观点交锋

利用自然语言语音绘制人体大脑皮层的语义地图

译者：李若愚，吴佳炜，刘知远 / 清华大学

摘 要：语义信息由大脑皮层中统称为语义系统的一组区域表示。然而，现有的研究不能全面地阐释语义信息在整个语义系统区域上的表示情况，同时关于语义系统上大部分区域的语义选择性的研究也相当匮乏。本文使用体积元建模的方法，处理被试收听有声叙述故事时采集到的大脑磁共振成像数据，全面系统地绘制了大脑皮层的语义选择性地图。建模结果说明语义系统的语义表示本身具有复杂的分布模式，但其分布模式在不同被试间似乎是一致的。之后利用全新设计的生成模型，得到详细的语义地图集。实验结果表明，语义系统中的大部分区域都表示其特定的语义模块，即包含相关概念的一组词汇；同时绘制的大脑语义地图集也详尽地说明了每个区域所具体表示的语义模块的种类。这项研究同时也说明，目前在人脑神经解剖和功能连接研究中常用的数据驱动的方法，提供了一种强大而高效的手段以获取大脑功能表示定位。

此前的神经影像研究已经识别出大脑中的一组极有可能表示语义信息的区域。这些区域统称为语义系统（semantic system），相比非词汇（non-words）、音韵学任务（phonological tasks）和随机无序的语言语音（temporally scrambled speech），语义系统对单词（words）、语义任务（semantic tasks）和自然语言语音（natural speech）有更显著的反应。针对语义系统中特定类型表示的研究表明，语义系统对具体或抽象的词汇、行为动词、社会性叙事等语义特征具有区域选择性。其他研究也发现，语义系统对特定的语义模块（semantic domain），既包括相关概念的一组词汇，例如单词 living things、tools、food 和 shelter 属于同一个语义模块，也具有区域选择性。然而，现有的研究仅仅在少量的刺激条件下进行，所以迄今为止并没有研究能够全面地阐释语义信息在整个语义系统上是如何表示的。

为试图解决上述问题，我们利用数据驱动的方法，对听到含有多个语义模块的有声记叙故事后大脑的反应进行建模。利用功能性磁共振成像（fMRI）技术，我们记录了 7 名被试在收听 2 个多小时的“飞蛾广播时间”（The Moth Radio Hour）故事节选的过程中全脑血氧浓度（BOLD，blood-oxygen-level-dependent）的变化。接着我们利用一种对复杂自然脑刺激建模的高效方法——体积元建模（voxel-wise modelling），估计在大脑每个体元区域上的语义选择性（见图 1a）。

体积元像素模型的估计和验证

在体积元建模的过程中，首先从刺激中提取感兴趣的特征，随后使用回归的方法来确定在每个体积元中，不同的特征是如何影响全脑血氧浓度的。我们使用词向量空间来确定故事中出现的每个单词的语义特征。我们选取了 985 个常见的英语单词（如“above”、“worry”、“mother”等词汇）作为基准词汇（即 985 维语义特征），在大规模的英语语料库中，通过计算每个单词和基准词汇的正则化共现度（normalized co-occurrence）构建词汇表示空间。属于相同语义模块的单词倾向于在相似的文本中出现，因此具有相似的共现度。例如，单词 month 和 week 在语义十分接近，其共现向量相关度为 0.74，而单词 month 和 tall 的共现向量相关度仅为 −0.22。

接着使用正则化线性回归（regularized linear

regression）估计出在每个被试体内，这 985 维的语义特征在不同的大脑皮层体积元上是如何影响全脑血氧浓度的（见图 1a）。为了描述一些由于实验刺激中的低级特征造成的影响，比如语速、音素等特征，在进行体积元建模估计时先引入了额外的回归因子，但在进一步的结果分析中将其舍弃。同时引入了额外的回归因子来描述生理和情感的因素，但这些因素对最终估计的语义模型没有影响。

