Multi-view and Cross-view Brain Decoding

August 13, 2022

COLING 2022

Subba Reddy Oota[1,2], Jashn Arora[2], Manish Gupta[2,3], Bapi Raju Surampudi[2]

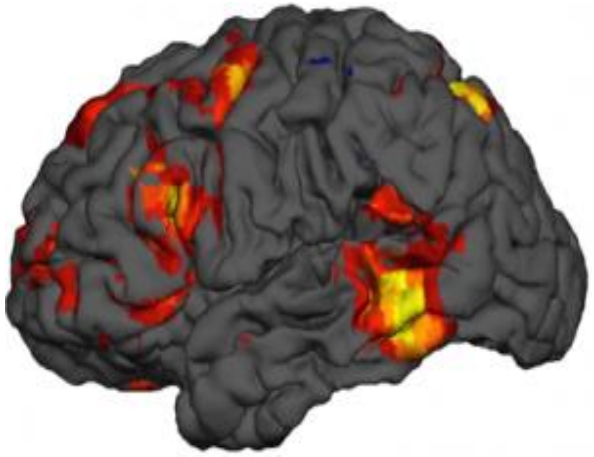
[1]Inria Bordeaux France, [2]IIIT-Hyderabad, [3]Microsoft India

What is fMRI?

A vision-language task in the scanner

(Bird)

Concept + Picture



fMRI Brain Activity

https://www.biopac.com/events/fmri-psych/

Encoding vs. Decoding

Stimulus Representation

Encoding

fMRI

At three o'clock precisely I was at Baker Street, but Holmes had not yet returned. The landlady informed me that he had left the house shortly after eight o'clock …

It was close upon four before the door opened, and a drunken-looking groom, ill-kempt and side-whiskered, with an inflamed face and disreputable clothes, walked into the room. Accustomed as I was to my friend's amazing powers in the use of disguises, I had to look three times before I was certain that it was indeed he.

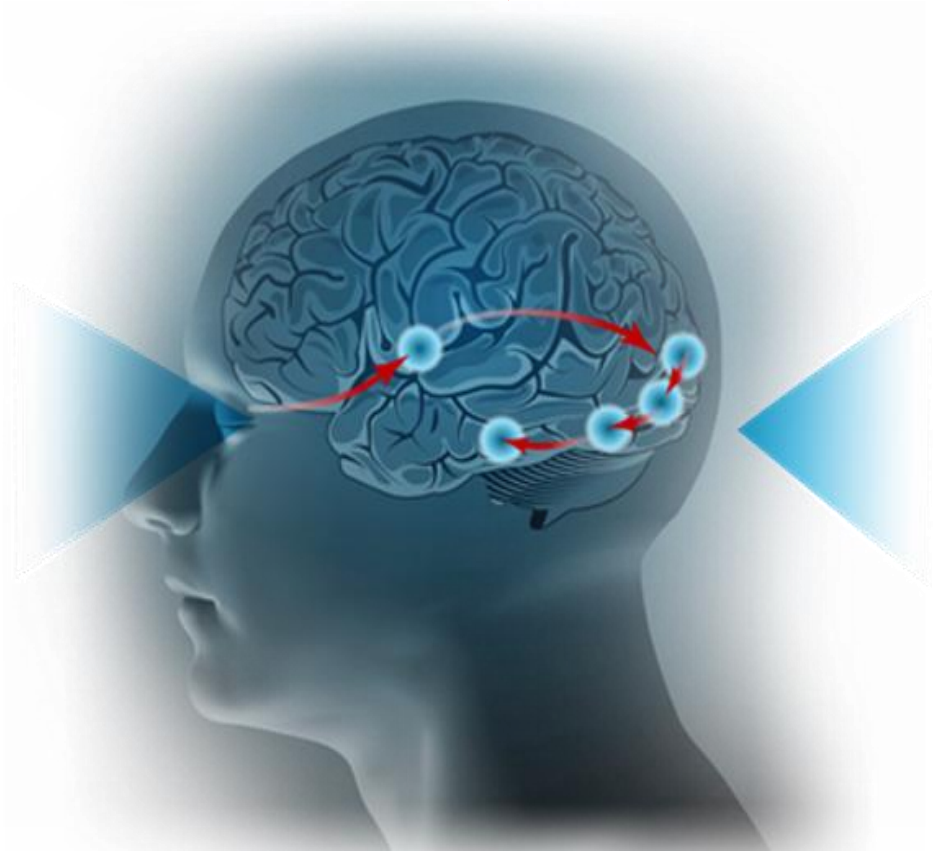
"Well, really!" he cried, and then he choked; and laughed again until he was obliged to lie back, limp and helpless, in the chair.

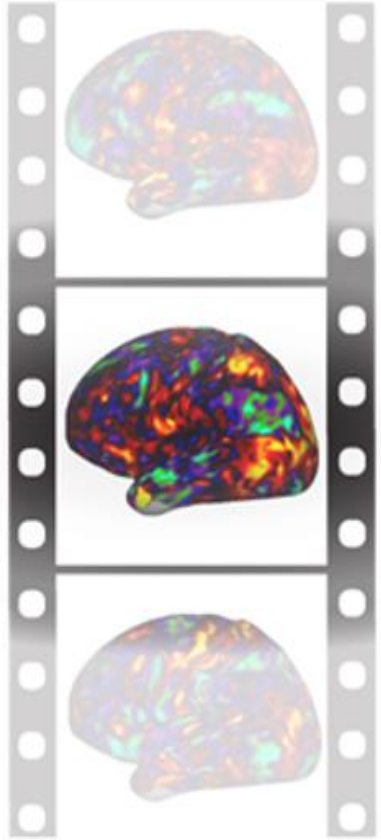
"What is it?"

"It's quite too funny. I am sure you could never guess how I employed my morning."

"I can't imagine. I suppose that you have been watching the habits, and perhaps the house, of Miss Irene Adler."

"Quite so; but the sequel was rather unusual. I will tell you, … I soon found Briony Lodge. It is a bijou villa, with a garden at the back, but built out in front right up to the road, …





