

Brain Intelligence and Artificial Intelligence

人脑智能与机器智能

Lecture 4 – Language processing

Quanying Liu (刘泉影)

SUSTech, BME department

Email: [liuqy@sustech.edu.cn](mailto.liuqy@sustech.edu.cn)

Lecture 4 – Language processing

- **What is language**
- **The language system in brain**
- **Aphasia**
 - Broca's Aphasia
 - Wernicke's aphasia
- **Decoding language from the neural data**
 - ECoG-based decoding
 - EEG-based decoding
- **NLP in machines v.s. NLP in human brain**
 - NLP models
 - Representation of semantics
 - Brain-AI alignment
- **Linguistic stimuli to modulate the brain activity**

What is language?

Language is a system for **representing** and **communicating** information that uses words combined according to **grammatical rules**.

Language can be expressed in a variety of ways, including *gestures, writing, and speech*.

Speech is an audible form of communication built on the sounds humans produce.

Speech comes *naturally* to humans: Even with no formal training, children raised in a normal language environment will invariably learn to understand spoken language and to speak.

Reading and **writing**, on the other hand, require years of formal training, and more than 10% of the world's population is illiterate.

Is language unique to humans?

Do animals communicate with others?

Apparently, yes! Passerine birds attract mates with songs, bees code the distance and direction to honey by dancing, and monkeys signal a desire for sexual contact or fear at the approach of an enemy with coos and grunts.

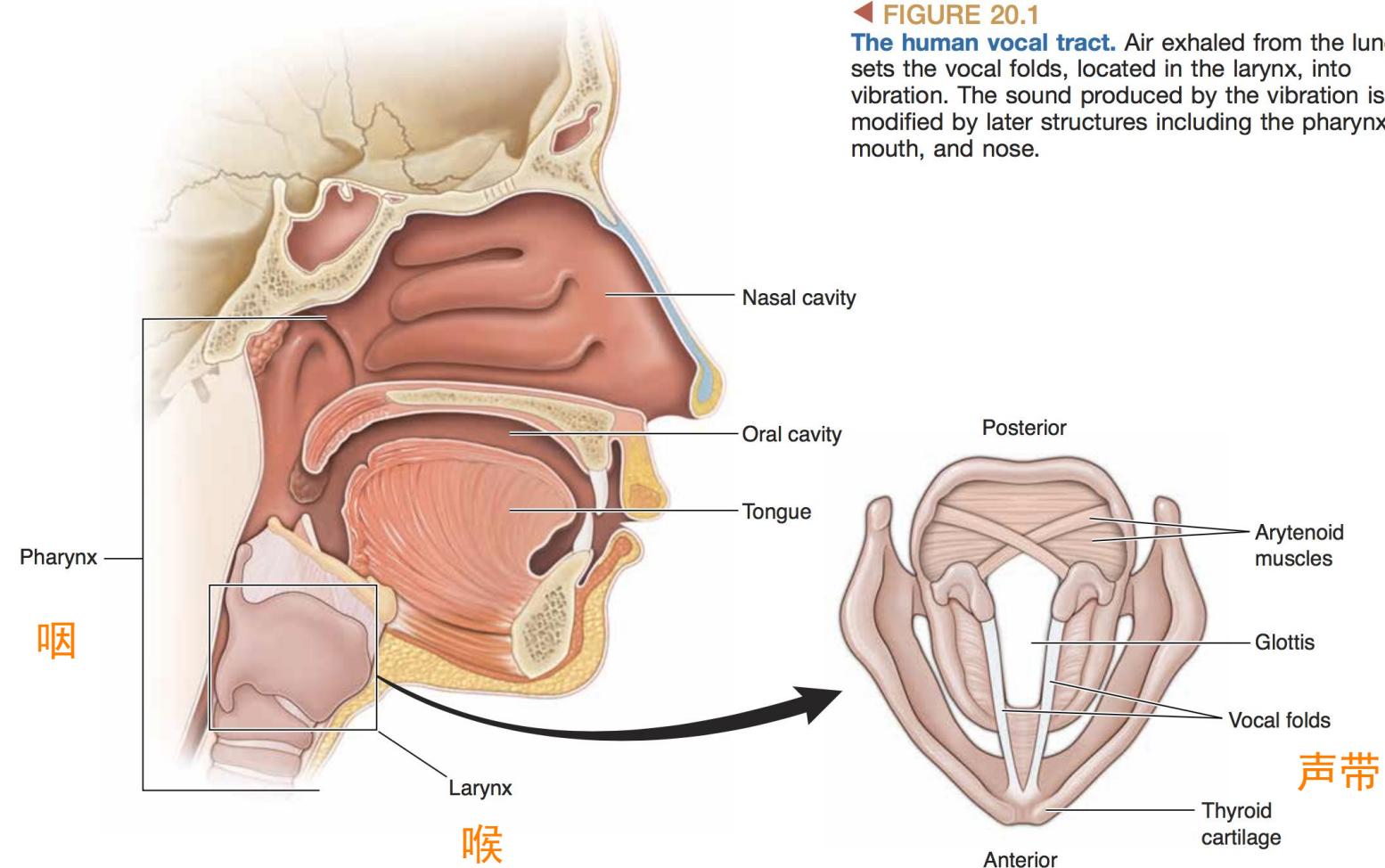
Do animals use language?

No. Compared to humans, non-human primates appear to have a very limited range of vocalizations.

Maybe Yes. In a recent study, scientists reported the results of analyzing thousands of gestures categorized into 66 types. Behavioral reactions in other chimpanzees viewing the gestures were used to infer the meaning and intended outcome of each gesture. Examples are “groom me,” “follow me,” and “stop that.” Some gestures appeared to have a particular intention, and others were used more flexibly. This is clearly a much more elaborate system of communication than the dance of a honey bee.

Human Sound and Speech Production

- Human speech involves a remarkable coordination of over **100** muscles ranging from those controlling the *lungs* to those of the *larynx* and the *mouth*.
- Question:
- Can we build a forward model from muscle activity to speech?



The language we speak impacts our thoughts

- Perception of masculine and feminine properties of inanimate objects
- Descriptions of color, time, and spatial location ...

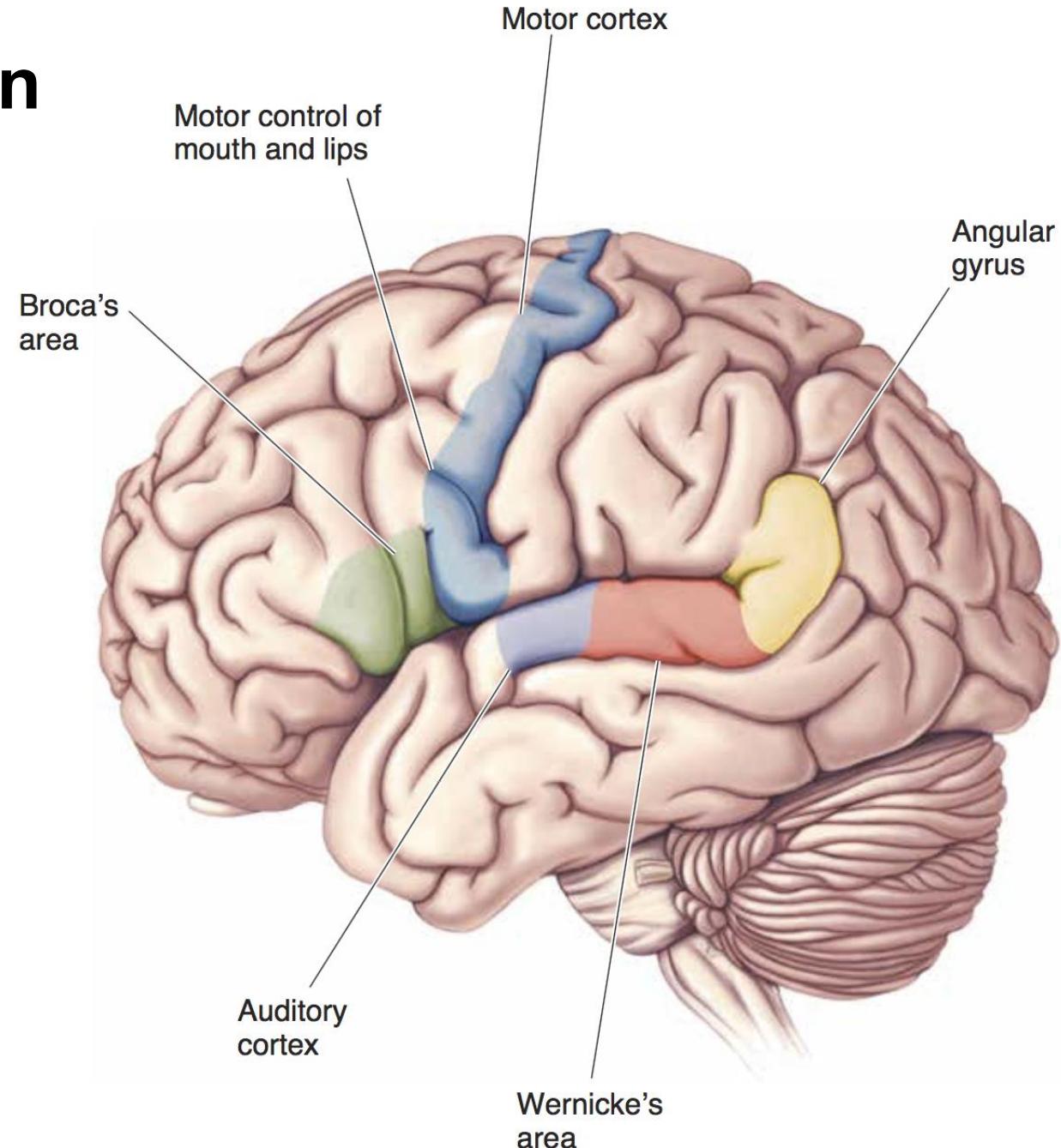


The language system in brain

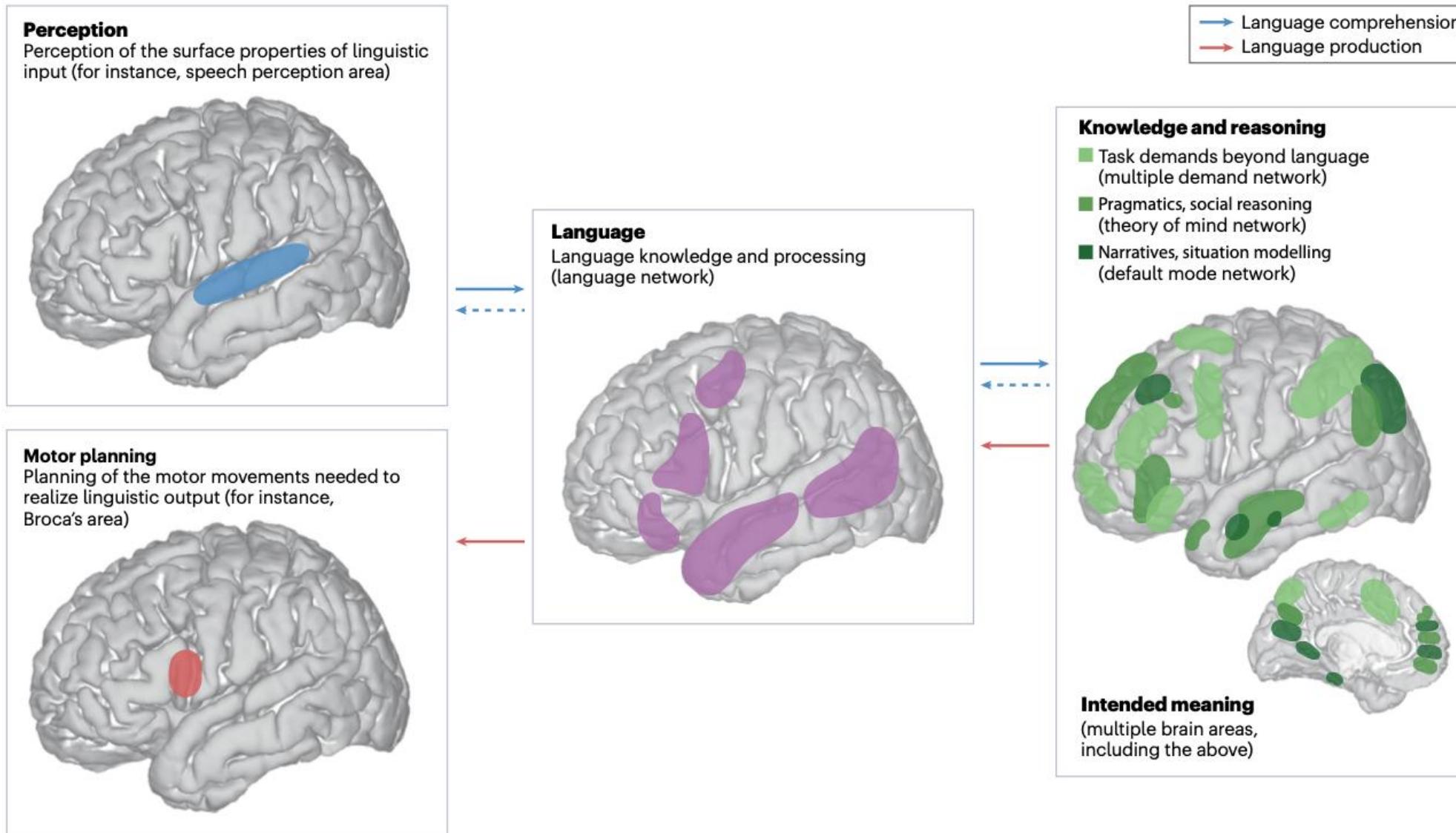
Key components of the language system in the left hemisphere.

In the frontal lobe, **Broca's area** lies next to the area that controls the mouth and lips in the motor cortex.

Wernicke's area, on the superior surface of the temporal lobe, is situated between the auditory cortex and the angular gyrus.



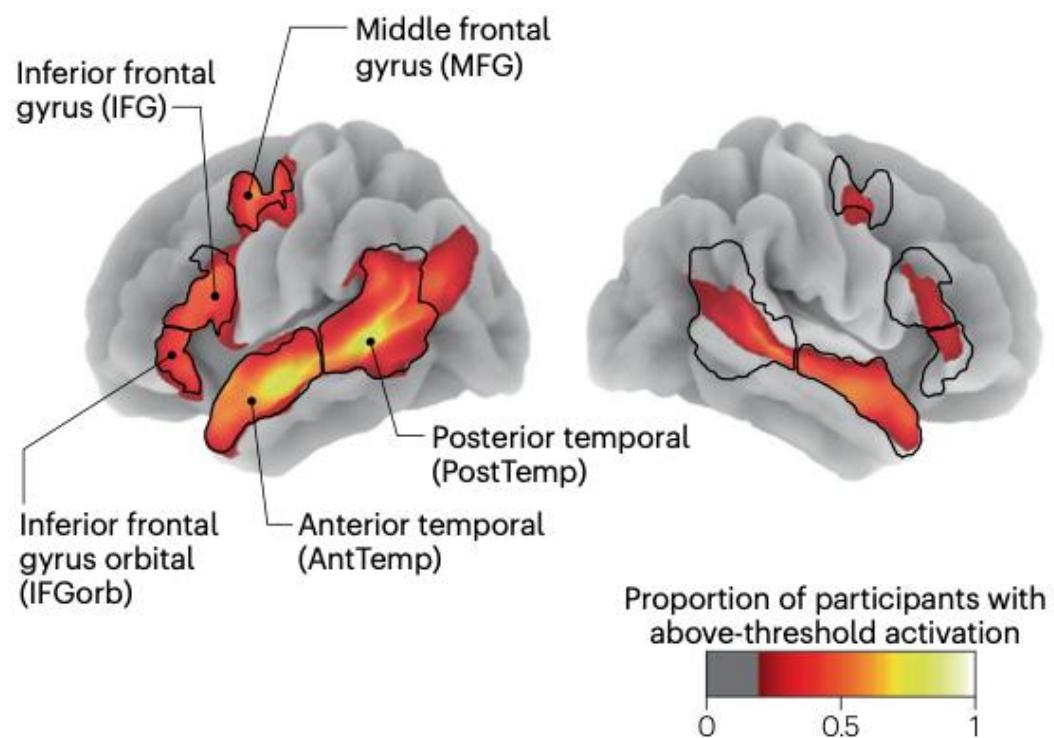
Brain systems of language comprehension & production



Brain systems of language comprehension & production

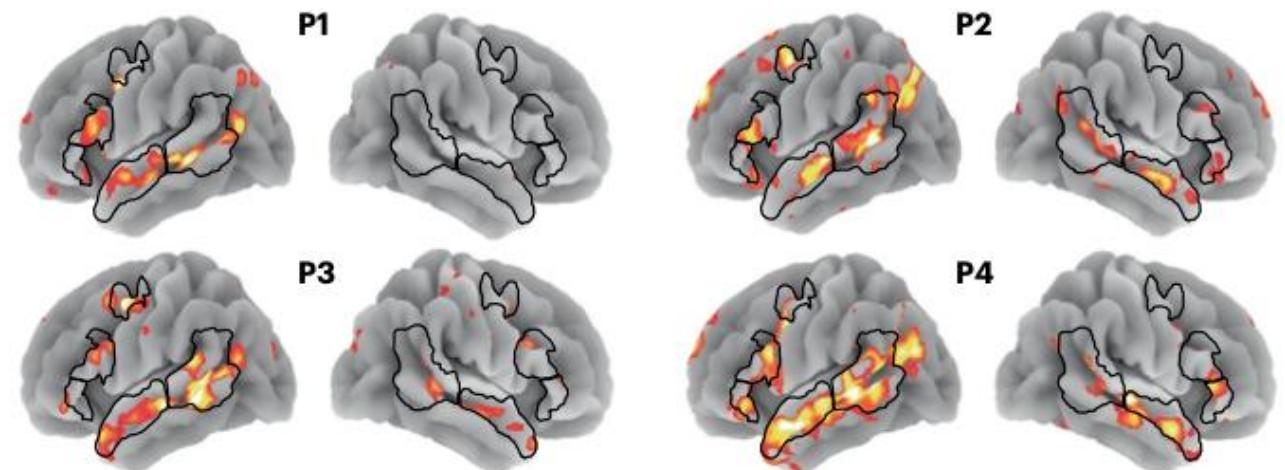
a Probabilistic map ($n = 806$)

Sentences > non-words

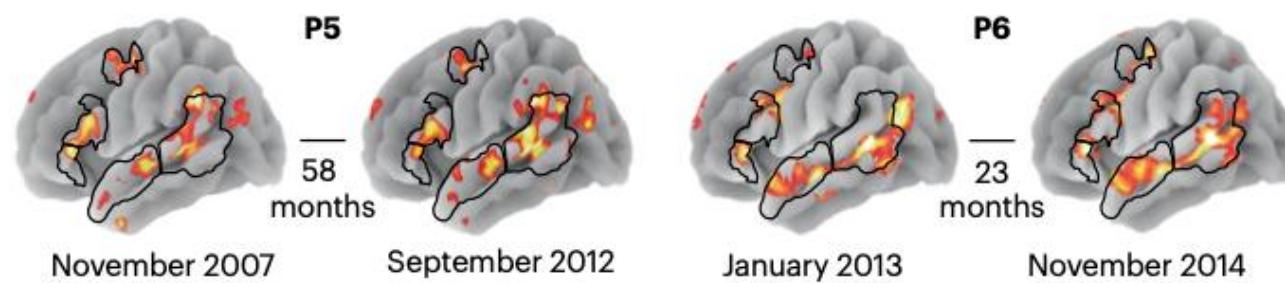


b Individual activation maps ($n = 4$)

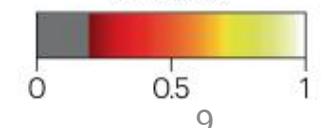
Sentences > non-words



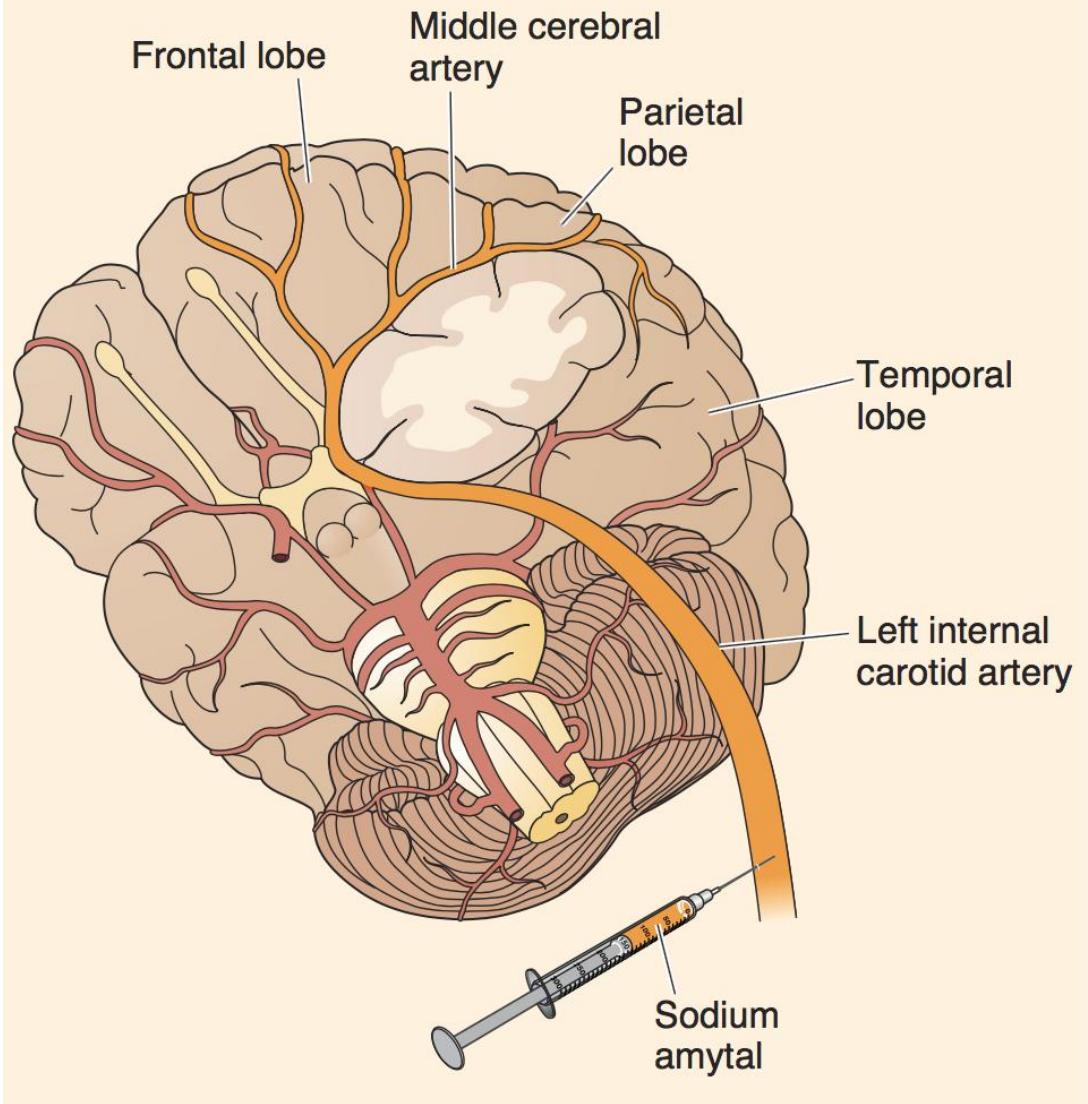
c Activation maps are stable across time ($n = 2$)



t values



Assessing Hemispheric Language Dominance



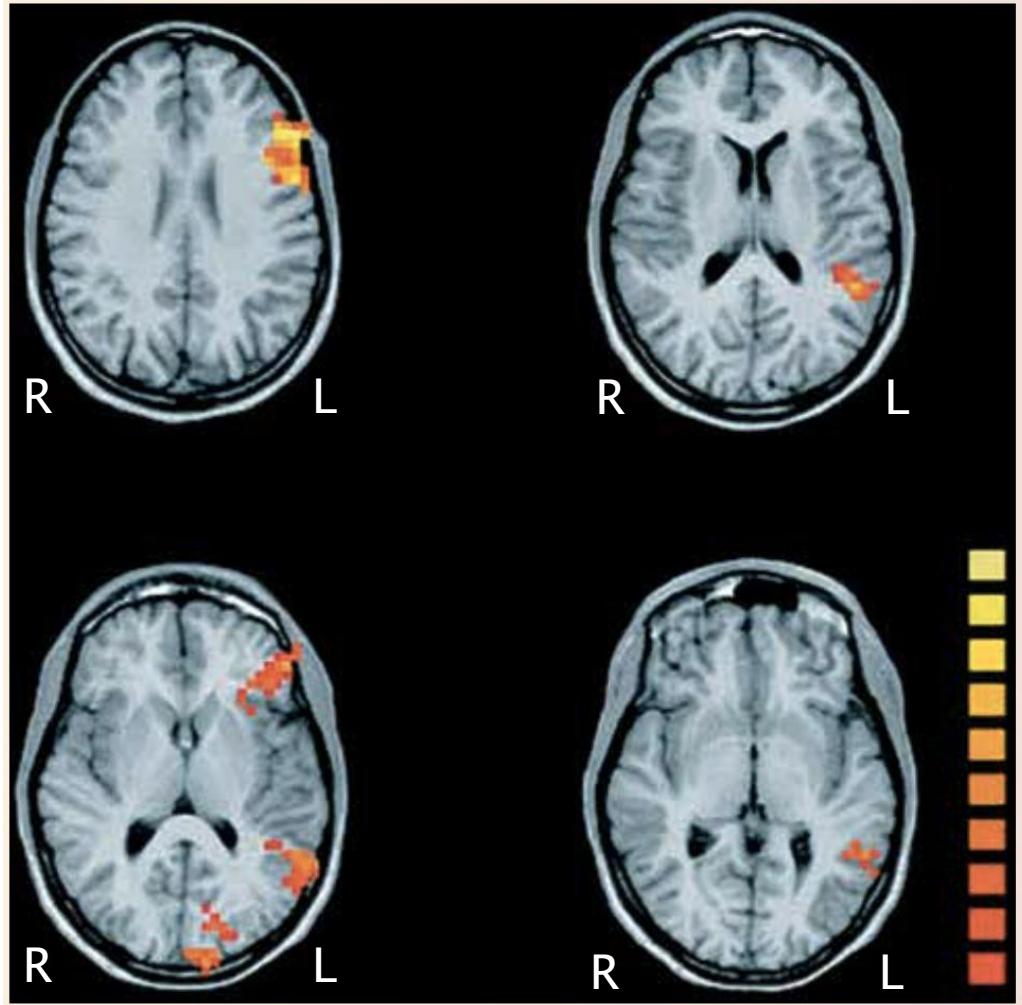
Wada procedure

A fast-acting barbiturate, such as sodium amytal, is injected into the carotid artery on one side of the neck.

The drug is preferentially carried in the bloodstream to the hemisphere ipsilateral to the injection, where it acts as an anesthetic for about 10 minutes.

