FMRI ENCODING AND DECODING METHOD WITH TEXTUAL REPRESENTATIONS

Shaonan Wang
https://wangshaonan.github.io/

Institute of Automation Chinese Academy of Sciences

Outline

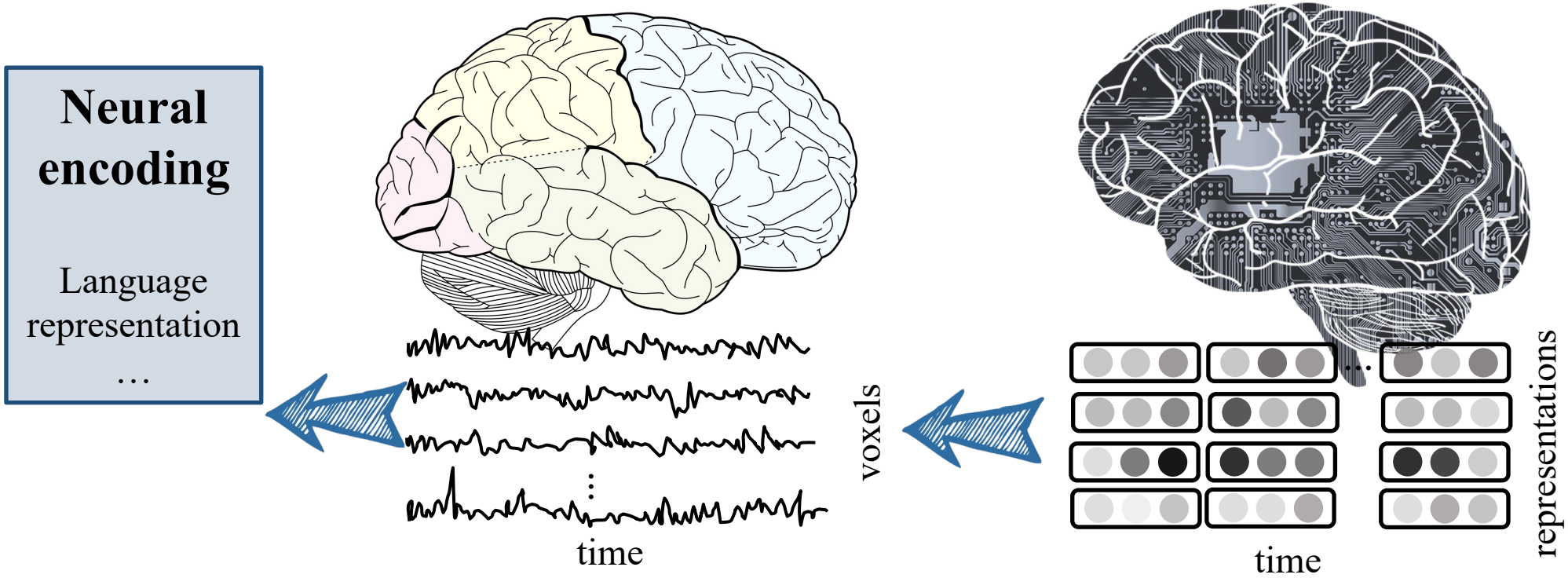
- **fMRI Encoding**
  - Framework
  - Probing Brain Activation Patterns by Dissociating Semantics and Syntax in Sentences
  - Probing Word Syntactic Representations in the Brain by a Feature Elimination Method
- **fMRI Decoding**
  - Framework
  - Cross-Modal Cloze Task: A New Task to Brain-to-Word Decoding
  - Towards Brain-to-Text Generation: Neural Decoding with Pre-trained Encoder-Decoder Models
- **Summary**

Outline

- **fMRI Encoding**
  - Framework
  - Probing Brain Activation Patterns by Dissociating Semantics and Syntax in Sentences
  - Probing Word Syntactic Representations in the Brain by a Feature Elimination Method
- **fMRI Decoding**
  - Framework
  - Cross-Modal Cloze Task: A New Task to Brain-to-Word Decoding
  - Towards Brain-to-Text Generation: Neural Decoding with Pre-trained Encoder-Decoder Models
- **Summary**

fMRI encoding framework

Computational language models are uninterpretable
Cannot explore how the brain encodes different linguistic features

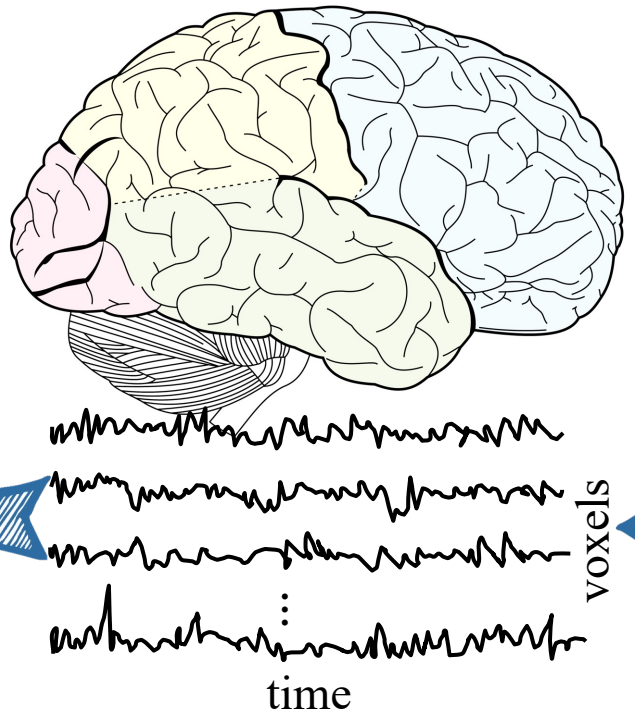


fMRI encoding with specialized representations

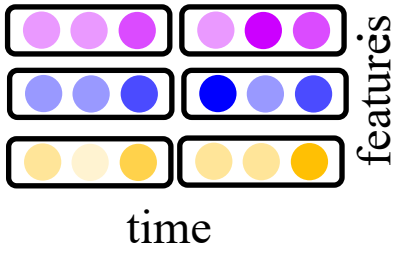
- *Shaonan Wang, Jiajun Zhang, Nan Lin and Chengqing Zong.* ***Probing Brain Activation Patterns by Dissociating Semantics and Syntax in Sentences****. AAAI-2020.*

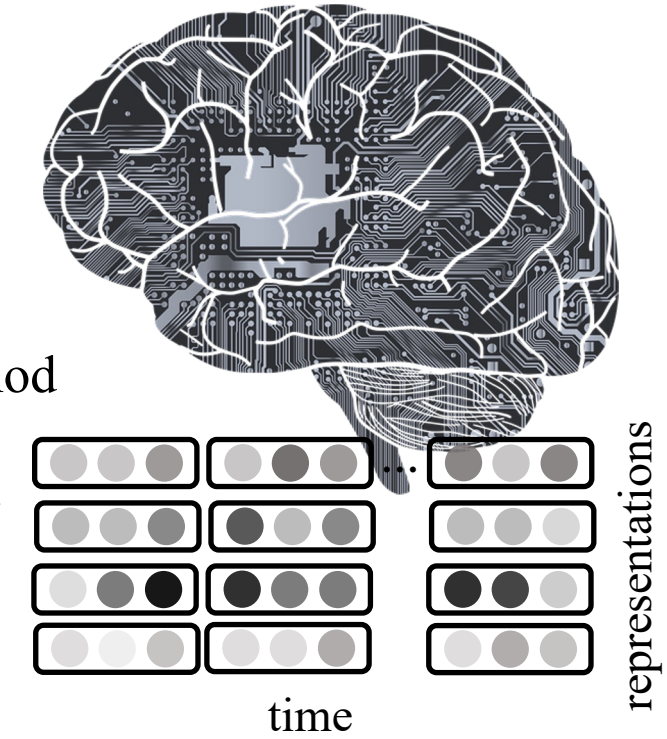
Neural encoding

word meaning
composition
syntax
attention
memory
…



- Semantics&syntax
  - Disentangled feature representation model
- Syntactic features
  - Feature Elimination Method





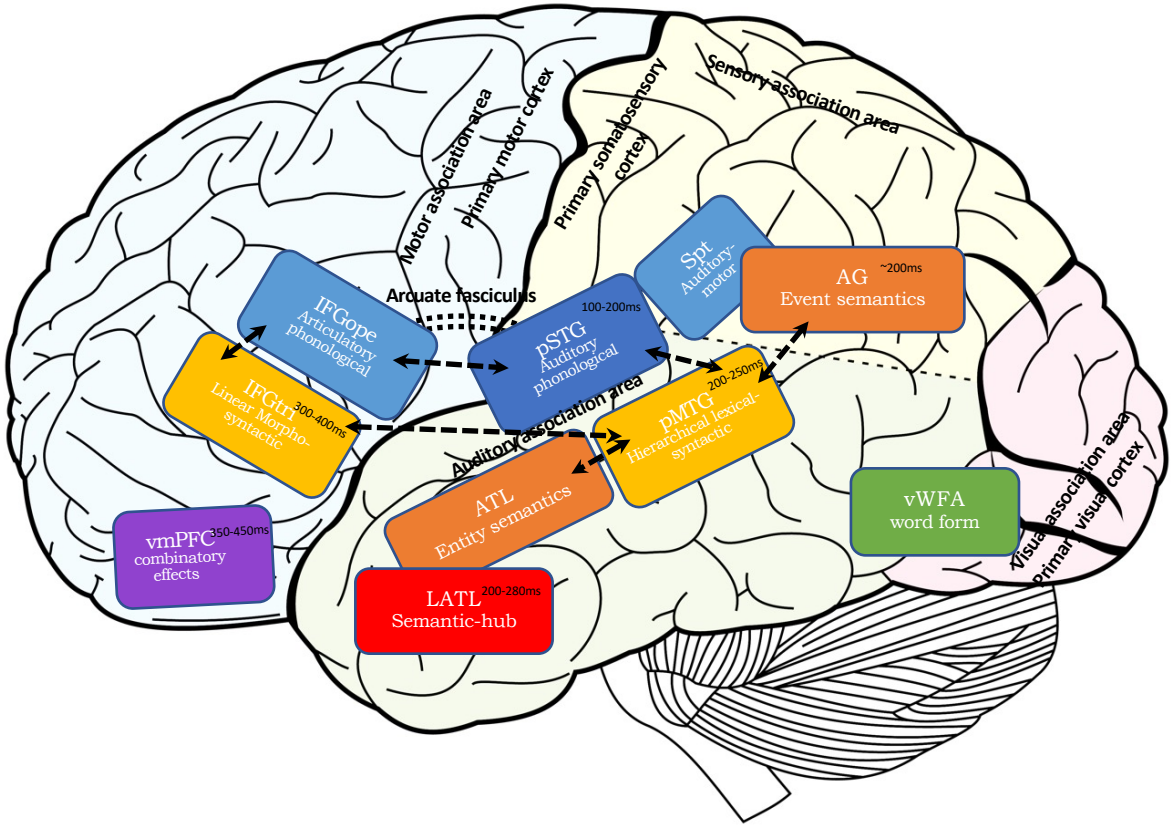
- *Xiaohan Zhang, Shaonan Wang, Nan Lin, Jiajun Zhang and Chengqing Zong.* ***Probing Word Syntactic Representations in the Brain by a Feature Elimination Method****. AAAI-2022*

Outline

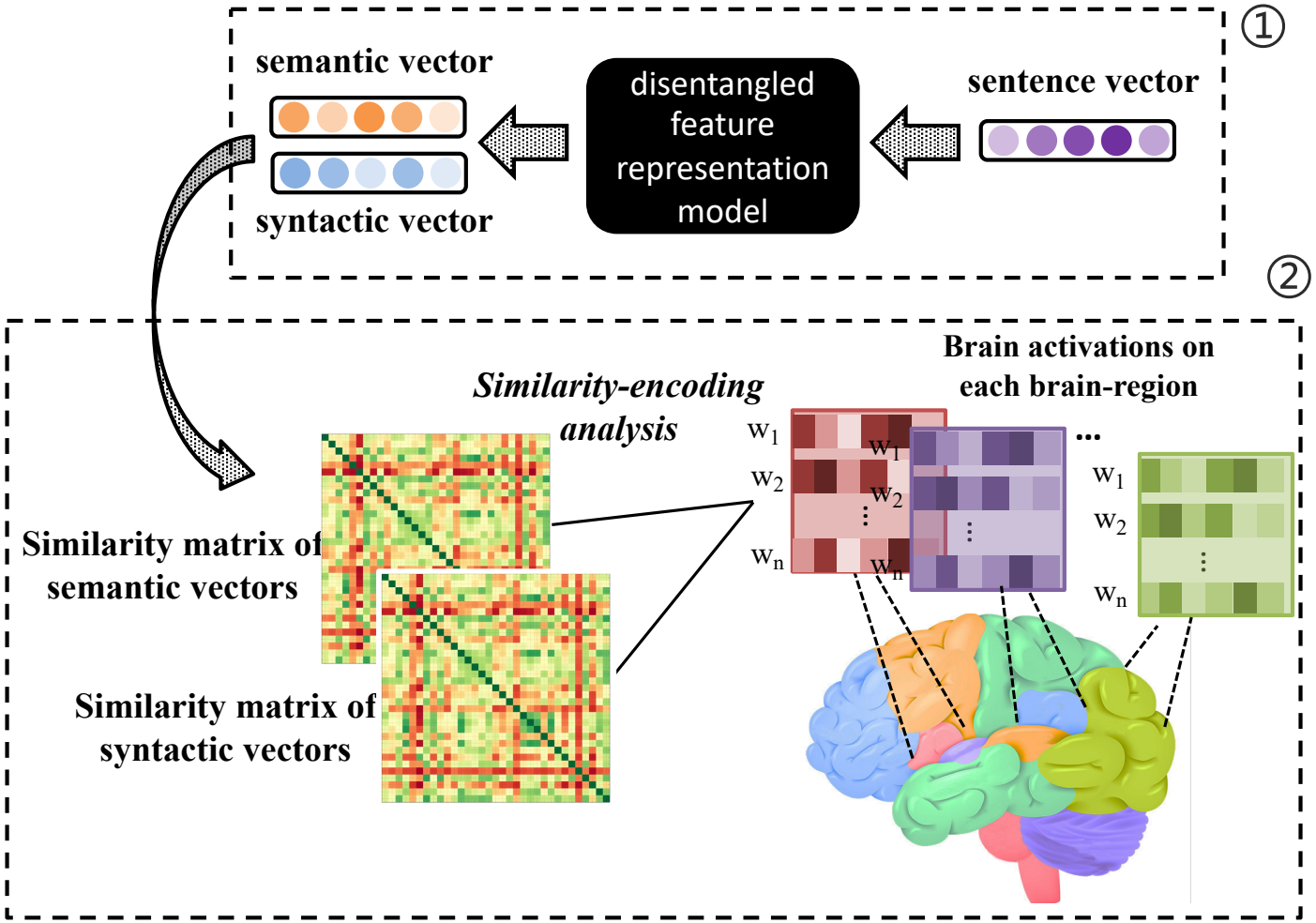
- **fMRI Encoding**
  - Framework
  - Probing Brain Activation Patterns by Dissociating Semantics and Syntax in Sentences
  - Probing Word Syntactic Representations in the Brain by a Feature Elimination Method
- **fMRI Decoding**
  - Framework
  - Cross-Modal Cloze Task: A New Task to Brain-to-Word Decoding
  - Towards Brain-to-Text Generation: Neural Decoding with Pre-trained Encoder-Decoder Models
- **Summary**

Probing Brain Activation Patterns by Dissociating Semantics and Syntax in Sentences

- What is the relation between semantics and syntax and where they are represented in the brain?