Stimulus Representation

Decoding

fMRI

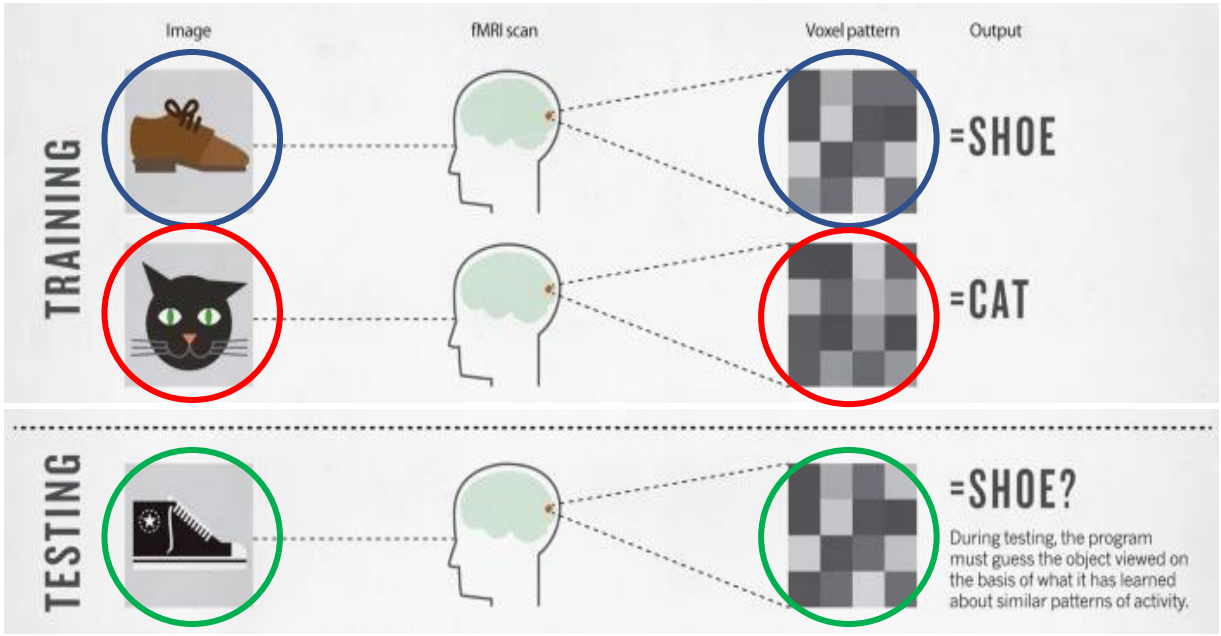
Haiguang Wen et al, 2017

Coling-2022

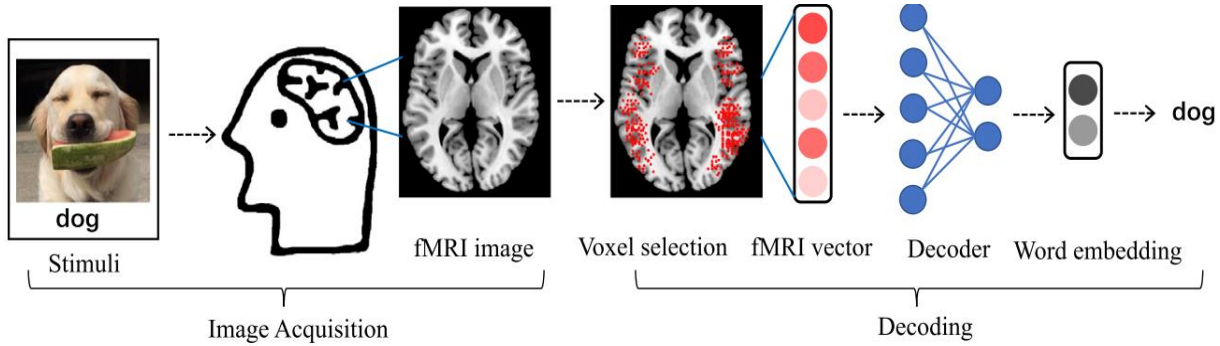
What is Brain Decoding?

- Can you read the mind with fMRI?
- Or at least tell what the person saw?

Visual Task



Language Task

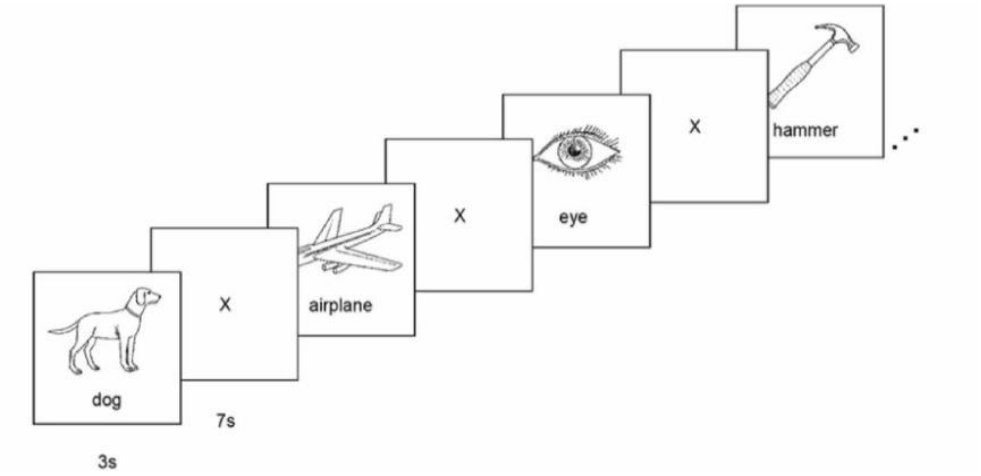


Smith et al., 2011, Wang et al. 2019

Coling-2022

Classical Decoders

- Classical decoding solutions extracting linguistic meaning from imaging data have been largely limited to
  - concrete nouns,
  - using similar stimuli for training and testing,
  - small number of semantic categories.

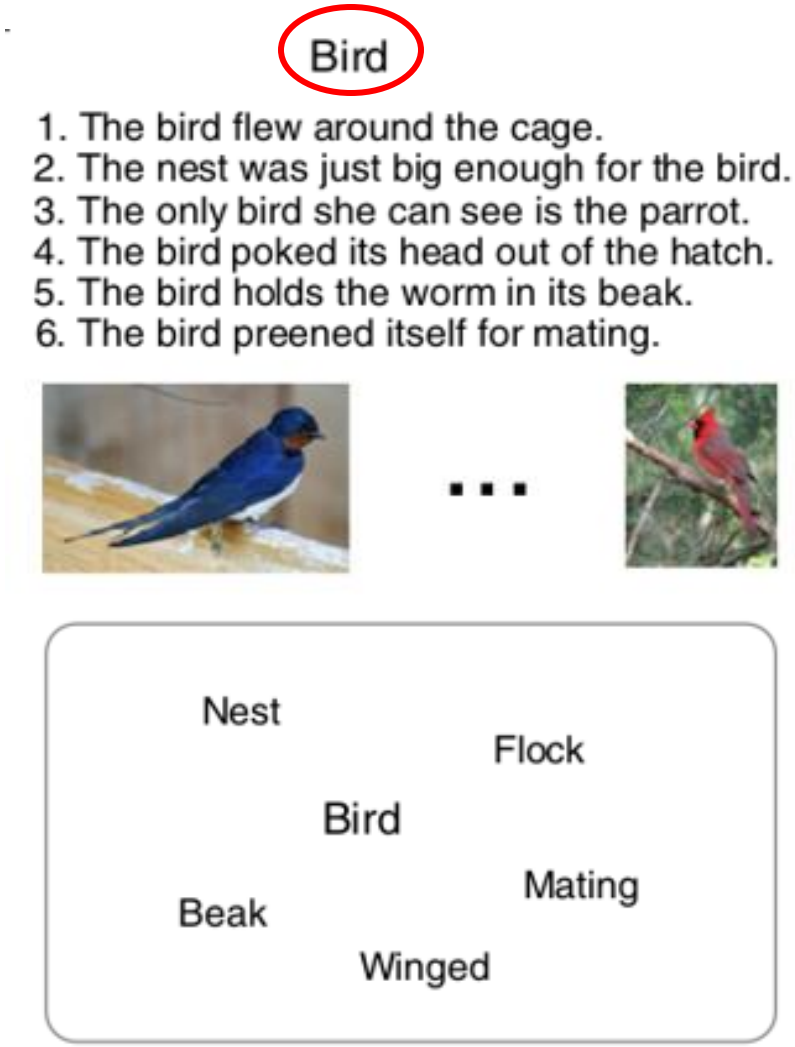


Category	Exemplar 1	Exemplar 2
animals	bear	cat
body parts	arm	eye
buildings	apartment	barn
building parts	arch	chimney
clothing	coat	dress
furniture	bed	chair
insects	ant	bee
kitchen utensils	bottle	cup
man made objects	bell	key
tools	chisel	hammer
vegetables	carrot	celery
vehicles	airplane	bicycle

Mitchell et al. 2008

Multi-view and Cross-view Brain Decoding

- Human brains have the unique capability of language acquisition:
  - the process of learning the language
  - understand the meaning of concepts from multiple modalities such as images, text, speech, and videos.
- Prior works focus on single-view brain decoding using traditional feature engineering.
- However, how the brain captures the meaning of linguistic stimuli across multiple views is still a critical open question in neuroscience.
- Consider three different views of the concept bird:
  - (1) sentence using the target word,
  - (2) picture presented with the target word label, and
  - (3) word cloud containing the target word along with other semantically related words.
- Earlier works have explored which of these three different views provides richer information to understand the concept.



Dataset Details (Experiment-1)

Concept Word

Bird

1. The bird flew around the cage.
2. The nest was just big enough for the bird.
3. The only bird she can see is the parrot.
4. The bird poked its head out of the hatch.
5. The bird holds the worm in its beak.
6. The bird preened itself for mating.

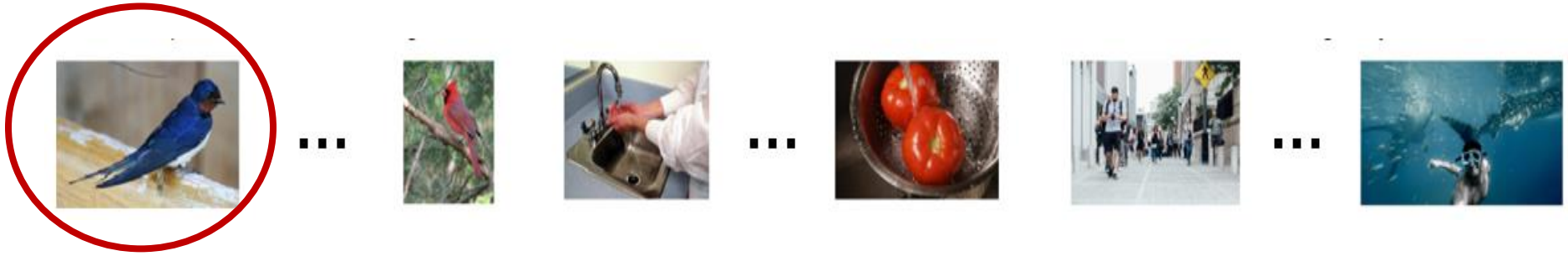
Wash

1. To make the counter sterile, wash it.
2. The dishwasher can wash all the dishes.
3. He likes to wash himself with bar soap.
4. She felt clean after she could wash herself.
5. You have to wash your laundry beforehand.
6. The maid was asked to wash the floor.

