

Diachronic Language Change Explains Apparent Age-Related Differences in Information-Theoretic Efficiency

Ellis Cain, Alton Chao, Rachel Ryskin (University of California, Merced)
ecain@ucmerced.edu

During production, speakers make choices which appear consistent with a pressure toward information-theoretic efficiency (ITE). For instance, some meanings can be expressed using multiple forms which differ in length (e.g., *A/C* and *air-conditioning*). Mahowald et al. (2013) previously showed that shorter forms were preferred (relative to long forms) in supportive than in neutral contexts. Previous work has argued that the ability to engage in prediction during language processing is reduced in older adults (Federmeier, 2010; Wlotko & Federmeier, 2012), which may influence their ITE during communication. Alternatively, age-related differences may be explained by exposure to changing language statistics (Ryskin & Nieuwland, 2023). Here, we examine how preferences differ across the lifespan and further explore how diachronic change in usage patterns (Michel et al., 2011) may explain these differences.

Methods: We recruited 126 English-speaking participants through Prolific (Age range: 20-60 years old, $M = 39.64$ y.o., 49% female). As in Mahowald et al. (2013), participants were presented with either a supportive or neutral sentence stem and had to choose between a short and long version of a word (e.g., 'A/C' or 'air-conditioning'; $n = 40$ word pairs). The context type of each newly written stem was verified using GPT-2 (Radford et al., 2019), such that the average surprisal of the full sentences for both short and long forms was lower in supportive ($M = 3.71$) than in neutral contexts ($M = 5.66$).

Results: Overall, the short form was more likely to be chosen in the supportive context, relative to the long form (Fig. 1). The frequencies of both the short and long forms varied over time (Fig. 2). We capture this change by the difference in short form frequencies between the 1960s and 2000s for each pair. Figure 3 shows the proportion of choosing short across the age range, grouped by change in short form frequency. Older adults (OA) appear less likely to choose the short form particularly for larger changes in frequency. We trained a Bayesian multilevel model¹ to predict whether participants chose the short form based on context, age, and change in frequency. Replicating Mahowald et al. (2013), participants were more likely to choose the short form when the context was supportive relative to neutral ($\beta_{\text{context}} = 0.38$, 95% CrI = [-0.04, 0.79]), though the credible interval of the effect includes zero. The model also confirmed that OA were less likely to choose the short form ($\beta_{\text{age}} = -0.51$, [-0.79, -0.25]). There was a modest interaction between age and context, with the effect of age being reduced in supportive relative to neutral contexts ($\beta_{\text{age}*\text{context}} = 0.08$, [-0.05, 0.20]). There was no main effect of frequency change, but there was an interaction with age ($\beta_{\text{age}*change} = -0.09$, [-0.19, 0.01]), such that OA became even less likely to choose the short form for larger changes in short form frequency. The 95% CrI for the interactions included zero. There was no evidence of a 3-way interaction.

Conclusion: Replicating and extending Mahowald et al. (2013), we found that lexical choices are driven by a pressure for information-theoretic efficiency, but the tendency to use shorter forms is increased in YA. We also found that this increased tendency may be tied to patterns of language change. When the short form was previously much lower in frequency than the long form, OA were less likely to use the short form, suggesting that OA demonstrate a similar pressure toward information-theoretic efficiency in communication as young adults but the lexical choices of YA may be influenced by usage statistics from recent decades. We plan to collect more data to replicate these findings.

References:

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and language*, 115(3), 149–161.

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: speakers choose shorter words in predictive contexts. *Cognition*, 126(2), 313–318.

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science (New York, N.Y.)*, 331(6014), 176–182.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners.

Ryskin, R., & Nieuwland, M. S. (2023). Prediction during language comprehension: what is next?. *Trends in cognitive sciences*, 27(11), 1032–1052. <https://doi.org/10.1016/j.tics.2023.08.003>

Wlotko, E. W., & Federmeier, K. D. (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology*, 49(6), 770–785.

