

基于现代符号学对大语言模型的认识论思考

马 援

(山西大学 马克思主义哲学研究所,太原 030006)

摘 要:作为探究人类语言生成机制以及关注人类符号化对象世界的现代符号学,可提供切入对大语言模型的认识视角:从“共时态”与“历时态”对大语言模型处理自然语言的分析;从“语言”与“言语”对大语言模型与人类自然语言的比较;从符号隐喻机制对大语言模型“学习”的认识,拉动大语言模型与人的符号活动的互动关联,为数字化人文的发展提供现代符号学认识论的思考方式。

关键词:现代符号学;大语言模型;自然语言;认识论

中图分类号:N02 **文献标识码:**A **文章编号:**1674-7062(2023)05-0058-06

大语言模型(Large Language Models, LLM)以语言建模的方式,通过人工智能(AI)算法,用于开采人类自然语言理解与生成,实现了在语言推理、文本生成和人机对话的突出表现,标志着计算机对自然语言处理从小模型训练到大模型预训练的巨大跨越。大语言模型是目前自然语言处理的高级化展现,而自然语言处理(Natural Language Processing, NLP)“属于人工智能领域的一个重要甚至核心分支,是计算机科学与语言学的交叉学科,又常被称为计算机语言学(Computational Linguistic, CL)……”同时,自然语言处理研究也为人们更深刻地理解语言的机理和社会的机制提供了一条重要的途径^[1]。一方面,大语言模型是在对人类自然语言充分认识的基础上而形成的对自然语言智能化掌握;另一方面,大语言模型同样是人类高级符号化世界的展现,在人类使用科学符号的语言建模中,可更好地认识人类自身的语言系统和延展人类的符号化活动。

作为探究人类语言生成机制的现代符号学,在分析语言符号内部系统的同时,关注人的现实的符号化活动,探究人类延绵历史中文化化实践的自然语言生成过程,形成从语言符号嵌入于政治、经济、文化和社会的考察,产生了不同面向和层级的符号

学,在最新发展中注入了对新科学技术的思考。正如李幼蒸所言,“AI理论虽然属于控制论和信息论的科学与技术领域,然而由于它直接处理语言记号系统问题,故在某些方面与语言符号学和一般符号学有重要的交接面,是今日一般符号学家不能不关注的新科学”^[2]⁴⁷⁴。现代符号学可与大语言模型建立彼此接连的关系通道。本文依据语言大模型背后的技术原理,尝试从基于人类语言符号的内在系统以及符号化对象世界的现代符号学,对大语言模型做出认识论思考。

一、以“共时态”与“历时态”对大语言模型处理自然语言的分析

大语言模型相比与原先处理自然语言的阶段,实现了从小规模到大规模、从专家知识库到语料库统计再语料库深度学习直至目前预训练模型。在大数据和大模型的加持下,大语言模型从海量语言库中充分无监督学习从而提取自然语言特征。而现代符号学认识论的关键之处,在于索绪尔开拓的“共时语言学”和“历时语言学”,之后叶姆斯列夫的语符学划分了“语言结构”与“语言运用”或者“聚合”与“组合”^[3],以及巴尔特称之为“语言的两根

【收稿日期】 2023-02-17

【基金项目】 教育部青年基金项目“‘结构—文化主义’范式的马克思主义话语隐喻研究”(20YJC720013);国家社科基金重点项目“马克思主义哲学视野中的维特根斯坦研究”(22AZX015)

【作者简介】 马 援(1981-),女,山西长治人,山西大学马克思主义哲学研究所副教授,研究方向为马克思主义符号学。

轴”^[4]⁴²,从而探究符号的意义生成以及符号之间的相互关联。现代符号学以符号折回自身的方式以及符号间的差异性原则,对语言系统做出“共时态”与“历时态”分析,从这一维度可更容易理解大语言模型是如何有效处理自然语言的。具体表现为:

