Journal of Memory and Language 126 (2022) 104337

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Examining the roles of regularity and lexical class in 18–26-month-olds' representations of how words sound

Charlotte Moore[a,b,*], Elika Bergelson[b]

[a] Department of Psychology, Concordia University, Canada
[b] Psychology & Neuroscience, Center for Cognitive Neuroscience, Duke University, United States

ARTICLE INFO

Keywords:
Lexical class
Irregular words
Verb learning
Spoken word comprehension
Language acquisition

ABSTRACT

By around 12 months, infants have well-specified phonetic representations for the nouns they understand, for instance looking less at a car upon hearing 'cur' than 'car' (Swingley and Aslin, 2002). Here we test whether such high-fidelity representations extend to irregular nouns, and regular and irregular verbs. A corpus analysis confirms the intuition that irregular verbs are far more common than irregular nouns in speech to young children. Two eyetracking experiments then test whether toddlers are sensitive to mispronunciation in regular and irregular nouns (Experiment 1) and verbs (Experiment 2). For nouns, we find a mispronunciation effect and no regularity effect in 18-month-olds. For verbs, in Experiment 2a, we find only a regularity effect and no mispronunciation effect in 18-month-olds, though toddlers' poor comprehension overall limits interpretation. Finally, in Experiment 2b we find a mispronunciation effect and no regularity effect in 26-month-olds. The interlocking roles of lexical class and regularity for wordform representations and early word learning are discussed.

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 and Werker, 1994; Werker and Tees, 1984). At the same time, they begin to understand common nouns referring to familiar people, foods, and body parts (Bergelson and Swingley, 2012;Tincoff and Jusczyk, 1999, 2012), with relatively robust comprehension in place just after the first birthday (Bergelson, 2020;Fernald et al., 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 and 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. Data supporting this kind of underspecification has been shown in 11-month-olds. In a headturn preference procedure, Hallé and de Boysson-Bardies (1996) report that infants did not show a preference between well-pronounced word lists and mispronunciations of the same words, suggesting that they either did not detect a difference between the two lists, or simply did not prefer one list to the other at 11 months.

A second view posits that infants have highly detailed representations of words they know early on, regardless of phonological neighborhood density (Gerken et al., 1995;Swingley and 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 subphonemic representations of the words they know (White and Morgan, 2008;Zamuner et al., 2016). A second line of work focusing on phonemic-level manipulations also supports this view, in what are usually dubbed "mispronunciation" studies.[1]

Mispronunciation studies are similar to standard looking-while-listening studies (Fernald et al., 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

* Corresponding author at: 7141 Sherbrooke St W, Montreal, Quebec H4B 1R6, Canada
E-mail address: charlotteemma.moore@concordia.ca (C. Moore).

[1] 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.

https://doi.org/10.1016/j.jml.2022.104337
Received 9 December 2020; Received in revised form 13 April 2022; Accepted 17 May 2022
Available online 24 June 2022

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 and Plunkett, 2007;Mani and Plunkett, 2010; Swingley and Aslin, 2002). 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 and Bergmann, 2018).

However, previous work in this area has two key gaps. First, it has focused largely on words that do not undergo morphological processes that change how the root of the word sounds (i.e. *regular* nouns). While a few irregular words have been tested in prior infant eyetracking studies (e.g. “foot” in Bergelson and Aslin, 2017;Tincoff and Jusczyk, 2012), they have not been examined systematically as potentially distinct from other words. Second, prior work has almost exclusively focused on noun representations, neglecting other lexical classes. In what follows, we first expound upon why these are important gaps to fill, and then begin to fill them through evidence from a suite of studies looking at regularity and wordform representations for nouns and verbs in 18–26-month-olds.

1.1. Lexical Class, Wordform Representations, and Learnability

One reason that the prior literature on wordform 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 et al., 2021), despite verbs actually appearing *more* frequently than nouns in infants’ environments (Goodman et al., 2008; *cf.* 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,Gentner, 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 and 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 and 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 wordform 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 “my”) allow 8-month-olds to segment novel nouns in a speech stream, mispronounced function words (e.g. *kes*) do not.[2] 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 et al., 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 novel word learning.

What about children’s early representations of the sounds in common verbs? Answering this question lets us examine the interrelatedness 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 Liu et al., 2013;Tsuji and Cristia, 2014), 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.

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 (e.g. 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

[2] Note that we concern ourselves in this paper primarily with *understanding* nouns and verbs as opposed to merely recognizing common sound patterns as in e.g. Shi and Lepage (2008). This difference in tasks is important to bear in mind when considering the results of our experiments below, which test both comprehension of open-class words and wordform specificity.

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 wordforms 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.

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 and 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. Corpus Analysis

English irregulars stem from a wide variety of historical and sociolinguistic forces (Anderwald, 2013; Miller, 2016), resulting in a varied assortment of past tense morphology. The present analysis focuses on a particular type of irregularity: one that leads to a vowel alternation across forms, as detailed below. 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 overregularized to similar degrees through early development (Marcus, 1995;Plunkett and Marchman, 1993). The equivalent rates of overregularization 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 and Marchman, 1991,Plunkett and Marchman, 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.

2.1. Method

Choosing words to include. Our goal was to compare the degree of irregularity that young children hear in verbs relative to nouns. To accomplish this, we sought to analyze a representative sample of highly-frequent, early-learned nouns and verbs in children's input. We used CHILDES, a database which contains transcribed child-directed speech (MacWhinney, 2000), to estimate the frequency and usage of common words in speech to young children. We focused our analysis on words from the Macarthur-Bates Communicative Development Inventory (MCDI), a widely-used vocabulary assessment questionnaire (Fenson et al., 1994). The Words and Sentences form of the MCDI includes 103 verbs. We used all 103 of these verbs in our analysis. For nouns, we included the 103 nouns with the highest reported production rates in 18-month-olds according to Wordbank (Frank et al., 2017) that were also present in the CHILDES corpus, in order to retain the earliest-learned nouns while keeping a constant *N* across lexical classes.[3]

We then classified the 103 nouns and verbs as either "irregular" or "regular." Nouns were classified as irregular if the stem's nuclear vowel alternates when the word is pluralized, e.g. *foot* ~ *feet*. All other nouns were classified as regular, e.g. *cat* ~ *cats*. Similarly, verbs were classified as irregular if they alternate at least their nuclear vowel when conjugated into past tense, e.g. "run" ~ "ran." Verbs could also change other sounds, e.g. *go* ~ *went* and still be classified as irregular. Verbs where the vowel sounds in the stem do not change when in past tense were classified as regular for analysis purposes.

Determining word frequency. We next extracted all the nouns and verbs in the North American English corpora in the CHILDES database

[3] An analogous analysis based on nouns and verbs in CHILDES that were not pre-selected for being on the MCDI revealed the same pattern of results, but led to the need to make many more decisions about what should "count" as a noun or verb (e.g. "right" is tagged as a highly common noun in CHILDES, despite its non-noun uses like "right here" and "right now").

(MacWhinney, 2000) using the childesr package (Braginsky et al., 2019, version 2018.1). We pulled tokens from all of the utterances spoken to children aged 3–36 months, excluding the child's own utterances. This left us with 3,363,486 total 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 S2 in the Supplemental Materials for details about all included corpora. From this dataset, we calculated the frequency of each of the selected noun and verb stems, where "stem" refers to the root morpheme of each word; this allowed us to calculate each word's frequency across inflections or other morphological transformations.

2.2. Results

All words included in the analysis can be found in the Supplementary Materials. The top 20 most frequent nouns and verbs are presented here in Table 1 for convenience. Fig. 1 displays the proportion of all 103 nouns and verbs by regularity.

