constitutive role in the next stage of language modelling.

Keywords generative artificial intelligence; formal grammar; large language models; deep learning

一、引　言

“生成语法”与“人工智能”（AI）在 1950 年代几乎同时问世。前者以乔姆斯基为代表，首次提出“语言何以能无限生成”的惊世之问（Chomsky 1957）；后者则以图灵关于“模仿游戏”的设想（Turing 1950）以及麦卡锡等人提出的“达特茅斯计划”（McCarthy et al. 1955/2006），宣告了将人类智能机械化的研究正式起步。此后 70 余年，语言学与人工智能一方面沿着各自的轨道发展，一方面也因共同关注的对象“自然语言”而有机会不时结伴同行。前者不断细化和扩展语音、词汇、句法、语义、语用等层面的理论分析模型，后者则在基于符号的规则方法、基于特征的统计学习和基于神经网络的深度学习等范式转型中屡屡突破。从研究的终极目标来看，现代语言学与人工智能研究都旨在探索人类语言能力的基本原理。然而，从实际成效来看，语言学距离完成其理论目标仍遥遥无期；而随着 2022 年 11 月底 ChatGPT 发布，生成式人工智能在工程层面第一次呈现出大规模、可迁移的跨语言交互能力，甚至被视为触及通用人工智能（AGI）的边缘。两条轨道的速度差如此之大，迫使学界不得不反思：在生成式人工智能时代，语言学应如何重估自身的方法论路径，与人工智能研究更好地互动？（袁毓林 2025）

本文尝试将镜头拉远，在历史纵深中考察语言学与人工智能研究的范式嬗变，从较为宏大的视角来对比这两个领域主流工作模式的差异，探讨语言学与人工智能融合发展的具体路径。第二节在两千年大历史尺度上概括语言学观念的变迁。第三节梳理人工智能技术范式从符号主义到经验主义再到联结主义的跃升。第四节参考人工智能的技术演进之路，反观人类语言学研究模式的优势与不足，进而提出由语言学知识驱动，构建语言能力评测任务及数据集的研究思路。主张通过严谨的语言能力评测任务设计，将语言学知识转化为评测数据集，专注于可验证的语言学实证研究，实现语言学数据可持续的规模扩展。第五节是结语，以积极态度展望语言学与人工智能融合之路的前景。

二、语言学的范式嬗变

把“语言”作为一个科学研究的对象，其实是很晚近的事情。人工智能的早期研究者之一特里·维诺格拉德在《语言作为认知过程》（Winograd 1983）一书中，借鉴托马斯·库恩的科学范式革命理论，把语言学的历史概括为 4 个范式：规范语言学，比较语言学，结构语言学，生成语言学。其中第一个范式严格来说并不适合称为“语言学”，而更应称之为“语文学”。这个阶段从古代一直延续到 17、18 世纪，人类在这漫长的时间里对语言研究关注的核心问题，一言以蔽之就是：**这句话是什么意思？**即重视对个别的、具体的字词句的意义阐释。

随着 14、15 世纪欧洲文艺复兴、宗教改革等一系列重大事件的发生，世界历史开始进入一个崭新的时期。大航海、殖民扩张、宗教和学术传播推动了世界范围的语言接触。比较语言学由此兴起，

并以一个全新的问题为中心：**语言如何演化？**语言学者开始在更广阔的时空尺度上比较不同语言，追溯它们的谱系关系和共同来源。威廉·琼斯关于梵语、希腊语和拉丁语同源性的著名论断，正是这种震撼性发现的缩影。为了重建语言家族树、构拟原始祖语，研究者不仅依赖历史文献，也开始系统记录活态语言的实际用法。对共时系统的兴趣不断上升，开始超过对语言的历时演变研究，为结构语言学的出现准备了条件。

结构语言学关注的问题可以概括为：**意思从何而来？**既不同于古典语文学着眼于个别字词句的释义，也不同于比较语言学关注语言演化的历时过程，结构主义语言学者更强调研究语言的当下和语言系统本身。结构主义语言学认为，只有建立起一套科学的分析方法，把语言看作一个符号系统，根据系统的组合规则，通过对比离析出基本单位，小的单位组合成大的单位，由简单到复杂，最终才能形成完整的解释语句意义的可靠程序。索绪尔首次提出“语言”和“言语”的区分，主张以前者替代后者作为现代语言学的理论研究对象，把语言符号跟其意义的联系，置于系统观念之下来审视，尝试用更稳定、更具规律性的依附于社会群体的“语言”系统，来驾驭表面上千变万化捉摸不定的个人的“言语”实践。

乔姆斯基于20世纪中叶创立的生成语言学，则进一步提出了更具震撼性的问题：**语言如何生成？**这彻底改变了现代语言科学的面貌。乔姆斯基在索绪尔关于“语言”和“言语”区分的基础上进一步提出“语言能力”和“言语表现”的对立。这一对立跟索绪尔理论的最大不同在于，“语言能力”的主体是个体人脑。语言研究应关注如何为人脑心智构建语言结构系统，该系统要能够解释人的无限句子生成能力，即人为什么能说出以前从未说过的话，为什么能听懂以前从未听过的话。

