Near-Zipfian Distribution is Prevalent in Infant Input

Brianna E. Kaplan (brianna.kaplan@austin.utexas.edu)
Department of Psychology and Center for Perceptual Systems, 108 E Dean Keeton Street
Austin, TX 78712 USA

Chen Yu (chen.yu@austin.utexas.edu)
Department of Psychology and Center for Perceptual Systems, 108 E Dean Keeton Street
Austin, TX 78712 USA

Abstract

Understanding infants' natural input is essential for advancing theories of cognitive development and learning. Recent research indicates that across modalities, infant input approximates a near-Zipfian distribution, with a large amount of input about a few items and substantially less about the rest. However, prior work has only examined aggregated distributions across subjects, focused on a single modality in isolation, and considered the input available to infants rather than what they actively select. We show that at both the corpus and individual levels, infant attention selection and the verbal input infants receive from parents follows a near-Zipfian distribution. Moreover, when integrating across modalities, the verbal input infants hear while attending to the same object becomes even more skewed than verbal input alone. Findings suggest that Zipfian-like structure is not only a property of infant environments but emerges through active selection, highlighting its potential role in shaping early learning.

Keywords: zipfian distribution; infant attention; multimodal input; learning input

Introduction

To understand early learning and development, past research has primarily focused on examining the internal mechanisms of the learning system. One effective approach has been to use laboratory experiments that provide learning input and systematically measure learning outcomes. By linking different types of input to outcomes, researchers gain insight into how the learning system operates. However, real-world learning depends not only on the learning system itself, but also on the nature of the input it receives. Even with the same learning system, variations in input lead to different outcomes. Recognizing this, recent research has shifted toward quantifying the learning input available to infants and children.

The first domain to adopt this approach was language. Studies have shown that child-directed speech, like natural human language, follows a heavily skewed pattern, known as a Zipfian distribution (Figure 1A), where a small number of words (e.g., "the," "a," "is") are used frequently while most words are used infrequently (Piantadosi, 2014; Zipf, 1946) (Hendrickson & Perfors, 2019; Lavi-Rotbain & Arnon, 2022, 2023; Wolters, Lavi-Rotbain, & Arnon, 2024). Mathematically, this distribution is characterized by a linear relation between log rank and log frequency (a power law; second plot in Figure 1A). This pattern has been observed across languages worldwide (Zipf, 1946) and even in some forms of sign language (Kimchi, Wolters, Stamp, & Arnon, 2023). However, natural language does not perfectly adhere to Zipf's law, with consistent deviations for high frequency words, and sometimes also for lower frequency words (Figure 1B; Montemurro, 2001; Piantadosi, 2014). As a result, researchers often refer to language distributions as "near-Zipfian," a term we adopt in this work.

This skewed distribution is not limited to language. A study using head-mounted cameras showed that the visual input available to infants—such as the objects they encounter during everyday activity—also exhibits a highly skewed distribution (Clerkin, Hart, Rehg, Yu, & Smith, 2017). This pattern extends to combinations of objects (Lavi-Rotbain & Arnon, 2021) and even the faces infants see (Jayaraman, Fausey, & Smith, 2015). These findings suggest that skewed

