Cognition 253 (2024) 105932

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Zipfian distributions facilitate children's learning of novel word-referent mappings

Lucie Wolters [a,*], Ori Lavi-Rotbain [b], Inbal Arnon [c]

[a] Department of Cognitive and Brain Sciences, The Hebrew University of Jerusalem, Mt Scopus, Israel
[b] The Edmond and Lilly Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Israel
[c] Department of Psychology, The Hebrew University of Jerusalem, Mt Scopus, Israel

ARTICLE INFO

Keywords:
Zipfian distribution
Word learning
Language acquisition
Predictability

ABSTRACT

The word-frequency distributions children hear during language learning are highly skewed (Zipfian). Previous studies suggest that such skewed environments confer a learnability advantage in tasks that require the learner to discover the units that have to be learned, as in word-segmentation or cross-situational learning. This facilitative effect has been attributed to contextual facilitation from high frequency items in learning lower frequency items, and to better learning under the increased predictability (lower entropy) of skewed distributions. Here, we ask whether Zipfian distributions facilitate learning beyond the discovery of units, as expected under the predictability account. We tested children's learning of novel word-referent mappings in a learning task where each mapping was presented in isolation during training, and did not need to be dicovered. We compared learning in a uniform environment to two skewed environments with different entropy levels. Children's learning was overall better in the two skewed environments, even for low frequency items. These results extend the facilitative effect of Zipfian distributions to additional learning tasks and show they can facilitate language learning beyond the discovery of units.

1. Introduction

Language learners extract distributional information from their environment to understand and use the language they hear around them. The structure of this environment can impact learning trajectories and outcomes. One striking commonality in the environment of learners is the way word frequencies are distributed (Zipf, 1949): A small number of words appear very frequently, most words appear infrequently, and frequency decreases exponentially, showing a power law relation between frequency and rank. Zipfian distributions are found across languages and parts of speech (Bentz, Kiela, Hill, & Buttery, 2014; Ferrer i Cancho, 2005; Mehri & Jamaati, 2017; Piantadosi, 2014). Importantly, they have also been identified in the learning environment of children: Both the words children hear (Lavi-Rotbain & Arnon, 2023) and the objects they see (Clerkin, Hart, Rehg, Yu, & Smith, 2017; Lavi-Rotbain & Arnon, 2021) follow a Zipfian distribution.

While the linguistic environment children are exposed to is consistently skewed, many lab-based studies of word learning present learners with a uniform distribution, where each word appears equally often. This enables researchers to control for frequency effects, but does not address the impact of distribution skew on language learning. Recently, a number of studies have explored the impact of Zipfian distributions on learning in lab based settings, and there is growing evidence that learning is improved in Zipfian distributions. Word segmentation is facilitated in both children and adults when word frequencies follow a Zipfian as opposed to a uniform distribution (Kurumada, Meylan, & Frank, 2013; Lavi-Rotbain & Arnon, 2019, 2022; Meylan, Kurumada, Borschinger, Johnson, & Frank, 2012). Similar effects were found for segmentation in the visual domain (Lavi-Rotbain & Arnon, 2021), and for learning novel grammatical categories and syntactic structures (Boyd & Goldberg, 2009; Casenhiser & Goldberg, 2005; Schuler, Reeder, Newport, & Aslin, 2017; Wonnacott, Brown, & Nation, 2017). Cross-situational word learning, where words appear with multiple potential referents and learners have to aggregate co-occurrence statistics to determine the correct mappings, was also facilitated in Zipfian distributions (Hendrickson & Perfors, 2019).

But which properties of Zipfian distributions drive the facilitative effect on learning? One possibility is that the small number of highly frequent items are quickly discovered, providing *contextual facilitation* for the discovery of low frequency items (Kurumada et al., 2013). This

* Corresponding author.
E-mail address: lucia.wolters@mail.huji.ac.il (L. Wolters).

https://doi.org/10.1016/j.cognition.2024.105932
Received 2 November 2023; Received in revised form 7 July 2024; Accepted 21 August 2024
Available online 31 August 2024

can facilitate learning in tasks like word-segmentation and cross-situational learning, where the learner has to discover the relevant linguistic units during learning (e.g. discover the correct words or word-referent mappings). Studies have shown improved segmentation when novel words are presented adjacent to familiar words, supporting the idea that language learners use contextual facilitation in natural language learning (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Dahan & Brent, 1999; Hollich, Jusczyk, & Brent, 2001; Kurumada et al., 2013). Similarly, in cross-situational learning, familiarity with the high frequency mapping allows the learner to decrease referential uncertainty by excluding the familiar referent as a potential mapping for unfamiliar words.

