W2B – word to brain
Predicting Neural Activity with Language Models: Incorporating Listening Comprehension Strategies

Yoav Meiri, Refael Tikochinsky, Roi Reichart

April 11, 2023

1 Motivation and background

Language transformers, like GPT-2, have demonstrated remarkable abilities to process text, and now constitute the backbone of deep translation, summarization, and dialogue algorithms. However, whether these models encode information that relates to human comprehension remains controversial.

More accurately, it is not yet clear if SOTA transformer based LMs can be related as cognitive models. A cognitive model refers to a computational model that simulates or describes human cognitive processes involved in a particular task or domain. Cognitive models attempt to explain how humans perform various mental tasks, such as perception, attention, memory, language comprehension, decision-making, and problem-solving.

In the context of transformer-based language models, a cognitive model would involve using the language model as a computational model of human language processing, to better understand the cognitive processes involved in tasks such as language comprehension, language production, and other related tasks. This work focuses on language comprehension while listening.

To extract meaning from natural speech, the human brain combines information from each word togather with previous words, or context. Without context, humans would be unable to understand homonyms (words that have identical spelling and pronunciation, whilst maintaining different meanings), parse phrases, or resolve coreferences (when two or more expressions refer to the same person or thing) [7]. **Contextual information must thus be represented in the human cortex.**

To achieve reasonable comprehension, learners often engage in mental mechanisms, which the literature refers to as 'strategies', following the use of the term in cognitive psychology [4]. Integrating cognitive tactics into embeddings produced by LMs can help us learn about the contribution of those tactics to listening comprehension in humans [4].

2 Problem definition

The objective of this project is to explore the viability of employing the GPT-2 transformer-based language model as a cognitive model for language processing in the human brain. In case it falls short of being a suitable cognitive model, we aim to identify potential approaches to bridge the gap. To achieve this aim, functional magnetic resonance imaging (fMRI) scans will be utilized to measure neural activity in participants as they listen to stories. Subsequently, the neural activity will be compared to the word embeddings generated by GPT-2.

There are some cognitive tactics that had been previously presented as ones that take place during listening comprehension [4]. These cognitive tactics can be used as tools for **manipulating stimuli embeddings created by LMs to predict brain responses recorded with fMRI to the best of our ability.**
The main 2 cognitive tactics that had been examined in this work are summarization and language modeling. [4].

We will use representations produced by a LM to incorporate context (manipulated using the cognitive tactics above) into encoding models that predict fMRI responses to natural, narrative speech
.

To this end, we analyzed 101 subjects recorded with functional Magnetic Resonance Imaging (fMRI) while listening to 70 min of short stories [10].
We will use common methods (Pearson correlation) to try and measure the correlation between the created embedding and the matching neural activity.

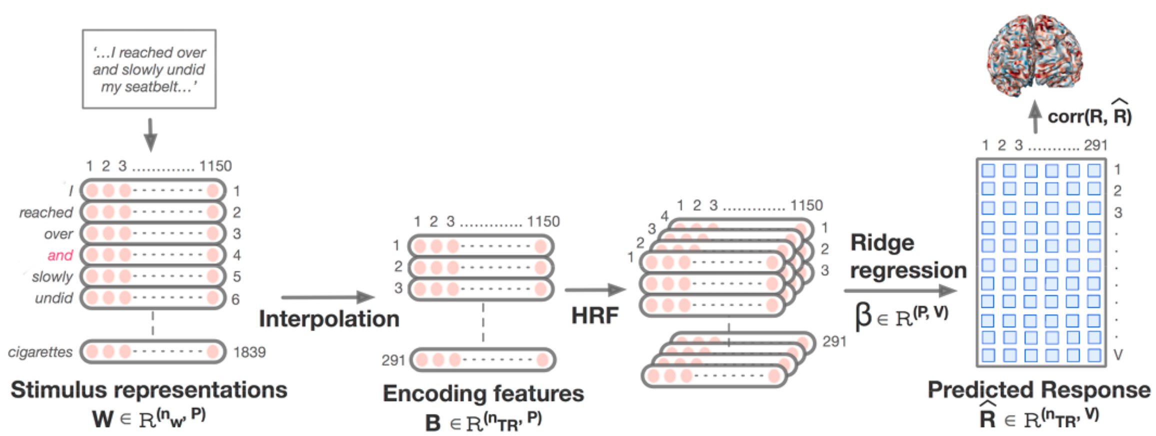
Our problem consists of two main parts:

- Building a robust baseline and a mechanism that predicts neural activity given word embeddings and then evaluates the correlation between the predicted and the real activity.
- Find a way to incorporate the previously discussed cognitive tactics into the word embeddings and compare the results to the baseline

we will conduct an evaluation of the predictive capacity of embedding vectors in relation to neural activity, utilizing ridge regression models. For each voxel, which represents an individual unit of brain activity that we are monitoring, we will employ a linear model and subsequently assess its goodness of fit through the calculation of Pearson correlation coefficients.