Models:

1. Chose short ~ context * age * change + (1 + context * age | pair) + (1 + context * change | subject)

Figures:

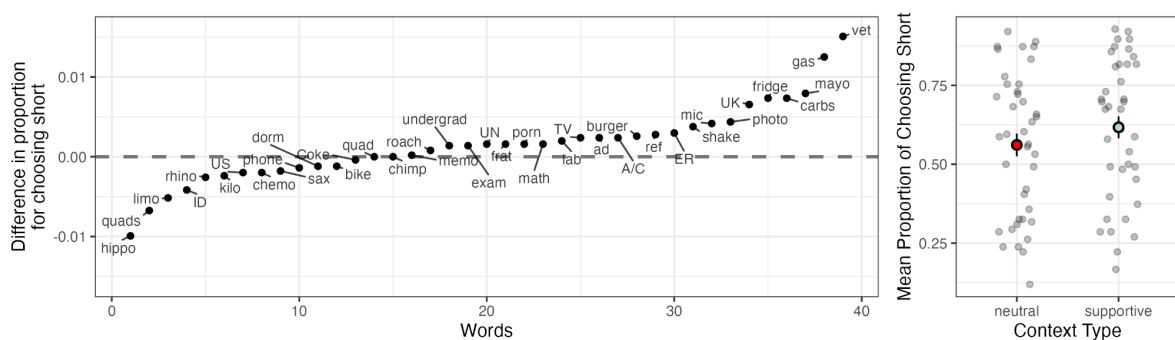


Figure 1: Left shows the difference in the proportion of choosing the short form in the supportive context relative to the neutral context, across the different word pairs. Right shows the overall average proportion of choosing the short form for each context type.

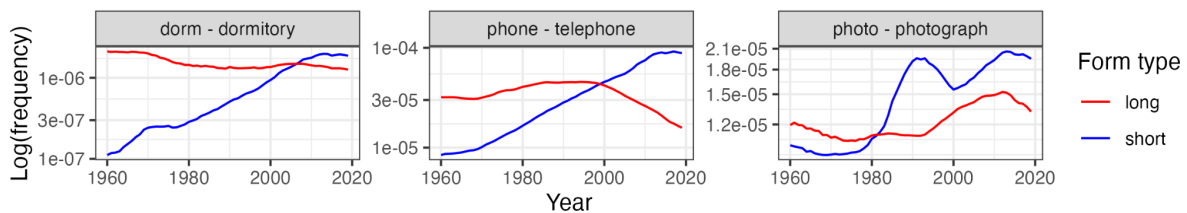


Figure 2: Diachronic trends in long and short form frequency for a small subset of the word pairs, gathered from the Google books corpus. For 31 out of 40 word pairs the short form was less frequent than the long form in the 1960s. 25 of the short forms increased in frequency between the 1960s and 2000s and 3 surpassed their long forms by the 2000s.

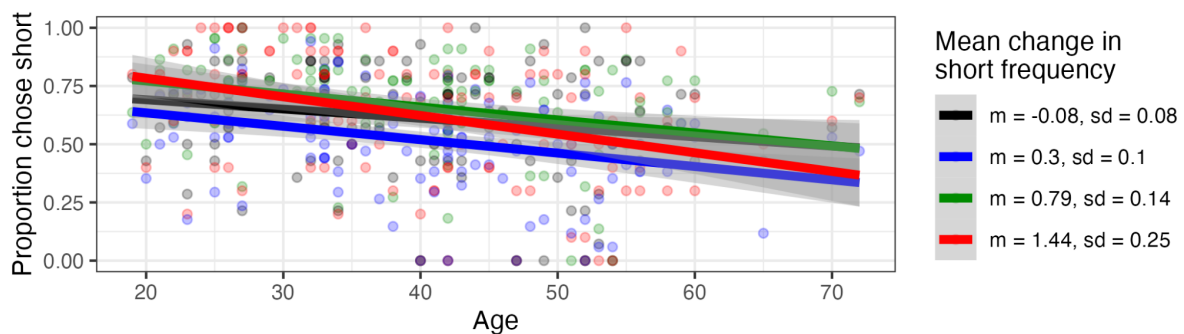


Figure 3: Proportion of choosing the short form in the supportive context, relative to the neutral context, by participant age. Color represents the bins for the average change in short form frequency.