Nouns. The early-learned nouns in our analysis made up 31% of the total noun tokens in North American English CHILDES. As context, the top 5 nouns overall were "baby," "book," "right," "car," and "thing"; our filter for MCDI words removes "right" and "thing". Critically, only 3 of the 103 nouns were classified as irregular ("foot," "mouse," and "tooth"). These irregulars occurred in their plural forms 53% of the time on average. Thus, based on these corpora, the vast majority of the most frequent concrete nouns young English-learning children hear are regular.

Verbs. The early-learned verbs in our analysis made up 59% of the verb tokens that children heard in CHILDES. We see a pattern previously reported in the literature: although nouns in English are generally learned earlier than verbs, verbs are more common in the input (Goodman, Dale, and Li, 2008). Consistent with this, we find the included verbs were much more frequent in our corpus analysis than the nouns. Again for context, the top 5 verbs overall were "do," "go," "want," "see," and "have"; our filter for MCDI words removes "do" and "want".

Turning to our central question regarding regularity, verbs showed a stark contrast to nouns. We found that 38 of these early-learned verbs (37%) 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 16% of the time when any of these verbs were spoken on average. The verb "fall" appeared in its irregular past tense form the most frequently (58% of the time), while "swing" did so the least often: only 2% of the time. Thus, while individual verbs varied, a large proportion of the most frequent action verbs were irregular, and children heard these verbs with different nuclear vowels relatively often in everyday language contexts.

Taken together, this corpus analysis provides strong evidence that common, early-learned verbs are more likely to be irregular than common concrete nouns in day-to-day language input to young children. While just a few of the included nouns are irregular, more than a third of the verbs are. Setting aside the systemic meaning difference between nouns vs. verbs (roughly speaking, objects vs. actions), the prevalence of words with multiple surface forms alone distinguishes these classes for the words 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.

Table 1
The most frequent 20 early-learned nouns (left) and 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. 'Prop. Plural' and 'Prop. Past Tense' show the proportion of total tokens of each word occured in the inflected form, where vowel changes occur for irregulars. Proportion data is missing for words where past and present tense are orthographically the same.

Rank	Noun	Noun Frequency	Prop. Plural	Verb	Verb Frequency	Prop. Past Tense
1	book	6,889	0.14	*go	28,242	0.09
2	baby	6,698	0.09	*see	23,225	0.04
3	car	4,648	0.12	have	20,764	0.11
4	ball	3,968	0.07	*get	19,437	0.27
5	house	3,555	0.03	put	18,056	NA
6	hand	3,003	0.38	*think	13,358	0.09
7	box	2,994	0.03	*say	13,278	0.29
8	chair	2,581	0.10	look	8,962	0.01
9	water	2,444	0.00	make	8,805	0.15
10	toy	2,355	0.65	like	7,858	0.02
11	truck	2,279	0.13	*take	7,128	0.12
12	hat	2,224	0.08	*eat	6,044	0.08
13	cookie	2,205	0.36	*give	4,766	0.16
14	bed	2,188	0.06	play	4,424	0.05
15	bear	2,180	0.12	*sit	4,358	0.05
16	train	2,095	0.11	*find	3,894	0.20
17	dog	2,072	0.16	*read	3,616	NA
18	*foot	2,062	0.51	help	2,220	0.04
19	juice	1,977	0.00	*throw	2,206	0.11
20	head	1,912	0.03	show	2,165	0.04

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.

2.3. 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).[4] An additional 8 total toddlers participated but were excluded due to one of three reasons: 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 ± 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 race. 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.

2.4. Materials

2.4.1. Questionnaires

Parents were asked to complete the Words & Gestures version of the MCDI (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

[4] 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.

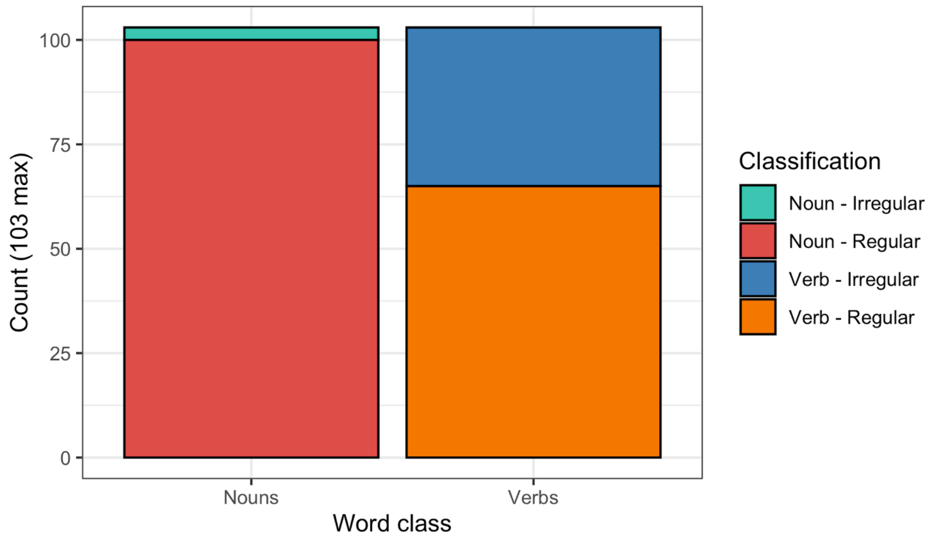

Fig. 1. Relative quantity of regulars and irregulars in the 103 verbs and 103 nouns on the MCDI. Color indicates regularity, with irregulars depicted above regulars in each bar.

day".

2.4.2. Eyetracking

Design

The eyetracking 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 Experiments 1 and 2 can be found at https://osf.io/q9uvd/.

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 and Siskind, 2001). See Table 2 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). Mispronounced stimuli were created by changing the nuclear vowel of each noun. We selected vowel mispronunciations that maximized the acoustic distance between 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 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.

In the testing room, toddlers and their caregivers were seated in front of a display screen (33.7 $\times$ 26.9 cm, 1280 $\times$ 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 2500 ms after the trial began. Participants then had an additional 4000 ms 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 or children if they had not completed the questionnaire prior to arrival. At the conclusion of the visit, families were compensated $5 or $10 for their

Table 2
Stimuli used in Experiment 1. Noun Type indicates whether the noun was regular or irregular. MP is the orthographic representation of the mispronunciation. IPA is the phonological transcription of the mispronunciation. Carrier Phrase 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 and Gestures version of the MCDI. Exp 1 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	IPA	Carrier Phrase	WB Comp.	Exp 1 Comp.	Mean Exp.
A	Tooth	Irreg.	Tath	tæθ	Look at the _	0.77	0.69	3.39
	Kitty	Reg.	Kotty	kɑri		0.89	0.55	3.08
B	Mouse	Irreg.	Mace	meis	Can you see the _	0.50	0.52	2.61
	Pig	Reg.	Pog	pɑg		0.76	0.66	3.05
C	Foot	Irreg.	Foat	foʊt	Find the _	0.93	0.76	4.37
	Dog	Reg.	Dag	dæg		0.96	0.90	4.58
D	Goose	Irreg.	Giss	gIs	Where's the _	0.33	0.38	2.24
	Whale	Reg.	Wile	waɪəɫ		NA	NA	2.16

travel to the lab (based on distance traveled) and toddlers received a book or other small gift as thanks for their participation.

2.5. 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.[5] 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. Parents reported that comprehension was higher in general for regular nouns, overall (mean comprehension for regulars was 70%, for irregulars 59%). 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 for by-item values.

Eyetracking Results

Eyetracking Analysis plan

Our dependent variable is the proportion of target looking during the target window (preregistered as 367 ms to 3000 ms after the target word onset, e.g. "mouse" in "Can you see the mouse?")[6] This looking window was selected to match the window chosen in White and Morgan (2008), who tested mispronunciations in toddlers of the same age. For each participant, their target looking proportion was computed for each of the 16 audio targets (i.e. the correct and mispronounced version of each noun). In sum, each participant contributed 16 datapoints to our analyses. For some of our analyses, these 16 datapoints were further aggregated into 4 summary values: one for correctly pronounced regular nouns, one for correctly pronounced irregular nouns, one for mispronounced regulars, and one for mispronounced irregulars.