The findings from this work will facilitate the assessment of the impact of incorporating cognitive strategies into the language model (LM) embeddings on the correlation with the actual neural activity. A higher correlation between the predicted activity and the observed activity would suggest a greater involvement of cognitive skills in the process of listening comprehension. Additionally, we will conduct comparisons of various cognitive tactics to determine their effects on specific areas of the brain, including whether they increase or decrease the correlation with neural activity. Figure 1 presents a visual representation of the main pipeline that defines our research problem.

Figure 1: A description of the working pipeline we will be using in this project for examining the influence of each cognitive tactic [8].
This process of model building and evaluating will be done seperately for each voxel in the dataset whose activity is averaged over all participants in the dataset.
HRF (hemodynamic response function) is the function we use to vectorize the neural activity



3 Preliminary Literature Review

This section will be divided into two parts, each representing a main branch of this work. The first part will describe the cognitive and meta cognitive strategies employed by individuals to achieve comprehension. The second part will provide an overview of prior efforts to integrate natural language and cognitive elements into language model (LM) representations for the purpose of exploring cognition using fMRI data.

3.1 Listening Comprehension and Cognitive Tactics

Nowadays we know that to achieve reasonable comprehension, learners often engage in mental mechanisms, which the literature refers to as 'strategies',

Cognitive strategies identified in the literature include inferencing, elaboration, prediction, translation, contextualisation (O'Malley et al., 1989; Oxford, 1990; Young, 1997; Ross, 1997) and visualisation (DeFillipis, 1980). Metacognitive strategies identified include self-monitoring, comprehension monitoring, selective attention and self-evaluation (O'Malley and Chamot, 1990; Bacon, 1992; Young, 1997).
Cognitive tactics are used to process utterances directly by transforming them into mental representations that could be stored and recalled.
The act of listening comprehension consists of 2 types of processes: automatic and controlled. Automatic processes are cognitive processes that have been well learned, make little or no demand on processing capacity and require no attention. Controlled processes, on the other hand, are conscious. They require attention and can be used flexibly in changing circumstances.
Comprehension processes that can become automatic include word recognition and syntactic analysis. When these low-level processes become automatic, more cognitive capacity is freed for higher-level processing, such as making inferences [4].

3.2 Usage of Language Models for fMRI Decoding

Under the assumption that humans hold cortical representation of things they are exposed to, one powerful tool for mapping these representations is encoding models, which use features extracted from stimuli to predict brain responses recorded with fMRI. Previous language encoding studies have successfully mapped word-level semantic representations using embedding vectors [7].

In a previous study it has been shown that **incorporating context** into encoding models that predict fMRI responses to natural, narrative speech perform significantly better at predicting brain responses than previously published word embedding models that don't use context [7].

This example for using a cognitive tactic (contextualization) for extraction of stimuli embeddings, that improves correlation to brain responses indicates that using different cognitive tactics can help improve correlation to brain responses even more.
On top of that, there is another aspect of brain activity while listening that has been previously addressed in research; Natural language contains information at multiple timescales, ranging from phonemes to narratives. The human brain processes language using a hierarchy of representations at different timescales. Early stages represent acoustic and word information at the sub-second scales, while at later stages information is combined over many seconds to derive meaning.
The new interpolation methods of investigating timescale representations in the brain improved timescale estimates across a variety of brain regions. Given that some cognitive tactics are associated with different processing time scale (understanding a paragraph compared to understanding a single word), we may choose regions of interest in the brain which match the processing time scale of the cognitive tactic we are examining.[8]

3.3 Effectiveness of Task Specific NLP Models for fMRI Encoding

This section relates to [11]. In this work, the pipeline was almost the same as the pipeline presented in figure 1, except for the pretrained models they used (they used multiple pretrained versions of BERT-Base). In that paper, they uncovered insights about the association between fMRI voxel activations and representations of diverse NLP tasks representations. Given Transformer models finetuned for various NLP tasks, they proposed the problem of finding which of these are the most predictive of fMRI brain activity for reading and listening tasks (for the listening part, they used the same dataset we used). They found both which are the tasks that yield the most predictive models for the neural encoding task, and how correlated are embeddings created by different models. Below we present a one of their results that is related to this work:

