

# **Optimal Scheduling of the Leaves of a Tree and the SVO Frequencies of Languages**

Check for  
updates

**摘要：**本文定义并研究了一个新的关于二叉树叶顺序调度的优化问题，并进行了寻找最优顺序的推广。我们假设调度过程从树的根开始，并并行地继续宽度优先，尽管可能有干预锁定和解锁步骤，这定义了调度成本。产生这个问题的动机来自于在大脑中建模语言生成。我们表明，这个问题中的最优性考虑为语言学中一个有趣的现象提供了一个新的解释，即句子中主语、动词和宾语的某些排序方式在世界语言中比其他语言中常见得多。

**关键词：**二叉树最优叶调度语言基本词序

## 1. 简介：叶片调度问题

考虑一个二叉树 $T$ ，e.g. 图中的一个。1-二叉树的意思是向下向树 $2n-1$ 节点，一个节点的学位（根）， $n-2$ 节点的三度，和 $n$ 节点（叶子），假设我们也给定一个顺序 $\sigma$ 的叶子，说顺序（SVO）在这个例子中。我们感兴趣的是根据以下规则为树的节点分配整数倍：

1. 根被分配的时间为0；
2. 一个非根节点 $i$ 要么被分配给时间 $t+1$ ，其中 $t$ 是分配给它的父节点的时间，要么它被它的父节点锁定；
3. 叶子被分配的时间按照给定的顺序严格增加， $\sigma$ ；
4. 如果有一片叶子 $l$ 被分配一个时间 $t$ ，那么一个锁定的节点 $i$ 可能被分配时间 $t+1$ ，在这种情况下我们说！解锁我。

我们说，如果一个分配满足这些规则，那么它是一个可行的时间表。直观地说，这种时间分配形式化了树的节点从根“火”开始，节点的子节点对Oc施普林格自然瑞士AG 2022

Oc Springer Nature Switzerland AG 2022

D. E. Simos et al. (Eds.): LION 2022, LNCS 13621, pp. 3–14, 2022.  
[https://doi.org/10.1007/978-3-031-24866-5\\_1](https://doi.org/10.1007/978-3-031-24866-5_1)

在他们的父母之后。例外的是，节点可以在自己触发时选择锁定它的一个或两个子节点。一片叶子，在触发时，可以解锁一个被锁定的节点。很明显，给定一棵树和它的叶子的顺序，总是有一个可行的时间表：总是锁定不按顺序导致下一个叶子的子节点，而任何触发的叶子都按顺序打开下一个叶子被锁定的祖先。

现在定义叶子调度问题如下：给定一个树 $T$ 和其叶子的顺序 $\sigma$ ，找到一个成本最小的可行调度，其中调度的成本是使用的锁定命令的数量（等价地，解锁命令的最小数量）。我们还可以通过为每个可能的锁定和解锁命令分配一个权重，并最小化这些权重的和来定义这个问题的加权版本。我们可以进一步定义一个更复杂的问题，称为最优叶阶问题，其中我们只给定一个树 $T$ ，并且我们寻找具有最小调度代价的阶 $\sigma$ 。

例如，图中SVO阶的最优叶形调度问题。1A是分配0给根，1给S，内部节点，2给V，3给O。也就是说，内部节点锁定O，而V解锁它。这个解决方案需要一个成本，因为使用了一个锁，而且很明显没有一个零成本的解决方案。实际上，S-V-O和S-O-V的顺序是最优叶订单，成本为1，而其他四个订单的最优成本为2。对于一个更复杂的例子，读者可能想要验证图中的树。1B的叶顺序从左到右，最优叶调度成本为3，而该树的最优顺序为ADBC和ADCB，成本为1。正如我们将在下一节中看到的，这两个算法问题都可以用贪婪算法来解决——除了具有权值的最优叶序问题，这是np困难的。

