Language and the Brain: Deep Learning for Brain Encoding and Decoding

Subba Reddy Oota[1], Manish Gupta[2,3], Raju S. Bapi[2]

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

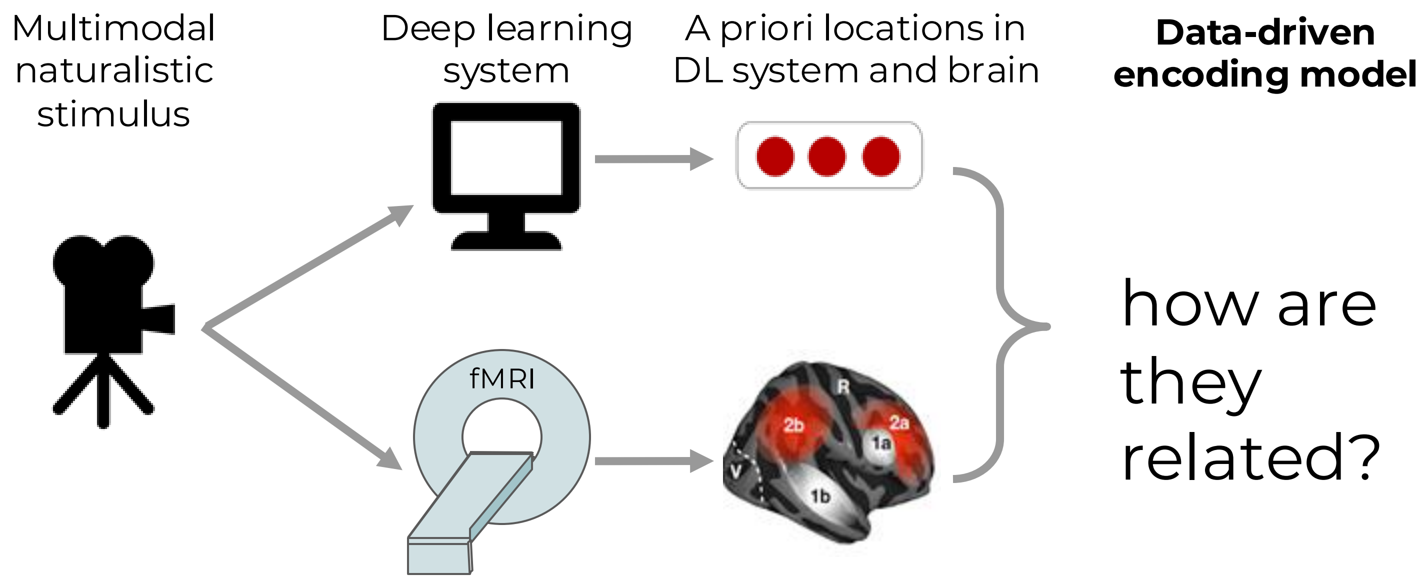
subba-reddy.oota@inria.fr, gmanish@microsoft.com, raju.bapi@iiit.ac.in



Agenda

- Introduction to Brain encoding and decoding [10 min]
- Text Stimulus Representations [30 min]
- **Deep Learning for Brain Encoding [40 min]**
- Deep Learning for Brain Decoding [30 min]
- Summary and Future Trends [10 min]

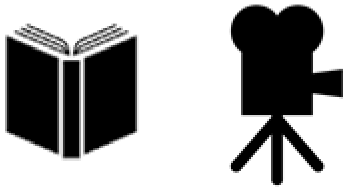
Data-driven encoding models evaluate the relationships between brains and deep learning models



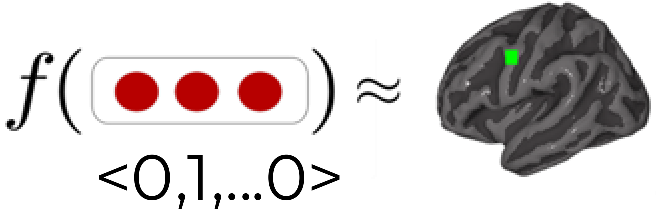
[Toneva, M., & Wehbe, L. (2019). Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain). Advances in Neural Information Processing Systems, 32.]()

Deep learning models enable data-driven encoding models for naturalistic stimuli

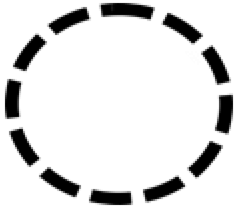
more naturalistic stimuli



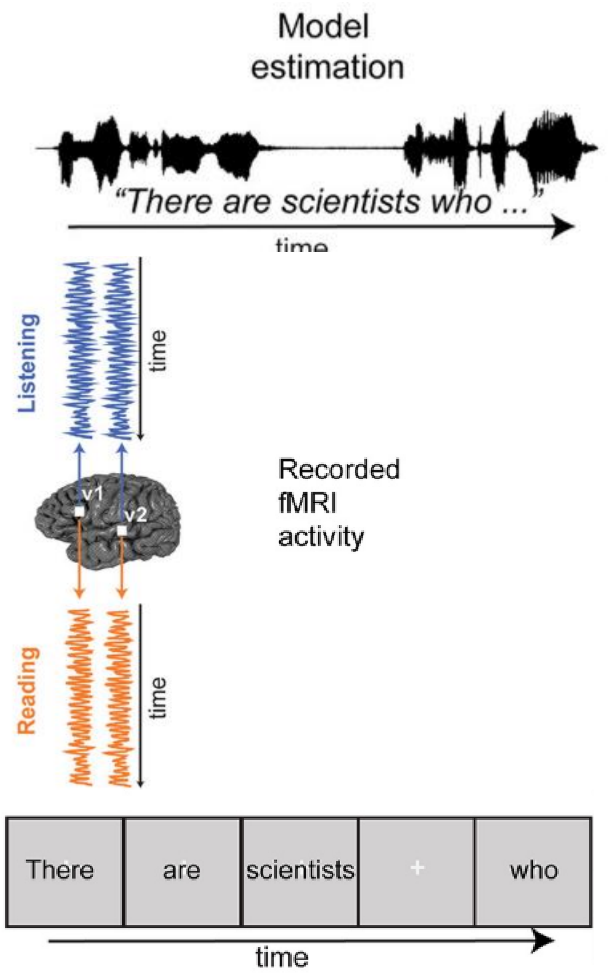
simple stim. representations explain less variance in brain activity

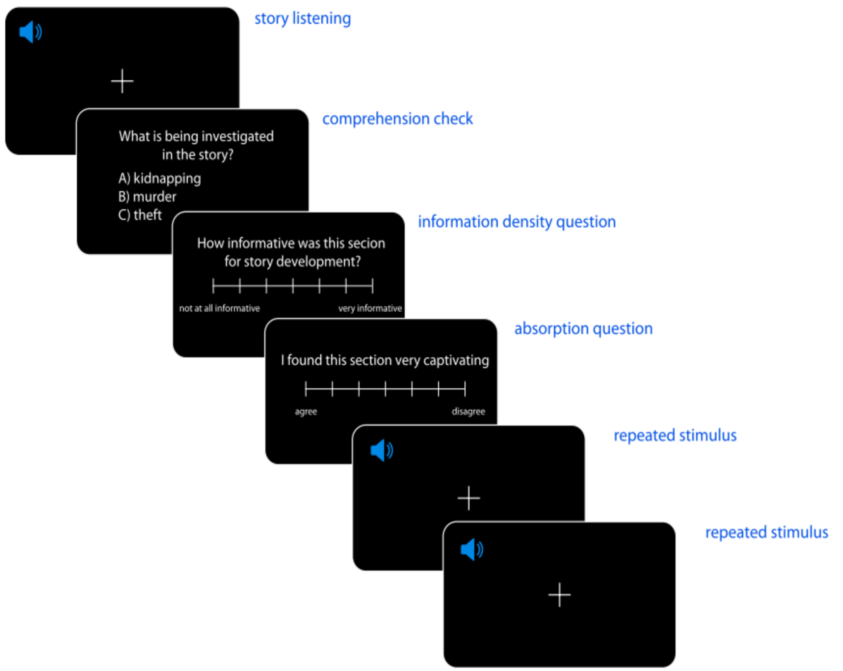


more stimulus properties that affect brain activity

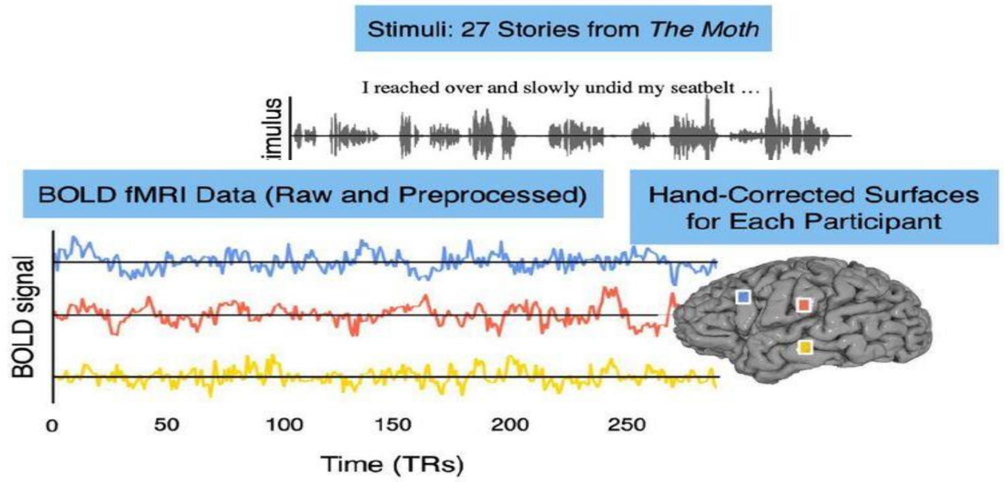


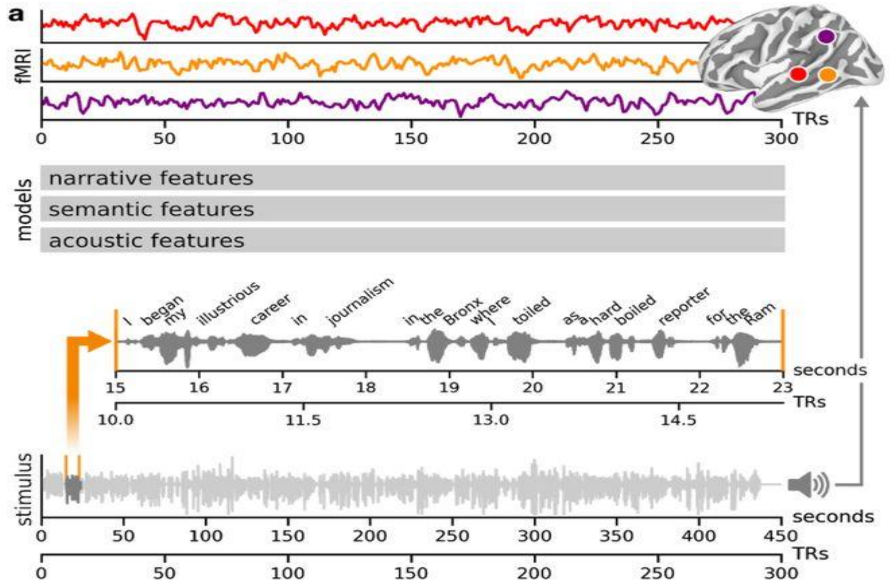
Deep learning models enable data-driven encoding models for naturalistic stimuli





	Participants			Data Collection	Stimuli			
Language	Number	Mean Age	Female	Location	Material	Length (s)	N Words	N Sentences
English	49	21.3	30	Cornell University, United States	The little prince EN audiobook	5632	15376	1499
Chinese	35	19.9	15	Jiangsu Normal University, China	The little prince CN audiobook	5954	16009	1577
French	28	24.4	15	NeuroSpin, France	The little prince FR audiobook	5828	15391	1480





[Fatma Deniz, Anwar O. Nunez-Elizalde, Alexander G. Huth and Jack L. Gallant. The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. Journal of Neuroscience, 2019.]

