

Variable Wordforms, Adaptable Learners: Evidence  
from Real-time Word Comprehension and  
Naturalistic Corpora

by

Charlotte Moore

Department of Psychology & Neuroscience  
Duke University

Date: \_\_\_\_\_

Approved:

---

Elika Bergelson, Advisor

---

Michael Tomasello

---

Elizabeth Marsh

---

Jeff Mielke

Dissertation submitted in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy in the Department of Psychology & Neuroscience  
in the Graduate School of Duke University  
2021

ABSTRACT

Variable Wordforms, Adaptable Learners: Evidence from  
Real-time Word Comprehension and Naturalistic Corpora

by

Charlotte Moore

Department of Psychology & Neuroscience  
Duke University

Date: \_\_\_\_\_

Approved:

\_\_\_\_\_  
Elika Bergelson, Advisor

\_\_\_\_\_  
Michael Tomasello

\_\_\_\_\_  
Elizabeth Marsh

\_\_\_\_\_  
Jeff Mielke

An abstract of a dissertation submitted in partial fulfillment of the requirements for  
the degree of Doctor of Philosophy in the Department of Psychology &  
Neuroscience  
in the Graduate School of Duke University  
2021

Copyright © 2021 by Charlotte Moore  
All rights reserved except the rights granted by the  
Creative Commons Attribution-Noncommercial Licence

# Abstract

When learning a language, typically-developing infants face the daunting task of learning both the sounds and the meanings of words. In this dissertation, we focus on a source of variability that complicates the one-to-one relationship between words and their meanings: wordform variability. In Chapter 1 we make a distinction between the micro timescale, where learning and acquisition can be observed during the comprehension of individual utterances, and the macro timescale, where a longitudinal lens can reveal insights about the speech that infants and toddlers can learn from. In Chapter 2, a corpus analysis confirms that infants hear more irregular verbs than irregular nouns. We then compare toddlers' phonological representations of irregular nouns and verbs to their regular counterparts in an eyetracking study. Toddlers demonstrate well-specified representations for verbs eight months later than they do for nouns, but the sounds in irregular words are represented with high fidelity at the same time as regular words in both syntactic categories. In Chapter 3, an eyetracking study with adults explores word comprehension when coarticulation cues have been manipulated. Adults take available referents into account when they hear a word whose coarticulation cues match an unfamiliar word. We also find that participants vary in whether they map an unfamiliar word to an available unfamiliar object if the unfamiliar word is sufficiently similar to a relevant known word. In Chapter 4, we return to early childhood, using annotated data from a longitudinal corpus of naturalistic recordings from the lives of 6-17-month-old infants.

In this corpus, we characterize and quantify wordform variability, and find that for high-frequency words, wordform variability may aid in word learning. Theoretical implications and next steps are discussed in Chapter 5. Taken together, this work suggests that throughout the lifespan, wordform variability is no problem for learners and can in some cases facilitate learning and comprehension.

To me a month ago when I didn't think I could do it. I did!

# Contents

Abstract	iv
List of Tables	xi
List of Figures	xii
List of Abbreviations and Symbols	xiii
Acknowledgements	xiv
<b>1 Introduction</b>	<b>1</b>
1.1 Word Comprehension on two scales . . . . .	2
1.1.1 Micro scale: insights from inside the lab . . . . .	3
1.1.2 Macro-scale: longitudinal and corpus methods . . . . .	8
1.2 Wordform Variability . . . . .	10
1.2.1 Infant-directed speech: a special case . . . . .	11
1.3 How wordform variability affects comprehension at both timescales .	12
<b>2 Examining the roles of regularity and lexical class in 18—26-month-olds’ representations of how words sound</b>	<b>15</b>
2.1 Introduction . . . . .	15
2.1.1 Lexical Class, Word-Form Representations, and Learnability .	17
2.1.2 Regularity in Nouns and Verbs . . . . .	19
2.1.3 Overview of the Present Work and its Hypotheses . . . . .	20
2.2 Corpus Analysis . . . . .	22

2.2.1	Method . . . . .	23
2.2.2	Results . . . . .	24
2.3	Experiment 1 . . . . .	26
2.3.1	Experiment 1 Methods . . . . .	27
2.3.2	Experiment 1 Results . . . . .	32
2.4	Experiment 1 Discussion . . . . .	35
2.5	Experiment 2a . . . . .	39
2.5.1	Experiment 2a Methods . . . . .	39
2.5.2	Experiment 2a Results . . . . .	44
2.6	Experiment 2a Discussion . . . . .	47
2.7	Experiment 2b . . . . .	49
2.7.1	Experiment 2b Methods . . . . .	49
2.7.2	Experiment 2b Results . . . . .	50
2.7.3	Experiment 2b Discussion . . . . .	52
2.8	Pooled Analysis Across Experiments 1 and 2 . . . . .	53
2.9	General Discussion . . . . .	54
2.9.1	Interlocking effects of age, regularity, word class, and wordform specificity . . . . .	56
2.9.2	Blocked versus degraded comprehension when words are mis- pronounced . . . . .	59
2.10	Conclusion . . . . .	60
<b>3</b>	<b>Listeners can use coarticulation cues to predict an upcoming novel word</b>	<b>62</b>
3.1	Introduction . . . . .	62
3.1.1	Mutual Exclusivity . . . . .	64
3.1.2	The Present Study . . . . .	65
3.2	Methods . . . . .	67

3.2.1	Participants . . . . .	67
3.2.2	Design . . . . .	67
3.2.3	Stimuli . . . . .	68
3.2.4	Procedure . . . . .	69
3.2.5	Data analysis plan . . . . .	69
3.3	Results . . . . .	71
3.3.1	Aim 1: Coarticulation effect . . . . .	71
3.3.2	Aim 2: Mutual Exclusivity . . . . .	73
3.3.3	Listener type to predict coarticulation processing . . . . .	75
3.4	Discussion . . . . .	76
3.5	Conclusion . . . . .	79
<b>4</b>	<b>Wordform variability in infants' language environment and its effects on early word learning</b>	<b>81</b>
4.1	Introduction . . . . .	81
4.2	Methods . . . . .	86
4.2.1	Participants . . . . .	86
4.2.2	Procedures . . . . .	87
4.2.3	Data Processing . . . . .	88
4.2.4	Noun Annotations . . . . .	88
4.2.5	Identifying wordplay in the corpus . . . . .	89
4.3	Results . . . . .	90
4.3.1	Wordplay and other changes . . . . .	91
4.3.2	Wordplay transformations . . . . .	94
4.3.3	Predicting Wordplay . . . . .	95
4.3.4	Wordplay does not vary as a function of age . . . . .	97
4.3.5	Wordplay across families . . . . .	99

4.3.6	Wordplay and learning . . . . .	102
4.4	Discussion . . . . .	104
4.4.1	Wordform variability boosts learning . . . . .	106
4.4.2	Diminutives, reduplicated forms, and shortened words: Popular for a reason . . . . .	107
4.5	Conclusion . . . . .	108
<b>5</b>	<b>Conclusion</b>	<b>109</b>
5.1	Summary of Studies . . . . .	109
5.2	How learners accommodate wordform variability . . . . .	112
5.3	Future Directions . . . . .	116
5.3.1	Limitations . . . . .	118
5.4	Conclusion . . . . .	121
<b>A</b>	<b>Appendix A</b>	<b>122</b>
A.1	Corpus Information . . . . .	122
A.2	Exploratory vocabulary analyses . . . . .	122
A.3	Exploratory motor analyses . . . . .	125
A.4	Item-level looking behavior . . . . .	127
	<b>Bibliography</b>	<b>129</b>
	<b>Biography</b>	<b>143</b>

# List of Tables

2.1	The top 25 concrete nouns and verbs from the North American portion of the CHILDES corpus . . . . .	24
2.3	Stimuli used in Experiment 1 . . . . .	30
2.4	Trial type-level means and statistical tests comparing to chance for Experiment 1 . . . . .	34
2.5	Experiment 2 stimuli and word knowledge rates. . . . .	42
2.6	Trial type-level means and statistical tests for Experiment 2 . . . . .	46
2.7	Pooled variance model regression output. . . . .	54
3.1	The 8 stimuli words included in the study by condition . . . . .	68
3.2	fixed effects of growth curve analysis . . . . .	71
4.1	Examples of the 8 most common types of wordplay . . . . .	95
4.2	Model estimates from logistic regression predicting wordplay prevalence	96
4.3	Fixed effects for mixed-effects model (comprehension) . . . . .	103
4.4	Fixed effects for mixed-effects model (production) . . . . .	104
A.1	Corpora included in the corpus analysis. . . . .	123

# List of Figures

2.1	Relative frequency of regular and irregular nouns and verbs in CHILDES	25
2.2	Proportion of target looking for each trial type in Experiments 1-2b	36
2.3	Item-level breakdowns of toddlers' responses across Experiments 1 and 2	37
2.4	Timecourse plots for Experiments 1-2b	38
2.5	Schematic of sample trial in Experiment 2	43
2.6	Plots exploring the significant interactions of the pooled model	55
3.1	Two example images presented as in a trial with the corresponding three auditory prompts	70
3.2	Looking over time to the unfamiliar object for both listener types	73
3.3	Histogram of unfamiliar label trial responses	74
3.4	Raw data in logit space and the predicted values from the growth curve model	76
4.1	The relative frequency of each lemma's change type	92
4.2	The relative proportions of each type of change	93
4.3	Visualizations of the significant interactions from the logistic regression	98
4.4	Violin plot showing the effect of infant age on wordplay frequency	99
4.5	Correlations between proportions and total input	101
4.6	Plots showing the predicted results of the mixed effects model	105
A.1	item-level values for target looking before and after target onset	128

# List of Abbreviations and Symbols

## Abbreviations

MCDI	MacArthur-Bates Communicative Development Inventory
WG	Words and Gestures
WS	Words and Sentences
ME	Mutual Exclusivity

# Acknowledgements

I would like to begin by thanking my committee. Erika, thank you for dedicating so much of your time, energy, and attention towards sculpting me into the scholar I have become. I would not be the person I am today without your guidance. I am eternally grateful that you chose me to be your first graduate student and that I got to develop my research interests with you. Further, thank you for ensuring that my professional life for the last five years has been filled with fun and camaraderie alongside intellectual rigor. I hope to be the mentor you were for me for my own students one day.

Mike, thank you for teaching me how to make a clear point, both with my research design and with my writing. Your mentorship throughout my degree clarified my thinking on countless occasions and expanded the topics I considered in my work.

Beth, thank you for your guidance and advocacy during my degree. Since first-year seminar, you have given me excellent advice at every step of my career. Your incisive questions and comments on my work have always improved the final draft.

Jeff, thank you for teaching me phonology in my first years of university. Your classes fostered my passion for language research. Thank you for teaching me so thoughtfully before I knew anything, and thank you for showing me areas of my work where I still don't know enough now. Your perspective on my research is invaluable and I'm grateful for it.

Many others were also instrumental in helping me complete this dissertation.

Thank you to my incomparable parents, Gary and Kathleen Moore, for always believing in me even when I didn't believe in myself, and for loving me no matter what. Thank you for getting excited about my professional milestones and for putting in the effort to understand my work. It means the world to me that I get to share this accomplishment with both of you.

I also feel tremendous gratitude to Brenda Yang, Paula Yust, Jane Tandler, Divya Subramanian, and Hannah Moshontz, who were an incredible support system throughout the dissertation-writing process. Thank you for the edits, the rallying cries, the surprise gifts, the pep talks, and the reality television show suggestions. I still can't believe that I lucked into the most delightful friend group on the planet.

Thank you also to my incredible lab mates who provided me with meaningful feedback and endless encouragement. It was an honour to build a community of practice with you. The collaborative spirit of the BLAB has been an enormous contributor to my happiness and success during my time at Duke.

I also acknowledge SSHRC for generously funding my work, and the Charles Lafitte foundation for enabling me to attend conferences that would have been out of my reach otherwise. I am a more well-rounded researcher because of these investments in my education.

Finally, I gratefully acknowledge the many research assistants who did necessary data cleaning and assisted with data collection, and the hundreds of families who participated in my research. Without their contributions, my work would be impossible.

# 1

## Introduction

When infants first hear their parents speak, they have very little knowledge about the language they are about to learn. Upon arriving into the world, typically-developing newborns are tasked with learning spoken language, a hierarchically organized code, from speech, a linear encoding system (Chomsky, 1959). Sounds make up the foundation of language, since it is the main signal that hearing infants receive. From these phonetic cues, infants build phonological representations of sounds (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Patricia K. Kuhl et al., 2006) associate meaning with those sounds (Bergelson & Swingley, 2012), come to use the syntactic structure of sentences to understand relationships between arguments (Strohner & Nelson, 1974), and develop high-level pragmatic reasoning about the meanings of utterances (Papafragou & Musolino, 2003), all in a few short years.

The timeline of phonemic perception is well-mapped from birth. Newborns begin with the ability to distinguish some phonemes from one another (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Between 6 and 12 months, infants narrow their perception of phonemes to include only contrasts relevant to native language comprehension (Kuhl et al., 2006; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). Word com-

prehension, where sound is attached to meaning, was originally thought to begin only once this phonemic refinement was well underway. However, the overlapping nature of language development has become clear in the last decades of study: infants learn the meanings of words at the same time as the sounds contained within them (Bergelson & Swingle, 2012). Similarly, toddlers use the syntax of sentences to learn the meaning of novel words (Naigles, 1990; Naigles, 1996), leveraging knowledge of a more abstract linguistic structure to help with word learning. Language learning is thus an interwoven process, with cascading gains between levels of abstraction. The work in this dissertation examines the seams between these different levels of analysis, focusing on the ways that higher levels of abstraction interact with speech comprehension to affect word learning and comprehension.

## 1.1 Word Comprehension on two scales

Infants quickly learn their first language without explicit instruction (Brown & Hanlon, 1970; Hirsh-Pasek, Treiman, & Schneiderman, 1984), suggesting that infants and toddlers come to the task of learning language with the necessary abilities to learn words. The literature has focused on discerning what those necessary abilities may be (Hollich, Hirsh-Pasek, & Golinkoff, 2000; Johnson, 2016), and the preliminary question of what exactly is available in the speech stream for infants to learn from (Soderstrom, 2007). These questions are fruitfully approached by examining familiar word comprehension from two complementary time scales. The first is the micro level, the act of understanding a single spoken word contained within an utterance. This occurs on the order of milliseconds, and the timeline of its development has been sketched out in-lab, particularly with eyetracking studies (Fernald, Zangl, Portillo, & Marchman, 2008).

The second approach to language comprehension is from a macro, longitudinal scale, answering the larger question of how one comes to build a lexicon (e.g. Bloom,

2001). At this scale, we look less at the moment-to-moment processes that lead a listener to recognize a referent, instead examining the types of information infants can leverage in their language environments, for example isolated words (Brent & Siskind, 2001; *cf* Swingley & Humphrey, 2018). At this scale, we can not only understand what young learners glean from the speech they hear, but also approximate the trajectory of typical vocabulary growth (Frank, Braginsky, Yurovsky, & Marchman, 2017). By examining word comprehension at both time scales, we develop a greater understanding of the mechanisms involved in acquiring language.

### *1.1.1 Micro scale: insights from inside the lab*

Observing the micro scale of word comprehension unveils some of the moment-to-moment processing that listeners must do upon hearing a spoken word in order to find its referent. By studying infants of varying ages, the field has been able to sketch out a timeline of when infants develop the necessary abilities to complete this task. These approaches use tightly-controlled stimuli, which enable researchers to assess early comprehension at a fine-grained level. Much of the experimental work in language acquisition relies on the use of gaze-based measures, like preferential looking paradigms (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Kemler Nelson et al., 1995), habituation (Cohen & Strauss, 1979), and eyetracking.

Eyetracking has a long history in the study of language comprehension across the lifespan. Once automated eyetracking became more widely available at the end of the 20th century, psycholinguists readily adopted the technology, as evidenced by the more than three thousand articles citing Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995). Their work was among the first to use eyetracking measures for psycholinguistic purposes (*c.f.* Cooper, 1974). With adults, eyetracking studies employ the Visual World paradigm (Allopenna, Magnuson, & Tanenhaus, 1998), a method that has provided many key insights for language development (see Section

1.1.1). Infancy research uses the Looking-While-Listening paradigm (Fernald, Zangl, Portillo, & Marchman, 2008). Both methods include a visual scene with a target item and one or more carefully-selected distractors. Participants hear a linguistic prompt while eye-gaze is measured. The linking hypothesis is that eye gaze is a cross-modal proxy for a listener’s attention, which is being directed by the scene in front of them as well as the speech they hear (Aslin, 2007; Huettig, Rommers, & Meyer, 2011).

The Looking-While-Listening paradigm has enabled the construction of a timeline of infants’ speech comprehension abilities. Infants begin to look more at named foods and body parts than at distractor images by 6-9 months of age (Bergelson & Swingley, 2012), suggesting that word comprehension begins early in development. At around 14 months, infants begin to show much more robust comprehension for familiar words, spending a higher proportion of each trial fixated on the named object (Bergelson, 2020).

By 18 months, toddlers have added some powerful word-learning strategies to their arsenal. Word learning constraints provide some heuristics for what constitutes a probable word referent (Markman, 1990). Without some limit on plausible meanings, infants would need to consider theoretically infinite options (Quine, 1960). The most important strategy to our current aims is the mutual exclusivity bias, which features extensively in Chapters 3 and 4. Mutual exclusivity refers to the phenomenon where at around 17 months, monolingual learners begin to prefer a one-to-one relationship between words and objects, and resist associating new novel words to objects whose labels are already known (Halberda, 2003). Mutual exclusivity is not always feasible as a learning constraint; for example, multilingual learners make less use of it due to having to learn a label in each language (Byers-Heinlein & Werker, 2009). However, a recent meta-analysis shows that mutual exclusivity acts as a robust constraint in experimental paradigms, increasing in strength as children develop (Lewis, Cristiano, Lake, Kwan, & Frank, 2020).

In order to use the mutual exclusivity bias, it helps for children to have a sense of what other speakers would reasonably be referring to. Listeners can therefore bolster their word learning by participating in what amounts to a social reasoning task. Social coordination skills are thus crucial to language development. Gaze following, a component of joint attention, is an early social coordination ability where infants direct their visual attention towards an object that an adult is looking at. Rudimentary gaze following can be observed in newborns (Farroni, Massaccesi, Pividori, & Johnson, 2004) but begins more robustly around 3 months and in earnest around 6 months (Gredebäck, Fikke, & Melinder, 2010). 6-month-olds follow gaze only when communicative intent is established by the adult (Senju & Csibra, 2008). The tight relationship between social coordination and gaze following is borne out in gaze following's effects on word learning: infants who are better at following an adult's gaze at 10.5 months have larger vocabularies at 2.5 years, controlling for general visual attentiveness (Brooks & Meltzoff, 2005). The role of joint attention thus cannot be overlooked when considering word learning strategies.

Another skill in toddlers' arsenal is the ability to use the syntax of a sentence to constrain novel word meanings. This process, known as syntactic bootstrapping (Gleitman, 1990; Landau & Gleitman, 1985), helps toddlers learn new words by using the syntactic structure of the sentence as a constraint on the meaning of a novel word in that sentence. For example, 18-month-olds distinguish verbs from nouns using sentence structure: when French-learning toddlers hear the equivalent of "look, it's a bamoule" they map "bamoule" to an object, but when they hear "look, she's bamouling" they associate "bamoule" with an action (De Carvalho, He, Lidz, & Christophe, 2019). Further, two-year-olds can distinguish between two possible verb meanings based on the argument structure of the verb. When 25-month-olds see two characters doing both an intransitive action (spinning their arms) and a transitive action (the duck character pushed the bunny character), toddlers look more to the

transitive action when they hear “the duck is gorging the bunny” and more to the intransitive action when they hear “the duck and the bunny are gorging” (Naigles, 1990). These findings suggest that novel word learning is supported by the syntactic structure of the novel word’s carrier sentence.

While infants learn to associate sounds with meanings, they simultaneously develop their phonemic representations for words. Through the use of eyetracking studies, we understand the timeline of when infants and toddlers begin to represent the sounds in familiar words with a sufficient level of detail to identify mispronunciations. At the earliest stages of word comprehension, infants do not seem to notice when words are mispronounced; 6-9-month-olds look equally long to a named object when its label is mispronounced relative to when it is correctly pronounced (Bergelson & Swingley, 2018). This finding suggests that initial word representations are not furnished with the phonological detail required to notice mispronunciations. However, by 12 months infants detect mispronunciations, looking less to target objects when either a consonant or vowel in the label is changed (Nivedita Mani & Plunkett, 2010). This effect is robust throughout toddlerhood (Von Holzen & Bergmann, 2018). By 18 months, toddlers are sensitive to the phonetic size of the mispronunciation. For example, toddlers look less to a target *cup* when they hear “vup” compared to when they hear “tup.” This graded looking behavior suggests that 18-month-olds are sensitive to sub-phonemic differences between segments (White & Morgan, 2008), seen here in the fact that “vup” and “tup” are both single-phoneme mispronunciations of *cup*, but “vup” has a different initial consonant by three features: the place, manner, and voicing of /v/ all differ from /k/. Meanwhile “tup” differs only in place of articulation of the initial consonant. While previous research on phonological representation has focused primarily on regular concrete nouns, we expand these findings by examining verbs and irregular words in Chapter 2.

Throughout early childhood, toddlers continue to develop their perception of

sub-phonemic information available in spoken words. By the time children are two years old, they are sensitive to anticipatory coarticulation cues, where sounds from upcoming speech sounds affect the articulation of sounds earlier in the utterance (Fowler, 1981). Toddlers use coarticulation cues to find familiar words (Zamuner, Moore, & Desmeules-Trudel, 2016) and to predict upcoming words (Mahr, McMillan, Saffran, Ellis Weismer, & Edwards, 2015), suggesting that by two years old, toddlers are adept at using fine-grained phonetic detail during word comprehension. Adults' ability to make use of these same cues in a novel context is explored in Chapter 3.

*Adult language comprehension is highly integrated*

Spoken word comprehension research on the micro timescale has revealed that by adulthood, small changes at the phonological or phonetic level can have big effects on word comprehension (Huettig, Rommers, & Meyer, 2011). For instance, as adults perceive a word, other words in the lexicon that share the same onset or rhyme are also activated (Allopenna, Magnuson, & Tanenhaus, 1998), suggesting that word comprehension does not only rule out possible word candidates as speech progresses, but also that listeners continuously process matches between incoming speech and known words. Adult word comprehension is affected by a word's overall frequency as well as the commonness of the sound combinations in the word (Magnuson, Dixon, Tanenhaus, & Aslin, 2007). Furthermore, adults look to semantically-related distractors more often than to unrelated ones, with incremental changes in semantic relatedness corresponding to incremental changes in looking behavior (Mirman & Magnuson, 2009).

We also know that adults are experts at using the visual context together with linguistic cues to predict upcoming words. For example, when faced with a scene containing multiple objects, listeners who hear “the boy will eat the cake” look to a cake more rapidly than when they hear “the boy will move the cake” because the

meaning of the verb “eat” helps to constrain the possible meanings of future words (Altmann & Kamide, 1999). Taken together, these results suggest that the adult lexicon is complexly interconnected through syntactic, semantic, and phonological relationships. In the context of this dissertation, findings with adults are not only important for couching the results of Chapter 3, but also for clarifying the metaphorical “destination” of language development. Infants eventually become adults, and thus infant language comprehension evolves into adult language comprehension.

### *1.1.2 Macro-scale: longitudinal and corpus methods*

Setting aside the moment-to-moment processes that lead a listener to recognize a referent during speech comprehension, macro-scale methods allow for exploration of the types of information infants have available in their language environments, and how that leads to the gradual development of a child’s vocabulary. Two methods that have helped illuminate this longer time scale have been corpus analysis and parental questionnaires. Corpus analysis begins with recording an infant’s environment, a practice that was first used for linguistic research in the 1950s (Lynip, 1951). Captured data can then be used to understand infants’ own speech production, or the language environment that infants experience. The CHILDES database of transcriptions contains children’s speech and that of their caregivers from dozens of corpora (MacWhinney, 2000). Furthermore, with recent advances in large-scale audio recording, automated data processing, and data storage, it is becoming increasingly feasible to capture full days of data from many infants’ environments and extract useful metrics like the amount of adult speech contained in the file (VanDam et al., 2016).

Alongside these naturalistic recordings, vocabulary questionnaires have also acted as an integral tool for understanding early comprehension. The Macarthur-Bates Communicative Development Inventory (MCDI) is a vocabulary checklist that asks

parents about hundreds of individual words and whether their child has started saying or understanding that word yet (Fenson et al., 1994). It is one of the most widely-used questionnaires in early language research and is appropriate for typically-developing children between 6 and 30 months. Furthermore, it has been translated and validated for use in dozens of languages (Frank, Braginsky, Yurovsky, & Marchman, 2021). Thousands of administrations of the MCDI have been collected and digitized through Wordbank (Frank, Braginsky, Yurovsky, & Marchman, 2017). In gathering multiple corpora and questionnaires into their own centralized repositories with hundreds or thousands of individuals represented, language researchers can make generalizations about the acquisition of single words, or about the frequency of those words in speech to infants, along with countless other research questions.

While the quality and quantity of input in the language environment varies widely between individual children, many aspects of that environment appear to remain stable across cultures. Between individual children, both the quantity and quality of input vary widely (Huttenlocher, Vasilyeva, Waterfall, Vevea, & Hedges, 2007). However, many aspects of the language environment appear to remain stable cross-culturally. For example, Bunce et al. (2020) show that the majority of speech input comes from women across five disparate cultures. Their study also corroborated Huttenlocher and colleagues' (2007) findings cross-culturally: the average amount of speech directed to children across cultures was fairly stable, with significant variability between individual children in how much speech they heard.

Much of the variability we observe in language development is at the individual level. Variability in both the amount and the diversity of speech that children hear robustly maps onto variability in both receptive and expressive vocabulary development, with the effects detectable longitudinally from infancy and into preschool (Anderson, Graham, Prime, Jenkins, & Madigan, 2021; Rowe, 2012; Weisleder & Fernald, 2013), highlighting the tight relationship between infants' input and their

vocabulary development over time.

If we instead look at data in aggregate, obscuring individual differences, strikingly stable patterns appear regarding the order that words are learned. There is even cross-linguistic consistency across the earliest words that infants begin to say. For example, “mommy,” “daddy,” “baby,” “yes,” “no,” and “bye” all appear in the first ten words acquired by infants in at least six language versions of the MCDI (Frank, Braginsky, Yurovsky, & Marchman, 2021). Shifting to comprehension, Frank and colleagues show that many of the earliest-understood words are nouns, with predicates and function words being acquired later, despite being higher-frequency word classes in the input (Goodman, Dale, & Li, 2008). Synthesizing these results, we find patterns of both variability and stability emerge when examining word-learning with a longitudinal lens, whether we examine comprehension or production.

Through both the micro- and macro-scale views on language development, some basic truths have revealed themselves. The first is that as learners get older, they become increasingly sophisticated word learners and comprehenders. The second is that while language input varies between individuals in its quantity and quality, learning progresses in a similar order for most infants. So far this survey of the literature has looked at word learning without explicitly making a distinction between the form of the word and its meaning. We explore this under-studied facet of word acquisition and comprehension below.

## 1.2 Wordform Variability

Consider the difference between “dog” and “doggy.” Here we need to distinguish between the phonetic wordform that a word takes as it is articulated and travels through the air to be heard, and the conceptual lemma that a word refers to in the mental lexicon. Wordform variability, the focus of this dissertation, refers to cases where the same lemma can be denoted by different wordforms. Even when the same

lemma is being talked about across multiple instances, the sounds in the wordform can change. For example, when infants hear “she ate an oat this morning, and she’ll eat another one tomorrow,” the two instances of the verb “eat” have different vowels – which change the tense of the verb – even though they refer to the same action: eating. Infants need to map both of these verbs onto the same concept, even though the sounds in the two words differ. Furthermore, infants need to learn that “eat” and “ate” are instances of the same concept but that “oat” has another meaning entirely, despite being as phonologically similar.

Inflection from morpho-syntactic processes like verb tense and noun pluralization is a major source of wordform variability. We tackle its effect on word comprehension directly in Chapter 2, asking if toddlers represent the vowels in more variable words with less specificity. Other sources of variability include indexical characteristics of the speaker like voice pitch (Atkinson, 1976), the speed at which the speaker is talking (Lindblom, 1990), the speech register that the speaker selects to convey the utterance, and the accent of the speaker (Magen, 1998). Furthermore, the sound context of a word has effects on its realization, like the phonetic context of the word (Fowler, 1981), and allophonic variants where the same phonemes are realized in different ways (Lieberman & Studdert-Kennedy, 1978).

### *1.2.1 Infant-directed speech: a special case*

The variability already mentioned is pervasive throughout adult-directed speech. However, to understand how infants build a lexicon, we need to focus on the particular type of speech that infants hear. Infant-directed speech is preferred by infants cross-culturally (Manybabies Consortium, 2020), potentially due to its unique prosodic properties relative to adult-directed speech. A slower speech rate, a higher mean pitch, and larger pitch changes than in adult-directed speech can be found cross-linguistically throughout infancy, even in genetically distinct languages (Fer-

nald et al., 1989; Narayan & McDermott, 2016). Beyond more dramatic prosodic contours, infant-directed speech also includes more variable phoneme realization. Some evidence suggests that in addition to its attention-grabbing qualities, infant-directed speech may be hyper-articulated to facilitate phonological learning (Kuhl et al., 1997). However, contrasting findings suggest that the vowels in infant-directed speech are made more variably (Miyazawa, Shinya, Martin, Kikuchi, & Mazuka, 2017) and that infant-directed syllables are less discriminable (Martin et al., 2015). Those properties increase the variability present in speech, on both the acoustic and lexical level (Guevara-Rukoz et al., 2018).

### 1.3 How wordform variability affects comprehension at both timescales

Thus, research converges on the fact that wordform variability exists, and may be even more prevalent in infant-directed speech. Learners are tasked with overcoming that variability in order to correctly identify wordforms and build a lexicon. One possible option that learners have available to them when faced with variability is to attempt to abstract away the invariable components of speech, building an abstract, idealized prototype of the word or sound (Kuhl, 1991; Oden & Massaro, 1978). The advantage of such a strategy would be that the remaining material would be simpler to categorize and process. However, potentially useful information from the speech stream would be lost. Furthermore, since wordforms are *so* variable in their realization, it is likely that in many cases nothing would be left once all variance was abstracted (Kleinschmidt & Jaeger, 2015; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

Another, more tractable option involves preserving representations of variability in speech to inform predictions about upcoming information and/or to use as redundant information to facilitate comprehension (Pierrehumbert, 2001). While the advantages to this approach include increased predictive power and higher confi-

dence in one’s perception, keeping a trace of all instances of variability requires more memory and attention. This dissertation examines three instances of wordform variability in detail, and provides explanations for how listeners use wordform variability to facilitate perception.

Each of the following chapters focuses on a specific source of wordform variability. In Chapter 2, we examine wordform variability from both timescales. We ask whether irregular nouns like “goose” are represented with less specificity than regular nouns like “whale” due to the increased wordform variability that occurs when irregulars are pluralized. We also explore phonological representations in verbs, a syntactic category with higher rates of irregular phonology. We begin by conducting a corpus analysis to quantify irregularity in nouns and verbs in input to infants. Next, we present the results of three eyetracking experiments that examine the effect of wordform variability. We compare toddlers’ sound representations for nouns and verbs that are irregular (i.e. that mark changes in number or tense with vowel changes instead of with adding a morpheme, e.g. *goose* ~ *geese* and *run* ~ *ran*) against their representations for regular nouns and verbs (e.g. *whale* ~ *whales* and *walk* ~ *walked*). These eyetracking studies examine this question at two ages, providing a longitudinal view of moment-to-moment word comprehension.

