

自然语言的语义表征与解码

宗成庆

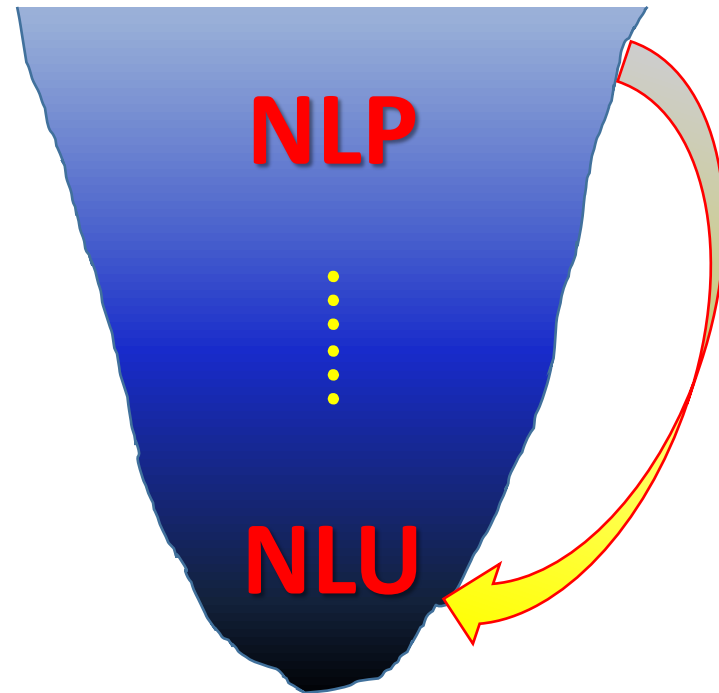
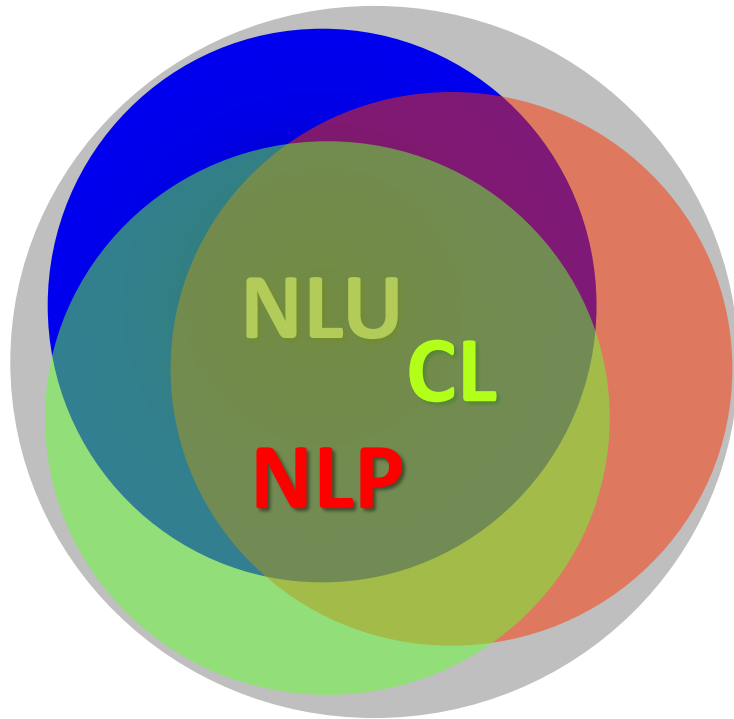
中国科学院自动化研究所

cqzong@nlpr.ia.ac.cn

提 纲

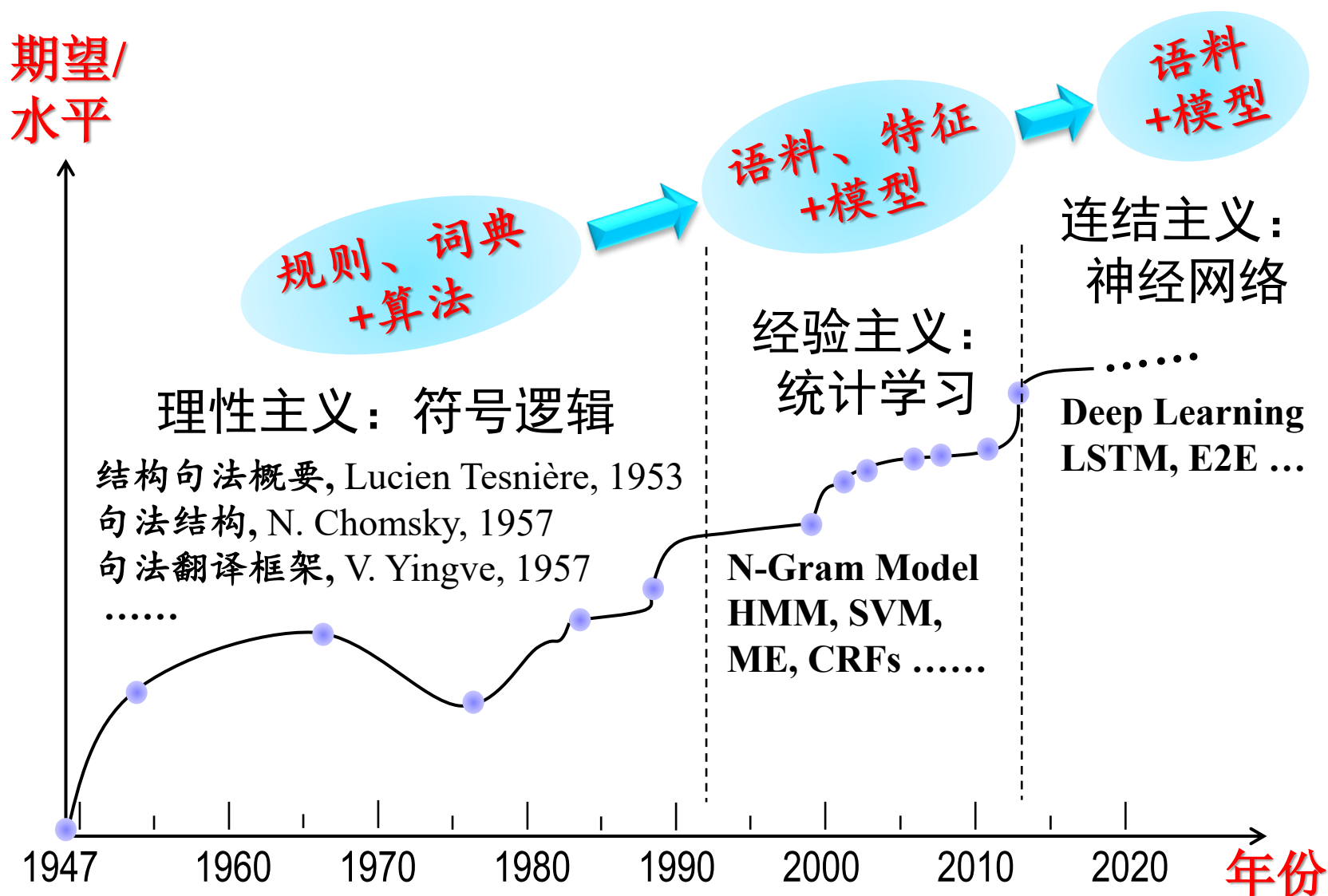
- 
1. 引言
 2. 语义表征
 3. 语义解码
 4. 结束语

1. 引言

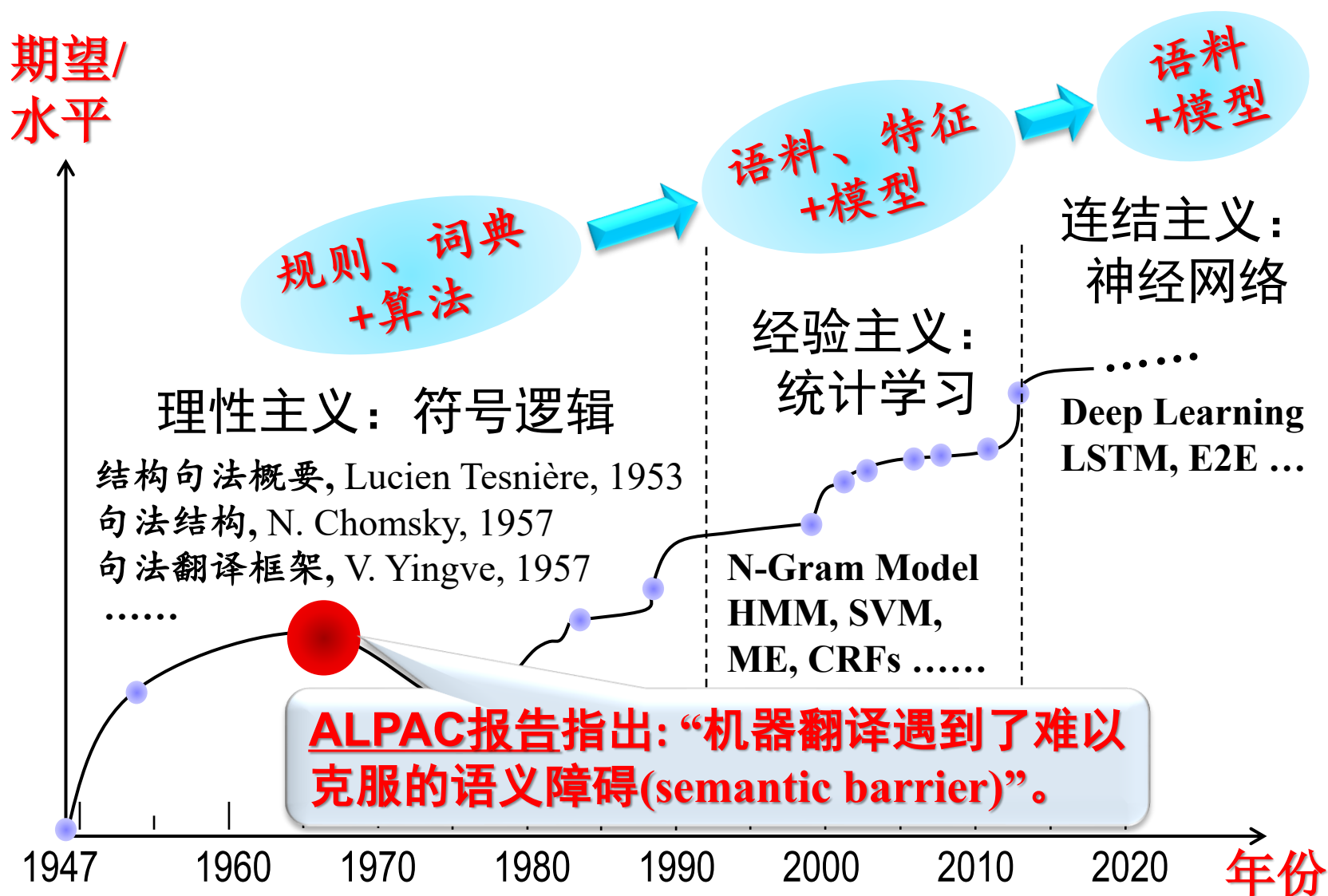


- NLU: natural language understanding (1950s)
- CL: computational linguistics (1960s)
- NLP: natural language processing (1970~80s)
- HLT: human language technology (1980s)

