CALaMo：建构主义语言模型评估

Ludovica PannittoCIMeC 特伦托大学
ludovica.pannitto@unitn.it

Aurélie HerbelotCIMeC/DISI 特伦托大学
aurelie.herbelot@unitn.it

arXiv:2302.03589v1 [cs.CL] 7 Feb 2023

摘要

本文提出了一个新颖的框架，采用建构主义方法评估神经语言模型的语言能力。基于使用的模型不仅符合神经架构背后的随机性哲学，还使语言学家能够在分析中将意义作为决定性因素。我们概述了该框架，并介绍了两种可能的应用场景。

1 引言

多年来，语言学家一直在思考语言是什么，以及如何对其进行最恰当的形式化描述。不同的研究方法有时目标相互矛盾，却由此产生了极其丰富的概念工具，用于描述语言现象。这些工具在解释语言的系统发生、个体发生或历史文化层面时发挥着不同作用，而且彼此之间往往有着密切的交织。在这一研究领域中，计算建模主要用于模拟和研究不同粒度层次上的说话者行为。计算社群中的一个特定领域称为（神经）语言建模，其目标是借助（伪）概率模型再现语言表层结构。由于神经网络架构具有灵活性，因此在这一研究分支中发挥着特殊作用。

在利用计算建模分析语言加工时，语言学中理论工具的极端复杂性会按数量级大幅缩减。例如，当人工神经网络相关讨论中提到语言一词时，它似乎常常被当作语法的单纯同义词：从更广泛的理论视角来看，这两个对象显然并不重合，但许多计算研究模糊了二者的区别。也就是说，那些本应在理论语言学中明确阐述的假设（例如，语法抽象如何融入语言这一概念），在计算研究中却未得到明确讨论；人们往往默认采用一套特定的语言描述方案。目前的大多数研究似乎还隐含着若干关于神经语言模型（以下简称NLMs）应当产生何种语法的假设，而这一潜在选择也常常反映在最常见的评估设置以及从此类实验得出的结论中。这些默认假设大多承袭自乔姆斯基式先天论传统和普遍语法（UG）框架（Chomsky，1986；Smithand Allott，2016），这一传统深刻影响了大量语法计算研究，近期关于神经网络模型的文献也仍受其影响。

颇具讽刺意味的是，主流计算方法中根深蒂固的先天论假设，与该领域所创建模型的本质恰恰相悖。神经网络模型本质上以模式学习为基础，对于要处理的数据的性质完全不设预设。认为语言可以从通用统计机制中抽象出来，这种观点更接近基于使用的观点（下文简称UB）研究路径（Barlow 和 Kemmer，2000年；Goldberg，2003；Tomasello，2003），而神经语言模型则能为这一理论脉络提供更为自然的试验平台。在认知语言学和基于使用的观点中，语言发展过程中对可预测性的利用（这里再次强调，发展涵盖系统发生、个体发生和文化演化这三个层面）是图式化、固化和分布分析等若干基本机制的根源（Lewkowicz 等人，2018）。从这些过程来看，语言被视为一个由构式构成的结构化构式库，并在代际传承中逐步形成（Cornish 等人，2017），也在说话者的一生中不断发展：共享的语言

材料，例如话语中的形态标记，使人们能够识别出特定的模式或构式，并将其视为承载意义的单位（Croft，2001）。

先天论传统与非先天论传统之间在计算建模方面看似存在的差距，可能源于历史因素。乔姆斯基学派及其形式化方法提出了一种语言定义，在过去很容易通过新兴的计算方法加以解释和实现。但这种偏向没有理由继续存在。本文主张采用基于使用的框架，分析人工神经网络中的语言习得。我们首先指出迄今为止影响神经语言模型评估的先天论所涉及的方面（§2）。随后，我们从构式主义视角出发，介绍一个用于定量和定性分析神经语言模型语言能力的框架（§3）。最后，我们展示使用所提出的形式化方法进行的一些初步分析（§4）。

2 语言习得的先天论与非先天论方法

所有关于语言使用与发展的理论都承认，人类语言能力的根基在于处理符号结构的能力。但这些理论对于说话者语言知识的具体内容、此类内容的习得方式，以及这种储存的知识在多大程度上影响语言能产性，意见并不一致（Bannard 等人，2009a）。各理论在三个方面存在分歧：输入、稳定性和系统性。对这三个方面各自采取的视角，会影响我们从神经语言模型对评估任务的响应中得出的结论。下文将依次讨论每个方面，并特别指出，由于未能明确关联研究问题中的实验层面与理论层面，计算模型的评估会因此产生偏差。

