The Effect of Context on Typing Time: Evidence from 100,000 TypeRacers

Robert Chen, Roger Levy, Tiwalayo Eisape (MIT)
robertcc@mit.edu

Context effects in human spoken language are well-documented and play a central role in language production. However, the role of context in written language production is far less well understood, even though a considerable proportion of the language produced by many people today is written. Here we use computational language models (LMs) to quantify the effect of context-based predictability on typing time in a subset of the data available on TypeRacer.com.

TypeRacer is a viral online typing game where players race against themselves, friends, or strangers in groups of up to 10 to complete a short text as quickly as possible (Fig. 1). With races in 50 languages and a wide variety of text genres, TypeRacer is an openly accessible, massive, and untapped dataset of typing times that contains data from over 100,000 users — across 35,000 distinct texts, and 70,000,000 races.

We take a random sample of 100 users from TypeRacer and a random sample of up to 100 races from each of those users, resulting in a total of 317,000 measures of word typing times (μ=49.1 words per race, σ=20.4). Of our sample of 100 users, 4 do not list their location, 41 are in the United States, followed by 9 in Canada, 5 in India, 4 in the United Kingdom, and 37 from other countries. 92 do not list their age, and the ages of the remaining range from 13 to 28 (μ=19.1, σ=4.8). 77 do not list their keyboard layout, 22 use Qwerty, and 1 uses Colemak.

We use LMs trained on the WikiText-2 dataset (Merity et al., 2016) to estimate in-context probability for the words in our dataset. The models we compare on this task are as follows. **Forward full surprisal**: An LSTM language model trained with a standard autoregressive language modeling objective. **Backward full surprisal**: A variant of **Forward full surprisal** trained to predict the text in WikiText-2 in reverse. **Forward bigram surprisal**: A variant of **Forward full surprisal** where, during training and inference, context is limited to only the previous word. **Backward bigram surprisal**: A variant of **Forward bigram surprisal** trained to estimate bigram probability in reverse. **Unigram surprisal**: negative log-frequency estimates based on data from the COCA corpus (Davies, 2010).

We use generalized additive models (GAMs; Wood 2006) to determine the functional form of the relationship between each of our context-based predictability estimates and typing time (Fig. 2). Furthermore, to capture both fixed and participant level effects we use a “two-stage” approach (Gelman, 2005) in which we fit a linear mixed-effects model with the above effects (plus word length and random by-word intercepts/slopes for all surprisal effects) for each participant individually, and then analyze the distribution of fitted coefficients (Fig. 3).

We find the same general shape of effects of word properties and context-based predictability on typing time as has been documented for word duration in spoken language production, but we also find key differences in the detailed patterns of sensitivity. Firstly, the results of our GAM analysis show that, in the predictor ranges where most of the data lie, typing times are roughly linear in word length and log-probabilities, with the notable exception of unigram surprisal (word frequency), which has a nonlinearity in the 10–15 bit range for which we do not yet have an explanation (Fig. 2). In our second analysis, the median and distribution of surprisal coefficients show that the effect of predictability based on left context is a stronger determinant of typing time than predictability based on right context (Fig. 3). This contrasts results in spoken language production that show the opposite effect (Bell et al., 2009) Furthermore, we find predictability based on local context (bigram surprisals) is a stronger determinant of typing time than predictability based on global context. Notably, word frequency (unigram surprisal) has no predictive value for typing time once surprisal effects are taken into account (Fig. 3).