1. 引言

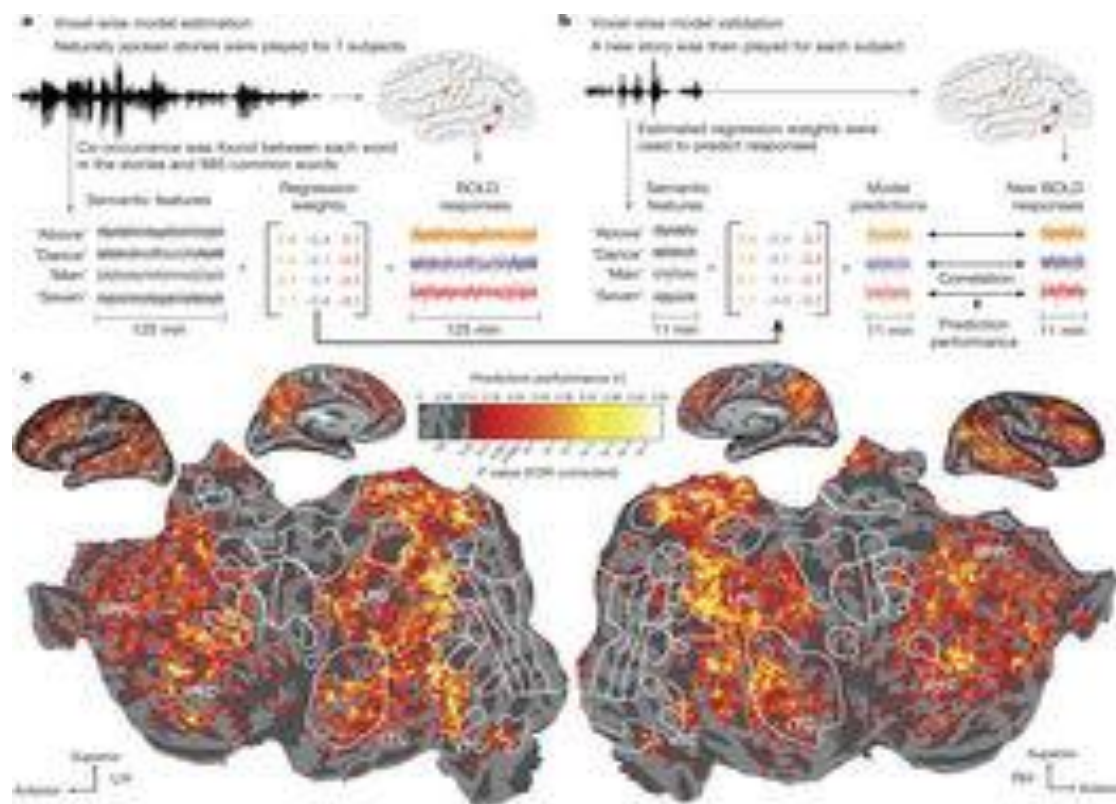


1. 引言



1. 引言

◆脑科学研究对我们的启发

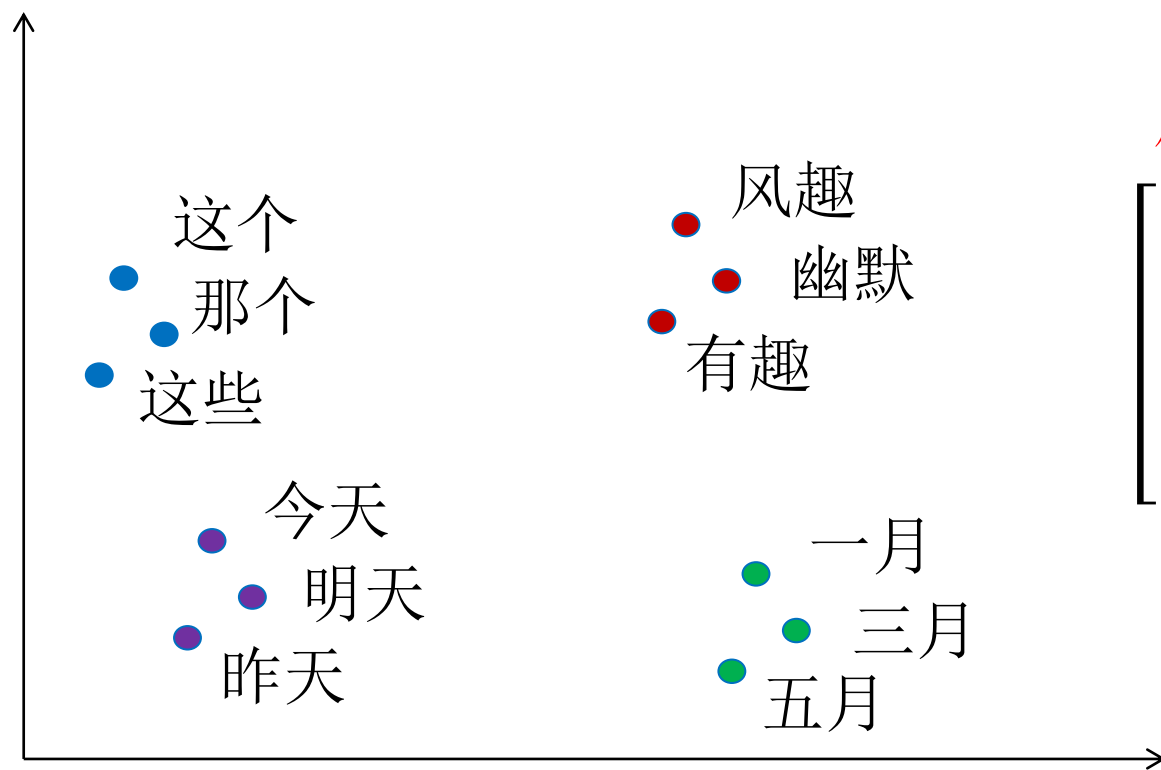


- 985个英语词汇在大脑中的分布图
- 语义加工在大脑左右半球基本上是对称的，打破了传统研究中认为语言加工左半球偏侧化的情况
- 不同个体大脑的整个语义网络“看起来特别相似”

Alezander G. Huth, Wendly A. Heer et al., Natural speech reveals the semantic maps that tile human cerebral cortex, Nature, April 28, 2016, 532: 453~ 458

1. 引言

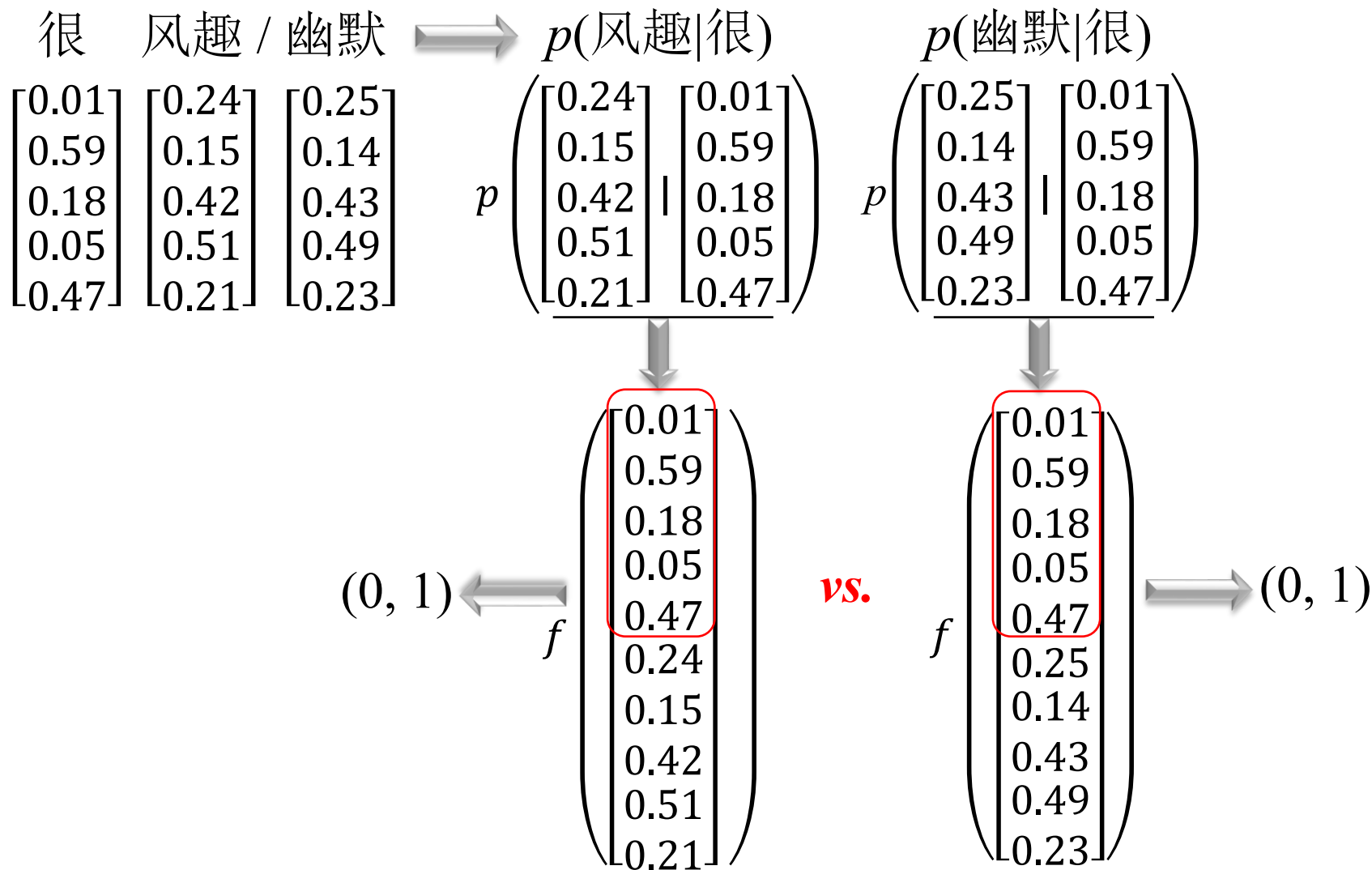
◆ 分布式表示



风趣	幽默	明天
0.24	0.25	0.56
0.15	0.12	0.78
0.42	0.39	0.42
0.51	0.46	0.66
0.21	0.26	0.86

低维、稠密的连续实数空间

1. 引言

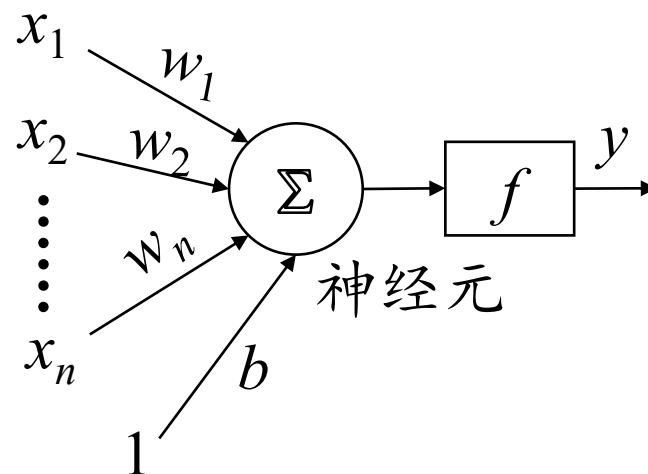


1. 引言

基于 FNN 的语言模型

$$p(\text{风趣}|\text{很}) = f \left(\begin{bmatrix} 0.01 \\ 0.59 \\ 0.18 \\ 0.05 \\ 0.47 \\ 0.24 \\ 0.15 \\ 0.42 \\ 0.51 \\ 0.21 \end{bmatrix} \right) \Rightarrow ? (0, 1)$$

$$p(\text{幽默}|\text{很}) = f \left(\begin{bmatrix} 0.01 \\ 0.59 \\ 0.18 \\ 0.05 \\ 0.47 \\ 0.25 \\ 0.14 \\ 0.43 \\ 0.49 \\ 0.23 \end{bmatrix} \right) \Rightarrow ? (0, 1)$$



词向量 ?
 $W, b = ?$

1. 引言

◆ 问题

- 如何获得准确的向量表示？
 - 如何解释基于不同模态信息获得的语义表示？
 - 不同的表示方法在人脑解码中的性能表现如何？
 - 如何从神经影像数据中重构相关的语言信息？
 - 不同类型的分布式表示模型与其对应文本的人脑神经表征的关系是什么？
 - 分布式表示模型编码的何种信息与人脑语言理解语言相关联？
 - 分布式表示模型能帮助我们揭示不同脑区编码的语义差别吗？
-

