

Correlation Analysis Summary

GitHub Link: [Available Here](#)

Part 1

Key Findings

meanItemRT vs word_length ($r = 0.280, p < 0.001$)

- Weak positive correlation: longer words are associated with slightly longer reading times

meanItemRT vs nItem ($r = 0.000, p = 0.921$)

- No correlation: word frequency has negligible effect on mean reading time

word_length vs nItem ($r = -0.295, p < 0.001$)

- Weak negative correlation: longer words tend to have lower frequency in the dataset

Word length is the only variable showing a significant relationship with reading time, though the effect is modest. Frequency (nItem) does not independently predict reading time in this dataset.

Part 2

Overall Model Comparison

- Model 1** (Freq + Length): $R^2 = 0.0857, MSE = 2059.10$
- Model 2** (-log(GPT3) + Length): $R^2 = 0.0299, MSE = 1407.34$

Model 1 with word frequency and length explains more variance in reading times.

Content vs Function Words

Content words:

- Best model: Freq + Length ($R^2 = 0.0910, MSE = 2084.98$)
- Outperforms -log(GPT3) + Length ($R^2 = 0.0361, MSE = 1410.36$)

Function words:

- Best model: -log(GPT3) + Length ($R^2 = 0.0526, MSE = 1099.35$)
- Notably better than Freq + Length ($R^2 = 0.0034, MSE = 1156.45$)

Word frequency predicts content word reading times, while predictability (GPT3 probability) better predicts function word reading times.

Part 3

Hypothesis 1: Root Frequency vs Surface Frequency

Model 1 (Surface): $RT \sim word_freq + word_length$

- $R^2 = 0.0782$, MAE = 32.15 ms, RMSE = 45.56 ms

Model 2 (Lemma): $RT \sim lemma_freq + lemma_length$

- $R^2 = 0.0840$, MAE = 32.12 ms, RMSE = 45.42 ms

Finding: Lemma frequency provides modest improvement ($R^2 + 0.58\%$), supporting the hypothesis that root frequency predicts reading times better than surface frequency. Word length remains the dominant predictor in both models.

Hypothesis 2: Pseudo-Affixed vs Real-Affixed Words

Category	Mean RT	SD	Avg Length	Avg Freq
Pseudo-affixed	368.89 ms	48.68	6.0	1.6
Real-affixed	352.36 ms	19.50	6.2	3.0

Difference: 16.54 ms (4.7% longer for pseudo-affixed)

- Cohen's d = 0.446 (small effect)
- $t(8) = 0.631$, $p = 0.546$ (not statistically significant)

Finding: Pseudo-affixed words show longer reading times, but the difference is not significant. The trend supports the hypothesis qualitatively, though lower frequency in pseudo-affixed words may be a confounding factor.

Overall Conclusion

Root frequency and morphological structure both influence reading times, but effects are modest compared to word length. The FOBS model demonstrates that lemmatization captures cognitive processing dynamics better than surface-level analysis.