L语言层级化为人类工作记忆提供了最优解决方案

局限

Luyao Chen[1,2*], Weibo Gao[3], Junjie Wu[4,5,6,7], Jinshan Wu[8*], Angela D. Friederici[2*#]香港理工大学人文学院语言科学与技术系，香港，中国[2]马克斯·普朗克人类认知与脑科学研究所神经心理学系，德国莱比锡[3]中国科学技术大学认知智能国家重点实验室，合肥

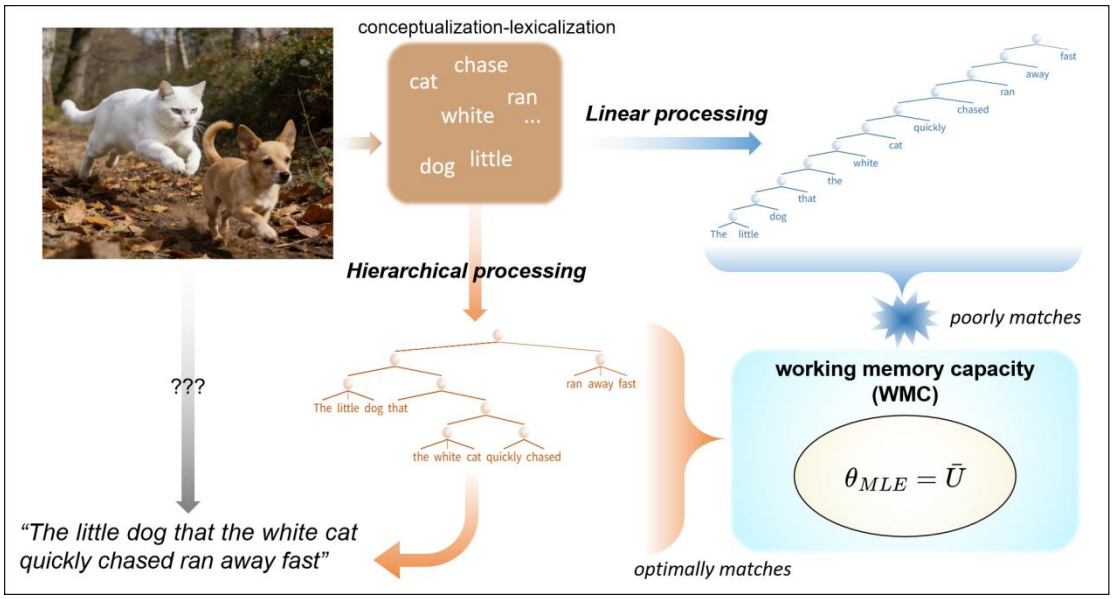
，

中国[4]教育部人文社会科学重点研究基地、天津师范大学心理学与行为学院，中国天津；[5]天津师范大学心理学院，中国天津。[6]麦考瑞大学医学、健康与人文科学学院语言学系，澳大利亚新南威尔士州悉尼。[7]天津市学生心理健康与智能评估重点实验室，中国天津[8]北京师范大学系统科学学院教育系统科学研究所，中国北京 *共同通讯作者：Luyao Chen 教授、博士，香港九龙红磡 Hung Lok Road 8 号，中国香港，harry-luyao.chen@polyu.edu.hk；Jinshan Wu 教授、博士，中国北京市海淀区新街口外大街 19 号，jinshanw@bnu.edu.cn；Angela D. Friederici 教授、博士，德国莱比锡 Stephan Straβe 1A，friederici@mpg.cbs.de # 高级作者

摘要

语言是人类独有的特征，它通过将词序列组织成句子中的层级结构[1-4]，高效地传递信息。一个核心问题始终悬而未决：为什么人类语言具有层级性？本研究表明，层级化能够最优地解决工作记忆容量有限这一挑战[5-7]。我们建立了一个似然函数，用于量化语言处理机制所涉及的平均单位数量与人类工作记忆容量（WMC）的吻合程度，并以直接 的方式进行衡量。该函数的最大似然估计（MLE） θ_{MLE}结果表明，这一数值就是单位的平均数量。通过对符号序列进行计算模拟，并对自然语言句子开展验证分析，我们发现，与线性处理相比，层级加工在序列／句子长度增加时，能更有效地将 θ_{MLE} 值限制在人类工作记忆容量上限以内。研究还显示，这一模式与儿童工作记忆容量发展相一致。这些结果表明，构建层级结构能够在不超出记忆约束的前提下，优化序列语言输入的加工效率，从而真正解释人类语言普遍具有层级性的原因[8]。