输入。先天论框架提出的主要论据之一是刺激贫乏：儿童接触到的输入并不充分，无法解释学习者身上观察到的习得概括（Crain 和 Pietroski，2001年）。这类理论假定，儿童在由先天约束界定的假设空间中进行探索（Eisenbeiß，2009）。相反，构式主义路径认为，语言是通过领域一般机制从输入中涌现出来的：这意味着输入经过特定塑造和偏置，以增强其可习得性（Boyd 和 Goldberg，2009）。一条已获充分确立的研究脉络表明，儿童是高效的统计学习者（Gómez 和 Gerken，2000年；Romberg 和 Saffran，2010；Christiansen，2019）。近期实验也显示，类语言结构可以从纯粹的线性信号中涌现，例如（Cornish 等人，2017）的研究证明，语言序列结构的重要方面可能源于对人类学习者和使用者认知局限的适应（Christiansen 和 Chater，2016b）。此处先天论与非先天论路径的关键区别，在于接收到的输入与习得的语言结构之间的关系有多紧密：如果我们认定输入仅仅是几乎预先决定的认知结构的触发因素，自然就会将注意力从输入的特征上移开，主要转向结构的特征。另一方面，要从输入结构本身推导出语言结构，就必须同时考察这两方面。迄今为止，大多数关于神经语言模型的研究都忽略了输入对实验结果的影响（Pannitto 和 Herbelot，2022）。

稳定性。 这种 连续性假设 最初由 Pinker(1984) 提出，旨在调和发展语言与生成框架的某些方面。该假设认为，成人与儿童语言结构之间的差异微不足道，仅仅是表现因素所致。与之相反，我们可以将 发展假说 概括为：习得的底层机制在一生的习得过程中始终不变，但这些机制所生成的结构和抽象会随时间演变。基于使用的模型也强调语言信号具有线性且依赖时间的特性（Christiansen 和 Chater，2016b；Cornish 等人，2017）。根据基于使用的观点，概括是逐渐形成的，能产性则由项目特定知识发展而来（Bannard 等人，2009b）。

稳定性的另一个方面是说话者之间的差异。普遍语法认为，所有说话者最终都会趋同于同一种语法（Lidz and Williams，2009；Crain 等人，2009）。然而，几乎在语法的各个领域都发现了个体差异，其成因包括环境因素在内的多种因素（Street 和 D.abrowska，2010）。这种“相同性”假设贯穿于

计算语言学文献中，在该领域，评估依据每项任务的单一“黄金标准”进行。对于情感分析或词语相似度评分等传统任务，人类受试者的标注结果会取平均值，系统则根据该平均值进行评估。对于语言建模，模型困惑度是根据大型语料库的统计特征计算得出的；该语料库汇集了数千名说话者的写作风格和语言习惯。尽管各种研究人员已开始批评这种现状，但它目前仍是既定做法。若将语言发展视为一个依赖说话者、且深受输入性质影响的过程，那么基于“平均说话者”的评估就实在难以令人满意。我们不能假定真实值存在，而必须依赖较为灵活的评估指标：显然，不同说话者的语言行为必须有足够的重叠，才能实现交流；但我们也希望在网络的输出中观察到人类所表现出的那种变异性。

系统性。运用有限手段理解和生成无限数量的新句子，被视为语言能力的标志之一。不过，这种系统性的边界在很大程度上仍不清楚：即使我们对可用的有限手段达成共识，说话者实际实现的可能性并非全部，且已实现的可能性并不都具有相同的认知或语言学地位。

理解系统性的一种方式是将其视为组合性，而最广为人知的版本可能要归功于Katz and Fodor (1963年)，他们将 Chomsky 的先天论引入语义学：要通过整合词语的意义来系统地构建句子的意义，就需要一套规则或约束。即便是 Montague 的组合性形式化方法（Montague，1970年）也依赖于乔姆斯基派思想，假定存在一个存储意义的稳定词库，以及一套精确的解释规则，使这些意义能够通过句法这一滤镜进行组合和调节。这两种观点的核心仍然非常以句法为中心（语义学必须与句法同构），几乎没有为不确定性、说话者之间的协商，以及与语言的互动和交际本质相关的其他方面留出空间（事实上，不同个体可以保留与同一词汇标签相关联的截然不同的概念——例如，拉博夫，1973年）。简而言之，如果我们从组合性的角度看待系统性，那么语言结构中的准规律性便是一个亟待克服的主要障碍。

相较之下，准规律性是能产性的引擎，体现为说话者能够运用所有可用的语言手段来传达预期含义。与组合性一样，能产性探讨的是某种语法模式在语言语境中可以如何运用而不失可解释性，也探讨语言中实际上可能出现什么，以及可接受性的边界应划在哪里。这种移位并不只是句法层面的：例如，在语义学中对系统性这两个方面进行形式化表征时，以组合为导向（Katzand Fodor，1963年）或以能产性为导向（Fillmore，1976年）的理论，对选择限制这一概念的构想各不相同。

