Journal of Language Evolution, 2022, 59–80
https://doi.org/10.1093/jole/lzac005
Advance Access Publication Date: 9 July 2022
Research article

A Cognitive Bias for Zipfian Distributions? Uniform Distributions Become More Skewed via Cultural Transmission

Amir Shufaniya * and Inbal Arnon

Department of Psychology, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem 9190501, Israel

*Corresponding author: amir.shufaniya@mail.huji.ac.il

Abstract

There is growing evidence that cognitive biases play a role in shaping language structure. Here, we ask whether such biases could contribute to the propensity of Zipfian word-frequency distributions in language, one of the striking commonalities between languages. Recent theoretical accounts and experimental findings suggest that such distributions provide a facilitative environment for word learning and segmentation. However, it remains unclear whether the advantage found in the laboratory reflects prior linguistic experience with such distributions or a cognitive preference for them. To explore this, we used an iterated learning paradigm—which can be used to reveal weak individual biases that are amplified overtime—to see if learners change a uniform input distribution to make it more skewed via cultural transmission. In the first study, we show that speakers are biased to produce skewed word distributions in telling a novel story. In the second study, we ask if this bias leads to a shift from uniform distributions towards more skewed ones using an iterated learning design. We exposed the first learner to a story where 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'). Over time, word distributions became more skewed (as measured by lower levels of word entropy). The third study asked if the shift will be less pronounced when lexical access was made easier (by reminding participants of the novel word forms), but this did not have a significant effect on entropy reduction. These findings are consistent with a cognitive bias for skewed distributions that gets amplified over time and support the role of entropy minimization in the emergence of Zipfian distributions.

Key words: Zipfian distributions; iterated learning; cultural transmission; learnability

1. Introduction

Despite the many differences between them, human languages share certain similarities. These similarities can provide a window onto our shared cognition and the ways in which cognitive biases impact the distribution of linguistic properties across the world's languages (e.g., Christiansen and Chater 2008; Culbertson and Kirby 2016; Gibson et al. 2019). One striking commonality between languages is the way word frequencies are distributed. Across languages, word frequencies follow a Zipfian distribution (Zipf 1936), showing a power-law relation between a word's frequency and its frequency rank. This reflects the presence of a small number of high frequency words, a large number of low frequency words, and the fact that frequency does not decrease in a linear fashion (i.e., the most frequent word is twice as frequent as the second and so on). The resulting distribution is right skewed and the appearance of words in it is more predictable compared with a uniform distribution, where all words are equally probable. Unlike other cross-linguistic tendencies, the presence of Zipfian distributions seems to be a foundational and recurrent property of the world's languages: it has been found consistently across many languages (Bentz and Ferrer-i-Cancho 2016; Ferrer-i-Cancho 2005; Mehri and Jamaati 2017; Piantadosi 2014), across linguistic categories (Piantadosi 2014), and in child-directed speech (Baixeries et al. 2013; Hendrickson and Perfors 2019; Lavi-Rotbain and Arnon 2019a, 2022).

Given its seeming universality, the source of Zipfian distributions in language has been studied extensively (see Ferrer-i-Cancho et al. (2022); Piantadosi (2014); and Semple et al. (2022) for reviews). Various explanations have been offered for the origin of these distributions, with ongoing debate about whether they reflect something meaningful about the nature of language and/or human cognition. The question arises since Zipfian distributions are also found across the physical world, where they are thought to reflect general mathematical principles not unique to language (e.g., scale-invariance; Chater and Brown 1999). However, their recurrence in language—a human creation—may nevertheless reflect fundamental properties of human cognition and communication, as the same pattern (power-law distributions) may have a different source in complex (and living) systems. Supporting their cognitive and/or communicative origin, Zipfian distributions have been suggested to provide an optimal trade-off between speaker and listener effort (Ferrer-i-Cancho and Sole 2003; Ferrer-i-Cancho et al. 2022); to be a form of optimal coding that minimizes cognitive costs (Ferrer-i-Cancho 2018); to enable an efficient hierarchical organization of word meaning (Manin 2008); and to facilitate communicative efficiency (Mahowald et al. 2020). While there is no agreed account of their source, their presence is often attributed to an interplay between multiple competing and converging cognitive and communicative pressures (see Semple et al. (2022) for one such comprehensive account).

Several recent proposals suggest that the recurrence of Zipfian distributions in language may be driven (at least in part) by learnability pressures (Bentz et al. 2017; Coupé et al. 2019; Lavi-Rotbain and Arnon 2020, 2022). One such account proposes that Zipf's law emerges from a tension between two principles: mutual information minimization and word entropy minimization, which together work to reduce the cognitive costs of communication (Ferrer-i-Cancho 2018; Ferrer-i-Cancho et al. 2022). An environment with lower word entropy is more predictable and may consequently be easier to learn from (e.g.,Clark 2013). This lower entropy/greater predictability could facilitate word segmentation and learning by making it easier to predict upcoming words and to use higher frequency words as anchors for learning lower frequency ones (Lavi-Rotbain and Arnon 2019a,b, 2020, 2022). While much work on the source of Zipfian distributions has centered on mathematical and corpus-based investigations, a growing number of studies test the postulated learnability advantage of such distributions experimentally. Across several tasks, learning was shown to be as good (if not better) in Zipfian distributions compared with uniform ones (word segmentation: Kurumada et al. (2013); grammatical category learning: Schuler et al. (2017); and cross situational word learning: Hendrickson and Perfors (2019)).

While these studies suggest that Zipfian distributions can improve learning, they do not spell out what exactly about the distribution is facilitative. A recent paper relates the learning advantage to word entropy minimization (Ferrer-i-Cancho 2016, 2018) and to the resulting lower unigram entropy of such distributions. Using a combination of corpus-based and experimental methods, the study examines the impact of unigram entropy reduction on word segmentation (Lavi-Rotbain and Arnon 2022). As a first step, Lavi-Rotbain and Arnon quantified unigram entropy across fifteen languages using the information-theoretic notion of efficiency, which indicates how predictable a distribution is relative to a uniform one with the same set size. They found that different languages have similar unigram entropy. They then modified existing word segmentation paradigms by changing the uniform distribution often used into a skewed distribution with language-like unigram entropy. Indeed, word segmentation was facilitated for both children and adults in language-like unigram entropy, compared with both uniform and less skewed distributions. Similar effects were found in the visual domain where Zipfian distributions facilitated segmenting a continuous stream of visual images into recurring units (Lavi-Rotbain and Arnon 2022).

Taken together, these findings suggest that exposure to Zipfian distributions has facilitative learnability consequences across various tasks. However, they do not tell us whether there is an underlying preference for such distributions and where this preference might come from. One possibility is that the beneficial effect on learning reflects prior experience and the fact that the experimental input is now more similar to real-world learning environments: That is, learning may be facilitated in Zipfian distributions because that is what learners are used to hearing. Alternatively, the facilitation could reflect the minimized cognitive costs associated with this distribution (Ferrer-i-Cancho 2018; Ferrer-i-Cancho et al. 2022; Lavi-Rotbain and Arnon 2022). In this study, we want to explore this second possibility by asking if learners will actively change their input to make it more skewed (thereby reducing word entropy). Finding they will, would provide preliminary support for the idea that such a preference plays a role in the propensity and preservation of Zipfian distributions in language and support accounts that attribute the propensity of the distribution to reducing cognitive costs (Ferrer-i-Cancho et al. 2022; Lavi-Rotbain and Arnon 2022). In raising this possibility, we build on the notion of cultural transmission and the way weak individual biases can impact language structure over time (Kirby et al. 2008; Kirby et al. 2014). During cultural transmission, an individual acquires behavior by observing a similar behavior in another individual who acquired it in the same way. Individual biases become amplified over time through a repeated cycle of use, observation, and induction, leading to changes in the learned behavior (Kirby et al. 2014).

While such processes are hard to examine in real-time (since they occur over a long time-span and we do not have access to much of the relevant data), their effect on language structure has been investigated computationally and experimentally using iterated learning (Kirby et al. 2014; Tamariz and Kirby 2016). Iterated learning studies simulate the process of cultural transmission by using a diffusion chain paradigm, which examines how a learned behavior changes when it is transmitted between learners (much like the game of 'Chinese whispers'). In this paradigm, the first participant is exposed to a target behavior they need to reproduce. The output behavior produced by the first participant becomes the input behavior for the second participant, and so on for several 'generations' of learners. Mathematical and computational models of iterated learning show that the structural properties of language can be shaped over time to better-fit individual learning biases (e.g., Griffiths and Kalish 2007; Kirby et al. 2007). In a highly influential study, Kirby et al. (2008) used such a paradigm experimentally with adult participants to show that non-structured artificial languages (with random mappings between forms and meanings) can become more learnable and more structured over time: Over ten generations, the languages became easier to learn and developed consistent mappings between meaning and form. These findings have been replicated for various linguistic and non-linguistic behaviors (see Tamariz and Kirby (2016) for a review) and across different populations of human and non-human learners (children: Kempe et al. 2015; Raviv and Arnon 2018; songbirds: Fehér et al. 2017; primates: Claidière et al. 2014). They show that structure can emerge through cultural transmission and that culturally transmitted languages are impacted by individual's cognitive and communicative biases. As a result, the product of transmission can provide insight about the biases themselves.

2. The current study

In the current study, we extend the iterated learning paradigm to study the emergence of Zipfian distributions in language. We base our work on the two main findings discussed above: (1) Zipfian distributions provide a facilitative environment for learning because of their lower unigram entropy and (2) languages transmitted through iterated learning change in ways that reflect learning biases and preferences. Building on these findings, we ask whether learners will turn uniform word distributions into more skewed ones through the process of cultural transmission, and in doing so, reveal a cognitive preference for more skewed distributions. As a first step, we ask whether speakers are biased to produce skewed word distributions when telling a story (Study 1). If so, we want to see whether this individual bias gets amplified over time, resulting in a shift from an initial uniform distribution to a more skewed one (Study 2). Finally, we ask how the preference for skewed distributions is impacted by lexical access demands (Study 3). Study 1 is motivated by findings mentioned in Piantadosi (2014) where speakers were asked to tell a story about six unfamiliar aliens to see whether the resulting word distribution is Zipfian. Indeed, even though all the characters were equally novel and were each introduced only once, the resulting frequency distribution (averaged over participants) was near-Zipfian. That is, speakers tended to mention some characters more often than others, resulting in a frequency–rank relation that showed a good fit to a power law distribution. This experiment provides preliminary evidence that speakers have a preference for Zipfian distributions in language production and also provide support for the principle of (word) entropy minimization (e.g., Ferrer-i-Cancho 2018). However, since only names were used, the results may have reflected a unique narrative-related need to choose a salient protagonist for the story. In Study 1a, we replicated the effect for names and in

Study1b, we generalized beyond names by presenting the novel words as nouns or verbs instead. In both parts of Study 1, we find that speakers produce highly skewed distributions, with a good fit to the power law distribution, across parts of speech.

