

Moving Beyond “Nouns in the Lab”: Using Naturalistic Data to Understand Why Infants’ First Words Include *Uh-Oh* and *Hi*

Kennedy Casey¹, Christine E. Potter^{1, 2}, Casey Lew-Williams¹, and Erica H. Wojcik³

¹Department of Psychology, Princeton University

²Department of Psychology, University of Texas at El Paso

³Department of Psychology, Skidmore College

Why do infants learn some words earlier than others? Many theories of early word learning focus on explaining how infants map labels onto concrete objects. However, words that are more abstract than object nouns, such as *uh-oh*, *hi*, *more*, *up*, and *all-gone*, are typically among the first to appear in infants’ vocabularies. We combined a behavioral experiment with naturalistic observational research to explore how infants learn and represent this understudied category of high-frequency, routine-based non-nouns, which we term “everyday words.” In Study 1, we found that a conventional eye-tracking measure of comprehension was insufficient to capture U.S.-based English-learning 10- to 16-month-old infants’ emerging understanding of everyday words. In Study 2, we analyzed the visual and social scenes surrounding caregivers’ and infants’ use of everyday words in a naturalistic video corpus. This ecologically motivated research revealed that everyday words rarely co-occurred with consistent visual referents, making their early learnability difficult to reconcile with dominant word learning theories. Our findings instead point to complex patterns in the types of situations associated with everyday words that could contribute to their early representation in infants’ vocabularies. By leveraging both experimental and observational methods, this investigation underscores the value of using naturalistic data to broaden theories of early learning.

Public Significance Statement

Across languages, many of the first words that infants understand and say are routine-based, social words, such as *uh-oh*, *hi*, *more*, *up*, and *all-gone*. The current work leverages experimental and observational methods to investigate infants’ early representations of these often-overlooked but foundational everyday words. We find that conventional approaches to testing early word comprehension via eye tracking conflict with the real-world contexts that surround everyday words. Specifically, early learning cannot just be a process of mapping labels onto consistent visual referents. This multimethod approach underscores the value of using naturalistic data to broaden theories of early learning.

Keywords: word learning, language input, naturalistic recordings, ecological validity, eye tracking

Supplemental materials: <https://doi.org/10.1037/dev0001630.supp>

Word learning is often viewed as a mapping problem. For instance, children must determine that *cup* refers to the object that holds their juice and does not refer to their spoon, kitchen table, or other neighboring objects in their home. Impressively, by 9 months, infants reliably show evidence that they know the names of objects

that they see and interact with frequently (Bergelson & Swingley, 2012, 2015; Kartushina & Mayor, 2019; Parise & Csibra, 2012; Tincoff & Jusczyk, 1999, 2012). Various proposals have been put forth to explain such learning, including those that emphasize infants’ use of social-attentional cues, such as speakers’ eye gaze,

This article was published Online First October 12, 2023.

Kennedy Casey  <https://orcid.org/0000-0001-5326-3080>

This work was supported by the National Institute of Child Health and Human Development (F32HD093139 to Christine E. Potter, R01HD095912 to Casey Lew-Williams) and by Princeton University (fellowship awarded to Kennedy Casey). The authors thank the participating families and the members of the Princeton Baby Lab. The authors declare that there is no conflict of interest. All experimental protocols, including procedures for obtaining informed consent, were approved by the Princeton University Institutional Review Board (Approval: 0000007117, Language Learning: Sounds, Words, and Grammar). The data that support the findings of this study are available at <https://osf.io/tdbqn/>.

Kennedy Casey served as lead for conceptualization, formal analysis, investigation, methodology, and writing—original draft. Christine E. Potter served in a supporting role for conceptualization and methodology. Casey Lew-Williams served in a supporting role for conceptualization and methodology. Erica H. Wojcik served in a supporting role for conceptualization and methodology. Christine E. Potter and Casey Lew-Williams contributed equally to funding acquisition. Christine E. Potter, Casey Lew-Williams, and Erica H. Wojcik contributed equally to writing—review and editing and supervision.

Correspondence concerning this article should be addressed to Kennedy Casey, Department of Psychology, Princeton University, 17 Peretsman Scully Hall, Princeton, NJ 08540, United States. Email: kcasey@princeton.edu

to determine intended object referents (Çetinçelik et al., 2021) and infants’ tracking of word–object co-occurrences (Smith & Yu, 2008; Stevens et al., 2017). However, existing accounts of word learning primarily focus on infants’ ability to map nouns onto concrete objects and thus address only a subset of the words that young children learn.

