Capturing effects of context during naturalistic story reading

Shannon McKnight & Albert Kim

Adapted from a talk presented at CUNY 2019

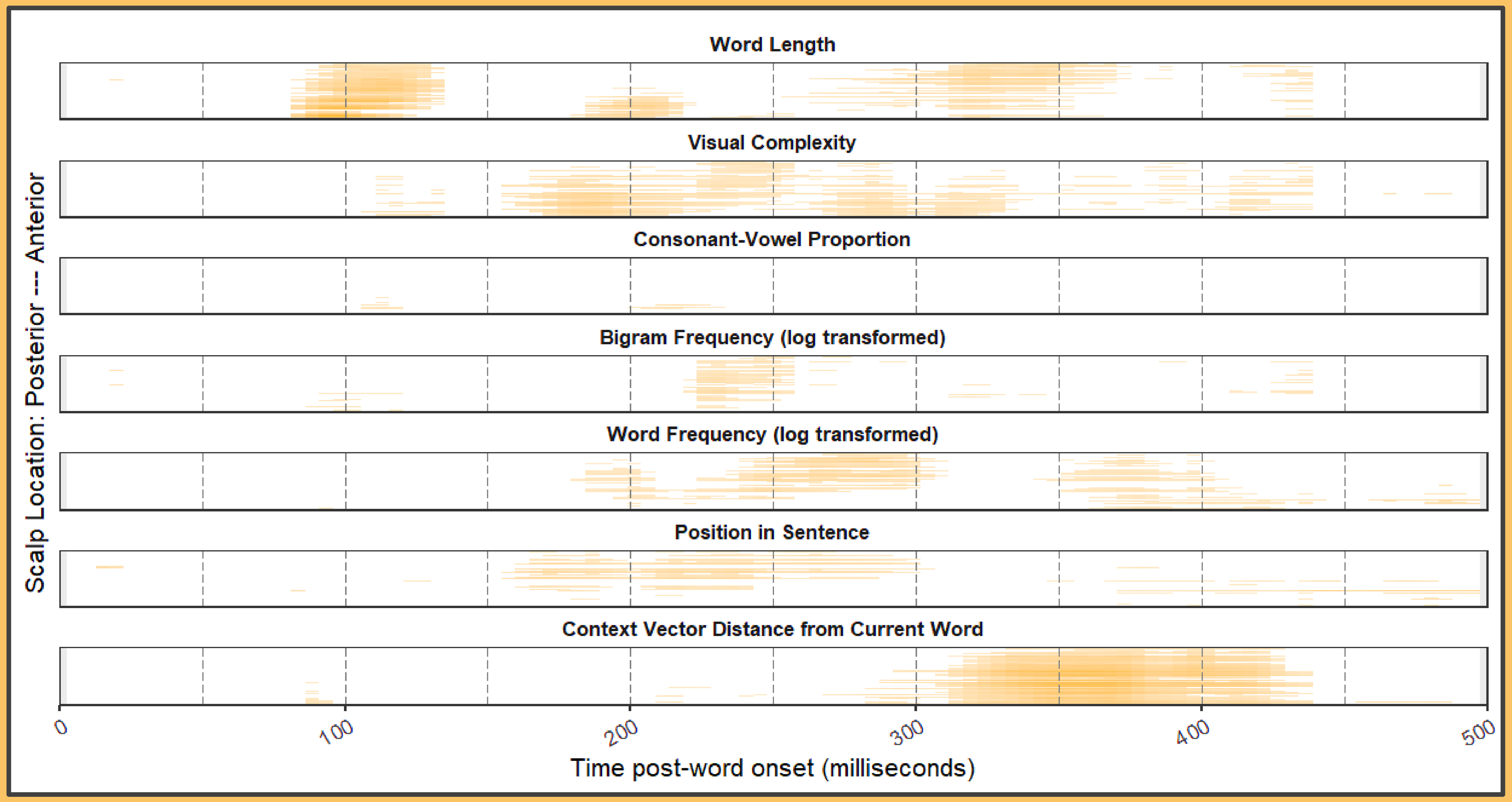
INTRO

- ERP research often focuses on the response to a single "target word" within a sentence that has been designed to violate our expectations based on contextual information.
- Not only does this experimental paradigm limit the amount of data we're able to collect in one session, it does not capture a complete picture of language processing "in the wild"
- Here we employed a self-paced story reading paradigm to capture the processing of rich context and how it may interact with established components of the visual word recognition system.