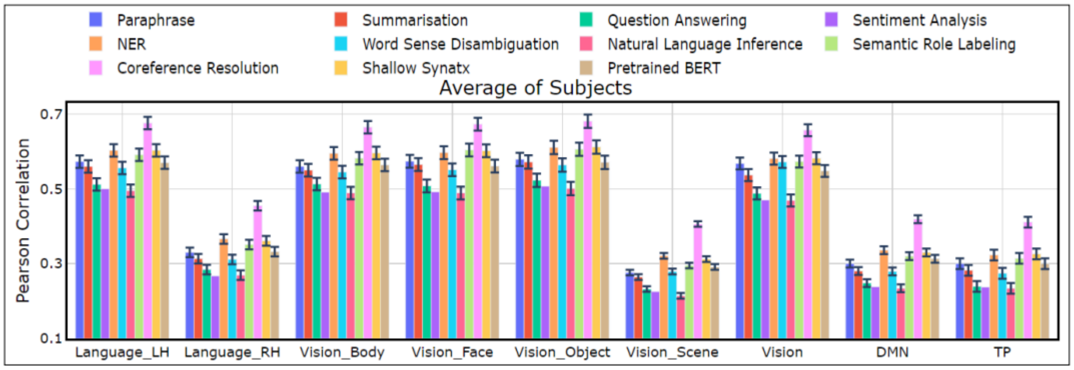


Figure 2: Pearson correlation coefficient between predicted and true responses across different brain regions using a variety of NLP tasks. Results are averaged across all participants. CR, NER, and SS perform the best.

These results show that there is indeed a difference between the different pretrained models. Under the assumption that there are no big differences in the amount and distribution of the data the multiple fine-tuned BERT-base were pretrained on, it can be concluded that incorporating the element of cognitive tactics in BERT helps neural activity prediction, and that paraphrase detection, summarization, and Natural Language Inference tasks display the best correlation during listening.

This result inspired this work to focus on summarization and language modeling as cognitive tactics that can improve neural activity prediction.

3.4 Prompting in Large Language Models

Prompting is a technique that leverages the pre-trained knowledge of large language models (LLMs) to perform various natural language processing (NLP) tasks without fine-tuning. Prompting involves constructing natural language queries or templates that elicit the desired output from the LLM, such as classification labels, entity types, or semantic relations. Prompting can be seen as a form of natural language interface that allows users to access the LLM's capabilities without requiring additional training data or specialized architectures.

One of the challenges of prompting is to design effective and robust prompts that can generalize to different domains and inputs. A promising direction is to use contextual information to guide the prompting process and improve the LLM's performance (like we will do in this work). Contextual information can include the task description, the input sentence, or external knowledge sources. Contextual information can help the LLM to disambiguate the meaning of ambiguous words, resolve coreference, and infer implicit relations and arguments.

Several studies have explored the use of contextual information for prompting in LLMs. For example, Shin et al. (2020) proposed a method that uses natural language explanations as context to improve the LLM's accuracy and interpretability for sentiment analysis. Gao et al. (2020) proposed a method that uses entity linking and knowledge graphs as context to enhance the LLM's ability to answer open-domain questions.

These studies demonstrate that prompting in LLMs can benefit from adding context to sentence analysis. Contextual information can help the LLM to better understand the input sentence and generate more accurate and relevant outputs. However, there are still many open questions and challenges in this area, such as how to automatically generate or select optimal prompts, how to incorporate multiple sources of context, and how to evaluate the quality and robustness of prompting methods.

4 Research Overview

4.1 General setup, notations and word embeddings

As previously mentioned, the experimental setup involves a participant undergoing MRI scans while listening to a story. MRI snapshots of the participant's brain activity are taken at intervals of 1.5 seconds. Additionally, the precise timing of each word heard by the participant is recorded. This allows for the reconstruction of the exact sequence of words heard by the participant between each MRI scan. The primary objective of this work is to develop a model capable of embedding the narrated text and predicting its corresponding neural activity. Here we provide the notations used in the paper for reference.

N : Number of participants

n : Number of words in the narrative

S : The narrative the participants listened to (a series of n words)

M : Transformer-based language model

$v(i)$: An embedding vector for S_i created using M

$I(i)$: The textual input to M that was used to embed S_i

$\cdot, +$: Concatenation operator

The following elements may vary between different participants:

T : Total number of fMRI scans taken at intervals of 1.5 seconds (number of TRs)

y : ($y \in R^{T \times \text{No. voxels}}$) Neural activity measured in all TRs for all voxels

$TR(t)$: The text associated with the time interval between the t-1'th and t'th MRI measurements (series of words)

$v_{TR}(t)$: An embedding vector for $TR(t)$ created using M

Initially, we assumed humans process language by concurrently considering two components: a fixed-size "window" comprised of recently heard words (considered short-term processing), and contextual information formed by priming knowledge or other cognitive processes, which may not solely rely on the aforementioned "window". For example, in the sentence:

The mitochondria is the powerhouse of the cell.