相比于传统的神经影像方法，体积元建模方法的一大优点在于，可以使用在模型估计过程中没有引入的神经刺激对全脑血氧浓度变化进行预测，从而对估计的模型进行验证。这使得我们可以计算分析估计得到的模型对实际情况的解释程度。我们利用一段 10 分钟长度在模型估计阶段没有使用的新的“飞蛾广播时间”故事（见图 1b），来测试体积元模型对 BOLD 的预测能力，结果发现我们估计的模型在整个语义系统，包括外侧颞叶皮层（LTC，lateral temporal cortex）、腹侧颞叶皮层（VTC，ventral temporal cortex）、外侧顶叶皮层（LPC，lateral parietal cortex）、内侧顶叶皮层（MPC，medial parietal cortex）、内侧前额叶皮层（medial prefrontal cortex）、顶上前额叶皮层（SPFC，superior prefrontal cortex）、顶下前额叶皮层（IPFC，inferior prefrontal cortex）区域的体积元上都具有较好的预测表现（见图 1c）。这表明语义系统的大部分区域都具有语义选择性。

大脑皮层中语义表示的定位

审视估计的语义模型，我们能够确定在每个大脑体积元中所能表示的特定语义模块。理论上，只要分别检测每个体积元，就能得到整个大脑的语义表示定位。然而，每个被试的实验数据都包含成千上万个体积元，使得这种方法不具有操作性。另一种可行的方法是将估计的语义特征映射到一个低维向量子空间，同时保留尽可能多的体积元上的语义表示信息。对于每个被试，我们采用主成分分析法（PCA，principal components analysis）分析估计的语义特征模型，得到了 985 个正交语义维度（主成分），并根据每个维度的方差在总方差中的占比进行排序。我们倾向于认为，其中只有一部分的语义维度是不同被试所共有的，其余的语义维度则反映了被试间的个体差异、fMRI 的噪音和截取的故事的统计特征。为了确定这些共有的语义维度，测试每一个语义维度是否在模型间相较于偶然因素表现出更大的方差，而这方差是由用于模型评价的激励矩阵主成分确定的。在一共 7 名被试的 6 人中，至少有四个维度表现出明显的方差（P< 0.001，Bonferroni-corrected bootstrap test），而对于最后一名被试，只有三个维度是显著的。这表明我们获取的 fMRI 数据在被试者中包含约 4 个在统计上显著的语义维度。

这 4 个共有的语义维度提供了一个直观的方式来总结每个体积元区域的语义选择性。为了解释体积元模型上映射的语义特征维度中的信息，需要理解语义信息在这四维语义特征维度空间是如何编码的。为了可视化语义空间，我们将故事中出现的 10 470 个单词从词汇表示空间分别映射到每一个语义特征维度上。然后使用 k-means 聚类方法定义了 12 种词汇目录，最终手工处理和标记了这些词汇种类。这 12 个词汇种类分别为“tactile（触觉）”（如“fingers”）、“visual（视觉）”（如“yellow”）、“numeric（数值）”（如“four”）、“locational（地点）”（如“stadium”）、“abstract（抽象）”（如“natural”）、“temporal（时间）”（如“minute”）、“professional（职业）”（如“meetings”）、“violent（暴力）”（如“lethal”）、“communal（公共）”（如“schools”）、“mental（精神）”（如“asleep”）、“emotional（情感）”（如“despised”）、“social（社会）”（如“child”）。

接下来，将这 12 种词汇目录在共享的语义特征空间上可视化（见图 2a）。每个目录标签通过 RGB 色彩来表示。其中红色、绿色、蓝色的通道值分别由第一个、第二个、第三个维度决定。在全部的 7 名被试的数据中，第一个语义维度的方差在总方差的占比最高。该语义维度的一端存在的词汇目录主要与人类和社会相互作用有关，包括“social（社会）”、“emotional（情感）”、“violent（暴力）”和“communal（公共）”。该语义维度的另一端存在的词汇目录则主要和触觉描述、数量表达、环境有关，包括“tactile（触觉）”、“locational（地点）”、“numeric（数值）”和“visual（视觉）”。这与此前的猜测是一致的，人类包含一个特别显著且表征强烈的语义模块。语义特征空间中方差占比靠后的语义维度通常所含信息量少，难于仔细阐释。