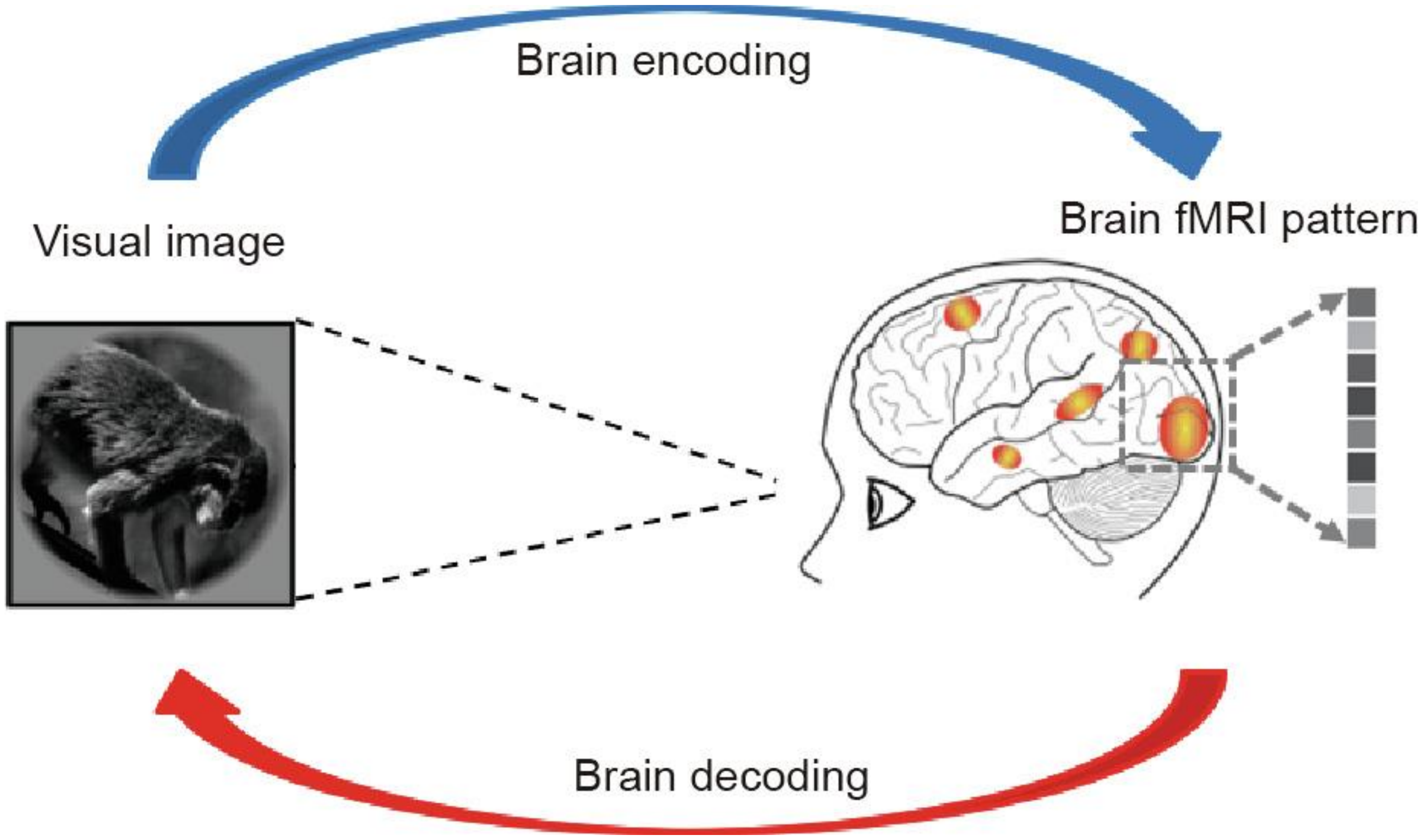
[Samuel A. Nastase. The "Narratives" fMRI dataset for evaluating models of naturalistic language comprehension. Nature, 2021.]

[Jixing Li. Le Petit Prince multilingual naturalistic fMRI corpus. Nature, 2022.]

[Amanda LeBel, Lauren Wagner, Shailee Jain, Aneesh Adhikari-Desai, Bhavin Gupta, Allyson Morgenthal, Jerry Tang, Lixiang Xu, Alexander G. Huth. A natural language fMRI dataset for voxelwise encoding models. Arxiv, 2022.]

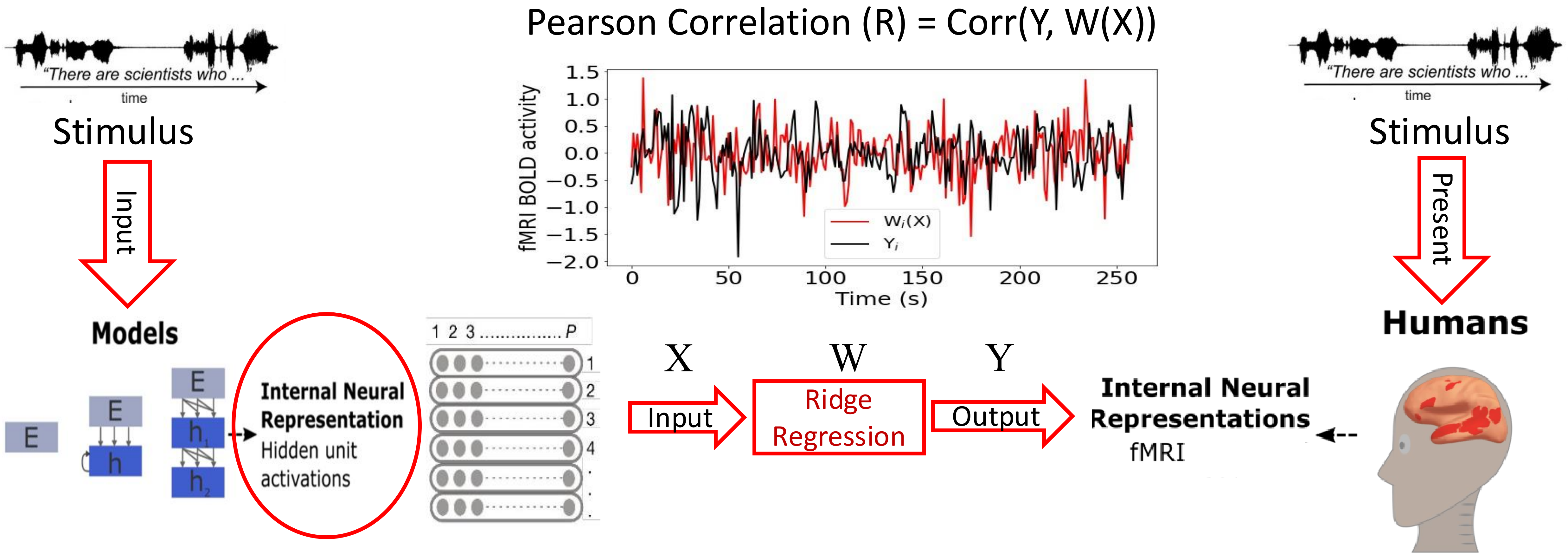
Encoding (Well-posed) vs Decoding (Ill-posed) in Neuroscience

- Encoding: How is the stimulus represented in the brain?
- Decoding: Can we reconstruct the stimulus, given the brain response?



Changde Du, Jinpeng Li, Lijie Huang, Huiguang He. Brain Encoding and Decoding in fMRI with Bidirectional Deep Generative Models. Science Direct 2019.

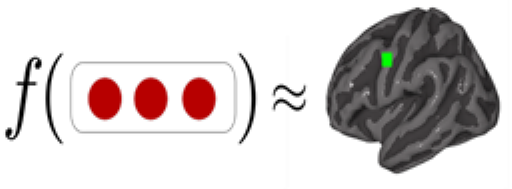
Brain Encoding?



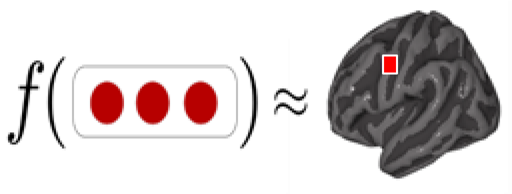
Encoding: training **independent** models

- Independent model per participant

P1

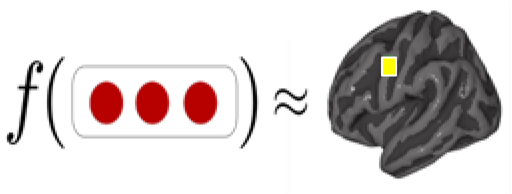


P2



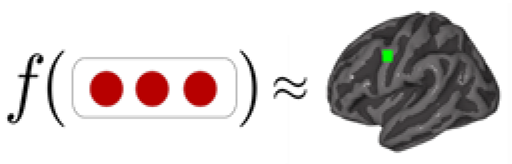
...

PN



- Independent model per voxel / sensor-timepoint

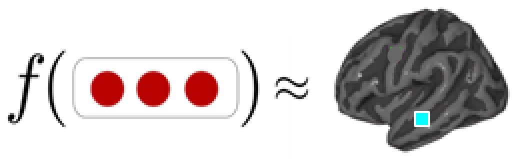
P1, v1



P1, v2

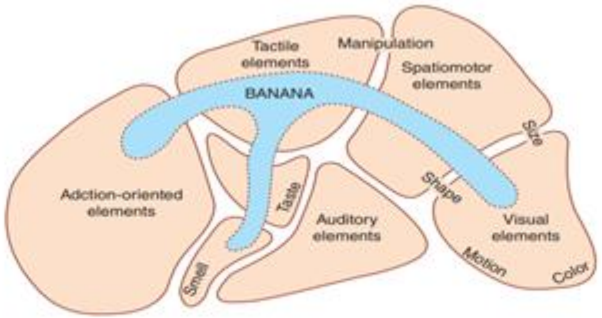
...

P1, vm

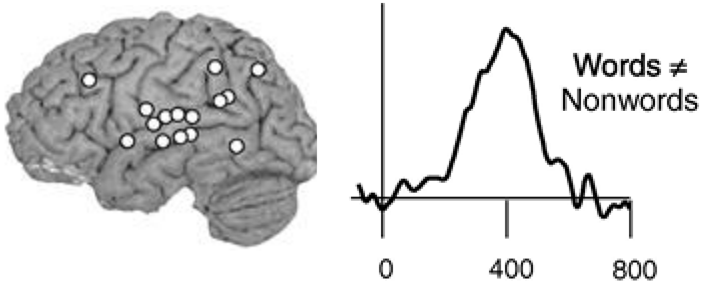


Mechanistic understanding of information processing in the brain: 4 big questions

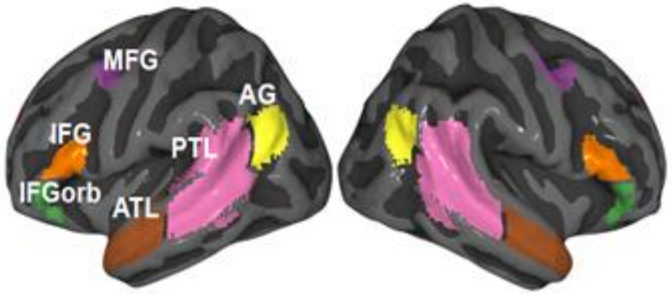
What



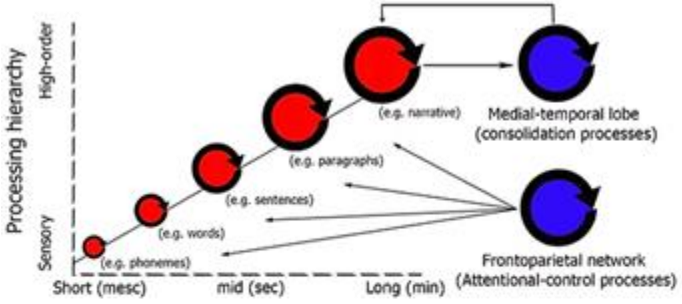
When



Where



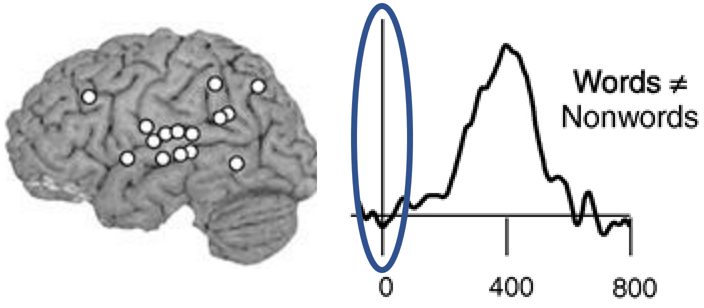
How



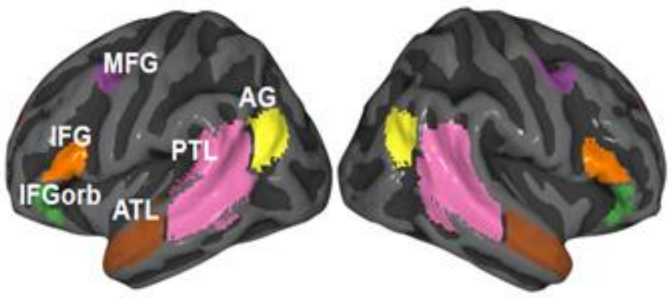
With MEG we can analyze sub-word time course

- MEG recording data at very fast temporal resolution
- So, we can look at sub-word process
- fMRI recording data at very high-spatial resolution