Another possibility is that language learning is facilitated by the greater predictability of Zipfian distributions. Language learning and processing involves forming predictions about which word, object or word-referent mapping will come next (Kray, Sommerfeld, Borovsky, & Häuser, 2024). Zifpian environments are more predictable than uniform or less skewed environments, making it easier to form predictions. This increased predictabiltiy is expected to be facilitative both for discovering linguistic units (as in word segmentation) and for learning about the properties of the units themselves (as in word learning tasks). A recent line of research has explored the effect of distribution predictability on word segmentation (Lavi-Rotbain & Arnon, 2019, 2022). Lavi-Rotbain and Arnon (2022) operationalized distribution predictability using normalized entropy (reffered to as efficiency by the authors), which normalizes word entropy by set size (the number of unique words in the distribution). Normalized entropy ranges from 0 to 1, with higher numbers indicating a less predictable environment: a uniform distribution has a normalized entropy of 1, with skewed distributions having lower values. Corpus work with fifteen languages from eight different language families revealed that normalized word entropy is highly similar across languages, ranging from 0.6 to 0.7 (Lavi-Rotbain & Arnon, 2022, 2023). This language-like entropy was found to be uniquely facilitative for word segmentation in the lab: Children and adults showed improved word segmentation at language-like entropy, compared both to uniform distributions, and to skewed distributions with entropy levels that were higher than those found in natural language (lower than 1, but higher than 0.7).

Existing findings show that distribution predictabiltiy impacts word segmentation, with better learning in distributions that have language-like entropy. However, we do not know if this effect is independent from that of contextual facilitation, and will also be found in tasks where the units do not need to be discovered. In this study, we want to further test the role of distribution predictability (measured using normalized entropy) by asking whether the increased predictability of Zipfian distributions provides a facilitative learning environment in a task where the correct linguistic units are easy to discover. A recent study suggests this may not be the case: Hendrickson and Perfors (2019) looked at the effect of Zipfian distributions on word learning. They found better learning in a Zipfian environment with a cross-situational setting, where participants are presented with multiple words and referents on each trial and need to discover the correct mappings, but not with an isolated training setting, where participants see only one word-referent mapping on each trial. The authors proposed that Zipfian distributions are facilitative only when there is ambiguity about the mappings. However, there is reason to further explore this effect. The entropy of the Zifpian distributions in Hendrickson and Perfors (2019) was lower than that of natural language, at the level where facilitation was not found in previous word segmentation studies (Kurumada et al., 2013; Lavi-Rotbain & Arnon, 2022). Potentially, Zipfian distributions do facilitate the learning of word-referent mappings in an isolated training setting when the distribution has language-like entropy: when it is predictable enough.

In the current study, we look at the effect of Zipfian distributions in a word learning task where only one word-referent mapping is presented on each training trial. In this task, the units that need to be learned are given, mimizing the need for contextual facilitation. We compare learning from a uniform distribution to two skewed distributions[1] with different entropy levels. We want to ask two questions: First, does the decreased entropy of Zipfian distributions have a facilitative effect on language learning that is independent from contextual facilitation? If Zipfian distributions are only facilitative because they provide contextual facilitation when discovering the units, then we should not see improved word learning in a Zipfian environment when each word-referent mapping is presented in isolation. This would be consistent with findings of Hendrickson and Perfors (2019). If, alternatively, distribution predictability has an additive effect on learning, we should see better word learning in Zipfian distributions also when each mapping is presented in isolation. As a second question, we ask whether learning is uniquely facilitated in language-like entropy in tasks where the units are given and do not need to be discovered. This pattern has been previously reported, but only for word segmentation (Lavi-Rotbain and Arnon, 2022). Here, we test this by comparing word learning in two Zipfian distributions: one with language-like entropy, and one with higher entropy than found in natural languages (less predictable than natural language). If learning is uniquely facilitated at language-like entropy, as was found in word segmentation, then we should only see better learning in a Zipfian distribution with language-like entropy. We test these predictions in children for two reasons: First, both accounts assume that the facilitative effect of skewed distributions impacts natural language learning, and therefore should be found in younger learners. Second, while the impact of skew on word segmentation was examined in both children and adults, the impact on word learning was only tested in adults: we do not know how distribution skew impacts children's learning in tasks where the units do not have to be discovered.

2. Learning novel word-referent mappings in uniform and Zipfian environments

We used a word learning experiment where each word-referent mapping is presented is isolation during training. We compared learning of novel word-referent mappings in four conditions that varied in distribution shape (uniform vs. Zipfian) and predictability (maximal entropy, reduced-entropy, and language-like entropy). In all conditions, participants had to learn the names of eight different aliens.

2.1. Materials and method

2.1.1. Participants

273 children (age range: from 4 to 12 years, mean age 7;7; 129 females) took part in the experiment. Children outside the age range of 5;6–9;5 years old were excluded from analysis ($N = 20$).[2] 53 more participants were excluded to have comparable participant numbers in all conditions, leaving 200 participants in the final analysis (mean age 7;6; 97 females).[3] Participants were recruited in the Living Lab at the Bloomfield Science Museum in Jerusalem (https://www.thelivi

[1] Importantly, in our design there is a one-to-one mapping between label and referent, such that the frequency of the meaning and the frequency of the word are the same. What we manipulate is how often each label-referent pairing appears.