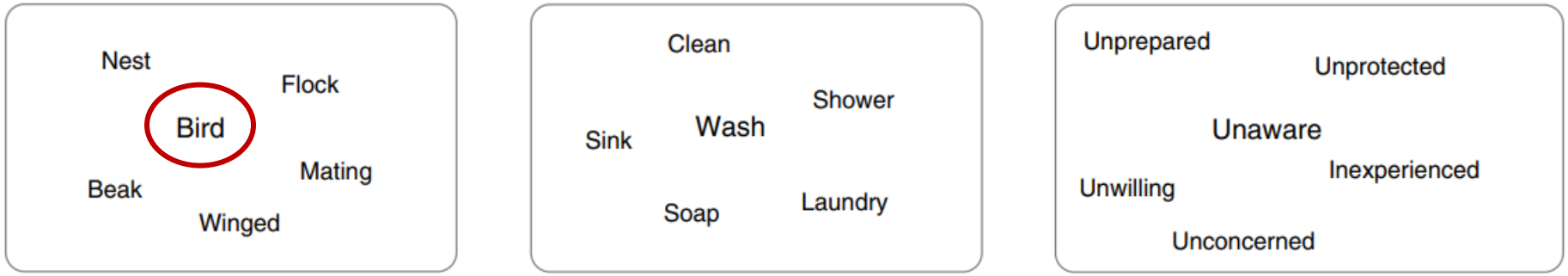
Unaware

1. She was unaware of how oblivious he really was.
2. She was unaware of her status.
3. Unprejudiced and unaware, she went full throttle.
4. Unaware of current issues, he is a terrible candidate.
5. He was unaware of how uninterested she was.
6. He was unaware of the gravity of the situation.

⇨ Concept + Sentence View



⇨ Concept + Picture View



⇨ Concept + Wordcloud View

Periera et al. 2018

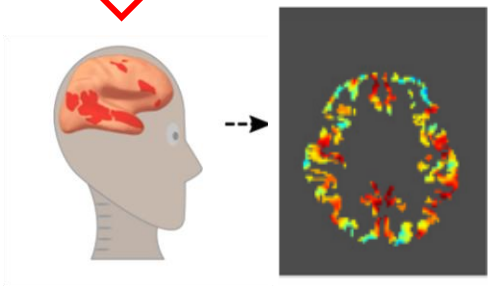
Coling-2022

Informative Voxel Selection

Stimulus: Apartment

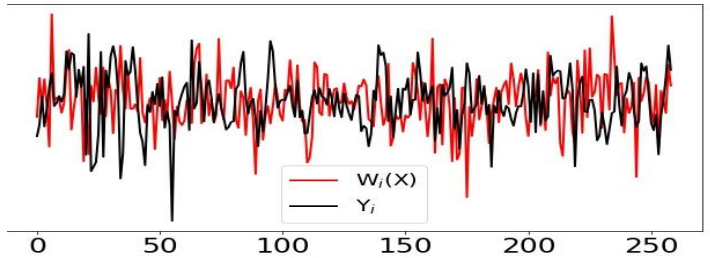
Present

3D Image



Correlation across feature dimensions

Pearson Correlation (R) = Corr(Y, W(X))



X W Y

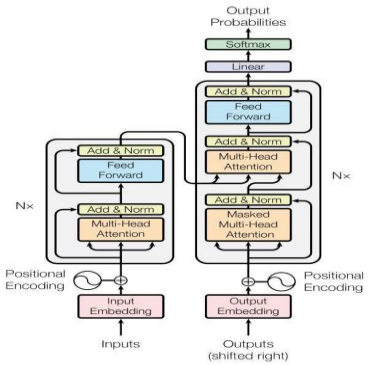
Input → Ridge Regression → Output

Text semantic vector for "apartment"

Voxel + 26 neighbors in 3D

Stimulus: Apartment

Present



BERT

$V_1 - R_1$
$V_2 - R_2$
....
$V_n - R_3$

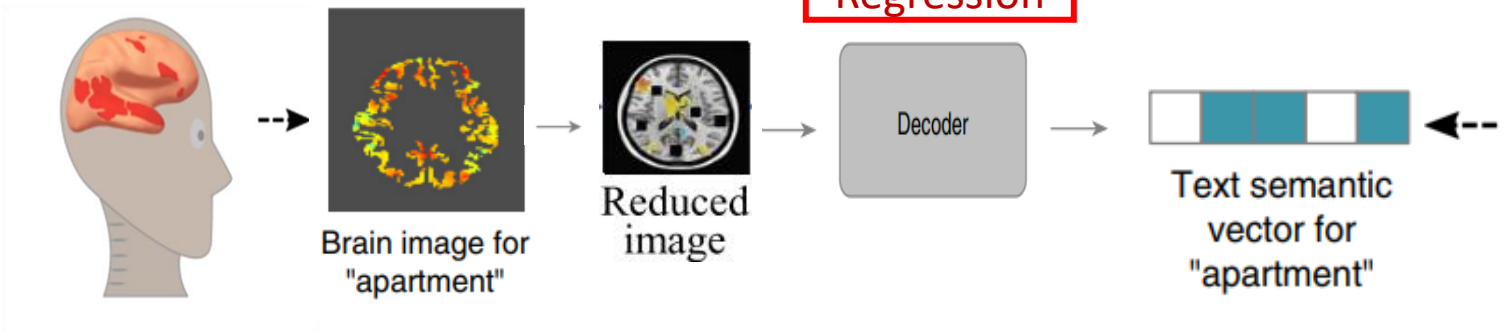
Select 5000 voxels based on top-5000 correlation scores

Brain Decoder Schematic? (concept+picture)

Stimulus:
Apartment

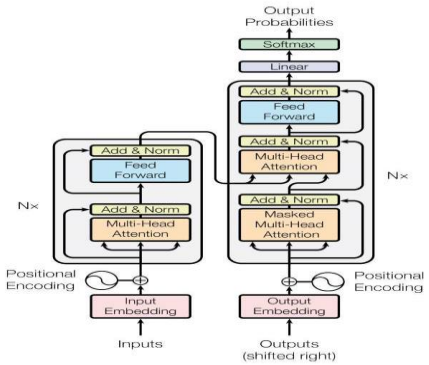
Present

Ridge Regression



Stimulus:
Apartment

Present



BERT

Periera et al. 2018, Devlin et al. 2018

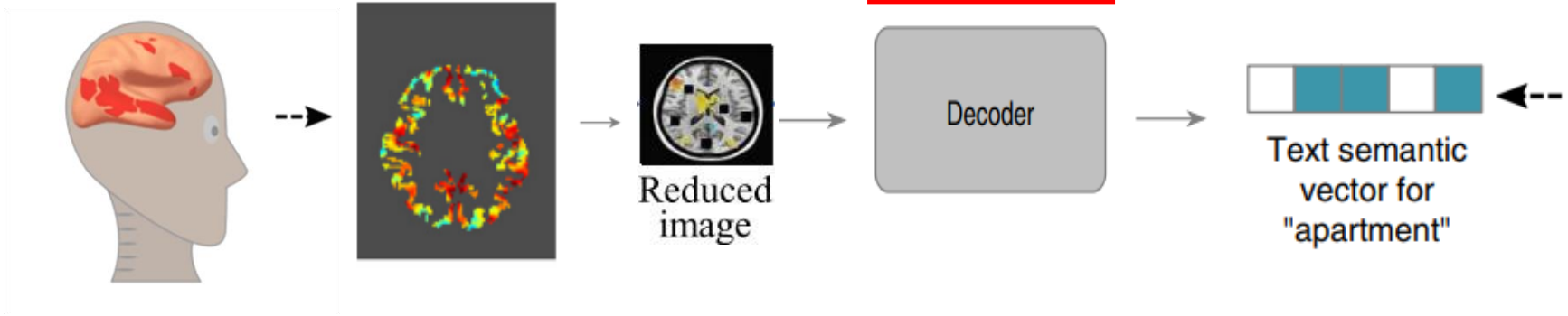
Coling-2022

Brain Decoder Schematic? (concept+sentence)

"An apartment is a self-contained home that is part of a building."