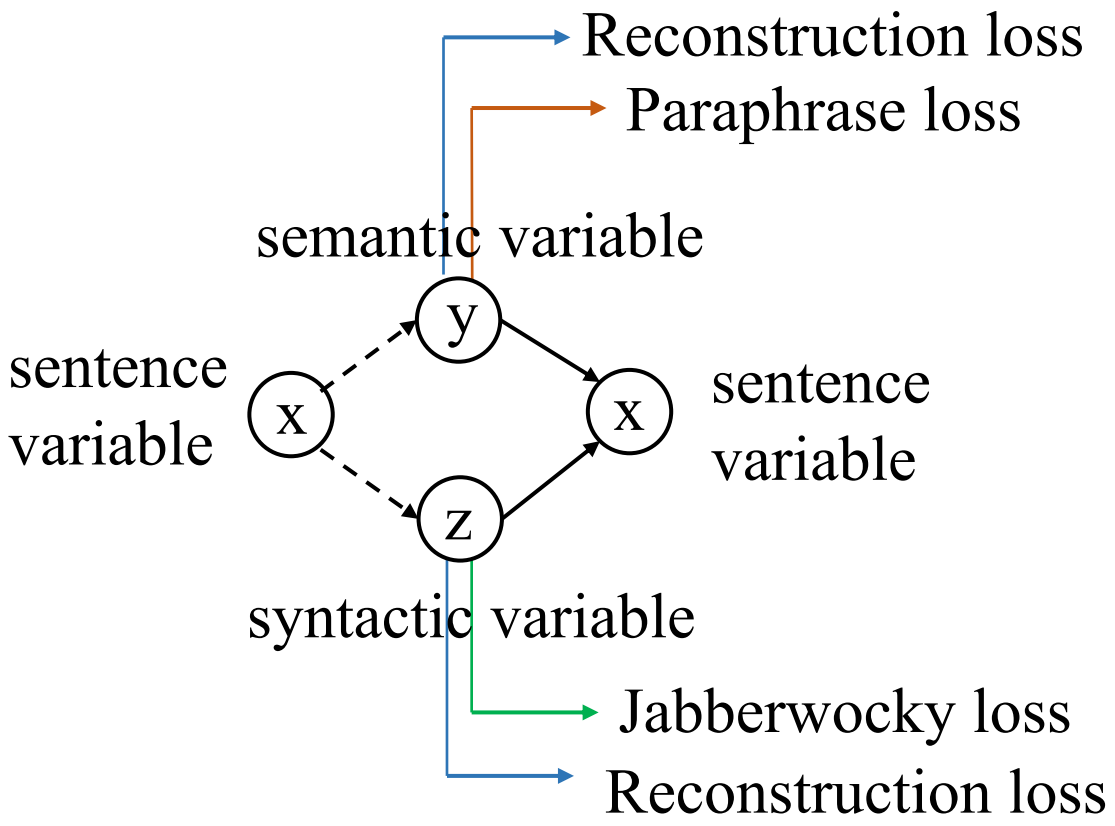
Method fMRI encoding with disentangled feature representation method



① Disentangled feature representation model

② Similarity encoding analysis

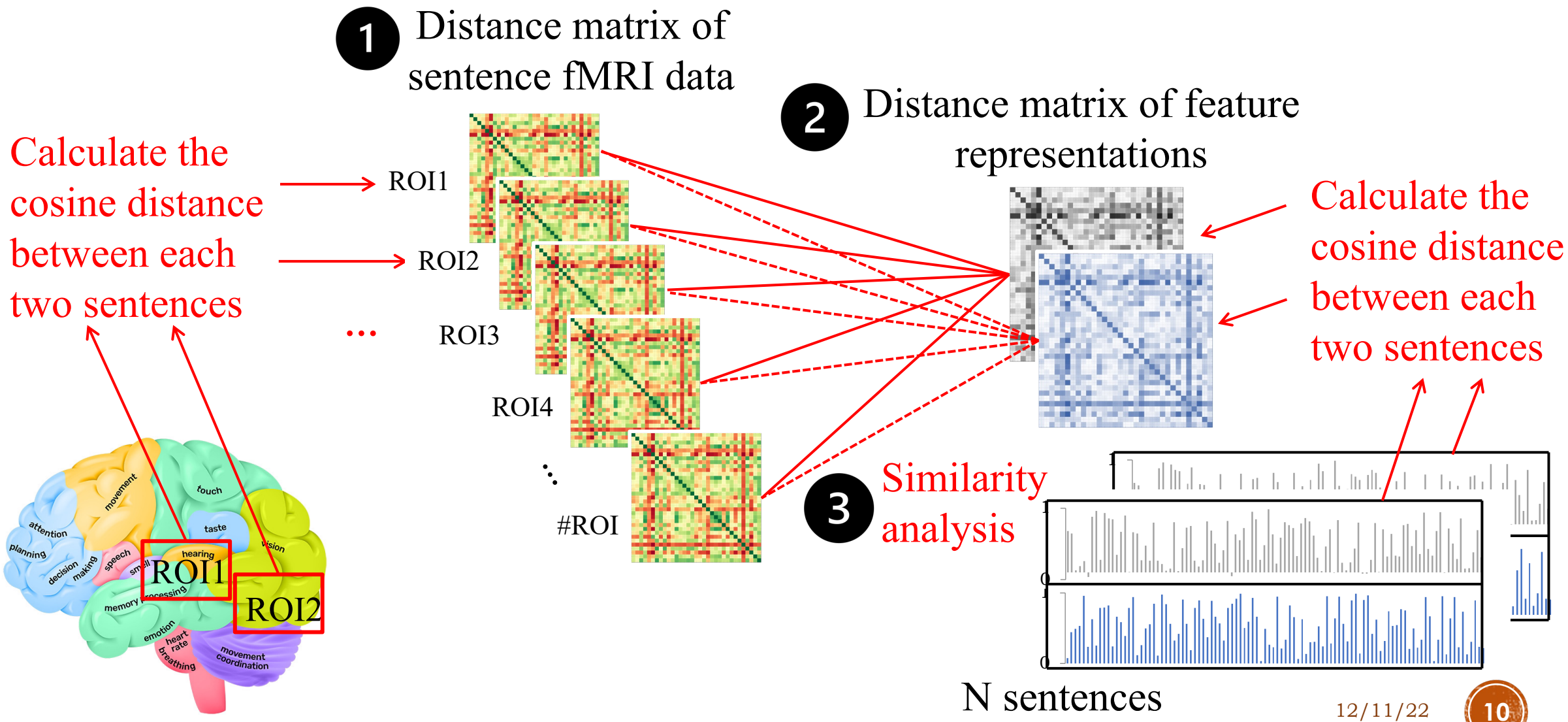
Method Disentangled feature representation model



Paraphrase:
a cat is running & there is a cat running

Jabberwocky:
a cat is running & two dogs are playing

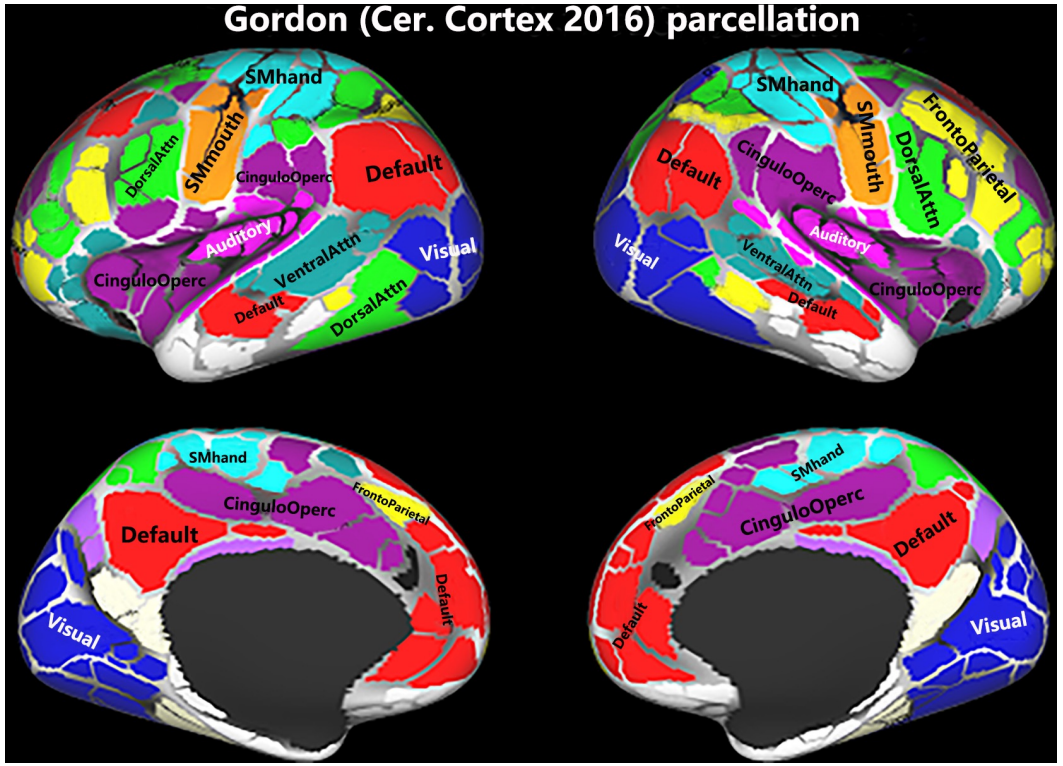
Method Similarity encoding analysis



Datasets

- DFRM model training
  - 500,000 paraphrases
- DFRM evaluation data
  - Semantic: sentence pair similarity
  - Syntactic: POS tagging & parsing
- fMRI data
  - Pereira et al. (2018)
  - 5 participants
  - Dataset 1: 384 sentences (96 text passages, each consisting of 4 sentences describing basic information of a particular concept)
  - Dataset 2: 253 sentences (72 passages, each consisting of 3 or 4 sentences about a particular concept)

333 cortical patches (ROI0-ROI332)



Result Evaluation of the DFRM method

	STS benchmark test (% Pearson correlation ↑)		Averaged STS tests (% Pearson correlation ↑)		Constituent Parsing (F1 ↑)		POS Tagging (%Acc. ↑)	
Random	39.7		42.5		19.2		12.9	
GloVe	41.0		47.9		27.3		23.9	
InferSent	67.8		61.0		28.0		25.1	
ELMo	57.7		60.3		30.4		27.8	
BERT	54.9		59.0		28.6		25.8	
WordAvg	71.9		64.8		25.5		21.4	
LSTMAvg	71.4		64.4		25.7		21.6	
	sem var.	syn var.	sem var.	syn var.	sem var.	syn var.	sem var.	syn var.
VGVAE	65.7	32.2	57.6	28.4	25.1	26.7	20.9	22.6
VGVAE+PLoss	72.5	24.0	**66.3**	28.5	24.2	29.0	19.6	26.3
VGVAE+WPLoss	69.4	8.5	61.1	18.9	24.4	35.7	19.8	33.2
VGVAE+JLoss	55.2	17.0	48.3	24.2	23.6	32.3	18.5	32.3
VGVAE+PLoss+JLoss	72.5	11.4	66.0	22.9	24.2	34.2	19.2	34.4
VGVAE+PLoss+WPLoss	71.2	15.0	65.9	22.6	24.0	34.6	19.3	32.7
DFRM (VGVAE+all)	**73.0**	8.5	65.9	18.4	24.2	**40.0**	19.5	**38.6**

- Our DFRM model are best at separate semantic and syntactic information

Result

Examples of most similar sentences to particular query sentences calculated by semantic or syntactic variables

Query sentence	Neighbor sentences by semantic var.	Neighbor sentences by syntactic var.
a cook is making food .	there is a cook preparing food .	a kid is playing keyboard .
the dog is chasing the geese .	one dog is chasing the other .	the cat is licking a bottle .
you can do it , too .	yes , you can do it .	you should prime it first .
it makes absolutely no difference .	i do n't think it makes much difference .	this is a big problem .
but the economy has n't shown signs of sustainable growth .	the economy , nonetheless , has yet to exhibit sustainable growth .	but the north korean nuclear crisis has dominated his time in office .

