

# Reading times and temporo-parietal BOLD activity encode the semantic hierarchy of language prediction

Lea-Maria Schmitt<sup>1</sup> (l.schmitt@uni-luebeck.de)

Julia Erb<sup>1</sup> (julia.erb@uni-luebeck.de)

Sarah Tune<sup>1</sup> (sarah.tune@uni-luebeck.de)

Anna Rysop<sup>2</sup> (rysop@cbs.mpg.de)

Gesa Hartwigsen<sup>2</sup> (hartwigsen@cbs.mpg.de)

Jonas Obleser<sup>1</sup> (jonas.obleser@uni-luebeck.de)

<sup>1</sup>Department of Psychology, University of Lübeck, Maria-Goeppert-Straße 9 A  
Lübeck, Germany

<sup>2</sup>Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1 A  
Leipzig, Germany

## Abstract:

When poor acoustics challenge speech comprehension, listeners are thought to increasingly draw on semantic context to predict upcoming speech. However, previous research focused mostly on speech material with short timescales of context (e.g., isolated sentences). In an fMRI experiment, 30 participants listened to a one-hour narrative incorporating a multitude of timescales while confronted with competing resynthesized natural sounds. We modeled semantic predictability at five timescales of increasing context length by computing the similarity between word embeddings. An encoding model revealed that short informative timescales are coupled to increased activity in the posterior portion of superior temporal gyrus, whereas long informative timescales are coupled to increased activity in parietal regions like the angular gyrus. In a second experiment, we probed the behavioral relevance of semantic timescales in language prediction: 11 participants performed a self-paced reading task on a text version of the narrative. Reading times sped up for the shortest informative timescale, but also tended to speed up for the longest informative timescales. Our results suggest that short-term dependencies as well as the gist of a story drive behavioral processing fluency and engage a temporo-parietal processing hierarchy.

a slip of the tongue) or relevant (e.g., after a change of subject). Here, we ask how the human brain orchestrates the multitude of semantic timescales underlying natural speech to build up predictions on upcoming speech. Hasson, Chen & Honey (2015) found that larger timescales of speech like paragraphs are processed in higher cortical areas but short timescales like words in lower cortical areas. As the predictive coding model of Friston (2009) proposes that predictions are fed back from higher to lower cortical areas, we hypothesized that listeners exploit the semantic timescales of speech to inform speech prediction. On the neural level, we expected that the timescales of semantic prediction are organized along an auditory dorsal processing hierarchy. Additionally, we expected predictive timescales to propel behavioral ease of language processing.

The present study investigated the neural underpinnings of semantic prediction in an fMRI listening task and the behavioral relevance of different semantic timescales in a self-paced reading task.

**Keywords:** semantic prediction; neural gradient; fMRI

## Methods

### Introduction

In natural speech processing, rich semantic context guides a listener's expectation on upcoming speech. At the same time, not all context is meaningful (e.g., after

### Participants

Sixty-three participants (18–78 years) took part in the fMRI listening study. Here, we analyzed a subset of 30 younger participants (16 female, 18–31 years). Another



11 participants (9 female, 20–27 years) took part in the self-paced reading study. All participants were right-handed healthy German native speakers.

## Materials and Design

Our stimulus set consisted of a spoken and a written version of the same narrative: We split a one-hour audio recording of a natural narrative told by a female German speaker in eight blocks and converted the spoken narrative to text. The narrative contained 9,446 words (4,451 content words).

In the fMRI study, participants listened to all eight blocks of the narrative embedded in a competing stream (SNR 0 dB) of resynthesized natural sounds. In the self-paced reading study, participants read the narrative word-by-word on a noncumulative, left-aligned display and pressed a button to reveal the next word. Participants were presented with four randomly chosen blocks. In both experiments, each block was followed by three multiple-choice questions on the plot of the story.

## Data Acquisition and Analysis

**Operationalization of semantic predictability.** We used an established stimulus set manipulating the semantic predictability of sentence-final keywords (Erb et al., 2012) to validate that the similarity of word embeddings (i.e., vector representations of meaning; Mikolov et al., 2013) is a measure of semantic predictability. Keywords of predictable sentences correlated stronger with the average vector of all preceding words when compared to unpredictable keywords ( $p < 0.0001$ ; Figure 1).

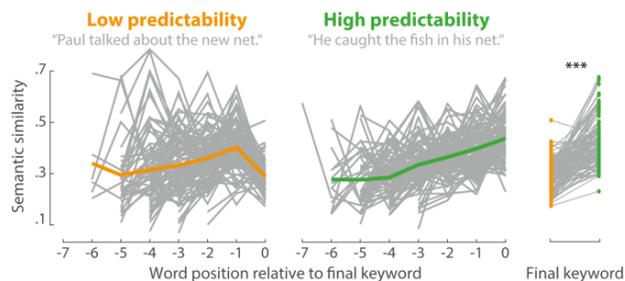


Figure 1: Correlations between average context and keyword embeddings differentiate predictable (green) from unpredictable sentences (orange); \*\*\*  $p < 0.0001$ .

We modeled semantic predictability at five timescales corresponding to a logarithmic increase in context length (i.e., 1–24 words) by computing the similarity

between the embedding of each content word in the story and each timescale's average word embedding.

