Wordform variability in infants' language environment and its effects on early word learning

Charlotte Moore^{1, 2} & Erika Bergelson²

¹ Concordia University

² Duke University

Author Note

Authors have no conflicts of interest to declare.

This work was supported by the NIH Office of the Director DP5-OD019812 to Erika Bergelson and a SSHRC doctoral fellowship to Charlotte Moore. We would also like to acknowledge the research assistants at Rochester University and Duke University who annotated the data. Code and data are available at https://osf.io/gr6ca/?view_only=1083f928c17144b5901cfd1f74f1c8e6

Correspondence concerning this article should be addressed to Charlotte Moore, 7141 Sherbrooke St W, Montreal QC, H4B 1R6. E-mail: charlotteemma.moore@concordia.ca

Abstract

Traditional views of language development suggest that noun learning involves creating a one-to-one mapping between concrete objects and their labels. In the current work, we provide evidence that real world language input to infants does not provide such tidy mappings. Instead, infants encounter many variant wordforms for familiar nouns (e.g. dog~doggy~dogs). We explore this wordform variability in 44 English-learning infants' naturalistic environments using a longitudinal corpus of infant-available speech. We looked at both the frequency and composition of wordform variability. We found two broad categories of variability: morpheme-adding changes, where words were pluralized or compounded (e.g. coat~raincoats); and wordplay, where words changed form without any associated change in meaning (e.g. bird~birdie). Wordplay occurred with a limited number of lemmas that were usually early-learned, highly-frequent, and shorter. When looking at all wordform variability, we found that individual words with higher levels of wordform variability were learned earlier than words with fewer wordforms, over and above the effect of frequency.

Wordform variability in infants' language environment and its effects on early word learning

Introduction

When we imagine an idealized word-learning scenario, we might conjure an image of a baby who sees an object like a bottle, then hears their mother say “bottle” pronounced clearly and at the opportune moment. This then leads the baby to draw a connection between the sounds they hear and the object they are attending to. Unfortunately, language learning in real world settings is rarely this straightforward. Setting aside the added difficulty of learning abstract words like verbs or function words, even learning concrete nouns requires both a correct representation of the wordform and the selection of the appropriate meaning from a theoretically infinite set of possible meanings (Quine, 1960).

Further complicating matters, there are often multiple accurate labels for speakers to choose from to refer to a single object. If we take *dog* as an example, we can use a synonym, as in “hound” or label at different levels of taxonomy, as in “labradoodle” or “mammal”. Similarly, for any given wordform there are often multiple possible lemmas, as is the case for homophones (where two words share sound but are unrelated in meaning, e.g. “pair”~“pear”) and polysemes (where two words share the same sounds and one is a semantic extension of the other, e.g. “sheet” for *sheet of paper* and also for *bed sheet*).

Each of these types of ambiguity has been maintained in productive language systems (homophones: Zipf, 1945; polysemes: Floyd, Goldberg, & Lew-Williams, 2020; Srinivasan, Berner, & Rabagliati, 2019; synonyms: Kipfer, 2011; taxonomic relations: Clark & Grossman, 1998; Diesendruck & Shatz, 2001), suggesting that language users are able to accommodate these types of relationships, despite the ambiguity. This may be because listeners can accommodate words that are unlikely to be confused, as in the case of homophones, which often belong to different syntactic categories or are used in very different semantic contexts (Piantadosi, Tily, & Gibson, 2012). For instance, 20-month-olds

learn novel meanings for known words only if the novel meanings are sufficiently distinct from the known meanings (Dautriche, Fibla, Fievet, & Christophe, 2018). In contrast, polysemes may be retained due to the close semantic ties between meanings. These extensions have been shown to help toddlers structure the acquisition of new related words by highlighting probable referents (Floyd et al., 2020). Thus, evidence suggests that learning multiple meanings for the same wordform may not be a challenge if the meanings are either very distant, or very similar.

This does not appear to hold true for synonym learning. Even synonyms with no phonological overlap (e.g. dog~hound) pose a learning challenge; 16-month-olds struggle to learn a second label for a familiar object but succeed above chance to learn a first label for a novel object. This synonym effect persists when toddlers are taught new labels for two known objects at 24 months (Liittschwager & Markman, 1994). Studies of mutual exclusivity also provide evidence that by 17 months, toddlers resist assigning two words to the same object if there is a novel referent available (Halberda, 2003). A meta-analysis of mutual exclusivity tasks suggests that this bias only strengthens during early childhood (Lewis, Cristiano, Lake, Kwan, & Frank, 2020). Taken together, these findings suggest that toddlers may struggle in learning situations where multiple, phonologically dissimilar wordforms map to the same object. However, these laboratory experiments rely on the existence of true synonyms, where the meaning of the known label and the new label are identical.

In laboratory word learning tasks, novel words are encountered in a highly constrained context, often with little information about the register, connotation, or taxonomic level of the novel word that might distinguish a real synonym from known words, which may add a layer of challenge to lab studies that does not exist in the real world. Contextual information has been shown to help two-year-olds map novel labels to familiar objects (Diesendruck & Shatz, 2001), suggesting that linguistic context supports

toddlers in learning multiple labels for the same objects. Further, Clark (1987) refutes the idea of complete synonyms, suggesting that multiple wordforms cannot have truly identical meanings. Instead she proposes the Principle of Contrast: every wordform must pull its semantic weight in order to remain in the lexicon, and thus form changes must correspond to meaning changes. Clark posits that if two forms do not denote different entities in the world, then they denote social factors like register or dialect. In this framing, learning synonyms and taxonomic relations requires learning meaning differences in tandem with wordform differences. While the principle of contrast may help determine which words are preserved in the lexicon and become conventional, here we provide evidence that many wordforms in early speech may not fully follow the Principle of Contrast. For many of the highest-frequency words in infants' environments, wordform variability is extensive and meaning differences are not always clear, even with contextual support.

We explore this facet of wordform variability by analyzing another type of wordform-to-meaning mismatch, one that is likely particular to infancy and early childhood: dubbed **wordplay**. We define wordplay as any unconventional wordform that is used to denote the same concept as a better-known conventional object word *and* the unconventional form shares some overlapping phonology with the conventional form. In this definition, "conventional" can be defined as the most frequent form in adult-directed speech, and/or the wordform one would find in a dictionary. Wordplay is similar to synonymy in that multiple wordforms denote the same concept, but unique in that all wordforms share at least some sounds. For example, "dog" and "doggy-woggy" can both be used to refer to the same physical entity and share most of their speech sounds, but "doggy-woggy" is a wordplay form, since it is certainly a less conventional wordform and shares overlapping phonetic material with "dog". Wordplay words also differ from synonyms in that they are not conventional words in the language and speakers are likely not choosing them to clarify their meaning but instead to mark their register as child-directed. Interestingly, multiple wordplay forms can exist for the same lemma;

“doggy-woggy” and “doggy-dog” are interchangeable, and seemingly do not differ in register nor dialect, providing a challenge to the notion that all words in the speech environment pull distinct semantic weight. The overlap of both sound and meaning in wordplay, as well as the variety of forms that wordplay takes without changing register, may have unique effects on learnability.