“Everyday Words”: An Understudied Subset of Infants’ Early Vocabularies

Not all of infants’ early words are object names. Young children learn verbs, adjectives, and sounds, along with other words that do not tidily fit into established lexical categories. In fact, *uh-oh* is the fourth most commonly produced English word at 16 months, preceded only by *mommy*, *daddy*, and *ball* (Frank et al., 2017). Other non-nouns, such as *hi*, *more*, *up*, and *all-gone*, occur frequently in infant-directed speech and appear consistently in infants’ earliest productions across languages, as observed in prior research using diverse methods, such as diary studies, parent surveys, and naturalistic observations (e.g., Bates et al., 1994; Bloom et al., 1993; Bowerman, 1978; Caselli et al., 1995; Frank et al., 2021; Gleason et al., 1984; Lieven et al., 1992; Nelson, 1981; Tardif et al., 2008; Tomasello, 1987). These routine-based, social words (hereafter—“everyday words”) make up approximately 25% of the first 50 words in English and 25% of the first 10 words across 15 languages (Frank et al., 2017, 2021). Despite this prevalence, routine-based words remain not only underexplored in developmental research but also virtually unacknowledged in theories of early word learning. This oversight raises the question of whether current proposals of learning mechanisms, generated primarily from lab-based studies on the learning of concrete nouns, can adequately explain early word learning.

“Nouns in the Lab”: The Standard Approach to Assessing Early Comprehension Via Eye Tracking

The tendency of word learning theories to overlook everyday words could reflect entrenched assumptions about nouns holding a privileged position in the early lexicon, but it also may simply reflect the difficulty in visually depicting their meanings—a necessity for standard lab-based comprehension tasks. That is, what exactly does *uh-oh* look like? Language scientists have largely depended on eye-tracking studies to assess young children’s recognition of familiar words. For example, in the LWL procedure (Fernald et al., 2008), infants are typically presented with two side-by-side images of common objects or animals, and researchers measure their ability to look at the appropriate image when it is labeled. Looking time to the correct image serves as a proxy for comprehension because children tend to look more reliably at the target referent as their lexical knowledge develops (Fernald et al., 2006). This method critically depends on researchers’ ability to depict imageable, common referents for the target words, making this approach particularly well suited for testing comprehension of concrete nouns with object-based meanings.

The LWL procedure has also been applied to assess young children’s understanding of imageable words other than nouns, such as familiar verbs (Golinkoff et al., 1987; Valteau et al., 2018) and adjectives (Fernald et al., 2010; Forbes & Plunkett, 2019). Two eye-tracking studies have included a few everyday words when testing comprehension of non-nouns (e.g., verbs, adjectives, proper names) and point to some evidence of comprehension in 9- to 16-month-olds

(Bergelson & Swingley, 2013; Syrnk & Meints, 2017). However, infants’ comprehension of everyday words has yet to be systematically examined, and existing findings are limited due to the small number of everyday words studied and/or imageability confounds.¹

The Current Studies: Everyday Words in the Lab and in the World

To address this gap in our knowledge about infants’ understanding of early-learned everyday words, we first conducted an in-lab study. Following current conventions and best practices, we tested infants’ comprehension of everyday words with an eye-tracking paradigm and preregistered our design and planned analyses (<https://osf.io/z5qxf>). However, unlike studies of early-learned nouns, our study failed to reveal reliable evidence of everyday word understanding. We report the details of an unsuccessful behavioral experiment and consider why this implementation of an eye-tracking study did not allow infants to demonstrate their emerging knowledge of everyday words.

Motivated by the lack of evidence for comprehension in Study 1, we took a different approach to exploring infants’ representations of everyday words in Study 2. Specifically, we conducted detailed descriptive analyses of a naturalistic video corpus. In doing so, we captured novel characteristics of the real-world visual and social contexts that surround everyday words in the lives of infants. The results of this ecologically-motivated work introduce theoretical and methodological challenges for future research.

Study 1: Behavioral Experiment

Using a standard LWL paradigm, we tested whether infants could associate each of 12 everyday words with a corresponding referent (e.g., *uh-oh*: fallen cup; *hi*: person waving; see Open Science Framework [OSF] for all stimulus images). Since the target words are highly frequent in infant-directed language (MacWhinney, 2000) and are reported to be understood, on average, by 60.5% of English-learning 12-month-olds (Frank et al., 2017) as well as 69.7% of the infants in our sample (see “Method” section), we predicted that conventional looking-time measures would provide the first experimental evidence of early comprehension of everyday words.