在这两种情况下，系统性知识都被视为说话者对其语言所拥有的隐性知识。然而，先天论方法主要关注组合性，因此神经语言模型也常以此为评估标准：给定语法规则和词库，哪些计算机制能够使它们组合起来？另一方面，基于使用的理论主要关注能产性：意义边界能够被拓展到何种程度？哪些机制能够促成语言创造性？当然，这也意味着，在基于使用的理论社群中，同样需要关注输入的表层属性：Croft 和 Alan Cruse (2004) 例如指出，最具图式化特征的构式，如 主语-动词-宾语，也是能产性最高的构式；而这也与它们作为一种类型的频率以及各自实例化形式的频率有关。

3 CALaMo

在我们提出的方法 CALaMo（语言模型的构式主义评估）中，我们将基于使用的视角贯穿三个方面：输入、稳定性和系统性。

就输入而言，CALaMo 与标准方法不同，它认为输入数据是决定学习者语法知识形态的重要因素。在传统情境中，输入仅起到触发因素的作用，其特征在分析中几乎不起作用。而从 UB 的角度来看，这种关系

——即输入的抽象语法结构与习得语法之间的关系，而习得语法又会约束学习者的产出——是严格的。

关于稳定性，根据对连续性假说所持的观点不同，我们可以将神经语言模型的语法能力视为二元属性或渐变属性。前一种情况下，我们测试网络是否能够处理某种语言现象；而在后一种情况下，正如 CALaMo 所主张的，我们关注的是在训练过程中，某种语言方面如何以及为何会在网络中变得越来越显著。

组合性与能产性的视角最终意味着语言知识具有不同的组织方式：主流的组合性视角倾向于将意义搁置一旁，并把词库视为一个有组织的意义库（因此，测试神经语言模型处理语义不通的句子的能力，或将已知规则扩展到完全未知的词汇项目，都是合理的）。相反，在能产性视角中，意义本就是这一过程不可或缺的一部分，也被视为语法的一个系统性方面。

3.1 语言习得

谈及神经语言模型的语言能力时，语言习得（A）的问题通常被形式化为：在某种计算复杂度 C 下，人工说话者接触某种类型的数据 I 后，能够习得多少语言 Λ：

$$A : C \times I \mapsto \Lambda \tag{1}$$

方程中的所有组成部分都是语言学争论的核心。然而，从这一基本形式化出发，我们确定了两个重点，并将在框架中专门探讨。首先，上述公式将习得描述为瞬时过程，但实际上更适合将其描述为一个过程 $A = (a_0, a_1, \cdots a_N)$（§3.2）。从认知视角来看，这一过程是完全连续的；而在人工情境中，输入数据通常是以“批次”的形式馈入的。不过我们可以设想，如果能够随意增加步骤的数量（即在保持数据量不变的同时，使 N 变大），便可以将步骤形式化得足够细致，从而使这两个过程具有可比性。

其次，语言通常被视为学习者已经习得并掌握了相关知识的某种东西。我们希望将语言学家观察者的作用重新纳入框架中，即语言学家观察者对说话者的语言行为进行抽象（§3.3）。由于说话者实际习得的知识无法被直接检测，只能通过元语言方式加以解释，而这种方式并不适用于神经网络（也就是说，我们无法询问神经语言模型它们知道哪些语言规律），因此必须考虑到这样一个事实：我们始终是在分析说话者接收到的语言输入，以及习得过程所产生的输出，而这两者都要通过语言学家观察者创造并使用的分析类别来进行。换言之，Λ 它并不是说话者的一种属性，而是语言学家观察者所执行的一项功能。它不会在习得过程中自行进化，而是帮助我们检测并刻画说话者能力的进化。

3.2 习得过程

方程1 中的所有元素都应随着习得过程的展开而随时间变化。

在现实情境中，学习者接触到的输入 I 会持续变化。因此，我们可以定义 $I = (\iota_0, \iota_1, \cdots, \iota_N)$，其中 ι_i 是学习者在 a_i 与 a_{i+1}之间接触到的输入数据集合。同样，理想情况下，只要 N 足够大，每个 ι_i 甚至都可以对应一个句子。随着语言知识被纳入习得函数，计算复杂度也会随之共同演变。在人类的情形中，初始状态不可观测；而在人工情境中，由于神经网络模型的初始化是随机的，初始状态通常并不重要。相反，在步骤 i，已将截至步骤 $i-1$ 的知识纳入其中的计算机制会接触到 ι_i。基于这些原因，我们定义 $C = (c_\emptyset, c_0, \cdots, c_{N-1})$。因此，$\Lambda$ 在整个习得过程中标识出不同的子集 $\lambda_0, \lambda_1, \cdots, \lambda_N$，即 $_N\Lambda = \bigcup \lambda_i i{=}0$

因此，更广泛流程的每个步骤 A 都可以定义为：

$$\begin{cases} a_0 : \iota_0 \times c_\emptyset \mapsto \lambda_0 \\ a_i : \iota_i \times c_{i-1} \mapsto \lambda_i \end{cases} \tag{2}$$

3.3 我们如何观察习得的语言?

