

Chunking explains frequency-dependent regularization

Charles Torres (chjtorres@ucdavis.edu) and Emily Morgan

University of California, Davis

Introduction: Previous work has shown that the ordering preferences for binomial expressions (e.g. *bread and butter* vs. *butter and bread*) incorporate both generative knowledge (e.g. short-before-long) and item-specific knowledge [3]. A component of this item-specific knowledge is a tendency towards extreme ordering preferences for frequent expressions (e.g. *facts and techniques* > *techniques and facts* but *bread and butter* >>> *butter and bread*), a pattern known as frequency-dependent regularization. However, why this regularization occurs is still an open question, standard models of language change (in which an ideal learner attempts to mirror the distribution of its input) do not recreate this phenomenon. We demonstrate via a novel computational model on an English corpus that a learner which incorporates frequency-dependent “chunking” recreates frequency-dependent regularization with qualitatively similar patterns as in natural data.

Background: Binomial ordering preferences can be partially predicted by a number of variables dependent on the phrase constituents. However, even when incorporating many generative preferences, item-specific preferences emerge where speakers strongly prefer orders at rates above what would be expected from the general predictors alone (e.g. fixed expressions like *salt and pepper*). Given data which is biased in this way an ideal learner should recreate the distributions they are exposed to. However, why would item specific preferences distinct from these generative preferences emerge to begin with? Previous models that seek to explain this emergence have incorporated a bias which assigns greater probability to extreme ordering preferences [4], but this bias is simply stipulated. We seek to explain this bias as a natural result of a model incorporating usage-based chunking [1, 2].

Method: We model language change via an iterated learning model with learners inferring a policy p relating meanings m to binomial ordering probabilities from provided data. After learners infer probabilities, they then use this to generate a new dataset for the next generation.

Data: The possible utterances for each learner are English binomials from Morgan & Levy [3] as well as the individual nouns used in these binomials, scaled to realistic estimates for the numbers of nouns and binomials a college-age speaker has experienced. Each unordered binomial expression, with meaning we denote $m = \{n_1, n_2\} = \{n_2, n_1\}$, has an alphabetical ordering preference dictated by its general features $g(\alpha|\{n_1, n_2\})$. We generate seed data from this corpus by first sampling meanings m and then, in the case of binomial expressions, sampling orders $\alpha(m)$.

Model Learner: Central to our results is the behavior of our model learner. We model a learner that upon encountering a binomial phrase with meaning $\{n_1, n_2\}$ in order α at least k times stores the ordered expression in their lexicon as a chunk c_α . When the learner creates the next generation’s training data the probability for generating orders α is $p(\alpha | \{n_1, n_2\}) = p(\{n_1\})p(\{n_2\})g(\alpha|\{n_1, n_2\}) + p(c_\alpha)$ where p are the maximum likelihood estimates of the meanings for individual words and/or binomials from the learner’s training data (with binomial probabilities marginalized over both possible orders). For the purposes of calculating p the chunks are considered not to contribute to the counts of their constituent nouns. For the results presented here we perform 20 iterations and look at the distribution learned by the final generation.

Results and Discussion: Results for the best performing value investigated, $k = 75$, show the emergence of frequency dependent regularization: for the highest frequency binomial phrases relative ordering preferences drift towards the extremes at a rate stronger than generative preferences would predict (Fig. 1). A one-sided t-test shows a slope significantly greater than 1.0 ($t = 5.15, p < .001$). In (Fig. 2) we show that we get a trimodal distribution also found in naturalistic data. Our results suggest that the chunking of input is a reasonable explanation for the emergence of frequency regularization in natural language, especially for binomial expressions.