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. Next, we report the results of a mixed-effects linear regression model testing our two independent variables (noun pronunciation and noun regularity). We also test the relationship between these two factors and toddlers' reported comprehension of the test words using model comparison.[7] We additionally 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. The average proportion of on-screen looks during the window of analysis in the final sample was 0.89.

Subject- and item-level comparisons to chance.

Statistical details over trial types, subjects, and items are provided in Table 3 and summarized below. By Shapiro–Wilk test, each set of trial type means was normally-distributed (all $ps \geqslant 0.23$). Using One-Sample *t*-tests, we find that when participants heard a regular noun, target-

Table 3
Means by pronunciation and regularity, 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. '(t(28)=' and 'p-value' show the results of One-Sample t-tests comparing proportions to chance. '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. '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

[5] 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 331/396 words).

[6] 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.

[7] Our preregistration stated that we would conduct ANOVAs, but a mixed effects model provides improved control for stimulus variation and individual differences (Judd et al., 2012). For completeness, we include the results analyzed with ANOVAs in the Supplementary Materials.

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 chance when the nouns were correctly pronounced; this pattern again held for most toddlers, 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 toddlers (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 Fig. 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 3 and Fig. 3 for details.

Comparing across trial types.

When creating our mixed-effects model, we used the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages in R. We began by finding the maximal random-effects structure that would converge in a null model with no fixed effects. That model contained random slopes capturing variance in participants' responses to the targets, in addition to random intercepts for target and participant. Once the maximal random structure was created, we added regularity and pronunciation to the model as fixed effects. The final model had the equation:

$$Looking\ proportion \sim Pronunciation^{*}\ Regularity + (1|Target : Subject)$$

Based on reviewer feedback, we also created a model including each participant's MCDI-reported comprehension of each target word as an exploratory additional fixed effect. However, this added predictor was not significant and model comparison revealed that its addition did not improve model fit overall ($\chi^2 = 1.62, p = 0.44$). Furthermore, including MCDI comprehension as a fixed effect required us to exclude trials where "whale" was the target word, since "whale" is not on the MCDI. For these reasons, we present a model including all of the data, without MCDI as a fixed effect.

The intercept represents mean target looking at a correctly pronounced, irregular noun. Pronunciation was a significant fixed effect (see Table 4); mispronunciations reduced target looking during the trial. The effect of regularity was marginal, where participants looked slightly more to the target when it was a regular noun compared to an irregular noun, but this effect did not reach statistical significance. There was no significant interaction. See Fig. 2 for the summary data and Fig. 3 for item-level trends.

To query regularity further, we conducted binomial tests by splitting the data in half by noun regularity. This showed that for irregular nouns, a significant majority of participants (22/29, $p = 0.01$) showed a reduction in looking when they heard a mispronounced noun. However, this did not hold for regular nouns alone (17/29, $p = 0.46$), suggesting that the main effect of pronunciation was weaker within the regular nouns, consistent with the marginal regularity effect in our model. Together, the results of the mixed-effects model and the binomial tests suggest that overall, mispronunciations reduced participants' looking to the target across noun types, but that numerically stronger comprehension and a smaller mispronunciation effect for the regulars rendered this effect more robust within irregular nouns, when data from the two noun types was examined separately.

For data transparency, we provide the timecourse of target looking across trial types in Fig. 4, where the patterns across pronunciation and regularity found when we collapse across our window of analysis can be readily observed across time.

Experiment 1 Discussion

Summarizing across our analyses, Experiment 1 established three results. First, 18-month-olds understood both our correctly pronounced regular and irregular nouns at above-chance rates by most analyses. Second, overall, 18-month-olds looked at the target image less when nouns were mispronounced than when they were correctly pronounced, just as we would expect from previous literature (e.g. Von Holzen and Bergmann, 2018). Third, evidence for an effect of regularity was modest at best. While our mixed-effects models (based on trial-level data) showed only a main effect of pronunciation (i.e. not regularity and no interaction), binomial tests (based on the number of participants with trial-type and pronunciation-type means >.5) showed more robust comprehension for correctly pronounced regulars than irregulars when data from each noun type was considered separately. Relatedly, we found that comprehension of regular nouns was only slightly worse (by 3%) when these nouns were mispronounced, while comprehension dropped more substantially (by 9%) when irregular nouns were mispronounced, relative to performance with their correctly pronounced counterparts. This meant that our overall mispronunciation effect (6% decrease in comprehension relative to correctly pronounced trials when collapsing across regularity) was smaller than that found for (mostly regular) nouns in prior work (e.g. approximately 12% in Swingley and Aslin, 2000). Why might this be?

One possibility is that our mispronunciation effect is smaller because the items we used were lower frequency and/or less well understood by 18-month-olds. Supporting this somewhat, the items in Swingley and Aslin (2000) all appear in the top 50 nouns heard by young children (see Table S1 in the Supplemental Materials), while only 6 of our 8 items do ("goose" and "whale" do not). This suggests that frequency may play a role in the robustness of phonetic representations. Indeed, frequency and age of acquisition have been documented as correlated (but distinct) characteristics for nouns (Goodman et al., 2008), providing further evidence that lower-frequency words are less likely to be understood by toddlers. Relatedly, our items too were reportedly less well understood than those used in prior studies (e.g. Swingley and Aslin, 2000). That is, all the items in Swingley and Aslin (2000) are reportedly understood by an average of 92% of 18-month-olds according to Wordbank, while ours were understood by 73% (Table 2.).

That said, we caution the reader against putting too much stock in these parent-reported comprehension measures for two reasons. First, they did not correspond to toddlers' performance in this task (nor did vocabulary predict performance in prior work finding larger pronunciation effects in more frequent regular nouns, Swingley and Aslin, 2000). This suggests that the relationship between reported and observed comprehension for individual items or across items may not be reliable, or reliably associated with corresponding phonetic specificity for the word. Second, consistent with this, the MCDI was designed to holistically estimate a child's vocabulary size, and is not intended as a measure of any individual child's knowledge of specific items on the questionnaire (Fenson et al., 1994). This makes it particularly ill-suited for assessing the relative accuracy of participants' reported comprehension of specific test words compared to their eyetracking performance. With that caveat in place, the overall lower frequency and reported comprehension for our item set as a whole versus those in Swingley and Aslin (2000) may nevertheless help explain why the reliable mispronunciation effect we find is smaller in magnitude than in that study.

