Contents lists available at ScienceDirect

Cognition



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

Infant-directed speech becomes less redundant as infants grow: Implications for language learning

Shira Tal^{a,b,*}, Eitan Grossman^c, Inbal Arnon^d

^a Department of Linguistics and English Language, University of Edinburgh, United Kingdom

^b Department of Cognitive Sciences, The Hebrew University of Jerusalem, Israel

^c Department of Linguistics, The Hebrew University of Jerusalem, Israel

^d Department of Psychology, The Hebrew University of Jerusalem, Israel

ARTICLE INFO

Keywords: Infant-directed speech Entropy rate Efficient communication Multi-word sequences Redundancy

ABSTRACT

Do speakers use less redundant language with more proficient interlocutors? Both the communicative efficiency framework and the language development literature predict that speech directed to younger infants should be more redundant than speech directed to older infants. Here, we test this by quantifying redundancy in infant-directed speech using entropy rate – an information-theoretic measure reflecting average degree of repetitiveness. While IDS is often described as repetitive, entropy rate provides a novel holistic measure of redundancy in this speech genre. Using two developmental corpora, we compare entropy rates of samples taken from different ages. We find that parents use less redundant speech when talking to older children, illustrating an effect of perceived interlocutor proficiency on redundancy. The developmental decrease in redundancy reflects a decrease in lexical repetition, but also a decrease in repetitions of multi-word sequences, highlighting the importance of larger sequences in early language learning.

1. Introduction

The way we talk is impacted by who we talk to. One of the hallmark examples of this is how speech directed to infants differs from the language used between adults (Kempe and Casillas, to appear; Schwab & Lew-Williams, 2016a; Soderstrom, 2007). For example, speakers use higher pitch, slower speech rate, and hyperarticulated vowels when they talk to young infants (Graf Estes & Hurley, 2013; Tippenhauer, Fourakis, Watson, & Lew-Williams, 2020; Uther, Knoll, & Burnham, 2007). Alongside these properties, infant-directed speech (IDS) also exhibits various forms of repetitiveness: it is characterized by lower lexical diversity (Soderstrom, 2007), frequently recurring phrases (Cameron-Faulkner, Lieven, & Tomasello, 2003), and successive utterances with partial self-repetitions (Küntay & Slobin, 1996; Lester et al., 2022; Tal & Arnon, 2018; Waterfall, 2006). While these characteristics reflect different types of repetitions, they may all be manifestations of the same over-arching principle: speakers being more redundant when talking to younger, and less proficient, learners.

Importantly, increased redundancy in speech directed to learners is also predicted by a different theoretical framework — that of

communicative efficiency. According to the communicative efficiency hypothesis, speakers' production choices reflect a trade-off between two competing pressures of reducing effort and maximizing understandability (Grice, 1975; Piantadosi, Tily, & Gibson, 2011b; Zipf, 1949). On the one hand, speakers aim to maximize understandability by providing enough information to successfully transmit their message. At the same time, speakers aim to minimize production effort, by exerting as little effort as possible. One way of balancing these pressures is to produce less linguistic material for more predictable messages: assigning longer signals to less expected messages and shorter signals to more expected messages can facilitate listener comprehension and balance the overall information rate of the interaction (Aylett & Turk, 2004; Levy & Jaeger, 2007; Pate & Goldwater, 2015; Shannon, 1948; Zipf, 1949). In line with this view, there is abundant evidence that speakers tend to reduce or omit more predictable elements, at many levels of linguistic analysis (syllable, word, construction: Aylett & Turk, 2004; Cohen Priva, 2015; Frank & Jaeger, 2008; Jaeger, 2010; Kravtchenko, 2014; Kurumada & Jaeger, 2015; Levy & Jaeger, 2007; Mahowald, Fedorenko, Piantadosi, & Gibson, 2013; Pate & Goldwater, 2015; Piantadosi, Tily, & Gibson, 2011a; Piantadosi et al., 2011b; Tily & Piantadosi, 2009). For example,

E-mail address: shira.tal1@mail.huji.ac.il (S. Tal).

https://doi.org/10.1016/j.cognition.2024.105817

Received 23 November 2022; Received in revised form 13 May 2024; Accepted 14 May 2024

0010-0277/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).



^{*} Corresponding author at: Centre for Language Evolution, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, 3 Charles Street, Edinburgh EH8 9AD, United Kingdom.

when choosing between shorter and longer lexical forms with the same meaning (e.g., *undergrad* and *undergraduate*), speakers tend to use the shorter forms in contexts where they are more predictable (Mahowald et al., 2013). In information-theoretic terms, speakers aim to minimize redundancy in their speech: when possible, they use fewer or shorter elements to convey a more predictable message.

According to the principles of efficient communication, the balance between effort and understandability can change depending on the comprehension difficulty within a conversation. Speakers are predicted to produce more effortful signals (i.e., ones containing more linguistic material) when understandability is at risk. Theoretically, understandability can be impacted by different factors (see discussion in Pate & Goldwater, 2015): predictability of the linguistic message itself, properties of the environment (e.g., noisy vs. quiet), and characteristics of the interlocutor. Indeed, a great volume of studies demonstrate that speakers are more likely to speak redundantly (i.e., increase effort) when the conveyed meaning is unpredictable in context (Aylett & Turk, 2004; Kurumada & Jaeger, 2015; Levy & Jaeger, 2007; Mahowald et al., 2013). For example, Japanese speakers are more likely to produce optional case marking when the thematic assignment is less predictable in context (e.g., the criminal arrested the police officer vs. the police officer arrested the criminal, Kurumada & Jaeger, 2015). Other studies document an effect of environmental noise on redundancy: talkers increase effort in the presence of background noise (a phenomenon known as The Lombard Effect, Lombard, 1911; Van Summers et al., 1988; Zhao & Jurafsky, 2009).

Finally, another potential source of comprehension difficulty is the interlocutor herself, who could face (or be perceived as facing) difficulty in comprehension, leading the speaker to increase effort. Speakers are perceptive of their audience's knowledge and modify their speech accordingly, a phenomenon known as audience design (Ariel, 1990; Arnold, 2008; Brennan & Hanna, 2009; Chafe, 1994; Clark & Murphy, 1983; Heller, Gorman, & Tanenhaus, 2012; Isaacs & Clark, 1987; Lockridge & Brennan, 2002). Audience design predicts that speakers will modify their speech based on the presumed properties of interlocutors. From the perspective of communicative efficiency, this modification should lead speakers to increase redundancy (increase effort) when their interlocutors misunderstand them (see also Grice, 1975). Indeed, several studies demonstrate this pattern (Buz, Tanenhaus, & Jaeger, 2016; Lockridge & Brennan, 2002; Roche, Paxton, Ibarra, & Tanenhaus, 2013). For example, speakers tend to hyperarticulate potentially confusable words when feedback from their interlocutors suggests that they have misunderstood previous words (Buz et al., 2016). Importantly, speakers can modify their speech not only on the basis of local misunderstandings, but also based on global estimations of their interlocutors' knowledge state and comprehension ease (Arnold, 2008). For instance, New Yorkers refer to landmarks differently when talking to other New Yorkers compared to people from other cities (Isaacs & Clark, 1987).

While global interlocutor properties can impact overall comprehension difficulty, their impact on speaker effort has been relatively under-studied. The theoretical prediction is that speakers should modify redundancy based on their interlocutors' language proficiency, with more redundancy when talking to less proficient interlocutors. Since IDS is inherently directed to addressees with lower language proficiency, this speech register should generally be characterized by more redundancy. In particular, speech directed to young infants, who are in the early stages of language learning, should be more redundant than speech directed to older children. Here, we investigate this prediction by comparing redundancy in IDS directed to younger and older infants. Beyond investigating redundancy in IDS, this serves as a test case for the more general prediction that speakers modify redundancy in their speech based on global properties of their interlocutors.

Importantly, different factors impacting understandability in communication (and modifications in redundancy in response) are very likely to interact. For example, if a message is unpredictable, it is very likely to cause a perceived difficulty for interlocutors. However, many studies only compare or manipulate the predictability of the message or environmental noise as sources for communication difficulty. Much fewer studies have investigated the direct impact of interlocutor properties on speaker effort while holding the message constant. This is, however, not a trivial task. Ideally, investigating the impact of an interlocutor's proficiency on production choices would require keeping the message identical. However, it is virtually impossible to compare completely parallel messages of naturalistic speech. One commonly-used way of dealing with this challenge (of controlling the message on the one hand but maintaining ecological validity on the other) is to hold constant the experimental context eliciting the conversations. Controlling the context can be done, for example, by eliciting stories using identical picture books (e.g., Berman & Slobin, 1994; Tal, Grossman, Rohde, & Arnon, 2023), or eliciting speech by asking participants to conduct the same task (Bard et al., 2000; Bard & Aylett, 2004; Pate & Goldwater, 2015; Rodriguez-Cuadrado, Baus, & Costa, 2017; Van Engen et al., 2010). Here, we will follow this logic to investigate the impact of interlocutors' language proficiency on speakers' effort.

The current study has two goals. The first, as mentioned above, is to ask whether redundancy is impacted by the interlocutors' perceived overall proficiency level, a core prediction of the communicative efficiency hypothesis. An additional goal is to provide a novel measure of global redundancy in IDS. While infant-directed speech is often described as repetitive, this is usually based on the analysis of individual properties (e.g., lexical types, frequent frames, variation sets), rather than on a holistic evaluation. Here, we use entropy rate, an informationtheoretic measure that allows us to capture the redundancy of a text as a whole (see details below). If speakers use more linguistic material with less proficient speakers, then speech directed to infants should decrease in redundancy with development. We investigate this hypothesis in two corpus studies. In Study 1 we use two developmental corpora to ask whether redundancy in IDS decreases with age. Having found an effect of age on redundancy, in Study 2 we ask which properties of the input impact this effect. Specifically, we ask whether the increased redundancy in speech directed to younger infants is impacted by changes in the repetition of multi-word sequences. While there is ample evidence showing that lexical diversity increases with age, less is known about the proportion of recurring multi-word sequences. Such sequences, however, are a salient property of IDS (Cameron-Faulkner et al., 2003; Goldberg, 2019) and have been shown to play an important role in learning various grammatical relations (Arnon, 2021; Arnon & Clark, 2011; Arnon & Ramscar, 2012; Siegelman & Arnon, 2015). Finding that increased redundancy in young children's input is influenced by the repetition of multi-word sequences will provide additional evidence for their role in early learning.