Method

Participants

Participants were 33 full-term, monolingual North American English-learning infants (13 female), ranging in age from 10 to 16 months ($M = 13.2$ months) and recruited through a database of local families in New Jersey. All infants had no reported hearing or

¹ Syrnk and Meints (2017) included three everyday words (*bye-bye*, *night-night*, *no*) out of 18 tested words in an eye-tracking comprehension study with 9- to 13-month-old infants but did not report individual item effects. Bergelson and Swingley (2013) tested infants’ recognition of 14 non-nouns (seven verbs, two adjectives, five everyday words—*all-gone*, *bye*, *hi*, *more*, *uh-oh*) and reported evidence of comprehension in two age groups: 10- to 13-month-olds and 14- to 16-month-olds. The significant comprehension effect for 10- to 13-month-olds was primarily driven by a single verb pair (*kiss/dance*), and the word pairs with the highest comprehension scores for the 14- to 16-month-old group also included highly imageable verbs (*kiss/dance*, *eat/hug*) rather than more abstract everyday words.

vision impairments and were exposed to English at least 85% of the time. Parents provided informed consent, and participants received a small gift in exchange for their participation. All experimental protocols, including procedures for obtaining informed consent, were approved by the Princeton University Institutional Review Board (Approval: 000007117, Language Learning: Sounds, Words, and Grammar). We preregistered a sample size of 50 to match that of a similar previous study (Bergelson & Swingley, 2013). However, data collection was prematurely stopped due to COVID-19. At this point, 52 infants had participated but 19 did not meet inclusion criteria for analysis based on preregistered guidelines due to fussiness ($n = 9$), equipment malfunction ($n = 1$), or failure to contribute data on at least 50% of test trials ($n = 9$).

Stimuli

All experimental stimuli are available on OSF (<https://osf.io/tdbqn/>). Infants were presented with 24 test images (two per target word), organized into yoked pairs (*uh-oh/hi*, *wow/all-gone*, *bye-bye/yum*, *no/night-night*, *up/more*, *shh/thank-you*). Target words were determined based on MacArthur-Bates Communicative Development Inventory (CDI) production norms at 16 months (Frank et al., 2017) as well as frequency counts from the Child Language Data Exchange System (CHILDES, MacWhinney, 2000) corpus for children up to 17 months of age. The final 12 items were selected among candidate words from these two sources based on researchers' intuitions about imageability (see Table S1 in the online supplemental materials for stimuli characteristics).

Following established traditions, visual referents were chosen based on researchers' intuitions about what infants around 12 months of age typically see when they hear the target words produced in naturalistic contexts. As in prior eye-tracking studies with young children, two referents were chosen per target word to account for variability in infants' natural visual experience. To better inform our intuitions, we also collected informal pilot data from three caregivers of 12-month-old children. Across a 3-day span, caregivers documented any instance when their child heard one of the 12 target words by photographing the corresponding visual scene. Target images were then selected to represent naturalistic, infant-perspective referent scenes, informally matched for visual salience, and yoked such that they would not be easily confusable (e.g., *bye-bye* was not matched to *hi* since both involve a person waving).

Auditory stimuli consisted of natural recordings of the 12 target words (two tokens per word), produced in infant-directed speech by a female native English speaker and left unedited to preserve typical intonation. Target words ranged from 410 to 1,240 ms in duration ($M = 820$ ms) and were normed to a standard mean intensity of 65 dB (Boersma & Weenink, 2016).

Procedure

Using the LWL procedure (Fernald et al., 2008), we tested infants' comprehension of everyday words. Participants sat on their parents' lap in a dimly lit testing room and viewed two images at a time on a 55-in. TV monitor display. Parents wore opaque sunglasses and were asked to avoid directing their child's attention during the study. All 12 target words were tested twice, each time with a different target image, resulting in 24 test trials. Trial order was pseudorandomized such that there were at least two trials between repetitions of a word ($M = 11$ trials between repetitions), and the target side was

counterbalanced to ensure an equal number of left and right target trials across the study. Participants were randomly assigned to one of two trial orders, where the second was the reverse of the first. Infants first saw two practice trials with common concrete nouns as targets (*ball* and *dog*). On each of the following test trials, images appeared on the screen for 2,500 ms before the onset of the auditory stimulus. Then, participants heard two repetitions of the target word in isolation, with a fixed 2,200 ms delay between the onset of the first and second repetition of the word. After the second repetition, images remained on the screen for an additional 2,000 ms, resulting in test trials that were 6,900 ms long (+500 ms intertrial interval). Infant-friendly, nonverbal videos or verbal reinforcement appeared between blocks of four test trials to maintain engagement.