The effects are sudden and dramatic. Within a matter of **seconds**, the limbs on the side of the body contralateral to the injection become paralyzed along with loss of somatic sensation.

Assessing Hemispheric Language Dominance



fMRI technique

The brain images in Figure were collected while a subject was given a word and asked to select a synonym from four word options.

The brain scans show that frontal, temporal, and parietal areas are activated exclusively in the **left** hemisphere, which is thus dominant for language in this person.

TABLE A **Hemispheric Control of Speech in Relation to Handedness**

Handedness	Number of Cases	Speech Representation (%)		
		Left	Bilateral	Right
Right	140	96	0	4
Left	122	70	15	15

(Source: Rasmussen and Milner, 1977, Table 1.)

Aphasia

Aphasia is the partial or complete loss of language abilities following brain damage, often [without](#) the loss of cognitive faculties or the ability to move the muscles used in speech.

TABLE 20.1 **Characteristics of Types of Aphasia**

Type of Aphasia	Site of Brain Damage	Comprehension	Speech	Impaired Repetition	Paraphasic Errors
Broca's	Motor association cortex of frontal lobe	Good	Nonfluent, agrammatical	Yes	Yes
Wernicke's	Posterior temporal lobe	Poor	Fluent, grammatical, meaningless	Yes	Yes
Conduction Global	Arcuate fasciculus Portions of temporal and frontal lobes	Good Poor	Fluent, grammatical Very little	Yes Yes	Yes —
Transcortical motor area	Frontal lobe anterior to Broca's	Good	Nonfluent, agrammatical	No	Yes
Transcortical sensory	Cortex near the junction of temporal, parietal, and occipital lobes	Poor	Fluent, grammatical, meaningless	No	Yes
Anomic	Inferior temporal lobe	Good	Fluent, grammatical	No	

Broca's Aphasia

- The person has difficulty speaking even though he or she can understand language heard or read.
- People with Broca's aphasia have difficulty saying anything, often pausing to search for the right word. The inability to find words is called **anomia** (literally meaning “no name”).
- Broca aphasics have difficulty repeating things spoken to them.

“I asked Mr. Ford about his work before he entered the hospital.

I'm a sig. . . no. . . man. . . uh, well, . . . again.” These words were emitted slowly, and with great effort. The sounds were not clearly articulated; each syllable was uttered harshly, explosively, in a throaty voice. With practice, it was possible to understand him, but at first I encountered considerable difficulty in this.

“Let me help you,” I interjected. “You were a signal. . .”

“A signal man. . . right,” Ford completed my phrase triumphantly.

“Were you in the Coast Guard?”

“No, er, yes, yes. . . ship. . . Massachu. . . chusetts. . . Coastguard. . . years.”

He raised his hands twice, indicating the number “nineteen.”

“Could you tell me, Mr. Ford, what you've been doing in the hospital?”

“Yes, sure. Me go, er, uh, P.T. nine o'cot, speech. . . two times. . . read. . . wr. . . ripe, er, rike, er, write. . . practice. . . get-ting better.”

“And have you been going home on weekends?”

“Why, yes. . . Thursday, er, er, er, no, er, Friday. . . Bar-ba-ra. . . wife. . . and, oh, car. . . drive. . . purnpoke. . . you know. . . rest and. . . tee-vee.”

“Are you able to understand everything on television?”

“Oh, yes, yes. . . well. . . almost.” Ford grinned a bit. (Gardner, 1974, pp. 60–61)

Wernicke's aphasia

- In Wernicke's aphasia, speech is fluent but **comprehension** is poor.
- Insight about the possible function of Wernicke's area is provided by its location on the superior temporal gyrus near the primary auditory cortex.
- Wernicke's area may play a critical role in relating incoming sounds to their **meaning**. In other words, it is an area specialized for **storing memories of the sounds that make up words**.

“What brings you to the hospital?” I asked the 72-year-old retired butcher 4 weeks after his admission to the hospital.

“Boy, I’m sweating, I’m awful nervous, you know, once in a while I get caught up, I can’t mention the tarripoi, a month ago, quite a little, I’ve done a lot well, I impose a lot, while, on the other hand, you know what I mean, I have to run around, look it over, trebbin and all that sort of stuff.”

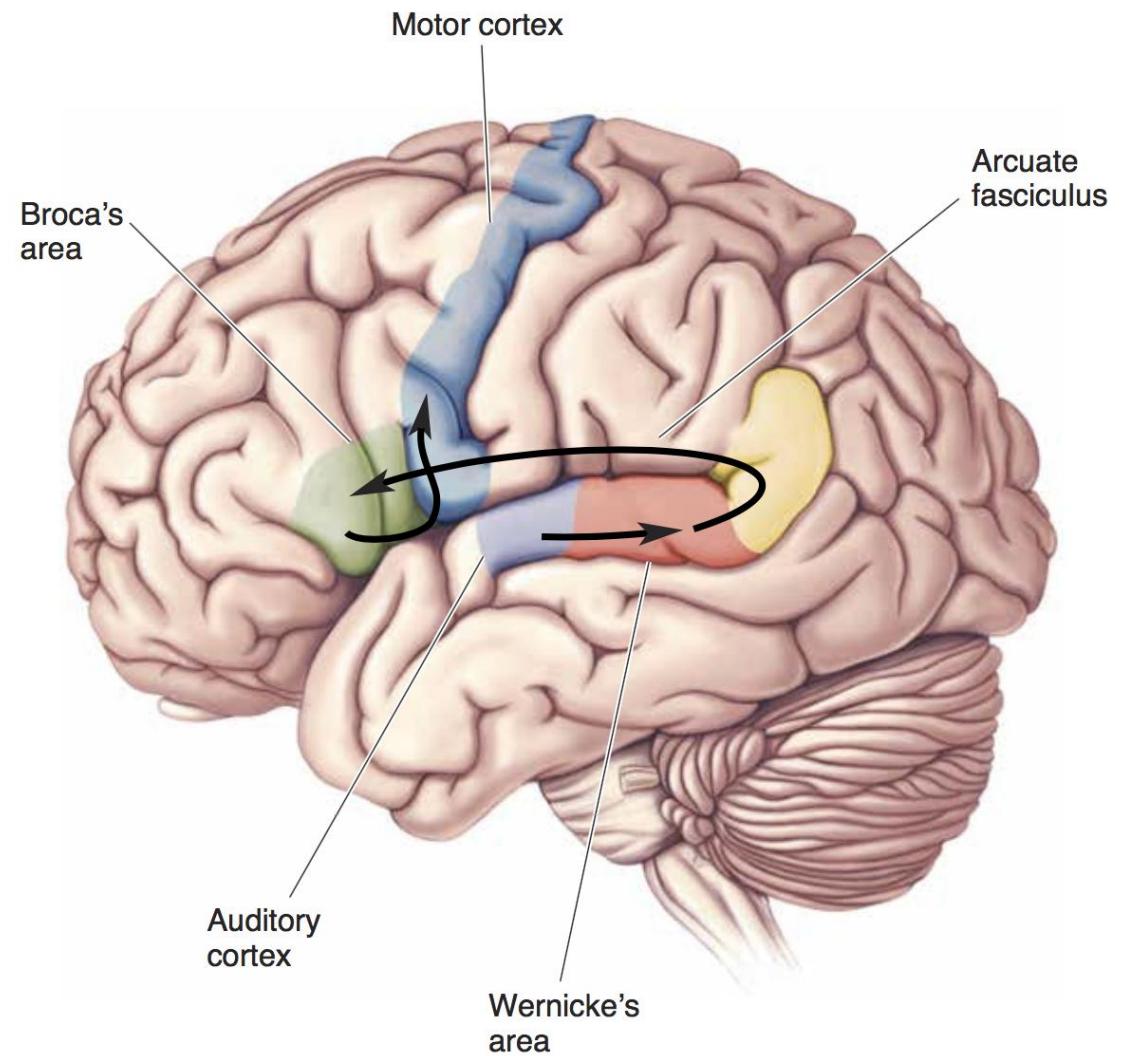
I attempted several times to break in, but was unable to do so against this relentlessly steady and rapid outflow. Finally, I put up my hand, rested it on Gorgan’s shoulder, and was able to gain a moment’s reprieve.

“Thank you, Mr. Gorgan. I want to ask you a few—”

“Oh sure, go ahead, any old think you want. If I could I would. Oh, I’m taking the word the wrong way to say, all of the barbers here whenever they stop you it’s going around and around, if you know what I mean, that is tying and tying for repucer, repuceration, well, we were trying the best that we could while another time it was with the beds over there the same thing. . . .” (Gardner, 1974, pp. 67–68)

Wernicke–Geschwind model of Language and Aphasia

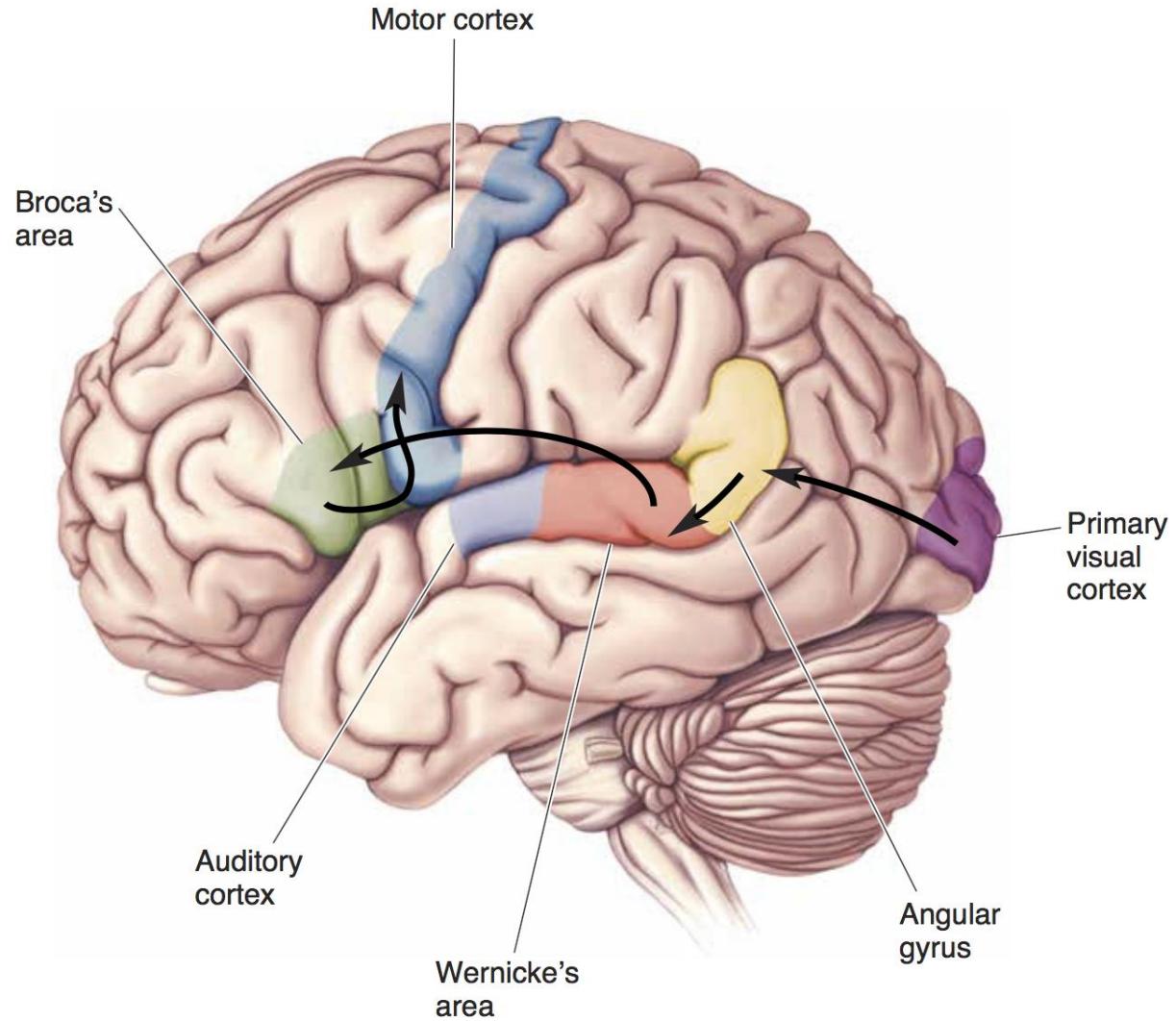
- Wernicke proposed a model for language processing in the brain. Later extended by Norman Geschwind at Boston University.
- This model is known as the **Wernicke–Geschwind model**.
- The key anatomical elements in the system are **Broca's area**, **Wernicke's area**, the **arcuate fasciculus** (a bundle of axons connecting the two cortical areas), and the **angular gyrus**.
- Two tasks: 1) repetition of spoken words; 2) reading written text aloud



The pathway involved in repeating a spoken word

Wernicke–Geschwind model of Language and Aphasia

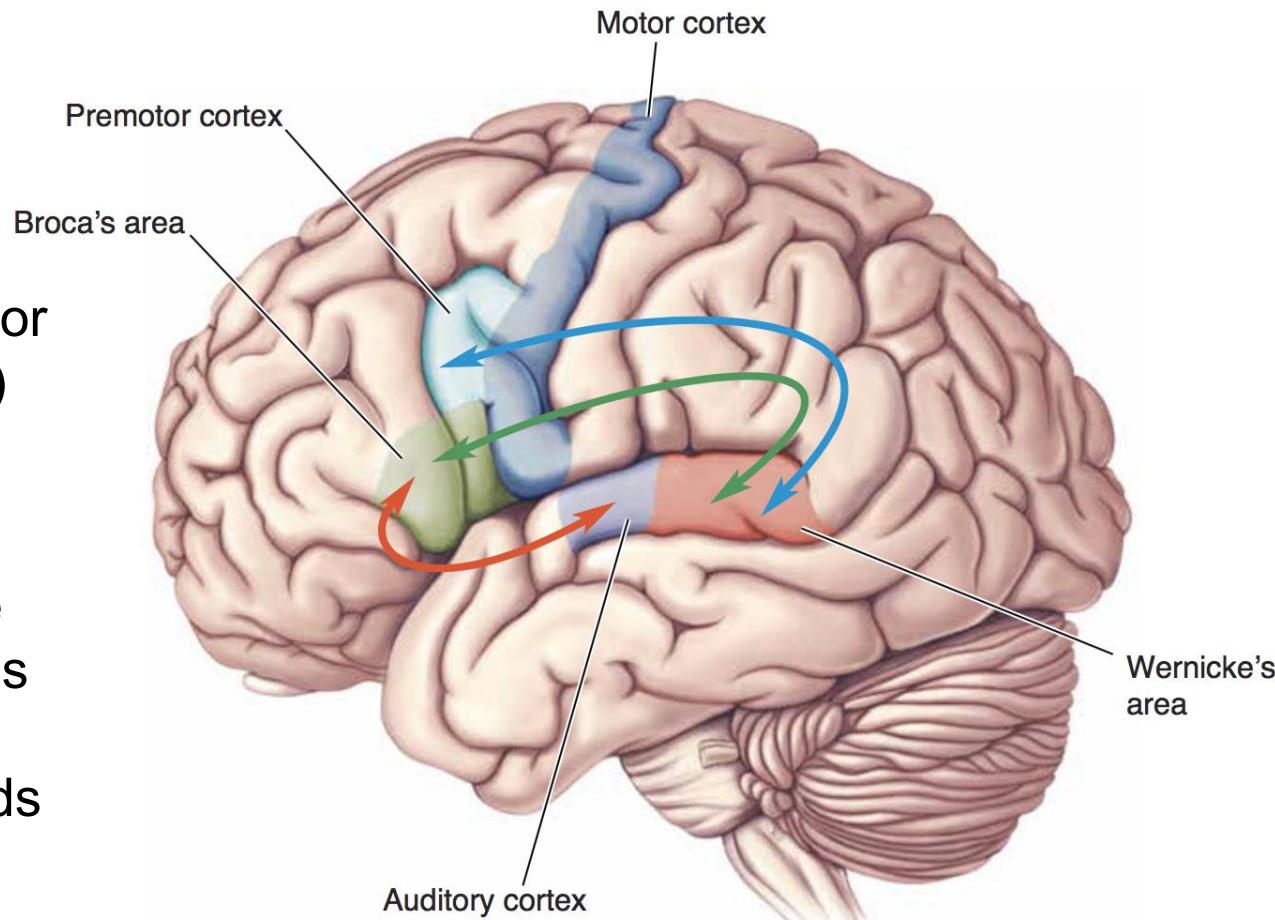
- Task 2: reading written text aloud
- The incoming information is processed by the **visual system** through the striate cortex and higher-order visual cortical areas.
- The visual signals are then passed to the **angular gyrus** at the junction of the occipital, parietal, and temporal lobes. In the cortex of the angular gyrus, it is assumed that a transformation occurs so that the output evokes the same pattern of activity in **Wernicke's area** as if the words were spoken rather than written.
- From this point, the processing follows the same progression as in the first example: Wernicke's area to Broca's area to motor cortex.



The pathway involved in repeating aloud a written word

Parallel language pathways in current model

- Current models of language processing emphasize **multiple streams of processing**.
- The model here includes **two** dorsal pathways and **one** ventral pathway.
- One dorsal pathway (**blue**) connects the superior temporal gyrus (Wernicke's and auditory areas) with premotor cortex and is involved in **speech production and repeating words**.
- The other dorsal pathway (**green**) connects the superior temporal gyrus with Broca's area and is thought to be involved in **processing complex syntactic structure**—that is, the analysis of words arranged according to a grammar system.
- The ventral stream (**red**) takes the sounds of speech and **extracts their meaning**.



Berwick et al., 2013

Aphasia – characteristics

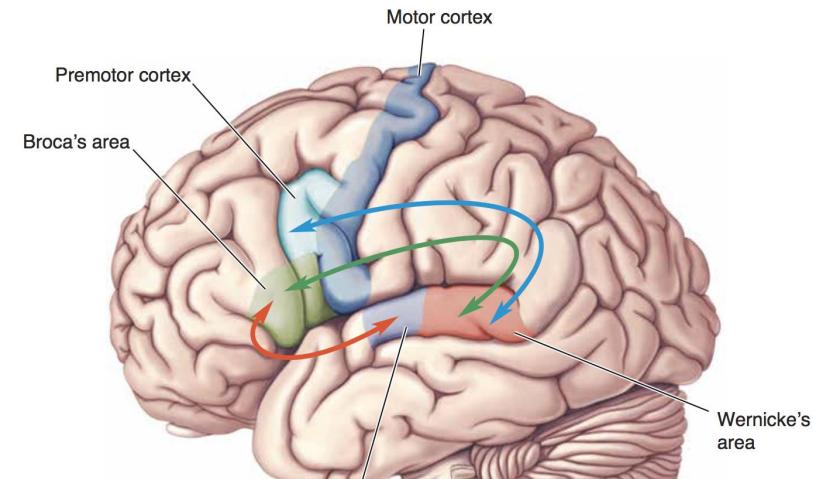
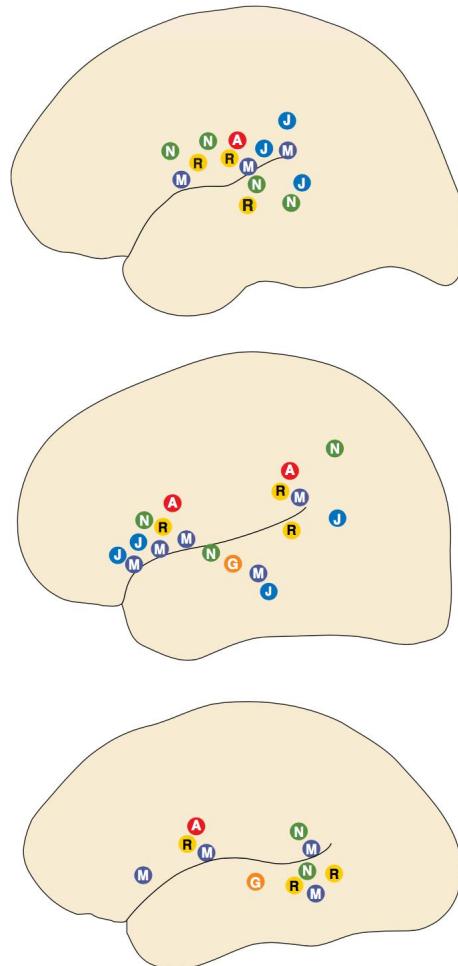


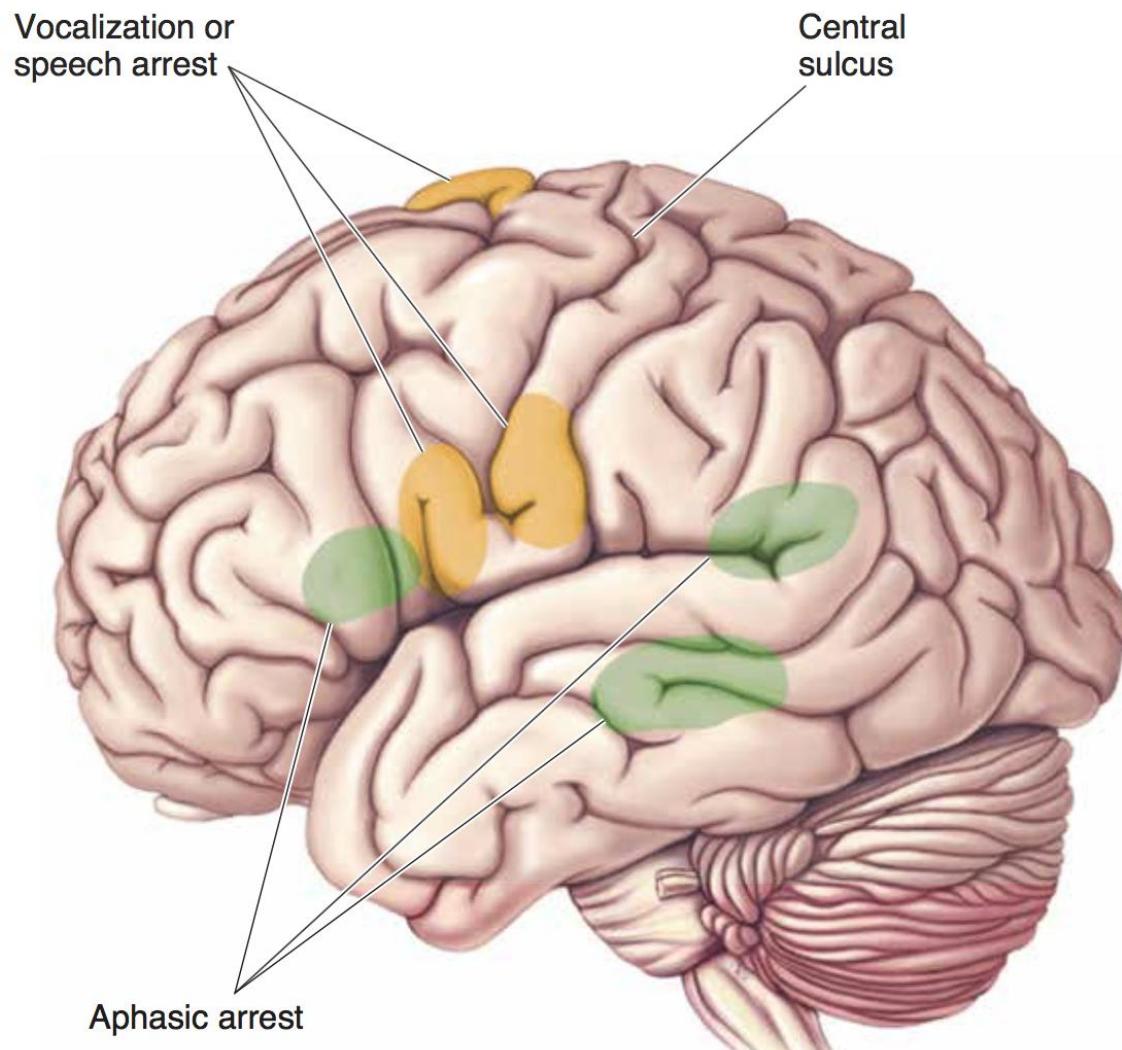
TABLE 20.1 Characteristics of Types of Aphasia

Type of Aphasia	Site of Brain Damage	Comprehension	Speech	Impaired Repetition	Paraphasic Errors
Broca's	Motor association cortex of frontal lobe	Good	Nonfluent, agrammatical	Yes	Yes
Wernicke's	Posterior temporal lobe	Poor	Fluent, grammatical, meaningless	Yes	Yes
Conduction	Arcuate fasciculus	Good	Fluent, grammatical	Yes	Yes
Global	Portions of temporal and frontal lobes	Poor	Very little	Yes	—
Transcortical motor area	Frontal lobe anterior to Broca's	Good	Nonfluent, agrammatical	No	Yes
Transcortical sensory	Cortex near the junction of temporal, parietal, and occipital lobes	Poor	Fluent, grammatical, meaningless	No	Yes
Anomic	Inferior temporal lobe	Good	Fluent, grammatical	No	

The Effects of Brain Stimulation on Language



▲ FIGURE 20.19
The effects of brain stimulation in three patients being treated for epilepsy.

