UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Children Learn Words Better in Low Entropy

Permalink

Journal

Authors

Lavi-Rotbain, Ori
Arnon, Inbal

Publication Date

2019

Children Learn Words Better in Low Entropy

Ori Lavi-Rotbain (oriedit.lavi@mail.huji.ac.il)
Edmond & Lily Safra Center for Brain Sciences, Edmond J. Safra Campus,
The Hebrew University Jerusalem, 9190401, Israel

Inbal Arnon (inbal.arnon@mail.huji.ac.il)
Department of Psychology, The Hebrew University of Jerusalem,
Mount Scopus, Jerusalem 91905, Israel

Abstract

During their first year, infants learn to name objects. To do so, they need to segment speech, extract the label and map it to the correct referent. While children successfully do so in the wild, previous results suggest they struggle to simultaneously learn segmentation and object-label pairings in the lab. Here, we ask if some of children's difficulty is related to the uniform distribution they were exposed to, since it differs from that of natural language, and has high entropy (making it less predictable). Will a low entropy distribution facilitate children's performance in these two tasks? We looked at children's (mean age=10;4 years) simultaneous segmentation and object-label mapping of words in an artificial language task. Low entropy (created by making one word more frequent) facilitated children's performance in both tasks. We discuss the importance of using more ecologic stimuli in the lab, specifically- distributions with lower entropy.

Keywords: Statistical learning; Multi-modal cues; Word segmentation; Word learning; Entropy; Children.

Introduction

During the first year of life, infants make their initial steps in learning language. One ability they acquire is naming objects. To do so, infants need to extract the segmented labels and map them onto the correct object. While infants learn some object-label mappings early on (Bergelson & Swingley, 2012), even older children seem to struggle with simultaneously learning segmentation and object-label pairings in the lab. Previous work examined children's ability to perform both tasks at the same time in a statistical learning paradigm (Lavi-Rotbain & Arnon, 2017). Children were exposed to an unsegmented speech stream where transitional probabilities served as a cue for word boundary (as in Saffran, Aslin, & Newport, 1996). The language had an additional visual cue to segmentation: each word was matched to an image of an object that appeared for the duration of the word (e.g., 'dukame' → blue star). The prediction was that the visual cue will assist segmentation and allow children to learn the object-label pairings, illustrating their ability to integrate multimodal cues. As predicted, the results showed that after a short exposure (under two minutes) 10;6-year-olds managed to learn both aspects (segmentation and object-label mapping). However, while children showed some learning of the object-label pairing (they were above chance, M=34.4%, chance=25%), their learning was relatively poor. Younger children (mean age: 7;8 years) did not learn the pairings at all (M=25.96%%, chance=25%), even though they are clearly capable of relating labels to objects in natural language. Why then do children struggle with this task in the lab? And what can we learn from their difficulty about the factors that impact children's language learning? Here, we ask how the distributional properties of the language may have impeded learning. In particular, we focus on the use of a uniform distribution – where all items are equally frequent. This was the distribution used in the previous study, and one that is used in most statistical learning studies.

A uniform distribution of stimuli, where every element (e.g. word) is presented the same number of times, differs from what is found in natural language. Words in natural language have a Zipfian distribution (Zipf, 1936) with few very frequent words, and most words having low frequency. The Zipfian distribution is a highly skewed distribution, with a narrowed peak for the small number of very frequent words, and a long tail for the rest of the words. Words show a Zipfian distribution across many languages, in both adult-to-adult speech (Zipf, 1936; Piantadosi, 2014) and child directed speech (Hendrickson & Perfors, 2019; Lavi-Rotbain & Arnon, submitted). Other aspects of language, like grammatical categories, also show a Zipfian distribution (Piantadosi, 2014; Lavi-Rotbain & Arnon, submitted). Interestingly, the objects that infants see also show a Zipfian distribution (Clerkin, Hart, Rehg, Yu, & Smith, 2017). That is, using a uniform distribution does not accurately reflect the distribution of words (or objects) that children are exposed to.

Moreover, uniform distributions are also less predictable than non-uniform distributions. One way to quantify the difference between them is to use Shannon's Entropy (Shannon, 1948). Entropy quantifies how unpredictable a distribution is as a whole, with higher entropy assigned to less predictable distributions. The uniform distribution is the least predictable - it is hard to guess which word will appear next when they all have equal probabilities - and consequently has high entropy. Non-uniform distributions, such as Zipfian distributions, are more predictable, and have lower entropies: it is easier to guess the next word when only a few are highly probable.