图文摘要



MAIN

人类交流主要依赖语言，使人们能够在短时间内高效处理大量信息。接收到一连串语言输入后，人类必须迅速将语言信号解码为语音、词语、短语，最终形成完整的句子[9]。正如 G. Miller[5] 几十年前所指出的：“我们的语言极其有用，能够将材料重新打包成少数几个信息丰富的信息块。”要理解一个句子，就需要将其各个组成部分组装成结构关系，而这又要求我们在有限的工作记忆（WM）中保持多个元素[10–15]。据推测，成年人的工作记忆容量（WMC）约为 4 或 7 个项目[5-7]。对于长度超出人类工作记忆容量限制的句子，如果逐词单独存储，或以线性方式存储，那么处理和理解这些句子将难以完成。因此，要处理更长的句子，就必须将词语整合为更大的组块，从而减少词语的有效数量，降低工作记忆负荷[16-17]。这一过程称为“层级化”，它是一种层次处理机制，通过将词语合并为更高阶的成分来保持加工效率[18]。例如，12 词句子“The little dog that the white cat quickly chased ran away fast.”可以划分为若干成分，如限定词短语“[the [white cat]],”、动词短语“[quickly chased],”以及关系从句“[that [[the [white cat]] [quickly chased]]]”（另见图 1a）。近期一项研究提出，层级结构可以通过全局和局部秩来实现；在神经层面，这两种秩分别编码为跨越二维空间的维度，以便高效维持语言信息[19]。尽管人们长期以来一直认为语言层级化与工作记忆容量等一般认知能力有关，但这种机制为何以及如何形成，仍是个谜[20]。人类与非人灵长类的差异，不仅在于工作记忆容量的总体大小（非人灵长类为2 ± 1[21]个项目），更重要的是，人类具备形成层级语言结构的能力，这是一种独特的人类语言能力[22-23]。以往研究发现，人类在处理语言时会采用“最省力”策略，这反映在某些语言特征中，例如尽量缩短依存距离；依存距离

用于量化一个词与其所依赖的另一个词相距多远（以两者之间间隔的词数衡量）[24-27]。一项最新研究提出，人类语言的结构应当能够最大限度地降低序列可预测性的复杂度，其衡量指标为超额熵[14]。然而，人类语言的结构或加工方式与领域一般认知能力之间的直接联系，至今尚未得到确立。具体而言，一个根本问题仍未解决：在人类工作记忆容量的限制下，人类语言的层级化是否是高效处理海量信息的最优方案?

为此，我们首先建立一个似然函数，即一种直接整合语言特征与工作记忆容量（WMC）的概率模型，用于检验以下假设：在相对宽泛但符合经验的句子长度范围内，语言层级化代表了应对人类工作记忆容量限制的最优方案[28]。该函数的核心是一个高斯函数，其中包含一个指数项，用于计算按照给定结构、通过统计每个词的开放节点数而获得的单位数量，与未知的工作记忆容量参数之间的标准化距离平方（在u_i 第 *i-th* 个词处）[29]；该参数记为 θ，有待估计。“开放节点”指句子加工动态过程中等待合并、并在工作记忆中主动维持的项目[29]。高斯函数假设u_i 与 θ 之间的差值服从正态分布。经过数学推导，该似然函数的最大似然估计（MLE）θ_{MLE}等于单位数量的均值（见图1a和方法）。因此，θ_{MLE} 越大，工作记忆负荷就越高。如图1a所示，加工“the little dog”需要θ_{MLE}，其值为2。值得注意的是，由于本研究中非人灵长类的工作记忆容量为 2 ± 1 个项目[21],，因此将人类工作记忆容量定义为4至9个项目。选择这一刻意放宽的范围，是为了避免围绕工作记忆“神奇数字”确切数值的争论[5-7]。我们的主要目标是探究，与分支结构固有的线性加工机制相比，语言层级化是否有助于降低工作记忆负荷（以 θ_{MLE}量化）。对抽象符号序列进行的模拟表明，语言层级化能够将 θ_{MLE} 值降低至人类工作记忆容量范围内，

与分支结构的线性加工相比，在自然语言句子中进行的进一步检验——包括英语经典作品的语料库、类型学上不同的语言语料库，以及英语儿童口语句子的语料库——验证了这些发现，并揭示了这一支撑语言层级的基本原则。

结果

首先进行了模拟，以确定将词语合并为各种多层级成分、从而模拟自然人类语言结构所生成的层级结构，也就是“层级加工机制”，是否优于分支结构的线性加工机制。分支结构的例子是“[...[[[[s_0],s_1]s_2]s_3]...s_n]”，它每次都将一个词线性地与此前的词合并（图1b）。层级结构通过定义两种合并机制生成（图1b）：（a）严格平衡二叉合并，要求每个层级上的父节点包含两个子节点（除非只剩一个子节点），例如“[[[s_0, s_1], [s_2, s_3]], [[s_4, s_5], [s_6, s_7]]]”。（b）宽松多节点合并，要求父节点随机包含1至4个子节点，例如“[s_0, [[[s_1, s_2], [s_3, s_4, s_5]], [s_6, s_7, s_8]]]”。二叉合并被认为是生成层级结构的最简单机制[18, 22]。对于每种序列长度（本研究范围为1至100个词），每种机制都随机生成了1,000个结构标记。如图1c所示，显然：（a）依赖这两种合并机制的语言层级化优于分支结构所依据的线性加工机制，即使序列长度极长，也能最大限度地减少WMC负荷；（b）其 θ_{MLE} 以对数方式增长；（c）在合理的句子长度范围（5至30个词）[28]内，相应的 θ_{MLE} 主要处于人类WMC范围内。此外，还计算了每种序列长度下各加工机制的香农熵，两种合并机制的熵均低于分支结构的线性加工机制（图1d）。