乔姆斯基之前的语言学关注的是已经说出来的句子——语言的外化成品，而乔姆斯基心目中的语言学研究对象，是产生句子的原因——人脑中的语言机制。乔姆斯基将人的语言能力高度概括为“有限手段的无限运用”，引入数学的形式化方法，使语言学的研究重心从如何理解句子扩展到如何生成句子，首次把语言研究跟人脑心智结构联系起来，主张语言学应超越对外在言语现象的描写，深入到心智之内，解释语言的无限生成性。从这个意义上讲，语言学跨越千年的范式变迁，一言以蔽之，是研究目标由外而内的逐步深化。

尽管乔姆斯基为语言学研究规划的人类普遍语法愿景无疑是值得探索的，但具体的路线图一直没有很好的着落。多数语言学研究是针对具体微观的语言现象，观察总结其中成分组合的约束条件，归纳语言符号形式和其意义之间的对应模式。人类语言研究者已经积累了大量的这类微观语言学研究成果，可是，这样的微观成果，如何聚沙成塔，形成一个整体的语言模型（普遍语法）呢？从微观到宏观，从局部语言知识到整体语言模型，中间似乎横亘着一个巨大的鸿沟。到底通过什么途径，才能从现实的语言现象观察和分析出发，到达为人类语言能力建模的理想彼岸呢？用乔姆斯基提出的“内部语言”和“外部语言”的概念来说，语言学研究的终极目标是为内部语言建模，可现实环境却只能是对外部语言的剖析。连接现实与终极目标的通路到底为何，一直是未解的谜题。

三、人工智能的范式嬗变

从1950年代生成语言学提出至今，人类语言学家仍未找到“语言如何生成”的确切答案。结果出人意料地，生成式人工智能的研究，提出了一个全新的候选答案。虽然未必是最终正解，但至少在

工程实践和应用效果上，机器已然能像人一样生成自然语言了。回顾过去近 70 年机器语言能力不断提升的历程，人工智能的技术方法大致可以概括为 3 个范式：（1）早期规则推理与专家系统主导的**符号主义范式**；（2）以浅层统计学习和特征工程为核心的**经验主义范式**；（3）近 20 年迅速崛起的以深度神经网络和表征学习为支点的**联结主义范式**。每一次范式升级都是在围绕“如何让机器具有人类语言能力”这一问题重写技术路线。

在符号主义范式下，研究者主张一切智能行为都可还原为对符号的操作（Newell & Simon 1976）。研究的核心任务是设计基于符号的显式语法与推理规则，即由知识工程师手工编码语言规则与领域知识。符号主义范式的优势在于推理链条清晰、结果可解释，但其知识获取严重依赖专家经验，难以覆盖开放世界的语言现象多样性，存在显著的“知识获取瓶颈”（ALPAC 1966；Lighthill 1973）。

1980—1990 年代，随着计算能力提升与大规模语料库的逐步建立，人工智能——尤其是语言技术——开始出现从符号主义到基于统计的经验主义范式的转向。以语音识别与机器翻译为代表的应用场景中，隐马尔可夫模型、n 元语言模型、最大熵模型等一批统计方法成为主流（Brown et al. 1990；Jelinek 1998）。新方法不再完全依赖小规模的人工规则，而是参考人类提供的先验特征，在大规模真实语料上进行参数估计。研究者开始把“模型在真实分布上表现如何”作为主要评价标准，用降低误差率来优化系统。这一阶段推动了语料库语言学和计算语言学的发展，句法树库等标注语料资源成为连接语言学理论与工程实践的重要桥梁。不过，从语言能力建模的角度看，基于统计机器学习的经验主义范式仍然停留在对语言符号“浅层分布”规律的学习层面：典型的 n 元统计模型只能在有限长度的窗口内建模词序列的共现约束，难以捕捉长距离依赖与复杂语义结构。解决这一问题的，则是接下来登场的基于深度学习的人工智能联结主义范式。

联结主义范式的建立并非一蹴而就，而是经历了“思想—算法—模型”长达 70 年的曲折攀升（见图 1）。其思想源头可追溯到感知器与早期神经网络（Rosenblatt 1958；Minsky & Papert 1969）；1980 年代“并行分布式处理”（PDP）框架把知识表征推进到“分布式表示”的方向（Rumelhart, McClelland & PDP Research Group 1986），而反向传播（梯度下降训练多层网络）则为多层网络的可训练性提供了关键的数学与工程基础（Rumelhart, Hinton & Williams 1986）。只是长期以来，受限于数据稀缺、算力不足与训练不稳定，神经网络难以在大规模真实任务上落地，使得联结主义方法在相当长时间内承受着“方向可能正确但难以规模化兑现”的质疑压力。