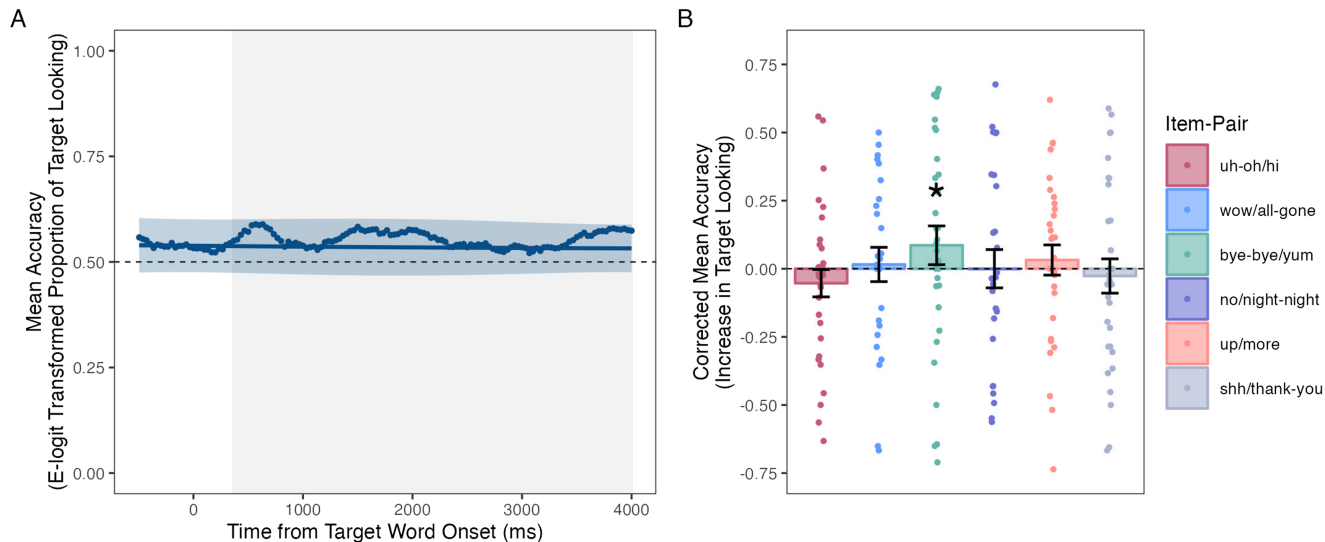
After the study, parents completed a vocabulary and image familiarity survey, where they were asked to report whether their child understood and/or said each of the 12 target words and to rate the visual stimuli (on a scale from 1 to 4) for whether they considered them to be congruent with infants' typical at-home experience with each word. Results from the vocabulary checklist indicated that all 12 everyday words were understood by a majority of participants. On average, 69.7% of infants reportedly understood ($SD = 13.3%$, range = 45.5%–84.8%) and 14.4% reportedly produced each target word ($SD = 10.1%$, range = 3.0%–30.3%). On average, parents attributed a relatively high degree of familiarity to our target images, meaning that the chosen scenes were thought to match infants' typical visual experience with the tested words. The mean image familiarity rating across words was 2.9/4 overall ($Mdn = 3.0$, $SD = 1.0$) and 3.1/4 for the subset of infants with reported target word comprehension ($Mdn = 3.0$, $SD = 0.9$). Additionally, children's global receptive and expressive vocabulary knowledge was assessed using the CDI: Words and Gestures (Fenson et al., 1994), but we found no relationship between vocabulary size and LWL task performance (see the online supplemental materials for details).

Infants contributed usable data on 19.4 out of 24 test trials, on average ($SD = 3.4$ trials). Videos of participants' eye movements were coded offline, frame by frame, in 33 ms intervals. For each frame, a trained coder, naïve to the target condition, determined whether the infant was looking at the left image, right image, away from both images, or shifting between images. Following our preregistered plan, trials were excluded if participants looked away from both images for 50% or more of the critical window (367–4,000 ms following target word onset; consistent with Bergelson & Swingley, 2013). To ensure reliability in gaze coding, 20% of videos were recoded by a second coder. Intercoder reliability was high, with coders agreeing on infants' gaze location within a single frame on 99% of frames overall, and for a more conservative measure, on 97% of frames surrounding shift events.