----- Inset Fig. 1 about here ----

自然语言中的验证进一步呈现出 θ_{MLE} 分布的一致模式。我们进行了若干验证步骤。首先，从九部随机选取的经典作品（以英语创作或译成英语）中抽取文本进行初步验证。其次，为了检验该机制的普遍性，我们进行了跨语言验证。我们选取了刘易斯·卡罗尔创作的《爱丽丝梦游仙境》（*1865*年），这部作品是世界上最受欢迎、最具代表性的儿童经典作品之一，并选取了它的七种其他语言译本，包括中文、法语、德语、俄语、日语、意大利语和西班牙语，此外还包括英语原文。第三，为了探索该机制对语言发展的有效性，我们在儿童发展情境中（年龄范围为3至10岁）进行了验证测试。我们使用Stanza[30]对这些自然语言语料库中的句子结构进行自动分析，并进行单元／开放节点计数。对经典语料库的验证测试显示，随着句子长度增加，语言层级化使 θ_{MLE} 保持在人体工作记忆容量范围内（平均值为 $\theta_{MLE} = 5$），这与模拟结果一致（图2a）。同样，这八种语言都呈现出相似的 θ_{MLE} 分布模式，其中语言层级化优于线性加工机制（平均值为 $\theta_{MLE} = 4$）（图2b-c）。

----- Inset Fig. 2 about here ----

此外，就“儿童口语语料库”而言，验证结果在所有年龄组中均揭示出一致的模式（图3a-b）：语言层级化降低了 θ_{MLE}，且各年龄组的平均 θ_{MLE}呈现出一条发展曲线（非线性回归：$R2 = 0.746$，$F = 14.657$，$p < .05$）（图 3c），与工作记忆容量的发展轨迹[31] 十分吻合。

----- Inset Fig. 3 about here ----

讨论

正如深刻的物理定律支配着苹果落地这一平凡现象，日常的人类语言交流也受固有的自然

原则支配：语言使我们能够将线性序列信息折叠成层级结构，并在理解和产出过程中加以运用。在本研究中，我们提出指标“θ_{MLE}”，以直接 量化人类语言结构与工作记忆容量之间的关系。我们首先分别对人工语言材料和自然语言材料进行探索性数据模拟和验证性数据模拟，并以新的实证证据揭示了一种潜在的语言定律：语言层级化是人类语言的一种显著机制，它受限于我们有限的工作记忆容量；与分支结构所依赖的线性加工机制相比，语言层级化能够将工作记忆负荷最优地降低至特定范围。我们将语言层级化的两条子定律概括如下：

(a) 效率： $\theta_{\mathrm{MLE}|\mathrm{H}} << \theta_{\mathrm{MLE}|\mathrm{L}}$ （即层级加工 [H] 比线性加工 [L] 所需的 [<<] 工作记忆负荷少得多）；

(b) 生态效度： $\theta_{\mathrm{MLE}|\mathrm{H}} \subset_{\approx}$ 人类工作记忆容量（即层级加工主要能够减轻工作记忆负荷，使句子长度适中时的负荷维持在人类工作记忆容量范围内）。

先前研究表明，层级加工与人类工作记忆容量之间存在关联[17, 25], ，尽管支撑这两种能力的神经基质可能彼此分离[34–38]。然而，这些结果无法回答一个更根本的问题：人类为何需要层级结构及其底层的层次处理机制？我们的初步假设是，为了在可加工的句子中“装入”更多信息，我们必须生成层级结构，以减轻工作记忆的负担。这与开篇引用的 G. Miller 的话相呼应[5]；但我们更进一步，强调语言层级化的重要性：在人类工作记忆容量有限的约束下，层级化尤其有助于实现高效的语言加工。为直接检验这一假设，我们构建了一个似然函数，用于量化语言结构与工作记忆容量之间的关系。模拟结果一致表明，随着 θ_{MLE}

句子长度增加，语言层级化在减少人类工作记忆容量范围内的[27]方面优于线性加工机制。这些结果或许还能进一步解释为何人们偏好较短的线性依存距离[24-25, 27]以及较低的预测努力/超额熵[26],：这些特征存在于线性序列中，而线性序列实际上是由内部层级结构外化而来的[1]。具体而言，语言层级化能够最优地缓解句子熵，从而提高加工效率（图 1d）。因此，θ_{MLE} 在直接联系语言层级化与工作记忆容量、解释加工效率与一般认知（此处即工作记忆容量）约束之间的张力方面发挥着关键作用。

