

# Frequency-dependent regularization arises from a noisy-channel processing system

Zachary Houghton<sup>1</sup> & Emily Morgan<sup>1</sup>

<sup>1</sup> University of California, Davis

This is an abstract.

*Keywords:* keywords

## Introduction

Speakers often have great flexible in their choices to convey a meaning. For example, speakers are often confronted with many different ways to express the same meaning. A customer might ask whether a store sells “radios and televisions”, but they could have easily asked whether the store sells “televisions and radios.” However, despite conveying the same meaning, speakers sometimes have strong preferences for one choice over competing choices. For example, speakers prefer *men and women* to *women and men* (Benor & Levy, 2006; Morgan & Levy, 2016a), and this preference is more extreme for higher frequency items (e.g., *bread and butter*). That is, higher-frequency items typically have more polarized preferences (Liu & Morgan, 2020, 2021; Morgan & Levy, 2015, 2016b). Specifically, Morgan and Levy (2015) demonstrated that more frequent binomial expressions (e.g., *bread and butter*) are more strongly regularized (i.e., are preferred in one order overwhelmingly more than the alternative). Further, Liu and Morgan (2020) demonstrated that this phenomenon holds true for the dative alternation in English as well (*give him the ball* vs *give the ball to him*).

How does this polarization for high-frequency items arise? One possibility is that it occurs as a consequence of imperfect transmission between generations. For example, Morgan and Levy (2016b) demonstrated that in an iterated-learning paradigm (Real & Griffiths, 2009), this frequency-dependent regularization can arise from an interaction between a frequency-independent bias and transmission across generations.

However, it is unclear what the process driving the frequency-independent bias in language could be. One possibility is that this is a product of noisy-channel processing (Gibson, Bergen, & Piantadosi, 2013). That is, listeners are confronted with a great deal of noise in the form of perception errors (per-haps due to a noisy environment) and even production errors

(speakers don’t always say what they intended to) (Gibson et al., 2013). In order to overcome these errors, a processing system must take into account the noise of the system, and this claim is backed up by experimental evidence (e.g., Felty, Buchwald, Gruenenfelder, & Pisoni, n.d.). For example, Felty et al. (n.d.) demonstrated that when listeners do misperceive a word, the word that they believe to have heard tends to be higher frequency than the target word. Gibson et al. (2013) further demonstrated two key points: First, when people are presented with an implausible sentence (e.g., *the mother gave the candle the daughter*), participants are more likely to interpret the sentence as being the plausible version (e.g., *the mother gave the candle to the daughter*) if there is increased noise (e.g., by adding errors to the filler items). Second, increasing the likelihood of implausible events increases the rate of interpretations of the implausible version of the sentence. Indeed these results are consistent with the predictions of their noisy-channel processing model (Equation 1), which operationalizes noisy-channel processing as a Bayesian process where a listener computes the probability that their perception matches the speakers intended utterance as being proportional to the likelihood of the intended utterance times the probability of the intended utterance being corrupted to the perceived utterance.

$$P(S_i|s_p) \propto P(S_i)P(S_i \rightarrow S_p) \quad (1)$$

It is possible that the frequency-dependent regularization we see in Morgan and Levy (2016b) is a product of listeners’ noisy-channel processing. Thus, the present study examines whether Gibson et al. (2013)’s noisy-channel processing model can account for frequency-dependent regularization.

## Dataset

Following Morgan and Levy (2016b), we use Morgan and Levy (2015)’s corpus of 594 binomial expressions. This corpus has been annotated for various phonological, semantic, and lexical constraints that are known to affect binomial ordering preferences. The corpus also includes estimated generative preferences for each binomial (i.e., compositional ordering preferences, estimated from the above constraints) and observed binomial orderings.

---

Correspondence concerning this article should be addressed to Zachary Houghton, . E-mail: znoughton@ucdavis.edu

### Model

Following Real and Griffiths (2009), we use a 2-alternative iterated learning paradigm. A learner hears  $N$  tokens of a binomial expression and then produces  $N$  tokens for the next generation. After hearing  $N$  tokens, they will infer the probability,  $\theta$ , of the alphabetical form of the binomial. Using that probability, they will then produce  $N$  tokens for the next generation, and this process will continue iteratively.