## 动机：自然语言中的Word顺序

这些问题有趣是因为它们与语言学中的一个经典问题有关，我们接下来会解释。在英语中，一个句子的主语通常出现在动词的前面，而宾语如果存在，则跟在两者后面：“狗追猫”。这种排序并不是通用的，因为其他语言采用了六种可能的排序中的任何一种，参见例如[4]。与英语中相同的顺序，即SVO，在法语、希伯来语、现代希腊语和罗马尼亚语中普遍存在，总体上约占世界语言的42%。SOV目更为常见，占各种语言的45%，包括印地语、乌尔都语、日语、拉丁语和古希腊语。VSO（9%）、VOS（2%）和OVS（1%）的订单要不常见得多，而OSV（< 1%）实际上被忽略了。在英语中，改变语言的SVO顺序会产生毫无意义的句子（“追逐猫，狗”），或者改变意思（“猫追逐狗”）。在其他语言中，如德语、俄语或现代希腊语，偏离标准顺序是可以容忍的，因为名词在这些语言中有一个例子，这使得它们的句法角色（主语vs宾语）很容易独立地识别

位置然而，许多语言学家认为，大多数语言都有一个占主导地位的、默认的词序。

大量的文献证明了基本词序的频率有很大的变化，见[1、8、10、11、14、17 -19, 22]。这些过去的解释是基于合理的语言原则，与交流意义的容易性，或学习语法[8, 9, 14, 22]的困难有关，而最近的解释考虑了语言中词序的变异性和进化[17 -19]。在这里，我们提出了一种不同的解释，基于在大脑中表达句子的难度。

事实上，人们可以假设，为了产生一个像“猫追狗”这样的句子，说话者必须首先通过他们大脑中的神经元回路，创建一个句子的树表示，如图所示。1A, C. 有认知证据[6]表明，这棵树是二元的（也就是说，没有超过两个孩子的节点），事实上，三个叶子“猫”、“追逐”和“狗”是有组织起来的，如图所示。1A, C（而不是替代品，e. g., 其中S和V先合并）；见图。3在[20]。现在有了这棵树，说话者必须清晰地表达它，这涉及到选择和实现六个词序中的一个。为了到达其中的一个命令，一个神经机制的“锁”和

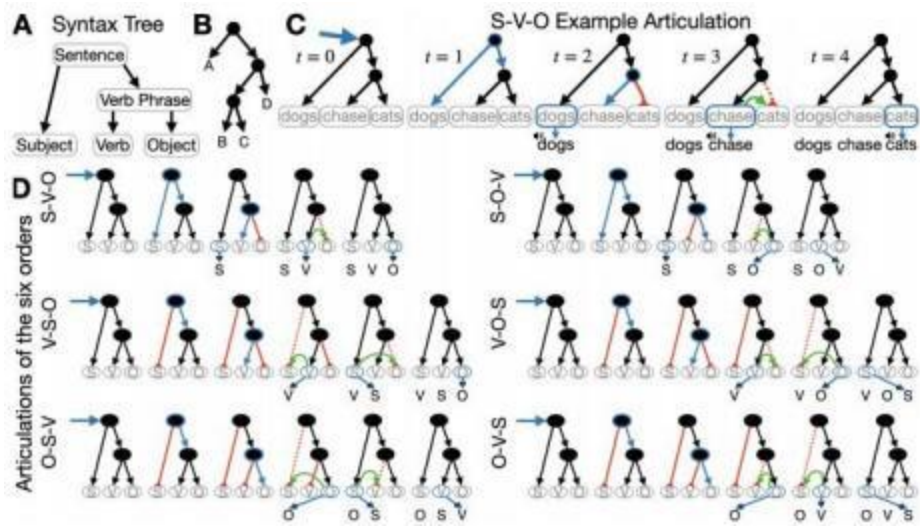


图1. (A) 基本语法树与“动词短语”内部节点和“主语”、“动词”和“对象”叶子。(B) 另一种二叉树示例，比基本的语法树更复杂。(C) 从句法树到SVO顺序语音的例子。黑色箭头处于非活动状态。蓝色箭头会激活它们在下一个时间步中所指向的对象。红色的抑制信号对它们所指向的物体保持着锁定。绿色箭头可移除锁。(D) 发音的所有可能的基本词序的顺序语音，从同一句法树开始。适当的锁定和解锁操作决定了基本的字序。