Many different types of wordplay have already been studied, but not as a single, unified category. Special vocabulary for use with infants is a cross-linguistic phenomenon of infant-directed speech: reduplicated words (repeating all or part of a word) and diminutives are common in the infant-directed speech of many languages (Ferguson, 1964; Gleason, Perlmann, Ely, & Evans, 1994; Savickien & Dressler, 2007). Onomatopoeia is particularly well represented in speech to infants relative to speech to adults (Laing, 2019) and in some languages, caregivers extensively use special vocabulary when speaking with their young children (Mazuka, Kondo, & Hayashi, 2008). Other proposed lexical features of infant-directed speech are phonological simplification and an increase in words with CVCV syllable structure (Soderstrom, 2007) as well as an increase in iconic speech (e.g. using *choo-choo* to refer to a train; Ota, Davies-Jenkins, and Skarabela (2018)). However, most of these lexical features were identified through parental report, a subjective measure of language use. Here we seek to characterize these types of changes more systematically and also capture any additional types of wordplay changes that previous methods did not unearth, perhaps because caregivers are not aware of all the wordform variability they employ.

In a corpus analysis of speech to infants aged 6-17 months, we begin by examining the overall prevalence of wordform variability in the corpus, including wordplay alongside meaningful changes like compounding and pluralization. We predict that wordplay will be less common than meaningful changes but will still occur frequently in the dataset, as many wordplay forms have been previously reported in other descriptions of infant-directed

speech, as noted above. We then explore if we find different kinds of wordplay than what other work has discussed, and explore which features of a word predict whether that word is used in wordplay in our corpus. Since wordplay is a hallmark of infant-directed speech, we predict that a word's frequency in our corpus, its age of acquisition, and its length may play a role in predicting wordplay prevalence. Finally, we explore the effects of wordform variability writ large on word learning, looking both at the learnability of specific words and at whether language outcomes differ for children who hear different proportions of wordplay. We also expand to look beyond wordplay at wordform variability more broadly, to explore whether concepts with more variable labels are learned later than words with more stable labels. Wordform variability includes not only wordplay but also other changes to a word's conventional form like pluralizing or compounding. Here too we have no directional predictions - on the one hand, increased variability could delay learning by creating many mutual exclusivity problems throughout learning. On the other hand, many features of wordplay have been shown to boost learning. We thus turn to the data to discover the answers.

Methods

Many aspects of the methods have been previously reported in more detail elsewhere (e.g. Bergelson, Amatuni, Dailey, Koorathota, & Tor, 2018). We provide critical aspects of the methods here as well for readers' convenience.

Participants

Families were recruited from a mid-size city in the Northeastern United States. Forty-six infants were recruited, but two families dropped out after the first month of data collection, leaving a final sample of 44 participants. All infants heard $\geq 75\%$ English in their environments, were full term, and had no known hearing or vision impairments. There was one pair of dizygotic twins. Families were compensated for their participation; compensation increased as participation continued. The sample was very well-educated

compared to national averages: 50% of mothers had a graduate degree, and 41% had an Associates or Bachelor's. The remaining 9% had at least a high school diploma. The sample was also very racially homogeneous: 95% of infants were white, and 5% were mixed-race.

Procedures

Parents provided consent at the beginning of the yearlong data collection effort. Every month from when infants were 6 months old, families completed demographics questionnaires and the Macarthur-Bates Communicative Development Inventory: Words and Gestures (MCDI:WG) form (Fenson et al., 1994). Participants also contributed one full day of audio recording (around 16 hours of data) and one hour-long video recording on separate two days each month from when children were 6 months until they were 17 months. Due to technical errors, one video and one audio file are missing from the final dataset, resulting in 1,054 total unique recordings.

To collect video recordings, researchers brought recording materials to infants' homes and set up a camcorder (Panasonic HC-V100 or Sony HDR-CX240) in the room that parents expected to be in most for the following hour. Parents were asked to bring the tripod with them if they switched locations. When possible, infants were provided with soft headbands or hats outfitted with two Looxcie cameras (22g each) to capture the infant's point of view. Occasionally, parents wore the camera headbands instead of their children, when infants refused. One camera was angled slightly up and the other slightly down to capture as much of the visual field as possible. After setup, researchers departed for the hour to ensure that their presence wouldn't disrupt the family's dynamics. All videos were approximately one hour long.

For the audio recordings, parents were provided with LENA recorders (LENA Foundation, Boulder, CO) and vests for infants to wear that contain a pocket for the recorder. LENAs are designed for capturing daylong recordings of infants and young

children. Parents were instructed to have their child wear the vest from when they woke up until bedtime, with exceptions for naps and bath-time. Final recordings were approximately 16 hours long. Parents were provided with the option to turn the recorder off at any time but were asked to capture the maximum possible amount of time throughout the day.

Data Processing

Detailed data pipeline and scripts are available at <https://app.gitbook.com/@bergelsonlab/s/blab/data-pipeline>. For each video recording, one combined multi-angle video file from all footage sources (i.e. Looxcies, camcorder) was created in Vegas and with in-house scripts. Videos were then annotated manually in Datavyu by research assistants. Audio recordings went through LENA’s proprietary data processing pipeline to identify utterances and periods of silence, and then were annotated manually in CLAN (MacWhinney & Wagner, 2010) by research assistants.

All videos were fully annotated for concrete nouns. For audio recordings, in months 6 and 7, the entire daylong recording was annotated. Silent periods (e.g. during naps) were tagged using a pre-processing script, then trained researchers listened to the remaining file and manually identified concrete nouns. In months 8-13, the four most talkative hours according to the LENA software were annotated. These talkative hours were identified as containing the most conversational turns and child vocalizations, determined with a 60-minute sliding window passed over the audio in five-minute increments until four non-overlapping hours were found. For months 14-17, the three most talkative hours were annotated. All told, researchers annotated ~80 hours of data per participant across audio and video for concrete nouns and a few of their properties, as detailed below..

Noun Annotations

Annotators tagged all concrete nouns that were deemed to be easily audible to the child but not necessarily directed at the child. The child’s own utterances were also

annotated in the same manner. Each annotation included the word as it was spoken by the speaker and the lemma (i.e. the conventional form) of the word. For example, if a child heard “doggies,” the annotation would include “doggies” as the wordform and “dog” as the lemma. For each annotation, the speaker (e.g. mother, brother), the type of utterance (e.g. question, declarative), and whether the object was present (yes/no) were also recorded. For noun-noun compounds e.g. “raincoat” both nouns were annotated as a single entry. Proper names for concrete objects (e.g. Little Blue Truck, a popular children’s book) were included in their entirety as lemmas. Each annotation was created by one researcher and then checked for accuracy by another. In case of disagreement, both researchers discussed until consensus was reached.

Identifying wordplay in the corpus

The process above generated 359,793 annotations. Of the total annotations, 109,839 contained at least one change between the lemma and spoken wordform (e.g. “dog” vs. “doggy”). The first author compared each unique realized word to its lemma and noted up to three phonological transformations that occurred to change the lemma into the realized word. See Table 1 for examples of phonological transformations.

While all changes were noted, a wordform was only categorized as wordplay if it met all of the following criteria:

- At least one phonological change exists between the spoken wordform and the lemma, i.e. at least one sound in the conventional wordform was added, removed, or changed.
- The lemma and the realized word could denote an identical referent.
- The realized word and the lemma shared some phonological material in common e.g. same first sounds or syllables.