[2] Our original intention was to collect a comparable number of participants for each age. However, data collection was stopped due to Covid-19. This resulted in sparse data for the younger and older ages. We excluded the younger and older ages since we wanted to have similarly sized samples across the ages and conditions to increase our ability to correctly estimate the effect sizes of the predictors in our mixed-effects models. However, the results do not change when all children are included. The anlayses including all participants can be found on our OSF page: https://osf.io/8ah3d.

[3] Due to a technical error, more participants were collected for one of the conditions. To have equal numbers of participants across conditions, we only included the initial 50 participants in each condition.

nglabjerusalem.com/). Parental consent was obtained for all participants. None of the children had known language or learning difficulties and all were native Hebrew speakers. Each child received a small prize for their participation.

2.1.2. Stimuli

The visual stimuli consisted of 8 images of aliens in 8 unique colors, as shown in the sample screen in Fig. 1. The audio stimuli consisted of 8 two-syllable novel alien-names, made up of 16 hebrew syllables: 'fee-nam', 'tzo-ked', 'su-leb', 'lil-jeen', 'pli-ret', 'dee-cha', 'cho-ral', and 'zor-ta'. The audio files were generated by the Mac OS software Speech Synthesis Manager.[4]

2.1.3. Procedure

The experiment consisted of a training and test phase.

Training phase. Participants were told they will see aliens and hear their names in a new language. They were instructed to pay close attention to the names, as they will be tested on them in the second part of the experiment. In each trial, an alien appeared on the screen, and after 800 ms participants heard its novel name. The next trial started after 400 ms. The mapping between aliens and names was randomized per participant, as was the order of presentation. Each alien always appeared with the same novel name during training.

Test phase. The test phase started immediately after training. In each trial, participants completed a 4-alternative-forced-choice task: First, four aliens were displayed on the screen in a two-by-two grid, and after 500 ms, they heard the name of one of the aliens. Participants were asked to select the alien whose name they heard by clicking on it. A sample screen is shown in Fig. 1. The three foil aliens for each trial were semi-randomly selected such that each alien appeared 9 times as a foil. Participants were tested three times on each of the eight aliens, resulting in 24 trials. The order of trials was randomized per participant. See Appendix A for the training and test instructions.

2.1.4. Conditions

Participants were randomly assigned to one of four training conditions (see Table 1): two uniform conditions (where each word-referent mapping appeared equally often) and two Zipfian conditions (where the frequency distribution of the mappings was skewed and approximated a power law). In all conditions, participants learned the names of eight different aliens. In the *uniform condition*, participants saw each word-referent mapping 10 times. This condition had a maximal normalized entropy of 1. In the *Zipfian reduced-entropy condition*, the mappings appeared in a skewed frequency distribution (item frequencies: 32, 15, 10, 8, 6, 4, 3, 2) with a normalized entropy of 0.83. This level of entropy is reduced compared to the uniform distribution but higher than the entropy levels found in natural language. In the *Zipfian language-like-entropy condition,* the mappings also appeared in a skewed frequency distribution (item frequencies: 50, 8, 6, 4, 3, 2, 2) but with a normalized entropy of 0.65. This condition is language-like in that its entropy is similar to that of word distributions in natural languages (Lavi-Rotbain & Arnon, 2022). The total exposure in the uniform and Zipfian conditions was the same: participants saw 80 word-referent mappings and the training phase lasted 2.15 min. What differs between them is the frequency distribution of the mappings.

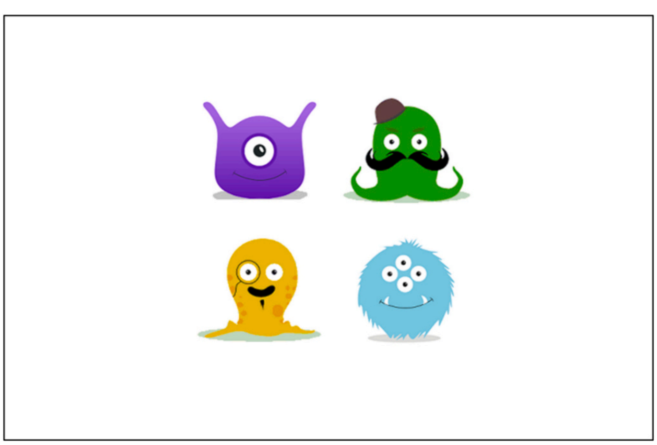


Fig. 1. Sample screen of a single test trial.

In the uniform condition, item frequency was higher than that of the lower frequency items in the Zipfian conditions (the mean frequency of the lower frequency items was 6.9 in the reduced-entropy condition and 4.3 in the language-like-entropy condition). This higher frequency could improve learning in the uniform condition, independent of distribution shape. To control for this, we added a *uniform matched-frequency* condition, where each item appeared only six times (similar to the average frequency of the low frequency items in the Zipfian conditions). This condition consisted of 48 trials, and had a duration of 1.21 min.

