Chunking Constrains Prediction during Language Comprehension

Abstract

Speech is a highly rapid sequence with complicated internal structures. Prediction and chunking are two possible mechanisms for the brain to efficiently process speech and they are often viewed as separate or even opposing mechanisms. Here, we investigate whether the two mechanisms interact and hypothesize that the chunk structure in speech modulates how the brain predicts basic linguistic items, i.e., morphemes. In three magnetoencephalography (MEG) experiments in Mandarin Chinese, we characterize neural prediction of morphemes using the neural response to morpheme surprisal and analyze how this response is modulated by chunks, i.e., major linguistic constituents. We demonstrate that the MEG surprisal response is significantly stronger for morphemes belonging to the ongoing chunk than morphemes across a chunk boundary. This chunk-boundary effect on morpheme prediction is further modulated by the certainty of a chunk boundary. The conclusions are also confirmed by analyzing open dataset of MEG responses to English narratives. In summary, these results strongly suggest that the brain employs a chunk-based prediction strategy and more precisely predicts items within a chunk.

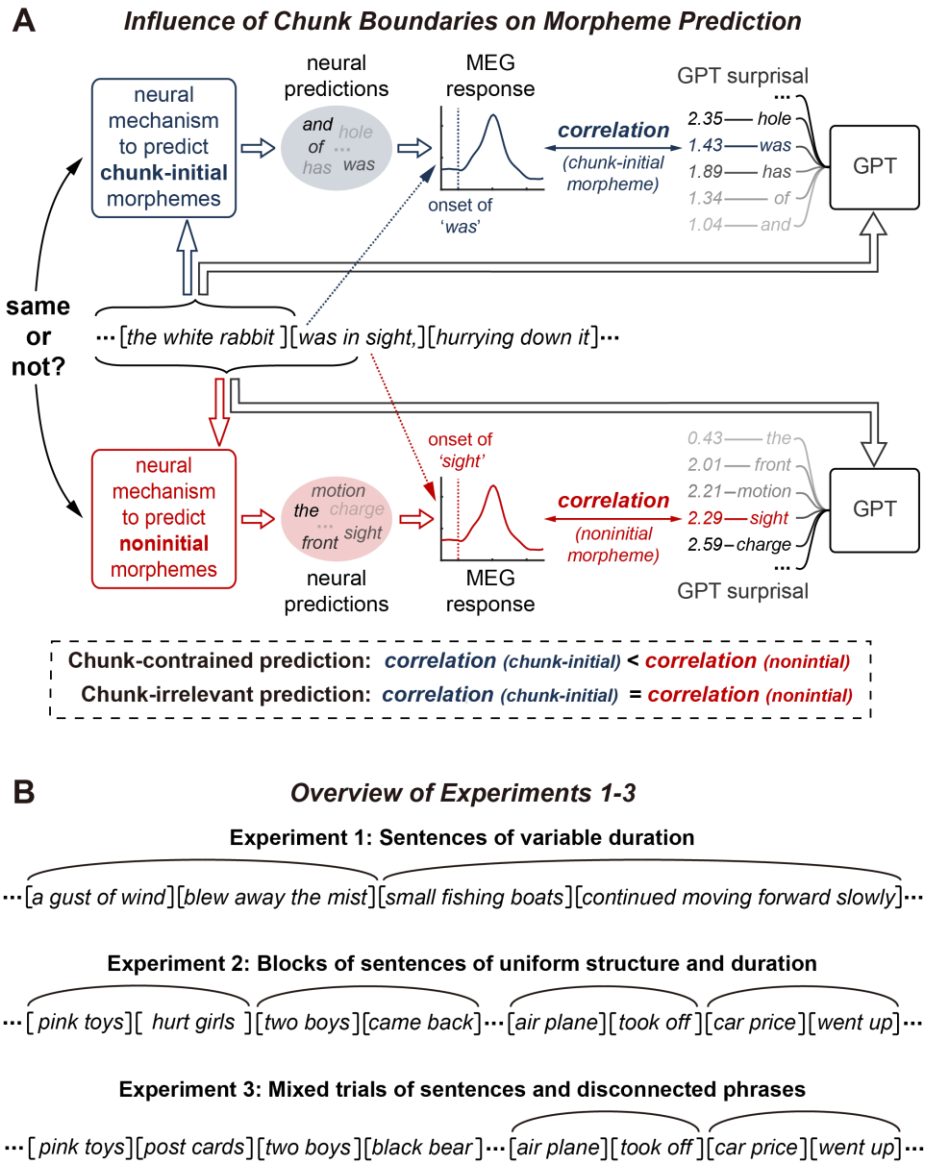

Figure 1. Chunk-constrained prediction hypothesis and experiment overview. (A) The study investigates how chunk boundaries, i.e., boundaries between major linguistic constituents, influence neural prediction of upcoming words/morphemes. In most word prediction theories and LLMs, each upcoming word is predicted using the same mechanism regardless of whether the upcoming word continues a linguistic chunk or initiates a new chunk. The chunk-constrained prediction hypothesis, however, posits that the brain more actively predicts words that continue a linguistic chunk than words that initiate a new chunk. We test the hypothesis by evaluating the MEG response encoding surprisal, which is quantified using GPT-2. (B) Experiment 1 presents sentences that are composed of two major chunks, i.e., the subject and the predicate, which both have variable durations. Experiments 2 and 3 present sentences that have more certain boundaries. In Experiment 2, sentence sequences are presented in blocks, while in Experiment 3, sentence sequences are mixed with sequences of disconnected phrases.