Based on our assumption and assuming a window size of $WS = 4$, if the participant has just heard the word "cell", the short-term processing component would be processing "powerhouse of the cell", while the long-term processing component would be handling prior context which considers the remainder of the sentence.

Building upon this assumption, we propose that an optimal text representation should take into account both short-term and long-term processing to best correlate with corresponding neural activity. Specifically, for the presentation of the word S_i, the plain input for M is: $I(i) = S_{i-WS} \cdot \ldots \cdot S_i$. Our baseline model considers only the short-term processing and does not take into account any further contextual information as model input.

This work presents two methods for incorporating context terms and manipulating the plain text input to incorporate both types of processing, while using the same model for representing each word. The manipulation is solely applied to the input text, and we do not alter the model used for representation (in contrast to the manipulation presented in 3.3).

4.2 Evaluation

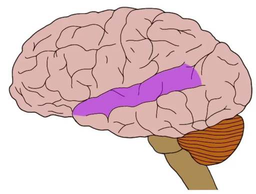
Let us examine a specific participant denoted as p within our study. Using our established model, we calculate the average embedding vector $v_{TR}(t)$ for each time point t within the range of T, by aggregating the embedding vectors of all the words in $TR(t)$. Subsequently, we employ a multi-linear ridge regression model to predict the hemodynamic response function (HRF) activity of every voxel in the brain (y), based on the embedding vector for the tth TR.

During the evaluation of the regression model, we calculate the coefficient of determination, denoted as R^2, which represents the Pearson correlation coefficient for each voxel. This coefficient elucidates the strength and direction of the correlation between the embedding vectors and the corresponding neural activity.

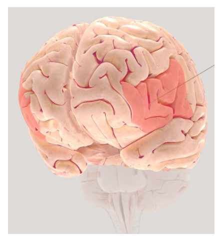
In our findings, we present the results utilizing a statistical map of the brain. Notably, our advantage lies in leveraging prior knowledge about the brain and regions of interest (ROIs) within the context of language comprehension, which enables us to better interpret the obtained outcomes.

In particular, we will evaluate the results w.r.t the R^2 scoring of ROIs relevant in the context of language comprehension and processing. The main regions of interest w.r.t this work are:

- Superior temporal gyrus (STG) - The superior temporal gyrus is an essential structure involved in auditory processing, as well as in the function of language in individuals who may have an impaired vocabulary, or are developing a sense of language. In particular we examine Wernicke's area that is an important region for the processing of speech so that it can be understood as language. [13]



- Inferior frontal gyrus (IFG) - It was shown that this area is involved in many verbal activities that include even the identification of emotional signs in speech.[3]



- Prefrontal cortex (PFC) - Various areas of the prefrontal cortex have been implicated in functions regarding speech production, language comprehension, and response planning before speaking.[2]

- Supramarginal gyrus (SMG) - Both the left and right supramarginal gyri of healthy, right-handed individuals are shown to be active when making phonological word choices. [5]

In all plots the threshold for voxel correlation values to appear is minimal absolute value of 0.25. All voxels had a correlation score associated with them, but we decided to ignore any correlation between -0.25 and 0.25. Voxels with correlations this 'weak' might not be informative enough to be considered and better be left out.

4.3 Baseline - Description and Results

As mentioned in 4.1, our way to represent the short-term processing is to include a fixed-size "window" comprised of recently heard words, according to a predetermined window side. There is no definitive answer to what is the ideal context length for listening comprehension, as it may vary depending on various factors, such as the difficulty and familiarity of the topic, the speed and clarity of speech, and the listener's prior knowledge and expectations. However, some studies have suggested that people can process about 7 ± 2 chunks of information at a time in their working memory [6], which may translate into about 20-30 words in a spoken text. This may serve as a rough estimate of the average context length that people can handle during listening comprehension, but it should not be taken as a fixed rule.

For creating $v(i)$ using the baseline, we set: $I(i) = S_{i-WS} \cdot \cdots \cdot S_i$. We tokenize $I(i)$ using GPT-2's tokenizer, feed it to M and obtain multiple vectors as the numbers of tokens in $I(i)$ (we used the vectors from the last hidden layer of GPT-2). Than we averaged all vectors to obtain a representation for S_i.