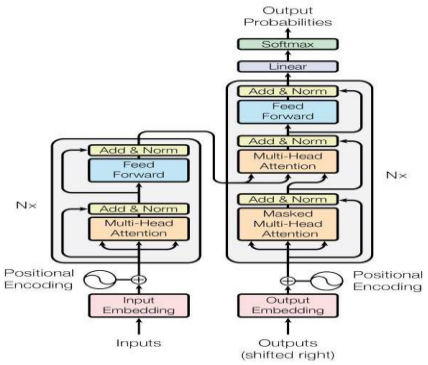
Stimulus:
Apartment

Present



Stimulus:
Apartment

Present

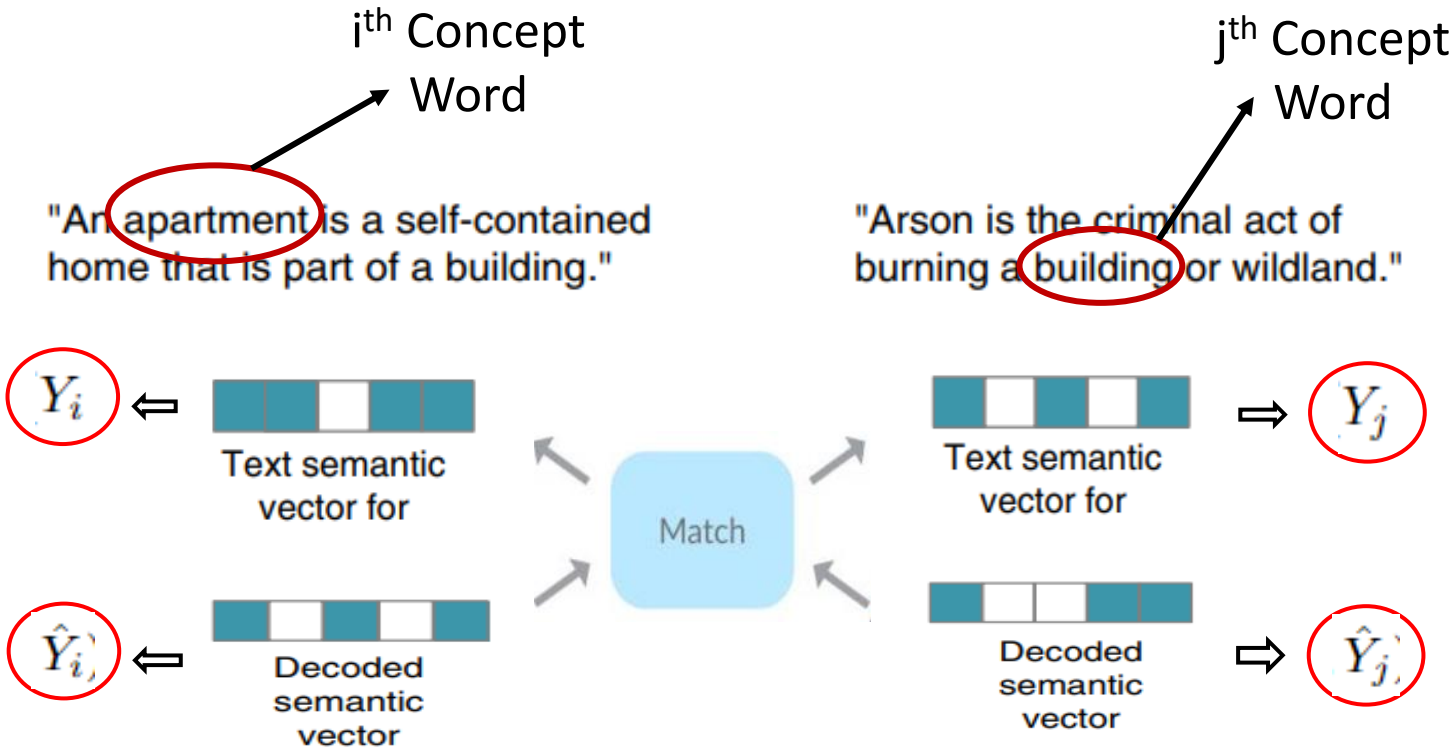


BERT

Periera et al. 2018, Devlin et al. 2018

Coling-2022

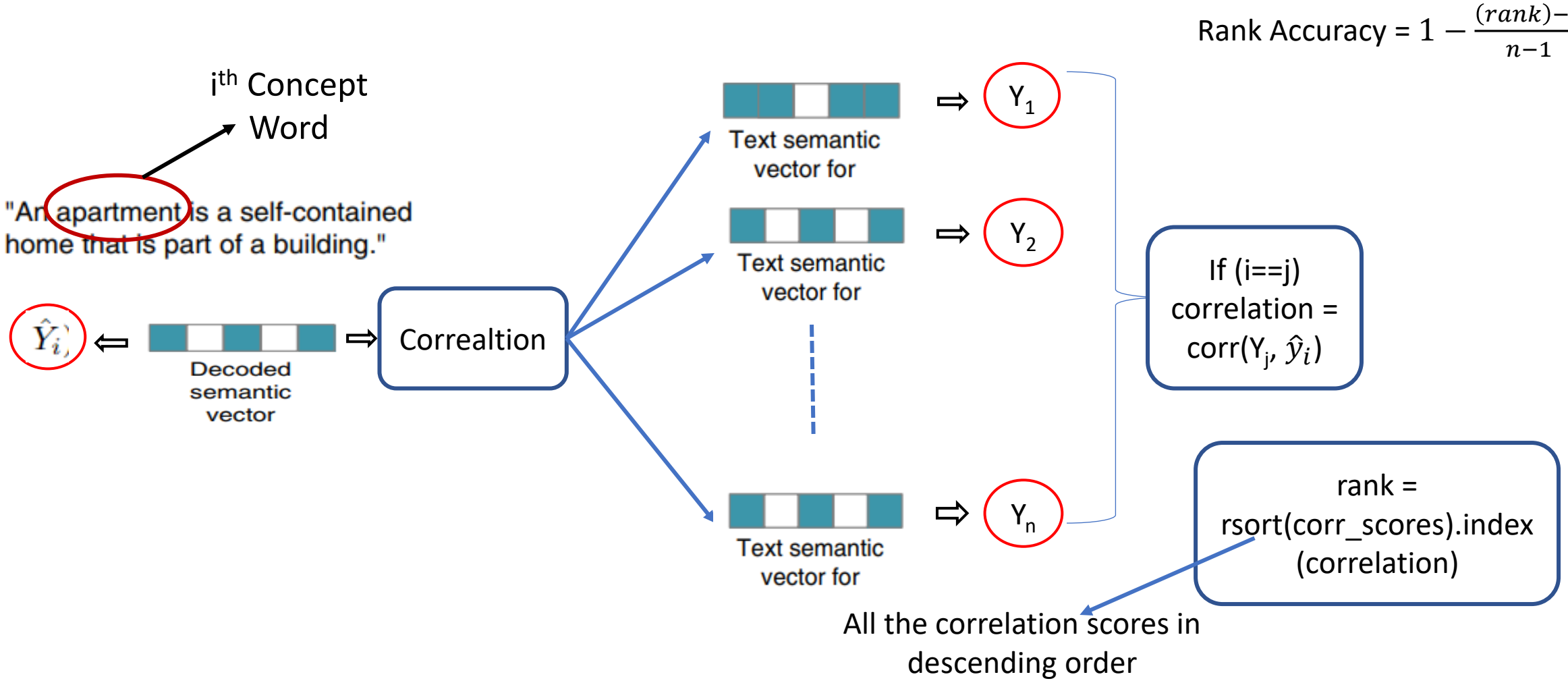
Evaluating Decoding Models: Pairwise Accuracy



Pairwise Accuracy = 1; corr(Y_i, $\hat{y}_i$) + corr(Y_j, $\hat{y}_j$) > corr(Y_i, $\hat{y}_j$) + corr(Y_j, $\hat{y}_i$)