One exception to the rules above was words that denoted private body parts or

descriptions of diaper/potty contents ($N = 8$ lemmas), where all euphemisms were bundled together with their lemmas due to the enormous variety in wordforms, and classified as wordplay.

Changes that did not qualify as wordplay included the addition of meaningful morphemes such as pluralization (e.g. “dogs” with lemma “dog”) or distinguishing information (e.g. “roller skates” with lemma “skate”). These types of changes were classified as “morpheme-adding.” When words had multiple transformations, if any of the transformations were wordplay then the word was classified as wordplay. For example, the word “doggies” has two transformations between it and “dog:” it has a -y suffix indicating diminutivization (wordplay), and it has an -s suffix to mark plurality (morpheme-adding). Thus, “doggies” was classified as a wordplay instance. Infants’ own utterances were excluded from this process, since our focus in this work is on the characteristics of the input, and very early productions are notoriously challenging to characterize (Moore, Dailey, Garrison, Amatuni, & Bergelson, 2019; Vihman & McCune, 1994). The result of this identification process was a corpus where all changes between the lemma and the realized form were identified and characterized, first as whether the word qualified as a wordplay word, and then the record of the transformations (whether wordplay like diminutivization, morpheme-adding like pluralization, or both).

Operationalizing lemma properties

We posited three variables that might interact with wordplay prevalence for a given word: age of acquisition, frequency, and word length. Age of acquisition is likely to predict wordplay since both early-learned words and wordplay are associated with infant-directed speech (Frank, Braginsky, Yurovsky, & Marchman, 2021; Ota et al., 2018). We thus predict that early-learned words will be more likely to license wordplay than later learned words. We operationalized early-learned words as those present on the Words and Gestures or Words and Sentences forms of the MCDI. Words that were not on the MCDI were classified

as later-learned. We used presence on the MCDI as our proxy for early-learnedness because the words on the MCDI were selected such that even the most difficult words were known by at least 5% of the children in the instrument’s age range (Fenson et al., 1994), suggesting that all of the words on the MCDI could conceivably be in a 30-month-old’s vocabulary. Thus early-learnedness was a binary variable, with a 1 for words on the MCDI and a 0 for words not on the CDI. We acknowledge that this is a very rough proxy for age of acquisition, but seemed like a reasonable approach for this exploratory analysis.

The second predictor was frequency, operationalized as overall tokens of any wordform that refer to a given lemma in the corpus, collapsing across all months and families. Given the prevalence of frequency effects across first language acquisition (Ambridge, Kidd, Rowland, & Theakston, 2015), we anticipate that here too we may find them such that more frequent words are more common targets for wordplay. We thus include frequency in the model to quantify the size of the effect and to examine any interactions.

Our final predictor was word length. We predict that shorter words are more likely to undergo wordplay, since many wordplay operations lengthen a word (e.g. reduplication and suffix adding) which is less articulatorily practical for long words (e.g. hippopotamus vs. cat) While some wordplay operations do shorten long words (e.g. hippo), we predict that the relative frequency of lengthening operations outweigh the effects of shortening operations, given the prevalence of relatively shorter words in English nouns that children hear. Lemma length was operationalized as the number of phonemes in the lemma according to the CMU pronunciation dictionary. Lemmas which were not in the CMU dictionary had their phonemes counted by a trained linguist.

Results

We begin by describing the prevalence of wordplay and morpheme-adding changes relative to words where no changes are made between the lemma and the wordform. This

enables us to address our prediction that wordplay will commonly occur in the dataset but will be less frequent than morpheme-adding changes. We then focus on characterizing wordplay words in aggregate, examining the frequency of different change types. After this, we present a logistic regression predicting which of the lemmas in the corpus will have attested wordplay forms. Finally using mixed-effects linear models we look at the effect of wordplay on learning, and then broaden our scope to examine the effects of all wordform variability on learning.

Changes from the lemma

There were three ways that any given wordform could potentially present itself in the corpus: 1) wordforms could show *no changes* between the lemma and the realized wordform, 2) words could undergo *morpheme-adding* morpho-phonetic changes like pluralization or compounding, or 3) words could undergo *wordplay* as characterized above. Words that underwent wordplay alone or wordplay in addition to morpheme-adding were still categorized as wordplay, e.g. both “doggy” and “doggies” were counted as wordplay instances even though “doggies” contains both a wordplay change and a morpheme-adding change. Thus, each word form is only ever counted in one of the 3 categories: no change, (solely) morpheme-adding, and wordplay.

In terms of token counts, wordforms with no changes from the lemma, like “book”, were the most prevalent: wordforms in the dataset appeared in their lemma forms 247,296 times, making up the bulk of tokens (69% of all tokens). In other words, the conventional forms of the nouns in the dataset were the most common. That said, lemmas where *all* tokens were in the lemma form (dubbed “frozen” words) were relatively rare; tokens of such lemmas occurred just 5.61 times per lemma on average ($SD = 17.11$). By types, 3,752 lemmas appeared exclusively in their lemma forms with no attested changes (58% of types). These frozen words included many proper names for people and media (e.g. *Big Bird*), as well as many numbers and letters of the alphabet. Frozen words also included the 1,669

Figure 1. Violin plot showing the relative frequency of each lemma change type (+0.1 to avoid infinite values). Y-axis shows log frequency. Each grey point represents one lemma’s number of appearances with that change type. Black points represent the mean number of of tokens with that change type per word. For demonstrative purposes, red points are the specific counts for the lemma ‘book’ when it appears with each change type. Examples of each type are presented next to each red point.

lemmas that occurred only once in the corpus. The most common frozen lemma was “snow” with 510 occurrences in the corpus. Figure 1 shows the relative frequency of these three change types (no change, morpheme-adding, wordplay) for each lemma. Figure 2 shows the relative proportions of each change type.

Morpheme-adding wordforms like “books” were also common in the dataset. By tokens, a total of 32,143 morpheme-adding tokens occurred over the whole dataset (25% of tokens). By types, 2,254 lemmas appeared with morpheme-adding changes at least once (35% of types). Each lemma that appeared with morpheme-adding changes words did so 27% of the time on average, suggesting that morpheme-adding changes were fairly common in lemmas that licensed them.

Figure 2. The relative proportions of each change type. Left shows the proportion of all tokens in the corpus that appeared with each change type. For example, ‘book’ would be counted in the ‘no change’ segment, ‘books’ in the morpheme-adding segment, and ‘booksies’ in the wordplay segment of the left bar. On the right is the proportion of all lemmas in the corpus where a wordform displayed each change type at least once. For example, the lemma ‘book’ is counted as a part of the wordplay segment of the plot because it licensed wordplay at least once in the dataset, ‘hair’ is in the morpheme-adding column because it occurs as ‘hair’ and ‘hairs’, and ‘snow’ is in the no change column because it is frozen (i.e. only occurs as ‘snow’).

Wordplay words, like “booksies”, were the rarest in the dataset. By tokens, 21,552 total wordplay tokens occurred in the corpus (6% of all tokens). On average, wordplay-licensing words appeared with their wordplay forms 9.8% of the time. Wordplay alternations co-occurred with morpheme-adding changes for 16% of tokens with wordplay. By types, only 423 lemmas ever showed any wordplay (out of the 6,429 different lemmas in the corpus, i.e. (7% of lemmas). Notably, these words were many of highest-frequency lemmas in the corpus. Exemplifying this, of the 50 most frequent nouns in the corpus, 43 showed wordplay. Of all the lemmas captured in the dataset, 323 (5% of lemmas) appeared at least once in all three categories: at least once with no change, at least once with only morpheme-adding, and at least once with wordplay.