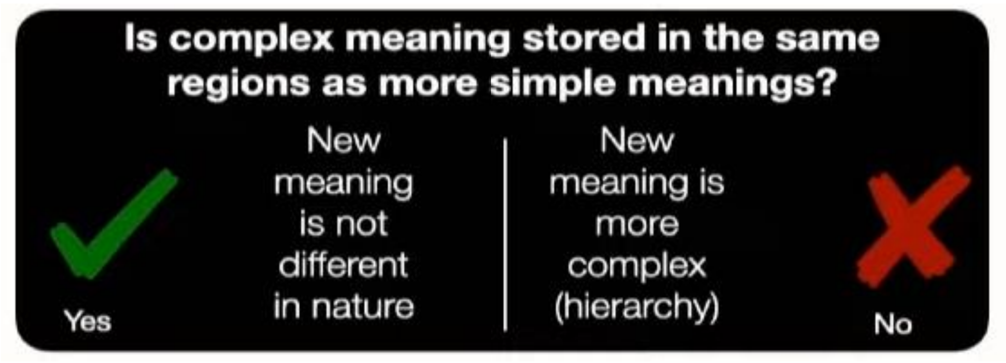
When

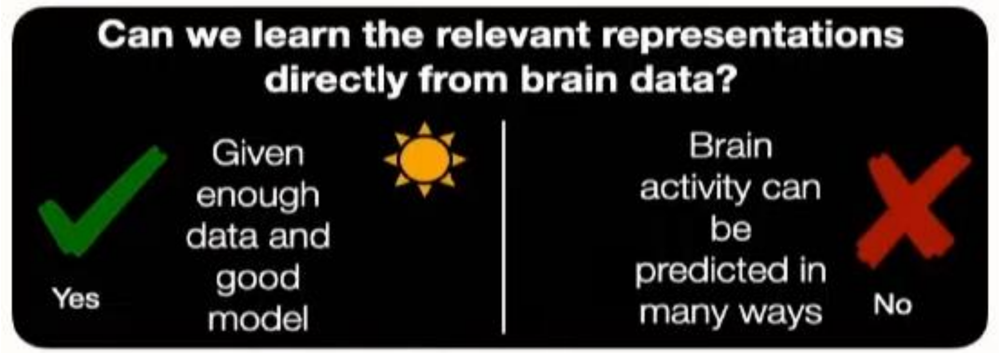


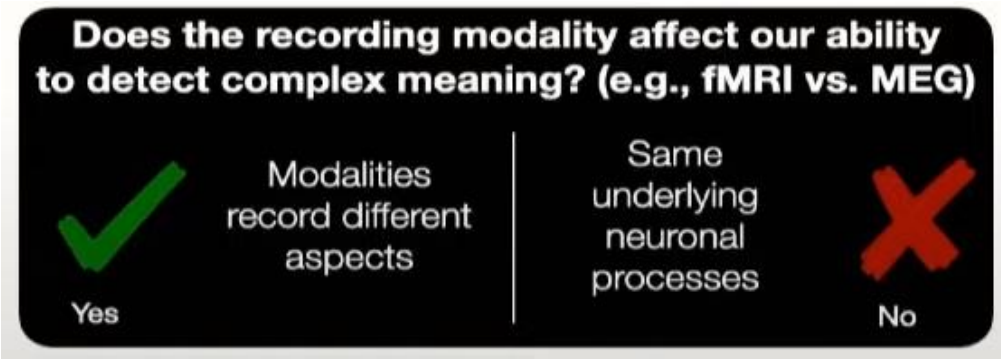
Where

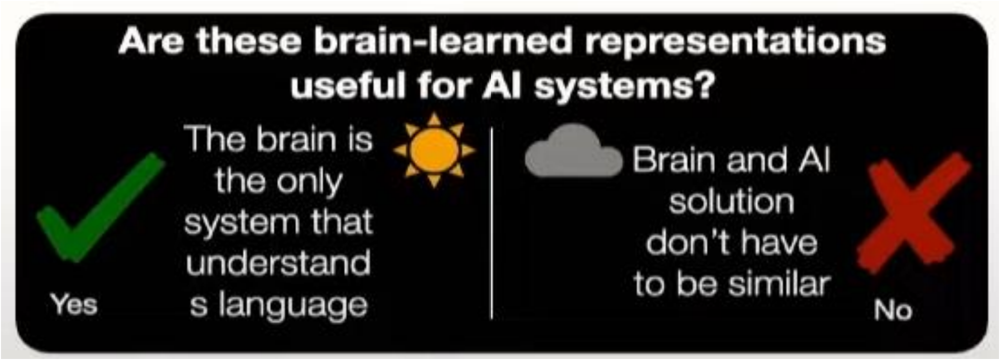


How does brain represents complex meaning? (Where, When and What)









Word Context

1-word context vehicle

2-word context [speed] vehicle   vehicle [down]

3-word context [high] [speed] vehicle   vehicle [down] [the]

4-word context [my] [high] [speed] vehicle   vehicle [down] [the] [road]

5-word context [drove] [my] [high] [speed] vehicle   vehicle [down] [the] [road] [today]

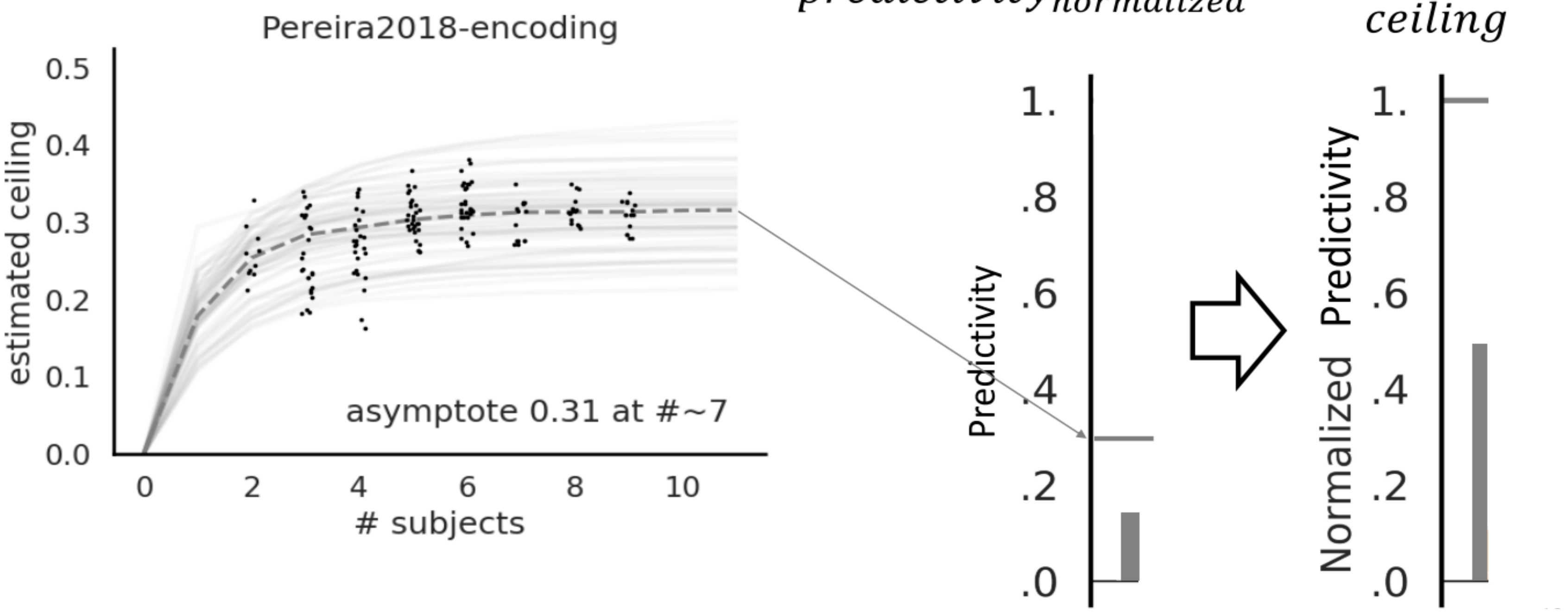
Past context   *Future context*

Normalized Predictivity

“how close are we” – **ceiling**

compute how well a pool of subjects predicts a held-out subject

$$predictivity_{normalized} = \frac{predictivity}{ceiling}$$



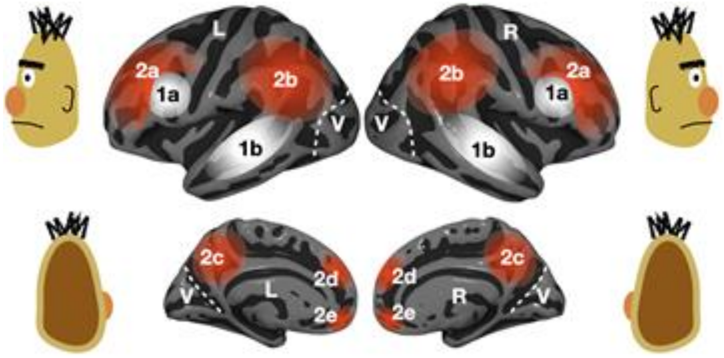
Schrimpf, Martin, Idan Asher Blank, Greta Tuckute, Carina Kauf, Eghbal A. Hosseini, Nancy Kanwisher, Joshua B. Tenenbaum, and Evelina Fedorenko. "The neural architecture of language: Integrative modeling converges on predictive processing." Proceedings of the National Academy of Sciences 118, no. 45 (2021): e2105646118.

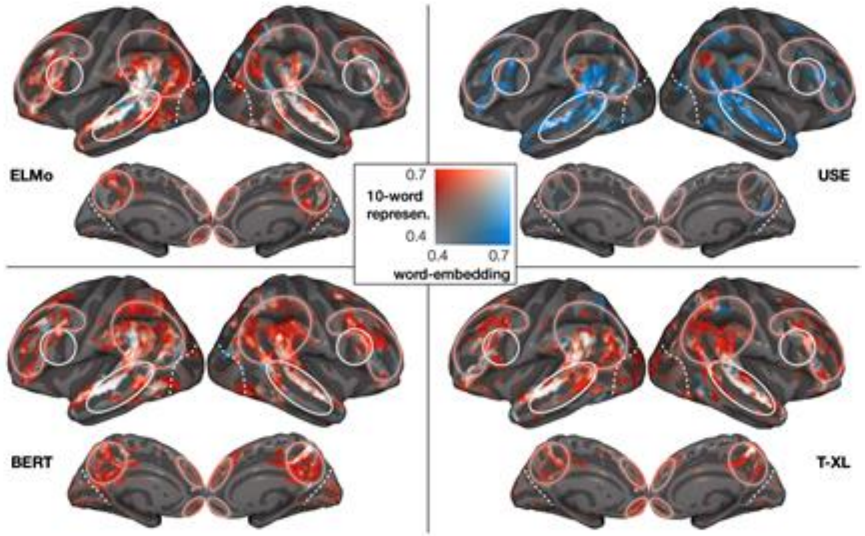
Recent work utilizing progress in LLMs for encoding

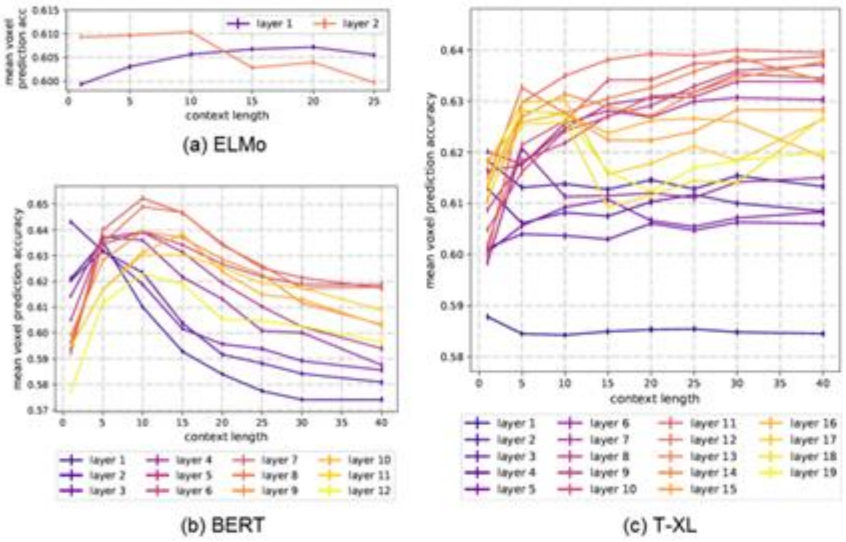
- Using representations of stimuli from deep learning systems
- **Language:**
  - Wehbe et al. 2014;
  - Jain and Huth, 2018;
  - Toneva and Wehbe, 2019;
  - Caucheteux and King, 2020/2022;
  - Schrimpf et al. 2020/2021;
  - Goldstein et al. 2021/2022;
  - Toneva and Wehbe, 2022/2023;
  - Khai et al. 2023
  - Oota et al. 2022/2023;

Language: work utilizing DL progress

- Stimuli: one chapter of Harry Potter
- Stimulus representation: derived from **pretrained** NLP systems
- Brain recording & modality: fMRI, reading







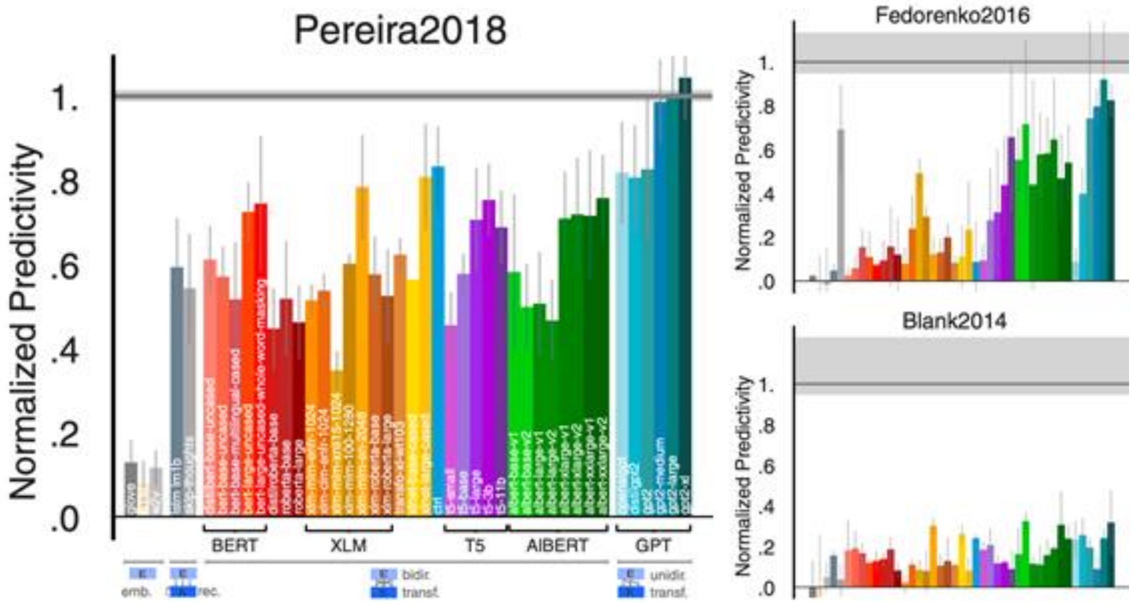
across several types of large NLP systems, best alignment with fMRI in middle layers

Toneva, M., & Wehbe, L. (2019). Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain). Advances in Neural Information Processing Systems, 32.