In Chapter 3, we look to adults on the micro timescale to understand how wordform variability interacts with a listener’s visual context. We ask whether listeners integrate the visual scene with sub-phonemic cues, and whether adults use mutual exclusivity when an unfamiliar word is highly similar to a salient known word. To answer these questions, we manipulate the coarticulation cues of a familiar word to cue a novel word, and provide a novel object as a possible referent. This design enables us to detect whether adults can integrate sub-phonemic changes in a wordform with the possible referents available to them. At the same time, this experiment tests where listeners place the boundary between a variant version of a familiar word

and an instance of a novel word. We do this by presenting adults with trials where participants are asked to find a novel object whose label is highly similar to a known word.

Finally, on the macro timescale, we examine wordform variability in a longitudinal corpus of speech to infants in Chapter 4. We ask how much wordform variability exists in speech to infants, what types of wordform variability are most prevalent, and what effects wordform variability has on learning at both the word and the family level. We exhaustively document instances of wordplay, where speakers choose unconventional wordforms to express lemmas in their infants' environments. In this chapter, we examine the effects of unconventional wordforms, as well as the relationship between wordform variability and learning in 6–18-months-olds.

Across these chapters, our overarching question is: Does wordform variability pose a challenge for learners at either the micro or macro timescale? In other words, do words with more wordform variability show different patterns of learning compared to words with less wordform variability?

We endeavour to answer that question in the chapters that follow.

## Examining the roles of regularity and lexical class in 18—26-month-olds’ representations of how words sound

### 2.1 Introduction

Before age one, infants have made progress on two key components of language acquisition: learning speech sounds and word meanings. Indeed, over the first year of life infants become increasingly good at discriminating their native-language vowels and consonants (Kuhl et al., 2006; Polka & Werker, 1994; Werker & Tees, 1984). At the same time, they begin to understand common nouns referring to familiar people, foods, and body parts (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999, 2012), with relatively robust comprehension in place just after the first birthday (Bergelson, 2020; Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998).

While these studies suggest early word comprehension is well underway in the first postnatal year, the nature of early word representations is still a topic of debate. One view holds that infants initially store words *holistically*, i.e. that phoneme-level or subsegmental representations only emerge to differentiate similar-sounding lexical

items (e.g. *hat* and *cat*) as vocabulary size increases (Charles-Luce & Luce, 1990, 1995; Storkel, 2002). On this account, it is only when a child learns similar-sounding words (i.e. “neighbors”) that fine-grained representations are necessary to keep lexical items distinct. Hallé and de Boysson-Bardies (1996) provide data supporting this kind of underspecification in 11-month-olds. They found that infants listened equally long to mispronounced familiar words as they did to correctly-pronounced ones in a headturn preference paradigm.

A second view posits that infants have highly detailed representations of words they know early on, regardless of phonological neighborhood density (Gerken, Murphy, & Aslin, 1995; Swingley & Aslin, 2002). On this view, early wordforms are automatically high-fidelity. One line of support for this view comes from studies revealing that toddlers have sub-phonemic representations of the words they know (White & Morgan, 2008; Zamuner, Moore, & Desmeules-Trudel, 2016). A second line of work focusing on phonemic-level manipulations also supports this view, in what are usually dubbed “mispronunciation” studies.<sup>1</sup>

Mispronunciation studies are similar to standard looking-while-listening studies (Fernald, Zangl, Portillo, & Marchman, 2008). Namely, infants or toddlers see two images onscreen (e.g. a pig and a cat) and are prompted to look at one (e.g. “Look at the pig!”). Correctly-pronounced trials establish how well infants understand the words being tested. Mispronunciation trials instead alter the wordform (e.g. “pog” for *pig*) to see whether this change leads to a decrease in looking at the named target image. Such a decrease is interpreted as evidence that children know what the word typically sounds like, and is found for both consonant- and vowel-mispronunciations by around 12-14 months (Mani & Plunkett, 2007, 2010; Swingley & Aslin, 2002).

---

<sup>1</sup> In the current work we focus on studies of *familiar* word comprehension as opposed to *novel* word learning. While the latter is fascinating in its own right, the focus of word learning studies is often on factors relevant to initial exposure and learning, rather than the longer-term representation we tackle here.

Supporting this finding, a recent meta-analysis finds a reliable effect across 32 studies using this approach, confirming that infants and toddlers spend significantly less time looking at an object when its label is mispronounced relative to when it is correctly pronounced (Von Holzen & Bergmann, 2018).

However, previous work in this area has two key gaps. First, it has been almost exclusively focused on noun representations, neglecting other lexical classes. Second, it has solely focused on “well-behaved” words that do not undergo morphological processes that change how the root of the word sounds (i.e. *regular* nouns). In what follows, we first expound upon why these are important gaps to fill, and then begin to fill them through evidence from a corpus analysis and two experiments looking at regularity and wordform representations for nouns and verbs in 18-26 month olds.

### *2.1.1 Lexical Class, Word-Form Representations, and Learnability*

One reason that the prior literature on word-form representations has focused on nouns is that nouns are acquired relatively early. In fact, cross-linguistically, nouns are acquired earlier than verbs (and predicates more broadly) (Frank, Braginsky, Yurovsky, & Marchman, 2021), despite verbs actually appearing *more* frequently than nouns in infants’ environments (Goodman, Dale, & Li, 2008; *c.f.* Frank et al., 2021 for discussion of possible exceptions e.g. Beijing Mandarin). Thus, by looking at wordform representations in nouns, we stand to find some of the earliest evidence not just about whether infants *discriminate* speech sounds, but whether they have strong expectations for how words sound in the context of understanding them.

But why should nouns be understood more readily than verbs in the first place? One explanation for this general noun advantage is that early-learned nouns may be easier to learn than other open classes like verbs because they usually denote concrete, perceptually cohesive objects, while even simple action verbs denote relational, often fleeting events (Gentner, 1982, 2006). A related explanation for why

nouns precede other predicates in acquisition is that while early nouns (and some other very concrete words) have meanings that can be observed readily from the world, verbs are generally far more ambiguous. Their learning thus relies more heavily on sentence structure to do some of the heavy lifting, through what is referred to as syntactic bootstrapping (Fisher & Gleitman, 2002; Gleitman, 1990). For instance, a given scene in the world will rarely help a learner differentiate *leading* vs. *following* or *putting* vs. *placing*. Compatible with these views, the earliest evidence of comprehension for abstract words (including some verbs) is at 10-13 months, lagging several months behind comprehension for concrete nouns (Bergelson & Swingley, 2013).

In any event (pun intended), verbs being learned later than nouns may have trickle-down effects on toddlers' phonetic representations of early-learned verbs. While we tackle this possibility directly in the experiments below, two prior studies looking at lexical class and word-form representations obliquely address it, and merit mention. In a study focused on word segmentation (rather than comprehension), Shi and Lepage (2008) find that while correctly-pronounced frequent function words in French (e.g. *mes*, meaning "mine") allow 8-month-olds to segment novel nouns in a speech stream, mispronounced function words (e.g. *kes*) do not. This suggests that even in the first year, infants use their expectations for how words sound to guide them in parsing the speech stream.

Toddlers too seem to consider lexical class when building their wordform representations. While French-learning 18-month-olds have no trouble learning a new noun that has a verb neighbor, they fail to learn a new noun that has a noun neighbor (Dautriche, Swingley, & Christophe, 2015). More concretely, to use an English example, these toddlers had no trouble learning a new noun like "kiv" which has a verb neighbor *give* but failed to learn a new noun like "tog" which has a noun-neighbor *dog*. These results suggest that for 18-month-olds, new noun learning is influenced by existing wordforms within the same lexical class, but not wordforms in

other lexical classes.

Taken together, prior research thus highlights two key points regarding lexical class, wordform representations, and learnability. First, nouns are generally learned earlier than other lexical classes like verbs. Second, while research on wordform representation in other lexical classes is sparse, the evidence to date suggests that infants and toddlers are able to draw upon existing phonetic representations of how non-nouns sound, at least in the context of word segmentation or learning.

What about children’s early representations of the sounds in common verbs? Answering this question lets us examine the inter-relatedness of speech-sound discrimination and comprehension. By the time children understand common verbs, they have many more months of experience hearing speech than when they first understand common nouns. This in principle should facilitate not only continuing refinement of their phoneme inventories in general (e.g. Tsuji & Cristia, 2014; Liu, Tsao, Chang, & Hsu, 2013), but also more opportunities to encode fine phonetic detail about the common verbs they hear. Experiment 2 below tests whether the added exposure (and continuing development) that children experience by the time they understand verbs leads to well-specified phonetic representations that can be called upon during verb comprehension. However, verbs throw one further wrench in the works relative to nouns, one that no account of verb comprehension in English can reasonably ignore: a large proportion of them are irregular.

### *2.1.2 Regularity in Nouns and Verbs*

Cross-linguistically, languages vary in the degree to which a set of root sounds is altered when words undergo morphological processes. For instance, in Semitic languages like Arabic or Hebrew, a 3-consonant root (e.g. k-t-b) has different vowel patterns applied to it, creating related words with meanings like *write* (and its various conjugations), *writer*, *letter* and *book*.

In English on the other hand, nouns and verbs with vowel alternations are broadly characterized as “irregular”. Regular English nouns simply add morphemes during common processes like pluralization or compounding, leaving word-internal vowels intact (e.g. *apple* ~ *apples*; *truck* ~ *firetruck*). While irregular nouns exist (e.g. *mouse* ~ *mice*), they are unusual (see corpus analysis below). In contrast, a large proportion of common irregular verbs have vowel alternations to indicate tense (e.g. *drink* ~ *drank*, *run* ~ *ran*).

In the studies below we focus on nouns and verbs that vary in whether they exhibit vowel alternations, i.e. whether the nuclear vowel in the word changes to mark pluralization for nouns and tense changes for verbs. Hereafter we refer to non-alternating words as “regulars” and alternating words as “irregulars”, while noting that the full morphological system of English is far more complex than this operationalization captures.

How might regularity influence phonetic specificity? If young children store word-forms with high phonetic specificity, irregulars may provide a unique challenge, since learners must map words with different nuclear vowels onto a single concept (e.g. *mouse/mice*, *drink/drank*). In contrast, if they store them with low phonetic specificity, then irregulars may pose no more of a challenge than regulars, since the alternating vowels within these words may not be encoded in the first place. By examining mispronunciation effects in regular and irregular nouns and verbs, we can deepen our understanding of the roles of both lexical class and regularity on word comprehension and phonetic specificity.

### *2.1.3 Overview of the Present Work and its Hypotheses*

In what follows, we have three broad research questions, tackled within a corpus analysis, an experiment testing for a mispronunciation effect within regular and irregular nouns (Experiment 1), and regular and irregular verbs (Experiment 2).

**First, how common are regular and irregular nouns and verbs in child language input in English?** While intuitively verbs seem more commonly irregular in English than nouns, this intuition is worth quantifying. In particular, examining rates of regularity across nouns and verbs within speech to infants and toddlers will provide grounding for the subsequent experiments looking at lexical class and regularity that follow.

**Second, does regularity play a role in toddlers' phonetic representations of common words?** Here, three broad patterns are possible. A *regulars-over-irregulars* pattern would show a stronger mispronunciation effect for regular nouns and verbs that toddlers understand relative to irregular ones. This would suggest that irregular words have poorer wordform representations in children's lexicons than regular words do, perhaps due to the alternating vowels that occur across their surface forms. In contrast, an *irregulars-over-regulars* pattern would reveal a stronger mispronunciation effect for irregulars relative to regulars, suggesting that perhaps irregulars' surface-form variability *heightens* children's sensitivity to the component sounds of these words relative to regulars. A final possible pattern is that regulars and irregulars have equally robust representations across the lexicon, leading to no differences in mispronunciation effects as a function of regularity.

**Finally, we ask whether wordform specificity differs between nouns and verbs.** Here we anticipate one of two plausible patterns, based on the established mispronunciation effect for nouns in the literature (Von Holzen & Bergmann, 2018). First, we may find a *nouns-over-verbs* pattern wherein toddlers exhibit a stronger mispronunciation effect for nouns than verbs. This pattern would suggest that toddlers make class-wide distinctions between nouns and verbs when building phonological representations, and potentially that the overall regularity of the lexical class plays a role (see first question above). While in principle we may find the opposite pattern, *verbs-over-nouns*, where the verb mispronunciation effect is stronger

than the noun one, this seems less likely: verbs are understood later than nouns, and comprehension is a necessary foundation for a mispronunciation effect. A final pattern we find more likely is no difference in mispronunciation effect across nouns and verbs, suggesting lexical class does not exert a detectable effect on wordform representations during spoken word comprehension. Notably, while we characterize these possible patterns separately for regularity and lexical class, it is of course a distinct possibility that these two factors interact. We begin by characterizing regularity across lexical classes in a corpus analysis.

## 2.2 Corpus Analysis

English irregulars arise from a wide variety of historical and sociolinguistic forces (Anderwald, 2013; Miller, 2016), resulting in a heterogeneous assortment of past tense morphology. As noted above, the present analysis will include only the subset of irregulars with vowel alternations. This focus enables us to draw clearer parallels between our findings here and the effects of the vowel mispronunciations we test below.

In English, irregular morphology has been a focus of a large body of work exploring how children may learn the complicated and heterogeneous past tense system (see Bybee, 1995 for an overview). Indeed, the adult lexicon of English is rife with irregular morphology, which poses a challenge for young learners. As such, nouns and verbs are both over-regularized to similar degrees through early development (Marcus, 1995; Plunkett & Marchman, 1993). The equivalent rates of over-regularization suggest that toddlers use similar mechanisms to determine the correct morphology across word classes. Relatedly, models of verb acquisition have shown that the proportion of irregulars in the training data is an important factor in correct past tense usage (Plunkett & Marchman, 1991, 1993). Thus, it is important to determine the rates of irregularity in the language children hear. It is possible that young children

hear mostly regular forms, though previous work reports consistent irregular input to children over time (Marcus et al., 1992). In this section we examine just how frequently infants and toddlers hear irregular nouns and verbs in child-directed English. To be clear, this analysis in no way relies on children understanding tense or plurality (which can be challenging to measure), but simply seeks to characterize the presence of irregular nouns and verbs in children’s input.

### *2.2.1 Method*

We conducted an analysis of North American English corpora in the CHILDES database (MacWhinney, 2000) using the *childesr* package (Braginsky, Sanchez, & Yurovsky, 2019, version 2018.1). We analyzed all of the utterances spoken to children aged 3-36 months, excluding the child’s own utterances. This left us with 3,363,486 tokens from 745 unique speakers across 34 different corpora. Each corpus in the sample contributed a median of 40.50 transcripts to the sample. See Table S1 in the Supplemental Materials for details about all included corpora.

From this sample, we calculated the frequency of each noun and verb stem, where “stem” refers to the root morpheme of each word; this allowed us to calculate word frequency across inflections or other morphological transformations. We then took the 25 most frequent stems for concrete nouns or action verbs and classified those words as either “irregular” or “regular”.<sup>2</sup> Nouns were classified as irregular if the stem’s nuclear vowel alternates when the word is pluralized. All other nouns were classified as regular. Similarly, verbs were classified as irregular if they alternate at least their nuclear vowel when conjugated into past tense, but could also change other sounds. Verbs where none of the sounds in the stem changed when in past tense were classified as regular for analysis purposes.

---

<sup>2</sup> Concrete nouns and action verbs were operationalized as those that could readily be depicted in an image or video, as these are most relevant to the Experiments that follow. Light verbs were also removed from analysis as they are often used with non-action meanings.

Table 2.1: The top 25 concrete nouns (left) and action verbs (right) from the North American portion of the CHILDES corpus, ranked by frequency. Irregular words (i.e. words with vowel alternations in the stem when inflected) are indicated with an asterisk.

Ranking	Noun	Noun Frequency	Verb	Verb Frequency
1	baby	7,242	see*	23,225
2	book	6,903	put	18,056
3	car	4,708	say*	13,278
4	ball	3,989	look	8,962
5	mommy	3,645	eat*	6,044
6	house	3,555	play	4,424
7	boy	3,494	tell*	4,388
8	hand	3,003	sit*	4,358
9	box	2,994	find*	3,894
10	picture	2,799	read*	3,616
11	chair	2,581	try	2,938
12	girl	2,546	turn	2,935
13	water	2,482	help	2,220
14	toy	2,355	throw*	2,206
15	truck	2,307	show	2,165
16	daddy	2,286	bring*	2,141
17	bear	2,243	fall*	2,118
18	hat	2,235	thank	2,079
19	cookie	2,205	hear*	2,069
20	bed	2,188	open	1,854
21	dog	2,171	call	1,841
22	train	2,118	wait	1,814
23	foot*	2,073	hold*	1,738
24	bird	1,848	pick	1,726
25	horse	1,509	push	1,689

### 2.2.2 Results

All words included in the analysis can be found in Table 2.1. Figure 2.1 displays the proportion of the top 25 nouns and verbs by regularity.

#### *Nouns*

The top 25 concrete nouns in CHILDES made up 17% of the total noun tokens in the corpus (the most common nouns being abstract words like “thing” or “way”).

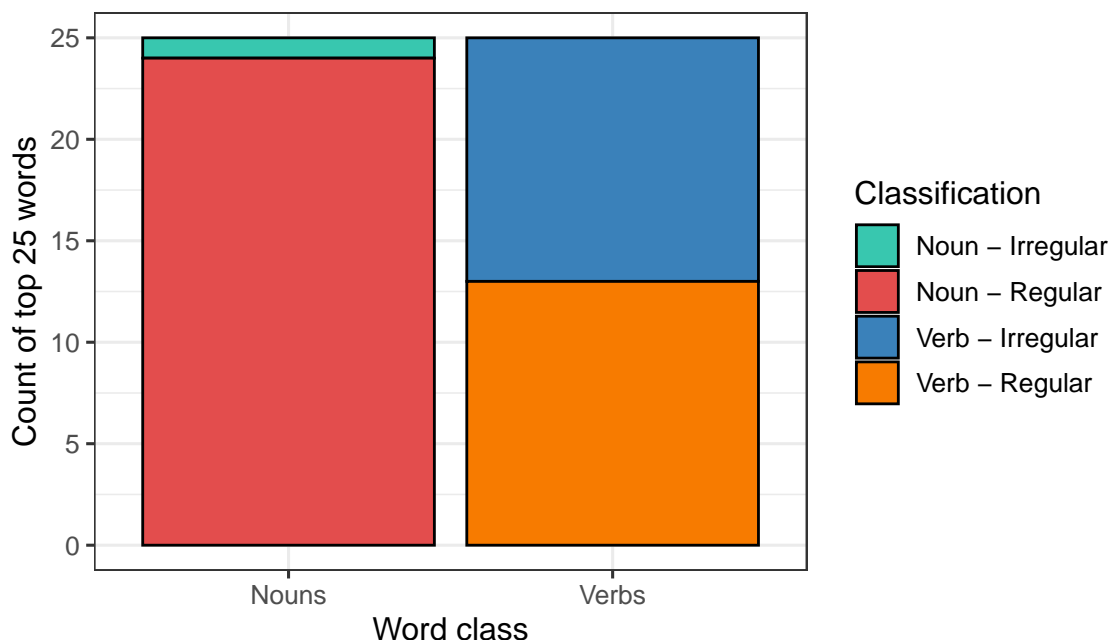


FIGURE 2.1: Relative quantity of regulars and irregulars in the top 25 most frequent concrete noun and action verb types heard by 3-36-month-olds in the North American portion of the CHILDES corpus. Color indicates regularity, with irregular depicted above regulars in each bar.

All but one of the top 25 nouns was classified as regular. The lone irregular noun was “foot”, which was ranked 23rd in frequency, and occurred in its plural form “feet” 51% of the time. Thus, based on these corpora, the vast majority of the most frequent concrete nouns young English-learning children hear are regular.

### *Verbs*

The most frequent 25 action verbs in the CHILDES corpus made up 28% of the verb tokens that children heard. In stark contrast to irregular nouns, 12/25 verbs, (48%) were irregular. To provide a more detailed understanding, we examined these verbs’ rates of occurrence in the past tense, the key context in which alternations occur. The past tense form occurred 17% of the time when any of these 25 verbs was spoken. The verb “fall” appeared in its irregular past tense form the most frequently

(58% of the time), while “hold” did so the least often: only 4% of the time. Thus, while individual verbs varied, nearly half of the most frequent action verbs were irregular, and children heard these verbs with different nuclear vowels relatively often in everyday language contexts. Children also had more data about the irregularity of some words (e.g. “fall”) than about others (e.g. “hold”) since participants heard many more past-tense instances of some words compared to others in our data.

Taken together, this corpus analysis provides strong evidence that common action verbs are more likely to be irregular than common concrete nouns in day-to-day language input to young children. While just one of the 25 top nouns is irregular, nearly half of the verbs are.<sup>3</sup> Setting aside the systemic meaning difference between nouns and verbs (roughly speaking, objects vs. actions), the prevalence of words with multiple surface forms alone distinguishes these classes, at least for the concrete word types young children hear most often. We now move to empirical studies testing potential consequences of this irregularity difference by examining toddlers’ phonological representations of regular and irregular nouns and verbs.

## 2.3 Experiment 1

In Experiment 1, we test whether toddlers show the same mispronunciation effect for irregular nouns as they have shown for regular nouns in previous work. We examine this by presenting 18-month-olds with correctly pronounced and mispronounced nouns and comparing their looking across pronunciation types. We do this for both regular and irregular nouns. Our preregistration can be found at <https://aspredicted.org/7qd4i.pdf>.

---

<sup>3</sup> In this analysis, we included only concrete nouns and action verbs, but if we expand to look at the 25 most common nouns and verbs of any type, the difference between word classes is wider; 0 of the nouns are irregular, while 60% of verbs are irregular.

### 2.3.1 Experiment 1 Methods

#### *Participants*

Our final sample included 29 typically-developing toddlers (10 girls) between 16 and 20 months old ( $M = 18.11$  months,  $SD = 1.26$  months).<sup>4</sup> An additional 8 toddlers participated but were excluded due to extreme fussiness ( $N = 1$ ), a parent-reported language delay ( $N = 1$ ), or not providing enough data based on the criteria described in our preregistration and in “Data Cleaning and Exclusion” below ( $N = 6$ ). All participants in the final sample were full term ( $40 \pm 3$  weeks), were exposed to only English (parents reported  $<25\%$  exposure to other languages), and were reported by parents to have typical hearing and vision. Participants were 79% White, 3% Black, 10% multiracial, and 7% unreported. Families were recruited to participate from the Research Triangle area of North Carolina through our participant database. Parents consented to participate on behalf of themselves and their toddlers, through a process approved by the Duke University IRB. For brevity, we refer to this group of 16–20-month-olds as 18-month-olds.

#### *Materials*

*Questionnaires* Parents were asked to complete the Words & Gestures version of the Macarthur-Bates Communicative Development Inventory (MCDI), a vocabulary checklist querying what words parents believe their children understand or say (Fenson et al., 1994), along with an optional demographics questionnaire. Finally, each parent completed a vocabulary exposure questionnaire to estimate how frequently their toddler heard each noun tested in the experiment. Parents responded on a 1-5 ordinal scale where 1 corresponded to “never” and 5 corresponded to “several times a day”.

---

<sup>4</sup> Our intended sample size was 32 but upon analyzing the data, fewer participants than anticipated met our pre-registered inclusion criteria. Additional participants could not be recruited due to the ongoing pandemic.

## *Eyetracking*

*Design* The eye-tracking study presented participants with pairs of images, one of which was named aloud, across 32 test trials (detailed below). The design manipulated two independent variables: noun type (regular vs. irregular) and pronunciation (correct vs. mispronounced). Irregular nouns were each yoked with a regular noun, resulting in 4 pairs, i.e. whenever a regular noun was the target, the same irregular noun was always the distractor and vice versa. Each pair occurred 8x, with each noun occurring as the target twice in each pronunciation. Trials were pseudo-randomized to ensure that the target never appeared on the same side of the screen >2x in a row, that targets of the same pronunciation type never occurred >2x in a row, and that the same word pair did not occur back-to-back. Trial order and target side were counterbalanced across participants. While mispronounced regular and irregular targets occurred 4x each on the left and right across trials as intended, a counterbalancing error resulted in each participant seeing correctly-pronounced irregular targets 5x on one side of the screen compared to 3x on the other side of the screen, and vice versa for correctly-pronounced regular targets. This error was counterbalanced, so half of participants saw the extra correctly pronounced irregular trial on the left and the other half saw it on the right.

*Items* Audio and visual test stimuli for all test items for Experiment 1 and 2 can be found at <https://osf.io/q9uud/>.

Due to their rarity relative to regular nouns, irregular nouns were selected first based on their imageability and familiarity to children. Regular nouns were then matched to each irregular noun first in animacy, then in frequency in the Brent-Siskind corpus (Brent & Siskind, 2001). See Table 2.3 for stimuli details.

*Audio Stimuli* The audio stimuli included 3 practice trial prompts (“Do you see the banana?”, “Can you find the apple?”, “Look at the cracker!”) and 16 test trial prompts (8 correctly pronounced, 8 mispronounced, see Table 2.3). Mispronounced stimuli were created by changing the nuclear vowel of each noun. We selected vowel mispronunciations that differed from the word’s correct vowel, any of its other vowel alternates (when applicable), and any words in children’s lexicons. For example, *mouse* was mispronounced as *mace*, which is a (low-frequency) word that is very unlikely to be known to toddlers.

Audio stimuli were recorded by a female native English speaker from the region as full sentences including each target word. They were recorded in a sound-attenuated booth and normalized so that participants would hear them at 60 dB.

*Visual Stimuli* In total, toddlers saw 3 warm-up images (banana, apple, and cracker; each displayed one at a time) and 8 target images (displayed in pairs; see Table 2.3 for list of stimuli). All images were composed of photographs or drawings of objects, which were standardized such that all objects appeared at roughly the same size on a 500x500 pixel gray background.

### *Procedure*

Before parents arrived, they were emailed the MCDI to complete. Upon arrival at the lab, parents and toddlers were taken to a waiting room where parents completed the consent form and remaining questionnaires. Then the procedure for the eyetracking component of the visit was explained and families were escorted to the eyetracking testing room.

Table 2.3: Stimuli used in Experiment 1. Noun Type indicates whether the noun was regular or irregular. MP is the orthographic representation of the mispronunciation. Carrier is the sentence used to introduce each target (for both items in each pair). WB Comp. shows the proportion of 18mos who are reported to understand each item based on Wordbank norms for the Words & Gestures version of the MCDI. E1 Comp. shows the proportion of our participants who understood each item by parental report on the MCDI. Mean Exp. shows the average exposure rating (out of 5) for each test item from the vocabulary exposure questionnaire (see Questionnaires).

Pair	Target	Noun Type	MP	Carrier	WB Comp.	E1 Comp.	M Exp.
A	Tooth	Irreg.	Tath	Look at the --	0.77	0.69	3.39
	Kitty	Reg.	Kotty		0.89	0.55	3.08
B	Mouse	Irreg.	Mace	Can you see the --	0.50	0.52	2.61
	Pig	Reg.	Pog		0.76	0.66	3.05
C	Foot	Irreg.	Foat	Find the --	0.93	0.76	4.37
	Dog	Reg.	Dag		0.96	0.90	4.58
D	Goose	Irreg.	Giss	Where's the --	0.33	0.38	2.24
	Whale	Reg.	Wile		NA	NA	2.16

In the testing room, toddlers and their caregivers were seated in front of a display screen (33.7×26.9 cm, 1280×1024 resolution) connected to an Eyelink 1000 Plus eyetracker in a dimly lit room with black curtains surrounding the eyetracker and participant chair to minimize distraction. Parents were asked to keep their eyes closed or wear a visor that blocked their ability to see the screen; compliance was monitored through a video feed. At the beginning of the experiment, a small target sticker was placed on participants' forehead or cheek to aid calibration. A five-point calibration was conducted at the beginning of the experiment and then as needed throughout the study. For example, toddlers occasionally removed the target sticker from their faces and then replaced the sticker elsewhere, which led to a recalibration. As needed, a 2s attention-getting video featuring shapes and harp music was played between trials to recapture the attention of distracted participants.

The experiment began with 3 practice trials followed by 32 test trials. Each of the 16 unique test trials was presented twice during the experiment. During practice trials, participants saw a single image in the center of the screen, while during test trials they saw two images side-by-side. On both practice and test trials, participants heard sentences prompting them to look at the target image. The target noun in each sentence occurred 2500ms after the trial began. Participants then had an additional 4000ms of looking time before the trial ended. In total, the eyetracking experiment lasted around 5 minutes.

After the eyetracking study, families were brought back to the waiting room for debriefing. Parents were also provided with the option of completing the MCDI while a research assistant watched their child(ren) if they had not completed the questionnaire prior to arrival. At the conclusion of the visit, families were compensated \$5 or \$10 for their travel to the lab (based on distance traveled) and toddlers received a book or other small gift as thanks for their participation.

### 2.3.2 Experiment 1 Results

#### *Questionnaire Responses*

All families in the final sample completed the demographics questionnaire, the MCDI vocabulary checklist and the noun exposure survey. On average, parents reported that their children understood 173 and produced 50 of the 396 words on the MCDI vocabulary checklist. Participants were reported to understand an average of 4.45 out of the 7 of our test nouns present on the MCDI (“Whale” is not on the MCDI), and say 2.07/7.<sup>5</sup> Parents reported that 64% of their toddlers understood the 7 test nouns on the MCDI (70% for regulars and 59% for irregulars.) Regarding noun exposure, parents estimated their toddlers heard our test words at least several times a week, on average ( $M = 3.18/5$ ,  $SD = 1.39$ ). See Table 2.3 for by-item values.

#### *Eyetracking Results*

*Eyetracking Analysis plan* Our dependent variable is the proportion of target looking during the target window (preregistered as 367ms to 3000ms after the target word onset, e.g. “mouse” in “Can you see the mouse?”).<sup>6</sup> This proportion was computed for each of the 16 audio targets (i.e. the correct and mispronounced version of each noun), for each participant. For each toddler, these proportions were then averaged to create four trial type means used for subsequent analysis (correctly pronounced regular nouns, correctly pronounced irregular nouns, mispronounced regular nouns, and mispronounced irregular nouns).

---

<sup>5</sup> While the Words & Gestures version of the MCDI is only designed to assess vocabulary up to 18 months, we chose to use it for our entire sample to simplify our analyses. None of our participants were at ceiling on the measure (the child with the largest vocabulary knew 331 words).

<sup>6</sup> Some studies, especially with younger infants, correct this proportion for pre-target baseline preferences. Following Swingley and Aslin (2002), Swingley and Aslin (2000) and others, we opted not to do so here, though we note that the reported pattern of results holds with a baseline-corrected version of the DV as well. For full data transparency, we direct interested readers to our Supplementary Materials where we provide figures depicting toddlers’ looking in the pre-target baseline for each item for all experiments.

We first compare participants' average target looking proportion in each of our four trial types to chance (which is 50% given the 2-picture display) using both t-tests and binomial tests, and provide descriptives of performance across infants and items. Finally, we report the results of a 2x2 factorial ANOVA testing our two independent variables (noun pronunciation and noun type) and pre-registered follow-up t-tests examining the effect of pronunciation type. We also pre-registered exploratory analyses examining the relationship between vocabulary size and mispronunciation sensitivity. These results were not particularly interesting or informative, and are thus reported in the Supplementary Materials.

*Data Cleaning and Exclusion* Following our preregistration, we excluded data from trials where toddlers did not look at the screen for at least 1/3 of our window of analysis (245 trials) and/or looked at only a single image for the entire trial (85 trials). After removing these trials (266 total, due to overlap across the criteria), we then removed participants if they failed to provide data for at least 16/32 trials. This led us to remove 6 participants.

See <https://osf.io/f6bdn/> for complete data processing code. After exclusions, the final sample contained 29 participants who provided usable data on an average of 26/32 trials.