可以使用“解锁”步骤。在教派。3我们指出，这在[20]中提出的大脑计算模型中可以很容易地完成。为了有效的信息传递，所有使用一种语言的人最终都使用相同的固定顺序，这是有道理的；尽管大多数语言在发音顺序上更加灵活，但许多语言学家认为每种语言的[5, 8, 17, 19, 22]都有占主导地位的顺序。但是这六个订单中的哪一个将被选为主导订单呢？我们认为，一个词序在大脑中的实现成本越小，选择这个顺序的可能性就越大。在[20]中的大脑计算模型中，并在第二节中进一步解释。3、树的每个节点都位于不同的大脑区域，远程抑制性神经元被用来锁定和解锁大脑区域，从而表达句子。

本文的其余部分组织如下。2我们在算法上研究了叶排序问题，而在第二节。3我们在[20]中阐述了大脑计算的模型及其实现句子生成的方法。这个模型最终为词序概率的差异提供了一个解释：两个最常见的词序对应于两个最优解，而其他四个

则滞后于后者。采用一种统计力学风格的模型来计算订单的频率，人们甚至可以预测各种锁和解锁步骤的成本差异，这将最好地解释这些频率。

## 2 贪婪算法

回想一下引言中定义的两个问题：叶排序问题寻求实现给定序列的锁定和解锁步骤的最优调度，而最优顺序问题希望找到使这个最优成本最小化的顺序。这两个问题都可以被加权。

### 定理1。

1. 通过贪婪算法，可以在  $O(n \log n)$  时间内求解叶片排序问题；加权情况也是如此。
2. 如果所有的锁定和解锁步骤都有单位代价，则最优顺序问题可以通过同一贪婪算法的自适应来解决。
3. 然而，如果解锁步骤有不同的成本，即使成本被限制为一个或两个，最优顺序问题也是  $np$  困难的。

证明。（1）我们非正式地描述了该算法。它需要连续触发树的所有节点，从根开始；触发从一个节点传播到它的子节点（由一个节点队列实现的宽度优先搜索）。具体来说，根在第一个平行时间步长被触发。在步骤  $t + 1$ ，在步骤  $t$  中父触发的内部节点将触发。此外，任何在  $t$  时刻被另一个叶子解锁的叶子都将在  $t + 1$  时刻触发（在任何时间步长中最多会有一个被解锁的叶子）。最后，如果触发的一个内部节点有任何叶子节点，那么每个子节点都被锁定，除非它是下一个要输出的叶。

为了跟踪叶子，我们维护一个由  $\sigma$  排序的单独锁定的叶子堆，最初为空，然后是一个索引，最初是1。如果在某个步骤中遇到正在处理的节点的叶子  $i$ ，有两种情况：如果下一个  $\sigma(i) =$ ，并且在这个步骤中没有输出其他叶子，那么叶子立即输出，下一个增加1。否则，接下来是  $\sigma(i) >$ ，并且我加入了被锁定的叶子堆。在并行步骤  $t$  的开始（处理深度  $t-1$  的树的节点的第一轮宽度搜索），我们检查堆的最小值，称为  $m$ ，接下来是否有  $\sigma(m) =$ 。如果是这样，那么我们输出  $m$ ，下一个输出增量。然后我们继续进行广度优先的搜索。当堆和队列都为空时，该算法将终止。

我们声称，该算法以  $\sigma$  的顺序输出叶子，并且它使用最少的锁定和解锁操作和最少的并行步骤。我们首先声称，根据平行时间，每一片叶子都尽可能早地输出。这可以从两件事得到：(a) 叶子  $i$  可以比时间  $T(i)$  更早输出，其中  $T(i)$  满足递归  $T(i) = \max\{T(\sigma^{-1}(\sigma(i)-1)) + 1, \text{深度}(i)\}$ ，如果我不是第一个叶子， $T(i) = \text{深度}(i)$ ，否则；(b) 算法实现了这个时间，可以通过  $\sigma(i)$  的归纳显示。我们还声称它实现了排列最少的锁，这遵循两个事实 (c) 最小可能的锁数是  $n-1$  巧合的数量，巧合是一个我的两个术语的递归定义是相同的，和 (d) 这样的巧合捕获和利用算法。