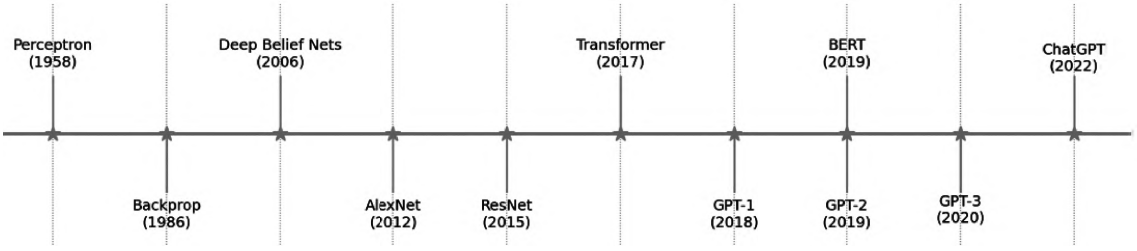

图 1　深度学习（神经网络）技术路线发展里程碑

进入 2010 年代，各项条件的成熟使这一路线仿佛在一夜间突然崛起：数字化与互联网浪潮的多年积累沉淀了可用的海量数据，自然语言处理公开基准与竞赛推动了可重复的评测文化；GPU 芯片将大规模矩阵运算变成廉价算力，叠加分布式训练、云计算与软件框架的成熟，使“更大数据 + 更大模型 + 更长训练时间”成为可持续的工程策略。深层网络首先在视觉任务上通过大规模数据与训练

结构优化（如残差连接、归一化等）解决了训练稳定性问题，确立了端到端学习的优势（Hinton et al. 2006；Krizhevsky et al. 2012；He et al. 2015）；随后这一方法论扩展到自然语言处理领域，Transformer 自注意力架构进一步把并行化与长距离依赖建模推到新水平（Vaswani et al. 2017），并催生出“大规模自监督预训练 + 下游任务适配（微调 / 指令化 / 对齐）”的主导路线（Devlin et al. 2019；Brown et al. 2020）。最终，模型参数规模、互联网级训练数据与并行算力三者耦合，使模型在复杂语义理解、跨任务迁移与生成能力方面大幅提升，再结合人类反馈对齐等方法，共同造就了生成式大语言模型的成功，也确立了联结主义人工智能的范式地位（Radford et al. 2018，2019）。从某种意义上讲，以 ChatGPT 为代表的生成式大语言模型实际上是把“从海量语言事实中归纳语言规律”这一经验主义语言学思想做到了极致。

综上所述，与语言学范式变迁聚焦于研究目标的变化不同，人工智能研究的范式嬗变并不涉及研究目标的变化，而是在如何实现目标的具体方法上不断提升。表 1 从模型假设、知识表示形式、知识来源（学习方式）、基础资源、优势与劣势等维度对 3 种范式的差异进行了概括呈现。

表 1　人工智能 3 种范式的对比概览

维度	符号主义范式	经验主义范式	联结主义范式
模型假设	语言是离散符号系统，依靠语法规则和逻辑进行推理与生成	语言是一个概率随机过程，依靠特征 + 概率分布模型刻画语言符号共现规律	语言符号映射到连续向量空间，语义可以用向量的空间关系来表达
知识表示形式	词典、产生式规则、逻辑公式	基于特征的统计模型参数（HMM、n 元模型、CRF 等）	高维向量、网络连接权重、中间隐藏层激活模式
知识来源（学习方式）	人工设计 无机器学习	人工设计 机器浅层（有监督）学习	无需人工设计 机器深度（半监督）学习
基础资源	基于小规模语料的专家知识	中等规模标注语料库，知识库、特征工程 + 普通算力	海量文本、代码、多模态数据 + GPU/TPU 高性能算力
优势	推理链条清晰，易于形式验证和解释	单任务性能强，鲁棒性好	兼具语言理解和生成能力，跨任务泛化能力强
劣势	知识获取成本高、覆盖范围有限	难以处理深层语义和长距离依赖，无跨任务泛化能力	内部机制不透明，难以精确控制和解释

人工智能的进阶方式是不断抽离对人脑的依赖，而依靠强大计算能力和存储能力，直接对人类数据（外部语言）进行全量建模：不像语言学那样区分不同层级的模块（词法、句法、语义、语用等），而是统一用预测下一个词元（token）的任务，来优化模型对任意符号分布位置的概率估计。这实际是对结构主义语言学分布思想（意义即用法）的终极实践：对任意一个语言字符（串）的意义理解，本质上就等于“记录”它的全部分布环境。

四、语言学与人工智能的互动：语言知识驱动的数据资源构建

生成语法与生成式人工智能都含有“生成”这个概念。从研究目标来讲，都是要实现能够无限生成语言的“生成”模型；但从研究实践来讲，语言学研究的“生成”模型，如果不涉及人脑生物层面