*Subject- and Item-level comparisons to chance* Statistical details over trial types, subjects, and items are provided in Table 2.4 and summarized below. By Shapiro-Wilk test, each set of trial type means was normally-distributed (all  $ps \geq 0.23$ ). Using one-sample t-tests, we find that when participants heard a regular noun, target-looking was significantly above chance (i.e. >50%). This was true both when regular nouns were correctly pronounced and when they were mispronounced, and held for most toddlers in each case. For irregular nouns, target-looking was also significantly above

Table 2.4: Trial type-level means and statistical tests comparing to chance for Experiment 1. Rows depict data for each Pronunciation type (Pron., correctly pronounced (CP) or mispronounced (MP)) and Noun Type (regular (reg) or irregular (irreg)). M shows the mean proportion of target looking across participants for each of these 4 trial types. The next two columns ( $t(28)=$  and p-value) show the results of one-sample t-tests comparing this proportion to chance (.50). The next two columns ('Ss>0.5' and 'Binom p-value') show how many participants had a proportion of target looking above 0.5 in each trial type, and the corresponding p-value from a binomial test. The final column (Items>0.5) shows how many of the 4 items in each trial type were looked at above chance when they were the target.

Pron.	Noun Type	M	$t(28)=$	p-value	Ss>0.5	Binom p-value	Items>0.5
CP	reg	0.60	5.17	<0.001***	24/29	<0.001***	4/4
CP	irreg	0.55	2.40	0.023*	16/29	0.711	3/4
MP	reg	0.57	5.05	<0.001***	24/29	<0.001***	3/4
MP	irreg	0.46	-1.69	0.101	8/29	0.024*	2/4

chance when the nouns were correctly pronounced; this pattern again held for most infants, though fewer than for regular nouns. In contrast, when irregular nouns were mispronounced, participants did not look to the target at greater than chance rates, and moreover, a significant majority of infants (by binomial test) looked more at the distractor than the target for this trial type, suggesting the mispronunciation may have derailed word comprehension here. See Figure 2.2. The pattern over items was largely similar to that over participants: target looking exceeded 50% (i.e. chance) for all four correctly pronounced regular trials, 75% of mispronounced regular and correctly pronounced irregular trials, and 50% of mispronounced irregulars. See Table 2.4 and Figure 2.3 for details.

*Comparing across trial types* Our 2x2 factorial ANOVA looking at the effects of noun type and pronunciation revealed a main effect of pronunciation ( $F(1,112) = 8.24$ ,  $p = .005$ ). Participants looked significantly more at target nouns when their labels were pronounced correctly. We also found a main effect of noun type ( $F = 17.85$ ,  $p < .001$ ). Participants looked more at regular nouns when they were the target

than they did at irregular nouns. There was no interaction between noun type and pronunciation ( $F = 2.52$ ,  $p = 0.12$ ).

Planned follow-up t-tests showed that for regular nouns, toddlers did not significantly change their looking behavior when labels were mispronounced ( $t = 1.01$ ,  $p = 0.32$ ). However, for irregular nouns, they looked more at the target image upon hearing words pronounced correctly ( $t = 2.88$ ,  $p = .006$ ). For regular nouns, 17/29 participants showed a mispronunciation effect: they looked less at the target noun when it was mispronounced compared to when it was correctly pronounced. This held for 3/4 items. For irregular nouns, 22/29 participants showed a mispronunciation effect, which held for 4/4 nouns.

For data transparency, we also provide the timecourse of target looking across trial types in Figure 2.4, where the patterns across pronunciation and regularity found when we collapse across our window of analysis can be readily observed across time.

## 2.4 Experiment 1 Discussion

Across our analyses we find that both regularity and pronunciation played a role in the degree to which 18-month-olds looked at named images of irregular nouns and frequency-matched regular nouns. Our ANOVA provided evidence for a mispronunciation effect, and revealed slightly better performance for regular nouns vs. irregulars, regardless of pronunciation. Our follow-up tests and initial comparisons to chance indicated that toddlers understood both correctly pronounced regular and irregular nouns at above-chance rates across most of our analyses, again with stronger performance on correctly pronounced regulars over irregulars. For the mispronounced trials, we found that performance on regular nouns was only slightly worse (by 3%) when these nouns were mispronounced. In contrast, performance dropped more substantially (by 9%) when irregular nouns were mispronounced, relative to performance

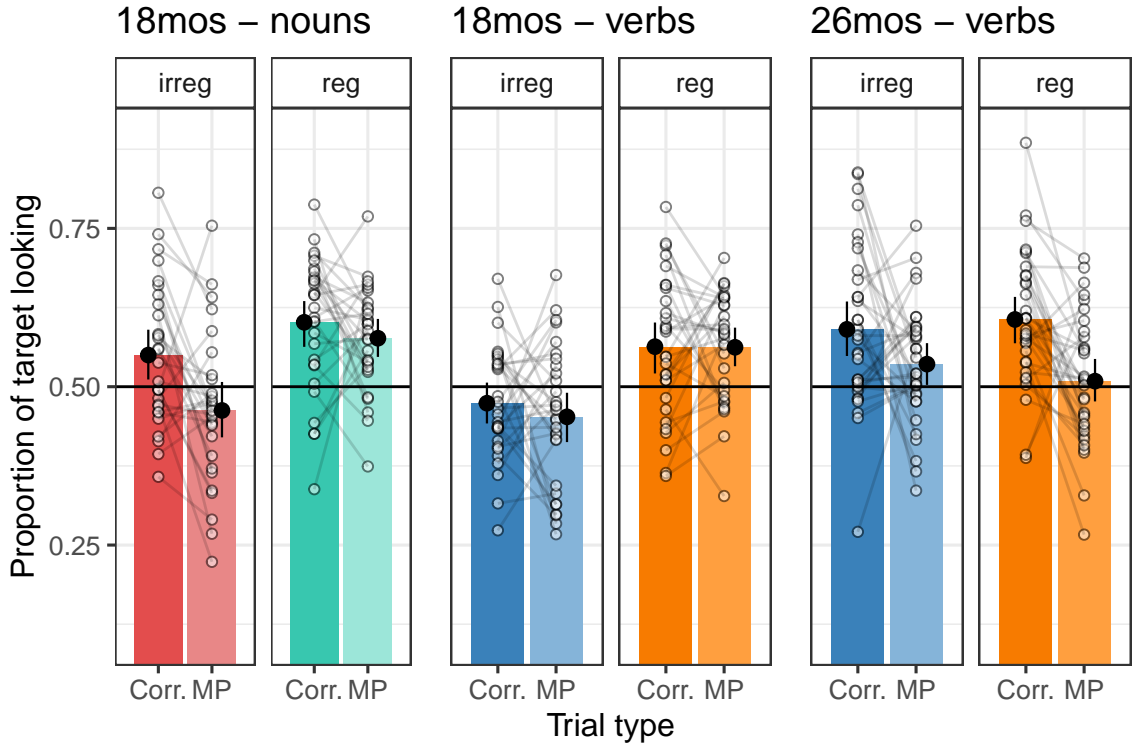


FIGURE 2.2: Proportion of target looking for each trial type in Experiment 1 (left), 2a (middle) and 2b (right). Chance is represented by a black line at 0.5. White dots represent each participant’s mean proportion of target looking for each trial type, with lines connecting participants’ values across pronunciation type; lines going down left to right indicate decreased target looking when words were mispronounced, and vice versa. Black point ranges represent the mean and 95% bootstrapped confidence intervals.

with their correctly pronounced counterparts.

Notably, the mispronunciation effect we find is smaller than others report at 18 months using more frequent regular nouns like e.g. “baby”, “car”, and “ball” used in Swingley and Aslin (2000). Since our items are lower-frequency than those in previous work, this suggests that frequency likely plays a role in the robustness of phonetic representations. The ideal pattern of results supporting a mispronunciation effect would be seeing that toddlers understand both correctly-pronounced words and mispronounced words at above chance rates, but show significantly better performance for correctly-pronounced words over mispronounced ones. This is not

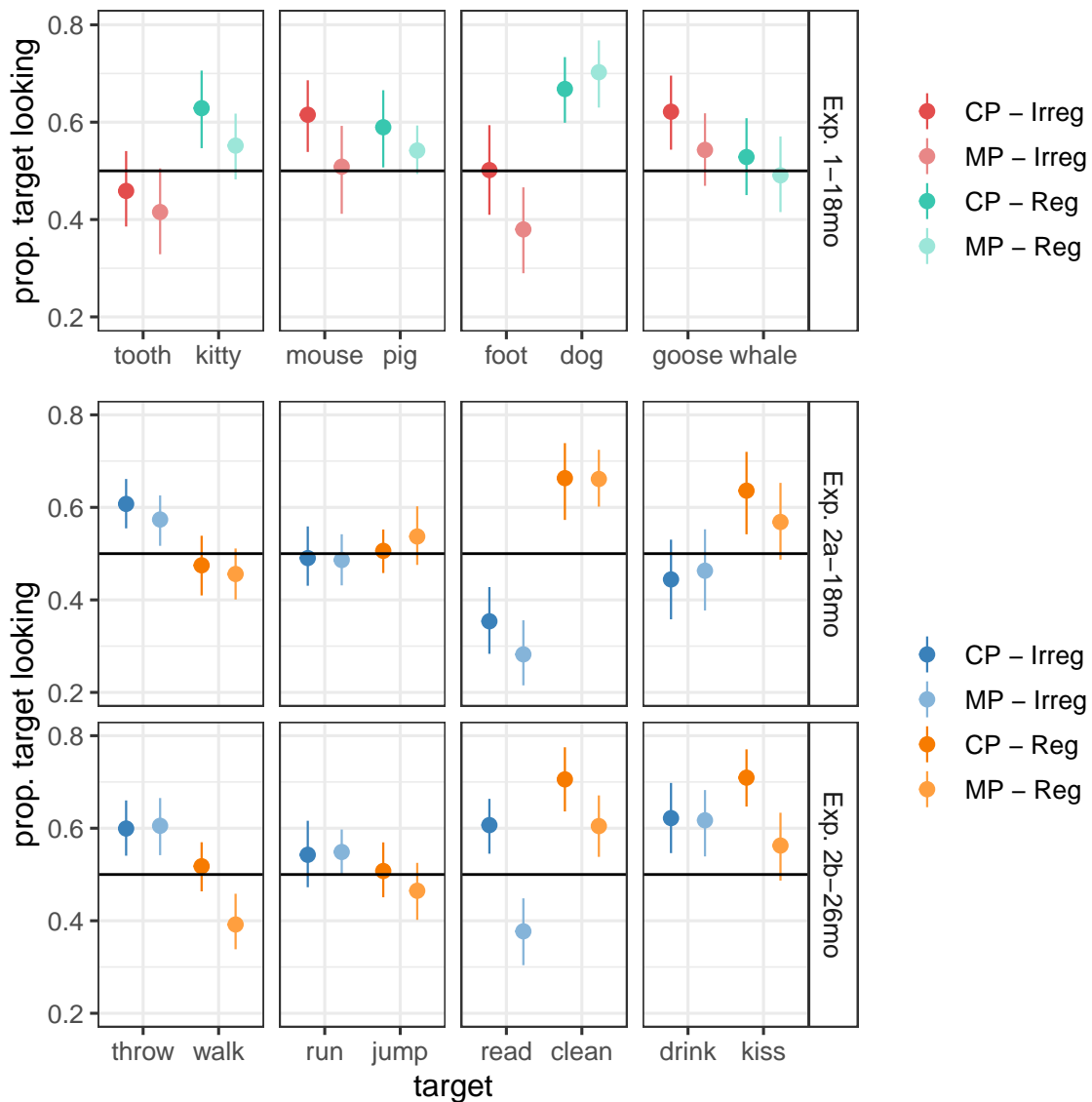


FIGURE 2.3: Item-level breakdowns of toddlers' responses across Experiments 1 and 2. Y-axis shows the proportion of time participants spent looking at the target after it was named. Colors indicate word type and pronunciation. Each panel shows one yoked pair. X-axis displays each specific item name.

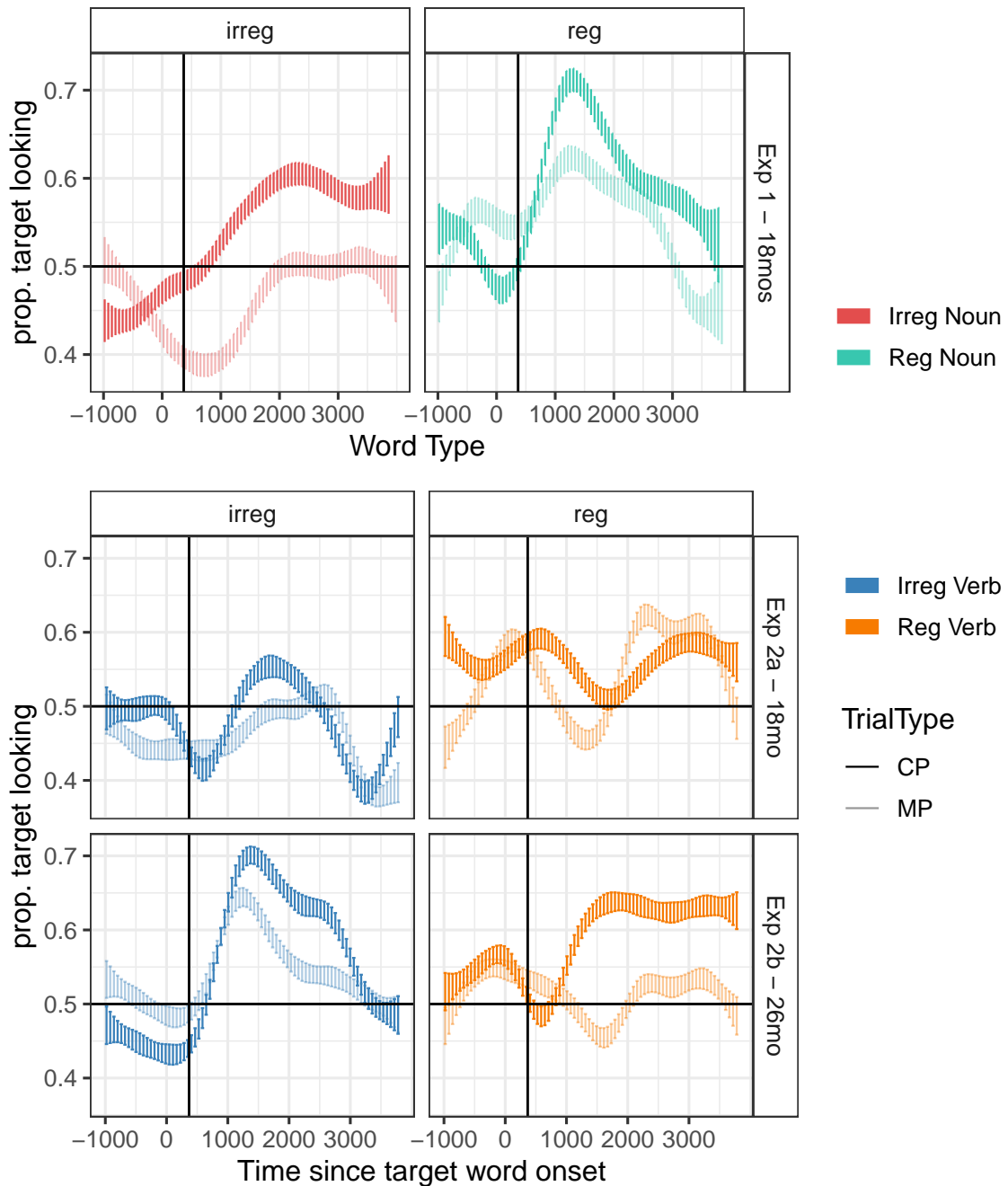


FIGURE 2.4: Timecourse plots for Experiment 1 (top) and Experiment 2 (bottom 4 panels). X-axis shows the time (milliseconds) since target word onset. Y-axis shows the mean proportion of looks to the target image at each timepoint. Each bar represents mean and 95% confidence interval for each 20ms bin, aggregated over participant and trial type.

what we see for either the regular or irregular nouns in this study, and we revisit this divergence in the General Discussion.

Finally, we note that the increased variability 18-month-olds experience in their input for (relatively rare) irregular nouns did not appear to hamper their ability to recognize which pronunciations were correct and which were not. If anything, the mispronunciation effect for irregulars was numerically stronger, providing some support for an *irregulars-over-regulars* pattern for mispronunciation effects. We return to this in the General Discussion as well. We next shift our focus to verbs, to determine whether regularity and phonetic representations show the same patterns on the same developmental timeline across word classes.

## 2.5 Experiment 2a

In Experiment 2a, we examine the same age group as in Experiment 1, but here we extend the mispronunciation paradigm to explore regular and irregular verbs. Testing a new group of 18-month-olds, we show participants videos of highly common action verbs and present them with correctly pronounced and mispronounced labels for the actions, then compare looking across pronunciation types.

### 2.5.1 Experiment 2a Methods

#### *Participants*

Participants in the final sample were  $N = 30$  typically developing toddlers (14 girls)<sup>7</sup> between 16—20 months old ( $M = 17.37$ ,  $SD = 1.33$ ) who met the same exclusion criteria as Experiment 1. An additional 5 participants were excluded due to fussiness (1), not contributing enough data (using the criteria as in Experiment 1 and detailed below) (3) or technical error (1). Participants were 80% White, 7% Black,

---

<sup>7</sup> As in Experiment 1, our goal sample size was 32 toddlers, but fewer participants than expected met our data-driven inclusion criteria, and unfortunately we could not run additional participants due to the ongoing pandemic.

7% multiracial, and 7% unreported. Participants were recruited and consented in the same process as in Experiment 1. One participant from Experiment 1 also participated in Experiment 2a. For brevity, we refer to this group of 16—20-month-olds as 18-month-olds.

### *Materials*

*Questionnaires* Questionnaires were the same as in Experiment 1, with the addition of a questionnaire used to assess early motor development (Libertus & Landa, 2013; Walle & Campos, 2014; questionnaire details can be found in Moore, Dailey, Garrison, Amatuni, & Bergelson, 2019). This motor questionnaire was included as an exploratory metric, to allow us to investigate the relationship between motor ability and comprehension for relevant verbs like “jump”.

### *Eyetracking*

*Design* Study design was identical to Experiment 1, except that the stimuli in Experiment 2 were videos of actions instead of images of objects, and trials included a more extended preview to familiarize children to the actions in the videos (see Procedure and Figure 2.5). Pseudo randomization and counterbalancing were the same as in Experiment 1 (including the counterbalancing error).

*Items* All test stimuli can be found at <https://osf.io/q9uvd/>. All verbs had one syllable, denoted actions, and were known by at least 63% of 18-month-olds, according to Wordbank norms for English-learning children (Frank, Braginsky, Yurovsky, & Marchman, 2017). See Table 2.5 for a summary of stimuli. In addition to the test and practice stimuli described below, display items also included four 3.5-second videos featuring a brightly colored shape (e.g. a star or heart) that changed color

and/or moved around the middle of the screen while instrumental music played on each trial (see Procedure).

*Audio Stimuli* The audio stimuli included 4 practice trial prompts (the same ones as in Experiment 1, plus “Where’s the cat?”) and 16 test trial prompts (8 correctly pronounced, 8 mispronounced; each was heard 2x). As with nouns in Experiment 1, mispronounced stimuli were created by changing the nuclear vowel of each verb to differ from the word’s correct vowel, any of its other vowel alternates (if applicable), and any words in children’s lexicons. All audio was recorded and normalized as in Experiment 1.

*Visual Stimuli* Toddlers saw 4 warm-up images (banana, apple, cracker, and cat) one at a time and 8 target videos in pairs. Videos faded in and out from black. All videos featured the same female actor. For each pair of videos, the actor wore a red shirt in one video and a blue shirt in the other. Both videos were filmed in the same setting, to control for environmental cues. All verbs occurred continuously during the video e.g. the “jump” video included continuous jumping for the duration of the video. The only exception was the video for “throw”, where the actor threw one ball only once during the video. See Table 2.5 for pairs. Each video was exactly 5 seconds long.

Table 2.5: Experiment 2 stimuli and word knowledge rates. WB Comp. and WB Prod. show the proportion of 18mos reported to understand and produce each word based on Wordbank norms of the MCDI (Frank et al., 2017) Words & Gestures norms are used for comprehension, Words & Sentences norms for production. 2a Comp. and 2b Prod. show the proportions of 16—20mos (2a) and 24—28mos (2b) whose parents reported they understood or produced each item, respectively.

Pair	Target	Type	MP	Carrier	WB Com.	2a Com.	WB Pro.	2b Pro.
A	Throw	Irreg.	Thraw	She’s gonna --	0.89	0.56	0.65	0.61
	Walk	Reg.	Wick		0.80	0.53	0.72	0.71
B	Run	Irreg.	Roan	Look! She can --	0.73	0.44	0.71	0.71
	Jump	Reg.	Joomp		0.63	0.28	0.73	0.71
C	Read	Irreg.	Rowd	Look! She can --	0.79	0.69	0.78	0.68
	Clean	Reg.	Cline		0.64	0.41	0.69	0.65
D	Drink	Irreg.	Droink	She’s about to -- it	0.93	0.69	0.83	0.77
	Kiss	Reg.	Koss		0.98	0.59	0.85	0.71

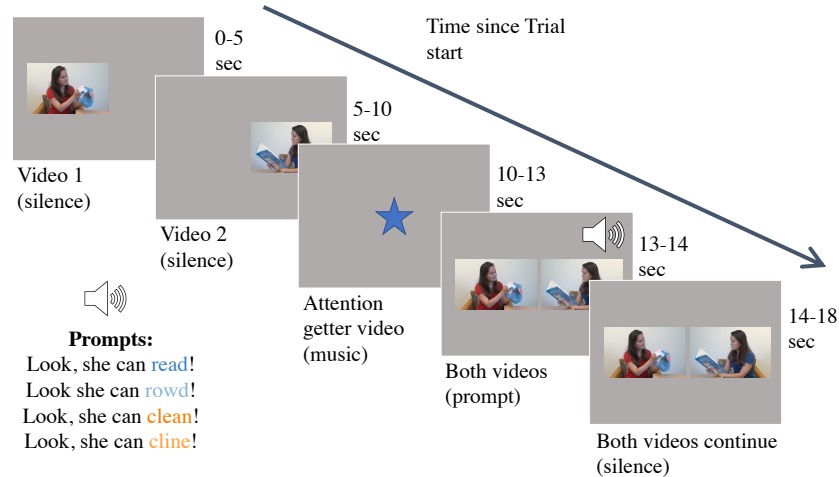


FIGURE 2.5: Schematic of a sample trial in Experiment 2. Time progresses from the top left to bottom right. Side of first video (left above) was counterbalanced across participants and trials.

### *Procedure*

Procedure was identical to Experiment 1, except for the experimental trials of the eyetracking experiment. Experimental trials were composed of two phases: familiarization, and test. In the familiarization phase, one 5-second video played on one side of the screen (counterbalanced across participants), then another video of the same woman doing a different action played on the other side of the screen. Between familiarization and test phases, a 3.5-second attention-getting video drew toddlers' attention back to the middle of the screen. During the test phase, both videos played simultaneously, in the same location they had played during familiarization. 1130ms into the test phase, the target word occurred in its carrier phrase (e.g. “run” or “roan” in “Look she can run/roan”). Looking behavior was recorded throughout the trial. The entire study took approximately 9 minutes. See Figure 2.5 for a schematic of a sample trial.

### 2.5.2 Experiment 2a Results

#### *Questionnaire responses*

In total, 29 families completed the vocabulary exposure measure, 26 families completed the demographics questionnaire, 27 families completed the MCDI, and 23 parents completed the motor questionnaire for their children within two weeks of their visit to the lab. On average, parents reported that their children understood 169.30/396 words and produced 48.85. Participants were reported to understand an average of 4.96/8 test verbs and produce 0.74. For the verb exposure questionnaire, the mean verb exposure score was 4.05/5 ( $SD = 1.17$ ). That is, parents estimated their children heard our test words at least once a day, on average. Motor survey results for both studies in Experiment 2 are presented in the Supplementary Materials.

#### *Eyetracking Results*

*Analysis plan* While not preregistered, we follow the same plan laid out in Experiment 1. The only exception to this was the analysis window, which began at 367ms after target word onset, as in Experiment 1, but was extended to the end of the trial (i.e. 3970ms after the target verb’s onset, since all videos were 5 seconds and target verbs occurred 1030ms into the video). Given the lack of prior mispronunciation studies with verbs from which to base a window empirically, using the maximal post-target window is a conservative approach. Further, Valteau, Konishi, Golinkoff, Hirsh-Pasek, and Arunachalam (2018) show that a longer analysis window is required to detect comprehension of dynamic verb scenes in toddlers of a similar age to our participants.<sup>8</sup> As in Experiment 1, we also tested for correlations between toddlers’

---

<sup>8</sup> As in Experiment 1, we report results from only this post-target window. However, we did notice that toddlers had a tendency to fixate the left side of the screen more before the target word was said. Fortunately, target side was counterbalanced across trials, rendering this idiosyncrasy largely irrelevant for the post-target analysis window presented here.

vocabulary and their eyetracking performance. These results are reported in the Supplementary Materials.

*Data cleaning and exclusion* Following the same trial exclusion criteria as Experiment 1, we excluded 97 trials with data for less than 1/3 of the analysis window and 54 trials where participants only looked at one of the videos during the entire test trial. This resulted in 122 total excluded trials (due to overlap across the criteria). A total of 3 participants were removed for contributing data to fewer than 16 trials. The final sample of 30 participants contributed an average of 27/32 trials.

*Subject- and item-level comparisons to chance* Statistical details over trial types, subjects, and items are provided in Table 2.6 and summarized below. For each trial type, we took the proportion of time toddlers spent looking at the target verb video in each of the four conditions, and compared that to chance looking (50%). By Shapiro-Wilk test, each set of trial type means was normally-distributed (all  $ps \geq 0.12$ ). Using one-sample t-tests, we find that when participants heard a regular verb, target-looking was significantly above chance (i.e.  $>50\%$ ). This was true both when regular nouns were correctly pronounced and when they were mispronounced, and held for a significant majority of toddlers (by binomial test) in each case.

In contrast, irregular verbs did not elicit above-chance looking from toddlers in either pronunciation condition. In fact, participants looked at the wrong video at above chance rates when hearing irregular verbs mispronounced. Across participants, for irregular verbs, fewer than half of participants looked to the right video, regardless of pronunciation type. The same pattern held across items: toddlers looked at the target  $>50\%$  of the time for 75% of regular verbs, and 25% of irregular verbs, regardless of pronunciation. See Figure 2.3 for a by-item breakdown and Table 2.6 for details.

Table 2.6: Trial type-level means and statistical tests for Experiment 2a (18mos) and Experiment 2b (26mos). Each row depicts data for each Pronunciation Type ('Pron.', correctly pronounced (CP) or mispronounced (MP)) and Verb Type (regular (reg) or irregular (irreg)). Mean shows the mean proportion of target looking across participants for each of these 4 trial types. The next two columns (Stat and p-value) show the results of one-sample t-tests comparing this proportion to chance (.50). The next two columns ('Ss>chance' and 'Binom p-value') show how many participants had a proportion of target looking >.5 in each trial type, and the corresponding p-value from a binomial test. The final column ('Items>0.5') shows how many of the 4 items in each trial type were looked at above chance when they were the target.

Age	Pron.	Verb Type	Mean	Stat	p-value	Ss>0.5	Binom p-value	Items>0.5
18mo	CP	reg	0.57	t(29) = 3.09	0.004**	22/30	0.016*	3/4
18mo	CP	irreg	0.47	t(29) = -1.58	0.125	12/30	0.362	1/4
18mo	MP	reg	0.56	t(29) = 3.89	<0.001***	21/30	0.043*	3/4
18mo	MP	irreg	0.45	t(29) = -2.36	0.025*	10/30	0.099	1/4
26mo	CP	reg	0.61	t(32) = 5.83	<0.001***	30/33	<0.001***	4/4
26mo	CP	irreg	0.59	t(32) = 4.07	<0.001***	25/33	0.005**	4/4
26mo	MP	reg	0.51	t(32) = 0.49	0.627	18/33	0.728	2/4
26mo	MP	irreg	0.54	t(32) = 2.16	0.038*	23/33	0.035*	3/4

*Comparing across trial types* A 2x2 factorial ANOVA revealed that pronunciation type did not have a significant effect on looking time ( $F = 0.37, p = .542$ ). That is, toddlers looked to the correct video the same amount whether its label was correctly pronounced or mispronounced. Due to this null result, we present no follow up t-tests. In contrast, verb type did have a significant effect on looking time ( $F = 29.33, p < .001$ ). Participants looked significantly more to the target video when they heard regular verbs compared to when they heard irregular verbs. The interaction of pronunciation and verb type was not significant ( $F = 0.31, p = .577$ ). In sum, 18-month-olds showed a significant preference for regular verbs in the target window, but pronunciation played no role in toddlers' looking behavior. For regular verbs, 15/30 participants showed a mispronunciation effect: they looked less at the target verb when it was mispronounced compared to when it was correctly pronounced. This held for 3/4 items. For irregular verbs, 13/30 participants showed a mispronunciation effect, which held for 3/4 nouns.

## 2.6 Experiment 2a Discussion

The results of this study were not what we anticipated. 18-month-olds looked more at the correct video when the target was a regular verb (versus an irregular verb), regardless of whether it was correctly or incorrectly pronounced. For irregular verbs, toddlers did not look at the target video in the correctly pronounced trials, and indeed looked slightly but systematically more at the distractor video (i.e. the regular verb) when the target was a mispronounced irregular verb. This pattern reflects neither strong comprehension of these words, nor the expected mispronunciation effect. Why might this be?

Although the Wordbank norms we based our item choices on predict that 80% of 18-month-olds understand our 8 test verbs overall (76% for regulars and 84% for irregulars), parents reported that only 52% of the participants in Experiment 2a

understood our 8 test verbs (45% for regulars and 59% for irregulars). However, low comprehension on its own would have resulted in chance levels of looking across both types of verbs, rather than the increased target looking for regulars that we observed.

Might toddlers simply have liked the regular verb videos better, regardless of what words they heard in the target sentences? To explore this, we measured whether toddlers already preferred the regular video in the part of the test trial before the target word was said. We found that they did not ( $t(59) = -1.53$ ,  $p = 0.13$ ), rendering this possibility unlikely. However, given the dynamic nature of videos, an idiosyncratic preference for the later portion of the regular videos could have played a role.

Another possibility, based on the lower reported comprehension for regulars vs. irregulars on the MCDI, is that toddlers had a tendency to look more at videos depicting events they do not know the words for. That said, parent-reported MCDI and lab-measured word comprehension tasks like this one do not always go hand in hand (Bergelson & Swingley, 2012; Houston-Price, Mather, & Sakkalou, 2007; Valteau et al., 2018), leaving it unclear how much stock to put in MCDI data for predicting comprehension for a specific set of items. A final possibility is that toddlers really do understand the regular verbs, but not the irregulars. If so, this would suggest that regulars may be easier to learn, perhaps because of their relative phonological stability across conjugations.

Notably, 18-month-olds do *not* show any evidence of a mispronunciation effect for either verb type. Interpreting this pattern is challenging given toddlers' poor performance with irregular verbs. Taken at face value, our findings would lend credence to the theory that early verb representations may be holistic, in line with a *nouns-over-verbs* mispronunciation account. This would suggest a large role for lexical class in phonetic representations, given that the literature on noun mispronunciation reflects fine-grained phonetic representation in children half a year younger than those tested

here (e.g. Mani & Plunkett, 2010), and that toddlers of the same age in Experiment 1 showed an overall mispronunciation effect for our set of irregular and relatively low-frequency nouns. To provide a fuller picture of the development of verb representations, we next test older toddlers. An older sample allows us to determine both when toddlers begin to understand irregular verbs, and whether they show a mispronunciation effect for either verb type once comprehension is more robust.

## 2.7 Experiment 2b

Due to the perplexing pattern of results found in Experiment 2a, we conducted a followup study to disentangle the roles of regularity, pronunciation, and participant comprehension. In order to ensure that participants understand the meaning of our common verbs, we conduct the same study as in Experiment 2a but with an older group of toddlers. In Experiment 2b, we test 26-month-olds, who have larger verb vocabularies and thus a greater chance of understanding our test verbs.