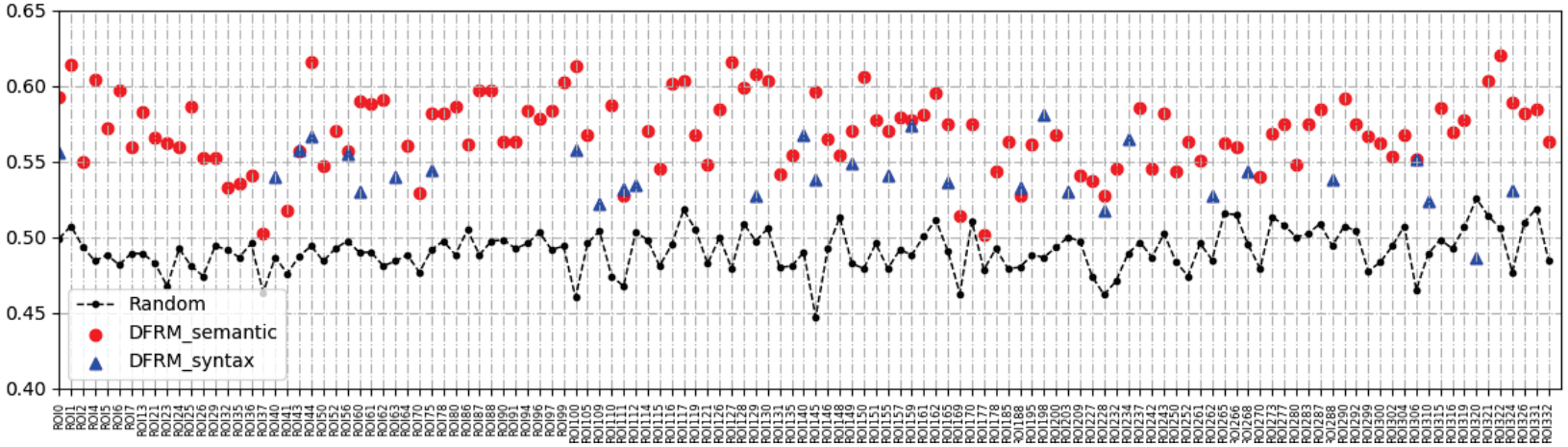
- Neighbor sentences by semantic variable have similar meaning
- Neighbor sentences by syntactic variable have similar syntactic structure

- Our DFRM can separate semantic and syntactic information

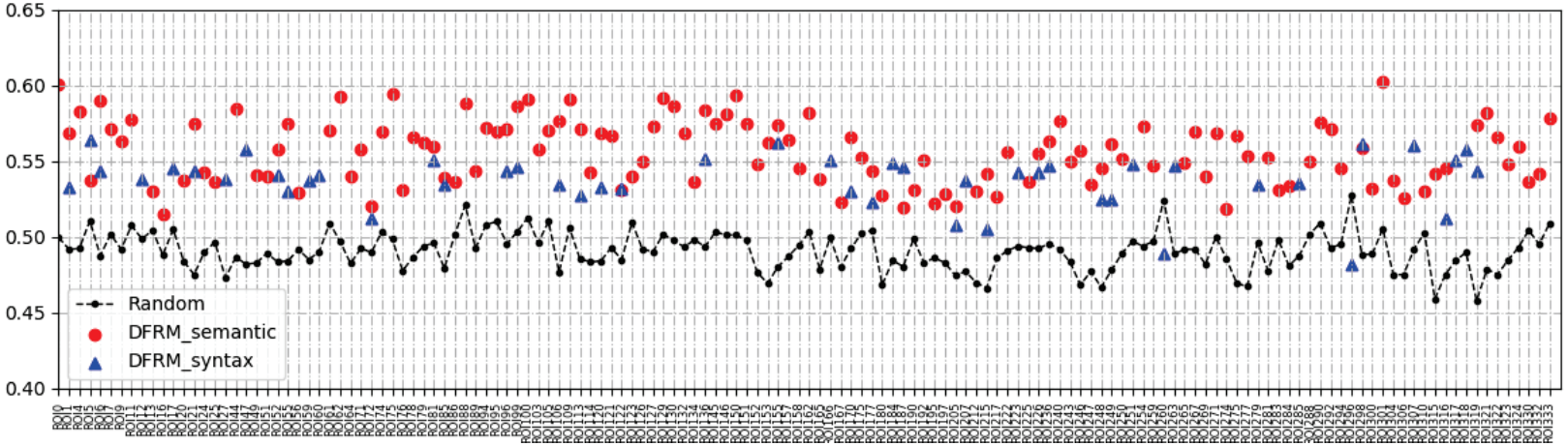
Result

Significant voxels of our DFRM model

Dataset 1



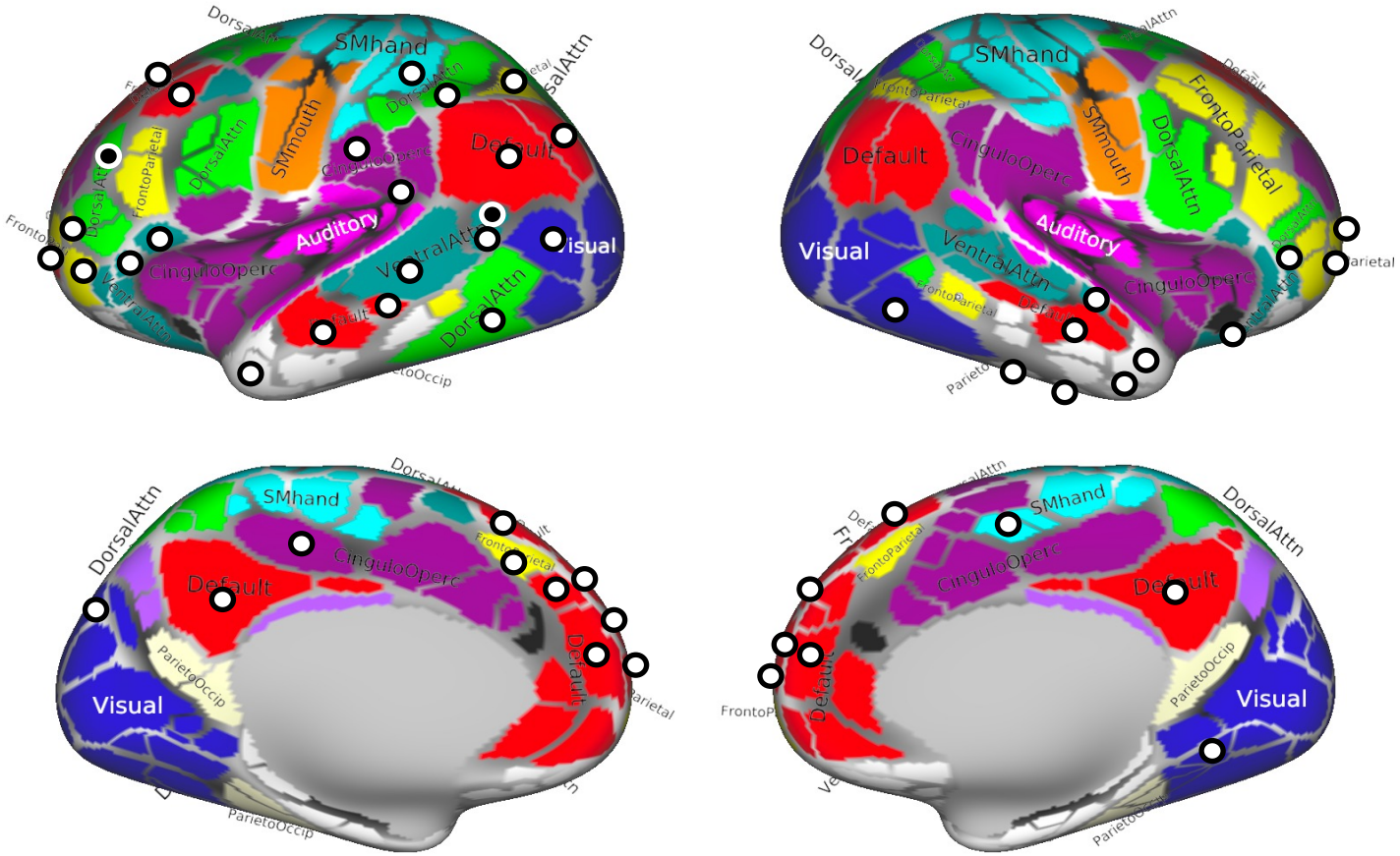
Dataset 2



- Compared with syntactic features, semantic features are more robustly represented in brain regions such as default-model networks, frontoparietal, and visual networks.
- There are overlaps and specific brain regions for semantic and syntactic representation. For instance, part of the frontal and temporal lobe is only sensitive to semantics, while part of the right superior frontal and right inferior parietal is only sensitive to syntax.

Result

Dataset 1&Dataset 2: statistically significant results on both dataset



Semantic feature

Syntactic feature

- Semantic and syntactic features are distributedly represented in the brain
- Semantic information are most robustly represented than syntactic information

Outline

- **fMRI Encoding**
  - Framework
  - Probing Brain Activation Patterns by Dissociating Semantics and Syntax in Sentences
  - Probing Word Syntactic Representations in the Brain by a Feature Elimination Method
- **fMRI Decoding**
  - Framework
  - Cross-Modal Cloze Task: A New Task to Brain-to-Word Decoding
  - Towards Brain-to-Text Generation: Neural Decoding with Pre-trained Encoder-Decoder Models
- **Summary**

Probing Word Syntactic Representations in the Brain by a Feature Elimination Method

- How fine-grained syntactic features are represented and whether the neural correlates of different syntactic features overlap or dissociate from each other?

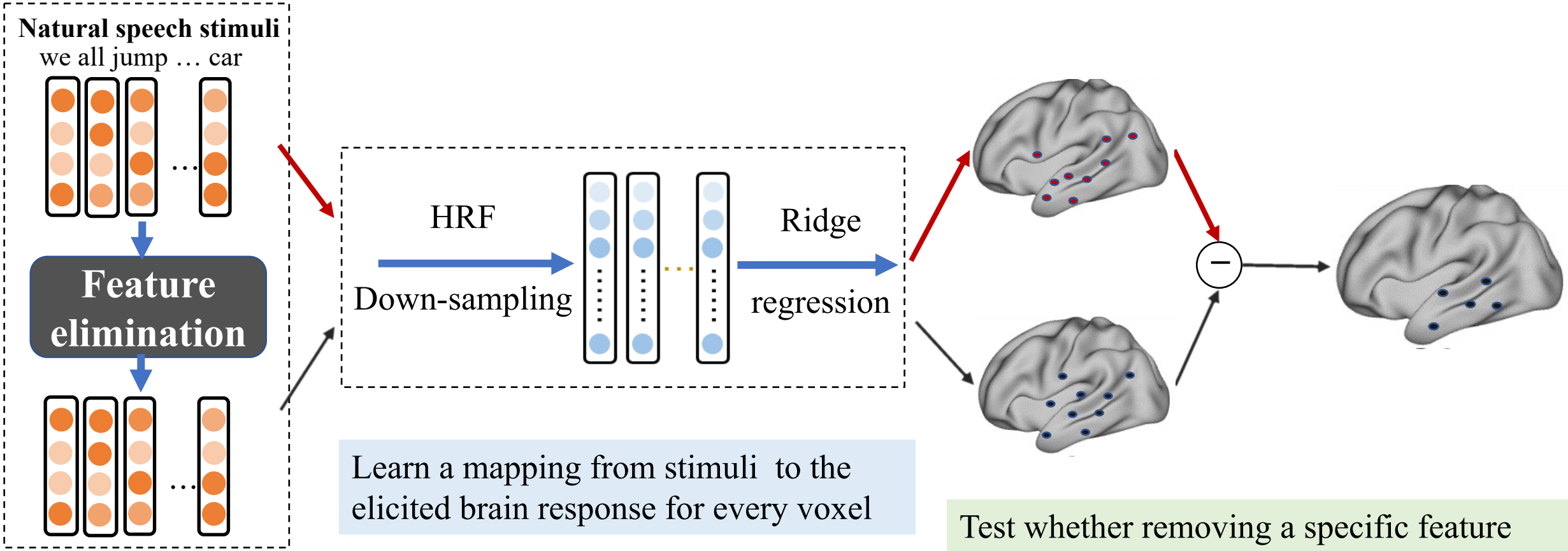
Word syntactic feature:

Part-of-Speech (POS)，Named Entity (NE)，Semantic Role (SR)，Dependency Relationship (DEP)

Sentence	When somebody wrote a story in the Washington Post on Friday morning …
POS	When [WRB] … Washington [NNP] Post [NNP] on [IN] Friday [NNP] morning [NN]
NE	When [*] … Washington [ORG] Post [ORG] on [*] Friday [TIME] morning [TIME]
SR	(wrote when) [ARGM-TMP] (wrote somebody) [ARG0] … (wrote story) [ARG2]
DEP	(wrote when) [advmod] (wrote somebody) [nsubj] (story a) [det] (wrote story) [dobj]

- WRB: Wh-adverb; NNP: Proper noun; IN: Preposition or subordinating conjunction; NN: Noun
- [*] means not an entity
- ARGM-TMP: when; ARG0: giver; ARG2: entity given to
- advmod:adverbial modifier; nsubj:nominal subject; det:determiner; dobj:direct object