对不同语言中的自然语言句子进行验证后，发现了一种一致的模式。本文测试的所有自然语言语料库都表明，语言层级化对于最大限度降低 θ_{MLE} 的数值，并使其保持在人类WMC范围上限[5-7]之下至关重要。此外，这种模式在类型学差异语言中普遍存在，这与近期一项研究[39],的发现一致。该研究发现，语言层级普遍性是人类语言中最显著的普遍性之一，说明了这一机制的普遍性。语言类型学参数对这些语言中 θ_{MLE} 数值的个体间差异所产生的影响，是一个极具前景的未来研究方向。此外，正如儿童口语语料库的验证结果所揭示的，θ_{MLE} 数值的变化与儿童工作记忆容量发展轨迹[31],高度吻合，这表明语言层级化能力很早就已出现，并与儿童工作记忆容量的发展密切相关[40]。值得注意的是，儿童的 θ_{MLE} 在7岁之后显著超过“3”（图3c），这令人联想到先前研究[4]提出的句法能力变化。大约在 7[4]岁之后，儿童便能够处理层级嵌入更深的句法复杂句。更重要的是，本分析提供了一种严谨的模拟方法，可以反驳儿童在口语句子生成中依赖线性加工机制这一假设。与语言层级化不同，将层级结构视为仅供线性处理的分支结构，无法描绘实际的工作记忆容量发展轨迹[31]。

总之，本文揭示，人类语言的层级本质并非随意形成，而是在面临人类工作记忆限制时不可避免的最优解决方案。通过建立似然函数并推导出度量指标“θ_{MLE}”，直接关联语言处理机制与工作记忆，我们基于对抽象序列和自然句子语料库进行的直接计算模拟，始终如一地为语言内在地趋向层级组织提供了一种普遍的机制解释。展望未来，任何语言演化模型都必须解释这种“必然性”；这意味着，未来的神经认知科学研究应将这一根本联系视为理解人类独特语言能力的发展与复杂性的核心要素。

方法

建立最大似然估计的似然函数

该函数表达了待加工单元数量与工作记忆容量（WMC，记为“θ”）之间的概率关系。如果二者在某一加工阶段完全匹配，例如 7 个单元与 WMC = 7 匹配，那么该结构便可得到高效加工，既不会浪费工作记忆资源（即待加工单元少于 WMC 值的情况），也不会使我们的 WMC 超负荷（即单元数量超过 WMC 上限的情况）。因此，假定某一阶段的完全匹配反映了最经济/最优的加工效率。我们首先采用开放节点计数方法[29] 遍历层级结构，统计每个单词中的单元数/开放节点数（u_i）。然后，对于每个 u_i，设定一个高斯函数，以确保结果范围为 0 到 1，其中 1 表示与给定 WMC 完全匹配：

$$e^{-\frac{1}{2}(\frac{u_i-\theta}{\sigma})^2}$$

因此，似然函数定义为：

$$L(u|\theta) = \prod_{i}^{n} e^{-\frac{1}{2}(\frac{u_i-\theta}{\sigma})^2}$$

其中，“n”是单词数。θ 的最大似然估计（MLE）为：

$$\theta_{MLE} = \bar{U} = \frac{1}{n}\sum_{i=1}^{n} u_i$$

此外，还计算了香农熵（另见图 1a）。对于每种处理机制（平衡二叉合并、左分支结构和随机多节点合并 [1–4节点]），我们首先使用一种考虑嵌套深度和子组件偏移量的层级遍历算法，计算序列中每个词对应的开放节点数。根据各词开放节点数所形成的分布，我们推导出概率质量函数 p_i，其定义为每种不同开放节点数的出现频率除以句子长度。随后，我们计算香农熵（以 2 为底，单位：比特），公式如下：

$$H = -\sum_{i} p_i \log_2(p_i)$$

其中，总和不包括零概率项，以避免数值奇异性。对于非确定性结构（左分支结构、多节点合并），我们对 1,000 次随机实例化得到的熵值取平均，以减小抽样方差；对于确定性的严格平衡二叉合并，每种句子长度只需计算一次。

人工符号序列中的模拟

由无意义符号组成的序列（例如用 s_1, s_2, 等表示词语）构成了后续模拟所用的“句子池”（最长序列包含 100 个词）。请注意，这些符号不包含语义信息，以避免语义促进；分析的仅是为每种给定长度的序列随机分配的结构。如正文所述，这些结构包括分支结构，以及由严格二叉合并和宽松多节点（1–4 个节点）合并机制生成的层级结构。每种序列长度均生成了 1,000 个结构标记。需要指出的是，宽松多节点合并的结构类型会随句长增加而增多，而另外两种类型除了长度变化外不会改变。结构标记指包含不同标记（符号）的结构。例如，“[[[s_0]s_1]s_2]”和“[[[s_3]s_4]s_5]”是两个结构标记，但它们共享同一种结构类型（即三词句子的左分支结构）。因此，严格二叉合并和左分支结构都会产生变化很小的固定结构类型。开放节点计数的计算方法是：将最外层列表指定为共同的父节点，并将其直接元素识别为直接子组件；第一个子组件中的词从 1 开始计数，每深入一个嵌套层级便递增 1；后续子组件中的词则在基于深度的计数上加上此前子组件的数量作为偏移量（另见图 1a）。由此得到不同序列长度下 θ_{MLE} 值的全局均值曲线。