Note that we averaged the neural activity of all participants and worked with a single y matrix throughout all experiments. Averaging neural activity helps denoising the neural activity (which is very noisy by nature), and emphasizing the effect.

For fixing a suitable window size we used the baseline method (only the text window as input for each time point) with different window sizes from $WS = 0$ where only the last heard word is embedded, to $WS = 128$ where the past 128 heard words are embedded and averaged.

As can be seen in figure 3 the baseline when using $WS = 16$ demonstrates the best correlations in all previously presented ROIs. Window sizes that were bigger and smaller than 16 both didn't preserve the ROIs that were well correlated with $WS = 16$ and also didn't show any other well correlated ROIs which were not well correlated in $WS = 16$.

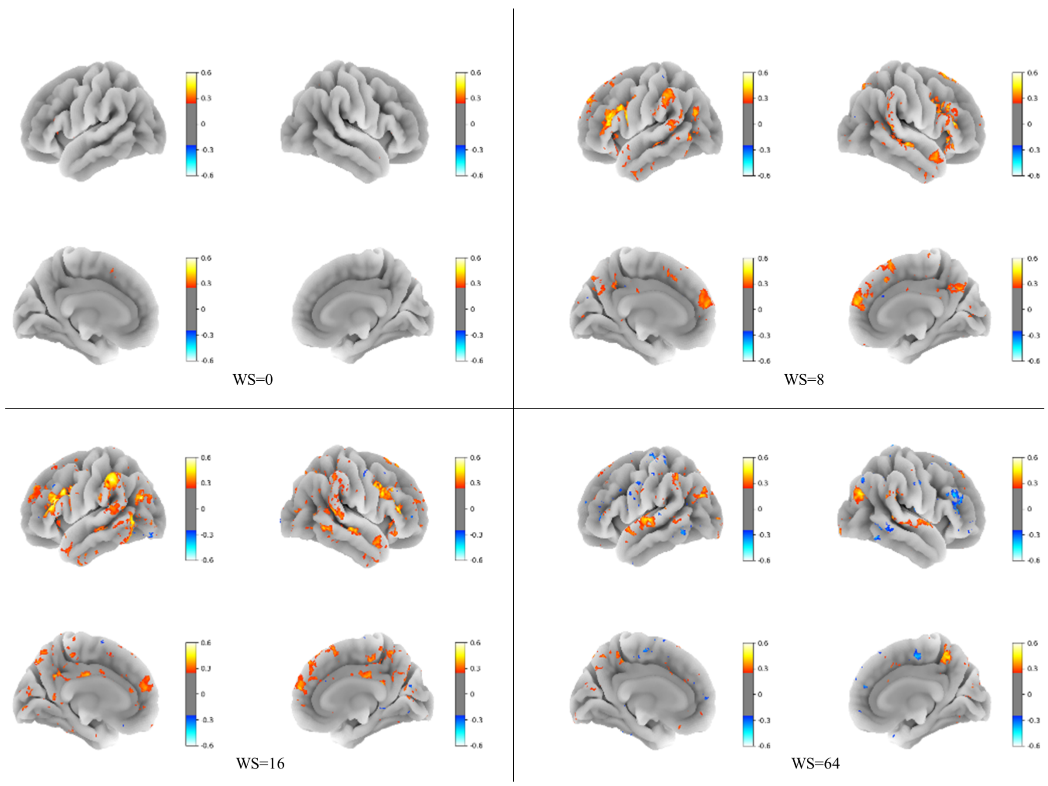


Figure 3: Pearson correlation coefficient between predicted and true responses across different brain regions using the baseline method with different window sizes (WS). Results are w.r.t the averaged activity of all participants.The threshold in each plot is absolute value of 0.25

We fixed $WS = 16$ and will maintain this value throughout all of our experiments. More specifically, in figure 4 we see remarkable correlation across the IFG (in both sides), and across the SMG (especially in the left side which is known to be more dominant). Also activations in Wernicke's area and the prefrontal cortex are well correlated to the embeddings.

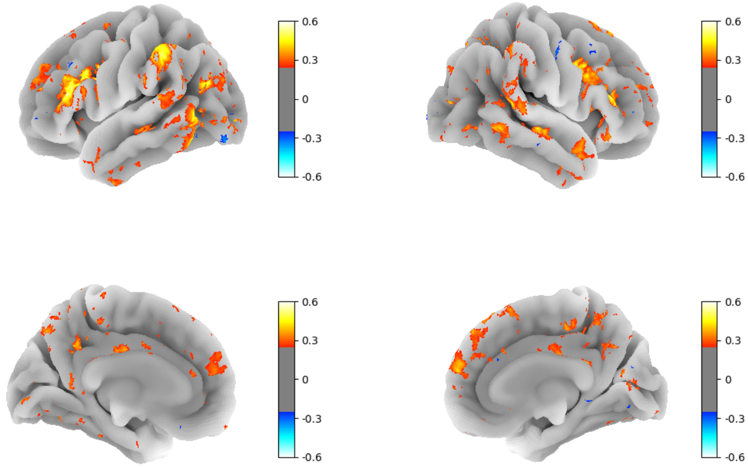


Figure 4: Pearson correlation coefficient between predicted and true responses across different brain regions using the baseline method with $WS = 16$. Results are w.r.t the averaged activity of all participants.The threshold is absolute value of 0.25