Here, we ask if children's simultaneous learning of segmentation and object-label pairings will be facilitated when using a distribution with low unigram entropy. Such a finding would have several important implications. First, it would indicate that children are sensitive to entropy, thereby expanding our understanding of the distributional properties

that impact learning. Second, it would highlight the importance of using stimuli that are more ecologically valid in their distributional properties: If children show better learning from a low entropy distribution, then previous conclusions about their ability to use multimodal cues may not be accurate. Under more natural conditions, children may show learning that was not previously detected. To give an example from another domain, children's knowledge of irregular plurals is much better when they are produced in familiar frames, as they are often produced in natural language (e.g. children produce "teeth" more accurately after "brush your---" compared to on its own, Arnon & Clark, 2011). Assessing children's morphological knowledge using single word elicitation under-estimated their true abilities, and could lead to inaccurate conclusions (e.g., that they have not learned the correct irregular form yet). Similarly, performance in artificial language learning studies improves when there are multiple cues to segmentation, as is found in natural language. Visual cues for word boundaries improve segmentation in adults (Cunillera, Camara, Laine, & Rodriguez-Fornells, 2010), as does the use of one-to-one mappings between words and objects (children: Lavi-Rotbain & Arnon, 2017; adults: Thiessen, 2010). Under these conditions, children and adults show better learning.

Will a reduction in entropy have a similar facilitative effect on learning? Looking at another domain, adults' cross-situational learning of novel object-label mappings was facilitated after exposure to a Zipfian distribution (with low entropy) compared to a uniform distribution (with high entropy) (Hendrickson & Perfors, 2019, Experiment 2). This facilitative effect was not found when words and labels were presented one at a time: the authors propose that Zipfian distributions are beneficial only when learners are faced with ambiguity. In such cases, the very frequent word can be learned early on and used to disambiguate later trials. Another reinforcement to the potential advantage of low entropy distributions comes from word segmentation studies: children's and adults' segmentation is facilitated when the input had low entropy (entropy was reduced by making one word more frequent than the other, Lavi-Rotbain & Arnon, 2018, 2019), and when it has a Zipfian distribution (Kurumada, Meylan, & Frank, 2013).

Here, we expand on these findings to look at the effect of reduced entropy on word learning in children: will lower entropy facilitate learning in a task that involves both segmentation and object-label mapping? The segmentation task is inherently ambiguous: since learners are exposed to an unsegmented stream, successfully segmenting one word can help in segmenting the rest. An additional facilitative effect can come from the overall greater predictability of the input: non-uniform distributions are more predictable and have lower entropy. If learners are sensitive to such measures of the environment, then learning may be facilitated even in non-ambiguous situations. Since both factors are relevant for the segmentation task, we hypothesized that segmentation will be better under low entropy. The predictions are less clear about learning the object-label mappings. On the one hand, this task does not involve ambiguity: the same object is always presented with the same label. At the same time, the overall predictability of the mappings is greater in the non-uniform distribution. If there is an effect of reduced entropy regardless of ambiguity, we should see a facilitative effect here as well. We hypothesized that learning the object-label mappings will also be facilitated under low entropy.

The current study

In the current study, we use the same artificial language learning paradigm used previously to examine children's learning of multimodal information (Lavi-Rotbain & Arnon, 2017). Children are exposed to an unsegmented speech stream containing four novel words, with consistent word-object pairings (each word is paired with an object: e.g., 'dukame' with a blue star). We ask if children will show better learning of both segmentation and object-label pairings when exposed to low entropy input compared to high entropy input. We focus our inquiry on words that have lower frequency. Frequency is known to affect word learning during infancy with more frequent words learned earlier (Goodman, Dale, & Li, 2008). Frequency, however, does not account for all the variance in a words' age-of-acquisition. It is easy to find examples of low frequency words among the early acquired ones: for example, the word ' cheek' is learned at 22 months (Frank, Braginsky, Yurovsky, & Marchman, 2017), but appears only 18 per million. Could the low entropy found for words in natural language help children learn low frequency words? Finding such a pattern in our experimental manipulation would open up new avenues for understanding how low frequency words are acquired.