2.2. Results

As a first check, we confirmed that participants in all experimental conditions had learned the word-referent mappings above chance (chance was 25% since it was a 4-AFC task). Accuracy was above chance in all four conditions: 68% in the language-like-entropy condition ($t(49) = 15.4$, $p < 0.001$), 67% in the reduced-entropy condition ($t(49) = 15.8$, $p < 0.001$), 54% in the uniform condition ($t(49) = 8.7$, $p < 0.001$), and 45% in the uniform matched-frequency condition ($t(49) = 8.3$, $p < 0.001$).

To examine the effect of distribution shape and predictability on learning, we analyzed the data using a mixed-effects logistic regression models, implemented using the lme4 library (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2023). The model's outcome variable was accuracy on a single test trial (binomial), and included fixed effects of training condition (factorial with 4 levels: 0 = uniform), age in years (centered) and log word frequency (centered). The model included random intercepts for participants and items.[5] We found an effect of condition on accuracy, indicating better learning in each of the Zipfian conditions compared to the uniform condition (reduced-entropy: $\beta = 0.75$, $SE = 0.20$, $z = 3.70$, $p < 0.001$; language-like-entropy: $\beta = 0.93$, $SE = 0.21$, $z = 4.51$, $p < 0.001$). To compare the Zipfian conditions to the uniform matched-frequency condition we releveled the condition variable using the uniform matched-frequency as the baseline (0 = uniform matched-frequency). We found better learning in the Zipfian conditions (reduced-entropy: $\beta = 1.00$, $SE = 0.20$, $z = 5.03$, $p < 0.001$; language-like-entropy: $\beta = 1.18$, $SE = 0.20$, $z = 5.90$, $p < 0.001$). We releveled the condition variable again to compare accuracy in the two Zipfian conditions (0 = reduced-entropy), and found no difference between the language-like-entropy and reduced-entropy conditions ($\beta = 0.18$, $SE = 0.20$, $z = 0.88$, $p = 0.4$). See Table 2 for full model results.

Comparing overall accuracy across the conditions is somewhat misleading since test items in the Zipfian conditions differ greatly in their training frequencies. To ensure that the improved learning in the Zipfian conditions is not just driven by better learning of the high frequency items, we grouped the test items into frequency bins, following Hendrickson and Perfors (2019). Mid frequency items appeared 8 or 10

[4] All images and audio files can be accessed on our OSF page: https://osf.io/8ah3d.

[5] We did not include trial number as a fixed effect in the reported models as we did not find an effect of trial number in any of the reported models. Models including a fixed effect of trial number can be accessed on our OSF page: https://osf.io/8ah3d.

Table 1
Training conditions.

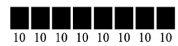
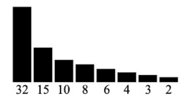
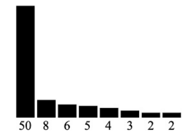
	Uniform		Zipfian	
	Matched frequency	Uniform	Reduced entropy	Language-like entropy
Normalized entropy	1	1	0.83	0.65
Unigram entropy	2.25	3	2.50	1.95
Number of tokens	48	80	80	80
Mean freq. of lower frequency items	6	10	6.9	4.3
Item frequency distribution (y-axis = 0, 50)	 6 6 6 6 6 6 6 6	 10 10 10 10 10 10 10 10	 32 15 10 8 6 4 3 2	 50 8 6 5 4 3 2 2

Table 2
Results of a mixed-effects logistic model on the complete data. The outcome variable was accuracy on a single test trial and the model had random intercepts for participants and alien-name. The intercept represents centered age, centered frequency and the uniform condition.

	Estimate	Std. Error	Z value	*p*-value
(intercept)	0.28	0.16	1.75	0.08
Age (centred)	0.26	0.09	3.00	< 0.01
Log frequency (centred)	0.48	0.06	8.39	< 0.001
Uniform matched-frequency condition	−0.25	0.20	−1.25	0.21
Zipfian reduced-entropy condition	0.75	0.20	3.70	< 0.001
Zipfian language-like-entropy condition	0.93	0.21	4.50	< 0.001

times and were similar in frequency to items in the uniform condition; low frequency items appeared less than 8 times (2, 3, 4, 5 or 6 times, mean: 3.7) and were similar in frequency to items in the uniform matched-frequency condition, and high frequency items appeared more than 10 times (15, 32 or 50 times, mean 32.3). Fig. 2 shows accuracy in the three frequency bins (high, mid, low) by training condition: Both the mid and low frequency items were learned better compared to the items of similar frequency in the uniform conditions (reduced-mid: 67%, language-like-mid: 70%, reduced-low: 63%, language-like-low: 63%).