(2) 对于最优顺序问题，我们首先注意到每个叶子  $i$  都可以在时间深度  $(i)$  处输出。其次，只有在叶子在可用的精确时间输出时，才可以在没有锁定/解锁步骤的情况下输出。否则，如果许多叶子有相同的深度，除了一个叶子外，所有的叶子都可以推迟到未来的任何时候，并通过在它之前输出的叶子来解锁。因此，下面的贪婪算法实现了最小的锁定/解锁步骤数：我们定义了一个从  $n$  个叶子到时间隙  $\{d, d, d+1, \dots, d+n\}$ ，其中  $d$  和  $D$  是树的一个叶子的最小和最大深度：首先，每个叶子  $i$  被映射到深度  $(i)$ ，这创建了一个由于碰撞而不是一对一的映射。然后我们从  $d$  开始反复通过时间槽，从小到大，并执行以下算法：对于任何时间段  $t$ ，如果有的话  $! > 1$  的叶子映射到它，选择  $! - 1$  中的这些叶子，并将它们分配给  $! - 1$  空时隙大于  $t$  且最接近  $t$ ，任意解析关系。很容易看出，该算法在前一段的意义上，选择具有最大巧合数量的叶子的排列（叶子是火的，恰好在它们可用的时候），从而选择最小可能的锁定和解锁步骤数。

(3) 最后，对于  $np$  硬度：假设这棵树是一个深度为  $n$  的完整二叉树，即  $n = 2d$  所有的叶子都会同时到达。然后所有的排列都是可用的，我们需要选择  $\sigma(1), \dots, \sigma(n)$  的排列...， $\sigma(n)$ ，这样  $\sum_{i=2}^n$  解锁成本  $(\sigma(i-1), \sigma(i))$  尽可能小。很容易看出，这是（开环）旅行推销员问题的一个通用实例，

这是np困难度的，即使边的长度是一个或两个[21]。这就完成了对第(3)部分和该定理的证明。

### 3. 在大脑中产生句子

它现在已经被神经科学家广泛接受，在大脑中，信息项目，如物体、想法、单词、事件等。是由大量的峰值神经元所代表的。这些种群被称为集合体。在[20]，提出了一个计算系统，其基本数据项是神经元的组装，其操作包括合并，创建一个与两个已经存在的组件具有强大突触连接的组件，以及抑制和解除抑制大脑区域的操作。请注意，通过重复应用合并操作，可以构建树。事实上，一个简单的句子，比如“狗追猫”，可以通过首先识别与词汇中的三个单词对应的三个集合来生成——它们被认为位于左内侧颞叶[7]。然后，这些词集合项目在颞上回韦尼克区域的不同子区域内创建三个新的集合，对应于主语、动词和对象的大脑区域。接下来，动词和对象程序集（在本例中分别对应于“chase”和“cats”）合并，从而在Broca的区域中创建一个动词短语程序集[7]。最后，主语和动词短语程序集合并，创建一个表示句子图的程序集。在布罗卡地区[7]的另一个分区。一个句子可能有许多其他的组成部分，如行列式、形容词、副词和命题短语，但在这里，我们只关注由它的三个基本句法部分组成的树：主语、动词和宾语。

通过首先将其中的三片叶子分组，可以构建三棵不同的二叉树。在语言学[5, 14, 17, 22]中有广泛的共识，以及来自认知实验[6]的证据中有广泛的共识，支持上面描述的基本树。1A与一个内部的动词短语节点，其成分是动词和对象。

一旦句子被生成，它就可以被发音，即转换为言语。这可以通过刺激树的根——句子集合——来实现，然后它将刺激树中的子节点，以此类推。最终，这三片叶子都会很兴奋。每一片叶子都可以调动运动程序，以表达每个单词，但这必须按顺序进行。因此，必须选择并执行这六个订单中的一个。也许是最简单和最简单的实现特定顺序的生物学现实机制涉及两个看似可信的原语，我们称之为锁和解锁。这些原语对应于熟悉的一个区域抑制（抑制性神经元群体的激活，将阻止该区域的兴奋性神经元放电）和去抑制（抑制性群体的抑制）的神经过程[3, 12, 16]。特别是，在射击时，树上的一个集合可以抑制它的一个子代的射击。其次，任何一片叶子，在燃烧时，都可以解除抑制任何其他的叶子。