我们所引入的语言概念包含语法，即我们叠加在语言流之上的分析类别，以便分析语言流及其随时间的展开。我们测试的并非说

话者的认知状态：我们将语言视为一组类别，观察者（即语言学家）认为这些类别与描述（人工）说话者产生的语言流相关。因此，语言流（无论是说话者感知到的输入，还是其产生的输出）与通过语言这一视角对其进行的表征之间，存在显著差异。

如果要使符号表示更加精确，我们应当承认这样一个事实：语言，即我们所指的 Λ，本身实际上是一个函数，它以某种语言流（某些可观察数据）为输入，并返回其表征。因此，我们可以将 a_i 的定义改写为a_i：$\iota_i \times c_{i-1} \mapsto \Lambda(o_i)$，其中 o_i 是说话者在习得步骤 a_i的作用下产生的语言流。

由于我们关注的是说话者习得并在语言理解和生成过程中运用的类别，定义$\lambda_{o_i} = \Lambda(o_i)$使我们能够对说话者所接触的输入 ι_i 应用相同的变换，从而获得 λ_{ι_i}，它在语言类别方面与λ_{o_i} 具有可比性。

坚持构式主义视角，同时尽可能少地假设语言知识的实际内容，我们假设语言由一个结构网络构成，该网络应当近似于构式。由于构式是形式-意义对，语法这一概念也包含了超越词汇层面的意义空间。这可以通过扩展向量空间模型这一概念来轻松实现；向量空间模型已在分布式语义学中得到广泛探索和应用（Lenci，2008；Erk，2012；Lenci，2018）。这与先天论方法和标准评估框架存在重大差异：意义不能从语法效应中剥离，习得框架必须说明意义在这一过程中的作用。如果我们要形式化任何 λ_i的内容，那么我们可以将其扩展为 $\lambda_i = \{(\kappa, \vec{\kappa}) \mid \kappa$是相对于某种语言流而言的构式$\}$

细说一下，我们的意思是，每个获得的构式库 λ_i 都是一个结构网络。这些结构的词汇化程度各不相同，其抽象程度可以作为衡量它们在网络中相互关联程度的指标，具体将在下一段中解释。每个构式都与一个分布向量相关联（图1），该向量表示其

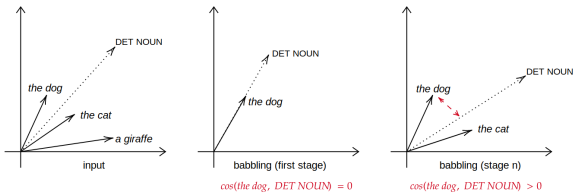


图1：假设 Λ 包含构式 限定词 名词 和这只狗，后者是前者的一个词汇化实例。在习得的不同阶段，这两个构式可能具有不同的意义，因此会关联到不同的分布向量。事实上，分布向量浓缩了有关特定文本片段中语言项目之间共现情况的信息。图中可见，在根据输入数据构建的构式库中，限定词 名词 周围聚集着一簇向量（最左侧面板）。这意味着该构式存在多种词汇化实例。在学习过程中，根据生成的输出构建的构式库中，该构式限定词 名词呈现出不同的分布。在中间面板中，限定词 名词 与这只狗 之间的余弦距离为 0，这意味着它们的分布语境（即共现情况）完全重合。而在最右侧面板中，随着另一个词汇化实例（即这只猫）的出现，两个向量之间的距离增大。在这种情况下，限定词 名词 出现的语境与这只狗出现的语境并未完全重合。

意义。

3.4 期望条件：由 Λ诱导的结构

我们将 Λ 定义为一个函数，它以语言流 τ 作为输入，并返回一个构式库 λ_τ：一个结构化的形式-意义对库。为了定义并探究构式库的内部结构，我们引入几个辅助函数和定义：

(i)：将意义定义为分布空间后，便可进行距离计算 $d(\kappa_i, \kappa_j)$，其中 $d: \Lambda \times \Lambda \mapsto [0,1]$。$d(\cdot, \cdot)$是一个度量函数，用于计算两个意义向量之间的距离。通常，$d(\kappa_i, \kappa_j) = 1 - cos(\vec{\kappa_i}, \vec{\kappa_j})$，其中 $cos(\vec{\kappa_i}, \vec{\kappa_j})$是与 κ_i 和κ_j相关联的两个向量之间的余弦相似度；

（ii）：网络中不同抽象层级的构式彼此相连。为了遍历该网络，我们引入函数$c(\kappa_i, \kappa_j)$，其中 $c: \Lambda \times \Lambda \mapsto \{0,1\}$ 是一个布尔函数，用于判断两个构式是否构成一条抽象链。例如，κ_i = nsubj、GIVE、iobj、dobj 和 κ_j =

nsubj、root、iobj、dobj 构成一条链，其中 κ_i 是κ_j的部分词汇化（因此抽象程度较低）实例。

3.5 使用场景

3.5.1 个体习得的时间进程

该框架可用于观察习得过程如何随时间展开。事实上，我们可以设定若干 n 步骤Λ，o_1并观察：(i) 比较各个 $\sim \Lambda_{(}o_2$步骤 $\sim \cdots \sim \Lambda_{(}o_n$，考察学习过程中语法形态如何变化，如下所示：（$o_1$） $\sim \Lambda_{(}o_2$（ $\sim \cdots \sim \Lambda_{(}o_n$）；(ii) 考察输入中的语法如何与说话者习得的语法进行比较，如下所示： Λ（I） $\sim$ Λ（o_n）。给定一 $K \subseteq$ Λ组I（[1]）[1] 有趣的构式，我们可以观察它们在学习过程中的表现。