We compared performance on the mid frequency items from the Zipfian conditions to the uniform condition, and performance on the low

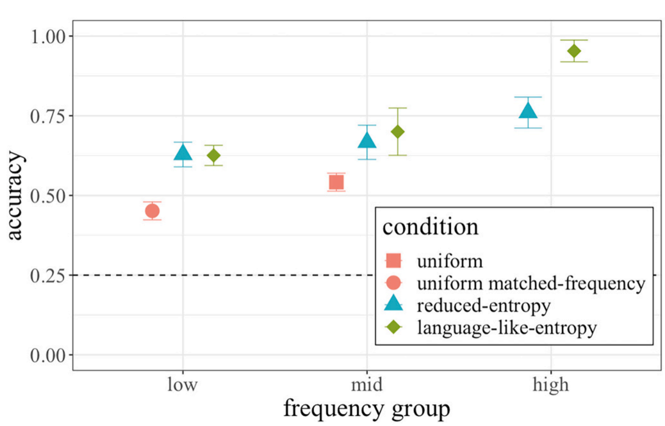


Fig. 2. Accuracy in the testing phase. The dashed line indicates chance performance of 25%. Error bars reflect 95% confidence intervals. Test items were grouped according to frequency: Mid-frequency items had a similar frequency as items in the uniform condition (8 and 10 times); low-frequency items appeared less than 8 times (2, 3, 4, 5 or 6 times, mean: 3.6), and high-frequency items appeared more than 10 times (15, 32 or 50 times).

frequency items from the Zipfian conditions to the uniform matched-frequency condition. To analyze performance on the mid frequency items, we used the same model as for the complete data but excluded the fixed effect of frequency, since the items were matched on frequency. Accuracy was higher in each of the Zipfian conditions (language-like-entropy: 70%, reduced-entropy: 67%) compared to the uniform condition (54%) (reduced-entropy: $\beta = 0.56$, SE = 0.27, z = 2.06, $p < 0.05$; language-like-entropy: $\beta = 0.73$, SE = 0.31, z = 2.36, p < 0.05). Releveling the condition variable (0 = reduced-entropy) revealed no difference in learning between the language-like-entropy and the reduced-entropy conditions ($\beta = 0.17$, SE = 0.33, z = 0.53, $p = 0.60$). See Table 3 for full model results.

We also compared performance on the low frequency items in the Zipfian conditions to the uniform matched-frequency condition. Here also, accuracy was higher in each of the Zipfian conditions (language-like-entropy: 77%, reduced-entropy: 66%) compared to the uniform matched-frequency condition (45%) (reduced-entropy: $\beta = 0.78$, SE = 0.21, z = 3.69, $p < 0.001$; language-like-entropy: $\beta = 0.76$, SE = 0.21, z = 3.70, p < 0.001). This occurred despite the lower mean item frequency in the Zipfian conditions (mean 3.7 vs. 6). See Table 4 for full model results. That is, accuracy was higher overall, for matched frequency items, and for lower frequency items in the two Zipfian conditions compared to the uniform ones.

Some of the test trials in the Zipfian conditions included the high frequency item as a foil. In these trials, familiarity with the high frequency item could have provided contextual facilitation in choosing the correct name during testing: Eliminating the high frequency foil reduces the options. To ensure that the improved performance in the Zipfian conditions is not driven by these trials, we looked at the effect of high frequency foils on the learning of the low and mid frequency items. To do so, we analyzed the Zipfian conditions with a logistic mixed-effects model that included fixed effects of age (centered) and high-frequency foil (binary), and random intercepts for participants and items. For the mid-frequency items, we found no effect of high frequency foil on accuracy ($\beta = 0.17$, SE = 0.25, z = 0.68, $p = 0.5$). However, we did find such an effect for the low-frequency items, indicating better performance on trials that had high-frequency foils ($\beta = 0.37$, SE = 0.12, z = 3.05, $p < 0.01$). To ensure that the better accuracy on the low-frequency

Table 3
Results of a mixed-effects logistic model on the mid frequency items (8 and 10). The outcome variable was accuracy on a single trial during testing and the model had random intercepts for participants and alien-name. The intercept represents centered age and the uniform condition.

	Estimate	Std. Error	Z value	*p*-value
(intercept)	0.29	0.19	1.51	0.13
Age (centred)	0.19	0.16	1.19	0.24
Zipfian reduced-entropy condition	0.56	0.27	2.06	< 0.05
Zipfian language-like-entropy condition	0.73	0.31	2.36	< 0.05

Table 4
Results of a mixed-effects logistic model on a subset of the data including the uniform matched-frequency condition and the low frequency items of the Zipfian conditions (< 8). The outcome variable was accuracy on a single trial during testing and the model had random intercepts for participants and alien-name. The intercept represents centered frequency, centered age and the uniform training condition.

	Estimate	Std. Error	Z value	*p*-value
(intercept)	−0.26	0.17	−1.53	0.13
Age (centred)	0.21	0.09	2.25	< 0.05
Log frequency (centred)	−0.09	0.14	−0.67	0.5
Zipfian reduced-entropy condition	0.78	0.21	3.69	< 0.001
Zipfian language-like-entropy condition	0.76	0.21	3.70	< 0.001

items (compared to the uniform matched-frequency condition) was not completely driven by high frequency foils, we repeated the analysis of the low-frequency items, this time excluding trials with a high frequency foil. The effect of condition remained significant: accuracy was higher in both Zipfian conditions compared to the uniform matched-frequency condition (reduced-entropy: $\beta = 0.77$, SE $= 0.22$, z $= 3.52$, $p < 0.001$; language-like-entropy: $\beta = 0.60$, SE $= 0.21$, z $= 2.88$, $p < 0.01$). See Table 5 for full model results.