Thus, the findings from this section suggest that wordplay makes up less than 10% of the input that children hear, but that it primarily affects the most high-frequency words in infants’ environments. Furthermore, changes from the lemma in general were fairly frequent, affecting 31% of the total tokens in the corpus. We now turn to look at the composition of wordplay changes in more detail.

Composition of wordplay changes

Wordplay took many forms in the dataset, but some transformations were much more prevalent than others (see Table 1). The top five alternations occurred in 17,164 tokens (69% of all wordplay tokens). We discuss these top 5 change types here.

By far the most frequent change both by type and by token was /i/-epenthesis, where an -/i/ suffix is added to the end of the wordform, as in “doggy” or “birdy”. This is one of many English morpheme options to diminutivize a word. Cross-linguistically, diminutive morphology like the -/i/ suffix does not necessarily suggest that the object being referenced is smaller than usual but instead provides an attitudinal marker of affection towards the listener (Kempe, Brooks, Mironova, Pershukova, & Fedorova, 2007). In contrast, for

example the -let morpheme (as in “owlet” which means a small owl) actually denotes smallness (Schneider, 2003). /i/-epenthesis diminutivization was pervasive and accompanied other changes like pluralization or word shortening 45% of the time. In second place by tokens were vowel changes, where one or more of the vowels in a word was replaced (e.g. “bonono” for banana). The high prevalence of this change was driven primarily by “cat” becoming “kitty,” a very high-frequency word.

The third most frequent change type was reduplication, where all or part of a wordform is repeated (e.g. “birdy-bird”). Both partial and total reduplication were common, and wordforms were sometimes repeated multiple times (see the example in Table 1). The fourth and fifth most frequent changes, weak syllable deletion and shortening, both consist of deleting some portion of the lemma. The main difference between the two categories is that with shortening, the syllable with primary stress in the lemma is deleted (e.g. “mac” for macaroni), whereas with weak syllable deletion, the syllable with primary stress is preserved. Many of the words that undergo shortening or weak syllable deletion have longer lemmas, potentially producing a subgroup of longer words that license wordplay (e.g. “vegetable”, “medicine”). We return to the potential reasons that these types of changes are so prevalent in the Discussion. A more complete list of all wordplay changes can be found in the Supplemental Materials.

Predicting which lemmas license wordplay

In this section we explore why some lemmas have attested wordplay forms and some do not. Is it simply random, or tied in a principled way to measurable properties of the noun itself? To assess this, we conducted a logistic regression predicting whether a lemma had and wordplay tokens in the corpus (1) or not (0). We sub-sampled our data to only include lemmas that occurred 10 or more times in the corpus to ensure that each included word had a sufficient chance to display wordplay in the dataset, since wordplay occurs on average 10% of the time in words that license it (i.e for lemmas that occurred with

Table 1

The 8 most common types of wordplay (see details about the top 5 types in text). Description shows a label for the change type. Lemma shows the dictionary form of the word. Example shows a prototypical example selected from the dataset. Token count shows the number of tokens in the dataset including changes of that change type. Type shows how many unique instances of that change type are attested in the dataset. Row 9 shows the total number of wordplay words there are in the dataset, and row 10 shows the total number of words in the dataset, for ease of comparison.

Description	Lemma	Example	Token Count	Type Count
y-epenthesis	dog	doggy	12,071	463
vowel change	apple	ooples	2,931	100
reduplication	bird	birdy-bird	1,981	278
weak syllable deletion	refrigerator	fridge	1,482	166
shorten	hippopotamus	hippo	1,127	45
simplify	bottle	baba	546	1
initialism	pajamas	pjs	543	10
euphemism	breast	booby	543	58
Total Wordplay			25,002	1,602
Total Words			358,651	6,429

Table 2

Model estimates from logistic regression predicting wordplay prevalence in lemmas which appeared 10 or more times in the corpus. Length is in phonemes. Frequency is in number of tokens. MCDI refers to whether the lemma was represented on either form of the MCDI.

Term	Log Odds	SE	Z =	p =
Intercept	-2.194	0.232	-9.461	<0.001***
frequency	0.003	0.001	2.760	0.006**
early-learned	1.493	0.516	2.896	0.004**
length	-0.048	0.039	-1.250	0.211
frequency*early-learned	-0.003	0.001	-4.539	<0.001***
frequency*length	0.001	0.000	3.655	<0.001***
early-learned*length	-0.102	0.105	-0.978	0.328

wordplay at all)¹. Predictors were early-learnedness (binary), word length (continuous), and word frequency (continuous) as operationalized in the Methods section. The three highest-frequency words classified as early-learned words were “dog”, “baby”, and “book”. The highest-frequency later-learned words were “poop”, “bag”, and “clothes”.

All two-way interactions were also included, resulting in the following final model formula: $Wordplay \sim (wordfrequency + MCDI + length)^2$ and included 1,844 lemmas (the exponent in the logistic equation indicates the inclusion of all possible two-way interactions but not the three-way interaction, which was omitted because it did not improve model fit ($F = 2.14$, $p = 0.14$)).²

¹ Conclusions are the same when we expand the dataset to include lemmas that occurred 5 or more times.

² Because we preemptively removed the long tail of the data’s word frequency range by sub-sampling based

The output of the logistic regression is presented in Table 2. The value of the intercept indicates that the log odds of a lemma exhibiting wordplay are low, corresponding to a predicted probability of 10% when all other predictors are set to 0. Two main effects emerged: frequency and age of acquisition. A significant main effect of frequency indicates that as a word’s frequency increases by 100, it becomes 100.28 times more likely that the lemma will license wordplay. The significant main effect of being early-learned (i.e. being a word on the MCDI) shows that such words are 4.40 times more likely to license wordplay. There was no main effect of word length.

These main effects were qualified by two significant interactions (see Figure 3). First, there was an interaction between word frequency and early-learnedness; for lower frequency nouns, early-learned words were more likely to license wordplay. However, as a word increased in frequency, later-learned words were more likely to license wordplay than early-learned words, as seen in the different slopes of the logistic curve in Figure 3. For the highest-frequency words in the data, early-learnedness had less predictive power, likely because of the main effect of frequency wherein almost all high-frequency words license wordplay. Second, there was a significant interaction between word frequency and length. At low frequencies, length appears to have a negligible effect, with shorter words showing a slightly higher baseline probability of licensing wordplay. However, as frequency increases, long words license wordplay than shorter words. This pattern aligns well with the prevalence of shortening operations we observed in the previous section, and makes intuitive sense: as a long word is said more, there seems to be increasing communicative pressure to shorten it.

on frequency, we do not log transform any of our predictors in this model; the full range of course shows a strong Zipfian distribution of frequency data so prevalent in language.

Figure 3. Significant interactions from the logistic regression using the predicted trends from the logistic regression model. Panel A shows how a word’s early-learnedness (colour) interacts with the word’s frequency (x-axis) to predict a given word’s probability of licensing wordplay. Panel B shows the relationship between word length in phonemes (colour) and lemma frequency (x-axis, log transformed for illustrative purposes) to predict a given word’s probability of licensing wordplay. Confidence bands represent the 95% confidence interval.

