Neural Language Taskonomy: Which NLP Tasks are the most Predictive of fMRI Brain Activity

July 13, 2022

NAACL-HLT 2022

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

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

What is 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?"

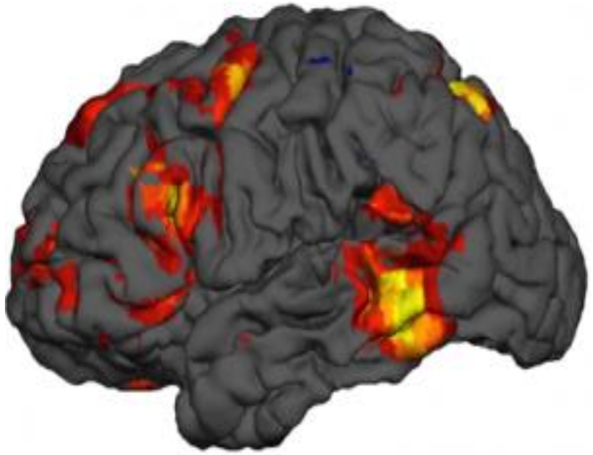
"It's quite too funny. I am sure you could never guess how I employed my morning."

"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, …

Text Corpus

A language task in the scanner



fMRI Brain
Activity

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

Brain 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.

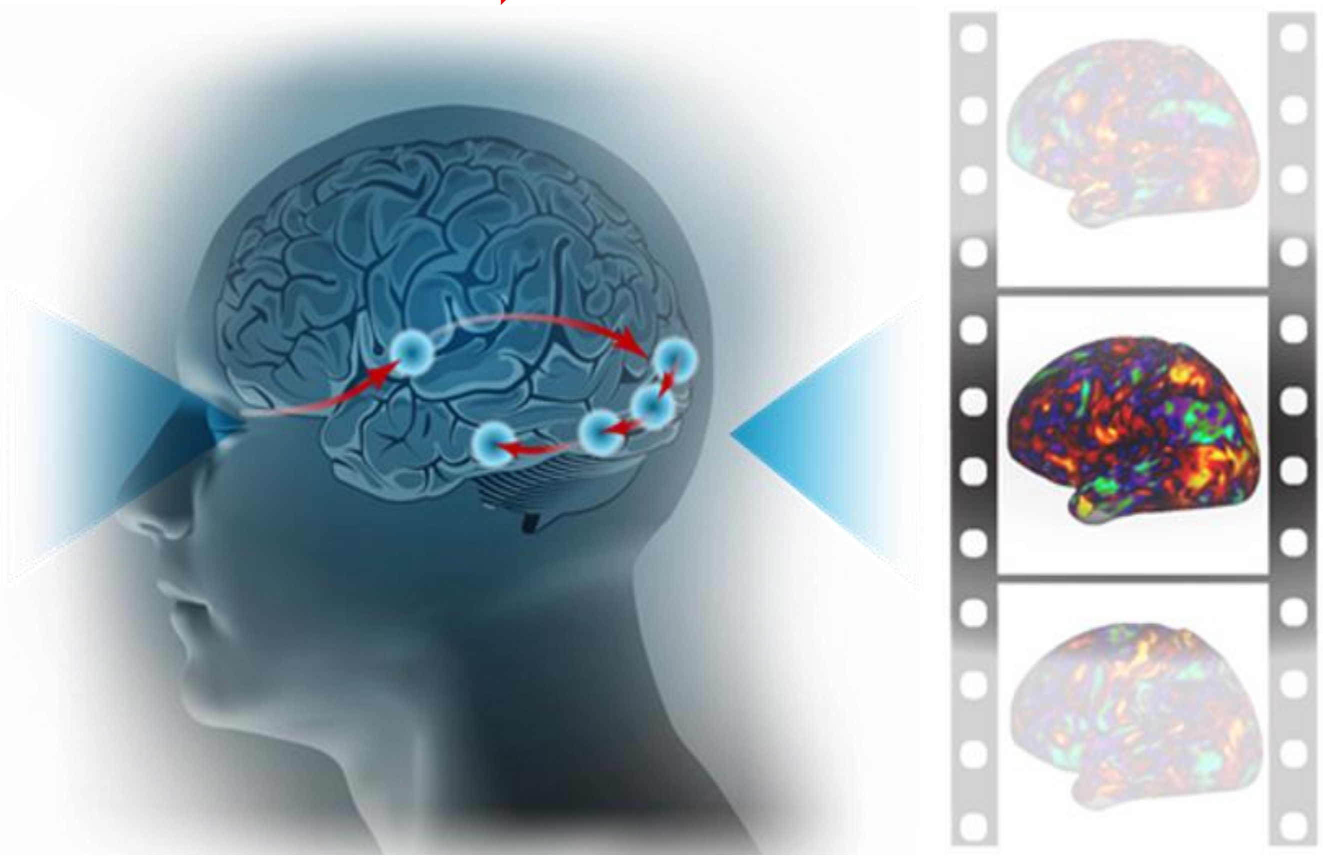
"Well, really!" he cried, and then he choked; and laughed again until he was obliged to lie back, limp and helpless, in the chair.

"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."

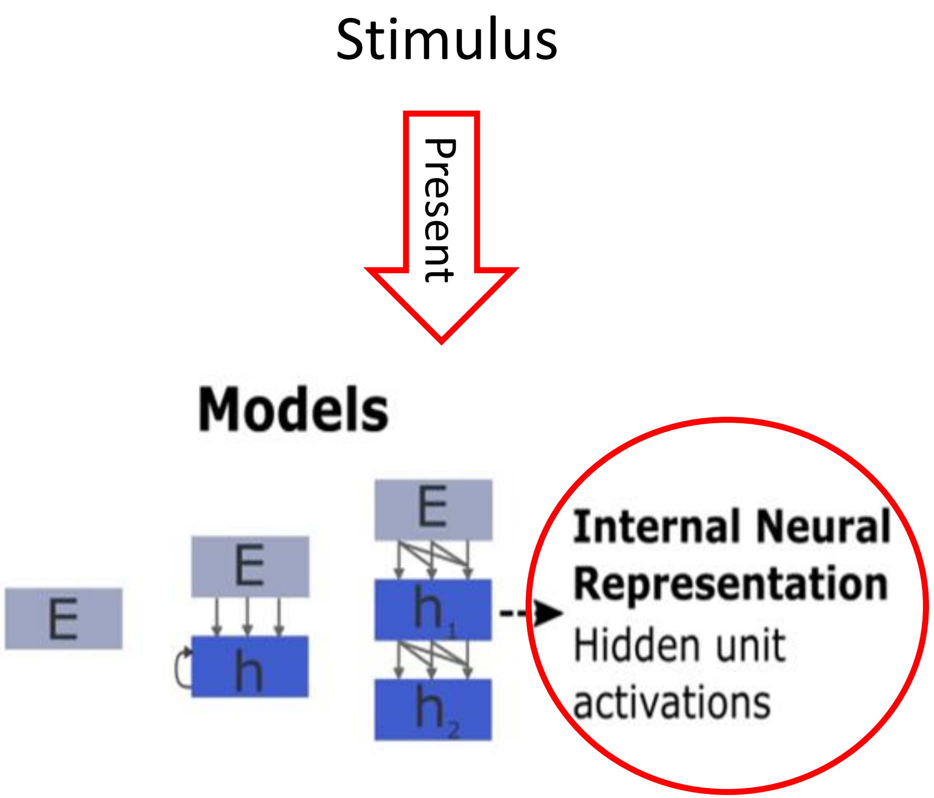
"Quite so; but the sequel was rather unusual. I will tell you, … I soon found Briony Lodge. It is a bijou villa, with a garden at the back, but built out in front right up to the road, …



Stimulus Representation ← Decoding ← fMRI

Haiguang Wen et al, 2017

What is Brain Encoding?



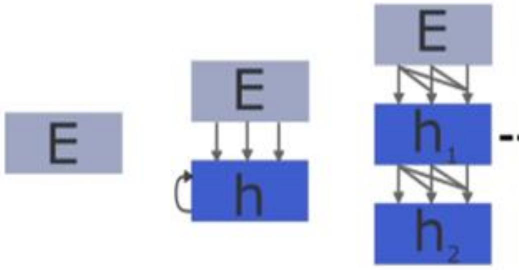
Schrimpf et al. 2021 fMRI

What is Brain Encoding?