Study 1 shows that speakers have a preference for skewed word distributions in language production, but it does not tell us whether this preference can shift word distributions over time. In Studies 2 and 3, we test the hypothesis that speakers will transform uniform distributions into skewed ones through the process of cultural transmission. If they do, this would illustrate how the production bias we observe at the individual level can get amplified over time to shift the input toward a more skewed distribution. To test this, we presented learners with a story including six novel words appearing equally often. We then asked participants to re-tell the story after reading it and used their version as the input for the next learner, for ten generations of learners in five different diffusion chains. We wanted to see whether the initial uniform distribution of the six words become more skewed (and hence more predictable) over time. We evaluated the predictability of the six-word distribution using entropy, which quantifies the information content of a random variable (the amount of uncertainty) and is lower when the variable is more predictable (Shannon 1948), see Equation (1) which shows the entropy of a given set of size N.

$$-\sum_{i=1}^{N} p(x_i) * \log_2 p(x_i). \tag{1}$$

We predict that the unigram entropy of the six words will become lower over generations. Because we are only looking at the distribution of six words in a short story, we do not have the sufficient number of word types or tokens to reliably estimate whether the distribution becomes more Zipfian over time. Instead, we ask whether the distribution becomes more skewed (lower unigram entropy) over time.

In Study 2, we had participants retell a story under two conditions. In the no-type-reduction condition, we allowed the transfer of stories that did not include all six novel words. In the allow-type-reduction condition, we required participants to use all novel words at least once, keeping the number of word types constant across generations. The comparison of these two conditions is important since entropy is inherently impacted by the number of unique word types (an environment is less predictable when it has more elements): For example, the unigram entropy of six words appearing in a uniform distribution is higher than the unigram entropy of four words appearing in a uniform distribution, even though both are not skewed. To preempt the results, distributions became more predictable over time in both conditions, even when the number of word types was not reduced. Study 3 asked how the increase in skew we found is impacted by lexical access: whereas in Study 2 participants had to remember the novel labels, here, we provided them with the labels during the re-telling. Our hypothesis was that part of the increase in skew may be related to difficulty with accessing/remembering the lower frequency forms. If this is the case, then that the increase in skew should be less pronounced when speakers do not have to remember the words. However, we did not find clear evidence for this hypothesis: the decrease in entropy was less significant than in the previous study, but the interaction between the two was not significant, meaning we cannot draw strong conclusions about the role of lexical access difficulty in the emergence of skewed distributions. Taken together, the findings support the presence of a cognitive preference for skewed distributions that is amplified over time.

3. Study 1

3.1 Method

In Study 1, participants were presented with a short prompt and asked to tell a story about six novel words. We then examined the distribution of the six words in their stories. The study aimed to replicate and expand an experiment briefly described in Piantadosi (2014) that examined speakers' preference for near-Zipfian distributions in language production. Study 1a was modeled closely after Piantadosi's experiment, presenting the novel words as names and using the same story length, while Study 1b examined the effect for two additional grammatical categories, nouns and verbs, in a shorter story. This was done to rule out the possibility of a unique narrative-related effect for names: choosing a salient protagonist for the story and then naturally mentioning this character more often. Finding that the resulting word distributions are near-Zipfian would indicate that speakers have a preference for skewed distributions in production. Furthermore, if similar results are obtained when using names, verbs, and nouns, then we may rule out a unique narrative-related effect for names, allowing us to use any of the three grammatical categories for Studies 2 and 3.

3.2 Participants

Eighty adult native English speakers (18 years old and above) from the USA participated in Study 1: twenty in Study 1a and sixty in Study 1b (20 in each of three conditions: nouns, names and verbs). The studies were conducted on Amazon's Mechanical Turk (MTurk) online platform. The tasks were published one after the other, without overlap, so that participants could only participate in one of the versions. Participants had to meet the following inclusion criteria: (1) located in the USA, (2) HIT (task submission) approval percentage >95% (to validate participants' reliability), and (3) being native English speakers (as verified by several English grammar questions). In addition, each participant had to read and sign our Explanation Sheet and Informed Consent Form in order to participate. Participants who met the inclusion criteria and followed the task's instructions received payment in return for their participation (calculated in compliance with the US federal living wage): $7 in Study 1a and $4 in Study 1b (participants were asked to tell longer stories in Study 1a, hence the payment difference).

3.3 Materials

All tasks were created using the MTurk platform and consisted of two main parts: (1) Three simple English grammar questions to verify that participants are truly native English speakers and (2) a story telling task, using six novel words. We used a fixed set of nonce words, taken from an artificial language learning study in Gómez (2002): 'Plizet', 'Nilbo', 'Skiger', 'Vamey', 'Chila', and 'Fengle'. The words were chosen out of twelve possible words based on their similarity to English. This was assessed by asking forty-five participants to rate the words on a scale of 1–5. All words were rated as equally similar to English (mean = 2.3, SD = 2.1).[1] Participants received the following prompt before telling the story (modeled on the work of Piantadosi 2014). The length of the requested story was 200 words, set to be the same as in Piantadosi (2014).

'This is John. John has a store on Main Street. Here are some people who visited John's store today: Chila, Fengle, Nilbo, Plizet, Skiger and Vamey. Please write a short story (no less than 2000 words) about John's day. Make sure to include all of the people in the list above more than once.'

In Study 1b, we wanted to see whether participants produce skewed distributions also when the nonce words are not names, but nouns and verbs. We also shortened the required story length to 400 words (to avoid high dropout rates). In this study, the nonce words belonged to different parts of speech in the three conditions and the prompt was modified accordingly: in the names condition they represented customers' names ('This is John. John has a store on Main Street. Some people visited John's store today: Chila, Fengle…'), in the nouns condition they represented items in the store ('This is John. John has a store on Main Street. Here are some of the items in John's store: a Chila, a Fengle…'), and in the verbs condition they represented actions ('This is John. John has a store on Main Street. Here are some things that John has to do in the store: to Chila, to Fengle…').

3.4 Procedure

Participants were asked to write a short story based on the prompt, using all six nonce words more than once. They were asked to produce 2000-word stories in Study1a and 400-word stories in Study 1b. While writing their stories, participants could see the prompt and the nonce words, so no memory demands were posed. One hour was allocated for telling the story in Study 1a and 35 min in Study 1b. After submission, the responses were manually reviewed and participants received payment if they met all inclusion criteria (see Appendix C) and followed the instructions.

3.5 Study 1: Results

Appendix A contains examples of stories produced by the participants in Study 1b where they had to produce 400-word stories (the R code for the analyses is provided in the following OSF link: https://osf.io/k83rn/). To assess whether the resulting nonce word distributions were near-Zipfian, we followed the method used by Piantadosi (2014). The frequency distribution of the six novel words was computed on the story produced by each participant. Because different participants could pick a different word as the most frequent one, we aggregated frequency counts by rank across subjects. That is, the frequency of the first ranked word was the summed frequency of the first ranked word for each individual subject, regardless of the word's identity. This was done to decrease noise since each subject is expected to use each word only a few times, and since we are not interested in the frequencies of specific words but in their resulting distribution. We then asked how well a linear line fits the log–log scale of frequency and rank. We expect it to show a high fit if it is near-Zipfian. We report the R^2 of this fit.

All word frequency distributions in Studies 1a and 1b showed a very good fit to a power-law distribution: Study 1a: $R^2 = 0.97$; Study 1b: Names: $R^2 = 0.98$, Nouns: $R^2 = 0.99$, and Verbs: $R^2 = 0.96$ (see Fig. 1), suggesting that speakers have a preference for producing skewed and near-Zipfian word distributions. The fit was similar across the different grammatical categories,

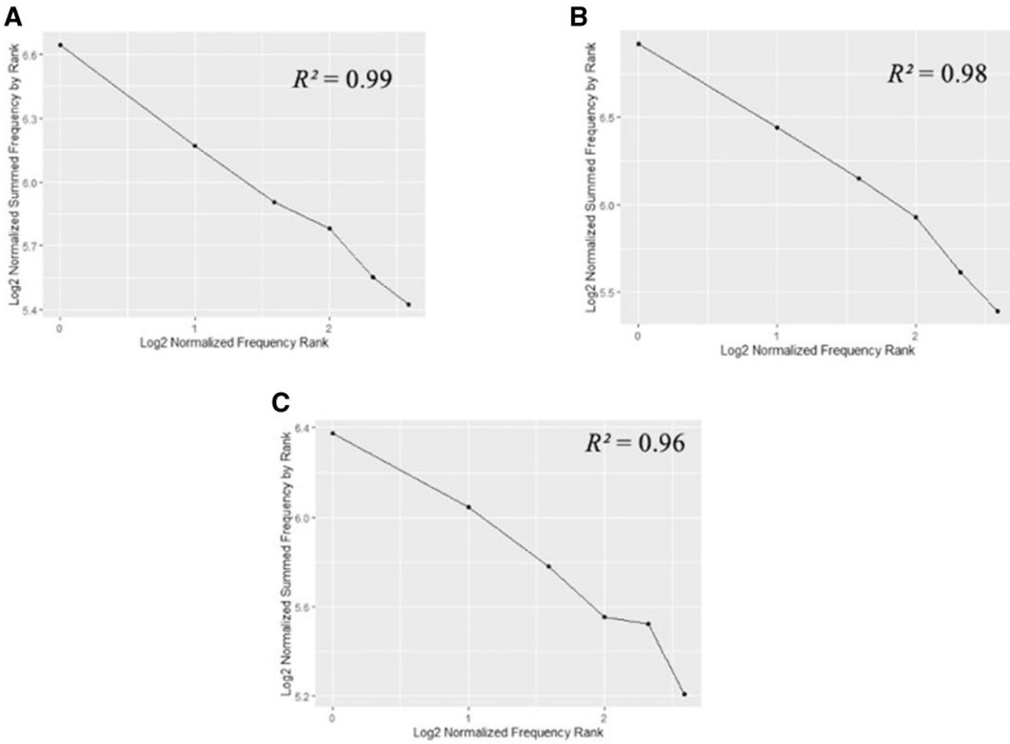

Figure 1. Log2 summed frequency across all participants (*y*-axis) and Log2 frequency-rank (*x*-axis) for the six nonce words for the nouns (**A**), names (**B**), and verbs (**C**) conditions in Study 1b. The R^2 indicates the fit to a power law distribution.

ruling out a name-specific explanation for the results of Piantadosi (2014) and Study 1a. Given that the fit was high for names, nouns, and verbs, we decided to use nouns for Studies 2 and 3, since they are easier to implement as nonce words.