Periera et al. 2018

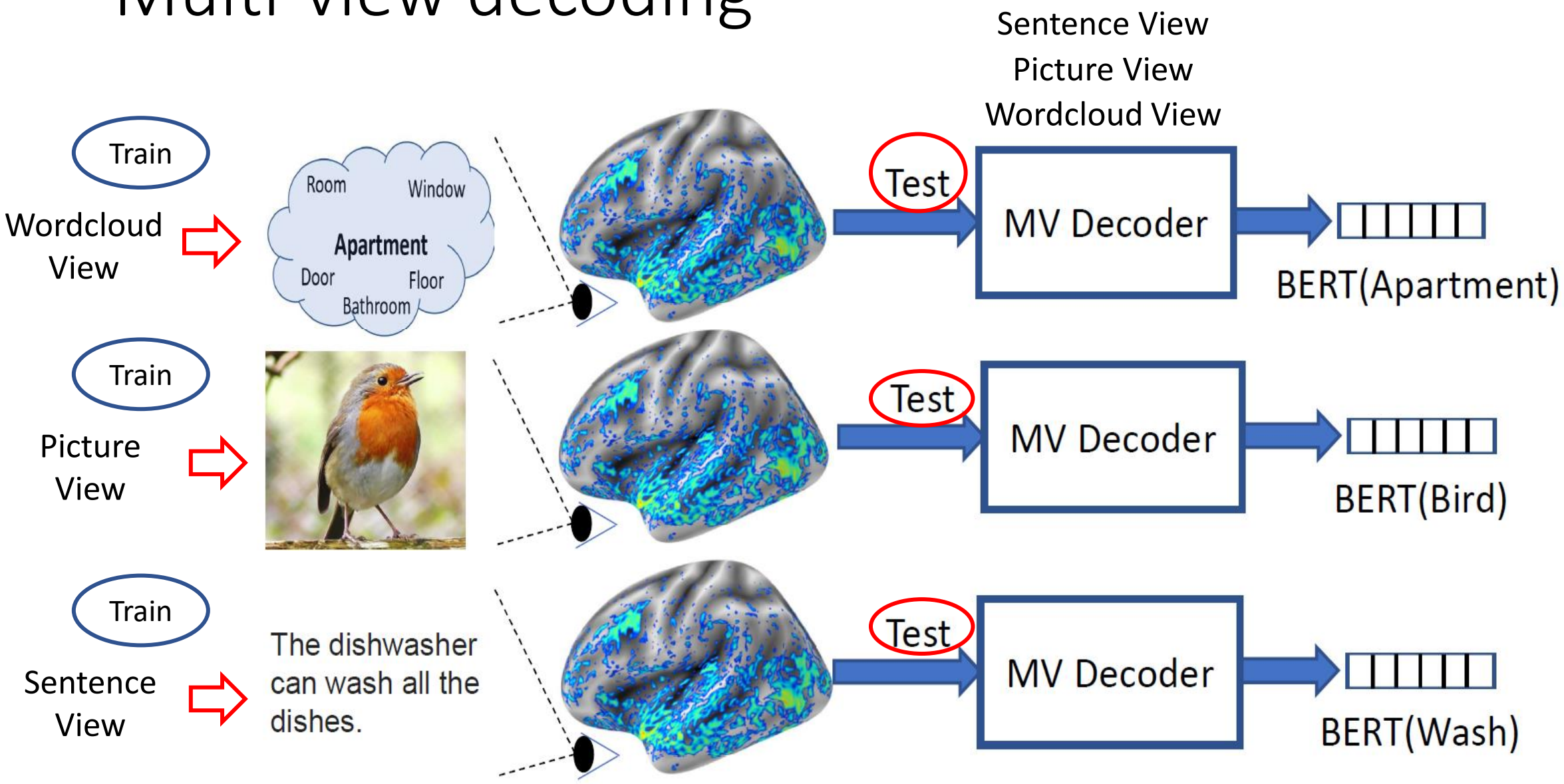
Evaluating Decoding Models: Rank Accuracy