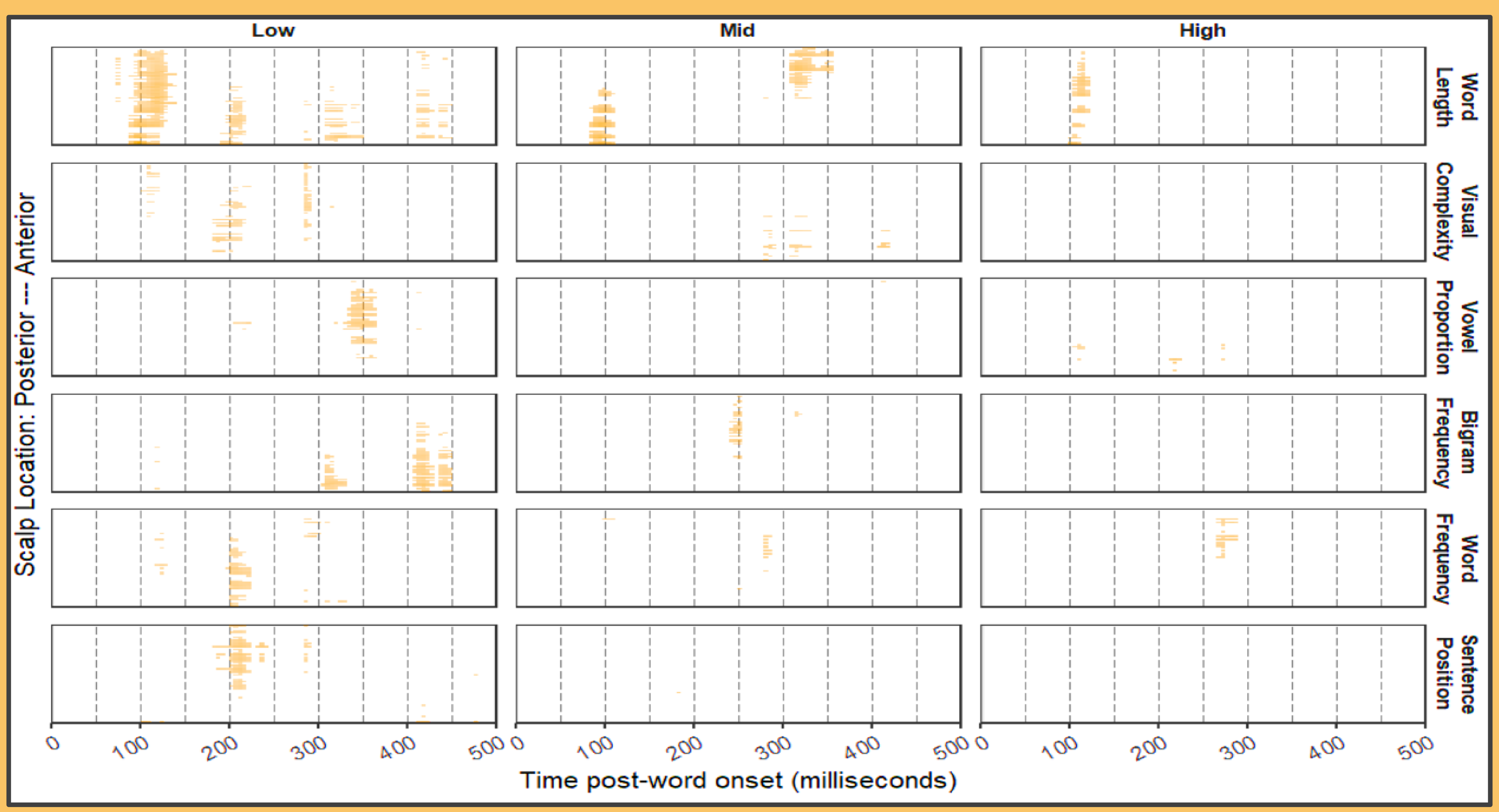
METHOD

- 30 participants read 9 short factual narrative stories during EEG recording
- Stories were presented one word at a time at a self-paced presentation rate
- Participants were instructed to read for comprehension and were probed with a set of multiple-choice questions at the end of each story
- Grand-average ERPs were constructed for each content word (~1400 words total) across all participants between 0-500ms post-word onset, relative to a 200ms pre-word baseline.

Vector representations of semantic similarity capture facilitation effects of context on the N400.



Interactions between Context Similarity (columns) and Lexical Factors (rows): This analysis suggests that contextual similarity has indirect effects before semantic similarity operations that occur at the N400. When contextual similarity is low, lexical factors are more highly correlated with the EEG than when contextual similarity is high. This suggests that physical and lexical information is not used as much during word recognition when a word fits in it's context.



MASS-MULTIVARIATE ANALYSIS

- ERPs were sampled every 5ms across 62 electrode sites
- The following model was evaluated at each combination of time x spatial location:

$$Voltage = \beta_0 + \beta_1 * Length + \beta_2 * Frequency + \beta_3 * ContextFit + \beta_4 * VisualComplexity + \beta_5 * BigramFrequency + \beta_6 * SentencePosition + \beta_7 * VowelProportion$$

- False-discovery rate multiple comparisons corrections were applied across all time x spatial-location combinations

CONTEXT BASED SEMANTIC SIMILARITY

GloVe vector space representations (Manning et al 2014) were collected for all words presented in the stories. To generate a context similarity measure, the vector representations for the 10-prior content words to a given were summed and compared to the current word vector by computing the cosine angle between the current word vector and the summed context vector.

CONTACT

shannon.mcknight@colorado.edu

University of Colorado Boulder

Cognitive Neuroscience of Language Lab