4. Study 2

In Study 2, we test the hypothesis that weak individual biases amplified over time would lead speakers to transform uniform word distributions into skewed distributions. We use an iterated learning design with a diffusion chain paradigm where participants read a story and had to re-tell it. The first generation of learners was exposed to the same story in all the different chains. All subsequent learners saw the output produced by the previous learner in their chain as their input. Our diffusion chains consisted of ten 'generations' of single participants, as is customary in iterated learning studies. The input we used for the first generation of learners was a short story written by us about 'John's day at the store' (see Appendix B for the full story). The story included six novel words, each appearing eight times throughout the story. Since all participants in generation 1 received a uniform distribution, any increase in skew would be the product of change implements through the diffusion chain. Finding that participants shift a uniform distribution into a more skewed one will provide evidence for speakers' cognitive preference for Zipfian distributions.

We had two conditions in this study: In the 'allow-type-reduction' condition, we allowed the transfer of stories that did not include all six novel words. That is, we allowed learners to reduce the number of unique word types: If a learner only used four of the six novel words in their re-telling, their story would be transmitted as is to the next learner. This condition is similar to the homonym condition in Kirby et al. (2008) where no filters were applied to learners' output: Participants could use the same label for multiple objects, thereby reducing the number of unique word types in the language. In the 'no-type-reduction' condition, we kept the number of word types constant across generations. This is important since a reduction in word types will lead to entropy reduction, even if the distribution stays uniform: A story with four words appearing equally often will have lower entropy that a story with six words appearing equally often. The 'no-type-reduction' condition allows us to see whether there are changes in entropy even when the number of word types remains the same. We did two things to ensure the number of types did not change over generations. First, we explicitly told participants that they have to use all six novel words in their re-telling (this was added as part of the instructions about the re-telling that participants heard after they read the story). In addition, we added a filtering procedure: If a participants' story did not include all six words at least once, we re-ran this generation on another participant, using the previous generation as the input story (e.g., if participant 3 did not produce all six novel words at least once, we ran another participant in this generation, who was given the story of participant 2 as input).

We ran five chains in each condition, with ten generations in each chain. To measure the effect of transmission on skew, we calculated the entropy of the distribution of the six novel words in each generation and asked whether it becomes lower (more predictable) across generations. This is our equivalent of the structure measure used in iterated learning studies.[3] We did not estimate the fit of the distribution to a power law since we only have a small number of tokens in each generation, rendering the assessment of R^2 meaningless (we had a much smaller number of word tokens than in Study 1 where we aggregated tokens over participants to create a larger sample).

4.1 Participants

One hundred adult native English speakers from the USA participated in the study through Amazon's Mechanical Turk (MTurk) online platform, with five chains in each condition and ten participants in each chain. To join the study, participants had to meet the same inclusion criteria described in Study 1: (1) located in the USA, (2) HIT (task submission) approval percentage >95% (to validate participants' reliability), and (3) be native English speakers (as verified by several English grammar questions). Each participant also had to read and sign our Explanation Sheet and Informed Consent Form in order to participate. Participants who met inclusion criteria and followed the instructions received $4 in return for their participation (calculated in compliance with the US federal living wage).

4.2 Materials

We used the same six nonce words used in Study 1 ('Plizet', 'Nilbo', 'Skiger', 'Vamey', 'Chila', and 'Fengle'). Participants read a story and were then were asked to retell it. The first generation in each chain was exposed to the same short story (~500 words), written by us, about 'John's day at the store'. The story started with a prompt introducing the situation and the six novel words in a list-like fashion.

'John woke up early on Monday morning, climbed on his bike and headed towards Main Street, where his store was. John arrived at the store early, so he would have time to organize all the items for sale: Chila, Fengle, Plizet, Skiger, Nilbo and Vamey'.

The story then described John's day, using the words as part of the text in a naturalistic fashion (e.g., 'First John had to unpack the Plizets. Plizets were the newest items on John's store, and he knew many customers would come today to get themselves a brand new Plizet', see the full story in Appendix B). The initial story included the six nonce words—representing items in John's store—appearing in a uniform distribution (eight times each). The words appeared in a batched presentation—meaning that the eight mentions appeared one after the other and were not distributed throughout the text. While this may differ from the way word use is interspersed throughout a conversation, it is consistent with the recurrence of variation sets (consecutive sentences where one word remains the same) in child-directed speech (e.g., Onnis et al. 2006; Shira and Arnon 2018). Subsequent learners in the same chain received the output story produced by the previous learner as their input story and so on for ten generations. The initial story was modeled on the stories produced by the participants in Study 1. A pilot was carried out to ensure that the content of the story and its length were not too easy or too hard to remember, and to measure the estimated time it takes to complete the task.

4.3 Procedure

The task had two parts: the eligibility evaluation was performed on MTurk and the story reading and retelling were carried out on the Qualtrics survey platform. Participants were first asked to answer three grammar questions verifying that they are indeed native English speakers. They were then exposed to the short story and told they would be asked questions about it later: they were not told beforehand that they would have to retell it. After reading the story, and before retelling it, we asked participants to rate the story they read on a scale of 0–5 on the following measures: how interesting it was, how coherent it was, and how well-written it was. We wanted to see if the ratings varied dependent on the skew of the distribution, and if so, could provide an indirect measure of comprehension ease. Participants then watched a short (unrelated) GIF for a few seconds, after which they were asked to re-tell the story they read earlier as accurately as possible, using all six novel words in their stories. In the second condition, participants were explicitly told they have to use each of the novel words at least once and that the task will not be considered complete if that were not the case. The expected duration of the task was approximately 30–40 min, but participants were given 1 h to complete it, to minimize the effect of time demands on performance. After submission, the responses were manually reviewed by the first author for eligibility and participants received payment if they met all inclusion criteria and followed instructions (for full inclusion and eligibility criteria, see Appendix C).

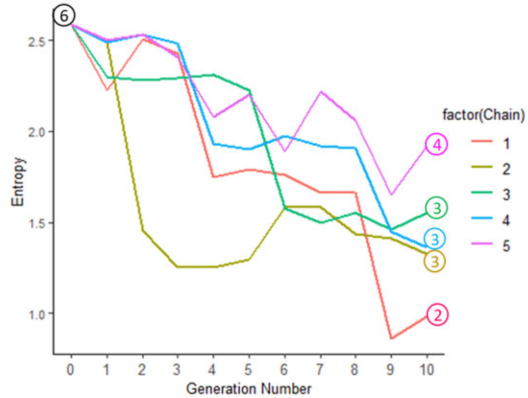

Figure 2. The decrease in entropy levels over generations in each of the five diffusion chains in the allow-type-reduction condition (where the reduction of word types was allowed). The circled numbers represent the initial and final number of unique word types in each chain.

4.4 Study 2: Results

We followed Raviv and Arnon (2018) in our statistical analyses. We analyzed the change in entropy using mixed effects regression models (and not *t*-tests comparing the first and last generation, as in Kirby et al. 2008) because such models allow us to examine the change across all ten generations; examine possible interactions; better control for possible differences between chains, and check for nonlinear trends in the data.[4] Our dependent variable was entropy and we had a fixed effect for generation (centered). All models had the maximal random effect structure justified by the data (Barr et al. 2013) and included a random intercept for chain (we had five different chains) and random by-chain slopes for the effect of generation. All regression models were run using the lme4 and pbkrtest packages in R (Bates et al. 2018; Halekoh and Højsgaard 2014; R Core Team 2016). P-values were obtained using the Kenward–Roger approximation,[5] which gives more conservative P-values for models based on a relatively small number of observations. We report the results for the two conditions separately since they differ in the number of word types (which inherently impacts their entropy levels). We provide examples of the stories told by participants in various generations in the two conditions in Appendix D.

In the 'allow-type-reduction' condition, where the reduction of word types was allowed, generation had a significant negative effect on skewness, with entropy levels decreasing as generations progressed ($\beta = -0.11$, $SE = 0.013$, $t = -8.93$, $P < 0.001$). Word distributions became more skewed over time across chains, with participants moving away from the initial uniform distribution toward a more predictable one in later generations (see Fig. 2). However, participants also reduced the number of word types over generations, resulting in a significant negative correlation between generation and number of word types ($r = -0.761$, $P < 0.0001$). By generation 10, the number of unique words dropped to as few as only two or three words in some chains, and entropy levels naturally dropped as well, since entropy is lower when there are fewer elements.

To determine which of the two factors, generation or number of word types, impacted the decrease in entropy, we ran an additional analysis using model comparisons to build a model that included number of types and then added generation. While number of types had a significant effect on entropy (compared with a model with only random variables, $\chi^2 = 125$, $P < 0.0001$), adding generation did not improve the model ($\chi^2 = 0.45$, $P = 0.49$), suggesting the decrease in entropy was primarily impacted by the reduction in the number of word types.

Importantly, there was a significant reduction in entropy also in the 'no-type-reduction' condition, where the number of word types stayed constant across generations (see Fig. 3). A model with entropy as the dependent variable, generation (centered) as a fixed effect, and the same random effect structure showed that the effect of generation was significant, with entropy levels decreasing as generations progressed ($\beta = -0.02$, $SE = 0.01$, $t = -2.49$, $P = 0.02$, $\chi^2 = 4.7$, $P < 0.05$ in model comparisons to model with only random variables). We can see from Fig. 3 that Chain 4 seems to differ from the others, having a U-shaped trajectory and ending approximately back at the initial entropy level of the uniform distribution. To ensure that this specific chain does not significantly alter the results, we performed an additional analysis using the same model, but excluding

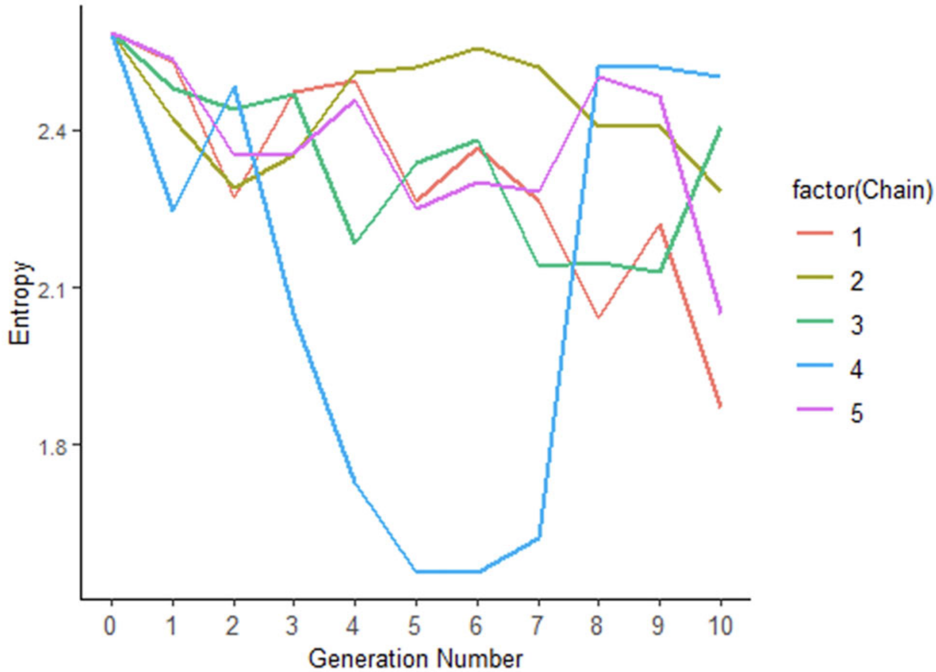

Figure 3. The decrease in entropy levels over generations in each of the five diffusion chains in Study 2b, where the number of word types was kept constant (six words).