1.1. Quantifying redundancy using entropy measures

Recent years have seen increased interest in the application of information-theoretic concepts and methods to the study of language (e. g., Bentz, Alikaniotis, Cysouw, & Ferrer-i-Cancho, 2017; Cohen Priva, 2015; Ferdinand, Kirby, & Smith, 2019; Juola, 2008; Piantadosi et al., 2011b; Zaslavsky, Kemp, Regier, & Tishby, 2018). These studies have provided novel insights into long-debated questions about language structure, learning, and use (see review in Gibson et al., 2019). Information-theoretic measures such as entropy (Shannon, 1948) quantify the amount of uncertainty in a distribution of linguistic elements and have been shown to impact real-time language processing (Jaeger, 2010; Linzen & Jaeger, 2015; Piantadosi et al., 2011a; Siegelman et al., 2020). Such measures can also be used to characterize language systems as a whole. For example, information-theoretic measures can be used to measure and compare the complexity of the same text across languages, with texts with lower entropy seen as less complex (Bentz et al., 2017; Bentz, Ruzsics, Koplenig, & Samardži, 2016; Ehret, 2021; Ehret & Szmrecsanyi, 2016, 2019; Juola, 2008; Koplenig, Meyer,

Wolfer, & Mu, 2017; Lupyan, 2019). One particularly relevant measure for our purposes is entropy rate, which quantifies redundancy in texts by calculating their average degree of repetitiveness. Unlike unigram entropy, entropy rate does not treat words independently and captures the sequential relation between words and the degree of repetitiveness in a text (for a discussion see Bentz et al., 2017). Specifically, measuring the degree of string repetition in a text provides a measure of its compressibility, which reflects the average information content of the text (Bentz et al., 2017; Gao, Kontoyiannis, & Bienenstock, 2008). Importantly, entropy rate provides us a way to assess redundancy in texts overall, without having to pre-suppose where that redundancy comes from (bigrams, trigrams, lexical repetitions, etc.). As such, it provides a more comprehensive measure of redundancy than merely looking at repetitions of a particular size. Another advantage of entropy rate is that it has been used previously to compare redundancy across different languages. Using this measure therefore allows us to tie together the literature on infant-directed speech and the typological literature. In particular, this measure was used in a recent study (Bentz et al., 2017) to compare translations of the same text in over 1200 languages. Entropy rate was found to be similar across languages, suggesting that languages show similar levels of complexity. Interestingly, no study to date has used entropy rate to examine redundancy across language registers within the same language (see Ehret, 2021 for a comparison between different English registers using a similar method). Similarly, while several language acquisition studies used informationtheoretic and other measures to capture repetitiveness and complexity in child-directed speech (Brodsky, Waterfall, & Edelman, 2007; Cameron-Faulkner et al., 2003; Elmlinger, Goldstein, & Casillas, 2022; Lavi-Rotbain & Arnon, 2023; Lester et al., 2022; Stoll et al., 2012), no study to date has used such a measure to capture the overall redundancy of speech directed to young children. Here, we will use entropy rate as a way to quantify redundancy in IDS, and ask whether it changes with development. In what follows we explain how entropy rate can be used to quantify redundancy.

The entropy rate of words provides the average information content of words conditioned on all preceding tokens. Entropy rate is affected by the *order* of words in a given text: it decreases when words are more predictable given their preceding contexts. Entropy rate is calculated in the following way (see Eq. (1)): for any given word¹ in the text, we look for the *longest* string it initiates that is found in the preceding text. For example, the longest string found for the underlined <u>the</u> in the sequence 'The girl saw <u>the</u> girl next door', will be the *the girl*, as the repetition of *the* alone is shorter than the repetition of *the girl* (see more details below).

$$\mathbf{h}(\mathbf{T}) = \frac{1}{N} \sum_{i=2}^{N} \frac{\log_2 i}{Li} \tag{1}$$

The entropy rate is the average of these match lengths where N is the overall number of tokens, $_i$ is the position in the string, and L_i is the longest match string for $_i$ plus one ($L_i=1$ + the longest match string). The average length of these strings reflects the redundancy in the text: the larger the average match string is, the more repetitive (and therefore, redundant) the text is. Mathematically, as the average match string grows, entropy rate becomes smaller. In other words, a lower entropy rate reflects a more redundant text. To give an example, we use the following short text (numbers index word order):

$$come_1 here_2 come_3 here_4$$
 (2)

In order to calculate entropy rate for this text, we start with the second token (*here*). Because there is no match for this word in the

preceding string (which contains only the word *come*), the match length for this token is 0. L₂ is 1 (1 + the length of the string match: 0), and the total information content for this word is $\frac{log_{2}2}{1+0} = 1$. Moving on to the next word (*i* =3), *come*₃, we find a match of length 1 in *come*₁. We then expand our search string, going one word forward from the word in question, and look for a match for the sequence *come*₃ *here*₄. We find a match for the two-word sequence in the preceding text, meaning that the match length for this token is 2, and the information content is: $\frac{log_{2}3}{1+2} = 0.528$. Finally, the longest matching string for the last word, *here*₄, is only one word long (*here*₂), and the information content is: $\frac{log_{2}4}{1+1} = 1$. The total entropy rate for this text is the average information content of all words, which is 0.632.

To illustrate how this measure captures different types of redundancies, we compare the following invented three short texts (after removing punctuation marks, disregarding any utterance boundaries):

- [A] where is the ball look at this banana
- [B] where is the ball where is the ball
- [C] where ball the where the ball is is

All three texts are composed of eight words. In text [A] all words are unique while Text [B] has only 4 unique words. Importantly, Text [B] is composed of two repetitive four-word sequences, making it more repetitive (i.e., redundant) than text [A]. Text [C] has the exact same words as text [B], but does not include repeated sequences. Text [B] and Text [C] have the same level of lexical repetition (four words, each repeating twice), but differ in the recurrence of multi-word sequences: text [C] is *less* repetitive, and therefore less redundant. Because of these differences, text [A] will have the highest entropy rate (it is the least redundant of the three texts), text [B] will have the lowest entropy rate (it is the most redundant) and text [C] will be somewhere in between. This is indeed the case: h(A) = 1.91, h(B) = 1.02, h(C) = 1.29.

Entropy rate provides us with a global measure of redundancy that is sensitive to the order of words; is based on repetitions of varying sizes; and can be used to compare different language registers. In Study 1 we compare entropy rate in samples of IDS directed to different ages. If speech to younger infants is more redundant, then entropy rate should increase with age. In Study 2, we explore the impact of lexical and sequential repetition on the decrease in redundancy. This investigation is motivated by the language development literature, where there is growing evidence for the importance of multi-word sequences in language acquisition (Abbot-Smith & Tomasello, 2006; Arnon, 2016, 2021; Arnon & Christiansen, 2017; Schwab & Lew-Williams, 2016b; Stoll, Abbot-Smith, & Lieven, 2009). Building on this literature, Study 2 examines the influence of changes in multiword repetitions on entropy rate.

2. Study 1

2.1. Method

2.1.1. The corpora

We used two developmental corpora, both taken from the CHILDES database (MacWhinney, 2000). To investigate our first question — whether redundancy is impacted by the interlocutors' perceived overall proficiency level, we used the NewmanRatner Corpus (Newman, Rowe, & Bernstein Ratner, 2016) which contains transcripts of 122 English-speaking mother-child dyads, who were recorded during lab visits as part of a longitudinal study. We chose this corpus because the recorded lab visits are all 15 minutes of free play with the same box of toys. This characteristic is an important and necessary feature for our purposes of comparing the effect of the interlocutor (specifically, children's age) on the redundancy in caregivers' speech: this corpus holds the convesational context constant while maintaining naturalistic caregiver speech.

The recordings in the NewmanRatner Corpus were taken at the ages

¹ A word is defined by white spaces, and is therefore identified orthographically by form, not lemma. That means *dog* and *dogs*, for example, are identified as different words. A string is any sequence starting from one word.

of 7,10,11,18 and 24 months, though not all children took part in all time points. We analyzed only three time points from this corpus (7-, 11- and 24-months), since the rest did not have enough data (see *Procedure* below). See Table 1 for a summary of IDS measures of this corpus. Note that unlike the more familiar developmental pattern (Schwab & Lew-Williams, 2016a; Soderstrom, 2007), type-token ratio does not increase with age in this corpus. We come back to this issue in Study 2.

To support the use of entropy rate as a viable measure of global redundancy in IDS, we wanted to make sure any developmental effects found for the NewmanRatner Corpus are also generalizable to other developmental corpora. We therefore conducted the same analyses on a second corpus. The second corpus we used is the Providence Corpus (Demuth, Culbertson, & Alter, 2006), accessed using the childesr package in R (version 2018.1, Sanchez et al., 2019). This corpus contains dense longitudinal recordings of six monolingual English-speaking children between the ages of 1-4 years, collected during naturalistic interactions with their caregivers at home. We analyzed the data of the three children that had the most data - Ethan, Lily and Naima. We created two developmental time points by dividing the recordings into two age bins (12-24 months and 24-36 months; beyond this age there was not enough data to match our purposes). Whereas the NewmanRatner Corpus has many dyads recorded under fairly controlled lab conditions, the Providence Corpus has few children recorded across multiple activities. This corpus therefore has less controlled experimental settings than the NewmanRatner Corpus. Together these two sets of corpora provide a rich representation of IDS in different contexts, thereby strengthening the utility of entropy rate to serve as a novel measure of global redundancy in IDS. See Table 2 for a summary of IDS measures of the Providence Corpus.