Language: work utilizing DL progress

- Stimuli: sentences, passages, short story
- Stimulus representation: derived from pretrained NLP systems (BERT, GPT-2, T5 , and XLM)
- Brain recording & modality: fMRI & ECoG, reading & listening

some NLP systems can predict fMRI and ECoG up to 100% of estimated noise ceiling



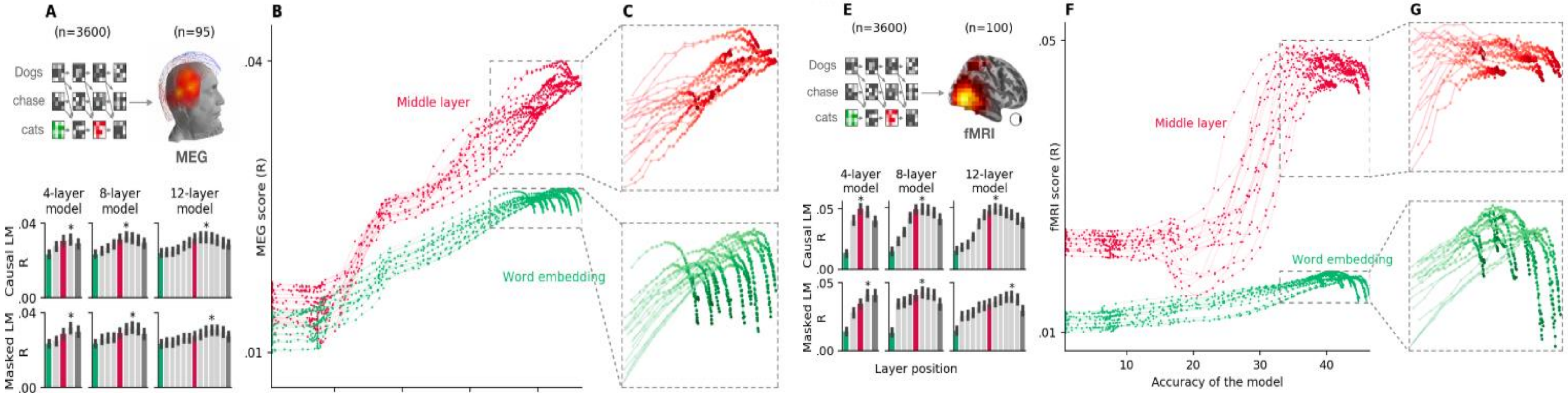
Schrimpf, Martin, Idan Asher Blank, Greta Tuckute, Carina Kauf, Eghbal A. Hosseini, Nancy Kanwisher, Joshua B. Tenenbaum, and Evelina Fedorenko. "The neural architecture of language: Integrative modeling converges on predictive processing." Proceedings of the National Academy of Sciences 118, no. 45 (2021): e2105646118.

Language: work utilizing DL progress

best alignment with fMRI & MEG in middle layers

better performance at predicting next word -> better prediction of fMRI & MEG

- Stimuli: sentences
- Stimulus representation: derived from pretrained NLP systems (BERT and GPT-2)
- Brain recording & modality: MEG & fMRI, reading

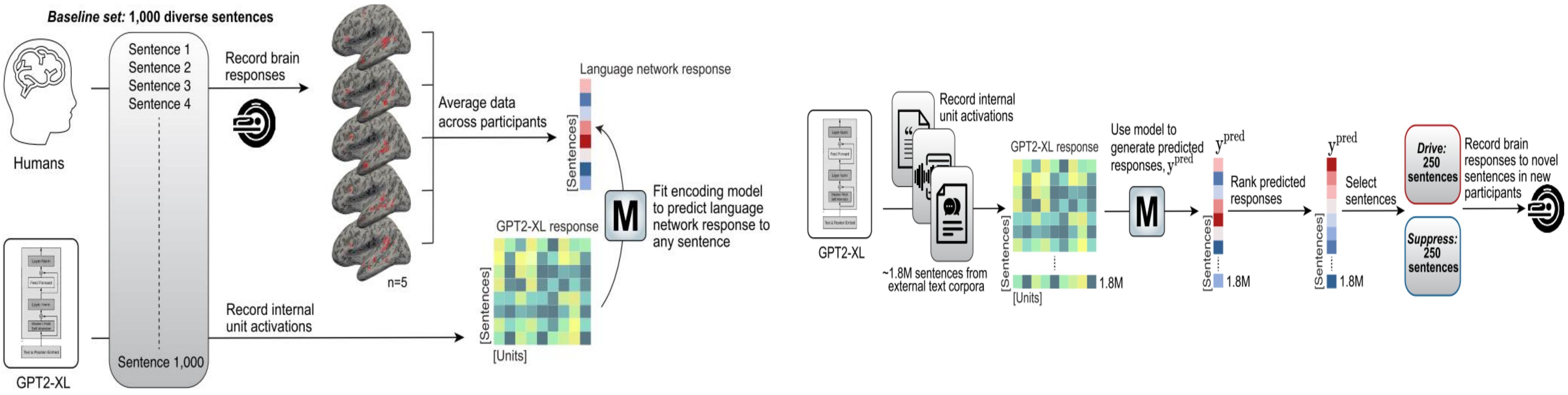


Caucheteux, Charlotte, and Jean-Rémi King. "Brains and algorithms partially converge in natural language processing." Communications biology 5, no. 1 (2022): 1-10.

Language: work utilizing DL progress

model-selected 'out-of-distribution' sentences indeed drive and suppress activity of human language areas in new individuals

- Stimuli: sentences
- Stimulus representation: derived from pretrained NLP systems (GPT-2 XL)
- Brain recording & modality: fMRI, reading



Greta Tuckute et al. 2023. "Driving and suppressing the human language network using large language models."

Challenges in using DL for cognitive science

- Not designed to specifically model brain processing

NLP systems: Designed to predict upcoming words

Harry never thought ***<u>???</u>***

Harry never thought he ***<u>???</u>***

Harry never thought he would ***<u>???</u>***

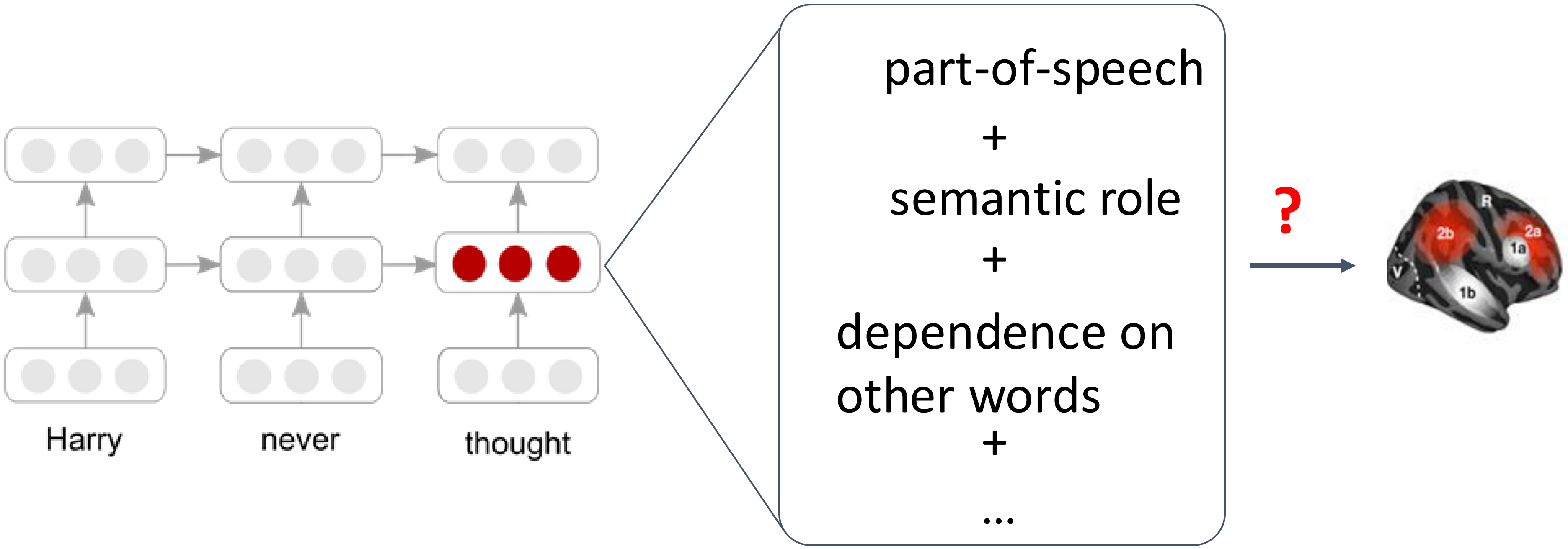
…

Challenges in using DL for cognitive science

- Not designed to specifically model brain processing
  - Training DL models using brain recordings
  - Task-based modeling

Challenges in using DL for cognitive science

- Not designed to specifically model brain processing
  - Training DL models using brain recordings
  - Task-based modeling
- Can be difficult to interpret due to multiple sources of information



Challenges in using DL for cognitive science

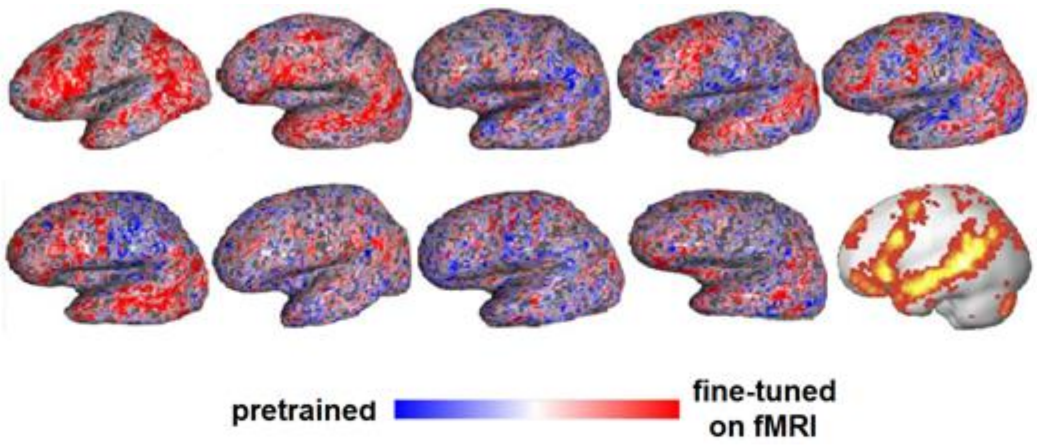
- Not designed to specifically model brain processing
  - Training DL models using brain recordings
  - Task-based modeling
- Can be difficult to interpret due to multiple sources of information
  - Disentangling contributions of different info sources to brain predictions

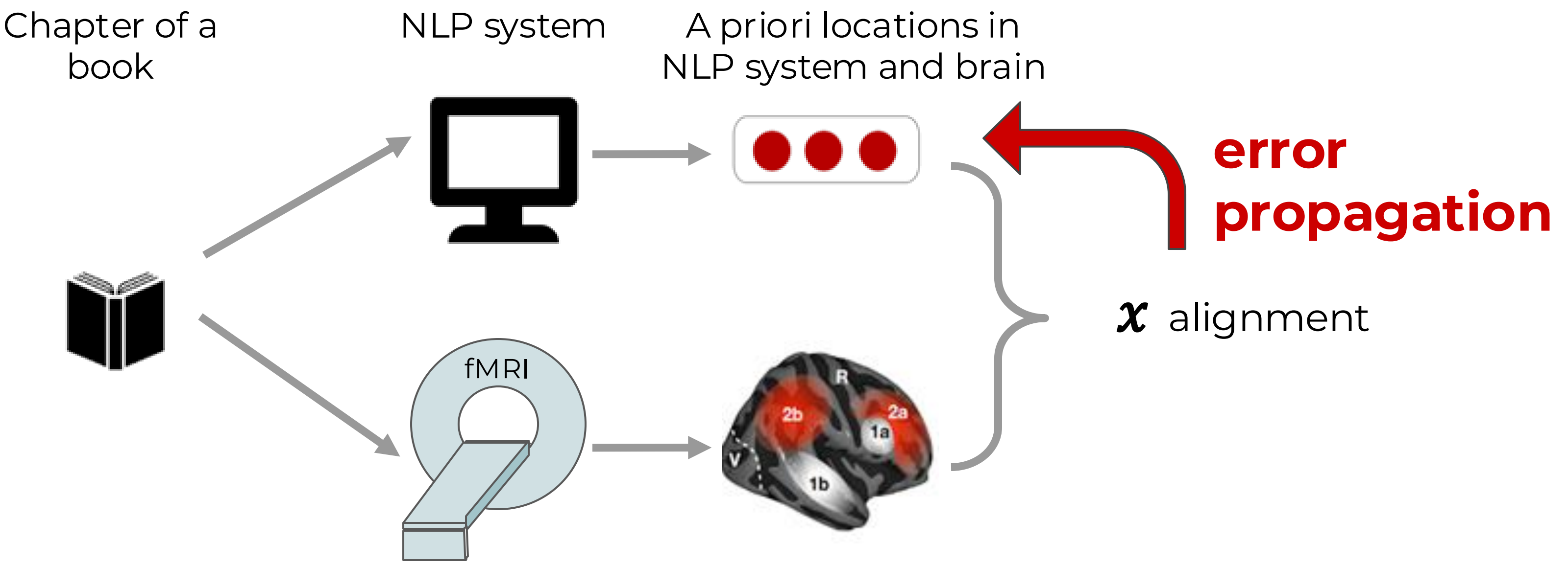
Challenges in using DL for cognitive science

