Early Interaction Effects of Word Type and Frequency on Lexical Decision Processes

Joshua Krause, Jelmer P Borst, Jacolien van Rij
Bernoulli Institute, University of Groningen

Background To make lexical decisions (LD), people progress through a series of cognitive processing stages. For example, by using a novel machine-learning method on EEG data, Berberyan and colleagues (2021, *Brain & Cogn.*) revealed that native Dutch speakers always complete a sequence of six stages to perform LD; independent of whether the stimuli were high- or low-frequent Dutch words, pseudo-words, or random character strings. Instead, these categories exclusively affected the *duration* of the fifth stage, hypothesized to reflect the main decision. However, previous evidence from ERP and eye-movement studies suggested earlier frequency effects (~ 100 ms post stimulus; e.g., Sereno et al., 1998, *Neuroreport*). **In the current EEG study we investigated the effect of continuous frequency of words *and* non-words on the duration of processing stages involved in LD.** In particular we were interested in revealing whether continuous frequency changes affect early processing stages and whether the effect of frequency differs between stages.

Method Native Dutch speakers (N=26) performed an LD task with 250 real Dutch words, 125 pseudo-words (derived from the real words), and 125 random character strings (“nonsense-words”; derived from shuffling the remaining real words) while EEG was recorded. Stimuli were selected from the DLP corpus (Keuleers et al., 2010, *Front. Psych.*). Because it allows to approximate non-word frequencies (cf. Hendrix & Sun, 2021, *JEP:LMC*), the Google result count was used as frequency measure for every stimulus (see Figure 1). We applied the machine learning method (HsMM-MVPA; Anderson et al., 2016, *Psych. Rev.*) used by Berberyan et al. (2021) to obtain processing stages for each LD trial. The duration estimates (log transformed) for each processing stage were analyzed using generalized additive mixed-effects models (GAMMs; Wood, 2017). We included (potentially nonlinear) effects of log-transformed frequency and word type as predictors, in addition to (nonlinear) random effects to account for variation in participants and stimuli.

Results We found six processing stages (see Figure 2) independent of word type, similar to Berberyan et al. (2021). Trial level results are visualized in Figure 3. The duration of early stages (0-70, 70-150, 150-240 ms) increased subtly but reliably with increases in frequency. While frequency also had an effect on the duration of the fourth stage, the word type effect is comparatively more pronounced in this stage. Longer durations were observed for non-words compared to words, independent of frequency scores – theoretically providing sufficient information to make the LD. Stage five revealed a complex significant non-linear interaction between frequency and word type: for nonsense-words stage duration increased linearly with frequency while for pseudo-words a more non-linear effect was visible. For words stage duration decreased non-linearly with frequency. The pattern visible in stage six is qualitatively similar to the pattern in the fifth stage.

Discussion Interestingly, we found that frequent stimuli prolong the earliest processes involved in LD, which we think are related to visual word recognition. For decision-related processes, the effect of frequency appears to depend on the goal of the decision. Compared to the pronounced stage duration differences between words and non-words frequency played a smaller role in stage four. This could be expected from a simple early discrimination between words and non-words where frequency might highlight unnecessary differences between non-words. Conversely, no clear word type distinction was observable in the “main” decision process attributed to the fifth stage. Instead frequency had a strong non-linear effect in this stage, possibly because more nuanced differences between stimuli are considered to reach a conclusive LD. The visible overlap in stage five durations for similarly frequent stimuli indeed suggests this more nuanced stimulus inspection to rely on frequency information. Concluding, frequency information is utilized at virtually every processing step involved in LD, yet the way it is utilized differs depending on the role fulfilled by each step.

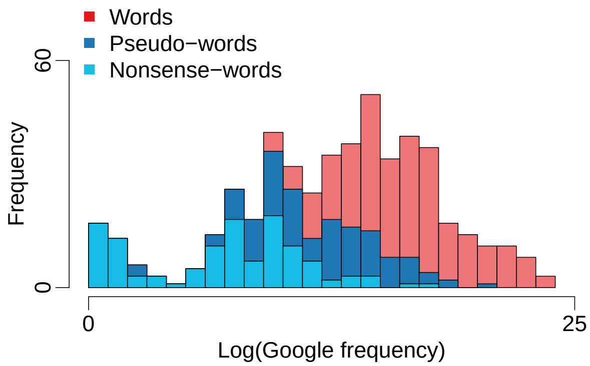

Figure 1: Google frequency of the stimuli.