提 纲

1. 引言

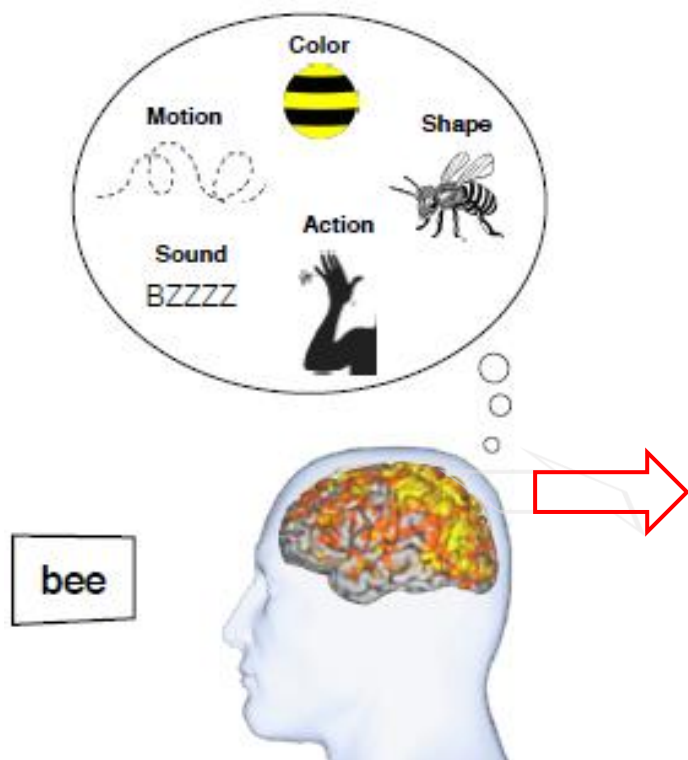
 2. 语义表征

3. 语义解码

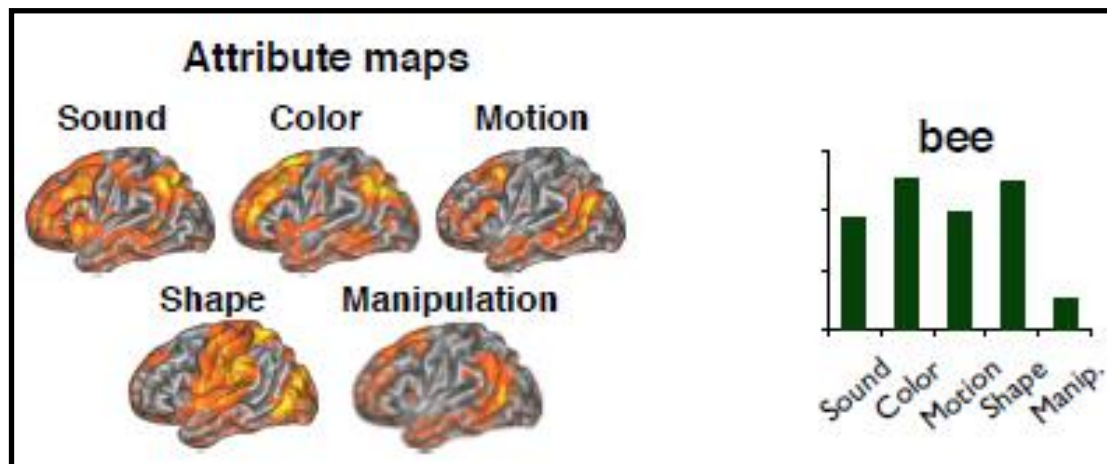
4. 结束语

2. 语义表征

◆多模态词汇表示的可解释性

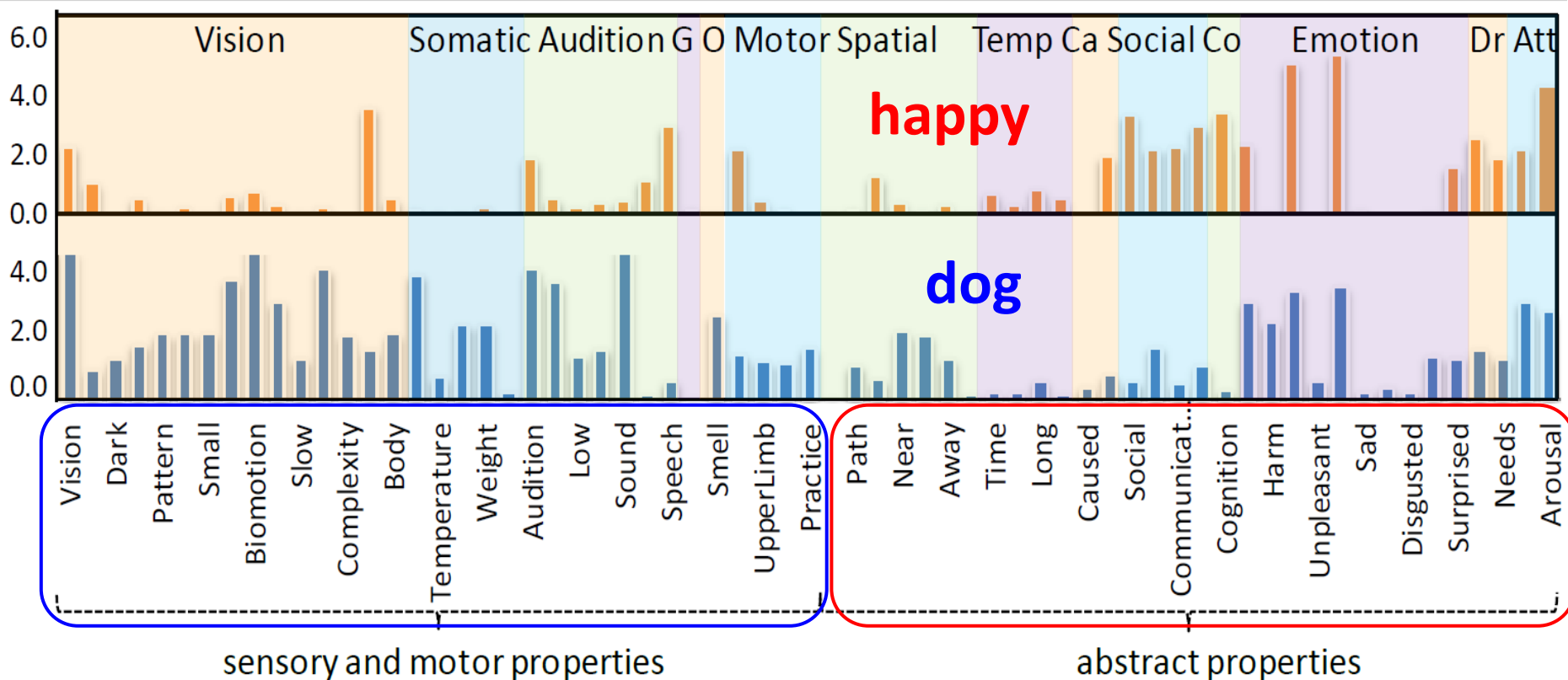


Brain semantic representation



2. 语义表征

基于人脑的成分语义表征



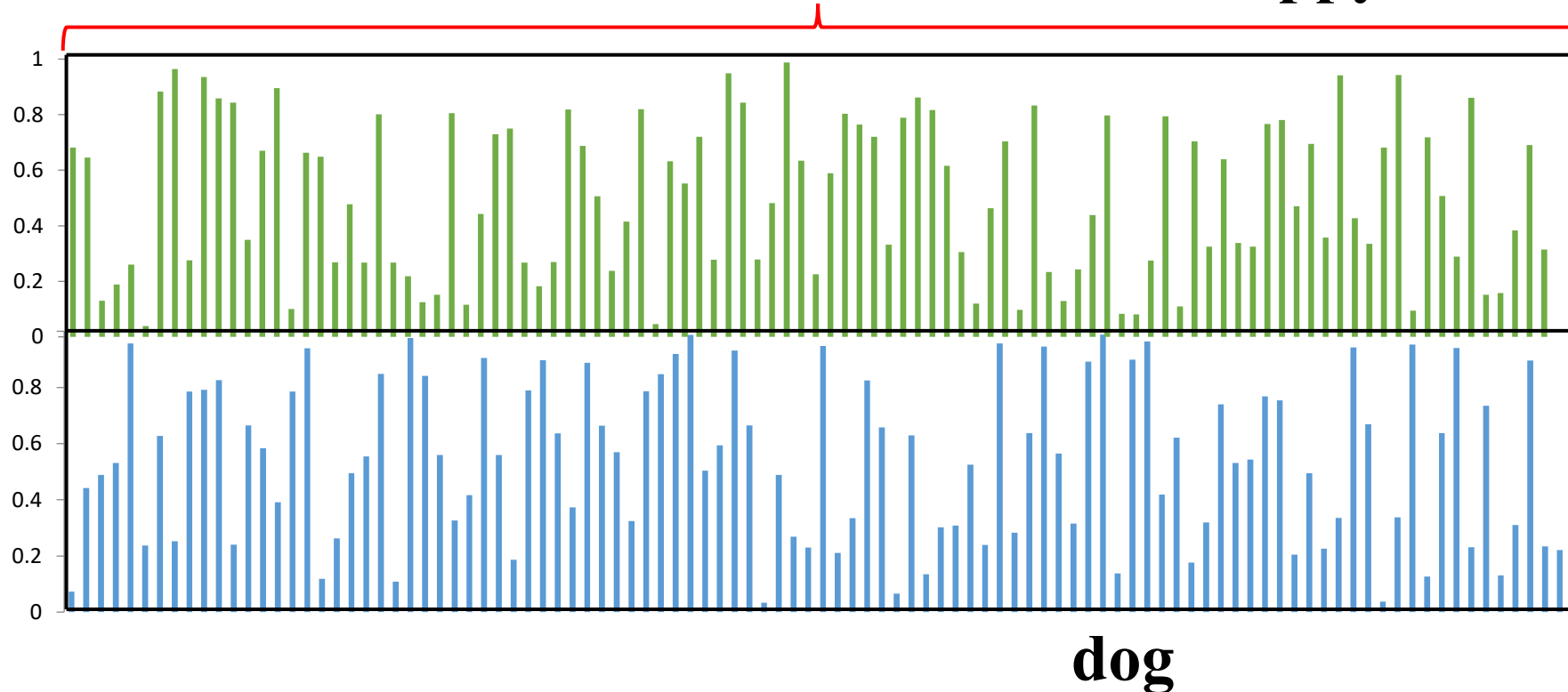
J. R. Binder *et al.* 2016. Toward a brain-based componential semantic representation.
Cognitive Neuropsychology, 33(3-4):130–174