### *2.7.1 Experiment 2b Methods*

#### *Participants*

Participants were 33 typically developing 24–28-month-olds (16 girls,  $M = 26.75$  months,  $SD = 1.57$  months) who met the same exclusion criteria as participants in Experiment 1.<sup>9</sup> An additional 11 participants were excluded due to fussiness ( $N = 10$ ) or technical error ( $N = 1$ ). Participants were 87.88% White, 3.03% Black, 6.06% multiracial, and 3.03% unreported. The fuss-out rate was notably higher for this age group, likely because the study was originally designed for younger toddlers. For brevity, we refer to this group of 24–28-month-olds as 26-month-olds.

---

<sup>9</sup> Our intended sample size was 32 but one more participant than expected met inclusion criteria.

## *Materials*

Materials were identical to those described in Experiment 2a, except parents completed the Words & Sentences version of the MCDI, which is designed for this age group. This version of the MCDI only queries production, not production and comprehension (Fenson et al., 1994). The Words & Sentences MCDI contains 680 words, 103 of which are verbs.

## *Procedure*

The procedure was identical to that described in Experiment 2a.

### *2.7.2 Experiment 2b Results*

#### *Questionnaire responses*

All families filled out the demographics questionnaire and vocabulary exposure measure. 26 parents completed the motor survey and 31 families completed the MCDI. On average, our participants were reported to produce 341.87 words. Participants were reported to produce an average of 5.55 out of our 8 test verbs.

On our vocabulary exposure measure, the mean verb exposure score was 4.47/5 ( $SD = 0.79$ ). That is, just as in the younger group, parents estimated their children heard our test words at least once a day, on average. See the Supplementary Materials for motor survey results.

#### *Eyetracking Results*

*Data cleaning and exclusion* The same data cleaning process was used as in Experiment 1. We therefore excluded 92 trials with data for less than 1/3 of the analysis window and 64 trials where participants only looked at one of the videos during the analysis window. This resulted in 126 total excluded trials (due to overlap across the criteria). A total of 3 participants were removed for contributing data to fewer than

16 trials. The final sample of 33 participants contributed an average of 28/32 trials.

*Subject- and Item-level comparisons to chance* Statistical details over trial types, subjects, and items are provided in Table 2.6 and summarized below. As in Experiment 2a, for each trial type, we took the proportion of time toddlers spent looking at the target verb video in each of the four conditions, and compared that to chance looking (50%). By Shapiro-Wilk test, each set of trial type means was normally-distributed (all  $ps \geq 0.15$ ). Using one-sample t-tests, we find that when 26-month-olds heard a correctly-pronounced verb, target-looking was significantly above chance (i.e.  $>50\%$ ) regardless of verb regularity. Indeed,  $>75\%$  of toddlers attained subject means above chance for all correctly pronounced trial-types. Similarly, 26-month-olds looked at the target at above-chance rates when regular verbs were mispronounced, with more than two thirds of toddlers attaining subject means above chance for this trial type. That said, results were numerically weaker for mispronounced regulars than for correctly pronounced verbs of either type. In contrast, for irregular mispronounced verbs, toddlers' looking to the target did not differ from chance, with just over half of participants attaining subject means above chance. The pattern over items mimicked that over subjects, with  $>50\%$  target looking for 100% of correctly pronounced verbs, 75% of mispronounced irregular verbs, and 50% of mispronounced regular verbs. See Figures 2.2 and 2.3 for subject- and item-level figures and Figure 2.4 for timecourse.

*Comparing across trial types* A 2x2 factorial ANOVA showed a main effect of pronunciation type, just as in Experiment 1 ( $F(1,128) = 16.32, p < .001$ ). Participants looked at the target video less when the target verb was mispronounced vs. correctly pronounced. In contrast to both Experiment 1 and 2a, there was no main effect of verb type ( $F = 0.09, p = .766$ ). That is, collapsing across pronunciations, looking

was equivalent for regular and irregular targets. Finally, just as in Experiment 1 and 2a, the interaction between verb type and pronunciation was not significant ( $F = 1.27, p = .262$ ).

Follow-up t-tests showed that for regular verbs, toddlers looked significantly more at the target when labels were correctly pronounced compared to mispronounced ( $t = 3.81, p = < .001$ ). For irregular nouns, the t-test was marginal ( $t = 1.98, p = .052$ ), suggesting that toddlers did not significantly reduce their looking when labels were mispronounced, but that looking behavior was trending in the same direction for both regular and irregular verbs. For regulars, 27/33 participants reduced their looking when labels were mispronounced. This held for 4/4 verbs. For irregulars, 17/33 participants and 2/4 verbs showed a mispronunciation effect.

### *2.7.3 Experiment 2b Discussion*

26-month-olds showed a markedly different pattern of results from 18-month-olds in the same study. In fact, their pattern of results was much more similar to the results from Experiment 1, i.e. the performance on regular and irregular nouns by 18-month-olds. Specifically, our ANOVA revealed that this older group of toddlers showed a significant mispronunciation effect overall, just as we saw in Experiment 1. This suggests that well-specified verb representations have emerged by this age, with stronger evidence from our results for regular verbs in particular. Importantly, while 26-month-olds did not show a significant mispronunciation effect for the irregular verbs alone, they clearly showed a stronger comprehension of irregular verbs relative to 18-month-olds in Experiment 2a. We take up these differences further in the General Discussion, after a pooled analysis across all three experiments.

## 2.8 Pooled Analysis Across Experiments 1 and 2

To model the effect of lexical class and participant age on looking behavior, we combined the data from both Experiment 1 and Experiment 2 into a linear regression model. Here we predicted proportion of target looking as a function of pronunciation (correctly pronounced or mispronounced) and regularity (regular or irregular). We also included experiment as a variable (1, 2a or 2b), which indexed both participant age and word class. Since none of the previous studies showed significant pronunciation×regularity interactions, we did not include them in this model. We did include pronunciation×experiment and regularity×experiment interactions. The final model was:

$$\hat{y} = \beta_0 + (\beta_1_{\text{Regularity}} + \beta_2_{\text{Pronunciation}}) * \beta_3_{\text{Experiment}}$$

Variables were contrast coded such that the intercept represents the grand mean proportion of target looking for correctly pronounced, irregular words. Experiment 1 was the reference group. See Table 2.7 for model estimates and statistics.<sup>10</sup>

The model predicted significant variance (adjusted- $R^2 = 0.17$ ,  $p < .001$ ). An ANOVA analyzing the effects also showed that all main effects and interactions were significant. On average across studies, participants looked 4.79% less at the target when its label was mispronounced and looked 5.89% more at regular words than at irregular words. Compared to Experiment 1, participants in Experiment 2a looked 6.60% less at the named target, whereas in Experiment 2b participants looked 6.70% more at the target.

In addition to these significant main effects, there was also a significant interaction between regularity and experiment. While toddlers' performance for regulars was roughly consistent across studies, their comprehension of irregulars differed by study.

---

<sup>10</sup> A model that also includes CDI score did not improve fit over this model by model comparison ( $F = 1.63$ ,  $p = 0.10$ ).

Table 2.7: Pooled variance model regression output

Term	Beta	Std. Error	t-value	p-value
Intercept	0.535	0.010	56.214	<0.001 ***
Pronunciation	-0.048	0.011	-4.361	<0.001 ***
Regularity	0.059	0.011	5.365	<0.001 ***
Exp. 2a	-0.066	0.024	-2.763	0.006 **
Exp. 2b	0.067	0.023	2.881	0.004 **
Pron. * Exp. 2a	0.045	0.027	1.645	0.101
Pron. * Exp. 2b	-0.020	0.027	-0.741	0.459
Reg. * Exp. 2a	0.017	0.027	0.609	0.543
Reg. * Exp. 2b	-0.089	0.027	-3.305	0.001 **

Estimated marginal means for regular words were approximately the same across studies (3% difference) while they varied more widely for irregulars (10% difference), see Figure 2.6. Taken together, this pooled analysis indicates that lexical class, regularity, and age each play a role in driving word comprehension.

Finally, the ANOVA revealed a significant interaction between pronunciation and study overall ( $F = 3.10$ ,  $p = .046$ ), though this interaction was not significant between individual levels of these factors (see Table 2.7). As visualized in Figure 2.6, while participants across all three samples had similar target looking for mispronounced words, looking varied more widely for correctly pronounced trials, with the largest pronunciation effect apparent for older toddlers (in Experiment 2b).

## 2.9 General Discussion

In the current set of studies, we investigated the role of regularity and word class on toddlers' phonetic representations of familiar nouns and verbs. Our corpus analysis confirmed that young children hear many more irregular verbs than irregular nouns. This set the stage for our two empirical studies looking at regularity, lexical class, and wordform specificity. In Experiment 1, we found that both regular and irregular nouns were readily understood by 18-month-olds when they were cor-

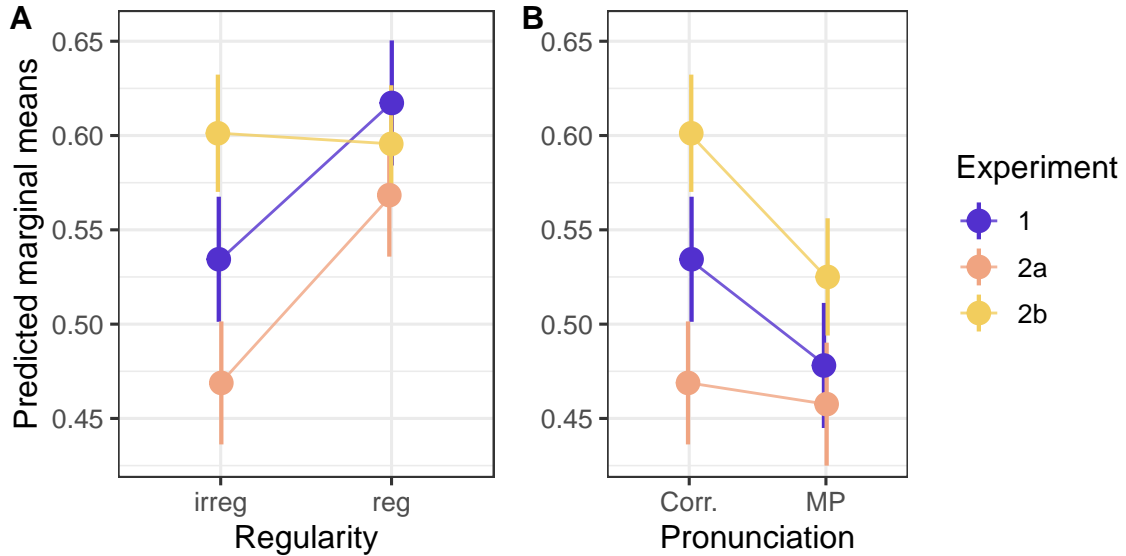


FIGURE 2.6: Plots of the significant interactions in the pooled model. X-axis shows predicted marginal effects for each experiment split by regularity (panel A) or by pronunciation type (panel B).

rectly pronounced. When these nouns were mispronounced, there was an (expected) overall decrement in performance across both regulars and irregulars, which was particularly robust for irregular nouns. Overall, there was an effect of both regularity and pronunciation in this experiment.

Experiment 2a showed a strikingly different set of results in toddlers of the same age. 18-month-olds showed no hint of a mispronunciation effect for verbs, but also showed generally poor comprehension, particularly for irregular verbs. Older toddlers in Experiment 2b looked far more similar to the 18-month-olds in Experiment 1, with some notable differences. Summarily, 26-month-olds showed a main effect of pronunciation, but not regularity. 26-month-olds both understood the tested verbs, and showed degraded comprehension when they were mispronounced, suggesting well-specified representations. However, in contrast to the 18-month-olds in Experiment 1, 26-month-olds' mispronunciation effect was particularly robust for regulars rather than irregulars.

Taken together, these results let us answer the three questions we posed in the introduction. First, we asked how common irregular nouns and verbs are in the input. We find that irregular verbs are highly common, making up nearly half of the most frequent action verbs young children hear. For nouns, irregularity was far less common: only one of the top 25 concrete nouns in young children’s input was irregular. This corpus analysis dovetails in an interesting way with our experimental results. Irregular nouns in English are relatively rare, meaning that toddlers have little experience with them. In our experiment, toddlers’ target looking fell to chance levels for mispronounced irregular nouns, suggesting that toddlers may have relatively brittle representations of how irregular nouns sound, such that changing the nuclear vowel blocks comprehension entirely. Verbs showed the opposite pattern: A large proportion of the common verbs young children hear are irregular, and in the group of toddlers who understood the test verbs (i.e. the 26-month-olds), performance on mispronounced *regular* verbs fell to chance, the marker of totally blocked lexical access.

This pattern of results suggests that rates of vowel alternation may shift toddlers’ expectations about the vowels in words for a given lexical class. That is, in a lexical class with few alternations (i.e. nouns), toddlers have difficulty with vowel changes in words that alternate, while in lexical classes with frequent alternations (i.e. verbs), comprehension takes a harder hit when non-alternating verbs occur with the wrong vowel.

### *2.9.1 Interlocking effects of age, regularity, word class, and wordform specificity*

We can now answer our remaining initial questions in the affirmative: regularity does play a role in early phonetic representations, and wordform specificity does indeed differ between nouns and verbs. However, the patterns that emerged were not simply characterizable as e.g. irregulars-over-regulars or nouns-over-verbs. Instead,

we found interlocking patterns across regularity and lexical class, their relationship to wordform representations, and development.

Why do 18-month-olds succeed at representing the sounds in irregular nouns but fail at even understanding irregular verbs? This result is not likely to be explained by overall vocabulary size, age, or demographic variables, since these factors (at least as we measured them) did not vary across the groups of toddlers in Experiment 1 and 2a.

Might 18-month-olds' difficulties with verb comprehension stem from challenges in selecting referents presented as videos instead of still images? Evidence against this possibility comes from studies testing infants' word comprehension with video rather than still image stimuli (Bergelson & Swingley, 2015; Tincoff & Jusczyk, 1999, 2012). These studies find evidence for comprehension at the same age with videos as with still images for a set of common words (mostly nouns). Thus, it is unlikely that stimulus type explains our results.

A more plausible explanation is that the nature of nouns and verbs and what it takes to learn them leads to the differences we find across experiments. We found that toddlers' abilities to both understand and represent the sounds in verbs improved between 18 and 26 months. Convergent research from a recent study finds that 22—24-month-olds (i.e. children between the ages tested here) understand verbs at above-chance rates, consistent with the trajectory for verb comprehension that we find for our correctly pronounced trials (Valleau et al., 2018). It seems likely that some aspect of linguistic development that may be particularly useful for verb learning comes online over the second half of year two. Then, once the verbs are understood, toddlers appear ready to represent their sounds with high fidelity. What types of skill improvements might fit the bill?

Two (potentially complementary) possibilities are changes in vocabulary and in grammatical knowledge (which have been theorized to both be indexes of the same

internal mechanism (Bates & Goodman, 1997)). More specifically, children add many more verbs to their vocabularies and gain morphological knowledge between 18 and 26 months. In terms of vocabulary growth, our finding that 18-month-olds succeed at representing nouns but flounder with verbs may be due in part to the fact that 18-month-olds (both in general, and in our sample) know many more nouns than verbs. Notably, given how common the verbs we tested are and the frequency that parents reported their toddlers were exposed to them, mere exposure does not appear sufficient to create robust wordform knowledge. Moreover, given that 18-month-olds were reported by their parents to understand roughly the same fraction of test items across experiments, toddlers might need to actually understand a critical mass of words within a lexical class (as older toddlers very likely do) before they can build robust sound representations for items in that class. By hypothesis, once toddlers hit that critical mass, robust representations may propagate throughout the word class. The details of such a mechanism await further research.

In terms of grammatical knowledge, toddlers take strides in understanding inflectional morphology around their second birthdays. For example, evidence from comprehension studies looking at auxiliaries and plurals shows that 24-month-olds have an emerging understanding of tense, aspect, and plurality (Kouider, Halberda, Wood, & Carey, 2006; Valian, 2006; Wagner, 2001). While research with children younger than 24 months is rare in this domain, Kouider et al. (2006) find that 24-month-olds understand plural marking while 20-month-olds do not.

On one hand, these results suggest that an important difference between our 18- and 26-month-olds in Experiment 2 may be improvements in early morphological processing. Toddlers may remain agnostic about the pronunciation of verbs until they begin to map each surface form onto a grammatically meaningful dimension (like tense). On the other hand, a morphology learning explanation does not explain our results with nouns given that toddlers detected irregular noun mispronunciations

at an age younger than they are likely to understand plural marking based on Kouider et al. (2006)’s result. Thus, even a morphology-learning account would need to posit differentiated mechanisms by lexical class.

### *2.9.2 Blocked versus degraded comprehension when words are mispronounced*

As noted above, in the two studies where we found a mispronunciation effect (Exp. 1 and 2b), mispronunciations appeared to *block* comprehension for irregular nouns and regular verbs, but only decrease it modestly for regular nouns and irregular verbs. This pattern contrasts from the general pattern reported in Von Holzen and Bergmann (2018)’s meta-analysis, where they find that for high-frequency regular nouns, mispronunciations degraded but did not eliminate comprehension. It may be that we did not see this pattern because we are testing toddlers just on the cusp of understanding our test words (i.e. 18 months for the nouns in Experiment 1, and 26 months for the verbs in Experiment 2). Lending credence to this possibility, Bergelson and Swingley (2018) also find an immature mispronunciation effect in 6-10-month-olds, who are just beginning to understand common nouns.<sup>11</sup>

A limitation of the current study is that while our results add important knowledge regarding the roles of regularity, lexical class, and word-form specificity, they still do not uncover an age at which toddlers show the “ideal” mispronunciation effect we’d expect. By testing older 2-year-olds, future work could discover the age at which young children robustly understand regulars and irregulars of both lexical classes, showing degraded but not eliminated comprehension when these words are mispronounced. Testing older children would also allow for more fine-grained manipulations of word frequency that are generally challenging to measure in younger children, whose vocabularies are relatively small. The present work also sets the

---

<sup>11</sup> While Von Holzen and Bergmann (2018) did not find an overall age effect, very few studies tested infants younger than 12 months

stage for research into wordform representations of other lexical classes, both open (e.g. adjectives) and closed (e.g. prepositions).

Extending our work to consonant mispronunciations would further illuminate the trajectory on which wordform specificity emerges. While Von Holzen and Bergmann (2018)’s meta-analysis finds evidence for both consonant and vowel mispronunciation effects with high-frequency regular nouns, testing consonant mispronunciations with irregulars would reveal whether toddlers have the same degree of specificity for the parts of an irregular that alternate (i.e. the vowels) and the parts that generally do not (i.e. the consonants). Relatedly, following up on this result in Semitic languages, where 3-consonant roots are pervasive and cross lexical class boundaries, would provide a particularly fascinating extension to our work.

The asymmetry we find across nouns and verbs further emphasizes the important role lexical class plays in early word representations. Recent findings suggest that by 14 months, infants already expect determiners to precede nouns and pronouns to precede verbs, but not vice versa (Babineau, Shi, & Christophe, 2020). Here we add to this broader literature, providing evidence that toddlers appear to be sensitive to morpho-phonological characteristics like vowel alternation, the prevalence of which differs across word classes.

## 2.10 Conclusion

In principle, toddlers’ representations of how nouns and verbs sound could have been based solely on how often these words tend to occur in daily life, wholly independent of both grammatical and conceptual differences between these word classes. In practice, we see that word class plays a large role in toddlers’ ability to detect mispronunciations. These results suggest that phonological development does not proceed all at once across the entire lexicon, but instead develops piecemeal, in tandem with word comprehension in a given word class.

Furthermore, we found that regularity mattered not only for toddlers' ability to detect mispronunciations, but for their comprehension of verbs in the first place. This in turn suggests that the process of mapping multiple surface forms onto a given concept, and perhaps beginning to understand systematic differences that vowel alternations represent, is part of a critical interface between phonology, morphology, and semantics. Critically, our results also underscore that being able to perceive the differences between phonemes is necessary but sorely insufficient for properly interpreting speech-sound changes during word comprehension. This work represents first steps into considering the influences of lexical class and regularity on toddlers' refinement of wordform representations in their growing lexicons.

## Listeners can use coarticulation cues to predict an upcoming novel word

### 3.1 Introduction

To achieve the speeds necessary for typical speech, articulators maximize efficiency by moving into position for upcoming sounds while earlier sounds are still being articulated. This preparation leads to a phonetic blending throughout an utterance, referred to as coarticulation (Fowler, 1981). Coarticulation cues can provide early acoustic signals to upcoming speech sounds before those sounds have been fully articulated.

Across eyetracking studies, findings suggest that coarticulation cues facilitate speech perception: participants find the target object more quickly when they hear useful coarticulation cues compared to when they hear no cues (Beddor, McGowan, Boland, Coetzee, & Brasher, 2013; Salverda, Kleinschmidt, & Tanenhaus, 2014). Similarly, misleading coarticulation cues can create lexical competition between the cued word and the realized one, delaying target looking relative to controls. For example, in Dahan, Magnuson, Tanenhaus, & Hogan (2001), participants were prompted

to find an object (e.g. a rod) with edited tokens where coarticulation cues in the vowel were either correct (e.g. “*ro(d)d*”) or misleading. When participants heard “*ro(k)d*”, where the vowel had coarticulation cues for the word “rock” but the final phoneme matched “rod,” they were slower to find the target compared to trials where cues were consistent (“*ro(d)d*”). This effect occurred whether or not a rock was present as a competitor image during the trial. In contrast, coarticulation cues that matched a novel word like in “*ro(p)d*” did not slow participants down. These findings suggest that the delay in target looking was caused by lexical competition between two known words; since *rop* was not a word participants knew, it did not create competition and thus did not influence participants’ target-looking. However, since this study only used familiar referents, it leaves open the possibility that referential context (i.e. the presence of novel objects) might interact with coarticulation cues to drive lexical competition, even for novel words. We explore this possibility below.

Studies with adults complement a robust literature looking at speech comprehension during development. Studies with infants often test their ability to perceive larger, phonemic changes (Von Holzen & Bergmann, 2018). For example, when hearing one-phoneme mispronunciations, infants look less to a target object than when the label is correctly pronounced (Mani & Plunkett, 2010). Even more relevantly, toddlers show gradient responses to mispronunciations: 1-feature changes disrupt comprehension less than 2- or 3-feature changes. Crucially, this gradient response only emerges when a plausible referent (e.g. a novel object) for a novel word (i.e. a mispronounced familiar word) is present (White & Morgan, 2008; White, Morgan, Wier, & Brown, 2004). This suggests that the greater pragmatic context plays a role even in early speech comprehension.

In what follows, we explore how adults respond when coarticulation cues in a familiar word point to a novel word while a possible referent for that novel word is visually available. In a context where a novel label is expected, listeners might

attend to coarticulation cues that do not match any known words. Do listeners need to know a word already for its coarticulation cues to interfere with word recognition? In order to answer this question, we presented participants with unknown objects alongside familiar ones, and manipulated coarticulation cues. Germane to this design is the assumption that new words go with new referents, i.e. mutual exclusivity.

### 3.1.1 *Mutual Exclusivity*

Mutual exclusivity (Carey & Bartlett, 1978; Markman & Wachtel, 1988) refers to learners' preference to map novel words onto novel objects, as opposed to mapping novel words onto objects whose names are already known. Older children use mutual exclusivity more reliably than younger children to match novel objects with novel labels, with the rate of use stabilizing in early childhood (Lewis, Cristiano, Lake, Kwan, & Frank, 2020). Studies with adults do not generally employ the same simple-choice mutual exclusivity tasks used with young children. However, cross-situational learning studies suggest that adults *do* use mutual exclusivity, and learn more novel words in conditions where they know the names of more distractor objects (Hendrickson & Perfors, 2019) or when a one-to-one mapping is maintained between objects and labels (Yurovsky & Yu, 2008).

Mutual exclusivity is less robust in some contexts (e.g. multilingual learners use it less; Byers-Heinlein & Werker (2009)). Further, mutual exclusivity cannot constrain learning in all contexts or it would bar learners from acquiring typical language features like synonyms or taxonomic relations. While most studies of mutual exclusivity use novel words that aren't neighbors of common nouns (e.g. *dax*), phonetic similarity may play a role. For instance, White, Yee, Blumstein, & Morgan (2013) showed that adults do not always use mutual exclusivity when perceiving newly-learned words. In an artificial language, they find that adults accepted one-feature mispronunciations of newly-learned words as correct even when a plausible alternate

referent was available (e.g. they selected the learned object *zad* when they heard “vad”). Adults were most likely to accept an incorrect label when they had only heard the new word once, suggesting that newly learned wordforms are particularly fragile. Here we extend this line of research beyond newly-learned words by examining whether adults use mutual exclusivity when considering novel words alongside highly similar early-learned nouns, i.e. words whose representations are ostensibly well-engrained in the lexicon.

While under-studied with regards to mutual exclusivity, phonological similarity may be particularly relevant for real-time word comprehension. Indeed, speech is notoriously variable in its realization (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), requiring listeners to perceive incoming speech adaptively (Kleinschmidt & Jaeger, 2015). Phonological similarity affects speech comprehension generally; words that share cohorts or rhymes create competition with one another (Alloppenna, Magnuson, & Tanenhaus, 1998). In the context of mutual exclusivity, a novel word that is highly similar to a known word may be difficult to learn because it may be perceived as a speech error. Put another way, if my friend mentions her “new dresh,” it is more likely that she is discussing the known object *dress* than a novel object *dresh*. However, the calculus of this situation changes if I see my friend holding an unknown object.

By testing familiar words and novel words that differ only in their final consonant, we can observe how adults perceive a novel word in a context where both a novel word and a familiar word are plausible. This allows us to test the relative contributions of phonetic cues and referential context to speech comprehension.

### *3.1.2 The Present Study*

Our study has two aims. The first (and primary) aim is to determine what happens while listeners perceive coarticulation cues in a context that makes a novel word

plausible. Do listeners take the greater context into account when understanding speech and consider any legal sound combination, or are listeners constrained by words they already know while attending to sub-phonemic cues like coarticulation? We assume that adults will look at familiar objects when hearing familiar labels (e.g. a photo of a crib when hearing “crib”), and novel objects when hearing novel labels (e.g. a photo of a new object when hearing “crig”) (Markman & Wachtel, 1988; Markman, Wasow, & Hansen, 2003). Here we ask what they do with misleading stimuli where the *vowel* cues a novel word, but the *final consonant* is consistent with a familiar one (e.g. *cri(g)b*). The answer to this question will tell us whether phonetic cues (i.e. coarticulation) interface with pragmatic cues during lexical competition and selection.

The second (auxiliary) aim of this study is to examine the relationship between mutual exclusivity and the phonological similarity of a novel word to a known word. Mutual exclusivity may only function as a word-learning heuristic when the novel label is sufficiently different from known words in the context. If this is the case, we would expect to find that listeners do not map a novel label like “crig” to a novel object when a crib is also an available referent; instead, listeners may assume they misheard the speaker or that the word was mispronounced. On the other hand, if listeners’ mutual exclusivity bias does not interact with a lower-level phenomenon like phonetic similarity, then we would expect to find that listeners use the mutual exclusivity bias at very high rates, even when the novel words presented are highly similar to known words. In this case we expect to see listeners looking significantly to e.g. a spiky dog toy when they hear “crig.”

To achieve these aims, we present an eyetracking study which simultaneously tackles questions of both coarticulation processing and mutual exclusivity. Taken together, the results highlight interactions between phonetic, lexical, and pragmatic considerations. While our main goal was to seek evidence that coarticulation percep-

tion interacts with the pragmatic context, this work also inadvertently highlighted some limits to the mutual exclusivity bias in adults.

## 3.2 Methods

### 3.2.1 *Participants*

We collected data in two samples. Sample 1 ( $N = 20$  participants) was a convenience sample of adults affiliated with a private university in the United States. The second sample ( $N = 37$ ) was recruited through this university’s undergraduate research program. Both consisted of fluent English speakers. Four participants had native languages other than English. Their exclusion did not change the results; they are retained in the analyses below. Race data was not collected.

### 3.2.2 *Design*

We use a modified Visual World paradigm akin to Looking While Listening (Allopenna, Magnuson, & Tanenhaus, 1998; Fernald, Zangl, Portillo, & Marchman, 2008). This study was a control for a study with infants, thus the design is geared towards younger participants, i.e we provide two images instead of four, and there is no overt selection response.

After 4 warm-up trials, participants saw 24 experimental trials. On each, they saw two images and heard one sentence. Sentences came from one of three conditions, which differed only at the target word. In the **match** condition, participants heard a label for the familiar object (e.g. “crib”). In the **mismatch** condition, participants again heard a label for the familiar object, but the coarticulation cues on the vowel did not match the expected final consonant and instead matched a novel word (e.g. the vowel /I/ in “crib” cued an upcoming /g/). The actual final consonant matched the familiar word (e.g. “cri(g)b”). In the **unfamiliar label** condition, participants heard a novel word which differed from the familiar word only in its final consonant

Table 3.1: The 8 stimuli words included in the study by condition. The sound in parentheses indicates the sound being coarticulated in the vowel.

Match	Mismatch	Unfamiliar
crib	cri(g)b	crig
foot	foo(p)t	foop
bed	be(g)d	beg
block	blo(p)ck	blop
egg	e(b)gg	ebb
dog	do(b)g	dob
hand	ha(ŋ)d	hang
sock	so(p)ck	sop

(e.g. “crig”). Each trial contained a familiar object and an unfamiliar object. Each pair of images was always presented together, making 8 unique pairs of images. All pairs were presented three times throughout the study, once with the audio prompt for each of the conditions listed above. See Table 1 and Figure 1.

### 3.2.3 Stimuli

Each yoked pair contained one photograph of a familiar, early-learned object (e.g. a crib) and one of an unfamiliar object with no well-known name (e.g. a spiky dog toy). All images were superimposed on a 500x500 pixel gray background. Auditory stimuli were created in Praat (Boersma, 2001). A female speaker produced all stimuli sentences. For unfamiliar tokens, stimuli were used as recorded. To create the match condition, we took the familiar word and spliced together the onset and vowel of one token with the final consonant of another token of the same word (e.g. *cri(b)b*). Stimuli were spliced at the closest zero-crossing to the closure of the stop consonant. For the mismatch condition, we spliced the onset and vowel of the novel word and the final consonant of the familiar word (e.g. *cri(g)b*). By splicing both the match and mismatch conditions, we could be sure that differences between the two conditions were not due to the splice itself. Stimuli are available at [bit.ly/36Ijy54](https://bit.ly/36Ijy54).

### 3.2.4 Procedure

Participants came to the lab and signed the consent form. They were then asked about their language background using a 3-question survey to determine when they learned English and if they spoke any additional languages. Participants in Sample 1 then proceeded to the eyetracking component. During data collection of Sample 1, we noticed that some participants were looking to the familiar object even when they heard the unfamiliar label. To ensure that this was not a result of a misunderstanding about the nature of the task, we collected an additional sample where we included more information about the study. In Sample 2, participants were read a script that stated that this study was designed as a control for an infant study, and reminded them that infants hear many words they may not know and that some of those words may sound very similar to words they already know.

Next, all participants were seated in front of a 1240x1028 pixel display connected to an Eyelink 1000 Plus eyetracker and calibrated using 5-point calibration. The eyetracking study itself contained 4 warm-up trials with a single image on the screen (e.g. a cracker) and a prompt to look at the image (e.g. “look at the cracker!”). After the warm-up trials, the test trials began. Each test trial featured two images on the screen, one of a familiar object and one of an unfamiliar object (See Figure 1). Participants first saw both images in silence for 2500ms, then heard one of the three corresponding prompts (coarticulation match, coarticulation mismatch, or unfamiliar label). Looking was recorded throughout the trial. After the experiment, participants were thanked and when applicable provided with course credit.