Method fMRI encoding with feature elimination Method



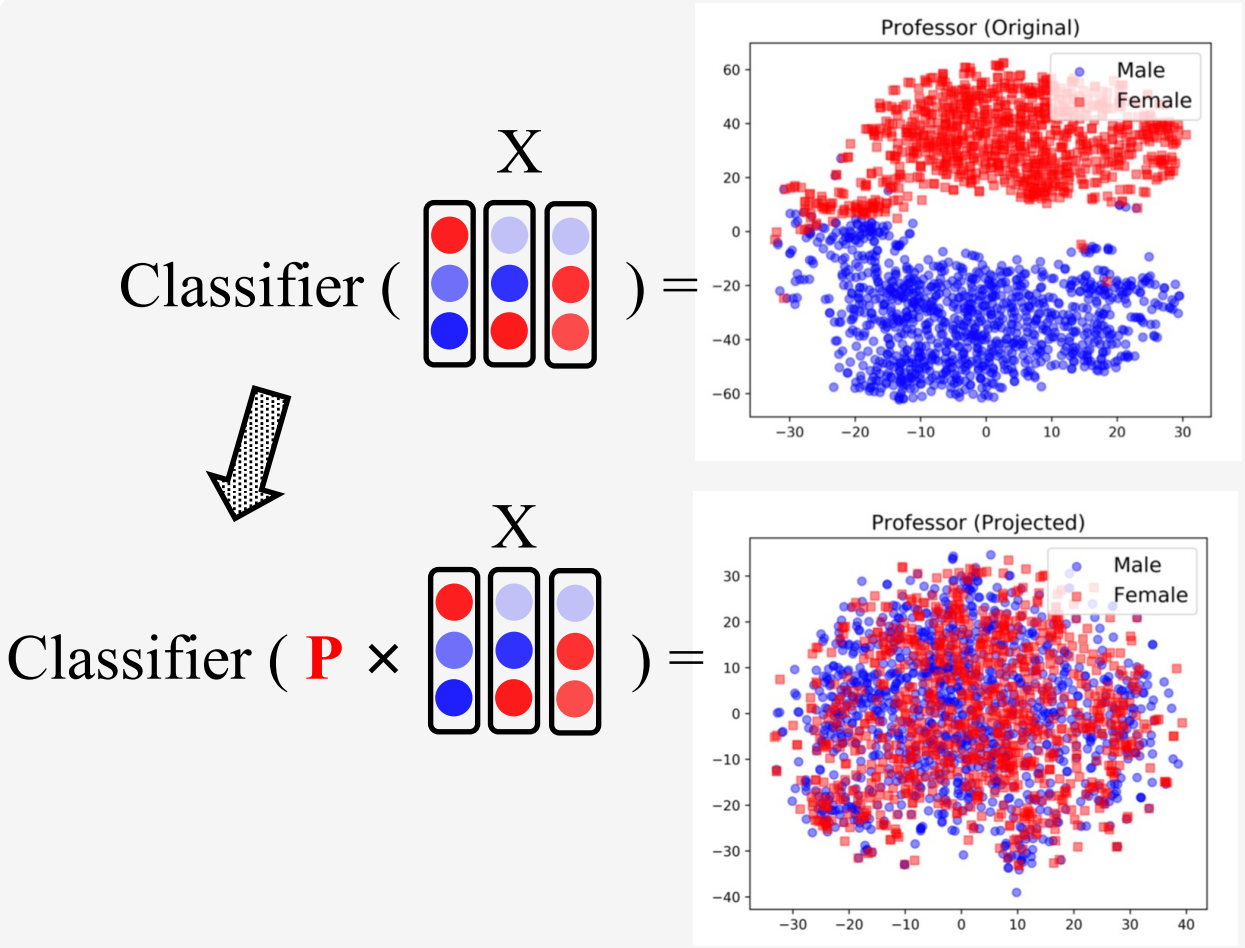
Learn a mapping from stimuli to the elicited brain response for every voxel

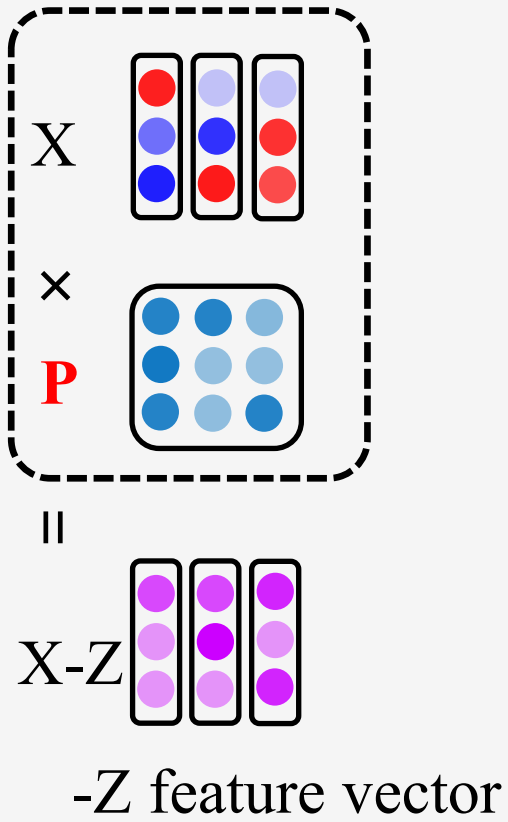
Test whether removing a specific feature causes a significant drop in encoding results

Word representations & Word representations that eliminate the information of a target feature

Method Feature elimination Method

Mean Vector Nullspace Projection





Datasets

- Feature elimination method training
  - OntoNotes 5.0 (POS, NE, WF, SR)
  - Universal dependencies treebank (DEP)

#words	POS	NE	WF	SR	DEP
Training	173,322	173,322	173,322	238,734	203,150
Testing	29,962	29,962	29,962	38,629	24,956
Validation	35,952	35,952	35,952	52,097	24,949

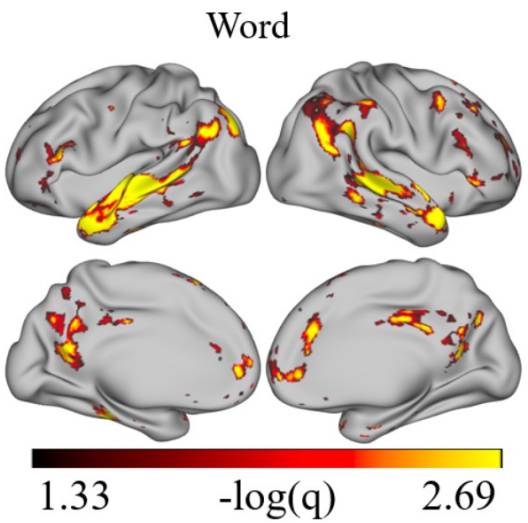
- fMRI dataset
  - Zhang et al. (2020) , 19 subjects listening to 52 different stories in total (2-3 stories per subject), then concatenate fMRI data across all stories and subjects.
  - 47,356 words

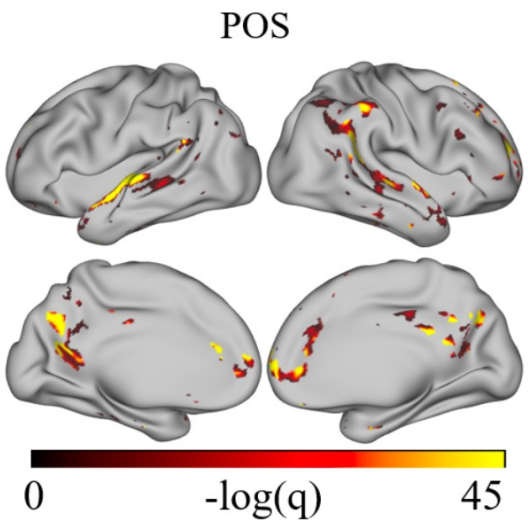
Results Evaluation of our MVNP method

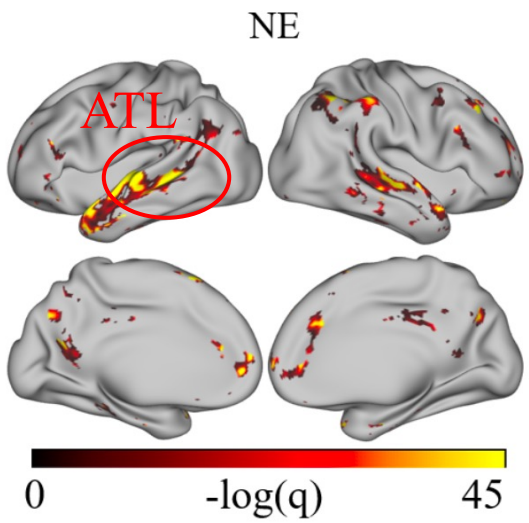
			POS	NE	SR	DEP
Random			11.77	19.54	42.43	12.30
ELMo	Word		97.65 ± 0.05	77.81 ± 0.99	77.49 ± 2.18	92.71 ± 0.15
	MVNP	Null POS	**21.70 ± 3.35**	78.98 ± 1.47	71.09 ± 0.28	87.93 ± 0.10
		Null NE	94.69 ± 0.10	**13.79 ± 8.77**	76.70 ± 0.15	92.24 ± 0.09
		Null SR	96.79 ± 0.05	80.23 ± 0.60	**22.03 ± 1.52**	88.28 ± 0.17
		Null DEP	77.79 ± 0.10	67.38 ± 0.67	62.15 ± 0.20	**15.98 ± 1.64**
BERT	Word		97.89 ± 0.06	78.89 ± 2.66	78.04 ± 0.37	93.21 ± 0.20
	MVNP	Null POS	**21.76 ± 2.68**	77.45 ± 0.64	72.80 ± 0.35	86.53 ± 0.15
		Null NE	95.23 ± 0.06	**12.30 ± 6.38**	77.23 ± 0.27	93.12 ± 0.18
		Null SR	96.86 ± 0.07	79.07 ± 0.84	**23.40 ± 1.93**	88.11 ± 0.14
		Null DEP	76.33 ± 0.11	63.10 ± 0.77	62.69 ± 0.40	**13.96 ± 1.80**

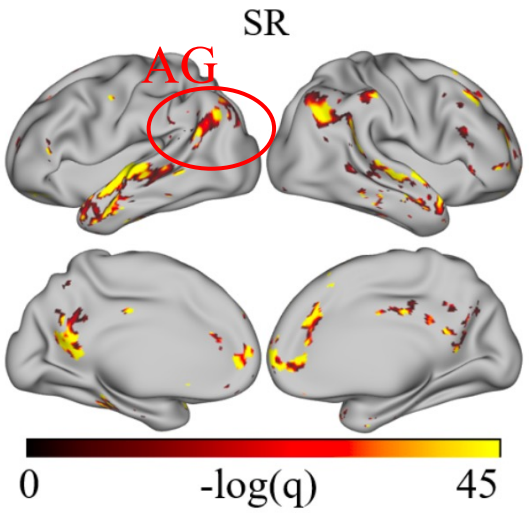
- Our MVNP can eliminate one feature effectively and has a smaller influence on other features when removing one feature from ELMo and BERT embeddings

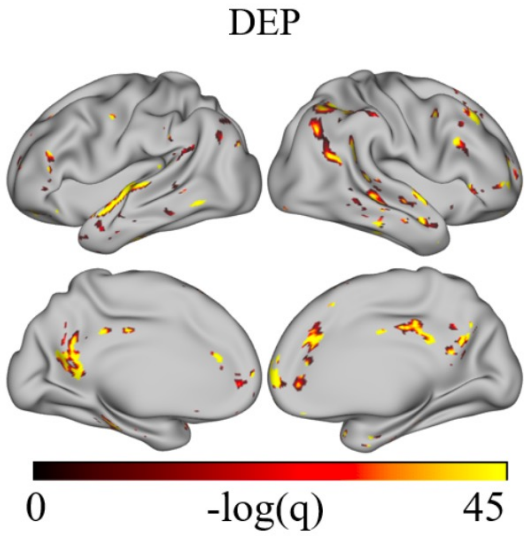
Results fMRI encoding of syntactic features