We manipulate entropy by making one word much more frequent: in the high entropy condition, all words appeared an equal number of times (each word appeared 32 times). In the low entropy condition, one word was much more frequent (appearing 214 times), while the other three appeared 19 times (half of the frequency of the words in the high entropy condition). We compare segmentation and word-object pairings for the low frequency words from the low entropy condition (which appeared 19 times) with the words from the high entropy condition (which appeared 32 times). If children are mostly sensitive to frequency, then learning should be better in the high entropy condition. However, if children are sensitive to more than mere frequency, in particular to the entropy of the distribution, than learning of the low frequency words should be better in the low entropy condition. If this happens regardless of ambiguity, then we should see better performance due to entropy reduction for both segmentation and learning the correct object-label pairings.

Method

Participants

61 children took part in this Experiment (age range: from 9;0 to 12;0 years, mean age: 10;4 years; 27 boys, 34 girls). We chose this age range since it matches the one used in the older

group of Lavi-Rotbain & Arnon (2017), where children showed poor learning of the object-label pairings. Participants were visitors at the Bloomfield Science Museum in Jerusalem and were recruited for this study as part of their visit to the Living Lab. 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.

Materials

Auditory stimuli
All participants were exposed to a familiarization stream corresponding to the condition they were assigned to. All streams were composed of the same four unique tri-syllabic synthesized words: "dukame", "nalubi", "kibeto", and "genodi". We used only four words to make the task learnable for children. As the results show, even this number proved challenging for children. The twelve different syllables making up the words were taken from Glicksohn & Cohen (2013). The syllables were created using the PRAAT synthesizer (Boersma & van Heuven, 2001) and were matched on pitch (~76 Hz), volume (~60 dB), and duration (250–350 ms). The four words were created by concatenating the syllables using MATLAB to ensure that there were no co-articulation cues to word boundary. The words were matched for length (average word length- 860ms, range=845-888ms). The words were then concatenated together using MATLAB in a semi-randomized order to create the auditory familiarization streams. Importantly, there were no breaks between words and no prosodic or co-articulation cues in the stream to indicate word boundaries. The only cue for word boundaries was transitional probabilities (TP's): TP's between words were lower compared to TP's within words.

Experimental conditions
We created two auditory sequences, corresponding to two levels of entropy: high and low. In the high entropy level, words followed a uniform distribution with each word appearing 32 times in a semi-randomized order (no word appeared twice in a row). The sequence had 128 tokens and lasted 1:50 minutes. TP's within a word were 1, and TPs between words were 0.333. In the low entropy level, words appeared with a skewed distribution: one word appeared 80% of the time (214 appearances) while each of the other three words appeared only 7% of the time (19 appearances for each word). The sequence had 271 tokens and lasted 3:50 minutes. The identity of the frequent word was counterbalanced across subjects in the low entropy condition to prevent item-specific effects. TP's within a word were 1, but the TP's between words varied depending on the next word (since the frequent word in this condition was more likely to occur). See Table 1 for full details of the experimental conditions.

Visual stimuli
While listening to the audio stream, participants saw shapes on the screen whose appearance was synchronized with word

Table 1: Different experimental conditions

	High entropy (Uniform)	**Low entropy**
Exposure length [minutes]	1:50	3:50
Number of tokens	128	271
Tokens per word	32	Frequent: 214 Infrequent: 19
Unigram entropy [bits]	2	1.1
TP's between words	0.33	For the frequent word: 0.75 For infrequent words: 0.08

boundaries. Shapes appeared at word onset and remained onscreen for the duration of the word. Each word appeared always with the same shape and vice versa ("dukame": blue star, "nalubi": green hexagon, "kibeto": purple heart, and "genodi": orange diamond). In the low entropy condition, when the same word appeared twice in a row, the shape disappeared briefly (for 200 ms) at the end of the first occurrence and reappeared with the second occurrence onset. The visual stimuli is modelled on the regular condition from Thiessen (2010) and Lavi-Rotbain & Arnon (2017), which was shown to facilitate segmentation in both adults and children. See Fig. 1 for an illustration.