- Not designed to specifically model brain processing
  - **Training DL models using brain recordings**
  - Task-based modeling
- Can be difficult to interpret due to multiple sources of information
  - Disentangling contributions of different info sources to brain predictions

Training DL models using brain recordings

- Stimuli: one chapter of Harry Potter
- Stimulus representation: brain-optimized NLP model
- Brain recording & modality: fMRI & MEG, reading

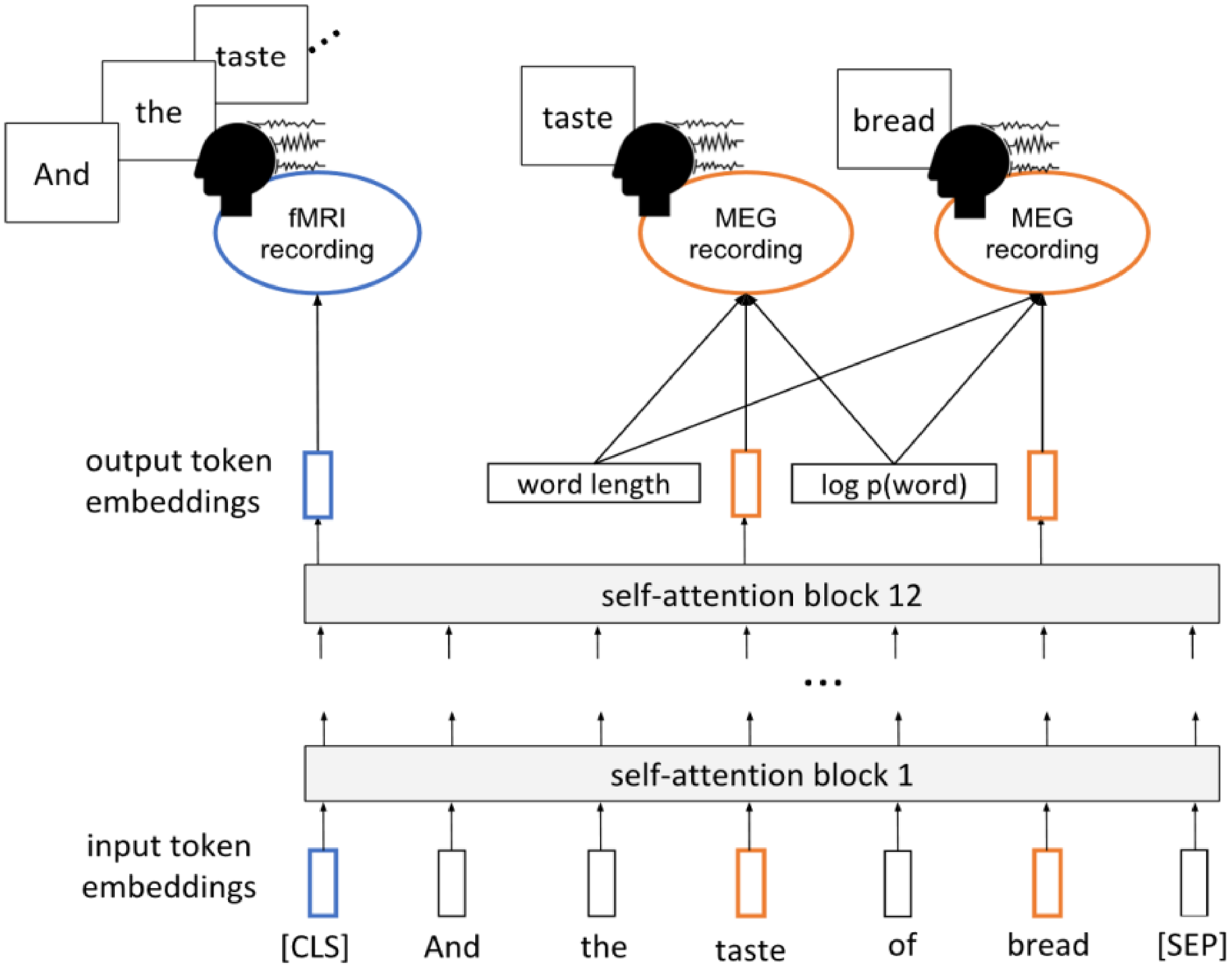




Brain-optimized NLP model predicts unseen fMRI recordings better, especially in canonical language regions

Schwartz, Dan, Mariya Toneva, and Leila Wehbe. "Inducing brain-relevant bias in natural language processing models." Advances in neural information processing systems 32 (2019).

Inducing Brain Relevant Bias



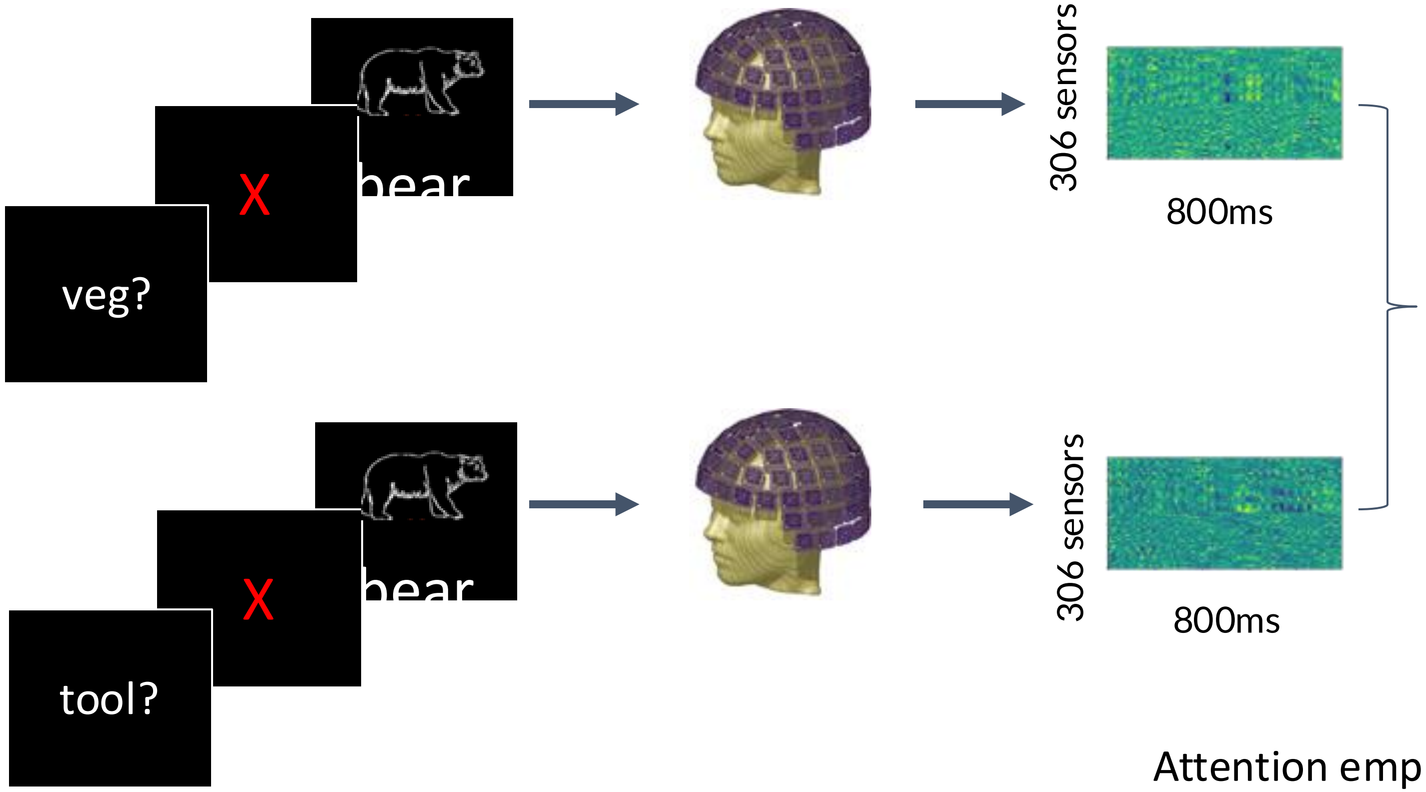
Metric	Vanilla	MEG	Joint
CoLA	57.29	57.63	**57.97**
SST-2	93.00	**93.23**	91.62
MRPC (Acc.)	83.82	83.97	**84.04**
MRPC (F1)	88.85	**88.93**	88.91
STS-B (Pears.)	**89.70**	89.32	88.60
STS-B (Spear.)	**89.37**	88.87	88.23
QQP (Acc.)	90.72	**91.06**	90.87
QQP (F1)	87.41	**87.91**	87.69
MNLI-m	83.95	**84.26**	84.08
MNLI-mm	84.39	84.65	**85.15**
QNLI	89.04	**91.73**	91.49
RTE	61.01	**65.42**	62.02
WNLI	53.52	**53.80**	51.97

Schwartz, Dan, Mariya Toneva, and Leila Wehbe. "Inducing brain-relevant bias in natural language processing models." Advances in neural information processing systems 32 (2019).

Challenges in using DL for cognitive science

- Not designed to specifically model brain processing
  - Training DL models using brain recordings
  - **Task-based modeling**
- Can be difficult to interpret due to multiple sources of information
  - Disentangling contributions of different info sources to brain predictions

Tasks affect processing



Systematic difference due to different question tasks

Can we model as a function of the task **AND** stimulus?

Attention emphasizes task-relevant information

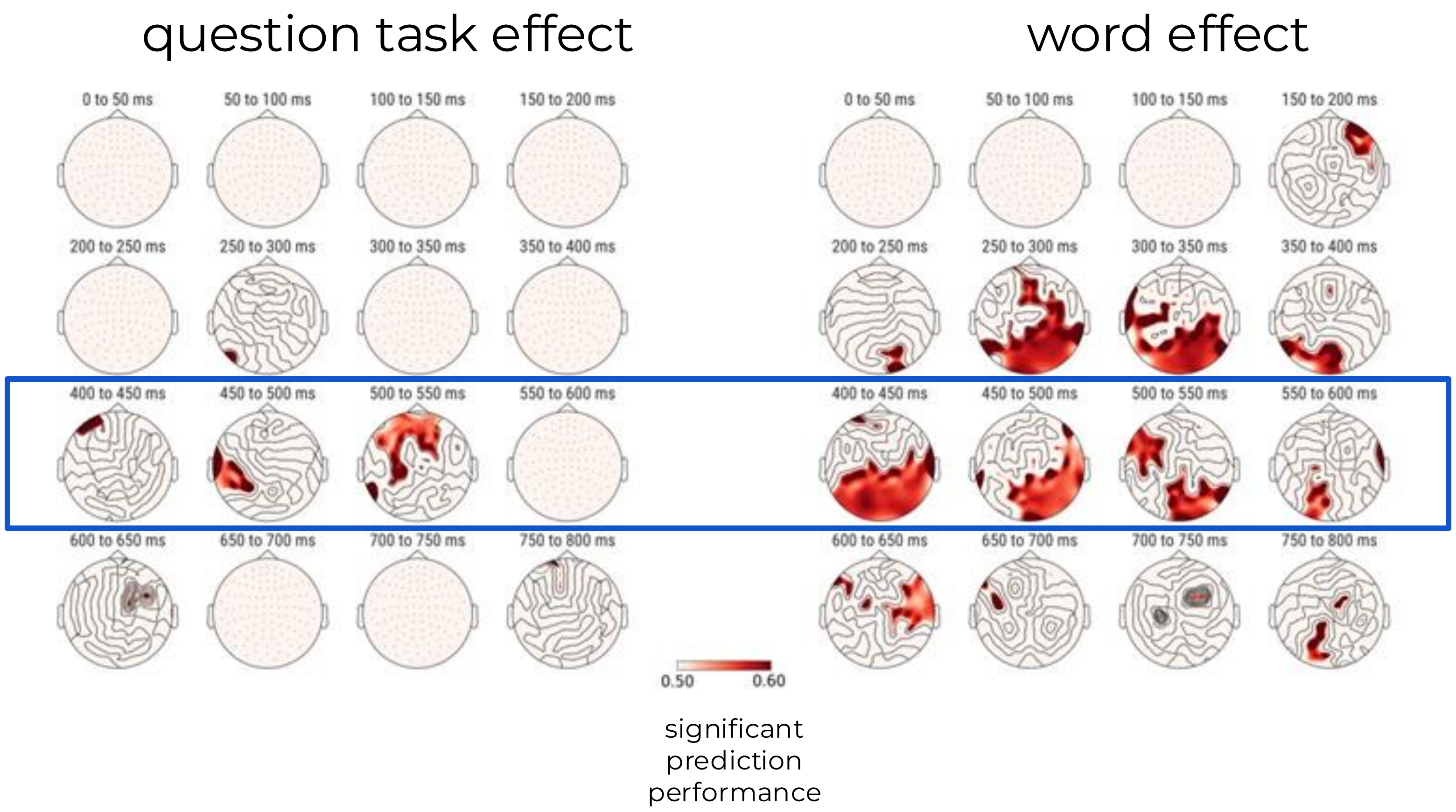
Mechanism?