Previous work also suggests that the "ideal" mispronunciation effect is where children 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 pattern suggests that mispronunciations do not totally block lexical access. For example, "giss" is still a better match for "goose" than it is for "whale," so listeners with well-specified wordform representations of these words should still look more to the goose than to the whale upon hearing "giss." This pattern is what we would expect to find in adults as well, as seen for novel words in White et al. (2013). While 18-month-olds in Experiment 1 did look less at the target overall when its label was mispronounced, they did not show this ideal looking

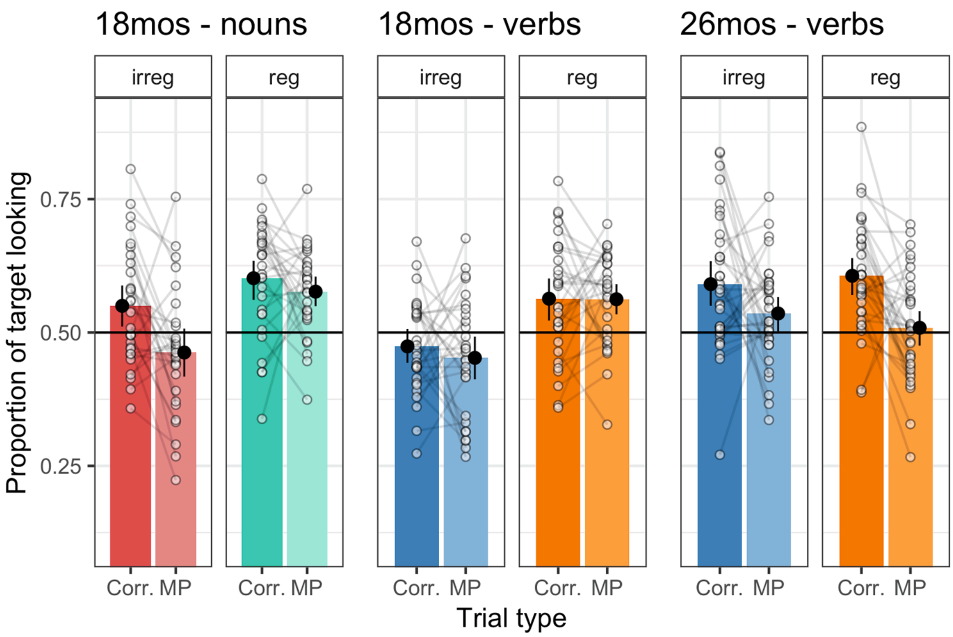

Fig. 2. Proportion of target looking for each trial type in Experiment 1 (left), 2a (middle) and 2b (right). Chance is represented by a black line at 0.5. White dots represent each participant's mean proportion of target looking for each trial type, with lines connecting participants' values across pronunciation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

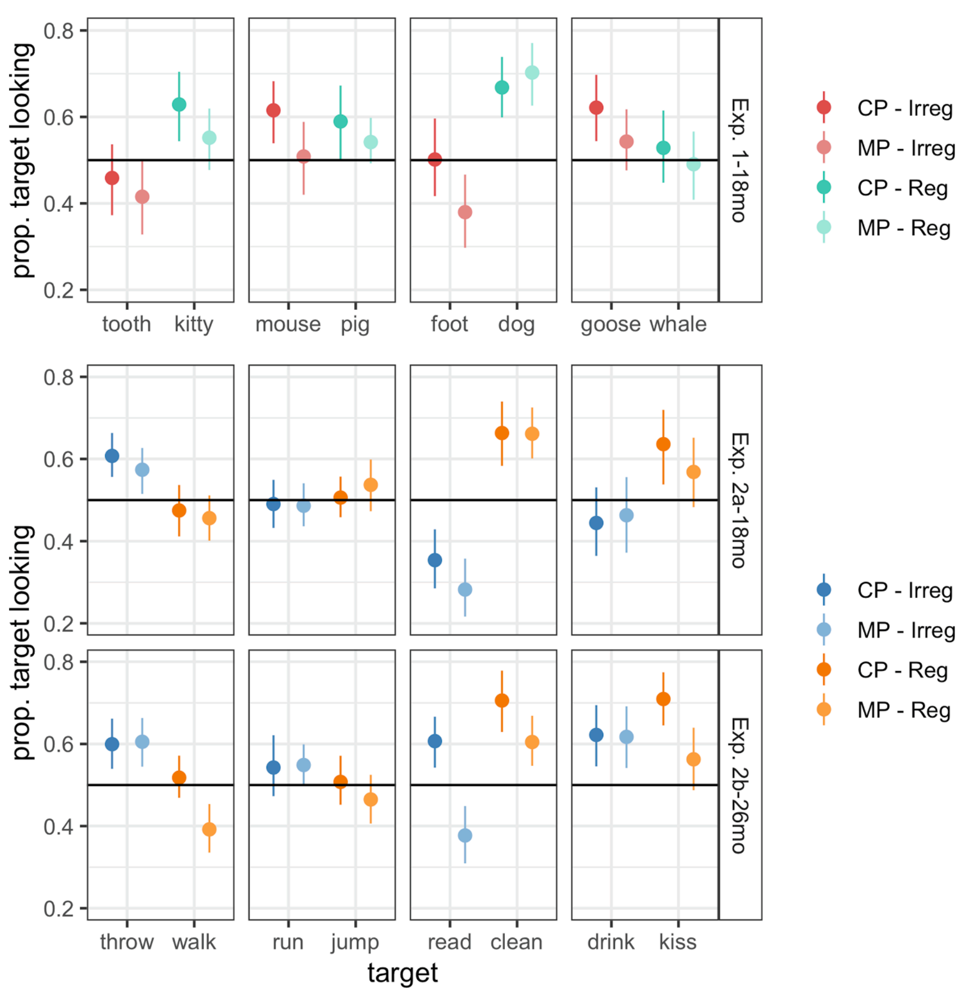

Fig. 3. Item-level breakdowns of toddlers' responses across Experiments 1 and 2. Y-axis shows the proportion of time participants spent looking at the target after it was named. Colors indicate word type and pronunciation. Each panel shows one yoked pair. X-axis displays each specific item name.

Table 4
Summary of fixed effects from the linear mixed effects model that describes Experiment 1. 'Est.' is the model estimates for each fixed effect. 'SE' is the standard error. '+' signifies marginal significance.

Term	Est.	SE	t=	df	p=
Intercept	0.55	0.02	26.22	426.71	<0.001***
Pronunciation	−0.09	0.03	−3.27	225.10	0.001**
Regularity	0.05	0.03	1.83	427.64	0.067+
Interaction	0.06	0.04	1.50	229.36	0.136