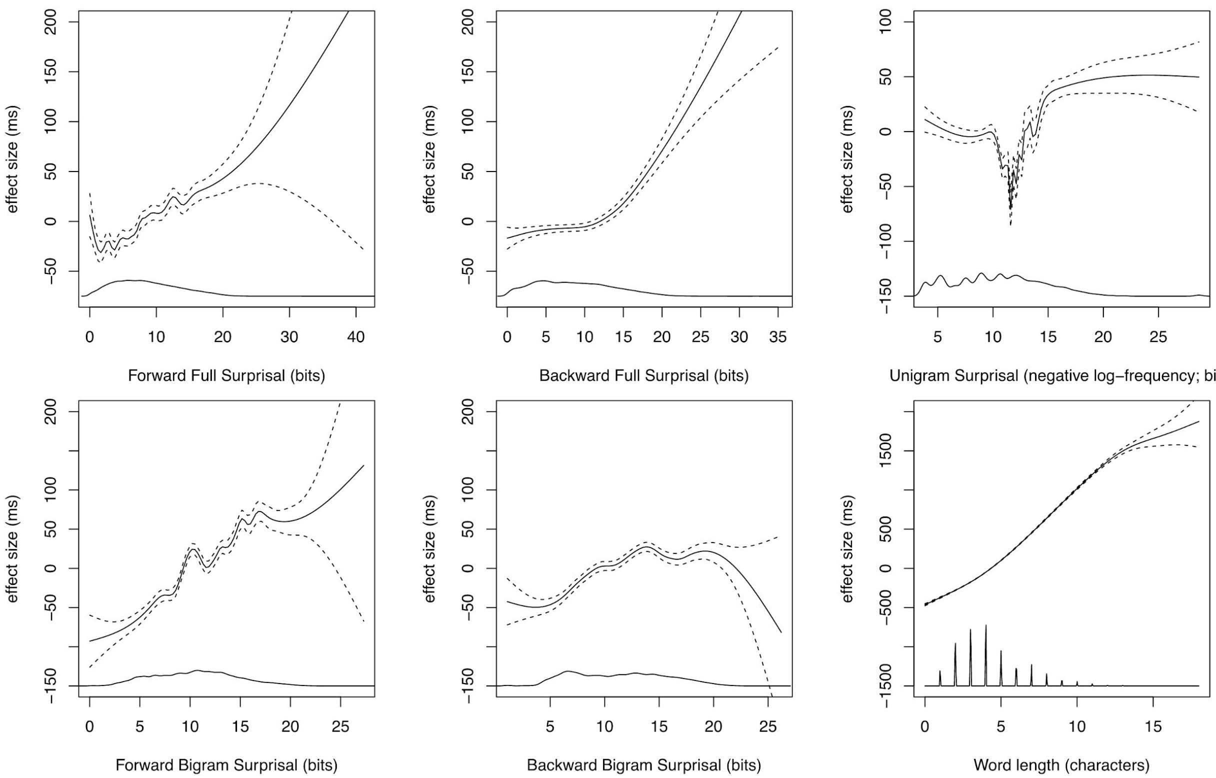

Figure 2: Relationship between various estimates of contextual probability (and frequency) and typing speed slowdown. Regression lines from fitted GAM models are shown as solid lines. Dashed lines indicate 95% confidence intervals but do not take into account the repeated-measures structure of the data or uncertainty in GAM hyperparameter values, and hence should be taken with a grain of salt. The marginal density of each predictor is shown at the bottom of each plot.

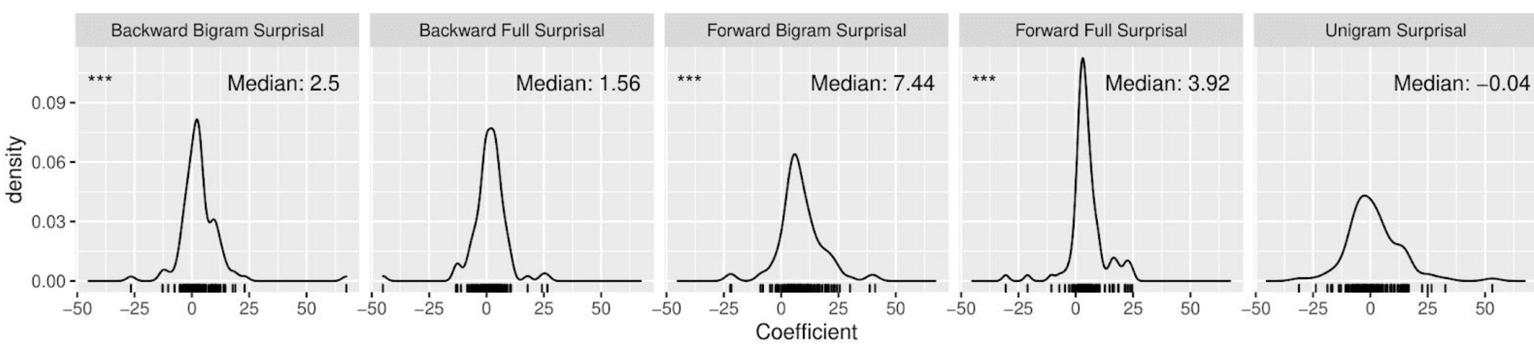

Figure 3: Median and distribution across participants of estimated predictor coefficients. Predictors marked with *** have mean above 0 at p<0.001 (t-test); other predictors have mean not significantly different from 0. For word length, all participants have estimated coefficient above 0, with median 153ms/character (not shown).

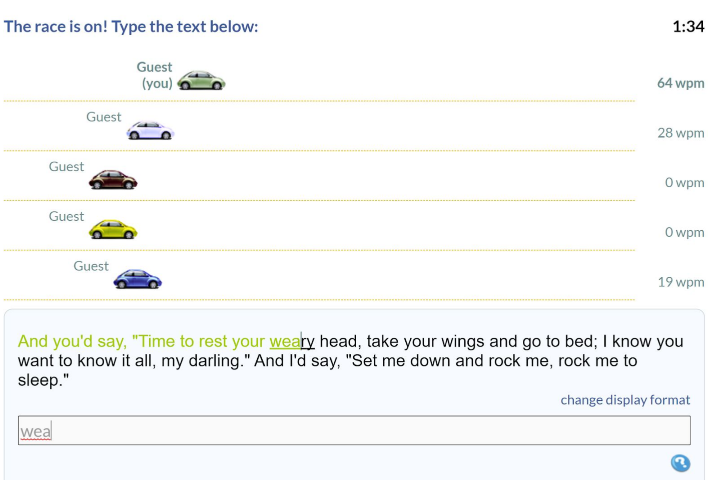

Figure 1: An in-progress race on TypeRacer.com. Racers are given up to 12 seconds to read the race prompt before the race starts.

References:

1. Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. Journal of Memory and Language, 60(1), 92-111.
2. Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. Literary and linguistic computing, 25(4), 447-464.
3. Gelman, A. (2005). Two-stage regression and multilevel modeling: a commentary. Political Analysis, 13(4), 459-461.
4. Merity, S., Xiong, C., Bradbury, J., & Socher, R. (2016). Pointer sentinel mixture models. arXiv preprint arXiv:1609.07843.
5. Wood, S. (2006). Generalized Additive Models: An Introduction with R. CRC Press.