Toneva, Mariya, Otilia Stretcu, Barnabás Póczos, Leila Wehbe, and Tom M. Mitchell. "Modeling task effects on meaning representation in the brain via zero-shot meg prediction." Advances in Neural Information Processing Systems 33 (2020): 5284-5295.

Tasks affect processing

The end of semantic processing of a word is task-dependent

- Stimuli: concrete nouns + line drawings
- Task: answer Yes/No questions about noun
- Stimulus representation: human judgments
- Brain recording & modality: MEG, reading

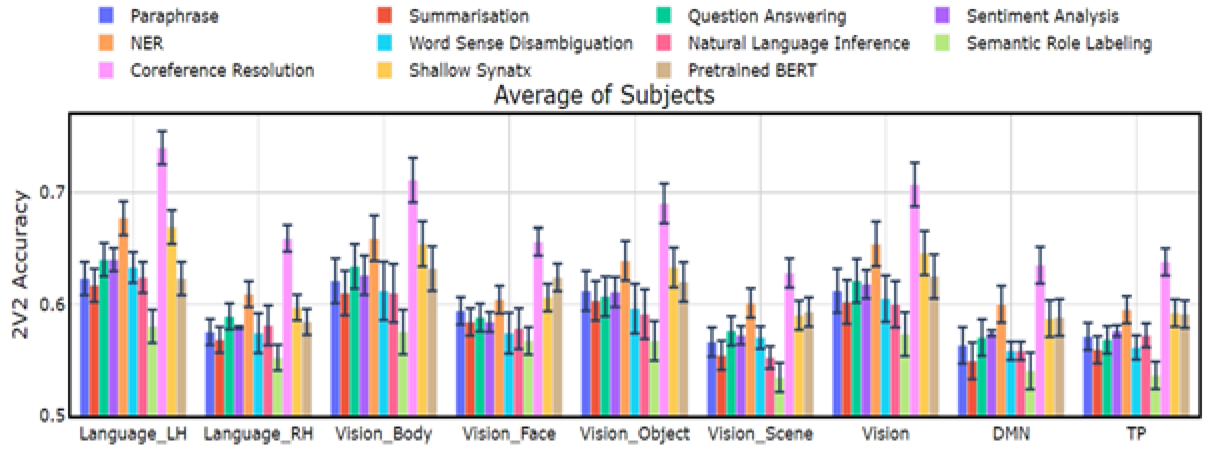


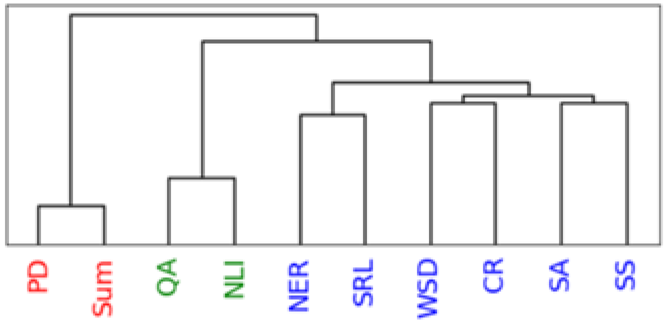
Toneva, Mariya, Otilia Stretcu, Barnabás Póczos, Leila Wehbe, and Tom M. Mitchell. "Modeling task effects on meaning representation in the brain via zero-shot meg prediction." Advances in Neural Information Processing Systems 33 (2020): 5284-5295.

Tasks affect processing

- Stimuli: passages and narratives
- Stimulus representation: task-optimized NLP models for a range of tasks
- Brain recording & modality: fMRI, reading & listening of different stimuli

Reading fMRI best explained by coref. resolution, NER, shallow syntax parsing
Listening fMRI best explained by paraphrasing, summarization, NLI





Oota, Subba Reddy, Jashn Arora, Veeral Agarwal, Mounika Marreddy, Manish Gupta, and Bapi Raju Surampudi. "Neural Language Taskonomy: Which NLP Tasks are the most Predictive of fMRI Brain Activity?." *arXiv preprint arXiv:2205.01404* (2022).

Challenges in using DL for cognitive science

- Not designed to specifically model brain processing
  - Training DL models using brain recordings
  - Task-based modeling
- Can be difficult to interpret due to multiple sources of information
  - **Disentangling contributions of different info sources to brain predictions**

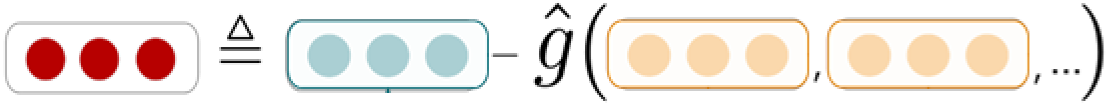
Disentangling contributions of different info sources to brain predictions

“Mary finished the apple”

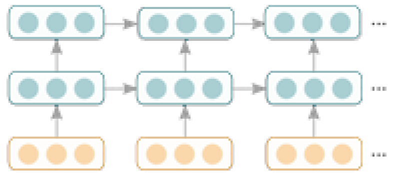
supra-word meaning may contain concept of:

- eating
- apple core
- …

Isolating supra-word meaning is a type of intervention



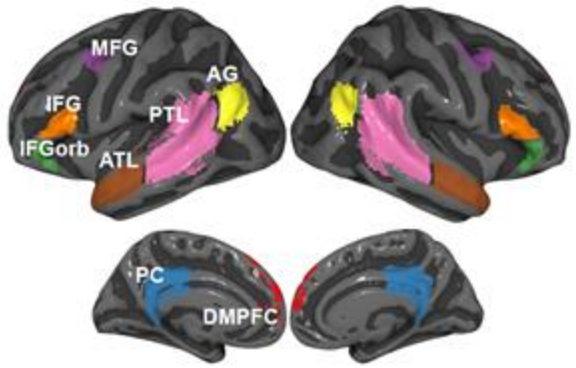
supra-word meaning



Toneva, Mariya, Tom M. Mitchell, and Leila Wehbe. "Combining computational controls with natural text reveals new aspects of meaning composition." BioRxiv (2020).

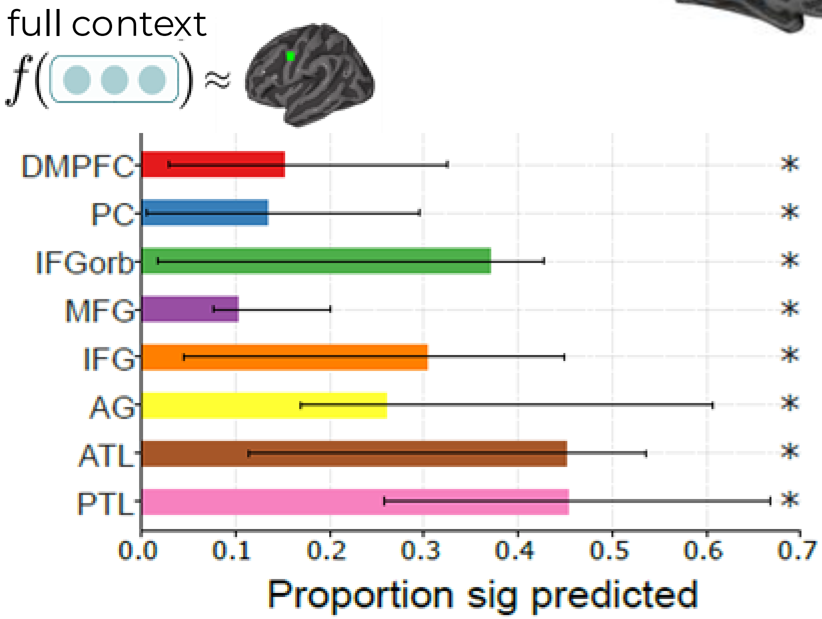
Disentangling contributions of different info sources to brain predictions

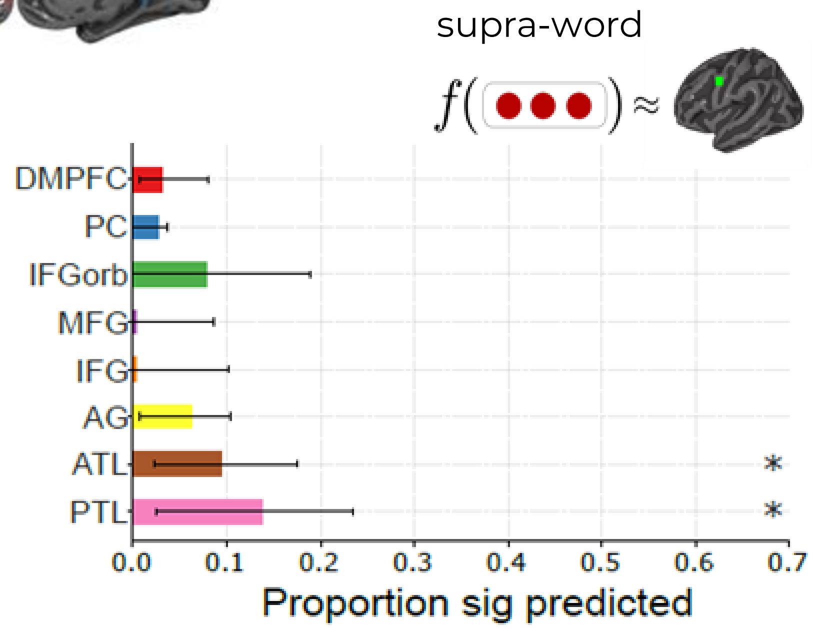
- Stimuli: one chapter of Harry Potter
- Stimulus representation: disentangled embeddings from pretrained NLP models
- Brain recording & modality: fMRI & MEG, reading



Bilateral PTL and ATL process supra-word meaning

Word-level information important for prediction of most language regions

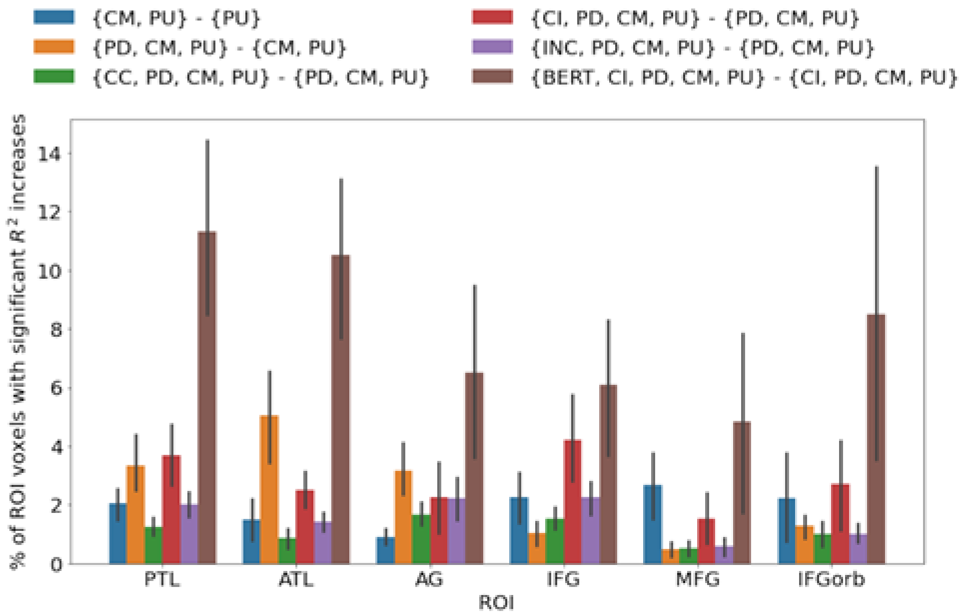


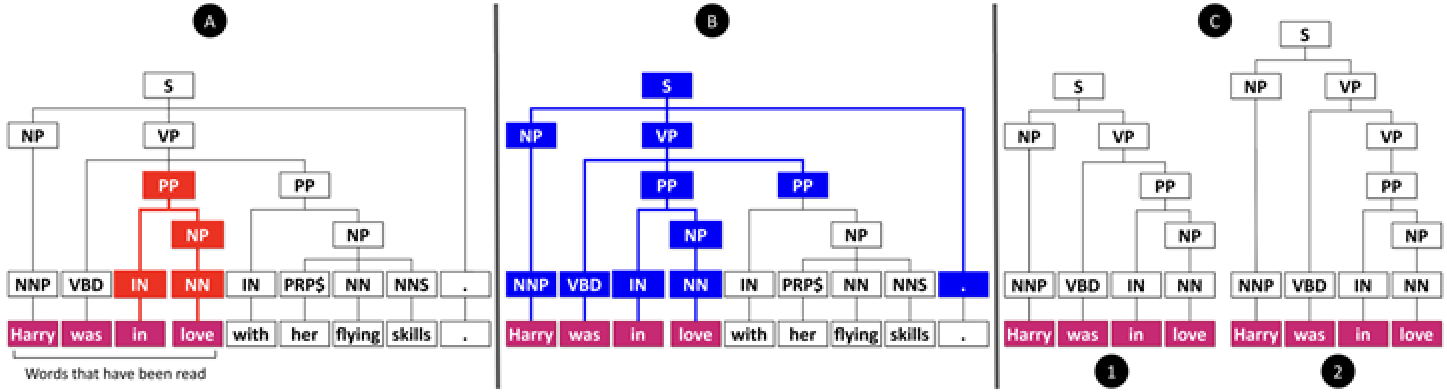


[Toneva, Mariya, Tom M. Mitchell, and Leila Wehbe. "Combining computational controls with natural text reveals new aspects of meaning composition." BioRxiv (2020).]

