Context-based learning of linguistic chimeras: an EEG study on the integration of novel semantic representations from minimal exposure to text

Fabio Marson, Giulia Loca, Marco Ciapparelli, Marco Marelli

Department of Psychology, University of Milano-Bicocca, Milan, Italy

Background

- Humans learn new words meaning through language even in **absence of direct experience** with the words' referent.
- Through language we can convey novel meanings and enable the **combination** of known concepts in novel semantic representations.
- **The linguistic context in which novel words are learned is thus crucial for a successful encoding of their meaning.**
- **Distributional semantic models (DSMs)** provide methods to quantify the fitting of a context in which novel words are learned. Such models have been found to predict human performances in evaluating similarity between novel concepts derived by integration of two concepts (chimeras) and similar/dissimilar concepts (Lazaridou et al., 2017).

Methods

Chimera = [Pivot concept] + [Compatible concept]

- Sentences for each concept extracted from itWac
- Each sentence was scored for its informativeness according to feature-based Italian norms (Montefinese et al., 2013)
- Labels are pseudowords adopted in the ICP (Amenta et al., 2025)

N = 54 concept
× 3 informativeness levels (low, mid, high)
× 3 condition (Congruent, Incongruent, Control conditions)

Encoding sentences **Testing sentences**

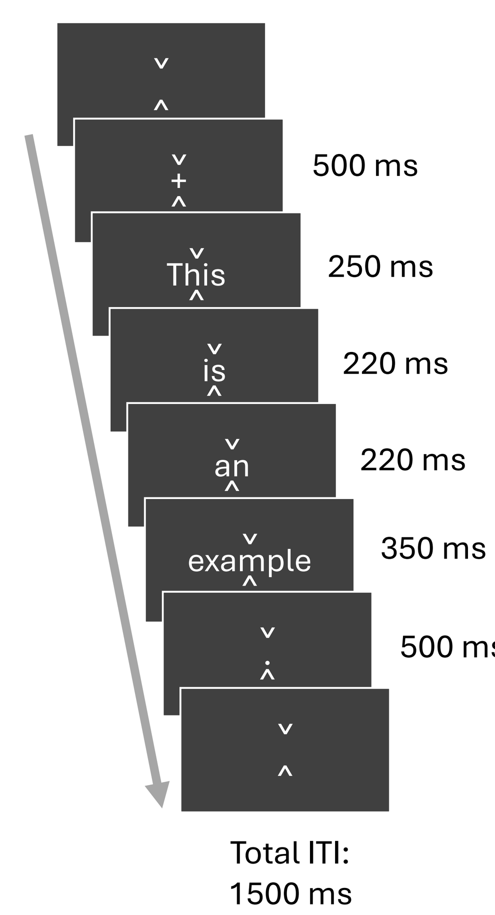
=

Final set of 972 natural sentences

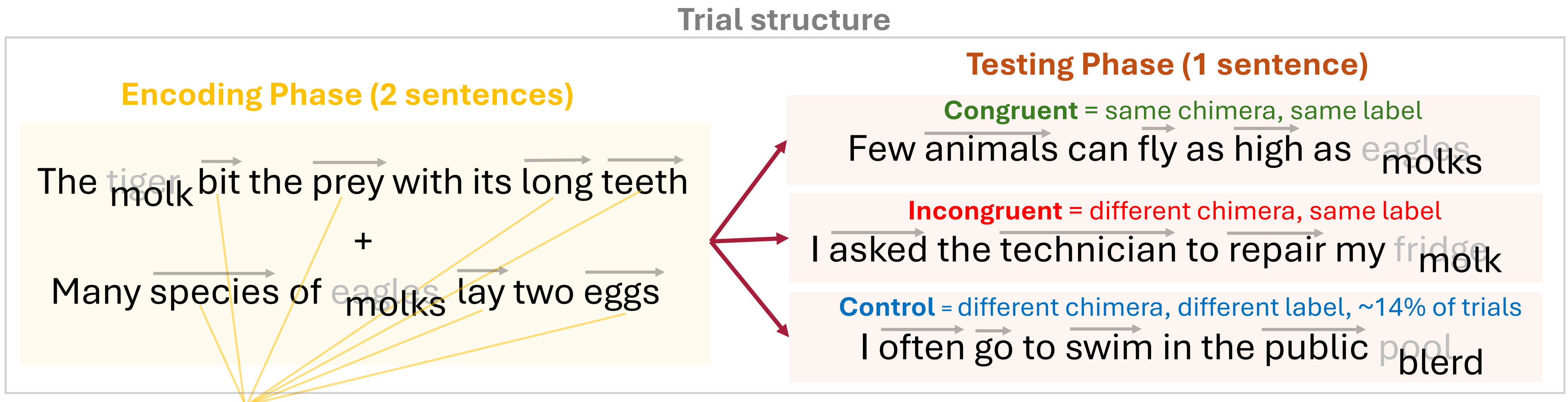
Examples of chimeras included in the current study