第二个语义维度的两端分别为与感知相关的词汇目录如“visual（视觉）”、“tactile（触觉）”，以及与感知无关的词汇目录，如“mental（精神）”、“professional（职业）”、“temporal（时间）”。排名第三和第四的语义维度则更加难以解释。

之前的研究发现语义系统包含在大脑的皮层区域中，但是并没有全面地阐释这些区域的语义选择性。通过将体积元模型映射到共有的语义特征维度上，能够可视化在整个大脑皮层区域上语义模块分布的区域选择性。图 2b 展示了将一名被试的体积元模型映射到前三个共有语义维度时的图像，其中 RGB 颜色的标注方法和图 2a 中相同。因此在图 2b 中，一个绿色的体积元能够对语义空间中呈绿色标记的词汇目录中的单词，比如“visual（视觉）”和 “numeric（数值）” 中的单词，产生更显著的血氧浓度变化。可视化结果显示，语义信息的表示区域在整个语义系统有复杂的分布模式，包括前额叶皮层、外侧颞叶皮层、内侧颞叶皮层、外侧顶叶皮层和内侧顶叶皮层的大部分区域。但同时，这些区域的分布模式在被试个体间又具有较强的一致性。

利用 PrAGMATiC 算法构建语义地图集

鉴于在不同的被试上，语义系统区域的语义选择性分布模式有很强的一致性，我们希望建立一个人类大脑皮层上的地图集来描述语义选择性功能区域的分布。为了实现这一目标，设计了一种新的贝叶斯算法 PrAGMATiC 来获得覆盖大脑皮层区域的概率生成模型。这个算法为体元精度下模拟的致密层叠的大脑功能同源区域地图的功能调谐模式建立了模型，同时考虑到了不同被试间大脑解剖结构和功能构造上的差异。不同功能区域的组织和选择性由使用 fMRI 数据通过类似对比散度 (contrastive divergence) 的最大似然估计法学习得到的参数决定。其中一些参数在不同被试间是共享的，这些参数描述了被试群体大脑皮层语义地图的共有性质。其他的参数则是被试间独立的，这些参数反映了被试间的差异。通过估计这些共享的和独立的参数，能够规避一般建模时要将不同被试间大脑解剖结构和功能的数据进行相互对应的问题。

PrAGMATiC 算法包括两个部分：安排模型（arrangement model）决定不同功能区域在皮质表层的位置；发散模型（emission model）在不同功能区的安排的基础上生成语义地图。安排模型模拟一个物理学上的弹簧网络结构，将每个功能区域的质心和相邻的质心连接起来。每个被试建模时都有相同的平衡弹簧长度，但是每个弹簧均可以独立的拉伸或压缩。功能区域的安排同时也受限于大脑功能区域上的界标（functional landmark），这些界标是在每个被试独立的数据中标记我们感兴趣的大脑已知区域得到的。这些限制条件保证了在不同被试间，绘制的语义地图是相似的，但同时保留了在功能区域的大小和细节的安排上被试间的个体差异。借助于安排模型，发散模型能够将每个皮质表层区域的顶点分配给最近的区域质心点，从而创建大脑功能同源区域。每个皮质表层区域顶点的语义功能值通过多元正态分布得到。而每个区域的的平均语义功能值在每个被试间保持一致，通过算法自学习得到。我们将语义功能值定义为一个四维向量，这个四维向量能够反映估计模型中大脑每个体积元在前文中提到的共有的四个语义维度的表示情况。