The prior probability of a binomial's ordering is operationalized using the beta distribution (following Morgan & Levy, 2016b). Specifically, we treat the generative preference for a given binomial as the prior for our model. This is operationalized using the beta distribution with  $\mu_{prior}$ , which is the generative preference for a given binomial, and  $\nu$ , which determines how tightly clustered around the means the draws are. In other words, a higher  $\nu$  corresponds to a stronger belief in the prior probability. Thus the prior probability for a given binomial in its alphabetical form is estimated as:

$$P(\theta_{prior}) = \frac{\mu_{prior} \cdot \nu}{(\mu_{prior} \cdot \nu) + (1 - \mu_{prior}) \cdot \nu} \quad (2)$$

and the probability of hearing it in non-alphabetical form is  $1 - P(\theta_{prior})$ .

This is then used as the prior in Gibson et al. (2013)'s noisy-channel processing model, such that if a listener hears the alphabetical form of a binomial they update the probability of hearing it in alphabetical form according to the following parametrization,

$$\hat{p}(\alpha) \propto \frac{\mu_{prior} \cdot \nu}{(\mu_{prior} \cdot \nu) + (1 - \mu_{prior}) \cdot \nu} \cdot p_{noise} \quad (3)$$

where  $1 - p_{noise}$  is a fixed parameter for every binomial. They also update the probability of hearing the non-alphabetical form:

$$\hat{p}(-\alpha) \propto 1 - \frac{\mu_{prior} \cdot \nu}{(\mu_{prior} \cdot \nu) + (1 - \mu_{prior}) \cdot \nu} \cdot (1 - p_{noise}) \quad (4)$$

If the listener hears the non-alphabetical form of the binomial, they update using the same equations, but with a slight change to the noise parameter:

$$\hat{p}(\alpha) \propto \frac{\mu_{prior} \cdot \nu}{(\mu_{prior} \cdot \nu) + (1 - \mu_{prior}) \cdot \nu} \cdot p_{noise} \quad (5)$$

and,

$$\hat{p}(-\alpha) \propto 1 - \frac{\mu_{prior} \cdot \nu}{(\mu_{prior} \cdot \nu) + (1 - \mu_{prior}) \cdot \nu} \cdot (1 - p_{noise}) \quad (6)$$

Once the learner finishes hearing  $N$  tokens, they then produce  $N$  tokens generated binomially, where  $\theta_1$  is their inferred probability of the alphabetical form of the given binomial:

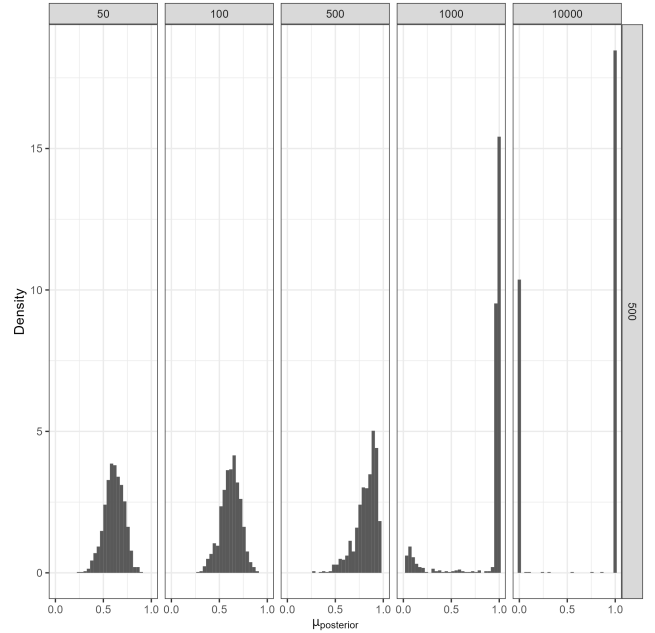
$$P(x_1|\theta_1) = \binom{N}{x_1} \theta_1^{x_1} (1 - \theta_1)^{N-x_1} \quad (7)$$

For the first generation of learners,  $\theta_1$  is set to 0.5.

Finally, when the learner produces the  $N$  tokens, there is a possibility the speaker will make an error. This is also generated binomially, with  $\theta_1$  being a fixed parameter, which is the probability that the producer makes an error.

### Results

First we show that the noisy-channel processing model can predict frequency-dependent regularization across generations, presented in Figure 1.