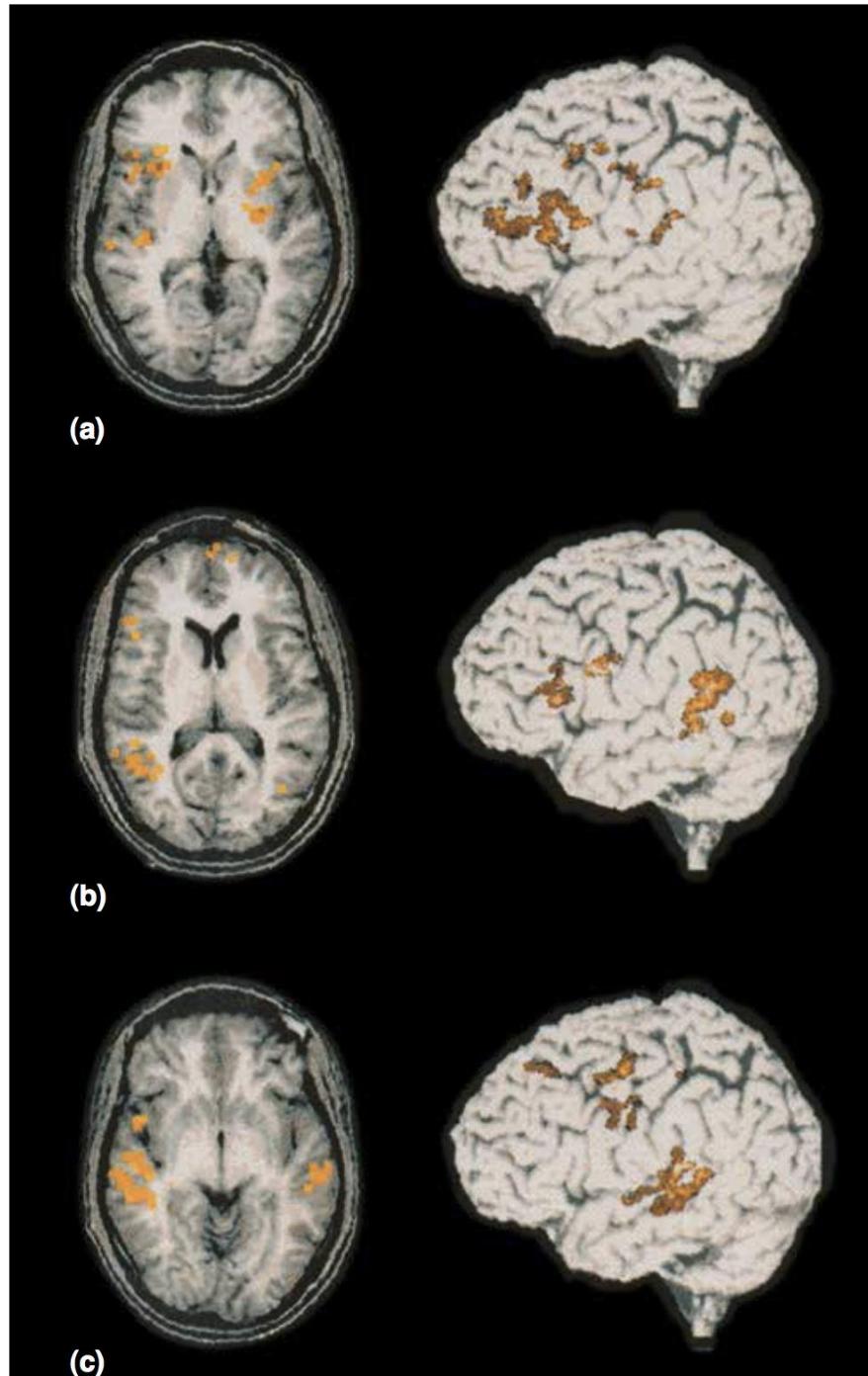


- Stimulation of **motor cortex** causes vocalizations or speech arrest by activating facial muscles.
- At **other sites**, stimulation causes an aphasic arrest in which language is agrammatical or anomia is observed.

Sites where electrical brain stimulation affects language ¹⁰

fMRI study on language processing

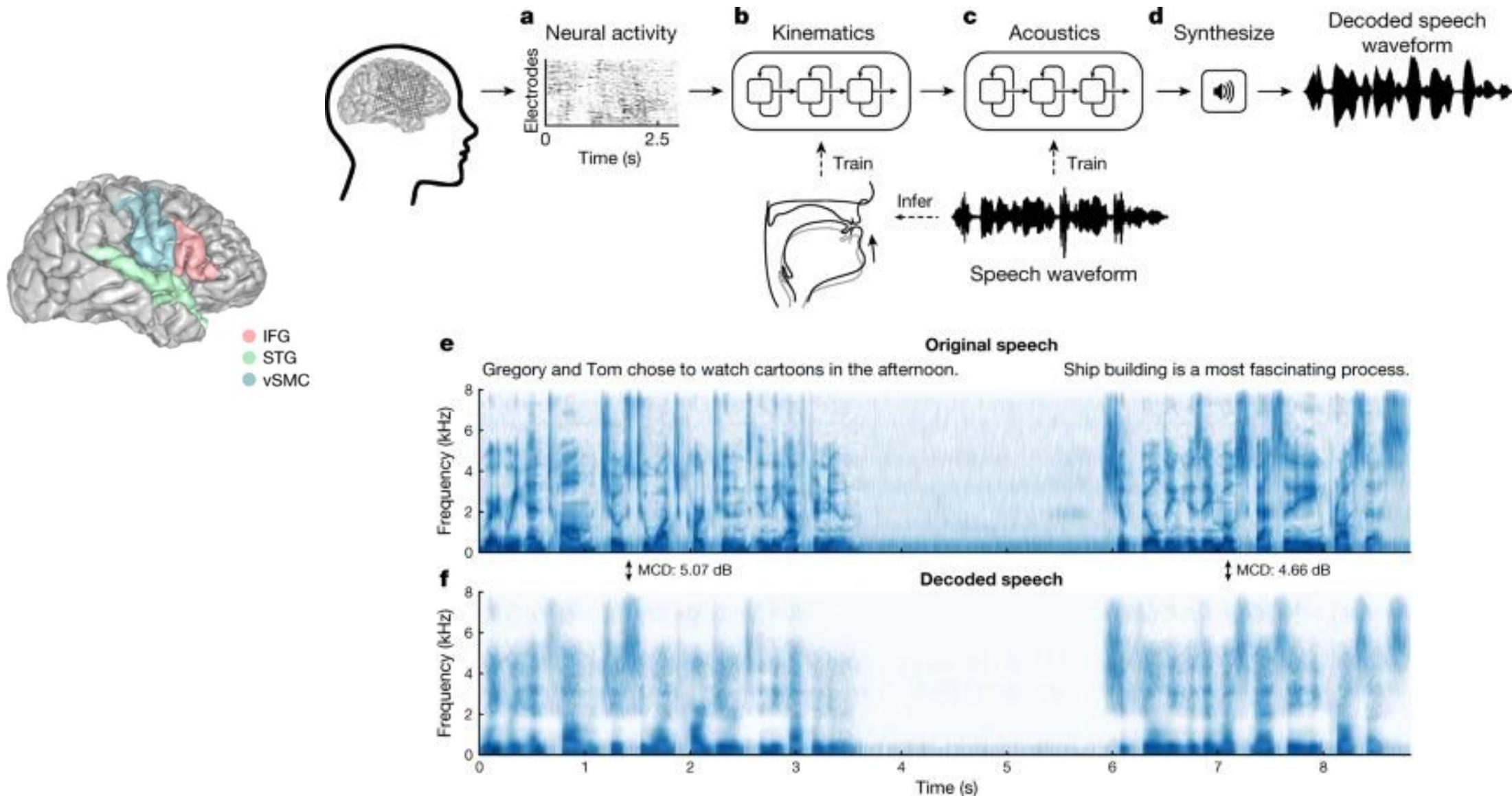
- **Bilateral brain activation shown by fMRI.**
- Based on a **Wada procedure**, the subject illustrated here had a strongly dominant left hemisphere for language.
- **fMRI** shows significant bilateral activation of language areas on a (a) word generation task, (b) passive story listening, and (c) silent sentence repetition.



Decoding language from the neural data

- ECoG-based decoding
- EEG-based decoding

Speech synthesis from neural decoding



Anumanchipalli et al (2019) Speech synthesis from neural decoding of spoken sentences, *Nature*

Speech synthesis from neural decoding

Speech synthesis from
neural decoding of spoken sentences



EEG based language decoding: Dataset – ZuCo

He served in the United States Army in World War II, then got a law degree from Tulane University.

Which university did he get his degree from?
 [1] Austin University
 [2] Tulane University
 [3] Louisiana State University

After this initial success, Ford left Edison Illuminating and, with other investors, formed the Detroit Automobile Company.

Does this sentence contain the *founder* relation? [1]=Yes, [2]=No

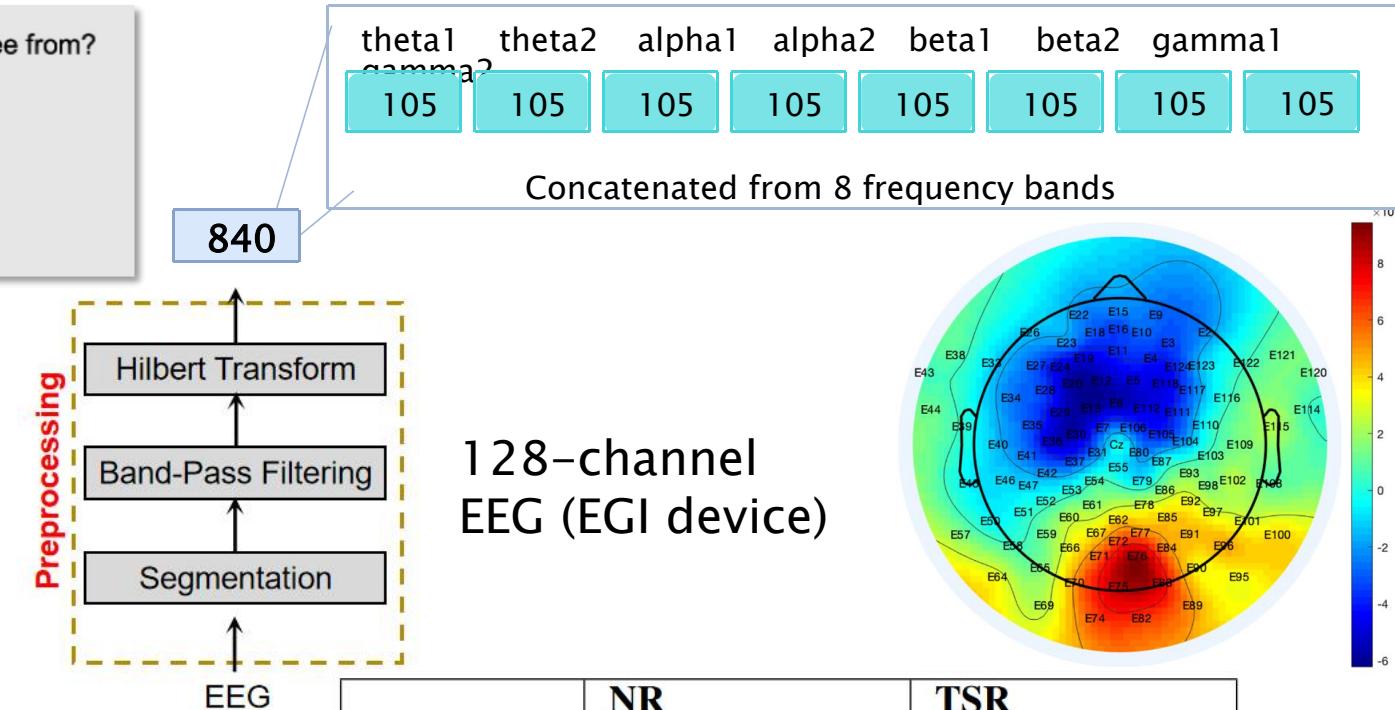
Subjects: ZuCo v1 ($n=12$), ZuCo v2 ($n=18$)

Task1: Sentiment Reading (SR) v1

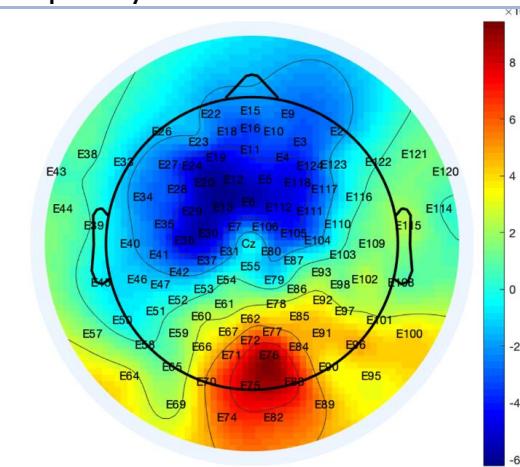
Task2: Normal Reading (NR) v1 & v2

Task3: Task-specific Reading (TSR) v1 & v2

Datasets: ZuCo (V1.0 & V2.0)



128-channel
EEG (EGI device)



	NR	TSR
sentences	349	390
sent. length	mean (SD), range 19.6 (8.8), 5-53	mean (SD), range 21.3 (9.5), 5-53
total words	6828	8310
word types	2412	2437
word length	mean (SD), range 4.9 (2.7), 1-29	mean (SD), range 4.9 (2.7), 1-21
Flesch score	55.38	50.76

Nora et al., Scientific data, 2018

Nora et al., LREC, 2020

EEG based language decoding: Dataset – ChineseEEG

For downloads

Openneuro: <https://openneuro.org/datasets/ds004952>

Science Data Bank 的汉脑交响社区

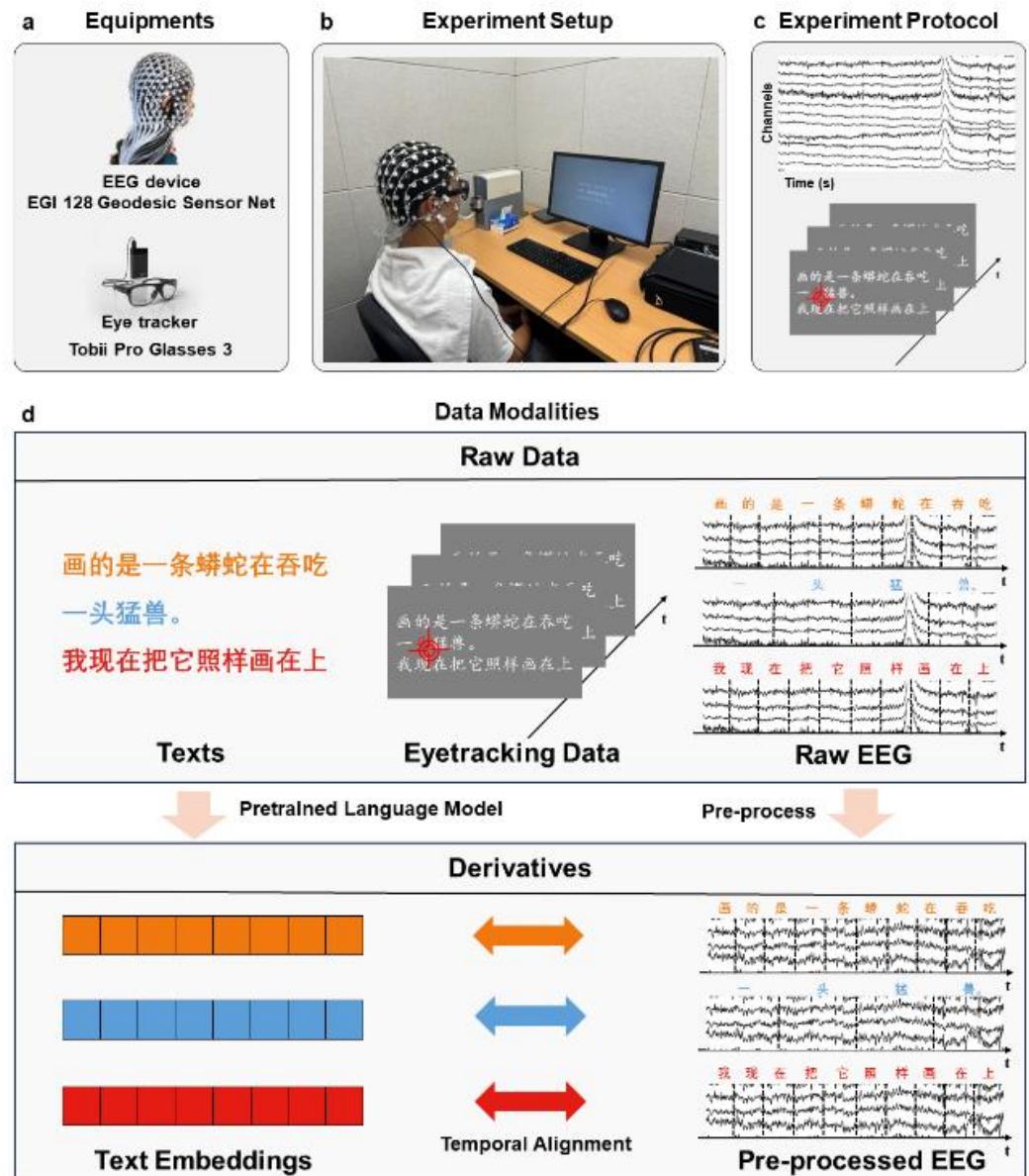
<https://doi.org/10.57760/sciencedb.CHNNeuro.00007>

Subjects: n=10

Task: Reading 12 hours/subject

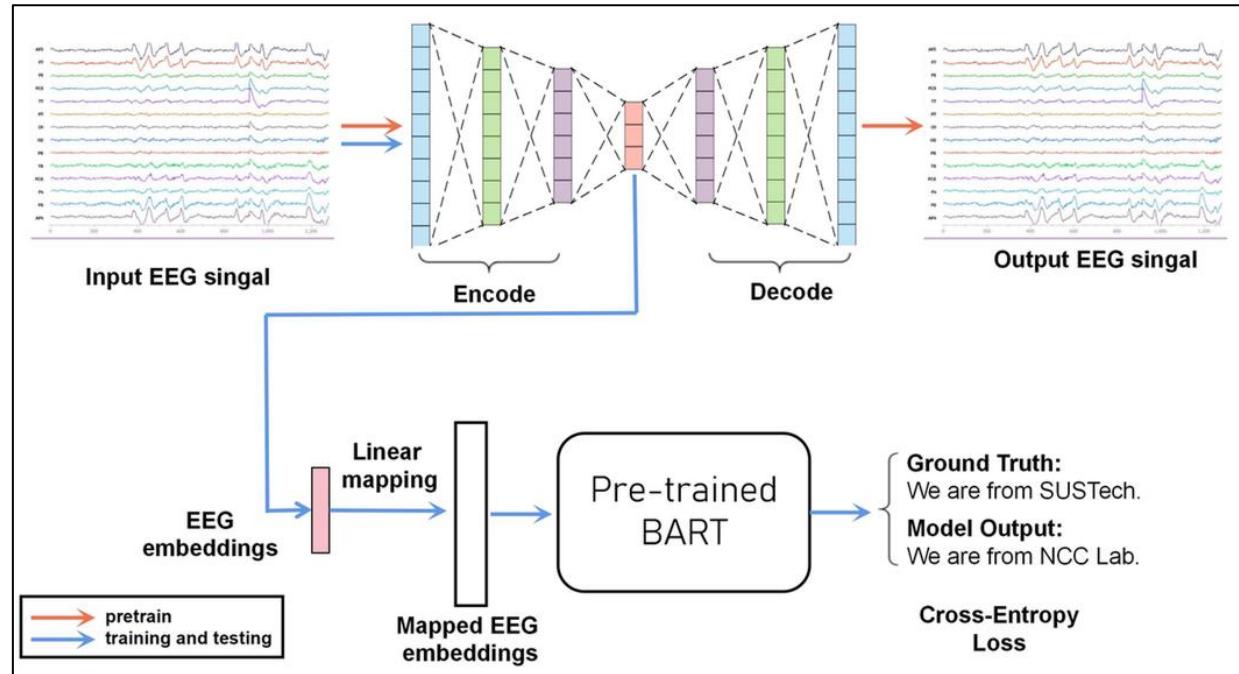
- 128-channel EEG raw/preprocessed data
- EEG source localization data
- Eye tracking data
- text embedding with a pre-trained model (BERT-base-chinese model)

Datasets: ChineseEEG



基于脑电的语言生成：EEG2TEXT

模型



数据

Reading Task	#Unique Sentence	#Training Sample	#Testing Sample
SR v1.0	400	3391	418
NR v1.0	300	2406	321
NR v2.0	349	4456	601
TSR v1.0	407	3372	350

Table 1: Dataset Statistics. SR: Normal Reading (sentiment), NR: Normal Reading (wikipedia), TSR: Task Specific Reading (wikipedia).

结果

Ground Truth: Jeb Bush was born in Midland, Texas, where his father was running an oil drilling company

SOTA Model: Bush, born in theway, Texas, and he **father was a a** oil company company

BTPAE Output: Bush **was** born in Newway, Texas, and he **father was a a** oil company company

Ground Truth: He was voted to five American Football League All-star teams (1961, 1963-1965)

SOTA Model: was born the the of Academy Hall All-Star Games.1962, 1963,1968,

BTPAE Output: was born the the of Academy Hall All-Star **teams**.1962, 1963,1966,

Ground Truth: He won a Nobel Prize in Chemistry in 1928.

SOTA Model: was the Pulitzer Prize for Literature in 1989.

BTPAE Output: was the Pulitzer Prize for **Chemistry in 1928**.

Table 2: EEG-to-Text translation examples on unseen test sentences. **Bold** words indicate exact match, Underline words indicate error match, and **word** words denote those words which can be translated by BTPAE rather than the current model.

Frameworks	BLEU-N(%)				ROUGE-1(%)		
	N = 1	N = 2	N = 3	N = 4	P	R	F
<i>SOTA</i>	40.1	23.1	12.5	6.8	31.7	28.8	30.1
<i>BTPAE</i>	41.3	24.1	13.3	7.6	35.7	29.1	31.9

Table 3: EEG-to-Text sequence-to-sequence model evaluation.

EEG/MEG-To-Text model: the history is accompanied with NLP

Enhancing EEG-to-Text Decoding through Transferable Representations from Pre-trained Contrastive EEG-Text Masked Autoencoder

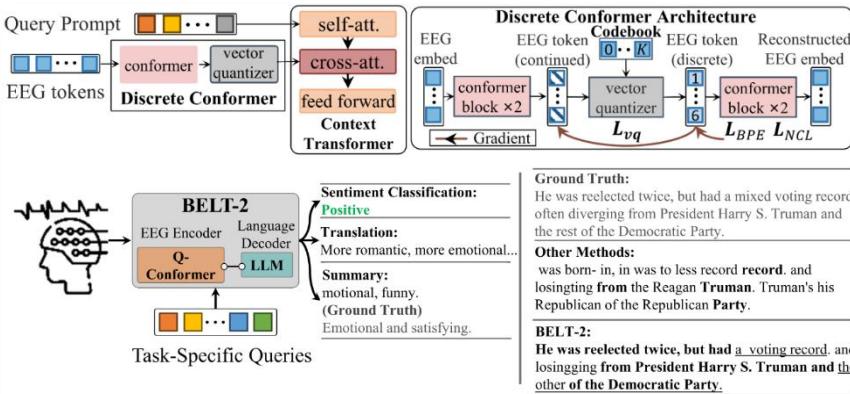
Jiaqi Wang^{1,2}, Zhenxi Song^{1*}, Zhengyu Ma², Xipeng Qiu³, Min Zhang^{1,2}, Zhiguo Zhang^{1,2†}

¹School of Computer Science and Technology, Harbin Institute of Technology Shenzhen, China

² Peng Cheng Laboratory, China

³School of Computer Science, Fudan University, China

mhwjq1998@gmail.com, {songzhenxi, zhangmin2021, zhiguo@hit.edu.cn
mazhy@pcl.ac.cn, xpqiu@fudan.edu.cn



E2T-PTR
arXiv 2024

Belt2
*rejected by ICLR
2024*

VQ-VAE
NeurIPS
2017

E2T-PTR
arXiv 2024

Belt2
*rejected by ICLR
2024*

VQ-VAE
NeurIPS
2017

BERT NAACL 2019

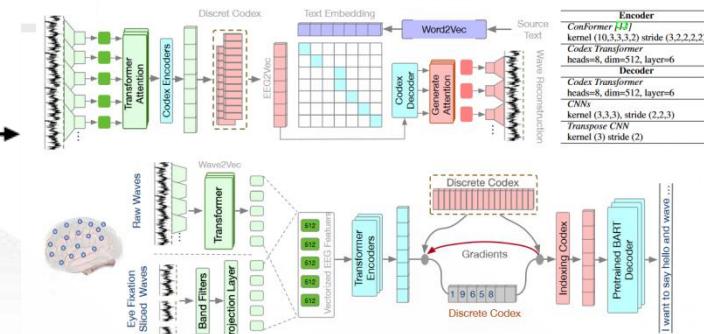
BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova
Google AI Language
{jacobdevlin, mingweichang, kentonl, kristout}@google.com

BART ACL 2020

BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension

Mike Lewis*, Yinhan Liu*, Naman Goyal*, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, Luke Zettlemoyer
Facebook AI
{mikelewis, yinhanliu, naman}@fb.com

DeWave NeurIPS 2023



ChatGPT

Transformer

Attention Is All You Need

Attention based models

Bi-directional LSTM

LSTM based models

RNN based models

Words Embeddings

Bag of Words

Natural language processing (NLP) in machines

1. Grammar based language models

- Symbolic NLP: 1950s – early 1990s

2. Statistical language models

- Statistical NLP: 1990s – 2010s

3. Neural network language models

- Neural NLP: present

Grammar based language models

In the early days, many language-processing systems were designed by symbolic methods, i.e., **the hand-coding of a set of rules**, coupled with a dictionary lookup: such as by writing grammars, or devising heuristic rules for stemming.

"Every time I fire a linguist, the performance of the system goes up."

—Frederick Jelinek (18 Nov 1932 – 14 Sep 2010)

Grammar based language models

In the early days, many language-processing systems were designed by symbolic methods, i.e., **the hand-coding of a set of rules**, coupled with a dictionary lookup: such as by writing grammars, or devising heuristic rules for stemming.

More recent systems based on **machine-learning algorithms have many advantages** over hand-produced rules:

- The learning procedures used during machine learning **automatically focus on the most common cases**, whereas when writing rules by hand it is often not at all obvious where the effort should be directed.
- **Automatic learning procedures** can make use of statistical inference algorithms to produce models that are robust to unfamiliar input (e.g. containing words or structures that have not been seen before) and to erroneous input (e.g. with misspelled words or words accidentally omitted).
- Systems based on automatically learning the rules can be made **more accurate simply by supplying more input data**. However, systems based on handwritten rules can only be made more accurate by increasing the complexity of the rules, which is a much more difficult task.³⁰

Statistical language models

- Statistical Language Model **learns the probability of word occurrence** based on examples of text.
- A good example is *speech recognition*, where audio data is used as an input to the model and the output requires a language model that interprets the input signal and recognizes each new word within the context of the words already recognized.

Unigram: the probability of each word only depends on that word's own probability in the document.

$$P(t_1, t_2, t_3) = P(t_1)P(t_2|t_1)P(t_3|t_1t_2) = P(t_1)P(t_2)P(t_3)$$

N-gram: In an n-gram model, the probability $P(w_1, \dots, w_m)$ of observing the sentence w_1, \dots, w_m is approximated as

$$P(w_1, \dots, w_m) = \prod_{i=1}^m P(w_i|w_1, \dots, w_m) \approx \prod_{i=1}^m P(w_i|w_{i-(n-1)}, \dots, w_{i-1})$$

Neural NLP models – word2vec

- Word2vec (Mikolov et al. 2013) is a framework for learning *word vectors*.
- The **word2vec** objective function causes **the words that occur in similar contexts to have similar embedding**.
- **Example:**
 1. The **kid** said he would grow up to be superman.
 2. The **child** said he would grow up to be superman
- The words **kid** and **child** will have similar word vectors due to a similar context.

How to obtain word2vec representation

- Given a large corpus of text
- Every **word** in a fixed vocabulary is represented by an **one-hot vector** (One bit ‘1’ and all others ‘0’). **Vector length** is the number of words in language.