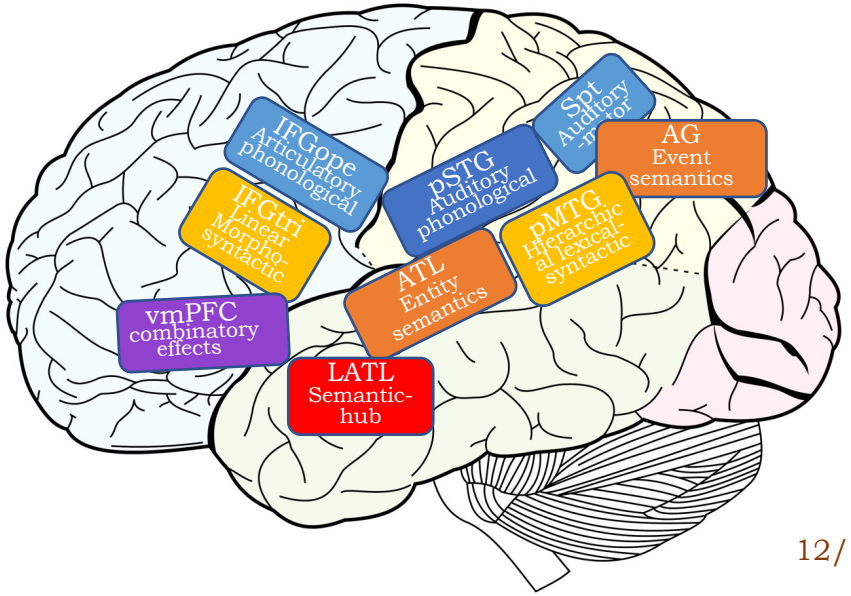






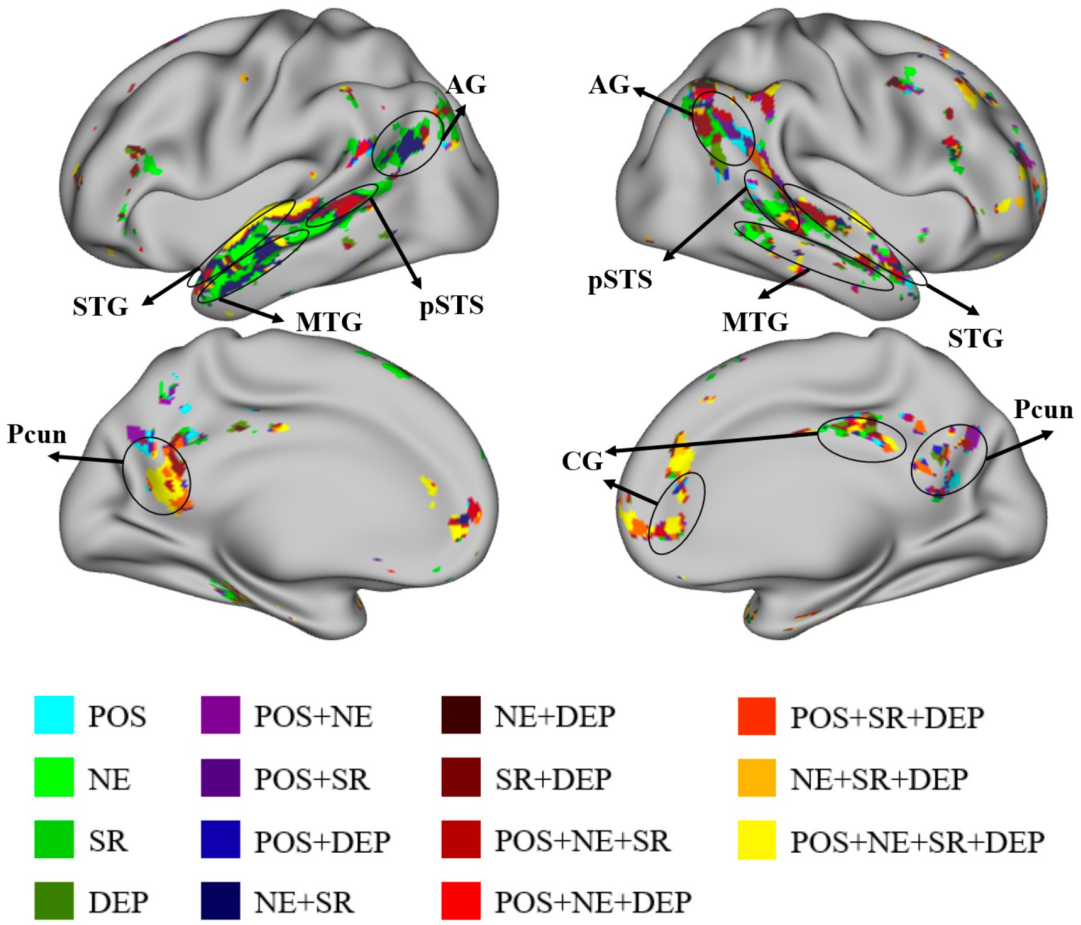


- Syntactic features are distributively represented in the brain
- Syntactic brain networks are largely overlapped with semantic brain networks



Results fMRI encoding of overlapped syntactic features

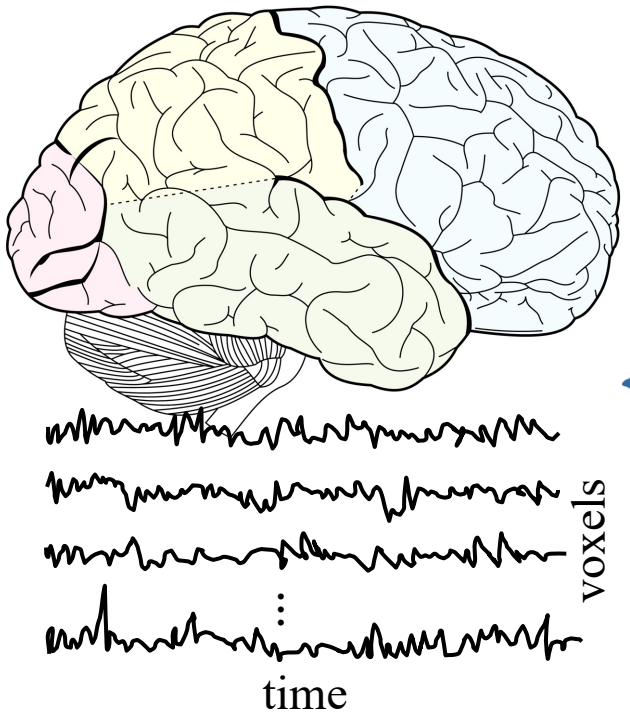
- Two brain regions are activated by syntactic features that have not been found in previous work, i.e., the precuneus (Pcun) and the cingulate (CG) gyrus, suggesting that the brain foundations of syntactic information processing might be broader than those suggested by classical studies.
- Different syntactic features are represented and integrated in a hierarchical brain system: there are some brain areas that encode several syntactic features, while other areas are only sensitive to one feature.

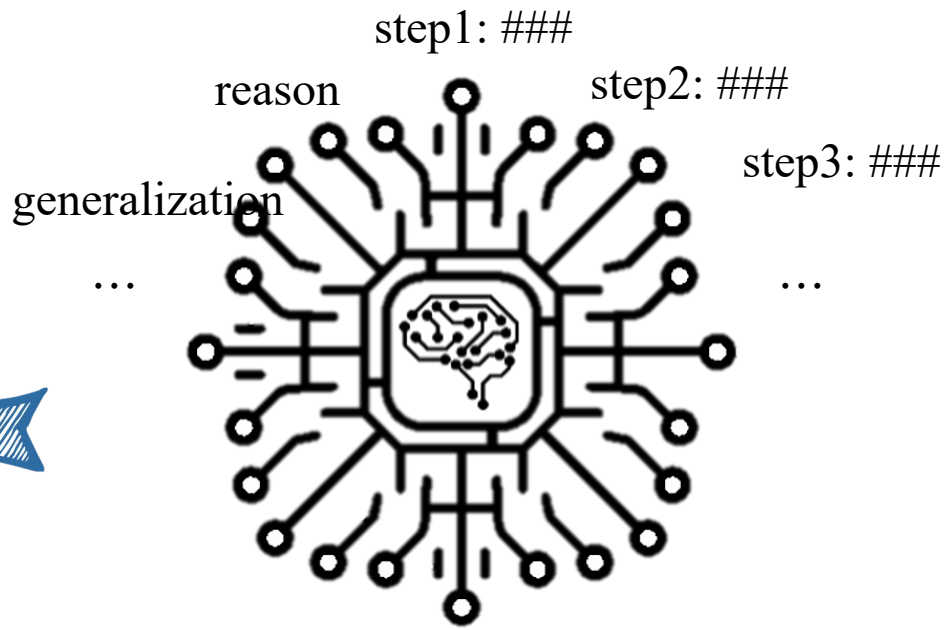


NEXT STEP: Interpretable Language Models

Language Understanding Mechanism

word meaning
composition
syntax
attention
memory
…





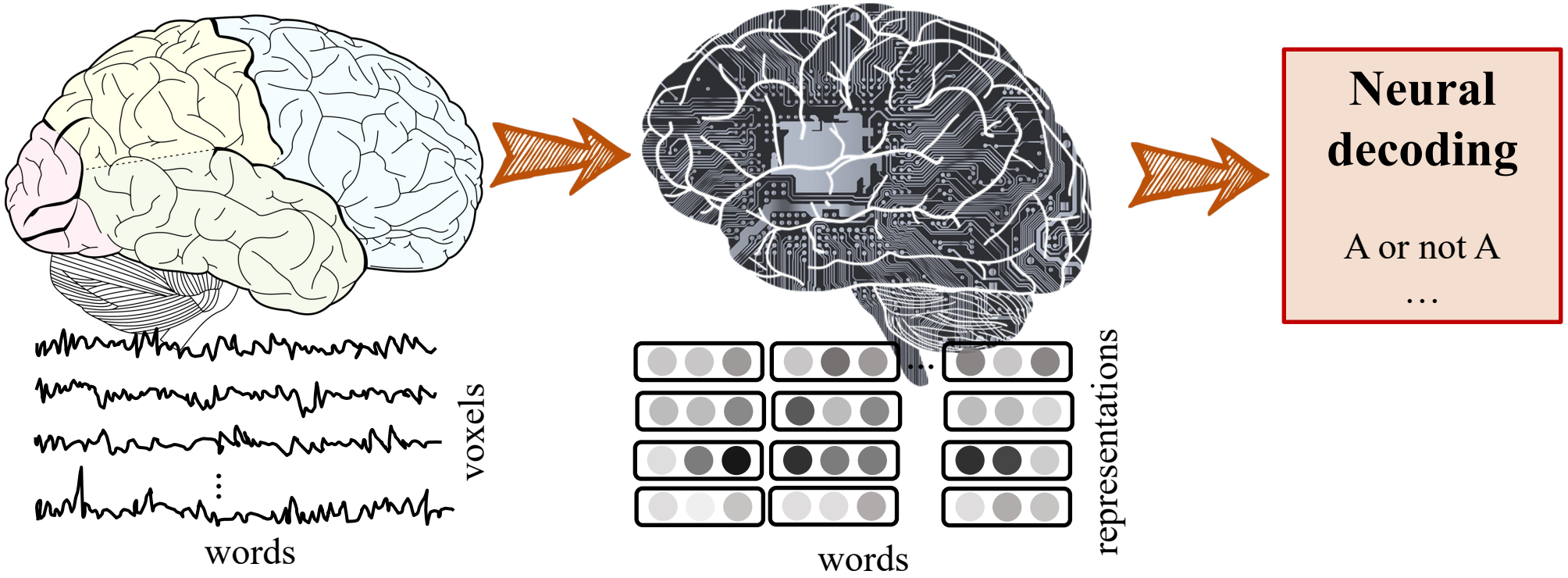
Interpretable and controllable language models

Outline

- **fMRI Encoding**
  - Framework
  - Probing Brain Activation Patterns by Dissociating Semantics and Syntax in Sentences
  - Probing Word Syntactic Representations in the Brain by a Feature Elimination Method
- **fMRI Decoding**
  - Framework
  - Cross-Modal Cloze Task: A New Task to Brain-to-Word Decoding
  - Towards Brain-to-Text Generation: Neural Decoding with Pre-trained Encoder-Decoder Models
- **Summary**

fMRI decoding framework

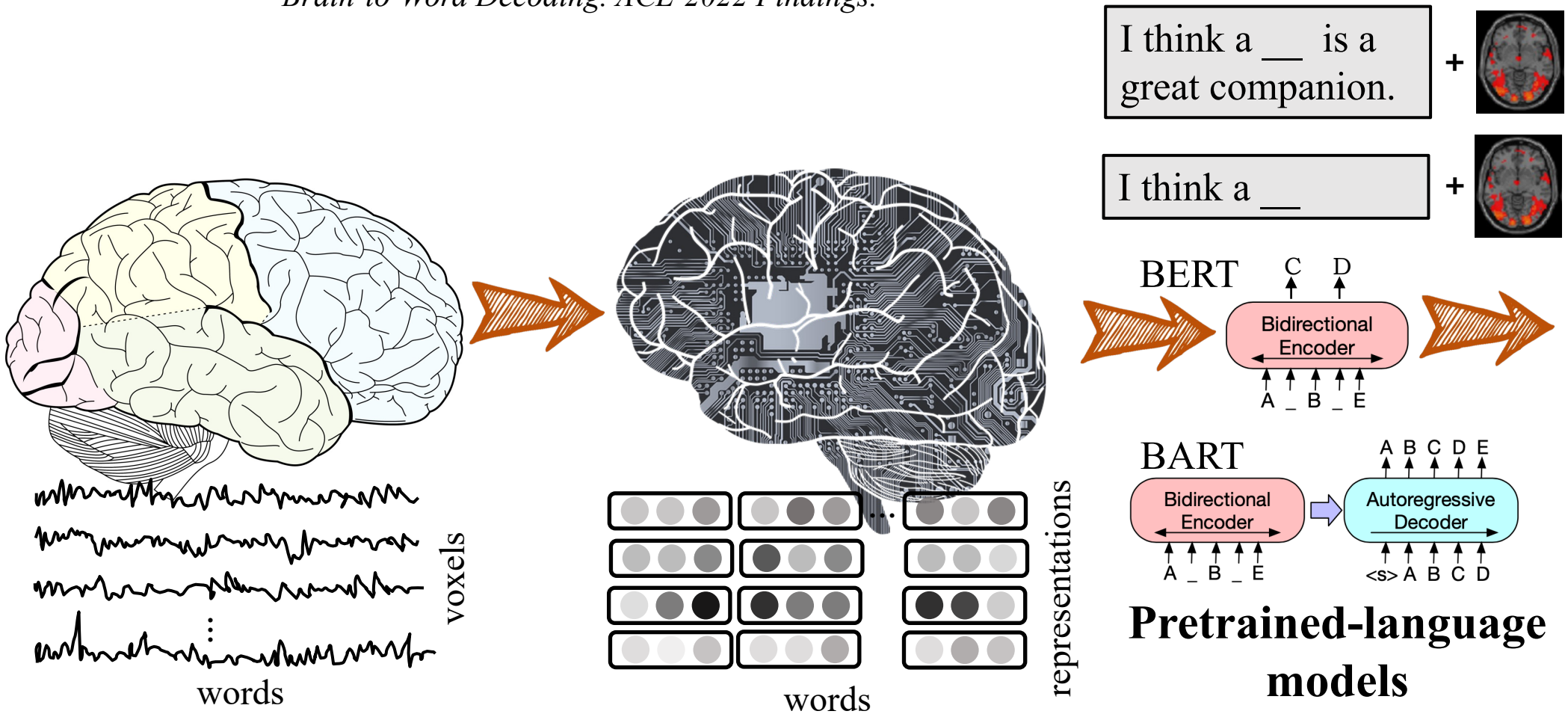
Cannot directly predict word from vocabulary
Cannot generate coherent text



Mitchell T M, Shinkareva S V, Carlson A, et al. Predicting human brain activity associated with the meanings of nouns. Science, 2008.

fMRI decoding with pretrained-language models

- *Shuxian Zou, Shaonan Wang, Jiiajun Zhang, Chengqing Zong. Cross-Modal Cloze Task: A New Task to Brain-to-Word Decoding. ACL-2022 Findings.*



Neural decoding

What should I eat after this boring story listening experiment …

- *Shuxian Zou, Shaonan Wang, Jiiajun Zhang, Chengqing Zong. Towards Brain-to-Text Generation: Neural Decoding with Pre-trained Encoder-Decoder Models. NeurIPS 2021 AI for Science Workshop.*

