Extending a neural network model of sentence processing to simulate N400 amplitudes during language comprehension and adaptation

Milena Rabovsky (rabovsky@stanford.edu)
Steven S. Hansen (sshansen@stanford.edu)
James L. McClelland (mcclelland@stanford.edu)
Stanford University

Neural networks have been used to implement theories of language comprehension and learning, and there is growing interest in linking these models to brain responses\textsuperscript{1–3}. In the current study, we re-implement the Sentence Gestalt (SG) model\textsuperscript{3} to simulate the most widely used electrophysiological indicator of meaning processing, the N400 component of the event-related brain potential (ERP). The model processes sequentially incoming words into a learned hidden-layer representation of sentence meaning (the Sentence Gestalt, Fig. 1A, Part 1) and uses this representation to respond to probe questions concerning the event described by the sentence (Fig. 1A, Part 2). Learning is based on back-propagated error on the responses to the probes. The model is probed concerning all aspects of the event (e.g., agent, action, patient, etc.) after every presented word, approximating a situation in which a language learner has a complete representation of an event available while listening to a sentence, and leading the model to form a predictive probabilistic representation of all aspects of the meaning of the sentence based on the words presented so far and the statistical regularities in the environment. Using the magnitude of the update of the the Sentence Gestalt layer activation induced by the next incoming word as the model’s N400 correlate, (i.e. $\text{Model N400} = |SG_i - SG_{f,t}|$), we have simulated a series of N400 effects (Rabovsky et al., \textit{Proc Cog Sci}, 2016). Here, we extend the model to address a limitation of the original, and at the same time to capture a close relationship between N400 amplitudes, prediction error, and adaptation in language comprehension, and open the door to further explorations of a new method for training predictive models like the SG model.

References

![Figure 1. A. Model architecture. B. Influences of semantic incongruency and repetition on the model’s N400 correlate.](image-url)