### 3.2.5 Data analysis plan

Our first question was whether mismatching coarticulation cues (e.g. *cri(g)b*) led listeners to look more at the unfamiliar object relative to match trials. To answer this, we fit growth curves for just the mismatch and match conditions. Our second



**Match:** Look at the cri<sub>(b)</sub>b!

**Mismatch:** Look at the cri<sub>(g)</sub>b!

**Unfamiliar:** Look at the crig!

FIGURE 3.1: Two example images presented as in a trial with the corresponding three auditory prompts. Prompts were heard aloud and are written here for demonstrative purposes.

question was based on a surprising discovery in our initial visualization of the data in Sample 1 ( $N = 20$ ). Namely, we found that counter to our predictions, not all participants readily looked at the unfamiliar object when hearing the unfamiliar word. In an exploratory analysis, we first categorized participants into two “listener types” based on their behavior on unfamiliar label trials. Next, we fit growth curves to their performance on these trials, and added listener type as a fixed effect to a base model to confirm that listener type helped explain variance in behavior. Finally, we returned to our original model but added listener type to look at any differences in performance on the mismatch and match conditions based on performance in the unfamiliar label condition.

For the growth curve analyses, we follow Mirman (2014). We conduct pseudologistic growth curve analyses using weighted empirical logits. We analyzed participants’ gaze to the *unfamiliar object* (the distractor image) in the match and mismatch

Table 3.2: Fixed effects of the growth curve analysis. ‘Int’ represents the intercept term. Terms labeled ‘ot’ represent time via orthogonal polynomials, with number indicating the order, e.g. ot1 is the linear term. Degrees of freedom (df) are approximate and estimated using Satterthwaite’s method.

Fixed Effects	Est.	SD	t =	df	p value
Int.	-1.36	0.16	-8.35	114	<0.001 ***
ot1	-5.31	0.30	-17.90	114	<0.001 ***
ot2	0.07	0.21	0.35	114	0.729
ot3	0.39	0.15	2.63	114	0.01 *
ot4	0.35	0.11	3.11	114	0.002 **
Cond.	-0.23	0.05	-5.06	114	<0.001 ***
Sample	-0.01	0.09	-0.11	114	0.916
ot1:Cond.	-1.00	0.30	-3.37	114	0.001 **
ot2:Cond.	0.58	0.21	2.82	114	0.006 **
ot3:Cond.	0.44	0.15	2.94	114	0.004 **
ot4:Cond.	-0.25	0.11	-2.19	114	0.03 *

conditions from 200-1500ms after the target word’s onset in each of 8 trials per condition. To accomplish this, each participant’s gaze data was divided into 20ms bins. For each participant, we summed looks (0 or 1) for each of the 20 ms bins across items, resulting in a condition-level value (with possible values ranging from 0-8) for each bin. These condition-level sums were then used to create the empirical logits and weightings used in the model.

### 3.3 Results

#### 3.3.1 Aim 1: Coarticulation effect

To determine whether participants were slower to find the target in the mismatch condition relative to the match condition, we used the lmerTest package to fit the following model:

$$\hat{Y} = (ot1 + ot2 + ot3 + ot4) * Condition + Sample + ((ot1 + ot2 + ot3 + ot4) || Subj \times Condition)$$

The fixed effects were Condition (match or mismatch), Sample (1 or 2) and orthogonal polynomials to model time. Time 0 on each trial corresponds

to target word onset (e.g. the start of *crib* in “Look at the *crib!*”). We use four orthogonal terms to model time in keeping with Mirman (2014), which establishes that individual terms are needed to appropriately capture each curve in the eye movement trajectories for this method. Thus, our data required four time terms (ot1 through 4).

The final model also included interaction terms between Condition and each of the 4 polynomial terms. The model including these interaction terms was significantly better than one without ( $\chi^2 = 723.73$ ,  $p < .001$ ), suggesting that looking to the unfamiliar object unfolded differently as a function of whether participants’ heard *cri(b)b*, *cri(g)*, or *cri(g)b*.

A summary of the fixed effects is in Table 2. All estimates are in logit space. The intercept estimates the log odds of looking to the unfamiliar object in the match condition for Sample 1. Sample was not a significant predictor, suggesting that our added instructions for Sample 2 did not influence looking behavior. There was a significant main effect of condition, indicating that participants looked at the unfamiliar object significantly more overall in the mismatch condition (e.g. *cri(g)* when hearing *cri(g)b*). Looking at the time estimates, the significant linear term indicates that looking to the unfamiliar object decreased overall as each trial unfolded. The significant cubic and quartic terms confirm curvilinear gaze dynamics, shown (untransformed) in Figure 2.

The slopes of the two conditions were also significantly different on the polynomial terms, corresponding to differences in the speed that participants looked away from the unfamiliar object across conditions; participants were significantly faster to look away from the unfamiliar object in the match condition, and looked less at it overall. In sum, coarticulation mismatches pulled participants’ looking away from the target across the entire analysis window, i.e. participants were mis-cued by the coarticulation on the vowel in mismatch trials. This in turn suggests that

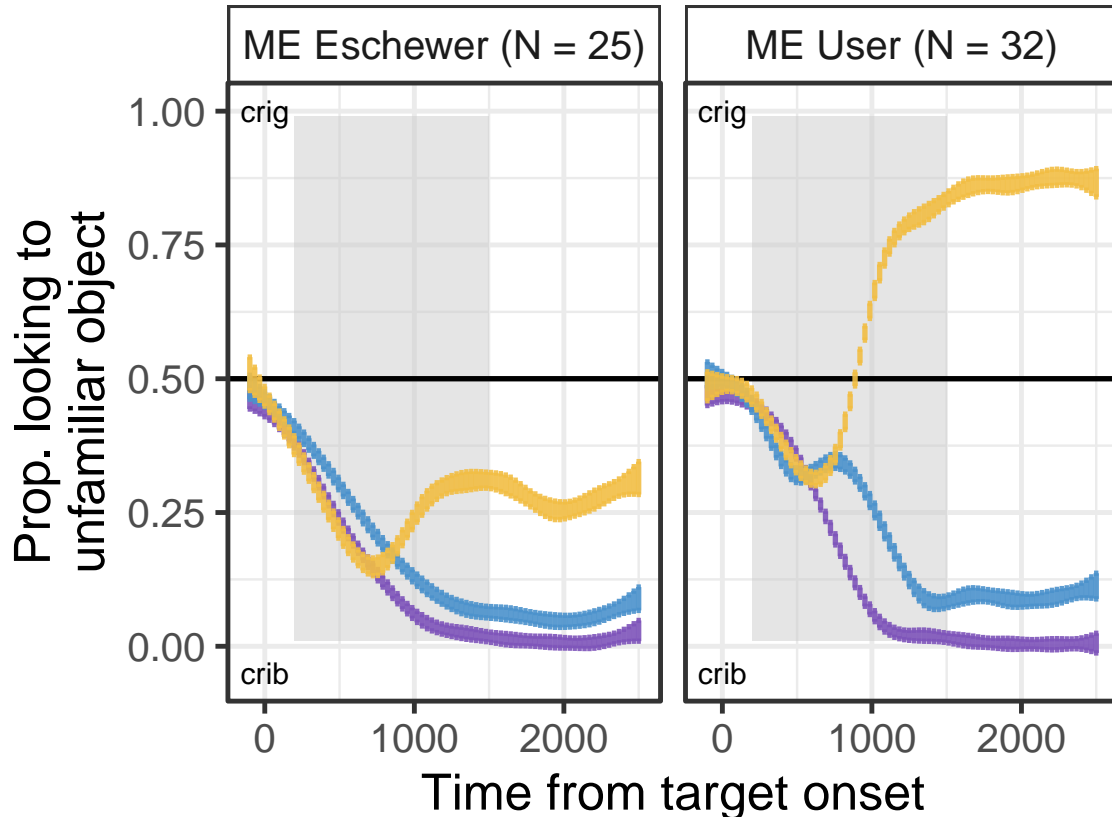


FIGURE 3.2: Looking over time to the unfamiliar object for both listener types. Horizontal line indicates chance. Shaded area indicates window of analysis for growth curve models. Width of band represents 95 percent CI.

the presence of an unfamiliar competitor object onscreen created lexical competition between the familiar word and the cued unfamiliar word

### 3.3.2 Aim 2: Mutual Exclusivity

Given the robustness of the mutual exclusivity (ME) constraint from early childhood on, we assumed that on hearing “*crig*”, participants would look at the unfamiliar object (e.g. the spiky dog toy) rather than the crib. To our surprise, a large proportion of our adult participants appeared not to do so, instead looking mostly at the crib upon hearing “*crig*”. Thus, in an exploratory analysis, we first categorize subjects into two groups, those who exhibited this unexpected behavior (ME Eschewers), and

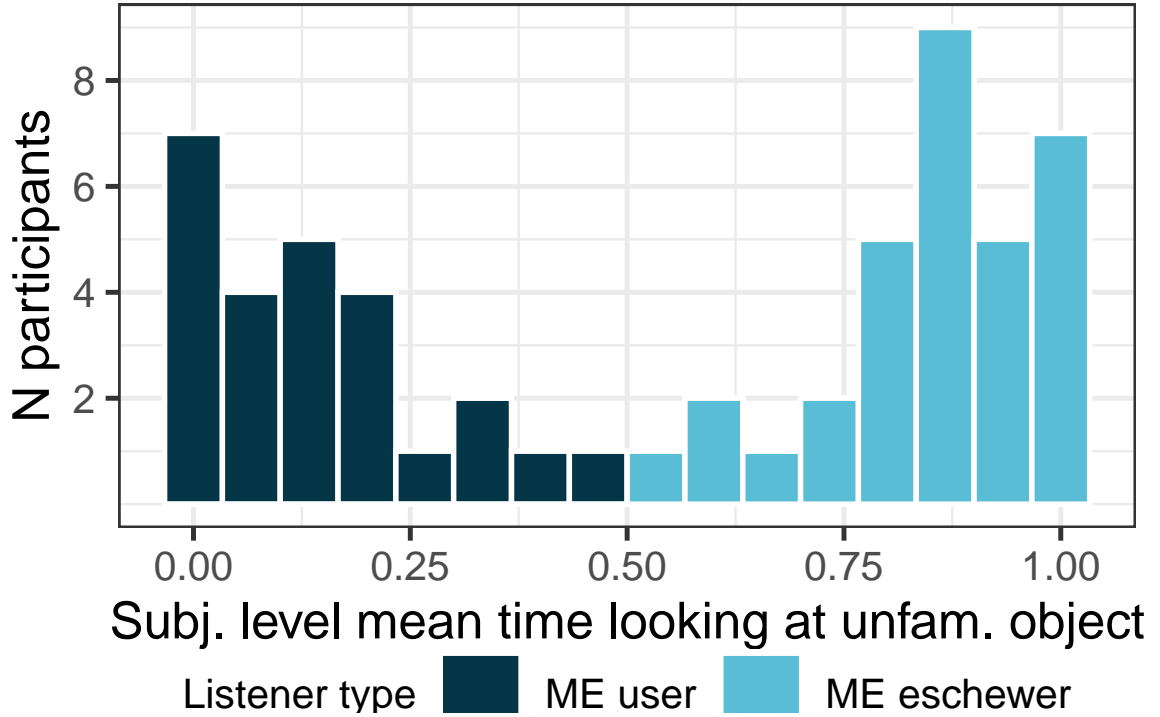


FIGURE 3.3: Histogram of unfamiliar label trial responses. X-axis shows subject-level mean proportion of time spent looking at the unfamiliar object 200-4000ms post-target word onset (full trial length). Higher proportions represent greater use of mutual exclusivity. Color indicates how participants were sorted in growth curve models.

those who behaved as expected (ME Users). Then we model their gaze using the same type of growth curves as in Aim 1, this time comparing these two groups' performance across conditions.

To quantify the two different patterns we observed across individuals, we began by aggregating the data for each participant's looking across all 8 of the unfamiliar label trials to determine what proportion of the time participants spent looking at the unfamiliar object. We used the entire length of the trial after the target word was uttered (from 200ms after target word onset to 4000ms after). We use the full window rather than making assumptions about the timeline of looking behavior given this unexpected result.

Participants who spent more than 50% of the trial after the target word's onset

looking at the unfamiliar object on unfamiliar label trials were categorized as ME Users ( $N = 32$ ). Participants who spent less than 50% of the time looking at the unfamiliar object were labeled ME Eschewers ( $N = 25$ ).<sup>1</sup> Neither group showed a preference for the unfamiliar object before the target word was spoken (50.4% for ME Eschewers and 49.8% for ME Users). See Figure 3.

After establishing the two distinct groups, we conducted an additional growth curve analysis predicting looking behavior in only the unfamiliar label condition using the same approach described above, comparing a baseline model to one with listener type (ME User vs. Eschewer).

The model with listener type as a fixed effect fit the data significantly better than one without it ( $\chi^2 = 15.12$ ,  $p < .001$ ). The final model was  $\hat{Y} = (ot1 + ot2 + ot3 + ot4) + Sample + ListenerType + ((ot1 + ot2 + ot3 + ot4)|Subj)$ . This suggests that in line with our holistic trial-level split, ME Users and Eschewers diverged in their dynamic responses to unfamiliar labels; we explore possible explanations in the Discussion. To see if this difference held beyond the unfamiliar trials, we next modeled the coarticulation effect in the match and mismatch conditions using listener type as a predictor.

### 3.3.3 *Listener type to predict coarticulation processing*

We extended the model in Aim 1, adding listener type as a fixed effect. This significantly improved model fit ( $\chi^2 = 12.98$ ,  $p < .001$ ), suggesting that in addition to differing in their responses in the unfamiliar label trials, the two listener-types had divergent looking trajectories in the two conditions where coarticulation cues were manipulated as well. Figure 4 highlights this difference: ME Users exhibited a characteristic ‘bump’ (from ~500 to 1000ms post-target onset) in the mismatch

---

<sup>1</sup> These distinct patterns were already present in the first half of the study and became stronger in the second half of trials, especially for the ME Users. See [bit.ly/36Ijy54](https://bit.ly/36Ijy54).

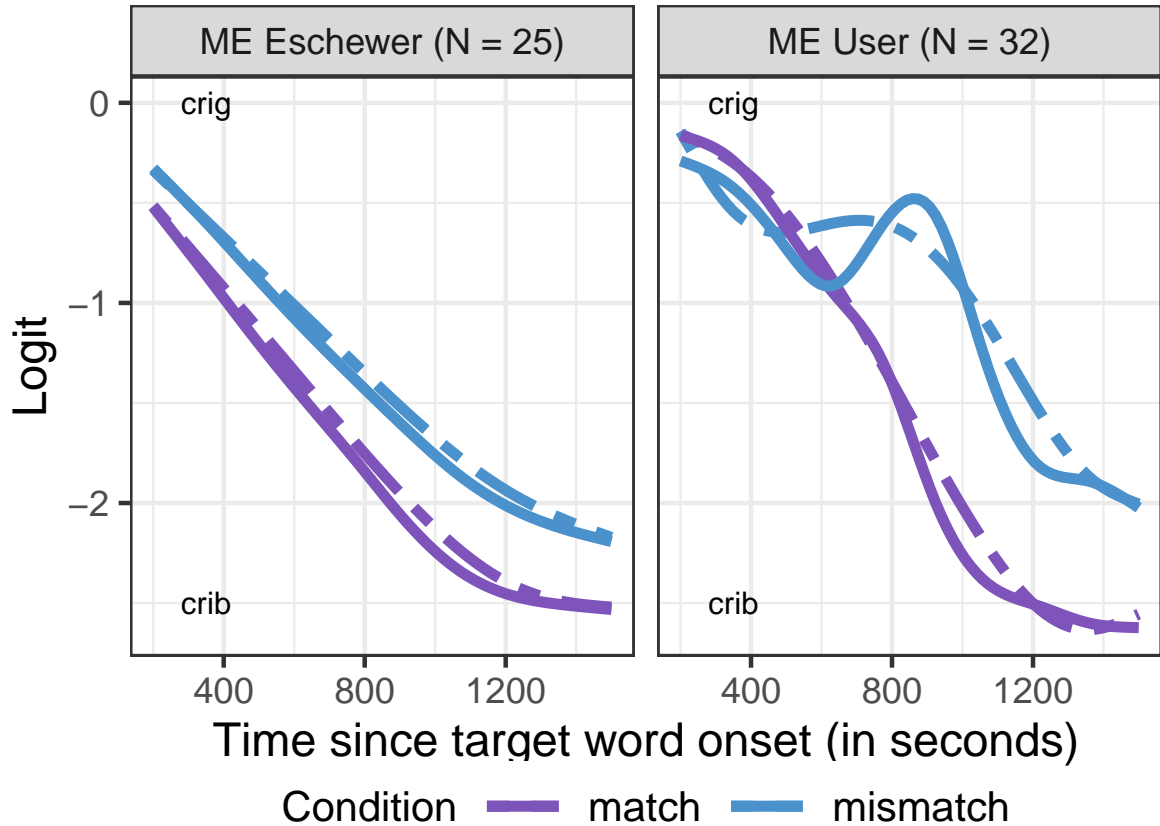


FIGURE 3.4: Comparison of the transformed raw data in logit space (solid lines) and the predicted values from the growth curve model which contains listener type as a predictor (dotted lines).

condition wherein they began looking at the unfamiliar object more when they heard the coarticulation cues in the vowel, then returned their gaze to the familiar object upon hearing the final consonant. In the ME Eschewers, this ‘bump’ is missing, suggesting that the coarticulation cues on the vowel did not create the same effect for these participants. Instead, ME Eschewers reduced their looking monotonically.

### 3.4 Discussion

We asked whether the referential context changes how coarticulation cues are used in speech comprehension. Previous work suggests that competition from within the lexicon is a prerequisite for coarticulation competition effects (Dahan, Magnuson,

Tanenhaus, & Hogan, 2001). Instead, we found that when listeners were provided with a plausible referent for an unfamiliar word, they were slower and less accurate at finding the target on mismatch trials compared to match trials. This suggests that listeners were not limited to the words in their lexicon when predicting the identity of a word from its vowel. Even ME Eschewers were sensitive to coarticulation mismatches, though the difference between match and mismatch trials was less pronounced for these participants. These findings provide evidence that coarticulation perception does not merely create competition between known words in the lexicon. Instead, coarticulation cues that correspond to novel words seem to guide behavior when a plausible referent for the novel word is available. This type of pragmatic inference where participants take context into account during speech comprehension may explain why we find a robust effect of coarticulation between our match and mismatch conditions where Dahan et al. (2001) did not.

While the divergence between the match and mismatch conditions is consistent with a lexical competition account, a possible alternative explanation is that upon hearing the mismatch stimuli (e.g. cri(g)b), participants were simply confused about the target. We find this explanation unlikely, given the systematic looking trajectories that participants provided across conditions; confusion would likely result in more random looking behavior than what our results have captured. That said, a lexical competition account would be bolstered by a replication of Dahan et al. (2001), to allow us to compare our findings to a condition with multiple familiar objects onscreen.

While we intended to examine the interaction of coarticulation cues with pragmatic considerations, along the way we unexpectedly discovered a limit to the mutual exclusivity bias. Participants fell into two distinct behavior patterns when faced with a novel word that was very similar to a known word. 56% of participants showed the pattern we would expect based on mutual exclusivity, i.e. looking to the unfamiliar

object when hearing the unfamiliar word. The other 44% showed the opposite pattern, looking at the familiar object for most of each trial even when they heard an unfamiliar word. Why might this be?

One possibility is that participants could not *detect* differences in final consonants across conditions. If this were the case, we would expect the timecourse of looking behavior in all three conditions to be equivalent; this is not what we find. Instead, all participants looked at the familiar object (e.g. the crib) least when they heard e.g. “crig.” Furthermore, participants looked less at the target on mismatch trials than on match trials, suggesting they detected fine-grained phonetic distinctions. It is therefore highly unlikely that half of our participants were unable to perceive the differences between each condition.

A Bayesian explanation in the vein of Kleinschmidt & Jaeger (2015) is that participants correctly heard each prompt, but because the prior probability of the familiar word was so much higher than a novel word, the winning candidate was the familiar word. Under this view, participants either interpreted the unfamiliar words as a mispronunciation of the familiar word instead of as a novel word, or assumed they misheard the prompt. These are plausible explanations for the data, since the unfamiliar words differed only in the final phoneme. Further, the mismatch condition likely added uncertainty, as those trials sounded unusual due to their conflicting cues. This explanation also helps us account for the emergence of two divergent approaches. ME Eschewers may have stronger priors to expect familiar words or less clear speech, whereas ME Users may have weaker priors in those domains.

To probe this possibility further, we took advantage of our randomized trial orders. Order 1 began with a familiar word while Order 2 began with an unfamiliar word. 20/29 Order 1 participants were ME Users, while only 12/28 Order 2 participants were ME Users. This suggests that the first trial participants saw may have biased them towards expecting familiar or unfamiliar words, in effect shifting

their priors. However, we interpret this finding with caution as a binomial test not significant ( $p = 0.06$ ).

Because this bifurcation of our participants was unexpected, these results require confirmatory follow-up beyond the second sample reported here. It would also be informative to vary the size of the phonetic difference between the match and unfamiliar conditions to see if the number of ME Eschewers scales with the size of the difference. Another fruitful avenue would be to query what features of the listener predict ME using/eschewing behavior (e.g. frequency of exposure to new words, frequency of exposure to unclear or accented speech, or personality factors like agreeableness.) Another important next step would be to add an overt response to our design, asking participants to click on the correct object in each trial. This would provide discrete evidence of the interpretation participants committed to, complementing the gradient gaze data.

Taken together, our results add evidence that the comprehension of sub-phonemic cues is subject to pragmatic factors outside of language. This further supports the inter-connectedness across levels of linguistic analysis, from global contextual effects to low-level phonetic cues. Further, we provide one of the first direct tests of mutual exclusivity in adults, showing that adults use mutual exclusivity to varying degrees when labels are sufficiently similar to known words.

### 3.5 Conclusion

Coarticulation cues are a fundamental component of speech perception, used to distinguish words in the lexicon from one another early in the speech stream. This study provides new evidence that coarticulation cues can also be used to distinguish known words from novel ones when a plausible novel referent is provided, and that this effect is present whether listeners map a novel word to an unfamiliar referent or not.

We also provided surprising evidence that individual differences play a large role in how participants perceive unfamiliar words that sound highly similar to known words. Taken together, these results show that coarticulation perception creates more than lexical competition - understanding this seemingly low-level phonetic cue involves a complex interplay between the context, the acoustics of the speech stream, and the possible intentions of the speaker.

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

### 4.1 Introduction

When we imagine an idealized word-learning scenario, we might conjure an image of a toddler who sees an object like a bottle, and then hears his mother say “bottle” pronounced clearly and at the opportune moment. This leads the toddler to draw a connection between the sounds he hears and the object he is attending to. Unfortunately, language learning in real world settings is rarely this straightforward. Setting aside the difficulty of learning abstract words like verbs or function words, even concrete noun learning requires both distinguishing wordforms from each other and selecting the appropriate meaning from a theoretically infinite set of possible meanings (Quine, 1960). Despite these obstacles, learners regularly pick out the correct word-object pairings.

Infants reliably find the right referent for wordforms that they encounter. In order to build a full, adult-like vocabulary, infants need to form a complex web connecting wordforms and their meanings to each other. For any given concrete object, there are

often multiple accurate labels for speakers to choose from. If we take the concept *dog* as an example, we can also refer to a dog with a synonym, as in “hound”, or label the dog at different levels of taxonomy, as in the subordinate label “labradoodle” or the superordinate “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 *bedsheet*). Each of these many-to-one relationships poses its own set of challenges to learners.

Homophones are thought to arise from a tension between the speaker’s need to minimize communication effort and the listener’s need to distinguish one message from another (Zipf, 1945). There are a finite number of phonotactically legal short words in any given language, so some particularly well-formed words end up being reused to reduce the demands of speaking (Zipf, 1945). While it is easier for speakers to produce short, well-formed words, ambiguity inevitably arises from unrelated words sharing the same phonological form. To disambiguate homophones, listeners can rely on context; homophones often belong to different syntactic categories or are used in very different semantic contexts (Dautriche, Fibla, Fievet, & Christophe, 2018; Piantadosi, Tily, & Gibson, 2012). Evidence suggests that this pattern of homophony corresponds to learning patterns: preschoolers resist learning new, unrelated meanings for known words, e.g. that “snake” also denotes an unrelated, novel object. However, they succeed at the same task when a known word’s new meaning is in a different syntactic category from its known meaning, e.g. that a novel object is called a “give” (Casenhiser, 2005).

While homophony arises from phonological pressures in a language, polysemy is a result of meaningful semantic extension. Polysemes are words that are used to refer to multiple conceptually similar objects, e.g. a cap on a water bottle and a baseball

cap worn on one’s head. Conceptually, these two labels are related in that they both cover the top of something (a bottle or a head). It is estimated that most high-frequency English words have multiple polysemous meanings (Floyd, Goldberg, & Lew-Williams, 2020; Srinivasan, Berner, & Rabagliati, 2019). Young children seem to be aware of these conceptual relationships; toddlers extend a polysemous word to novel instances that are attested in other languages, but not the child’s language, at rates above chance (Floyd, Goldberg, & Lew-Williams, 2020). When adults and preschoolers were taught a label for a novel object, they preferred to extend the novel label to the material as opposed to another object of the same shape, suggesting that polysemy may help learners structure the acquisition of new related words by highlighting probable referents (Floyd, Goldberg, & Lew-Williams, 2020). Taken together, the findings from homophony and polysemy research show that learning multiple meanings for the same wordform is possible if meanings are either sufficiently distinct or semantically related, creating an “uncanny valley” effect. In other words, research seems to suggest that if a word’s two meanings are very related, or completely unrelated, then learning is possible. However, if the relationship between two word meanings lies between those poles, then learning is a challenge.

We turn now to look at relationships where multiple words map to the same concept. First, synonyms are two words in the same language where the meanings of each word significantly overlap (e.g. *couch* and *sofa*). Synonyms do not need to share any sounds in common and overlap only in meaning. The adult English lexicon is rife with synonymy, with Roget’s thesaurus touting over 350,000 entries (Kipfer, 2011). Despite their frequency, synonyms seem to pose a learning challenge; 16-month-olds easily learn a first label for a novel object, but struggle to learn a second label for a familiar object. This effect persists when learning two novel labels 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). However, children clearly do eventually come to learn multiple labels for the same object, since they exist in the adult lexicon. Savage and Au (1996) find that half of four- and five-year-olds override their mutual exclusivity bias to learn two novel labels for the same object. In that study, children also entertained both provided words before settling on a single label, suggesting that with sufficient evidence of two labels, synonym learning occurs in preschoolers. Two-year-olds also show a willingness to accept a second basic-level label for a known object when provided with a new novel label (Mervis, Golinkoff, & Bertrand, 1994).

Taxonomic relations cause a problem for a one-to-one correspondence between concept and word; a single object can take different names at different levels of analysis, e.g. a dog can also be a dalmatian (subordinate) or a mammal (superordinate). Evidence suggests that these types of relations are difficult for children to learn, with preschoolers struggling to correctly answer questions that require use of both a subordinate and basic level label for the same series of objects (Inhelder & Piaget, 1969). However, two- and three-year-olds can make use of pragmatic directions from adults to learn that one novel object label is superordinate to another and thus may benefit from caregiver instruction (Clark & Grossman, 1998; Diesendruck & Shatz, 2001). Preschoolers also seem to benefit from cross-situational statistics when determining the meaning of an unknown word, differentiating between subordinate and basic-level meanings for a novel label depending on the learning exemplars presented (Wang & Trueswell, 2019).

Each of these types of ambiguity has been maintained in productive language systems, suggesting that language users are able to accommodate these types of relationships. Despite these many-to-one relations between wordforms and objects, Clark (1987) refutes the idea that multiple words can have identical meanings. Instead she proposes the principle of contrast: every word must pull its semantic weight

in order to remain in the lexicon, thus any 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. While the principle of contrast likely governs what is maintained in the lexicon, here we provide evidence that for many of the highest-frequency words in infants' environments, wordform variability is extensive and meaning differences are not always clear.

In the following study, we present evidence that the principle of contrast may not always help learners distinguish the meanings of multiple wordforms. We do this by exploring a fifth type of mismatch, one that is likely particular to infancy and early childhood: **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. 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 – especially relative to the adult lexicon – and shares overlapping phonetic material with “dog.” The overlap of both sound and meaning in wordplay may have unique effects on learnability. 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.

Many different types of wordplay have already been studied, but not as a single, united 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 IDS are phonological simplification and an increase in words with consonant-vowel-consonant-vowel 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 have been identified through parental report, a subjective measure of language usage. 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 systematically examining the overall prevalence of wordplay in the corpus. We then describe the features of a word likely to be used in wordplay and contrast that with words that do not appear in wordplay contexts. Next, we examine whether wordplay is a widespread phenomenon or whether it is idiosyncratic to specific families. Finally, we explore the effects of wordplay 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.

## 4.2 Methods

### 4.2.1 *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 impair-

ments. Families were compensated for their participation; compensation increased as participation continued. The sample was primarily white and well-educated: 75% of mothers had at least a Bachelor's degree and 95% of infants were white.

#### *4.2.2 Procedures*

Consent was obtained at the beginning of the yearlong study. Every month starting 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 daylong audio recording and one hour-long video recording on separate days 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.

#### *4.2.3 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.

For each recording, trained researchers listened to parts of the file that had been flagged in pre-processing as containing speech and identified concrete nouns. All videos were completely annotated. For audio recordings, in months 6 and 7, the entire recording was annotated. In months 8-13, the four most talkative hours according to the LENA software were annotated. 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.

#### *4.2.4 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. 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 sentence (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. To date, files from months 6-15 have been checked by two annotators.

#### *4.2.5 Identifying wordplay in the corpus*

The process above generated 348,820 annotations. Of the total annotations, 106,659 contained at least one change between the lemma and spoken wordform. The first author compared each unique realized wordform to its lemma and noted up to three phonological transformations that occurred to change the lemma into the realized wordform. See Table 4.1 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 were words that denoted private body parts or excrement, where all euphemisms were bundled together with their lemmas due to the enormous variety in wordforms.

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 -z suffix to mark plurality (morpheme-adding). Thus, “doggies” is 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 for infants. Further, the intended wordforms for very early productions are notoriously challenging to characterize.

The result of this identification process was a corpus where all changes that infants heard between the lemma and the realized form were identified and characterized. First we noted whether the word qualified as a wordplay word, and then we created a record of the transformations between lemma and wordform.

### 4.3 Results

We begin by describing the frequency of wordplay and morpheme-adding changes relative to words where no changes are made between the lemma and the wordform. We then focus on characterizing wordplay words in aggregate, examining the frequency of different change types. Next, we present a logistic regression predicting which of the lemmas in the corpus will have attested wordplay forms. After this, we pivot to examining family level variables with zero-order correlations. Finally we examine the role of wordform variability at the word level with a mixed-effects model.

### 4.3.1 Wordplay and other changes

There were three ways that any given word 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. We do not distinguish between words that only underwent wordplay changes and words that involved a combination of changes, e.g. both “doggy” and “doggies” were counted in the same way even though “doggies” contains both a wordplay change and a morpheme-adding change.

See Figure 4.1 for the relative frequency of these three change types and Figure 4.2 to see the relative proportions of each change type. Wordforms with no changes from the lemma, like “book,” were the most prevalent: by tokens, lemmas in the dataset appeared in their lemma forms 242,882 times in the corpus, making up the bulk of tokens (70% of all tokens).

Lemmas where we *never* observed any changes between the lemma and the wordform were low-frequency with a mean of just 5.49 tokens per lemma ( $SD = 16.60$ ). By types, 3,696 lemmas appeared exclusively in their lemma forms with no attested changes. These types of “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 1,653 lemmas that occurred only once in the corpus. The most common lemma that never changed was “snow” with 510 occurrences in the corpus.