2. 语义表征

计算模型得到的语义表征

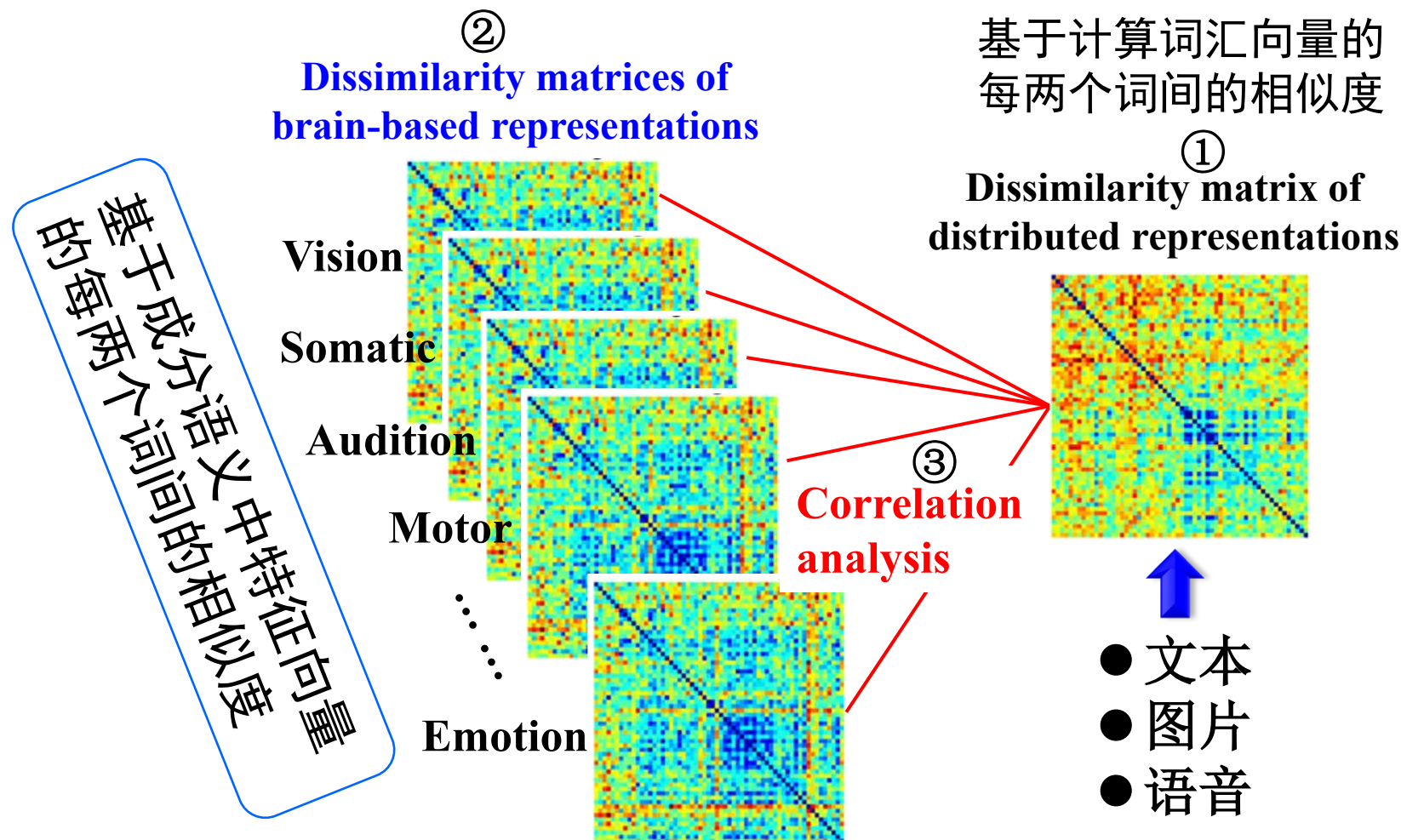
维度的含义未知

happy



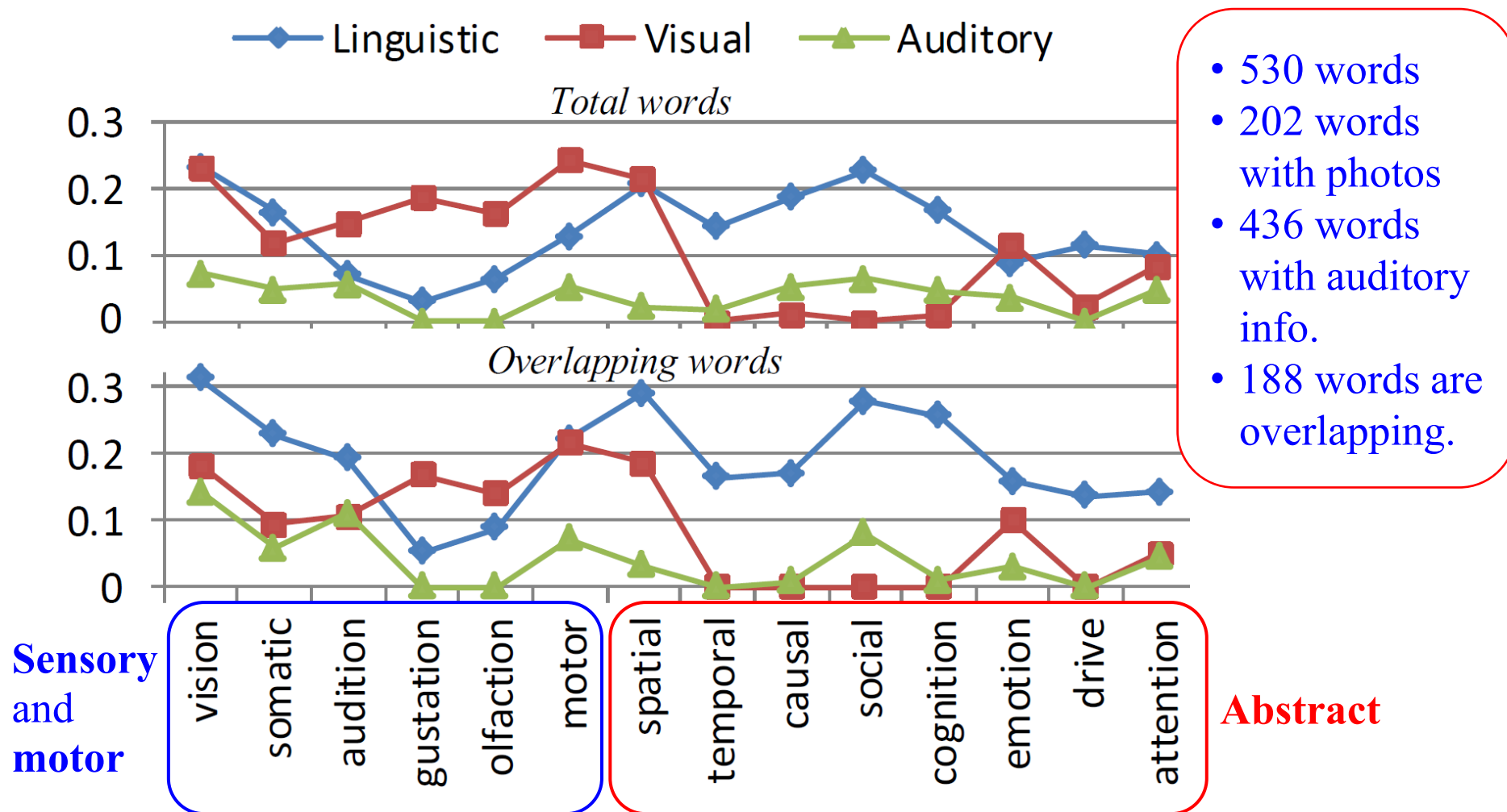
2. 语义表征

用表征相似性分析方法和大脑的成分语义表征解释向量表示



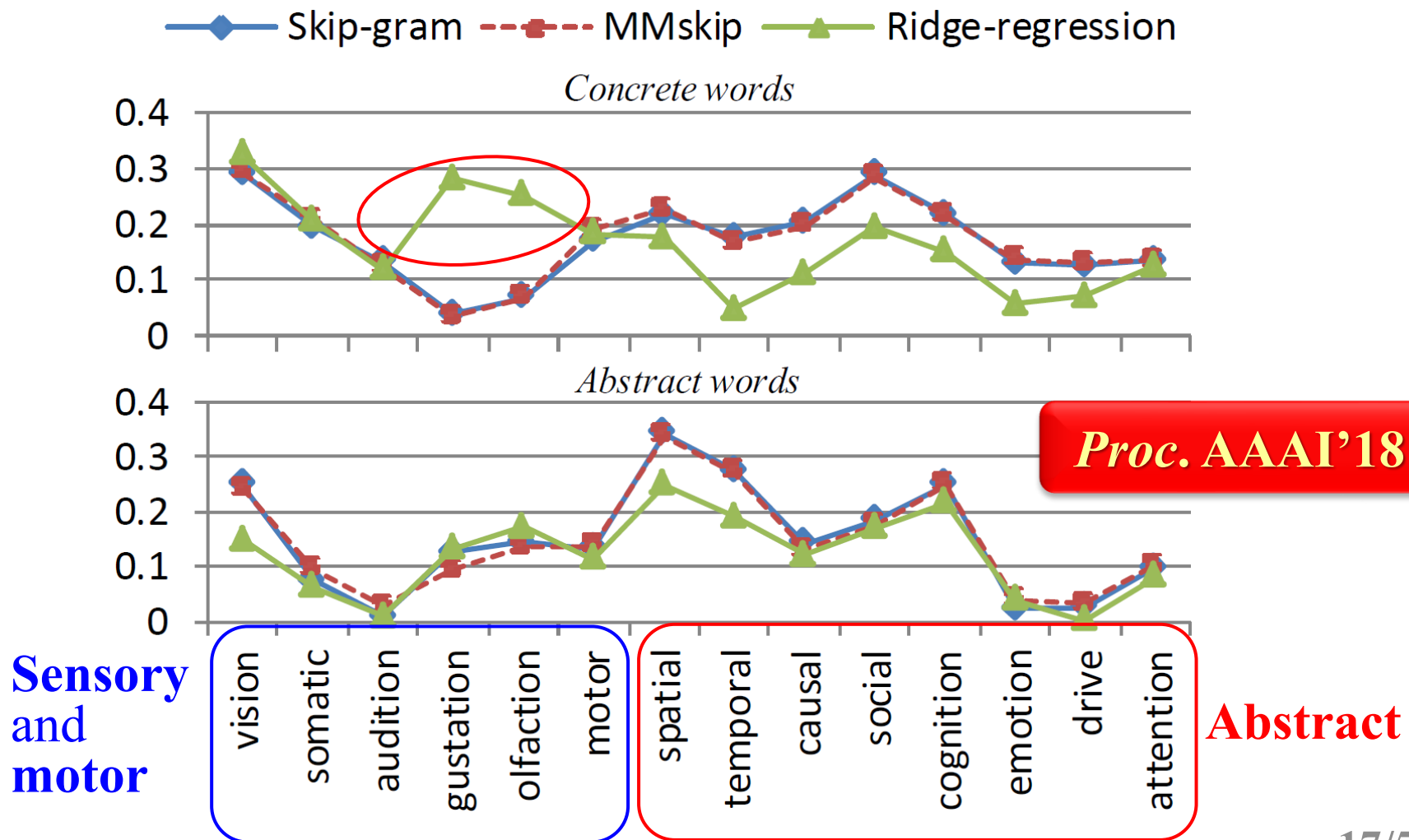
2. 语义表征

■ **发现1:** 不同模态得到的词汇表示编码了不同类型的属性



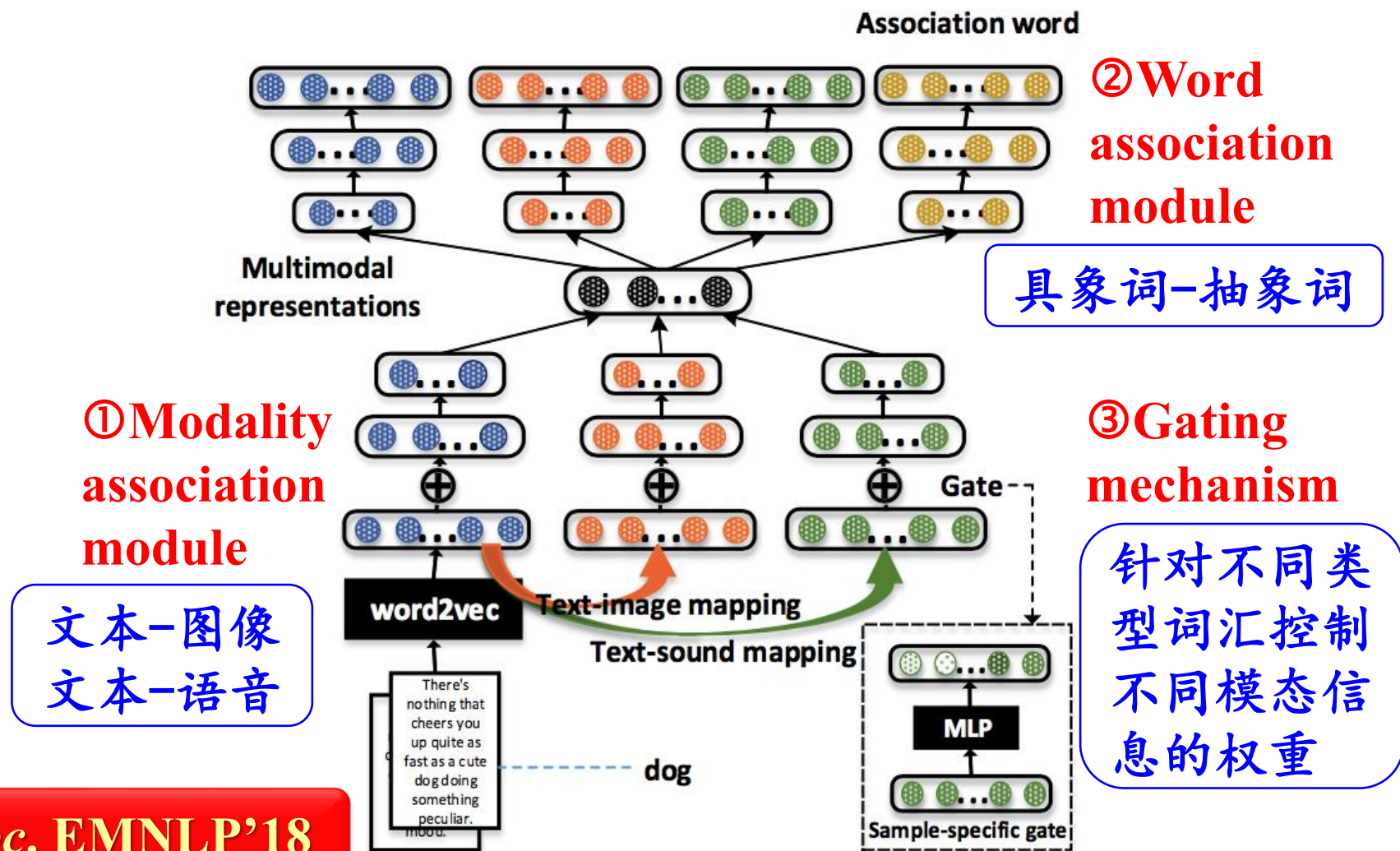
2. 语义表征

■ **发现2:** 词汇语义表示模型在实体词和抽象词上的表现不同



2. 语义表征

◆ 多路通道融合的编码方法

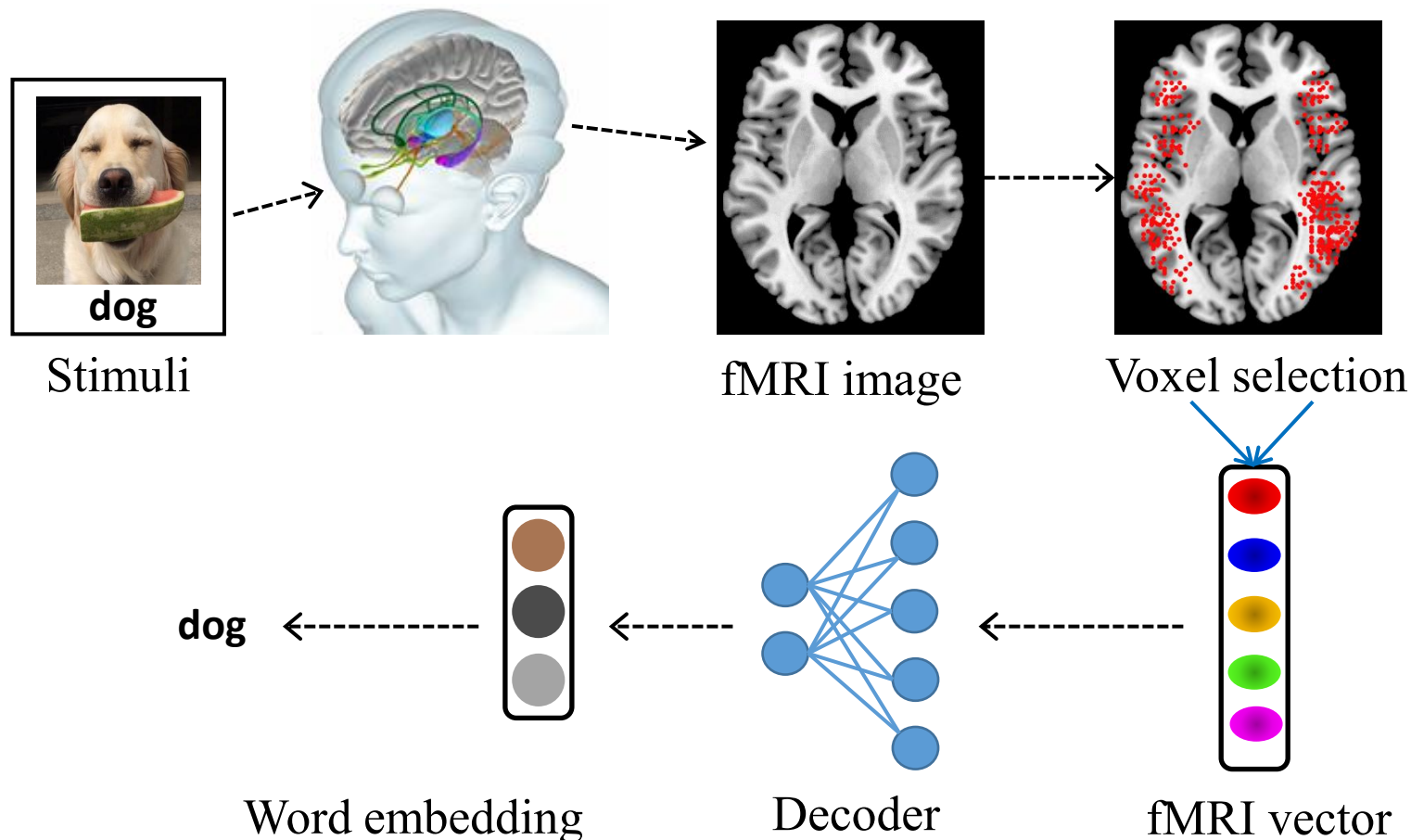


提 纲

1. 引言
2. 语义表征
-  3. 语义解码
4. 结束语

3. 语义解码

(1) 研究对比了不同类型分布式句子表示在人脑语言解码(brain decoding)上的性能表现



3. 语义解码

● 已有的研究

- 成功地从脑功能成像(fMRI)中解码出了受试者观察到的图像和视频信息[Haynes et al., 2006, Kamitani et al., 2005; Shirer et al., 2012]
- 在量化解析大脑处理语言（如词汇、短语和句子）语义时的活动方面，还没有令人非常满意的结果
- 已有的句子解码研究中使用的分布式表示方法是简单的向量平均法，即将句子中每个词对应的词向量加和并求平均值
 - 句子中不同的词对于整句的语义贡献程度不同
 - 句子的语义与其结构（如词序等）密切相关

3. 语义解码

● 我们的动机

①研究对比两类应用最广泛的解码方法

- **相似度解码**: 建立大脑活动(fMRI)与文本表示的统一相似性度量
- **回归解码**: 训练回归模型拟合大脑活动(fMRI)与文本表示之间的关系

3. 语义解码

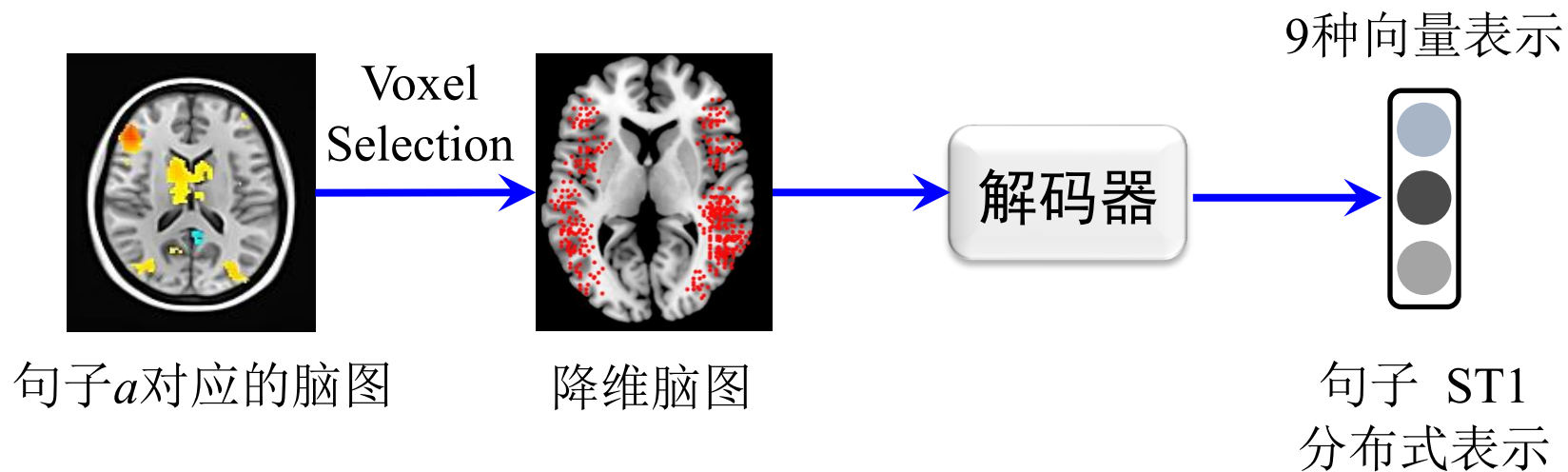
②研究对比不同的句子表示方法（9种）

模型分类		名称	结构基础
非结构化模型	朴素池化	① 平均池化(AVG)	平均池化
		② 最大池化(MAX)	最大池化
		③ 联合池化(AVMA)	拼接平均池化和最大池化
	复杂池化	④ SIF	加权平均
		⑤ FastSent	平均池化 + 语言模型
结构化模型	无监督模型	⑥ Skip-thought	LSTM
		⑦ Quick-thought	LSTM
	监督模型	⑧ GenSen	BiLSTM
		⑨ InferSent	LSTM