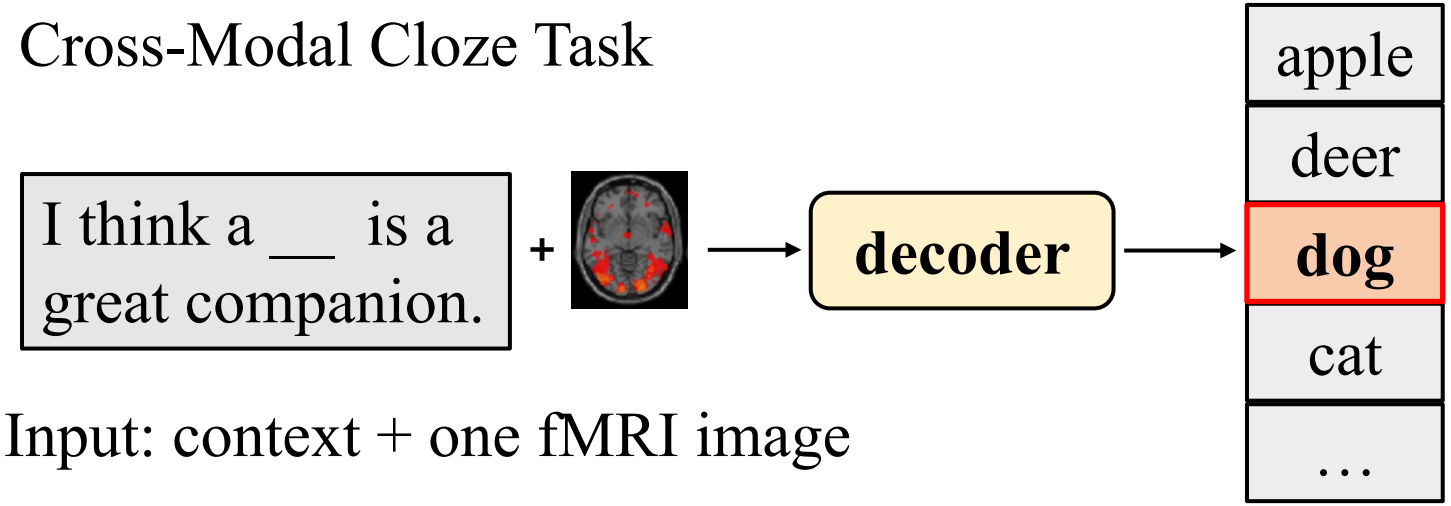
Outline

- **fMRI Encoding**
  - Framework
  - Probing Brain Activation Patterns by Dissociating Semantics and Syntax in Sentences
  - Probing Word Syntactic Representations in the Brain by a Feature Elimination Method
- **fMRI Decoding**
  - Framework
  - Cross-Modal Cloze Task: A New Task to Brain-to-Word Decoding
  - Towards Brain-to-Text Generation: Neural Decoding with Pre-trained Encoder-Decoder Models
- **Summary**

Cross-Modal Cloze Task: A New Task to Brain-to-Word Decoding

- With the help of pre-trained language encoding model, can we directly generate words from fMRI?

Cross-Modal Cloze Task

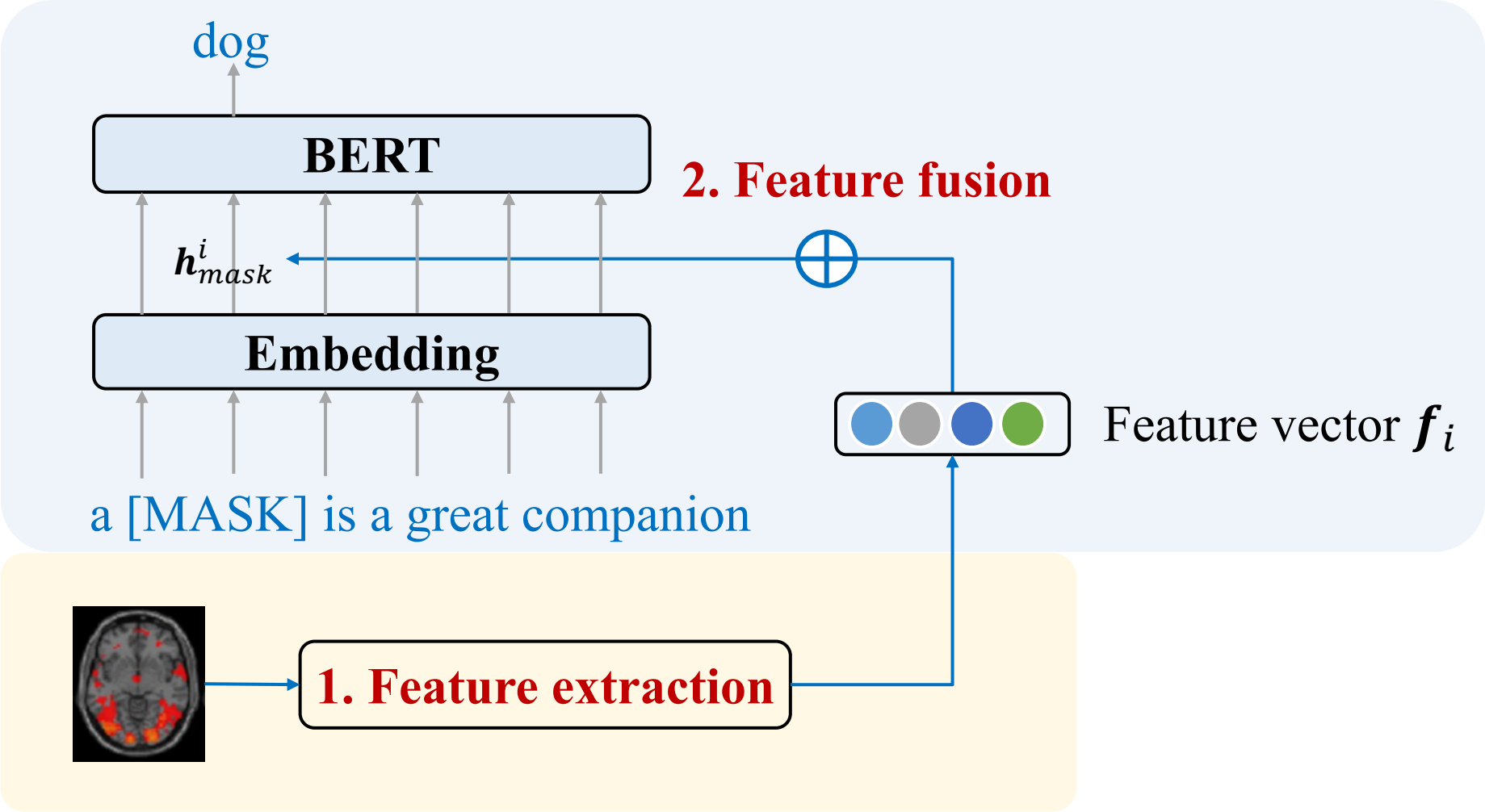


Input: context + one fMRI image

Output: target word

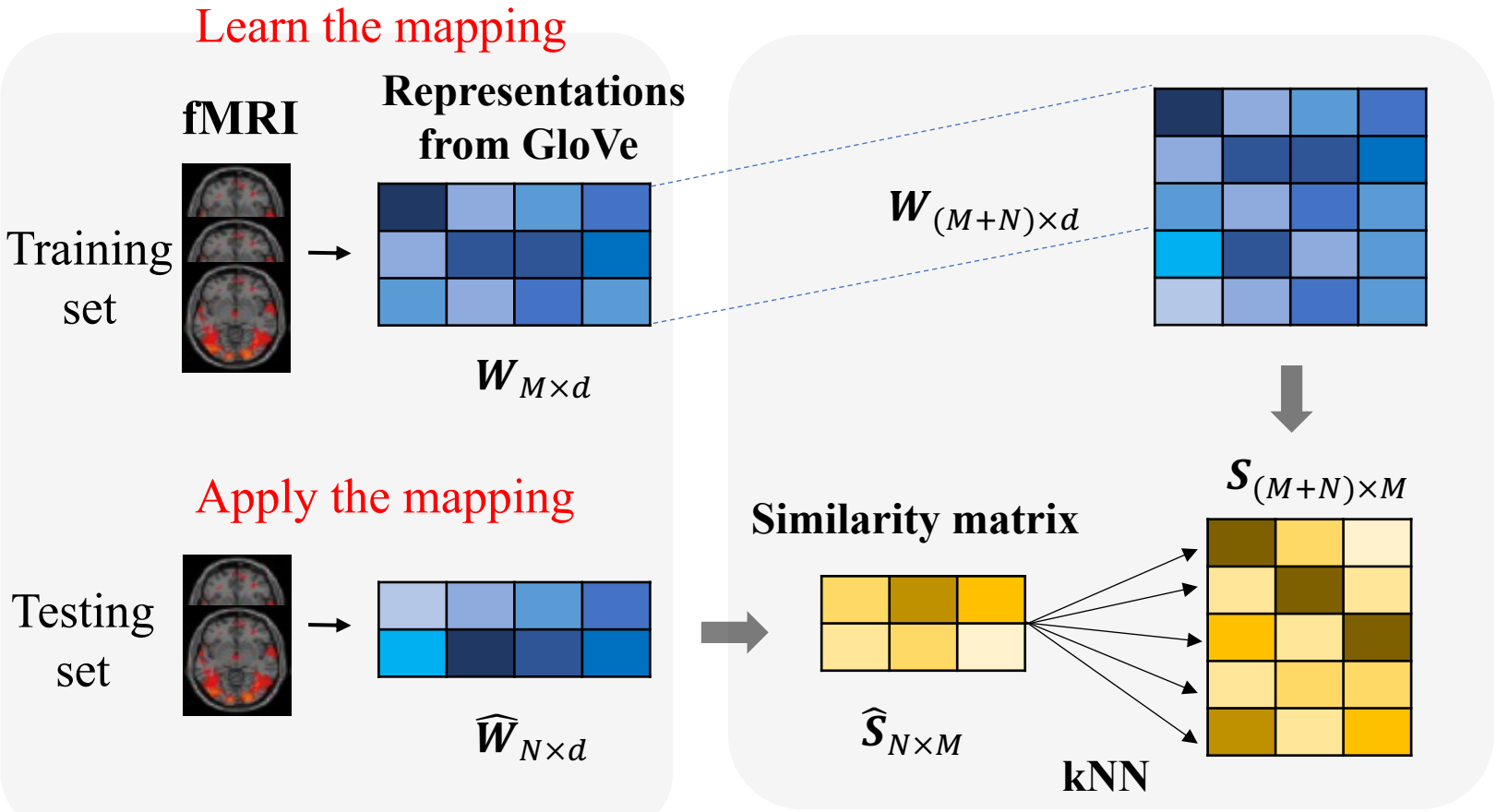
Evaluation: if the generated word is the target word, score.

Method Brain-word decoding with pretrained encoders



Method Feature extraction

1. Cross modal mapping 2. Representational similarity retrieval 3. Feature vectors

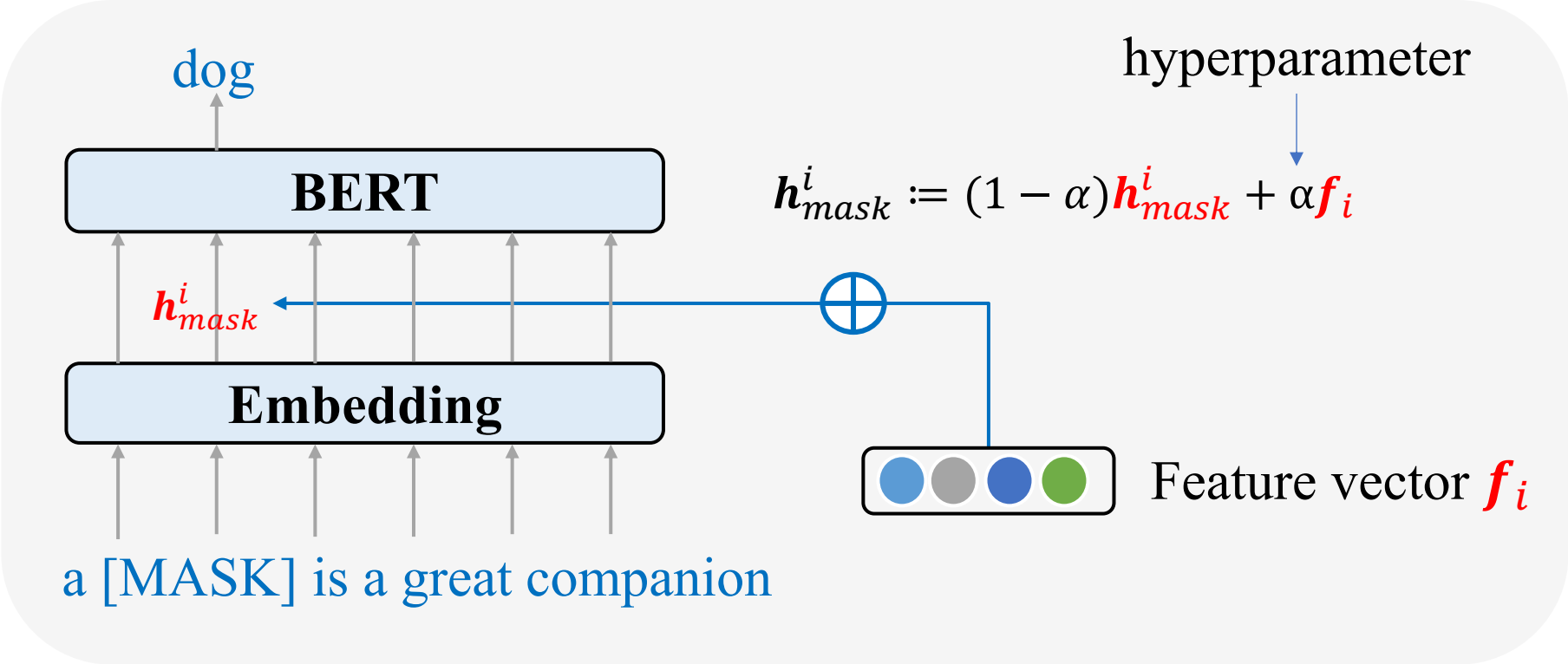


$$f_i = \frac{1}{k}\sum_{t=1}^{k} w'_{j_t}$$

- w' is embeddings from **BERT**
- $j_1, \ldots, j_k$ denotes the index of the top-k retrieved words
- k is hyperparameters

Calculate the similarity between the N predicted or (M+N) ground-truth embedding and all M words in the training set

Method Feature fusion



Dataset fMRI images

fMRI60 Mitchell et al. (2008)

- 9 participants
- 60 words

carrot, dog, hammer, igloo, skirt

fMRI180 Pereira et al. (2018)

- 15 participants
- 180 words

ability, big, damage, experiment, seafood

dog

A **dog** is a great companion.

gerbil cat
dog
chow
animal pet

Dataset Cross-modal cloze task ({fMRI, context}, target word)

Construct 6 contexts for each target word (each with 4-13 words, 7 words on average)

	Words	Context
fMRI60_CMC	carrot	the [MASK] is his favorite vegetable.
	hammer	she puts the [MASK] down on the ground.
fMRI180_CMC	ability	he has the [MASK] to cultivate creativity.
	damage	the accident left some serious [MASK].