2.1.2. Procedure

We calculated entropy rate using the Hrate package (Bentz et al., 2017) in R (all analyses were conducted using R Version 3.6.1, R Core Team, 2019). Since this is the first developmental investigation of entropy rate measures, we were confronted with several challenges. First, how do we get reliable entropy rate estimates? This is crucial, since the calculation of entropy rate is reliable only for samples of sufficient sizes (Bentz et al., 2017). Second, how do we compare entropy rate across age groups? Previous studies compared entropy rate (and similar measures) across languages using parallel texts, like the Bible, where the same content is translated into many languages. When the text is held constant, differences can more easily be attributed to structural differences between the languages and not to differences in content (Bentz et al., 2017, 2016; Ehret & Szmrecsanyi, 2016; Juola, 2008; Koplenig et al., 2017; Lupyan, 2019; Lupyan & Dale, 2010). However, we do not have such parallel texts at our disposal when doing developmental research (nor is it clear what such parallel texts would be for naturalistic interactions). Our third challenge was to ensure that differences in entropy rate between age groups will indeed reflect the effect of age, and not other factors. In what follows we outline our solutions in brief, and elaborate on each in subsequent sections. First, we used a stabilization criterion to determine the minimal sample size needed for reliable entropy rate estimates. In other words, we chose a sample size beyond which entropy rate remained almost unchanged. Each sample was composed of multiple conversations (a conversation is one unit of recording in the original corpora). In order to compensate for the fact

Table 1

Summary of IDS properties in the NewmanRatner Corpus (values represent mean per interaction).

Age in months	Number of words	Type-token ratio	Mean length of utterance
7	759.5	0.32	3.37
11	759.23	0.3	3.28
24	1160.1	0.25	3.8

Table 2

Summary of IDS properties in the Providence Corpus (values represent mean per interaction).

Child	Age in months	Number of words	Type-token ratio	Mean length of utterance
Naima	12–24	4075.2	0.17	5.25
	24–36	3738.1	0.19	5.66
Lily	12–24	4606.75	0.18	4.55
	24–36	4553.51	0.2	5.65
Ethan	12–24	3066.18	0.19	4.18
	24–36	3688.82	0.19	4.75

that different age groups are not "parallel texts", we took multiple samples from each age group – the number of samples was also determined using a stabilization criterion. Finally, to make sure that any differences found between the samples from different ages indeed reflect age, rather than a sampling confound, we created an additional sampling procedure (the mixed-age condition), where we followed the same procedure but took samples from mixed ages: if entropy rate is impacted by the infants' age, then the change in entropy rate should disappear in this condition. The next sections describe these processes in detail.

2.1.2.1. Division into age groups. The NewmanRatner Corpus is already divided into different age groups, corresponding to different lab visits. In the Providence Corpus, however, children were recorded weekly/ monthly. We could not measure entropy rate for each recording since it does not provide enough data for calculating reliable estimations of entropy rate (see next section).² To overcome this, we grouped the recordings into two age bins: 12-24 months and 24-36 months, resulting in two age points per child (The mean age across all transcripts and children is 18.25 months (SD = 3.51) in the younger age group and 29.84 months (SD = 3.32) in the older age group). The sampling procedures described below were applied separately to each age group in the NewmanRatner Corpus (3 groups in total), and to each age group for each child in the Providence Corpus (6 groups in total). After having established the age bins, we made them similar in size: if one age group had more conversations than another age group, we randomly selected a subset of the larger corpus, to equate it in size with the smaller one (see Tables 3 and 4 for total corpus size from which samples were taken for the different age groups).

2.1.2.2. Sample structure. We created each sample by randomly selecting conversations from each age group, and aggregating them in a random order. We only included utterances spoken by caregivers. Importantly, since entropy rate is affected by the order words appear in, we kept each conversation intact.³ The resulting sample was one long

Table 3

Number of words in each age group from which samples were taken in the NewmanRatner Corpus: each sample contained 10,000 words.

7 months	11 months	24 months
92,659	85,794	90,081

² Looking at each conversation separately is not only too short in terms of data, but also from a theoretical perspective: it is unlikely that one conversation is a proper unit of analysis of IDS from the infant's perspective.

³ Except for one conversation in each sample: Since the determined sample size was fixed for all samples and all age groups, the sample size was the cut-off point of the sample, even if it was in the middle of a conversation. For example, if the sample size determined for a corpus was 10,000 words, then each sample was of exactly 10,000 words. Because of this fixed word limit, each sample is very likely to have one conversation cut somewhere in the middle. Other than that, all conversations remained intact.

Table 4

Number of words in each age group from which samples were taken in the Providence Corpus: each sample contained 20,000 words.

Ethan		Lily		Naima	Naima	
1–2 years	2–3 years	1-2 years	2–3 years	1–2 years	2-3 years	
84,991	79,052	81,456	80,343	82,726	81,647	

text of aggregated conversations (see Fig. 1 for an illustration of the sampling procedure). Sample size was the same for all age groups (based on the stabilization criteria detailed in the next sections).

2.1.2.3. Determining sample size. As mentioned, the calculation of entropy rate is reliable only for samples of sufficient sizes (Bentz et al., 2017). We followed the stabilization criterion used in Bentz et al., 2017 to determine sample size. In that paper, entropy rate was calculated for a gradually growing sample size. The sufficient sample size was defined as one where the SD of entropy rate in the next ten samples was smaller than or equal to 0.1, that is, where entropy rate did not change much with growing sample size. We applied the same procedure to our data. For each age bin, we gradually increased sample size, where each time the sample was 1000 words larger than the previous (1000, 2000, 3000...), until we reached the total number of words in that age bin. We then calculated entropy rate for each of these sample sizes, and calculated the SD of the entropy rate of each ten consecutive sample sizes, looking for the minimal corpus size (in words) where entropy rate did not change with further increase. This procedure was done for each age group separately. The final sample size was the one for which all age



Fig. 1. Illustration of the sampling procedures. Each dot represents IDS extracted from one conversation. Different colors represent different age groups (X,Y,Z). Solid frames represent samples taken from different age groups: each frame is a random assembly of conversations taken from the same age group (the separate-age condition). The dashed frame represents a sample taken from the mixed-age condition: composed of randomly chosen conversations taken from different age groups (the mixed-age condition).

groups reached stabilization (e.g., if 10,000 words were needed for a reliable calculation at 7-months, but 20,000 were needed for 24-months, our selected sample size was 20,000 for all age bins). In practice, all age groups in both corpora required the same sample size, so no such adaptation was needed. The stabilization criterion procedure was implemented using the Hrate package (Bentz et al., 2017).

2.1.2.4. Comparing entropy rate between age groups. To make sure samples are representative of the age group they are taken from, we took multiple samples from each age group.⁴ This method was inspired by previous work that compared the complexity of non-parallel texts by taking multiple samples from each text (e.g., learner essays of beginners vs. advanced English learners, Ehret & Szmrecsanyi, 2019). The number of samples was determined using a similar procedure as the one used to determine the minimal sample size. After finding the minimal reliable sample size (see previous section), we created 1000 samples of this size from the same age bin, increasing the number of samples by 10 each time (10 samples, 20 samples, 30 samples...). We then calculated the average entropy rate for each number of samples (e.g., average entropy rate for 10 samples: 5.57, average entropy rate for 20 samples: 5.6, and so on). Finally, we calculated the SD of the average entropy rate of each five consecutive N of samples (10-50,60-100,100-150...). The sufficient number of samples was defined as one where the SD of the average entropy rate in the next five samples was smaller than 0.01. For both corpora, the determined number of samples for each age bin was 100. We therefore randomly chose 100 samples out of the 1000 initial samples created for each age group. See supplementary material for the stabilization of entropy rates for both corpora.

2.1.2.5. Variables of interest and hypothesis testing. After determining the sample size and the number of samples, we calculated entropy rate for each sample. To ensure that any difference in entropy rate stems from an effect of age, and is not simply the result of comparing multiple samples, we conducted another comparison where we aggregated conversations from different ages into the same sample. That is, we took the same number of samples (with the same sample size) but we allowed the mixing of conversations from different ages. We call this the mixed-age condition, in contrast with samples taken from separate ages – which we call the separate-age condition. The mixed-age condition had the same number of groups and the same number of samples as the separate-age condition. However, the groups in the mixed-age condition were not selected based on having the same age. In both conditions the procedure was the same: each sample was consisted of random sampling of conversations without replacement (that is, a conversation could not appear more than once in a given sample). If entropy rate is affected by the infant's age, then we should not see a developmental effect in the mixedage condition. See Fig. 1 for an illustration of the sampling procedure in

⁴ Instead of taking multiple samples from each age group, we could also treat all conversations within each age group as one big sample, such that we would only have one sample for each age group. Indeed, we repeated our sampling procedures with one big sample from each age group and the patterns of our results remain unchanged. This method is however problematic for two reasons: a. It doesn't allow for a statistical comparison between age groups, and b. Since entropy rate is affected by the order in which words appear, the way conversations are concatenated together can potentially impact the result. By taking multiple samples we control for that possible confound, since in each sample different conversations are grouped together in different order. This allows us to make sure that any age differences found cannot be driven by the particular order in which conversations were grouped together within each age group. We nevertheless conducted an additional analysis where each conversation appeared in only one sample. This method resulted in only a few samples per each age group (which is why we did not use it as our main analysis), but ensured that samples were maximally different from one another. We report the results of this alternative sampling procedure for each corpus below. Importantly, the results held in both sampling procedures.

both conditions.