其一,大语言模型是“大数据+大模型”的双相配合,大数据能够保障大语言模型对自然语言完成细颗粒度处理和有效整合自然语言的海量信息,其逻辑关系表现了“共时态”与“历时态”的双向维度。大数据承载了自然语言的巨大存储库,所收集的自然语言数据并非线性时间或历时延绵秩序,多数情况下是时空叠加而生的信息。大语言模型在大数据的支撑下,以“历时态”与“共时态”的互动对庞大信息进行整合,改变传统历时跟踪研究的程式化模式。同时,它以强大的信息处理功能,突破时空叠加造成碎片化和非连贯海量信息的无序和凌乱,并根据实际需求调控时间长度与地域距离,锁定一定时间和一定范围内所需采集数据的有效性,以切换焦点和焦距的方式对宏观测定和微观描述给予准确化的信息。大语言模型按照人类的符号语言处理自然语言,呈现词项间相互勾连的关系图式。

正像《作为经验探索的计算机科学:符号和搜索》一文中,提到“第一个是符号系统观念的形成。第二个是启发式搜索观念的形成。对于理解信息是如何加工的,以及智能是如何获得的,这两个概念有其深刻的意义”^[5]。符号无论对于人类的智性运动还是对于人工智能化来说都是其重要的载体。符号化对于实现人类将机器作为一种代理者而言是其中的关键途径。大语言模型对自然语言的操作勾连起人类符号与机器符号的关系机制。

对大语言模型而言,需要首先获取大规模的文本数据,为预训练提供“保质”+“保量”的支持。“保质”在于保障由高到低语料之间细颗粒化的差异性关联,从而减轻低质量语料的混入,即以“共时态”聚合充分实现对语料间相关度的分析与统计;与此同时,“保量”在于获取丰实和充沛语料信息,即以“历时态”组合建立语料之间上下文的连接。“保质”+“保量”的大语言模型在“共时态”与“历时态”交互下,达到拓扑学意义上语言网状化图景。

例如,在大语言模型中输入“美洲驼”时,可得到高于原先5倍量的相关参数,形成“+聊天数据”“+任务数据”“+合成数据”的对话联动,并在继续预训练、模型继承和数据继承相互作用下进行指令调优,以多通道模型实现对下游任务的良好表现。

以现代符号学来看,词项“美洲驼”,一方面是与“无峰驼”“大羊驼”“驼马”之类词项,处在同一聚合结构的“共时态”关系中;同时另一方面,又与“生长在安第斯山脉高原地区”“是秘鲁国徽图案中的一部分”等,置于不同组合结构的“历时态”关系中,“美洲驼”意义的确定是由这两个部分相互作用而形成的。同样,大语言模型以词项“共时”差异性原则辨别“美洲驼”这一词项,并以词项间“历时”组合原则丰富“美洲驼”的语用层级,在双向关系得到意义的确定。大语言模型有效分析与定位某一语词的意义,就遵循了类似于现代符号学对自然语言展开的双向维度,在共时聚合与历时组合之中接连词项和对话,在多通道模型中对具有多语言提示的多语言任务进行信息微调,以概率统计分析自然语言,完成编码与解码的指令。

现代符号学以两重维度:一个是共断面上的结构秩序,另一个是历史延绵的时间秩序,呈现了人类自然语言的编织系统。《信息检索与语言哲学》一文指出,“语言确实有结构,但它不是传统意义上的语法;而且这种结构是动态的——就像一个正在玩的游戏,而不是一套需要遵循的指令”^[6],说明自然语言中“共时态”结构与“历时态”事实的辩证关系。现代符号学聚焦“能指”与“所指”的结成关系、能指链条的滑动、意指实践活动等符号运动,通过类比、聚类、迁移和组合,将符号的形式、结构、层级与符号的内容、事实、材料互动关联,探寻人类对对象世界认识、理解和加工的符号化过程。同样,大语言模型实现的背后代表了通用人工智能(Artificial General Intelligence, AGI)以“大数据+大模型”的方式,经过对海量数据的梳理、清洗、筛选和计算,实现从庞杂数据到信息再到知识的转化,在搭建像人一样的神经网络,最终使机器具有自主学习能力,从而加剧对自然语言“历时组合”与“共时聚合”,达至目前自然语言处理能力。