4.4 Proposed Methods - Description and Results

As previously stated, this work employs the same embedding model, GPT-2, for both the baseline and the two methods we introduce. However, the distinction between the baseline and the proposed methods lies in the model input at each time point. While the baseline solely includes the 16 most recently heard words as input at each time point, the proposed methods augment this input with prompts designed to capture more intricate and long-term cognitive processes. These additional prompts are expected to enhance the correlation with neural activity, ultimately leading to improved results.

In order to generate the prompts, our work will specifically emphasize two cognitive skills that have been previously discussed: summarization and continuation prediction (also known as language modeling), which are relevant in the field of natural language processing (NLP). We hypothesize that incorporating these cognitive skills into the process of listening comprehension can facilitate the model's ability to embed the text in a manner that exhibits stronger correlation with the corresponding neural activity.

4.4.1 Summarization

For this method we use a separate language model we denote by M_{summ}. M_{summ} is a Bart-large based model fine tuned on the BookSum dataset for the summarization task [12]. We denote the summarization result for the text S using M_{summ} by $M_{summ}(S)$.
We will define the input in each time point using recursion. First we define: $summ(0) = ""$. The model input for creating embedding of S_i for all $i > 0$:

$$I(i) = summ(i) + " : " + S[i - WS : i] \text{ [1]}$$

$$\text{Where: } summ(i) = \begin{cases} M_{summ}(summ(i-1) + " : " + S[i - summ\_WS : i]), & i(mod)15 = 0 \\ summ(i-1), & otherwise \end{cases}$$

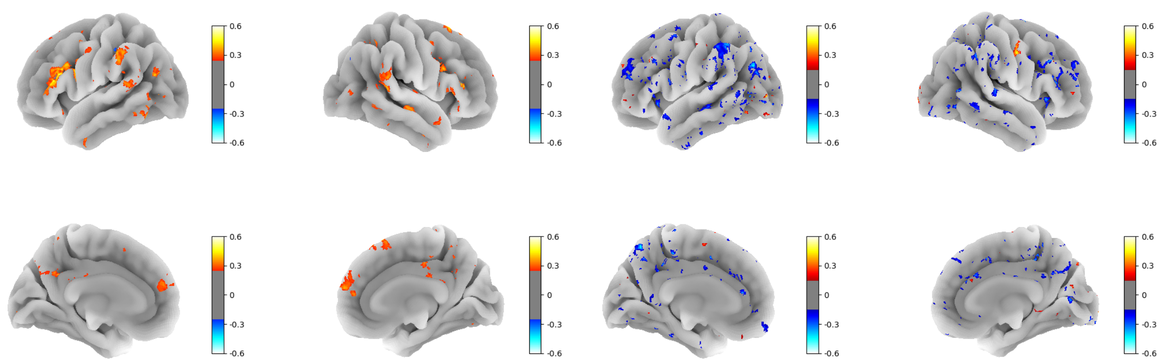
[2]

Note that the summary is updated at intervals of 15 words, which is considered reasonable based on the assumption that each update should introduce enough new information to make the new summary more informative than the previous one. Additionally, we have set the value of $summ_W S$ to 500. We have treated summarization as a cognitive skill that involves longer-term processing, and thus it is

[1] $S[i - x : i] = S_{i-x} \cdot \dots \cdot S_i$

[2] $summ\_WS$ - a predetermined window of previous words used for summarization

appropriate to incorporate long-term context when implementing it. Furthermore, it is important to highlight that each new summary is generated based on the last summary as input, which serves the purpose of cumulative summarization, aiding in the preservation of context for as long as possible and serving as an information compression technique to include as much textual information as possible within the prompt without excessively lengthening the input data.



(a) Pearson correlation coefficient using the summarization method, with threshold of absolute value of 0.25

(b) Subtraction of the correlation coefficients in the summarization and the baseline, with threshold of absolute value of 0.2

Figure 5: Using the summarization method with $WS = 16$. Results are w.r.t the averaged activity of all participants.