Results

Overall performance on the LWL task was at chance level ($M = 0.51$, $p = .96$, Wilcoxon test). That is, infants did not look reliably more to labeled target referents, relative to unlabeled distracters, during the critical window from 367 to 4,000 ms following target word onset. To better understand the dynamics of looking behavior, as planned in our preregistration, we analyzed the time course data using growth curve analysis (GCA, Mirman, 2014; see the online supplemental materials for model details). Infants' accuracy in looking to target images was above chance based on the intercept term

Figure 1
Study 1 (Behavioral Experiment) Results



Note. (A) Growth curve estimates of empirical logit-transformed mean proportion of target looking during the critical window, offset by 0.5. Points reflect model-predicted means over participants and items for each 33 ms time bin in the critical window. (B) Saliency-corrected mean accuracy across item pairs. Accuracy was calculated as the proportion of time spent looking at a given test image when it appeared as the labeled target image, relative to the proportion of time spent looking at the same image when it appeared as the unlabeled distracter. Points reflect means over participants. Chance-level performance (0.0) is indicated by the dashed line. Error bars reflect standard errors of the mean, and asterisks denote above-chance performance after correcting for multiple comparisons ($p < .008$). See the online article for the color version of this figure.

(Estimate = 0.52, $SE = 0.06$, $t = 8.29$, $p < .001$; Figure 1A). However, both the linear and quadratic time terms were not significant, suggesting that participants' looks to the target image did not increase significantly after target word onset (Estimate = 0.01, $SE = 0.19$, $t = 0.03$, $p = .98$) and looking behavior did not follow a parabolic trajectory (Estimate = 0.11, $SE = 0.12$, $t = 0.89$, $p = .38$). While the significant intercept term points to some weak evidence of everyday word comprehension, the profile of target looking in the current study does not match the profile that is seen for early learned concrete nouns at the same age (e.g., Bergelson & Swingley, 2015). Moreover, exploratory analyses of item effects (Figure 1B; Figure S2 in the online supplemental materials) as well as analyses of age, parent-reported target word comprehension, and target image familiarity effects found no reliable evidence of comprehension of everyday words (see the online supplemental materials, pp. 6–13).

Discussion

Results of Study 1 revealed that, on average, infants did not significantly increase their looking to the target image after word onset. While we found no overall evidence of comprehension, it is possible that the LWL paradigm may be able to capture infants' comprehension of certain everyday words (in this case, *bye-bye/yum*, though potential visual saliency effects may still be at play). Conclusions from this study are limited due to the small sample size, but even the oldest infants in our sample did not show reliable recognition of the target words. Conventional looking-time measures provided little to no evidence of comprehension even though the tested words were reported to be understood by a

majority of participants and the test images were considered by caregivers to match infants' typical visual experience. These null effects could be due to a true lack of understanding of everyday words, but that possibility conflicts with evidence of early comprehension and production obtained from thousands of parent-report surveys (Frank et al., 2017).

Because caregivers reliably report that infants, both generally and in the current study, do understand everyday words, the failure of this behavioral experiment could be explained, in part, by stimulus limitations. While our analyses control for saliency differences between images (consistent with Bergelson & Swingley, 2012, 2013), the potential for detecting above-chance comprehension is diminished if the images within yoked pairs are not similarly salient (e.g., *hi* = person waving vs. *uh-oh* = fallen cup may not attract equal attention). Additionally, static images may fail to fully represent the meanings of routine-based words, whereas dynamic stimuli (e.g., Bergelson & Swingley, 2013; Szyrak & Meints, 2017) could better match naturalistic referential scenes.

The failure of this behavioral experiment to provide evidence of comprehension could also stem from poor adult intuitions about infants' representations of the target words. Although our visual stimuli for Study 1 were validated by parent report, it is nevertheless possible that adults' assumptions about common referents failed to resemble infants' typical visual experience associated with these words (i.e., *uh-oh* might not regularly occur in the presence of a fallen cup). This mismatch is particularly possible given the routine-based nature of everyday words. Whereas the experimenter-intuited referents of early-learned concrete nouns (e.g., *cup* or *ball*) are likely to match the common tangible objects seen by infants when these words are uttered (Custode & Tamis-LeMonda, 2020), the meanings

of everyday words may not be as adequately expressed by unimodal, static images chosen by developmental scientists. Thus, we next conducted a naturalistic observational study, one goal of which was to identify better stimuli—either static or dynamic—that could be used to evaluate infants' comprehension of everyday words.

Study 2: Video Corpus Analysis

Current theories of word learning primarily explain infants' learning of concrete nouns with stable visual referents and are supported by evidence of robust noun–object co-occurrence in naturalistic word learning environments (Custode & Tamis-LeMonda, 2020; Pereira et al., 2014). However, without knowing the characteristics of the input surrounding everyday words, we cannot yet determine whether the same principles used to account for the learning of concrete nouns also extend to other early-learned words, including everyday words.² One prior study found that some abstract words, including several routine-based words like *uh-oh* and *bye-bye*, were less likely than concrete nouns to be uttered in the presence of their referents (Bergelson & Swingley, 2013), but these data only describe whether or not the researcher-intuited referent was present when the word was uttered. This top-down approach, also used in Study 1, contrasts with the primarily bottom-up approach used in the next part of our investigation.

