Linear (not logarithmic) effects of lexical predictability: A challenge to surprisal theory

While all contemporary models of language comprehension accept the important role of sentence context, there are disagreements over the cognitive mechanisms linking contextual predictability and word processing difficulty. Specifically, different models of anticipatory language processing make divergent predictions over whether the behavioral function linking word predictability and reading time is *linear* or *logarithmic.*

According to surprisal theory (Hale, 2001; Levy, 2008) the primary determinant of online processing difficulty is the negative *log*-probability of a word given the preceding context (i.e. *surprisal*). This logarithmic relationship arises naturally from an incremental, predictive comprehension architecture in which comprehension difficulty reflects the magnitude of the shift of prior message-level probabilities after encountering a new word. A logarithmic linking-function between word predictability and reading time is also predicted in the Bayesian Reader model (Norris, 2009; Bicknell & Levy, 2010), which claims that word identification unfolds by optimally combining contextual and perceptual information though the incremental application of Bayes rule.

In the present study, we tested the empirical predictions of these models in two high-powered experiments. Unlike previous studies, which have used correlational approaches (Smith & Levy, 2013, Luke & Christiansen, 2016), here we parametrically manipulated lexical predictability while holding other, potentially confounding lexical variables constant. Word predictability was operationalized using an offline sentence completion task (*cloze*), which has been shown to better account for behavioral reading times than corpus-based measures (Smith & Levy, 2011). By sampling heavily at the low end of the probability scale (0-20% cloze) and including a large number of items and participants, we were able to obtain sufficient statistical power to adjudicate between *linear* and *logarithmic* models of lexical probability effects.

<u>High cloze (91%)</u>: *Her vision is terrible and she has to wear* ***<u>glasses</u>*** *in class.*
<u>Moderate cloze (20%)</u>: *She looks very different when she has to wear* ***<u>glasses</u>*** *in class.*
<u>Low cloze (1%):</u> *Her mother was adamant that she has to wear* ***<u>glasses</u>*** *in class.*

In Experiment 1, 216 participants recruited using Mechanical Turk were asked to read individual sentences for comprehension, like the examples above, using a moving window self-paced reading paradigm. In Experiment 2 (N = 36) participants were asked to name a picture (*glasses*) appearing at the end of auditory sentence contexts, while naming latencies were recorded. As expected, increasing levels of word predictability led to reductions in processing time in both experiments. Critically, in both experiments this relationships was strongly linear, with significantly better model fits for linear compared to logarithmic models (see Figure 1). In addition, we conducted a separate meta-analysis of eight, previously published eye-tracking while reading studies (N = 223). This analysis also demonstrated a clear linear relationship between predictability and reading time, suggesting this relationship is robust and replicable.

We believe this linear relationship is most consistent with a model in which readers pre-activate semantic features according to their contextual probability, which in turn facilitates lexico-semantic access. We discuss the implications of our results for surprisal theory and Bayesian Reader, including which theoretical assumptions are (and are not) compatible with a linear relationship between predictability and reading time. Finally, we provide methodological recommendations for researchers interested in investigating or controlling for the effects of predictability during sentence comprehension.