Chain 4. The results without Chain 4 remained the same, with a significant effect of generation on entropy ($\beta = -0.03$, SE = 0.01, $t = -3.3$, $P < 0.05$). The reduction in entropy in Chain 4 between generations 2 and 3 reflects an increase in skew (the frequency distribution of the six words in generation 2 was: 7, 5, 4, 4, 3, 3 while the frequency distribution in generation 3 was: 13, 3, 2, 2, 2, 2). The increase in entropy from generations 7 to 8 reflects a reduction in the number of mentions and a decrease in the frequency of the frequent word (generation 7: 11, 1, 1, 1, 1, 1; generation 8: 2, 1, 1, 1, 1, 1).[6]

To exemplify that the frequency distribution of the six words did become more skewed over time, we looked at the final generation (generation10) in the 'no-type-reduction' condition, aggregated across chains (Fig. 4). To generate this number, we ranked each of the six words in generation 10 and then summed the frequency of all the first ranked words, second ranked words, and so on, to get an aggregated number of mentions across the five chains. As can be seen, the distribution is no longer uniform but instead shows a skew similar to that found in natural language.

We wanted to further explore the variability and consistency in the identity of the frequent words across chains/generations. As a first step, we wanted to see if the increase in skew is driven only by recency/primacy effects: The order of mention was identical in the initial story across chains (e.g., the word Plizet was the first to be mentioned and the word Nilbo was the last, see the full story in Appendix A). If participants simply used the first/last word more often than the others (because it was more salient due to its position in the story), this could introduce a skew whose sources would be less relevant to actual language learning, where the order of mention of lexical items is not fixed. To check if this is the case, we counted how many of the frequent words were the first/last in the previous generation. In Study 2a (where word types were allowed to decrease), 56% of the frequent words at any given generation were not the first/last in the previous study. In Study 2b, 40% of the frequent words were not the first/last in the previous generation. That is, it is not the case that the frequent words were always the first/last in the previous generation. More importantly, the order of mention did not stay constant between generations, further reducing recency/primacy effects: Across generations and chains, there were over thirty-five different forms serving as the first/last word. This number is much higher than the six types in each generation, since participants introduced novel forms in their re-telling (e.g., after hearing the initial story with its six novel words, participant 1 in Chain 4 of Study 2a used the novel form Wilba as one of their

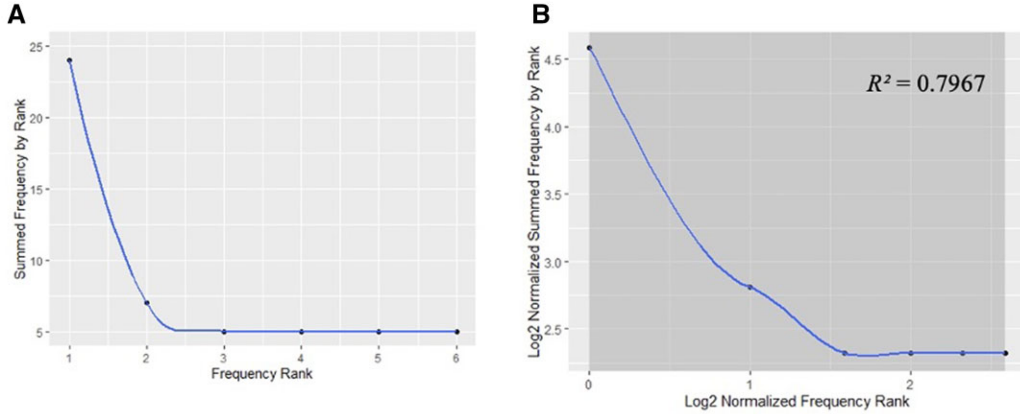

Figure 4. Summed frequency (y-axis) by rank (x-axis) (**A**) and logged frequency-rank (**B**) for the six nonce words in Generation 10 in the 'no-type-reduction' condition (aggregated across the five chains). The R^2 indicates the fit to a power law distribution.

six words). Interestingly, the identity of the frequent word also changed between generations, especially in the early generations where the skew was less pronounced. About half of the time, the frequent word in generation $x+1$ was not the frequent word in generation x. The shift in the identity of the frequent word happened more during earlier generations, where the skew was less prominent: Once one word became much more frequent it was more likely to be kept frequent in subsequent generations. In Study 2a, only 48% of the frequent words in each generation were also the frequent word in the previous generation. This number was higher in later generations: Of the frequent words that remained the same between generations, 35% appeared in generations 1–4 and 65% appeared in generations 5–10. A similar pattern is found in Study 2b where 54% of the frequent words was also the frequent word in the previous generation. Of these, 22% appeared in generations 1–4 and 78% appeared in generations 5–10. Taken together, these analyses demonstrate the variability in the identity and consistency of the frequent word across generations/chains (it is important to note that this variability was probably magnified because participants did not have to learn a consistent mapping between meaning and form, we return to this in Section 6).

We had also asked participants to rate the story they read on a scale of 0–5, for how interesting, how coherent, and how well-written it was, to see whether the ratings are affected by the skewness of the distribution. We had hoped that these ratings could serve as a learnability measure: an indirect way to ask if the skew of the story impacts its comprehension. However, while the ratings tended to correlate with one another (Study 2a: well-written: interesting, $r = 0.42$, $P < 0.01$; well-written: coherent, $r = 0.6$, $P < 0.0001$; interesting: coherent, $r = 0.16$, $P > 0.1$; Study 2b: well-written: interesting, $r = 0.7$, $P < 0.0001$; well-written: coherent, $r = 0.73$, $P < 0.0001$; interesting: coherent, $r = 0.55$, $P < 0.0001$), they did not correlate with entropy (all P-values > 0.1, expect that of coherent: entropy, which was 0.055), suggesting that skew alone did not impact the judgments.

In line with our predictions, the findings from the two conditions show that speakers shift a uniform distribution to a more skewed one through the process of transmission. In the next study, we wanted to start exploring the factors that could give rise to the change, and in particular, the role of memory demands. One possibility is that some words were used less often because they were not remembered as well/were harder to access. If the increase in skew is related to difficulty accessing/selecting words (e.g., Ferrer-i-Cancho and Sole 2003), then the shift may be smaller if participants are provided with the novel words before the re-telling and do not have to recall them. Alternatively, the shift may reflect individual learners' production biases (as seen in the story telling in Study 1) and be unaffected by memory demands. To contrast these two possibilities, we ran an additional study, where participants were given the novel labels during the re-telling.

5. Study 3

5.1 Method

We used the same design as that of Study 2 with the following change: We now provided participants with the nonce words during the re-telling, instead of having them recall them. We used the same initial story (see Appendix B), the same nonce words, and ran the same number of chains (five). We used the same instructions and filtering procedure as in the 'no-type-reduction' condition in Study 2 to ensure that the number of types remained constant across generations. The only difference was that participants saw the six novel words before re-telling the story.

5.2 Participants

Fifty adult native English speakers from the USA participated in Study 3 through MTurk (five chains, ten participants in each chain). Inclusion criteria, requirements, and payment were identical to Study 2 (see 'Participants' under Study 2).

5.3 Study 3: Results

We ran a mixed-effect model with entropy as the dependent variable, generation (centered) as a fixed effect, random intercepts for chain, and a random by-chain slopes for the effect of generation (the same model used to analyze the second condition of Study 2). In contrast with the no-type-reduction condition in Study 2, we did not find a significant decrease in entropy across generations, though it was trending in that direction ($\beta = -0.02$, SE = 0.01, $t = -1.86$, P = 0.11). When participants did not have to remember the words, they did not consistently shift their input toward more skewed distributions. This difference is also reflected in the overall higher entropy (averaged across generations and chains) of the word distributions in this study compared with the parallel no-type-reduction condition in Study 2 ($t(\mathrm{df} = 98) = -2.70$, P < 0.01). As can be seen in Fig. 5, entropy did decrease in three of the five chains between the initial and final generation (meaning that the distribution did not stay uniform). This shift is also evident when we look at the frequency distribution in Generation 10 (aggregating across chains and participants): As in Study 2, the distribution shows a right-skew that resembles natural language (Fig. 6).

To further understand the impact of not having to recall and access the novel labels, we ran an additional model on the combined data from the no-type-reduction condition in Studies 2 and 3. The model included entropy as the dependent variable, the interaction between generation (centered) and study (2b and 3) as a fixed effect, random intercepts for chain (numbered 1–10 for the ten chains we had across conditions), and a random by-chain slopes for the effect of generation. The effect of generation was significant ($\beta = -0.02$, SE = 0.01, $t = -2.71$, P < 0.05), showing that entropy decreased overtime. Importantly, the interaction between generation and study was not significant ($\beta = 0.0006$, SE = 0.02, $t = 0.04$, P = 0.97; $\chi^2 = 0.002$ in model comparisons), indicating that the general direction—of a shift toward more skewed distributions—is not significantly different across studies, though perhaps less consistent when memory demands are reduced. Finally, as in Study 2, the ratings were not impacted by the skew: while the ratings tended to correlate with one another (well-written: interesting, $r = 0.52$, P < 0.001; well-written: coherent, $r = 0.40$, P < 0.01; interesting; coherent, $r = 0.19$, P > 0.1), they did not significantly increase or decrease in correlation with entropy (all P-values > 0.1), suggesting that skew alone did not impact the judgments.

As in Study 2, we wanted to further explore the variability and consistency in the identity of the frequent words across chains/generations. Here also, over half of the frequent words at any given generation were not the first/last word in the previous generation (52%), suggesting that the skew was not driven only by recency/primacy effects. More importantly, the order of mention did not stay constant between generations, further reducing recency/primacy effects: Across generations and chains, each of the six unique words appeared at least once as the first/last word in the story (unlike in the previous study, no novel word forms were introduced since participants were given the words before the retelling). Interestingly, the identity of the frequent word also changed between generations, especially in the early generations where the skew was less pronounced. The shift in the identity of the frequent word happened more during earlier generations, where the skew was less prominent: Once one word became much more frequent it was more likely to be kept frequent in subsequent generations. In Study 3, only 48% of the frequent words in each generation were also the frequent word in the previous generation. This number was higher in later generations: Of the frequent words that remained the same between generations, 30% appeared in generations 1–4 and 70% appeared in generations 5–10.

In sum, the effect of memory demands, as operationalized here, was inconclusive: While entropy did not decrease significantly when the word forms were given, the analysis of the combined data from the parallel 'no-type-reduction' condition in Studies 2 + 3 showed a significant decrease of entropy and no interaction. That is, learners' shift toward more skewed distributions even when the word forms do not have to be recalled. We discuss the limitation of our memory manipulation and its relevance to actual language learning, in Section 6.