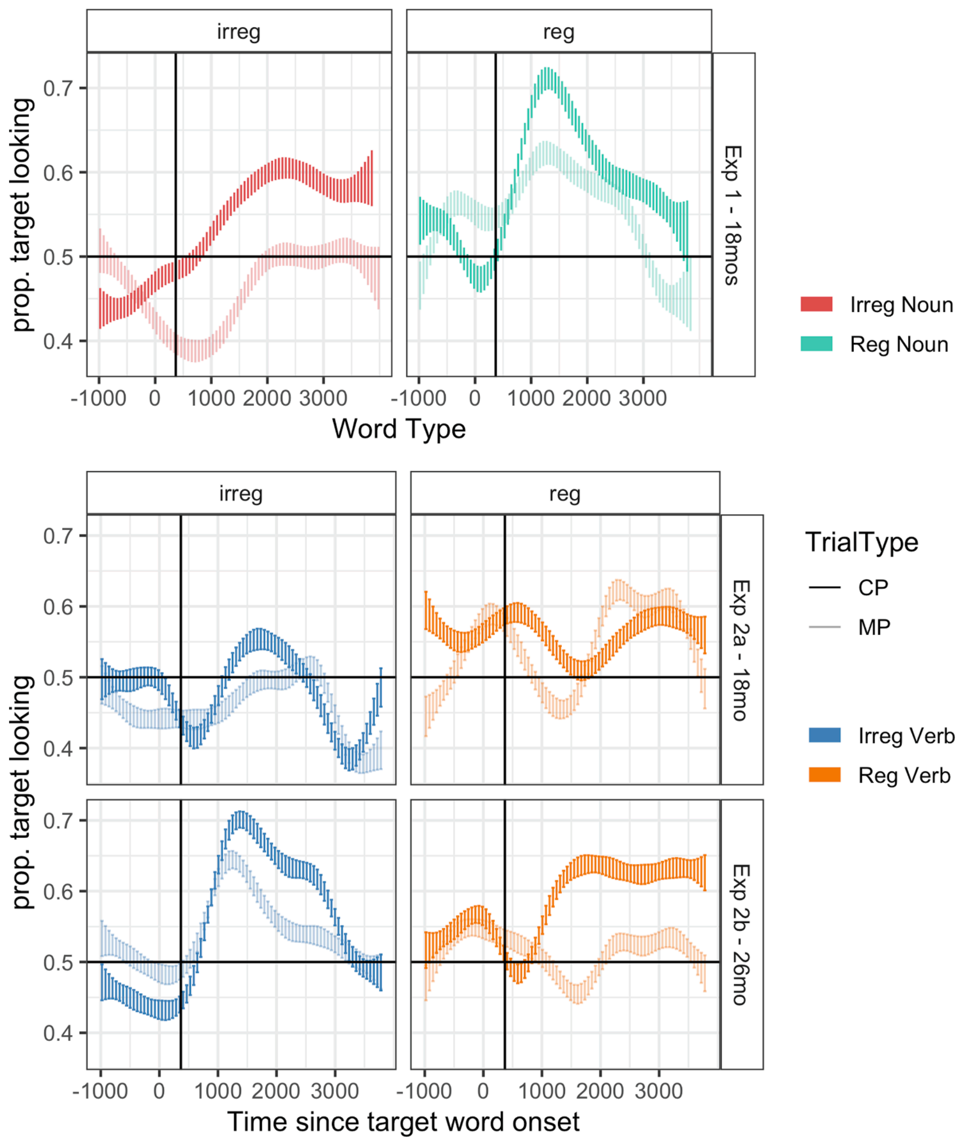

Fig. 4. Timecourse plots for Experiment 1 (top) and Experiment 2 (bottom 4 panels). X-axis shows the time (milliseconds) since target word onset. Y-axis shows the mean proportion of looks to the target image at each timepoint. Each bar represent mean and 95% confidence interval for each 20 ms bin, aggregated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

pattern of Correct > Mispronounced > Chance for either regulars or irregulars. We revisit this 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, participants were more intolerant of mispronunciations for irregular nouns, though comprehension was slightly lower overall for this verb type.

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.

3. 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

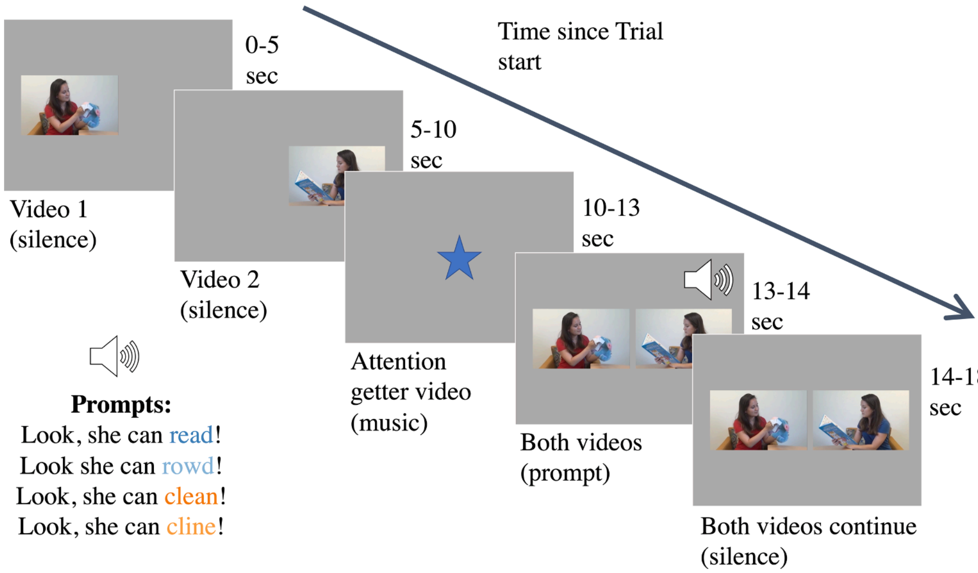

Fig. 5. Schematic of a sample trial in Experiment 2. Time progresses from the top left to bottom right. Side of first video (left above) was counterbalanced across participants and trials.

Table 5
Experiment 2 stimuli details. WB Comp. and WB Prod. show the proportion of 18mos reported to understand and produce each word based on Wordbank's MCDI norms (Frank et al., 2017) for the 'Words and Gestures' and 'Words and Sentences' forms, respectively. 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. (N.B., the 'Words and Sentences' form only queries production.)

Pair	Target	Verb Type	MP	IPA	Carrier	WB Comp.	2a Comp.	WB Prod.	2b Prod.
A	Throw	Irreg.	Thraw	θɹɑː	She's gonna _	0.89	0.56	0.65	0.61
	Walk	Reg.	Wick	wɪk		0.80	0.53	0.72	0.71
B	Run	Irreg.	Roan	ɹoʊn	Look! She can _	0.73	0.44	0.71	0.71
	Jump	Reg.	Joomp	d͡ʒu:mp		0.63	0.28	0.73	0.71
C	Read	Irreg.	Rowd	ɹaʊd	Look! She can _	0.79	0.69	0.78	0.68
	Clean	Reg.	Cline	klɑɪn		0.64	0.41	0.69	0.65
D	Drink	Irreg.	Droink	dɹoiŋk	She's about to _ it	0.93	0.69	0.83	0.77
	Kiss	Reg.	Koss	kɑs		0.98	0.59	0.85	0.71

with correctly pronounced and mispronounced labels for the actions, then compare looking across pronunciation types.

3.1. Experiment 2a Methods

Participants. Participants in the final sample were 30 typically developing toddlers (14 girls)[8] 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, 7% multiracial, and 7% unreported race. 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 and Landa, 2013;Walle and Campos, 2014; questionnaire details can be found in Moore et al., 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 Fig. 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 et al., 2017). See Table 5 for a summary of stimuli. In addition to the test and practice stimuli described below, display items also included four 3.5-s 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 in a way that maximized the acoustic distance between the word's correct vowel, any of its other vowel

[8] 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.

Table 6
Means by pronunciation and regularity, and statistical tests comparing to chance for Experiment 2a (18mos) and Experiment 2b (26mos). Rows depict 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. '(t(28)=' and 'p-value' show the results of One-Sample *t*-tests comparing proportions to chance. '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. '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 $p=$	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

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 5 for pairs. Each video was exactly 5 seconds long.

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-s 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-s 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. 1130 ms 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 min. See Fig. 5 for a schematic of a sample trial.

3.2. 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 367 ms after target word onset, as in Experiment 1, but was extended to the end of the trial (i.e. 3970 ms after the target verb's onset, since all videos were 5 seconds and target verbs occurred 1030 ms 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, Valleau et al. (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.[9] As in Experiment 1, we also tested for correlations between toddlers' 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. The average proportion of on-screen looks during the window of analysis in the final sample was 0.91.

Subject- and item-level comparisons to chance.

Statistical details over trial types, subjects, and items are provided in Table 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 $ps \geqslant 0.12$). 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 Fig. 3 for a by-item breakdown and Table 6 for details.

Comparing across trial types

We conducted the same linear mixed effects regression analysis on these data as we did in Experiment 1. Again, we found that adding MCDI-reported comprehension as a fixed effect to the model did not show it to be a significant predictor nor did its inclusion improve model fit ($\chi^2 = 0.68, p = 0.41$). Additionally, using this model would require us to omit data from the participants for whom we lacked MCDI data. For

[9] 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.

Table 7
Summary of fixed effects for Experiment 2A. 'Est.' is the model estimates for each fixed effect. 'SE' is the standard error.

Term	Est.	SE	$t=$	df	$p=$
Intercept	0.47	0.02	24.58	453.47	<0.001***
Regularity	0.09	0.03	3.38	454.22	<0.001***
Pronunciation	−0.02	0.03	−0.84	230.41	0.399
Interaction	0.01	0.04	0.31	236.23	0.754

these reasons, we present a model including all of the data, without MCDI as a fixed effect.