The results depicted in Figure 5a indicate that, for the most part, the Regions of Interest (ROIs) examined in the baseline analysis exhibit strong correlations. However, Figure 5b reveals that, in nearly all ROIs, including the IFG, STG, and SMG, the correlations are comparatively weaker in comparison to the baseline. Despite previous evidence suggesting that summarization can enhance listening comprehension quality as a cognitive skill [9], our findings demonstrate that the manner in which we incorporated summarization in our input did not contribute to improved prediction of brain activity in the context of this work.

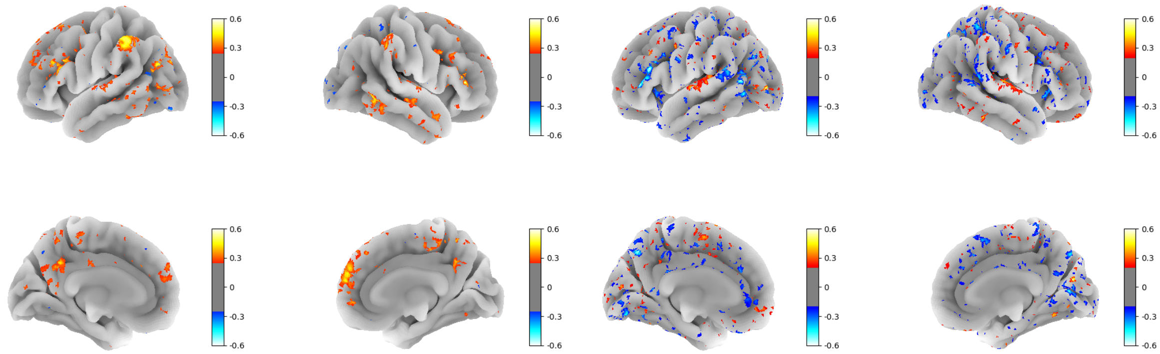
4.4.2 Next Sentence Prediction

For this method we use a separate language model we denote by M_{pred}. M_{pred} is another GPT-2. GPT-2's training task was language modeling, therefore it is suitable for the task of next sentence prediction. We denote the prediction result for the text S using M_{pred} by $M_{pred}(S)$.
The model input for creating embedding of S_i for all $i > 0$:

$$I(i) = S[i - WS : i] + M_{pred}(S[i - pred\_WS : i])$$ [3]

In contrast to the summarization approach, the prediction task necessitates updates with each new word due to the nature of the task. This is imperative in order to uphold grammatical accuracy and readability of the combined text, comprising the original input and the generated predictions. To balance the computational constraints, we determined a value of $pred_W S = 30$ for the prediction window, taking into account the limitations imposed by runtime considerations. We reasoned that this window size would be sufficiently large to yield informative and meaningful continuation predictions.

[3] $pred\_WS$ - a predetermined window of previous words used for predicting the rest of the text (between 1 and 2 next sentences)



(a) Pearson correlation coefficient using the prediction method, with threshold of absolute value of 0.25

(b) Subtraction of the correlation coefficients in the prediction method and the baseline, with threshold of absolute value of 0.2

Figure 6: Using the continuation prediction method with $WS = 16$. Results are w.r.t the averaged activity of all participants.

The findings presented in Figure 6a demonstrate that the proposed method yields substantial correlation in the SMG and the prefrontal cortex, while exhibiting a notable reduction in correlation in the inferior frontal gyrus (IFG), as evidenced by Figure 6b. This unexpected decrease in correlation in the IFG is noteworthy, given that previous research has implicated the IFG in neural mechanisms associated with predictive coding [1]. However, it is worth noting that there is a marked increase in correlation in the primary auditory cortex (PAC), which is known to play a key role in speech processing.

5 Discussion and Future Work

In conclusion, this work aims to predict brain activity using language model embeddings. We utilized a baseline approach where each word was embedded using the combined embedding of a window preceding the word. Additionally, we proposed two methods to incorporate cognitive tactics into the input in an effort to improve the correlation to brain activity. The first method involved adding a cumulative summary to the input before creating the embeddings, while the second method entailed adding a prediction of the text continuation after the original text before creating the embeddings.

Our findings revealed that both methods partially preserved the correlations observed in the baseline approach. However, despite these efforts, none of the methods resulted in a significant improvement in the correlation to neural activity compared to the baseline. Our results highlight the challenges of predicting brain activity using language model embeddings and the limitations of incorporating cognitive tactics into the input to enhance predictive accuracy.