Stimulus

Present

Models



Internal Neural Representation
Hidden unit activations

Stimulus

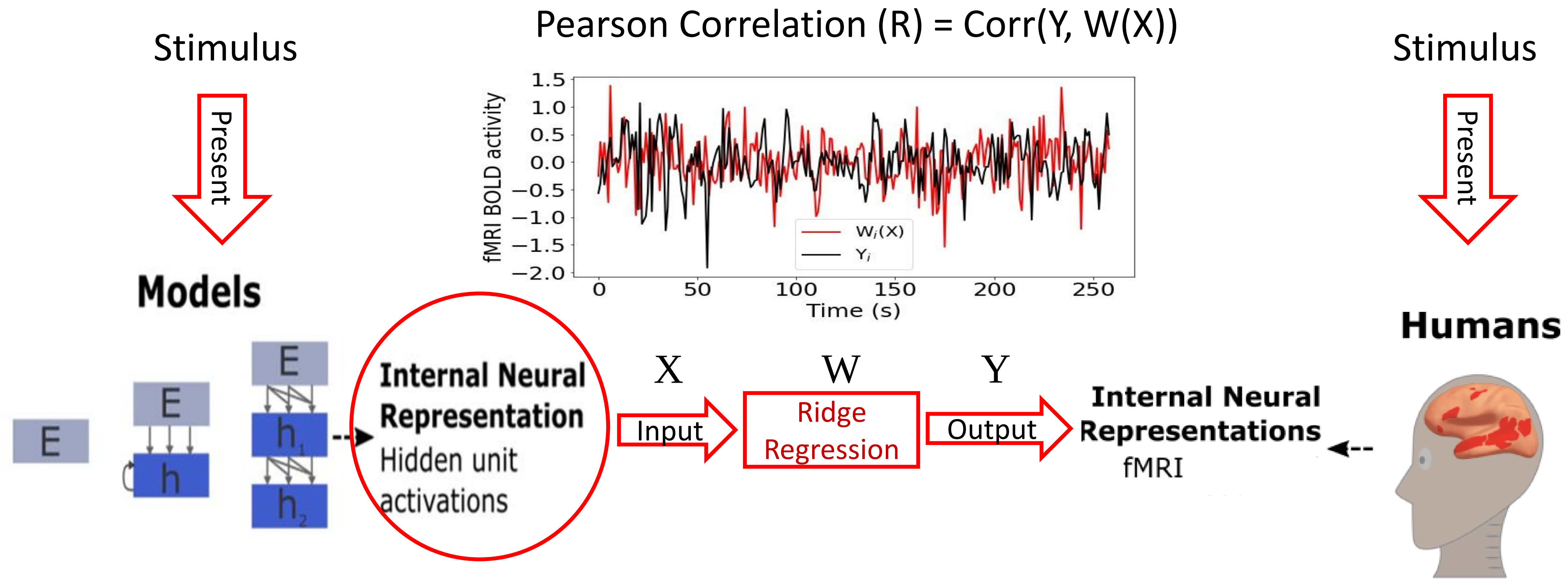
Present

Humans

Internal Neural Representations
fMRI

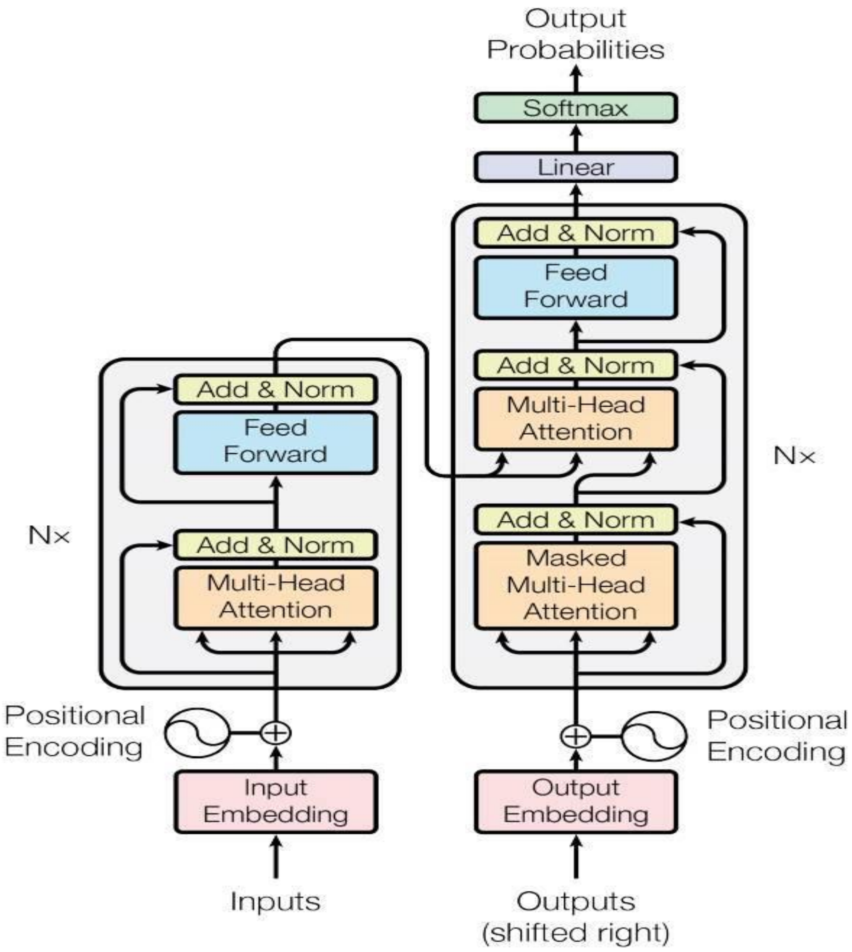
Schrimpf et al. 2021 fMRI

What is Brain Encoding?



Schrimpf et al. 2021 fMRI

Most popular language models are Transformers



Transformer language models
(BERT, XLM, GPT,…)

Vaswani et al. 2017

Pretrained language models accurately predict brain activity

The neural architecture of language: Integrative modeling converges on predictive processing

Martin Schrimpf[a,b,c,1], Idan Asher Blank[a,d,2], Greta Tuckute[a,b,2], Carina Kauf[a,b,2], Eghbal A. Hosseini[a,b], Nancy Kanwisher[a,b,c,1], Joshua B. Tenenbaum[a,c,3], and Evelina Fedorenko[a,b,1,3]

[a]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139; [c]Center for Brains, Minds and Machines, Massachusetts Institute of Technology, Cambridge,

Contributed

The neuro with an i brain func many com this appro anisms in taking this processing most pow

Linking artificial and human neural representations of language

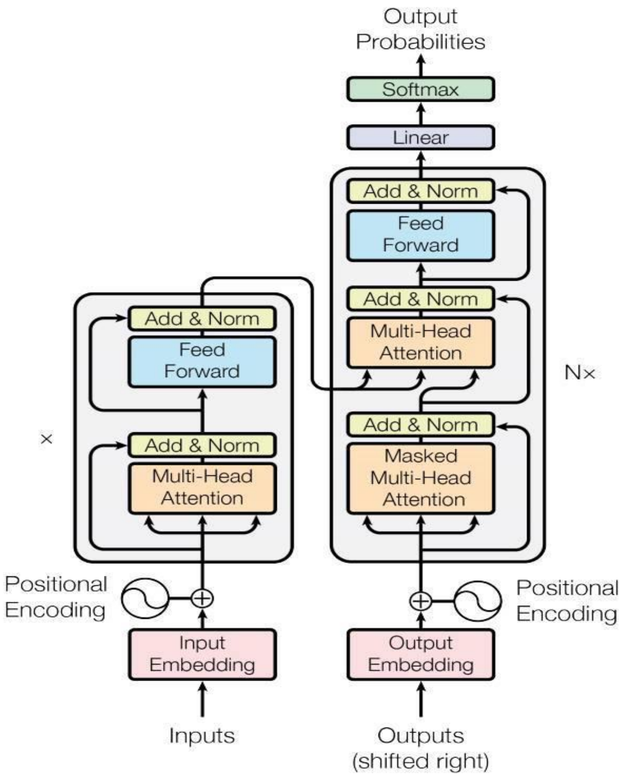
Jon Gauthier and **Roger P. Levy**
Massachusetts Institute of Technology
Department of Brain and Cognitive Sciences
jon@gauthiers.net, rplevy@mit.edu

Abstract

What information from an act of sentence understanding is robustly represented in the human brain? We investigate this question by comparing sentence encoding models on a brain decoding task, where the sentence that an

theories of language understanding, many are specified at too high a level of analysis to plausibly map onto neural structures without serious further revision (Poeppel, 2012).

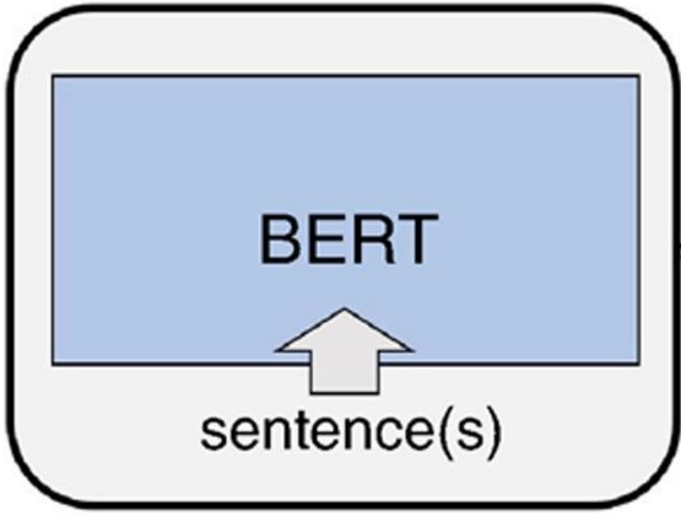
Studies which draw on these high-level representations must therefore also assume some link



Transformer language models (BERT, XLM, GPT,…)

Vaswani et al. 2017, Gauthier et al. 2019, Schrimpf et al. 2021

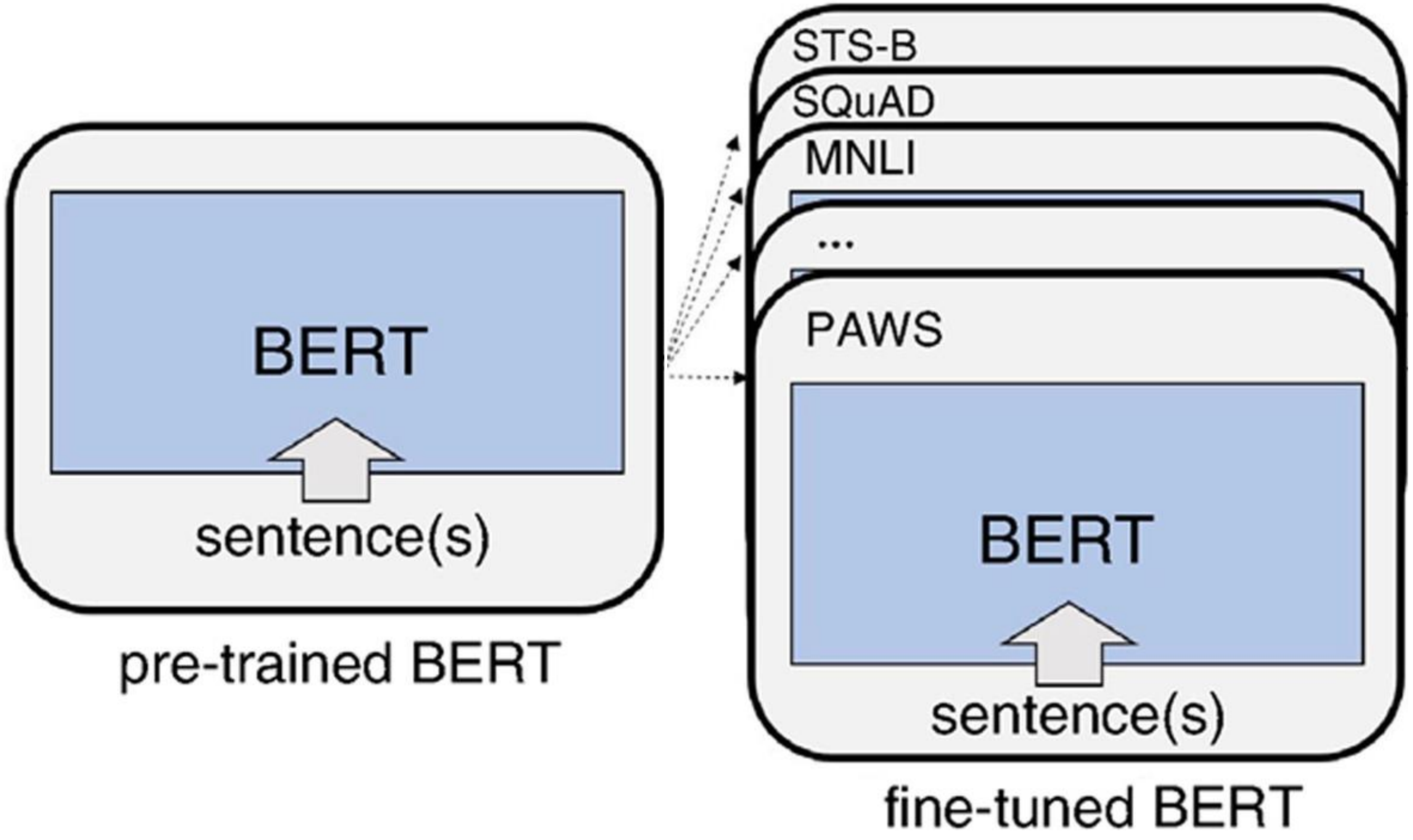
Can task-specific language models better predict fMRI brain activity?



pre-trained BERT

Devlin et al. 2019

Can task-specific language models better predict fMRI brain activity?