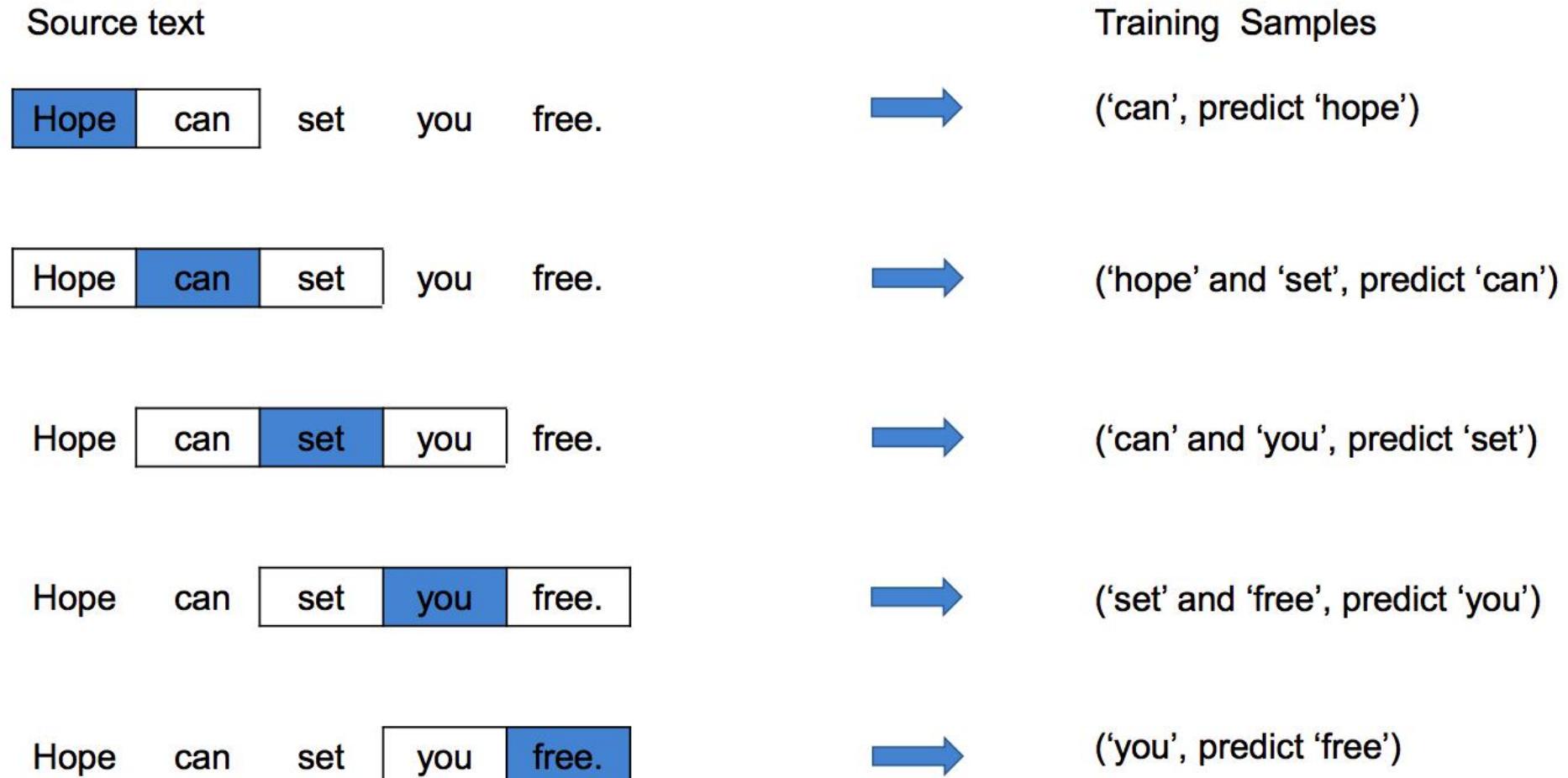
1. Continuous bag of words (CBOW): Predict the target word from the context.

- Go through each position t in the text, which has a center word c and context words o. Given context words o, predict center word c.

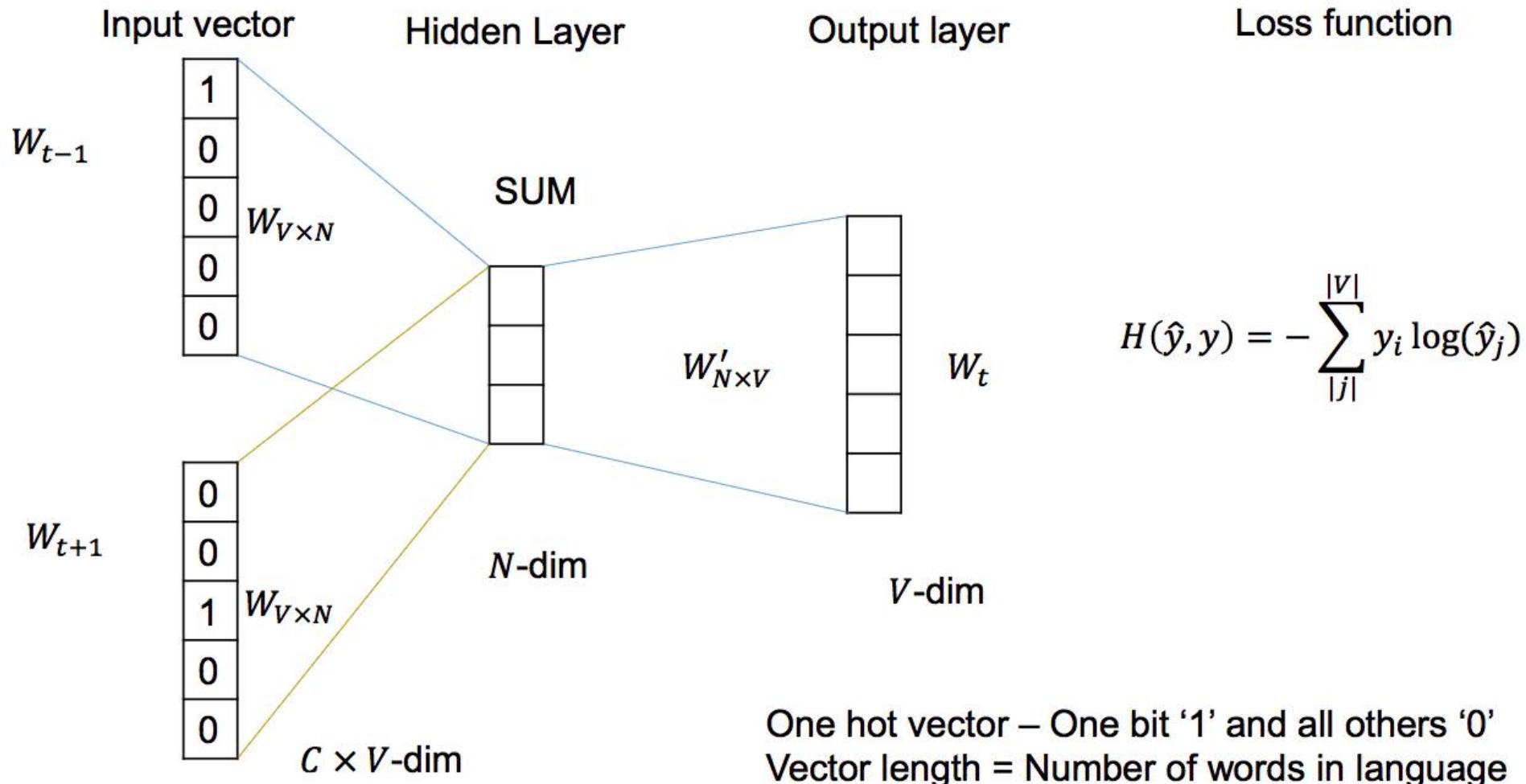
2. Skip-Gram: Predict the context words from target.

- Given center word c, predict context words o.
- Keep adjusting the word vectors to **maximize** this probability.

Continuous bag of words (CBOW)



Continuous bag of words (CBOW)



Skip-Gram

Source text

Hope can

set you free.



Training Samples

('hope', predict 'can')

Hope can set

you free.



('can', predict 'hope' and 'set')

Hope can set you

free.



('set', predict 'can' and 'you')

Hope can set you

free.



('you', predict 'set' and 'free')

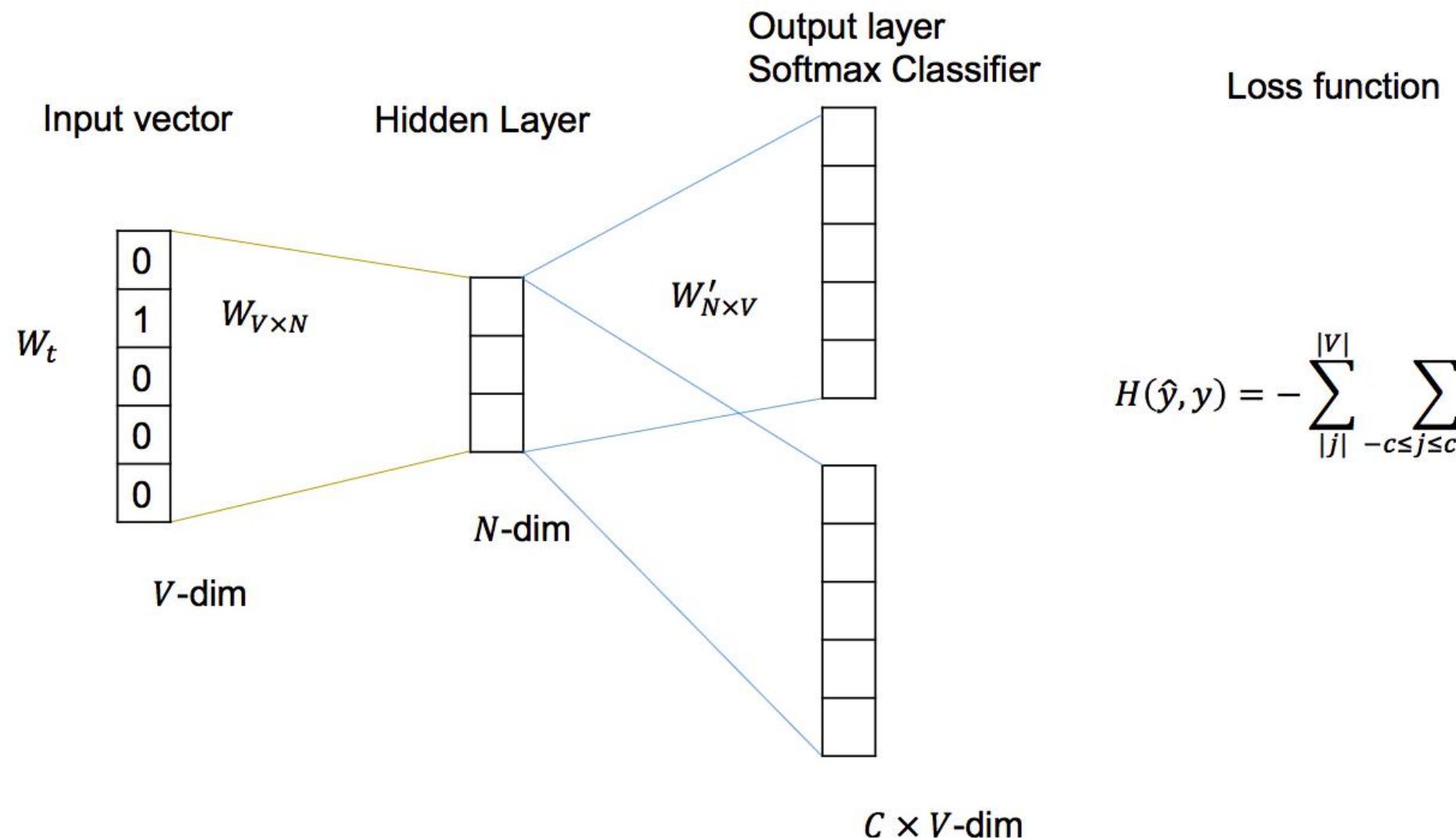
Hope can set you

free.



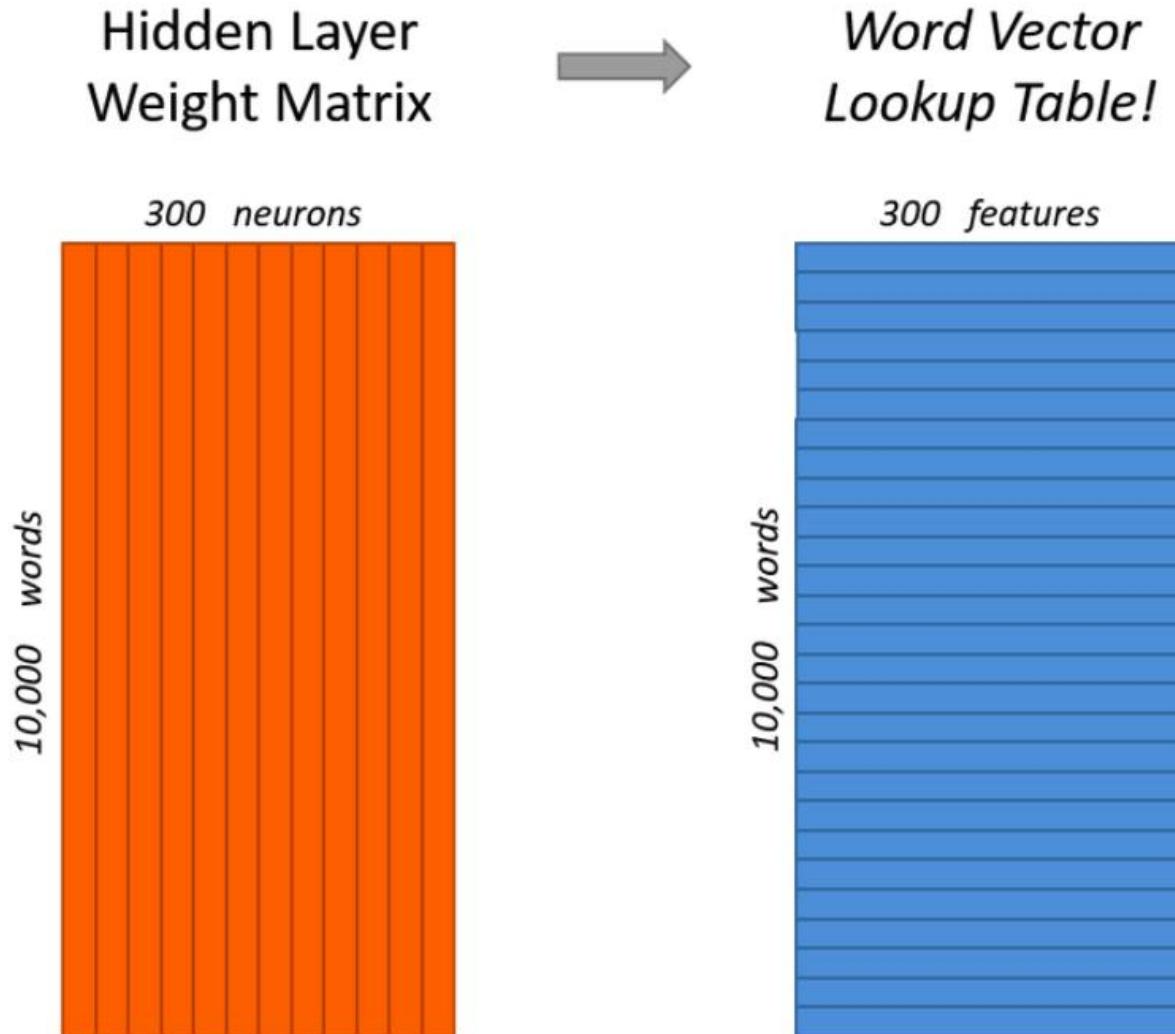
('free', predict 'you')

Skip-Gram



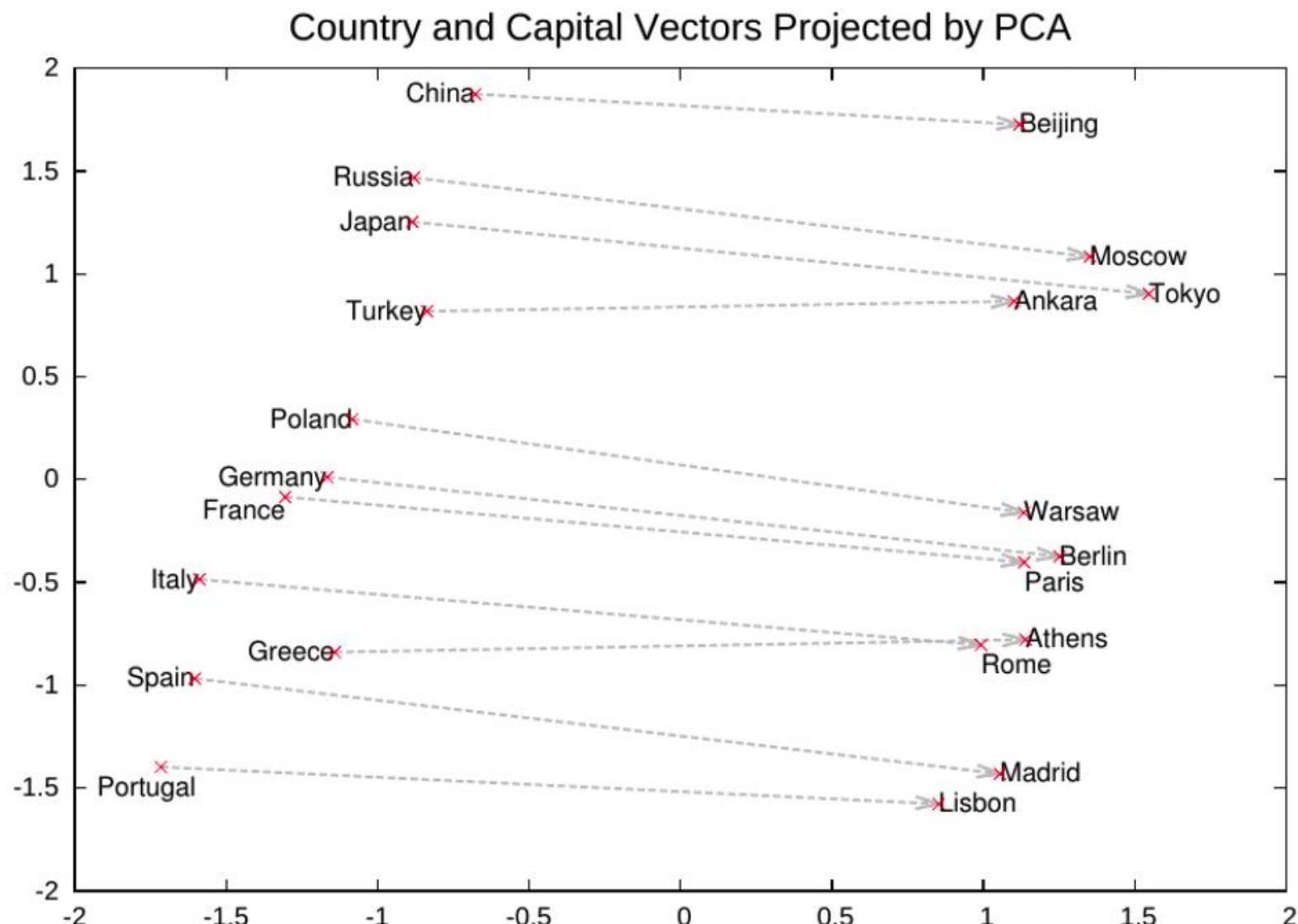
$$H(\hat{y}, y) = - \sum_{|V|} \sum_{-c \leq j \leq c, j \neq 0} y_i \log(\hat{y}_j)$$

Hidden layer

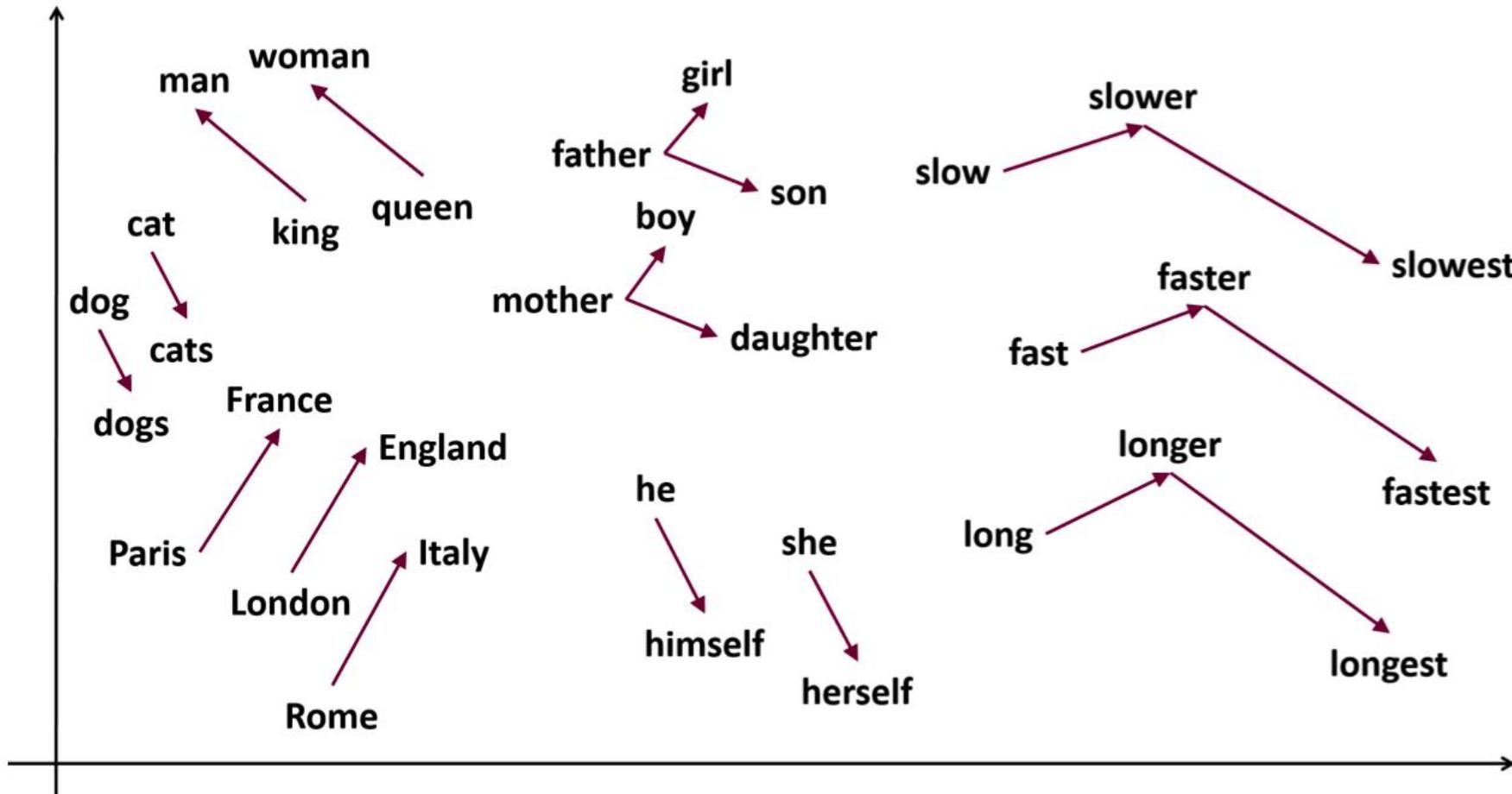


$$[0 \ 0 \ 0 \ 1 \ 0] \times \begin{bmatrix} 13 & 7 & 1 \\ 6 & 23 & 4 \\ 18 & 13 & 6 \\ 10 & 5 & 19 \\ 7 & 17 & 25 \end{bmatrix} = [10 \ 5 \ 19]$$

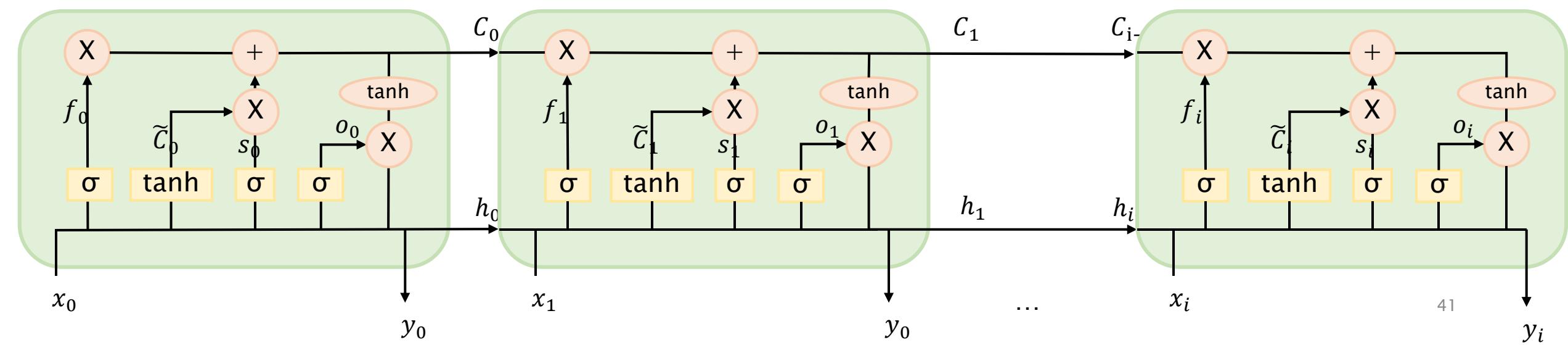
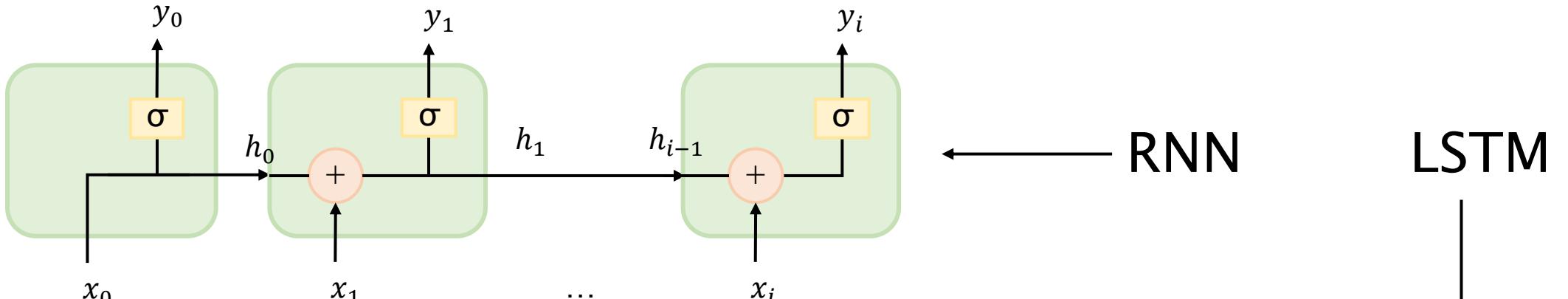
Word2vec realization examples