例如，一种广为流传的构式主义假说（Goldberg，2006）指出，构式的意义（例如双及物模式*Subj VObjObj2*）及其能产性，源于它与学习者所接收输入中特定词汇项的关联（例如双及物模式中的*give*）：部分词汇意义仍附着于句法模式的意义上，因此其分布属性也随之保留。假设说话者已经习得了某个构式 κ（例如双及物构式）。当他们能够以能产且灵活的方式使用该构式时（即在比输入中该构式与*give* 紧密关联的语境更多样的语境中使用），我们便可以运用所提出的框架，检验两个构式的分布意义 κ_i， $\kappa_j \in \Lambda$（I）与 c（κ_i， κ_j） $= 1$（即 κ_i 是κ_j的较低抽象程度实例）κ_j 作为独立构式的可习得性是否会受到影响。

前文引入的抽象链 这一概念有助于我们检验这一假设，因为我们可以在每个时间步检查该链（κ_i， κ_j）的行为。我们可以将构式记作 $\kappa_i^{\lambda_k}$ $\kappa_i \in \lambda_k$，并类似地将构式记作 $\kappa_j^{\lambda_k}$ $\kappa_j \in \lambda_k$。通过分布分析，我们可以捕捉 κ_i 和κ_j 所处语境的变化，以及这种变异是否与语法概括相关。事实上，我们预计，在习得过程中， d（κ_i， κ_j）会增加：

$$d(\kappa_i^{\lambda_a}, \kappa_j^{\lambda_a}) \leq d(\kappa_i^{\lambda_b}, \kappa_j^{\lambda_b}) \ \forall\, a, b \mid a \leq b \quad (3)$$

如果κ_j 是在并非完全重合的语境中产出的与产出 κ_i 的语境相比，这表明说话者已掌握结构κ_j的产出性用法，并将其识别为独立于 κ_i 的结构。反之，如果在习得过程中二者的距离缩小，我们或许可以推断，说话者已认识到κ_j 没有必要，因为其适用范围被限制在 κ_i的适用范围内。

3.5.2 语言作为说话者群体的表达

我们通常希望根据说话者之间可视为共享的语言知识来界定语法。事实上，构式语法的一个核心方面，是将语言主要视为一种社会性、外在的现象，这与关注语言内在本质的先天论相对。借助这一框架，我们可以将语法分析为对一个群体的语言产出的抽象： P说话者Π $=$ （σ_1， σ_2， $\cdots$， σ_P）。我们可以将社群所采用的语法惯例定义为Π $\Lambda_\Pi =$（λ_{σ_1}， λ_{σ_2}， $\cdots$， λ_{σ_P}）。这使我们能够对不同说话者的习得过程中的变异进行建模。说话者 σ_i 接触到的输入材料序列可能独一无二$\iota_0^{\sigma_i}$， $\cdots$， $\iota_N^{\sigma_i}$，未必与说话者σ_j接触到的相同。

在这种设定下，我们可以探究哪些内容是无论输入是什么都能学到的，哪些内容则是每位说话者所特有或独有的。我们可以定义：

$$G_{\geq p} = \left\{ \kappa \mid \sum_{i=0}^{P} X(\kappa, \sigma_i) \geq p \right\} \quad (4)$$

为我们能够在 p 位或更多说话者的语言产出中观察到的构式集合。其中：

$$X(\kappa_i, \sigma_j) = \begin{cases} 1 & \text{if } \kappa \in \Lambda^{\sigma_j} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

这是一个辅助函数：如果构式 κ 出现在说话者 σ_j 的产出中，则其值为 1，否则为 0（这只是帮助我们统计有多少说话者能有效地使用构式 κ ）。G_P 例如，可以是群体中所有说话者共有的构式集合，因此可将其确定为 Λ^{Π}中的核心构式集合。相反，当 $p \ll P$时，我们观察到的是群体中并非大量说话者共有的构式，其

[1]实际上，我们必须确保 $K \subseteq \Lambda$（I） $\cap\, \lambda_0 \cap \lambda_1 \cap \cdots \cap \lambda_n$