的机制问题[①]，仅从符号层面来说，其实质是从对象语言抽象出元语言，即设计一套更为精简的、有限的符号表征体系，来推导现实世界中无限的、多样的言语表现。这作为研究理念没有问题，但在具体工作方法层面却大有问题。（1）要为极其复杂的自然语言设计一套元语言表征体系，实际工作都是由语言学者的个体经验驱动的，即通过对局部微观语言现象的观察，来归纳针对局部语言现象规律的抽象规则。这些局部规则承载的语言学知识，其实很难拼接起来成为一个整体模型，去覆盖无限的对象语言集合。（2）随着时间的推移，尽管考察的局部语言现象在不断增多，但个体语言学者的努力并不能形成规模效应。即从工程视角，很难看到量变引起质变，找到大一统的元语言模型的现实路径。大量分散的局部语言学知识并不能自然地整合为全局的语言学知识。

而现代生成式人工智能跟语言学的工作方法迥然不同：放弃清晰可解释的元语言体系的构建，转而用纯数学、纯计算的方法，将自然语言符号映射到高维向量空间，把“意义”转化为词元在空间中的分布模式，实现“你有来言，我有去语”的无限词语接龙能力。在广度上，生成式人工智能对自然语言的观察，是“上帝”全局视角而非人类语言学家的局部微观视角。在深度上，生成式人工智能是符号意义的向量化多维联结重构，而不是从元语言到对象语言的规则映射。

基于上述对比，我们有以下两点认识。（1）在观察语言符号的表层分布规律这个层面，机器比人有显著的优势；但在透过现象看本质、挖掘表层分布没有充分表达的深层认知规律方面，人比机器有优势。（2）人对局部语言现象的深入分析所获得的成果，不应止步于语言学知识的表征形式层面，而应该探索如何进一步将语言学知识转化为语言数据。这样才能以数据的形式，实现语言学知识的持续规模扩展，同时便于在人工智能应用场景中对知识加以验证。

所谓“上帝的归上帝，恺撒的归恺撒”，我们的思路是让人脑和人工智能各自发挥优势，探索出一条可行的优势互补之道。下面分两部分来阐述这个思路。先通过示例展示人在微观语言现象的观察和分析方面的优势，再讨论将语言学知识转化为语言数据的具体做法。

（一）语言符号形式意义对应规律的微观研究

语言学家日常的研究工作主要是探究具体微观的语言现象中形式和意义之间的对应规律，大体可分为“异形同义”和“同形异义”两方面讨论。下面4组示例中，前两组是异形同义，后两组是同形异义。

（1）a. 吴丰连忙写了张五百块钱的领条递过去。会计转身打开**身后**的保险柜，先将两张条子放进去，又随手取了一张条子，一边递给吴丰，一边说……

b. 吴丰连忙写了张五百块钱的领条递过去。会计转身打开**面前**的保险柜，先将两张条子放进去，又随手取了一张条子，一边递给吴丰，一边说……

（2）a. 我不是公论家，**有**上帝一般决算功过的能力。

b. 我不是公论家，**没有**上帝一般决算功过的能力。

（3）a. 张老师出差回来带给我们**一人一本书**。

b. 张老师出差回来带给我们**三人一本书**。

（4）a. The city councilmen refused to give the demonstrators a permit for a demonstration because **they**

① 限于研究条件，大多数语言学研究工作无法触及语言的“生物属性”。关于人类语言模型的研究，最接近“内部语言”的，是在形式化层面构建抽象的语言符号规则表征体系，而不是生物学意义上的神经表征。

feared violence. (市议员们拒绝给示威者发放举行示威活动的许可，因为他们担心暴力。)

b. The city councilmen refused to give the demonstrators a permit for a demonstration because **they** advocated violence. (市议员们拒绝给示威者发放举行示威活动的许可，因为他们鼓吹暴力。)

例（1a）和（1b）两句中“身后”和“面前”这两个词不同，但两句整体语义并无差异，可以表达相同的场景（会计打开保险柜时保险柜都在其面前）。例（2a）和（2b）两句中“有”跟“没有”是对立的，但两句整体语义却无差异，都表示“我没有决算功过的能力”。例（3a）和（3b）中的“一人一本书”和“三人一本书”虽然具体数字“一”和“三”不同，但从模式层面来说是完全相同的，都是“数＋名＋数＋量＋名”模式；然而这两句的语义解释却完全不同——例（3a）是每一个人得到一本书，例（3b）是三人一起得到一本书。例（4a）和（4b）中的 they 指代句中不同的名词，例（4a）中指代 councilmen（市议员），例（4b）中指代 demonstrators（示威者）（此例出自 Winograd 1971）。