Tasks

- Paraphrase
- Summrisation
- Question Answering
- Sentiment Analysis
- NER
- Word Sense Disambiguation
- Natural Language Inference
- Semantic Role Labeling
- Coreference Resolution
- Shallow Syntax
- Pretrained BERT

Syntactic

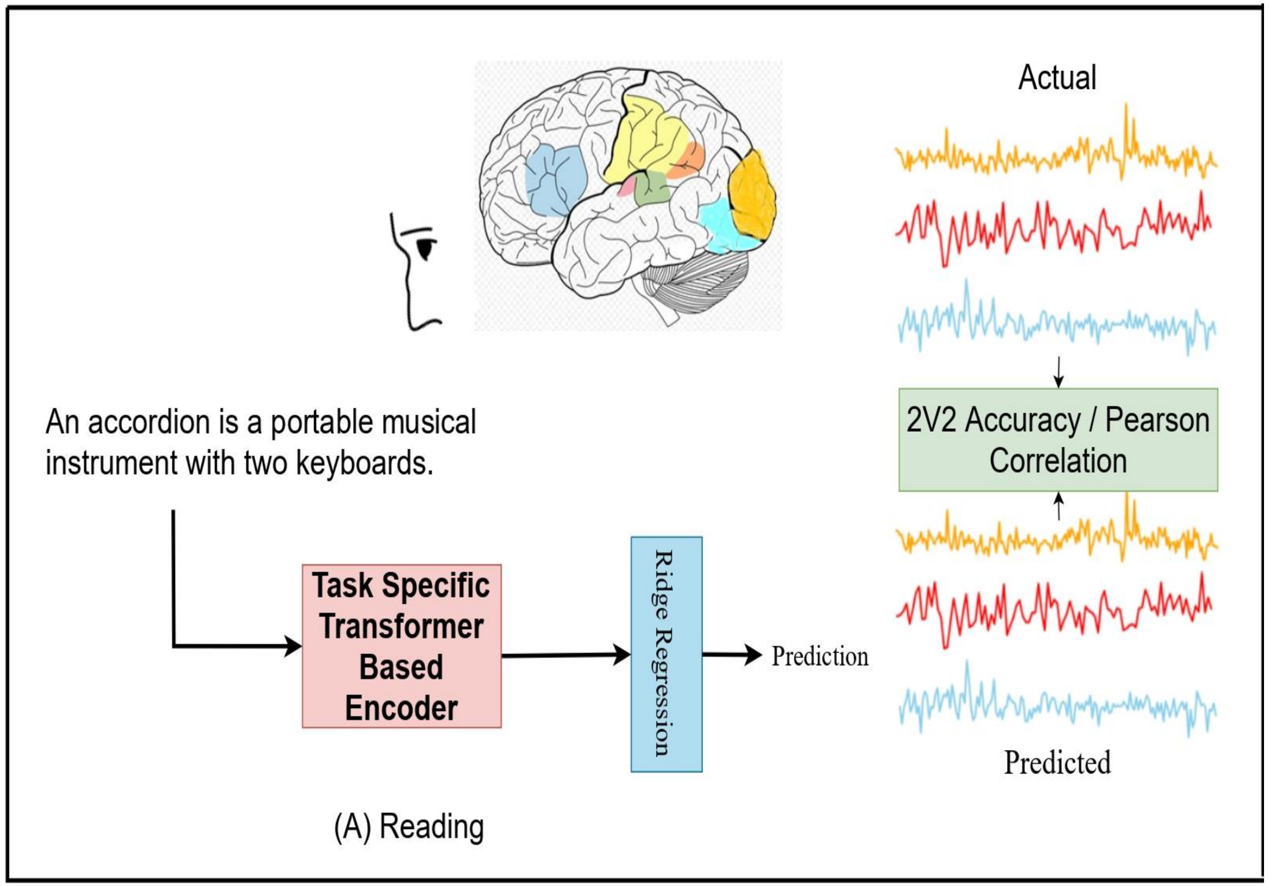
Devlin et al. 2019, Bowon et al. 2020

Task-specific Models (10) + Pretrained BERT

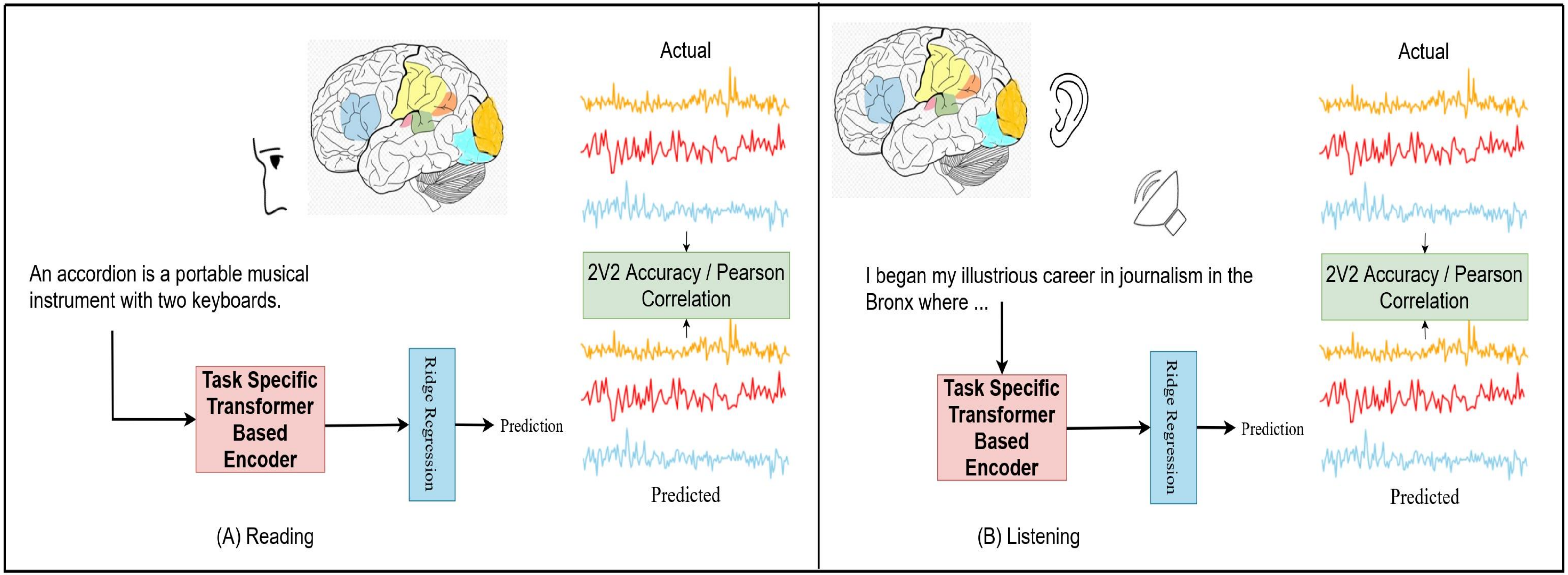
Common underlying model
Bert-base (768 dimension)

Task	HuggingFace Model Name	Dataset	URL
NLI	bert-base-nli-mean-tokens	Stanford Natural Language Inference (SNLI), MultiNLI	https://huggingface.co/sentence-transformers/bert-base-nli-mean-tokens
PD	bert-base-cased-finetuned-mrpc	Microsoft Research Paraphrase Corpus (MRPC)	https://huggingface.co/bert-base-cased-finetuned-mrpc
SS	bert-base-chunl	CoNLL-2003	https://huggingface.co/vblagoje/bert-english-uncased-finetuned-chunk
Sum	bart-base-samsum	SAMSum	https://huggingface.co/lidiya/bart-base-samsum
WSD	bert-base-baseline	English all-words	https://github.com/BPYap/BERT-WSD
CR	bert_coreference_base	OntoNotes and GAP	https://github.com/mandarjoshi90/coref
NER	bert-base-NER	CoNLL-2003	https://huggingface.co/dslim/bert-base-NER
QA	bert-base-qa	SQUAD	https://huggingface.co/docs/transformers/model_doc/bert#bertforquestionanswering
SA	bert-base-sst	Stanford Sentiment Treebank (SST)	https://huggingface.co/barissayil/bert-sentiment-analysis-sst
SRL	bert-base-srl	English PropBank SRL	https://s3-us-west-2.amazonaws.com/allennlp/models/bert-base-srl-2020.02.10.tar.gz

Can task-specific language models have similar predictive performance in reading and listening?



Can task-specific language models have similar predictive performance in reading and listening?



Reading data target: human brain recordings

- We use Periera dataset
  - reading sentences
  - 5 subjects
  - 627 sentences (experiment 2 + 3)

Example: "A clarinet is a woodwind musical instrument"

Experiment 2:

Musical instruments (clarinet)

A clarinet is a woodwind musical instrument.
It is a long black tube with a flare at the bottom.
The player chooses notes by pressing keys and holes.
The clarinet is used both in jazz and classical music.

Musical instruments (accordion)

An accordion is a portable musical instrument with two keyboards. One keyboard is used for individual notes, the other for chords. Accordions produce sound with bellow that blow air through reeds. An accordionist plays both keyboards while opening and closing the bellows.

Musical instruments (piano)

The piano is a popular musical instrument played by means of a keyboard. Pressing a piano key causes a felt-tipped hammer to hit a vibrating steel string. The piano has an enormous note range, and pedals to change the sound quality. The piano repertoire is large, and famous pianists can give solo concerts.

Experiment 3:

Skiing (passage 1)

I hesitantly skied down the steep trail that my buddies convinced me to try. I made a bad turn, and I found myself tumbling down. I finally came to a stop at a flat part of the slope. My skis were nowhere to be found, and my poles were lodged in a snow drift up the hill.

Skiing (passage 2)

A major strength of professional skiers is how they use ski poles. Proper use of ski poles improves their balance and adds flair to their skiing. It minimizes the need for upper body movements to regain lost balance while skiing.

Skiing (passage 3)

New ski designs and stiffer boots let skiers turn more quickly. But faster and tighter turns increase the twisting force on the legs. This has led to more injuries, particularly to ligaments in the skier's knee.

Gambling (passage 1)

When I decided to start playing cards, things went from bad to worse. Gambling was something I had to do, and I had already spent close to $10,000 doing it. My friends were sick of watching me gamble my savings away. The hardest part was the horror of leavin a casino after losing money I did not have.

Gambling (passage 2)

Good data on the social and economic effects of legalized gambling are hard to come by. Some studies indicate that having a casino nearby makes gambling problems more likely. Gambling may also be associated with personal bankruptcies and marriage problems.