### 1. 3. 调度成本解释了SVO的频率

我们已经看到，在这六个订单中，只有两个可以通过一个锁和一个解锁操作来实现，而其他所有订单都需要两个锁和两个解锁操作。1D. 换句话说，这个简单的模型立即预测了频率统计量的“最高阶位”，即SVO和SOV序的流行率。除了这两个顺序之外，所有其他顺序都需要额外的抑制和去抑制，这些已知的原体需要显著的大脑能量消耗[2, 3, 15]。此外，额外的操作使发音过程变得更加复杂，并可能使学习者学习语言的这一方面对他们来说更加困难。

### 3. 2叶片调度成本作为能源

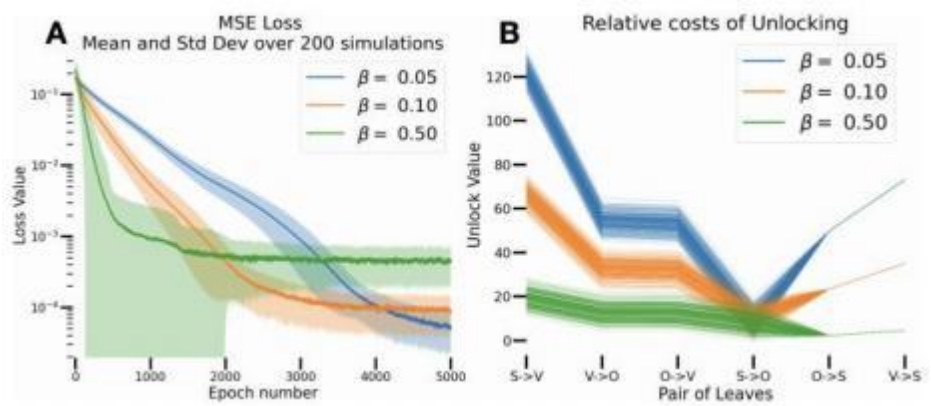
在文献[5, 8, 17-19]中，有人认为，语言在其历史上经历了转变，其中的词序发生了变化，因此当前的频率反映了这种动态过程的动态平衡。这一观点激发了一个简单的统计力学公式，将基本词序的频率视为玻尔兹曼分布[13]，其中能级为L的状态普遍存在，其概率与 $e$ 成正比 $- \beta L$ ，对于一个温度参数 $\beta$ 。为了简单起见，我们在这个考虑中使用 $\beta=1$ （但在我们的实验中，我们对 $\beta$ 图使用了广泛的值。2）。我们的模型的状态是六个基本的词序，相关的能量是每个发音选择所需的操作次数。SVO或SOV的最佳选择能量较低，只需要两个操作（一个锁和一个解锁），而其他四个最佳选择能量较高，需要四种操作（两个锁和两个解锁）。1D. SVO、SOV、VSO、VOS、OSV、OVS六个目的患病率与数字 $e$ 成正比 $-2, e^{-2}, e^{-4}, e^{-4}, e^{-4}, e^{-4}$ 各自地SVO和SOV订单将比其他订单高 $e$ 倍 $2 \approx 7.4$ 、预测频率（.39, .39, .055, .055, .055, .，一个很好的经验分布的一阶近似（.45, .42, .0902, ...01, .01）。

一个大脑区域锁定和解锁另一个区域的真正成本可能会有所不同，这取决于所涉及的两个大脑区域之间的距离和它们的神经连接的强度，以及目标区域锁定状态的逐步持续时间。通过引入这些超参数，除了 $\beta$ ，并将它们与观测数据进行拟合，我们实际上可以预测它们的值。也就是说，通过抑制性的神经连接，预测大脑区域之间的连接。结果证明，这些预测对包括 $\beta$ 在内的各种超参数都是稳健的。玻尔兹曼分布模型为估计这六种基本词序的频率提供了基础。我们将这些频率[方程2]与经验观测相等，并数值求解了六个非线性方程组。我们注意到，方程显示出解析简并度，这也可以从模拟中恢复；具体来说，六个参数中的四个只能确定为一个共同的加法