3. 语义解码

● 实验流程

➤ 解码器训练



3. 语义解码

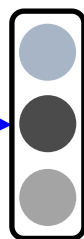
➤ 解码器测试

句子b1对应的脑图

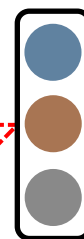


解码器

解码所得的
语义向量S1

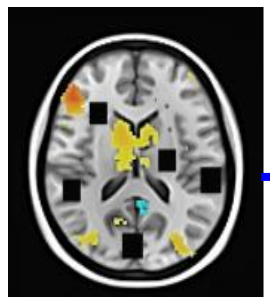


句子b1对应
的句子向量E1



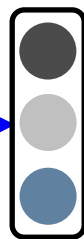
计算相关系数

句子b2对应的脑图



解码器

解码所得的
语义向量S2



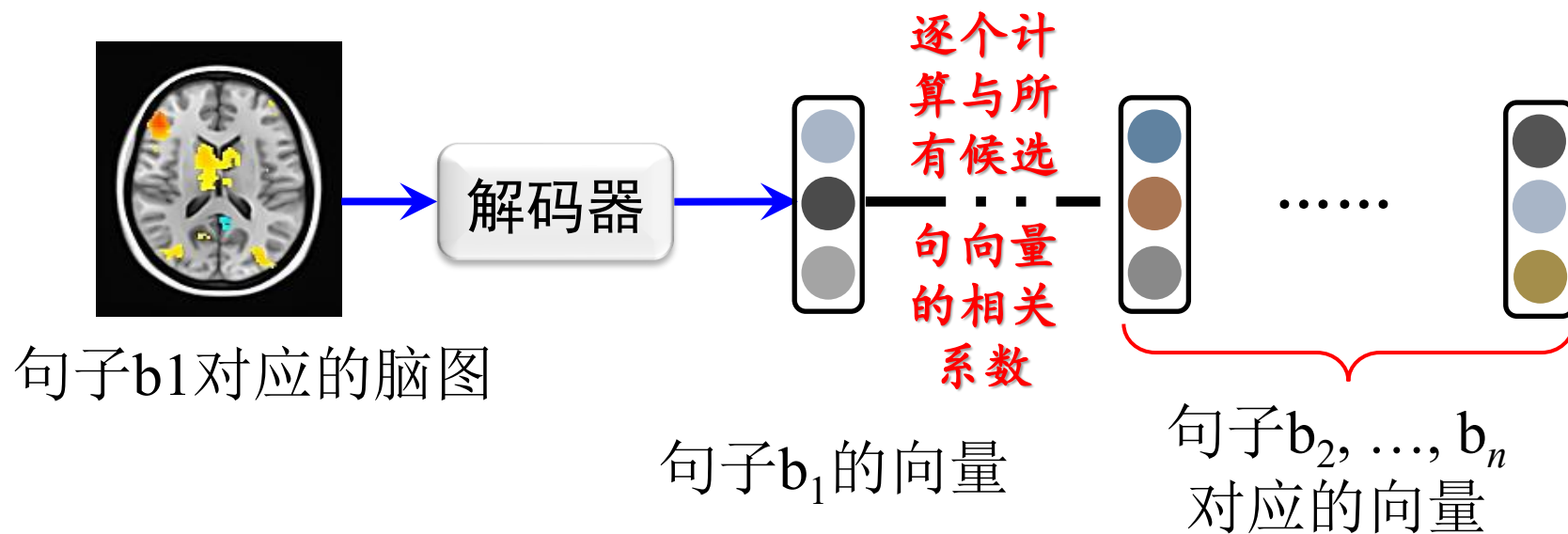
句子b2对应
的句子向量E2



如果
 $\text{corr}(S1, E1) + \text{corr}(S2, E2) > \text{corr}(S1, E2) + \text{corr}(S2, E1)$,
那么, 匹配成功。

3. 语义解码

➤ 排序任务



3. 语义解码

➤ 数据样例

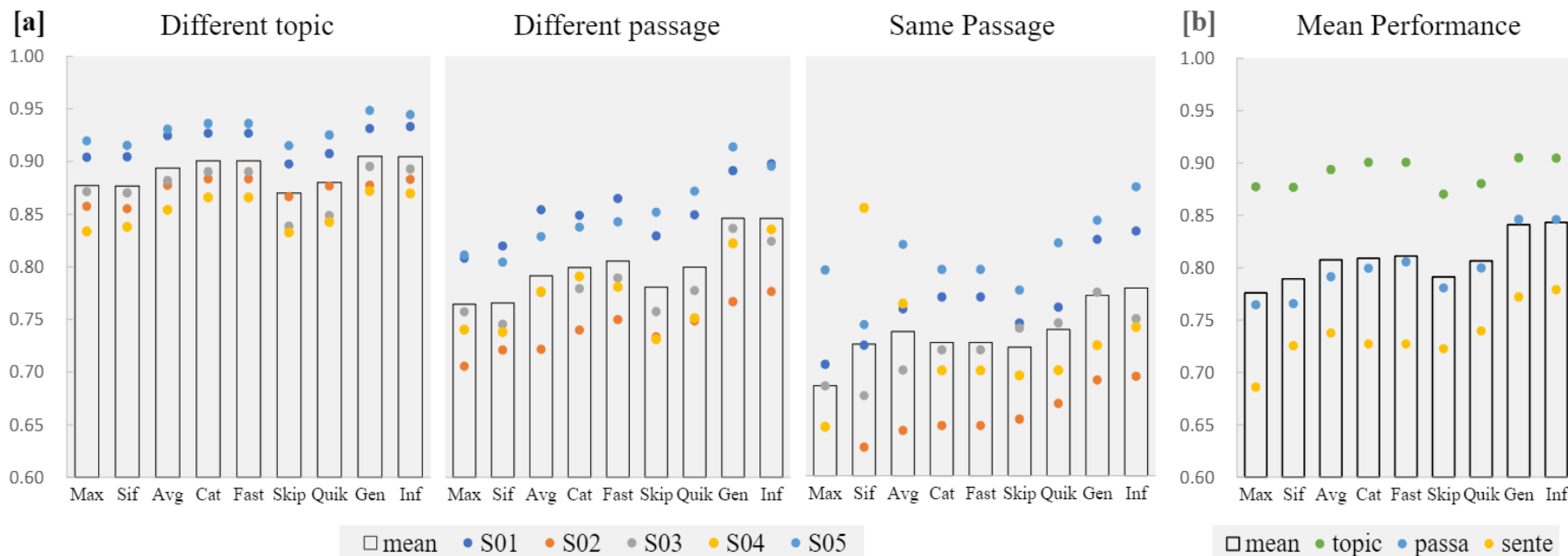
呈现给受试者的句子(sentence stimulus)按照**主题(topic)**—**段落(passage)**—**句子(sentence)**的层次组织, 包含:

- 48个主题, 每个主题3~5个段落, 每个段落3~4个句子, 每个句子长10-20个词; 共637个句子。以乐器(Musical Instruments)主题为例:

Topic	Passages	Sentences
Musical Instruments	Piano	1. The piano is a popular musical instrument...
		2. Pressing a piano key causes a felt-tipped hammer...
		3. The piano has an enormous note range.
	Accordion	1. A clarinet is a woodwind musical instrument...
		2. It is a long black tube with a flare at the bottom
		3. The player chooses notes by pressing keys and holes.
	Clarinet	1. An accordion is a portable musical instrument.
		

3. 语义解码

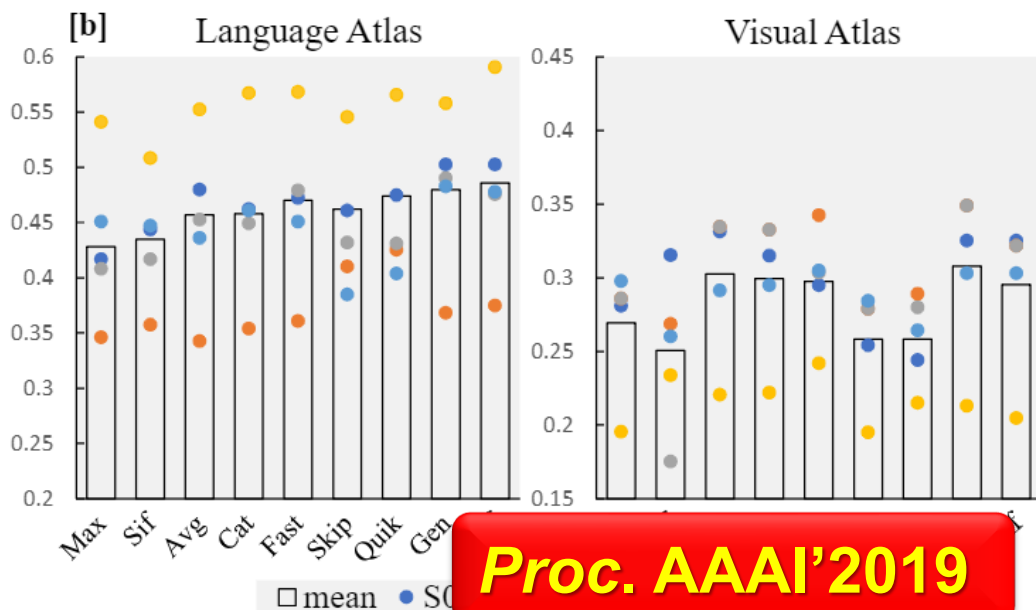
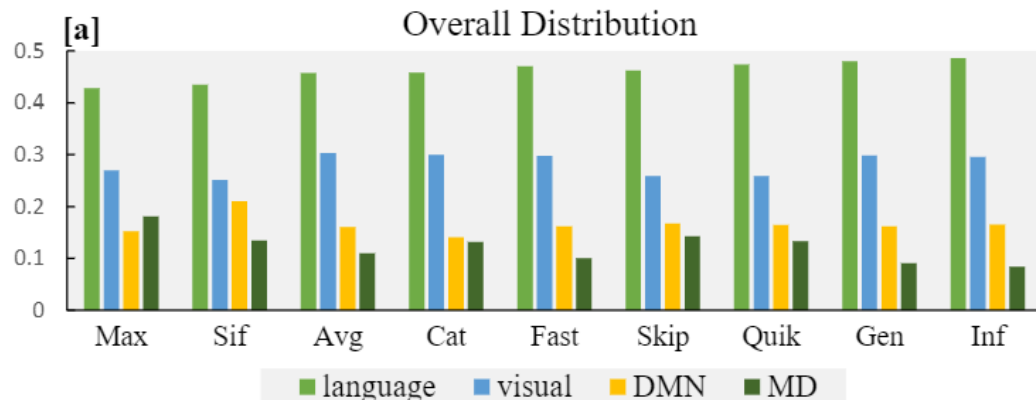
● 实验结果1：相似度方法用于神经解码任务



- 非结构化模型中，最大池化表现最差；将最大池化和平均池化结合，可取得超过单一模型的表现；
- 有监督结构化模型的表现均显著优于其他模型；无监督结构化模型在跨主题和段落的解码任务上表现与非结构化模型相差无几，而在相同段落解码任务上优势较为明显。

3. 语义解码

● 实验结果2: 与句向量相关度较高的Voxel在功能脑区上的分布



- 与受试句向量相关的像素大部分分布在人脑的语言区；
- 在解码任务上表现较好的模型与其相关的像素在人脑语言区的分布更密集；
- 尽管句子的阅读同样涉及视觉的运用，但相关像素在人脑视觉区的分布仅为30%左右。

3. 语义解码

(2) 大脑语言神经解码：如何从神经影像数据中重构相关的语言信息？

- 基于词汇语义表示和线性回归方法方法，建立不同类型的词汇与大脑激活图(fMRI)的对应关系，重构大脑激活图对应的词汇，实现词汇级别的思维解码。
- 词汇级别的大脑语言解码的最佳方式是什么？即在何种实验范式下，利用哪一种词汇表示模型、针对什么词汇类型可以得到最佳的解码效果？

3. 语义解码

- 三种刺激 (Stimulus)

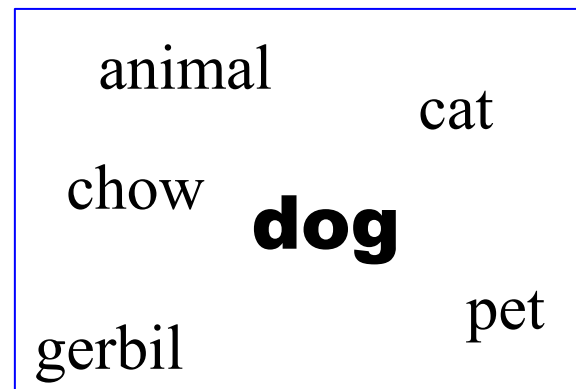


dog

A dog is a great companion.