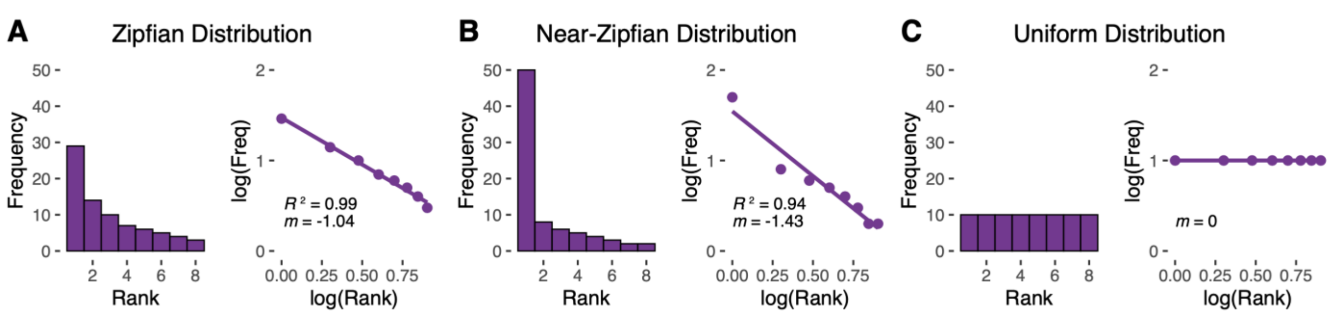

Figure 1. Examples of possible types of distributions. The R^2 for each comes from the linear model between the logs of frequency and rank; *m* represents the slope of the line. A) True Zipfian distribution. B) The typical distribution of human language, often referred to as near-Zipfian. C) Uniform distribution. Note that R^2 is not calculated for unform distributions.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

distributions may be a fundamental property of infants' natural input across multiple domains.

Moreover, increasing evidence suggests that skewed distributions are not incidental but serve a functional role in learning. Traditional studies of learning in adults, children, and infants often use uniform distributions of input (Figure 1C) to control for frequency effects (e.g., Saffran, Aslin, & Newport, 1996). However, recent research demonstrates that learners often achieve better outcomes when exposed to input that follows a skewed or Zipfian distribution. This advantage has been observed across a range of inputs, including word-referent mappings, word segmentation, and grammatical categories (Hendrickson & Perfors, 2019) (Kurumada, Meylan, & Frank, 2013; Lavi-Rotbain & Arnon, 2021; Wolters et al., 2024). Computational models also show superior performance when trained on data with a skewed distribution compared to a uniform one (Bambach, Crandall, Smith, & Yu, 2018).

These advances raise several open questions. First, most existing studies have analyzed input at the corpus level, aggregating data across many individuals. As a result, it remains unclear whether the input to individual infants follows a skewed distribution or whether the observed skew emerges only at the population level. Second, prior research has only examined the distribution of a single modality at a time, despite the fact that infants simultaneously acquire information across multiple modalities—e.g., visual, auditory, and tactile. It is unknown whether in the same moments, input in all modalities follows the same distribution or whether skew varies across modalities and contexts. Finally, past research has focused on learning environments as a proxy for learning input. For instance, studies such as Clerkin et al. (2017) have demonstrated that the availability of infants' visual input is skewed. However, attention gates what enters the learning system, meaning that learning input is not just what is present in the environment but what infants actively select. Accordingly, one goal of this study was to quantify what information in the environment is selected by learners (i.e., what they actually choose to look at and touch).

The current study addresses these gaps by examining the distribution of infants' natural input across multiple modalities and levels of analysis. The data were collected in a home-like laboratory environment (Figure 2) in which parents and infants played together with 27 toys. Both members of the dyad wore head-mounted eye trackers, allowing us to precisely measure how infants selected visual information (Figure 3). Additionally, we coded infants' object handling and parent speech to capture input across the visual, auditory, and tactile modalities. This work has three aims:

1. Examine the distribution of infants' active attention selection. We predicted that both visual and manual attention selection follow a skewed (near-Zipfian) distribution.
2. Analyze the distribution of the verbal input infants receive from their parents. We predicted that the distribution would mirror the near-Zipfian skew found in previous studies of child-directed speech (Lavi-Rotbain & Arnon, 2023). Additionally, we investigated whether this skew remained when focusing on content-related input, such as nouns and action verbs, removing high-frequency function words. We predicted that the distribution of nouns and action verbs would remain skewed, albeit less Zipfian than overall speech.
3. Investigate how the distribution changes when integrating input from multiple modalities. Infants do not learn one modality at a time, nor do they process information from a single modality in isolation. For example, language learning requires mapping words to visual referents. We predicted that the distribution of words infants hear while attending to the same object remains skewed.

For each aim, we analyzed both corpus-level distributions and individual-level variability, providing a comprehensive understanding of the statistical structure of infants' natural input.

Method

Participants

The study included both cross-sectional and longitudinal data, with a total of 35 parent-child dyads (15 female children) participating in one to four sessions, spaced approximately three months apart. In total, 57 play sessions were conducted. Among the participants, 19 dyads attended a single session, 11 attended two sessions, four participated in three sessions, and one dyad completed all four sessions. Children's ages ranged from 14 to 31 months (M = 21.25 months, SD = 4.70). According to parental reports, 86% of children were as White, 6% Asian, 3% Native Hawaiian or Pacific Islander, and 6% multi-racial. Eighty percent of

Figure 2. Parent and infant playing in the Home Lab.

children were reported as non-Hispanic, and 20% were reported as Hispanic.