Gambling (passage 3)

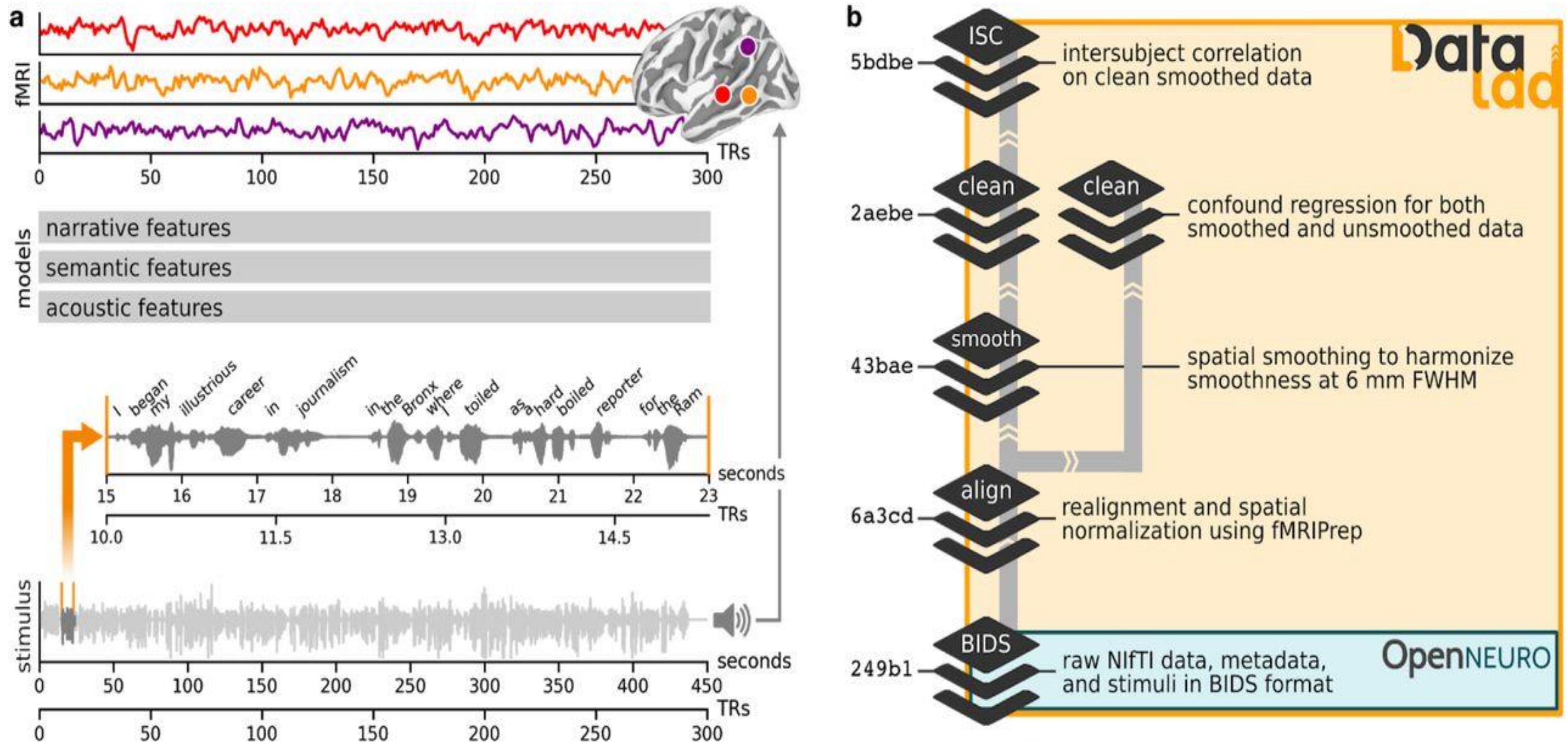
Over the past generation, there has been a dramatic expansion of legalized gambling. Most states have instituted lotteries, and many have casinos as well. Gambling has become a very big but controversial business.

Periera et al. 2018 fMRI

Listening data target: human brain recordings

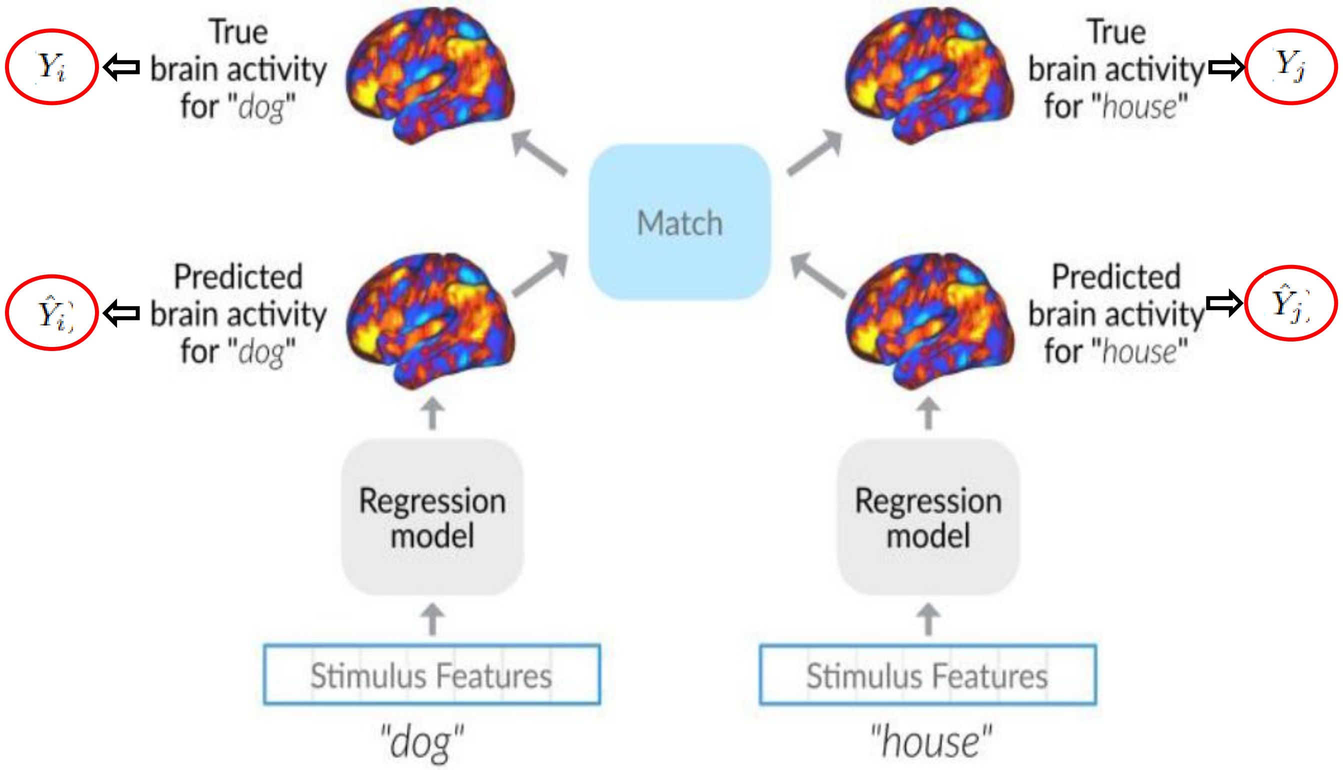
- We use Pieman story listening:
  - 82 subjects,
  - 282 TRs (repetition time)
  - here it is 1.5 sec.

Example: "I began my illustrious carrier in journalism…"



Nastase et al. 2021 fMRI

Evaluation Metrics: 2V2 and Pearson



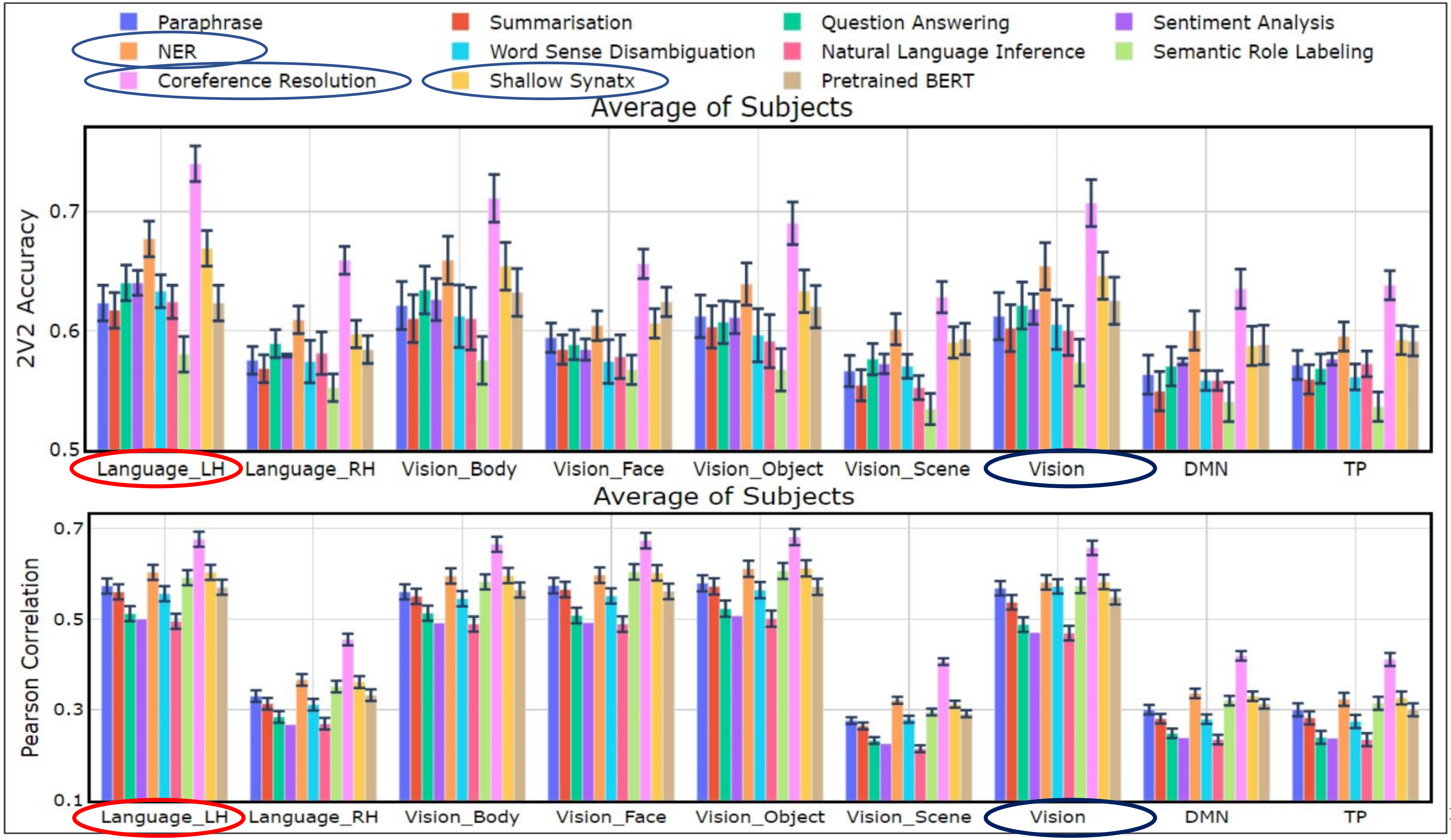
2V2 Accuracy

$$\text{2V2 Accuracy} = \frac{1}{N_{C_2}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} I[\{cosD(Y_i, \hat{Y}_i) + cosD(Y_j, \hat{Y}_j)\} < \{cosD(Y_i, \hat{Y}_j) + cosD(Y_j, \hat{Y}_i)\}]$$