Segmentation test
This test asked how well children segmented the continuous stream into words using 16 two alternative forced-choice trials. The visual stimuli did not appear on screen during test. Participants heard two words and were asked to decide which belonged to the language they heard. We used non-words as foils ("dunobi", "nabedi", "kilume", and "gekato", average length: 860ms; range 854-868ms), created by taking three syllables from three different words, while keeping their original position. We used non-words (instead of part-words) as foils since these are easier to distinguish from 'real' words. Since children struggle with this task, we chose to focus only on the "easier" non-word vs. word distinction. Each of the four words appeared once with each of the four foils to create 16 trials (in a random order, with the constraint that the same word/foil did not appear in two consecutive trials). The order of words and foils was counter-balanced so that in half the trials, the real word appeared first and in the other half, the foil appeared first.

Fig. 1: Audio-video illustration

Word–shape correspondence test

This test asked how well children learned the correspondence between the words and the shapes. In each trial, children saw the four shapes on the screen and heard one of the four words. Then, they had to choose the shape corresponding to the word. Each word was repeated four times on non-consecutive trials, to create 16 trials that appeared in a random order between subjects.

Procedure

After receiving parental consent, children were seated in front of a computer station with a noise-blocking headset next to an experimenter. The children were told they are about to hear an alien language, and that they need to pay attention to what they will see and hear and try to learn it as best as they can. Each child was randomly assigned to one of the two experimental conditions. After the exposure phase, children completed a segmentation test and a word-shape correspondence test. The instructions were identical in all conditions.

Results

Children were divided as follows between the two conditions: high entropy, N=28; low entropy, N=33. Age did not differ across entropy conditions (F(1)=0.195, p=0.66). In the low entropy condition, the frequent word was counterbalanced across subjects. A one way ANOVA revealed that performance was not impacted by which word was the frequent one in the segmentation test (F(3)=2.326, p=0.1), or in the recognition test (F(3)=0.52, p=0.67). Consequently, in all subsequent analyses we collapsed the data across the different frequent words.

Segmentation analysis

Children showed learning (were above chance) in both conditions (low entropy condition: M=73.9%, t(32)=7.69, p<0.001; high entropy condition: M=65.2%, t(27)=4.83, p<0.001). However, this success rate includes both the frequent and the infrequent word for the low entropy condition. Since the frequent word had much higher frequency (214 appearances) than the other words (19 appearances) in the low entropy condition, it does not make sense to include the frequent word in our analysis. In order to examine the effect of entropy on low frequency words alone, we looked only at trials where the correct answer was a low
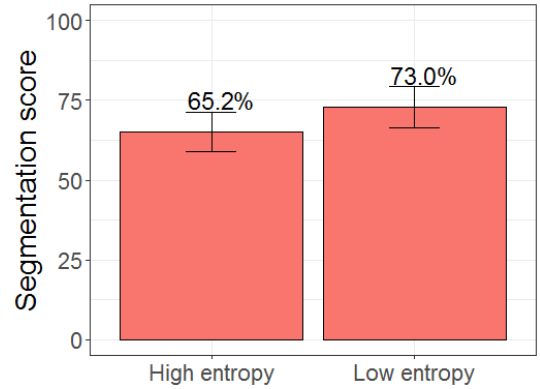

Fig. 2: Mean segmentation score of low frequency words by entropy condition with 95% confidence intervals

frequency word (appearing 19 times during exposure). This left 12 trials per participant. In this subset of the segmentation test, children showed learning above chance of the infrequent words (low entropy condition: M=73.0%, t(32)=7.0, p<0.001) (see Fig. 2). We will now compare this mean to the one from the high entropy condition (73.0% versus 65.2% respectively).

We used mixed-effect linear regression model to examine the effect of entropy level on segmentation of infrequent words. Following Barr et al. 2013, the models had the maximal random effect structure justified by the data that would converge. Our dependent binominal variable was success on a single trial of the segmentation test. We had entropy condition (high entropy condition as baseline) as a fixed effect, as well as: age-in months (centered); gender; trial number (centered); and order of appearance in the test (word-first trials vs. foil-first trials). The model had random intercepts for participants and for item (Table 2). To examine the overall effect of entropy, we used model comparisons.

As predicted, entropy level impacted segmentation of low frequency words (chi(1)=3.2, p=0.07). Participants showed better segmentation of low frequency words in the low entropy condition compared to the high entropy condition, despite appearing half the times (β=0.42, SE=0.23, p=0.07). Order of appearance in the test significantly affected segmentation, with better accuracy on trials where the word appeared before the foil (β=0.39, SE=0.16, p<0.05), as has been found in previous studies (Lavi-Rotbain & Arnon, 2017; Raviv & Arnon, 2018). Since the order of presentation of

Table 2: Mixed-effect regression model for segmentation of infrequent words. Variables in bold were significant. Significance obtained using the lmerTest function in R.