PrAGMATiC 算法中的一个重要超参数是覆盖大脑皮层的区域数量。我们使用交叉验证的方法在每半个脑区选择合理的覆盖区域数量，同时测试这些区域是否具有语义选择性。首先使用 6 名被试的数据对 PrAGMATiC 算法模型进行估计，然后在仅获知第 7 名被试的大脑皮层生理结构和功能区域上的界标的条件下，对其大脑语义地图进行预测。使用预测得到的语义地图对第 7 名被试的血氧浓度进行预测，和实验中收集的真实数据进行比较，从而来观察 PrAGMATiC 算法模型在不同被试间的泛化能力。随着覆盖大脑皮层的区域数量从 8 上升到 128，预测准确度快速上升，128 之后仍有缓慢提高（见图 3b）。在大脑左半球部分，预测准确度在区域数量超过 192 后没有统计学意义上的明显提升（FDR>0.1，使用包含被试个体随机差异影响的 Tukey 事后检验法）。在大脑右半球部分，预测准确度在区域数量超过 128 后没有统计学意义上的明显提升。然而，由于 PrAGMATiC 算法模型对大脑皮层全覆盖处理，先前的这些区域中同时包含具有语义选择性和不具有语义选择性的区域。为了挑选具有语义选择性的区域，同时去除没有语义选择性的区域，在每个区域检测该区域的平均体积元语义模型是否能够比仅利用语速、音率、音素等低级语

言特征的平均预测模型更好地预测血氧浓度变化。通过这种方式，能够去除对语义和低级语言特征没有选择性的区域，比如运动和视觉皮层区域；同时也能够去除没有专一语义选择性的区域，如 Broca 区域，因为在这些区域语义模型权重的不确定性较大。

图 3c 展示了一名被试映射到大脑皮层表面的语义地图集的图像。左半球大脑包括 77 个语义区域（FDR<1/192，bootstrap 检验），右半球大脑包括 63 个语义区域（FDR<1/128，bootstrap 检验）。结果显示外侧顶叶皮层（LPC）、内侧顶叶皮层（MPC）和顶上前额叶皮层（SPFC）上覆盖有多个表示不同语义模块的区域。在外侧顶叶皮层和内侧顶叶皮层的中央区域（二者分别靠近角形脑回和顶下沟）对“social（社会）”词汇目录具有语义选择性，而它们的周边区域则对“numeric（数值）”、“visual（视觉）”和“tactile（触觉）”词汇目录具有语义选择性。在顶上前额叶皮层，中心区域主要对“social（社会）”词汇目录具有语义选择性，而背外侧的区域的语义选择性则更加多样。外侧顶叶皮层、内侧顶叶皮层和顶上前额叶皮层同属于默认模式神经网络（DMN，default mode network），通常认为与内省、沉思和意识思维有关。那么存在一种有趣的可能——我们的模型中确认的语义区域同样可以表示大脑进行意识思维活动时的语义模块。这也许表明人思考的内容和内心言语可以尝试使用体积元模型进行解码。相比外侧顶叶皮层、内侧顶叶皮层和顶上前额叶皮层三个区域，外侧颞叶皮层（LTC）在我们的语义地图集中只覆盖了极少的语义选择性区域。考虑到外侧颞叶皮层同属于默认模式神经网络，并且被认为在语言理解中发挥重要作用，这一实验结果非常令人惊讶。然而，由于实验中颞叶前部记录到的 fMRI 信号质量较差，因此外侧颞叶皮层可能包含其他使用我们现有技术方法不能覆盖到的语义选择性区域。外侧顶叶皮层、内侧顶叶皮层、顶上前额叶皮层、外侧颞叶皮层、腹侧颞叶皮层（VTC）、顶下前额叶皮层（IPFC）以及鳃盖骨皮层（opercular cortex）和岛叶皮层（insular cortex）上语义表示的详细分析、相关讨论和与早期神经影像、损伤实验结果的比较都可以在补充材料中获取。