其二,在大语言模型中,大模型的建模方式可与现代符号学“共时态”与“历时态”相互关系加以对比理解。大语言模型在Transformer“现代自然语言的主力模型”^[7]的架构体系下,以少量样本实现强大通用性的能力,使得参数量跃升为千亿级,从而有效完成下游任务。在通过对大规模非标注数据和一般通用任务的预训练中,机器以无监督学习的方式,获得泛化性的有效参数,让机器能够充分处理自然语言,从而在面对下游任务时,通过微调参数完成迁移,反馈用户较好体验。大语言模型体现了“预训

练”与“微调”的优化组合,将人类自然语言符号得到了有效归置。大语言模型在基于自然语言经验事实的基础上,对各类相关符号信息从从属地位、中心/边缘、层级分化和嵌入依赖等结构关系,进行对比、配对和聚类等操作,这与现代符号学以“共时态”挖掘语言符号的底层逻辑与结构有相似之处。

在语言“共时态”与“历时态”的两根语言轴上,除了第一根组合延展关系的“历时态”言语链之外,还有一根聚合联想关系的“共时态”语言结构层,也就语言符号的“聚合面”或者“系统面”,它构成了词项间的“记忆库”,具有“潜在的记忆系统”^[4]⁴³。大语言模型的运作非常类似于语言两根轴的运动,庞大参数的神经网络把自然语言进行信息压缩,将所有语料的存储信息加以记忆。也就是,无论是可利用的还是冗余的语料,大语言模型先记忆存储下来,然后,通过自学习的方式,不仅获得语言表层的内容与事实,而且发掘出自然语言深层的结构与规律。大语言模型在大数据与大模型的综合作用下,以处理自然语言“共时聚合”与“历时组合”两根轴的相互关系,实现对自然语言的高阶仿真。

大语言模型在对海量语料的认知加工中,从言语事实获取自然语言理性认知的分析框架,实现“动态”历时层面的“组合”和“邻接”与“静态”共时层面的“聚合”和“聚类”的交互,使得在不同时间状态下,形成对多个层次和层次关系的语言处理。大语言模型在庞大“共时态”与“历时态”人类自然语言的语料库中,将所有与之相关的各种形式的、不同时间阶段的信息加载在一起,不再局限于单线条或者时间秩序的推导方式,而形成更为多层次、立体化的关系序列。因此,大语言模型会在问答、翻译、分类和摘要等不同用户指令中做出了令人满意的答案,很重要的一方面在于,它将对话置于一种动态和相互作用的过程中,以足量的自然语言事实为依据,获取或者输出的语言也更加符合自然语言的方式。

大语言模型将自然语言归属到 Transformer 所架构的模块,用“组合”和“聚合”的形式化方式,对海量语料信息进行分布式与结构层级处理。它包含了对语言模型建模的结构和层级关系,体现了人类对机器建立的形式化系统。同样,现代符号学在历时态的秩序中对意义变化进行共时态结构层的探讨,以形式化语言研究为基础来发掘结构的普遍规律,从而在以人的符号化对象世界的探寻中摸索人的智性运动。符号学以“聚合”“组合”“邻接”和“聚类”等原则对经验事实形式化分析,形成表层现

象与深层结构间交互对话的途径。大语言模型从一个侧面充分展现了人类形式化语言的高级阶段。某种意义上讲,形式化语言或者作为符号语言的数学“被看作是理智世界的真正中心”^[8]。

