Supplemental Material for “Hierarchical structure guides rapid linguistic predictions during naturalistic listening”

Jonathan R. Brennan*, University of Michigan
John T. Hale, Cornell University

November 2, 2018

*Corresponding Author: jobrenn@umich.edu
Figure S1: Example sentences showing word-by-word surprisal estimates from each of the three models.
Figure S2: **All significant whole-head effects** for surprisal terms from the whole-head analysis, divided by whether they are fit against function-word data (A), or content-word data (B). Black lines represent the time-series of regression fits ($\beta \pm CI_{95}$) for the indicated surprisal term. Dark grey shading indicates significant time-points and inset shows significant channels and coefficient averages across significant time-points. It appears as if the various function-word effects have significant temporal and electrode overlap. Model comparisons (Tables 3 and S1) indicate that *Ngram* and *SRN* surprisals significantly improve model fits over *CFG* for these electrodes and time-windows. Despite the specificity to function-words in this whole-head analysis, such improvements do not interact with function-vs-content word-class when such an interaction is explicitly tested.
Table S1: **Auxilliary model comparison results** including 8 participants who did not meet the pre-specified behavioral criterion (total N = 41). The table shows WAIC difference scores (± standard error) for changes in model fit across six ROIs (columns). Each set of rows tests a different statistical question using step-wise model comparison. Terms that are being evaluated are indicated to the left of “>”; interactions with word-class are indicated with “:WC”. For each, the baseline model includes all control covariates along with the indicated surprisal term(s) and as any interactions between word-class and those surprisal terms. The values are scaled so that positive numbers represent improvements for the larger model, while negative numbers indicate that the added complexity of the larger model is not justified. Bold-face indicates WAIC improvements that are more than two standard errors from zero.
### Table S2: Model goodness-of-fit summaries from the set of models fit against each of the 6 ROIs examined in the model-comparison analysis.

**RHat** summarizes the concordance between chains used in the Monte-Carlo estimation procedure and should be below 1.1. **R2** is a Bayesian formulation of the proportion of variance explained in the single-trial EEG data (*bayes_R2()* function from the *rstantools* package). **AR(1)** indicates that the first-order auto-correlation of the residuals is low.

| Time Interval | Rhat  | R2    | AR(1)  |
|---------------|-------|-------|--------|
| 216–554 ms    | 0.999 | 0.0004| 0.187  |
|               |       | 1.003 | 0.0005 | 0.171  |
| 174–420 ms    | 0.999 | 0.0005| 0.144  |
| 102–158 ms    | 0.999 | 0.0005| 0.068  |
| 174–252 ms    | 0.999 | 0.0005| 0.081  |
| 210–310 ms    | 0.999 | 0.0006| 0.052  |