A summary of the fixed effects can be found in Table 7. The intercept represents the estimated target looking for an irregular, correctly pronounced verb. Only verb type was a significant main effect: participants spent more time fixated on regular verbs when they were the target compared to irregular verbs. There were no other significant main effects and no interaction. 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.

To query regularity further, we conducted binomial tests by splitting the data in half by verb regularity. Binomial tests corroborate the pattern found in the mixed model. For regular verbs, only 15/30 participants showed a mispronunciation effect (i.e. reduced target looking for mispronounced trials relative to correctly pronounced ones, $p = 1$). This held for 3/4 items. For irregular verbs, 13/30 participants showed a mispronunciation effect ($p = 0.58$), which held for 3/4 verbs.

4. 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 above chance 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 (Frank et al., 2017) 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. Further, just as in Experiment 1, parent-reported comprehension measures did not improve model fit, which calls into question how well real-time comprehension (as indexed by the eyetracking results) aligns with parental reports. This pattern seen across Experiment 1 and 2a further bolsters the notion that parent-reported comprehension is difficult to compare to eyetracking data at this scale and granularity. We next consider some hypotheses that may help explain this unexpected pattern of eyetracking results.

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, as noted above MCDI data did not improve model fit in predicting the proportion of target looking in the eyetracking study, a decoupling seen too in other work (Bergelson and Swingley, 2012;Houston-Price et al., 2007; Valleau et al., 2018). Again, this leaves unclear how much stock to put into MCDI data for predicting expected real-time word comprehension for a specific set of items.

If we entertain the idea that MCDI data may be misleading in this case, then 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. Participants' early preference for the regular images/videos in Experiments 1 and 2a are consistent with this possibility (as seen in the timecourse plots, Fig. 4). This pattern may suggest that participants prefer to look at referents whose labels are more stable. However, given the challenges of interpreting this pattern of results here, paired with low reported verb comprehension overall, we hesitate to draw strong conclusions regarding this possibility.

Notably, 18-month-olds do *not* show any evidence of a mispronunciation effect for either verb type, in contrast to what we find in nouns. 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 and 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.

5. 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.

5.1. 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.[10] 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 race. 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.

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.

[10] Our intended sample size was 32 but one more participant than expected met inclusion criteria.

5.2. 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. The average proportion of on-screen looks during the window of analysis in the final sample was 0.92.

Subject- and Item-level comparisons to chance

Statistical details over trial types, subjects, and items are provided in Table 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 $ps \geqslant 0.15$). 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 correctly pronounced trials. 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, toddler's 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 Figs. 2 and 3 for subject- and item-level figures and Fig. 4 for timecourse.

Comparing across trial types

We conducted the same mixed-effects analysis as in the previous two experiments. Here, adding MCDI comprehension was not an option as the Words and Sentences version of the MCDI only queries production. We therefore added item-level production as a fixed effect this time, which did improve model fit ($\chi^2 = 5.02, p = 0.02$).[11] We thus include it in the model we present here, though note that due to missing MCDI data, only 31/33 participants' data were able to be analyzed in this model.

A summary of fixed effects can be found in Table 8. The intercept represents the proportion of looking time for trials where participants were reported not to know the target verb and where the target verb was irregular and correctly pronounced. There was a significant main effect of pronunciation, such that participants looked significantly less to the target verb when its label was mispronounced. In contrast to Experiment 2a, there was no main effect of regularity; participants spent an equivalent amount of time looking at regular and irregular verbs when they were the targets. We also found a significant effect of production based on the MCDI, indicating that participants spent more time looking to target verbs whose labels they were reported to produce than to those they did not yet say. There was no significant interaction between pronunciation and regularity.

Table 8
Fixed effects for Experiment 2b. 'Est.' is the model estimates for each fixed effect. 'SE' is the standard error. The 'Production' row shows the estimated increase in a child's looking time to the target for words they are reported to say on the MCDI. 'Interaction' refers to the tested Pronunciation-by-Regularity interaction.

Term	Est.	SE	$t=$	df	$p=$
Intercept	0.56	0.02	23.38	403.96	<0.001***
Regularity	0.01	0.03	0.51	479.42	0.609
Pronunciation	−0.06	0.03	−2.30	245.59	0.022*
Production	0.05	0.02	2.24	247.38	0.026*
Interaction	−0.04	0.04	−1.16	246.59	0.248

Again, to query regularity further, we conducted binomial tests by splitting the data in half by verb regularity. These tests included participants whose families did not complete the MCDI as well. These tests revealed evidence for a mispronunciation effect within toddlers and items that was robust within regular verbs alone but not within irregular verbs alone. For regulars, 27/33 participants reduced their looking when labels were mispronounced ($p < .001$). Across items, across-participant looking decreased for 4/4 verbs. For irregulars, 17/33 participants ($p > .999$) and 2/4 verbs showed a mispronunciation effect.

5.3. 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 mixed-effects model 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 wordform representations have emerged by this age across this set of verbs, with particularly robust evidence of a mispronunciation effect within the subset of data from regular verbs. Importantly, while 26-month-olds did not show a significant mispronunciation effect in the half of the data testing irregular verbs alone, they clearly showed a stronger comprehension of irregular verbs relative to 18-month-olds in Experiment 2a.

We also see an effect of parent-reported productive vocabulary emerge in this age group, which was not found in the 18-month-olds in Experiment 1 or 2a (who had a limited productive vocabulary). Our interpretation of this result is that parents' ability to veridically assess their children's word knowledge for specific items is greater based on production than on comprehension. We now look at the role of participant age and lexical class together in a pooled analysis across all three experiments.

6. 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 separate, pooled, mixed-effects 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. We

[11] Given this, we also considered a fixed-effect based on MCDI production rather than comprehension for Exp. 1 and 2a, but just as when we added the MCDI comprehension data, we did not find improved model fit by model comparison ($ps \geqslant 0.44$). This is likely due to the quite small number of test words 18-month-olds were producing.

Table 9
Output for the ANOVA run over the pooled mixed effects model. Degrees of freedom are estimated using Satterthwaite's method.

Factor	SumSq	MeanSq	F =	df	p =
Pronunciation	0.93	0.93	24.02	1, 735	<0.001 ***
Regularity	0.91	0.91	23.35	1, 742	<0.001 ***
Experiment	0.46	0.23	5.86	2, 741	0.003 **
Pron. * Exp.	0.25	0.12	3.17	2, 735	0.043 *
Reg. * Exp.	0.60	0.30	7.77	2, 741	<0.001 ***

Table 10
Fixed effects for the pooled model. 'Est' is the model estimate for each fixed effect. The final four rows show the interaction terms.

Term	Est.	Std. Error	*t*-value	*p*-value
Intercept	0.537	0.010	53.907	<0.001 ***
Pronunciation	−0.051	0.010	−4.901	<0.001 ***
Regularity	0.058	0.012	4.832	<0.001 ***
Exp. 2a	−0.064	0.025	−2.565	0.01 **
Exp. 2b	0.069	0.024	2.846	0.004 **
Pron. * Exp. 2a	0.044	0.026	1.665	0.096
Pron. * Exp. 2b	−0.019	0.026	−0.730	0.466
Reg. * Exp. 2a	0.016	0.030	0.521	0.603
Reg. * Exp. 2b	−0.090	0.029	−3.052	0.002 **