Procedure

Parents and children played in a home-like laboratory environment with 27 toys (animals, vehicles, and food items). Dyads wore head-mounted eye-trackers (Pupil Labs NEON) and parents were told to play like they normally would at home. Sessions lasted 10 minutes or until children removed the eye tracker (M = 8.22 minutes, range = 1.68 – 12.28). Each eye tracker consisted of three cameras, one that was centered on the forehead and recorded the participant's field of view and two that recorded the participant's eyes. Calibration was done automatically by the eye tracker after a small online correction by the experimenter.

Behavioral Coding

Following the play session, infants' visual attention and manual actions were coded using Datavyu (Figure 3). Fixations were automatically detected by Pupil Labs software and subsequently imported into Datavyu for coding. Coders classified each fixation according to the region of interest (ROI); fixations directed elsewhere were excluded from analysis. Manual actions were coded separately for the left and right hands, with any physical contact with an object included. Parental speech was automatically transcribed using Whisper software and then reviewed and corrected by coders.

Object Tracking

We retrained the last recognition layer of a pre-trained deep learning model (YOLOv8) to automatically detect individual objects in view (Farhadi & Redmon, 2018). The trained model provided up to 27 bounding boxes per frame. Four toys (all dark in color) did not provide reliable tracking data (not detected in more than 40% of child ROI instances) and were removed from all analyses.

Data Analysis

We compared the structure of infants' natural input for each modality to that of a near-Zipfian distribution, following the methods of Lavi-Rotbain and Arnon (2023). We first estimated the parameters of the distribution using maximum likelihood estimation:

$$f(r) \propto \frac{1}{(r+\beta)^{\alpha}}$$

Here, r is the word's rank; α is the exponent of the power law, which determines the steepness of this distribution; and β is a correction term added by Mandelbrot (1953) to improve the fit to real-world data. Using the fitted values of α and β, we computed the expected frequency of each rank under a near-Zipfian distribution. We then calculated the Pearson correlation between the observed and expected frequencies. We applied this approach both to the group-level data and to individual participants.

Results

Distribution of Infant Attention Selection

We first examined whether infants' active visual and manual attention selection follows a near-Zipfian distribution.

Visual Object Selection To analyze visual attention selection, we determined the proportion of the session each object was in view and then calculated the proportion of that