讨 论

我们获得的大脑皮层语义地图中一个引人注目的发现是语义选择区域相对对称地分布在大脑皮层的左右两半球侧，这与之前大脑损伤实验得到的语义表示偏向于存在大脑左半球的结果并不一致。然而，大量 fMRI 实验表明语义表示只有中等程度的偏侧性，同时一项利用叙述性故事为实验素材的研究发现了与本文类似的两侧分布结果。这项结果表明，相比于 fMRI 研究中常用的单词和短语素材，右半球大脑可能对叙事性刺激具有更强的反应。当然，我们需要进一步的研究才能确定大脑的左右半球中的语义区域在理解语言中发挥的作用。

另一个有趣的发现是大脑语义选择性区域的分布在不同被试间具有高度的一致性，这可能是因为大脑的先天生理解剖结构联结或者大脑皮层的细胞结构限制了高级语义表示的组织。当然，这也有可能是被试共同的生活经验导致的，因为他们都在西方工业社会出生并接受教育。我们还需要进一步研究具有不同成长背景的被试，从而确定这样的被试间语义选择性脑区组织一致性在多大程度上分别受到大脑结构和被试经验影响的。

实验中我们采用的 PrAGMATiC 算法存在一个限制性假设：每一个区域都具有功能同源性。虽然这是一个在神经成像研究中实验设计和数据分析过程中非常常见的假设，但是，包括视觉皮层区域的语义地图在内的许多大脑皮层表示地图，常常具有平滑的梯度变化的表示。因此，PrAGMATiC 算法应该进行进一步的优化，来对功能梯度表示进行更为精细的建模。这使其能够更加客观的说明，语义地图最佳的描述方式究竟是功能同源区域还是功能梯度区域。

数据驱动方法在人类神经解剖学和静息态脑网络的研究中已经非常普遍，但是在功能成像研究中才刚刚兴起。我们的研究表明了数据驱动方法在大脑功能区域定位研究中的强大与高效。虽然我们的实验设计中仅仅简单地让被试收听语音故事，但是获得的丰富实验数据仍然足够描绘出一个全面的大脑语义选择性的图集。并且，我们的数据驱动的研究框架具有普遍意义。通过构建能反映语音、句法结构等信息的语言特征，这些语言信息也能够在脑区上得到定位，甚至可以使用本项研究的数据集。融合词汇共现向量的复杂语义模型也能够进行量化

的评测和比较。这些模型的泛化能力则可以使用非自传体式故事作为大脑刺激来进行测试。有时，我们很难将数据驱动的实验结果和假说驱动的实验结果结合起来，但是随着方法和理论的进一步发展，两者将能很好地建立联系。希望这项研究中的语义地图集能够对研究语言的神经生物学基础的研究人员有一定帮助；也希望通过综合未来的研究成果，使语义地图集能够更准确并得到进一步扩展。因此，我们创建了一个在线的互动版本的语义地图集，读者可以在 http://gallantlab.org/huth2016 进行探索实验。