3. Results

Since we applied our sampling procedure separately to each corpus, and since each had a different division into age groups, we report the results for each corpus separately.

3.1.1. The NewmanRatner Corpus

Our stabilization criteria resulted in samples containing 10,000 words, and in taking 100 samples from each age group. We calculated entropy rate for each sample. As can be seen in Fig. 2, entropy rate seems to increase with age: while entropy rate doesn't differ between 7-months (5.15) and 11-months (5.15), it is higher for 24 months (5.32). A one way ANOVA showed a significant effect of age (F(2) = 116.7, p < 0.001, $\eta 2 = 0.44$), indicating that, as predicted, parents speak less redundantly to older children. Simultaneous Tukey tests found the differences between 24 months and 7 months and between 24 months and 11 months to be significant, both at p < 0.001 level. There was no difference, however, between 7 and 11 months (p = 0.99).⁵ Importantly, no change in entropy rate was found in our mixed-age condition (5.29, 5.29, 5.28, F (2) = 0.92, p = 0.4, $\eta 2 = 0.006$), indicating that the differences in entropy rate reflect an effect of age, and are not a byproduct of comparing multiple samples.

3.1.2. The Providence Corpus

The Providence Corpus differs from the NewmanRatner Corpus, containing longitudinal naturalistic interactions between a small number of caregivers and children, engaged in various activities. Here, we investigate the entropy rate trajectory for each child separately. For this corpus, the stabilization criteria resulted in samples of 20,000 words, and in taking 100 such samples from each group (recall there are six groups in this corpus: two age points for three different children). We calculated entropy rate for each sample. As can be seen in Fig. 3, entropy rate increases with age for all three children (Ethan: 5.58, 5.89; Naima: 5.77, 5.89; Lily: 5.87, 6.09). We ran a two-way ANOVA to compare the effect of child, age and their interaction on entropy rate. There was a significant effect of age (F(1) = 1007.4, p < 0.001, $\eta_p^2 = 0.63$), indicating that entropy rate increases with development; a significant effect child $(F(2) = 446.04, p < 0.001, \eta_p^2 = 0.6)$, indicating individual variation between children; and a significant interaction (F(2) = 66.96, p < 0.001, $\eta_p^2 = 0.18$), indicating individual variation in the developmental change in entropy rate.⁶ In line with our prediction, entropy rate increased with age: speech directed to infants was less redundant as they were older. We next analyzed the results of the mixed-age condition. For the Providence corpus, the mixed-age condition was created by taking conversations from the same child, but from different ages. No differences in entropy rate were found between the mixed-age groups (Ethan: 5.86,

5.85; Lily: 5.86, 5.86; Naima: 5.85, 5.86; F(1) = 0.075, p = 0.78, $\eta_p^2 = 0.0001$), or between children (F(2) = 1.8, p = 0.16, $\eta_p^2 = 0.006$). Again, finding no differences in the mixed-age condition indicates that the differences in entropy rate obtaining between age groups indeed stems from age (see Fig. 3).

3.2. Discussion

In line with our prediction, we found an increase in entropy rate in the NewmanRatner Corpus, suggesting that parents use less redundant speech with their children as they grow older. These findings suggest that speakers are sensitive to the proficiency level of their interlocutors, and increase redundancy with less proficient interlocutors, as would be predicted by the communicative efficiency hypothesis (Buz et al., 2016; Jaeger & Buz, 2017; Pate & Goldwater, 2015). Importantly, we found the same decrease with age in both corpora, illustrating that parents use less redundant speech when talking to older children across different conversational settings. Taken together, these findings also provide a new way to assess the overall level of redundancy in IDS: instead of investigating individual properties of IDS (lexical diversity, syntactic diversity), entropy rate offers a holistic evaluation of how repetitive and therefore, predictable — is the input that young children hear.

But what speech characteristics underly the increased redundancy in younger ages? IDS has lower lexical diversity compared to ADS, with lexical diversity increasing with age (Schwab & Lew-Williams, 2016a; Soderstrom, 2007). The decrease in redundancy we found could reflect the increase in lexical diversity that occurs with development. However, IDS has another property that could contribute to the increase in entropy rate: IDS has substantial repetition of multi-word sequences, in the form of frequent frames or repeated chunks (Arnon, 2016; Cameron-Faulkner et al., 2003; Ferrier, 1978; Stoll et al., 2009). These larger sequences are claimed to play an important role in language learning by providing preferred patterns for multi-word production (Bannard & Matthews, 2008; Lieven, Behrens, Speares, & Tomasello, 2003; Lieven, Salomo, & Tomasello, 2009), and serving as early building blocks for learning morphological and syntactic regularities (Abbot-Smith & Tomasello, 2006; Arnon & Clark, 2011; Arnon, McCauley, & Christiansen, 2017; Reali & Christiansen, 2007; Skarabela, Ota, O'Connor, & Arnon, 2021). Such units are predicted to be relied on less in later development and in second language learning (Arnon & Christiansen, 2017; Siegelman & Arnon, 2015). Importantly, both types of repetitions (lexical and multiword) will make the input more redundant and will be reflected in a lower entropy rate. Recall the comparison of the three texts in Section 1.1: text [A] has a lower entropy rate than text [B] because it has more repetitions of words and multi-word sequences, while text [C] has a higher entropy rate than text [B] only because of multi-word sequences: both texts have the same lexical diversity, but text [B] consists of two multi-word chunks while text [C] does not. Importantly, while there is ample evidence in the developmental literature documenting a decrease in lexical repetitiveness with age, there is no evidence, to our knowledge, for a similar decrease in multi-word repetitiveness, even though such a decrease is predicted under usage-based approaches to language acquisition (Tomasello, 2003). Moreover, such a decrease is directly predicted, but has not been tested, under the Starting Big approach to language learning (see Arnon, 2021 for a review).

In Study 2 we investigate which properties underlie the differences in redundancy we found. Specifically, we examine the role of multi-word sequences in these differences. We do this by creating an additional comparison condition where in each sample words are shuffled within each sentence. If the increase in entropy rate is driven only by an increase in lexical diversity, then the effect should not be impacted by disrupting the *order* in which words appear in sentences. If, however, input to younger children is more redundant also due to having more repetitions of multi-word sequences, then the difference between age groups should become smaller (or disappear) once words within sentences are shuffled (since this shuffling will disrupt the repetition of

⁵ As mentioned in Footnote 4, we conducted an additional analysis where each conversation appeared in only one sample. This resulted in considerably fewer samples (8 per each age group). We calculated and compared entropy rate for each sample, and the pattern of results remains the same: no difference between 7 months (5.16) and 11 months (5.15), and an increase with age for 24 months (5.28). This was confirmed by a one way ANOVA (F(2) = 6.58, p = 0.006, $\eta 2 = 0.38$).

 $^{^6}$ Here also we conducted an additional analysis where each conversation appeared in only one sample. This resulted in 3 samples per each age of each child. We calculated entropy rate for each sample, and the pattern of results remained the same (Ethan: 5.54, 5.92; Naima: 5.78, 5.93; Lily: 5.9, 6.09). A two way ANOVA confirmed an effect of age (F(1) = 40.53, $p < 0.001, \eta_p^2 = 0.77$) and child (F(2) = 16.26, $p < 0.001, \eta_p^2 = 0.73$). The interaction between them was marginally significant (F(2) = 3.23, $p = 0.07, \eta_p^2 = 0.35$).



Shuffled age

Fig. 2. Entropy rate as a function of age in the NewmamRatner Corpus where (A) each group of samples are taken from a different age (separate-age condition) and (B) where age is shuffled between the groups (mixed-age condition). Note that the x-axis in (B) is made similar to that of (A) just for demonstration (i.e., there are no age differences between samples in the mixed-age condition). Each dot represents a sample, numbers represent group means.

multi-word sequences).

4. Study 2

Study 2 set out to investigate the impact of lexical vs. multi-word repetition on the increase in entropy rate found in Study 1. We can contrast three different predictions. First, since lexical diversity (measured by type-token ratio) increases with age (Schwab & Lew-Williams, 2016a; Soderstrom, 2007), the effect we found in Study 1 could reflect parents' use of a more diverse vocabulary with their children as they grow older. If this is the only effect driving the results in Study 1, then shuffling words within each sentence should not impact the entropy-rate effect: diversity of lexicon is not impacted by the order in which words appear in a sentence. A second possibility is that the results of Study 1 are driven by a developmental decrease in repetitions of multi-word sequences, rather than a decrease in lexical repetitions. If the results of Study 1 are driven only by multi-word repetitions, then an increase in entropy rate should not be found when the words in each sentence are shuffled. Finally, if the results of Study 1 are driven by both types of repetitions, then disrupting the order in which words appear in sentences should still result in an increase in entropy rate, but entropy rate measures should be overall higher compared to the non-shuffled text, since one of the factors contributing to redundancy in the input (repetitions of multi-word sequences) will no longer be there.

It is important to note that when using this method to gauge the

influence of multi-word sequences on entropy rate, utterance length needs to be taken into account. Shorter utterances contain more chances for repeated multiword sequences in their shuffled version. For example, compare the possible word permutations in the utterances "look at the bunny" and "look at the teeny weeny doggy". Possible frequent multiword sequences such as "look at" and "look at the" are more likely to remain intact after shuffling when they are part of the first utterance compared to the second. This issue is not of concern in the current study, since MLU increased in less than a word over development in the corpora we used (see Tables 1&2), but it should be taken into account when applying this measure to other corpora.

4.1. Method

4.1.1. The corpora

We used the same corpora as in Study 1.