Wordplay varies by family, but not by age

Next we ask whether wordplay usage is stable across time. Since our data is longitudinal, we can examine whether infant age has an effect on caregivers’ use of wordplay. To examine whether infants heard different amounts of wordplay as they got older, we correlated the child’s age in months with the number of different wordplay types present in that month’s recordings. Two potential confounds were taken into account when planning this analysis. First, the number of annotated hours per recording varied as a function of the child’s age, as stated in the Methods section. Further, in months where only a subset of the recording was annotated, hours were selected to contain the most talk; this precluded a simple “wordplay-per-hour” conversion. Second, the language density in video

is significantly higher than in audio recordings (Bergelson et al., 2018), leading the video data to have a potentially outsized effect on token and type counts in months where fewer hours of audio are annotated. To circumvent these potential confounds, we opt here to limit our scope to only video when examining the effect of age on wordplay. See Figure 4 for a visualization of the distribution of wordplay types across time for each family.

Figure 4. X-axis shows the age of the infants in months. Y-axis shows the number of unique wordplay types that each infant heard in the one-hour video recording taken each month. Individual points represent the total for one infant in that month. Violins show the distribution. Point ranges show the mean and the 95% confidence interval

This analysis did not reveal a relationship between children’s age and the amount of wordplay they hear (types: $r = 0.07$, $p = 0.09$, tokens: $r = 0.03$, $p = 0.50$). We thus collapse across child age going forward, using data from the entire year of data collection when calculating per-child summary statistics in subsequent analyses.

All families heard at least some wordplay, from 3% to 11% of the unique noun types in the input per family. Family differences were more accurately obvious when looking at the range of tokens of wordplay: while 2% of noun tokens were wordplay in the family with the least wordplay, but up to 11% of noun tokens were wordplay in the family with the most

Table 3

Fixed effect estimates (Est.) from the comprehension mixed effects model predicting when a lemma will first be understood based on its log frequency in each child's environment, and the number of wordforms each child heard to refer to that lemma.

Term	Est.	Std. Error	t =	df	p =
Intercept	15.67	0.58	27.23	297	<0.001***
log frequency	-0.29	0.08	-3.70	280	<0.001***
unique wordforms	-0.17	0.02	-8.45	5324	<0.001***

wordplay. This finding provides two insights. First, while wordplay is a relatively rare phenomenon, it does occur in the language input of each child in our sample. Second, families vary in the degree to which they use wordplay.

Wordform variability and learning

We next step beyond wordplay and ask whether wordform variability more broadly has an effect on infants' word learning. We answer this question by looking at the role of wordform variability on the learning of individual words by each child. We do this with mixed-effects linear models predicting the earliest age that words were reported to be understood or produced by the children in our sample.

Our outcome variable for the mixed-effects linear models used each child's MCDI results for each word, with separate models for comprehension and production vocabulary). That is, we take parental report of whether the child understood or produced a word in a given month as our dependent variable. The predictor variables were the number of unique wordforms each lemma appeared with for each child and the total log frequency of the lemma in the entire corpus; both models contain random intercepts for child and for

lemma. A wordform was counted if it differed from the conventional form in any way, including routine morphological changes like pluralization and compounding. For example, for the conventional form “baby”, “babies” and “robo-baby” would both be counted as instances of wordform variability. Log frequency was used here since frequency data is logarithmically distributed. If an infant was never reported to understand/produce a particular word, or never heard it in the corpus, then the datapoint was excluded from analysis. Our final datasets for this analysis included reported vocabulary data for all distinct lemmas per child that were both present on the MCDI and heard/said by the child. This resulted in an average of 121.8 lemmas included per child in the comprehension model, and 37.1 in the production model. We want to clearly note that MCDI data at the item level goes beyond the tool’s intended use. At the same time, Wordbank data suggests that individual MCDI items can be a fruitful avenue for predicting a child’s overall vocabulary (Frank et al., 2021), providing support for our approach here.

Comprehension. The model for comprehension contained 5,357 observations, with each individual observation representing one child’s minimum age of comprehension for one lemma. Including an interaction term did not improve model fit ($\chi^2 = 1.35$, $p = .246$) thus the final model contains our two predictors (lemma’s log frequency and the number of wordforms used to refer to the lemma within the home recordings) with no interaction term. A summary of the fixed effects can be found in Table 3.

The intercept in this model represents the minimum predicted age of comprehension for a lemma that occurred once with a single wordform. We found a significant main effect for both the lemma’s log frequency and the number of wordforms used to refer to the lemma. As expected, lemmas that were more frequent in infants’ input were reported to be understood earlier. As a lemma increases in frequency by one order of magnitude, children were predicted to understand that word nearly 9 days earlier. Furthermore, lemma that appeared with more wordform variability were also reported to be understood earlier. For each additional wordform a word has, comprehension is predicted around 5 days earlier.

Table 4

Fixed effect estimates (Est.) from the production mixed effects model predicting when a lemma will first be spoken based on its log frequency in each child's environment, and the number of wordforms each child heard to refer to that lemma.

Term	Est.	Std. Error	t =	df	p =
Intercept	17.15	0.49	34.96	453	<0.001***
log frequency	-0.06	0.06	-0.98	422	0.329
unique wordforms	0.38	0.17	2.19	716	0.029*
interaction	-0.06	0.02	-2.73	678	0.007**

See Figure 5.

Production. The model for production contained 1,446 observations (37.1 per child on average). This is many fewer observations than in the comprehension model because infants' reported productive vocabularies were smaller than their receptive ones throughout the study (as expected for our infant age range). The base model was the same as above but predicted parent-reported production on the MCDI instead of comprehension. Here the inclusion of an interaction term between our two predictors did improve model fit ($\chi^2 = 6.96$, $p = .008$) and so was retained in the final model. A summary of the fixed effects can be found in Table 4.

In contrast to the model for comprehension, frequency did not significantly predict variance on its own. Furthermore, the number of unique wordforms had a significant main effect, but the direction of effect was opposite from in the comprehension model. Here more wordforms predicted later learning. However, this main effect was qualified by a significant frequency by wordform interaction. The effect of wordform variability was dependent on the frequency of the word, with more wordform variability contributing to earlier

Figure 5. Plots showing the predicted effects of lemma frequency and the number of attested wordforms per lemma on when infants are reported to understand (left) or say (right) words on the MCDI:WG. Lines represent the predicted data for the mean frequency value, one standard deviation above, and one standard deviation below the mean. Note that only the production model shows a significant interaction. Comprehension model’s interaction is also plotted for comparison purposes.

production in high-frequency words, but weaker wordform variability effects for lower frequency words. See Figure 5. Taking the comprehension and production results together, this pattern suggests that wordform variability may be broadly helpful for word learning, particularly when infants have access to a sufficiently large pool of tokens to learn from, including wordplay tokens. The effect of multiple wordforms is less clear for low-frequency words: variability seems to speed word learning across the board for comprehension, but may have no effect or even slow vocabulary uptake when words are lower-frequency. The differences for comprehension and production may also reflect differences in the instruments’ utility at the item level for these two types of parental vocabulary estimate.

