

Supplemental Materials

Frequency and predictability effects for line-final words

Adam J. Parker & Timothy J. Slattery

This Supplemental Materials includes additional resources to support the article *Frequency and Predictability Effects for Line-Final Words*.

Distribution of Lexical Variables

Figure 1 shows the distributions of lexical variables for intra-line and line-final words. From this figure, the distributions of these variables are highly similar.

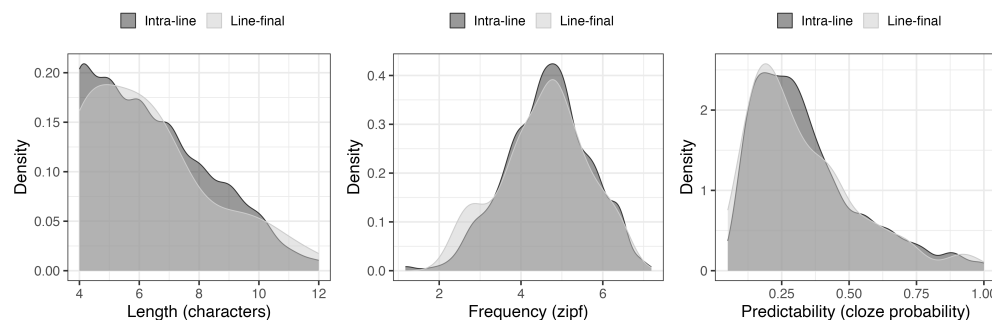


Figure 1: Density plots illustrating the distribution of word lengths, frequency ratings, and cloze probability values for observations entering our analysis of the Provo Corpus. The distributions are shown in dark grey for intra-line words and light grey for line-final words.

Variance Inflation Factors

Inspection of Table 1 in the main article suggest a strong correlation between word length and frequency for analysed intra-line and line-final words. To assess the amount of multicollinearity in each of our linear mixed-effects analyses for the Provo Corpus, we used the `check_collinearity()` function from the *performance* package (version 0.10.8; Lüdtke et al., 2021) to calculate a variance inflation factor (VIF) for each fixed effect within the models fitted to single-fixation duration and gaze duration. Following Menard (2001), we interpret a $VIF > 5$ as a cause for concern and $VIF > 10$ as problematic.

As can be seen in table 1, the VIFs are all below 5. This indicates that for our analysis of the Provo Corpus, multicollinearity is not a concern within our modelling framework.

Table 1: Variance Inflation Factors for linear mixed-effects models
fitted to single-fixation and gaze durations from the Provo Corpus.

Model	Fixed effect	Single fixation duration			Gaze duration		
		VIF	2.5% CI	97.5 % CI	VIF	2.5% CI	97.5 % CI
Comparison model	Frequency (F)	1.44	1.43	1.46	1.41	1.40	1.43
	Predictability (P)	1.38	1.37	1.40	1.46	1.44	1.47
	Length (L)	1.56	1.54	1.58	1.52	1.50	1.53
	Position	1.18	1.17	1.20	1.15	1.14	1.16
	F x P	1.54	1.53	1.56	1.60	1.58	1.62
	F x L	1.21	1.20	1.23	1.19	1.18	1.20
	P x L	1.57	1.56	1.59	1.61	1.59	1.63
	F x Pos	2.21	2.18	2.24	2.00	1.97	2.02
	P x Pos	1.64	1.62	1.66	1.63	1.61	1.65
	L x Pos	1.95	1.92	1.97	1.91	1.89	1.94
	F x P x L	1.42	1.40	1.44	1.41	1.39	1.42
	F x P x Pos	1.81	1.79	1.83	1.67	1.65	1.69
	F x L x Pos	2.59	2.55	2.62	1.93	1.91	1.96
	P x L x Pos	2.00	1.97	2.03	1.80	1.78	1.83
	F x P x L x Pos	2.80	2.76	2.85	2.25	2.22	2.28
Line-final words	Frequency (F)	1.57	1.49	1.66	1.56	1.49	1.64
	Predictability (P)	1.39	1.33	1.47	1.39	1.34	1.46
	Length (L)	1.40	1.34	1.48	1.41	1.35	1.48
	F x P	1.58	1.50	1.67	1.61	1.54	1.69
	F x L	1.84	1.74	1.95	1.87	1.78	1.97
	P x L	1.55	1.47	1.63	1.56	1.49	1.64
	F x P x L	2.15	2.03	2.29	2.19	2.08	2.32