现代符号学一个重要意义在于在历经“语言转向”之后,从实体论证模式转向语言系统与对象世界“相关关系”的认识论模式,将事物置于整个世界发展的结构中。“我们都生活在一个非线性的世界中,一个充满复杂性的现实的世界与一个同样的语言系统的世界。”^[9]无论是现实的世界还是人类的语词世界都以非线性的方式存在,符号系统中“共时性”与“历时性”所搭建的“相关关系”正展现了人类语言的构造关系,同时,也体现了人类对对象世界的认知方式。大语言模型某种程度上是对人类自然语言的高效集成,可对任一问题做到最庞大大他者的回答,以“共时态”与“历时态”把无数个他者的语言信息编织在一起。

二、从“语言”与“言语”对 大语言模型与人类自然语言的比较

大语言模型试图实现人工智能的自然语言化,那么,大语言模型与人类自然语言生成系统究竟有什么本质区别?可借助现代符号学中“语言”与“言语”的这对概念,对比人类自然语言的符号机制,从而对此问题加以说明。

现代符号学最为重要的一组概念就是对“语言”与“言语”的划分。在这一划分中,语言(lan-gue)被视为社会的、整体的、同质的和具有规则的,而言语(parole)则是私人的、堆积的、异质的和难以把握一般原则的。现代符号学之父索绪尔区分这组概念的目的就在于将语言从含混的语言现象中,获取语言规律的一般法则,对人类语言的结构、系统和规律做出分析。

事实上,大语言模型对自然语言的重新建模,正是在于它能够寻找到人类自然语言的规则和逻辑关系。人类自然语言中存在着一定的形式与结构,大语言模式通过计算机科学实现对以外在于人类自身语言规律的研究与仿效;而现代符号学则通过科学认知符号的内在形式与结构,来建立人类语言的符号系统,两者都力图实现对人类语言符号认识规律的印证。在现代语言,把相对稳定的结构看作是语言的部分,那些不确定和变化的部分放置在言语活动中,通过语言的形式和结构来把握人类语言无尽变化的规则。因此,巴尔特在索绪尔对“语言”与

“言语”的区分上,进一步提出“语言结构既是一种社会性的制度系统,又是一种值项(valeurs)系统”^{[2]4}。值项系统意味着一定的固定性,可被一定结构和数值关系所确定。因此,从“语言”的值项来看,人类语言存在着一定深层结构和内在系统,而这种结构和系统会在人类的言语活动中反复出现。

这就会看到,尽管现在大语言模型被视为巨大黑箱,其突出表现被看成是难以琢磨的现象学层面问题,但可明确的是,在海量的人类自然语言库中,大语言模型通过被大量“投喂”和一定的预训练,确实是能够找到人类语言表达式中的规律。在现代符号学中,“语言”与“言语”的作用关系表现在,“言语是由一些相同记号的反复结合形成的……每一个记号都成为语言结构的一个成分”^{[4]5},也就是在说话者以组合式串联成的言语链条中,存在着聚合式语言结构代码的底层逻辑。大语言模型处理自然语言的方式与现代符号学对自然语言中“语言”与“言语”的分析具有一致性。大语言模型一方面按照自然语言中言语的组合方式,也就是相同记号反复结合的大概率组合;另一方面通过算法和大量参数的预训练,也就是相同记号所构成的语言结构,将言语组合和语言结构一同运作起来而产生如同自然语言的生产。

说到此,会出现另一个问题,就是如果按照现代符号学的认识论中“语言”与“言语”的二元关系,大语言模型似乎也是如此,因为,一方面,大语言模型在大量的人类自然语言中提取出人类语言的逻辑秩序或形式系统;另一方面,在聚集海量人类言语的语料库中,大语言模型通过概率统计的方式,转变成大概率预测下的仿真人类自然语言的输出。