Study 2 bypassed adult assumptions about word meanings by providing a detailed descriptive analysis of the actual input surrounding everyday words in longitudinal, at-home recordings of five 1-year-olds from the Providence corpus (Demuth et al., 2006). Prior studies have described examples of some early meanings of everyday words (e.g., *up* referring to a child being picked up, Tomasello, 1987; or *bye-bye* being associated with people leaving, Gleason & Weintraub, 1976). Here, the primary aim of our naturalistic observational research was to describe *all* attested meanings in children's input and thereby obtain a comprehensive picture of the possible referents associated with everyday words.

We first quantified the consistency with which everyday words co-occurred with different visual referents. However, after quickly discovering the vast variability of the visual referents associated with everyday words in a first coding pass, we also defined “situational contexts”—broader categories of routine-based usage that collapse across basic visual features of referent scenes (e.g., the exact objects present) to acknowledge and describe some level of consistency in children's input. As an example, if the word *uh-oh* was produced when a child dropped their cup, the visual description would be “cup falling,” and the situational description would be “object falling.” A secondary aim was to assess the validity of our approach in Study 1, and we did this with respect to both the visual and situational appropriateness of our stimuli. By describing the characteristics of infants' real-world input, we can evaluate whether current theories of word learning can explain infants' early learning of everyday words.

Method

Video Corpus

We annotated longitudinal, at-home recordings for five monolingual North American English-learning children with video data available in the Providence corpus (Demuth et al., 2006). Video recordings began at the age of first-word onset (11–16 months) and continued for up to 3 years. All recording sessions up to 24

months ($N = 114$) were annotated in the present study. Children and their household members were recorded for ~1 hr every 1–2 weeks during this period. The videos capture spontaneous, highly naturalistic interactions that span a variety of activity contexts (e.g., free play, book reading, mealtime) and physical locations within children's homes (e.g., living room, bedroom, kitchen, yard).

Coding Procedure

All utterances containing one of the 12 target everyday words were found using the PhonBank online database (Rose & MacWhinney, 2014; <https://sla.talkbank.org/TBB/phon/Eng-NA/Providence>). We watched the video accompanying each utterance to determine the visual scene and usage context of the target word. As needed, we relied on information from preceding or following utterances to better understand the intended usage. Tokens were marked as “indeterminable” and excluded from analyses if the usage could not be reasonably inferred from the available media (5.8% of the data set), most often due to the speaker/target child being out of frame during the moment of target word production.

We extracted 11,920 total tokens across all 12 target everyday words (Table S4 in the online supplemental materials). Frequency varied considerably across items and children, likely reflecting natural variation along with sampling biases. For instance, *night-night* had the lowest token count, at least in part because recording sessions primarily took place during daytime hours, and counts for *yum* varied drastically across children because some families did not record during mealtime. Crucially, though, even for low-frequency items, we extracted over 150 total tokens, and 11 of 12 items included tokens sampled from all five households ($M = 200.2$, range = 30.6–633.8 tokens per household).

For each token, we annotated eight variables of interest, including (1) the speaker who produced the target word, (2) whether the speech was directed to the target child or overheard, (3) whether the speaker (if not the target child) was in view of the camera, (4) whether the target child was in view of the camera, (5) the visual referent, (6) whether the token was a visual match to our experimental stimuli, (7) the situational context, and (8) whether the token was a situational match to our experimental stimuli. Variables 5–8 are the targets of analysis in the present study.

Visual referents were coded as the exact objects, actions, and/or people involved in the referent scene (e.g., *uh-oh*: cup falling, block tower falling, child falling, crayon breaking, phone ringing, etc.). Because poor video quality occasionally compromised our ability to determine the speaker/target child's view when the word was uttered, our coding reflects the “idealized” referent for each token, determined based on a combination of the available video and audio. For instance, the referent of *uh-oh* could be coded as “cup falling” even if the video did not capture the dropping event or if there was temporal displacement between the dropping event and utterance of the target word, so long as the audio unambiguously reflected this intended usage. Situational contexts refer to broader clusters of usage that describe, at a level higher than the specific visual features of the scene, the

² There is an existing disagreement surrounding whether more abstract, routine-based words (e.g., *hi*, *bye-bye*) are referential words (e.g., Gleason & Weintraub, 1976; Lieven et al., 1992). Even if everyday words are interpreted as a qualitatively different verbal signal, description of the input surrounding production is still critical for determining *how* they become different words from the perspective of the learner.

intended meaning of the target word (e.g., *uh-oh*: any object falling). Situational context categories were generated by grouping together similar referents from the visual-level coding (e.g., *uh-oh*: cup falling and block tower falling were coded as the same situational context since both involved an object falling; see Table 1 for the top three situational contexts by item). To evaluate the validity of our experimental stimuli, we also characterized referent scenes according to whether they matched the visual referents and/or situational contexts depicted in the test images from Study 1 (e.g., *uh-oh*: cup falling = “visual match,” any object falling = “situational match”).