**fMRI listening study.** During the listening task, we acquired continuous whole brain 3-T fMRI data (2.5 mm isotropic voxels, TR = 947 ms, TE = 28 ms). Initially, we calculated the intersubject correlation of the blood-oxygen-level dependent (BOLD) signal (Nastase et al., 2019) across the whole cortex to determine brain regions consistently engaged in speech processing. All further analyses were limited to those parcels (Glasser et al., 2016) along the temporo-parietal pathway with at least 60 % of vertices in the top 30 % of correlation coefficients. Next, we used vertex-wise ridge regression within a fourfold cross-validation scheme (6 training blocks, 2 testing blocks) to project the BOLD signal onto the semantic timescales of speech prediction. Individual best-timescale maps were derived from the encoding model, smoothed and submitted to a group-level cluster-permutation test.

**Behavioral self-paced reading study.** Like in the fMRI data analysis, we used ridge regression to project reading times onto the semantic timescales of speech prediction.

## Results

On the behavioral level, participants read faster when the shortest and—less consistently—the longest timescales were more predictive of a presented word. However, predictive power of medium timescales did not affect reading times (Figure 2).

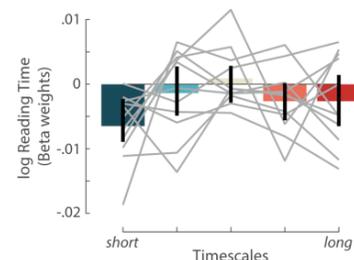


Figure 2: Beta weights represent the effect of different semantic timescales on log-transformed reading times; grey lines indicate single-participant reading time estimates, black lines indicate bootstrapped confidence intervals.

On the whole cortex, the intersubject correlation showed consistent modulation across participants for those brain areas implicated in the language network (Figure 3A). Along the auditory dorsal pathway, we found two distinct clusters ( $p < 0.0001$ ): Increased activity in the posterior portion of superior temporal

gyrus was coupled to short timescales informative of the next word, whereas parietal regions like the temporo-parietal junction and angular gyrus were most responsive to predictive long timescales (Figure 3B).

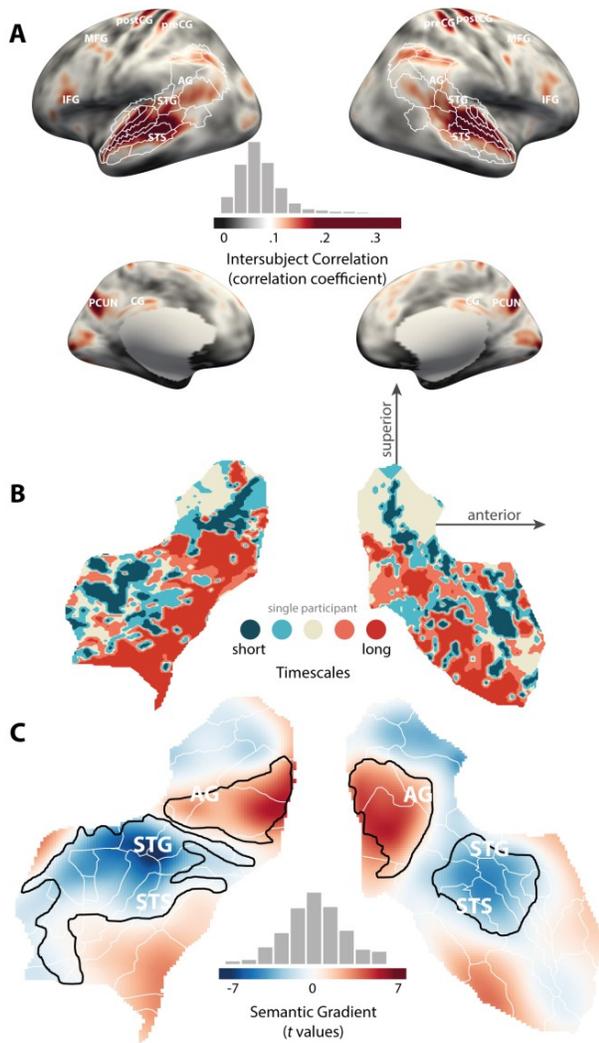


Figure 3: (A) Intersubject correlations during the fMRI listening task; white outlines indicate temporo-parietal region of interest. (B) Exemplary single-participant map of vertex-wise winning timescales. (C) Grand-average statistics calculated on best-timescale maps; blue regions indicate increased activity to informative short timescales, red regions indicate increased activity to informative long timescales, black outlines indicate significant clusters.

## Conclusions

In this study, we showed that the timescales of semantic prediction are organized along an auditory dorsal processing hierarchy; posterior temporal regions code for concrete and temporally close semantics predictive of speech, whereas parietal convergence zones code for abstract and temporally more distant dependencies (~ 0.5 s) as well as the gist of a story (~ 25 s) drive behavioral processing fluency. Next, we will use a measure of semantic predictability fine-tuned to the unique hierarchical structure underlying the context of each word by incorporating a deep neural network trained to determine the probability of an upcoming word given the semantics at a timescale.

## Acknowledgements

This research was supported by the German Research Foundation (DFG; grant 341673897 to Gesa Hartwigsen and Jonas Obleser).

## References

- Erb, J., Henry, M., Eisner, F., & Obleser, J. (2012). Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. *Neuropsychologia*, 50, 2154-2164.
- Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, 13, 293-301.
- Glasser, M. F., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536, 171-178.
- Hasson, U., Chen, J., Honey, C. J. (2015). Hierarchical process memory: memory as an integral component of information processing. *Trends in Cognitive Sciences*, 19, 304-313.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv*, 1301.3781.
- Nastase, S. A., Gazzola, V., Hasson, U., & Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *bioRxiv*.