Discussion

In this paper, we explored wordform variability in children’s everyday language input for a sample of English-learning children in North America. In particular, we took a close look at wordplay, defined as wordforms that not only mean something more or less identical to their conventional counterpart, but have substantial overlap in speech sounds as well. We explored the prevalence of wordplay in a longitudinal corpus of infant-available speech. We showed that while wordplay was a heterogeneous phenomenon with many different forms, some changes were prevalent. When wordplay occurred, it was mostly likely to feature word shortening, vowel changes, y-epenthesis, and/or reduplication. Further, it was most likely to occur in high-frequency words, though early-learnedness and word length also interacted with frequency to predict wordplay. We found evidence of wordplay in every family to varying degrees, suggesting that wordplay is widespread in infants’ environments. However, wordplay was low-frequency overall, though some families used it more than others. More broadly, wordform variability - which includes wordplay among other changes - was much more common in the nouns infants heard and significantly predicted a lemma’s age of acquisition: high-frequency words which had more wordform variability were understood and produced earlier than words with fewer wordforms by 18 months.

Wordform variability associated with earlier learning

Despite the literature showing that multiple wordforms for the same lemma can pose a learning challenge, (Halberda, 2003; Liittschwager & Markman, 1994), we find here that wordform variability supports learning, particularly for the highest-frequency nouns infants heard. Why might this be? One possibility is that when infants are learning words, they are simply not listening very closely to the details of the wordform, and are instead capturing a more holistic version of the wordform (Charles-Luce & Luce, 1990; Jusczyk, 1993). Under this view, infants hearing wordform variability might ignore the differences between wordforms, and the benefit of wordform variability would be the increased token

count. However, this explanation contradicts a robust literature shows that infants have highly-specified sound representations for many concrete nouns by 12 months (e.g. Mani & Plunkett, 2010; Von Holzen & Bergmann, 2021). Infants represent even irregular nouns robustly at 18 months (Moore & Bergelson, 2021), suggesting that wordform variability does not disrupt the creation of detailed speech sound representations. Thus, it seems likely that young children can not only detect the differences between wordforms, but that these differences are helpful for learning in some way. Our results suggest that sufficient access to the conventional form of a word may play an important role: high frequency conventional forms may help children bolster these forms’ phonetic representations. This in turn may help infants identify them when encountering wordform variability; the conventional form is the one they hear most often. This could explain why we find that wordform variability facilitates earlier production in high-frequency wordforms in particular.

But why is wordform variability helpful at all, when synonymy, a related phenomenon, poses a challenge for learners (Liittschwager & Markman, 1994)? The differences between these two constructs are informative. Notably, wordform variability isn’t directly comparable to previous work on true synonyms, in that synonyms typically have no phonological overlap between the wordforms (e.g. couch/sofa); listeners cannot rely on sound similarity to tie the meanings together. Instead, listeners must use semantic and contextual information to tie synonyms to the same meaning; this may get increasingly challenging as it would require overriding the otherwise useful mutual exclusivity heuristic, which emerges by around 17 months (Halberda, 2003). In contrast, wordplay and other forms of wordform variability provide the listener with all of the information they receive in the case of synonyms *and* additional scaffolding from phonological overlap between the forms. Much how polysemy is argued to improve learning by leveraging semantic extension (Floyd et al., 2020), wordform variability may help improve learning by leveraging phonological extension.

The overlap of both sound and meaning content—and the context and distributional information this results in—may be particularly potent in helping infants appropriately group related forms of words. That is, grouping together words that simply sound similar would likely lead to misgroupings: based on speech sounds alone “dog” and “doggy” should be as likely to be versions of the same word as “red” and “ready”, yet the latter words are unrelated. Infants may be able to tell that wordplay wordforms carry overlapping meaning with conventional forms because in addition to the sound and meaning overlap, the contexts of usage for both wordforms are very similar, both in the world and in language. These co-occurrence statistics are a rich source of semantic knowledge (Griffiths, Steyvers, & Tenenbaum, 2007; Huebner & Willits, 2018; Unger & Fisher, 2021), and may help infants bundle related wordforms. While “red” and “ready” are both adjectives, they are unlikely to occur in the same semantic or syntactic contexts. In contrast, felicitous contexts for “dog” and “doggy” (especially in speech that young children hear) are essentially identical. Learners are likely to hear about “dogs” and “doggies” both in the presence of dogs, potentially helping infants to map both wordforms onto the same concept. Further, the content words surrounding “dog” and “doggy” are likely to be the similar (e.g. “cat”, “bone”, etc.) The surrounding words have been shown to be integral to building word meanings in semantic networks used in natural language processing (e.g Mikolov, Chen, Corrado, & Dean, 2013). Since “dogs” and “doggies” can be used in free variation in infant-directed linguistic contexts, infants may be able to use that information to discern that they are variants of the same word. While out of scope for the current paper, future work could fruitfully examine semantic network analysis across wordplay forms to see whether indeed semantic networks for e.g. “dog” and “doggy” show high overlap as expected.

Another potential cue to semantic relatedness for infants - beyond distributional overlap - is temporal bunching. A future direction of this work would be to examine how closely in time two wordforms occur to each other, as this temporal clustering may also

serve as a cue that the wordforms refer to the same lemma. Previous work suggests that different wordforms of the same lemma occur closer together in time than unrelated minimal pairs, potentially providing a cue to the listener that the variant wordforms stem from the same lemma (Peperkamp, Hegde, & Carbajal, 2019). For example, in speech to a French child, words like “tricycle” appeared in two different surface forms (both with and without the final /l/, roughly /trisik/ vs. /trisikl/). Peperkamp et al. (2019) found that both of these forms appeared in the corpus an average of only 12 utterances apart from each other. In contrast, two words that were minimal pairs differentiated by the same size of difference (i.e. with and without the final consonant in a cluster) appeared on average 39 utterances apart. This distribution of wordform variants compared to minimal pairs suggests that temporal clustering may help children understand which wordforms are variants of each other. In sum, different wordforms of the same lemma share syntactic, semantic, and temporal distributions; by hypothesis (and hopefully future empirical work), similar wordforms of different lemmas seem less likely to do so.

Low frequency phenomena are worth considering

We find that wordplay makes up roughly 10% of concrete nouns infants hear in everyday interactions with their caretakers. One may wonder whether such a relatively small-scale phenomenon is worth taking seriously as a possible contributor to word learning over infancy. We argue that it is, given that (a) wordplay contributes to overall wordform variability, which in turn accounts for variance in reported receptive and productive vocabulary even in our relatively small and homogeneous sample, as elaborated above; (b) other relatively low-frequency linguistic phenomena too have been found to link to learnability in prior work, and (c) wordplay adds to notions of variability currently not captured in theories of language development. To give a concrete example of (b), the prevalence of words in isolation in everyday speech to children (e.g. “Ball!”), for instance, has been shown to boost noun learning, yet makes up only 9% of maternal utterances on

average (Brent & Siskind, 2001; Swingley & Humphrey, 2018). Regarding (c), wordplay is emblematic of a variety of phenomena in everyday language input (like polysemy) that are generally swept under the rug in favor of an idealized, readily segmentable, perfectly produced unambiguous lemma that children are delivered while learning words. This simply isn't a fair characterization of the input or the messy learning problem children solve. Wordplay words are heterogeneous, with many different possible wordforms per lemma. Many individual wordplay instances occur only once or a few times in our corpus; others many dozens. Despite this, infants sift through wordplay instances to learn the conventional wordforms that are used in adult-directed speech. That is, wordplay is a component of infants' early experience that theories of learning need to account for.