Morpheme-adding wordforms like “books” were also common in the dataset. By tokens, a total of 87,308 morpheme-adding tokens occurred over the whole dataset (25% of tokens). By types, 2,555 lemmas appeared with morpheme-adding changes at least once (40% of types). Each lemma that appeared with morpheme-adding

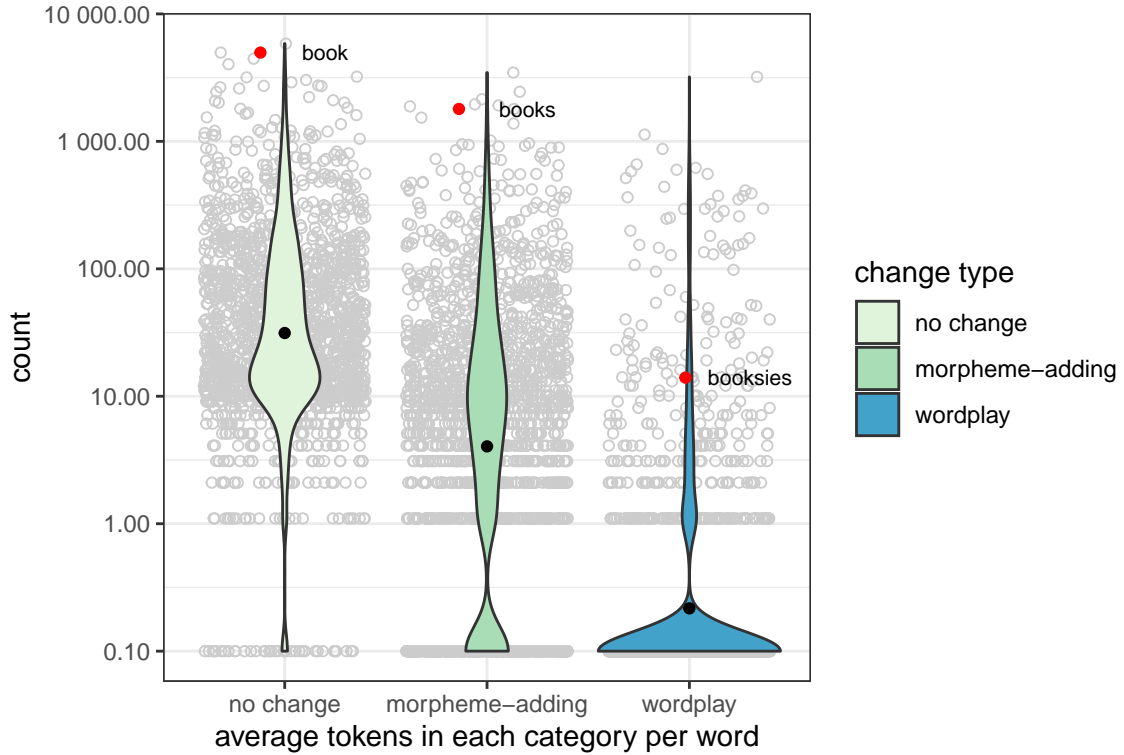


FIGURE 4.1: Violin plot showing the relative frequency of each lemma’s change type (+0.1 to avoid infinite values). Y-axis shows log frequency. Each grey point represents one lemma’s number of appearances with that change type. Black points represent the mean number of of tokens with that change type 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.

changes words did so 27% of the time on average, suggesting that morpheme-adding changes were not particularly rare for lemmas that licensed them.

Wordplay words, like “booksies,” were the rarest in the dataset. By tokens, 18,630 total wordplay tokens occurred in the corpus (5% of all tokens). Wordplay-licensing lemmas appeared with their wordplay forms 8.7% of the time, on average. By types, only 422 lemmas ever showed any wordplay of the 6,342 different lemmas identified in the dataset (7% of lemmas in the corpus). These words were many of highest-frequency lemmas in the corpus. Of the 50 most frequent words in the corpus, 43 showed wordplay.

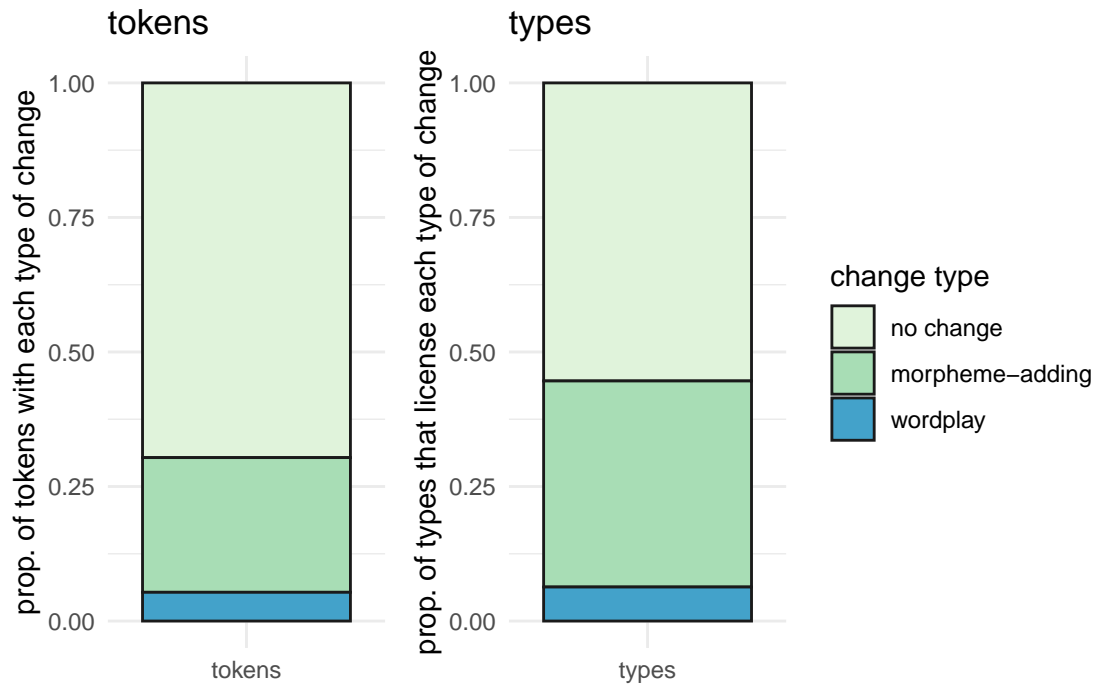


FIGURE 4.2: The relative proportions of each type of change. On the left is the proportion of all tokens in the corpus that appeared with each change type. For example, ‘book’ would be counted in the ‘no change’ segment, ‘books’ in the morpheme-adding segment, and ‘booksies’ in the wordplay segment of 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

Of all the lemmas captured in the dataset, 319 (5% of lemmas) appeared at least once in all three categories: once with no change, once with only morpheme-adding, and 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. Further, changes from the lemma in general were fairly frequent, affecting 30% of the total tokens in the corpus.

### 4.3.2 *Wordplay transformations*

Wordplay took many forms in the dataset, but some transformations were much more prevalent than others (see Table 4.1). The top three alternations occurred in 13,181 tokens (67% of all wordplay tokens). By far the most frequent change both by type and by token was *y*-affixation, 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, but was by far the most prevalent in this corpus. In English, 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. In contrast, the *-let* morpheme (as in “owlet” which means a small owl) actually denotes smallness (Schneider, 2003). *Y*-affixation diminutivization was pervasive and accompanied other changes like pluralization or word shortening 32% of the time.

The second 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 (e.g. “kitty-kitty-kitty”). The third and fourth 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 <sub>1</sub>maca'roni), 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.

Table 4.1: The 8 most common types of wordplay. 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.

Description	Lemma	Example	Token Count	Type Count
y-affixation	dog	doggy	9,592	449
reduplication	kitty	kitty-kitty	2,040	274
weak syllable deletion	vegetable	veggies	1,419	161
shorten	hippopotamus	hippo	1,098	45
simplify	bottle	baba	644	1
vowel change	apple	ooples	578	86
initialism	pajamas	pjs	531	10
euphemism	breast	booby	525	56

### 4.3.3 Predicting Wordplay

In this section we ask why some lemmas have attested wordplay forms while most do not. Wordplay occurrences may emerge in a structured way, which would suggest that caregivers selectively use wordplay instead of changing wordforms at random throughout their speech. To determine whether we can find some characteristics that capture wordplay’s occurrence in the data, we run a logistic regression. For this analysis, we subsampled our data to only include lemmas that occurred 10 or more times in the corpus. This subsampling was to ensure that each word included had a sufficient chance to display wordplay in the dataset, since wordplay occurs on average 9% of the time in words that license it. Therefore, we conducted a logistic regression predicting a binary outcome for each word. Words that had any tokens labeled as wordplay received a 1, while words with no tokens that classified as wordplay received a 0.

Table 4.2: Model estimates from logistic regression predicting wordplay prevalence in lemmas which appeared 10 or more times in the corpus. Length is in phonemes. Freq (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	-2.166	0.245	-8.835	<0.001***
freq	0.003	0.001	2.496	0.013*
MCDI	1.443	0.517	2.793	0.005**
length	-0.061	0.042	-1.441	0.15
freq*MCDI	-0.003	0.001	-4.030	<0.001***
freq*length	0.001	0.000	3.518	<0.001***
MCDI*length	-0.079	0.104	-0.755	0.45

Predictor variables were:

- If the word was early-learned, operationalized as presence on either the Words and Gestures or Words and Sentences forms of the MCDI.
- Each lemma’s total token count, operationalized as overall tokens of any word-form that refers to that lemma in the corpus, collapsing across all months and families.
- Each lemma’s length, operationalized as the number of phonemes in the lemma according to the CMU pronunciation dictionary. Lemmas which were not in the CMU dictionary (e.g. “allosaurus”) were excluded from the analysis.

All two-way interactions were also included. Three-way interactions were not calculated. The final model was  $\text{Wordplay} \sim (\text{word frequency} + \text{MCDI} + \text{length})^2$  and included 1,707 lemmas.<sup>1</sup> Note that the exponent in the logistic equation does not actually square any of the terms but instead instructs the model to include all possible two-way interactions without calculating the three-way interaction.

<sup>1</sup> Conclusions are the same when we expand the dataset to include lemmas that occurred 5 or more times. Because we artificially removed the long tail of the data’s word frequency range by subsampling based on frequency, we do not log transform any of our predictors in this model.

The output of the logistic regression is presented in Table 4.2. The intercept shows that the log odds of a lemma exhibiting wordplay are low, corresponding to a base rate of  $\sim 10\%$  when all other predictors are set to 0. The significant main effect of frequency indicates that as a word increases in frequency, it also becomes more likely that the lemma will license wordplay. The significant main effect of MCDI shows that words which are present on the MCDI are more likely to license wordplay. Word length was not a significant main effect.

Two interaction terms were also significant (see Figure 4.3 for visualizations of the interactions). First, there was an interaction between word frequency and its presence on the MCDI; when nouns were on the MCDI, they were much more likely to license wordplay if they were higher frequency. In contrast, for words that weren't on the MCDI, frequency played less of a role in distinguishing wordplay words from other words. Next, there was a significant interaction between word frequency and length, where if a word was both short and high-frequency, it was much more likely to license wordplay than a word that was just one or the other. In sum, words had a much higher chance of licensing wordplay if they were more frequent, on the MCDI, and shorter all at the same time.

#### 4.3.4 *Wordplay does not vary as a function of 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

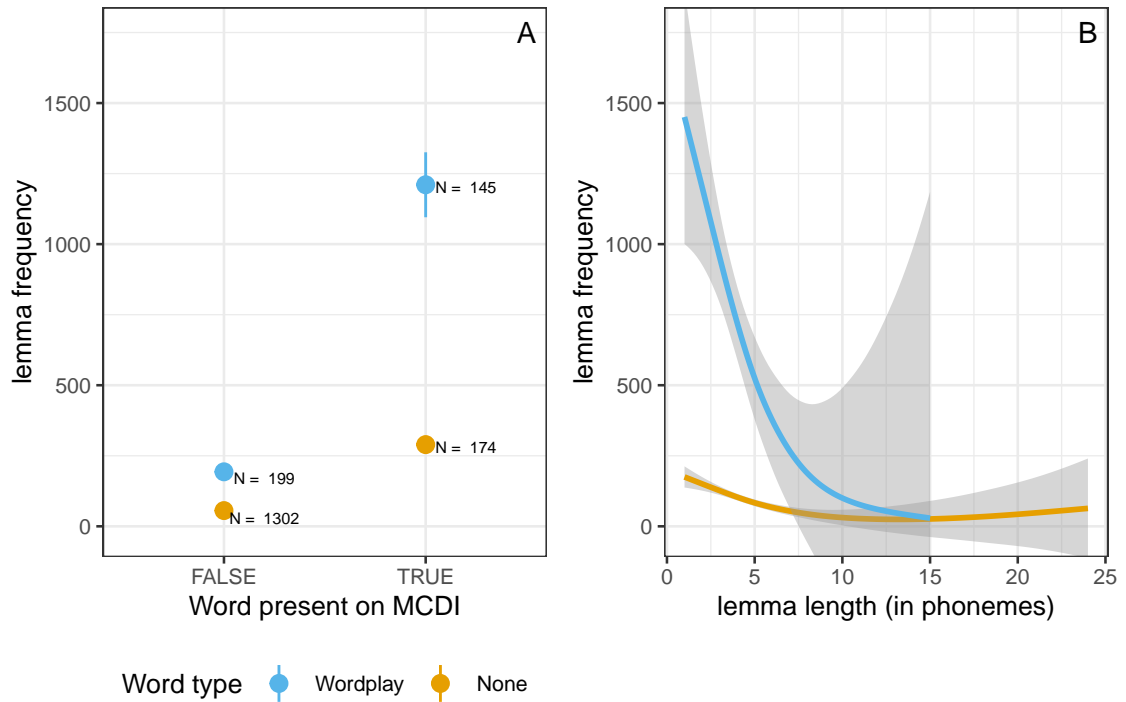


FIGURE 4.3: Visualizations of the significant interactions from the logistic regression. Panel A shows how a word’s presence on the MCDI (X-axis, FALSE means the word is not on the MCDI) interacts with the word’s frequency (Y-axis) to predict if the word licenses wordplay. Pointranges represent the 95% confidence interval. The number of observations represented by each pointrange is presented on the figure. Panel B shows the interaction between word length (X-axis) and frequency (Y-axis) to predict whether the word licenses wordplay. Confidence bands represent the 95% confidence interval.

be more talkative. Thus, even a metric of wordplay-per-hour would be affected by this sampling method. Second, the language density in video is significantly higher than in audio recordings (Bergelson, Amatuni, Dailey, Koorathota, & Tor, 2018), meaning that videos will have an outsize 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. Since all videos are annotated for the full hour, we have access to the same quantity and quality of data consistently across ages with this sampling method. See

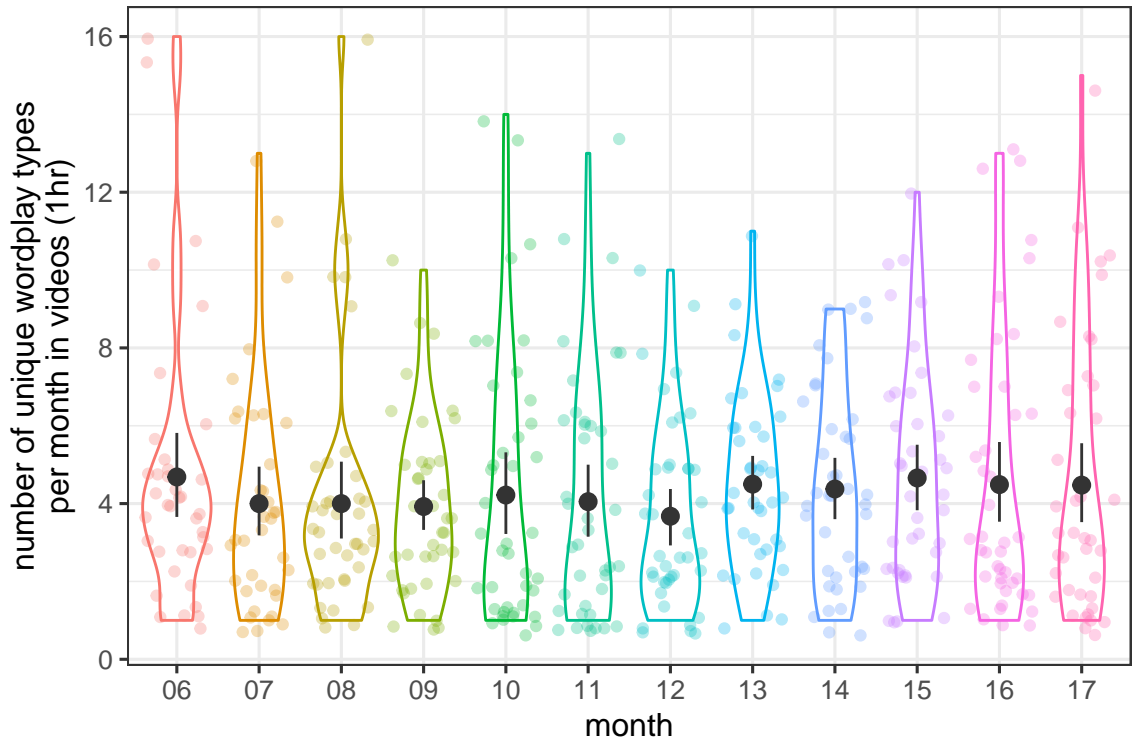


FIGURE 4.4: Violin plot showing the effect of infant age on wordplay frequency. X-axis shows the age of the infants in months. Y-axis shows the number of unique wordplay types that each infant heard in the one-hour video recording taken each month. Individual points represent the total for one infant in that month. Violins show the distribution. Pointranges show the mean and the 95% confidence interval

Figure 4.4 for a visualization of the distribution of wordplay types across time for each family.

When looking at only the collected videos, we find no relationship between children’s age and the amount of wordplay they hear (types:  $r = 0.04$ ,  $p = 0.39$ , tokens:  $r = -0.01$ ,  $p = 0.76$ ). We thus collapse across time for family-level analyses.

#### 4.3.5 Wordplay across families

To examine whether the amount of wordplay that infants heard varied between families, we conduct a series of correlations. We begin by correlating the number of wordplay tokens that each child heard with the number of tokens they heard that

had no changes. Tokens that had no changes were always in the conventional, lemma form. These provide a measure of frequency to correlate with that does not include any tokens with changes to the wordform, thus avoiding an artificially inflated correlation coefficient by correlating one measure with another measure that contains it.

We find a positive correlation between wordplay and the number of tokens children hear that match the lemma ( $r = 0.46$ ,  $p < .001$ ). Children who hear more wordplay types also hear more word types in total ( $r = 0.57$ ,  $p < .001$ ). These two positive correlations suggest that children who hear more words also hear more wordplay.

While raw counts can be informative, counts are best interpreted in the context of a base rate, as is supplied by proportions. Wordplay as a proportion of total input tells a different story. If wordplay simply scales with input, we should expect to find no correlation between the amount of speech children hear and the proportion of that input that constitutes wordplay, because the proportion should be the same in every family. In other words, if we expect one in every  $\sim 10$  words to be wordplay in all infant-directed speech, as suggested by summary statistics at the word level, then the proportion of wordplay should be roughly 10% in all families, regardless of the total amount of input.

Instead we find negative relationships, statistically significant both for types ( $r = -0.35$ ,  $p = .021$ ) and for tokens ( $r = -0.34$ ,  $p = .023$ ). This relationship suggests that families who provide a higher quantity of input to their infants use less wordplay than we would expect if wordplay scaled perfectly with the total amount of speech in the home, and that families who use more types (i.e. more lexical diversity) around their infants similarly use fewer wordplay forms. See Figure 4.5.

How can wordplay both increase as a function of total input and also represent less of the total input for families who talk more? To make sense of this seemingly

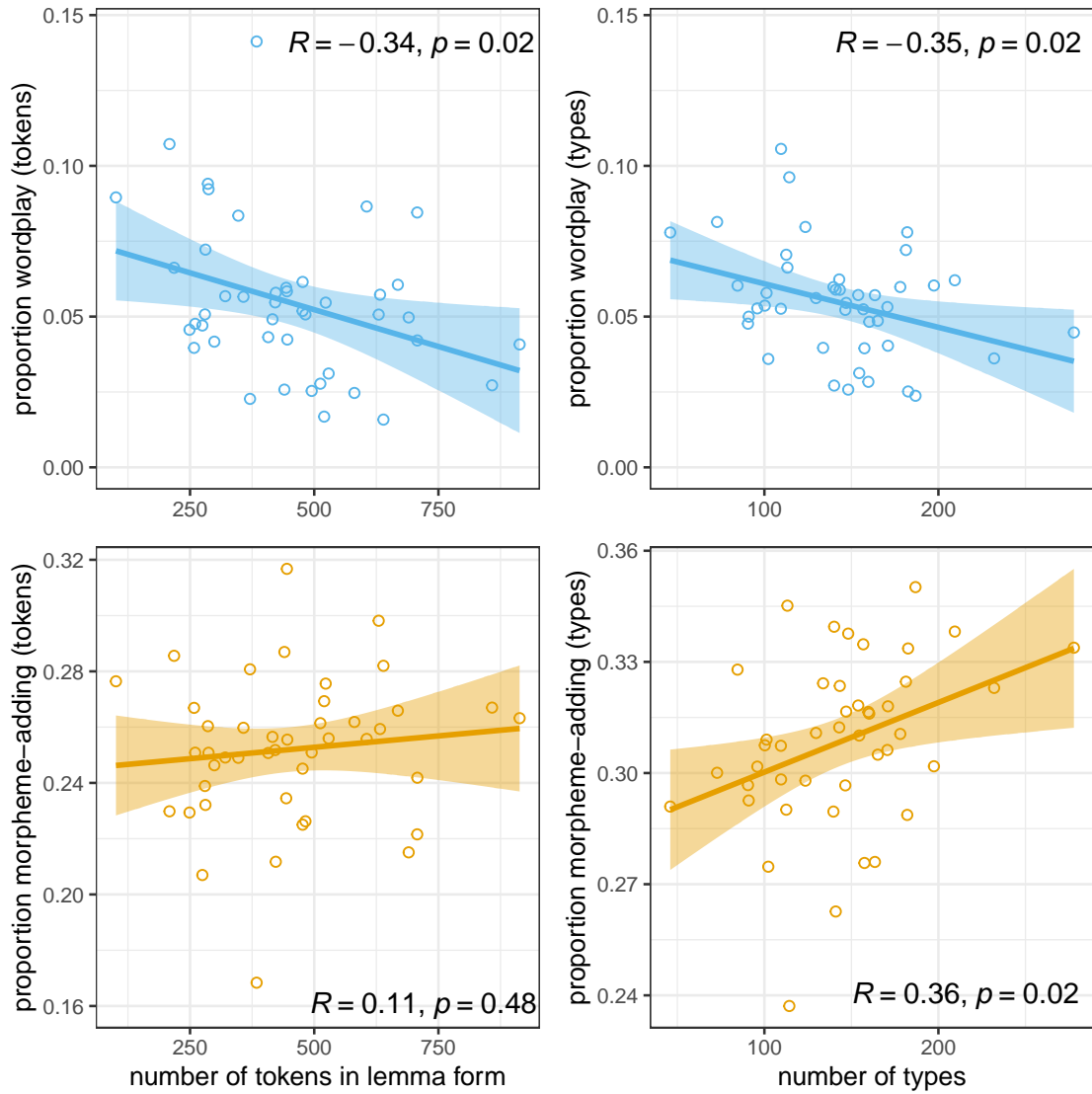


FIGURE 4.5: Correlations between proportions and total input. Top row depicts correlations with wordplay. Bottom row depicts correlations with morpheme-adding words. Left column shows correlations by tokens, where the x-axis depicts the number of tokens that each infant heard where the lemma matched the wordform. Right column shows correlations by type, where the x-axis depicts the number of unique lemmas the child heard. Each individual point represents one infant’s values, averaged across time. Confidence bands around the regression line show the 95% confidence interval.

paradoxical finding, we can contrast the wordplay findings with the same analyses for morpheme-adding changes. Here we find that average total monthly input and average total morpheme-adding words are correlated so strongly as to be markers of the same construct ( $r = 0.95$  for tokens,  $r = 0.96$  for types). In this context, our wordplay-to-input correlations are relatively modest, suggesting a decoupling of wordplay and total input in a way that differs from how morpheme-adding words are distributed. Further, there is no significant relationship between the proportion of morpheme-adding words each child hears and the number of words they hear with no changes from the lemma for tokens ( $r = 0.11$ ,  $p = 0.48$ ), just as we would expect if one metric scales linearly with the other, and again contrasting with what we find for wordplay. For types, we find a significant positive correlation between the number of unique morpheme-adding types and the number of lemmas ( $r = 0.36$ ,  $p = .02$ ), suggesting that a larger diversity in words also results in a larger diversity of morpheme-adding wordforms for those words.

Taken together, these findings show that the proportion of total input that is classified as wordplay decreases as infants hear more words, whereas the proportion of morpheme-adding words in the input either stays consistent between families or increases as a function of total input. See Figure 4.5.

#### 4.3.6 *Wordplay and learning*

We next 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 a mixed-effects linear model predicting the earliest age that words are learned and said by the infants in our sample.

Our outcome variable for the mixed-effects linear models use each child's MCDI results for each word as the dependent variable. The predictor variables count the

Table 4.3: 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	16.50	0.40	40.81	254	<0.001***
unique wordforms	0.79	0.14	5.70	5347	<0.001***
log frequency	-0.47	0.05	-10.40	5348	<0.001***
interaction	-0.13	0.02	-7.30	5346	<0.001***

number of unique wordforms each lemma appears with and the total log frequency of the lemma. If an infant never understood a particular word, or never heard it in the corpus, then the datapoint is excluded from analysis. We run separate models for comprehension and production. Our final dataset included learning data for 204 distinct words.

The model for comprehension contained 5,390 observations, with each individual observation representing one child’s minimum age of comprehension for one word. We began with a base model that predicted parent-reported comprehension on the MCDI based on the word’s log frequency in the corpus alone. We also included a random intercept for each child. Adding the number of unique wordforms as a predictor significantly improved model fit ( $\chi^2 = 168.49$ ,  $p < .001$ ), suggesting that wordform variability is a meaningful predictor of word learning above and beyond lemma frequency. A summary of the fixed effects can be found in Table 4.3.

The model for production contained 1,460 observations. This is many fewer observations than in the comprehension model because children’s productive vocabularies were smaller than their receptive ones throughout the study. The base model was the same as above but predicted parent-reported production on the MCDI instead of comprehension. Adding the number of unique wordforms as a predictor significantly improved model fit ( $\chi^2 = 59.58$ ,  $p < .001$ ), suggesting that wordform variability

Table 4.4: 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	16.69	0.43	38.41	591	<0.001***
unique wordforms	0.68	0.14	4.77	1434	<0.001***
log frequency	-0.05	0.05	-0.83	1431	0.405
interaction	-0.10	0.02	-5.58	1435	<0.001***

is a meaningful predictor of word learning above and beyond lemma frequency. A summary of the fixed effects can be found in Table 4.4.

For the comprehension model, both the lemma’s log frequency and the number of wordforms used to refer to the lemma were significant main effects. For the production model, only the number of unique wordforms were a significant main effect. However, in both models, frequency and wordform variability were involved in a significant interaction. Visualizations of the interactions are presented in Figure 4.6. The effect of wordform variability was dependent on the frequency of the word, with more wordform variability contributing to earlier production and comprehension in high-frequency words, but having less of an effect in lower frequency words. This suggests that wordform variability may be helpful for word learning 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 may have no effect or slow uptake when words are lower-frequency, particularly in production.

#### 4.4 Discussion

In this paper, we defined wordplay as wordforms where both the meaning and phonology of two wordforms overlap. We explored the prevalence of wordplay in a longi-

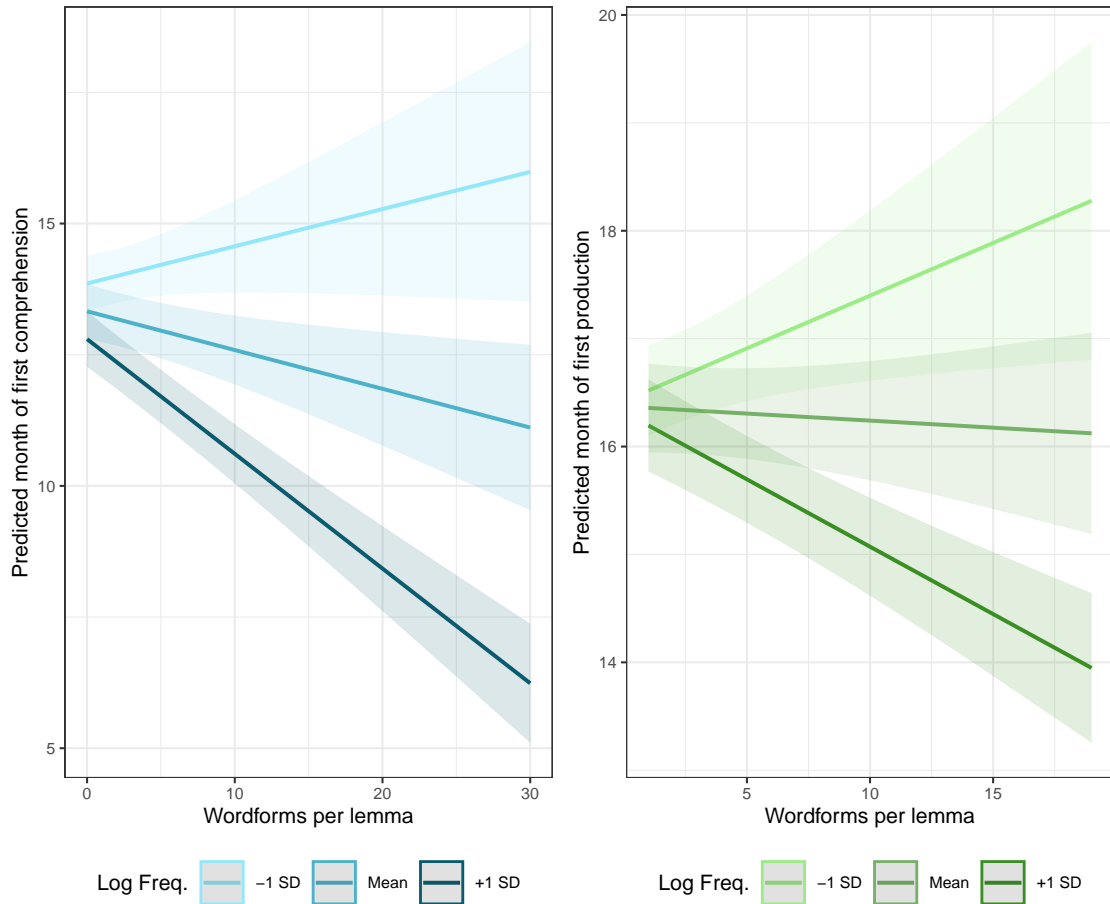


FIGURE 4.6: Plots showing the predicted effects of lemma frequency and the number of attested wordforms per lemma on when infants are reported to understand (left) or say (right) words on the MCDI:WG. Lines represent the predicted data for the mean frequency value, one standard deviation above, and one standard deviation below the mean.

tudinal corpus of infant-directed speech. We showed that while wordplay was a heterogeneous phenomenon with many different forms, some changes were prevalent. Word shortening, y-affixation, and reduplication were the most frequently observed. We found evidence of wordplay in every family to varying degrees (from 2% to 11% of the unique noun types in the input per family), suggesting that wordplay is an established feature of infant-directed English. While wordplay was low-frequency overall, wordform variability (of which wordplay is a part) did predict when infants learned

individual words. High-frequency words which had more wordform variability were learned and spoken earlier than words with fewer wordforms by 18 months.