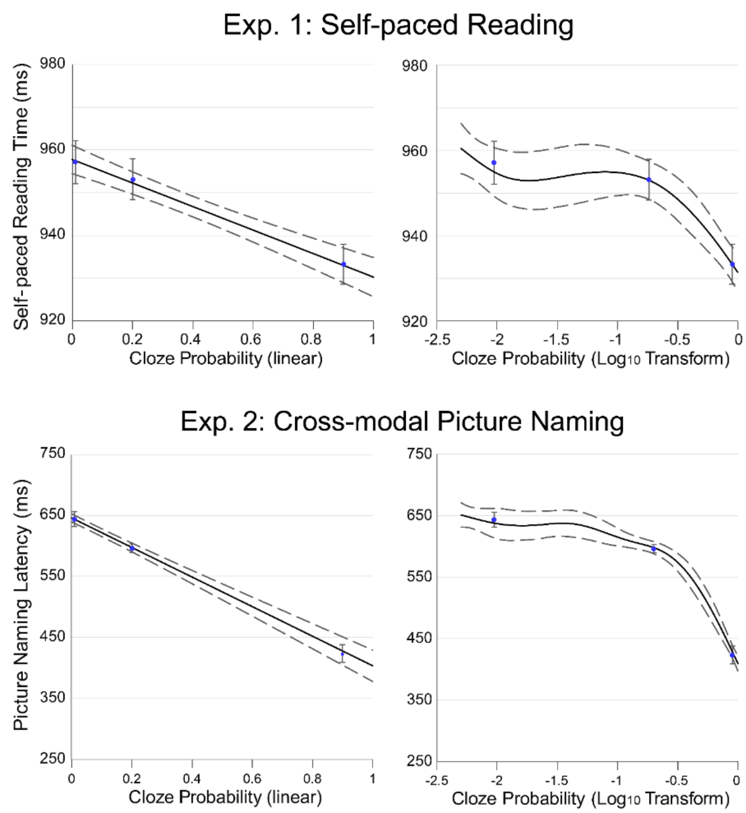
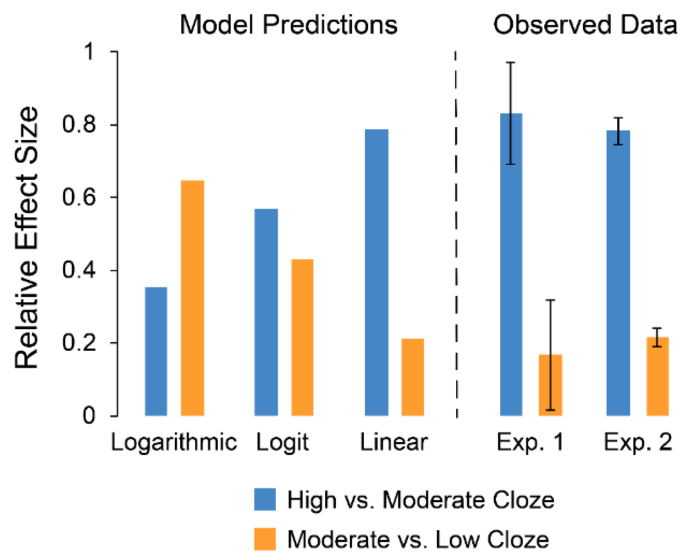


Figure 1. The relationship between word predictability and self-paced reading times (upper), and picture naming latencies (lower). Reading times are reported for a three-word critical region (following Smith & Levy, 2013). Blue dots represent mean RTs for low, moderate, and high cloze items, with error bars representing 95% confidence intervals. Note the linear relationship between cloze probability and processing time across both experiments (left). When cloze was log-transformed, the predictability-RT relationship became strongly non-linear (right).

Figure 2. Relative size of the word predictability effect when comparing high-cloze vs. moderate cloze words (.91 vs. .2) and moderate-cloze vs. low-cloze words (.2 vs. .01). Results predicted by logit, logarithmic, and linear models are plotted on the left, observed data right. Error bars represent within-subject SEMs.



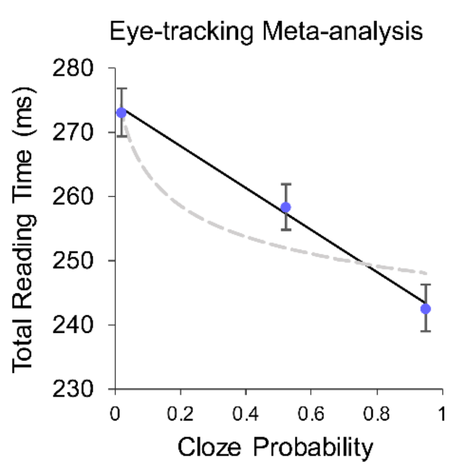


Figure 3. Total reading time results from a meta-analysis of eight eye-tracking while reading experiments (Rayner & Well, 1996; Rayner, Reichle, Stroud & Williams, 1996; Rayner Li Juhasz & Yan 2005; Soreno, Hand, Shahid, Yao, O'Donnell, 2017; Paul, *unpublished* dissertation) Best linear fit is shown in black, while the logarithmic fit are shown in grey. Error bars represent 95% confidence intervals.