自然语言验证测试

我们在自然语言句子中验证了数值模拟结果。首先，从某在线电子书网站（https://www.gutenberg.org/）下载量前100名列表中随机选取了 9 部英语经典作品。这些作品包括《傲慢与偏见》（简·奥斯汀，1813）、《小妇人》（路易莎·梅·奥尔科特，1868）、《罪与罚》（费奥多尔·陀思妥耶夫斯基，1866）、《简·爱》（夏洛蒂·勃朗特，1847）、《双城记》（查尔斯·狄更斯，1859）、《了不起的盖茨比》（F. Scott Fitzgerald，1925）、《汤姆·索亚历险记》（马克·吐温，1876）、《战争与和平》（列夫·托尔斯泰，1869）以及《远大前程》（查尔斯·狄更斯，1861）。随后，我们基于“经典语料库”，分析了线性加工机制（将句子解析为分支结构）和层次处理机制各自的全局均值 θ_{MLE} 曲线。其次，为了检验我们假设的跨语言普适性，我们选取了著名的儿童经典作品《爱丽丝梦游仙境》（刘易斯·卡罗尔，1865）[33]。考虑到译本的版权问题，并为统一翻译风格，我们请 Deepseek API（3.2 版，非思考模式）将英语原作翻译成另外 7 种语言，包括中文、法语、德语、俄语、日语、意大利语和西班牙语，从而构建“爱丽丝语料库”。基于该语料库，我们分别针对两种加工机制计算了各语言的 θ_{MLE} 曲线和全局均值 θ_{MLE} 曲线。最后，我们还希望检测语言发展过程中 θ_{MLE} 值的变化，因为儿童工作记忆容量会随年龄增长而发展[32],，其语言输出应反映其在每个年龄阶段的工作记忆容量（限制）。我们从 CHILDES儿童语言数据交换系统（https://talkbank.org/childes/）中，按年龄组随机选取了各年龄段的儿童英语口语句子（例如 $3 \leq$ 年龄 < 4 岁，记作 3–4 Y；3–10 岁各年龄组各 3,000 句），以构建“儿童口语语料库”。我们计算了各年龄段的 θ_{MLE} 曲线和全局均值 θ_{MLE}曲线，并使用非线性回归估算θ_{MLE} 值的变化。需要指出的是，所有语料库中的自然语言句子均按照句长标准进行清理：句长的第 25 百分位数（*Q1*）− 1.5 IQR $\leq$ 有效句长 $\leq$句长的第 75 百分位数（*Q3*）+ 1.5 IQR（IQR：“四分位距” = *Q3* − *Q1*）。各语料库的描述性统计见补充信息 表 S1。

致谢

我们衷心感谢Liping Feng教授和Emiliano Zaccarella博士在本研究中给予的启发性讨论和建设性意见。

作者贡献

Luyao Chen： 概念构思、正式分析、软件、撰写初稿、审阅与编辑，以及经费获取；**Weibo Gao：** 正式分析、软件及撰写初稿。**Junjie Wu：** 审阅与编辑。**Jinshan Wu：** 概念构思、正式分析及审阅与编辑。**Angela D. Friederici：** 概念构思、正式分析、软件、撰写初稿及审阅与编辑。

利益冲突

作者声明，本研究不存在利益冲突。

经费支持

本研究得到科技创新2030—重大项目+2021ZD0200500的支持。

数据和代码可用性

所有计算模型及分析代码均可获取，并将在我们的 GitHub 账户上发布。小型语料库可在提出合理请求后获取。

参考文献

1. Berwick, R. C., Friederici, A. D., Chomsky, N., & Bolhuis, J. J. 进化、大脑与语言的本质。《认知科学趋势》17, 89–98 (2013).

2. Everaert,M. B. H., Huybregts, M. A. C., Chomsky, N., Berwick, R. C. & Bolhuis, J. J. 结构，而非字符串：语言学作为认知科学的一部分。《认知科学趋势》19, 729–743 (2015). 3. Friederici, A. D. 语言加工的大脑基础：从结构到功能。Physiol

.