This final point ties into wordplay's place in theories that posit a one-to-one mapping between words and their meanings, like the principle of contrast (Clark, 1987). A strict interpretation of the principle of contrast may seem to clash with our findings; many of the words infants hear have multiple wordforms, unaccompanied by changes in meaning, or are all said (often to the same child) within an infant-directed register or context in our recordings (e.g. "dog", "doggy-woggy", and "doggo"). However, our wordplay results are not at odds with existing theory when we take its ephemeral nature seriously. While some of the processes that form wordplay persist across instances with some transformations appearing often, any individual wordplay word is rare when compared to the number of conventional forms children also hear. Thus, instead of being thought of as separate words that children need to learn and add to the lexicon, wordplay instances can be thought of as creative extensions of the conventional form, arising from rule-governed morphological combinations, just like other wordform variability like plurals. In wordplay instances however, the meanings of the morphemes are not distinctive. Wordplay thus seems to highlight the possibility that a language might maintain the use of a heterogeneous set of transformations that do not serve distinctive semantic or affective purposes, at least in input to very young children (as we discuss below, such forms seem to largely leave the

input by later toddlerhood). Our findings thus create an interesting potential caveat for a strictly interpreted principle of contrast.

Diminutives, reduplicated forms, and shortened words: Popular for a reason

Why might wordplay morphemes may be maintained, if not for their semantic or register-specific value? In our corpus, wordplay consisted of mostly reduplication as in “car-car”, addition of the *-/i/* morpheme as in “birdy” and “doggy”, and word shortening like in “cado” for “avocado”. All of these changes occurred in multiple words, suggesting that they were part of a set of productive wordplay operations. We speculate that these wordform operations may be particularly common because they provide infants with useful information about the sounds and edges of a word. For instance, reduplication gives listeners a second chance to hear the sounds in a wordform through repetition, and more information about word boundaries based on where the reduplicated segment repeats. Indeed, reduplication has been shown to facilitate word segmentation at 9 months (Ota & Skarabela, 2018) and novel word learning at 18.5 months (Ota & Skarabela, 2016). This learning bias is at odds with the statistics of adult-directed language, where repeated consonants are dispreferred (MacNeilage, Davis, Kinney, & Matyear, 2000). This is consistent with the ephemeral nature of wordplay forms in speech to young children.

Adding a diminutive suffix like *-/i/* regularizes the right edge of the wordform, potentially making it easier for infants to find word boundaries, especially at younger ages: previous work has shown that 5- and 8-month-olds fail to segment an artificial language composed of words of varying lengths, suggesting that transitional probabilities alone are not sufficient for word segmentation (Johnson & Tyler, 2010). Regularized word endings could provide an additional cue that infants could leverage for segmentation. This suggests that diminutivization, which regularizes the length and prosodic contour of words, may make it easier for infants to find word boundaries in the speech stream. Evidence from adults supports this: word ending invariance has been found to be associated with better

learning of nonce words extracted from speech in unfamiliar languages (Kempe, Brooks, & Gillis, 2005; Kempe, Brooks, Gillis, & Samson, 2007).

Word shortening may similarly increase the learnability of long words by creating wordforms that follow the strong-weak trochaic stress pattern of most English words. English-learning 8-month-olds have been shown to value this trochaic pattern over transitional probabilities as a cue to word boundaries (Johnson & Jusczyk, 2001), suggesting that adding an /i/ suffix or shortening a word into a CVCV trochee may make it easier for infants to segment wordplay wordforms. Shortening words also makes them easier to produce since there is less to articulate. Indeed, infants' first words are often composed of open CV syllables or CVCV bisyllables (Demuth, 1995), suggesting that shortening and /-i/ suffixes may make more complex words easier for infants to say themselves.

In infancy, higher rates of reduplication and /i/-epenthesis in speech to 9-month-olds have been found to predict a steeper slope of vocabulary growth between 9 and 21 months, despite similarly low base-rates in their data to what we find here (Ota et al., 2018), suggesting that these lexical features of infant-directed speech are beneficial to listeners. Further evidence to this effect, from Casey and Casillas (2022), shows that caregivers persist in using wordplay forms in a majority of instances until their children are ~28 months old (almost a year older than the infants in this corpus), then reduce their usage. This suggests that—with or without caretakers' awareness—wordplay may be most common in caretakers' speech when infants are more novice language users: its usage seems to lessen in toddlerhood as children become more adept at understanding words.

Conclusion

Wordplay, and more broadly wordform variability, are features of the speech environment that infants must learn from as they develop language. In the present work we take a first look at how much wordform variability infants hear, and what forms it takes.

We find that wordplay is present at low levels in every family in our corpus, is created through a heterogeneous collection of morphological processes, and is a stable feature of infant-available speech. When we look at the effects of wordform variability writ large, we find that it predicts earlier learning at the word level. Taken together, these findings suggest that, rather than posing a problem for word learning, wordform variability may instead may be a helpful and fundamental part of the early word learning process.

References

- Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, 42(2), 239–273. <https://doi.org/10.1017/S030500091400049X>
- Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2018). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, 0(0), e12715. <https://doi.org/10.1111/desc.12715>
- Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81(2), B33–44.
- Casey, K., & Casillas, M. (2022). From doggy to dog: Developmental shifts in children’s use of register-specific words. *Proceedings of the Annual Conference of the Cognitive Science Society, Cognitive Science Society*, 8.
- Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children’s lexicons. *Journal of Child Language*, 17(1), 205–215.
- Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In *Mechanisms of language aquisition* (pp. 1–33). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Clark, E. V., & Grossman, J. (1998). Pragmatic directions and children’s word learning. *Journal of Child Language*, 25, 1–18. <https://doi.org/10.1017/S0305000997003309>
- Dautriche, I., Fibla, L., Fievet, A.-C., & Christophe, A. (2018). Learning homophones in context: Easy cases are favored in the lexicon of natural languages. *Cognitive*

- Psychology*, 104, 83–105. <https://doi.org/10.1016/j.cogpsych.2018.04.001>
- Demuth, K. (1995). *Markedness and the Development of Prosodic Structure*. 25, 15.
- Diesendruck, G., & Shatz, M. (2001). Two-year-olds' recognition of hierarchies: Evidence from their interpretation of the semantic relation between object labels. *Cognitive Development*, 16(1), 577–594. [https://doi.org/10.1016/S0885-2014\(01\)00047-8](https://doi.org/10.1016/S0885-2014(01)00047-8)
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5), 1-173; discussion 174-185.
- Ferguson, C. A. (1964). Baby Talk in Six Languages. *American Anthropologist*, 66(6_PART2), 103–114. https://doi.org/10.1525/aa.1964.66.suppl_3.02a00060
- Floyd, S., Goldberg, A. E., & Lew-Williams, C. (2020). Toddlers recognize multiple meanings of polysemous words. *Proceedings of the Annual Conference of the Cognitive Science Society, Cognitive Science Society*.
- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. (2021). *Variability and Consistency in Early Language Learning*. Cambridge, MA: MIT Press. Retrieved from <https://langcog.github.io/wordbank-book/>
- Gleason, J. B., Perlmann, R. Y., Ely, R., & Evans, D. W. (1994). The baby talk register: Parents' use of diminutives. *Handbook of Research in Language Using CHILDES*, 50–76.
- Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, 114, 211–244. <https://doi.org/10.1037/0033-295X.114.2.211>
- Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87(1), B23–34. [https://doi.org/10.1016/s0010-0277\(02\)00186-5](https://doi.org/10.1016/s0010-0277(02)00186-5)
- Huebner, P. A., & Willits, J. A. (2018). Structured Semantic Knowledge Can Emerge Automatically from Predicting Word Sequences in Child-Directed Speech. *Frontiers in Psychology*, 9.
- Johnson, E. K., & Jusczyk, P. W. (2001). Word Segmentation by 8-Month-Olds: When