4.1.2. Procedure

We used the same sampling procedure as in Study 1 (taking 100 samples from each age bin in each corpus), but now in each sample we shuffled the words within each sentence before calculating entropy rate.⁷ That is, within each sample (taken from multiple conversations

⁷ We used the corpora's existing segmentation to sentences.

group means.



Fig. 3. Entropy rate as a function of age and child in the Providence Corpus where (A) each group of samples within each child is taken from a different age (separate-age condition) and (B) where age is shuffled within each child (mixed-age condition). Note that the x-axis in (B) is made similar to that of (A) just for demonstration (i.e., there are no age differences between samples in the

mixed-age condition). Each dot represents a sample, numbers represent

within each age group; 10,000 words in the NewmanRatner Corpus and 20,000 words in the Providence Corpus) we shuffled the words in each sentence (but sentences still occurred in the same order as in the original conversation they belonged to). Importantly, the samples were taken anew (that is, the samples taken in Study 2 were not identical to the ones taken in Study 1). The shuffling procedure was conducted using the *sample()* function in R (R Core Team, 2019). By shuffling words in this way, we removed their sequential position in a sentence, but left their lexical diversity intact.

4.2. Results

4.2.1. The NewmanRatner Corpus

Fig. 4 shows entropy rate across age bins when words were shuffled within each sample. As can be seen, while there seem to be differences in entropy rate between the three ages, these differences are smaller when compared to Study 1 (7 months: 5.76, 11 months: 5.69, 24 months: 5.81). In addition, entropy rates are higher overall compared to those found in Study 1: by disrupting the order in which words appeared in sentences in the original conversations, the texts were now less predictable and therefore less redundant. A one-way ANOVA revealed an effect of age (F(2) = 72.61, p < 0.001, $\eta 2 = 0.2$). Simultaneous Tukey tests found the differences between 24 months and 7 months and between 24 months and 11 months to be significant, both at p < 0.001level. There was no difference, however, between 7 and 11 months (p =0.44). Note that although the difference between 7 and 11 months is not significant, there seems to be a trend of a decrease in entropy rate between these ages. We believe this pattern reflects an idiosyncratic property of this corpus: the fact that lexical diversity in the corpus as a whole (measured by type-token ratio) does not increase with age (see Table 1), in contrast with the more common developmental pattern (Soderstrom, 2007). The lack of a developmental increase in lexical diversity in this corpus is most likely due to its specific characteristics: in this corpus, mothers and children of different ages were invited to the lab to play with the same box of toys. This setup most likely confines lexical diversity, thereby placing an upper bound on its increase.

We directly compare the impact of age on entropy rate in Study 1 and Study 2, by running a two-way ANOVA on a pooled dataset of both samples. Samples were coded as "non-shuffled" (Study 1) or "shuffled" (Study 2). A two-way ANOVA with study (non-shuffled vs. shuffled), age and their interaction revealed a significant effect of age (F(2) = 200.3, p < 0.001, $\eta_p^2 = 0.4$), indicating that entropy rate is impacted by age. There was also a significant effect of study (F(1) = 7734.5, p < 0.001, η_p^2



Fig. 4. Entropy rate as a function of age in the NewmanRatner Corpus when each sentence in each sample is shuffled for words (Study 2) compared to when no such shuffling occurs (Study 1). Each dot represents a sample, numbers represent group means.

= 0.93), showing that entropy rate was higher in Study 2 (where words in each sample were shuffled within each sentence), indicating the samples were less predictable. There was also a significant interaction between age and study, (F(2) = 28.65, $p < 0.001, \eta_p^2 = 0.09)$, indicating that the impact of age on entropy rate is stronger in Study 1.

In sum, the findings from the NewmanRatner Corpus show that entropy rate increases with age when the words within sentences are shuffled, but the effect of age is smaller than in Study 1. This suggests the decrease in redundancy is not only affected by an increase in lexical diversity, but also by a decrease in repetitions of multi-word sequences.

Overall, the fact that the NewmanRatner corpus shows no decrease in lexical repetitiveness (in contrast to the more typical state of affairs), contributes an interesting and important angle to the current findings: we find an increase in entropy rate despite the fact that this corpus does not show the typical decrease of lexical repetitiveness. This highlights even further the decrease of redundancy over age, and in particular the role of multi-word repetitiveness in contributing to this decrease. Taken together, these results point to the role of multi-word units in the developmental decrease in redundancy in IDS.

4.2.2. The Providence Corpus

Fig. 5 shows entropy rate after word-shuffling for the Providence corpus. As in the NewmanRatner corpus, we still find here an increase of entropy rate with age, albeit the increase seems to be slightly smaller (Ethan: 6.15,6.42; Lily: 6.42,6.58; Naima: 6.33,6.42). Entropy rates were higher overall compared to the ones found in Study 1, indicating that the input becomes less predictable (and therefore, less redundant) as words are divorced from their original sequential position in the sentence. A two-way ANOVA with child, age and their interaction as factors confirmed there was a significant effect of age (F(1) = 1014.6, p< 0.001, η_p^2 = 0.63), indicating a developmental increase in entropy rate; a significant effect of child (F(2) = 516.3, p < 0.001, $\eta_p^2 = 0.63$), indicating individual variation in entropy rate; and their interaction (F (2) = 103.2, p < 0.001, $\eta_p^2 = 0.26$), indicating a different effect of age on entropy rate for the different children. To directly compare Study 1 and Study 2, we ran an ANOVA on the pooled dataset, coding each sample for the study it came from (non-shuffled (Study 1) or shuffled (Study 2)). We used a three-way ANOVA to compare the effect of study (nonshuffled vs. shuffled), age, child and their interactions on entropy rate.



Fig. 5. Entropy rate as a function of age and child in the Providence Corpus when each sentence in each sample is shuffled for words (Study 2, darker shades) compared to when no such shuffling occurs (Study 1, lighter shades). Each dot represents a sample, numbers represent group means.

There was a significant effect of age (F(1) = 1997, p < 0.001, $\eta_p^2 = 0.63$), indicating a developmental increase in entropy rate across studies; a significant effect of child F(2) = 941.85, p < 0.001, $\eta_p^2 = 0.61$), indicating individual variation; and a significant effect of study (F(1) =15,049.2, p < 0.001, $\eta_p^2 = 0.93$), indicating higher entropy rates for Study 2 (as, again, shuffling the words within sentences render the input less redundant). Finally we found all the interactions were significant: age and child: F(2) = 161.14, p < 0.001, $\eta_p^2 = 0.21$, indicating a different age effect for different children; age and study: F(1) = 23.43, p < 0.001, $\eta_p^2 = 0.02$ – indicating that the effect of age is larger in Study 1 compared to Study 2 (though note the small effect size); and child and study: F(2) = 5.2, p = 0.006, $\eta_p^2 = 0.008$, indicating children had different entropy rates in the different studies. Note that even though Lily shows the largest increase of MLU with age - slightly over one word (from 4.55 to 5.65, see Table 2), she does not show the largest increase in entropy rate between those ages. This further indicates that changes in MLU do not affect the current results.

In sum, for two of the three children, entropy rate still increased with age when the words in the samples were shuffled, but the effect of age was smaller than in Study 1. In other words, the increase in redundancy seems to be affected both by an increase in lexical diversity and by a decrease in repetitions of multi-word sequences.

4.3. Discussion

Study 2 set out to investigate what underlies the decrease in redundancy found in both corpora in Study 1. By shuffling words within each sentence, we removed the possible contribution of multi-word repetitiveness, allowing us to tease apart the effect of lexical repetitiveness and multi-word repetitiveness (both characteristic of IDS; Cameron-Faulkner et al., 2003; Schwab & Lew-Williams, 2016a, 2016b; Soderstrom, 2007). Our results suggest that the developmental increase in entropy rate found in Study 1 reflects not only an increase in lexical diversity, but also a decrease in repetitions of *multi-word* sequences. In both the NewmanRatner Corpus and the Providence Corpus, there was still an increase in entropy rate, but a smaller one, suggesting that lexical repetitiveness underlies some of the effect in Study 1, but not all of it. That is, repetitions of multi-word sequences contribute to the developmental effect found in Study 1 for both corpora.

5. General discussion

Speakers are known to modify their speech on the basis of global estimations of their interlocutors (Arnold, 2008; Loy & Smith, 2020). However, very few studies have investigated whether speakers adapt the redundancy in their speech based on such global properties. Our study set out to ask whether speakers speak more redundantly with interlocutors that are perceived to have difficulty in understanding, specifically, infants learning their first language. We examined redundancy developmentally in IDS, with two goals in mind: (1) Testing whether redundancy is impacted by the interlocutors' perceived overall proficiency level, and (2) Providing a novel measure of global redundancy in IDS. We used an information-theoretic measure called entropy rate, which was previously used to compare language complexity crosslinguistically (Bentz et al., 2017), but has not been used to compare different types of speech within a language. This measure provided us with a global estimate of redundancy that is sensitive to the order of words: texts where words are more predictable given previous words will have a lower entropy rate, reflecting greater redundancy.

Applying this measure to samples of IDS taken from infants of different ages, we found that, in line with our predictions, parents tend to speak more redundantly to their children when they are younger. These results are compatible with recent findings in the phonetic domain, showing that adult speakers provide more redundant acoustic signals when talking to infants compared to proficient adult speakers (Pate & Goldwater, 2015; Tippenhauer et al., 2020; Uther et al., 2007).

Our results expand these findings in two important ways. First, they broaden the scope of speakers' adaptations from the phonetic level to the word and multi-word level. Second, they show that differences in redundancy are not only found when talking to infants as opposed to adults (a salient distinction that is likely to be computationally easy for the speaker), but also when talking to children of different ages. They illustrate speakers' fine-grained sensitivity to the proficiency level of their interlocutors. These results are compatible with other findings showing that adult modifications in IDS are impacted by infants' age (for example, re-using children's words more the younger they are, Snow, 1972; Yurovsky, Doyle, & Frank, 2016).