配 图

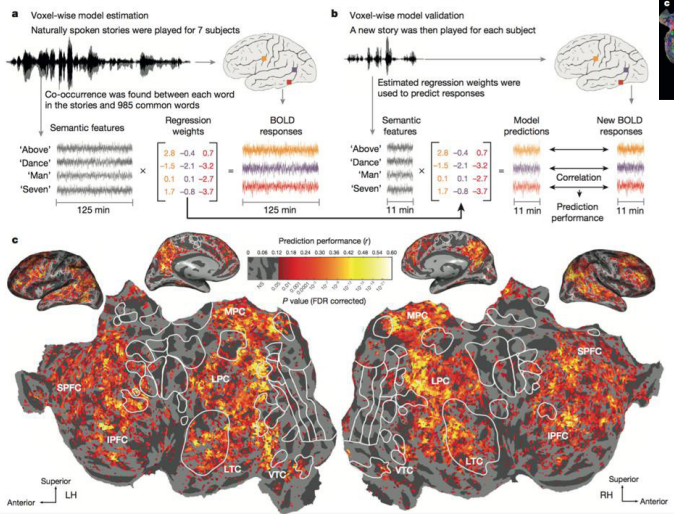

图 1 体积元建模（a，在 7 名被试收听 2 个多小时的有声记叙文故事节选的过程中，利用 fMRI 测量血氧浓度变化。通过计算在大规模语料库中故事中出现的每个单词和 985 个基准词汇的词汇共现统计信息，将这些单词映射在 985 维的词汇向量表示空间中。对每个被试的数据使用有限冲激响应（FIR）回归模型来估计每个体积元。体积元模型中每个单词的权重表示这个单词在故事中对血氧浓度信号的影响程度。图 b，利用一段 10 分钟长度的在模型估计阶段没有使用的语音故事，对估计的模型进行测试。使用模型预测的血氧浓度变化和实际测得的血氧浓度变化的相关系数来衡量模型的预测能力。图 c，估计得到的体积元模型对一名被试的血氧浓度变化预测结果。估计的语义模型在大量脑区上有准确的预测效果，包括 LTC/VTC/LPC/MPC/SPFC/IPFC。这些区域已经被之前的研究确认为大脑语义系统的一部分。LH 为左半球；RH 为右半球。）

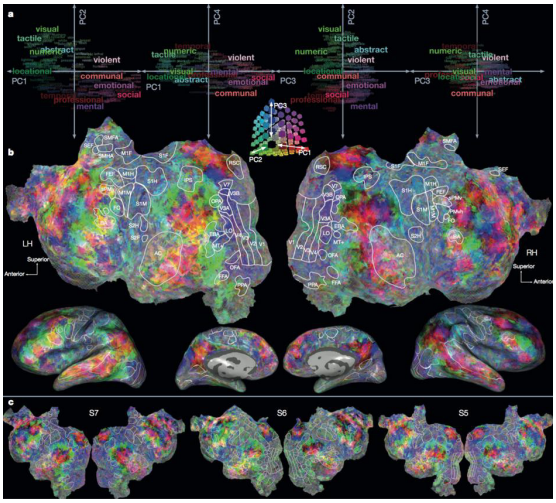

图 2 体积元语义模型的主成分（a~c 显示出了对体积元模型的权重进行主成分分析，得到 4 个大脑中共有的语义特征维度。a，基于语义空间的前 3 个语义特征维度，利用 RGB 的描绘方式给单词和体积元着色。将最能够符合 4 个语义特征维度的单词选出，然后利用 K-Means 聚类算法将这些单词归为 12 个单词目录。每个目录进行手工标记。这 12 个单词目录标记（图中字体较大的单词）和选择出的 458 个单词（图中字体较小的单词）绘制在图中的 4 组以语义特征维度为坐标轴的坐标系中。方差占比最高的轴即表示第一个维度，两侧的词汇目录分别为与感知、物质相关的词汇目录（如“tactile”、“locational”）和与人相关的词汇目录（如“social”、“emotional”、“violent”）。PC 为主成分。b，体积元模型中的单词权重被映射到选择出的语义特征维度中，然后同样使用 RGB 方式着色。展示在图中的是 3 号被试的大脑皮层的着色语义地图。语义信息在语义系统上的表示呈现出复杂的分布模式。c，3 个被试的大脑的语义特征主成分维度着色平展图。比较这些平展图，可以看出，被试个体间共有许多语义表示分布模式。图示中的缩写详见 Methods 一节。）