Picture + word

A sentence



Word cloud

- (1) 三种类型的词：名词，动词，形容词
- (2) 三种实验范式：图+词，句子，词图
- (3) 采用分布式词表示

3. 语义解码

具体地，

- 从 29,805 个词汇中选取出了180个词汇，其中，131 个名词，22个动词，21个形容词，6个副词（最终没有参与实验）
- 这180个词汇被分成两组，各90个，经过两轮实验，对每一位被试分别采用不同的范式进行实验

(a)图片+词汇：每个词汇对应6幅图片，共 $6*180$ 个图片+词汇对；

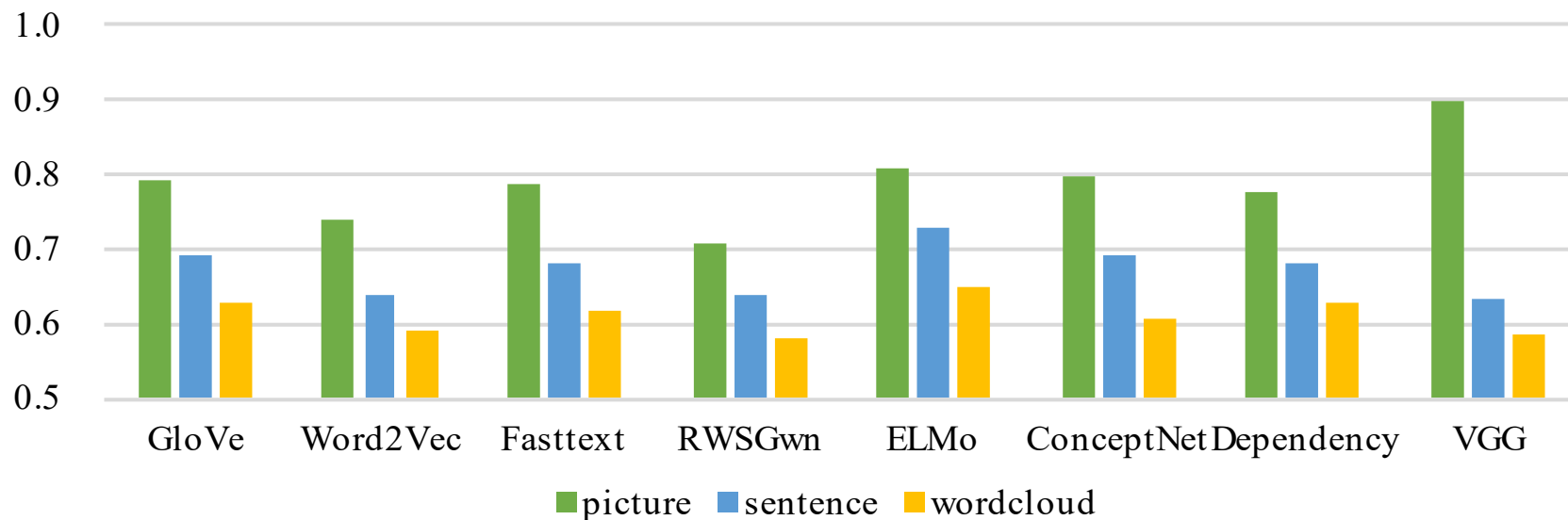
(b)句子：每个词对应6个句子，其中句子是人工撰写的描述该概念的，长度为4~11个词，共 $6*180$ 个句子+词汇对；

(c)词图：每个词汇对应6个词图，每个词图包括1个目标词和5个相关词。

Each stimulus was presented for 3s followed by a 2s fixation period. The details of the experimental setup, materials and presentation scripts are available at <https://osf.io/crwz7/wiki/home/>
The acquisition parameters and processing pipelines are described in detail in the following paper, F. Pereira, B. Lou, B. Pritchett et al. Toward a universal decoder of linguistic meaning from brain activation, *Nature Communications*, 2018, 9(1):963

3. 语义解码

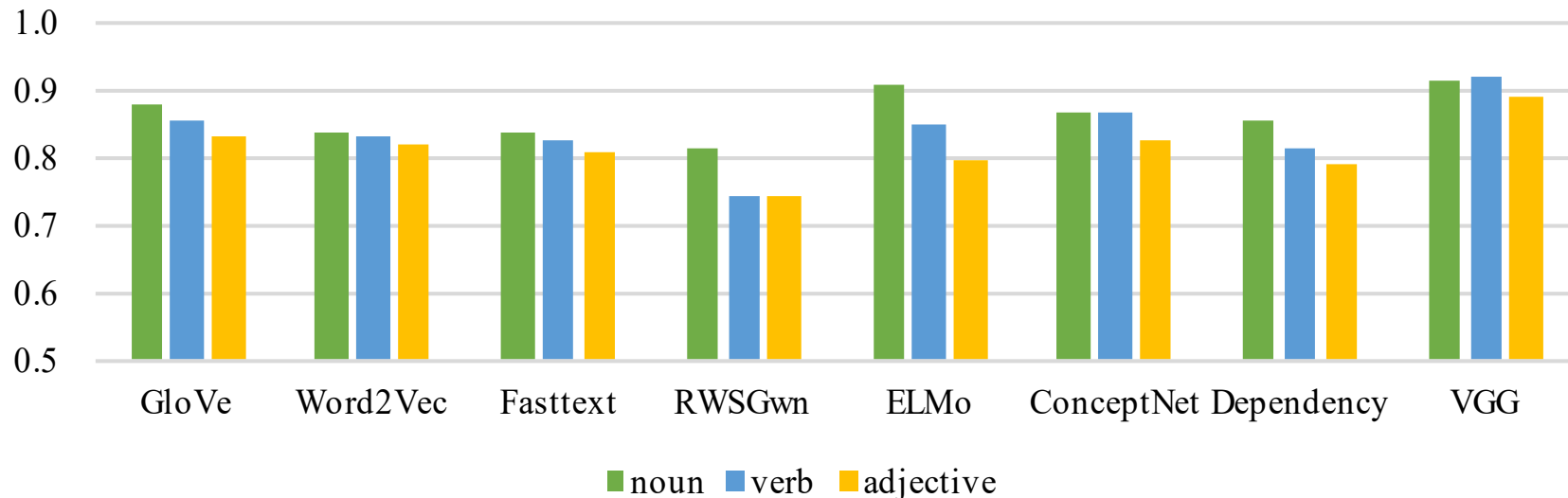
- 不同实验范式下解码的准确率（解码器采用简单的岭回归方法）



- “图片+词汇”范式最稳定，且得到了最好的平均结果
- “图片+词汇”与VGG文本表示模型搭配使用可获得最好的效果
- 句子和词图两种实验范式与ELMo搭配效果最好

3. 语义解码

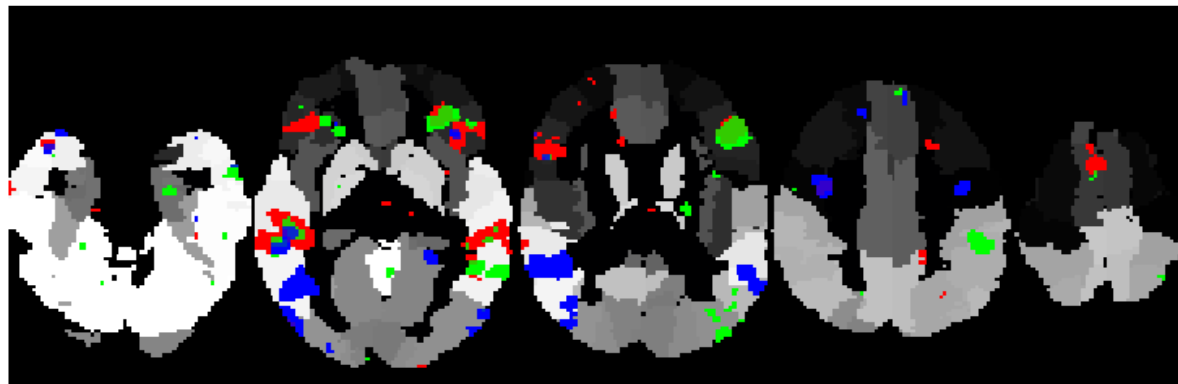
● 使用不同文本表示方法的解码准确率（8种文本表示模型）



- 对于不同类型的词汇，VGG模型得到了最好的平均结果，其次是ELMo模型；
- 名词与ELMo模型搭配，动词和形容词与VGG模型搭配时效果最好；
- 不同类型的词应该用不同的词向量表示方法。

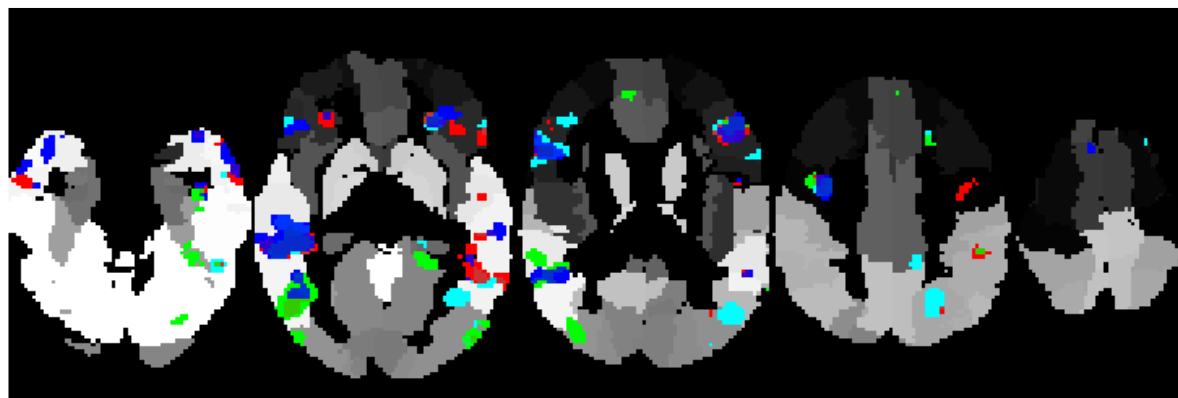
3. 语义解码

- 不同模态的刺激激活不同的脑区 (5张图表示5个不同的侧面)
不同的文本表示模型与不同的脑区最相关



5000 voxels, GloVe

- 红色表示“图+词”
- 绿色表示句子
- 蓝色表示词图



5000 voxels, three reps

- 红色代表GloVe
- 绿色代表VGG
- 蓝色代表ELMo
- 黄色代表Dependency

Information Sciences, 507 (2020) 256–272

<https://doi.org/10.1016/j.ins.2019.08.043>

3. 语义解码

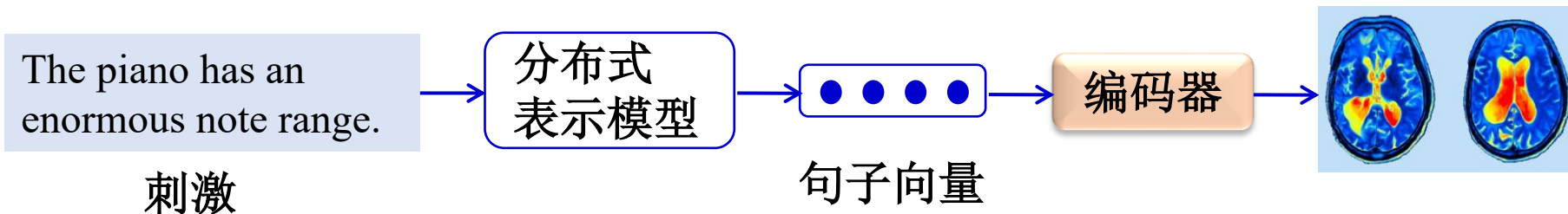
(3) 分布式表示与人脑之间的关系

- 不同类型的分布式表示模型与其对应文本的人脑神经表征的关系是什么？
- 分布式表示模型编码的何种信息与人脑语言理解语言相关联？
- 分布式表示模型可以帮助我们揭示不同脑区编码的语义差别？

3. 语义解码

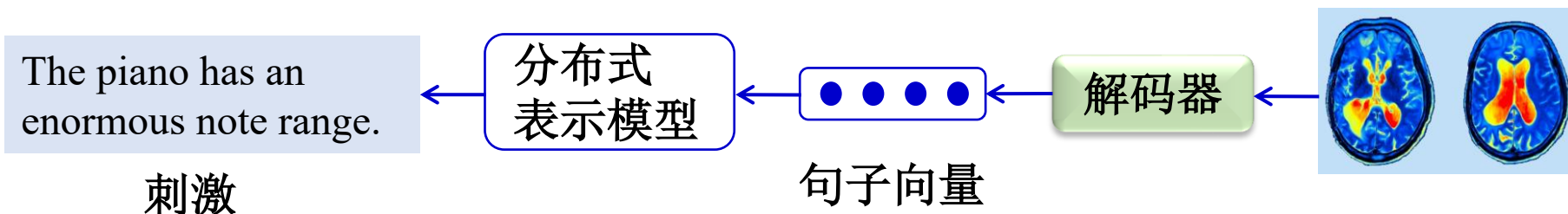
❖ 神经编码

目标：通过语言刺激的变化来预测对应各脑区、体素点神经活动，进而帮助解析人脑的语言表征机制；



❖ 神经解码

目标：从脑神经活动中重建出人所感知到的刺激，是脑机接口的计算基础之一。



3. 语义解码

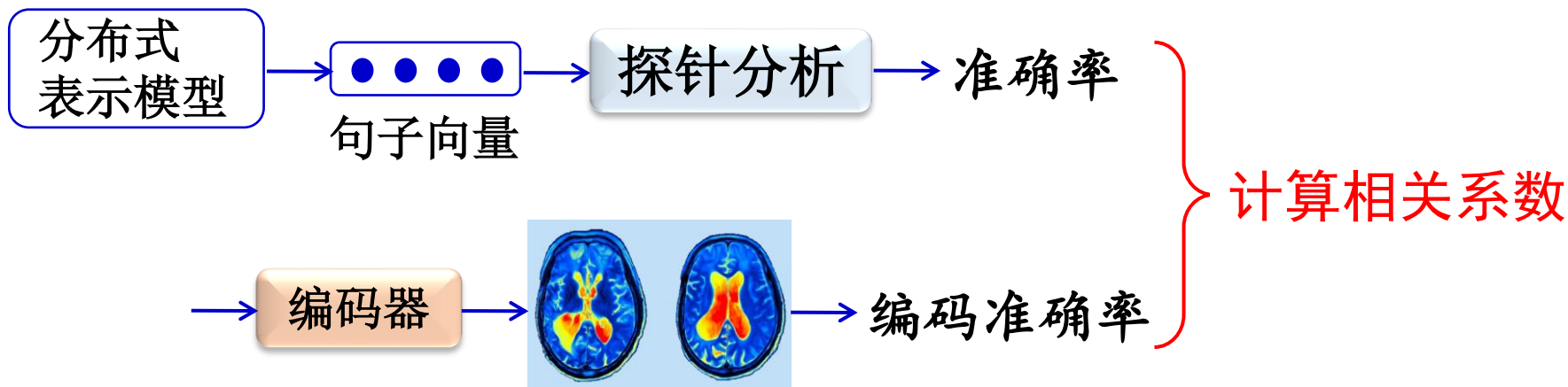
● 分布式文本表示

模型分类		名称	结构基础
非结构化模型	朴素池化	① 平均池化(AVG)	平均池化
		② 最大池化(MAX)	最大池化
		③ 联合池化(AVMA)	拼接平均池化和最大池化
	复杂池化	④ SIF	加权平均池化
结构化模型	无监督模型	⑤ Skip-thought	LSTM
		⑥ FairSeq	CNN
		⑦ ELMo	CNN-BiLSTM
		⑧ BERT	Transformer
		⑨ GPT2	
		⑩ RoBerTa	
	监督模型	⑪ GenSen	BiLSTM
		⑫ InferSent	CNN-BiLSTM