- Speech Cues Count More Than Statistics. *Journal of Memory and Language*, 44(4), 548–567. <https://doi.org/10.1006/jmla.2000.2755>
- Johnson, E. K., & Tyler, M. D. (2010). Testing the Limits of Statistical Learning for Word Segmentation. *Developmental Science*, 13(2), 339–345. <https://doi.org/10.1111/j.1467-7687.2009.00886.x>
- Jusczyk, P. W. (1993). Some reflections on developmental changes in speech perception and production. *Journal of Phonetics*, 21, 109–116.
- Kempe, V., Brooks, P. J., & Gillis, S. (2005). Diminutives in child-directed speech supplement metric with distributional word segmentation cues. *Psychonomic Bulletin & Review*, 12(1), 145–151.
- Kempe, V., Brooks, P. J., Gillis, S., & Samson, G. (2007). Diminutives facilitate word segmentation in natural speech: Cross-linguistic evidence. *Memory & Cognition*, 35(4), 762–773. <https://doi.org/10.3758/BF03193313>
- Kempe, V., Brooks, P. J., Mironova, N., Pershukova, A., & Fedorova, O. (2007). Playing with word endings: Morphological variation in the learning of Russian noun inflections. *British Journal of Developmental Psychology*, 25(1), 55–77. <https://doi.org/10.1348/026151006X100675>
- Kipfer, B. A. (2011). *Roget's International Thesaurus, 7th Edition* (7th ed. edition). New York: Collins Reference.
- Laing, C. (2019). A role for onomatopoeia in early language: Evidence from phonological development. *Language and Cognition*, 11(2), 173–187. <https://doi.org/10.1017/langcog.2018.23>
- Lewis, M., Cristiano, V., Lake, B. M., Kwan, T., & Frank, M. C. (2020). The role of developmental change and linguistic experience in the mutual exclusivity effect. *Cognition*, 198, 104191. <https://doi.org/10.1016/j.cognition.2020.104191>
- Liittschwager, J. C., & Markman, E. M. (1994). Sixteen- and 24-month-olds' use of mutual exclusivity as a default assumption in second-label learning. *Developmental Psychology*,

- 30(6), 955–968. <https://doi.org/10.1037/0012-1649.30.6.955>
- MacNeilage, P. F., Davis, B. L., Kinney, A., & Matyear, C. L. (2000). The Motor Core of Speech: A Comparison of Serial Organization Patterns in Infants and Languages. *Child Development*, 71(1), 153–163. <https://doi.org/10.1111/1467-8624.00129>
- MacWhinney, B., & Wagner, J. (2010). Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository. *Gesprachsforschung : Online-Zeitschrift Zur Verbalen Interaktion*, 11, 154–173.
- Mani, N., & Plunkett, K. (2010). Twelve-Month-Olds Know Their Cups From Their Keps and Tups. *Infancy*, 15(5), 445–470. <https://doi.org/10.1111/j.1532-7078.2009.00027.x>
- Mazuka, R., Kondo, T., & Hayashi, A. (2008). Japanese Mothers' Use of Specialized Vocabulary in Infant-Directed Speech: Infant-Directed Vocabulary in Japanese. In N. Masataka (Ed.), *The Origins of Language: Unraveling Evolutionary Forces* (pp. 39–58). Tokyo: Springer Japan. https://doi.org/10.1007/978-4-431-79102-7_4
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv:1301.3781 [Cs]*. Retrieved from <https://arxiv.org/abs/1301.3781>
- Moore, C., & Bergelson, E. (2021). *Examining the roles of regularity and lexical class in 18–26-month-olds' representations of how words sound*. <https://doi.org/10.31219/osf.io/kp7tv>
- Moore, C., Dailey, S., Garrison, H., Amatuni, A., & Bergelson, E. (2019). Point, walk, talk: Links between three early milestones, from observation and parental report. *Developmental Psychology*.
- Ota, M., Davies-Jenkins, N., & Skarabela, B. (2018). Why Choo-Choo Is Better Than Train: The Role of Register-Specific Words in Early Vocabulary Growth. *Cognitive Science*, 42(6), 1974–1999. <https://doi.org/10.1111/cogs.12628>
- Ota, M., & Skarabela, B. (2016). Reduplicated Words Are Easier to Learn. *Language Learning and Development*, 12(4), 380–397.

<https://doi.org/10.1080/15475441.2016.1165100>

Ota, M., & Skarabela, B. (2018). Reduplication facilitates early word segmentation*.

Journal of Child Language, 45(1), 204–218.

<https://doi.org/10.1017/S0305000916000660>

Peperkamp, S., Hegde, M., & Carbajal, M. J. (2019). Liquid Deletion in French Child-Directed Speech. *Interspeech 2019*, 3574–3578. ISCA.

<https://doi.org/10.21437/Interspeech.2019-2838>

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122(3), 280–291.

<https://doi.org/10.1016/j.cognition.2011.10.004>

Quine, W. V. O. (1960). *Word and Object*. Cambridge, UNITED STATES: MIT Press.

Savickien, I., & Dressler, W. U. (2007). *The Acquisition of Diminutives: A cross-linguistic perspective*. John Benjamins Publishing.

Schneider, K. P. (2003). *Diminutives in English*. Walter de Gruyter.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, 27(4), 501–532.

<https://doi.org/10.1016/j.dr.2007.06.002>

Srinivasan, M., Berner, C., & Rabagliati, H. (2019). Children use polysemy to structure new word meanings. *Journal of Experimental Psychology: General*, 148(5), 926–942.

<https://doi.org/10.1037/xge0000454>

Swingley, D., & Humphrey, C. (2018). Quantitative linguistic predictors of infants' learning of specific English words. *Child Development*, 89(4), 1247–1267.

<https://doi.org/10.1111/cdev.12731>

Unger, L., & Fisher, A. V. (2021). The emergence of richly organized semantic knowledge from simple statistics: A synthetic review. *Developmental Review*, 60, 100949.

<https://doi.org/10.1016/j.dr.2021.100949>

Vihman, M. M., & McCune, L. (1994). When is a word a word?*. *Journal of Child*

Language, 21(3), 517–542. <https://doi.org/10.1017/S0305000900009442>

Von Holzen, K., & Bergmann, C. (2021). The Development of Infants' Responses to Mispronunciations: A Meta-Analysis. *Developmental Psychology*, 57(1), 18.

<https://doi.org/10.1037/dev0001141>

Zipf, G. K. (1945). The Meaning-Frequency Relationship of Words. *The Journal of General Psychology*, 33(2), 251–256. <https://doi.org/10.1080/00221309.1945.10544509>