#### *4.4.1 Wordform variability boosts learning*

Despite the literature showing that synonyms are a challenge to learn, we find here that wordform variability aids infants' learning for the highest-frequency words in the input. Why might this be? One possibility is that wordform variability acts in much the same way as the other highly variable features of infant-directed speech. Just as others have found large degrees of variability in infant-directed pitch contours (Fernald & Simon, 1984) and vowel realization (Miyazawa, Shinya, Martin, Kikuchi, & Mazuka, 2017), high wordform variability may similarly help capture infants' attention, thus leading to better learning. However, structural features of infant-directed speech may not be as attention-grabbing as the prosodic features (Segal & Newman, 2015).

Another 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. Evidence for this comes from failure-to-discriminate findings (Stager & Werker, 1997), as well as evidence that in lexical classes with higher wordform variability, similar wordforms are not discriminated until much later (See Chapter 4). Counter-evidence comes from a robust literature showing that infants do in fact build robust representations of nouns by 12 months (e.g. Mani & Plunkett, 2010; Von Holzen & Bergmann, 2021).

How can infants learn that “dog” and “doggy” are not meaningfully different while “red” and “ready” are? Infants may be able to tell that wordplay wordforms carry overlapping meaning with conventional forms because the contexts of usage for both wordforms would be very similar, both in the world and in language. Learners may hear about “dogs” and “doggies” both in the presence of dogs, leading infants

to map both wordforms onto the concept. The language context is also a helpful cue, as in semantic networks used in natural language processing where the surrounding linguistic context is used to build word meanings (e.g Mikolov, Chen, Corrado, & Dean, 2013). Since “dogs” and “doggies” can be used in identical linguistic contexts, infants may be able to use that information to discern that they are variants of the same word.

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. Different wordforms of the same lemma occur more tightly clustered 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/). Both of those 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 be a key to understanding which wordforms are variants of each other.

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

Wordplay consisted of mostly reduplication as in “car-car,” addition of the -y morpheme as in “birdy” and “doggy,” as well as word shortening, like in “cado” for *avocado*. These wordform variants may be particularly common because they provide infants with useful information. Reduplication gives listeners a second chance to hear the sounds in a wordform through repetition. Adding a diminutive suffix like “-y” to the ends of words regularizes the right edge of the wordform, potentially mak-

ing it easier for infants to find word boundaries. Shortening words makes them easier to produce since there is less to articulate, possibly making them more accessible to infants. Reduplication and y-affixation have both been previously associated with better learning outcomes (Ota, Davies-Jenkins, & Skarabela, 2018), suggesting that their presence in infant-directed speech is beneficial to listeners. This is compatible with our findings that wordform variability in general is beneficial for learning.

We did find that wordform variability only led to earlier word learning when words were highly frequent, suggesting that infants may need access to many tokens of a lemma in addition to wordform variants in order to build representations. This hypothesis needs to be tested in a more confirmatory study. Wordform variability did not have a detectable negative effect on word-learning, suggesting that even if it is not helpful, wordform variability does not block infants' learning of words.

## 4.5 Conclusion

Wordplay is a feature of infant-directed speech that contributes to wordform variability. Despite the fact that wordplay creates multiple wordforms that map onto a single lemma, infants seem to learn just as well in families with high levels of wordplay as in low levels. Further, we find evidence that for the most frequent words in the corpus, wordform variability helps infants learn and say words earlier. Taken together, these findings suggest that wordform variability is not a problem for infants and instead may be helpful in early word learning.

## Conclusion

This dissertation explored the messy correspondence between sounds and their meanings from two complementary perspectives. We zoomed in on the moment-to-moment unfolding of word comprehension using eyetracking methods, and we took an expansive look at learning over the course of a year using corpus analysis and questionnaires. The findings of each study highlight the relationship between the sounds in a word and higher levels of representation, like pragmatics, syntax, and semantics. Each study provides insight into a different aspect of wordform variability.

### 5.1 Summary of Studies

In Chapter 2, we asked whether toddlers accurately represented the sounds in verbs on the same timeline as nouns. Additionally, we asked whether the sounds in irregular nouns and verbs are represented with the same specificity as the sounds in regular nouns and verbs. Verbs were of interest because previously, the bulk of the literature on sound representation had investigated only concrete nouns. Irregulars were also examined because they have multiple correct wordforms; irregulars in English have multiple nuclear vowel alternations that young children need to learn, poten-

tially without an understanding of tense or plural marking. We began by looking at the macro scale: a corpus analysis confirmed the intuition that irregular verbs are far more common than irregular nouns in speech to young children, meaning that toddlers have to contend with wordform variability more in the case of verb representation than in noun representation. This sound difference between syntactic categories meant that toddlers may have used different strategies to represent verbs than nouns, to accommodate the high prevalence of irregulars.

With the knowledge that toddlers hear more irregular verbs than nouns, we tested toddlers' comprehension and wordform knowledge in a suite of eyetracking experiments. We found that a word's syntactic category had a significant effect on whether toddlers understood the word, and consequently whether they noticed mispronunciations. At 18 months, toddlers did show a mispronunciation effect for irregular nouns, looking less to a goose when they heard "giss" than when they heard "goose." However, toddlers did not alter their looking behavior at all when they heard mispronounced verbs of either type, looking equivalently when they heard "run" and "roan." We thus additionally studied two-year-olds, finding that once toddlers showed evidence of robust verb comprehension, they also demonstrated well-specified phonological representations, just as they did – months earlier – for nouns.

Regularity had some effect for noun representations, with participants looking slightly more to the target when they heard regular nouns like "pig" compared to trials with irregular nouns like "mouse". In contrast, regularity had no detectable effect on verb representations once participants showed robust verb comprehension at 26 months. This pattern suggests that toddlers did not struggle to form well-specified phonological representations for verbs whose vowels change relative to regular words in the same syntactic category. The findings from this study suggest that while the sounds in one lexical class can be represented months before the sounds in another,

a similar delay is not observed for words with multiple correct surface forms.

In Chapter 3, we delved into an even finer level of sound representation when we examined the role of sub-phonemic cues on word comprehension in adults. Here we asked whether the phonology of an unfamiliar word interacted with adults' use of the mutual exclusivity constraint, and whether adults take the visual context into account when listening to coarticulation cues. To do this, we took familiar words and changed the sounds they contained. We either created novel words by changing the last phoneme in the word, or created edited versions of the familiar word where the coarticulation cues in the vowel of the word and the final consonant of the word did not match. These edited stimuli helped us determine the role of coarticulatory cues more readily than would be possible with un-edited stimuli, since coarticulation cues always match the phonemes in naturally-produced speech.

Participants' looking behavior showed that adults looked more to the unfamiliar object on trials where they heard a coarticulation mismatch than they did when cues matched. This suggests that listeners integrated the visual context when parsing coarticulatory cues – the word's coarticulation cued listeners to an unfamiliar object. We additionally found a bimodal distribution of behavior in the unfamiliar word condition. Roughly half of participants did not make use of the mutual exclusivity constraint. This pattern suggests that the border between perceiving stimuli as a new wordform or a highly deviant version of a known wordform is flexible based on the context in which the word is heard.

In Chapter 4, we zoomed out from the micro-scale and looked at both the prevalence and consequences of wordform variability in speech to infants. We found evidence that wordplay, i.e. unconventional wordforms that overlap with conventional ones in both sound and meaning, occurred in every household to varying degrees. Wordplay introduces unpredictable variability into the wordforms used to denote some early-learned, highly frequent concepts in infants' lives. However, we also find

that this increased variability either has no effect on a word’s learnability, or is actually associated with earlier word learning. Higher wordform variability was associated with an earlier age of acquisition for high-frequency words, and there was no evidence of negative effects of wordform variability at either the word or the family level. This suggests that wordform variability may help learners understand words on the macro time scale.

Synthesizing across the two chapters focused on toddlers in the dissertation, we find that wordform variability does not delay word learning for the most frequent words that infants and toddlers hear. In Chapter 2 we saw that irregular verbs were highly frequent in infant-directed English, and were phonologically represented on a similar timeline to regular verbs. Further, we saw that irregular nouns were much more uncommon than regular nouns. Despite this difference in the relative frequency of irregularity in each syntactic category, nouns were represented robustly by 18 months, though toddlers looked less at them overall, likely due to their lower frequency. In Chapter 4, we saw that the lemmas with more wordform variability were also learned earlier when it came to highly frequent nouns.

## 5.2 How learners accommodate wordform variability

Why doesn’t a word with high wordform variability enter the vocabulary later or show a delay in the timeline of its phonological representation? It may be tempting to suggest an explanation for this pattern in toddlers that aligns with holistic theories (Charles-Luce & Luce, 1990, e.g. 1995; Hallé & de Boysson-Bardies, 1996): maybe toddlers cannot detect the differences between wordforms during comprehension because early wordforms are under-specified. However, extensive evidence supports the existence of high-fidelity sound representations early in development (Von Holzen & Bergmann, 2021; White & Morgan, 2008; Zamuner, Moore, & Desmeules-Trudel, 2016). We thus move forward with the assumption that listeners can and do detect

the wordform variability they hear. We present a few possible accounts to make sense of learners' incredible robustness to wordform variability.

First, it is possible that wordform variability is not a problem for word learning because infants always expect meaningful variations, as proposed by (E. V. Clark, 1987), but are not immediately concerned with learning the underlying meanings of those variations. For example, children may expect that “eat” and “ate” share some similar meaning because of their overlapping sounds and the crucial fact that both words are used in similar contexts. Learners might then go on to assume that the difference in vowels between “eat” and “ate” corresponds to a difference in meaning, even if infants do not yet know what that meaning difference is. The same type of thinking would similarly apply in contexts with wordplay, where the mapping between changes in the wordform and changes in meaning is less clear. Infants could merely assume that a difference in meaning will become apparent and thus learn the separate wordforms, even if the meaning changes do not subsequently reveal themselves. The results of Chapters 2 and 4 support this possibility, since learners accommodated wordform variability in the input.

An alternative possibility is that when faced with wordform variability, the expectation of one-to-one relationships is overridden by other, more salient, information. Wordform variability is only theoretically an issue because it flies in the face of easy one-to-one concept-word relationships, but it is possible that one-to-one relationships are only the expectation in the absence of any other cues to word meaning. Previous work supports this point of view in the case of homophony (Dautriche, Chemla, & Christophe, 2016). Preschoolers and adults apply the same novel label to two distinct sets of animals when the novel word was taught with a bimodal distribution of referents, e.g. during training, the same word was used to label both monkeys and snakes, creating a novel homophone (Dautriche, Chemla, & Christophe, 2016). This finding suggests that when there is sufficient evidence from the learning environment

that a word has multiple meanings, learners across the lifespan can accommodate a many-to-one relationship between meanings and their wordforms.

The work presented in this dissertation is also consistent with a model of word learning where contextual information like the visual scene and social information like joint attention are more highly-valued cues to a wordform's lemma. Under this view, infants can override their baseline preference for one-to-one lemma-to-wordform mappings if the learning context provides sufficient additional support for a many-to-one correspondence. This possibility would explain the findings of Chapters 3 and 4, which provided evidence for significant limits on the use of mutual exclusivity. In Chapter 3, half of the participants didn't use mutual exclusivity even though under the traditional view of mutual exclusivity, its use would be helpful in learning a name for the unfamiliar object. In Chapter 4, infants learned words with higher levels of wordform variability earlier than words with fewer wordforms, suggesting that the differences between wordforms did not delay infants in learning the word.

An additional limitation on learners' use of mutual exclusivity may emerge when words are too phonologically similar to one another. Wordplay words in Chapter 4 and the minimal pairs presented in Chapter 3 both shared significant phonological content with their conventional forms. This provides evidence that learners can deactivate the mutual exclusivity constraint in word learning contexts where novel words are similar to known words, instead assuming that words that sound too similar are more likely to stem from the same lemma. This proposal rests on a theory of mutual exclusivity where listeners begin with the assumption that all sound changes correspond to meaning changes, and thus the mutual exclusivity constraint should automatically be applied as a baseline.

This strategy would solve a lot of the problems in navigating wordform variability, allowing learners to quickly bundle similar-sounding words together with their uniting lemma. One limitation to this strategy is that it would also block the uptake of

novel words with overlapping phonology to known words that are not part of the same lemma. For example, if a learner hears “goose” and “geese” and determines that these words are sufficiently similar that the mutual exclusivity constraint should not apply, then we should find evidence of children mapping “gas” to the lemma for goose as well, since the phonological distance between these words is the same. Thus, this account requires that word learning is supported via additional processes that incorporate top-down, contextual information. Taking our example again, learners are much more likely to hear both “goose” and “geese” in a context where the birds are present e.g. a park, while “gas” is likely to emerge in very different circumstances e.g. during a drive.

Further evidence that the mutual exclusivity constraint is only activated in appropriate circumstances comes from Weatherhead et al. (2021). They found that both monolinguals and bilinguals took the context (in this case the social group of the speaker) into account when deciding whether a novel word was a synonym for a known object or a word for a feature of that object. For example, 18-month-olds mapped “zabe” to the known object *dog* when the speaker was of an unfamiliar race, but mapped “zabe” to the color of the dog when the speaker was of a familiar race. This pattern was the same for monolinguals and bilinguals, suggesting that learners use mutual exclusivity when they are confident that the speaker is from the same language community as them, but less likely to use it when the speaker may be from a different language community. This study, in concert with the findings presented here, suggest that mutual exclusivity is simply one tool of many that learners have access to when learning words, and can be used judiciously even in toddlerhood.

It is evident that higher levels of processing like syntax and pragmatics interact with phonology, even on the macro scale. Chapter 2 shows this clearly where toddlers developed strong wordform representations, but instead of showing evidence that this occurred over the whole lexicon, phonological representation appeared to

proceed one syntactic category at a time. In other words, phonological representation did not develop monolithically for all wordforms as one would expect if phonology operated independently from syntax. Similarly, Chapter 3 showed that the effect of coarticulatory cues was mediated by the objects present in the visual scene, suggesting that phonetic cues were being actively integrated with visual cues during comprehension. The corpus created in Chapter 4 provides rich data to potentially explore a theory where extra-linguistic cues like temporal closeness may help guide infants to connect the appropriate wordforms with their lemmas, as in Peperkamp, Hegde, & Carbajal (2019). In sum, we provide a view from two timescales of both the amount of wordform variability that exists and the ways that syntax and social reasoning may aid listeners in understanding words despite it.

### 5.3 Future Directions

In Chapter 2, evidence suggests that shortly after their second birthdays, toddlers can tell when familiar irregular nouns and verbs are mispronounced with a vowel that never occurs in that word in naturalistic speech. However, this finding does not indicate whether infants recognize both correct vowels in those words as acceptable. From our current data, we do not know whether toddlers would only accept the most common form of an irregular noun or verb as the correct form. Do toddlers look as much to “mouse” as they do to “mice?” In order to get traction on this question, we propose a follow-up study to examine whether toddlers accept the attested vowels in irregular words as correct wordforms. For example, in the future we could add a new condition to the existing study using phrases like “look at the mice” for a single mouse or “she’s gonna ran” with the verb stimuli. If we find that toddlers spend equivalent amounts of time looking at the target in both of these conditions, then we can conclude that toddlers know which vowels are acceptable in each word, but not the contexts in which they are acceptable. In contrast, if toddlers treat the attested-

but-ungrammatical vowels as mispronunciations, then we know that toddlers are also tracking the appropriate grammatical context of irregular wordforms. This experiment would thus help us to determine if toddlers know what the correct vowels are in irregular words. It would also help us discover whether toddlers understand when each vowel is appropriate, providing more direct evidence that toddlers are in fact representing the phonology of both forms of these words.

Along the same lines, Chapter 3 raises questions about the specific circumstances that give rise to the effects we observe in the unfamiliar word condition. Is there a certain degree of phonetic distance required between the known word and the unfamiliar word before mutual exclusivity begins to function as expected for all participants? Finding that inflection point would be useful for understanding the relative contributions of visual context and phonetic information. It would also be interesting to see if we still find that listeners still ignore the phonemic difference between the novel and unfamiliar words in a study without coarticulation mismatches. It's possible that by introducing an intermediate condition (e.g. trials like "cri(g)b") between the two control conditions ("crib" and "crig"), participants were simply more confused than they ever would be when listening to naturally produced speech. While this proposal seems unlikely due to the highly variable nature of speech in the wild, it would be worth ruling out with empirical evidence.

One additional unanswered question raised by this dissertation is how infants distinguish between cases where they are encountering the wordform for a new lemma and cases where they are learning an additional wordform for a known lemma. For example, how do infants know to group "eat" and "ate" together but not "oat?" Many possible cues exist for distinguishing these scenarios: visual context, knowledge of morphosyntactic processes transforming one wordform into another, the surrounding words, whether or not the conventional form of the lemma has been used in the same conversation, etc. In order to determine which of these cues matter and their

relative importance as learners build their vocabularies, more research is needed at both time scales. The role of visual context is particularly well-suited to examination through eyetracking studies with infants. In a future study, we would ask where infants or toddlers draw the line between a variant wordform of a known word and a novel word, just as we did with adults in Chapter 3. To test this, participants could inspect scenes containing known objects and novel objects. Then participants would hear either correct labels for the familiar objects, or unattested but plausible wordform variants of the conventional name for each known object. If toddlers mapped these novel wordforms onto novel objects, we would be able to conclude that phonology alone is sufficient to distinguish one word from another for toddlers. If we find that toddlers instead select the familiar objects, we would have evidence that toddlers, like half of adults in Chapter 3, assume that the speaker meant the familiar object, grouping the variants together with one single lemma. The results of an experiment like this could shine light on the mechanisms that protect infants from confusion during wordplay.

### *5.3.1 Limitations*

While this work examines wordform variability from three different perspectives, it is still limited in its scope. First, all of our studies were conducted exclusively in English, a language where words are easy to delineate and categorize cleanly into different lexical classes. English is over-represented in language development research, and findings in English findings may not replicate in other languages, particularly polysynthetic languages like Mohawk where the semantic content of an entire English clause can be conveyed within a single “word” (Kelly, Wigglesworth, Nordlinger, & Blythe, 2014). The consequences of wordform variability in a language with more productive combinatorial morphology is demonstrably different (Xanthos et al., 2011). analyzed child-directed speech in 9 languages that vary in the morphological com-

plexity of their noun and verbs. They find that in languages with more variable morphology, children produce more morphologically rich utterances earlier, a finding that highlights the potential benefits of wordform variability. A future direction of the current work entails a thorough examination of the equivalent phenomena in other, genetically distinct languages and language families, since language of exposure has already been shown to interact with rate of acquisition (Xanthos et al., 2011).

Along the same lines, cultural considerations need to be taken into account. All of these studies were conducted with primarily affluent, primarily white participants in the United States, a demographic that is not representative of all language learners, even when focusing exclusively on the acquisition and comprehension of English. Considering the role that social reasoning and visual context appear to play in word representation and comprehension, it would be reasonable to expect that children reared in different social contexts would have different strategies for approaching wordform variability. For example, cross-cultural differences exist around what preschoolers determine to be the salient aspects of a scene (Kuwabara & Smith, 2012) and around rates of overimitation, a marker of cultural transmission (Stengelin, Hepach, & Haun, 2020). It is thus crucial that future work expand beyond the confines of a white, educated, industrialized, rich, and democratic sample to more accurately capture how word learning and comprehension functions for all learners, not merely for the most privileged among us (Henrich, Heine, & Norenzayan, 2010).

In addition to looking at other languages, these same questions also need to be examined from the perspective of a multilingual learner. Bilinguals have already been shown to use less mutual exclusivity than monolingual learners do (Byers-Heinlein & Werker, 2009), lending additional credence to the theory that mutual exclusivity is used judiciously. However, bilingual learners need to solve the same problems of wordform variability that monolinguals do. Furthermore, bilinguals need

to manage this while acquiring two separate phonological inventories at the same rate that monolinguals develop one (Fabiano-Smith & Barlow, 2010), further increasing phonological variability. Future work should address the same questions we asked in this dissertation with bilingual samples.

In a bilingual replication of Chapter 2, we would be able to determine whether bilinguals have a similar timeline to monolinguals for irregular word comprehension by testing bilingual infants of the same age. We would also be able to assess whether participants' additional language had an effect on this timeline in English. If participants who speak a language with more vowel alternations showed robust phonological representations earlier, then we would have additional reason to conclude that exposure to variability speeds learning, as in (Xanthos et al., 2011), but across linguistic barriers.

A bilingual replication of Chapter 3 would clarify whether bilingual adults show the same bifurcation in approach that we observed in monolingual adults when faced with a minimal pair where one of the words was novel. Based on multilingual toddlers' low rates of mutual exclusivity usage in early childhood (Byers-Heinlein & Werker, 2009), we predict that mutual exclusivity rates would be lower in the bilingual participants than in our monolingual participants. The results of this followup study would provide insights into whether the trends in word learning strategies that bilingual toddlers use continue into adulthood, or whether monolinguals and bilinguals converge on their strategies as they become adults.

Finally, a replication of Chapter 4 in bilinguals would clarify how much more wordform variability there is for each lemma in bilingual homes compared to in monolingual homes, and the effects that has on learning each lemma. In Chapter 4, we found that wordform variability within a single language could fruitfully predict when a monolingual infant learned a word. In a corpus of bilingual infants, we would be able to predict lemma learning based on not only within-language variability

but also between-language variability. For example, do French-English bilingual infants learn “dog” earlier when they hear more of both “doggy” and the translation equivalent “chien?” This finding would provide us with insight into the transfer of knowledge between languages as bilinguals build their lexicons.

## 5.4 Conclusion

When it comes to the nouns and verbs that young children learn, a neat correspondence between wordforms and their meanings is the exception, not the rule. In this suite of studies, we found that words with more phonological variability brought on by morphological changes did not delay toddlers in representing the sounds in those variable words (Chapter 2). We found that in adults, subphonemic mispronunciations provided a sufficient signal to listeners that they considered an alternative referent for familiar words (Chapter 3). Finally, we found that single lemmas that are represented with multiple different wordforms were learned earlier than words with fewer wordforms, at least for the highest-frequency words in the corpus (Chapter 4). Together, these findings show that instead of muddying the waters for learners, wordform variability is at worst neutral and at best beneficial for comprehension. We discovered that learners are robust to variability, and listeners can use it to correctly identify target words in real time. This research program opens many exciting research avenues to determine exactly how language users harness variability to support word learning and comprehension throughout the lifespan.

# Appendix A

## Appendix A

### A.1 Corpus Information

This appendix supplements the information provided in Chapter 2. See Table A.1 for a list of all included corpora.

### A.2 Exploratory vocabulary analyses

Correlation tests were conducted to examine potential links between toddlers' vocabulary and the size of their mispronunciation effects (i.e. the difference in looking time between correctly pronounced and mispronounced trials). While the presence of such links in prior work has been mixed (cf. Swingley & Aslin, 2000 and relevant discussion in Von Holzen & Bergmann, 2018), we were curious whether we would find them for the lower-frequency nouns and the verbs we test here.

Using the vocabulary scores we obtained from parents' reports on the MCDI, we measured Pearson correlations between mispronunciation effect size based on receptive vocabulary (Exp. 1 and 2a) or productive vocabulary (all Experiments). *N.B.* This exploratory analysis was preregistered for Experiment 1, but we extend it

Table A.1: Corpora included in the corpus analysis. ‘Corpus’ lists the corpus name. ‘Children Included’ lists the total number of children whose data were included per corpus. Note that the children’s own productions were excluded from analysis. ‘Tokens per Corpus’ provides the token counts. ‘Nouns’ and ‘Verbs’ show the number of tokens in each relevant lexical class, respectively.

<b>Corpus</b>	<b>Children Included</b>	<b>Tokens per Corpus</b>	<b>Nouns</b>	<b>Verbs</b>
Bates	30	37,637	4,559	5,195
Bernstein	9	31,176	4,637	4,016
Bliss	2	651	76	82
Bloom70	3	150,977	18,670	20,377
Bohannon	3	27,724	3,472	3,613
Braunwald	1	1e+05	11,339	13,164
Brown	3	152,259	19,489	19,012
Clark	1	122,183	16,197	15,674
Cornell	6	18,448	2,460	2,548
Demetras1	1	23,453	2,841	3,376
Demetras2	3	47,616	6,503	6,868
Feldman	1	27,586	4,461	2,870
Gleason	9	65,636	7,548	9,074
Higginson	3	40,051	4,868	5,022
Kuczaj	1	52,455	6,089	7,624
MacWhinney	1	47,406	5,259	6,639
McCune	7	34,847	4,710	5,141
McMillan	2	456	71	56
Morisset	82	46,626	4,796	7,474
Nelson	1	10,992	1,604	1,457
NewEngland	49	144,417	18,147	21,204
Peters	1	106,799	14,179	16,863
Post	3	121,875	14,888	18,392
Providence	6	1,197,626	179,098	148,941
Rollins	3	56,742	7,076	8,016
Sachs	1	42,264	6,197	5,167
Snow	1	67,457	10,094	7,868
Soderstrom	2	98,333	10,127	13,001
Suppes	1	145,447	21,359	18,624
Tardif	24	52,252	7,465	7,159
Valian	1	124,907	16,231	15,865
VanHouten	20	12,791	1,366	2,264
Warren	9	11,820	1,413	1,548
Weist	5	138,348	16,020	18,414

to both experiments below.

### *Experiment 1*

For Experiment 1, we did not find a correlation between mispronunciation effect size and vocabulary size for productive vocabulary ( $r = -0.14$ ,  $p = 0.47$ ), or for receptive vocabulary ( $r = -0.25$ ,  $p = 0.18$ ). This is in keeping with prior studies with high-frequency regular nouns (e.g. Swingley & Aslin, 2000).

### *Experiment 2a*

Just as in Experiment 1, we did not find a correlation between mispronunciation effect size and vocabulary size for productive vocabulary ( $r = 0.21$ ,  $p = 0.28$ ) or receptive vocabulary ( $r = 0.22$ ,  $p = 0.26$ ). This is not particularly surprising, since this age group failed to show a mispronunciation effect overall.

### *Experiment 2b*

For 26-month-olds, the MCDI only includes productive vocabulary. In contrast to Experiment 1 and 2a, we did find a significant correlation between mispronunciation effect size and productive vocabulary size ( $r = 0.41$ ,  $p = 0.02$ ). To further explore this relationship, we conducted correlations between vocabulary and the mispronunciation effect for regular and irregular verbs separately. The significant relationship only held for the irregular verbs ( $r = 0.42$ ,  $p = 0.02$ ), but not regular verbs ( $r = 0.11$ ,  $p = 0.54$ ). This pattern may be linked to the numerically larger decrement in comprehension toddlers showed for regular verbs when they were mispronounced, relative to irregular verbs. I.e. most toddlers had a (perhaps overly!) robust mispronunciation effect for regulars, potentially leaving less variance to be explained by vocabulary. That said, we limit speculation and potential overinterpretation of this result given the highly exploratory nature of this analysis.

### A.3 Exploratory motor analyses

In the motor survey, four of our eight test verbs were queried. The measure asked about toddlers' ability to jump with both feet, run short distances, throw a ball from standing position, and walk. For walking, the survey asked for an estimated date that children started walking. If a date was provided, the toddler was coded as a yes for walking. If no date was provided, the child was coded as a no. For the other three verbs, the motor survey provides respondents with 5 possible response choices for each question. We coded a response of "Probably shows behavior" or "Sure child shows behavior (remember particular instance)" as a yes, and responses of "Sure child does NOT show behavior", "Probably NOT", or "Unsure" as a no.

#### *Experiment 2a*

By parent report, 22% of our participants could jump, 78% could run, 87% could throw, and 100% could walk. As an exploratory analysis, we created a subset of the data containing participants' proportions of target looking on only the 4 test verbs on the motor survey: jump, walk, throw, and run. "Walk" was removed from analysis because all participants were reported to walk. Motor abilities and motor verb comprehension did not correlate (Spearman's  $\rho = -0.18$ ,  $p = 0.22$ ).

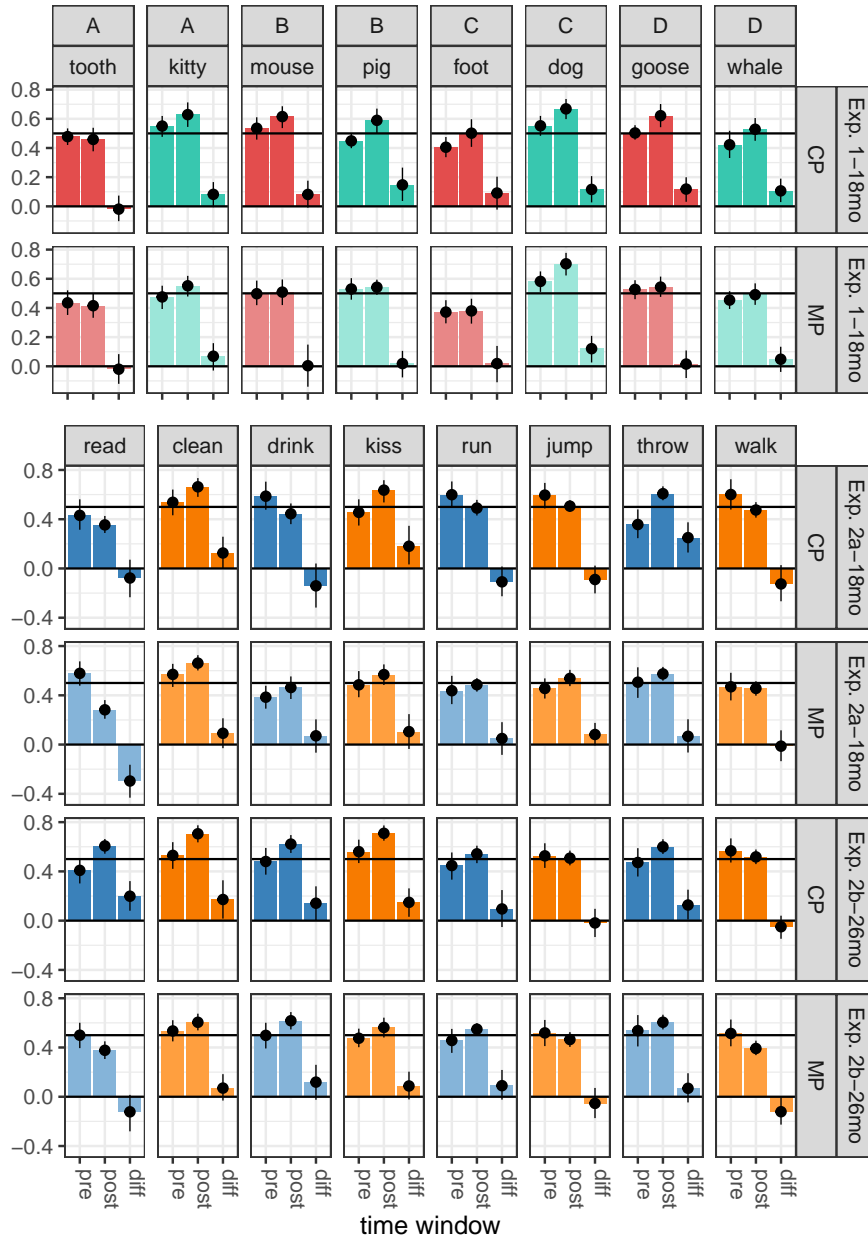
#### *Experiment 2b*

By parent report, 85% of our participants were able to jump with both feet, 96% were able to run short distances, 96% could throw a ball from a standing position, and 100% had taken at least 3 steps unassisted. Conducting the same exploratory analysis as in Experiment 2a, we found a marginal correlation between motor ability and comprehension that was relatively small in magnitude, and did not reach significance ( $\rho = 0.27$ ,  $p = 0.054$ ). That is, 26-month-olds who could do more actions had a (non-significant) tendency towards higher proportions of target-looking for the three motor

verbs included in this analysis.

## A.4 Item-level looking behavior

FIGURE A.1: Figure shows item-level values for target looking before (pre) and after (post) the target word was named, then the difference between those two bars (diff). Values are separated by Pair (A through D), target word, pronunciation type (CP for correct pronunciation, MP for mispronunciation) and experiment.