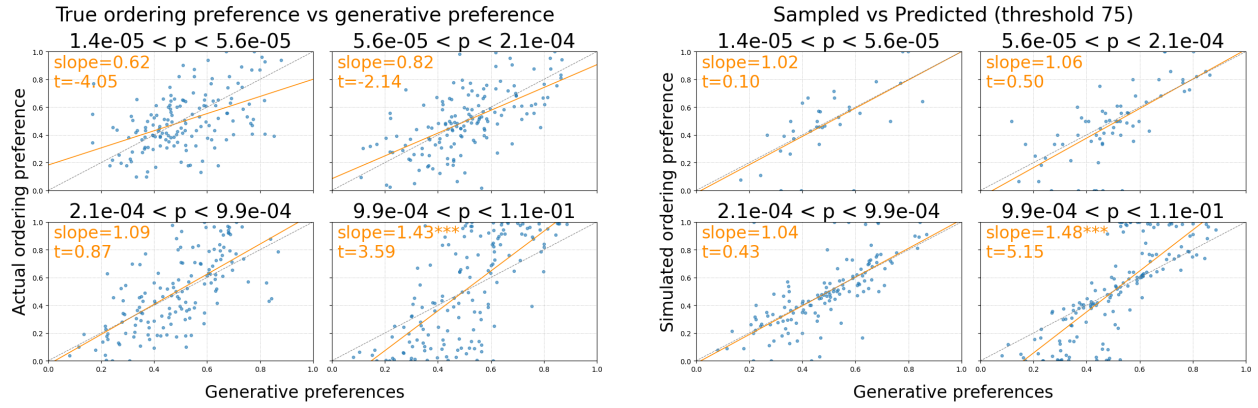


Figure 1: **Left:** Predicted generative ordering preferences plotted against actual frequency they are observed at. **Right:** Predicted generative ordering preferences against our simulated data after 20 generations ($k = 75$). Note, the most frequent quartile (bottom right) show a similar extreme ordering preference above generative preference predictions, that is a slope > 1 . This is because chunking pushes probabilities towards extreme arrangements (0 and 1). Some data was not observed after iterated learning leading to missing samples, especially in the least frequent quadrant.

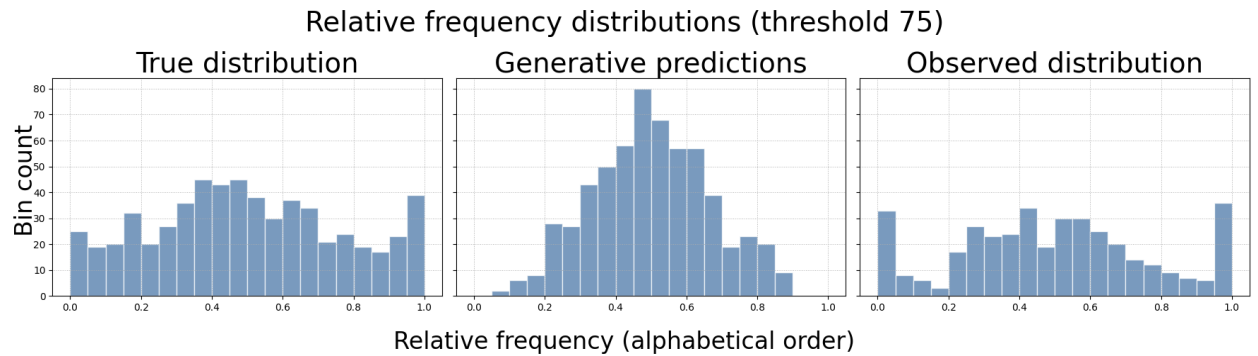


Figure 2: **Left:** Histogram of real human data. **Middle:** Histogram predicted by generative preferences alone. **Right:** Histogram after 20 generations of iterated learning using a chunking method.

References

- [1] Beckner, C. & Bybee, J. (2009). A usage-based account of constituency and reanalysis. *Language Learning*, 59, 27–46.
- [2] Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in second language acquisition*, 24(2), 215–221.
- [3] Morgan, E. & Levy, R. (2015). Modeling idiosyncratic preferences: How generative knowledge and expression frequency jointly determine language structure. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 37.
- [4] Morgan, E. & Levy, R. (2016). Frequency-dependent regularization in iterated learning. In *The Evolution of Language: Proceedings of the 11th International Conference (EVLANG11)*.