SEMANTIC REPRESENTATIONS WITH VARYING CONTEXT DURING LANGUAGE COMPREHENSION

Anuja Negi[1,2,*], Xue Gong[3], Fatma Deniz[1,2]

1. Technische Universität Berlin, Berlin, Germany; 2. Bernstein Center for Computational Neuroscience Berlin, Berlin, Germany; 3. Department of Neuroscience, UC Berkeley

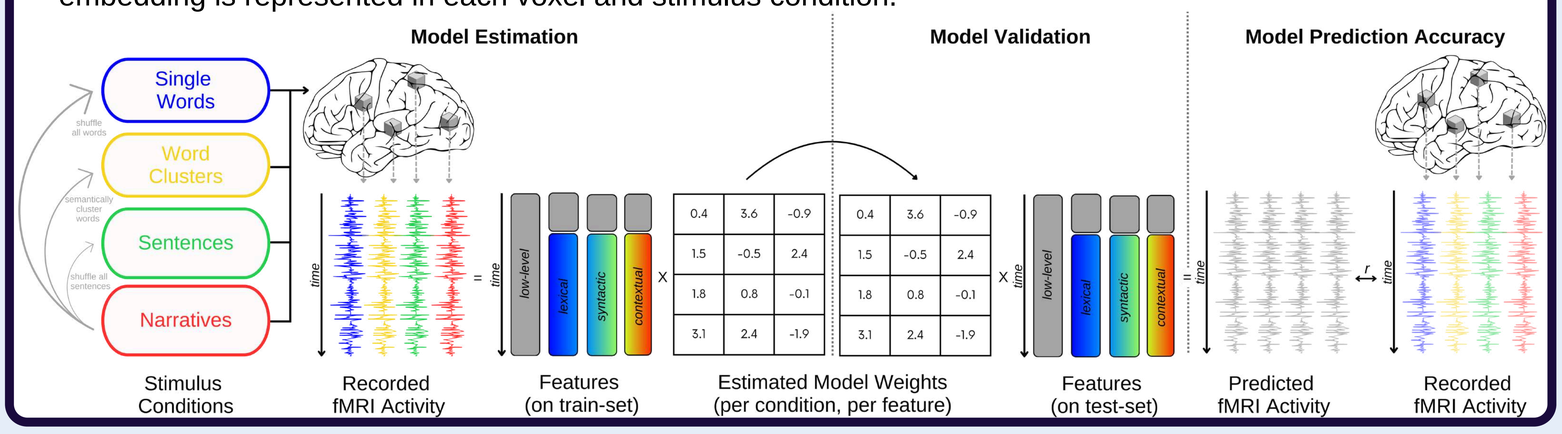
Introduction

"How does the brain comprehend language with increasing context?"