To ensure reliability in descriptive coding, a randomly selected five tokens per target word per child were recoded by a second coder. High reliability was achieved through an iterative training process (see OSF for coding manual). This process involved discussion of discrepancies between the primary and naïve coders on a separate training set (five tokens per target word per child), with initial estimates of intercoder reliability at 84%. Following training on the descriptive coding scheme, coders agreed on 91% of decisions.

Results

In the sections that follow, we report on all 11,920 observed instances of the 12 target words. Our primary analyses combine child-produced ($N = 3,086$, 25.9% of the data set) and other-produced tokens ($N = 8,834$, 74.1%), as we found no statistically significant differences in the reported measures across child versus adult speakers (see the online supplemental materials, pp. 20–21, 25, 28 for details). We analyzed all other-produced tokens, including those directly addressed to the target child ($N = 8,127$, 91.9%) and directed to others nearby ($N = 715$, 8.1%), as inclusion of non-child-directed tokens did not change any of the reported effects. Given the relatively low proportion of “indeterminable” use contexts (5.8% overall, 12.8% when the target child was not in frame, 8.3% when a non-target-child speaker was not in frame), we retained all tokens, regardless of which interactants were in view of the camera.

Everyday Words Do Not Reliably Co-Occur With Consistent Visual Referents

We first determined the frequency with which everyday words appeared with different visual referents (e.g., *uh-oh*: cup falling, block tower falling, child falling, crayon breaking, etc.). Our

analyses revealed that the visual input surrounding everyday words exhibits high variability within and across households, over time, and in both infants’ and adults’ productions.

We found that the most common visual referent for each everyday word in the naturalistic corpus (e.g., child falling for *uh-oh*, cereal for *more*, or Play-Doh for *wow*; see Table 2) captured a small fraction of the input. The top visual referent co-occurred with only 9.0% of word tokens on average (range = 1.4%–20.4%; Figure 2A) and never matched for all five children. In fact, out of 4,120 total unique visual referents (summed across all words), only 13 appeared at least once for all children in the corpus (Table S6 in the online supplemental materials). Although this cross-household variability stems in part from sampling limitations of the corpus, it suggests that individual children’s early representations may be idiosyncratic and underscores the visual inconsistency of everyday words, which is reflected in children’s productions (Figure S11 in the online supplemental materials) and persists across developmental time (Figures S15–S17 in the online supplemental materials).

Everyday Words Appear in Diverse Situational Contexts, But There Are One or More Dominant Contexts for Each Word

In a second set of analyses, we determined the frequency with which everyday words appeared with different situational contexts (e.g., *uh-oh*: any object falling, any person falling, or any object breaking). It is possible that despite being visually inconsistent, the input surrounding everyday words is more stable when considering the general context of the event-based scene. Although we still found vast variability at this broader level of the input, everyday words co-occurred more reliably with consistent situational contexts than with visual referents. On average, the most common single situation accounted for 48.1% of the input (range = 12.8%–94.0%; Figure 2B). The top three situational contexts captured a majority of tokens for 10 of 12 items, with proportions ranging from 31.5% to 100.0% of tokens ($M = 72.4%$). This pattern also appeared across children. The top situation accounted for 45.2% of each child’s input, on average (range = 39.8%–94.0%; Figure S14 in the online supplemental materials), and four of five children shared the same top situation, on average across words (range = 2–5). The dominance of the top situations also held across time

Table 1
Top Situations by Target Word

Target word	Most common situation	2nd most common situation	3rd most common situation
<i>all-gone</i>	Done playing	Done eating	Person leaving
<i>bye-bye</i>	Person leaving	Done interacting with animal	Putting away other inanimate object
<i>hi</i>	Initiating interaction with person	Answering phone	Initiating interaction with animal
<i>more</i>	Eating/requesting more food	Playing with/requesting more toys	Requesting more of another inanimate object
<i>night-night</i>	Going to bed	Putting toy to sleep	Pretending to sleep
<i>no</i>	Not wanting to do something	Doing something not allowed	Not wanting someone else to do something
<i>shh</i>	Someone sleeping	Someone screaming	Someone crying
<i>thank-you</i>	Exchanging object	Being helpful	Following directions
<i>uh-oh</i>	Object falling	Person falling	Object breaking
<i>up</i>	Referring to location (e.g., “up here”)	Cleaning up	Picking up person
<i>wow</i>	Playing with toy	Performing impressive action	Referring to other inanimate object
<i>yum</i>	Eating/referring to food	Pretending to eat	N/A

Note. N/A = not applicable.