In sum, we found improved learning of word-referent mappings in the two Zipfian conditions compared to the uniform ones. Accuracy was higher in the Zipfian conditions even after taking into account the possible effect of the high frequency foil.[6]

3. Discussion

Children are exposed to Zipfian frequency distributions during language learning, which has been found to facilitate various aspects of language language. However, it is still unclear which properties of Zipfian distributions drive this facilitative effect and whether it is restricted to certain learning tasks. This study set out to further explore the effect of Zipfian distributions on word-referent learning by children. In particular, we asked (1) whether Zipfian distributions are facilitative when the lingustic units are given and do not need to be discovered, and (2) whether, in such tasks, Zipfian distributions will uniquely facilitate learning at entropy levels similar to those found across natural languages (as was found for word segmentation).

Our results show that Zipfian distributions facilitate learning in a task where the linguistic units are given: Children learned word-referent mappings better when exposed to Zipfian distributions compared to uniform ones. Importantly, while item frequency positively affected accuracy, the effect of distribution held after controlling for frequency: When grouping the test items according to their training frequency, we found better learning of low and mid frequency items in the Zipfian distributions compared to items with comparable frequency in the uniform distributions. This effect is unlikely to reflect contextual facilitation from high-frequency items, for two reasons: The training design minimized the need for contextual facilitation (since each mapping was presented in isolation), and the effect of distribution was still significant when looking only at the test trials that did not have a high-frequency foil. These findings suggest that distribution predictability has a facilitative effect on learning that is independent from that of contextual facilitation. In contrast with previous findings, we did not find support for increased facilitation from language-like entropy: Zipfian distributions with language-like entropy improved learning to a similar degree as less predictable Zifpian distributions.

Our findings differ from prior research in several interesting ways. In terms of entropy levels, our results differ from those found for word segmentation, where learning was only facilitated by language-like entropy (Lavi-Rotbain & Arnon, 2022). The level of distribution entropy seems to matter for some tasks (word segmentation), but not others (word learning). One possibility is that different learning objectives benefit from different ranges of entropy: in particular, tasks that involve unit discovery may benefit from more skewed distributions than tasks that only involve learning about the properties of the units. In other words, discovering units and learning about them at the same time (which is how actual language learning happens), requires a distribution that is skewed enough, whereas for just learning about the units, any skew is better than a uniform distribution. This possibility can be tested by exploring the effect of different entropy levels on a task that involves both unit identification and mapping the units to objects, like in cross-situational learning.

Our results differ from a previous word learning study, where Zipfian distributions with entropy levels similar to our reduced-entropy condition improved accuracy in a cross-situational setting, but not in an isolated training setting (Hendrickson & Perfors, 2019). What can explain this difference? One possibility is that the effect of distribution is modulated by age and task type: we tested children, while Hendrickson and Perfors (2019) tested adults. It is possible that adults do not benefit from skewed distributions in a relatively simple learning task, while children do. However, this possibility does not seem likely given that adults did find the task difficult: they were far from ceiling in all conditions (55% accuracy in the uniform condition and 46% accuracy in the Zipfian conditions). A closer look at the results suggests that the lack of difference between the uniform and the Zipfian conditions in the isolated training setting stems from improved performance in the uniform condition in comparison to the uniform condition in the cross-situational setting: the accuracy in the Zipfian distributions was similar in the cross-situational and isolated training setting (43% vs. 48%), while accuracy in the uniform condition improved significantly (34% vs. 56%). This is somewhat surprising since the task should have been easier in both distributions (we would have expected accuracy levels to increase relative to the cross-situational setting in both distributions). More work is needed to see if adults really do not benefit from Zipfian distributions in tasks that do not require contextual facilitation. Regardless of the specific results, future research should explore whether children benefit more from the greater predictability of Zipfian distributions in comparison to adults, and whether there may be

Table 5
Results of a mixed-effects logistic model on a subset of the data including the uniform conditions and the low frequency items of the Zipfian conditions (< 8) but excluding all trials with a high frequency foil. The outcome variable was accuracy on a single trial during testing and the model had random intercepts for participants and alien-name. The intercept represents centered frequency, centered age and the uniform training condition.