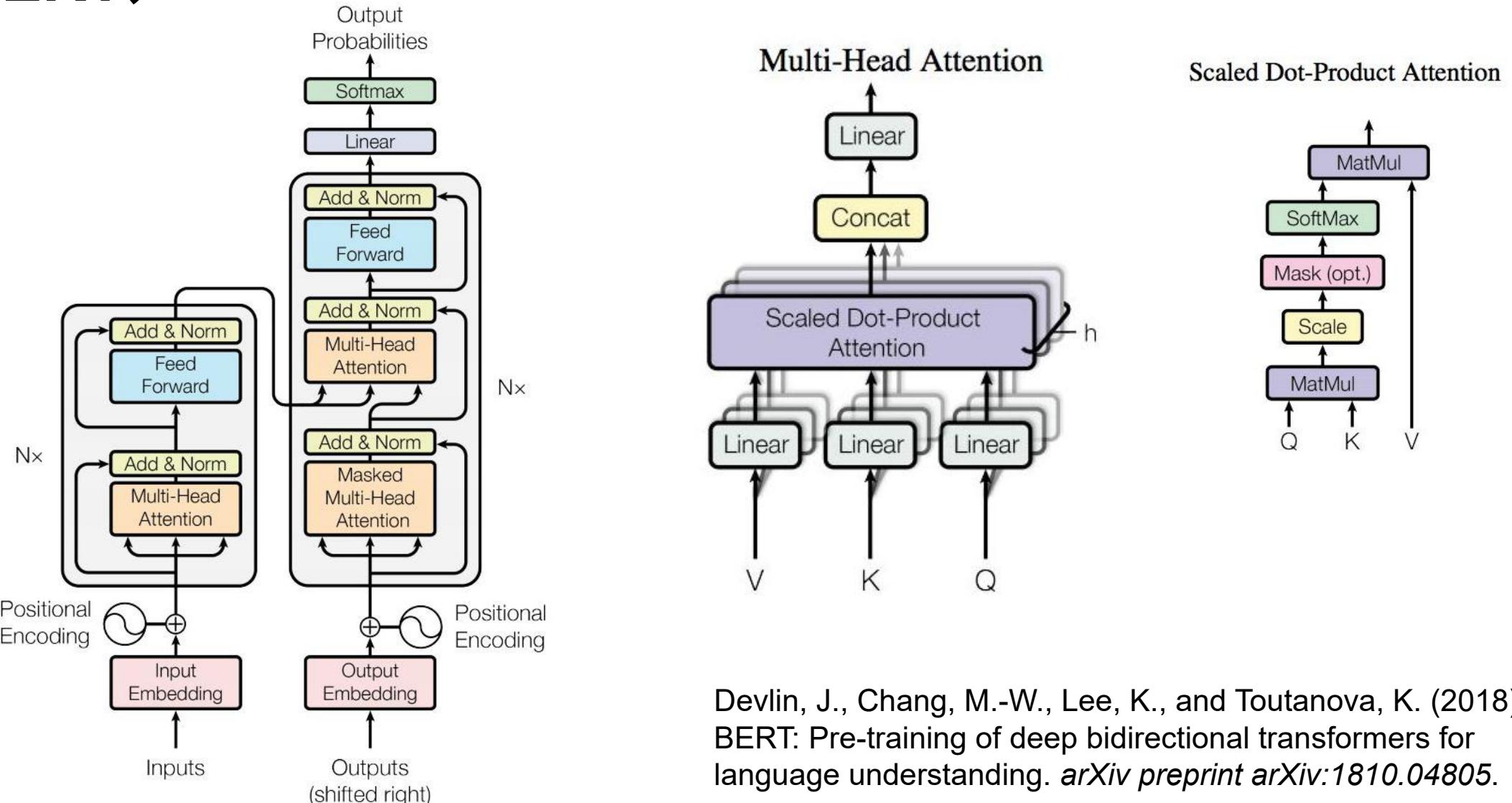
Word2vec realization examples



RNN & LSTM

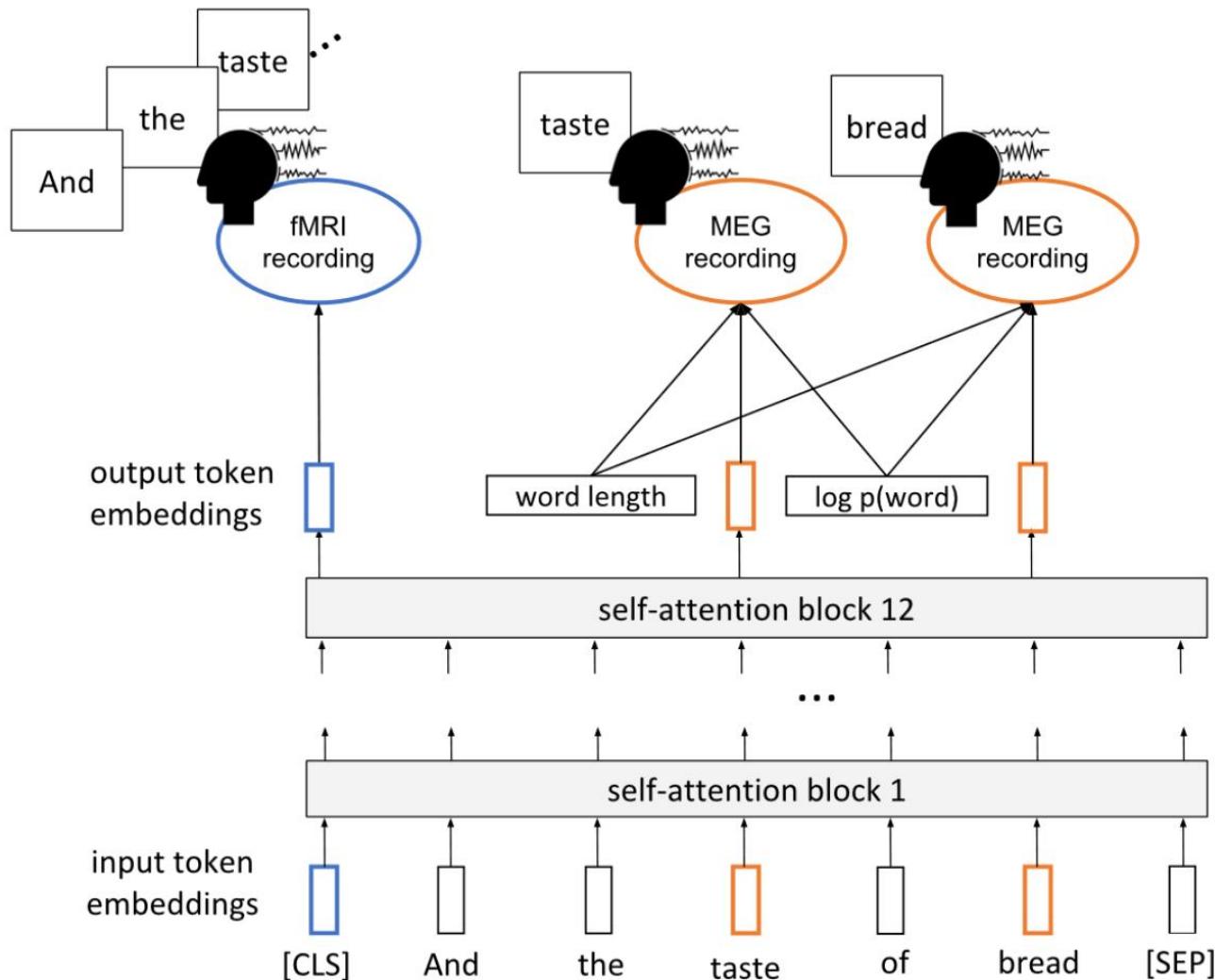


Bidirectional Encoding Representation from Transformers (BERT)



Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Fine-tuning BERT using neural data



General approach for fine-tuning BERT using fMRI and/or MEG data.

A linear layer maps the output token embeddings from the base architecture to brain activity recordings.

Only MEG recordings that correspond to content words in the input sequence are considered.

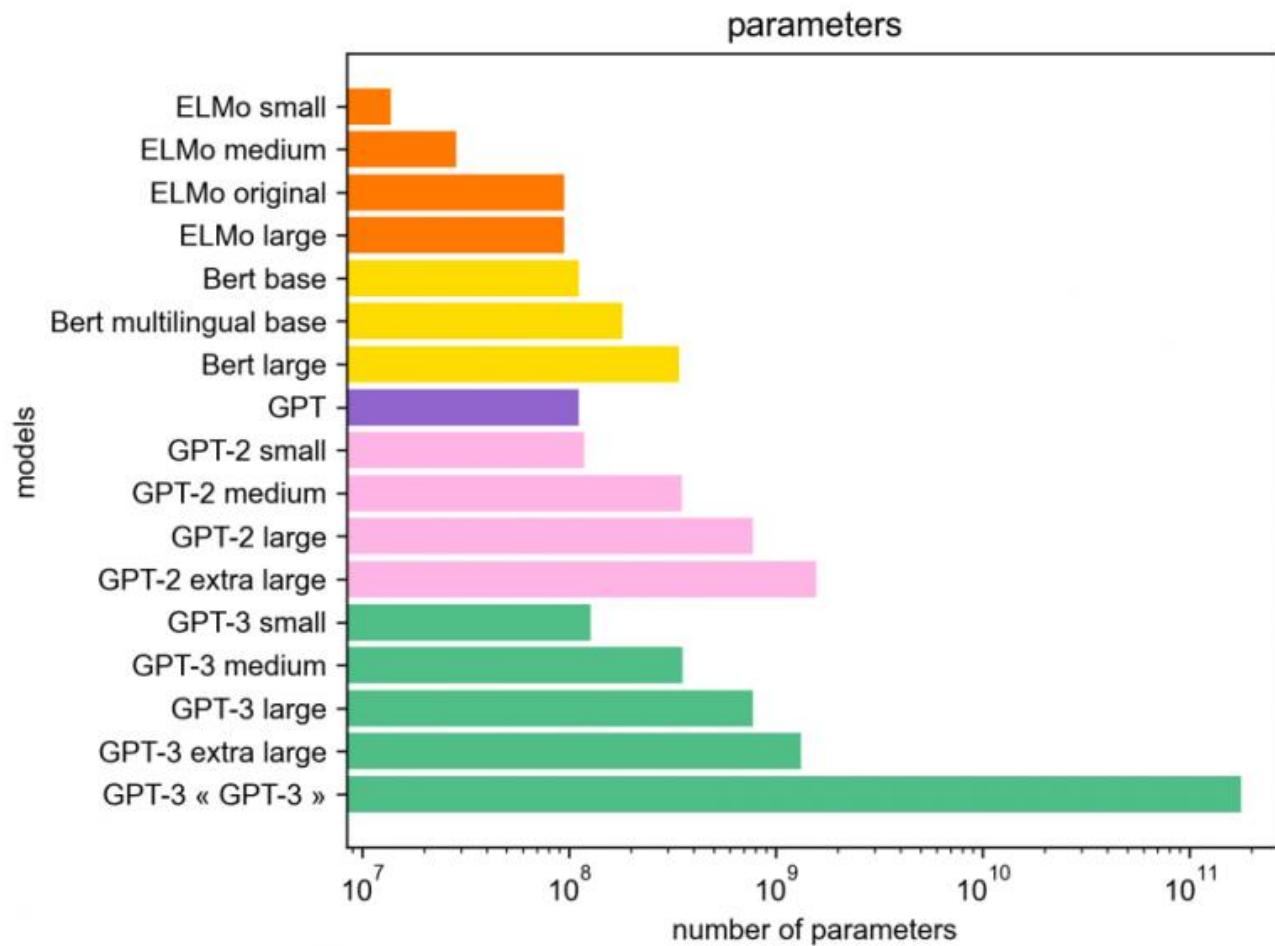
We include the word length and context-independent log-probability of each word when predicting MEG.

fMRI data are predicted from the pooled embedding of the sequence, i.e. the [CLS] token embedding.

Generative Pre-trained Transformer (GPT)

It is the third-generation language prediction model in the GPT-n series (and the successor to GPT-2) created by OpenAI, a San Francisco-based artificial intelligence research laboratory. GPT-3, which was introduced in May 2020, and was in beta testing as of July 2020, is part of a trend in natural language processing (NLP) systems of pre-trained language representations.

GPT-3 Training Data		
Dataset	# Tokens	Weight in Training Mix
Common Crawl	410 billion	60%
WebText2	19 billion	22%
Books1	12 billion	8%
Books2	55 billion	8%
Wikipedia	3 billion	3%



GPT (2018)

Improving Language Understanding by Generative Pre-Training

Alec Radford
OpenAI
alec@openai.com

Karthik Narasimhan
OpenAI
karthikn@openai.com

Tim Salimans
OpenAI
tim@openai.com

Ilya Sutskever
OpenAI
ilyasu@openai.com

Abstract

Natural language understanding comprises a wide range of tasks such as textual entailment, question answering, sentiment analysis, and document classification. Although large unlabeled datasets exist for learning these specific tasks, training discriminatively trained models to perform adversarial tasks on these datasets gains on these tasks can be realized by *generalization* on a diverse corpus of unlabeled text, followed by fine-tuning on the specific task. In contrast to previous approaches, our approach requires no explicit transformations during fine-tuning to achieve minimal changes to the model architecture. We evaluate our approach on a wide range of benchmarks. Our general task-agnostic model outperforms state-of-the-art models on 9 out of the 12 tasks studied, achieving improvements of 8.9% on commonsense reasoning (RACE), 1.5% on text

OpenAI is an AI research and deployment company. Our mission is to ensure that artificial general intelligence (AGI) benefits all of humanity.



Ilya Sutskever
1986 -- now

GPT2 (2019)

Language Models are Unsupervised Multitask Learners

Alec Radford *¹ Jeffrey Wu *¹ Rewon Child¹ David Luan¹ Dario Amodei **¹ Ilya Sutskever **¹

Abstract

Natural language processing tasks, such as question answering, machine translation, reading comprehension, and summarization, are typically approached with supervised learning on task-specific datasets. We demonstrate that language models begin to learn these tasks without any explicit supervision when trained on a new dataset of millions of webpages called WebText. When conditioned on a document plus questions, the answers generated by the language model reach 55 F1 on the CoQA dataset - matching or exceeding the performance of 3 out of 4 baseline systems without using the 127,000+ training examples. The capacity of the language model is essential to the success of zero-shot task transfer and increasing it improves performance in a log-linear fashion across tasks. Our largest model, GPT-2, is a 1.5B parameter Transformer that achieves state-of-the-art results on 7 out of 8 tested language modeling datasets in a zero-shot setting but still underfits WebText. Samples from the model reflect these improvements and contain coherent paragraphs of text. These findings suggest a promising path towards building language processing systems which learn to perform tasks from their naturally occurring demonstrations.

competent generalists. We would like to move towards more general systems which can perform many tasks – eventually without the need to manually create and label a training dataset for each one.

The dominant approach to creating ML systems is to collect a dataset of training examples demonstrating correct behavior for a desired task, train a system to imitate these behaviors, and then test its performance on independent and identically distributed (IID) held-out examples. This has served well to make progress on narrow experts. But the often erratic behavior of captioning models (Lake et al., 2017), reading comprehension systems (Jia & Liang, 2017), and image classifiers (Alcorn et al., 2018) on the diversity and variety of possible inputs highlights some of the shortcomings of this approach.

Our suspicion is that the prevalence of single task training on single domain datasets is a major contributor to the lack of generalization observed in current systems. Progress towards robust systems with current architectures is likely to require training and measuring performance on a wide range of domains and tasks. Recently, several benchmarks have been proposed such as GLUE (Wang et al., 2018) and decaNLP (McCann et al., 2018) to begin studying this.

Multitask learning (Caruana, 1997) is a promising framework for improving general performance. However, multitask training in NLP is still nascent. Recent work re-

GPT3 (2020)

Language Models are Few-Shot Learners

Tom B. Brown* Benjamin Mann* Nick Ryder* Melanie Subbiah*

Jared Kaplan† Prafulla Dhariwal Arvind Neelakantan Pranav Shyam

Girish Sastry Amanda Askell Sandhini Agarwal Ariel Herbert-Voss

Gretchen Krueger Tom Henighan Rewon Child Aditya Ramesh

Daniel M. Ziegler Jeffrey Wu Clemens Winter

Christopher Hesse Mark Chen Eric Sigler Mateusz Litwin Scott Gray

Benjamin Chess Jack Clark Christopher Berner

Sam McCandlish Alec Radford Ilya Sutskever Dario Amodei

Abstract

We demonstrate that scaling up language models greatly improves task-agnostic, few-shot performance, sometimes even becoming competitive with prior state-of-the-art fine-tuning approaches. Specifically, we train GPT-3, an autoregressive language model with 175 billion parameters, 10x more than any previous non-sparse language model, and test its performance in the few-shot setting. For all tasks, GPT-3 is applied without any gradient updates or fine-tuning, with tasks and few-shot demonstrations specified purely via text interaction with the model. GPT-3 achieves strong performance on many NLP datasets, including translation, question-answering, and cloze tasks. We also identify some datasets where GPT-3's few-shot learning still struggles, as well as some datasets where GPT-3 faces methodological issues related to training on large web corpora.

Is NLP a road to AGI?

Table 1: Downstream tasks for LLMs

category	task	description	dataset
understanding	language modeling	predict the next word or character in a sequence	WikiText-103 ⁴⁷ , Penn Treebank ⁴⁸ , The Pile ⁴⁹ , LAMBADA ⁵⁰
	part-of-speech tagging	label each word in a sentence with its part of speech	Penn Treebank ⁴⁸ , Ritter ⁵¹ , UD ⁵²
	named entity recognition	identify named entities in a text	CoNLL-2003 ⁵³ , WNUT-2017 ⁵⁴ , OntoNotes ⁵⁵ ,
	textual entailment	determine if a sentence logically follows from another one	GLUE ⁵⁶ , MNLI ⁵⁷ , RTE ⁵⁸
	sentiment analysis	identify and categorize opinions expressed in a text	IMDb ⁵⁹ , Yelp ⁶⁰ , GLUE ⁵⁶ , SST-2 ⁶¹
	question answering	answer questions based on a given context	Natural Questions ⁶² , TriviaQA ⁶³ , HotpotQA ⁶⁴ , WikiQA ⁶⁵ , SQuAD ⁶⁶
generation	machine translation	translate sentences between languages	WMT ⁶⁷ , WIT ⁶⁸
	text summarization	generate a short summary from a longer text	CNN/Daily Mail ³⁰ , GigaWord ³¹ , X-Sum ⁶⁹
	dialogue generation	conduct a conversation	PersonaChat ⁷⁰ , UDC ⁷¹
	code generation	generate code based on a natural language description	Human Eval ⁴⁰ , APPS ⁷² , SPoC ⁷³
reasoning	knowledge completion	fill or predict missing information in a knowledge units	FB15k ⁷⁴ , WikiFact ⁷⁵
	knowledge reasoning	reason over structured knowledge	CSQA ³² , StrategyQA ⁷⁶
	symbolic reasoning	reason over symbols following formal rules	Big-bench ⁷⁷ , PAL ⁷⁸ , TabFact ⁷⁹
	mathematical reasoning	solve mathematical problems based on text description	MMLU ⁸⁰ , GSM8k ⁸¹ , SVAMP ⁸² , MathQA ⁸³ , AQUA-RAT ⁸⁴

Table 2: Cognitive task evaluation in models

type	model	cognitive function	cognitive task	result
RNN	RNN	decision-making	perceptual decision-making	RNN exhibits representations highly similar to the biological brain ⁸⁹ .
	RNN	timing	time production task	RNN demonstrates effective capturing of flexible timing in time intervals ⁹⁰ .
	RNN	navigation	path integration task	RNN exhibits strong ability in path integration and can effectively model the neural responses of grid cells ⁹¹ .
Task-specific models	RNN	reward-based learning	value-based tasks	RNN captures experimental observations from diverse cognitive and value-based tasks ⁹² .
	CNN	vision	object recognition tasks	RNN proves highly predictive of neural responses in visual cortex ⁹³ .
	CNN	auditory	auditory tasks	CNN demonstrates strong fitting to auditory processing-related cortical areas ⁹⁴ .
	SNN	decision-making	two alternative forced choice task	SNN exhibits excellent performance and dynamic properties in the two alternative forced choice task ⁹⁵ .
LLMs	GPT-3	decision-making	gambling	GPT-3's performance falls short of human performance ⁴² .
	GPT-3	information search	horizon task	GPT-3 exhibits the capacity to make rational decisions when provided with option descriptions ⁴² .
	GPT-3	deliberation	two-step task	exhibits a preference for intuitive answers ⁴² .
	GPT-3	causal reasoning	causal reasoning tasks ⁹⁶	GPT-3 has difficulties with causal reasoning ⁴² .
InstructGPT, LLaMA, GPT-3, GPT-4	GPT-3	reasoning with probabilities, decision-making	Wason selection task, multi-armed bandit	GPT-3 outperforms humans in decision-making tasks ⁴³ .
	GPT-3	decision-making	lexical decision	the semantic activation patterns of GPT-3 are similar to humans ⁹⁷ .
	PaLM, PaLM 2	memorization	memory test	the memory capacity of PaLM 2 is inferior to that of PaLM when the repetition count is less than three ⁹⁸ .
	GPT-4	theory of mind	false belief test, emotion understanding	GPT-4 demonstrates a certain level of theory of mind capability, being able to infer the mental states of others ³⁶ .
SelfAware dataset	InstructGPT, LLaMA, GPT-3, GPT-4	self-knowledge	SelfAware dataset	while GPT-4 surpasses GPT-3, InstructGPT, and LLaMA in self-awareness capability, it still falls short of human-level self-knowledge recognition ⁹⁹ .

NLP in machines

v.s.

NLP in the brain

NLP in the brain

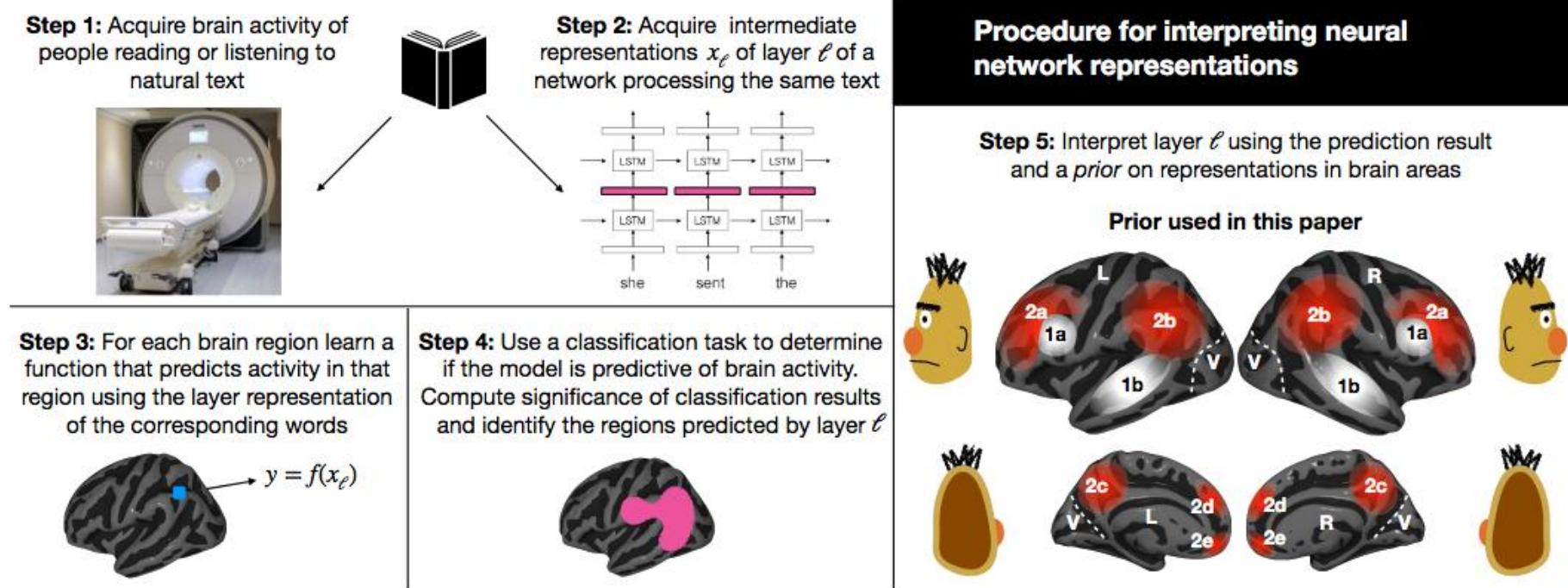


Figure 1: Diagram of approach and prior on brain function. The prior was constructed using the results of Lerner *et al.* (2011): regions in group 1 (white) process information related to isolated words and word sequences while group 2 (red) process only information related to word sequences (see Section 1.1). V indicates visual cortex. The drawing indicates the views of the brain with respect to the head. See supplementary materials for names of brain areas and full description of the methods.

NLP in the brain

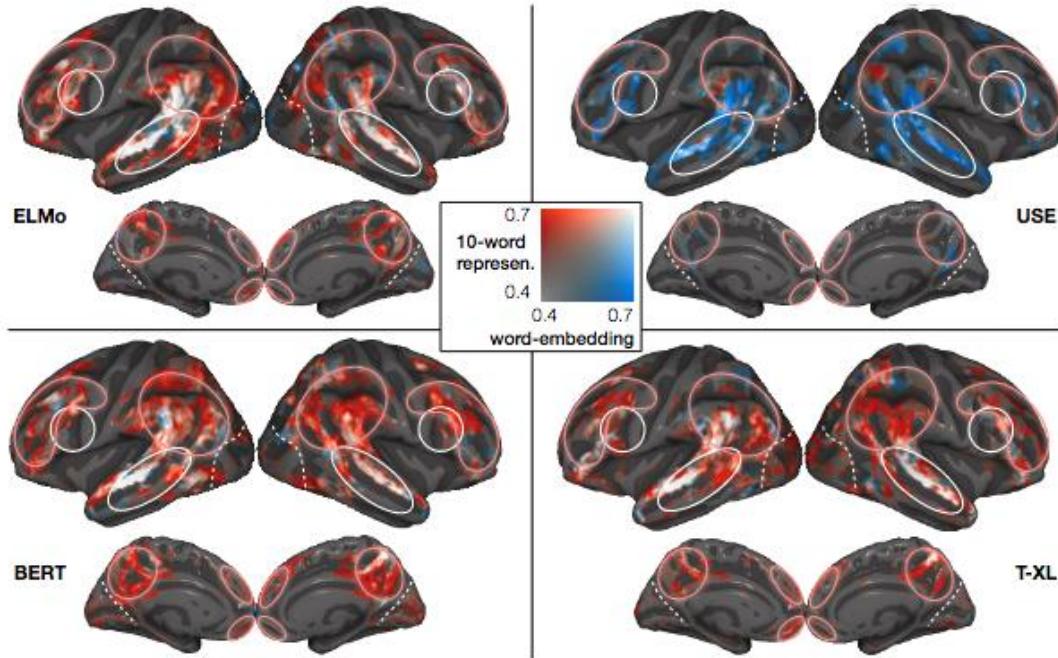


Figure 2: Comparison between the prediction performance of two network representations from each model: a 10-word representation corresponding to the 10 most recent words shown to the participant (Red) and a word-embedding corresponding to the last word (Blue). Areas in white are well predicted from both representations. These results align to a fair extent with our prior: group 2 areas (red outlines) are mostly predicted by the longer context representations while areas 1b (lower white outlines) are predicted by both word-embeddings and longer context representations.