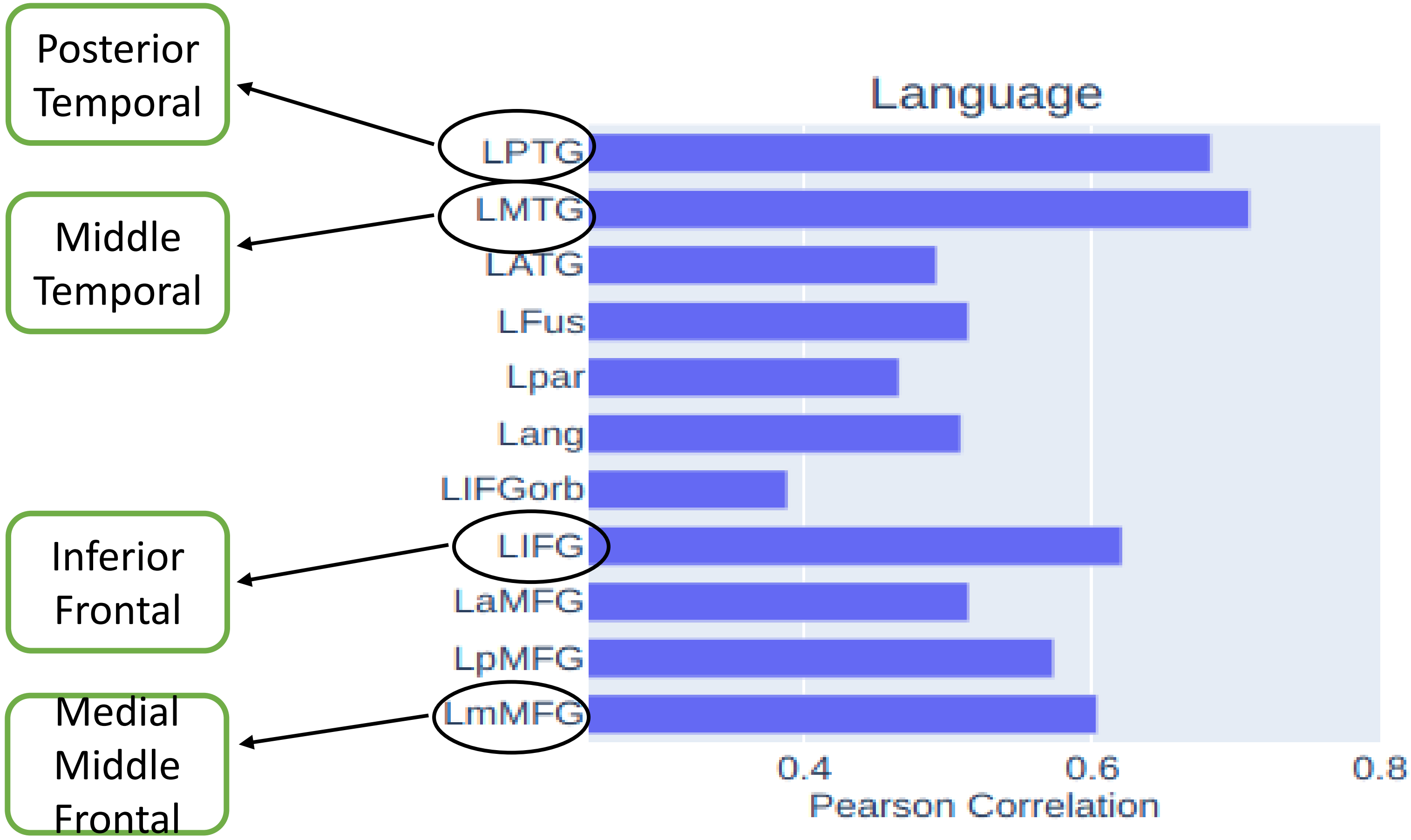
Cosine distance

Toneva et al. NeurIPS-2020

Encoding Performance (Reading)

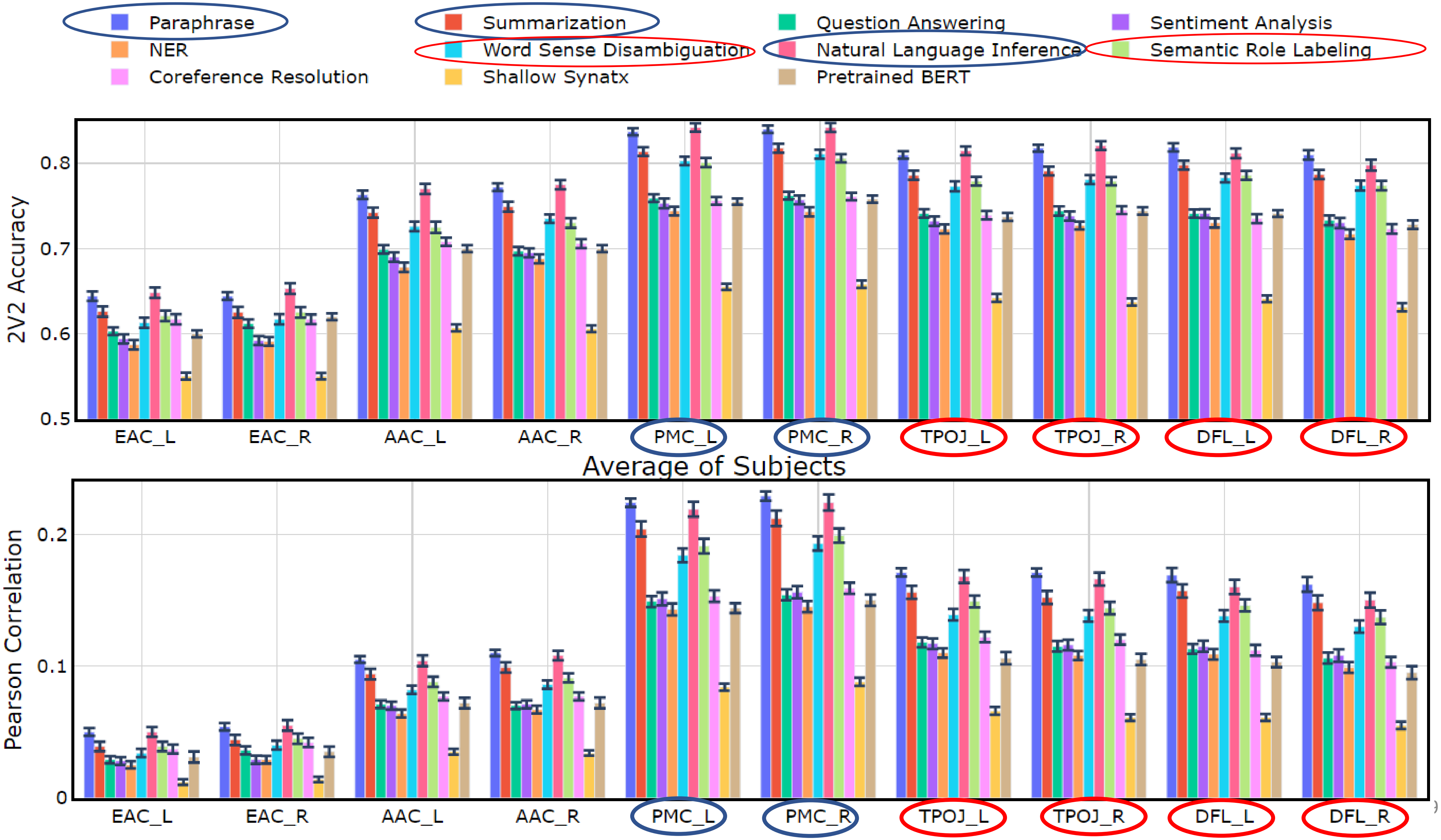


Which language sub regions have higher predicitivity?

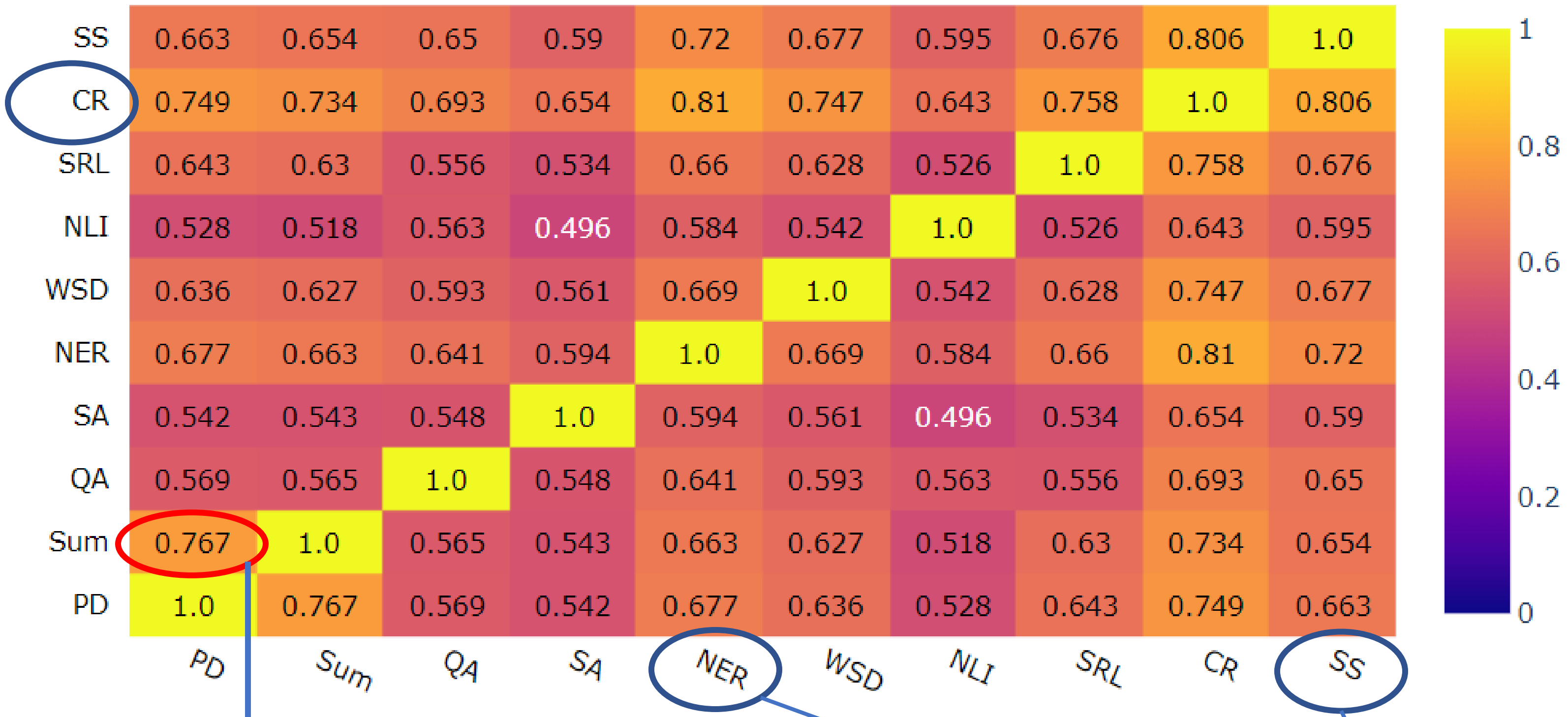


Corefernce Resolution Task

Encoding Performance (Listening)