因此，其使用可能取决于特定的输入实例，或说话者亚群的倾向。遵循同样的逻辑，我们当然也可以将 $G_{(\sigma_i,\sigma_j)}$ 定义为两位说话者 σ_i 和 σ_j 共有的构式。借助 $\widetilde{G}$，我们可以将 Λ_G 定义为函数 Λ的近似形式，它只使用在 G中保留的类别 $\widetilde{\ }$。 $\Lambda_{G_{\geq P}}$ 例如，可以是一个只考虑整个群体 Π共享的语言知识 $\widetilde{\ }$的函数，而 Λ_{σ_i} 则仅限于构式库 λ_{σ_i}。对于说话者 σ_i 和σ_j及其各自产生的语言输出 O_{σ_i} 和 O_{σ_j}，我们可以生成 $\widetilde{\ }$ $\widetilde{\ }$并比较 $\Lambda_{G_{\sigma_i}}(O_{\sigma_j})$ 和 $\Lambda_{G_{\sigma_j}}(O_{\sigma_i})$：分别表示说话者 σ_i能够从 O_{σ_j}中提取的内容，以及说话者σ_j 能够从 O_{σ_i}中提取的内容。

说话者使用相同构式κ 来构建其语言产出，当然并不能确保相应的意义 κ彼此一致。[2] 不同说话者会因接触到的输入不同，以及计算机制存在一定程度的随机性，而将不同的意义空间与同一构式关联起来。给定两位说话者 σ_i 和 σ_j，以及一个句子 s，因此我们可以比较 λ_{σ_i} 和 λ_{σ_j} 中被激活、用于对句子 s进行语言学（解）构的意义空间部分。

4 探索性实验

为了探索第 §3节所述框架的潜在应用，我们使用 CHILDES语料库（MacWhinney，2000年）作为输入数据，构建了一个简单实例 I，并采用标准的基于字符的LSTM（Hochreiter and Schmidhuber，1997年）作为计算机制 C。在这一简单设置下，我们探究了两个方面：（i）测试 λ_I 中的分布相似性是否会影响构式库的习得 λ_1，$\cdots$，λ_n；（ii）尝试描述语法如何从说话者群体中涌现。我们通过 链(Osborne et al.，2012)来近似表示构式：它们是从依存句法分析的句法表示中提取的子树（见图 2）。

4.1 抽象语法过度训练

我们首先复现Pan- nitto和 Herbelot(2020)中介绍的一项分析。作者在 CHILDES语料库上训练了一个基于字符的LSTM，并在其习得过程中固定了7个步骤

[2]这确保了 G_{σ_i} 不会与 λ_{σ_i}重合

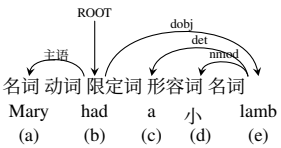


图2：句子 *Maryhad a littlelamb* 的依存表示，其中标注了形态句法和句法信息。在这一结构中，我们可以识别出以下链：a、b、c、d、e、ab、abce、abde、abcde、abe、bce、bde、be、ce、de、cde。其他可能的选择包括字符串 （例如，a、ab、abc、.....b、bc、.....e）或成分 （即，a、abcde、c、d、cde）。

κ_i	移位	余弦	κ_j	移位	余弦
@主语 @根节点 所以	0.18	0.43	更多 @根节点	0.2	0.21
@主语 only @根节点	0.18	0.41	_AUX 知道 @obj	0.19	0.66
什么 @根节点 @obj	0.18	0.39	@advmod 告诉	0.17	0.64
什么 @advmod _动词	0.16	0.19	@aux 知道 @obj	0.16	0.71
仅 @root	0.16	0.38	@advmod can _动词	0.15	0.76
更多 @root	0.16	0.23	知道 @obj	0.15	0.62
@root 它 @xcomp	0.15	0.61	一个 _名词	0.13	0.52
@det 分钟	0.15	0.25	可能 @root	0.13	0.70
_代词 only @根节点	0.15	0.53	_代词 @根节点 不	0.12	0.53
_动词 _DET 分钟	0.15	0.33	@根节点 那 _动词	0.12	0.65
_代词 @根节点 如此	0.14	0.54	_动词 'll @ccomp	0.12	0.71
_限定词 minute	0.134	0.33	_动词 me @obl	0.12	0.76

表1：平均偏移最高的构式 .

，每经过5个训练轮次执行一步。在我们的形式化表示中，这相当于7个构式库 λ_1 到λ_7。通过统计同一句子中构式之间的共现，得到每个 λ_i 的分布空间。然后，我们可以考察 I 中的抽象链（κ_i、κ_j）（即在 Λ（CHILDES）中），并为每条抽象链计算 d（$\kappa_i^{\lambda_7}$、$\kappa_j^{\lambda_7}$）$-d$（$\kappa_i^{\lambda_1}$、$\kappa_j^{\lambda_1}$），也就是第7步与第1步之间的余弦相似度差。按 κ_i 和κ_j对所有抽象链分组后，即可计算平均分布偏移，如表1所示（即，对于每个 κ_i，考察其更抽象的实例；对于每个κ_j，考察其更具体的实例）。