# Bibliography

- Alloppenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998), “Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models,” *Journal of Memory and Language*, 38, 419–439.
- Altmann, G. T. M. and Kamide, Y. (1999), “Incremental Interpretation at Verbs: Restricting the Domain of Subsequent Reference,” *Cognition*, 73, 247–264.
- Anderson, N. J., Graham, S. A., Prime, H., Jenkins, J. M., and Madigan, S. (2021), “Linking Quality and Quantity of Parental Linguistic Input to Child Language Skills: A Meta-Analysis,” *Child Development*, 92, 484–501.
- Anderwald, L. (2013), “Natural Language Change or Prescriptive Influence?: Throve, Dove, Pled, Drug and Snuck in 19th-Century American English,” *English World-Wide*, 34, 146–176.
- Aslin, R. N. (2007), “What’s in a Look?” *Developmental science*, 10, 48–53.
- Atkinson, J. E. (1976), “Inter- and Intraspeaker Variability in Fundamental Voice Frequency,” *The Journal of the Acoustical Society of America*, 60, 440–445.
- Babineau, M., Shi, R., and Christophe, A. (2020), “14-Month-Olds Exploit Verbs’ Syntactic Contexts to Build Expectations about Novel Words,” *Infancy*, 25, 719–733.
- Bates, E. and Goodman, J. C. (1997), “On the Inseparability of Grammar and the Lexicon: Evidence from Acquisition, Aphasia, and Real-Time Processing,” *Language and Cognitive Processes*, 12, 507–584.
- Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., and Brasher, A. (2013), “The Time Course of Perception of Coarticulation,” *The Journal of the Acoustical Society of America*, 133, 2350–2366.
- Bergelson, E. (2020), “The Comprehension Boost in Early Word Learning: Older Infants Are Better Learners,” *Child Development Perspectives*, n/a.
- Bergelson, E. and Swingle, D. (2012), “At 6-9 Months, Human Infants Know the Meanings of Many Common Nouns,” *Proceedings of the National Academy of Sciences*, 109, 3253–3258.

- Bergelson, E. and Swingley, D. (2013), “The Acquisition of Abstract Words by Young Infants,” *Cognition*, 127, 391–397.
- Bergelson, E. and Swingley, D. (2015), “Early Word Comprehension in Infants: Replication and Extension,” *Language Learning and Development*, 11, 369–380.
- Bergelson, E. and Swingley, D. (2018), “Young Infants’ Word Comprehension Given An Unfamiliar Talker or Altered Pronunciations,” *Child Development*, 89, 1567–1576.
- Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., and Tor, S. (2018), “Day by Day, Hour by Hour: Naturalistic Language Input to Infants,” *Developmental Science*, 0, e12715.
- Bloom, P. (2001), “Précis of How Children Learn the Meanings of Words,” *Behavioral and Brain Sciences*, 24.
- Boersma, P. (2001), “Praat, a System for Doing Phonetics by Computer,” *Glott. Int.*, 5, 341–345.
- Braginsky, M., Sanchez, A., and Yurovsky, D. (2019), *chilDesr: Accessing the ‘CHILDES’ Database*, R package version 0.1.2.
- Brent, M. R. and Siskind, J. M. (2001), “The Role of Exposure to Isolated Words in Early Vocabulary Development,” *Cognition*, 81, B33–44.
- Brooks, R. and Meltzoff, A. N. (2005), “The Development of Gaze Following and Its Relation to Language,” *Developmental Science*, 8, 535–543.
- Brown, R. and Hanlon, C. (1970), “Derivational Complexity and Order of Acquisition in Child Speech,” in *Cognition and the Development of Language*, Wiley, New York.
- Bunce, J., Soderstrom, M., Bergelson, E., Rosemberg, C., Stein, A., Alam, F., Migdalek, M., and Casillas, M. (2020), “A Cross-Cultural Examination of Young Children’s Everyday Language Experiences,” .
- Bybee, J. (1995), “Regular Morphology and the Lexicon,” *Language and Cognitive Processes*, 10, 425–455.
- Byers-Heinlein, K. and Werker, J. F. (2009), “Monolingual, Bilingual, Trilingual: Infants’ Language Experience Influences the Development of a Word-Learning Heuristic,” *Developmental Science*, 12, 815–823.
- Carey, S. and Bartlett, E. (1978), “Acquiring a Single New Word,” .
- Casenhiser, D. M. (2005), “Children’s Resistance to Homonymy: An Experimental Study of Pseudohomonyms,” *Journal of Child Language*, 32, 319–343.

- Charles-Luce, J. and Luce, P. A. (1990), “Similarity Neighbourhoods of Words in Young Children’s Lexicons\*,” *Journal of Child Language*, 17, 205–215.
- Charles-Luce, J. and Luce, P. A. (1995), “An Examination of Similarity Neighbourhoods in Young Children’s Receptive Vocabularies\*,” *Journal of Child Language*, 22, 727–735.
- Chomsky, N. (1959), “Review of Verbal Behavior,” *Language*, 35, 26–58.
- Clark, E. and Grossman, J. (1998), “Pragmatic Directions and Children’s Word Learning,” *Journal of child language*, 25, 1–18.
- Clark, E. V. (1987), “The Principle of Contrast: A Constraint on Language Acquisition,” in *Mechanisms of Language Acquisition*, pp. 1–33, Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US.
- Cohen, L. B. and Strauss, M. S. (1979), “Concept Acquisition in the Human Infant,” *Child Development*, 50, 419–424.
- Consortium, T. M. (2020), “Quantifying Sources of Variability in Infancy Research Using the Infant-Directed-Speech Preference,” *Advances in Methods and Practices in Psychological Science*, 3, 24–52.
- Cooper, R. M. (1974), “The Control of Eye Fixation by the Meaning of Spoken Language: A New Methodology for the Real-Time Investigation of Speech Perception, Memory, and Language Processing,” *Cognitive Psychology*, 6, 84–107.
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., and Hogan, E. M. (2001), “Subcategorical Mismatches and the Time Course of Lexical Access: Evidence for Lexical Competition,” *Language and Cognitive Processes*, 16, 507–534.
- Dautriche, I., Swingle, D., and Christophe, A. (2015), “Learning Novel Phonological Neighbors: Syntactic Category Matters,” *Cognition*, 143, 77–86.
- Dautriche, I., Chemla, E., and Christophe, A. (2016), “Word Learning: Homophony and the Distribution of Learning Exemplars,” *Language Learning and Development*, 12, 231–251.
- Dautriche, I., Fibla, L., Fievet, A.-C., and Christophe, A. (2018), “Learning Homophones in Context: Easy Cases Are Favored in the Lexicon of Natural Languages,” *Cognitive Psychology*, 104, 83–105.
- De Carvalho, A. A., He, A. X., Lidz, J., and Christophe, A. (2019), “Prosody and Function Words Cue the Acquisition of Word Meanings in 18-Month-Old Infants,” *Psychological Science*, 30, 319–332.

- Diesendruck, G. and Shatz, M. (2001), “Two-Year-Olds’ Recognition of Hierarchies: Evidence from Their Interpretation of the Semantic Relation between Object Labels,” *Cognitive Development*, 16, 577–594.
- Eimas, P., Siqueland, Jusczyk, P. W., and Vigorito (1971), “Speech Perception in Infants,” .
- Fabiano-Smith, L. and Barlow, J. A. (2010), “Interaction in Bilingual Phonological Acquisition: Evidence from Phonetic Inventories,” *International journal of bilingual education and bilingualism*, 13, 81.
- Farroni, T., Massaccesi, S., Pividori, D., and Johnson, M. H. (2004), “Gaze Following in Newborns,” *Infancy*, 5, 39–60.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., and Pethick, S. J. (1994), “Variability in Early Communicative Development,” *Monographs of the Society for Research in Child Development*, 59, 1–173; discussion 174–185.
- Ferguson, C. A. (1964), “Baby Talk in Six Languages,” *American Anthropologist*, 66, 103–114.
- Fernald, A. and Simon, T. (1984), “Expanded Intonation Contours in Mothers’ Speech to Newborns,” *Developmental Psychology*, 20, 104–113.
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., and Fukui, I. (1989), “A Cross-Language Study of Prosodic Modifications in Mothers’ and Fathers’ Speech to Preverbal Infants\*,” *Journal of Child Language*, 16, 477–501.
- Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., and McRoberts, G. W. (1998), “Rapid Gains in Speed of Verbal Processing by Infants in the 2nd Year,” *Psychological Science*, 9, 228–231.
- Fernald, A., Zangl, R., Portillo, A. L., and Marchman, V. A. (2008), “Looking While Listening: Using Eye Movements to Monitor Spoken Language Comprehension by Infants and Young Children,” in *Developmental Psycholinguistics: On-Line Methods in Children’s Language Processing, Language Acquisition and Language Disorders*, pp. 97–135, John Benjamins Publishing Company, Amsterdam, Netherlands.
- Fisher, C. and Gleitman, L. R. (2002), “Language Acquisition,” in *Stevens’ Handbook of Experimental Psychology*, American Cancer Society.
- Floyd, S., Goldberg, A. E., and Lew-Williams, C. (2020), “Toddlers Recognize Multiple Meanings of Polysemous Words,” *Proceedings of the Annual Conference of the Cognitive Science Society, Cognitive Science Society*.

- Fowler, C. A. (1981), "Production and Perception of Coarticulation among Stressed and Unstressed Vowels," *Journal of Speech & Hearing Research*, 24, 127–139.
- Frank, M. C., Braginsky, M., Yurovsky, D., and Marchman, V. A. (2017), "Wordbank: An Open Repository for Developmental Vocabulary Data," *Journal of Child Language*, 44, 677–694.
- Frank, M. C., Braginsky, M., Yurovsky, D., and Marchman, V. (2021), *Variability and Consistency in Early Language Learning*, MIT Press, Cambridge, MA.
- Gentner, D. (1982), "Why Nouns Are Learned Before Verbs : Linguistic Relativity versus Natural Partitioning," in *Language, Thought and Culture*, vol. 2 of *Language Development*, pp. 301–334, Lawrence Erlbaum, Hillsdale, NJ.
- Gentner, D. (2006), "Why Verbs Are Hard to Learn," in *Action Meets Word*, pp. 544–564, Oxford University Press.
- Gerken, L., Murphy, W. D., and Aslin, R. N. (1995), "Three- and Four-Year-Olds' Perceptual Confusions for Spoken Words," *Perception & Psychophysics*, 57, 475–486.
- Gleason, J. B., Perlmann, R. Y., Ely, R., and Evans, D. W. (1994), "The Baby Talk Register: Parents' Use of Diminutives," *Handbook of research in language using CHILDES*, pp. 50–76.
- Gleitman, L. (1990), "The Structural Sources of Verb Meanings," *Language Acquisition*, 1, 3–55.
- Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., and Gordon, L. (1987), "The Eyes Have It: Lexical and Syntactic Comprehension in a New Paradigm," *Journal of Child Language*, 14, 23–45.
- Goodman, J. C., Dale, P. S., and Li, P. (2008), "Does Frequency Count? Parental Input and the Acquisition of Vocabulary," *Journal of Child Language*, 35, 515–531.
- Gredebäck, G., Fikke, L., and Melinder, A. (2010), "The Development of Joint Visual Attention: A Longitudinal Study of Gaze Following during Interactions with Mothers and Strangers," *Developmental Science*, 13, 839–848.
- Guevara-Rukoz, A., Cristia, A., Ludusan, B., Thiollière, R., Martin, A., Mazuka, R., and Dupoux, E. (2018), "Are Words Easier to Learn From Infant- Than Adult-Directed Speech? A Quantitative Corpus-Based Investigation," *Cognitive Science*, 42, 1586–1617.
- Halberda, J. (2003), "The Development of a Word-Learning Strategy," *Cognition*, 87, B23–34.

- Hallé, P. A. and de Boysson-Bardies, B. (1996), “The Format of Representation of Recognized Words in Infants’ Early Receptive Lexicon,” *Infant Behavior and Development*, 19, 463–481.
- Hendrickson, A. T. and Perfors, A. (2019), “Cross-Situational Learning in a Zipfian Environment,” *Cognition*, 189, 11–22.
- Henrich, J., Heine, S. J., and Norenzayan, A. (2010), “The Weirdest People in the World?” *Behavioral and Brain Sciences*, 33, 61–83.
- Hirsh-Pasek, K., Treiman, R., and Schneiderman, M. (1984), “Brown & Hanlon Revisited: Mothers’ Sensitivity to Ungrammatical Forms\*,” *Journal of Child Language*, 11, 81–88.
- Hollich, G., Hirsh-Pasek, K., and Golinkoff, R. M. (2000), “I. What Does It Take to Learn a Word?” *Monographs of the Society for Research in Child Development*, 65, 1–16.
- Houston-Price, C., Mather, E., and Sakkalou, E. (2007), “Discrepancy between Parental Reports of Infants’ Receptive Vocabulary and Infants’ Behaviour in a Preferential Looking Task\*,” *Journal of Child Language*, 34, 701–724.
- Huettig, F., Rommers, J., and Meyer, A. S. (2011), “Using the Visual World Paradigm to Study Language Processing: A Review and Critical Evaluation,” *Acta Psychologica*, 137, 151–171.
- Huttenlocher, J., Vasilyeva, M., Waterfall, H. R., Vevea, J. L., and Hedges, L. V. (2007), “The Varieties of Speech to Young Children,” *Developmental Psychology*, 43, 1062–1083.
- Inhelder, B. and Piaget, J. (1969), “The Early Growth of Logic in the Child,” .
- Johnson, E. K. (2016), “Constructing a Proto-Lexicon: An Integrative View of Infant Language Development,” *Annual Review of Linguistics*, 2, 391–412.
- Kelly, B., Wigglesworth, G., Nordlinger, R., and Blythe, J. (2014), “The Acquisition of Polysynthetic Languages,” *Language and Linguistics Compass*, 8, 51–64.
- Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., and Gerken, L. (1995), “The Head-Turn Preference Procedure for Testing Auditory Perception,” *Infant Behavior and Development*, 18, 111–116.
- Kipfer, B. A. (2011), *Roget’s International Thesaurus, 7th Edition*, Collins Reference, New York, 7th ed. edition edn.
- Kleinschmidt, D. F. and Jaeger, T. F. (2015), “Robust Speech Perception: Recognize the Familiar, Generalize to the Similar, and Adapt to the Novel,” *Psychological Review*, 122, 148–203.

- Kouider, S., Halberda, J., Wood, J., and Carey, S. (2006), “Acquisition of English Number Marking: The Singular-Plural Distinction,” *Language Learning and Development*, 2, 1–25.
- Kuhl, P. K. (1991), “Human Adults and Human Infants Show a “Perceptual Magnet Effect” for the Prototypes of Speech Categories, Monkeys Do Not,” *Perception & Psychophysics*, 50, 93–107.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992), “Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age,” *Science*, 255, 606–608.
- Kuhl, P. K., Andruski, J., Chistovich, A., Chistovich, L., Kozhevnikova, E., Ryskina, V., Stolyarova, E., Sundberg, U., and Lacerda, F. (1997), “Cross-Language Analysis of Phonetic Units in Language Addressed to Infants,” *Science*, 277, 684–686.
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., and Iverson, P. (2006), “Infants Show a Facilitation Effect for Native Language Phonetic Perception between 6 and 12 Months,” *Developmental Science*, 9, F13–F21.
- Kuwabara, M. and Smith, L. B. (2012), “Cross-Cultural Differences in Cognitive Development: Attention to Relations and Objects,” *Journal of Experimental Child Psychology*, 113, 20–35.
- Laing, C. (2019), “A Role for Onomatopoeia in Early Language: Evidence from Phonological Development,” *Language and Cognition*, 11, 173–187.
- Landau, B. and Gleitman, L. R. (1985), *Language and Experience: Evidence from the Blind Child*, Harvard University Press, Cambridge, MA.
- Lewis, M., Cristiano, V., Lake, B. M., Kwan, T., and Frank, M. C. (2020), “The Role of Developmental Change and Linguistic Experience in the Mutual Exclusivity Effect,” *Cognition*, 198, 104191.
- Liberman, A. M. and Studdert-Kennedy, M. (1978), “Phonetic Perception,” in *Perception*, eds. S. M. Anstis, J. Atkinson, C. Blakemore, O. Braddick, T. Brandt, F. W. Campbell, S. Coren, J. Dichgans, P. C. Dodwell, P. D. Eimas, J. M. Foley, R. Fox, L. Ganz, M. Garrett, E. J. Gibson, J. S. Girgus, M. M. Haith, Y. Hatwell, E. R. Hilgard, D. Ingle, G. Johansson, B. Julesz, M. Konishi, J. R. Lackner, E. Levinson, A. M. Liberman, L. Maffei, T. Oyama, A. Pantle, E. Pöppel, R. Sekuler, C. F. Stromeyer, M. Studdert-Kennedy, H.-L. Teuber, R. K. Yin, R. Held, H. W. Leibowitz, and H.-L. Teuber, *Handbook of Sensory Physiology*, pp. 143–178, Springer, Berlin, Heidelberg.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967), “Perception of the Speech Code,” *Psychological Review*, 74, 431–461.

- Libertus, K. and Landa, R. J. (2013), “The Early Motor Questionnaire (EMQ): A Parental Report Measure of Early Motor Development,” *Infant Behavior & Development*, 36, 833–842.
- Liittschwager, J. C. and Markman, E. M. (1994), “Sixteen- and 24-Month-Olds’ Use of Mutual Exclusivity as a Default Assumption in Second-Label Learning,” *Developmental Psychology*, 30, 955–968.
- Lindblom, B. (1990), “Explaining Phonetic Variation: A Sketch of the H&H Theory,” in *Speech Production and Speech Modelling*, eds. W. J. Hardcastle and A. Marchal, pp. 403–439, Springer Netherlands, Dordrecht.
- Liu, H.-M., Tsao, F.-M., Chang, C.-J., and Hsu, L.-L. (2013), “The Development of Speech Discrimination in Preschool and School-Aged Children: Association with Word Comprehension. [The Development of Speech Discrimination in Preschool and School-Aged Children: Association with Word Comprehension.]” *Bulletin of Educational Psychology*, 45, 221–240.
- Lynip, A. W. (1951), “The Use of Magnetic Devices in the Collection and Analysis of the Preverbal Utterances of an Infant,” *Genetic Psychology Monographs*, 44, 221–262.
- MacWhinney, B. (2000), *The CHILDES Project: Tools for Analyzing Talk*, Lawrence Erlbaum Associates, Mahwah, NJ, 3 edn.
- MacWhinney, B. and 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.
- Magen, H. S. (1998), “The Perception of Foreign-Accented Speech,” *Journal of Phonetics*, 26, 381–400.
- Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., and Aslin, R. N. (2007), “The Dynamics of Lexical Competition During Spoken Word Recognition,” *Cognitive Science*, 31, 133–156.
- Mahr, T., McMillan, B. T., Saffran, J. R., Ellis Weismer, S., and Edwards, J. (2015), “Anticipatory Coarticulation Facilitates Word Recognition in Toddlers,” *Cognition*, 142, 345–350.
- Mani, N. and Plunkett, K. (2007), “Phonological Specificity of Vowels and Consonants in Early Lexical Representations,” *Journal of Memory and Language*, 57, 252–272.
- Mani, N. and Plunkett, K. (2010), “Twelve-Month-Olds Know Their Cups From Their Keps and Tups,” *Infancy*, 15, 445–470.

- Marcus, G. F. (1995), “Children’s Overregularization of English Plurals: A Quantitative Analysis\*,” *Journal of Child Language*, 22, 447–459.
- Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., and Clahsen, H. (1992), “Overregularization in Language Acquisition,” *Monographs of the Society for Research in Child Development*, 57, i–178.
- Markman, E. M. (1990), “Constraints Children Place on Word Meanings,” *Cognitive Science*, 14, 57–77.
- Markman, E. M. and Wachtel, G. F. (1988), “Children’s Use of Mutual Exclusivity to Constrain the Meanings of Words,” *Cognitive Psychology*, 20, 121–157.
- Markman, E. M., Wasow, J. L., and Hansen, M. B. (2003), “Use of the Mutual Exclusivity Assumption by Young Word Learners,” *Cognitive Psychology*, 47, 241–275.
- Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., and Cristia, A. (2015), “Mothers Speak Less Clearly to Infants Than to Adults: A Comprehensive Test of the Hyperarticulation Hypothesis,” *Psychological Science*, 26, 341–347.
- Mazuka, R., Kondo, T., and Hayashi, A. (2008), “Japanese Mothers’ Use of Specialized Vocabulary in Infant-Directed Speech: Infant-Directed Vocabulary in Japanese,” in *The Origins of Language: Unraveling Evolutionary Forces*, ed. N. Masataka, pp. 39–58, Springer Japan, Tokyo.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. (1988), “A Precursor of Language Acquisition in Young Infants,” *Cognition*, 29, 143–178.
- Mervis, C. B., Golinkoff, R. M., and Bertrand, J. (1994), “Two-Year-Olds Readily Learn Multiple Labels for the Same Basic-Level Category,” *Child Development*, 65, 1163–1177.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013), “Efficient Estimation of Word Representations in Vector Space,” *arXiv:1301.3781 [cs]*.
- Miller (2016), “Introduction English Plural Formation:,” *Journal of English Linguistics*.
- Mirman, D. (2014), *Growth Curve Analysis and Visualization Using R*, CRC Press.
- Mirman, D. and Magnuson, J. S. (2009), “Dynamics of Activation of Semantically Similar Concepts during Spoken Word Recognition,” *Memory & Cognition*, 37, 1026–1039.

- Miyazawa, K., Shinya, T., Martin, A., Kikuchi, H., and Mazuka, R. (2017), “Vowels in Infant-Directed Speech: More Breathier and More Variable, but Not Clearer,” *Cognition*, 166, 84–93.
- Moore, C., Dailey, S., Garrison, H., Amatuni, A., and Bergelson, E. (2019), “Point, Walk, Talk: Links between Three Early Milestones, from Observation and Parental Report.” *Developmental psychology*.
- Naigles, L. (1990), “Children Use Syntax to Learn Verb Meanings,” *Journal of Child Language*, 17, 357–374.
- Naigles, L. (1996), “The Use of Multiple Frames in Verb Learning via Syntactic Bootstrapping,” *Cognition*, 58, 221–251.
- Narayan, C. R. and McDermott, L. C. (2016), “Speech Rate and Pitch Characteristics of Infant-Directed Speech: Longitudinal and Cross-Linguistic Observations,” *The Journal of the Acoustical Society of America*, 139, 1272–1281.
- Oden, G. C. and Massaro, D. W. (1978), “Integration of Featural Information in Speech Perception,” *Psychological Review*, 85, 172–191.
- Ota, M., Davies-Jenkins, N., and Skarabela, B. (2018), “Why Choo-Choo Is Better Than Train: The Role of Register-Specific Words in Early Vocabulary Growth,” *Cognitive Science*, 42, 1974–1999.
- Papafragou, A. and Musolino, J. (2003), “Scalar Implicatures: Experiments at the Semantics–Pragmatics Interface,” *Cognition*, 86, 253–282.
- Peperkamp, S., Hegde, M., and Carbajal, M. J. (2019), “Liquid Deletion in French Child-Directed Speech,” in *Interspeech 2019*, pp. 3574–3578, ISCA.
- Piantadosi, S. T., Tily, H., and Gibson, E. (2012), “The Communicative Function of Ambiguity in Language,” *Cognition*, 122, 280–291.
- Pierrehumbert, J. B. (2001), “Exemplar Dynamics: Word Frequency, Lenition and Contrast,” in *Typological Studies in Language*, eds. J. L. Bybee and P. J. Hopper, vol. 45, p. 137, John Benjamins Publishing Company, Amsterdam.
- Plunkett, K. and Marchman, V. (1991), “U-Shaped Learning and Frequency Effects in a Multi-Layered Perception: Implications for Child Language Acquisition,” *Cognition*, 38, 43–102.
- Plunkett, K. and Marchman, V. (1993), “From Rote Learning to System Building: Acquiring Verb Morphology in Children and Connectionist Nets,” *Cognition*, 48, 21–69.

- Polka, L. and Werker, J. F. (1994), “Developmental Changes in Perception of Non-native Vowel Contrasts,” *Journal of Experimental Psychology: Human Perception and Performance*, 20, 16.
- Quine, W. V. O. (1960), *Word and Object*, MIT Press, Cambridge, UNITED STATES.
- Rowe, M. L. (2012), “A Longitudinal Investigation of the Role of Quantity and Quality of Child-Directed Speech in Vocabulary Development,” *Child Development*, 83, 1762–1774.
- Salverda, A. P., Kleinschmidt, D., and Tanenhaus, M. K. (2014), “Immediate Effects of Anticipatory Coarticulation in Spoken-Word Recognition,” *Journal of Memory and Language*, 71, 145–163.
- Savage, S. L. and fong Au, T. K. (1996), “What Word Learners Do When Input Contradicts the Mutual Exclusivity Assumption,” *Child Development*, 67, 3120–3134.
- Savickien?, I. and 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.
- Segal, J. and Newman, R. S. (2015), “Infant Preferences for Structural and Prosodic Properties of Infant-Directed Speech in the Second Year of Life,” *Infancy*, 20, 339–351.
- Senju, A. and Csibra, G. (2008), “Gaze Following in Human Infants Depends on Communicative Signals,” *Current Biology*, 18, 668–671.
- Shi, R. and Lepage, M. (2008), “The Effect of Functional Morphemes on Word Segmentation in Preverbal Infants,” *Developmental Science*, 11, 407–413.
- Soderstrom, M. (2007), “Beyond Babytalk: Re-Evaluating the Nature and Content of Speech Input to Preverbal Infants,” *Developmental Review*, 27, 501–532.
- Srinivasan, M., Berner, C., and Rabagliati, H. (2019), “Children Use Polysemy to Structure New Word Meanings.” *Journal of Experimental Psychology: General*, 148, 926–942.
- Stager, C. L. and Werker, J. F. (1997), “Infants Listen for More Phonetic Detail in Speech Perception than in Word-Learning Tasks,” *Nature*, 388, 381–382.
- Stengelin, R., Hepach, R., and Haun, D. B. M. (2020), “Cross-Cultural Variation in How Much, but Not Whether, Children Overimitate,” *Journal of Experimental Child Psychology*, 193, 104796.

- Storkel, H. L. (2002), “Restructuring of Similarity Neighbourhoods in the Developing Mental Lexicon,” *Journal of Child Language*, 29, 251–274.
- Strohner, H. and Nelson, K. E. (1974), “The Young Child’s Development of Sentence Comprehension: Influence of Event Probability, Nonverbal Context, Syntactic Form, and Strategies,” *Child Development*, 45, 567–576.
- Swingley, D. and Aslin, R. (2000), “Spoken Word Recognition and Lexical Representation in Very Young Children,” *Cognition*, 76, 147–166.
- Swingley, D. and Aslin, R. (2002), “Lexical Neighborhoods and the Word-Form Representations of 14-Month-Olds,” *Psychological Science*, 13, 480–484.
- Swingley, D. and Humphrey, C. (2018), “Quantitative Linguistic Predictors of Infants’ Learning of Specific English Words,” *Child development*, 89, 1247–1267.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995), “Integration of Visual and Linguistic Information in Spoken Language Comprehension,” *Science*, 268, 1632–1634.
- Tincoff, R. and Jusczyk, P. W. (1999), “Some Beginnings of Word Comprehension in 6-Month-Olds,” *Psychological Science*, 10, 172–175.
- Tincoff, R. and Jusczyk, P. W. (2012), “Six-Month-Olds Comprehend Words That Refer to Parts of the Body,” *Infancy*, 17, 432–444.
- Tsuji, S. and Cristia, A. (2014), “Perceptual Attunement in Vowels: A Meta-Analysis,” *Developmental Psychobiology*, 56, 179–191.
- Valian, V. (2006), “Young Children’s Understanding of Present and Past Tense,” *Language Learning and Development*, 2, 251–276, eprint: <https://doi.org/10.1207/s15473341lld0204.2>.
- Valleau, M. J., Konishi, H., Golinkoff, R. M., Hirsh-Pasek, K., and Arunachalam, S. (2018), “An Eye-Tracking Study of Receptive Verb Knowledge in Toddlers,” *Journal of Speech, Language, and Hearing Research : JSLHR*, 61, 2917–2933.
- VanDam, M., Warlaumont, A., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, P., and MacWhinney, B. (2016), “HomeBank: An Online Repository of Daylong Child-Centered Audio Recordings,” *Seminars in Speech and Language*, 37, 128–142.
- Von Holzen, K. and Bergmann, C. (2018), “A Meta-Analysis of Infants’ Mispronunciation Sensitivity Development,” p. 6.
- Von Holzen, K. and Bergmann, C. (2021), “The Development of Infants’ Responses to Mispronunciations: A Meta-Analysis,” *Developmental Psychology*, 57, 18.

- Wagner, L. (2001), “Aspectual Influences on Early Tense Comprehension,” *Journal of Child Language*, 28.
- Walle, E. A. and Campos, J. J. (2014), “Infant Language Development Is Related to the Acquisition of Walking.” *Developmental Psychology*, 50, 336–348.
- Wang, F. and Trueswell, J. C. (2019), “Spotting Dalmatians: Children’s Ability to Discover Subordinate-Level Word Meanings Cross-Situationally,” *Cognitive psychology*, 114, 101226.
- Weatherhead, D., Kandhadai, P., Hall, D. G., and Werker, J. F. (2021), “Putting Mutual Exclusivity in Context: Speaker Race Influences Monolingual and Bilingual Infants’ Word-Learning Assumptions,” *Child Development*.
- Weisleder, A. and Fernald, A. (2013), “Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary,” *Psychological Science*, 24, 2143–2152.
- Werker, J. F. and Tees, R. C. (1984), “Cross-Language Speech Perception: Evidence for Perceptual Reorganization during the First Year of Life,” *Infant Behavior & Development*, 7, 49–63.
- White, K. S. and Morgan, J. L. (2008), “Sub-Segmental Detail in Early Lexical Representations,” *Journal of Memory and Language*, 59, 114–132.
- White, K. S., Morgan, J. L., Wier, L., and Brown (2004), “When Is a Dar a Car ? Effects of Mispronunciation and Referential Context on Sound-Meaning Mappings,” .
- White, K. S., Yee, E., Blumstein, S. E., and Morgan, J. L. (2013), “Adults Show Less Sensitivity to Phonetic Detail in Unfamiliar Words, Too,” *Journal of Memory and Language*, 68, 362–378.
- Xanthos, A., Laaha, S., Gillis, S., Stephany, U., Aksu-Koç, A., Christofidou, A., Gagarina, N., Hrzica, G., Ketrez, F. N., Kilani-Schoch, M., Korecky-Kröll, K., Kovac´, M., Laalo, K., Palmovic´, M., Pfeiler, B., Voeikova, M. D., and Dressler, W. U. (2011), “On the Role of Morphological Richness in the Early Development of Noun and Verb Inflection,” *First Language*, 31, 461–479.
- Yurovsky, D. and Yu, C. (2008), “Mutual Exclusivity in Cross-Situational Statistical Learning,” *Proceedings of the Annual Meeting of the Cognitive Science Society*, 30, 7.
- Zamuner, T. S., Moore, C., and Desmeules-Trudel, F. (2016), “Toddlers’ Sensitivity to within-Word Coarticulation during Spoken Word Recognition: Developmental Differences in Lexical Competition,” *Journal of experimental child psychology*, 152, 136–148.

Zipf, G. K. (1945), "The Meaning-Frequency Relationship of Words," *The Journal of General Psychology*, 33, 251–256.

# Biography

Charlotte Emma Moore completed her undergraduate degree from the University of Ottawa in 2014, receiving joint honours in Linguistics and Psychology. Charlotte completed her master's degree in Linguistics at the University of Ottawa in 2015. Her PhD in Developmental Psychology was earned at Duke University in 2021. Charlotte was funded in her master's degree and during her PhD with fellowships from the Social Sciences and Humanities Research Council. She will be a postdoctoral researcher continuing her work on variability and language acquisition at Concordia University in Montreal.