《心理学评论》91，1357–1392（2011年）。4. Friederici, A. D.《我们大脑中的语言：独特人类能力的起源》。麻省理工学院出版社（2017年）。5. Miller, G. A.《神奇的数字七，加减二：关于我们信息加工能力的一些限制》。《心理学评论》63，81–97（1956年）。6 Cowan N.（2001年）。《短时记忆中的神奇数字4：重新思考心理存储容量》。《行为与脑科学》24(1)，87–185。7 Cowan N.（2010年）。《神奇的神秘数字四：工作记忆容量受到怎样的限制，以及为什么？》。《心理科学当前方向》19(1)，51–57。8. Chomsky, N., Seely, T. D., Berwick, R. C. 等。《合并与强最简主义论题》。剑桥大学出版社（2023年）。9. Chomsky, N.《句法结构》。海牙：Mouton（1957年）。10. Baddeley, A., Hitch, G. J. & Allen, R. J.《句子回忆中的工作记忆与绑定》。《记忆与语言杂志》61，438–456（2009年）。11. Caplan, D. & Waters, G. S.《言语工作记忆与句子理解》。《行为与脑科学》22，77–126（1999年）。12. Christiansen, M. H. & Chater, N.《“当下即逝”瓶颈：语言的一项基本约束》。《行为与脑科学》39，e62（2016年）。13. Hitch, G. J., Allen, R. J. & Baddeley, A. D.《工作记忆多成分模型问世五十年后》。《实验心理学季刊》78，222–239（2025年）。14. Futrell, R. & Hahn, M.《序列信息加工瓶颈所产生的语言结构》。《自然·人类行为》（2025年）。15. Thalmann, M., Souza, A. S. & Oberauer, K.《组块化如何帮助工作记忆？》J

.

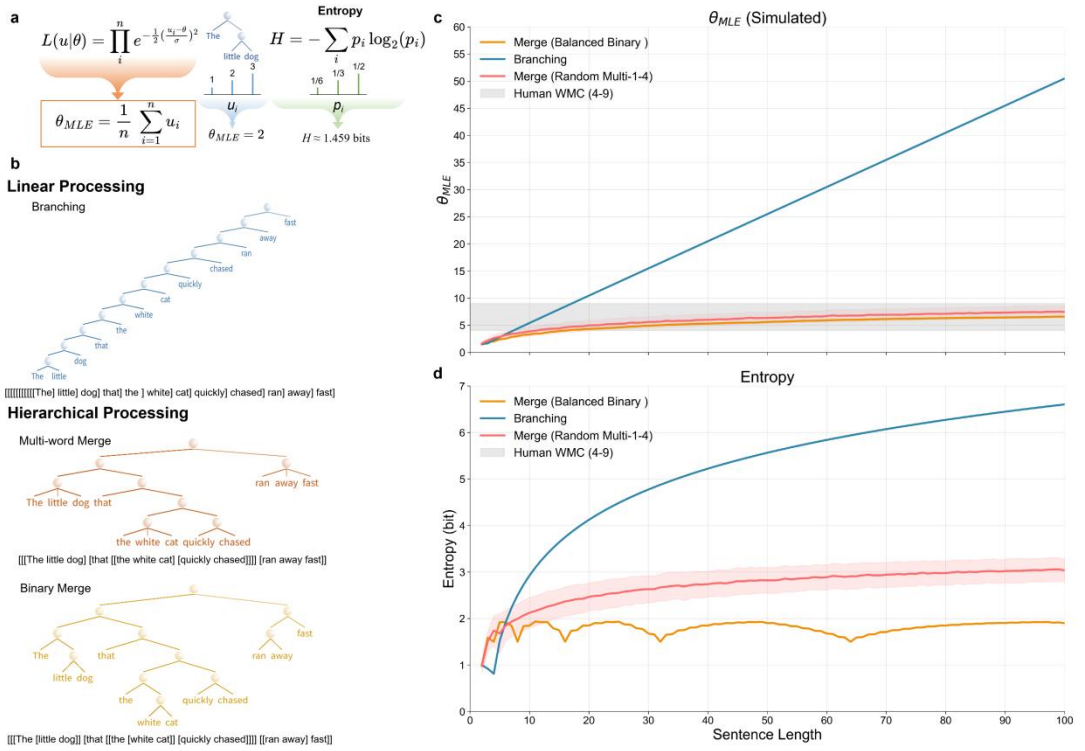
实验心理学：学习、记忆与认知 45，37–55（2019年）。

16. Bonhage, C. E., Fiebach, C. J., Bahlmann, J. & Mueller, J. L. 句子结构工作记忆的脑特征：增强编码与促进维持。《认知神经科学杂志》26，1654–1671（2014年）。17. Rogalsky, C. & Hickok, G. 布罗卡区在句子理解中的作用。《认知神经科学杂志》23，1664–1680（2011年）。18. Chomsky, N.《最简方案》。麻省理工学院出版社（1995）。19. Fan, Y., Wang, M., Fang, F., Ding, N. & Luo, H. 二维神经几何支撑人类工作记忆中序列的层级组织。《自然·人类行为》9，360–375（2025年）。20. Arnon, I. 等。人类语言何以可能？一个生物文化框架。《科学》390，6775（2025年）。21. Read, J. C. A. 等。非人灵长类的工作记忆。《神经科学与生物行为评论》136，104576（2022年）。22. Berwick, R. C. & Chomsky, N.《为何只有我们？》。麻省理工学院出版社（2016）。23. Hauser, M. D., Chomsky, N. & Fitch, W. T. 语言官能：它是什么、谁拥有它，又是如何演化的？《科学》298，1569–1579（2002年）。24. Gildea, D. & Temperley, D. 语法是否会最小化依存长度？《认知科学》34，286–310（2010年）。25. Gibson, E. 语言复杂性：句法依存的局部性。《认知》68，1–76（1998年）。26. Futrell, R., Mahowald, K., & Gibson, E. 37种语言中依存长度最小化的大规模证据。《美国国家科学院院刊》112，10336–10341（2015年）。27. Liu, H., Xu, C. & Liang, J. 依存距离：研究自然语言句法模式的新视角。《物理生命评论》21，171–193（2017年）。28. Zhou, Y., Jiang, J. & Liu, H. 句子长度中的语言共性：比较10种语言的句子长度分布。《认知科学》49，e70115（2025年）。29. Nelson, M. J. 等。句子加工过程中短语结构构建的神经生理动力学。《美国国家科学院院刊》114，E3669–E3678（2017年）。

