Divide & Concur: A Predictive Coding Account of the N400 Event-Related Potential Component

Samer Nour Eddine[1], Trevor Brothers[1,2], Lin Wang[1,2], Michael Spratling[3], Gina Kuperberg[1,2]; [1]Tufts University, [2]Massachusetts General Hospital, [3]King's College London

Predictive coding is a prominent theory of cortical function that is increasingly invoked to explain aspects of language processing. According to this theory, the brain infers high-level structure from low-level sensory data through iterative cycles of top-down prediction, bottom-up prediction error, and incremental belief updating. In language comprehension, the N400 response has often been linked to “prediction error”, but there has been no prior attempt to explicitly model this ERP component within a predictive coding framework. Here, we developed a computational model of word comprehension based on predictive coding principles. Our goal was to determine if the activation dynamics of this model could accurately reproduce the time-course of the N400 response and its sensitivity to a variety of lexical and contextual factors. Based on a modified interactive-activation architecture (Spratling, 2016), our model includes three hierarchical levels of linguistic representation: orthographic, lexical, and semantic, with distinct state units and error units at each level. On each iteration, state units at a given level predict states at the level below. Any mismatch between these predictions and the true state generates a prediction error (PE), which is passed up to “correct” the state that generated this incorrect prediction. Over time, errors are minimized as the model settles on a correct lexical and semantic state that can accurately explain the bottom-up orthographic input. We operationalized the N400 as the summed lexical and semantic PE produced by the model, averaged in a 10-iteration window around the error’s peak. Lexical Simulations: We selected 512 four-letter words that orthogonally varied in frequency, concreteness and orthographic neighborhood size (ON). Concreteness was implemented as the number of semantic features associated with each word (9 vs. 18). ON was measured as the mean Levenshtein-distance between each word and the 20 nearest neighbors in the model lexicon, and frequency was implemented by biasing the model’s feedback weights based on each word’s corpus frequency. Results: As expected, model PEs reproduced the characteristic rise-and-fall of the N400 response. Consistent with human ERP data, the lexico-semantic PE was enhanced for words with additional semantic features, words with more orthographic competitors, and words with lower frequencies. Contextual Simulations: To simulate the effects of prior linguistic context, we presented word-pairs that were either repetitions (LIME–LIME), semantic associates (SOUR–LIME) or unrelated (BANK–LIME). To simulate contextual predictability, the higher-level state units associated with each word were clamped to activations proportional to their predictability(i.e. cloze), before presenting any bottom-up input. Results: Similar to human readers, lexico-semantic PEs were attenuated by both word repetition and semantic priming. We also observed a graded reduction in PEs as a function of word predictability. Finally, this model also reproduced interactions between lexical and contextual factors, with smaller effects of frequency and concreteness for predictable words. Discussion: Together, these findings suggest that predictive coding can provide a parsimonious and biologically-motivated account of evoked neural activity during word comprehension. Moreover, this approach can potentially simulate behavioral responses, such as lexical decision times, and make novel empirical predictions for neuroimaging studies. References: Spratling, Cogn Proc, 2016