然而,现代符号学在区分“语言”与“言语”概念的同时,更为重要的是,对两者辩证关系的认识。也就是,在索绪尔看来,语言“是言语活动的一个确定部分,……它既是言语机能的社会产物,又是社会集团为了使个人有可能行使这机能所采用的一整套必不可少的规约”^[10]。语言与言语之间是一个辩证运动的过程,按巴尔特所讲,没有言语就没有语言结构,没有语言结构也就没有言语,只有把两者结合起来才有意义。梅洛-庞蒂也强调只有“语言”与“言语”之间的交互关系才能达到真正意义的语言实践。在我们人类的语言中,并不会将哪一部分有意识地标记为语言结构,而另一部分被归为言语活动,而是自然而然的生成过程。语言所承载的形式、结构和规范,是所在文化群体中的每个成员通过言语

实践集约而成的;言语所展开的内容、事实 and 材料,是受到浸润在一定较长时段、一定文化中的语言内在规约性所制约的。

因此,人类语言有自身生成的一整套机制,是索绪尔语言与言语共同作用的结果,一方面,人类拥有特有的语言机能,它是人类源源不断获得言语符号重组的能力;而另一方面,言语行动者不停歇地调动语言机能,产生出丰富变化的言语行为,使内部语言转变成用于表达和交流的外部言语,实现内部语言与外部言语的双向结合,使得人类自然语言始终散发着勃勃生机。叶尔姆斯列夫从图式层、规范层和用法层,对“语言”与“言语”进行了进一步规定,除了图式层即纯形式的语言结构、规范层即质料形式的语言结构之外,还存在着用法层,而这一层是与广袤的人类社会的历史文化和礼仪习惯密切连接的,它代表着人类对对象世界符号化的认识方式和实践方式。人类的自然语言标识人类历史文化演绎中特有的人类文化基因。

就目前而言,虽然大语言模式表现在对人类自然语言的有效处理上,产生了我们过去难以想象的现象学效果,但是,它并没有像人类一样拥有对现实世界的身心体验,不会真正理解人类自然语言所表达的意义,更多像是对庞大人类自然语言语料嚼碎了一样的概率重组。

三、以符号隐喻机制对 大语言模型“学习”的认识

大语言模型实现了突破语言逻辑为基础机器输入知识的阶段,进入了以互联网为路径的机器无监督学习的发展阶段。现代符号学以符号隐喻认知机制,实现了对人类认知活动和符号化活动的跟进。大语言模型的“学习”体现为三个方面:情境学习、思维链和自然指令学习。从现代符号学形成的一系列符号隐喻策略,可为大语言模型智能化的学习特点提供认识论视角。

第一,从符号隐喻的关系机制,看待情境学习的多轮对话形式。情境学习所展开的多轮对话形式,以更新对话状态的方式,将所需问题的所有相关信息全部记录下来,包括历史对话状态、解析模块当前对话状态和背景知识库,作为上文全部提供给大语言模型,从而使大语言模型根据具体提示语句(prompt)可游刃有余完成下游任务。大语言模型营造出的交互体系,不能封闭于单一化的场景,而会形成不同文化、不同历史、不同地理位置的场域的融合

互联,打破原先时空的线性发展模式,形成不同时、空的位移、叠加和重置的效应。而现代符号学的隐喻关系机制,带领我们的思维从习惯性时间秩序和空间秩序的有限屏障,以符号隐喻的关系机制,冲破时间和地理位置的束缚,走进人与人、人与物、物与物、物与非物之间不同节点而组建的物质关系网络中,建立网络式和交互性的思维方式。现代符号学正趋向于更大的隐喻场域中,“一系列概念——配置、代理、组合、网络、链、表演性、混合……——提出了思考一种没有本体论划分的联系方式,其中‘物质’和‘非物质’元素在同一平面上相互作用以形成结构并在这一过程中立即受到这些相互作用的约束”^[11]。现代符号学的隐喻关系机制,激活我们形成人与机器、虚拟与现实之间纷繁勾连的网状图景。