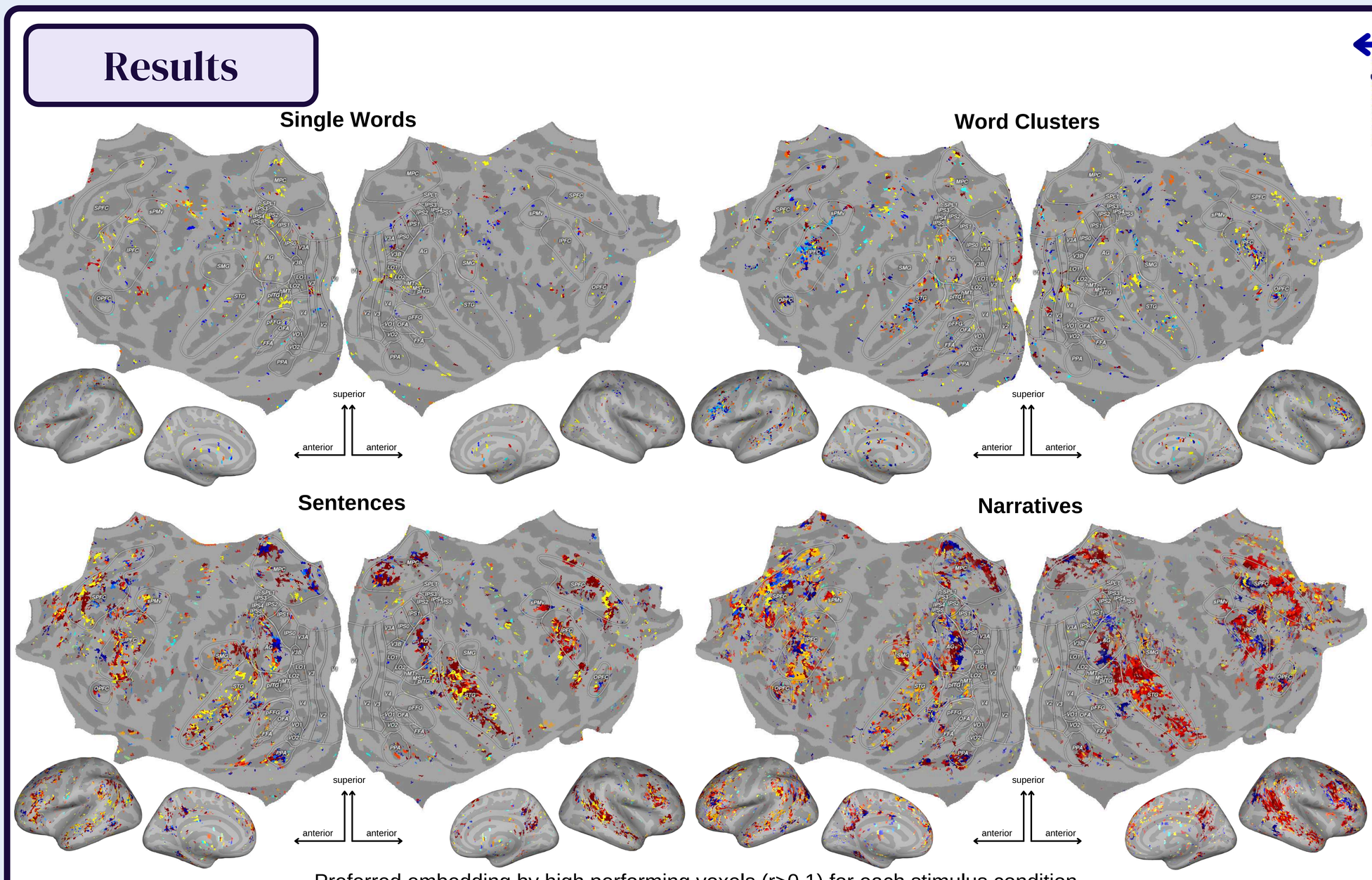
- Semantic representations are affected by the amount of context.
- In previous work, lexical embeddings were used to show an increased representation of semantic information with increasing context [1]. In addition, contextual embeddings from transformer models have been shown to be better predictors of human language processing [2].
- In this work, voxelwise encoding models based on low-level linguistic, syntactic, lexical, and contextual embeddings are compared across four stimulus conditions with varying contexts: single words, word clusters, sentences, and narratives.