语言学者的工作模式是从上面这些具体的例句扩展出去，追问用什么样的语言知识来解释同类现象的共同规律。比如：在什么条件下“面前”和“身后”的方位对立义会消失？在什么条件下“有”和“没有”的真值对立会消失？在什么条件下“数＋名＋数＋量＋名”模式会分化出不同的语义解释？如何判定句子中的第三人称代词（如 they）指称哪个成分？为回答上述问题，语言学者就要在对象语言（即这些例句或同类句子的显性成分）中寻找线索，或者假设一些元语言范畴（句子中未出现的隐含成分）来进行推理。比如“面前”和“身后”对立消失的条件，很显然需要引入“时间”范畴，即主体转身前（对应方位“身后”）和转身后（对应方位“前面”）这两个不同时点，是导致“面前”和“身后”指向同一个方位的决定性因素。在这类异形同义现象中，“转身、扭头”类动作是语境中必不可少的伴随现象。而“有”和“没有”对立的消失，则可以归结为相关谓语成分与主语成分本身的概念义是否匹配。比如例（2a）中“有上帝一般决算功过的能力”是跟其近距离主语“公论家”绑定匹配的，而例（2b）中“没有上帝一般决算功过的能力”不能跟“公论家”匹配，只能跟其远距离主语“我”绑定匹配。由此可以推理得到：无论例（2a）还是（2b），对于“我”来说，都“没有上帝一般决算功过的能力”。为让语言知识推理可计算、可自动化，就需要定义严格的元语言符号体系，将这样的知识进行结构化和形式化表达。图 2 和图 3 简要展示了解释例（3）和（4）中 a、b 句的语言学知识。

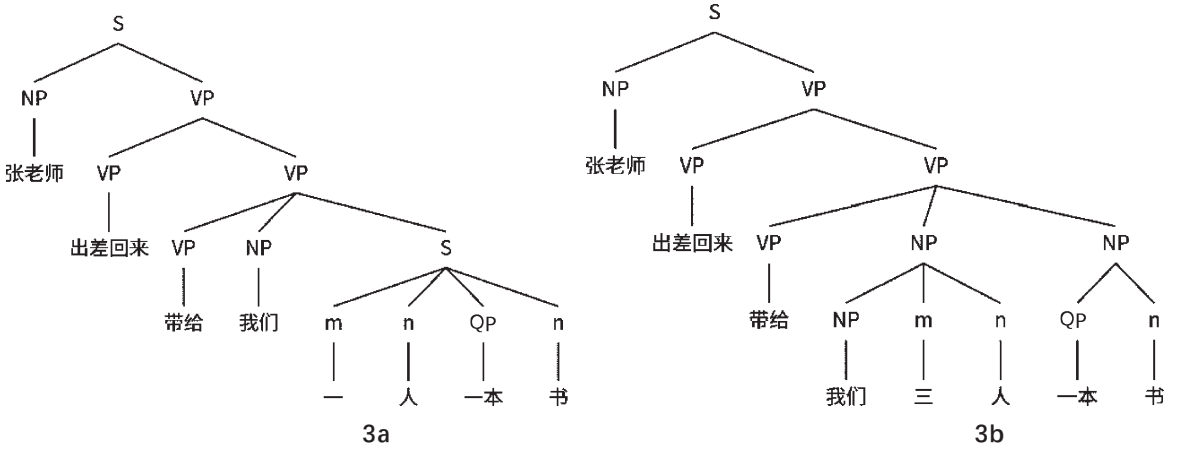

图 2　例（3a）和（3b）的句法结构分析树图

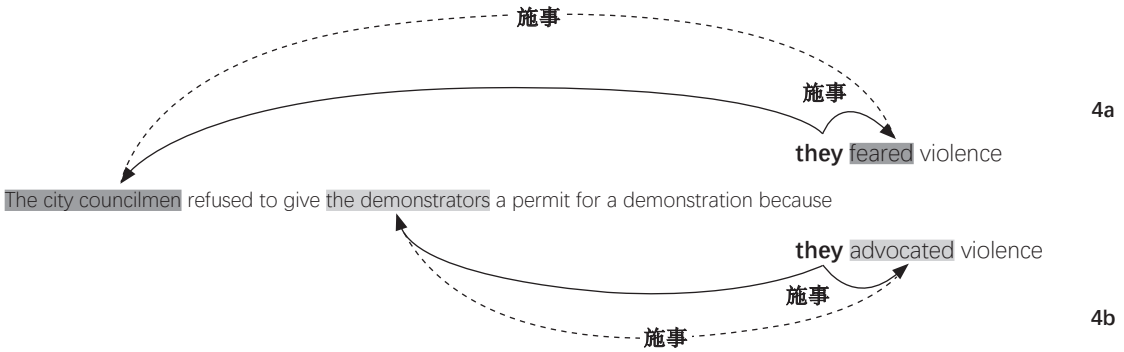

图 3 例（4a）和（4b）的语义指向分析图

图 2 中 S（句子），NP（名词性短语），VP（动词性短语），QP（数量短语），m（数词），n（名词）等，以及由它们组织起来的树状层次（嵌套）结构，图 3 中的“施事”，都是语言学假设的元语言范畴体系。语言学通过设计这些元语言范畴及其相互之间的组合约束，来实现无限生成对象语言并给出对象语言的合理语义解释的目标。尽管这一模式可以在局部语言现象上运转工作，但却很难铺排开来，拼装成一个完整的、宏观的，能够灵活地生成真实自然语言的大一统模型。语言学在大量微观语言现象上提炼归纳的离散的语言知识（元语言系统），其实缺少有效的黏合剂来完成总装。而且元语言知识与对象语言之间的映射关系无论在覆盖面还是准确度上，都存在大量的偏差或不确定性。比如 n、v 的集合如何定义，“施事”关系对应哪些具体的 n-v 组合，等等。仅依靠语言学者的个体经验，哪怕是以团队作战愚公移山的精神去尝试，目前来看，也难以实现积跬步以至千里的元语言知识累积效应。