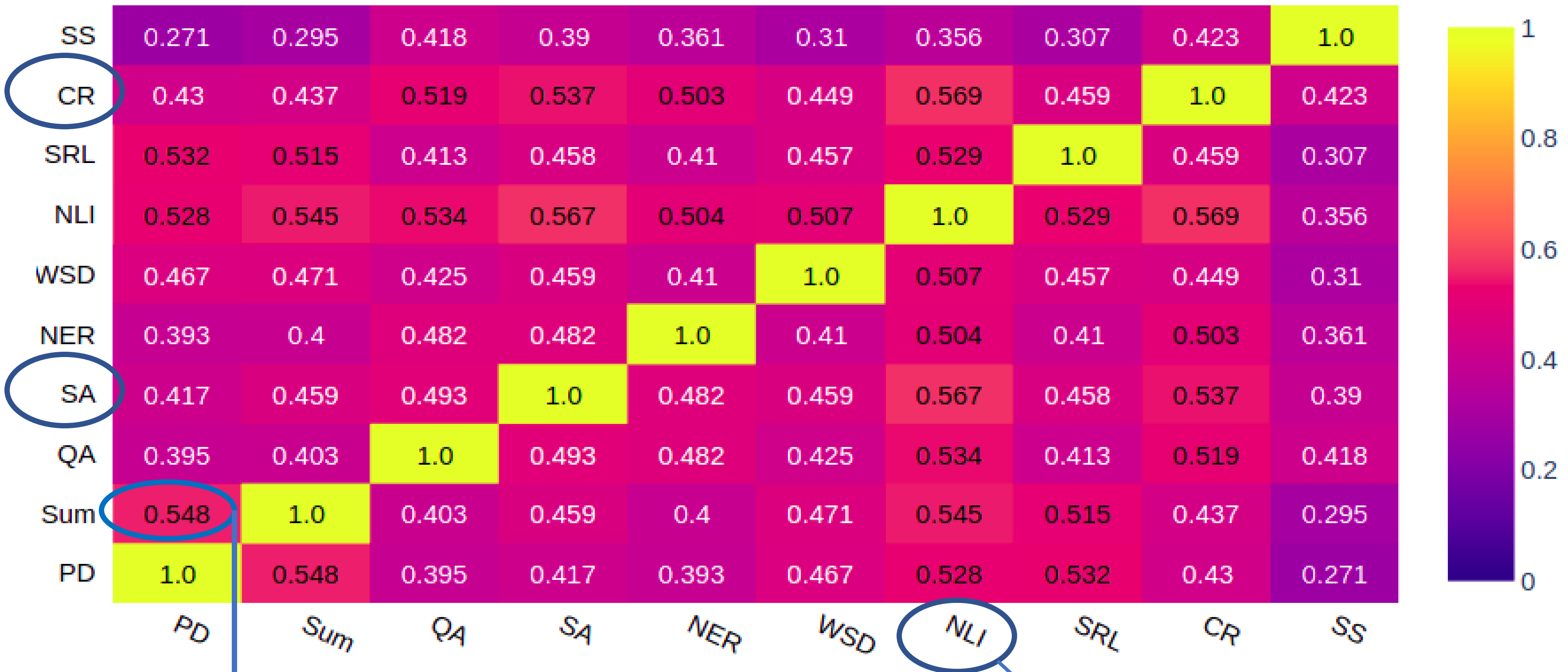
Task Similarity - Reading



Correlation between brain activity predicted with Summarization task and Paraphrase detection task

CR with NER, and CR with SS have high similarity match.

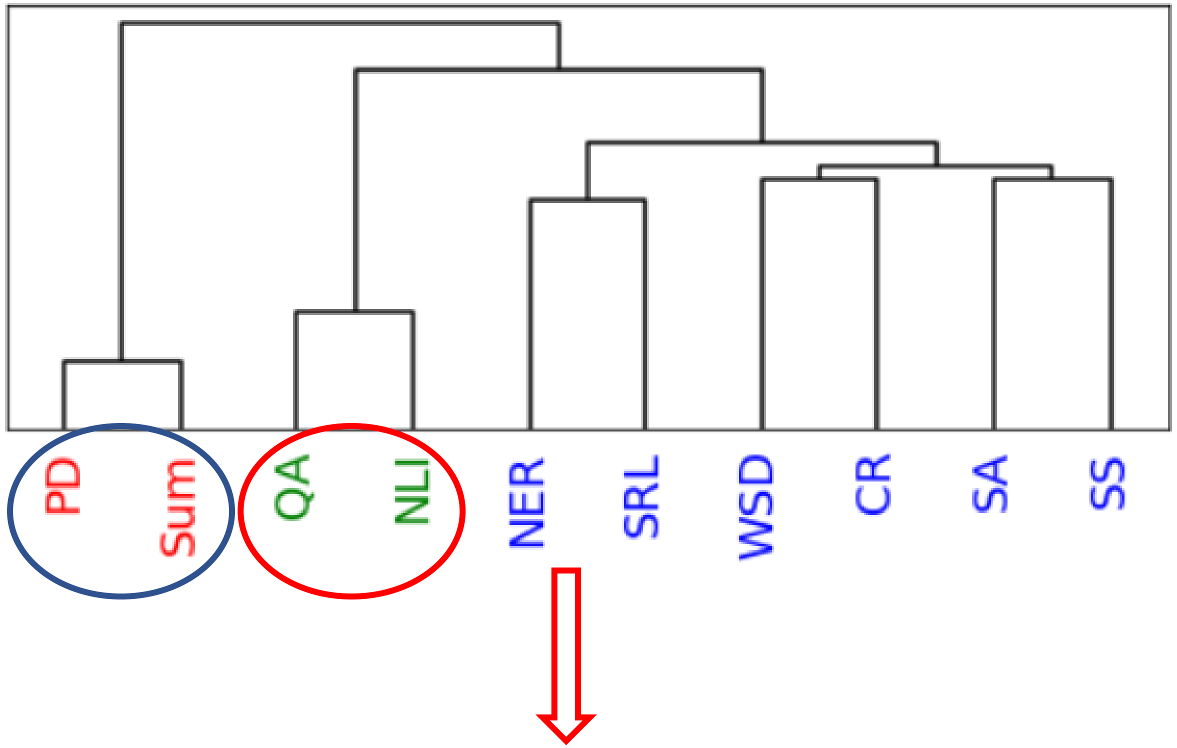
Task Similarity - Listening



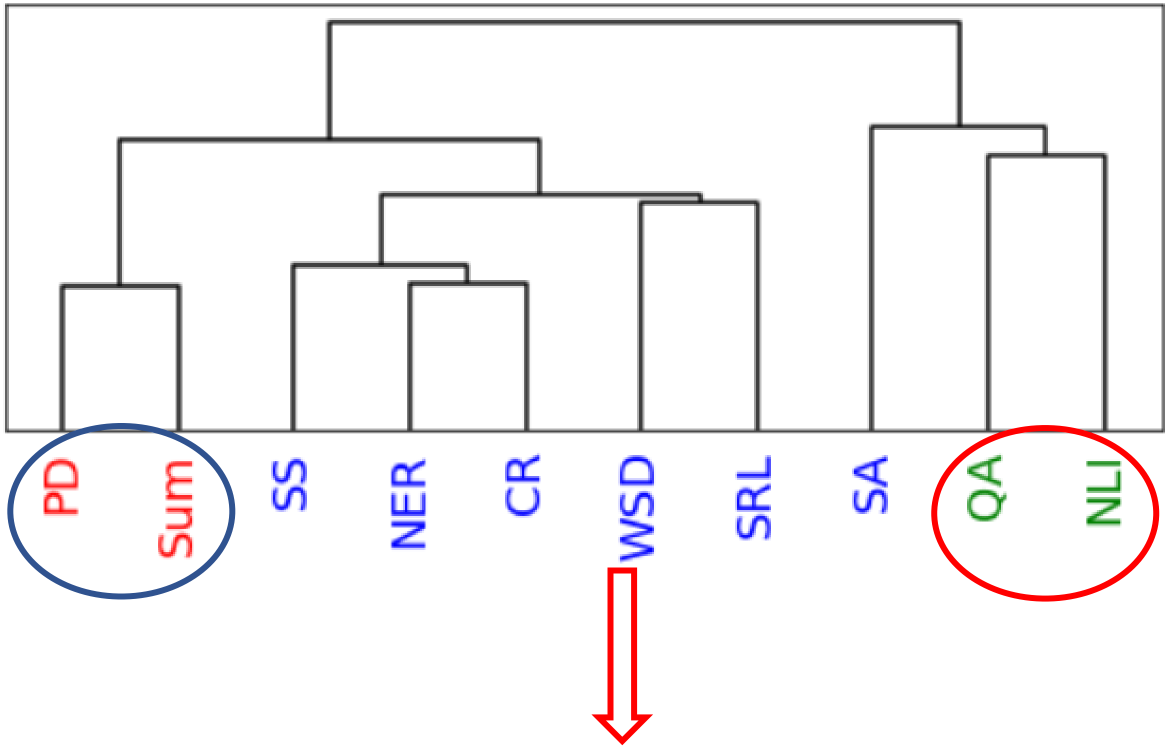
Correlation between brain activity predicted with Summarization task and Paraphrase detection task

NLI with CR, and NLI with SA have high similarity match.