6. Discussion

Zipfian distributions are prevalent in language and are in fact one of the most consistent commonalities

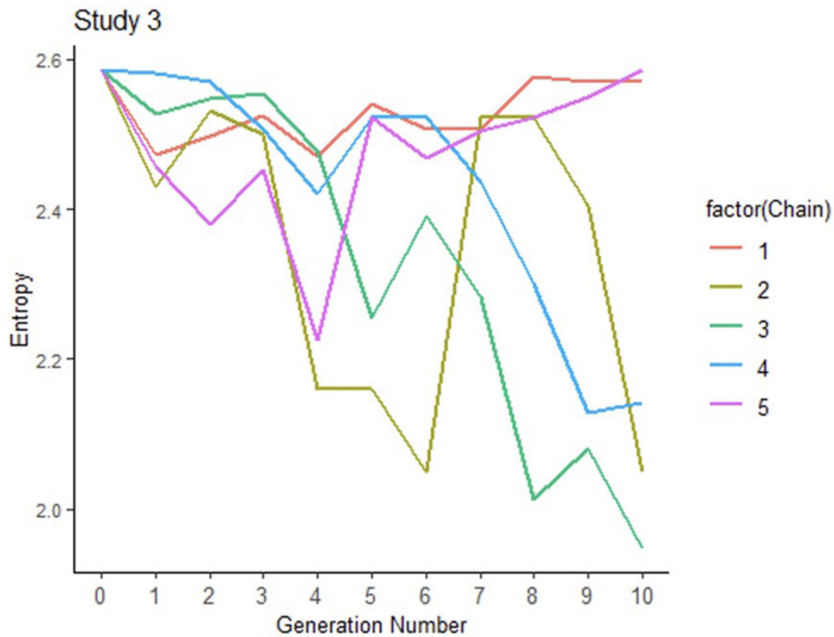

Figure 5. The decrease in entropy levels over generations in each of the five diffusion chains in Study 3 (where labels were given before the re-telling). Entropy decreased in chains 2–4, while in chains 1 and 5 it remained almost unchanged.

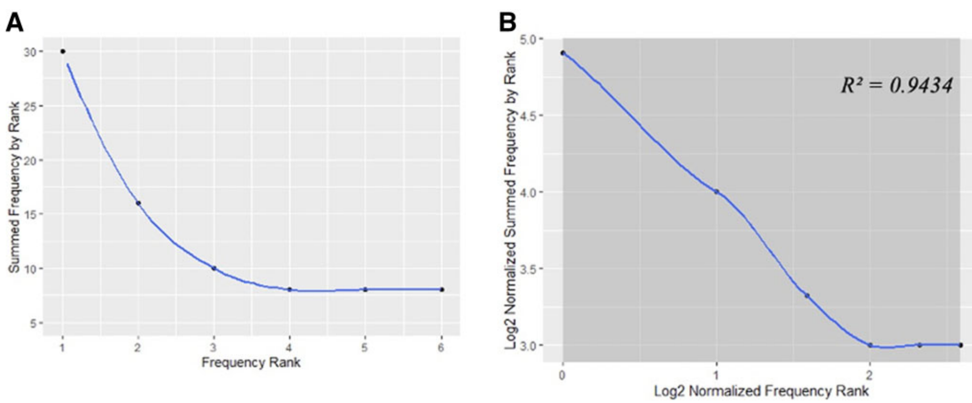

Figure 6. Summed frequency (*y*-axis) by rank (*x*-axis) (**A**) and logged frequency-rank (**B**) for the six nonce words in Generation 10 in Study 3. The R^2 indicates the fit to a power law distribution. The distribution shows a good fit to a power law distribution.

between languages, generating ongoing debate about their source (Ferrer-i-Cancho and Sole 2003; Ferrer-i-Cancho 2018; Piantadosi 2014). Recent work suggests that such distributions can provide a facilitative environment for learning (e.g., Kurumada et al. 2013; Lavi-Rotbain and Arnon 2020, 2022), raising the possibility that learnability pressures play a part in their recurrence and maintenance in language (Bentz et al. 2017; Ferrer-i-Cancho 2018; Ferrer-i-Cancho et al. 2022; Lavi-Rotbain and Arnon 2022; Semple et al. 2022). Here, we explore this idea by using an iterated learning paradigm to see if learners will shift uniform word distributions toward more skewed ones through the process of cultural transmission, and in doing so, reveal a cognitive preference for more predictable distributions. As a first step, we showed that speakers are biased to produce skewed word distributions when telling a story using novel names, nouns, and verbs: Their aggregated productions were right skewed and showed a good fit to the power-law distribution, across parts of speech. We then asked whether participants will actively change their input to create skewed distributions out of uniform ones. Such a finding would provide stronger evidence for an underlying cognitive preference and is harder to attribute only to prior linguistic experience.

We used an iterated learning design to examine the change in skew of word frequency distributions in a story transmitted over ten generations of learners in two conditions: when the number of word types could be reduced (allow-type-reduction) and when it was kept constant (no-type-reduction). We found that the entropy of the word distributions decreased over time, even when the number of unique word types was not reduced. The reduction in the number of word types we saw in the allow-type-reduction condition is a common finding in iterated learning studies when no other constraints are imposed (when the reduction is not artificially prevented and when there are no additional communicative pressures, Kirby et al. 2015). Finally, we asked whether the decrease in entropy is influenced by lexical access demands by running another no-type-reduction condition where the novel labels were given to participants during the re-telling instead of having them recall them. The findings here were inconclusive: Entropy levels in this study did not decrease significantly over time, though the distribution did become right skewed. However, further analyses comparing these results to those of the parallel 'no-type-reduction' condition in Study 2 revealed a main effect of generation and no interaction: the shift toward more skewed distributions is found in both studies. The similarity is also seen when we compare entropy levels in the two 'no-type-reduction' conditions (with and without giving the labels): The entropy of the tenth generation (summed over participants and chains) was lower than the entropy of the first generation in both ('no-type-reduction condition in Study 2: $2.48 > 2.22$; no-type-reduction condition in Study 3: $2.53 > 2.37$). Interestingly, the entropy of the first generation was similar to that of the novel stories from Study 1 (2.45) and lower than, though close to, a uniform distribution with six word-types (2.58). That is, the skew (measured in entropy) of the transmitted stories was larger than that of a first telling (Study 1).

These findings provide a first illustration of the emergence of skewed word distributions via cultural transmission, with implications for our understanding of the learnability pressures involved in the propensity of Zipfian distributions in language. They show that skewed distributions can emerge in an iterated learning design, and in doing so, extend the impact of cultural transmission to another aspect of language structure. More importantly, in combination with findings showing that languages evolve in ways that maximize their learnability (Kirby et al. 2008), they lend support to the proposal that Zipfian distributions confer a learnability advantage and that their recurrence in language is driven, at least in part, by learnability pressures (Bentz et al. 2017; Lavi-Rotbain and Arnon 2019a,b, 2020, 2022). Specifically, they support the prediction that speakers aim to minimize word entropy. While such minimization has been attributed to cognitive costs associated with communication (Ferrer-i-Cancho 2018), the current results—which are based on a task that does not directly involve communication—suggest that minimizing entropy can also be the product of individual-level biases (memory biases and lexical access dynamics).

How would such pressures impact language over time? One possibility is that an individual bias for producing skewed distributions—reflecting perhaps the nature of the world and of our communicative interests (we want to talk more about certain things)—would be amplified over time by the learnability and communicative benefits that such skewed distributions present, resulting in their recurrence and maintenance in language. The particular shape of the Zipfian distribution—with its non-linear decrease of frequency—could result from the convergence of learnability and other pressures, such as expressivity, efficient hierarchical organization of word meaning (Manin 2008), and wanting to balance speaker and listener effort (Ferrer i Cancho and Sole 2003). Importantly, the current study does not (and did not aim to) provide a comprehensive account of the source of Zipfian distributions in language. Instead, it offers a first step in experimentally testing the impact of converging and competing cognitive and communicative pressures on the emergence of such distributions. Using iterated learning paradigms allows us to contrast theoretical accounts and test the interplay between different pressures. For example, future work could test the proposal that Zipfian distributions produce an optimal balance between speaker and listener effort (Ferrer i Cancho and Sole 2003) by adding a more direct manipulation of communication (modeled after Kirby et al. 2015) to the iterated learning paradigm used here. In the current study, participants were not told their stories would have to be understood for another participant and did not have to complete a communication task based on them (i.e., there was no pressure to make the stories easier to understand). Such communicative pressures could be added to the paradigm. If the balance between speaker/listened effort indeed plays a role in the emergence of Zipfian distributions, we may expect the skew to emerge more rapidly (or be larger) in communicative contexts versus individual transfer.

Our findings about the possible role of lexical access demands in the shift toward more skewed distributions were inconclusive and should be interpreted with caution: A combined analysis of the two studies showed that entropy decreased in both, but the effect was not significant when looking only at Study 3. There are several possible explanations for this pattern. One possibility is that the current design was under-powered to detect the impact of lexical access demands. Participants remembered the words pretty well even when the labels were not given: Remembering six novel words may not be a hard enough task to reveal memory effects. This explanation is somewhat supported by a recent study that illustrates the effect of memory demands on the emergence of linguistic structure (Cornish et al. 2017). In this study, random letter strings acquired language-like structure (i.e., become more similar to the statistics of letter strings in participants' L1) through the process of transmission, more so when they were harder to remember (Cornish et al. 2017). This suggests that memory demands can amplify existing biases (e.g., for regularization or structure), a pattern our study may reflect if more words are used. Another possibility is that the way we manipulated memory demands—by either providing the labels or not—does not capture the relevant aspect of memory pressures on the emergence of Zipfian distributions. Unlike in our study, in real language, words have meaning, and low frequency words need to be learned, making it unlikely that they are used less simply because they are not remembered. What is more likely is that their reduced use is driven by lower activation threshold, which our manipulation did not simulate. Future work is needed to investigate more systematically the impact of lexical access/memory demands on the emergence of skewed distributions, for example by increasing the number of nonce words, or making them harder to remember (e.g., longer, more phonologically similar to one another, etc.).