	Estimate	Std. Error	z value	p-value
(Intercept)	0.69584	0.21439	3.246	**<.01 ****
Age (centered)	0.23516	0.14220	1.654	=0.098 .
Low entropy condition	0.42079	0.23321	1.804	**=0.07 .**
Gender (male)	-0.03948	0.11747	-0.336	>.1
Trial number (centered)	-0.02390	0.01705	-1.401	>.1
Order of appearance (word)	0.19385	0.07844	2.471	**<.05 ***

words and foils was counter-balanced this could not reflect a preference for pressing 1 or 2, and is in line with the "interval bias" which is often found in 2AFC tests (Yeshurun, Carrasco, & Maloney, 2008). Age almost reached significance: older children were slightly better than younger ones (β=0.24, SE=0.14, *p*=0.098). Trial number and gender did not affect segmentation (trial number: β= -0.02, SE=0.02, *p*>0.1; gender: β= -0.08, SE=0.23, *p*>0.1).

The beneficial effect of low entropy on segmenting low frequency words cannot be attributed to learning only the frequent word, and ruling out foils due to syllables they share with the frequent word. To see if there is a difference between trials in the low entropy condition where the foil shared one syllable with the frequent word (M=72.7%) and trials where it didn't (M=77.3%) we used a linear regression model with success on a single trial as the dependent binominal variable, and "is foil frequent" (assigned '1' for trials in which the foil shared any of its three syllables with the frequent word and '0' when it didn't) as a fixed effect, as well as log frequency (centered), gender, trial number (centered); and order of appearance in the test. The model had random intercepts for subjects and for items. "Is foil frequent" was not a significant predictor of accuracy (β= -0.27, SE=0.25, *p*>0.1), neither all the other fixed effects. That is, children in the low entropy condition indeed learned the low frequency words better.

Recognition analysis

Children showed learning of object-label pairings (were above chance) in the low entropy condition (M=49.3%, chance=25%, t(32)=5.16, *p*<0.001). In contrast, they were not above chance in the high entropy condition (M=32.7%, chance=25%, t(27)=1.66, *p*=0.11). The accuracy in the high entropy condition is similar to that from Lavi-Rotbain & Arnon (2017), using the same task and uniform distribution (M=34.4%). While children did show learning in the previous study (just above chance), their performance was still poor, indicating difficulty in learning the mappings from a uniform distribution. How well did children learn the infrequent words in the low entropy condition? As in the segmentation test, the mean accuracy in the low entropy includes also recognition of the frequent word. In order to look at recognition of low frequency words, we looked only at trials where the correct word was a low frequency word (12 trials per child). Since children learned the frequent word quite well (M=64.6%), we assume that chance level on each trial is not 25% but 33% (since they could rule out the shape corresponding to the frequent word). Note however that this is a quite rigid assumption: children did not show complete learning of the frequent word (they were incorrect 35% of the time), and there was very large variance in accuracy (SD=37.9%), meaning that for some trials they were picking between four options. Nevertheless, we put our prediction to a stringent test and assume that chance is 33% for the low frequency words. As predicted, children learned the infrequent words above chance in the low entropy condition (M=44.0%, chance=33%, t(32)=2.14, p<0.05) (see Fig. 3). This means that while in the uniform condition children did
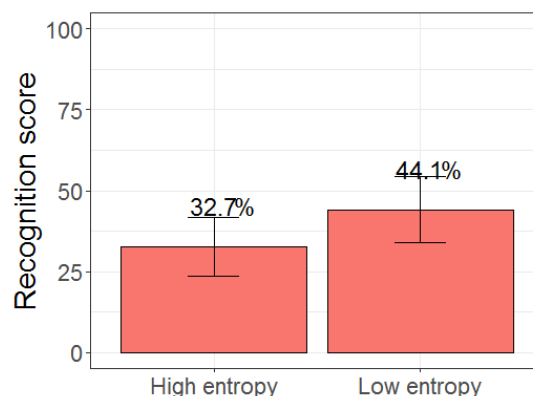

Fig. 3: Mean recognition score of low frequency words by entropy condition with 95% confidence intervals

not show learning of the object-label pairings (were not above chance), children in the low entropy condition did show learning even of the infrequent words, despite appearing half the number of times and despite the rigid chance level.