Reading Task: Dendrogram

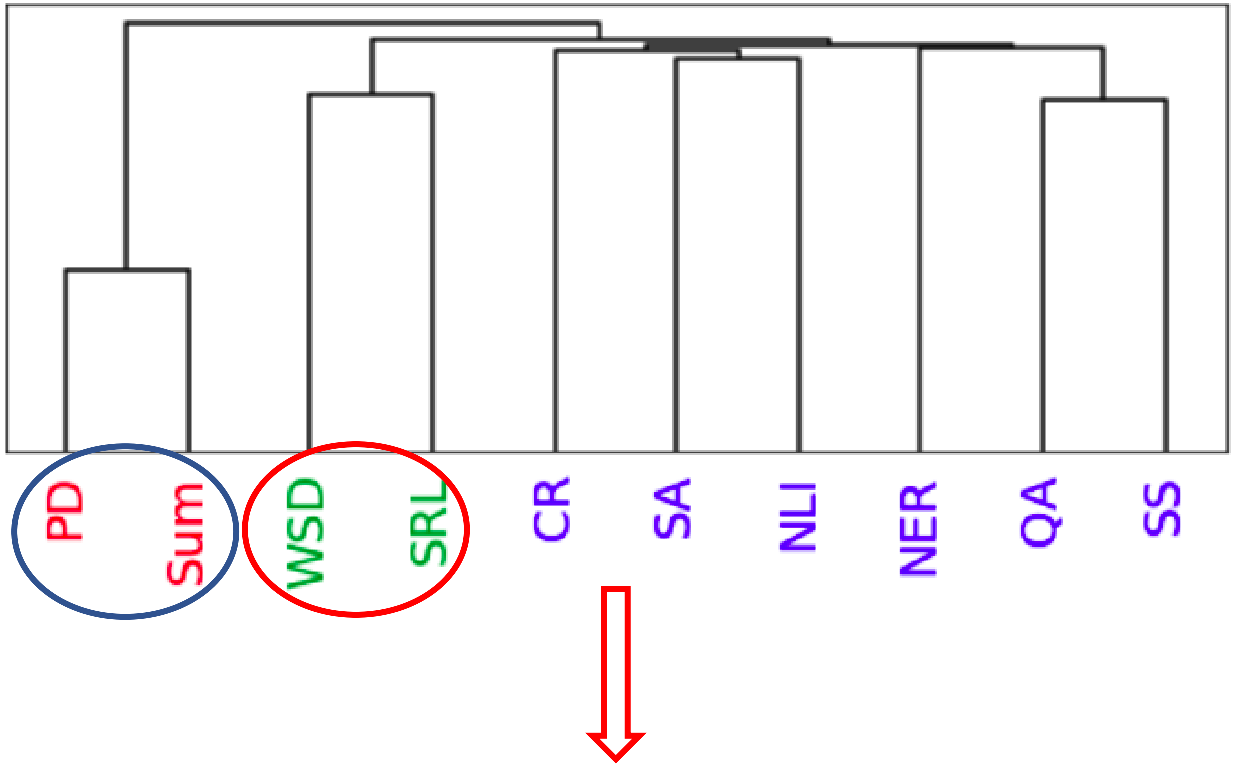


Similarity matrix with task-specific stimulus representations.

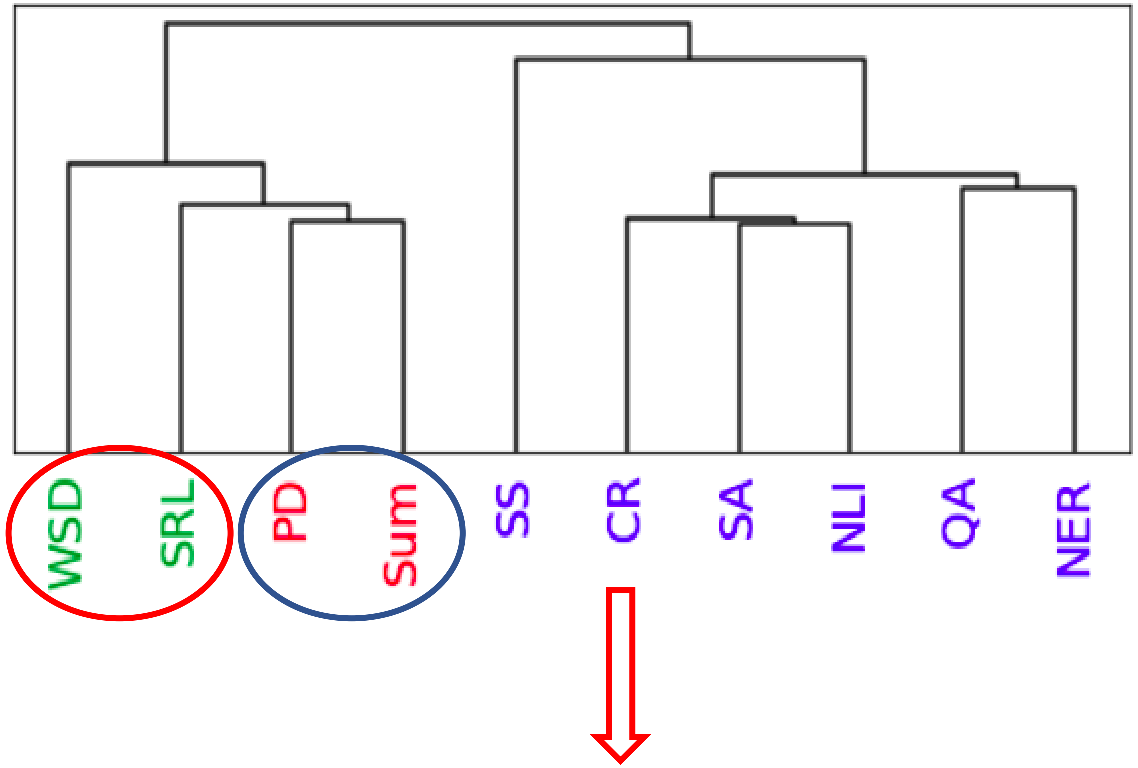


Similarity matrix with task-specific predicted brain activity.

Listening Task: Dendrogram

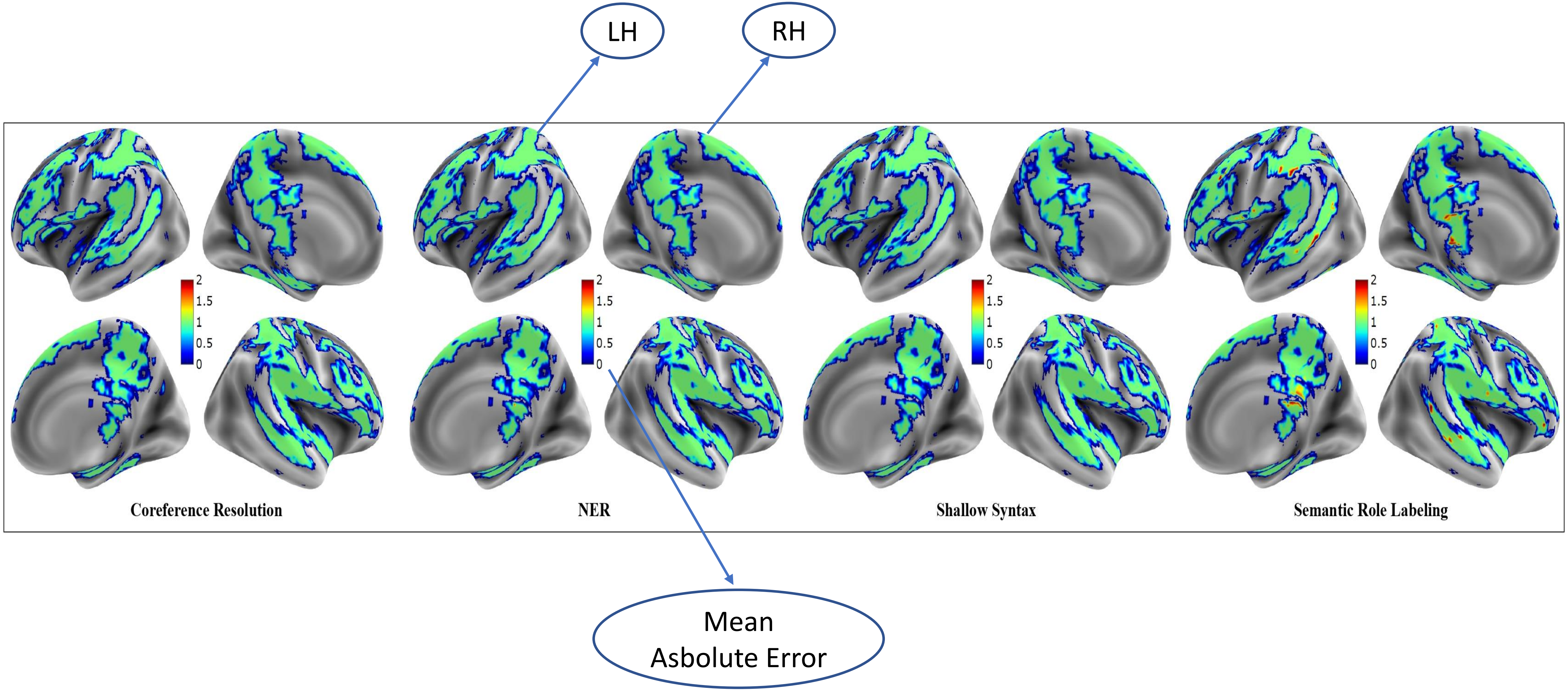


Similarity matrix with task-specific stimulus representations.