Pivot	Compatible	Category
Horse	Rabbit	*Animals*
Vase	Ashtray	*Furnishings*
Fork	Scissors	*Kitchenware*
Train	Truck	*Vehicles*
Shirt	Skirt	*Cloths*
Pine	Cherry tree	*Plants*

RSVP paradigm



Trial structure



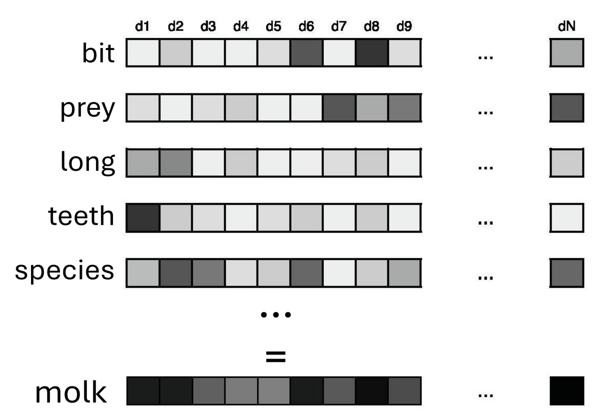
Analysis 1

Event-related potentials

Cluster-based Permutation Analysis

Cong vs **Incong**

Hypothesis: stronger N400 in Incongruent than in Congruent



Example of w2v based vectorial representation of a chimera

Computational measures:

- **Word2vec semantic similarity**: static vectorial representation of word meanings
- **Minerva surprisal**: log probability of finding a word in a given location in sentence
- **BERT cosine similarity**: context-sensitive vectorial representation of word meanings

Analysis 2

Test whether computational measures can predict ERP's amplitude (comparing fit of lmer with null model)

Hypothesis: ERPs amplitude predicted by similarities (positive slope) and surprisal (negative slope)