针对上述语言现象，普通母语者不需要任何显式的元语言知识就能自如地运用语言。也许母语者的大脑中存在某种形式的跟上面分析例（1）—（4）所谈到的元语言知识类似的知识，但母语者头脑中是所谓的“默会知识”，而非清晰的基于元语言范畴体系的结构化知识。普通母语者可以很容易地对例（1）—（4）这 4 组句子中 a、b 两句的语义进行理解，在交际活动中做出正确的反应，但未必能像语言学家那样分析，给出专业解释（基于元语言知识体系）。类似地，现在的生成式人工智能也能对这些句子进行语义理解并做出响应，但是还无法完全达到母语者水平，可能在响应中会出现错误判断。如人工智能不知道例（1）中 a、b 两句其实描述的是相同的场景，也不知道例（2a）和（2b）其实同义。从这个角度看，人工智能就很需要像例（1）—（4）这样的高质量语言数据，即经过精心设计和组织的外部语言，来帮助它获取更高阶的语言认知能力和超越符号表层分布规律的深层语义理解能力。

（二）语言学知识驱动的人工智能语言能力评测任务设计及数据集研制

微观语言现象的观察和分析是人类语言学者的强项。在研究过程中，语言学者自觉地调用专业分析能力，基于语言事实开展可控对比实验，可以挖掘出一般母语者习焉不察的、更深刻的语言使用规律。而其中的一些规律，在目前人工智能训练模式下，即便有海量自然语言数据，可能也难以从中自动获得（如上节 4 组例句）。语言学者如果能基于自身的语言学洞察力，充分挖掘微观语言现象中蕴含的深层语义问题，制作系统的、结构化的测试题，就可以把语言知识转化为高质量语言数据，用统一的数据接口与大语言模型对话。一方面，这类测试题有可能可以稳定地揭示模型在特定语义机制上的短板，并为针对性训练提供靶点；另一方面，持续积累的高质量题库也可以成为研究人类语言能

力机制（包括元语言模型）的可复用材料。在生成式人工智能时代，**本文认为有必要在语言学原有的“解释充分”目标基础上，再往前跨出一步，提倡一种更符合当下科研生态的研究目标——“生成充分”（Generative Adequacy）**。跟对语言现象的“解释”相比，生成符合特定约束条件的真实语例，并据此制作测试题，难度更大。特别是在工程上要求自动大批量生成符合特定条件的高质量真实语例，难度之大，是前沿大语言模型也难以胜任的。当语言学者针对特定语言现象的分析能够稳定地产出一致性高、低噪声、可复现的测试语料时，这种研究模式不仅可直接用于检测并优化人工智能模型能力，也在方法论上提升了语言学知识本身的“**可操作性**”和“**可验证性**”。

事实上，计算语言学领域早已有类似这一思路的工作。其中非常典型的例子就是受到例（4a）和（4b）对照句启发而研制的维诺格拉德指代消解挑战任务（Winograd Schema Challenge，WSC）。该任务以代词指代消解测试为外壳，评估机器的常识推理能力。2012 年正式公布的英文 WSC 试题共 273 题（Levesque 2011；Levesque et al. 2012）。这一范式也已有中文评测——CLUEWSC 2020（Xu et al. 2020）。[①] 相比于英文 WSC 数据集，中文 WSC 数据集规模更大（训练集、验证集、测试集题量分别是 1244 题、304 题、290 题，共 1838 题），但语料形式不再是像例（4a）和（4b）那样的“双胞胎句”——两句只有一处词语差异，该差异导致代词指代翻转。从这点上讲，中文 WSC 降低了语料收集难度。表 2 是中文 WSC 测试题示例。

表 2　中文 WSC 测试题示例

文本	片段 1	片段 2	同指
裂开的**伤口**涂满尘土，里面有碎石子和木头刺，我小心翼翼把它们剔除出去。	伤口	它们	否
裂开的伤口涂满**尘土**，里面有碎石子和木头刺，我小心翼翼把它们剔除出去。	尘土	它们	否
裂开的伤口涂满尘土，里面有**碎石子和木头刺**，我小心翼翼把它们剔除出去。	碎石子和木头刺	它们	是

这 3 道题共享一个完全相同的文本，其中“它们”指代“碎石子和木头刺”，而非指向“伤口”或“尘土”。通过对比设计，这 3 道题相当于提供了指代消解任务的正样本和负样本。

像 WSC 这样的评测数据集，同时也是专项语料库，即以某种语言学知识范畴（比如指代消解）为主题的语料集合。跟人工智能的前两种范式“符号主义范式”和“经验主义范式”时代的语言学知识库和标注语料库明显不同的是，WSC 数据集让语言学知识（元语言）隐身幕后，前台交互完全是自然语言（对象语言）形式。这是语言学与大语言模型人工智能打交道更合适的接口。像例（3a）和（3b）对照展示的汉语构式“数 + 名 + 数 + 量 + 名”的多义现象，在传统知识库或语料库标注中，是显式地用“分配构式义”标签来区分例（3a）和（3b）的差异，并配合像图 2 所示的句法分析树图。这种表征形式必然基于大量的、显式的语言学知识假设。但更适合大语言模型的数据形式，则是通过自然语言问答句对来呈现对具体句子语义的理解。无论是大语言模型还是普通人，都可以不依赖对语言学元语言知识体系的理解来完成回答。