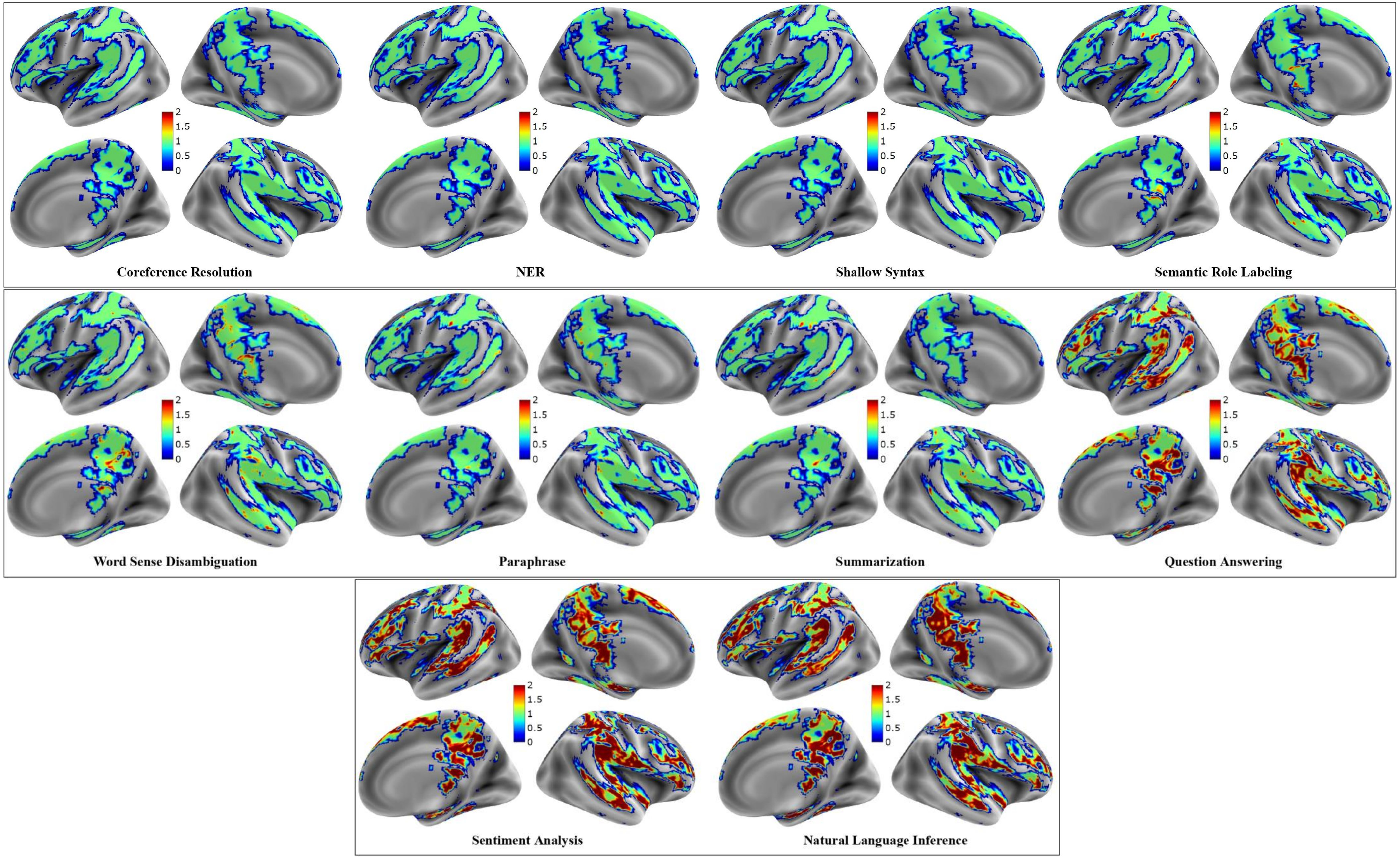


Similarity matrix with task-specific predicted brain activity.

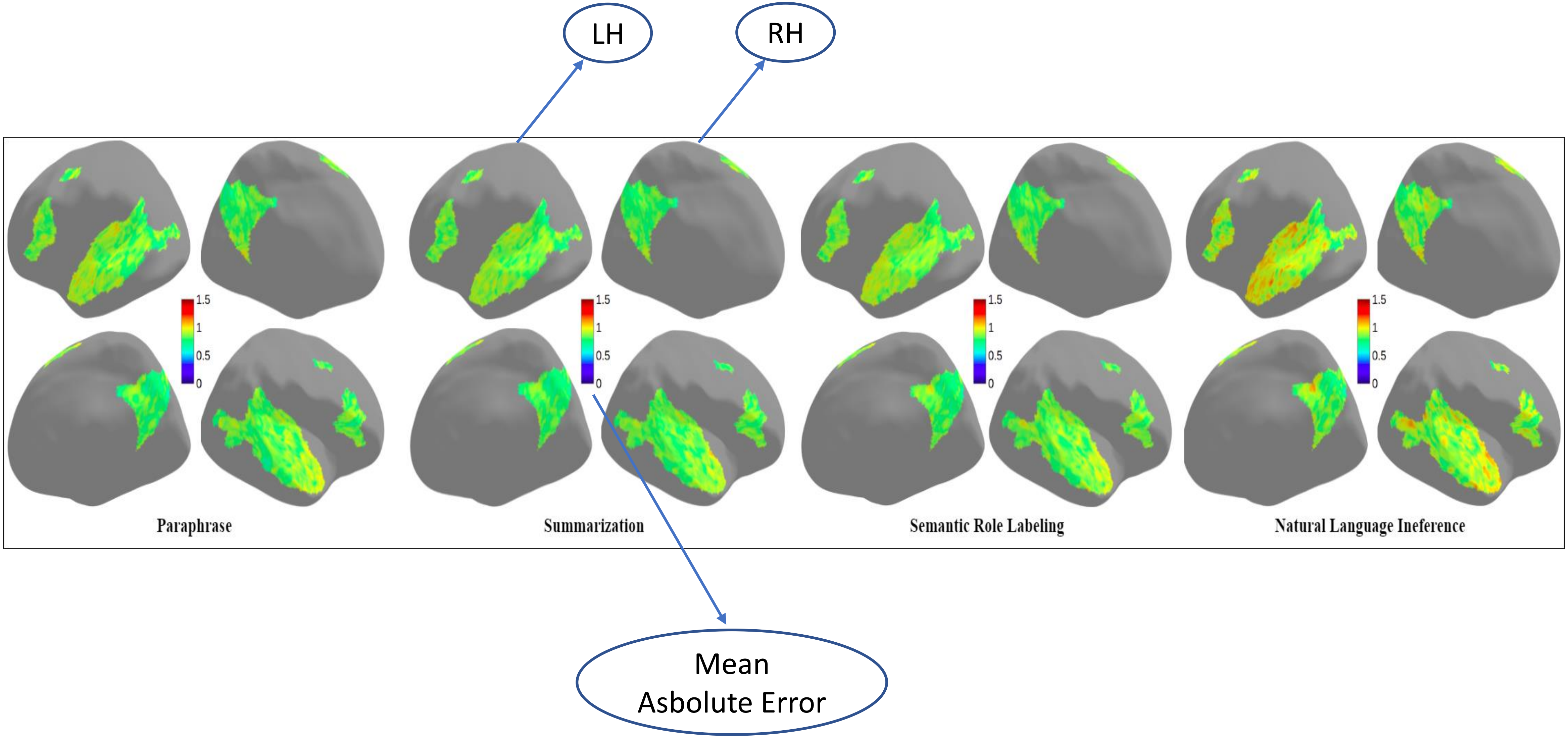
Brain Maps (Reading)



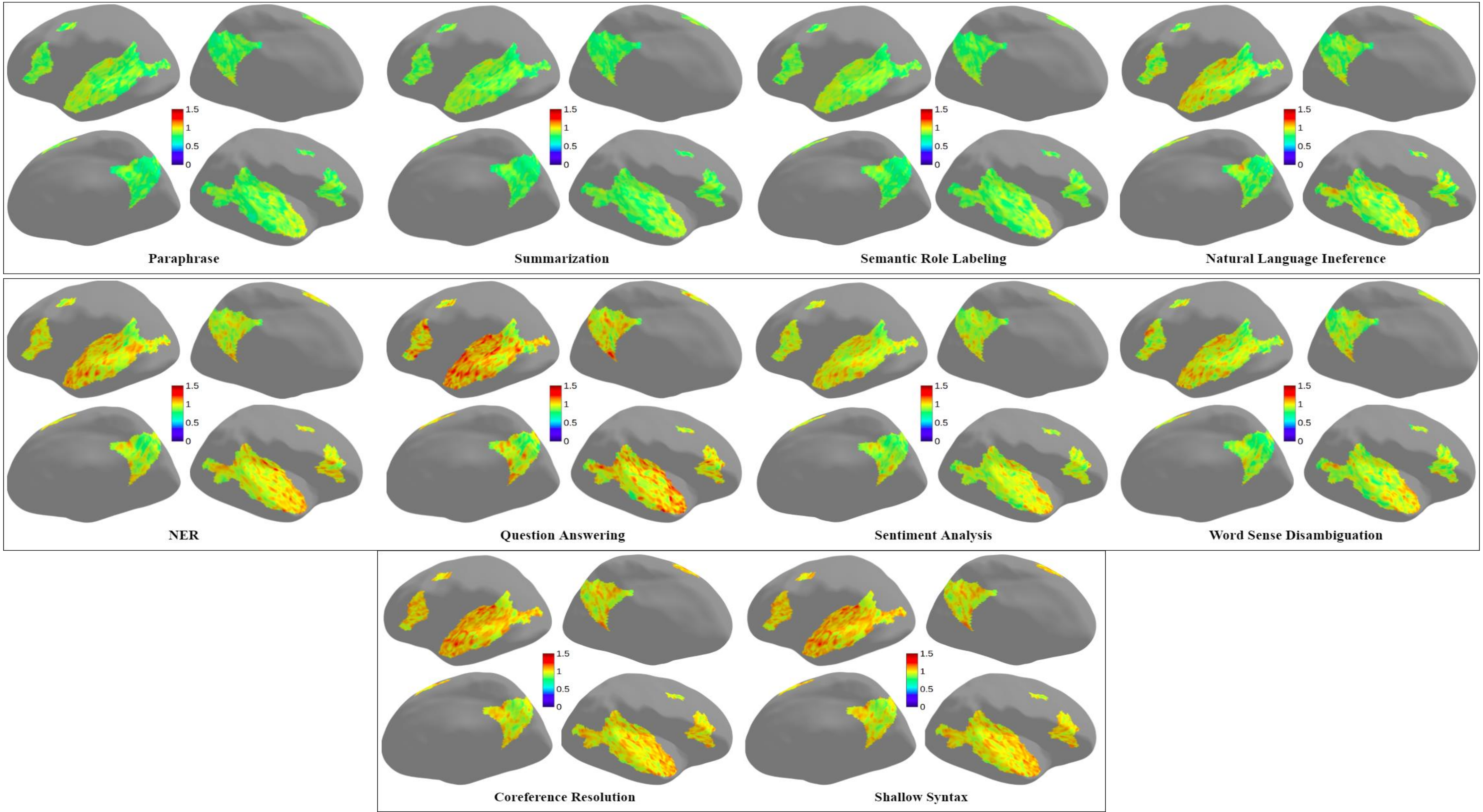
Brain Maps (Reading)



Brain Maps (Listening)



Brain Maps (Listening)



Limitations & Future Works

- We leveraged models finetuned using datasets of different sizes across tasks.
- While a fair comparison of dataset sizes across tasks is impossible,
  - we understand that this could have resulted in some bias in our results.
- The differences in task-specific model performances across reading and listening are mainly due to the learned stimulus representations,
  - other factors such as experimental conditions, the text domain of the stimuli or number of voxels also effect the model performance.

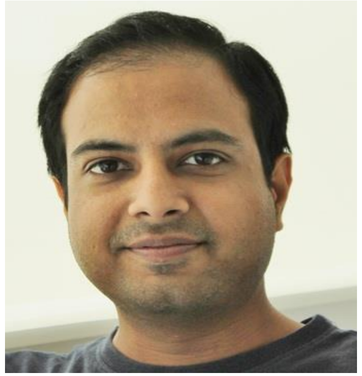
Collaborators

Subba Reddy Oota

Jashn Arora

Veeral Agarwal

Mounika Marreddy



Manish Gupta

Bapi Raju Surampudi

Acknowledgement

- We thank Dr. Mariya Toneva for her very helpful comments on this paper.
- We thank Google for the travel grant support.