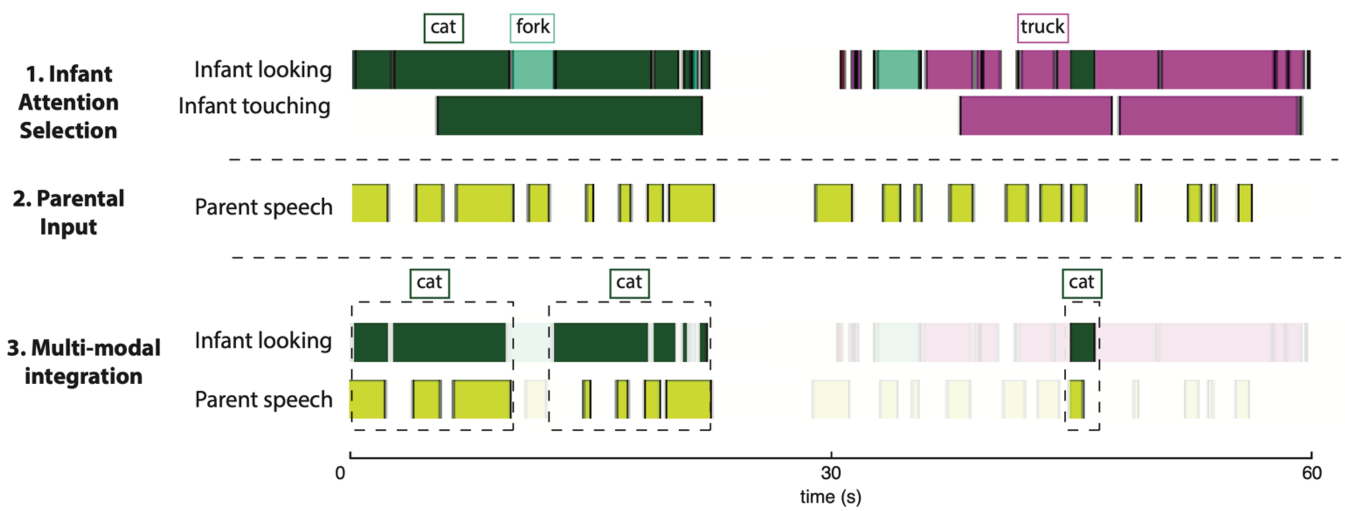

Figure 3. Data streams of coded behaviors for a single dyad. Infant looking and touching were coded to analyze the distribution of active attention selection. Parent speech was coded to examine the distribution of parental input. Nouns and verbs were also noted. To assess multi-modal integration, infant looking and parent speech were combined to analyze the distribution of the nouns and verbs infants heard while visually attending to each object. The third section illustrates an example of this cross-modal combination for a single object (cat). Only the nouns and verbs the child heard while visually attending to the cat were counted.

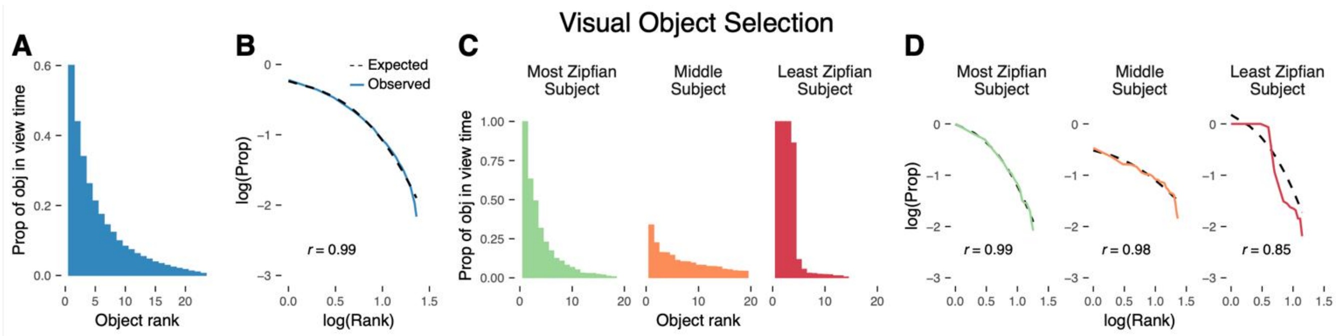

Figure 4. A) Distribution of visual object selection. B) Observed and expected distributions on a log-log scale. C) Distributions of individual participants—most Zipfian, median, and least Zipfian. D) Log-log plots of rank and proportion for individual participants.

time during which infants attended to the object (i.e., looking time relative to availability). Objects were ranked by attention for each child, and for corpus-level analyses, attention was averaged across participants for each object rank.

As shown in Figure 4A, visual attention selection was highly skewed. The mean (13%) and median (6%) of looking time were notably different and the measure of skewness (1.66) exceeded one, indicating a substantial skew. Figure 4B compares the observed distribution with the expected shape of a near-Zipfian distribution. The two curves closely align, reflecting a strong correspondence between the observed and predicted frequencies. This is further supported by a high Pearson correlation (r = 0.99), indicating that the estimated parameters provided an excellent fit to the data. The only slight deviation from the expected data occurred at the extreme end, a pattern consistent with natural language distributions (Montemurro, 2001; Piantadosi, 2014).

Individual-level analyses reinforced these findings. Figures 4C and 4D show the observed and expected distributions for the most Zipfian subject (r = .99; green), the median subject (r = .98; orange), and the least Zipfian subject (r = .85; red). Even the least Zipfian participant demonstrated a strong correlation with the expected near-Zipfian distribution. Overall, the average correlation across all subjects was r = .98 (SD = .02), indicating a robust fit for nearly all participants.

Infants rarely looked at an object for the entire duration it was in view. Across all subjects, only 20 out of 1,311 objects were attended to for 100% of their in-view time, and half of these were viewed for less than 15 seconds in total. Even for the highest-ranked object per subject, infants attended to it only 60% of the time it was visible. These findings suggest that while headcam data capture what is visually available to infants, they overestimate the input infants actually attend to. Nonetheless, the distribution of attended objects remains near-Zipfian.