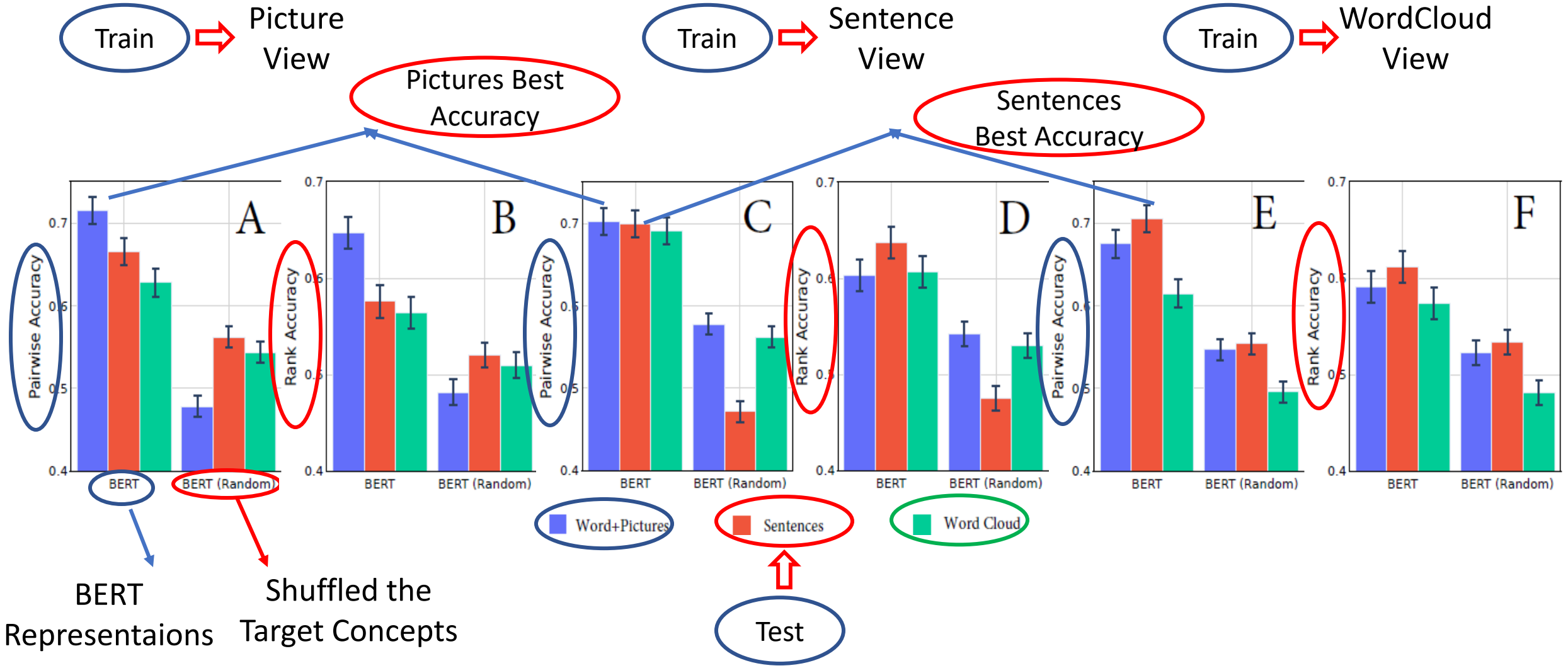
Periera et al. 2018

Coling-2022

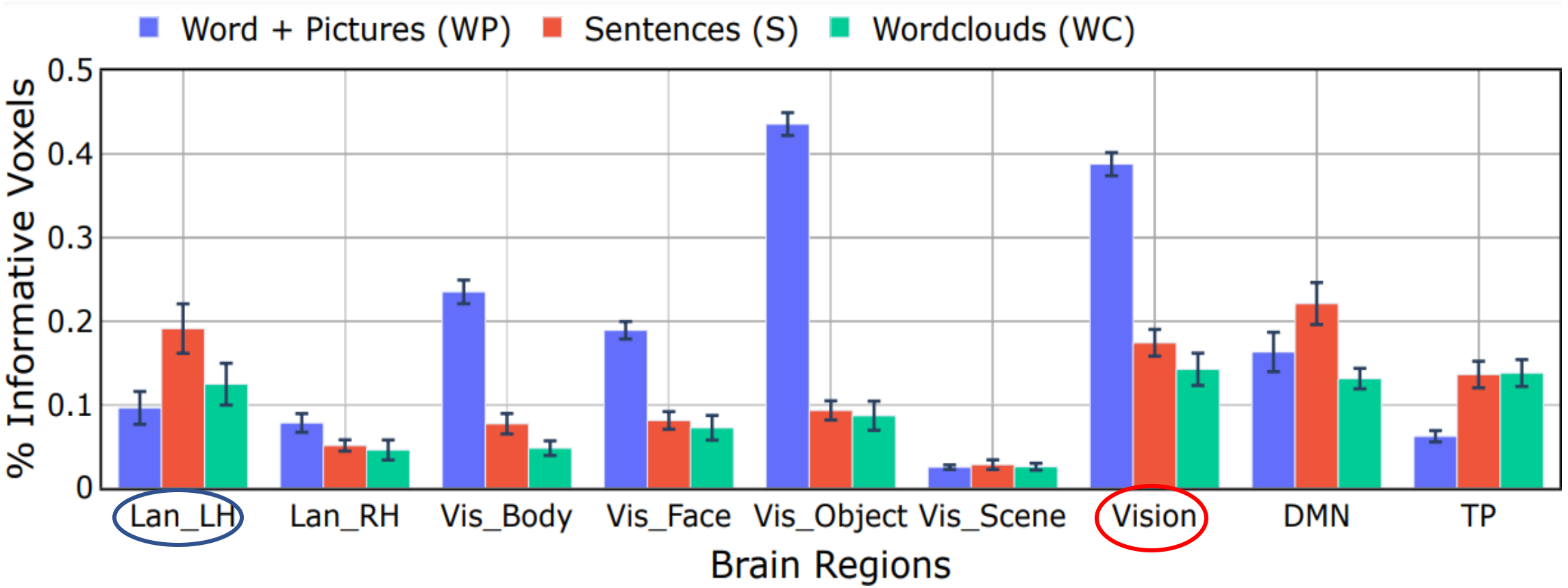
Multi-view decoding



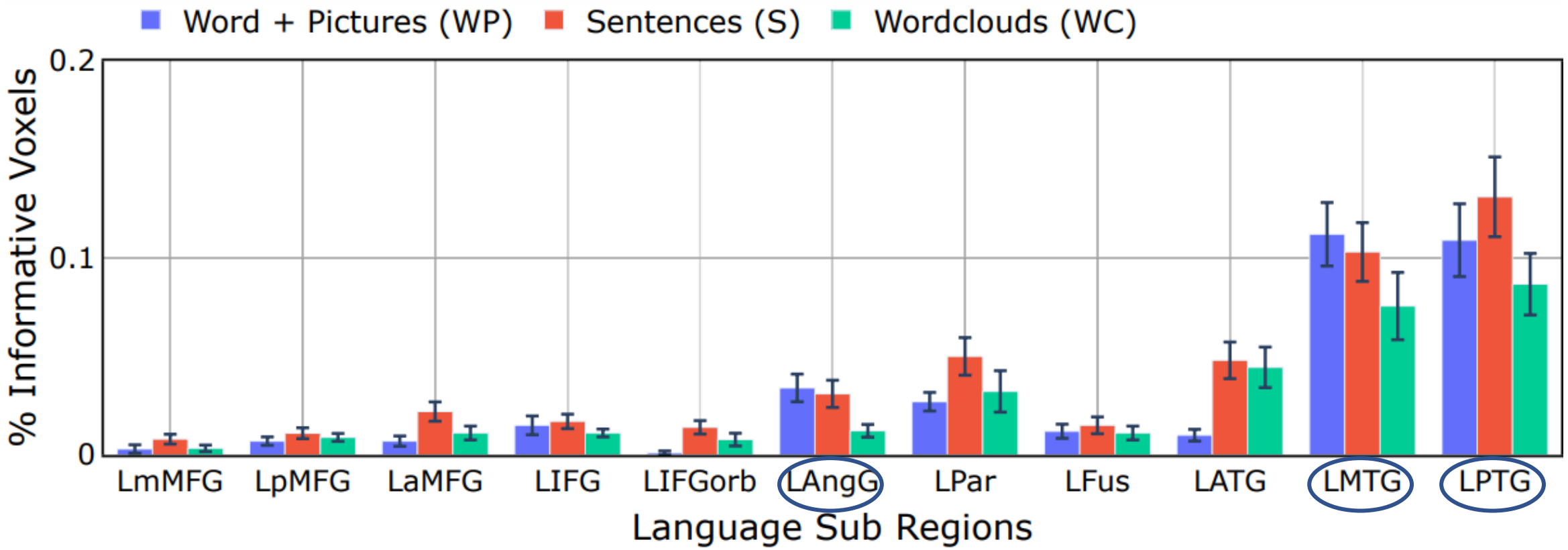
Multi-view decoding results



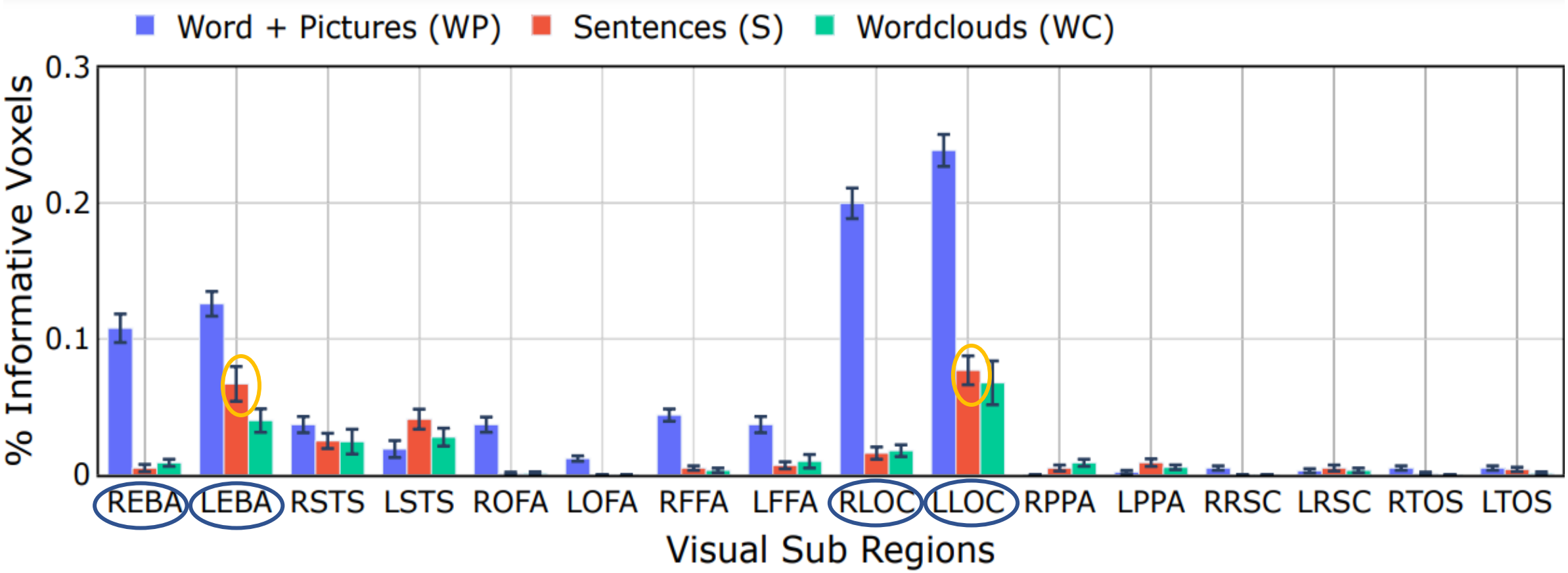