Despite the lack of significant improvements in correlation, our study contributes to the growing body of research on utilizing language models for understanding neural mechanisms. The partial preservation of correlations suggests that there may be potential for further exploration and refinement of these methods. Future research could explore alternative cognitive tactics or modifications to the input and embedding approaches to enhance the predictive accuracy of brain activity using language model embeddings.

Moving forward, there are several potential avenues for future work. Firstly, alternative cognitive tactics could be explored to further refine the input and embedding approaches. For instance, incorporating other contextual cues or domain-specific knowledge may improve the predictive accuracy. Secondly, additional feature engineering or model architecture modifications could be explored to enhance the performance of the language model embeddings. Thirdly, investigating larger datasets or different types of brain activity data could provide further insights and potentially lead to improved correlations.

In conclusion, our findings provide valuable insights into the complexities of predicting brain activity using language model embeddings and highlight the need for continued investigation in this field.

6 Ethical Statement

We reused a publicly available dataset for this work: Narratives. We did not collect any new dataset. Narratives dataset can be dowloaded from here. Please read their terms of use for more details. We do not foresee any harmful uses of this technology.

References

[1] Linda Ficco et al. "Disentangling predictive processing in the brain: a meta-analytic study in favour of a predictive network". In: *Scientific Reports* 11.1 (Aug. 2021), p. 16258. ISSN: 2045-2322. DOI: 10.1038/s41598-021-95603-5. URL: https://doi.org/10.1038/s41598-021-95603-5.

[2] J D Gabrieli, R A Poldrack, and J E Desmond. "The role of left prefrontal cortex in language and memory". en. In: *Proc Natl Acad Sci U S A* 95.3 (Feb. 1998), pp. 906–913.

[3] Morton Ann Gernsbacher and Michael P Kaschak. "Neuroimaging studies of language production and comprehension". en. In: *Annu Rev Psychol* 54 (June 2002), pp. 91–114.

[4] Christine C.M Goh. "Exploring listening comprehension tactics and their interaction patterns". In: *System* 30.2 (2002), pp. 185–206. ISSN: 0346-251X. DOI: https://doi.org/10.1016/S0346-251X(02)00004-0. URL: https://www.sciencedirect.com/science/article/pii/S0346251X02000040.

[5] Gesa Hartwigsen et al. "Phonological decisions require both the left and right supramarginal gyri". en. In: *Proc Natl Acad Sci U S A* 107.38 (Aug. 2010), pp. 16494–16499.

[6] C. Izawa and N. Ohta. "Human Learning and Memory: Advances in Theory and Applications: The 4th Tsukuba International Conference on Memory". In: Taylor & Francis, 2014, pp. 158–160. ISBN: 9781135617844. URL: https://books.google.co.il/books?id=GMB5AgAAQBAJ.

[7] Shailee Jain and Alexander Huth. "Incorporating Context into Language Encoding Models for fMRI". In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio et al. Vol. 31. Curran Associates, Inc., 2018. URL: https://proceedings.neurips.cc/paper/2018/file/f471223d1a1614b58a7dc45c9d01df19-Paper.pdf.

[8] Shailee Jain et al. "Interpretable multi-timescale models for predicting fMRI responses to continuous natural speech". In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., 2020, pp. 13738–13749. URL: https://proceedings.neurips.cc/paper/2020/file/9e9a30b74c49d07d8150c8c83b1ccf07-Paper.pdf.

[9] Iraj Khoshnevis and Sorour Parvinnejad. "The Effect of Text Summarization as a Cognitive Strategy on the Achievement of Male and Female Language Learners' Reading Comprehension". In: *International Journal of Learning and Development* 5 (Oct. 2015). DOI: 10.5296/ijld.v5i3.8271.

[10] Samuel Nastase et al. "The "Narratives" fMRI dataset for evaluating models of naturalistic language comprehension". In: *Scientific Data* 8 (Sept. 2021), p. 250. DOI: 10.1038/s41597-021-01033-3.

[11] Subba Reddy Oota et al. "Neural Language Taskonomy: Which NLP Tasks are the most Predictive of fMRI Brain Activity?" In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, 2022. DOI: 10.18653/v1/2022.naacl-main.235. URL: https://doi.org/10.18653%5C%2Fv1%5C%2F2022.naacl-main.235.

[12] Peter Szemraj. *led-large-book-summary (Revision 38be53c)*. 2022. DOI: 10.57967/hf/0101. URL: https://huggingface.co/pszemraj/led-large-book-summary.

[13] Marc Vander Ghinst et al. "Left Superior Temporal Gyrus Is Coupled to Attended Speech in a Cocktail-Party Auditory Scene". en. In: *J Neurosci* 36.5 (Feb. 2016), pp. 1596–1606.