The current study is only a first step in investigating the possible cognitive sources of Zipfian distributions in language and as such has limitations and leaves many questions unanswered. One limitation is that we did not measure learnability directly. We did not ask if more skewed distributions are also more learnable, as predicted by the learnability based perspective on their presence in language. The standard measure of learnability in iterated learning paradigm, which is the distance between the input and output strings (e.g., Levenshtein distance), was not applicable in our design for several reasons. First, participants were not taught a mapping between objects and strings, so we could not systematically pair output and input words to assess their similarity. We are currently running additional studies to see how distributions change when participants are also learning to associate the novel words with novel objects. Second, participants were asked to reconstruct the story and not only learn the words, making it hard to assess how similar the stories were. We initially considered evaluating the similarity between the events in the input and output story, but we decided against it since participants were not instructed to replicate all the events. However, such global measures should be used in future studies: the literature on story re-telling could provide a fruitful place to look for such measures (though it tends to focus on affect or subjective ratings of information content). Another limitation is the possible transfer of frequencies from English, participants' native language. The reason we used made-up words instead of real English words was to avoid the impact of the existing frequency distribution on participants' productions. That is, we wanted to limit the effect of a word's existing frequency on its use in the story. However, even though our novel words did not have meaning attached, participants could have matched them to existing words, with an impact on their use. A third limitation is that the number of tokens in the stories was relatively small (each novel word appeared only eight times in the initial story and the entire story was 500 words), making it hard to distinguish between more and less frequent words (there isn't much space for contrast if the maximal number of mentions is eight). Future studies exploring the emergence of word frequency distributions should consider these challenges when designing iterated learning experiments.

In sum, this study investigated a possible cognitive preference for more predictable word frequency distributions as an explanation for their propensity in language, by asking whether learners will turn uniform distributions into more skewed ones via cultural transmission. The results show that speakers are biased to produce skewed word distributions when telling a novel story and that this bias gets amplified over time, as skew increases with transmission. While this study is only a first step in investigating the cognitive sources for the emergence of more predictable distributions in language, these findings provide further support to the growing body of knowledge proposing that Zipfian distributions confer a learnability advantage and that their recurrence in language may be driven by learnability pressures.

Acknowledgements

We wish to thank Limor Raviv and Ori Lavi-Rotbain, whose papers provided the basis and inspiration for the current paper for helpful comments, suggestions, and discussions throughout the writing process. We also thank Shira Tal for insightful comments on the manuscript. We also wish to thank Shira Tal and

Ohad Stossel for their help in programing the experiments and analyzing the data.

Funding

The research was funded by a grant from the Israeli Science Foundation (ISF grant 445/20) awarded to the second author, as well as a fellowship from the Jerusalem Brain Community awarded to the first author (JBC Golden Opportunity Fellowship, October 2018—September 2019). All study materials and results can be found on OSF (link: https://osf.io/k83rn/?view_only=f5d539eef21542be977fe5d8f5354b3e).

Conflict of interest statement. None declared.

Notes

1. The word 'Wiffle' was rated the most similar to English by far (4.1) and was removed, assumed as being too close to 'Waffle'. The word 'Puser' also got a high rating (2.33) but was removed so that there would only be one word starting with the letter 'P' ('Plizet' got a higher rating).
2. The Zipfian distribution is determined by two parameters: α determines the slope of the distribution and β is a correction added by Mandelbrot (1950) that improves the fit to natural language. Some previous studies estimate alpha and beta from the data (and compare those across languages). However, we did not estimate the α and β of the distribution because we did not have enough data to do so reliably (estimating α is reliable only when there are around 50,000 tokens, while our aggregated stories only had between 400 and 1,000 mentions of the six entities in total, see discussion in Lavi-Rotbain and Arnon (submitted).
3. Iterated learning studies usually also assess transmission error between generations as a way to examine changes in learnability. However, the standard measure of learnability, which is the distance between the input and output strings, is not informative in the current design because participants are not learning form/meaning mappings.
4. Following Beckner et al. (2017) and Raviv and Arnon (2018), we checked whether the effect of generation is better fit by non-linear terms using the poly function for all reported models. Since no non-linear effects were found, we do not report these models.
5. The Kenward–Roger approximation provides a more accurate number of degrees of freedom for P-values and other inferential statistics in linear mixed models, in cases where such statistics may be biased due to a small number of clusters.
6. The disproportionate impact of one participant on a chain is one of the disadvantages of iterated learning paradigms: individual participants whose performance differs or deviates from that of the rest can change the course of an entire chain.

References

Baixeries, J., Elvevåg, B., & Ferrer-i-Cancho, R. (2013). The evolution of the exponent of Zipf's law in language ontogeny. *PloS one*, 8/3:e53227.

Barr, D. J. et al. (2013) 'Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal', *Journal of Memory and Language*, 68/3: 255–78.

Bates, D. et al. (2018) *Package 'lme4'*. Vienna: R Foundation for Statistical Computing.

Beckner, C., Pierrehumbert, J. B., and Hay, J. (2017) 'The Emergence of Linguistic Structure in an Online Iterated Learning Task', *Journal of Language Evolution*, 2/2: 160–76.

Bentz, C., & Ferrer Cancho, R. (2016). Zipf's law of abbreviation as a language universal. In: *Proceedings of the Leiden workshop on capturing phylogenetic algorithms for linguistics* (pp. 1–4). University of Tübingen.

—— et al. (2017) 'The Entropy of Words—Learnability and Expressivity across More than 1000 Languages', *Entropy*, 19/6: 275.

Chater, N., and Brown, G. D. (1999) 'Scale-Invariance as a Unifying Psychological Principle', *Cognition*, 69/3: B17–24.

Claidière, N. et al. (2014) 'Cultural Evolution of Systematically Structured Behaviour in a Non-Human Primate', *Proceedings of the Royal Society B: Biological Sciences*, 281/1797: 20141541.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36/3: 181–204. https://doi.org/10.1017/S0140525X12000477

Cornish, H. et al. (2017) 'Sequence Memory Constraints Give Rise to Language-Like Structure through Iterated Learning', *PLoS ONE*, 12/1: e0168532.

Coupé, C. et al. (2019) 'Different Languages, Similar Encoding Efficiency: Comparable Information Rates across the Human Communicative Niche', *Science Advances*, 5/9: eaaw2594.

Christiansen, M. H., and Chater, N. (2008) 'Language as Shaped by the Brain', *Behavioral and Brain Sciences*, 31/5: 489–509.

Culbertson, J., and Kirby, S. (2016) 'Simplicity and Specificity in Language: Domain-General Biases Have Domain-Specific Effects', *Frontiers in Psychology*, 6: 1–11.

—— (2012) 'Typological Universals as Reflections of Biased Learning: Evidence from Artificial Language Learning', *Language and Linguistics Compass*, 6/5: 310–29.

Fehér, O., Ljubičić, I., Suzuki, K., Okanoya, K., & Tchernichovski, O. (2017). Statistical learning in songbirds: from self-tutoring to song culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372/1711: 20160053.

Ferrer-i-Cancho, R., and Sole, R. V. (2003) 'Least Effort and the Origins of Scaling in Human Language', *Proceedings of the National Academy of Science of the United States of America*, 100/3: 788–91.

—— (2005). The variation of Zipf's law in human language. *The European Physical Journal B-Condensed Matter and Complex Systems*, 44/2:249–257.

—— (2005b) 'The Variation of Zipf's Law in Human Language', *European Physical Journal B*, 44: 249–57.

—— (2016) 'Compression and the Origins of Zipf's Law for Word Frequencies', *Complexity*, 21/S2: 409–11.

—— (2018) 'Optimization Models of Natural Communication', *Journal of Quantitative Linguistics*, 25: 207–37.

——, Bentz, C., and Seguin, C. (2022) 'Optimal Coding and the Origins of Zipfian Laws', *Journal of Quantitative Linguistics*, 29/2: 165–94.

Gibson, E. et al. (2019) 'How Efficiency Shapes Human Language', *Trends in Cognitive Sciences*, 23/5: 389–407.

Gómez, R. L. (2002) 'Variability and Detection of Invariant Structure', *Psychological Science*, 13/5: 431–6.

Griffiths, T. L., and Kalish, M. L. (2007) 'Language Evolution by Iterated Learning with Bayesian Agents', *Cognitive Science*, 31/3: 441–80.

Halekoh, U., and Højsgaard, S. (2014) 'A Kenward–Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models—the R Package Pbkrtest', *Journal of Statistical Software*, 59/9: 1–32.

Hendrickson, A. T., and Perfors, A. (2019) 'Cross-Situational Learning in a Zipfian Environment', *Cognition*, 189: 11–22.

Kempe, V., Gauvrit, N., and Forsyth, D. (2015) 'Structure Emerges Faster during Cultural Transmission in Children than in Adults', *Cognition*, 136: 247–54.

Kirby, S., Dowman, M., and Griffiths, T. L. (2007) 'Innateness and Culture in the Evolution of Language', *Proceedings of the National Academy of Sciences of the United States of America*, 104/12: 5241–5.

——, Tamariz, M., Cornish, H., and Smith, K. (2015) Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141, 87–102.

——, Cornish, H., and 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 of the United States of America*, 105/31: 10681–6.

——, Griffiths, T., and —— (2014) 'Iterated Learning and the Evolution of Language', *Current Opinion in Neurobiology*, 28: 108–14.

Kurumada, C., Meylan, S. C., and Frank, M. C. (2013) 'Zipfian Frequency Distributions Facilitate Word Segmentation in Context', *Cognition*, 127/3: 439–53.

Lavi-Rotbain, O., and Arnon, I. (2019a) 'Children Learn Words Better in Low Entropy', In: *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, pp. 631–637. Cognitive Science Society..

——, and —— (2019b) 'Low Entropy Facilitates Word Segmentation in Adult Learners', In: *CogSci*, pp. 2092–7.

——, and —— (2020) 'The Learnability Consequences of Zipfian Distributions: Word Segmentation Is Facilitated in More Predictable Distributions'. Manuscript submitted for publication.

——, & —— (2022). The learnability consequences of Zipfian distributions in language. *Cognition*, 223:105038.

Mahowald, K. et al. (2020) 'Efficient Communication and the Organization of the Lexicon'. Unpublished manuscript, https://psyarxiv.com/4an6v.

Manin, D. Y. (2008) 'Zipf's Law and Avoidance of Excessive Synonymy', *Cognitive Science*, 32/7: 1075–98.

Mehri, A., & Jamaati, M. (2017). Variation of Zipf's exponent in one hundred live languages: A study of the Holy Bible translations. *Physics Letters A*, 381/3:2470–2477.

Onnis, L., Waterfall, H. R., & Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition*, 109/3:423–430.

Perfors, A. (2012) 'When Do Memory Limitations Lead to Regularization? An Experimental and Computational Investigation', *Journal of Memory and Language*, 67/4: 486–506.

Piantadosi, S. T. (2014) 'Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions', *Psychonomic Bulletin & Review*, 21/5: 1112–30.

Core Team R. (2016) *A Language and Environment for Statistical Computing*. Vienna, Austria: *R Foundation for Statistical Computing*.

Raviv, L., and Arnon, I. (2018) 'Systematicity, but Not Compositionality: Examining the Emergence of Linguistic Structure in Children and Adults Using Iterated Learning', *Cognition*, 181: 160–73.

Schuler, K. D. et al. (2017) 'The Effect of Zipfian Frequency Variations on Category Formation in Adult Artificial Language Learning', *Language Learning and Development*, 13/4: 357–74.

Semple, S., Ferrer-I-Cancho, R., and Gustison, M. L. (2022) 'Linguistics Laws in Biology', *Trends in Ecology and Evolution*, 37/1: 53–66.