Disentangling contributions of different info sources to brain predictions

- Stimuli: one chapter of Harry Potter
- Stimulus representation: syntactic tree representations & pretrained NLP model
- Brain recording & modality: fMRI, reading





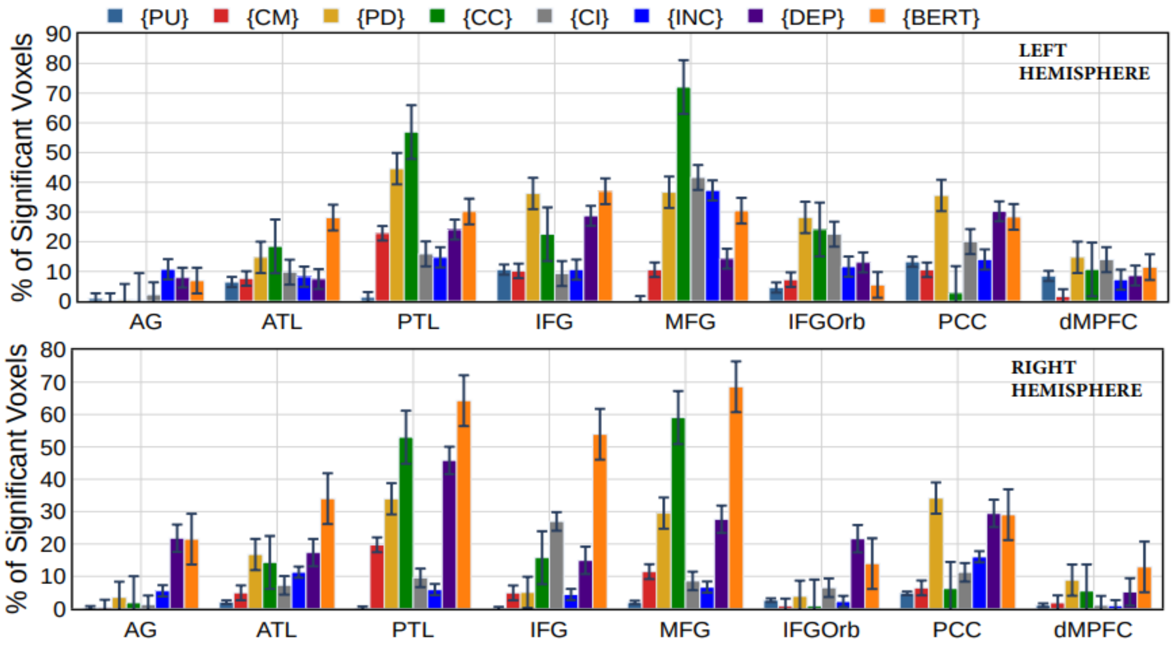
Syntactic structure-based features explain additional variance in language regions over complexity metrics

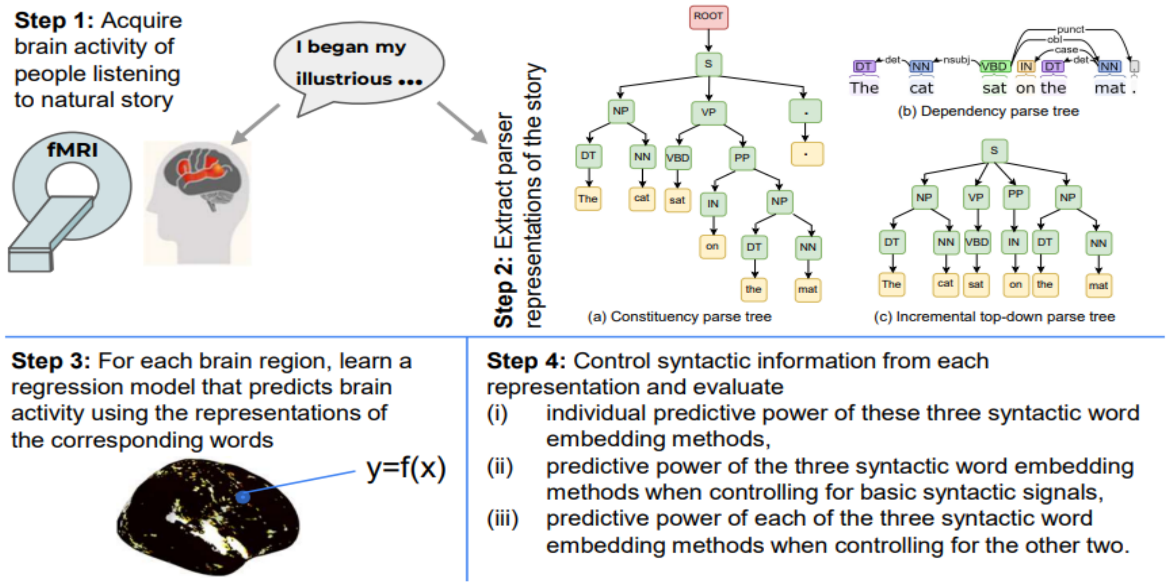
Regions predicted by syntactic and semantic are difficult to distinguish

Reddy, Aniketh Janardhan, and Leila Wehbe. "Can fMRI reveal the representation of syntactic structure in the brain?." Advances in Neural Information Processing Systems 34 (2021): 9843-9856.

Disentangling contributions of different info sources to brain predictions

- Stimuli: Narratives
- Stimulus representation: syntactic tree representations & pretrained NLP model
- Brain recording & modality: fMRI, listening





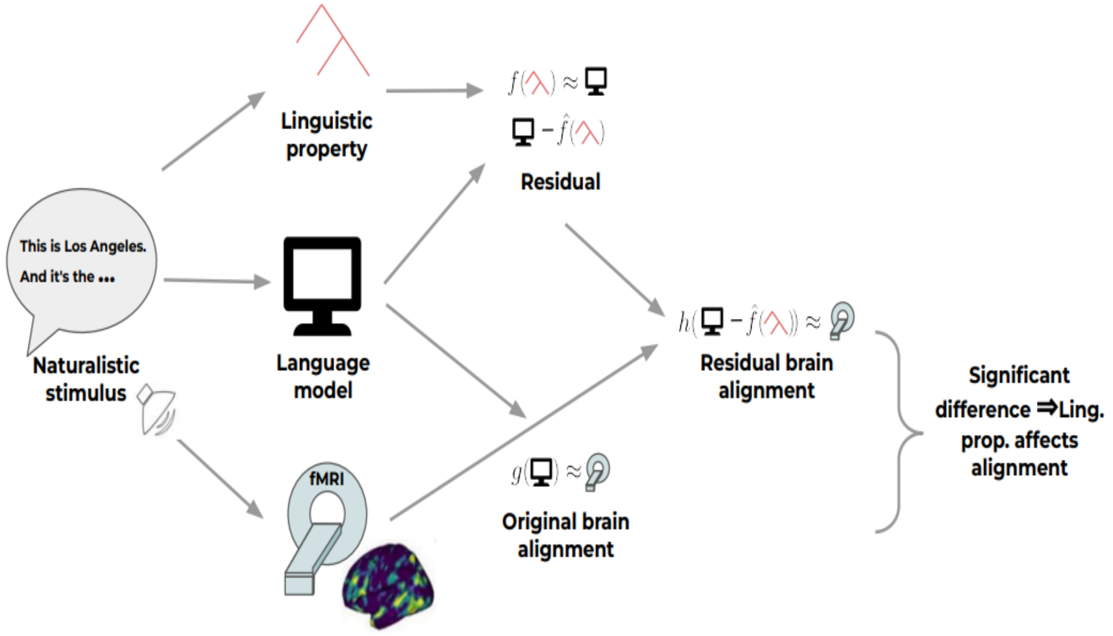
Constituency tree structure is better in temporal cortex and MFG, while Dependency structure is better in AG and PCC,

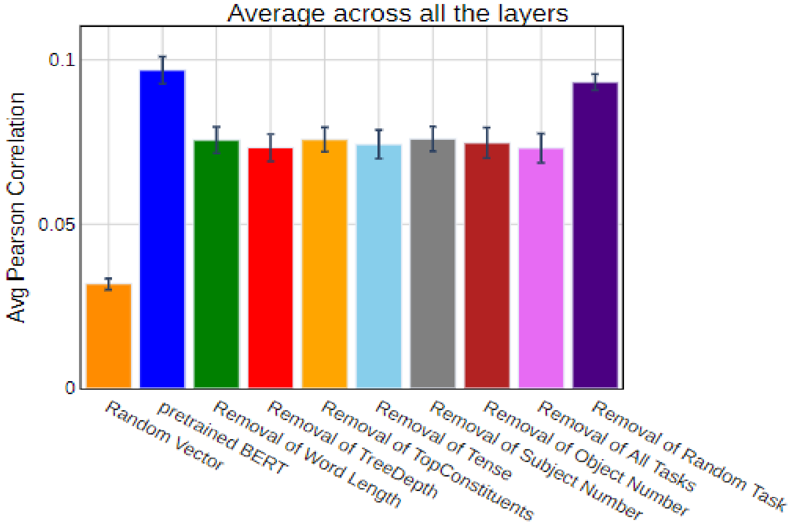
Regions predicted by syntactic and semantic are difficult to distinguish

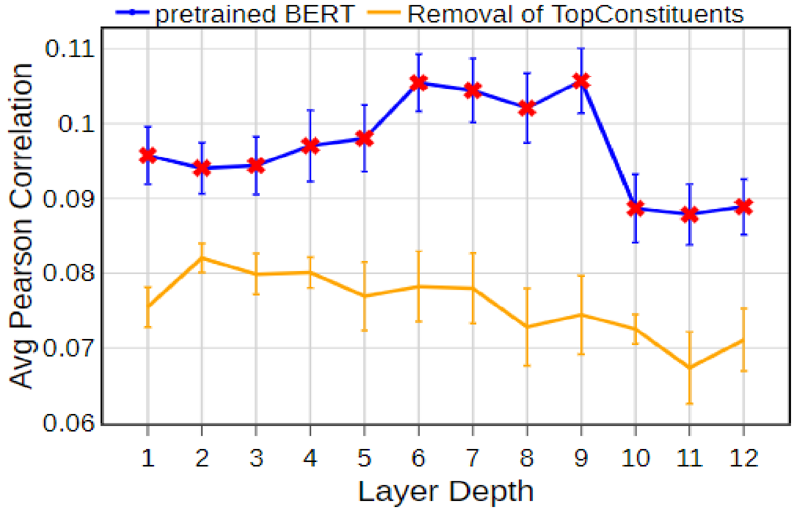
Oota, Subba Reddy et al. 2022 "How distinct are Syntactic and Semantic Representations in the Brain During Sentence Comprehension?" SNL 2022

Disentangling contributions of different info sources to brain predictions

- Stimuli: Narrative Stories
- Stimulus representation: pretrained NLP model and removal of linguistic properties
- Brain recording & modality: fMRI, Listening
- Questions: What linguistic properties underlie brain alignment, across all layers but also specifically in middle layers?





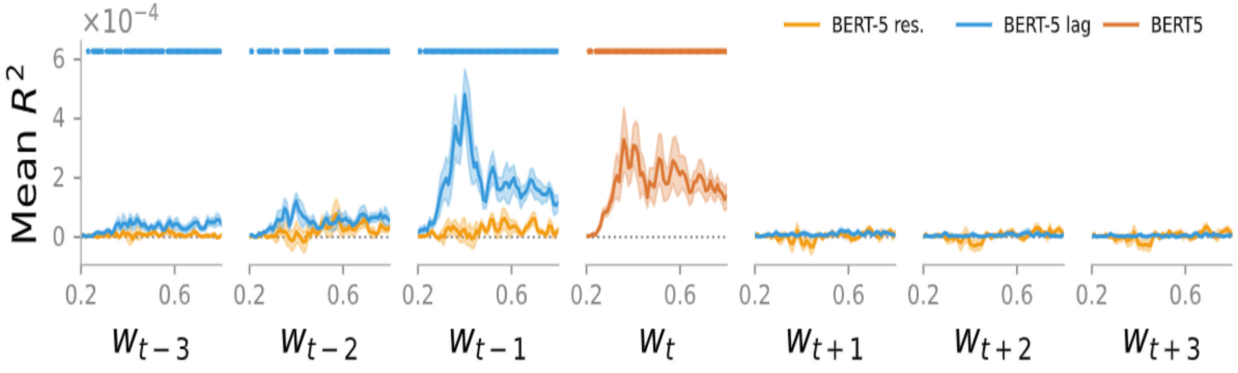


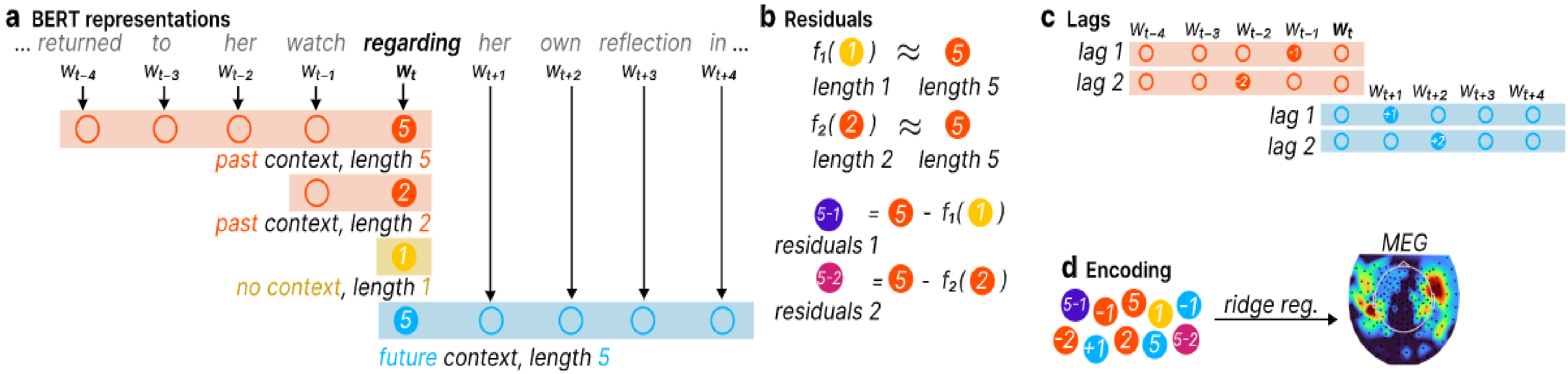
Top constituents and Tree Depth contribute the most to the alignment trend across layers

Oota, Subba Reddy, Gupta, Manish and Toneva, Mariya. "Joint processing of linguistic properties in brains and language models" 2022 arXiv.