	Participants	Sentences	Words	Context	fMRI
fMRI60_CMC	9	360	60	360	60
fMRI180_CMC	15	1080	180	1080	180

Results fMRI decoding on the cross-modal cloze task

(%)	fMRI60_CMC		fMRI180_CMC	
	Top-1 acc	Top-5 acc	Top-1 acc	Top-5 acc
BERT	27.50	45.56	26.02	50.93
BERT-Direct Fusion	27.78	49.54	26.51	51.78
BERT-Retri Fusion (random)	24.81	44.01	25.91	50.83
BERT-Retri Fusion	**31.08**(+3.58)	**55.99**(+10.43)	**27.60**(+1.59)	**53.11**(+2.19)

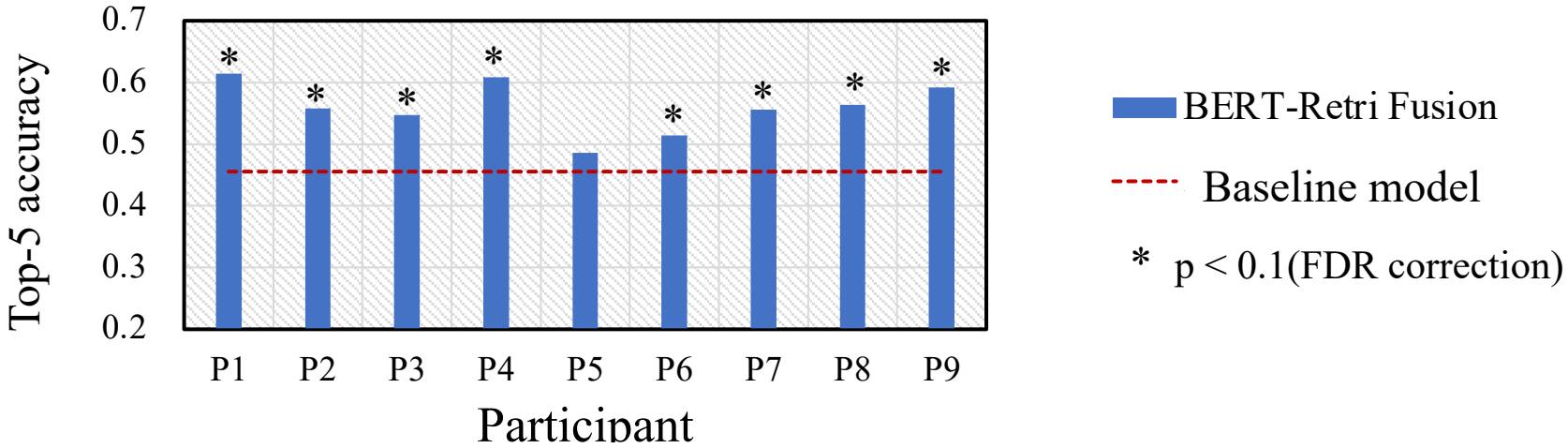
- fMRI feature vectors encode semantic information which could guide word prediction in BERT
- This task could serve as a bridge from decoding individual words to decoding continuous sentences, paving the way to build a practical neural language decoder.

Results fMRI decoding for single participants

fMRI60_CMC

Best participant:

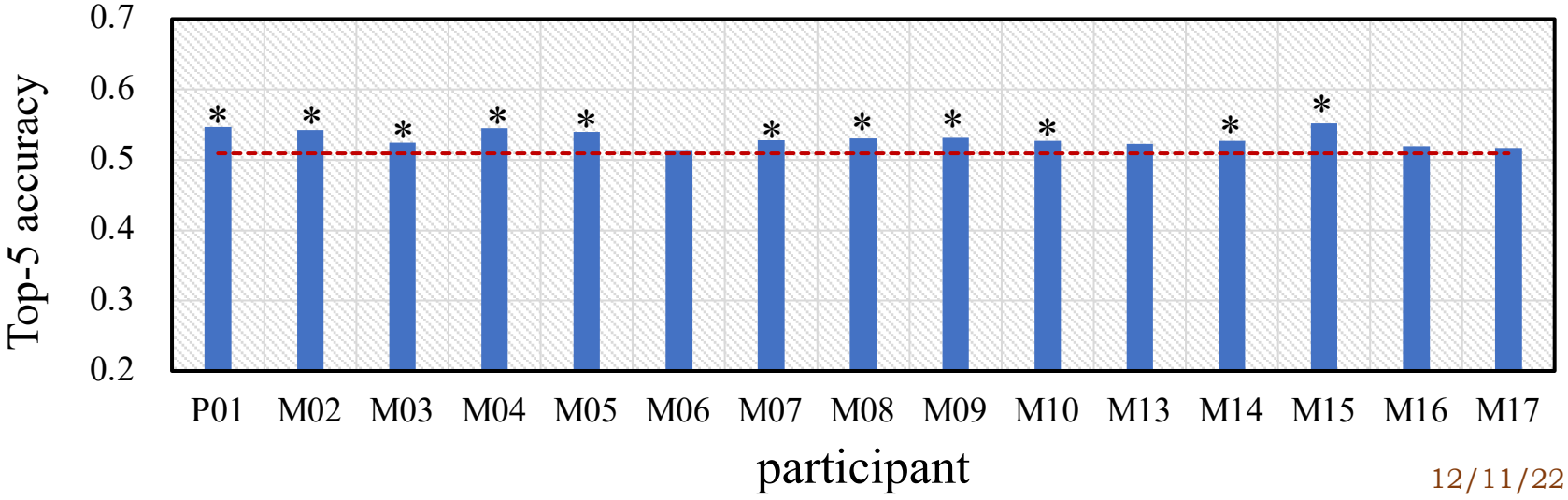
61.39% (+15.83)



fMRI180_CMC

Best participant:

55.19% (+4.26)



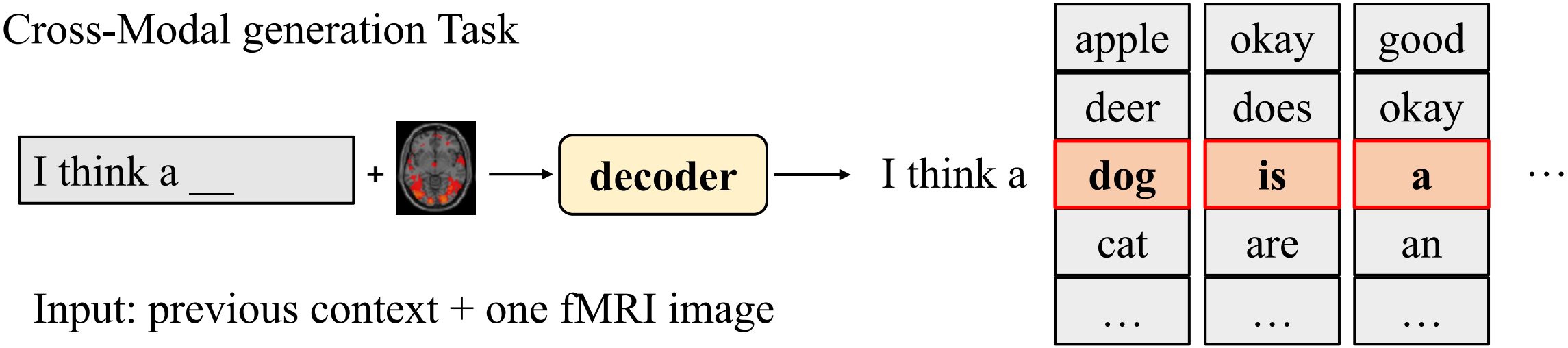
Outline

- **fMRI Encoding**
  - Framework
  - Probing Brain Activation Patterns by Dissociating Semantics and Syntax in Sentences
  - Probing Word Syntactic Representations in the Brain by a Feature Elimination Method
- **fMRI Decoding**
  - Framework
  - Cross-Modal Cloze Task: A New Task to Brain-to-Word Decoding
  - Towards Brain-to-Text Generation: Neural Decoding with Pre-trained Encoder-Decoder Models
- **Summary**

Towards Brain-to-Text Generation: Neural Decoding with Pre-trained Encoder-Decoder Models

- With the help of pre-trained language generation model, can we directly generate coherent text from fMRI?

Cross-Modal generation Task

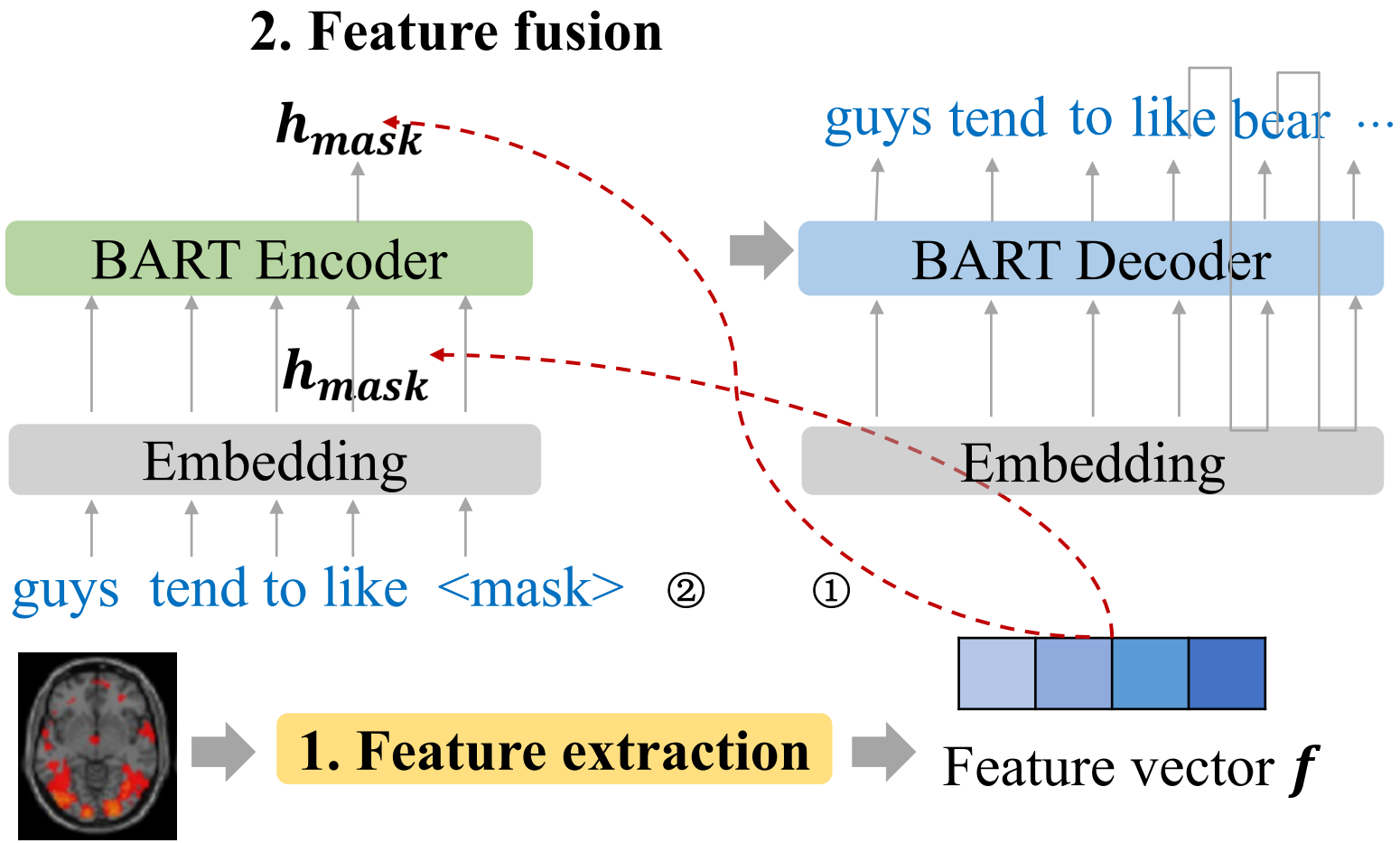


Input: previous context + one fMRI image

Output: one sentence that contains the target word

Evaluation: if the generated sentence contains the target word, score.

Method Neural decoding with pre-trained encoder-decoder models



1 feature extraction: cross-modal retrieval

$$f = \frac{1}{k}\sum_{t=1}^{k} w_{j_t}$$

2 feature fusion

$$h_{mask} := f$$

Where to fuse fMRI feature vector ?

