DEPENDENCY LOCALITY OPTIMIZED FOR COMPUTATIONAL RATIONALITY

Weijie Xu (UC Irvine), Richard Futrell (UC Irvine)
weijie.xu@uci.edu

Language processing requires active engagement of the computational resources of the cognitive system. For example, when comprehending a long-distance dependency, as in (1), the antecedent "key" is maintained active in working memory to be retrieved later at "was". However, computational resources are often limited, acting as an evolutionary pressure that constrains the overall length of syntactic dependencies in language, an efficiency principle termed *dependency locality* [1]. The current study argues that the principle of dependency locality is optimized for computational rationality in language processing [2].

Hypothesis. As an instantiation of computational rationality, memory resources are strategically allocated: More resources are allocated to encode novel and unpredictable information, yielding a more robust memory representation against interference [3]. For syntactic dependencies, we hypothesize that antecedents less predictable from context may get better memory representation, and can be more accurately retrieved later [4-5]. Therefore, we predict that less predictable antecedents should be able to tolerate longer dependency length since they have been strategically given more memory resources to process.

Method. We examined the hypothesis in English using the Georgetown University Multilayer (GUM) corpus[6] taken from Syntactic Universal Dependencies project (SUD)[7]. We obtained from GPT-3 neural language model[8] the surprisal of each word (the negative log probability of the word given a context $-\ln p(w|c)$), with all the text that precedes the target word in the corresponding document fed into the model as the context. We then collected all the syntactic dependencies (*N*=89099) contained in each sentence. The dependency length is calculated as the number of intervening words between the head and the dependent.

Result. We fit a linear mixed effect model as in (2), with antec-surpr (antecedent surprisal) as the critical fixed effect, and antec-deprel (syntactic relation of the antecedent in the dependency) as the random effect. We also included three control variables: sent-pos (sentence position in the text), antec-pos (antecedent position in the sentence), and sent-len (word counts of a sentence). Crucially, we find a significant positive effect of antecedent surprisal on the dependency length ($\hat{\beta}$=0.16, p<0.001), whereby the dependencies with antecedents of higher surprisal bear longer dependency length (Figure 1). For control variables, longer sentences have longer dependency length ($\hat{\beta}$=0.54, p<0.001); antecedents appearing earlier in the sentence have longer dependency length ($\hat{\beta}$=-0.52, p<0.001); there is no effect of sentence position.

Conclusion. We find that the dependency locality is modulated by the predictability of antecedents, which is proportional to the amount of computational resources required for processing. This supports that dependency locality is further optimized for computational rationality in language processing. Broadly speaking, our result suggests that cognitive constraints such as memory limitations act as an evolutionary force that shapes the structure of human language [9].

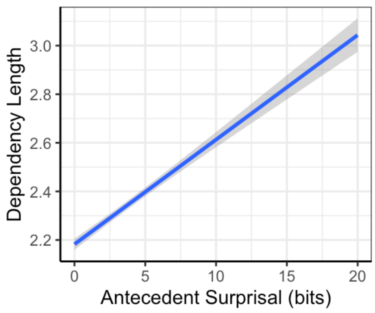

Figure 1: Correlation between antecedent surprisal and dependency length

(1) The key to the cabinet was in the drawer.

(2) dep-length $\sim$ sent-pos + antec-pos + sent-len + antec-surpr + (antec-surpr | antec-deprel)

References. [1] Futrell et al. (2020) *Language*; [2] Gershman et al. (2015) *Science*; [3] Hahn et al. (2022) *PNAS*; [4] Hofmeister (2011) *Lang Cogn Process*; [5] Xu & Futrell (2022) *AMLaP*; [6] Zeldes (2017) *Lang Resources and Eval*; [7] Kahane et al. (2021) *TLT* [8] Brown et al. (2020) *NeurIPS*; [9] Gibson et al. (2019) *Trends in Cog Sci*