3. 语义解码

- 基于探针任务和消融测试的编解码分析算法

- 应用探针任务

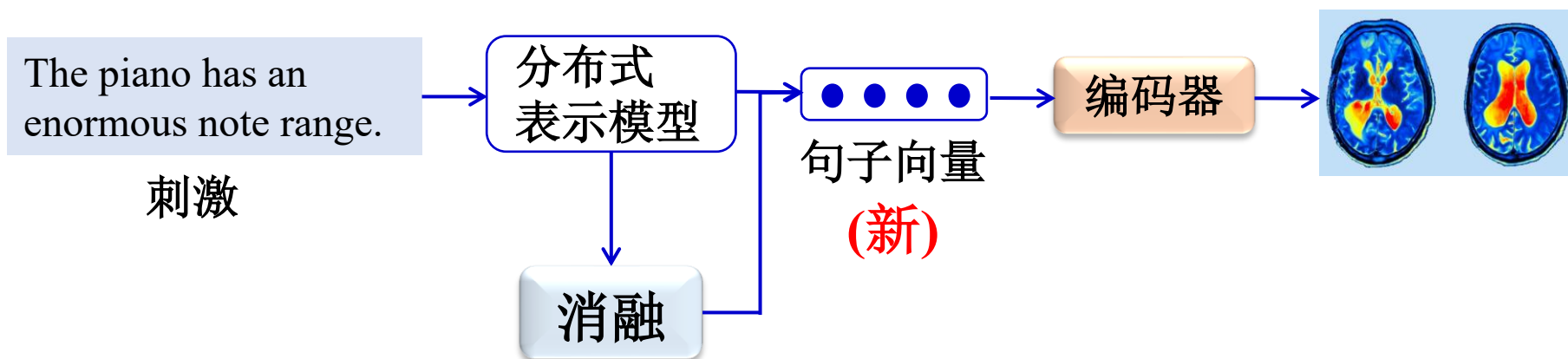


- 探针任务可针对语义、语法、形态学等句子特征设计；
- 通过探针任务，能够测试预分布式表示编码了何种特征，从而导致了在神经编解码中有对应的表现。

3. 语义解码

➤ 应用消融测试

消融测试用于削弱模型编码某些特征的能力，以此验证探针任务的发现。



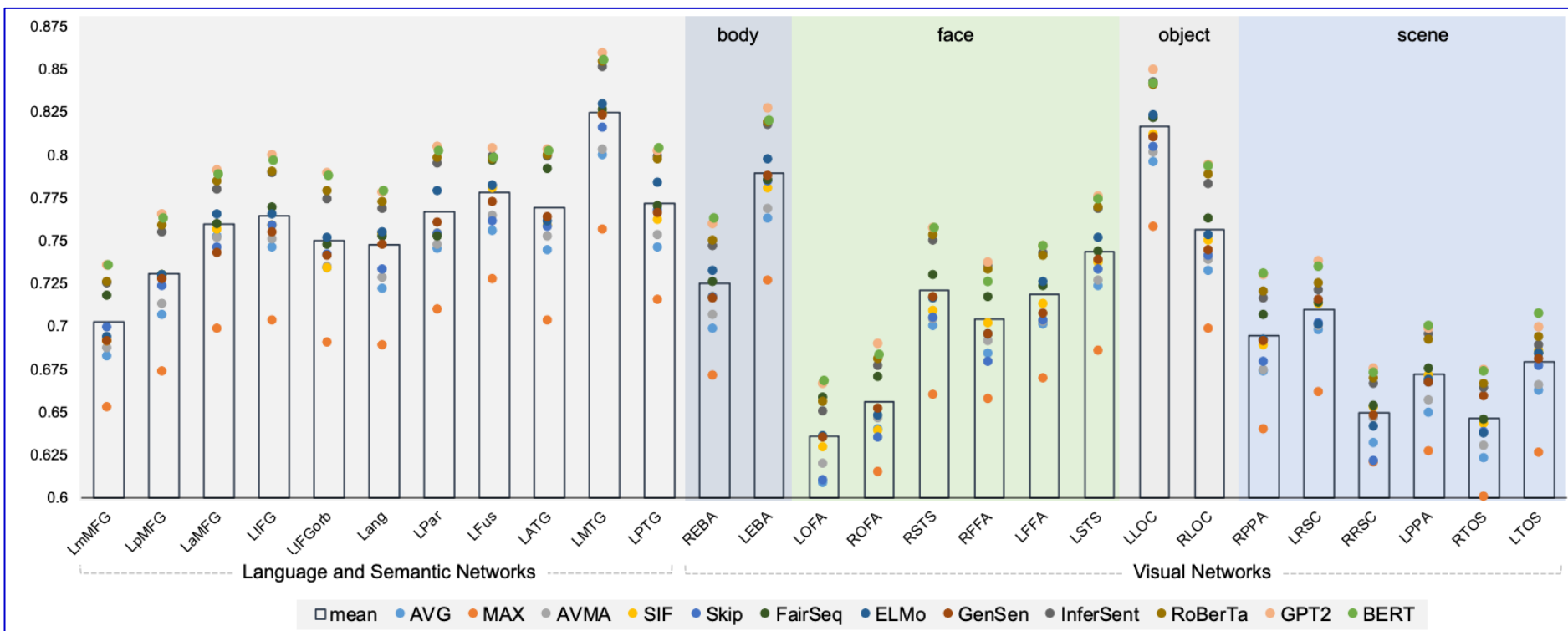
3. 语义解码

❖ 句子刺激结构及样例

Topic	Passages	Sentences
Musical Instruments	Piano	1. The piano is a popular musical instrument...
		2. Pressing a piano key causes a felt-tipped hammer...
		3. The piano has an enormous note range.
	Accordion	1. A clarinet is a woodwind musical instrument...
		2. It is a long black tube with a flare at the bottom
		3. The player chooses notes by pressing keys and holes.
	Clarinet	1. An accordion is a portable musical instrument.
		2. One keyboard is used for individual notes and the other for chords.
		3. Accordions produce sound with bellow that blow air through reeds
		

3. 语义解码

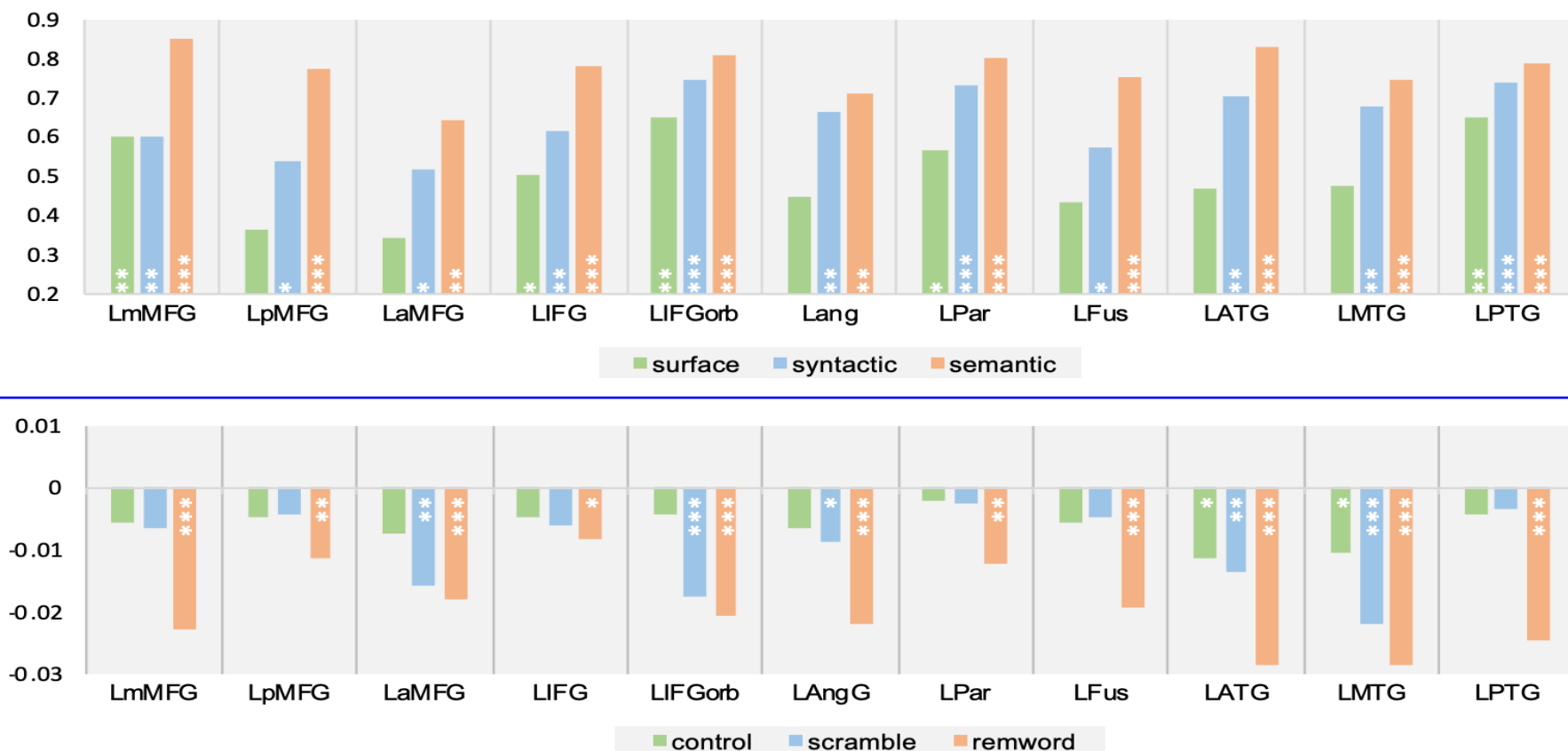
❖ 分布式表示模型在脑语言及视觉网络上的编码准确率



- 候选模型在大脑左颞叶中回(LMTG)和左枕叶复合体(LLOC)上编码表现突出；
- 基于Transformer的模型在语言网络上的编码准确率显著优于其他候选模型；
- 应用最广泛的平均方法，其编码表现仅排在所有候选模型的中游。

3. 语义解码

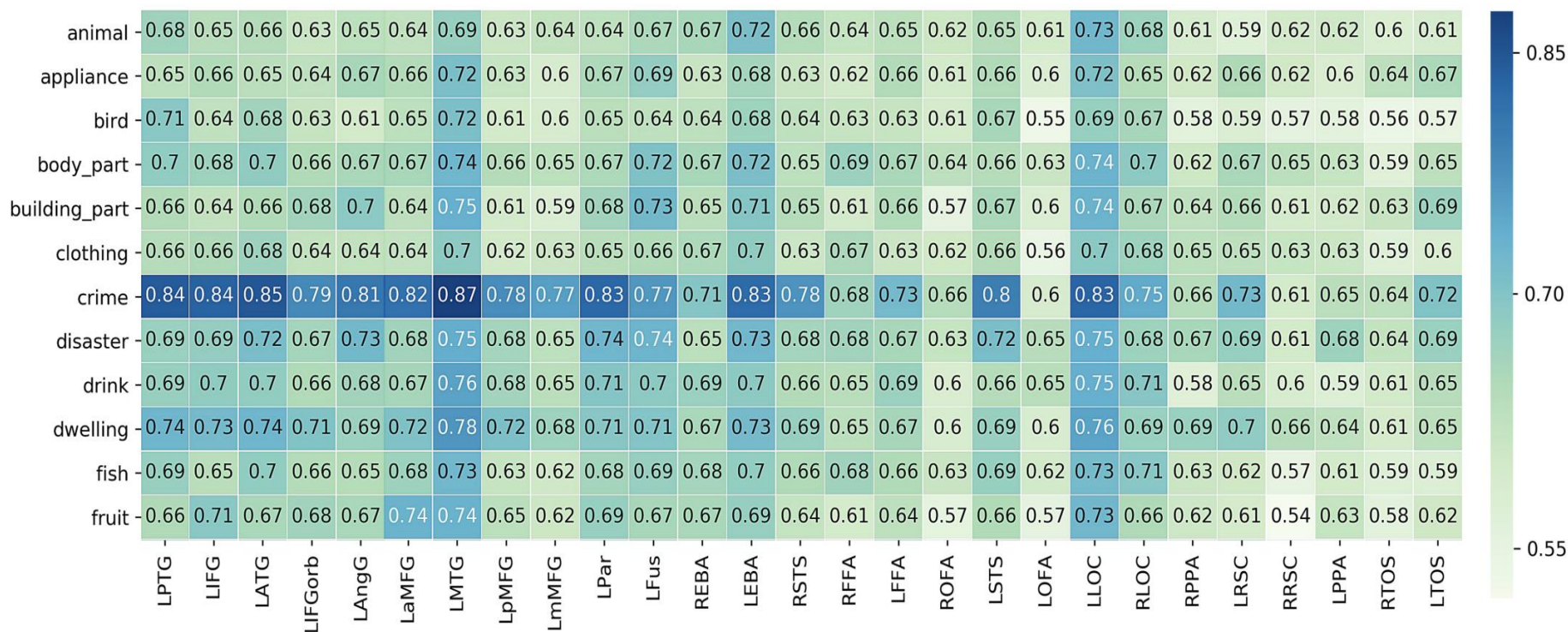
❖ 利用探针任务及消融测试解释表示模型的编码表现



- 借助探针分析方法量化了分布式表示捕捉语义、语法和形态学特征的能力，并计算编码准确率与探针任务分数的相关系数，我们发现分布式表示建模句子语义相关性的能力与编码准确率显著相关。这一发现受到了消融测试结果的支持。