Is there a correlation between segmentation and recognition scores? Previous results showed positive correlation in adults' performance between these two tasks, indicating that better segmentation went along with better word learning (Lavi-Rotbain & Arnon, 2017; Thiessen, 2010). However, such a correlation was absent in children's performance (Lavi-Rotbain & Arnon, 2017). Here, we found a positive correlation only in the low entropy condition: children who performed well in the segmentation test, were also good at mapping labels to objects (R^2=0.4, t(31)=2.42, *p*<0.05), highlighting the connection between segmentation and word learning found in natural language. Such a correlation was not found in the high entropy condition (R^2=0.21, t(26)=1.12, *p*>0.1).

Discussion

We set to ask whether children's ability to segment and learn object pairings for low frequency words will be better when learning from low entropy input compared to high entropy input. To do so, we examined children's performance in an artificial language across two levels of entropy (high and low), in two tasks: segmentation and object-label pairing. Entropy was reduced by making one word more frequent than the rest, so that it appeared 80% of the time. We focused on children's performance on low frequency words (that appeared only 19 times in the low entropy condition, versus 32 in the high entropy condition). Our results show that entropy reduction is beneficial for children's segmentation, (see also Lavi-Rotbain & Arnon, 2018, 2019), as well as for their learning of object-label mapping. In addition, we found a positive correlation between segmentation and mapping only under the low entropy condition. Based only on findings from the uniform conditions from this study and from Lavi-Rotbain & Arnon (2017), one could conclude that children are not able to simultaneously learn segmentation and object-

label mapping (at least in lab conditions). However, the low entropy condition offers an alternative explanation: when exposed to more predictable and ecological input, children show evidence of learning both tasks at the same time. Importantly, children's object-label accuracy was still not good, raising the need to find ways to make the task easier: we predict that the effect of reduced entropy will be stronger once that is done. We are currently running a similar study with the younger age group (that showed no learning of the object-label mappings in the previous study), to see if entropy reduction will have a similar facilitative effect on this age group and will enable them to learn both the segmentation and object-label pairing.

Why did the low entropy condition facilitate learning? Several inherent properties of low entropy distributions may be facilitative for learning. First, creating a low entropy in the way we did drastically increases the frequency of one or more of the words. These highly frequent words can be learned relatively early on and later serve as an anchor for learning other words, similar to presenting words in isolation prior to presenting the unsegmented stream (Cunillera, Càmara, Laine, & Rodríguez-Fornells, 2010). In addition, TP's between the frequent and infrequent words can be lower and hence be more salient for learning. However, we suggest that there is more to the low entropy condition that facilitates learning than anchoring and lower TP's. Language learners may be sensitive to the overall predictability of the input, and learn better from input with lower entropy. Such an account predicts that entropy reduction will also facilitate learning when there is less ambiguity. Our results provide some support for this, by showing that learning was improved also for the non-ambiguous object-label pairings (contra the prediction made in Hendrickson & Perfors, 2019). This prediction is also supported by findings showing that adults' word segmentation is facilitated in a low entropy condition compared to a medium entropy one, despite both having similar anchoring and TP cues (Lavi-Rotbain & Arnon, 2019). Further work is needed to understand what exactly about low entropy is facilitative and how that relates to the input that children are actually exposed to.

From a methodological perspective, our results highlight the importance of creating experimental stimuli that better reflect the input children hear. In particular, most SL studies use a uniform distribution during exposure, although the distribution itself is not relevant for their research question. However, by doing so, we may be introducing unnecessary difficulties for our participants that may interfere with our assessment of their abilities. For children, who find artificial language learning experiments harder to begin with, such factors may impact performance more, and more easily. Theoretically, the findings point to the importance of studying the impact of entropy on language learning. Entropy has been studied across many domains of language, including language processing, use and structure (e.g., Cohen Priva, 2017; Linzen & Jaeger, n.d.; Piantadosi, Tily, & Gibson, 2011). For example, there is evidence that the entropy of single words is restricted to a small range of values across

many languages, suggesting that speakers have similar preferences for how predictable their languages are (Bentz, Alikaniotis, Cysouw, & Ferrer-i-Cancho, 2017). In addition, there is a trade-off between unigram and trigram entropy over time across many languages, indicating that speakers maintain a relatively constant information rate (Cohen Priva & Gleason, 2016). Children also show sensitivity to such measures: two-year-olds show better repetition of unfamiliar four-words sequences when the final word "slot" has higher entropy (Matthews & Bannard, 2010). These studies show that language users are sensitive to entropy and other information-related measures, and suggest that languages are shaped by constrains arising from these measures. However, the role of entropy on language learning is understudied. Our results show that entropy effects are found in children and impact learning of both segmentation and word labels.

