

## A COGNITIVE BIAS FOR ZIPFIAN DISTRIBUTIONS? UNIFORM DISTRIBUTIONS BECOME MORE SKEWED VIA CULTURAL TRANSMISSION

Shufaniya, Amir<sup>1</sup>, Arnon, Inbal 2\*<sup>1</sup>

\*Corresponding Author: [inbal.arnon@mail.huji.ac.il](mailto:inbal.arnon@mail.huji.ac.il)

<sup>1</sup>Psychology Department, Hebrew University, Jerusalem, ISRAEL

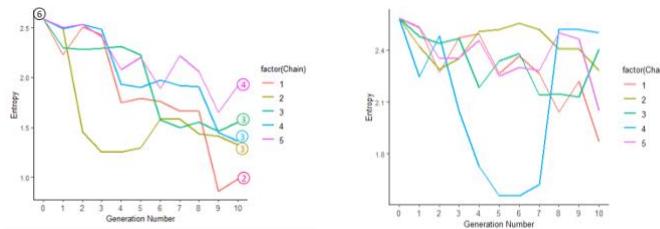
Despite the many differences between them, human languages share certain similarities. One of them is the Zipfian distribution of word frequencies (Zipf, 1949). Across languages, word frequency follows a power law distribution with many low frequency words, few highly frequent words and a non-linear decrease in frequency (e.g., Piantadossi, 2014). The presence of such skewed distributions is a recurring feature of natural language, with ongoing debate about their source and whether they reflect foundational properties of human language/cognition/communication (e.g., Ferrer i Cancho & Sole, 2003). Here, we use an iterated learning paradigm – which shows how weak individual biases impact language structure over time (e.g., Kirby et al. 2008) - to ask whether learners have a cognitive bias for Zipfian distributions, reflected in a tendency to shift uniform distributions into skewed ones during cultural transmission.

Recent work suggests the recurrence of Zipfian distributions in language may be driven (at least in part) by learnability pressures (e.g., Bentz et al., 2017; Lavi-Rotbain & Arnon, 2020, 2022). In particular, the lower unigram entropy of such distributions (compared to uniform and less skewed ones) may benefit learning (Ferrer i Cancho, 2018; Lavi-Rotbain & Arnon, 2022). Indeed, growing evidence suggests Zipfian distributions facilitate word segmentation and learning (Kurumada, Meylan & Frank, 2013; Hendrickson & Perfors, 2019), and that they do so because of their lower unigram entropy (Lavi-Rotbain & Arnon, 2022).

Importantly, existing findings do not tell us whether there is a cognitive preference for skewed distributions: the facilitation may reflect learners' prior experience with such distributions rather than a cognitive preference for them. Here, we explore the existence of such a preference by asking if learners will change their input to make it more skewed. In the first study, we show that learners are biased to produce skewed word distributions in telling a novel story: participants were given a short prompt including six nonce words introduced as names, nouns or verbs (Noun-prompt: "*John has a store with six objects in it: a plizet, a nilbo, a skiger, a vamey, a chila, and a fengle. Write a short story about John's day*"). Aggregated over participants, the nonce words followed a Zipfian

distribution for all three parts of speech (high fit of summed frequency by rank to a power law distribution for nouns ( $R^2=.98$ ), verbs ( $R^2=.98$ ), and names ( $R^2=.97$ )).

In the second study, we ask if this bias leads to a shift from uniform distributions towards more skewed ones using an iterated learning paradigm. We exposed the first learner to a story where the six nonce words appeared equally often, and asked them to re-tell it. Their output served as input for the next learner, and so on for a chain of ten learners (or "generations"). We looked at two conditions: (1) where the number of unique words could decrease (five chains) and (2) where it couldn't (participant had to use all words, five chains). Over time, word distributions became more skewed (as measured by lower levels of entropy), even when the number of word types remained constant (significant in model comparisons of mixed-effect regression models, also when excluding differently shaped chain 4). These findings suggest speakers have a cognitive bias for skewed distributions that gets amplified over time, lending support to the idea that their recurrence in language may be driven in part by learnability pressures.



**Figure 3.** The decrease in entropy levels over generations in each of the five diffusion chains in condition 1 (left) and 2 (right)

#### References

- Ferrer i Cancho, R., & Sole, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, 100(3), 788–791.
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31), 10681–10686.
- Kurumada, C., Meylan, S. C., Frank, M. C. (2013). Zipfian frequency distributions facilitate word segmentation in context. *Cognition*, 127, 439–453.
- Lavi-Rotbain, O. & Arnon, I. (2022). The learnability consequences of Zipfian distributions, *Cognition*, 223, xx-x
- Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, 21(5), 1112–1130.
- Retrieved from
- Zipf, G. K. (1949). Human behavior and the principle of least effort. In *Human behavior and the principle of least effort*. Oxford, England