大语言模型正体现了人类符号语言的隐喻机制,以处理和分析人的符号化对象世界的自然语言,从人类自然语言的内容和模式中提取人类符号隐喻机制,即通过对比、配对、聚类、类比、分离、交互和迁移的隐喻策略,形成“先验图式”与“共享意义”,“形式”与“内容”,“能指”与“所指”之间的关系论证,以符号隐喻的各种姿态呈现“主体”与“对象世界”“不同文化群体”“不同对象世界”之间的交互性对话。现代符号学把隐喻关系机制作为理解世界的重要途径,即以符号隐喻的方式,将意义的主导观念引向意义的使用和意义与生活接连,调动日常语言与抽象思维、文化现象与社会深层结构、主体经验与主体嵌入的结构之间的互联关系。实际上,大语言模型情境学习的多轮对话形式,展现了人类通过符号隐喻机制产生与对象世界、对文化他者和对不同观念形成对话的提取方式。

目前人工智能在语言大模型上成效显著,未来还会在多模态的大模型中,推进图像、视频等多模态信息的处理。人工智能不断朝向从技术层面根本改变原先线性化的实体论证模式,产生实体空间与虚拟网络空间的交互关联,形成现实空间的万事万物、虚拟空间所对应现实之物的平面投影、以及未来预想的“立体虚拟”^[12]三种空间的接连方式。这大大延展了人与物的场景关系,形成了网络空间对现实空间场景关系的压缩和排序,产生了人与人、人与物、物与物之间,高效节点而组建的新空间关系,并由新空间关系诱发生成新的结构关系,会形成更加多层次关系的符号隐喻场域。

第二,从符号隐喻的认知机制,理解思维链对复杂问题的分界与降级。思维链是大语言模型面对复

杂问题分解为多步骤简单问题的重要方式。思维链用中间推理步骤增加标准提示,包括算术推理、常识推理和符号推理,从而提升大语言模型在复杂推理上的性能。在接受任务、形成计划、推进计划和最后输出结果的思维链中,看似单一线性的链条结构,实际上,多了一个隐层,也就是计划、行动和反馈的内循环,形成了大语言模型与现实世界交互的一个重要通道,产生由外部环境对任务计划的有效“反馈”,从而在内循环不断反馈中,生成和优化执行结果。这样在外部环境的交互中,大语言模型以“思想树”的形式,从事实知识、上下文和数学逻辑等,将复杂推理转换成多步骤推理,从而实现自我评估并参与决策过程。

这类似于人类思维方式的隐喻认知机制,包括临近原则和差异性原则,通过这些原则构建已知项与未知项的接连,从简入繁、推陈出新从而产生新的认知活动。大语言模型的无监督学习不是被动给机器输入一般的推理程序或者罗列规律,而是依靠机器强大的统计功能,以统计方式捕捉人的符号化活动的规律,从而实现机器的自主学习。大语言模型以思维链的方式,将复杂问题通过海量预训练,转换成一个个被引导出来的推理,在遵循人类思维的隐喻认知机制中,产生不同语词、事物和事件之间的临近原则和组合关系。

第三,从符号隐喻的内生系统类比自然指令学习。自然指令学习目的在于在有限的指令中,通过大模型的微调,泛化到解决更多的任务,实现机器学习的泛化能力。这样一来,自然指令学习摆脱了原先烦琐的、依靠大量标记数据的机器学习的做法,而直接源自对人类自然语言原指令的学习,使得机器更容易接近人们的真实意图,更加符合人类自然语言的表达。自然指令学习直接从指令中发掘自然语言事实背后的结构规律,使得机器能够像人一样做出推理与判断,实现机器的心智运动,达至智能化过程。并且,自然指令学习对大模型进行适度微调,模拟人的思维结构和认知结构,这大大减轻了原先的计算难度和运算量,充分对大数据采集信息有效利用。自然指令学习有效对接人类自然语言,从海量语料中直接洞察和捕捉人的符号化活动的规律,通过人类自然语言中符号隐喻的内生系统,推断人是如何思考、做出分析和采取行动的。目前人工智能研究也正在积极借助儿童认识世界的过程来探索这一问题,从而搭建像人一样的“感觉系统”和“传感系统”。自然指令学习的最佳目标,最终使机器不