Results

Analysis 1

Congruent vs **Incongruent**

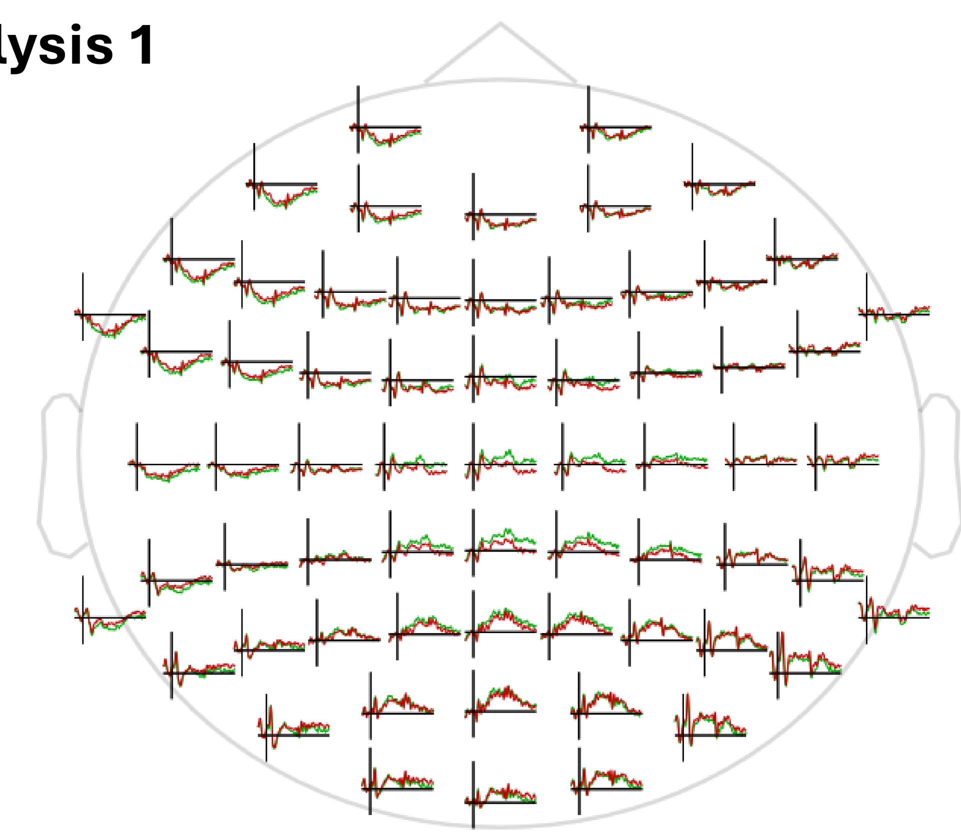
Two centroparietal clusters:

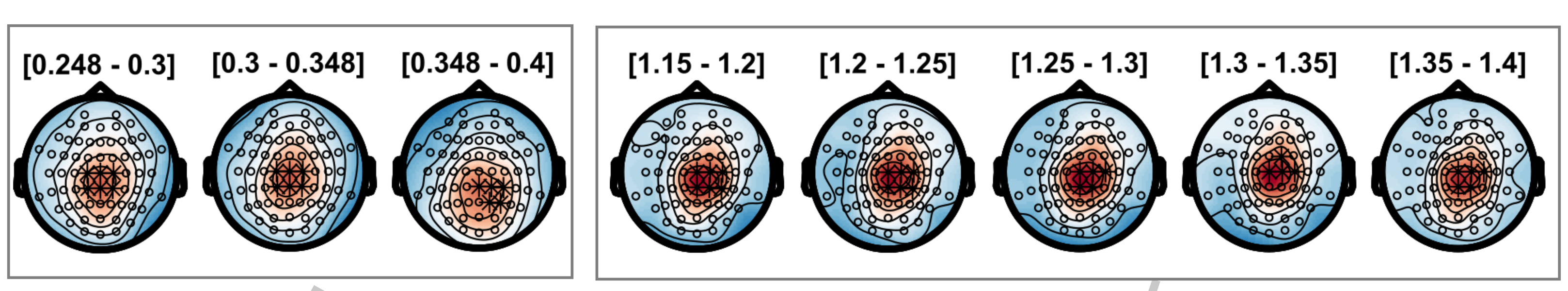
- 224 ~ 416 ms

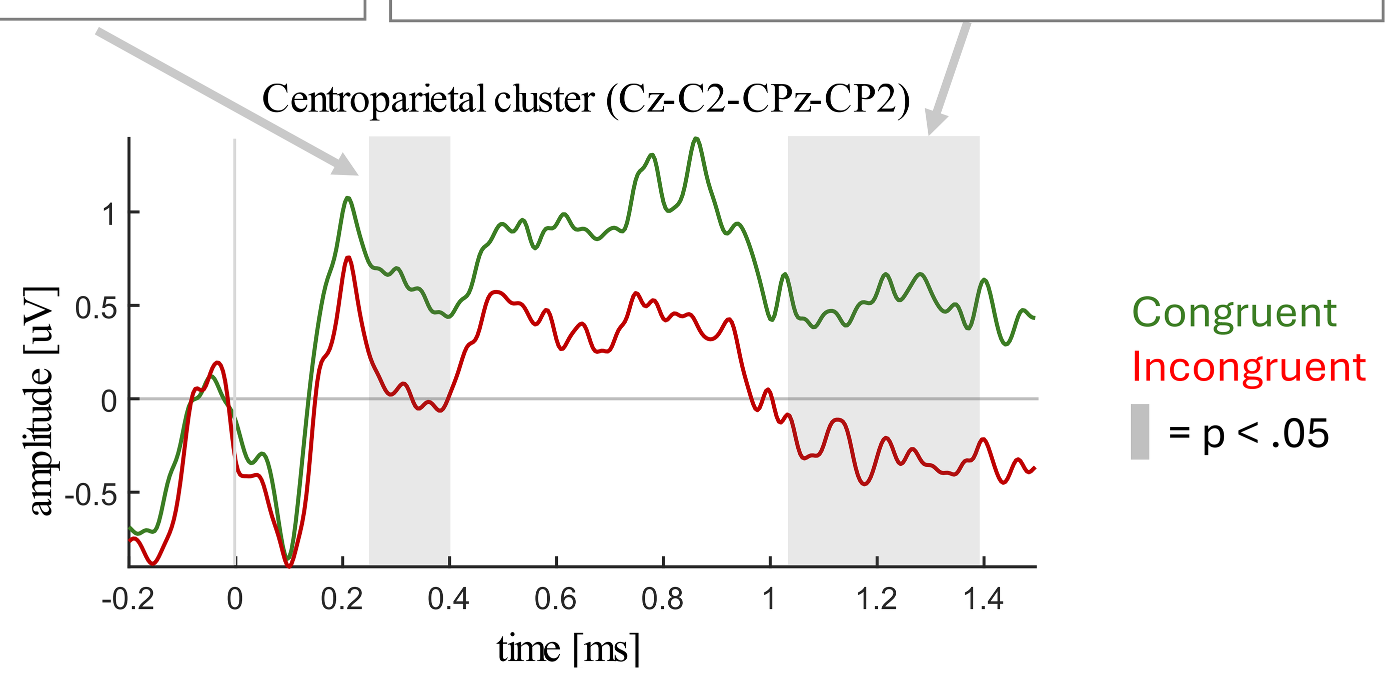
N400 (Berkum et al., 1999)

- 1148 ~ 1428 ms

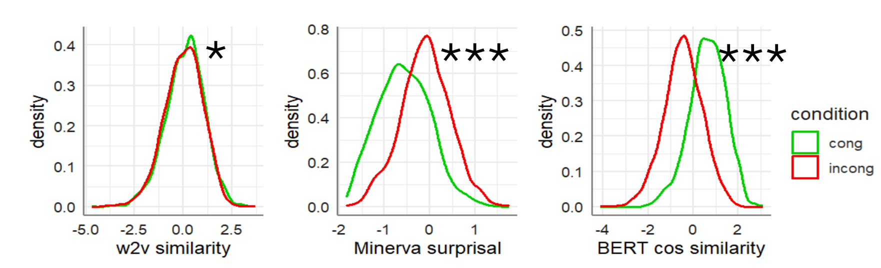
LNC (Friederici et al., 1999)