3. 语义解码

❖ 分布式表示模型的在不同脑区、不同主题上的编码准确率

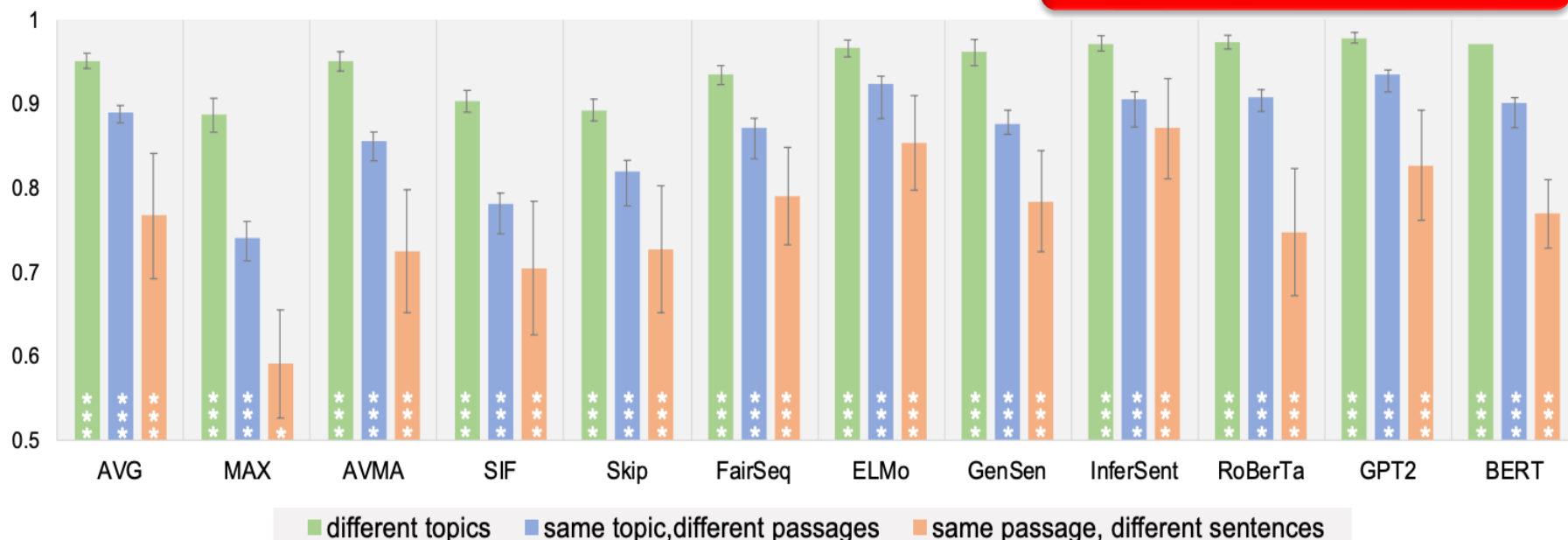


- 在主题层次上，左颞叶中回(LMTG)比其他语言和语义网络中的兴趣区(ROI)的预测准确率更高；左枕叶复合体(LLOC)比其他视觉网络中的ROI的预测准确率更高。
- 一个主题可以在多个ROI得到准确的预测，同时在一个特定ROI上可以预测出多个主题。这说明在人脑中，某个主题并不仅仅被一个或者几个脑区表示，而是广泛分布于多个脑区之中。

3. 语义解码

❖ 分布式表示模型的神经解码准确率

IEEE TNLS 2020



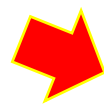
- 在神经解码准确率上，结构化模型并不始终显著地优于非结构化模型；
- 结构化模型在处理细粒度的语义差别上效果较好，而非结构化模型在处理粗粒度的语义差别上可以取得较为满意的表现；
- 值得注意的是，有监督的结构化模型在细粒度的语义解码任务上表现突出；监督任务引入的额外语义信息如何帮助了解码任务，值得进一步探讨。

提 纲

1. 引言

2. 语义表征

3. 语义解码

 4. 结束语

4. 结束语

◆词义向量表示是 NLP 的关键和基础

- 基于多模态信息构建更加准确的词向量表示
- 解释自然语言的编解码方法
- 构建大脑语义地图

◆由词构成短语、由短语构成句子甚至段落和篇章，现有的向量计算方法无法胜任

$$\left\{ \begin{array}{l} \text{蓝} + \text{天} \neq \text{蓝天} \\ \text{白} + \text{云} \neq \text{白云} \end{array} \right. \quad \left\{ \begin{array}{l} \text{疾} + \text{病} \approx \text{疾病} \\ \text{图} + \text{书} \approx \text{图书} \end{array} \right.$$

语义计算不是简单的数学运算，也不是逻辑运算，更不是“赌徒式”的概率推算，而应该是可以追踪、可以解释、可以举一反三和产生新概念的新的运算系统。

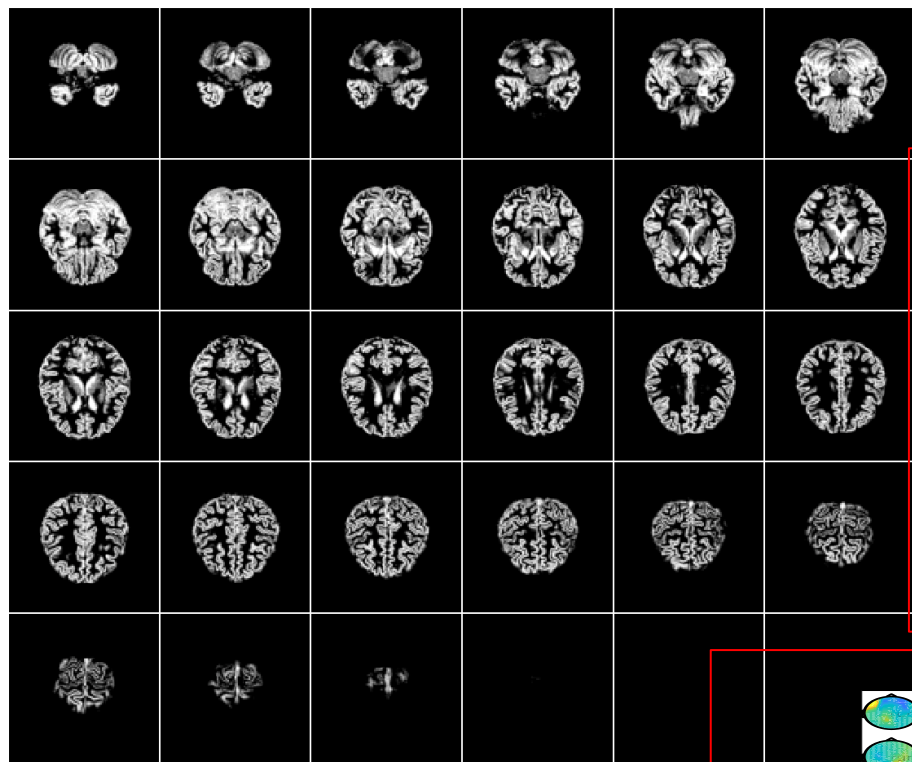
4. 结束语

◆正在开展的工作：构造汉语的大脑语义表征图谱

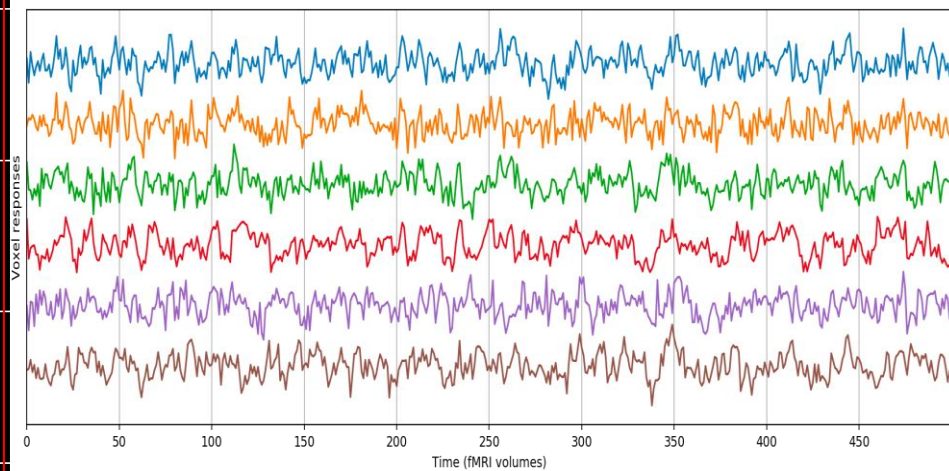
- 筛选了12个中文故事，内容为讲述个人经历的完整故事（包括1200-1700个中文单词，总计17887个单词，其中词典大小为3242，每个词出现1-20次）
- 请专业播音专业人员进行录音，每个故事音频长度10-12分钟
- 预实验：包括5个被试，采集其fMRI和MEG数据（已经完成）
- 最终收集：15人左右；分别采集fMRI数据和MEG数据（一周之后）

- 237癌症 李兰妮完成版.mp3
- 232舞蹈 刘岩完成版.mp3
- 208作家 吴晓波完成版.mp3
- 204经历 董明珠完成版.mp3
- 190遗憾 毕淑敏完成版.mp3
- 179博物馆 郑欣淼完成版.mp3
- 163学校 李希贵完成版.mp3
- 82故乡 阿来完成版.mp3
- 38医学 李晓鹰完成版.mp3
- 34建筑 崔愷完成版.mp3
- 14管道 张文伟完成版.mp3

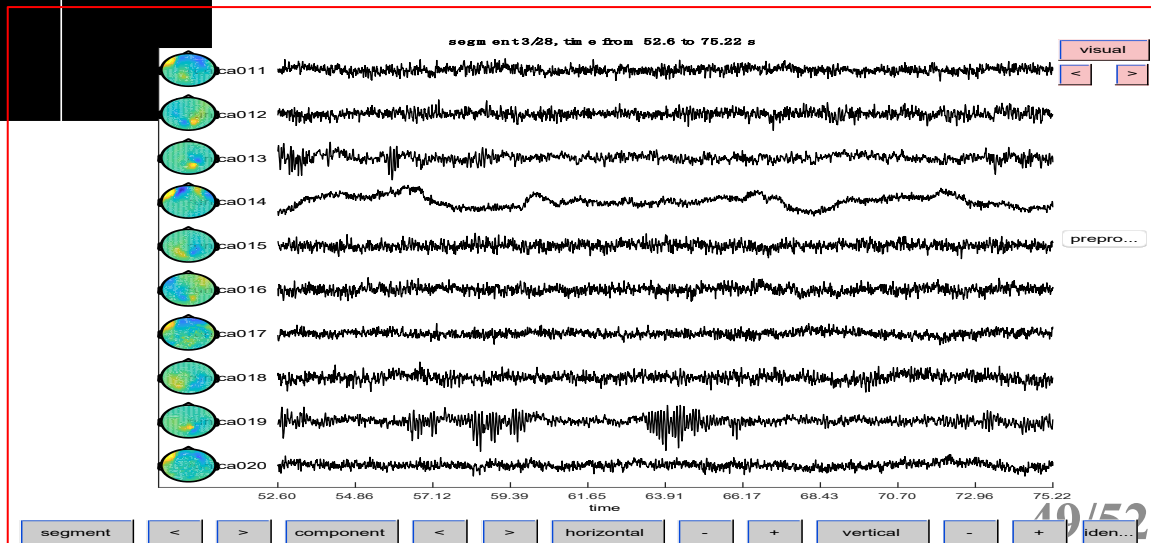
4. 结束语



fMRI: 功能磁共振成像



MEG: 脑磁图



相关工作

1. S. Wang et al. Fine-grained Neural Decoding with Distributed Word Representations. *Information Sciences*, 507 (2020) 256–272
2. S. Wang et al. Empirical Exploring Word-Character Relationship for Chinese Sentence Representation. *ACM TALLIP*, Vol. 17, No. 3, Article 14, 18 pages
3. S. Wang et al. Comparison Study on Critical Components in Composition Model for Phrase Representation, *ACM TALLIP*, Vol. 16, No. 3, Article 16, 25 pages
4. J. Sun et al. Towards Sentence-Level Brain Decoding with Distributed Representations. *Proc. AAAI'2019*, pp. 7047-7054
5. S. Wang et al. Investigating Inner Properties of Multimodal Representation and Semantic Compositionality with Brain-based Componential Semantics. *Proc. AAAI'2018*
6. S. Wang et al. Learning Multimodal Word Representation via Dynamic Fusion Methods. *Proc. AAAI'2018*, pp. 5973-5980

相关工作

7. S. Wang et al. Associative Multichannel Autoencoder for Multimodal Word Representation. *Proc. EMNLP'2018*, pp. 116-124
8. S. Wang et al. Exploiting Word Internal Structures for Generic Chinese Sentence Representation. *Proc. EMNLP'2017*, pp. 298-303
9. S. Wang et al. Learning Sentence Representation with Guidance of Human Attention. *Proc. IJCAI'2017*, pp. 4137-4143
10. S. Wang et al. Probing Brain Activation Patterns by Dissociating Semantics and Syntax in Sentences. *Proc. AAAI'2020*
11. J. Sun et al. Neural Encoding and Decoding with Distributed Sentence Representations, *IEEE Transactions on Neural Networks and Learning Systems* (TNNLS)
(http://www.nlpr.ia.ac.cn/cip/ZongPublications/2020/TNNLS_wangshaonan.pdf)

Thanks!

谢谢!