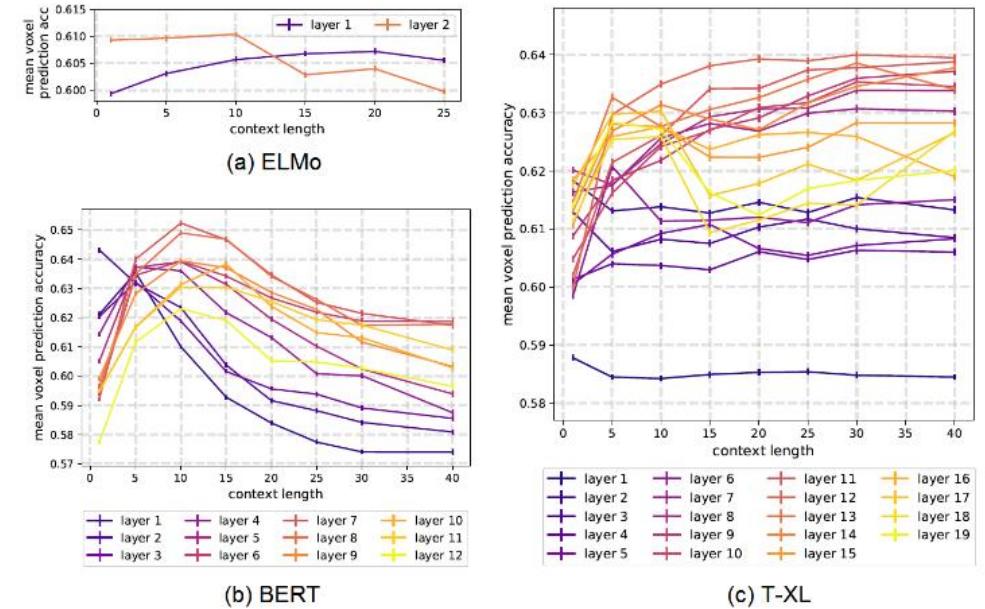


Figure 4: Performance of encoding models for all hidden layers in ELMo, BERT, and T-XL as the amount of context provided to the network is increased. Transformer-XL is the only model that continues to increase performance as the context length is increased. In all networks, the middle layers perform the best for contexts longer than 15 words. The deepest layers across all networks show a sharp increase in performance at short-range context (fewer than 10 words), followed by a decrease in performance.

NLP in AI vs. NLP in brain

Language Stimuli

Pereira2018

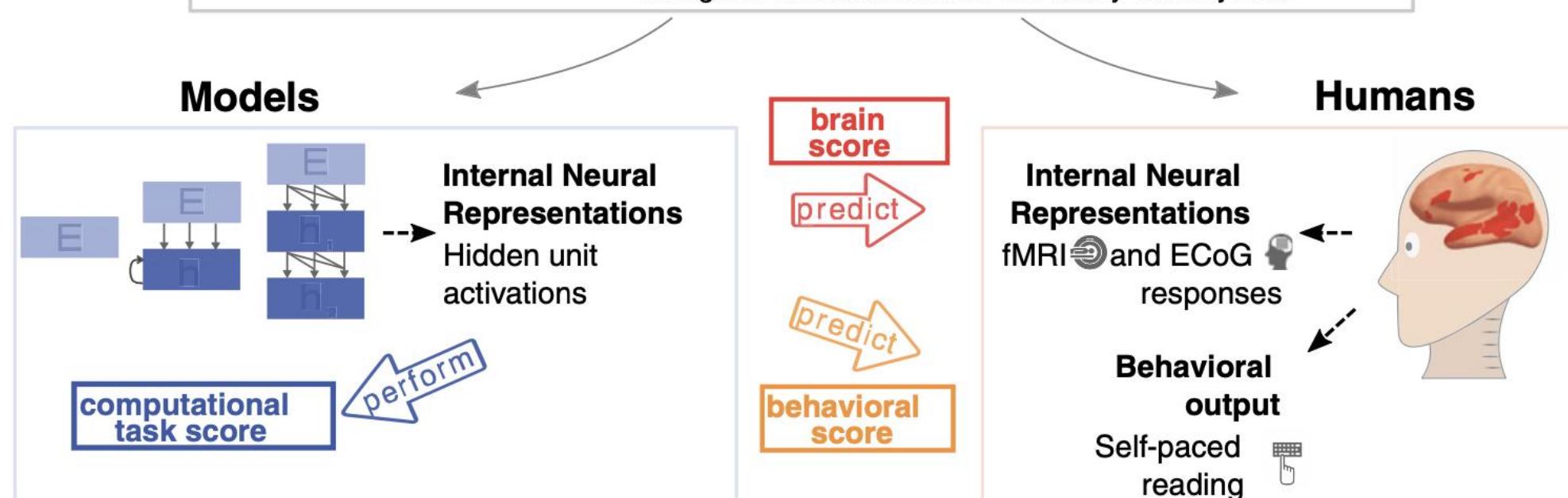
"Beekeeping encourages the conservation of local habitats. It is in every beekeeper's interest..."

Fedorenko2016

"Alex was tired so he took a nap."

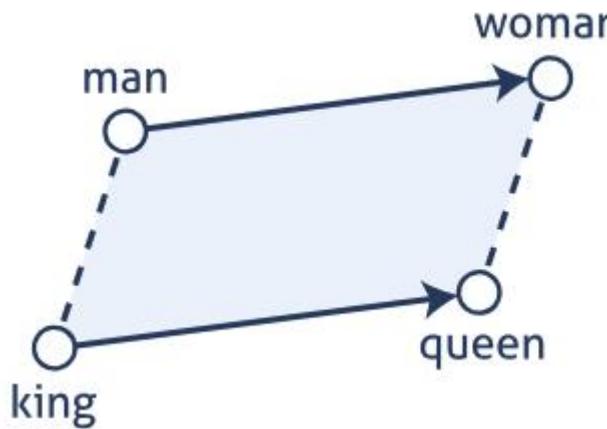
Blank2014

"If you were to journey to the North of England, you would come to a valley that is surrounded by moors as high as mountains. It is in this valley where you..."



Words' relations in human are different with AI (untill 2020)

The parallelogram model



$$\text{vec}(\text{"Madrid"}) - \text{vec}(\text{"Spain"}) + \text{vec}(\text{"France"})$$

$$\text{vec}(\text{"King"}) - \text{vec}(\text{"man"}) + \text{vec}(\text{"woman"})$$

$$\text{vec}(\text{"teaching"}) - \text{vec}(\text{"teach"}) + \text{vec}(\text{"learn"})$$

$$\text{vec}(\text{"Paris"})$$

$$\text{vec}(\text{"Queen"})$$

$$\text{vec}(\text{"learning"})$$

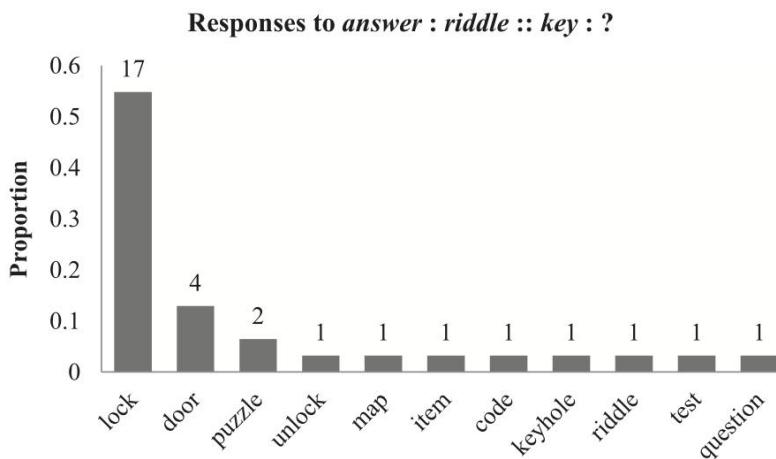
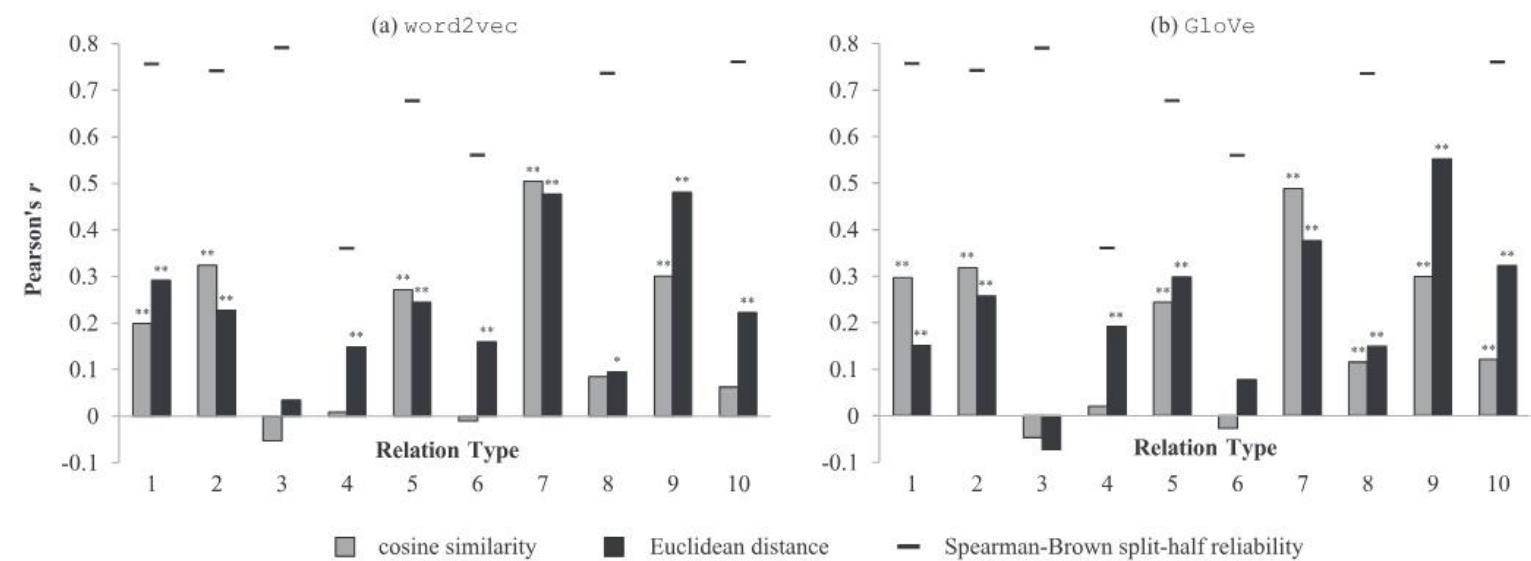
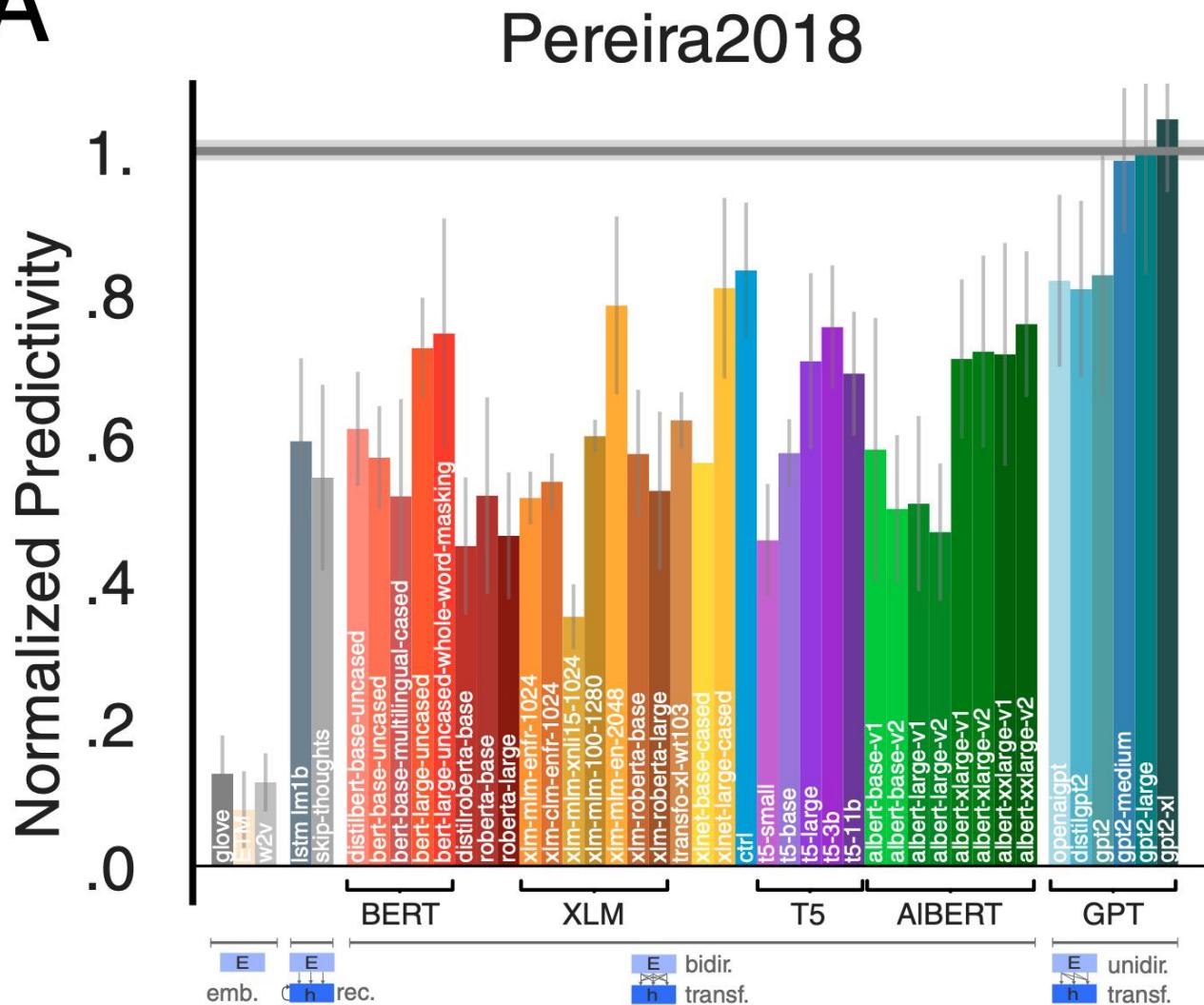
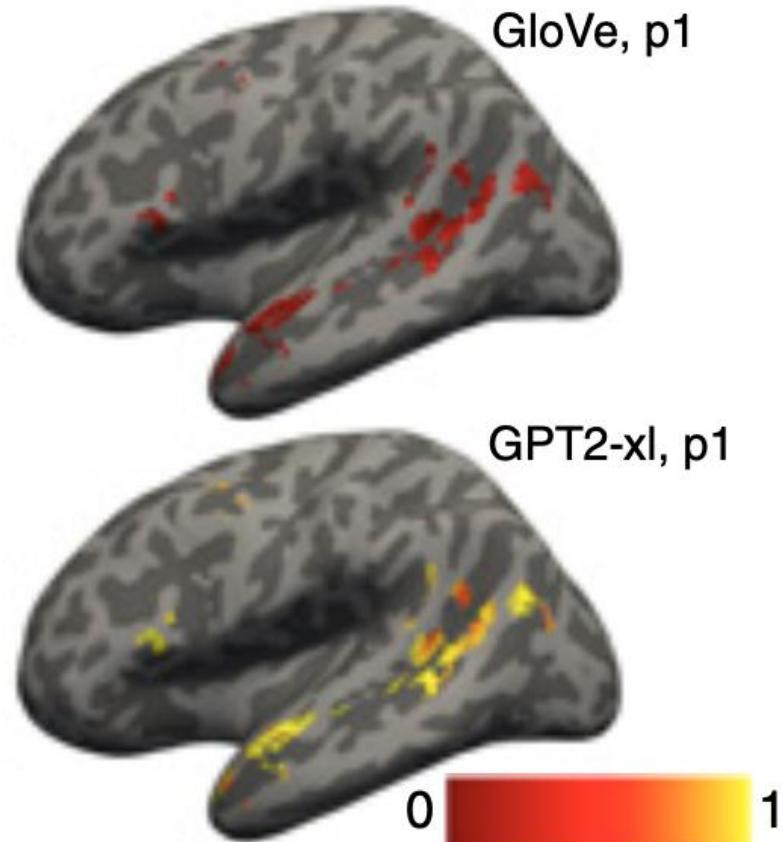


Fig. 3. Response proportions and frequencies (shown above the bars) for the question *answer:riddle::key:?*, a cross-domain question from Experiment 1a.



The better NLP performance, the higher similarity to humans.

A**B**

Next-word prediction is important !

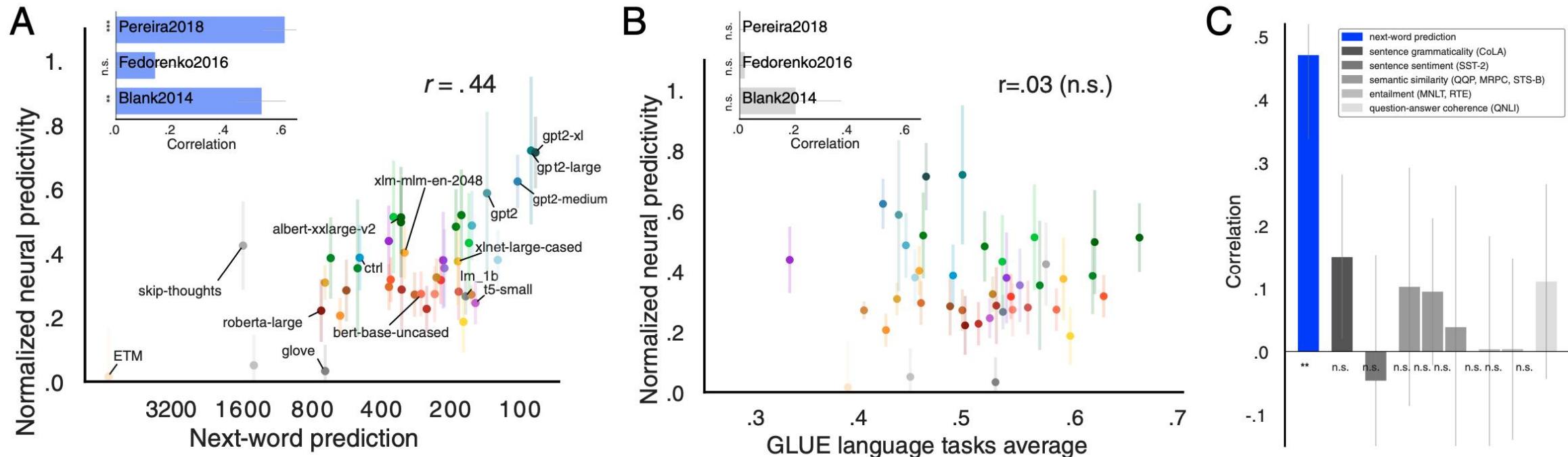
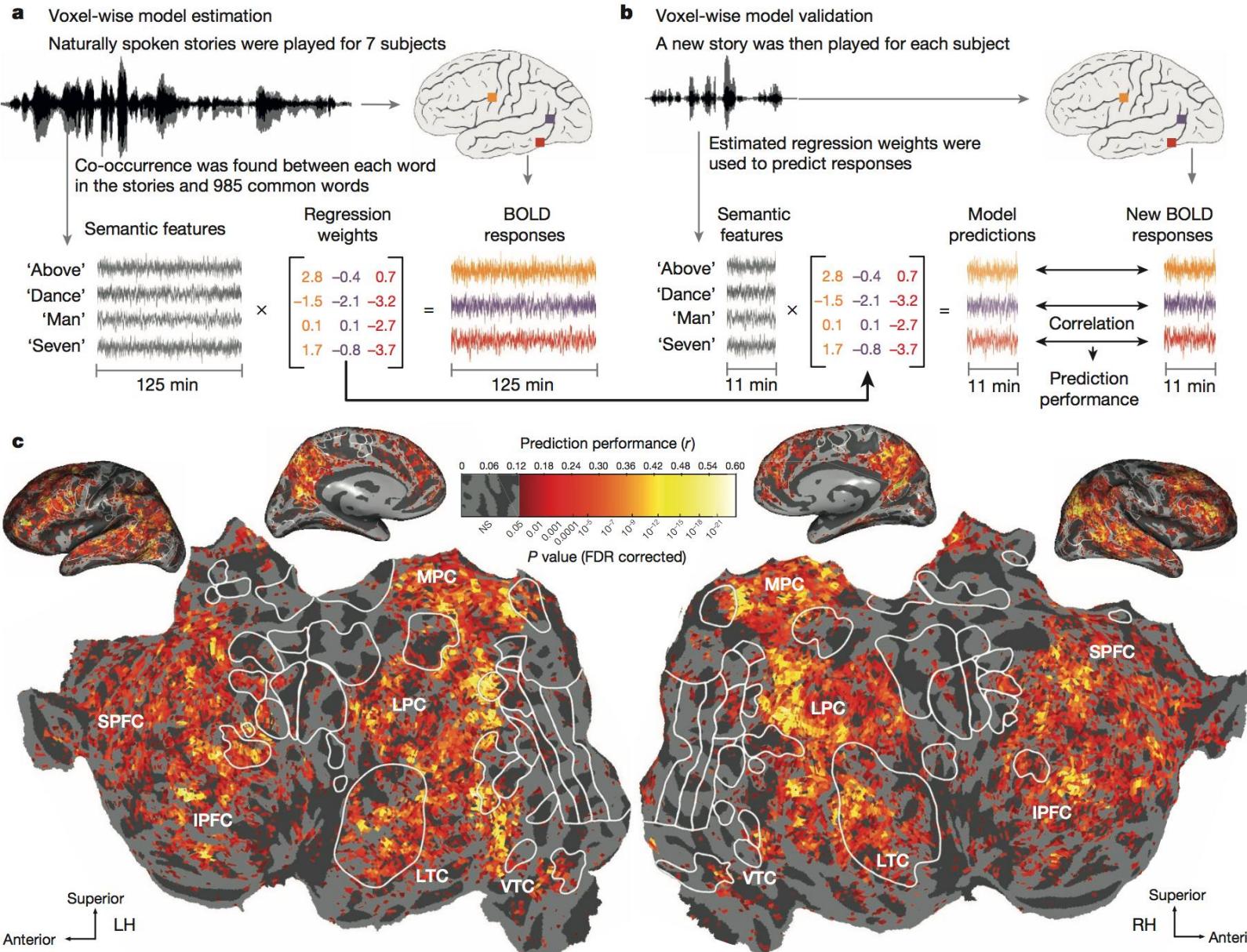


Fig. 3. Model performance on a next-word-prediction task selectively predicts brain scores. (A) Next-word-prediction task performance was evaluated as the surprisal between the predicted and true next word in the WikiText-2 dataset of 720 Wikipedia articles, or perplexity (x axis, lower values are better; training only a linear readout leading to worse perplexity values than canonical fine-tuning, see *Methods*, section 8). Next-word-prediction task scores strongly predict brain scores across datasets (*Inset*: this correlation is significant for two individual datasets: *Pereira2018* and *Blank2014*; the correlation for *Fedorenko2016* is positive but not significant). (B) Performance on diverse language tasks from the GLUE benchmark collection does not correlate with overall or individual-dataset brain scores (*Inset*; *SI Appendix*, *SI-2*; training only a linear readout). (C) Correlations of individual tasks with brain scores. Only improvements on next-word prediction lead to improved neural predictivity.

Semantic maps in human brain

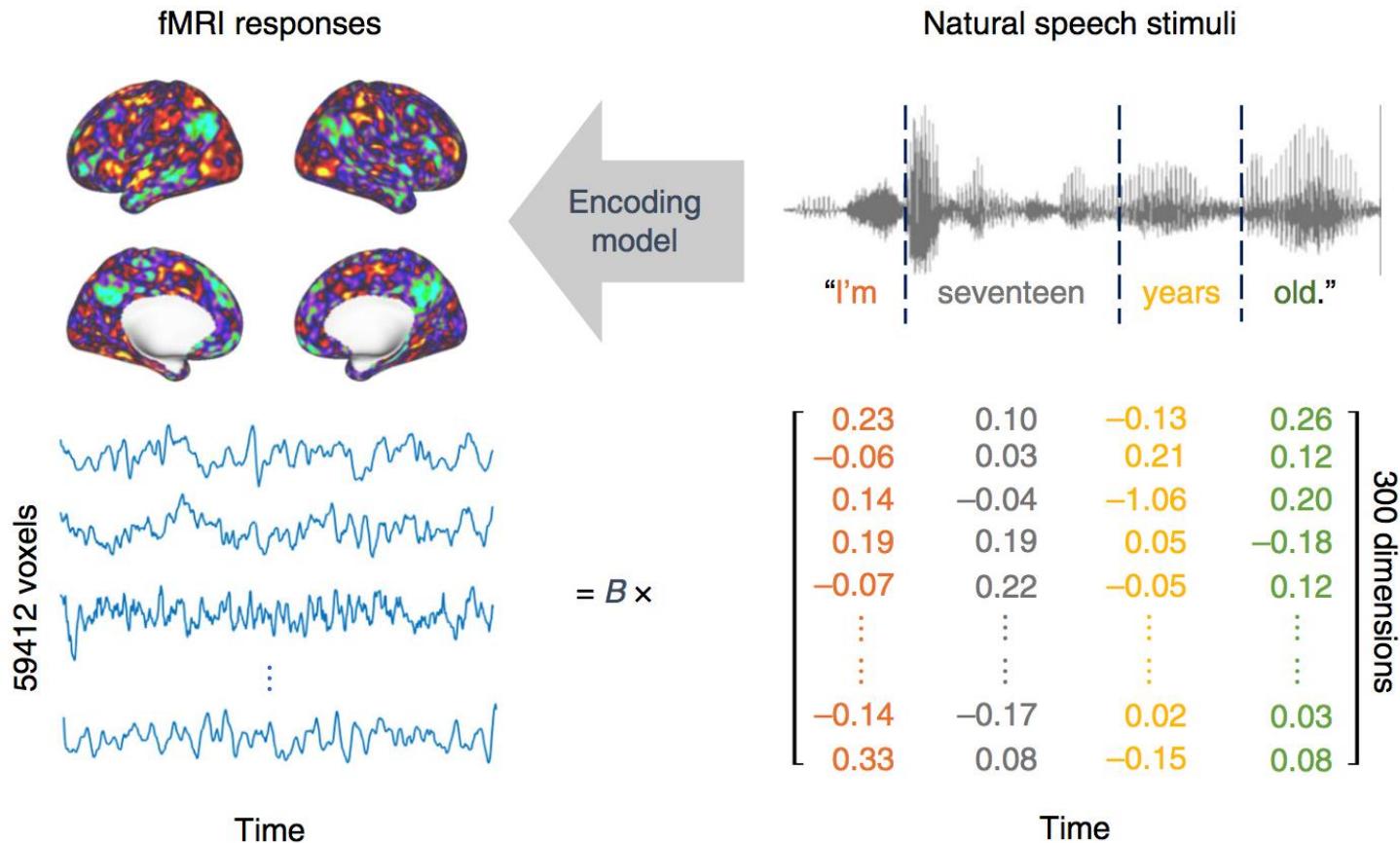


Jack Gallant group, UC Berkeley

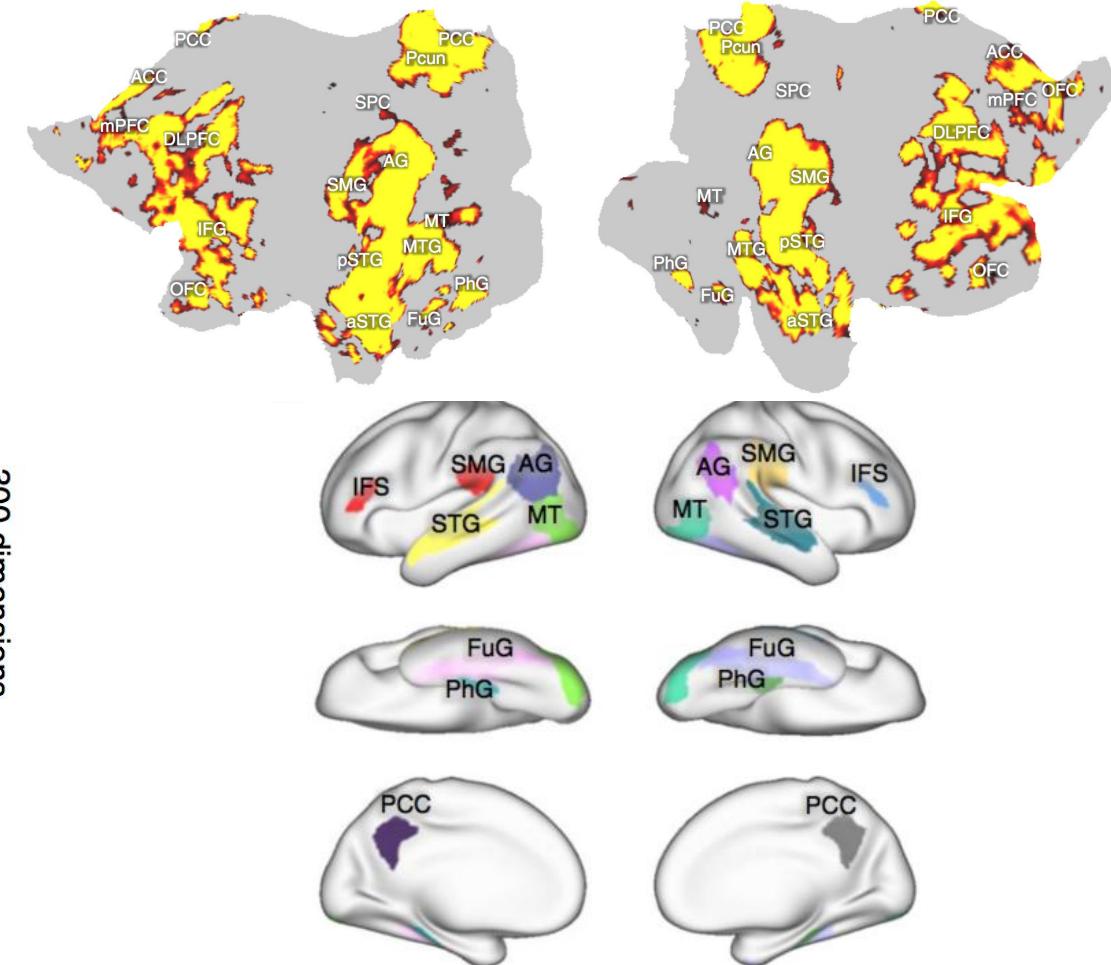
The embedding space was constructed by computing the **normalized co-occurrence** between each word and a set of **985 common English words** (such as 'above', 'worry' and 'mother') across a large corpus of English text.