Analysis 2

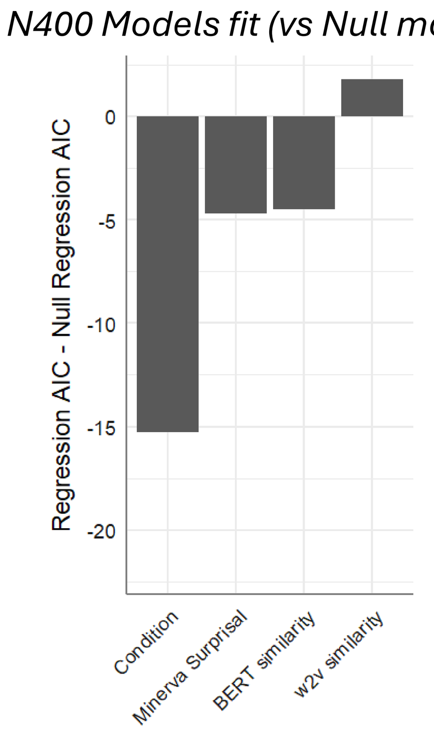


All computational measures captured incongruency at the sentence level

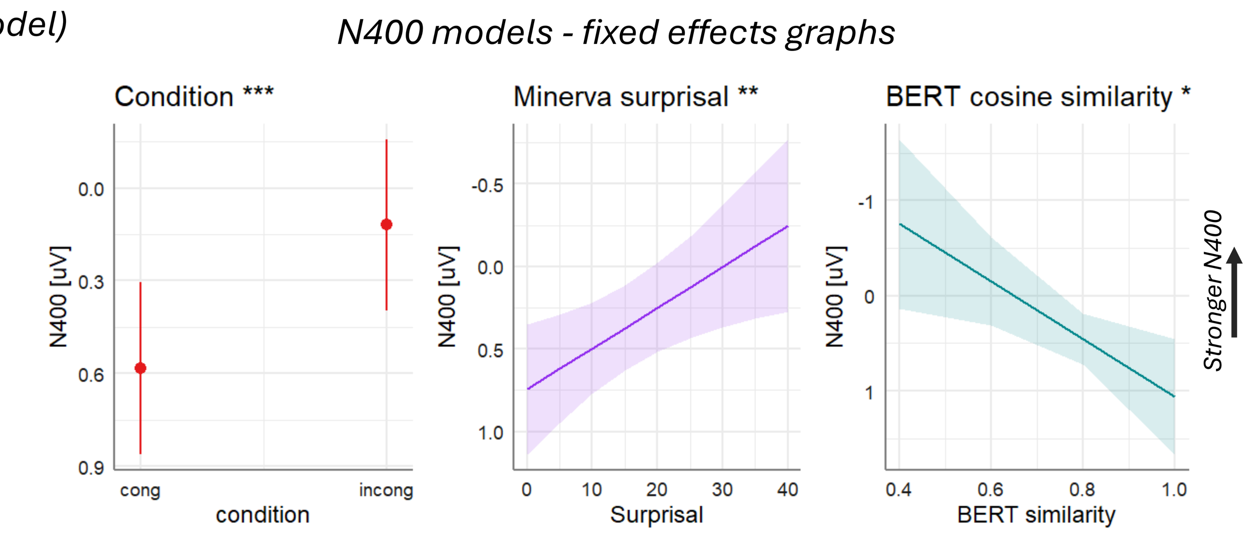
We compared the fit of a null model with different models capturing the effect of: *Condition*, *w2v similarity*, *Minerva surprisal*, *BERT cosine similarity*

We fit separate models for the **N400** and for the **LNC** at centroparietal roi

N400 Models fit (vs Null model)