持续不断的这种简并性在等式中很明显1，因为这四个参数不能与其他两个参数进行比较。

方程组没有解析解，但这六个参数可以用梯度优化来近似。该方法对不同的系数  $\beta$  图的值得到了相同的定性结果。2A和其他超参数的不同值。这些计算的结果足够可靠，足以支持关于从另一个图中解除一个大脑区域的相对成本的某些预测。2B. 更具体地说，我们发现：



**图2** . 不同  $\beta$  值在梯度下降时期绘制的 (A) 损失值。这些线代表平均损失，阴影区域代表超过 200 个初始化的标准偏差。(B) 从一个叶子到另一个叶子的解锁操作的相对成本。颜色代表具有不同  $\beta$  值的模型。每一行代表一个模型的优化参数。注意前四个叶对的解的简并性：这些线的区别只有一个加性常数。

$$US \rightarrow V \rightarrow UV \rightarrow O \text{ 封 } UO \rightarrow V \rightarrow US \rightarrow O \text{ 和 } UV \rightarrow S \rightarrow UO \rightarrow S, \quad (1)$$

其中  $U_{北} \rightarrow y$  是解除抑制组件  $y$  和组件  $x$  的成本。

### 3 3a. 统计学上的力学论证

在统计力学中，一个系统处于给定状态的概率取决于它的能量和温度参数。玻尔兹曼分布提供了一种估计系统所有状态的热平衡结构的方法。一个能量为  $E$  的状态出现的概率  $i$  与  $p$  成正比  $i \propto \exp(-\beta E_i)$ ，其中  $\beta$  是一个比例因子，与温度成反比，而  $E_i$  这取决于各自的解锁成本  $U_{北} \rightarrow y$  状态  $i$ 。我们所描述的系统有六种状态，因此，每种状态的概率是：

$$p_i = \frac{\exp(-\beta E_i)}{\sum_j \exp(-\beta E_j)}, \tag{2}$$

对于*i, j* ∈ {SVO, SOV, VSO, VOS, OVS, OSV }。

这些公式只是一种启发式的方法，与物理原理相一致，用来建模复杂性如何影响概率；然而，我们也注意到，在神经科学文献中(见e. g. [15])详细讨论了关于热能的代谢成本。因此，我们选择我们的态的能量为*E<sub>i</sub>*~1，我们假设是β元1。

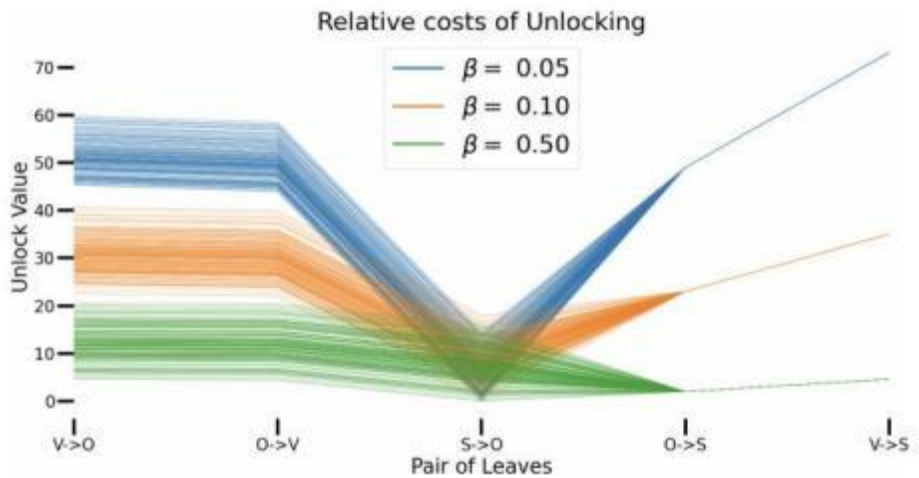


图3 . 预测主语和动词的首要性是否是OSV和OVS阶低频率的原因。请注意，对U没有预测S → V.