根据平均分布移位，划分为三组：我们的假设是，训练期间移位幅度最大的构式，在输入分布空间中表现出的相似度处于中等水平。事实上，输入相似度极高的链不太可能表现出抽象性：按照构式主义直觉，它们的分布相似性意味着，构式库 中的构式是图式化程度最低的构式，因此没有必要创建更具图式化特征（因而抽象）的类别。另一方面，相似度较低的配对可能只包含彼此无关的构式。

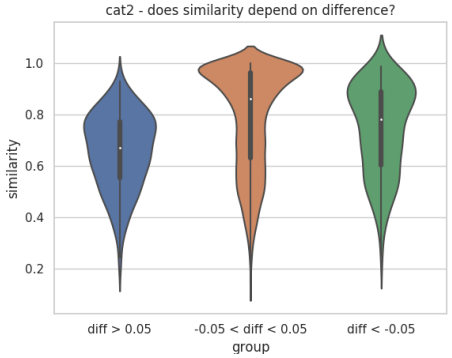


图3： $kappa_j$三组的平均余弦相似度分布，分别显示低、中、高平均偏移。

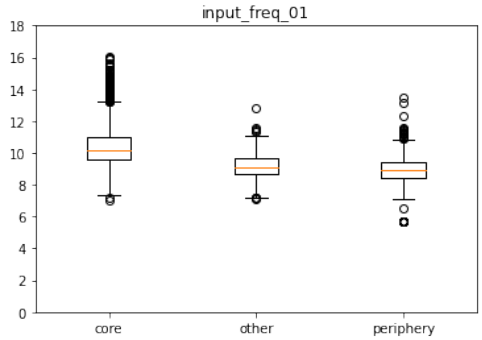


图 4：三组构式的输入频率差异：核心指所有说话者共有的构式，边缘指半数或更少说话者共有的构式，其他指其余构式。

这三个组呈现出不同的分布[3] 如图3所示。

4.2 人工说话者群体

在 Pannitto 和 Herbelot（2020）的实验基础上，我们考察了一个由10名说话者组成的群体，这些说话者分别由10个标准的基于字符的LSTM建模，并使用从 C HILDES 语料库中随机抽取的样本进行训练（每个样本包含100万词）。

在这一设定下，我们试图确定不同说话者之间变异的所在，并假设某些“核心”构式必须为所有个体共享，而其他构式对成功交流则不那么重要。

我们将分析范围限定在所有10名说话者都通过输入接触过的构式（11051个构式），并将 G_{10}设为核心构式的集合，将 $G_{\leq 5}$ 设为边缘构式的集合，也就是由半数或更少说话者共享的构式。由于使用从同一分布中随机抽取的样本进行训练，说话者共享了大多数构式（11051个中有9086个）。不过，我们预计，当输入语言沿着更精细的社会语言学维度发生变化时，这些数字也会显著改变。

我们还针对所有说话者检验了，核心组的构式与边缘组的构式在提供给各说话者的输入中，频率是否存在显著差异。如图4所示，三组之间的差异具有显著性，尽管其表现并没有预期的那么明显。

最后，我们通过语料库探究了输入 $\widetilde{\Lambda_{G_{10}}}$ 和 $\widetilde{\Lambda_{G_{\leq 5}}}$，如表2所示：两种表征（通过核心构式和通过边缘构式获得的表征）都凸显了句子中的有意义模式，但只有前一种可以被视为群体共享的语法表征。

语料库	Core	边缘
does	AUX	-
n’t	n’t	-
that	that	PRON
seem	@根节点	VERB
kind	-	ADV
of	-	-
愚蠢	ADV	ADV

表2：如果我们将范围限制为仅包含核心 （中间列）或边缘（右侧列）构式，那么句子将如左侧列所示。

5 结语

语言发展的语言学理论中，语言表征的性质是一个核心问题。我们认为，神经语言模型文献忽视了这一方面，因此提出一种将理论洞见重新纳入考量的方法。在此，我们采用基于用法的构式框架，但并非将其视为人类语言习得的理想模型，而是将其作为一套足以解释神经语言模型所生成语言的工具和范畴。

既然学习一门语言在很大程度上与学习加工输入相重叠，那么针对某些类型构式的加工偏向必定与这些构式在语言输入中的分布有关（Christiansen 和 Chater，2016a）。由于经验奠定了语言知识的基础，分布属性便成为决定语言表征内容的关键因素。在这一框架中，语言不被视为一个自主的认知系统。相反，语法习得被视为与其他概念化过程相同的过程，而语言知识

[3]我们进行了 Kruskal-Wallis单因素方差分析，结果显示数值具有显著性。

则是在使用中形成的。

总之，采用构式主义方法来观察神经语言模型的语言能力将大有裨益。评估可以在多个层面进行，既包括语言信号所描述情境的属性，也包括语言信号本身的属性。事实上，UB框架或许能为分析人工语言学习者的统计属性提供有用的类别；更重要的是，它使我们能够并行考察语义层和句法层，既考察学习者接收到的输入，也考察其生成的随机输出。