Shannon, C. E. (1948) 'A Mathematical Theory of Communication', *The Bell System Technical Journal*, 27/3: 379–423.

Shira, T. A. L., & Arnon, I. (2018). SES effects on the use of variation sets in child-directed speech. *Journal of child language*, 45/6:1423–1438.

Tamariz, M., and Kirby, S. (2016) 'The Cultural Evolution of Language', *Current Opinion in Psychology*, 8: 37–43.

Zipf, G. (1936) *The Psychobiology of Language*. London: Routledge.

Appendix

Appendix A: Example stories Study 1b (400-word stories)

1. John showed up for work on Monday morning to find that he had received a delivery of new inventory. He quickly opened the boxes and started putting the merchandise on the shelves. John was pleased to see a box of Chilas, a kind of warm hat from Peru that had been a very popular seller that winter. Lots of mothers had been buying the colorful Chilas for their children, along with matching Fengles for their feet and hands. Unfortunately, the shop was still low on Nilbos. John worried that customers would soon become irate if they couldn't get any more Nilbos for their dogs and cats. He pondered that problem as he folded a silky blue Pilzet and added it to the towering display. Nobody seemed interested in Pilzets this year, although they had been last year's hottest sellers. Last winter it was not uncommon to see people walking past the shop wrapped up in an exotic patterned Pilzet with matching Fengles. The bell above the door tinkled and John looked up to greet one of his regular customers, Mr. Henderson. 'Any new Nilbos today?' asked Mr. Henderson, stuffing his Chila into the pocket of his green tweed Skiger. 'Not today', said John, 'but I think we have a few Varneys left'. 'That will do', said Mr. Henderson. 'Archie liked that last Varney, but not as much as the Nilbos'. Archie was Mr. Henderson's labrador retriever, who often accompanied him around town. John and Mr. Henderson talked for a while about the differences between Varneys and Nilbos, and how funny dogs could be. Mr. Henderson paid for the Varney, and on impulse grabbed a new yellow Chila for his grandson before putting his wallet back into his Skiger pocket. He went out the door, promising to drop in again next week to check for Nilbos. John looked at the clock. His assistant, Nancy, would be in soon. He decided to ask her if she had any creative ideas for promoting these Pilzets. Maybe they had fallen completely out of style, but perhaps there were a few people in town who still liked them. Or maybe they could come up with a new way to use them that would get people interested in them again. At any rate, he had seventy-two Pilzets now and he had to sell them somehow. He took a long look around his shop and felt satisfied.
2. One morning, John had just opened his store and was tidying up in preparation for customers to arrive. He had decided over the weekend that he would put his Chilas, Fengles, Nilbos, and Plizets on sale, but he would mark up the Skigers and Vameys. John had made these decisions after analyzing recent supply and demand. He felt like this would be fair and would maximize his profit. He had set up a display of the Chilas, Fengles, Nilbos, and Plizets to make sure that they would be right at the door when customers came in. He had also moved the Skigers and Vameys over near the cash register. He was hoping that customers would not notice the small increase in price on those. After all, he had only raised the price of the Skigers by 10% and that of the Vameys by 15%. On the same token, he had marked down the Chilas, Fengles, Nilbos, and Plizets by almost 25%! When customers finally started arriving, John eagerly waited to see what would happen. HE hoped his sales techniques would be successful. Sure enough, by lunch time he had sold a noticeable increase in the Chilas, Fengles, Nilbos, and Plizets! All four were practically booming. John felt very confident about his new sales technique. On the same token, he had only sold a few of the Skigers and Vameys, whereas normally he would have sold at least a dozen by lunchtime. But he had expected that. Several customers had even seemed very angry that the price had risen. One had threatened to shop somewhere else instead. John felt upset by this, but he figured it was part of doing business. By the end of the day, John was very pleased. He had sold out almost entirely on his Chilas and Fengles, and he had sold a good bit of the Nilbos and Plizets. While he had received complaints about the rising price of the Skigers and Vameys, overall he still felt like his strategy had been a success. John decided that the economics for small business class he had taken at the community college had indeed been helpful. He had been skeptical when he first signed up for it, but his daily earnings were demonstrating that indeed it was working!

Appendix B: Initial story for iterated learning studies (2 and 3)

John woke up early on Monday morning, climbed on his bike and headed toward Main Street, where his store was. John arrived at the store early, so he would have time to organize all the items for sale: Chila, Fengle, Plizet, Skiger, Nilbo, and Vamey. First John had to unpack the Plizets. Plizets were the newest items on John's store and he knew many customers would come today to get themselves a brand new Plizet. John put up the sign that said ‘New Plizets!’ on the window and opened the store. As expected, the first couple of customers that came in asked for Plizets. John gave them each a brand new Plizet for 50$ and they thanked John and left. After some minutes, an elderly man came into John’s store asking for a Chila. John handed the man a Chila from the shelf and said ‘that would be 20$please’. ‘20$?!’ the man complained, ‘Last week I bought three Chilas for 30$ just across the street!’ ‘I don’t know where you bought those Chilas’ John replied, ‘but these are original Chilas, and their price is final’. The man frowned and handed the Chila back to John, then stormed out of the store. John wasn’t bothered by the incident, since at least once a day a customer would argue with him over prices. Next, a young woman came in. She was looking through the Skigers, then accidentally dropped one on the floor. Skigers were very expensive items. John ordered Skigers from France, where the Skiger was invented. The woman apologized deeply and even asked to pay for the broken Skiger, but John decided not to charge her for it. The woman thanked him and left and John swept the broken Skiger pieces off the floor, then took his half an hour lunch break. When John reopened, a mother came in with her boy. The boy was crying, asking for a Vamey. She told him that a Vamey was too expensive, but he kept on crying: ‘Vamey! Vamey!’ The woman looked at John helplessly. John reached for the shelf and handed the woman a medium sized Vamey. ‘Here, these are not as expensive, and I’ll give you a discount’ John said, smiling. The woman thanked John for his kindness, bought the Vamey, and went her way. Just before closing, a teenage boy came in, and went straight to the Fengle section. Fengles were very popular with kids his age. ‘How much for one Fengle?’ the boy asked. ‘10$’ John replied. ‘Can I try one first?’ the boy asked, and John nodded. The boy took one of the Fengles and tried using it, then he put it back and tried a different Fengle. Finally, after trying a few Fengles, the boy left without buying any. John was not surprised, as he never expected kids to actually buy an item. It was getting late and time to close the store. Before closing, John checked the inventory and found out he had only one Nilbo left, and needed more Nilbos. Nilbos were the most popular items on John’s store, since they came in every color and shape, and each day he would sell at least one Nilbo. John decided that tomorrow he would call his Nilbo supplier and order more Nilbos. On his way home, John thought about the nice day he had in the store, selling Plizets, Fengles, Nilbos, Chilas, Vameys, and Skigers.

Appendix C: Story eligibility criteria and coding criteria for Studies 2 and 3

Task approval and payment criteria for participants in Studies 2 and 3

1. Proven native English speakers, as assessed by three English grammar questions.
2. Followed task’s instructions.
3. Provided ICF and task completion codes from Qualtrics.
4. Did not copy and paste the story, or parts of it, instead of retelling it.
5. Did not copy and paste a story from the Internet.
6. Used the specified number of nonce words as requested (in specific cases, participants still received half the payment for their participation if determined a misunderstanding).
7. Used novel words, as instructed, and not real words (in specific cases, participants still received half the payment for their participation if determined a misunderstanding).

Story eligibility criteria for coding and analysis in Studies 2 and 3 (stories that did not meet the following criteria were accepted and paid for, but not included in our studies)

1. Story is legible and uses native English grammar.
2. Used at least 50% of the events used in the original story, including at least one event containing a novel word.
3. Retelling the story rather than telling ABOUT the story (e.g., ‘In the story there were three events: 1…’).
4. No extensive use of first person and personal comments (e.g., ‘I think I remember that John went to…’; minor cases were permitted, if used only once or twice throughout the story).
5. In Study 2a: at least two of the novel words used.

Manual amendments made to the stories in Studies 2 and 3

To ensure the flow of the story chain, meeting the analysis criteria and that not too much time is wasted on reading, the first author has made some manual amendments to the stories before republishing them for the next participants. The changes were minor and minimal and did not affect meaning or outcome measures. The following changes were made:

- Punctuation corrections (spaces, periods, etc.).

- In two cases, use of a real word instead of a nonce word that was judged accidental, was changed, while making sure the string distance measure is not changed (e.g., ‘Bonbons’ was changed to ‘Bomboms’).
- Minor register changes, in cases some words were archaic and uncommon.
- Minor grammatical changes, if the issue was specific and judged unintended.
- Minor deletions where participants used first person, in cases where the use was minimal and specific (‘I guess...’).

Coding criteria for Studies 2 and 3

In each story, we counted the number of occurrences of each unique word type in the story. We included the following forms: singular (e.g., Bonzo), plural (e.g., Bonzos), clear spelling errors/variations (e.g., Bonzoes), and clear substitutions (e.g., ‘the item’, clearly referring to a previously mentioned novel word). Otherwise, other lexical NPs were not counted since we did not count them as part of the original novel word count in our story (six types × eight appearances each).

Appendix D: Example stories Studies 2 and 3

1. Study 2, allow-type-reduction, chain 1, generation 1. John owned a store and carried several oddly named items, of varying prices, appeal, and popularity. On this morning, he went into his shop and began what would end up being a fairly typical day. The items he sold were given names of comparable absurdity to: Plizets, Spigots, Nibos, Vendels, and Zegals. Of course, these items were all quite different from one another and some varied in properties such as color, size, and shape to a higher degree than others. In particular, Nibos came in the most varieties of colors and shapes, and this made them the most popular item in his store; he managed to sell a Nibo more or less every day. It would come to pass later in the evening, as John closed his store, that he would notice that he had only a single Nibo left in stock, and so he needed to be sure to place an order promptly the next day. He took note of this diligently, of course. Several different customers came in this day—as many customers as there were items for sale that were described, meaning each person was interested in a different item...—and the interactions never ended in the same way, for example, a sale, a broken item, a rejected item (by an angry customer, one of which John would endure every single day, and so he’d come to the point that any hostilities displayed by a customer failed to emotionally distract him in the least), a somewhat superfluous or feigned shopping excursion by a teenager (naturally seeking the popular Nibo, looking and trying every single one in stock, then leaving the store without making the $10 purchase that John had offered as, characteristically, a discount), and a mother with child, among possible others. Plizet was the first item to be inspected by the earliest customer, a man who sought and bought one at a discounted price that John graciously offered. These discounts and favors would appear to be the norm if this day was indeed a general representation of the shop’s long-term trends. After selling the Plizet, the daily contrarian, a middle-aged man, came to argue about what he considered an unsavory, abnormally high price for a Spigot. John made clear that the price would not change and this was the exception to the rule for the day. Perhaps this was solely a result of the argumentative and solipsistic nature of the dissatisfied customer... Later in the day, a woman with a child would accidentally drop a Zegal, which was the most valuable item John had for sale in his store. Zegal’s were pricey and rare. The woman, feeling guilty, offered to pay for the broken Zegal, but in the end John would not allow it, and the woman was very grateful. She left the store without making a purchase. A man did come in later on to inspect and ultimately purchase a Vendel; this would be the final purchase of the day and only one of two that the store would see.
2. Study 2, allow-type-reduction, chain 1, generation 4. John owned a shop where he sold rare items such as Spigots, Zendas, Piros, and Zygots. His store was very popular, mostly because John was such a generous and kind man. He had many loyal customers. The objects in his shop were somewhat expensive, but that did not stop him from doing a lot of business. Piros especially had to be constantly restocked, as they were his most profitable item. Every day, John did a brisk business and many loyal customers would stop by to see what was new. John often offered discounts to new customers, it was rare for him not to do so. Every once in a while, for some reason, John would fail to do so. Was he distracted? Or in a bad mood? Did he not like the customer? No one knew ... Most of the time, however, John was quick to shower attention and special prices on new customers to entice them to buy. Other stores in town sold some of the same type of objects as John