Words related to the same semantic domain tend to occur in similar contexts, and so have similar co-occurrence values.

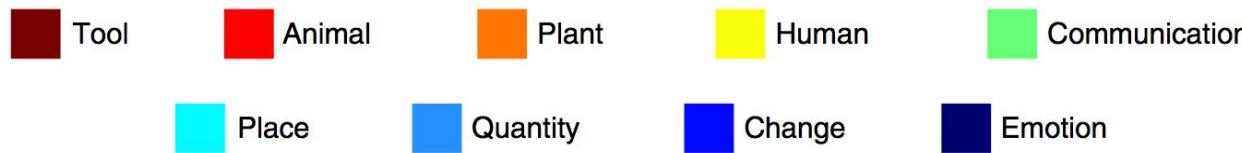
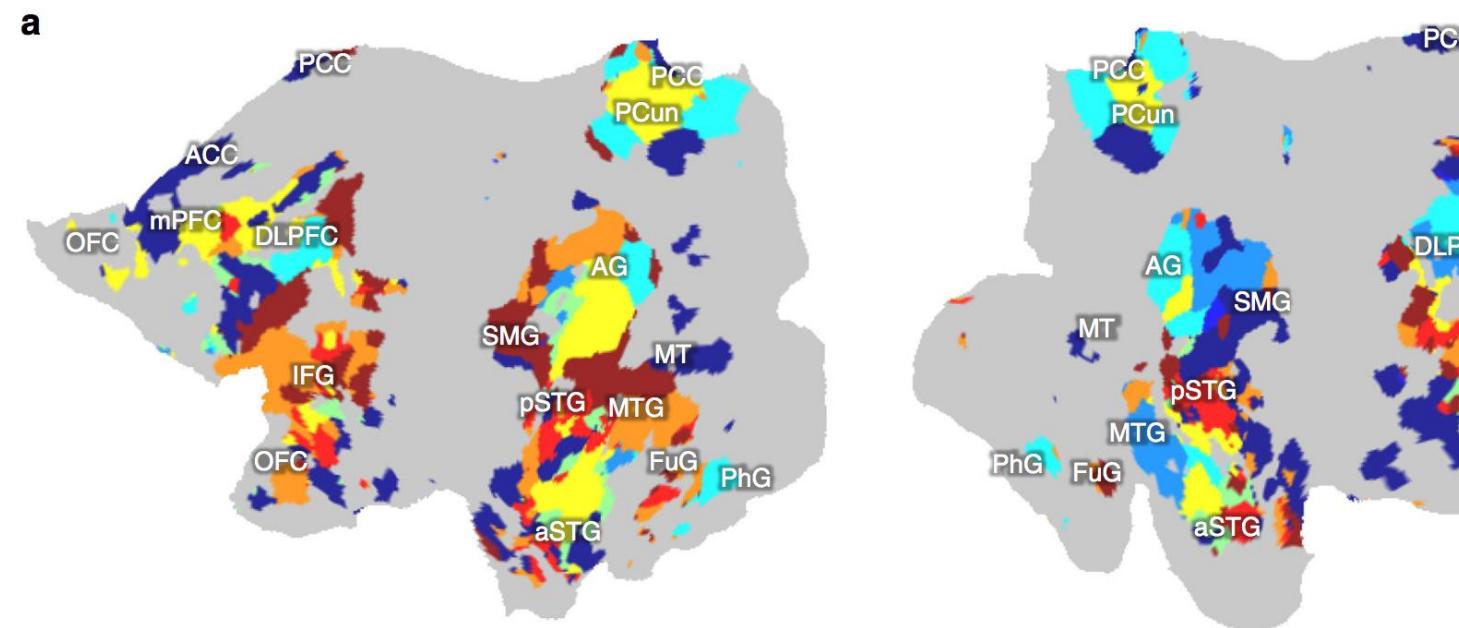
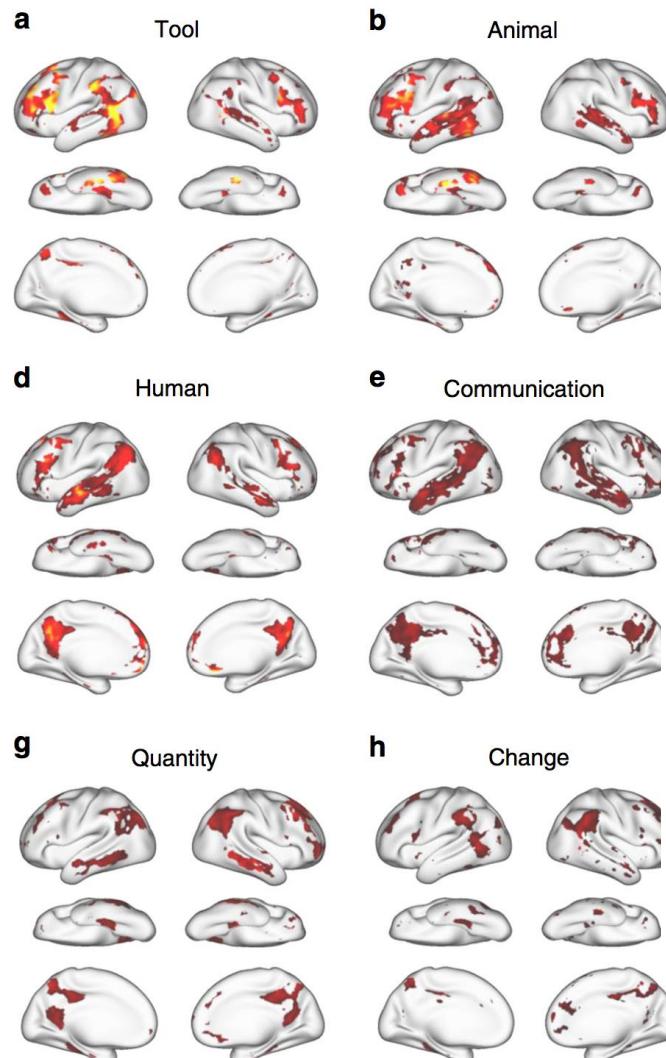
Reconstruction of semantic relations in human brain



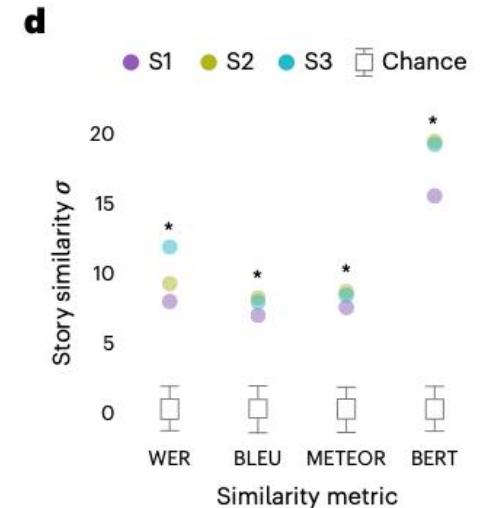
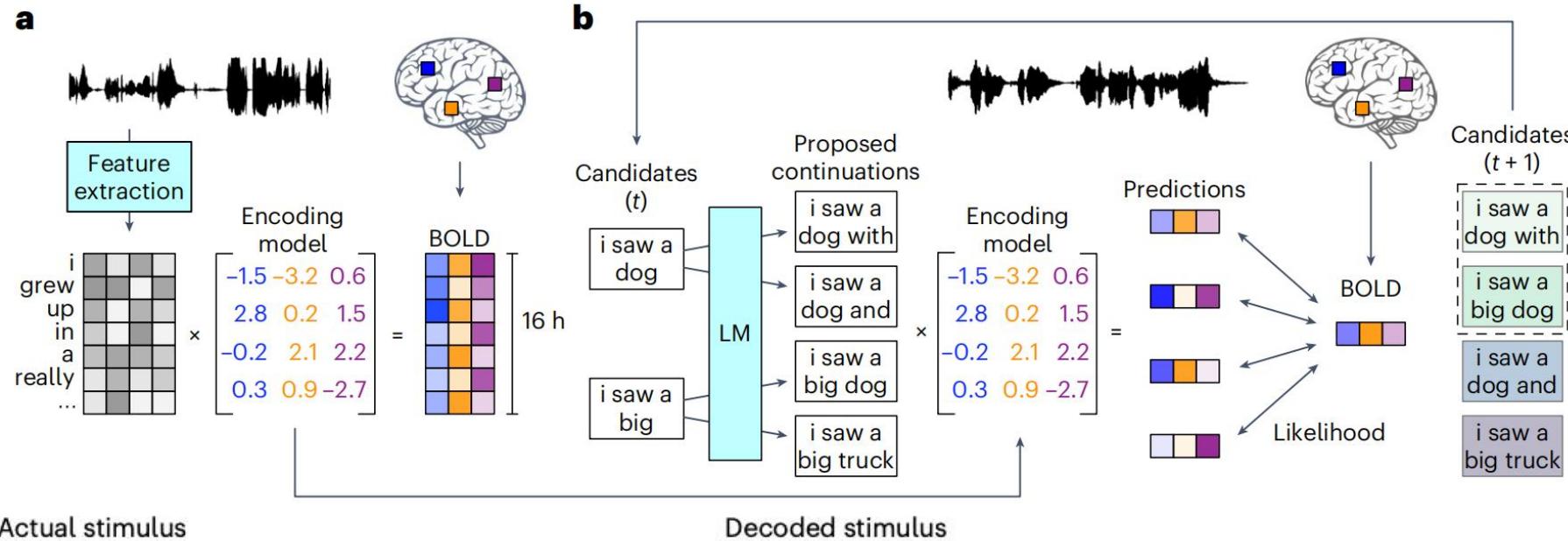
A map of the semantic system in brain



Semantic reconstruction of continuous language from fMRI



Semantic reconstruction of continuous language from fMRI



Brain-AI alignment

at semantic level

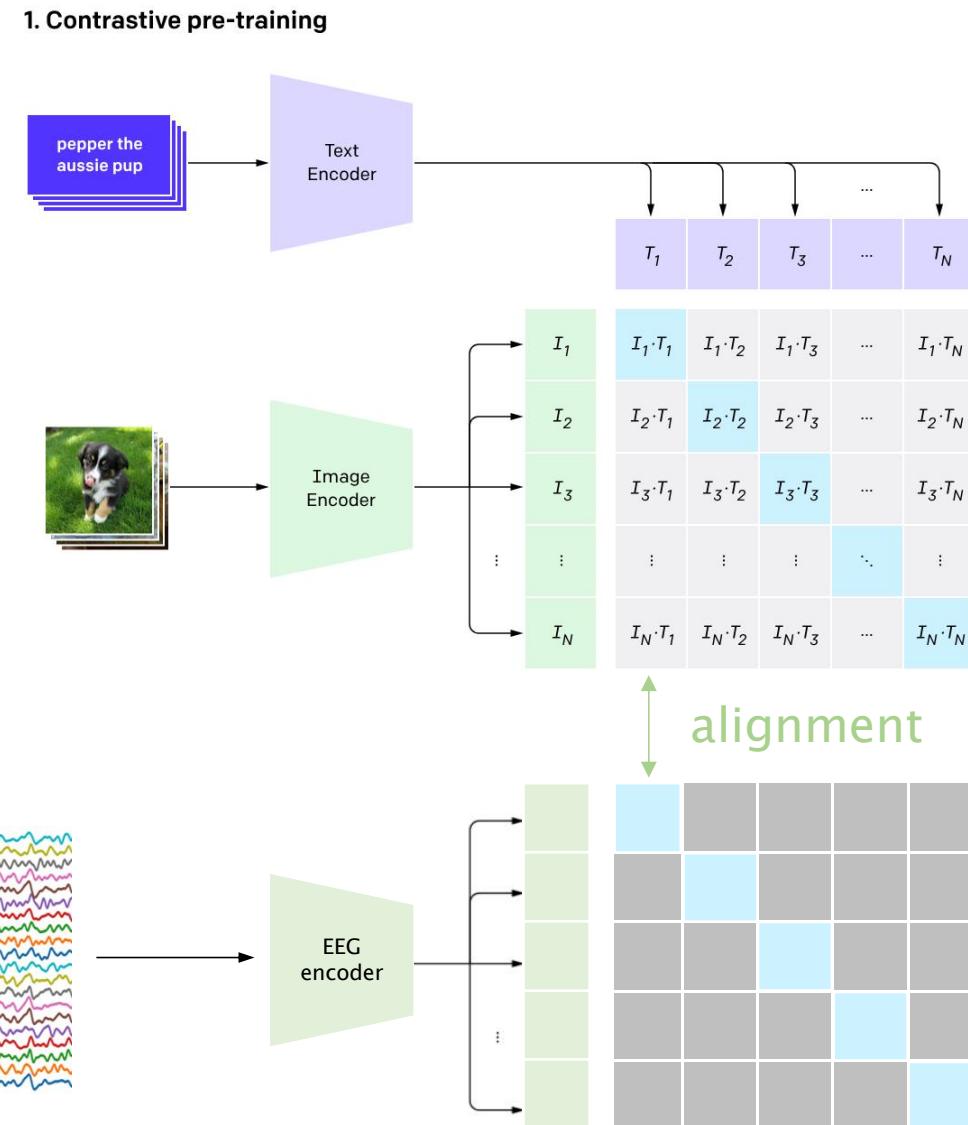
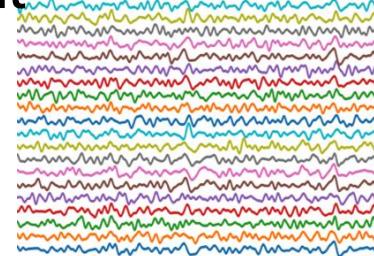
Brain-AI alignment: at semantic level

CLIP

一只小猫咪



**Brain-AI
alignment**



Benefits for AI

- 1) better abstraction
 - 2) better generalizability
 - 3) Interpretability
 - 4) AI safety
- ...

Benefits for neuroscience:

- 1) less neural data
 - 2) multiple downstream tasks
 - 3) zero-shot / few-shot capability
 - 4) Virtual experimental platform
 - 5) New science discovery
- ...

EEG based language decoding: models (EEG-to-text models)

EEG-to-text

Open Vocabulary Electroencephalography-to-Text Decoding and Zero-Shot Sentiment Classification

Zhenhailong Wang, Heng Ji

University of Illinois at Urbana-Champaign
 {wangz3, hengji}@illinois.edu

AAAI 2022

DeWave

DeWave: Discrete EEG Waves Encoding for Brain Dynamics to Text Translation

NeurIPS 2023

Yiqun Duan*, Jinzhao Zhou, Zhen Wang, Yu-Kai Wang, Chin-Teng Lin[†]

Belt-2

BELT-2: Bootstrapping EEG-to-Language representation alignment for multi-task brain decoding

Charles Chau, Yiqun Duan, Yu-Cheng Chang, Thomas Do, Yu-Kai Wang, Chin-teng Lin
 rejected by ICLR 2024

Belt

BELTO: Bootstrapping Electroencephalography-to-Language Decoding and Zero-Shot Sentiment Classification by Natural Language Supervision

arXiv 2023

Jinzhao Zhou, Yiqun Duan, Yu-Cheng Chang, Yu-Kai Wang, Chin-Teng Lin*

CET-MAE

Enhancing EEG-to-Text Decoding through Transferable Representations from Pre-trained Contrastive EEG-Text Masked Autoencoder

Jiaqi Wang^{1,2}, Zhenxi Song^{1*}, Zhengyu Ma², Xipeng Qiu³, Min Zhang^{1,2}, Zhiguo Zhang^{1,2†}

¹School of Computer Science and Technology, Harbin Institute of Technology Shenzhen, China

² Peng Cheng Laboratory, China

arXiv 2024

³School of Computer Science, Fudan University, China

mhwjq1998@gmail.com, {songzhenxi, zhangmin2021, zhiguo Zhang}@hit.edu.cn
 mazhy@pcl.ac.cn, xpqiu@fudan.edu.cn

Decode Neural signal as Speech

arXiv 2024

Yiqian Yang^{*1} Yiqun Duan^{*2} Qiang Zhang¹ Renjing Xu^{1†} Hui Xiong^{1†}

EEG based language decoding: models (EEG-to-text models)

Open Vocabulary Electroencephalography-to-Text Decoding and Zero-Shot Sentiment Classification

BELTO: Bootstrapping Electroencephalography-to-Language Decoding and Zero-Shot Sentiment Classification by Natural

These model all face a major problem: **teacher forcing (TF)**

It leads to **data leakage**.

The trained EEG-to-text models with teacher forcing cannot be used in real life.

BELT-2: Bootstrapping EEG-to-Language representation alignment for multi-task brain decoding

Charles Chau, Yiqun Duan, Yu-Cheng Chang, Thomas Do, Yu-Kai Wang, Chin-teng Lin
rejected by ICLR 2024

Decode Neural signal as Speech

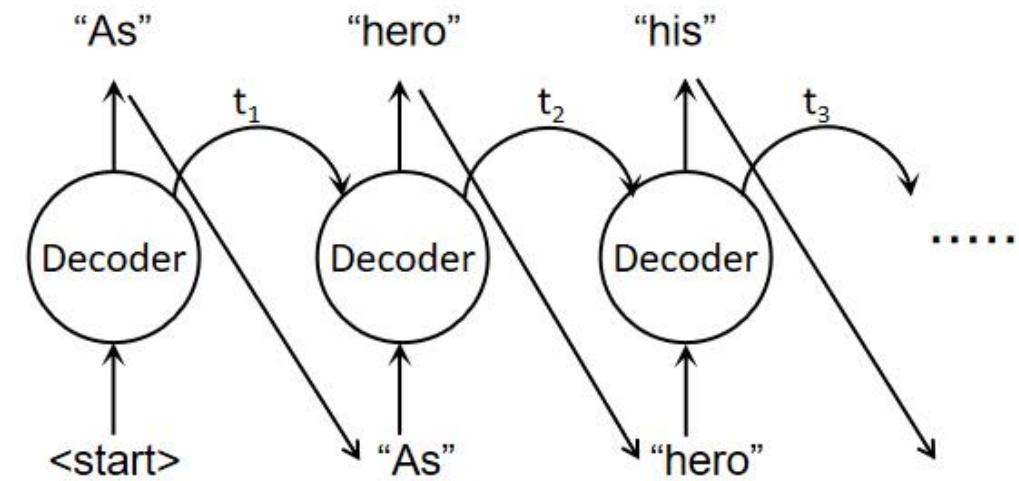
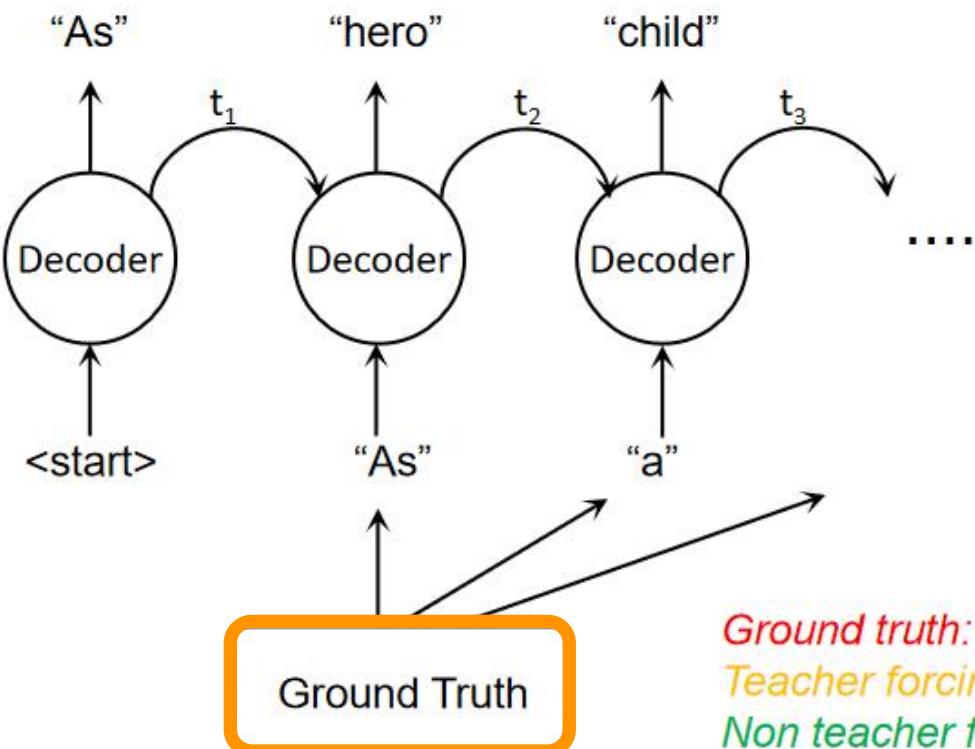
arXiv 2024

Yiqian Yang^{*1} Yiqun Duan^{*2} Qiang Zhang¹ Renjing Xu^{1†} Hui Xiong^{1†}

EEG/MEG-To-Text (teacher forcing / data leakage)

- Problem: EEG decoder aligned with language models (e.g., BART/T5 models) trained with teacher forcing relies on **the correct external prompt** during test.

Teacher forcing



Ground truth: As a child, his hero was Batman.

Teacher forcing: As hero, child was a Batman.

Non teacher forcing: As hero, his interests is music.

Our tests on the EEG-to-Text model v.s. Nat Neuro 2023 paper

EEG-To-Text (our experiments):

noise as input	teacher-forcing	BLEU-1↑	ROUGE-1f↑
yes	yes	0.2747	33.62
no	yes	0.2784	33.77
yes	no	0.0923	13.99
no	no	0.0887	13.56

target string: With his interest in race cars, he formed a second company, the Henry Ford Company.

predicted string : that current in mind and and he must a large part that which entire da.. It

target string: During this period, he personally drove his Quadricycle to victory in a race against Alexander Winton, a well-known driver and the heavy favorite on October 10, 1901.

predicted string : re dr re re re center re re center dr re re re j

target string: Ford was forced out of the company by the investors, including Henry M. Leland in 1902.

predicted string : and also to of the state to the end. and a and. I. and the. and it current is forcedined to a and It

fMRI-To-Text (NN 2023):

Table 1 | Language similarity scores

	WER	BLEU-1	METEOR	BERTScore
Null	0.9637	0.1908	0.1323	0.7899
Subject 1	0.9407	0.2331	0.1621	0.8077
Subject 2	0.9354	0.2426	0.1677	0.8104
Subject 3	0.9243	0.2470	0.1703	0.8116
Translation	0.7459	0.4363	0.3991	0.8797

c

Actual stimulus

i got up from the air mattress and pressed my face against the glass of the bedroom window expecting to see eyes staring back at me but instead finding only darkness

i didn't know whether to scream cry or run away instead i said leave me alone i don't need your help adam disappeared and i cleaned up alone crying

that night i went upstairs to what had been our bedroom and not knowing what else to do i turned out the lights and lay down on the floor

i don't have my driver's license yet and i just jumped out right when i needed to and she says well why don't you come back to my house and i'll give you a ride i say ok

Decoded stimulus

i just continued to walk up to the window and open the glass i stood on my toes and peered out i didn't see anything and looked up again i saw nothing

started to scream and cry and then she just said i told you to leave me alone you can't hurt me anymore i'm sorry and then he stormed off i thought he had left i started to cry

we got back to my dorm room i had no idea where my bed was i just assumed i would sleep on it but instead i lay down on the floor

she is not ready she has not even started to learn to drive yet i had to push her out of the car i said we will take her home now and she agreed

Exact

Gist

Error

Our tests on the SOTA EEG-to-Text model v.s. Nat Neuro 2023 paper

EEG-To-Text (our experiments):

noise as input	teacher-forcing	BLEU-1↑	ROUGE-1f↑
yes	yes	0.2747	33.62
no	yes	0.2784	33.77
yes	no	0.0923	13.99
no	no	0.0887	13.56

fMRI-To-Text (NN 2023):

Table 1 | Language similarity scores

	WER	BLEU-1	METEOR	BERTScore
Null	0.9637	0.1908	0.1323	0.7899
Subject 1	0.9407	0.2331	0.1621	0.8077
Subject 2	0.9354	0.2426	0.1677	0.8104
Subject 3	0.9243	0.2470	0.1703	0.8116
Translation	0.7459	0.4363	0.3991	0.8797

The problem has not been solved yet.