30. Qi, P., Zhang, Y., Zhang, Y., 等。Stanza：面向多种人类语言的 Python 自然语言处理工具包。arXiv 预印本 arXiv:2003.07082，2020年。31. Alloway, T. P.《改善工作记忆：支持学生学习》。伦敦：Sage出版社，2010年。32. Soni, A. & Frank, M. J. 自适应组块改善前额叶皮层与基底神经节回路中的有效工作记忆容量。eLife 13, RP97894（2025年）。33. Malik-Moraleda, S. 等。对45种语言和12个语系的研究揭示了普遍语言网络。《自然神经科学》25, 1014–1019（2022年）。34. Gao, C. 等。连续θ短阵刺激显示了语言网络特异性对句法加工的因果效应。《神经影像》306, 121014（2025年）。35. Fiebach, C. J. 等。重新审视布罗卡区在句子加工中的作用：句法整合与句法工作记忆。《人脑图谱》24, 79–91（2005年）。36. Fedorenko, E., Behr, M. K. & Kanwisher, N. 人脑高级语言加工的功能特异性。《美国国家科学院院刊》108, 16428–16433（2011年）。37. Makuuchi, M. 等。将人类语言的核心计算能力与工作记忆分离。《美国国家科学院院刊》106, 8362–8367（2009年）。38. Yang, M. 等。对左侧额下回进行经颅光生物调节，可提升普通话第一语言和第二语言复杂句加工表现。《脑与语言》256, 105458（2024年）。39. Verkerk, A., Shcherbakova, O., Haynie, H. J., Skirgård, H., Rzymski, C., Atkinson, Q. D., Greenhill, S. J. & Gray, R. D.（2025年）。贝叶斯空间系统发育分析揭示的语法持久约束。《自然-人类行为》，10.1038/s41562-025-02325-z。在线优先发表。40. Newport, E. L. 语言学习的成熟度约束。《认知科学》14, 11–28（1990年）。

