

**What the brain tells us about language comprehension:
A hierarchical generative framework
Gina Kuperberg MD PhD**

I will discuss neuroimaging evidence that supports a hierarchical dynamic generative framework of language comprehension. This framework posits a central role of *generative models* — the set of linguistic and non-linguistic representations that, at any given time, the comprehender believes can best explain the statistical properties of the bottom-up input that she has encountered, given her beliefs about the broader statistical structure of her current environment and her communicative goals. (Kuperberg & Jaeger, 2015; Kuperberg, 2016). I will suggest that, within a given generative model, prediction error at different levels of representation manifests in the brain as distinct spatiotemporal neural signatures. Specifically, neural activity within the left anterior temporal cortex, observed between 300-500ms after the onset of unpredicted inputs, and corresponding to the N400 ERP effect, may reflect prediction error at the level of semantic features, while later activity within the left inferior frontal cortex, observed between 400-700ms and corresponding to a later anteriorly distributed negativity ERP effect, may index prediction error at the level of event structures (representations of ‘who does what to whom’). Incremental modulation of this temporal-frontal neural network may reflect iterative cycles of probabilistic prediction and inference that proceeds until prediction error across the entire generative model is minimized and the comprehender has converged upon the particular message-level representation that best explains the bottom-up input. I will further argue that these generative models are not fixed, and that comprehenders can modify their structure, or switch to (or infer) alternative, previously stored models, in rapid response to changes in the statistical structure of their broader communicative environment and/or their communicative goals. Such adaptation manifests in the brain as an additional set of spatiotemporal neural signatures, corresponding to a series of late positivity ERP effects, which are distinguished by the levels of representation at which adaptation takes place. I will conclude by briefly discussing the implications of this dynamic generative architecture for linking comprehension, production and learning in healthy individuals, and how these links might break down in neuropsychiatric disorders such as schizophrenia (Brown & Kuperberg, 2015).

Kuperberg GR, Jaeger TF. *Language, Cognition & Neuroscience*. 2015

Kuperberg GR. *Language, Cognition & Neuroscience*. 2016.

Brown M, Kuperberg GR. *Frontiers in Human Neuroscience*. 2015.