CCL-CUE（Construction Understanding Evaluation）是我们依托北京大学 CCL-CxnBank 构式数据库正在构建的中文构式语义理解能力多任务评测基准数据集（詹卫东 2021）。[②] 具体做法是：通过自编

① WSC 任务英文和中文数据集示例可访问 https://ccl.pku.edu.cn/static/evaldata/index.html 查看。

② CxnBank 构式数据库，https://ccl.pku.edu.cn/ccgd。CCL-CUE 多任务构式语言能力评测数据集，https://ccl.pku.edu.cn/ccgd/llmeval/。

例句或改编真实语料中的构式用例，采用“最小形式扰动”策略，设计分组可对照的选择题，旨在考查大语言模型对构式整体意义的理解，对构式表达和同形普通短语之间意义差异的辨识能力，对多义构式的区分能力，以及对构式适用语境的判断能力。该数据集以多个任务联动方式来全方位地评估大语言模型是否真正掌握构式语义和用法，将构式知识库内容转化为评测数据集，探测大模型语言能力的短板，同时也有助于加深我们对中文构式知识的理解。

表 3 是 CCL-CUE 数据集的示例。以其中第 1 题为例，2025 年 10 月我们选取了国内外 8 个前沿大语言模型进行测试，其中 4 个模型（Claude-sonnet、Claude-opus、Deepseek-reasoner、GPT-4o）选取的答案是 B、C。错选答案 B 说明侧重推理训练的大语言模型存在过度推理现象。

表 3　CCL-CUE 中文构式理解能力多任务评测基准数据集示例

文本	问题	选项	答案
张老师出差回来带给我们三人六本书。	针对 text，下列说法正确的是？	A. 张老师一共带回十八本书。 B. 张老师给每人带回两本书。 C. 张老师给三人带回来六本书。 D. 张老师给每人带回来六本书。	C
张老师出差回来带给我们一人一本书。	针对 text，下列说法正确的是？	A. 张老师一共带回一本书。 B. 张老师给每人带回一本书。 C. 张老师给三人带回来三本书。 D. 以上都不对。	B
张老师出差回来带给我们一人两本书。	针对 text，下列说法正确的是？	A. 张老师一共带回两本书。 B. 张老师给我们带回了两本书。 C. 张老师出差回来时带了书给我们。 D. 张老师带回的书每个人分到两本。	CD

值得一提的是，近几年我们还设计了一系列中文文本空间语义信息理解任务，构建了相应的 SpaCE 评测数据集（詹卫东，等 2024）。其中比较传统的语言学知识导向的任务设计，比如识别文本中空间语义角色（Identifying Spatial Semantic Roles，ISR），对大语言模型反而没有挑战性。SpaCE2024 的参赛队伍中，总成绩前 6 名的系统在 ISR 这项评测中得分均超过 91 分，最高成绩为 94.29 分（Xiao et al. 2024）。这也在一定程度上说明语言学者应探索新的任务形式。比起用人类假设的元语言知识来考大语言模型，更具挑战性的测试方式是让语言学知识退居幕后当导演，指导前台生成自然语言数据形式来进行交互，在模拟人类日常沟通的语境中，考察大语言模型在实际语言应用场景中类似普通人的语言理解能力。

五、结　语

尽管大语言模型不依靠人类语言学知识，实现了多语言、多任务能力，在许多自然语言任务场景中表现出色，但人工神经网络的可解释性差这一问题仍很突出。从这个角度看，人类的语言学研究仍有必要推进。无论是解释人脑的语言能力奥秘，还是解释大语言模型语言能力的内在机制，语言学研究都责无旁贷，应该勇于去迎接挑战。

乔姆斯基提出了语言学要为“语言的无限生成性”提供理论解释的宏大目标。而**科学解释并不是为了解释而解释，其最终归宿必须落实到行为预测**。人工智能时代的语言学，应该在技术飞跃发展的