Disentangling contributions of different info sources to brain predictions

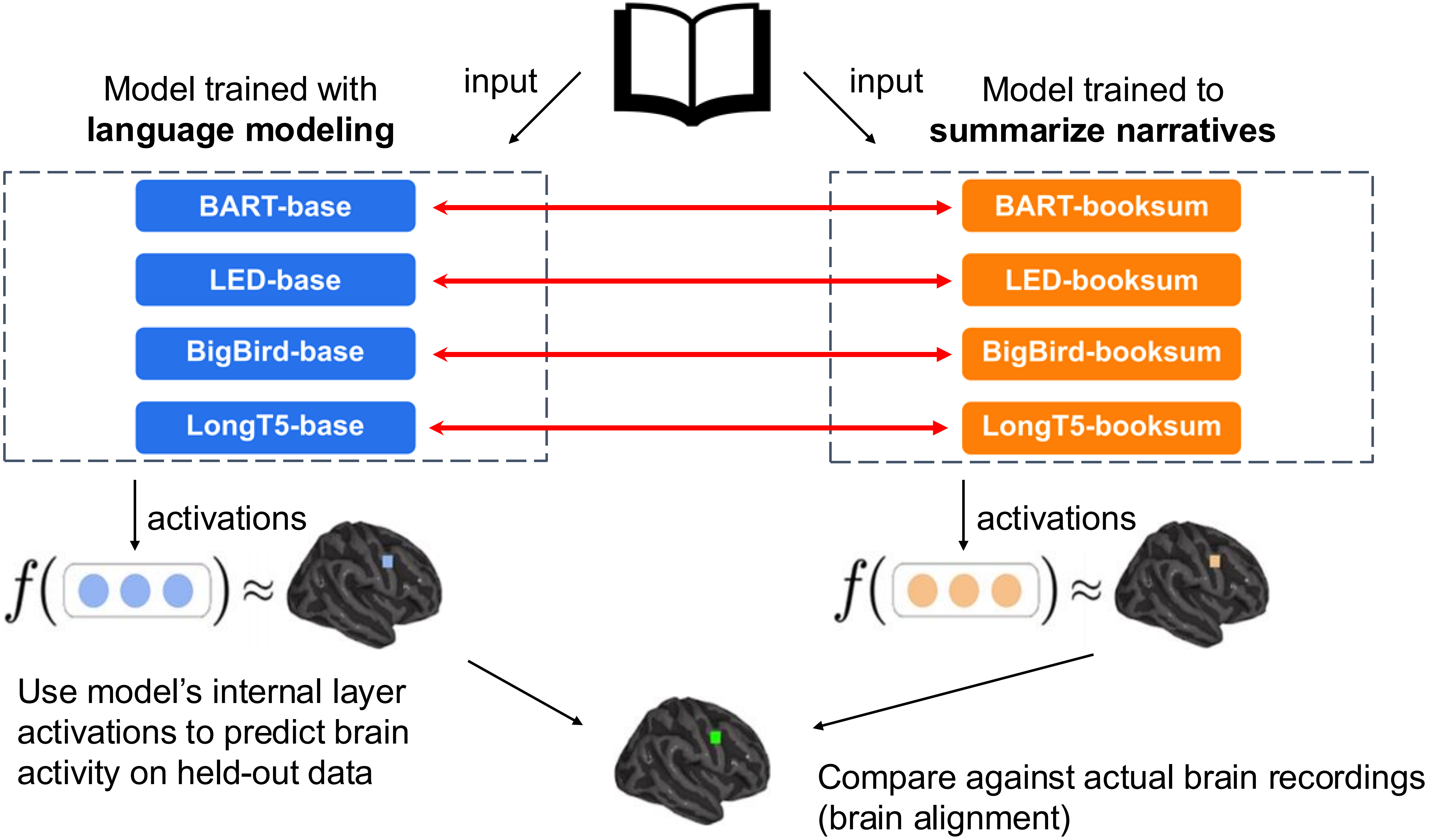
- Stimuli: four naturalistic stories
- Stimulus representation: basic syntactic tree representations & pretrained NLP model
- Brain recording & modality: MEG, Listening





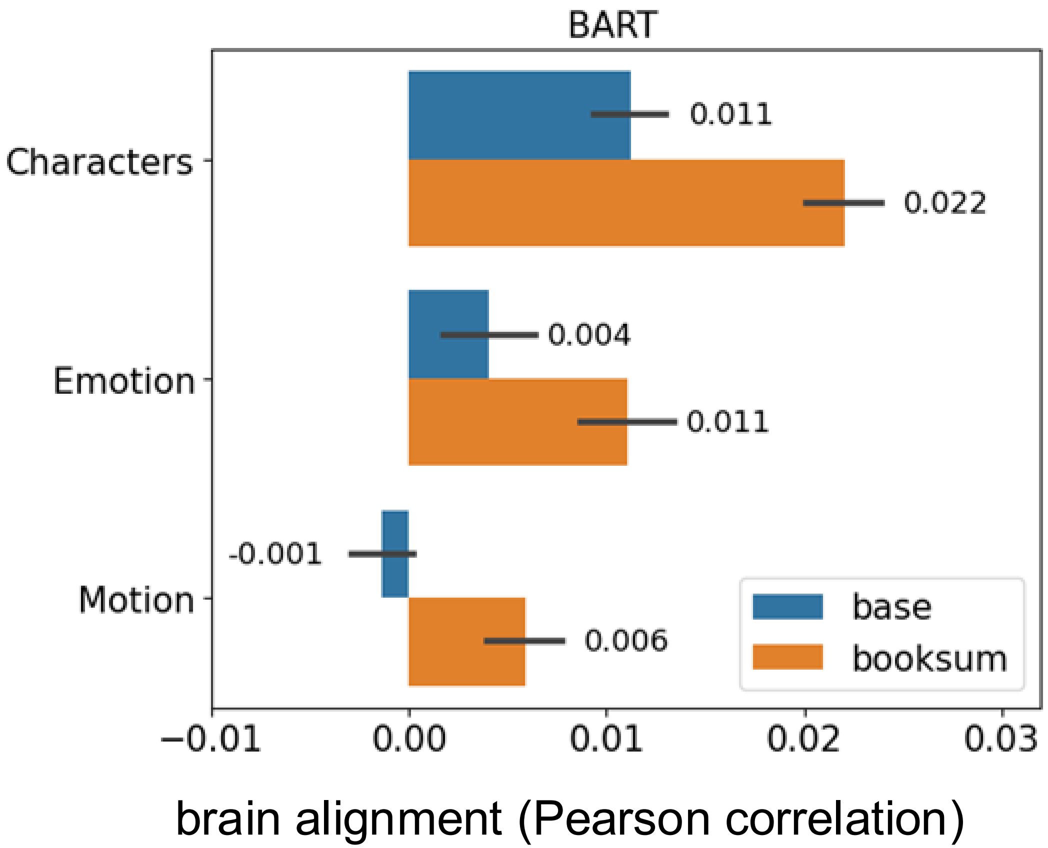
Past word context is crucial in obtaining significant results.

Oota, Subba Reddy et al. 2023. "MEG Encoding using Word Context Semantics in Listening Stories."



Model trained with
language modeling
input
input
Model trained to
summarize narratives
BART-base
LED-base
BigBird-base
LongT5-base
BART-booksum
LED-booksum
BigBird-booksum
LongT5-booksum
activations
activations
$f(\ \cdots\ ) \approx$
$f(\ \cdots\ ) \approx$
Use model's internal layer activations to predict brain activity on held-out data
Compare against actual brain recordings (brain alignment)

Result: Brain alignment improves for all discourse features



Booksum models' representations of Characters, Emotions and Motions are more aligned to the brain than the base models' representations.

- Stimuli: Narrative Stories
- Stimulus representation: pretrained NLP model and speech models
- Brain recording & modality: fMRI, Reading, Listening
- **Questions:** Is the choice of stimulus modality (reading vs. listening) important for the study of brain alignment?
- Are all naturalistic fMRI datasets equally good for brain encoding?
- How does the type of model (text vs. speech and encoder vs. decoder) affect the resulting alignment?

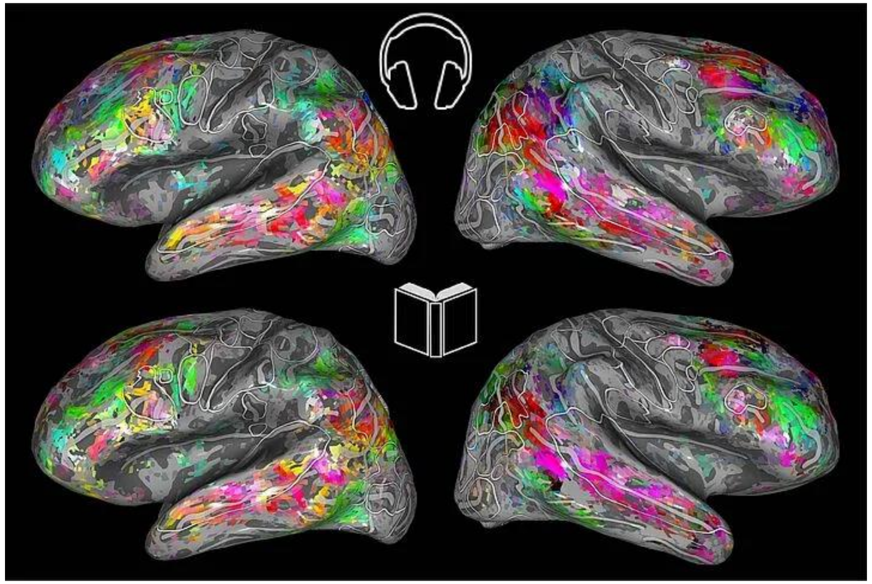


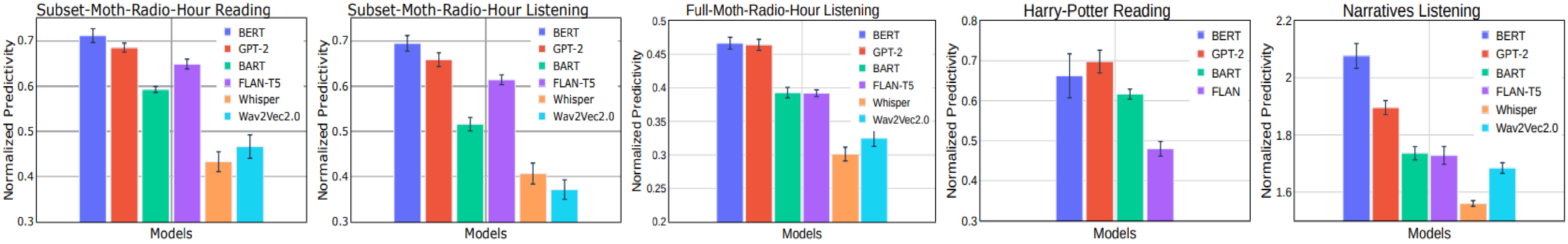
Table 1: Naturalistic Stories Datasets

Dataset	Modality	Subj	1-TR	# TRs
Full-Moth-Radio-Hour	Listening	8	2.0045s	9932
Subset-Moth-Radio-Hour	Reading	6	2.0045s	4028
Subset-Moth-Radio-Hour	Listening	6	2.0045s	4028
Narratives (21st-Year)	Listening	18	1.5s	2250
Harry-Potter	Reading	8	2s	1211

Table 2: Neural Pretrained Transformer Models

Model Name	Pretraining	Type	Layers
BERT-base-uncased	Text	Encoder (Bidirectional)	12
GPT2-Small	Text	Decoder (Unidirectional)	12
BART-base	Text	Encoder-Decoder	12
FLAN-T5-base	Text	Encoder-Decoder	24
Wav2Vec2.0-base	Speech	Encoder	12
Whisper-small	Speech	Encoder-Decoder	24

Text models predict fMRI recordings significantly better than speech models



Oota, Subba Reddy, and Toneva, Mariya. "What aspects of NLP models and brain datasets affect brain-NLP alignment?" 2023 arXiv.

A big thank you!

Tutorial, Code and Material:

Deep Learning for Brain Encoding and Decoding, Cogsci-2022

https://tinyurl.com/DL4Brain

Upcoming Tutorials:

- Deep Neural Networks and Brain Alignment: Brain Encoding and Decoding, IJCAI-2023 (A* conference)