In Study 2 we set out to investigate which properties of IDS drive the developmental decrease in redundancy. IDS is characterized both by lower lexical diversity (Soderstrom, 2007), and by having frequently recurring multi-word sequences (Arnon, 2016; Cameron-Faulkner et al., 2003; Stoll et al., 2009). While previous studies document a developmental decrease in lexical repetitiveness (Schwab & Lew-Williams, 2016a; Soderstrom, 2007), there is no parallel evidence for a decrease in repetitions of multi-word units. By measuring entropy rate on samples where words within each sentence were shuffled, we could tease apart the effect of lexical and multi-word repetitiveness on entropy rate. We found that in both tested corpora, the developmental decrease in redundancy was not a mere reflection of an increase in lexical diversity. Rather, the decrease in redundancy was also impacted by a decrease in repetitions of multi-word sequences. This finding serves as the first indication that repetitions of multi-word sequences decrease with infant's age, and moreover, that this decrease makes the input globally less redundant. In what follows, we discuss the relevance of our findings for communicative efficiency and language development.

5.1. Communicative efficiency

According to the communicative efficiency hypothesis, speakers should speak more redundantly when there is increased comprehension difficulty. This may happen when the message is unpredictable in context (Aylett & Turk, 2004; Kurumada & Jaeger, 2015; Levy & Jaeger, 2007; Mahowald et al., 2013); when there is high environmental noise (Van Summers et al., 1988; Zhao & Jurafsky, 2009); or when the interlocutor has difficulty decoding the message (Buz et al., 2016). Several studies demonstrate the impact of the interlocutor's perceived difficulty in local cases of misunderstanding within a conversation (Buz et al., 2016; Lockridge & Brennan, 2002; Roche et al., 2013). However, only a handful of studies show an increase in redundancy dependent on global characteristics of interlocutors, and they are mostly focused on the phonetic domain (Pate & Goldwater, 2015; Tippenhauer et al., 2020; Uther et al., 2007). Our results indicate that the tendency to reduce or increase redundancy is impacted by the interlocutor, and that this impacts speech beyond articulation. Specifically, speakers seem to increase redundancy when conversing with younger language learners.

Importantly, a direct test of this hypothesis would require comparing how speakers encode the exact same message when talking to interlocutors with varying levels of proficiency. However, it is virtually impossible to hold the message constant when using naturalistic speech. In this study, we followed the commonly-held practice of holding constant the experimental setting which elicits the conversation (Bard et al., 2000; Bard & Aylett, 2004; Berman & Slobin, 1994; Pate & Goldwater, 2015; Rodriguez-Cuadrado et al., 2017; Tal et al., 2023; Van Engen et al., 2010). To this end, we used the NewmanRatner corpus that records free play under the same circumstances with the same box of toys. Note, however, that this method is still a compromise between the need to control the message and maintain ecological validity at the same time. The current findings therefore show that caregivers speak more redundantly to younger children. They, however, do not directly show that parents are encoding exactly the same messages with greater redundancy when conversing with younger children. Caregivers might be using more complicated messages when talking to older children, while

engaging in the same activity with the same prompts. When looking at naturalistic infant-directed speech as a whole, it is impossible to adjudicate between these two interpretations. The current results are however in line with existing findings from other linguistic domains, where the message can be held constant (e.g., articulation, phonetics), showing increased redundancy in speech directed to younger children (Pate & Goldwater, 2015; Tippenhauer et al., 2020; Uther et al., 2007). Future work should broaden the existing findings by testing the communicative efficiency hypothesis on other linguistic domains, in which the message could be better controlled (at the expense of naturalistic speech). For example, other domains, such as morphosyntax, where communicative efficiency has been shown to impact production choices, could also be modified when talking to younger learners. For instance, we might see fewer morphological reductions (e.g, isn't vs. is not, Frank & Jaeger, 2008) or reduced omission of optional syntactic words (e.g., optional case marking, Kurumada & Jaeger, 2015) in speech directed to younger children compared to speech directed to older children or adults. We are currently conducting several corpus-studies to test the impact of the interlocutor on redundancy in morphosyntactic domains.

A second open question has to do with how proficiency is perceived and how interlocutor type impacts changes in redundancy. In the current study, we tested the impact of conversing with young language learners on redundancy, where their lower proficiency is deduced from their age. Additional properties of interlocutors could also impact the trade-off between minimizing effort and maximizing understandability. Within a developmental setting, other factors beyond age could indicate lower proficiency: for example, speech directed to late talkers could be more redundant than speech directed to more proficient talkers of the same age. Between adults, conversations between friends are typically based on more common knowledge compared to conversations between strangers (Fussell & Krauss, 1989), which could make them less redundant. More broadly, if increased redundancy is driven by the perceived proficiency level of the interlocutor, rather than interlocutors' age, then we should find similar changes in redundancy in speech directed to adult learners. The literature provides two contrasting predictions for whether we modify redundancy similarly in speaking with children and adult L2 learners. On the one hand, we may not produce more redundant speech when conversing with adult L2 learners: adult learners have cognitive capacities similar to those of adult native speakers, and might therefore be perceived as more competent, despite their low proficiency level in the language in question. In addition, a recent hypothesis predicts such differences on the basis of the idea that children and adults might have different learning constraints: the Linguistic Niche Hypothesis suggests that child learners benefit from redundancy, but adult learners do not, or at least not to the same degree (Dale & Lupyan, 2012; Lupyan, 2019; Lupyan & Dale, 2010). Following this hypothesis, speech directed to child learners is predicted to be more redundant than speech direct to adult learners, because of speakers' adaptation to these different learning constraints. On the other hand, the impact of the interlocutor on redundancy might be strictly related to perceived language proficiency, regardless of age. If this is the case, then the differences found between age groups in the current study should be replicated with adult learners. Specifically, speech directed to adult learners should be more redundant compared to speech directed to adult proficient speakers. However, it is not easy to find suitable corpora to test this comparison since the calculation of entropy rate requires samples of sufficiently large sizes (Bentz et al., 2017), of which there are very few for L2-directed speech. Moreover, the comparison requires parallel corpora of speech directed to adult learners vs. proficient speakers (for example, two sets of corpora that contain similar conversational content, but differ in the proficiency level of the interlocutors). However, a separate study we conduct on a different sort of redundancy supports the prediction above. In this study we compare participants' descriptions of the same picture book to child learners of different ages, adult learners and adult proficient speakers. We find that speakers use more redundant references (e.g., the boy rather than he) when their

interlocutors are learners compared to proficient speakers. Importantly, the same pattern is found for both child and adult learners, illustrating the effect is not (only) driven by general cognitive maturation processes of the listeners, but rather by language proficiency (Tal et al., 2023).

5.2. Language development

The language acquisition literature highlights the prevalence and the facilitative role of different types of repetitions in IDS: lexical repetition, frequently recurring phrases, variation sets and repeated chunks (Arnon, 2016; Cameron-Faulkner et al., 2003; Küntay & Slobin, 1996; Schwab & Lew-Williams, 2016a; Stoll et al., 2009). Importantly, these properties are typically investigated separately. Here, we provided a novel measure to assess the overall redundancy in IDS, and to investigate its developmental trajectory. Having shown that IDS becomes less redundant with development, we found that this effect is influenced by repetitions of multi-word units. This finding joins the growing literature on the prevalence and importance of multi-word sequences in language development (Abbot-Smith & Tomasello, 2006; Arnon, 2016, 2021; Arnon & Christiansen, 2017; Stoll et al., 2009). There is growing evidence that multi-word sequences are building blocks for language learning (Arnon & Clark, 2011; Bannard & Matthews, 2008; Fernald & Hurtado, 2006; Skarabela et al., 2021). Children and adult learners are claimed to differ in their reliance on multi-word sequences in learning, a tendency that can explain some of the difference between L1 and L2 learning (Arnon, 2021; Arnon & Christiansen, 2017; Arnon & Ramscar, 2012; Siegelman & Arnon, 2015). Consequently, multi-word sequences are predicted to be more common in the inventory of L1 learners compared to adult L2 learners (Arnon & Christiansen, 2017; McCauley & Christiansen, 2014). This prediction is supported by the current study: we show here for the first time that repetitions of multi-word units in IDS decrease as the child grows older. Taken together, this finding provides additional evidence for the role of multi-word sequences in IDS. The higher frequency of multi-word repetitiveness at younger ages might aid young children's reliance on such constructions in language learning (Goldberg, 2006; Tomasello, 2003). Investigating the contribution of multi-word sequences to the increased redundancy in children's input provides a new way to test theoretical predictions regarding their prominence in language acquisition. Further work is needed to assess the relative magnitude of lexical vs. multiword repetition in changes in redundancy in speech directed to children.