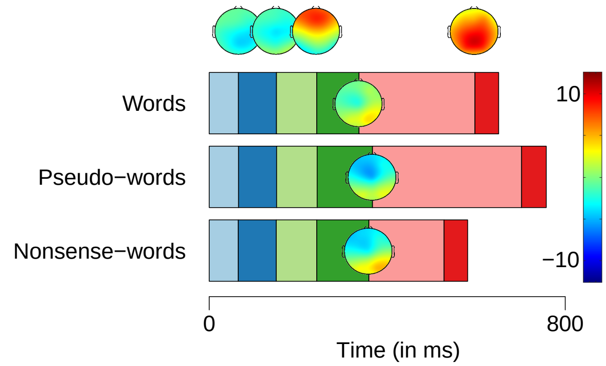

Figure 2: Average HsMM-MVPA stage duration

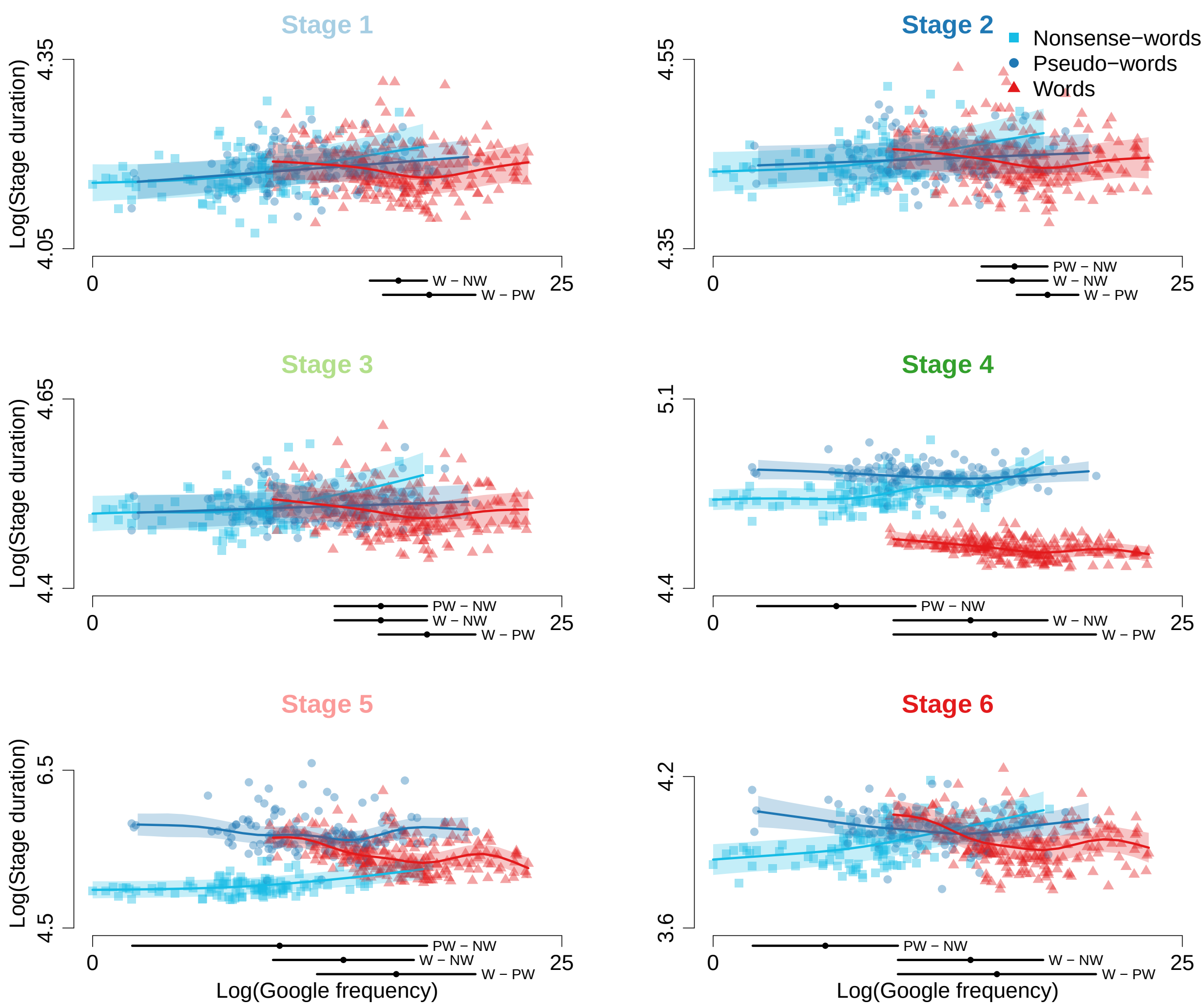

Figure 3: Estimated effects of frequency on stage duration per word type and processing stage. Since stages differ in duration, the y-scale is adjusted per stage. Colored lines reflect log-transformed stage duration GAMM estimates. Shaded areas reflect GAMM standard error estimates. Dots reflect average (over subjects and items) log-transformed stage duration values for a particular frequency score. Horizontal labeled black bars reflect range of frequency for which the difference in predicted stage duration between word types is significantly different from zero according to a 95% certainty interval.