Distribution of Informative Voxels



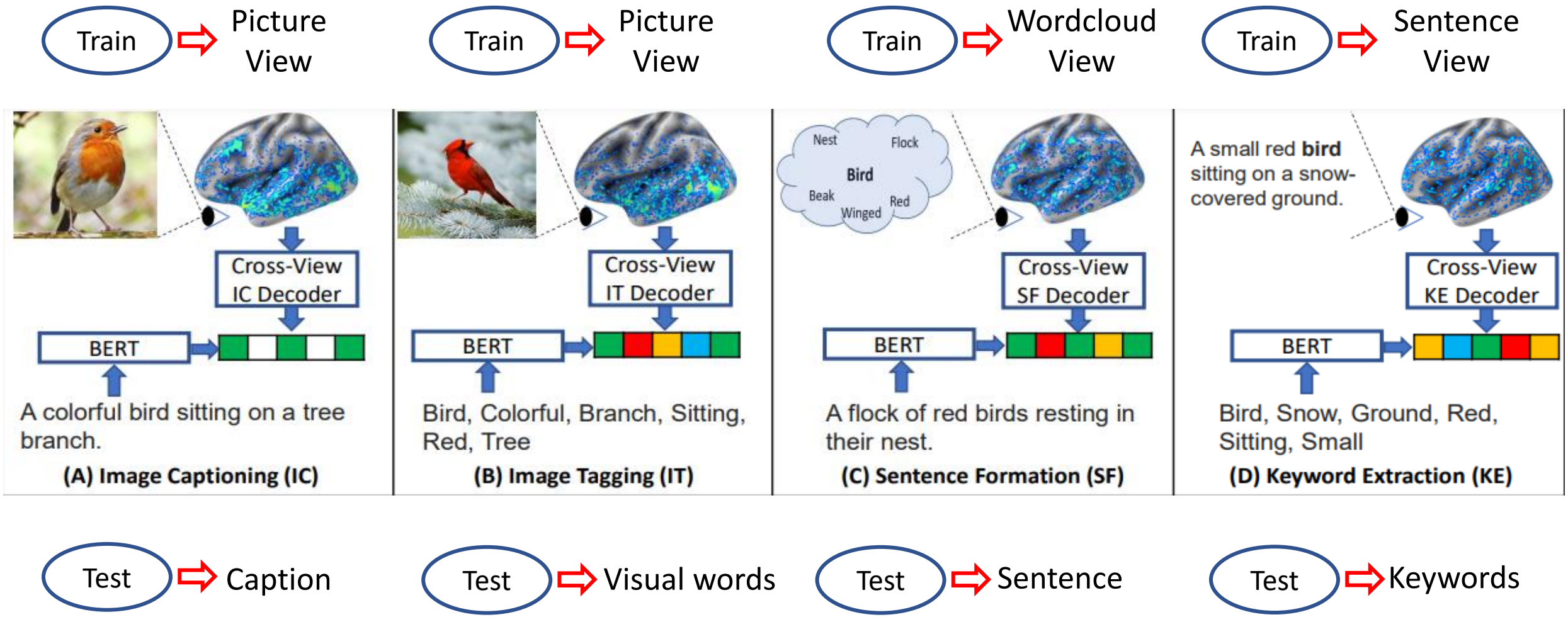
Distribution of Information Voxels: Language Network



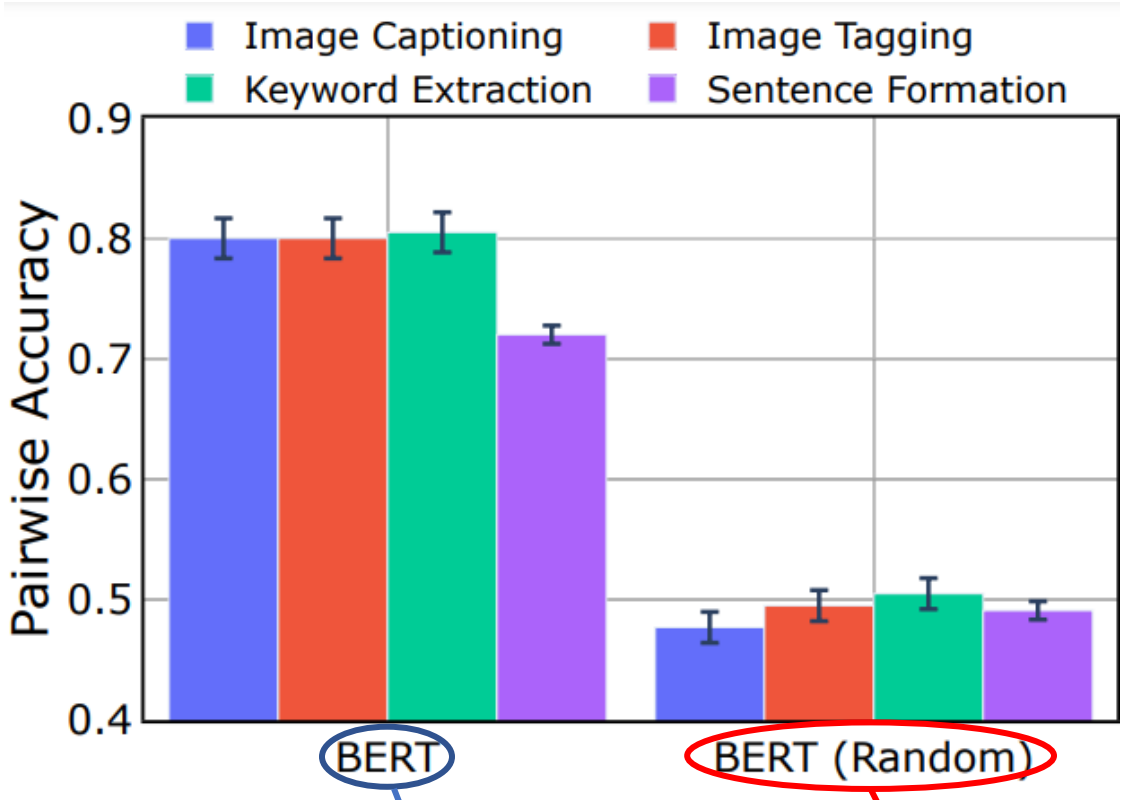
Distribution of Informative Voxels: Visual Network