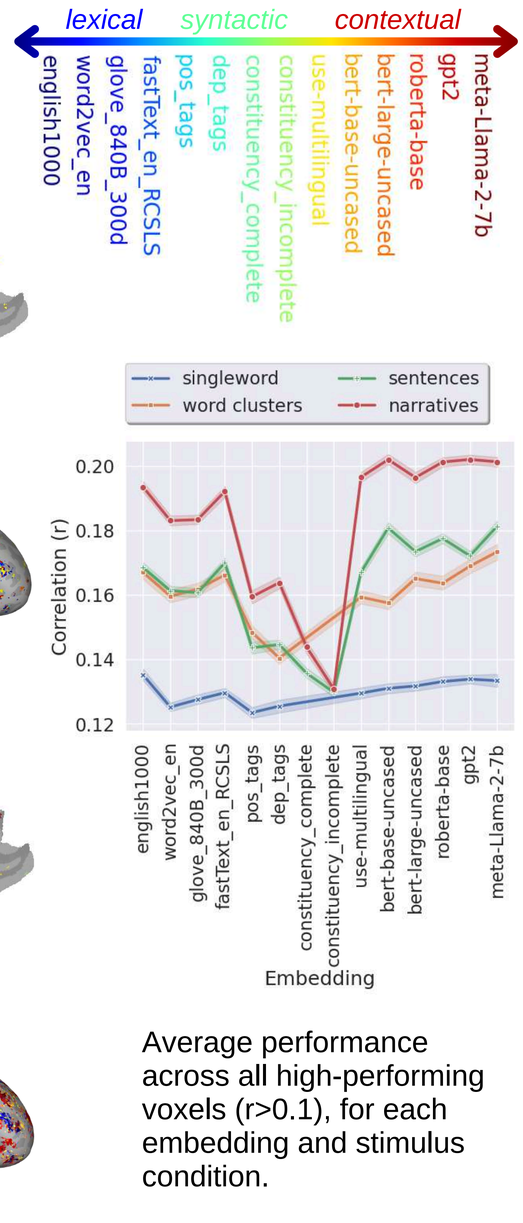
Methodology

- Stimuli for all four conditions were generated from 10 spoken stories from The Moth Radio Hour (used in [3]).
- Features are extracted from different embeddings, for the stimulus words in each condition separately;
  - Low-level linguistic embeddings: word rate, number of letters, letter count, and word length variation per TR.
  - Syntactic embeddings: PoS tag, dependency tag, constituency complete, constituency incomplete.
  - Lexical embeddings: English1000, Word2vec GloVe, and fastText.
  - Contextual embeddings: Layer-by-layer representations from LLMs such as USE, BERT, RoBERTa, GPT and Llama.
- Voxelwise encoding modeling approach with banded ridge regression [4, 5] is used to determine how each embedding is represented in each voxel and stimulus condition.



Results





Average performance across all high-performing voxels (r>0.1), for each embedding and stimulus condition.

Preferred embedding by high performing voxels (r>0.1) for each stimulus condition.

Conclusion

1. All embeddings predict brain responses more accurately when the stimulus includes more context.
2. During language comprehension, the brain prefers semantic and contextual information over syntactic structure.
3. The left hemisphere is more involved in processing lexical and short-range context, while the right hemisphere plays a larger role in long-range context processing.
4. Current LLM or embeddings do not fully capture all aspects of language comprehension in the brain, as distinct stimulus conditions and regions process different linguistic information.

References

1. Deniz, Fatma, et al. "Semantic representations during language comprehension are affected by context." Journal of Neuroscience 43.17 (2023): 3144-3158.
2. Schrimpf, Martin, et al. "The neural architecture of language: Integrative modeling converges on predictive processing." Proceedings of the National Academy of Sciences 118.45 (2021): e2105646118.
3. Huth, Alexander G., et al. "Natural speech reveals the semantic maps that tile human cerebral cortex." Nature 532.7600 (2016): 453-458.
4. La Tour, Tom Dupré, et al. "Feature-space selection with banded ridge regression." NeuroImage 264 (2022): 119728.
5. Nunez-Elizalde, Anwar O., Alexander G. Huth, and Jack L. Gallant. "Voxelwise encoding models with non-spherical multivariate normal priors." Neuroimage 197 (2019): 482-492.

1.
2.
Technische Universität Berlin
bccn berlin
Berkeley UNIVERSITY OF CALIFORNIA

***Contact:** anuja.negi@tu-berlin.de
@anujanegi_1
anujanegi.me