但产生类似于人类自然语言的表达,而且趋向人类真实意图和依循人的符号化活动相一致的发展。

数字化时代的开启必然会联动人文符号与机器符号的交互。人的符号化活动在这一时代被烙上了数字化、技术化和虚拟化的烙印,引发了“从人文到(数字化的)后人文”^[13]的到来。从现代符号学,借助历时与共时的交错、横断面上聚类与纵断面上组合、语言与言语的辩证系统以及符号隐喻机制的综合分析,有助于对大语言模型系统化阐释,为数字化时代所引发的思维革命提供认识论意义的思考。

除此之外,现代符号学不仅关注语言符号,同时包括对图像、手势、信号和标志等非语言符号,以及动态视频、时时监控画面等更加复杂、多立体层面符号集合的处理。同样,从人工智能(AI)的未来发展来看,也不止停留在对自然语言处理上,会拓展至不同模态之间相互关联的多模态发展,这更加需要充分建立现代符号学与人工智能的关系通道。在数字化时代发展中,以现代符号学对人类社会分环勾连文化符号图式的分析,推进现代符号学的作用,促进数字技术与人类符号化活动的交互,为人文符号与机器符号的联动思考提供新思考方向。

【参 考 文 献】

- [1] 车万翔,郭江,崔一鸣.自然语言处理[M].北京:中国工信出版集团,2021:2.
- [2] 李幼蒸.理论符号学导论[M].北京:中国社会科学出版社,1993.
- [3] 叶姆斯列夫.叶姆斯列夫语符学文集[M].陈琪龙,译.长沙:湖南教育出版社,2006:40.
- [4] 巴尔特.符号学原理[M].李幼蒸,译.北京:中国人民大学出版社,2008.
- [5] 博登.人工智能哲学[M].刘西瑞,王汉琦,译.上海:世纪出版股份有限公司,2006:115.
- [6] BLAIR D. Information retrieval and the philosophy of language[J]. The computer journal, 1992(3): 205.
- [7] 帕斯特,阿拉萨尼帕莱.自然语言处理实战[M].王书鑫,李锐,丁舶洋,等译.北京:机械工业出版社,2022:146.
- [8] 郭鸿.现代西方符号学纲要[M].上海:复旦大学出版,2008:174.
- [9] 乔瑞金.非线性科学思维的后现代诠释[M].太原:山西科学技术出版社,2015:106.
- [10] 索绪尔.普通语言学教程[M].北京:商务印书馆,2017:16.
- [11] BENNETT T, JOYCE P. Material powers: cultural studies, history and the material turn [M]. London: Routledge, 2010: 35.
- [12] 王骥.新未来简史:区块链、人工智能、大数据陷阱与数字化生活[M].北京:中国工信出版集团,2019:175.
- [13] 李河.从“代理”到“替代”的技术与正在“过时”的人类?[J].中国社会科学,2020(10):133.

Epistemological Thinking on Large Language Models Based on the Modern Semiotics

MA Yuan

(*Institute of Marxist Philosophy, Shanxi University, Taiyuan 030006, China*)

Abstract: Possessing the characteristics of exploring the mechanism of human language generation and focusing on the world of human symbolic objects, the modern semiotics can provide an understanding perspective on Large Language Models with analyses of processing of natural language by Large Language Models from “synchronic” and “diachronic”, a comparison of Large Language Models with human natural language from “langue” and “parole”, and an understanding on the “learning” of Large Language Models from symbolic metaphor mechanisms. In consequence, promoting the interactive association between Large Language Models and human symbolic activities can provide an epistemological thinking mode for the development of digital humanities based on the modern semiotic.

Key words: modern semiotics; Large Language Models; natural language; epistemology

(责任编辑 殷杰)