Manual Object Selection We applied a similar process to analyze infants' manual object selection by calculating the proportion of the session during which infants touched each object. Proportions rather than raw time were used to facilitate comparisons across subjects with different session lengths. Because infants could touch multiple objects simultaneously, values did not sum to one.

Figure 5A shows that manual object selection was also highly skewed. The mean (4%) and the median (2%) of object touching differed, and the skewness measure exceeded one (1.87). As with visual object selection, the observed and expected distributions closely aligned (Figure 5B), with a high correlation (r = .99), indicating a near-Zipfian fit. Deviations again emerged at the tail end of the distribution, with children interacting with the least-touched objects less than expected.

Figure 5. A) Distribution of manual object selection. B) Observed and expected distributions on a log-log scale. C) Distributions of individual participants—most Zipfian, median, and least Zipfian. D) Log-log plots of rank and proportion for individual participants.

Individual-level analyses confirmed this pattern. Figures 5C and 5D depict participant distributions, with even the least Zipfian subject (r = .95) still exhibiting a skewed pattern and alignment with the expected curve. Across participants, the average correlation was .98 (SD = .02), highlighting the robustness of this pattern.

Distribution of Parent Speech

We next asked if parents' speech followed a near-Zipfian distribution.

All Verbal Input To examine the distribution of all parents' speech, we counted the word frequency from each dyad's transcript, again dividing by total words per session to control for variations in trial duration across participants.

As expected, the distribution of parents' speech was very similar to that of other natural language corpora (Figure 6A compared to Figure 1A) and child-directed speech in particular (Lavi-Rotbain & Arnon, 2023). The mean and median differed substantially (normalized: 0.08% and 0.01%; non-normalized: 23.72 and 3) and the skewness value was very high (9.52; Figure 6A). The correlation between the observed and expected distributions was strong (r = 0.99), indicating a near-Zipfian fit.

In addition, without the rank bounds of the 23 toys used in the study, the true scale of this type of distribution can be seen. Parents used a total of 29,597 words across the 57 sessions. The most common word, "you" (likely due to the frequent direct address of infants), appeared 1,505 times, whereas 430 different words were spoken only once.

Individual-level analyses again showed consistent Zipfian structure (Figures 6C–D). Even the least Zipfian subject displayed a strongly skewed distribution that closely matched the expected pattern (r = .95) and the average correlation across participants was still high (r = .98; SD = .01).

Nouns and Verbs To assess whether the skewed distribution persisted when restricting the analysis to content words, we retained only concrete nouns and action verbs, which are target candidates for early language learning. Word counts were again normalized based on total nouns and verbs per session.

Even after removing repetitive helper words, the distribution remained highly skewed (Figure 6E). The mean (0.3%) and median (0.01%) differed, and the skewness measure (4.89) was still substantial. The overall correlation between the observed and expected distributions remained strong (r = .95), though slightly lower than for all speech. At the individual level, many subjects showed distributions that were even more Zipfian than the group average. The most Zipfian subject had a correlation of r = .99 (Figure 6G–H), and the average correlation across participants was .97 (SD = .02).