时势推动下，自觉地迈上一个新的台阶：不能仅仅满足于解释现有的由人类社会自然生成的语料，而是要如乔姆斯基当初所畅想的那样，真正能够从理论模型出发，生成出符合特定目的的语料。这也可以说是基于实证的方法来展示纯理论语言学研究真本事的最有效途径。基于人类语言学通过微观研究工作模式所获得的洞见，再依靠人工智能不断提升的技术能力辅助，我们有可能在观察充分、描写充分、解释充分的基础上，进一步探索“生成充分”的实践路径。如何由理论研究成果驱动，由纯手工到半自动再到全自动地生成能与人工智能直接交互的语言数据，是人工智能时代语言学者要认真思考的问题。人工智能时代的语言学者应该如诺贝尔物理学奖得主理查德·费曼所说的那样，持这样一种信念：**非经亲手造，何敢言真知**。（What I cannot create, I do not understand.）如果一个语言学理论研究成果真的有效的话，就应该能基于该理论，生成（创造）出大量的能满足某种特定意图的自然语言实例（即包括正例和负例的语言数据）。这一研究模式的直接效果是提高了理论可验证性、数据规模可扩充性。在人工智能工程能力的加持下，语言学研究有望从解释到生成，从内涵到外延，从组合到分布，内部语言与外部语言兼修，以理论知识驱动数据生成、以生成效果反馈理论验证，直至觅得形式与意义互证的通用工具，抵达“为内部语言建模”的终极目标。

参考文献

袁毓林　2025《描写还是解释：由 ChatGPT 反思语言学的两种目标》,《语言战略研究》第 1 期。

詹卫东　2021《构式的形式与意义表征——语言数据资源建设视野下的构式研究》,《语言学论丛》第 2 期。

詹卫东，孙春晖，肖力铭　2024《语言学知识驱动的空间语义理解能力评测数据集研究》,《语言战略研究》第 5 期。

ALPAC. 1966. Language and machines: Computers in translation and linguistics. Washington, DC: National Academy of Sciences / National Research Council.

Brown, P. F., J. Cocke, S. A. Della Pietra, et al. 1990. A statistical approach to machine translation. *Computational Linguistics* 16(2), 79–85.

Brown, T. B., B. Mann, N. Ryder, et al. 2020. Language models are few-shot learners. *Advances in Neural Information Processing Systems* 33, 1877–1901.

Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton / De Gruyter Mouton.

Devlin, J., M.-W. Chang, K. Lee, et al. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019*, 4171–4186.

He, K., X. Y. Zhang, S. Q. Ren, et al. 2015. Deep residual learning for image recognition. arXiv:1512.03385.

Hinton, G. E., S. Osindero & Y.-W. Teh. 2006. A fast learning algorithm for deep belief nets. *Neural Computation* 18(7), 1527–1554.

Jelinek, F. 1998. *Statistical Methods for Speech Recognition*. Cambridge, MA: MIT Press.

Krizhevsky, A., I. Sutskever & G. E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In *Proceedings of Advances in Neural Information Processing Systems 25*, 1097–1105.

Levesque, H. J. 2011. The Winograd Schema Challenge. American Association for Artificial Intelligence, AAAI 2011 Spring Symposium: Logical Formalizations of Commonsense Reasoning.

Levesque, H. J., E. Davis & L. Morgenstern. 2012. The Winograd Schema Challenge. In *Proceedings of the Thirteenth International Conference on Principles of Knowledge Representation and Reasoning* (KR 2012), 552–561.

Lighthill, J. 1973. Artificial intelligence: A general survey. In *Artificial Intelligence: A Paper Symposium*, 1–21. Science

Research Council.

McCarthy, J., M. L. Minsky, N. Rochester, et al. 1955/2006. A proposal for The Dartmouth Summer Research Project on artificial intelligence. *AI Magazine* 27(4), 12–14.

Minsky, M. & S. Papert. 1969. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, Mass: MIT Press.

Newell, A. & H. A. Simon. 1976. Computer science as empirical inquiry: Symbols and search. *Communications of the ACM* 19(3), 113–126.

Radford, A., K. Narasimhan, T. Salimans, at al. 2018. Improving language understanding by generative pre-training. OpenAI Technical Report.

Radford, A., J. Wu, R. Child, et al. 2019. Language models are unsupervised multitask learners. OpenAI Technical Report.

Rosenblatt, F. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* 65(6), 386–408.

Rumelhart, D. E., G. E. Hinton & R. J. Williams. 1986. Learning representations by back-propagating errors. *Nature* 323(6088), 533–536.

Rumelhart, D. E., J. L. McClelland & PDP Research Group. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press.

Turing, A. M. 1950. Computing machinery and intelligence. *Mind* 59(236), 433–460.

Vaswani, A., N. Shazeer, N. Parmar, et al. 2017. Attention is all you need. *Advances in Neural Information Processing Systems* 30, 5998–6008.

Winograd, T. 1971. *Procedures as a Representation for Data in a Computer Program for Understanding Natural Language*. Ph.D. dissertation. Massachusetts Institute of Technology.

Winograd, T. 1983. *Language as A Cognitive Process*. *Volume 1. Syntax*. Addison-Wesley Publishing Company.

Xiao, L. M., N. Hu, W. D. Zhan, et al. 2024. Overview of CCL24-Eval task 3: The fourth evaluation on Chinese spatial cognition. In the *Proceedings of the 23rd China National Conference on Computational Linguistics*, 122–134.

Xu, L., H. Hu, X. W. Zhang, et al. 2020. CLUE: A Chinese language understanding evaluation benchmark. COLING 2020. Retrieved from https://github.com/CLUEbenchmark/CLUEWSC2020.

责任编辑：韩　畅