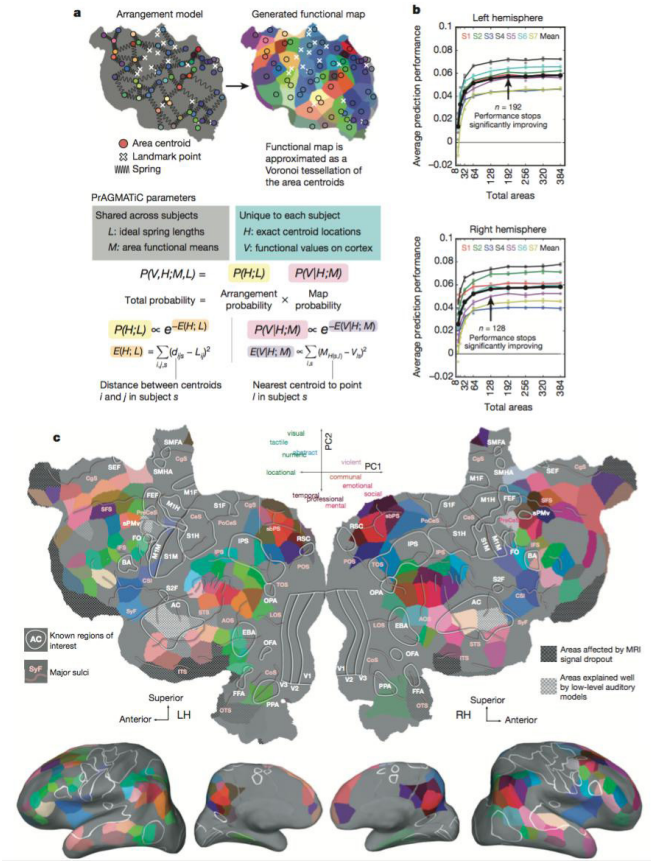

图 3 PrAGMATiC：大脑地图的生成模型（a~c，为了建立一个人类大脑皮层上的地图集来描述语义选择性功能区域的分布，我们设计了一种新的 PrAGMATiC 算法来获得覆盖大脑皮层区域的概率生成模型。a，PrAGMATiC 算法包括安排模型和发散模型两部分。安排模型和连接临近区域质心的物理弹簧系统相似。为了保证不同被试间绘制地图的相似性，19 个我们感兴趣的区域被单独分割标记，并用弹簧连接。发散模型将最近区域质心的语义功能值分配给皮层上的每一点，形成泰森多边形图（Voronoi tessellation）。弹簧的长度和区域的语义功能值均值在被试间保持一致，每名被试的区域具体位置是独立的。这些参数均使用最大似然估计法进行估计。b，使用留一法进行交叉验证来确定每个大脑半球区域的数目。首先使用 6 名被试的数据对 PrAGMATiC 算法模型进行估计，然后对第 7 名被试的血氧浓度变化进行预测。预测能力直到大脑左半球区域数为 192，大脑右半球区域数为 128 前随区域数量的增加有显著提高。c，用 7 名被试的数据估计得到的语义地图集。将语义特征模型预测能力低于低级语言特征（如语速、音素）模型预测能力的区域从语义地图集上移除。剩余的语义区域使用图 2 中相同的 RGB 着色法在一名被试的皮质表面绘出。受到 MR 信号衰减影响的区域用黑色阴影标出，同时低级语言特征模型预测效果好的区域用白色阴影线标出。绘制的语义地图集说明语义系统中功能区域的组织在不同被试间较为一致。）

（选自 Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. Nature, 532(7600), 453-458.）

李若愚

清华大学计算机科学与技术系本科生。

吴佳炜

清华大学医学院本科生。

刘知远

博士，清华大学计算机系助理研究员。在 AAAI、IJCAI、ACL 等人工智能领域的著名国际期刊和会议发表相关论文 20 余篇，Google Scholar 统计引用超过 850 次。曾获清华大学优秀博士学位论文、中国人工智能学会优秀博士学位论文、清华大学优秀博士后等称号。主要研究方向为表示学习、知识图谱和社会计算。