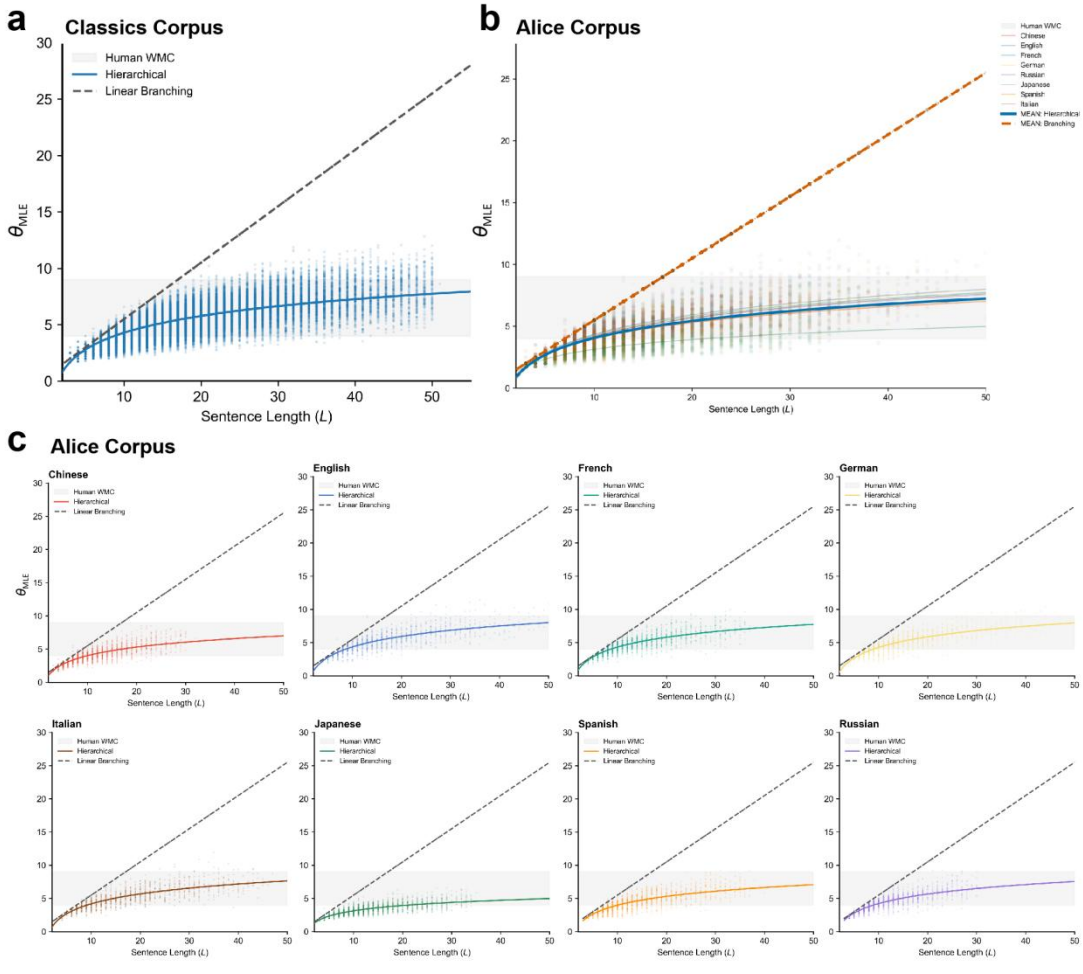
图注

图 1. 数值模拟结果。



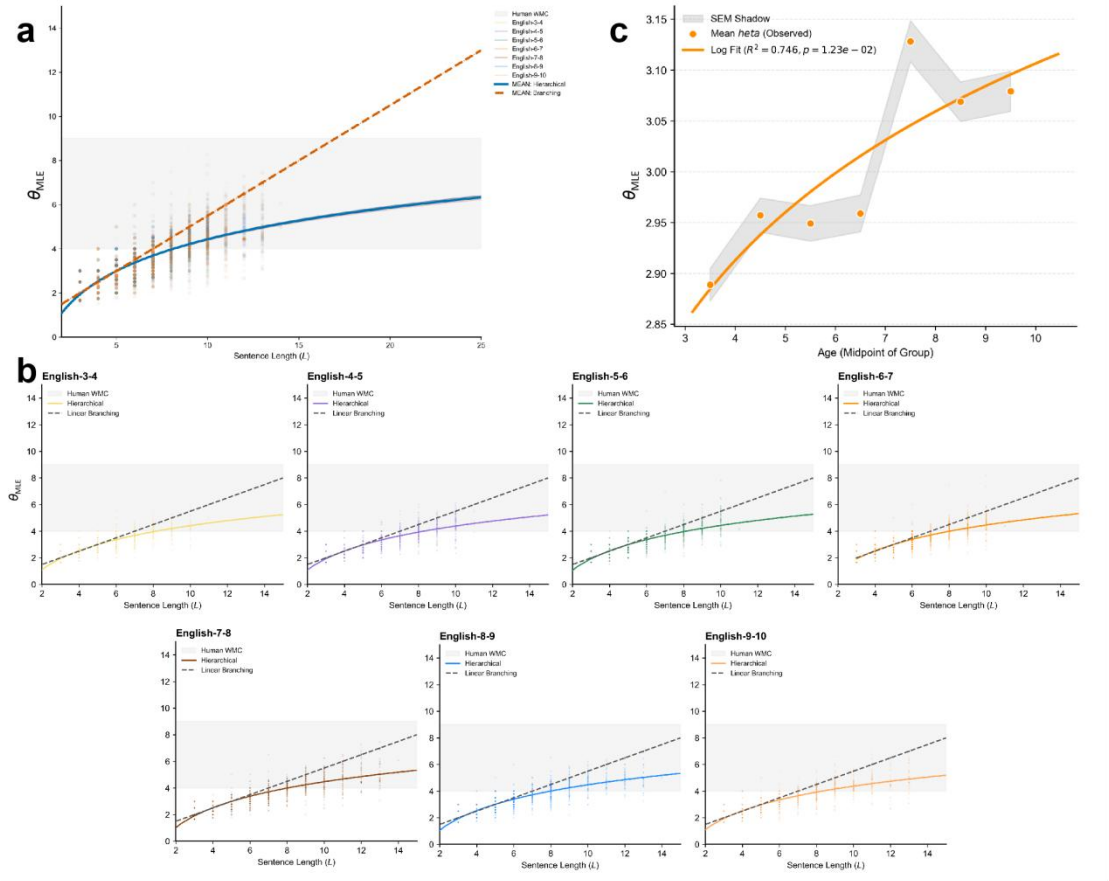
a.本研究采用的核心公式。提供了计算 θ_{MLE} 和熵（H）的示例。**b.** 线性处理机制和层次处理机制示意图。**c.** θ_{MLE}的模拟结果。**d.** 熵的模拟结果。

图2. 书面自然语言语料库中的验证结果。



a. “经典语料库”中的验证测试结果。**b.** “爱丽丝语料库”中所有语言的全局均值验证测试结果。**c.** “爱丽丝语料库”中每种语言的模拟结果。

图3. 儿童口语语料库中的验证结果



a. “儿童口语语料库”中所有年龄组的全局均值验证测试结果。**b.** “儿童口语语料库”中每个年龄组的验证测试结果。**c.** θ_{MLE} 与年龄组之间的非线性回归结果。