18 Additional Data Visualisation

19 Figure 2 depicts the frequency $\times$ length $\times$ position interaction for gaze duration in the Provo Corpus. For
 20 both intra-line and line-final words, it appears that word length effects are stronger for low-frequency words
 21 than for high-frequency words. Further inspection of the Figure suggests that the three-way interaction is
 22 driven by the effect of word length for low-frequency words being stronger for line-final relative to intra-line
 23 words.

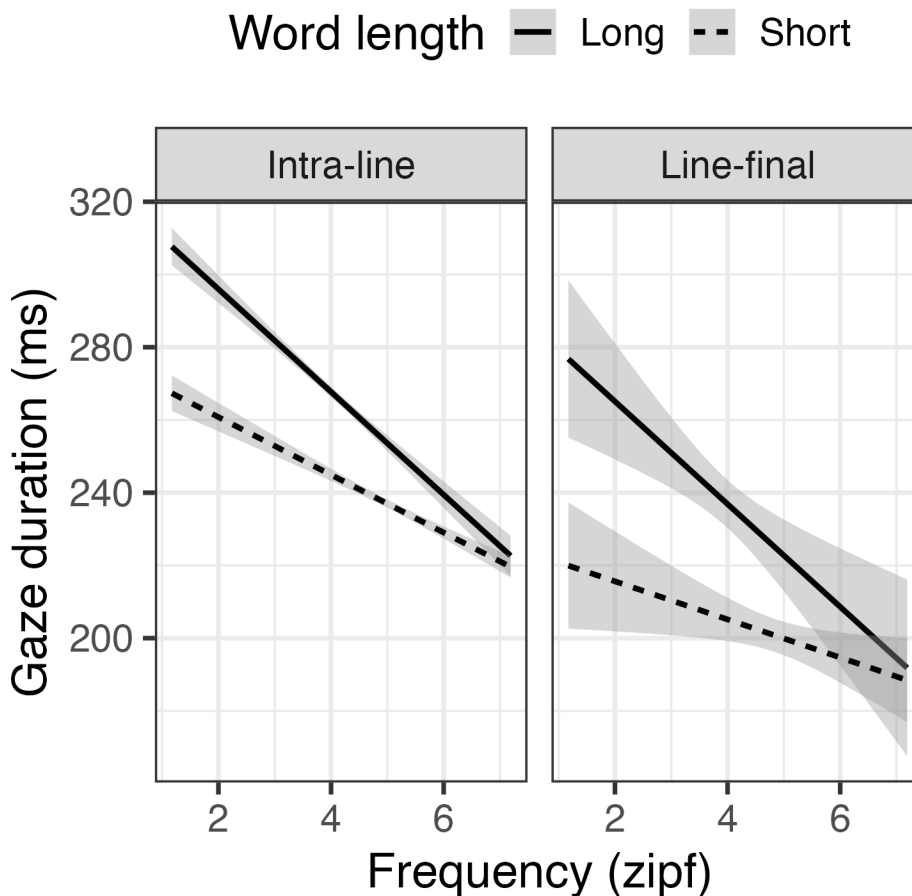


Figure 2: The frequency $\times$ length $\times$ position interaction for gaze duration in the Provo Corpus. The median of word length was used to categorise words as long and short, where the dashed line indicates gaze durations for short words and the solid line indicates gaze durations for long words. The gray bands represent the 95% confidence intervals.