Finally, the current study provides an important first step in examining redundancy developmentally in IDS using an information theoretic framework. As such, it has several limitations and raises additional questions to pursue in future work. First, while we looked at two types of developmental corpora - one with few controlled lab-recordings of many caregiver-child dyads, and one with dense recordings across multiple activities of three children - these are still only two sets of corpora. Although applying this measure to developmental corpora is challenging - because entropy rate is reliable only for samples of sufficient sizes, and because the compared conversations should be as similar as possible — future work should seek to replicate these findings in additional corpora. Second, our oldest children were 36 months old. At some point, speech directed to children should reach similar redundancy rates to speech directed to adults. When are children perceived as proficient enough for adult-like redundancy? To get an initial insight into this question, we can look at a previous study that calculated entropy rate for adult language: Bentz et al. (2017) looked at translations of the same high-register written texts in various languages, and found that entropy rates across languages display relatively narrow distributions. English (the language we looked at here) seems to be close to the mean value: 5.97. Interestingly, this value is very close to what we found for the 24-36 months in the Providence Corpus (5.89-6.09), so it could be tempting to suggest that by the age of three children are starting to be perceived as adult-like in terms of their comprehension. However, we are hesitant to draw such conclusions (even preliminary or

impressionistic ones), due to the large differences between the type of texts looked at in the current study (spontaneous spoken conversations between caregivers and infants), and the ones used by Bentz et al., 2017 (e.g., the Bible). A comprehensive comparison of entropy rate between IDS and adult speech should involve texts and contexts that are as similar as possible. Otherwise, it would be difficult to assess whether differences found are attributed to the different interlocutors or some other property of the conversation (Ehret & Szmrecsanyi, 2016, 2019; Juola, 2008). Another important way to broaden the current findings is to look at the developmental trajectory of entropy rate in languages other than English. Using a global measure of redundancy such as entropy rate provides us with a novel way to make cross-linguistic comparisons and form novel predictions. For example, it would be instructive to look at morphologically complex languages, where, compared to English, more information is expressed within words rather than using multi-word constructions. While we should still expect to see an overall decrease in entropy rate with development, we predict to find a different trade-off between lexical and multi-word repetitiveness driving this decrease. We are currently investigating these predictions. Finally, future work should compare entropy rate with other novel measures of complexity in IDS (Brunato & Venturi, 2023; Ehret, Berdicevskis, Bentz, & Blumenthal-Dramé, 2023). This would further promote our understanding of what aspects of language complexity different measures capture, and in what way IDS changes with age.

6. Conclusion

In this study we asked whether conversing with learners results in more redundant language. We tested this by quantifying redundancy in infant-directed speech using entropy rate – an information-theoretic measure reflecting the average redundancy of texts. We found that, as predicted by studies of infant-directed speech and by principles of efficient communication, speakers use more redundancy in speech directed to younger infants. The increased redundancy in younger ages was found to be influenced not only by repetitions of single words, but also by repetitions of multi-words sequences. This serves as a first indication for a decrease in such sequences over age, and joins a great volume of literature pointing at the prominent role of multi-word sequences in language learning.

CRediT authorship contribution statement

Shira Tal: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. Eitan Grossman: Writing – review & editing, Supervision, Methodology, Conceptualization. Inbal Arnon: Writing – review & editing, Supervision, Methodology, Conceptualization.

Data availability

I have shared the link to my data/code in the manuscript

Acknowledgments

We would like to thank Christian Bentz, Henry Brice, Haim Dubossarsky and Noam Siegelman for helpful discussions. We thank Noam Siegelman and Haim Dubossarsky for feedback on a previous version of this paper. This research was supported by the Israeli Science Foundation (grant number 584/16 awarded to IA, grant number 2765/21 awarded to EG and Postdoctoral grant number 106/23 awarded to ST). ST was additionally funded by the Mandel Scholion Center at the Hebrew University of Jerusalem.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.

org/10.1016/j.cognition.2024.105817.

References

- Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *Linguistic Review*, 23(3), 275–290. https://doi.org/10.1515/TLR.2006.0011
- Ariel, M. (1990). Accessing noun-phrase antecedents. Routledge.
- Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. Language & Cognitive Processes, 23, 495–527. https://doi.org/10.1080/ 01690960801920099
- Arnon, I. (2016). The nature of child-directed speech in Hebrew: Frequent frames in a morphologically rich language. In R. A. Berman (Ed.), *The Acquisition of Hebrew* (pp. 201–223). John Benjamins.
- Arnon, I. (2021). The starting big approach to language learning. Journal of Child Language, 1–22. https://doi.org/10.1017/S0305000921000386
- Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1–L2 differences. *Topics in Cognitive Science*, 9(3), 621–636. https://doi. org/10.1111/tops.12271
- Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, 7(2), 107–129. https://doi.org/10.1080/15475441.2010.505489
- Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, 92, 265–280. https://doi.org/10.1016/j.jml.2016.07.004
- Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, 122(3), 292–305. https://doi.org/10.1016/j.cognition.2011.10.009
- Aylett, M., & Turk, A. (2004). The Smooth Signal Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1), 31–56.
- Banard, C., & Matthews, D. (2008). Stored word sequences in language learning of fourword combinations. Psychological Science, 19(3), 241–248.
- Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42, 1–22. https://doi.org/10.1006/jmla.1999.2667
- Bard, E. G., & Aylett, M. (2004). Referential form, word duration, and modeling the listener in spoken dialogue. In J. C. Trueswell, & M. K. Tanenhaus (Eds.), Approaches to studying world-situated language use: Bridging the language-as-product and languageas-action traditions (pp. 173–191). MIT Press.
- 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. https://doi.org/10.3390/e19060275
- Bentz, C., Ruzsics, T., Koplenig, A., & Samardži, T. (2016). A comparison between morphological complexity measures: Typological data vs. language corpora. In Proceedings Of the workshop on computational linguistics for linguistic complexity (pp. 142–153). https://www.aclweb.org/anthology/papers/W/W16/W16-4117/.
- Berman, R. A., & Slobin, D. I. (1994). Relating events in narrative: A crosslinguistic developmental study. Lawrence Erlbaum Associates.
- Brennan, S. E., & Hanna, E. (2009). Partner-specific adaptation in dialog. Topics in Cognitive Science, 1, 274–291. https://doi.org/10.1111/j.1756-8765.2009.01019.x
- Brodsky, P., Waterfall, H., & Edelman, S. (2007). Characterizing Motherese: On the computational structure of child-directed Language. In Proceedings of the Annual Meeting of the Cognitive Science Society.
- Brunato, D., & Venturi, G. (2023). Why is this language complex? Cherry-pick the optimal set of features in multilingual treebanks. *Linguistics Vanguard*, 9(s1), 59–72. https://doi.org/10.1515/lingvan-2021-0017
- Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers ' subsequent pronunciations. *Journal of Memory and Language, 89*, 68–86. https://doi.org/ 10.1016/j.jml.2015.12.009
- Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, 27(6), 843–873. https://doi.org/10.1016/ j.cogsci.2003.06.001
- Chafe, W. L. (1994). Discourse, consciousness, and time. Chicago University Press.
- Clark, H. H., & Murphy, G. L. (1983). Audience design in meaning and reference. In J. F. LeNy, & W. Kintsch (Eds.), *Language and comprehension* (pp. 287–299). North-Holland Publishing Co.
- Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. Laboratory Phonology, 6(2), 243–278. https://doi.org/10.1515/lp-2015-0008
- Dale, R., & Lupyan, G. (2012). Understanding the origins of morphological diversity: The linguistic niche hypothesis. Advances in Complex Systems, 15(3), 1150017-1-1150017-16.
- Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis, and coda licensing in the acquisition of English. *Language and Speech*, 49(2), 137–174.
- Ehret, K. (2021). An information-theoretic view on language complexity and register variation: Compressing naturalistic corpus data. Corpus Linguistics and Linguistic Theory, 17(2), 383–410. https://doi.org/10.1515/cllt-2018-0033
- Ehret, K., Berdicevskis, A., Bentz, C., & Blumenthal-Dramé, A. (2023). Measuring language complexity: Challenges and opportunities. *Linguistics Vanguard*, 9, 1–8. https://doi.org/10.1515/lingvan-2022-0133
- Ehret, K., & Szmrecsanyi, B. (2016). An information-theoretic approach to assess linguistic complexity. In R. Baechler, & G. Seiler (Eds.), *Complexity, isolation, and variation* (pp. 71–94). de Gruyter.

- Ehret, K., & Szmrecsanyi, B. (2019). Compressing learner language: An informationtheoretic measure of complexity in SLA production data. *Second Language Research*, 35(1), 23–45. https://doi.org/10.1177/0267658316669559
- Elmlinger, S. L., Goldstein, M. H., & Casillas, M. (2022). Immature vocalizations simplify the speech of Tseltal Mayan and U.S. caregivers. *Topics in Cognitive Science*, 15, 315–328. https://doi.org/10.1111/tops.12632

Ferdinand, V., Kirby, S., & Smith, K. (2019). The cognitive roots of regularization in language. *Cognition*, 184, 53–68. https://doi.org/10.1016/j.cognition.2018.12.002

Fernald, A., & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science*, 9(3), 33–40.

Ferrier, L. J. (1978). Some observations of error in context. In N. Waterson, & C. Snow (Eds.), The development of communication (pp. 301–309). Wiley.

- Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *The 30th annual meeting of the cognitive science society* (pp. 939–944).
- Fussell, S. R., & Krauss, R. M. (1989). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, 19(6), 509–525. https://doi.org/10.1002/ejsp.2420190603
- Gao, Y., Kontoyiannis, I., & Bienenstock, E. (2008). Estimating the entropy of binary time series: Methodology, some theory and a simulation study. *Entropy*, 10(2), 71–99. https://doi.org/10.3390/entropy-e10020071
- Gibson, E., Futrell, R., Piandadosi, S. T., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, 23(5), 389–407. https://doi.org/10.1016/j.tics.2019.02.003

Goldberg, A. E. (2006). Constructions at work: The nature of generalization in language. Oxford University Press.