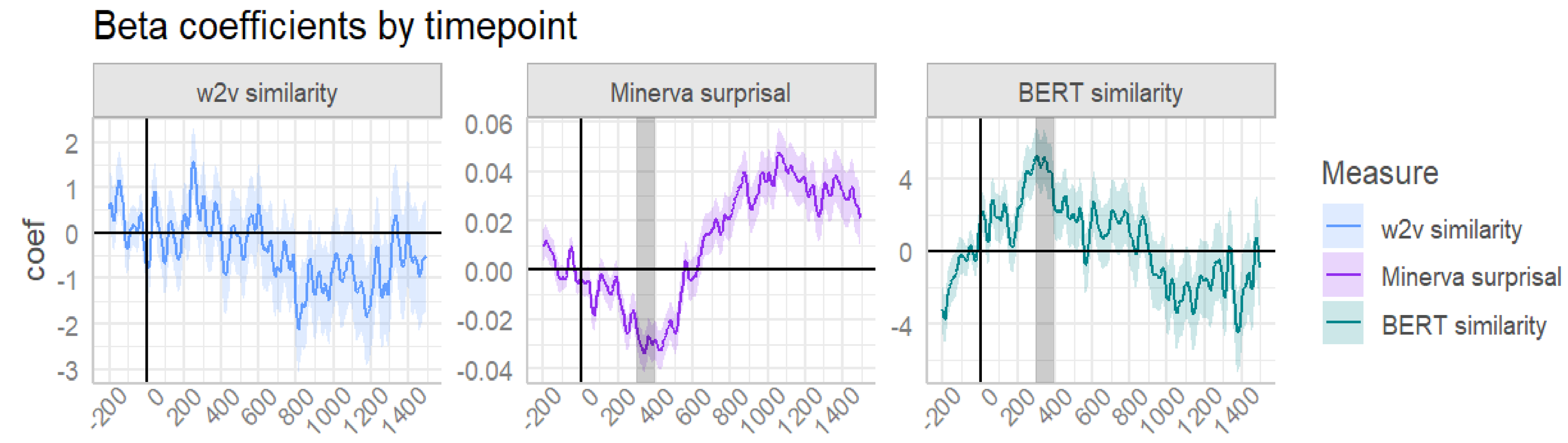


N400 models - fixed effects graphs



Condition, **BERT cosine similarity** and **Minerva surprisal** were found to predict amplitude in the **N400** time-window

Beta coefficients by timepoint



Discussion

We observed a significant effect of semantic congruency in the N400 and a Late Negative Component. The first can relate to difficulty of integration during lexical access (Kuperberg et al., 2020), the latter reflects effort of semantic reinterpretation of sentences' meaning following closure verification (Friederici et al., 1999). These indexes reflect successful semantic integration of novel word meanings.

Context-sensitive measures of semantic similarity and surprisal significantly predicted N400 amplitudes.

In detail, lower similarity between encoding and testing sentences, as well as higher surprisal related to the label in the testing sentence given the encoding context, were found to elicit stronger N400. Interestingly, the LNC was modulated by condition only, not by computational measures. Results are in line with literature on existing words in context (Michaelov et al., 2024) and add important information about the process involved in how we learn novel words through language.

Conclusions

1) Semantic integration of novel meanings through minimal textual exposure elicit cerebral responses detectable by EEG

2) Context-sensitive distributional semantic measures predict cerebral activity in a semantic-sensitive early time window

Take Home Message:

During novel word learning, neural signatures of semantic integration are modulated by linguistic context and can be predicted by context-sensitive computational models of distributional language use

Funded by the European Union
European Research Council Established by the European Commission
The work was supported by the EU
ERC-COG-2022
BraveNewWord (101087053)

Lazaridou A, Marelli M, & Baroni M (2017). Multimodal word meaning induction from minimal exposure to natural text. *Cognitive science*, *41*, 677-705.
Montefinese M, Ambrosini E, Fairfield B, & Mammarella N (2013). Semantic memory: A feature-based analysis and new norms for Italian. Behavior research methods, 45, 440-461.
Amenta S, de Varda AG, Mandera P, Keuleers E, Brysbaert M, & Marelli, M. (2024, September 10). Italian Crowdsourcing Project. Retrieved from osf.io/e4x7w
Friederici AD, Steinhauer K, & Frisch S (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory & cognition*, *27*(3), 438-453.
Berkum JJV, Hagoort P, & Brown CM (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of cognitive neuroscience*, *11*(6), 657-671.
Michaelov JA, Bardolph MD, Van Petten CK, Bergen BK, & Coulson S (2024). Strong Prediction: Language model surprisal explains multiple N400 effects. Neurobiology of language, 5(1), 107-135.
Kuperberg GR, Brothers T, & Wlotko EW (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of cognitive neuroscience*, *32*(1), 12-35.