did, but they were never quite as rare or special as the ones from John's shop. His items were just a bit rarer, just a bit more exciting than the goods of other shopkeepers. Kids would routinely stop by after school to stare and fantasize. Sometimes a child was able to purchase a treasured item and was envied by his peers. John paid special attention to the children; they were future customers as well. John's most expensive item was the Zygot. They were very fragile. One was kept in a locked showcase in the shop. John had never been robbed, but if he was, it would be the Zygot they were after. He sold them maybe once every 2 months. It was always an event with whomever was in the store at the time. It was like a performance or a ritual. John would take the Zygot from the locked case and wrap it carefully for the trip home. All in the shop would watch with reverence. The customer would beam with pride and ownership, sensing their status rise with their fellow customers. John loves his work at the shop, although he thinks he should view it as just a job. He should have other priorities in his life besides the shop. But his heart tells him that he will probably never retire. His job at the shop means too much to him. He could sell it to someone else and retire, but then what would he do? How would he fill his hours? What about his loyal customers? Would he be letting them down? John finds more than just a way to make a living with his shop. His customers find more than just items for sale. John think this is an excellent way make a living and to help people and make them happy. His customers think the shop is the most interesting in town. It does not carry necessities, only rare objects, but this makes the shop all the more necessary to its customers.

3. Study 2, no-type-reduction, chain 4, generation 3. John owns a candy store that sells six items of particular note: Gurples, Blombies, Pibbles, Gurplesses, Pebbles, and Tribbles. He wakes up each morning and starts the day by taking the long walk to his store. The walk is largely uneventful, but it gives him time to think of all the things he needs to do for the coming business day. Once there he unlocks the doors, steps inside and begins to take inventory and arrange the shelves. Gurples have been selling really well so he puts up a sign that reads 'Gurples Here!'. Sure enough not much later a man comes in and purchases a Gurple for $30 before leaving. A short time later another man comes in and attempts to buy a Gurple for $20. 'Another store sells 3 for $30' the man argues. Haggling is not uncommon in John's store and the man ends up buying the Gurple for $20. Pebbles sell almost as well Gurples and another customer comes in and buys one for $20. Later on a lady and her son enter the store and the child is absolutely excited to get a Gurple. 'Gurple!, Gurple! Gurple!' is all the child can say, but the mother explains she doesn't have the money. Before the mother can list off all the reason why she can't buy her son the candy, John graciously gave the child a Gurple for free. The mother thanks John and she and her son on their way. Later on a lady enters and begins to browse the store. As she inspects the shelves, she unfortunately and carelessly knocks over a container containing a Tribble. Shocked and even somewhat embarrassed she offered to pay the listed price of $40, but John being the honest man that he is refused politely. He swept up the shattered glass that was once the container and safely disposed of it in the refuse bin, while the woman ended up leaving without having purchased anything. Later on a lone child entered the store and requested the off brand version of Gurples, which are called Gurplesses. The child inspected the item, looked over it, and even played with it a bit before putting it back on the shelf and ultimately leaving without buying anything. This was not an odd occurrence, however, as children are rarely able to buy anything on their own and John offered a conciliatory thought of 'at least the child had fun'. John later made a note that Blombies, despite being pretty cheap, haven't sold at all today. He briefly pondered why this was, but he refused to become discouraged. He also noted that he only had 1 Pibble remaining in stock and so filled out all of the necessary paperwork to order more. As the day reached its end John closed up the shop and headed home. The walk home was largely uneventful, and he spent the rest of the evening resting and relaxing, but eager to start the next day.

4. Study 2, no-type-reduction, chain 4, generation 10. A candy shop owner was perhaps among the most charitable in the world. Even though it didn't make any real business sense, he cared more about the needs of his customers than the profitability of his own business. While this might have harmed his success, it overjoyed him all the same. Being able to provide people with a little bit of happiness was all the success he would ever need. His primary and possibly only hope was that his customers would leave his shop happier and more fulfilled than when they entered it. There are many examples of his kindness in this regard. For one, when a particular customer told the shop owner of his ability to purchase the Blombies that he was interested in elsewhere for a

much cheaper price, the shop owner responded most unusually. He offered the Blombies to the customer at the same cheaper price as could be found elsewhere even though it meant a loss of earnings to the owner. This loss didn't bother the owner in the least. In fact, the owner was thrilled to have contributed to the happiness and satisfaction of the customer. Seeing his smile as he left the shop with the candy he had desired meant more to the owner than anything. Another example is a mother and son that were too poor to afford the Dambies that they had always looked at so admiringly in the shop window. The owner had often seen them and offered some of the candies to them on the same day. The mother and son were exceedingly joyful and being able to partake in what they had viewed as a delicacy. Seeing them so pleased was amazing for the owner and he was convinced that his approach was the best route to take as a shop owner. There was, as well, a small boy that always looked at the Domblets and Parxels even though he never purchased any. The shop owner recognized the joy that it brought the boy and allowed him to continue. Sometime after all this on the same day, a distracted woman knocked over a display that featured Melrops and Trinkels. The woman offered to pay and cover the cost of the damage, but the shop owner was uninterested in taking her money. He didn't want to take what was hers and was completely eager to cover it himself. He'd clean it all by himself and wouldn't even begin to concern himself with worsening the happiness of the poor woman. At the end of the day, after all that occurred, he walked home happily. There was absolutely nothing that he would have changed. Even though he might have lost profit and the success of his business wasn't the best, being able to help and satisfy so many people brought him every comfort imaginable. He wouldn't trade this existence for the world and he felt himself undeniably contented with his lot in life. Such was the wonderful life of this meager shop owner, all the world a delight in his eyes.

5. Study 3, no-type-reduction + labels given, chain 5, generation 2. John, the owner of a store, decided while getting ready for work to ride his bicycle to work. Upon arrival at the store, John decided that he needed to make sure he had Plizet ready as this was popular with customers. Sure enough, right after opening the store customers entered the store, and one of them bought a Plizet for $50. John was very thankful and expressed this to the customer. The next customer, an old man, who asked for a Chila, was irate when he found out that he was being charged $20 for a single Chila. He told John that another store, one of good reputation, down the block charged half the price for three Chilas; he only paid $30 for all three! With that the customer left the store without buying the Chila. John chalked this up to the one a day price complaint by a customer. A little while later another eager customer arrived and wanted a Skiger, an expensive product John ordered from France. While this older woman was holding one of the Skigers, she dropped it on the floor and it shattered into many pieces. The Skiger was a loss. The woman was extremely apologetic and offered to pay for the Skiger, though John refused to take her money. She expressed how kind John was for his generosity. After John swept up the pieces and cleaned the floor, he decided to take his well-deserved lunch break. For 30 min he ate and relaxed and prepared himself for a busy afternoon. He hoped he would make some sales this afternoon and would add to his income from the morning. Early in the afternoon, a lovely lady and her beautiful child entered the store. The child, a girl, asked her mom for a Fengle many times, again and again, however; the mom told the child she could not afford to buy one. John overheard this and offered to sell a small piece for a discounted price. The mom was quite thankful for his kindness. The next customer to arrive in the afternoon was a young kid who wanted to buy a Vamey, though he first wanted to try on some of them on his head. John said that was fine with him, so with John's assistance, the kid spent a good amount of time trying several of them on his head. After trying them on and spending all that time the kid decided not to buy any of them. John was not surprised as John knows from experience, kids typically look and don't buy, especially younger ones. The end of the workday had finally come and John closed up the shop. He rode his bicycle home. Boy was he tired after such a busy day. After settling in for the night, he thought about ordering some Nilbos as this was one of his store's best sellers. He typically sold at least one a day. After giving it some thought he went ahead and placed the order for the glorious Nilbos.
6. Study 3, no-type-reduction + labels given, chain 5, generation 7. John is the sole owner and employee of a household furnishing store. The items he sells are unique and quite expensive. The only income John receives is the tiny sliver of profit from his sales in the store. Today is an extremely slow day, and most days consist of customers coming and going but not

buying anything often due to the high prices. One of John's most popular inventions was the Plizet. It sold well in the morning but not through the rest of the day though. The rest of the day he didn't have any luck making a sale. One customer came in and seem genuinely interested in buying a Fengle though, but the customer accidentally dropped it. John is a kind, understanding owner so he told the customer it was okay and didn't charge her for it even though he probably should have. This was not at all a rational response, he eventually realized, as if he couldn't make enough sales he couldn't keep the store up and running. If he couldn't keep the store open then he couldn't provide excellent customer service. He seemed a bit remorseful as he could have really used the sale and believed customers should generally pay for things they break even if it was an accident and the customer was remorseful. If you break it you buy it. After the customer left, he cleaned up all the shattered pieces, went on lunch, and tried to take on a more optimistic outlook. He wanted to believe the evening would see more customers. Not too long after returning from work, a third customer came in and seemed interested in a Chila but ultimately decided against it due to its high price. It was also available at a nearby store for a cheaper price. Despite this John didn't offer the customer a discount. The fourth customer came in with a daughter who seemed excited and interested in a Skiger, but the parent thought it was too expensive to actually purchase it. This time John was willing to be more flexible in hopes of making the sale and decided to try working out a deal with the customer. He offered a discount, but they still ended up passing on it. It was just was not enough of a discount the family. The last customer was a young boy who needed help trying on John's Vameys. John suspected the boy wasn't actually interested in purchasing the Vameys though, and ultimately was proven correct. This made him feel sad during the transition. John closed shop for the day, riding his bike home. He spent some time reflecting on the details of that day. Eventually John remembered how popular the Nilbos were. The Nilbos they sold well enough in the past and he decided to add them back to the catalog. This way he should be able to sell at least one per day as before, which would help make up the loss of the broken Fengle and provide a better inventory selection.