- Goldberg, A. E. (2019). Explain me this: Creativity, competition, and the partial productivity of constructions. Princeton University Press.
- Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. Infancy, 18(5), 797–824. https://doi.org/10.1111/infa.12006
- Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), Vol. 3. Syntax and semantics (pp. 41–58). Academic Press.
- Heller, D., Gorman, K. S., & Tanenhaus, K. (2012). To name or to describe: Shared knowledge affects referential form. *Topics in Cognitive Science*, 4, 290–305. https:// doi.org/10.1111/j.1756-8765.2012.01182.x
- Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. Journal of Experimental Psychology, 116(1), 26–37.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. Cognitive Psychology, 61(1), 23–62. https://doi.org/10.1016/j. cogpsych.2010.02.002
- Jaeger, T. F., & Buz, E. (2017). Signal reduction and linguistic encoding. In E. Fernández,
 & H. Cairns (Eds.), *The handbook of psycholinguistics* (pp. 38–81). John Wiley & Sons.
- Juola, P. (2008). Assessing linguistic complexity. In M. Miestamo, K. Sinnemäki, & F. Karlsson (Eds.), Language complexity: Typology, contact, change (pp. 89–107). Benjamins. https://doi.org/10.1075/slcs.94.07juo.
- Kempe, V., Casillas, M. (n.d.). Studying Child-Directed Speech. In L. Raviv, C. Boeckx (Eds.), The Oxford Handbook of approaches to language evolution. http://www.nbe r.org/papers/w16019.
- Koplenig, A., Meyer, P., Wolfer, S., & Mu, C. (2017). The statistical trade-off between word order and word structure – Large-scale evidence for the principle of least effort. *PLoS One*, 12(3). https://doi.org/10.7910/DVN/8KH0GB
- Kravtchenko, E. (2014). Predictability and syntactic production: Evidence from subject omission in Russian. In Proceedings of the 36th Annual Conference of the Cognitive Science Society (pp. 785–790). http://people.ucsc.edu/~ekravtch/Kravtch enko_CogSci2014.pdf.
- Küntay, A. C., & Slobin, D. I. (1996). Listening to a Turkish mother: Some puzzles for acquisition. In D. I. Slobin, J. Gerhardt, A. Kyratzis, & J. Guo (Eds.), Social interaction, social context, and language: Essays in honor of Susan Ervin-Tripp (pp. 265–286). Lawrence Erlbaum Associates.
- Kurumada, C., & Jaeger, T. F. (2015). Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language*, 83, 152–178. https://doi.org/10.1016/j.jml.2015.03.003
- Lavi-Rotbain, O., & Arnon, I. (2023). Zipfian distributions in child-directed speech. Open Mind, 7, 1–30. https://doi.org/10.1162/opmi_a_00070
- Lester, N. A., Moran, S., Küntay, A. C., Allen, S. E. M., Pfeiler, B., & Stoll, S. (2022). Detecting structured repetition in child-surrounding speech: Evidence from maximally diverse languages. *Cognition*, 221(December 2021), Article 104986. https://doi.org/10.1016/j.cognition.2021.104986
- Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schlökopf, J. Platt, & T. Hoffman (Eds.), Advances in neural information processing systems (NIPS) 19 (pp. 849–856). MIT Press.
- Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. Journal of Child Language, 30(2), 333–370. https://doi.org/ 10.1017/S0305000903005592
- Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, 20(3), 481–507. https://doi.org/10.1515/COGL.2009.022
- Linzen, T., & Jaeger, T. F. (2015). Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science*. https://doi.org/ 10.1111/cogs.12274
- Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, 9(3), 550–557.
- Lombard, E. (1911). Le signe de l'elevation de la voix. Annales de Maladies de L'oreille et Du Larynx, 37, 101–119.

Loy, J. E., & Smith, K. (2020). Syntactic adaptation may depend on perceived linguistic knowledge: Native English speakers differentially adapt to native and nonnative confederates in dialogue. https://doi.org/10.31234/osf.io/pu2qa

Lupyan, G. (2019). Larger languages have higher entropy. In Interaction and the evolution of linguistic complexity.

- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS One*, 5(1). https://doi.org/10.1371/journal.pone.0008559 MacWhinney, B. (2000). *The CHILDES-project: Tools for analyzing talk*. Lawrence
- Erlbaum.
 Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2), 313–318. https://doi.org/10.1016/j.cognition.2012.09.010
- McCauley, S. M., & Christiansen, M. H. (2014). Acquiring formulaic language: A computational model. *The Mental Lexicon*, 9(3), 419–436. https://doi.org/10.1075/ ml.9.3.03mcc

Newman, R. S., Rowe, M. L., & Bernstein Ratner, N. (2016). Input and uptake at 7 months predicts toddler vocabulary: The role of child-directed speech and infant processing skills in language development. *Journal of Child Language*, 43(5), 1158–1173. https://doi.org/10.1017/S0305000915000446

- Pate, J. K., & Goldwater, S. (2015). Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language*, 78, 1–17.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011a). The communicative function of ambiguity in language. *Cognition*, 122(3), 280–291. https://doi.org/10.1016/j. cognition.2011.10.004
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011b). Word lengths are optimized for efficient communication. Proceedings of the National Academy of Sciences of the United States of America, 108(9), 3526–3529. https://doi.org/10.1073/pnas.1012551108

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing.

- Reali, F., & Christiansen, M. H. (2007). Word chunk frequencies affect the processing of pronominal object-relative clauses. *Quarterly Journal of Experimental Psychology, 60* (2), 161–170. https://doi.org/10.1080/17470210600971469
- Roche, J. M., Paxton, A., Ibarra, A., & Tanenhaus, M. K. (2013). From minor mishap to major catastrophe: Lexical choice in miscommunication. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the annual meeting of the cognitive* science society (pp. 3303–3308).
- Rodriguez-Cuadrado, S., Baus, C., & Costa, A. (2017). Foreigner talk through word reduction in native/non-native spoken interactions. *Bilingualism: Language and Cognition*, 1–8. https://doi.org/10.1017/S1366728917000402
- Sanchez, A., Meylan, S. C., Braginsky, M., Macdonald, K. E., Yurovsky, D., & Frank, M. C. (2019). childes-db: A flexible and reproducible interface to the child language data exchange system. *Behavior Research Methods*, 51, 1928–1941.
- Schwab, J. F., & Lew-Williams, C. (2016a). Language learning, socioeconomic status, and child-directed speech. Wiley Interdisciplinary Reviews: Cognitive Science, 7(4), 264–275. https://doi.org/10.1002/wcs.1393
- Schwab, J. F., & Lew-Williams, C. (2016b). Repetition across successive sentences facilitates young children's word learning. *Developmental Psychology*, 52(6), 879–886.
- Shannon, C. E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27, 379–423.
- Siegelman, N., & Arnon, I. (2015). The advantage of starting big: Learning from unsegmented input facilitates mastery of grammatical gender in an artificial language. *Journal of Memory and Language*, 85, 60–75. https://doi.org/10.1016/j. jml.2015.07.003
- Siegelman, N., Rueckl, J. G., Steacy, L. M., Frost, S. J., van den Bunt, M., Zevin, J. D., ... Morris, R. D. (2020). Individual differences in learning the regularities between orthography, phonology and semantics predict early reading skills. *Journal of*

Memory and Language, 114, Article 104145. https://doi.org/10.1016/j.jml.2020.104145

Skarabela, B., Ota, M., O'Connor, R., & Arnon, I. (2021). 'Clap your hands' or 'take your hands'? One-year-olds distinguish between frequent and infrequent multiword phrases. *Cognition*, 211, Article 104612. https://doi.org/10.1016/j. cognition.2021.104612

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, 43 (2), 549–565.

- Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, 27, 501–532. https://doi.org/ 10.1016/j.dr.2007.06.002
- Stoll, S., Abbot-Smith, K., & Lieven, E. (2009). Lexically restricted utterances in Russian, German, and English child-directed speech. *Cognitive Science*, 33, 75–103. https:// doi.org/10.1111/j.1551-6709.2008.01004.x
- Stoll, S., Bickel, B., Lieven, E., Paudyal, N. P., Banjade, G., Bhatta, T. N., ... Rai, N. K. (2012). Nouns and verbs in Chintang: Children's usage and surrounding adult speech. *Journal of Child Language*, 39(2), 284–321. https://doi.org/10.1017/ S0305000911000080
- Tal, S., & Arnon, I. (2018). SES effects on the use of variation sets in child-directed speech. Journal of Child Language, 45(6), 1423–1438. https://doi.org/10.1017/ S0305000918000223
- Tal, S., Grossman, E., Rohde, H., & Arnon, I. (2023). Speakers use more redundant references with language learners: Evidence for communicatively-efficient referential choice. *Journal of Memory and Language*, 128, Article 104378. https://doi. org/10.1016/j.jml.2022.104378
- Tily, H., & Piantadosi, S. (2009). Refer efficiently: Use less informative expressions for more predictable meanings. In Proceedings of the workshop on the production of referring expressions: Bridging the gap between computational and empirical approaches to reference.
- Tippenhauer, N., Fourakis, E. R., Watson, D. G., & Lew-Williams, C. (2020). The scope of audience design in child-directed speech: parents' tailoring of word lengths for adult versus child listeners. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(11), 2163–2178. https://doi.org/10.1017/CB09781107415324.004
- Tomasello, M. (2003). Constructing a language: A usage-based theory of language acquisition. Harvard University Press.
- Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH ? A comparison of foreigner- and infant-directed speech. Speech Communication, 49, 2–7. https://doi.org/10.1016/j.specom.2006.10.003
- Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native-and foreign-accented english: Communicative efficiency across conversational dyads with varying language alignment profiles.
- Language and Speech, 53(4), 510–540. https://doi.org/10.1177/0023830910372495 Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., Stokes, M., & A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal* of the Acoustical Society of America, 84(3), 917–928.
- Waterfall, H. (2006). A little change is a good thing: Feature theory, language acquisition and variation sets (Doctoral dissertation). University of Chicago
- Yurovsky, D., Doyle, G., & Frank, M. C. (2016). Linguistic input is tuned to children's developmental level. In *Proceedings of the 38th annual meeting of the cognitive science* society (pp. 2094–2098).
- Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, 115(31), 7937–7942. https://doi.org/10.1073/pnas.1800521115
- Zhao, Y., & Jurafsky, D. (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*, 37, 231–247. https://doi.org/10.1016/j. wocn.2009.03.002
- Zipf, G. K. (1949). Human behavior and the principle of least effort: An introduction to human ecology. Addison-Wesley Press.