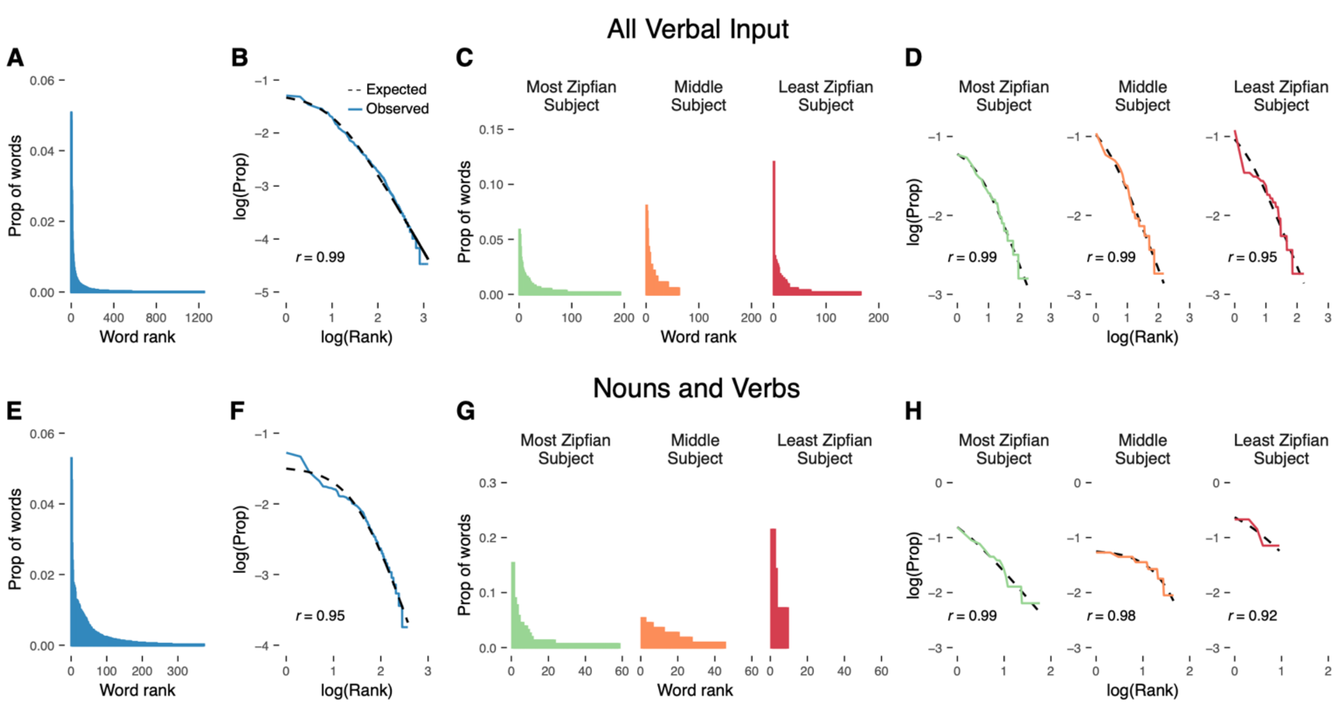

Figure 6. A) Distribution of all parent speech. B) Observed and expected distributions on a log-log scale. C) Distributions of individual participants— most Zipfian, median, and least Zipfian. D) Log-log plots of rank and proportion for individual participants. E) Distribution of nouns and action verbs. F) Observed and expected distributions on a log-log scale. G) Distributions of individual participants. H) Log-log plots of rank and proportion for individual participants.

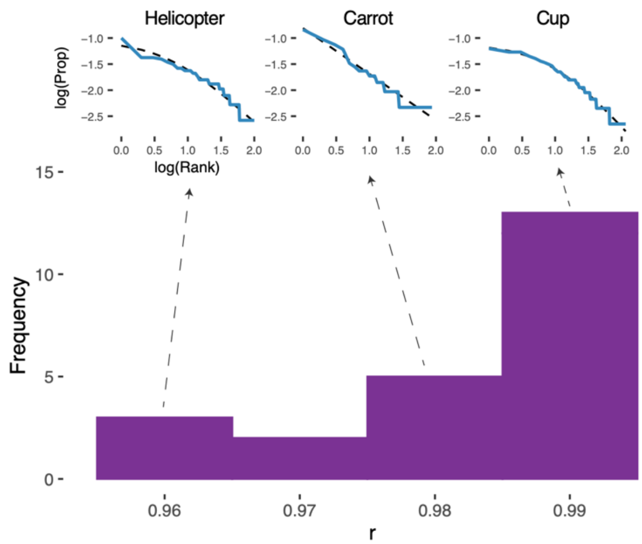

Figure 7. Distribution of correlation values between expected and observed distributions for the proportion of words infants heard when attending to the same object. Callouts highlight distributions for individual objects in the first, middle, and last bins.

Multi-Modal Integration of Infant Attention and Parent Speech

Lastly, we examined whether the skewed distribution persisted when combining multiple modalities—infants' active visual selection of objects and parents' verbal input. We aggregated all the nouns and verbs infants heard while attending to the same object across the session (Figure 3). For example, the distribution for "cat" only included the nouns and verbs heard during moments when the infant was looking at the cat. This created a group-level distribution for each object as opposed to a single distribution for the sample. Figure 7 displays a histogram of the correlation values for all 23 toys. Individual-level analyses were not conducted due to insufficient parent speech about individual objects to yield stable distributions.