Dataset Cross-modal generation task ({fMRI, previous context}, target word)

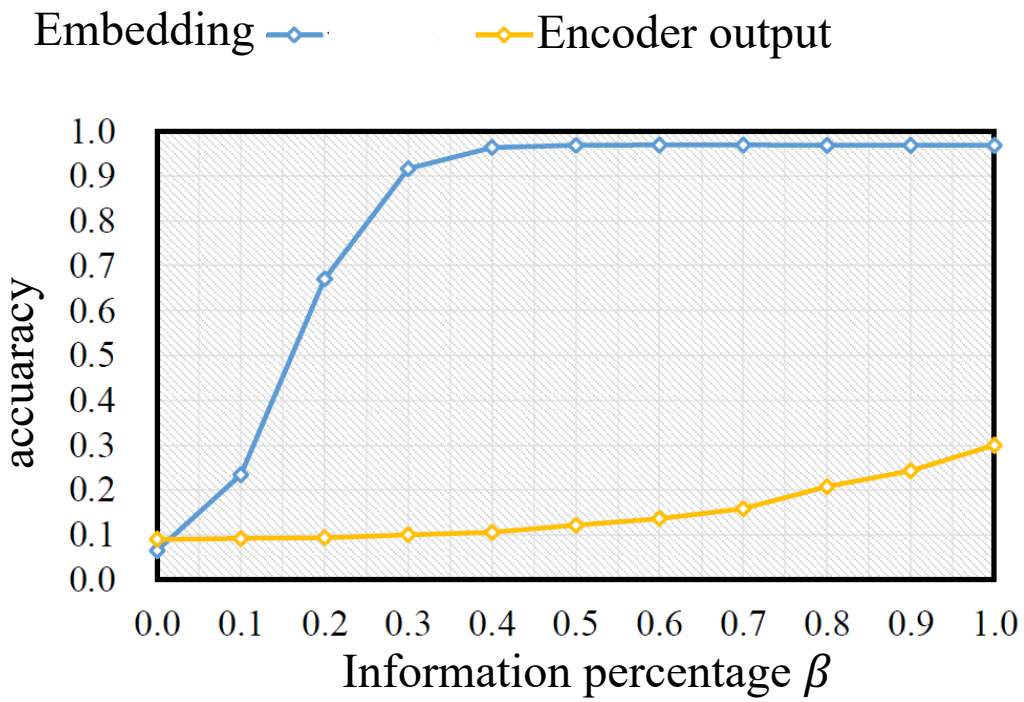
Construct 6 contexts for each target word and remove words after target

	Words	Sentence	Previous context
fMRI60_CMC	carrot	the **carrot** is his favorite vegetable.	the <MASK>
	hammer	she puts the **hammer** down on the ground.	she puts the <MASK>
fMRI180_CMC	ability	he has the **ability** to cultivate creativity.	he has the <MASK>
	damage	the accident left some serious **damage**.	The accident left some serious <MASK>

	Participants	Sentences	Words	Previous context	fMRI
fMRI60_CMC	9	360	60	360	60
fMRI180_CMC	15	1080	180	1080	180

Result Investigate the controllability of embedding and encoder output layer

	(%)	Acc	ΔAcc
BART		9.17	
Encoder	White noise	8.98	-0.19
	Ground truth	30.00	+20.83
	Wrong vector	8.98	-0.19
Embedding	White noise	6.48	-2.69
	Ground truth	96.85	+87.68
	Wrong vector	1.20	-7.97

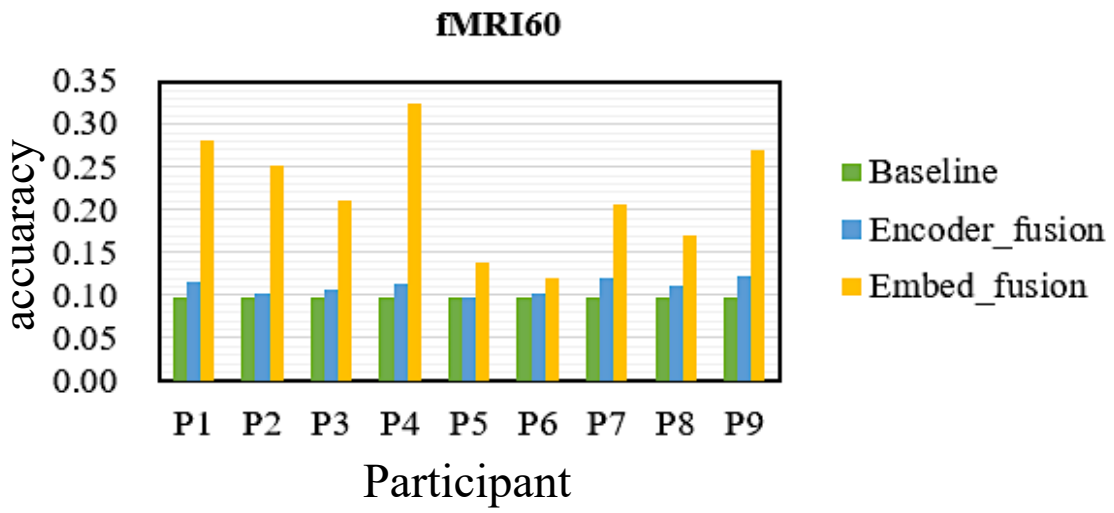


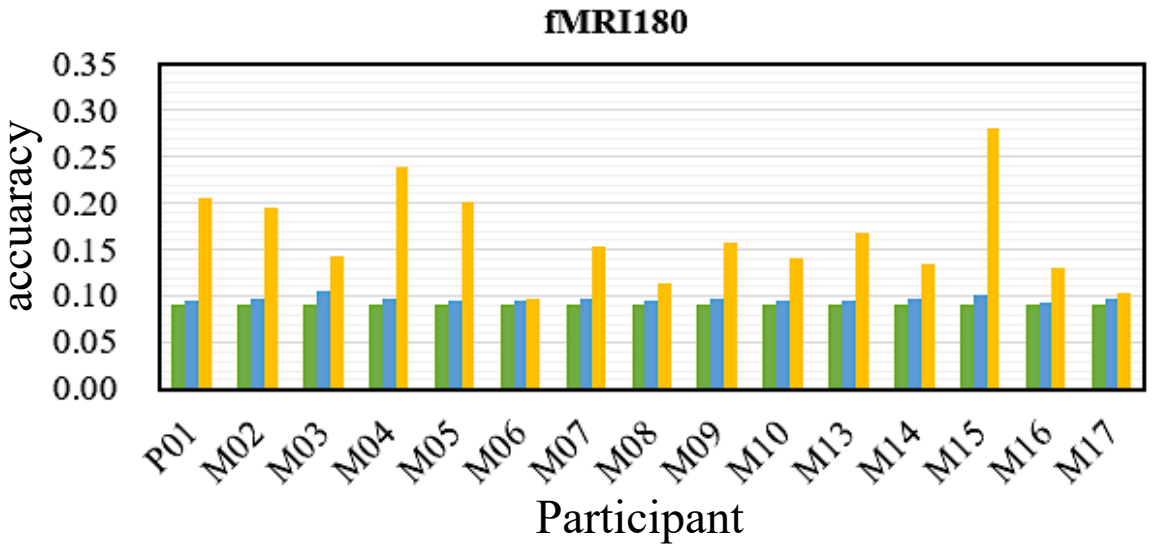
- Compared to Encoder output layer, the embedding layer is more controllable
- Compared to Encoder output layer, the embedding layer have low noise sensitivity

Result

fMRI decoding on the cross-modal generation task

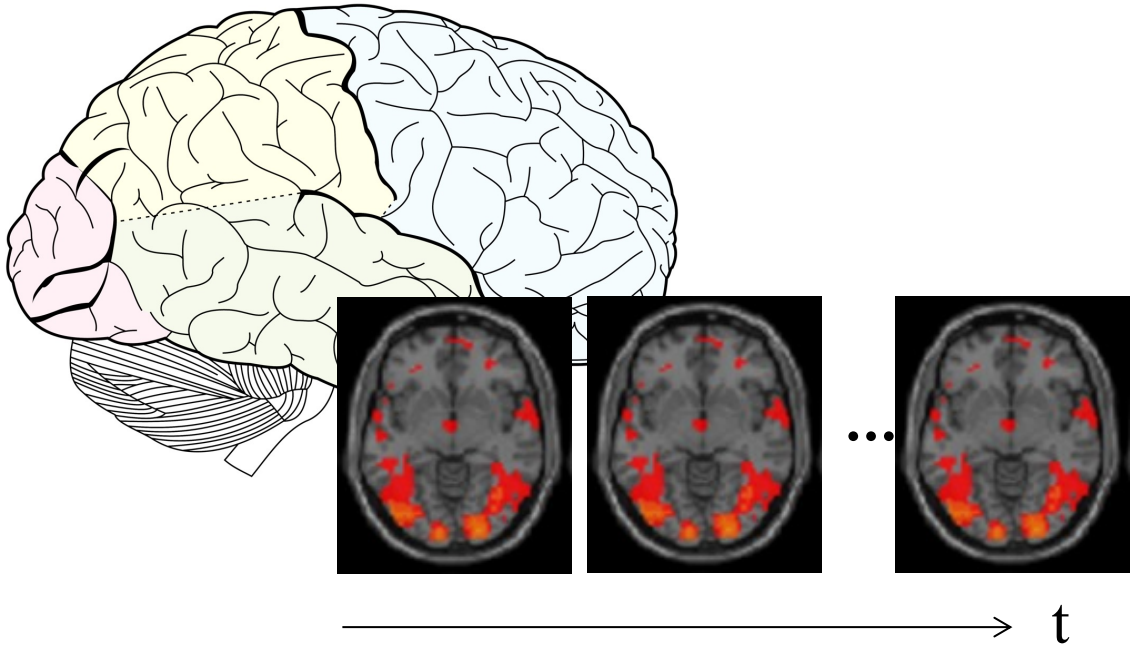
(%)	fMRI60		fMRI180	
	Acc	ΔAcc	Acc	ΔAcc
BART	9.72		9.17	
Encoder_fusion	11.02	+1.30	9.70	+0.53
Embed_fusion	**21.91**	**+12.19**	**16.45**	**+7.28**
Embed_fusion (random)	4.07	-5.65	3.56	-5.60





- BART embedding layer is the right place to fuse fMRI information
- fMRI data encode word semantic information which can guide text generation in BART
- It's possible to generate coherent text from fMRI

NEXT STEP: Real-time neural decoding



Decoder

What should I eat after this boring story listening experiment …

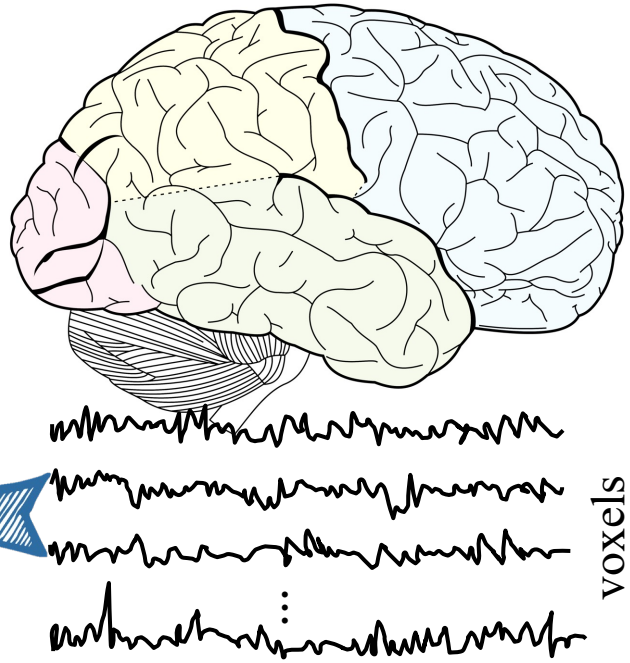
Outline

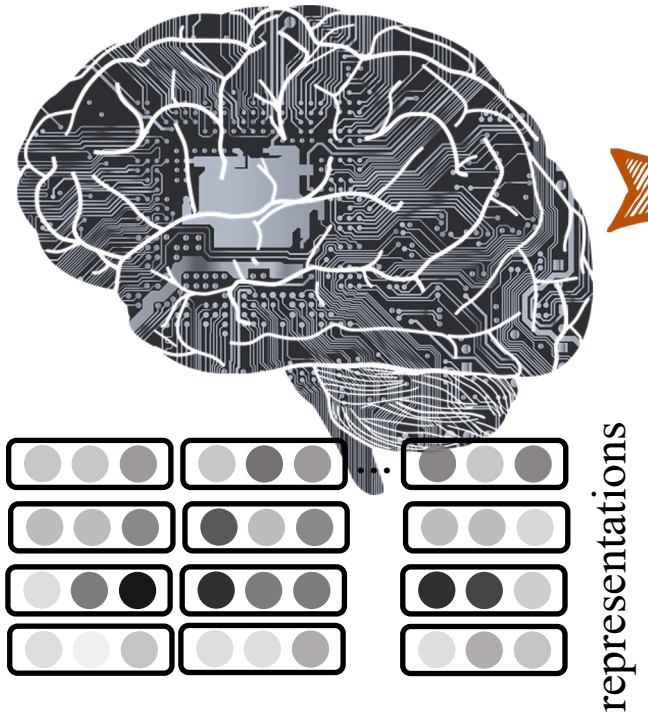
- **fMRI Encoding**
  - Framework
  - Probing Brain Activation Patterns by Dissociating Semantics and Syntax in Sentences
  - Probing Word Syntactic Representations in the Brain by a Feature Elimination Method
- **fMRI Decoding**
  - Framework
  - Cross-Modal Cloze Task: A New Task to Brain-to-Word Decoding
  - Towards Brain-to-Text Generation: Neural Decoding with Pre-trained Encoder-Decoder Models
- **Summary**

Towards Brain-to-Text Generation: neural decoding with pretrained language models

Neural Encoding

word meaning
composition
syntax
attention
memory
…





Neural Decoding

What should I eat after this boring story listening experiment
…

Probing brain activation patterns by using specialized computational representations

THANK YOU FOR LISTENING