	Estimate	Std. Error	Z value	*p*-value
(intercept)	−0.20	0.17	−1.18	0.24
Age (centred)	0.19	0.09	2.18	< 0.05
Log frequency (centred)	0.03	0.18	0.17	0.87
Zipfian reduced-entropy condition	0.77	0.22	3.52	< 0.001
Zipfian language-like-entropy condition	0.60	0.21	2.88	< 0.01

[6] All data and analyses can be accessed on our OSF page:https://osf.io/8ah3d.

developmental changes in the sensitivity to skewed distributions.[7]

Our study provides the first evidence that Zipfian distributions can facilitate learning when there is no need to discover the linguistic units. However, more research is required to understand what mechanisms drive facilitation in such contexts. Error-based theories of language acquisition suggest that children, like adults, continuously make and evaluate predictions in order to acquire and use language (Clark, 2018; Kray et al., 2024). In this framework, encountering unexpected input leads to prediction error, which leads to changes in the relevant representations. Recent work has explored the impact of prediction error on language learning (Fazekas, Jessop, Pine, & Rowland, 2020; Gambi, Pickering, & Rabagliati, 2021; Reuter, Borovsky, & Lew-Williams, 2019). For example, Fazekas et al. (2020) found enhanced learning for the same syntactic structure when it appeared in surprising as opposed to predictable context, in both adults and children. Another study (Reuter et al., 2019) found that prediction error supported the learning of novel word-referent mappings in 3- to 5-year olds, via efficient redirection of attention (but see Gambi et al., 2021 for conflicting results). Seen from this perspective, Zipfian environments provide more opportunities for making predictions and for learning from incorrect predictions since they contain a mix of high frequency items (predictable) and lower frequency ones (less predictable). In our study, exposure to a Zipfian distribution of word-referent pairings can lead the learner to predict a high frequency pairing on a next trial. If that is not the case (i.e. they see a lower frequency pairing), this can generate a prediction error, potentially boosting the learning of the unexpected pairing. That is, the presence of few low frequency items amongst many high frequency items will lead learners to form expectations about upcoming items and encounter items that violate those expectations. This framing generates a prediction to be tested in future work: We expect that learning from a Zipfian distribution generates more prediction error than learning from a uniform distribution. Morover, we expect that the low-frequency items in a Zipfian distribution will generate more prediction error than higher frequency items.

A better understanding of the properties that make Zipfian distributions beneficial for learning can direct research on their emergence and persistence in natural language. Language is shaped by a pressure to be learnable by being repeatedly transmitted from one generation of language learners to another generation of langauge learners (Kirby, Cornish, & Smith, 2008; Raviv, Meyer, & Lev-Ari, 2019). Zipfian distributions may be frequent in our learning environment precisely because they facilitate learning. This proposal is supported by recent work showing that Zipfian distributions emerge in an iterated sequence learning experiment, and that their emergence increases the reproduction accuracy of the sequence sets (Arnon & Kirby, 2024). This suggests that learning may play an important role in shaping the statistical structure of language. However, it is important to note that the facilitative properties of Zipfian distributions discussed in this paper apply to all skewed distributions. One property that is specific to power-law distributions is their scale-invariance, which means that the shape of the frequency distribution characterizes the distribution at different scales (Chater & Brown, 1999; Kello et al., 2010). The exact distribution shapes observed across languages are likely the result of an interaction between multiple pressures, including pressures from learning or from communication.

CRediT authorship contribution statement

Lucie Wolters: Formal analysis, Methodology, Writing – original draft. **Ori Lavi-Rotbain:** Conceptualization, Methodology. **Inbal Arnon:** Conceptualization, Writing – review & editing.

Declaration of competing interest

None.

Data availability

The data and analyses can be found at https://osf.io/8ah3d.

Acknowledgment

We thank the Living Lab staff and the Bloomfield Science Museum in Jerusalem, as well as the parents and children who participated. The research was funded by the Israeli Science Foundation grant 445/20 awarded to the second author.

Appendix A. Experiment instructions for participants (translated from Hebrew into English)

Instructions training phase: "Welcome to the game of smiles where you will meet new friends! Each time you see an alien on the screen and hear its name over the headphones. Try to get to know the aliens as well as possible, because in the second part you will be tested on them. Press the space bar when you are ready to start. Good luck!"

Instructions test phase: "Now let's see how much you know about our aliens… Each time four different aliens will appear on the screen and you will hear one of their names over the headphones. Click on the alien whose name you heard. If you do not know – guess. Press the spacebar when you are ready to start. Good luck!"

References

Arnon, I., & Kirby, S. (2024). Cultural evolution creates the statistical structure of language. *Scientific Reports, 14*, 5255.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4 (arXiv:1406.5823). *arXiv*. https://doi.org/10.48550/arXiv.1406.5823

Bentz, C., Kiela, D., Hill, F., & Buttery, P. (2014). Zipf's law and the grammar of languages: A quantitative study of Old and Modern English parallel texts. *Corpus Linguistics and Linguistic Theory, 10*(2), 175–211. https://doi.org/10.1515/cllt-2014-0009

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16*(4), 298–304. https://doi.org/10.1111/j.0956-7976.2005.01531.x

Boyd, J. K., & Goldberg, A. E. (2009). Input effects within a constructionist framework. *The Modern Language Journal, 93*(3), 418–429. https://doi.org/10.1111/j.1540-4781.2009.00899.x

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science, 8*(6), 500–508. https://doi.org/10.1111/j.1467-7687.2005.00441.x

Chater, N., & Brown, G. D. A. (1999). Scale-invariance as a unifying psychological principle. *Cognition, 69*(3), B17–B24. https://doi.org/10.1016/S0010-0277(98)00066-3