Mind reading from EEG/fMRI has long way to go.

Exact

Gist

Error

predicted string : and also to of the state to the end. and a and. I. and the. and it current is forcedined to a and It

Jerry Tang, Amanda LeBel, Shailee Jain, Alexander G. Huth (2023). Nature Neuroscience

Sementic stimulation
to modulate neural activity

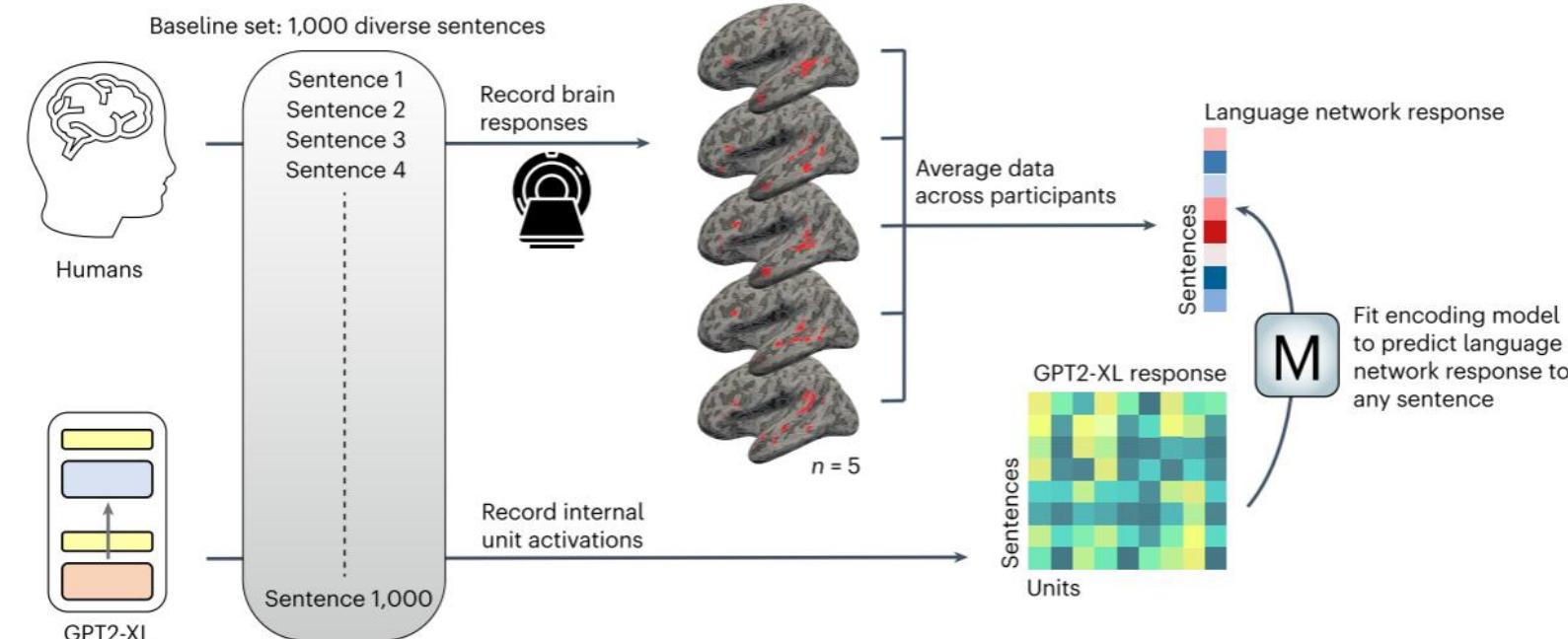
Sementic stimulation

to modulate neural activity

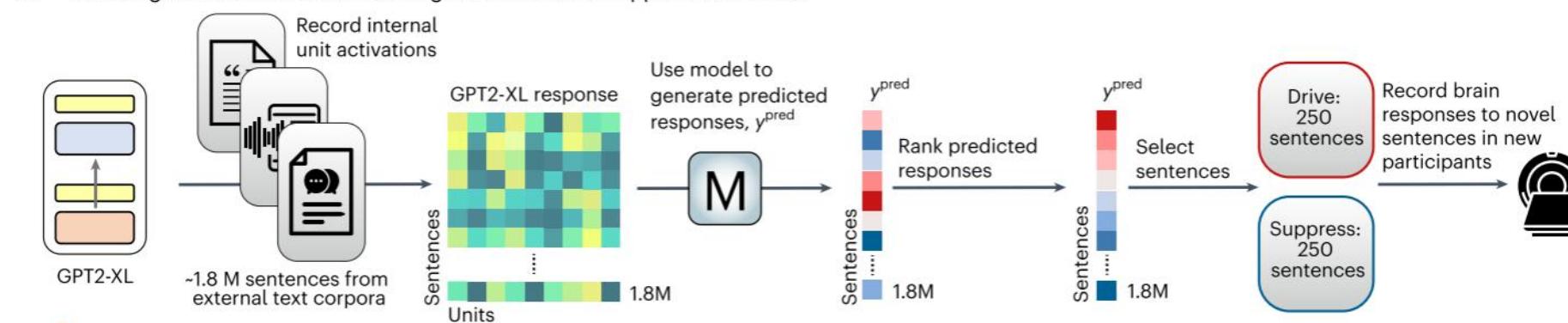
例如，夸夸大师 vs PUA大师

Stimulus selection to modulate neural activity

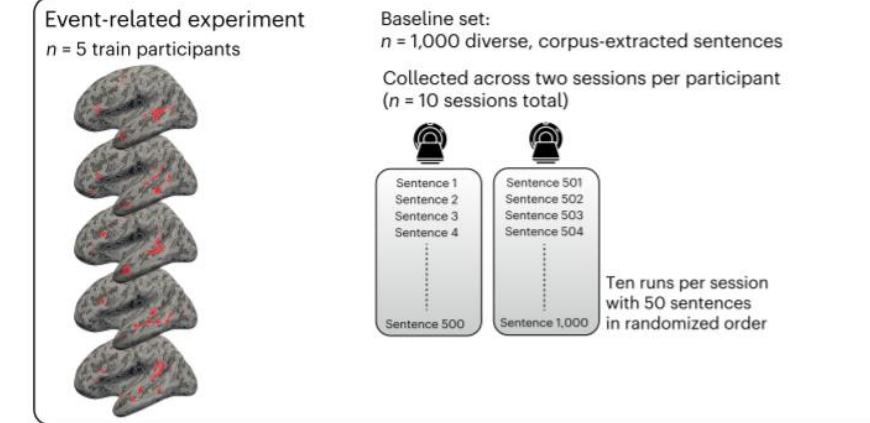
a Encoding model development: fitting a predictive model of the language network



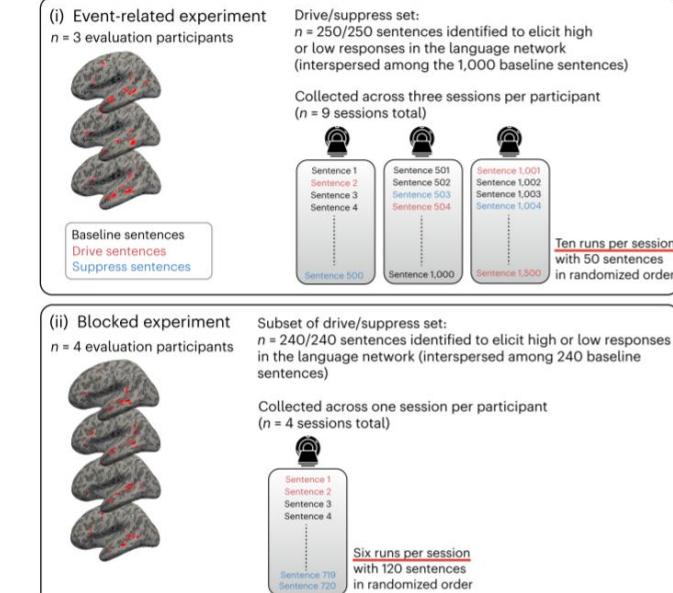
b Encoding model evaluation: selecting novel drive and suppress sentences



a Encoding model development



b Encoding model evaluation



Drive v.s. Suppress

a Brain ROIs

Masks used for defining the language network in individual participants

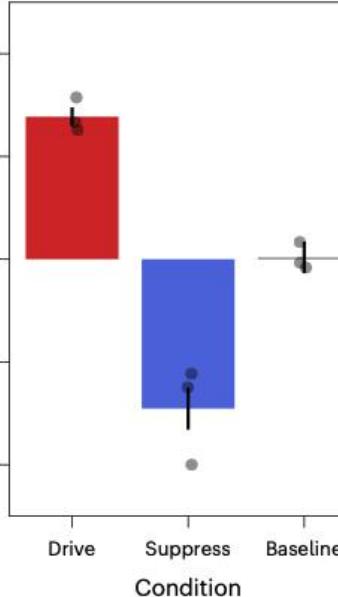


b Event-related design

$n = 3$ participants



z-scored BOLD response (mean \pm within-participant s.e.m.)

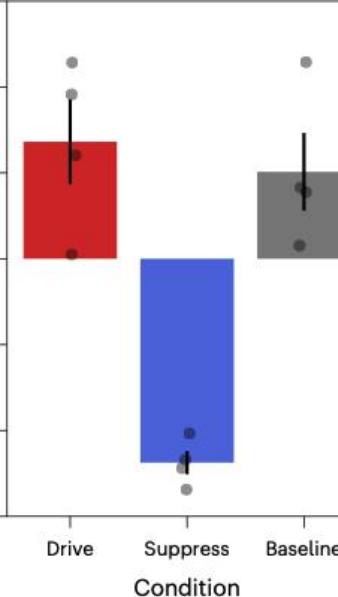


c Blocked design

$n = 4$ participants



z-scored BOLD response (mean \pm within-participant s.e.m.)



d Sentence examples from each condition

Drive

Sentences identified to elicit maximal response in the language network

Changing PhD group: Yes or Not?
Notice how you reacted to WTF.
Add, some feminists are call male.
Jiffy Lube of -- of therapies, yes.
People on Insta Be Like, "Gross!"
Buy sell signals remains a particular.
Turin loves me not, nor will.
URL right, or report reviewing Vimeo.

Suppress

Sentences identified to elicit minimal response in the language network

We were sitting on the couch.
That is such a beautiful picture!
They stood there for a moment.
They went up the stairs together.
Inside was a tiny silver sculpture.
They walked out onto the balcony.
Cas gazed up at the sky.
What else is there to do?

Baseline

Sentences sampled from diverse naturalistic text corpora

The judge spoke, breaking the silence.
How to create a personalized chart.
SEO- short for search engine optimization.
Were all the vampire stories true?
A king must have an heir.
Her wisdom and foresight is evident.
Wet hair clinging to her cheeks.
He also admitted killing Mrs. Hengesbach..

Drive v.s. Suppress

Sentence-level brain responses versus predictions

Functionally defined LH language network



(Masks used for defining the language network illustrated)

They made it into the hallway.

There's a tub in the kitchen.

She wore a short black dress.

The neatly trimmed beard and moustache.

He felt his stomach for blood.

Oh Hjalmar Wentz must have been.

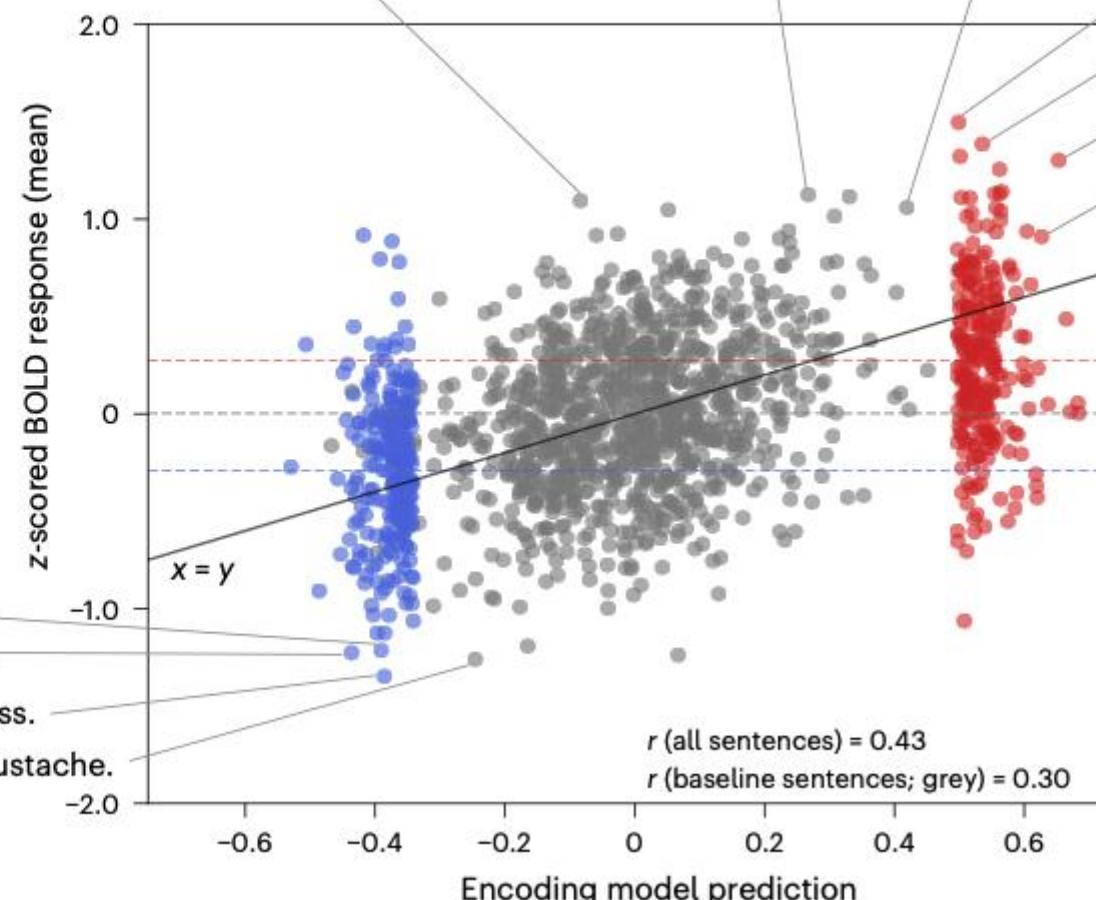
Both mentally and physically, you're attracted.

I'm progressive and you fall right.

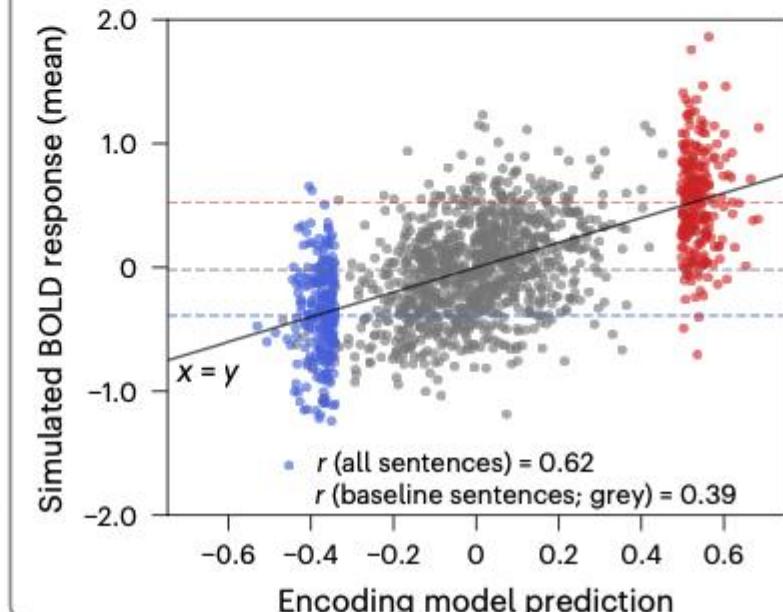
You collaborate request shows So expect!

Domain Wikileaks gone; access is NOT..

Guest started a topic LIVER Connection??

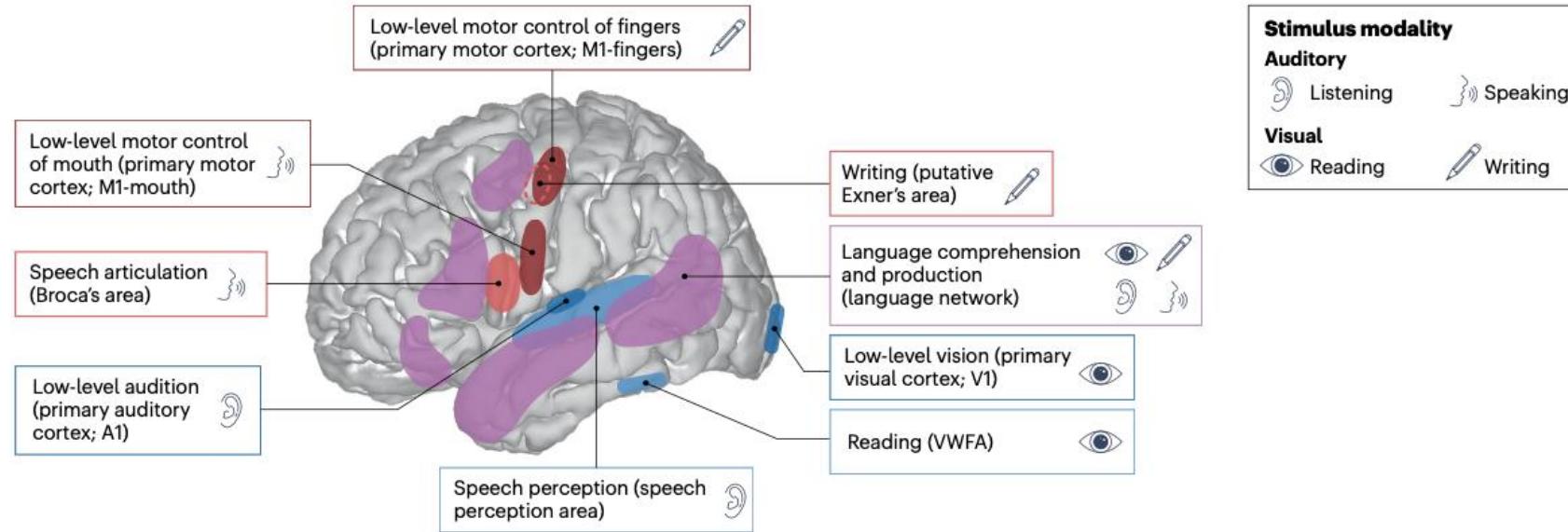


Simulated brain responses versus predictions

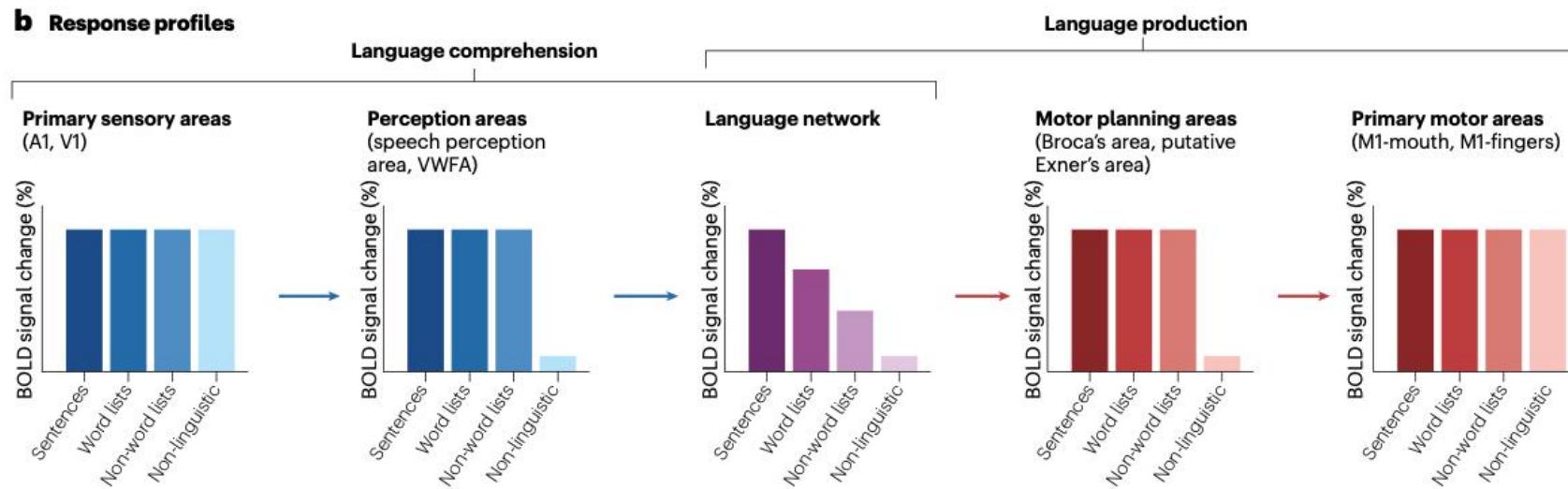


Brain areas responding to linguistic stimuli

a Brain areas that respond to linguistic stimuli



b Response profiles



Do birds have language?

I and my collaborator (a bird neuroscientist) would like to get published in Nature based on the idea of this birds' vocal communication project. What should be our scientific question? How can we design the study to answer this scientific question to be able to get published in Nature?

Formulating the Scientific Question

To get published in a high-impact journal like *Nature*, your study should address a groundbreaking and broadly significant scientific question. Here's a potential question:

Scientific Question:

Do birds use structured, rule-based vocal communication analogous to human language, and can we identify the neural correlates of this communication?

Key Aspects of the Scientific Question:

1. **Language Analogy:** The question draws a parallel between bird vocal communication and human language, which is a topic of broad interest in both neuroscience and linguistics.
2. **Structured Communication:** It focuses on whether birds use vocalizations in a structured, rule-based manner, moving beyond mere calls to something that might be considered proto-language.
3. **Neural Correlates:** Investigating the brain mechanisms underlying this communication can provide insights into the evolution of language and cognition, making it highly relevant to both neuroscience and evolutionary biology.
4. **Interdisciplinary Appeal:** The question bridges the fields of neurobiology, ethology, linguistics, and AI, appealing to a wide range of *Nature* readers.

Study Design to Answer the Scientific Question

1. Hypothesis Development**

- **Primary Hypothesis:** Birds exhibit structured, rule-based vocal communication that is functionally analogous to human language, and specific neural circuits underlie this communication.
- **Sub-Hypotheses:**
 - Birds' vocalizations follow syntactic rules.
 - Bird responses to specific vocalizations are predictable and consistent, indicating comprehension.
 - There are distinct neural correlates associated with specific vocal patterns and their comprehension.

2. Species Selection**

- Choose a bird species known for its complex vocalizations, such as zebra finches or European starlings.
- The species should have a well-documented repertoire of vocal behaviors and a social structure where communication plays a critical role.

3. Experimental Phases**

Phase 1: Data Collection**

- **Vocal and Behavioral Recordings:**
 - Collect extensive vocal and behavioral data from both wild and captive birds to capture natural communication.
 - **Neural Recordings:**
 - Use techniques such as in vivo electrophysiology, calcium imaging, or optogenetics to record neural activity in awake, behaving birds during vocalization and listening.

Phase 2: AI-Based Analysis**

- **Model Training:**
 - Train AI models on the vocal data to identify patterns, sequences, and potential rules in the vocalizations.
 - Use models like transformers or recurrent neural networks to analyze and generate vocalizations.
- **Latent Space Exploration:**
 - Use the latent space of the trained models to generate new, controlled vocalizations that vary in specific ways (e.g., syntax, tone, rhythm).

Phase 3: Behavioral and Neural Testing**

- **Playback Experiments:**
 - Play the AI-generated vocalizations to other birds and record their behavioral and neural responses.
- **Predictive Modeling:**
 - Use machine learning to predict behavioral and neural responses based on the structure of the vocalizations.

Phase 4: Experimental Manipulation**

- **Syntactic Disruption:**
 - Test birds' responses to syntactically altered or disrupted vocalizations to see if they recognize the change.
- **Neural Circuit Manipulation:**
 - Use optogenetics or pharmacological methods to disrupt specific neural circuits and observe changes in vocal behavior and communication.

4. Statistical and Computational Analysis**

- **Behavioral Analysis:**
 - Quantify the consistency and specificity of birds' responses to structured vs. random or disrupted vocalizations.
- **Neural Analysis:**
 - Identify and analyze neural patterns associated with specific vocal sequences, testing for rule-based processing.
- **Cross-Species Comparisons:**
 - Compare findings with known aspects of human language processing to highlight potential evolutionary parallels.

5. Broader Implications**

- **Evolution of Language:** Discuss how findings may provide insights into the evolution of language and cognitive processing in non-human animals.
- **AI and Neurobiology:** Highlight the interdisciplinary nature of the study, showing how AI models can be used to uncover new aspects of animal communication and cognition.

Ensuring Publication in Nature**

- **Novelty:** Ensure the study presents novel insights that significantly advance the understanding of animal communication and its parallels to human language.
- **Interdisciplinary Impact:** Emphasize the study's relevance across multiple disciplines, including neuroscience, linguistics, AI, and evolutionary biology.
- **Rigorous Methodology:** Use state-of-the-art techniques and rigorous experimental design to ensure the robustness and replicability of the findings.
- **Broader Relevance:** Frame the findings in a way that addresses big questions in science, such as the origins of language and cognition.

By addressing a fundamental question about the nature of communication and cognition, and by using innovative AI and neurobiological methods, your study has a strong potential for publication in *Nature*.

Summary Lecture 4 – Language processing

- **What is language**
- **The language system in brain**
- **Aphasia**
 - Broca's Aphasia
 - Wernicke's aphasia
- **Models of Language processing**
 - Wernicke–Geschwind model
 - Parallel language pathway model
- **Language models in NLP**
 - Statistical Language model
 - Neural network language model
- **NLP in AI v.s. in brain**

Reading materials

- *Neuroscience: Exploring the brain* (3rd ed), **Chapter 19 – Language**, pp685-717
- Nature communications 2020, Connecting concepts in the brain by mapping cortical representations of semantic relations
- Nature neuroscience 2021, Visual and linguistic semantic representations are aligned at the border of human visual cortex
- Schrimpf et al (2021) PNAS, The neural architecture of language
- Tuckute et al (2024) Nature Human Behaviour
- Evelina Fedorenko, Anna A. Ivanova, Tamar I. Regev (2024) Nature Reviews Neuroscience