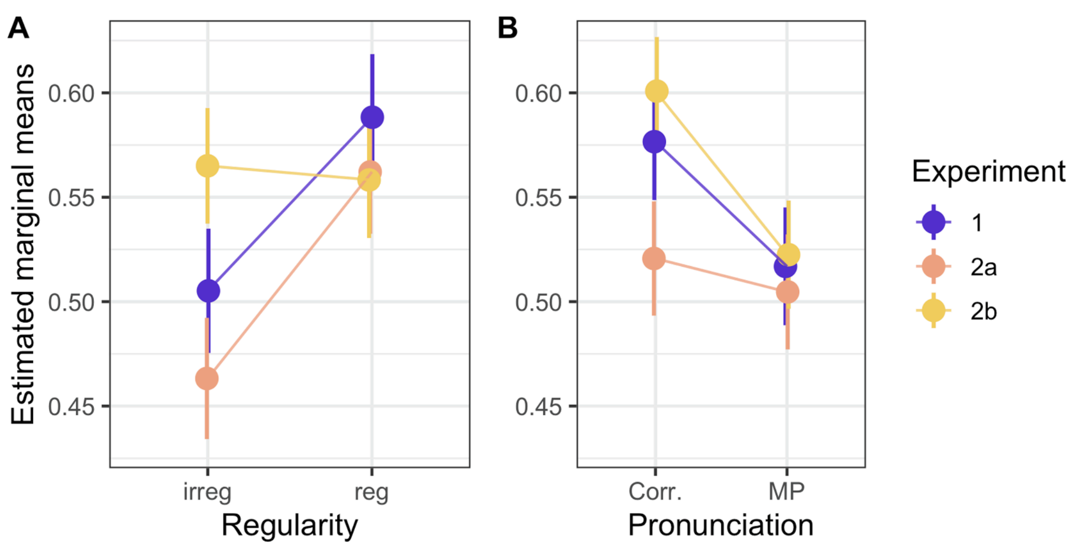

Fig. 6. Plots of the significant interactions in the pooled model. X-axis shows predicted marginal effects for each experiment split by regularity (panel A) or by pronunciation type (panel B).

used the same random effects structure in this analysis as in all previous models. The final model's equation was:

$$Looking\ prop. \sim (Regularity + Pronunciation)^{*}\ Experiment + (1|Target : Subject)$$

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 10 for model estimates and statistics.[12]

We first confirmed that our model predicted significantly more variance than a dummy model with the same random effects structure ($\chi^2 = 78.04, p < .001$). We next examined results from an overall ANOVA run on the mixed-effects model output (Table 9), which found each main effect (Regularity, Pronunciation, Experiment) and interaction (Regularity × Experiment; Pronunciation × Experiment) were significant (though not each level of each individual factor varied from baseline, see Table 10).

For the main effects, on average across studies, participants looked 5.13% less at the target when its label was mispronounced and looked 5.84% more at regular words than at irregular words. Compared to Experiment 1, participants in Experiment 2a looked 6.40% less at the named target, whereas in Experiment 2b participants looked 6.90% more at the target.

These main effects were qualified by the two significant interactions. The Regularity × Experiment interaction revealed that while toddlers' performance for regulars was roughly consistent across studies, their comprehension of irregulars differed by study. Estimated marginal means for regular words were approximately the same across studies (3% difference) while they varied more widely for irregulars (10% difference), see Fig. 6, panel A.

The significant interaction between pronunciation and study revealed that while participants across all three samples had similar

[12] A model that also includes how many words the child was reported to produce on the MCDI did not improve fit over the model by model comparison ($F = 2.35$, $p = 0.13$).

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); see Fig. 6, panel B. Taken together, this pooled analysis indicates that lexical class, regularity, and age (indexed by Experiment) each play a role in driving word comprehension.

7. 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 types of 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 correctly 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, noun comprehension was affected primarily by the noun's pronunciation rather than its regularity.

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. Because the mispronunciation paradigm assumes that participants know the meanings of the tested words, we hesitate to draw strong conclusions about verb representations in this younger age group. Older toddlers in Experiment 2b looked far more similar to the 18-month-olds in Experiment 1, with some notable differences. Summarily, like the toddlers in Experiment 1, 26-month-olds showed a main effect of pronunciation, but not regularity. Further, 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 a significant proportion of the most frequent action verbs young children hear (37% of the early-learned verbs we analyzed). In contrast, for nouns irregularity was far less common: <3% of the early-learned nouns in young children's input were irregular.

This corpus analysis dovetails nicely 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 relatively large proportion of the common verbs young children hear are irregular, and in the group of toddlers who understood the test verbs overall (i.e. the 26-month-olds in Experiment 2b), performance on mispronounced *regular* verbs fell to chance, the marker of failed lexical access.

This pattern of results suggests that rates of vowel alternation in the input 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. However, input frequency does not always drive children's expectations regarding underlying forms, as seen in 12-month-olds' representations of stops and taps (Sundara et al., 2021). At the same time, there is prior work supporting the idea that toddler's consideration of novel words differs based on whether they would have phonological neighbors in the same lexical class or a different one (Dautriche et al., 2015). To us, this suggests toddlers' wordform expectations may indeed factor in alternation rates in the input differentially for nouns and verbs.

7.1. Interlocking effects of age, regularity, word class, and wordform specificity

We can now answer our remaining initial questions regarding regularity and wordform specificity in the affirmative. Namely, we find that regularity does appear to play a role in early phonetic representations when considering nouns and verbs together (as in our pooled analysis). As to our wordform specificity query, we find that wordform specificity does indeed differ between nouns and verbs, though comprehension in the first place varies across these lexical classes as a function of age as well. That said, the patterns that emerged were not simply classifiable as e.g. *irregulars-over-regulars* or *nouns-over-verbs*. Instead, we found interlocking patterns across regularity, 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 and Swingley, 2015; Tincoff and 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 which 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 and 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.

7.2. 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.[13]

A limitation of the current study is that while our results add important knowledge regarding the roles of regularity, lexical class, and wordform specificity, they still do not uncover an age at which toddlers show the mispronunciation effect pattern the literature suggests represents adult-like representations. 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 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 et al., 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.

8. 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 find that regularity matters not only for toddlers' ability to detect mispronunciations across our experiments as a whole, 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 an 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.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the NIH Office of the Director DP5-OD019812 to Elika Bergelson and a SSHRC doctoral fellowship to Charlotte Moore. We would also like to acknowledge the undergraduate research assistants at Duke University who helped collect the data.

All materials, data and code for this manuscript can be found at https://osf.io/q9uvd/.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jml.2022.104337.

References

Anderwald, L. (2013). Natural language change or prescriptive influence?: Throve, dove, pled, drug and snuck in 19th-century American English. *English World-Wide, 34*(2), 146–176. https://doi.org/10.1075/eww.34.2.02and

Babineau, M., Shi, R., & Christophe, A. (2020). 14-month-olds exploit verbs' syntactic contexts to build expectations about novel words. *Infancy, 25*(5), 719–733. https://doi.org/10.1111/infa.12354

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using Lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bates, E., & 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*(5–6), 507–584. https://doi.org/10.1080/016909697386628

Bergelson, E. (2020). The Comprehension Boost in Early Word Learning: Older Infants Are Better Learners. Child Development Perspectives, n/a(n/a). https://doi.org/10.1111/cdep.12373.