Clark, A. (2018). A nice surprise? Predictive processing and the active pursuit of novelty. *Phenomenology and the Cognitive Sciences, 17*(3), 521–534. https://doi.org/10.1007/s11097-017-9525-z

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1711), 20160055. https://doi.org/10.1098/rstb.2016.0055

Dahan, D., & Brent, M. R. (1999). On the discovery of novel wordlike units from utterances: An artificial-language study with implications for native-language acquisition. *Journal of Experimental Psychology: General, 128*, 165–185. https://doi.org/10.1037/0096-3445.128.2.165

Fazekas, J., Jessop, A., Pine, J., & Rowland, C. (2020). Do children learn from their prediction mistakes? A registered report evaluating error-based theories of language acquisition. *Royal Society Open Science, 7*(11), Article 180877. https://doi.org/10.1098/rsos.180877

Ferrer i Cancho, R. (2005). The variation of Zipf's law in human language. *The European Physical Journal B - Condensed Matter and Complex Systems, 44*, 249–257. https://doi.org/10.1140/epjb/e2005-00121-8

Gambi, C., Pickering, M. J., & Rabagliati, H. (2021). Prediction error boosts retention of novel words in adults but not in children. *Cognition, 211*, Article 104650. https://doi.org/10.1016/j.cognition.2021.104650

[7] We did not see such developmental changes in our data. Adding an interaction between age and condition did not improve our models when looking at the entire data set or when looking only at the frequency matched items. These analyses can be accessed on our OSF page: https://osf.io/8ah3d.

Hendrickson, A. T., & Perfors, A. (2019). Cross-situational learning in a Zipfian environment. *Cognition, 189*, 11–22. https://doi.org/10.1016/j.cognition.2019.03.005

Hollich, G., Jusczyk, P. W., & Brent, M. R. (2001). How infants use the words they know to learn new words. In *Proceedings of the 25th annual Boston University conference on language development* (pp. 353–364).

Kello, C. T., Brown, G. D. A., Ferrer-i-Cancho, R., Holden, J. G., Linkenkaer-Hansen, K., Rhodes, T., & Van Orden, G. C. (2010). Scaling laws in cognitive sciences. *Trends in Cognitive Sciences, 14*(5), 223–232. https://doi.org/10.1016/j.tics.2010.02.005

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*(31), 10681–10686. https://doi.org/10.1073/pnas.0707835105

Kray, J., Sommerfeld, L., Borovsky, A., & Häuser, K. (2024). The role of prediction error in the development of language learning and memory. *Child Development Perspectives, 1–14*. https://doi.org/10.1111/cdep.12515

Kurumada, C., Meylan, S. C., & Frank, M. C. (2013). Zipfian frequency distributions facilitate word segmentation in context. *Cognition, 127*(3), 439–453. https://doi.org/10.1016/j.cognition.2013.02.002

Lavi-Rotbain, O., & Arnon, I. (2019). Children learn words better in low entropy. *Cognition, 223*, 1–15.

Lavi-Rotbain, O., & Arnon, I. (2021). Visual statistical learning is facilitated in Zipfian distributions. *Cognition, 206*, Article 104492. https://doi.org/10.1016/j.cognition.2020.104492

Lavi-Rotbain, O., & Arnon, I. (2022). The learnability consequences of Zipfian distributions in language. *Cognition, 223*, Article 105038. https://doi.org/10.1016/j.cognition.2022.105038

Lavi-Rotbain, O., & Arnon, I. (2023). Zipfian distributions in child-directed speech. *Open Mind, 7*, 1–30.

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*(31), 2470–2477. https://doi.org/10.1016/j.physleta.2017.05.061

Meylan, S., Kurumada, C., Borschinger, B., Johnson, M., & Frank, M. C. (2012). Modeling online word segmentation performance in structured artificial languages. *Proceedings of the Annual Meeting of the Cognitive Science Society, 34*(34).

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review, 21*(5), 1112–1130. https://doi.org/10.3758/s13423-014-0585-6

R Core Team. (2023). _*R: A Language and Environment for Statistical Computing*_. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Compositional structure can emerge without generational transmission. *Cognition, 182*, 151–164. https://doi.org/10.1016/j.cognition.2018.09.010

Reuter, T., Borovsky, A., & Lew-Williams, C. (2019). Predict and redirect: Prediction errors support children's word learning. *Developmental Psychology, 55*(8), 1656–1665. https://doi.org/10.1037/dev0000754

Schuler, K. D., Reeder, P. A., Newport, E. L., & Aslin, R. N. (2017). The effect of Zipfian frequency variations on category formation in adult artificial language learning. *Language Learning and Development, 13*(4), 357–374. https://doi.org/10.1080/15475441.2016.1263571

Wonnacott, E., Brown, H., & Nation, K. (2017). Skewing the evidence: The effect of input structure on child and adult learning of lexically based patterns in an artificial language. *Journal of Memory and Language, 95*, 36–48. https://doi.org/10.1016/j.jml.2017.01.005

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Addison-Wesley Press.