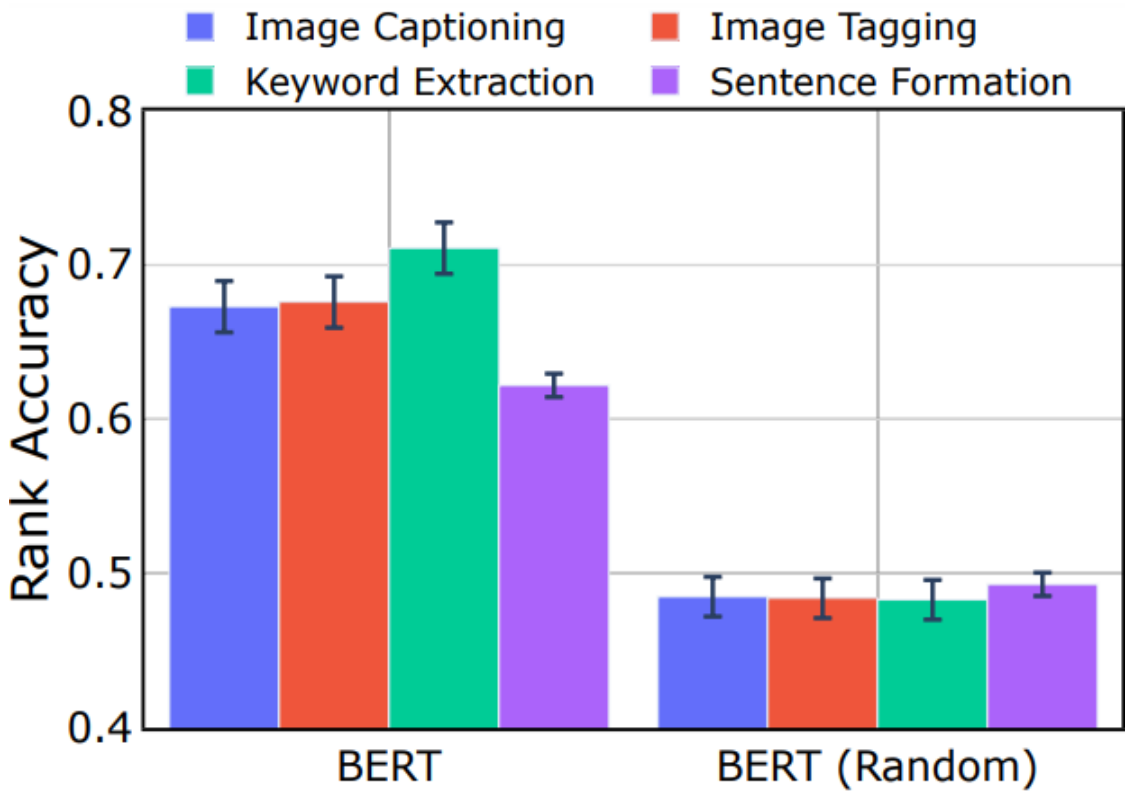


Cross-view Decoding



Cross-view Decoding results

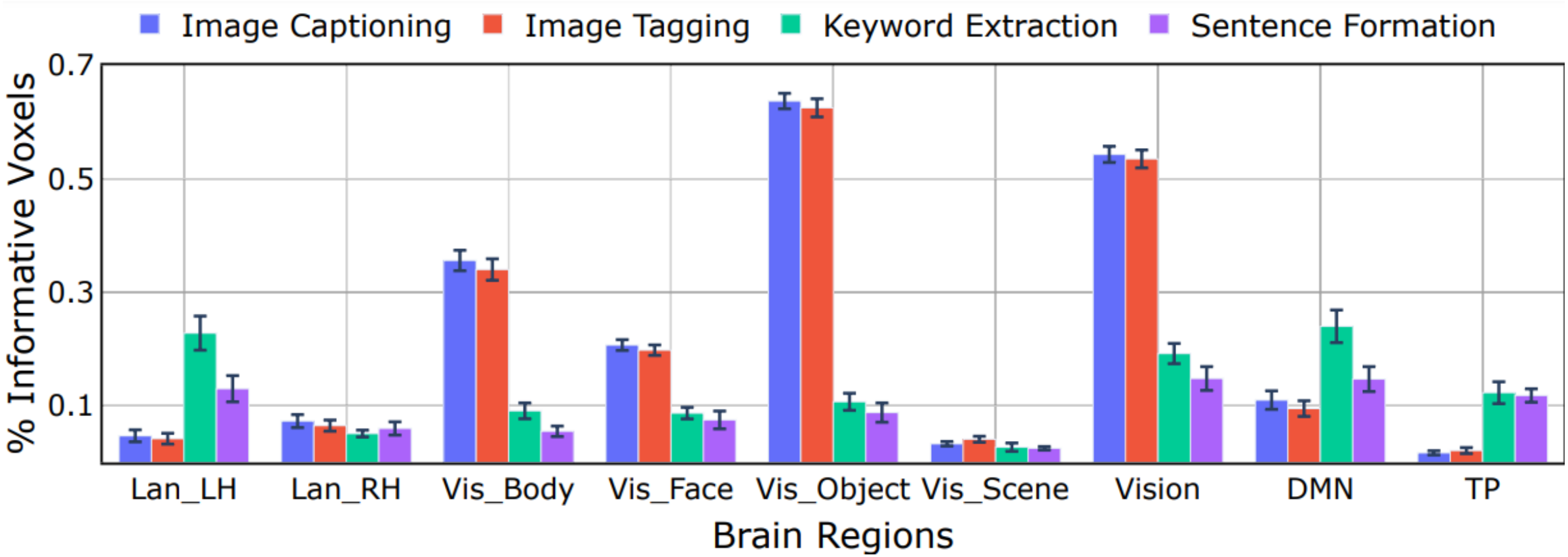




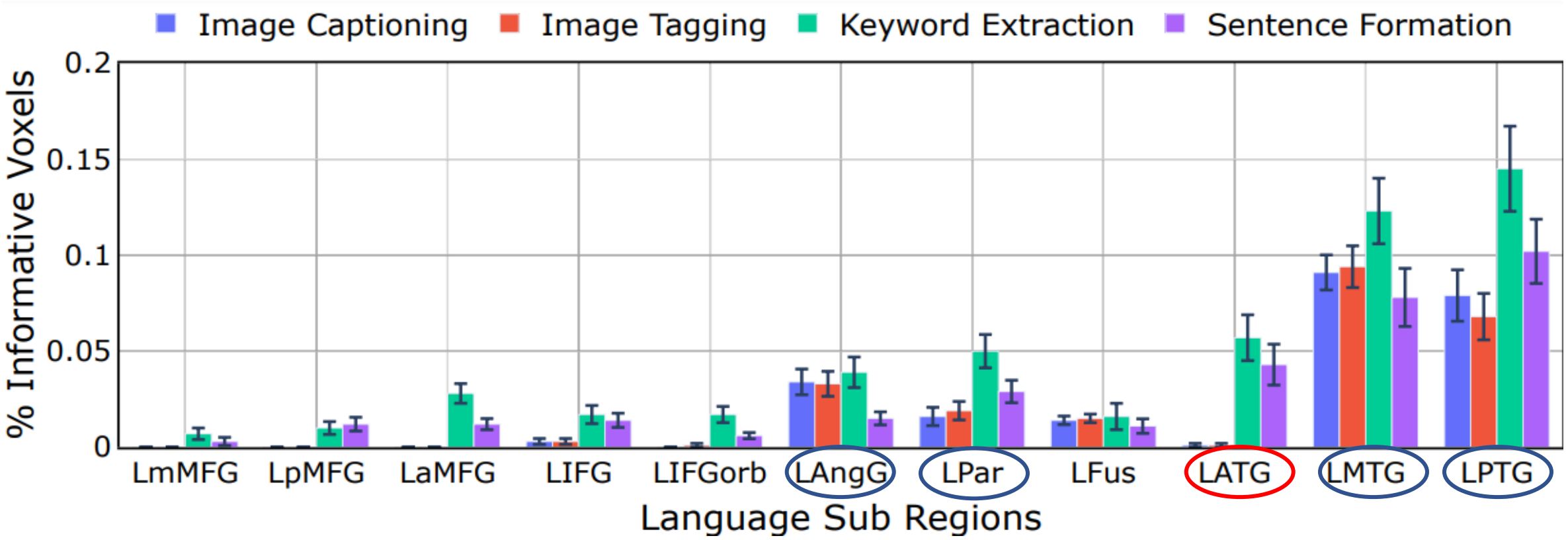
BERT Representaions

Shuffled the Target Concepts

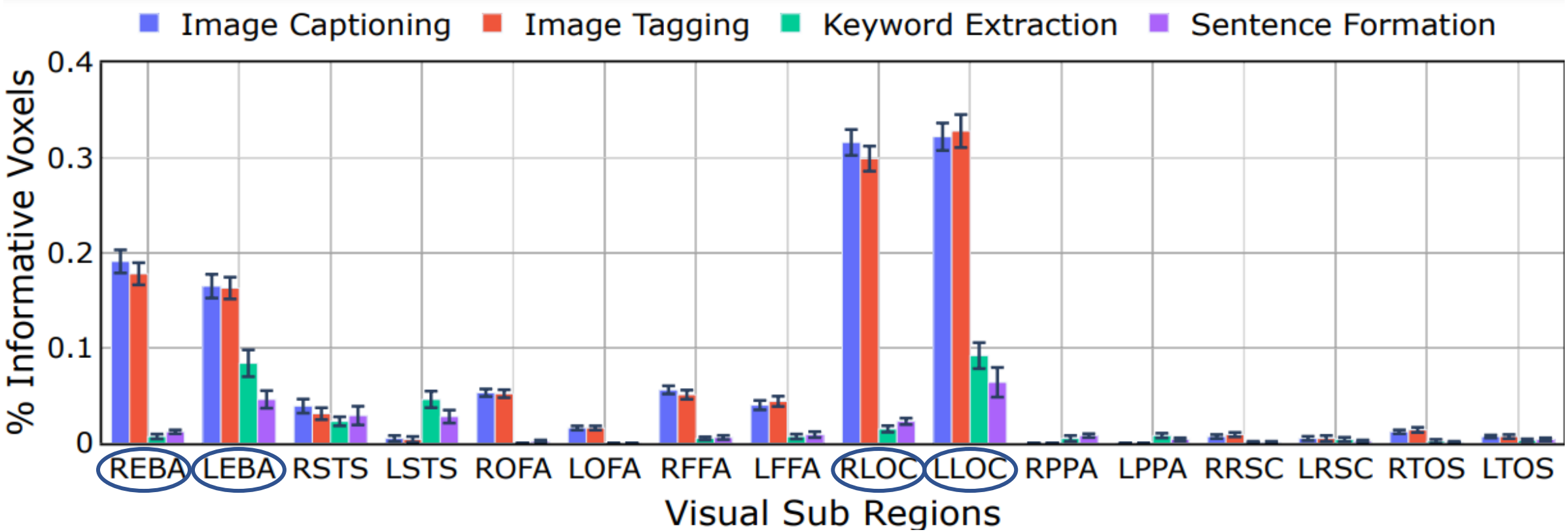
Distribution of Informative Voxels



Distribution of Informative Voxels: Language Network



Distribution of Information Voxels: Visual Network



Insights

- Without any apriori location constraints,
  - nearly half of the voxels selected by the models fall into the language atlas.
  - Conclusion: brain regions active in language processing also highly correspond to these representations.

Sun et al. 2019

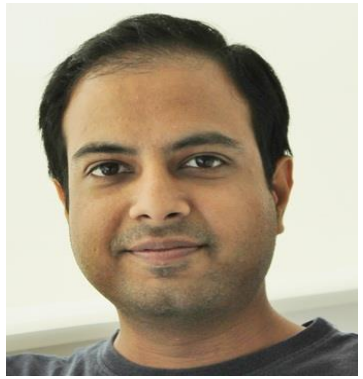
Collaborators

Subba Reddy Oota

Jashn Arora

Veeral Agarwal

Mounika Marreddy



Manish Gupta

Bapi Raju Surampudi