参考文献

Colin Bannard、Elena Lieven 和 Michael Tomasello。2009a。儿童早期语法知识建模。《美国国家科学院院刊》，106(41)：17284–17289。

ColinBannard、Elena Lieven 和 Michael Tomasello。2009b。儿童早期语法知识建模。《美利坚合众国国家科学院院刊》，106(41)：17284–17289。

Michael Barlow 和 Suzanne Kemmer。2000年。基于使用的模型。《语言》。斯坦福。

Jeremy K Boyd 和 Adele E Goldberg。2009年。构式主义框架中的输入效应。《现代语言学期刊》，93(3):418–429。

诺姆·乔姆斯基。*1986*年。语言知识：其本质、起源与运用。格林伍德出版集团。

Morten H Christiansen。2019年。隐性统计学习：两种文献的故事。《认知科学专题》，11(3):468–481。

Morten H Christiansen 和 *Nick Chater*。*2016a*年。创造语言：整合进化、习得与加工。麻省理工学院出版社。

Morten H Christiansen 和 Nick Chater。2016b年。“非此即彼”瓶颈：语言的一项根本制约。《行为与脑科学》，39:e62。

Hannah Cornish、Rick Dale、Simon Kirby 和 Morten H Christiansen。2017年。序列记忆约束通过迭代学习产生类语言结构。《公共科学图书馆·综合》，12(1)。

Stephen Crain 和 Paul Pietroski。2001年。天性、教养与普遍语法。《语言学与哲学》，24(2)：139–186。

Stephen Crain、Rosalind Thornton 和 Keiko Mura-sugi。2009年。捕捉难以捉摸的被动句。语言习得，16(2)：123–133。

William Croft。2001年。激进构式语法：类型学视角下的句法理论。牛津大学出版社。William Croft 和 D Alan Cruse。2004年。认知语言学。剑桥大学出版社。Sonja Eisenbeiß。2009年。语言学习的生成式方法。47(2)：273–310。Katrin Erk。2012年。词义与短语意义的向量空间模型：一项综述。语言与语言学指南，6(10)：635–653。Fillmore。1976年。框架语义学与语言的本质。纽约科学院年刊。Adele E Goldberg。2003。构式：一种新的语言理论方法。认知科学趋势，7(5)：219–224。

Adele E Goldberg。2006年。《构式的运作：语言概括的本质》。牛津大学出版社按需出版。

Rebecca L Gómez 和 LouAnn Gerken。2000年。婴儿的人工语言学习与语言习得。《认知科学趋势》，4(5)：178–186。

Sepp Hochreiter 和 Jürgen Schmidhuber。1997年。长短期记忆。《神经计算》，9(8)：1735–1780。

Jerrold J Katz 和 Jerry A Fodor。1963年。语义理论的结构。《语言》，39(2)：170–210。

William Labov。1973年。词语边界及其意义。分析英语变异的新方法。

Alessandro Lenci。2008年。语言学与认知研究中的分布语义学。意大利语言学杂志，20(1):1–31。

Alessandro Lenci。2018年。词义的分布模型。语言学年度评论，4:151–171。

David JLewkowicz、MarkA Schmuckler 和 Diane M J Mangalindan。2018年。层级序列模式的习得在婴儿期开始显现。发展心理生物学，60(3):243–255。

Jeffrey Lidz 和 Alexander Williams。2009年。假日上的构式。20(1):177–189。

Brian MacWhinney。2000年。*CHILDES*项目：谈话分析工具。第三版。Lawrence Erlbaum Associates。

Montague。1970年。英语作为一种形式语言。《社会与技术中的语言》。

Timothy Osborne、Michael Putnam 和 Thomas Groß。2012 年。链：介绍一种新颖的句法分析单位。*Syntax*，15(4)：354–396。

Ludovica Pannitto 和 Aurélie Herbelot。2020 年。递归咿呀学语：评估如何从有限的输入数据中习得语法。载于第 *24* 届计算自然语言学习会议论文集，第 165–176 页。

Ludovica Pannitto 和 Aurelie Herbelot。2022 年。循环神经网络能否验证基于使用的语法习得理论？《心理学前沿》，13。

Steven Pinker。1984 年。语言可习得性与语言发展。哈佛大学出版社，剑桥，马萨诸塞州。

Alexa R Romberg 和 Jenny R Saffran。2010 年。统计学习与语言习得。《*Wiley* 跨学科评论：认知科学》，1(6)：906–914。

Neil Smith 和 Nicholas Allott。2016 年。《乔姆斯基：思想与理想》。剑桥大学出版社。

James A Street 和 Ewa D.abrowska。2010 年。语言掌握中的更多个体差异：成年母语者对被动句和量词的了解程度如何？《*Lingua*》。国际普通语言学评论。《国际普通语言学评论》，120(8)：2080–2094。

Michael Tomasello。2003。《建构语言》：基于使用的语言习得理论。哈佛大学出版社。