Our results may offer a possible explanation for how children acquire low frequency words at a relatively early age. Words in natural language show a Zipfian distribution, in which most of the words have low frequencies. Under a low entropy distribution, such as the Zipfian distribution, the disadvantage of low frequency can turn into an advantage: the few frequent words can serve as an anchor for learning the low frequency ones. We are currently conducting a series of studies to examine the role of entropy in natural language learning, and in predicting variance in age-of-acquisition.

Acknowledgments

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68(3), 255–278.

Bentz, C., Alikaniotis, D., Cysouw, M., & Ferrer-i-Cancho, R. (2017). The entropy of words-Learnability and expressivity across more than 1000 languages. Entropy, 19(6), 1–32.

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. Proceedings of the National Academy of Sciences, 109(9), 3253–3258.

Boersma, P., & van Heuven, V. (2001). Speak and unSpeak with Praat. Glot International, 5(9–10), 341–347.

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.

Cohen Priva, U. (2017). Not so fast: Fast speech correlates with lower lexical and structural information. Cognition,

160, 27–34. Retrieved from

Cohen Priva, U., & Gleason. (2016). Simpler structure for more informative words : a longitudinal study. Proceedings of the 38th Annual Conference of the Cognitive Science Society, (2012), 1895–1900.

Cunillera, T., Camara, E., Laine, M., & Rodriguez-Fornells, A. (2010). Speech segmentation is facilitated by visual cues. Quarterly Journal of Experimental Psychology (2006), 63(2), 260–274.

Cunillera, T., Càmara, E., Laine, M., & Rodríguez-Fornells, A. (2010). Words as anchors: Known words facilitate statistical learning. Experimental Psychology, 57(2), 134–141.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. Journal of Child Language, 44(3), 677–694.

Glicksohn, A., & Cohen, A. (2013). The role of cross-modal associations in statistical learning. Psychonomic Bulletin & Review, 20(6), 1161–1169.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. Journal of Child Language, 35(3), 515–531.

Hendrickson, A. T., & Perfors, A. (2019). Cross-situational learning in a Zipfian environment. 189(May 2017), 11–22.

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. (submitted). Zipf's Law in Child-Directed Speech.

Lavi-Rotbain, O. & Arnon, I, (2018, November). Frequency or predictability? The effect of entropy on statistical learning in children and adults. Poster session presented at the Boston University Conference on Language Development, Boston, MA.

Lavi-Rotbain, O., & Arnon, I. (2019). Low Entropy Facilitates Word Segmentation in Adult Learners, Proceedings of the 41st Annual Conference of the Cognitive Science Society. Cognitive Science Society

Lavi-Rotbain, O., & Arnon, I. (2017). Developmental Differences Between Children and Adults in the Use of Visual Cues for Segmentation. Cognitive Science.

Linzen, T., & Jaeger, T. F. (2014). Investigating the role of entropy in sentence processing. ACL 2014, 10.

Matthews, D., & Bannard, C. (2010). Children's production of unfamiliar word sequences is predicted by positional variability and latent classes in a large sample of child-directed speech. Cognitive Science, 34(3), 465–488.

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.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. Proceedings of the National Academy of Sciences of the United States of America, 108(9), 3526–3529.

Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: modality-based differences in the effect of age. Developmental Science, 21(4), 1–13.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. Science (New York, N.Y.), 274(5294), 1926–1928.

Shannon, C. E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27, 379–423.

Thiessen, E. D. (2010). Effects of Visual Information on Adults' and Infants' Auditory Statistical Learning. Cognitive Science, 34(6), 1093–1106.

Yeshurun, Y., Carrasco, M., & Maloney, L. T. (2008). Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. Vision Research, 48(17), 1837–1851.

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