*Figure 1.* A plot of the distribution of simulated binomials at the 500th generation, varying in frequency. The top value represents  $N$ . On the x-axis is the predicted probability of producing the binomial in alphabetical form. On the y-axis is probability density. Speaker noise was set to 0.001, listener noise was set to 0.01, the generative preference was 0.6, and  $\nu$  was set to 10. 1000 chains were run. Note how for the binomials with large  $N$ , the ordering preferences tend to be more extreme.

Interestingly this regularization disappears if the listener's noise parameter is less than or equal to the speaker's noise parameter (Figure 2).

### Conclusion

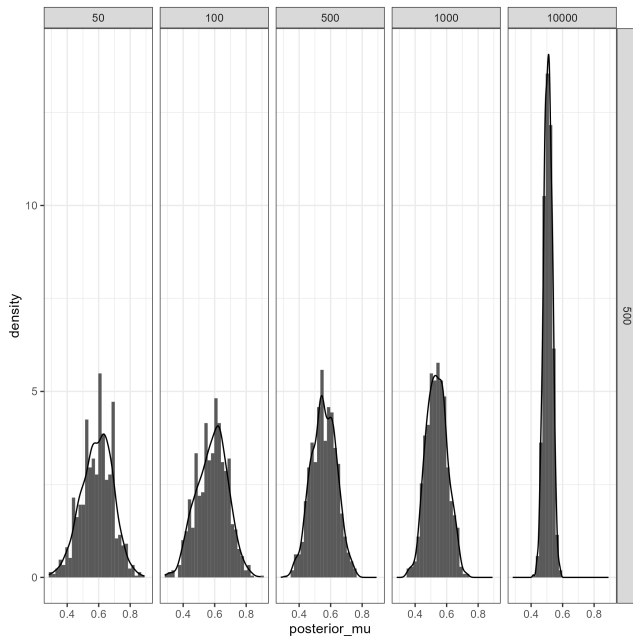


Figure 2. A plot of the distribution of simulated binomials at the 500th generation, varying in frequency. The top value represents N. On the x-axis is the predicted probability of producing the binomial in alphabetical form. On the y-axis is probability density. Speaker noise was set to 0.01, listener noise was set to 0.001, the generative preference was 0.6, and nu was set to 10. 1000 chains were run. Note how regularization does not appear to be present in this graph.

## References

- Benor, S. B., & Levy, R. (2006). The chicken or the egg? A probabilistic analysis of english binomials. *Language*, 233278. Retrieved from [https://www.jstor.org/stable/4490157?casa\\_token=nPxM9z86cbUAAAAA:](https://www.jstor.org/stable/4490157?casa_token=nPxM9z86cbUAAAAA:)
- cLIZ78yyFzblh6KJgwHgexpDRnUfo1wHk1\_rFYxcLnsARSdBKtGIHkm75tgFQH09tQtWPE92pt\_N9sTWdnASq45rsS83g07MWkgI2qAoMcSFJZ2WgDE
- Felty, R. A., Buchwald, A., Gruenenfelder, T. M., & Pisoni, D. B. (n.d.). *Misperceptions of spoken words: Data from a random sample of american english words*. Retrieved from <https://www.robfelty.com/academic-files/docs/FeltyEtAl2013.pdf>
- Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, 110(20), 8051–8056. <https://doi.org/10.1073/pnas.1216438110>
- Liu, Z., & Morgan, E. (2020). *Frequency-dependent regularization in constituent ordering preferences*. Retrieved from <https://www.cognitivesciencesociety.org/cogsci20/papers/0751/0751.pdf>
- Liu, Z., & Morgan, E. (2021). *Frequency-dependent regularization in syntactic constructions*. 387389. Retrieved from <https://aclanthology.org/2021.scil-1.41.pdf>
- Morgan, E., & Levy, R. (2015). *Modeling idiosyncratic preferences : How generative knowledge and expression frequency jointly determine language structure*. 1649–1654.
- Morgan, E., & Levy, R. (2016a). Abstract knowledge versus direct experience in processing of binomial expressions. *Cognition*, 157, 384–402. <https://doi.org/10.1016/j.cognition.2016.09.011>
- Morgan, E., & Levy, R. (2016b). Frequency-dependent regularization in iterated learning. *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANGX11)*, (2015).
- Real, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, 111(3), 317–328. <https://doi.org/10.1016/j.cognition.2009.02.012>