#### 4讨论

语言现象应该在新的见解下不断地重新解释，包括我们对大脑中语言处理的理解或理论的进步。尽管最近在这方面取得了进展，但由于认知和系统神经科学在规模和焦点上都存在很大的差距，阐明涉及语言功能的神经过程的限制并不容易。我们的工作试图利用装配微积分的计算框架来弥补这一差距，从而提供了一个新的解释：语言中六种基本词序的频率差异，难以从句子的基本语法树生成顺序。

我们的模型的最简单的版本定性地匹配了观察到的基本词序频率，而最复杂的版本可以被调整来预测准确的频率。然而，我们怀疑后一种计算可能构成过拟合，因为在确定这些频率时可能会考虑其他因素，包括对通信效率的语言考虑

和可学习性。这些其他因素是迄今为止唯一用于这一目的的因素。我们的模型并不是要替换这些参数，而是添加它们，它为打破基本词序的对称性提供了额外的基础。

我们相信，对词序现象的最终解释将结合语言和神经计算的证据，也许还有易学性的考虑，以及更多的证据。如何做到的一个例子，让我们把语言订单的极端罕见的主要原因订单“0”可能不是解锁的困难主题或动词子区域的对象区域根据我们的模型，但相对子句语义作用的对象在一个句子，相比首要的主题和动词[14, 17, 22]。面对这一点，我们可以决定OSV和OVS阶的低频率在语言学的基础上得到充分的解释，并通过相应的方程来解释剩下的四个频率。这导致了4个有5个参数的方程(因为 $US \rightarrow V$ 不再进入画面中)。为了平衡方程和参数的数量，我们可以确定参数的比值 $UO \rightarrow S$ 和 $UV \rightarrow S$ (在图中不受简并性影响的那些。2)，并通过梯度下降求解。结果如图所示。3. 我们注意到我们的预测，你 $V \rightarrow O$ 封 $UO \rightarrow V > US \rightarrow O$ 和 $UV \rightarrow S > UO \rightarrow S$ 是稳定的，而我们之前的预测是 $US \rightarrow V$ 是非常大的消失，因为这个解锁操作只在OSV顺序中起作用，我们忽略其频率。换句话说，预测是“ $US \rightarrow V$ “非常大”被认为是OSV和OVS的频率较小的原因，这一现象现在有了另一种基于语言学原则的因果解释。也许它仍然是你的世界上 $S \rightarrow V$ 可能会很大，可能是因为这种大脑连接很少被使用，但可能需要使用一个不同的模型或实验证据来解决这个问题。

## References

1. Aitchison, J.: Word order universals. John A. Hawkins Quantitative Analysis of Linguistic Structure Series. Academic Press, New York, London, Sydney, 1983–1986. [https://doi.org/10.1016/0024-3841\(86\)90039-2](https://doi.org/10.1016/0024-3841(86)90039-2)
2. Attwell, D., Laughlin, S.B.: An energy budget for signaling in the grey matter of the brain. *J. Cerebral Blood Flow Metab.* **21**(10), 1133–1145 (2001). <https://doi.org/10.1097/00004647-200110000-00001>. PMID: 11598490
3. Buzsáki, G., Kaila, K., Raichle, M.: Inhibition and brain work *Neuron* **56**(5), 771–83 (2007). <https://doi.org/10.1016/j.neuron.2007.11.008>, <https://www.sciencedirect.com/science/article/pii/S0896627307009270>
4. Dryer, M.: Order of subject, object, and verb. In: Martin, H., Matthew S.D., David, G., Bernard, C. (eds.) *The World Atlas Language Structures*, chap. 81, pp. 330–333. Oxford University Press, Oxford (2005)
5. Dunn, M., Greenhill, S.J., Levinson, S.C., Gray, R.D.: Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* **473**(7345), 79–82 (2011). <https://doi.org/10.1038/nature09923>



21. Papadimitriou, C.H., Yannakakis, M.: The traveling salesman problem with distances one and two. *Math. Oper. Res.* **18**(1), 1– 11 (1993).  
<http://www.jstor.org/stable/3690150>
22. Tomlin, R.S.: Basic word order. *Functional principles*. No. 1, Cambridge University Press (CUP) (1988). <https://doi.org/10.1017/s0022226700011646>.  
<https://www.cambridge.org/core/journals/journal-of-linguistics/article/abs/russell-s-tomlin-basic-word-order-functional-principles-london-croom-helm-1986-p-308/7542AFB4A8B28D651F6E109B810F4C04>