Bergelson, E., & Aslin, R. (2017). Semantic specificity in one-year-olds' word comprehension. *Language Learning and Development: The Official Journal of the Society for Language Development, 13*(4), 481–501. https://doi.org/10.1080/15475441.2017.1324308

[13] While Von Holzen and Bergmann (2018) did not find an overall age effect, very few studies tested infants younger than 12 months

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition, 127*(3), 391–397. https://doi.org/10.1016/j.cognition.2013.02.011

Bergelson, E., & Swingley, D. (2015). Early Word Comprehension in Infants: Replication and Extension. *Language Learning and Development, 11*(4), 369–380. https://doi.org/10.1080/15475441.2014.979387

Bergelson, E., & Swingley, D. (2018). Young Infants' Word Comprehension Given An Unfamiliar Talker or Altered Pronunciations. *Child Development, 89*(5), 1567–1576. https://doi.org/10.1111/cdev.12888

Braginsky, M., Sanchez, A., & Yurovsky, D. (2019). Childesr: Accessing the 'CHILDES' database. Retrieved from https://CRAN.R-project.org/package=childesr.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81*(2), B33–44.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*(5), 425–455. https://doi.org/10.1080/01690969508407111

Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons*. *Journal of Child Language, 17*(1), 205–215. https://doi.org/10.1017/S0305000900013180

Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies*. *Journal of Child Language, 22*(3), 727–735. https://doi.org/10.1017/S0305000900010023

Dautriche, I., Swingley, D., & Christophe, A. (2015). Learning novel phonological neighbors: Syntactic category matters. *Cognition, 143*, 77–86. https://doi.org/10.1016/j.cognition.2015.06.003

Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D.J., & Pethick, S.J. (1994). Variability in early communicative development. Monographs of the Society for Research in Child Development, 59(5), 1-173; discussion 174-185.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid Gains in Speed of Verbal Processing by Infants in the 2nd Year. *Psychological Science, 9*(3), 228–231. https://doi.org/10.1111/1467-9280.00044

Fernald, A., Zangl, R., Portillo, A.L., & 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 (pp. 97–135). Amsterdam, Netherlands: John Benjamins Publishing Company. https://doi.org/10.1075/lald.44.06fer.

Fisher, C., & Gleitman, L.R. (2002). Language Acquisition. In Stevens' Handbook of Experimental Psychology. American Cancer Society. https://doi.org/10.1002/0471214426.pas0311.

Frank, M., Braginsky, M., Yurovsky, D., & Marchman, V. (2021). *Variability and Consistency in Early Language Learning*. Cambridge, MA: MIT Press. Retrieved from https://langcog.github.io/wordbank-book/.

Frank, M., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language, 44*(03), 677–694. https://doi.org/10.1017/S0305000916000209

Gentner, D. (1982). Why Nouns are Learned Before Verbs: Linguistic Relativity versus Natural Partitioning. In *Language, thought and culture* (Vol. 2, pp. 301–334). Hillsdale, NJ: Lawrence Erlbaum.

Gentner, D. (2006). Why verbs are hard to learn. In *Action Meets Word* (pp. 544–564). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195170009.001.0001.

Gerken, L., Murphy, W. D., & Aslin, R. N. (1995). Three- and four-year-olds' perceptual confusions for spoken words. *Perception & Psychophysics, 57*(4), 475–486. https://doi.org/10.3758/BF03213073

Gleitman, L. (1990). The Structural Sources of Verb Meanings. *Language Acquisition, 1*(1), 3–55.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language, 35*(3), 515–531. https://doi.org/10.1017/S0305000907008641

Hallé, P. A., & de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development, 19*(4), 463–481. https://doi.org/10.1016/S0163-6383(96)90007-7

Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task*. *Journal of Child Language, 34*(4), 701–724. https://doi.org/10.1017/S0305000907008124

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54–69. https://doi.org/10.1037/a0028347

Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Acquisition of English Number Marking: The Singular-Plural Distinction. *Language Learning and Development, 2*(1), 1–25. https://doi.org/10.1207/s15473341lld0201_1

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Libertus, K., & Landa, R. J. (2013). The Early Motor Questionnaire (EMQ): A parental report measure of early motor development. *Infant Behavior & Development, 36*(4), 833–842. https://doi.org/10.1016/j.infbeh.2013.09.007

Liu, H.-M., Tsao, F.-M., Chang, C.-J., & 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(2), 221–240.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language, 57*(2), 252–272. https://doi.org/10.1016/j.jml.2007.03.005

Mani, N., & Plunkett, K. (2010). Twelve-Month-Olds Know Their Cups From Their Keps and Tups. *Infancy, 15*(5), 445–470. https://doi.org/10.1111/j.1532-7078.2009.00027.x

Marcus, G. F. (1995). Children's overregularization of English plurals: A quantitative analysis*. *Journal of Child Language, 22*(2), 447–459. https://doi.org/10.1017/S0305000900009879

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in Language Acquisition. *Monographs of the Society for Research in Child Development, 57*(4), i–178. https://doi.org/10.2307/1166115

Miller. (2016). Introduction English Plural Formation: Journal of English Linguistics. https://doi.org/10.1177/00754249922004615.

Moore, C., Dailey, S., Garrison, H., Amatuni, A., & Bergelson, E. (2019). Point, walk, talk: Links between three early milestones, from observation and parental report. *Developmental Psychology*.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perception: Implications for child language acquisition. *Cognition, 38*(1), 43–102. https://doi.org/10.1016/0010-0277(91)90022-V

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48*(1), 21–69. https://doi.org/10.1016/0010-0277(93)90057-3

Polka, L., & Werker, J. F. (1994). Developmental Changes in Perception of Nonnative Vowel Contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 16. https://doi.org/10.1037//0096-1523.20.2.421

Shi, R., & Lepage, M. (2008). The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science, 11*(3), 407–413. https://doi.org/10.1111/j.1467-7687.2008.00685.x

Storkel, H. L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language, 29*(2), 251–274. https://doi.org/10.1017/S0305000902005032

Sundara, M., White, J., Kim, Y. J., & Chong, A. J. (2021). Stem similarity modulates infants' acquisition of phonological alternations. *Cognition, 209*, 104573. https://doi.org/10.1016/j.cognition.2020.104573

Swingley, D., & Aslin, R. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*(2), 147–166. https://doi.org/10.1016/S0010-0277(00)00081-0

Swingley, D., & Aslin, R. (2002). Lexical Neighborhoods and the Word-Form Representations of 14-Month-Olds. *Psychological Science, 13*(5), 480–484. https://doi.org/10.1111/1467-9280.00485

Tincoff, R., & Jusczyk, P. W. (1999). Some Beginnings of Word Comprehension in 6-Month-Olds. *Psychological Science, 10*(2), 172–175. https://doi.org/10.1111/1467-9280.00127

Tincoff, R., & Jusczyk, P. W. (2012). Six-Month-Olds Comprehend Words That Refer to Parts of the Body. *Infancy, 17*(4), 432–444. https://doi.org/10.1111/j.1532-7078.2011.00084.x

Tsuji, S., & Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology, 56*(2), 179–191. https://doi.org/10.1002/dev.21179

Valian, V. (2006). Young Children's Understanding of Present and Past Tense. *Language Learning and Development, 2*(4), 251–276. https://doi.org/10.1207/s15473341lld0204_2

Valleau, M. J., Konishi, H., Golinkoff, R. M., Hirsh-Pasek, K., & Arunachalam, S. (2018). An Eye-Tracking Study of Receptive Verb Knowledge in Toddlers. *Journal of Speech, Language, and Hearing Research: JSLHR, 61*(12), 2917–2933. https://doi.org/10.1044/2018JSLHR-L-17-0363

Von Holzen, K., & Bergmann, C. (2018). A Meta-Analysis of Infants'. *Mispronunciation Sensitivity Development, 6*.

Wagner, L. (2001). Aspectual influences on early tense comprehension. *Journal of Child Language, 28*(03). https://doi.org/10.1017/S0305000901004792

Walle, E. A., & Campos, J. J. (2014). Infant language development is related to the acquisition of walking. *Developmental Psychology, 50*(2), 336–348. https://doi.org/10.1037/a0033238

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language, 59*(1), 114–132. https://doi.org/10.1016/j.jml.2008.03.001

White, K. S., Yee, E., Blumstein, S. E., & Morgan, J. L. (2013). Adults show less sensitivity to phonetic detail in unfamiliar words, too. *Journal of Memory and Language, 68*(4), 362–378. https://doi.org/10.1016/j.jml.2013.01.003

Zamuner, T. S., Moore, C., & 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. https://doi.org/10.1016/j.jecp.2016.07.012