As shown in Figure 7, each object exhibited a strongly skewed, near-Zipfian distribution. The average correlation was .98 (*SD* = .01), and the average skewness was 4.63 (*SD* = 1.51). Notably, every object-level correlation exceeded the overall correlation for nouns and verbs when attention was not considered (Figure 6F). Thus, it seems that combining the input infants select for themselves with the input given by parents results in an even more skewed distribution.

Discussion

Our findings demonstrate that the natural input infants receive and select across multiple modalities follows a near-Zipfian distribution. And individual analyses show that this structure is not merely an artifact of aggregated data. At both the corpus and individual levels, infants' visual and manual attention, as well as the verbal input they receive from parents, exhibited the characteristic skewed pattern observed in prior research on child-directed speech and visual availability (Clerkin et al., 2017; Jayaraman et al., 2015; Lavi-Rotbain & Arnon, 2023). And when integrating across modalities—specifically examining the verbal input infants received while attending to specific objects—the distribution became even more skewed than verbal input alone. This suggests that infants' own active selection plays a role in shaping the structure of their input.

Both computational models and human learners across development perform best in experimental tasks when input is skewed (Hendrickson & Perfors, 2019; Kurumada et al., 2013; Lavi-Rotbain & Arnon, 2021; Wolters et al., 2024). One key reason may be that early learning of high-frequency items scaffolds the learning of lower-frequency items, allowing for efficient information processing and generalization (Lavi-Rotbain & Arnon, 2021, 2022). However, it remains an open question whether the presence of this skew reflects a fundamental property of human cognition or is simply another instance of natural phenomena exhibiting Zipfian distributions (Lavi-Rotbain & Arnon, 2023). Regardless, identifying how the internal learning system interacts with the statistical structure of real-world input is essential for advancing our understanding of cognitive development.

Acknowledgements

This work was supported by NIH R01DC017925 and R01HD104624 to CY. We thank the Developing Intelligence Lab at UT Austin, specifically Cat Bianco, Melina Knabe, Jingwen Pang, and Jocelyn Thomas for their help in collecting and coding the data.

References

Bambach, S., Crandall, D., Smith, L., & Yu, C. (2018). Toddler-inspired visual object learning. *Advances in neural information processing systems, 31*.

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*, 20160055.

Farhadi, A., & Redmon, J. (2018). *Yolov3: An incremental improvement.* Paper presented at the Computer vision and pattern recognition.

Hendrickson, A. T., & Perfors, A. (2019). Cross-situational learning in a Zipfian environment. *Cognition, 189*, 11-22.

Jayaraman, W., Fausey, C. M., & Smith, L. B. (2015). The faces in infant-perspective scenes change over the first year of life. *PLoS ONE*.

Kimchi, I., Wolters, L., Stamp, R., & Arnon, I. (2023). Evidence of Zipfian distributions in three sign languages. *Gesture, 22*(2), 154-188.

Kurumada, C., Meylan, S. C., & Frank, M. C. (2013). Zipfian frequency distributions facilitate word segmentation in context. *Cognition, 127*(3), 439-453.

Lavi-Rotbain, O., & Arnon, I. (2021). Visual statistical learning is facilitated in Zipfian distributions. *Cognition, 206*, 104492.

Lavi-Rotbain, O., & Arnon, I. (2022). The learnability consequences of Zipfian distributions in language. *Cognition*, *223*, 105038.

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

Mandelbrot, B. (1953). An informational theory of the statistical structure of language. *Communication Theory*, *84*(21), 486-502.

Montemurro, M. A. (2001). Beyond the Zipf–Mandelbrot law in quantitative linguistics. *Physica A: Statistical Mechanics and its Applications, 300*(3-4), 567-578.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin & review, 21*, 1112-1130.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-olds infants. *Science, 274*, 1926-1928.

Wolters, L., Lavi-Rotbain, O., & Arnon, I. (2024). Zipfian distributions facilitate children's learning of novel word-referent mappings. *Cognition*, *253*, 105932.

Zipf, G. K. (1946). The psychology of language. In *Encyclopedia of psychology*: Philosophical Library.