Table 2
Top Referents by Target Word

Target word	Most common referent	2nd most common referent	3rd most common referent
<i>all-gone</i>	Crayons gone	Done reading	Cookies gone
<i>bye-bye</i>	Dad leaving	Camera (done recording)	Ending phone call
<i>hi</i>	Mom greeting child	Answering phone	Answering pretend phone call
<i>more</i>	More cereal	More beans	More cheese
<i>night-night</i>	Goodnight book	Animals going to bed in book	Eyes closing in flap book
<i>no</i>	Child not wanting to read	Child not wanting to share toys with mom	Child not wanting specific book
<i>shh</i>	Wheels on the bus song—parents on the bus say <i>shh</i>	Hush song—little babies love to sleep	Child screaming
<i>thank-you</i>	Child giving mom Play-Doh	Child giving mom block	Child giving mom bubble wand
<i>uh-oh</i>	Child falling	Microphone falling off	Broken crayon
<i>up</i>	Mom picking child up	Pop-up toy	Pop-up book
<i>wow</i>	Play-Doh	Bubbles	Child falling
<i>yum</i>	Yogurt	Lunch (unspecified)	Cheese

(Figures S15–S17 in the online supplemental materials) and was reflected in both infants' and adults' productions (Figure S13 in the online supplemental materials). Thus, while the visual information associated with everyday words is highly variable, the situational-level input appears more stable, both within and between households and over time.

Experimenter-Intuited Stimuli From Study 1 Depicted Plausible Visual Referents and Generally Matched Dominant Situational Contexts

To assess the validity of our experimental stimuli, we determined the frequency with which the visual referents and situational contexts depicted in the images from Study 1 appeared in the naturalistic corpus. Unsurprisingly given the high degree of visual variability associated with everyday words, we found that visual matches to our experimental stimuli were rare across words ($M = 7.7\%$, range = 0.0%–17.4%; Figure S7 in the online supplemental materials). However, the exact Study 1 visual referents appeared at least once for 11 of 12 everyday words, providing evidence that our experimenter-intuited scenes were plausible referents. Moreover, the top situations that emerged in the naturalistic corpus largely conformed with our broad intuitions about everyday word meanings. Situational matches to Study 1 stimuli were common ($M = 46.4\%$, range = 7.3%–81.9%) and even mapped onto the top situational context for 9 of 12 items. Together, these findings suggest that infants' failure to look to the appropriate images in Study 1 cannot necessarily be attributed to implausible test scenes and instead demonstrate that everyday words may not easily be captured by a single visual token, both within and across infants.

Exploratory Analyses: Visual and Situational Stability May Contribute to Early Learnability

Next, we explored how the consistency with which everyday words co-occurred with different visual referents and situational contexts may affect children's learning. That is, do children learn an everyday word earlier if it appears with consistent visual referents and situational contexts? We hypothesized that children whose visual and situational input for a given everyday word was more stable would be more likely to produce that word, relative to infants whose input was more variable.

We fit a linear mixed-effects model predicting children's rate of production of individual target everyday words, relative to their total rate of production, on the basis of visual and situational stability of their own input.³ Visual stability was calculated as the proportion of tokens in the child's input (excluding productions from the participating child) corresponding to a repeated visual referent for each item and child (e.g., for *uh-oh*, what proportion of the time did each child hear this word in the presence of a visual referent that had appeared previously?). Situational stability was measured as the proportion of tokens (excluding productions from the participating child) mapping onto the top situational context for each item and child (e.g., for *uh-oh*, what proportion of the time did each child hear this word used to refer to any object falling?). We found that both visual (Estimate = 1.11, $SE = 0.44$, $t = 2.53$, $p = .01$) and situational (Estimate = 0.72, $SE = 0.35$, $t = 2.10$, $p = .04$) stability positively predicted production. Notably, visual and situational stability were not highly correlated (e.g., even though *uh-oh* occurred in the situational context of an object falling with some reliability, the specific object(s) involved—the visual-level information—varied widely). This result suggests that infants who heard a given word in the presence of more consistent visual referents and more consistent situational contexts were more likely to produce the word, relative to infants who experienced more variable visual and situational input. Thus, stability at multiple levels of children's input could support their learning.

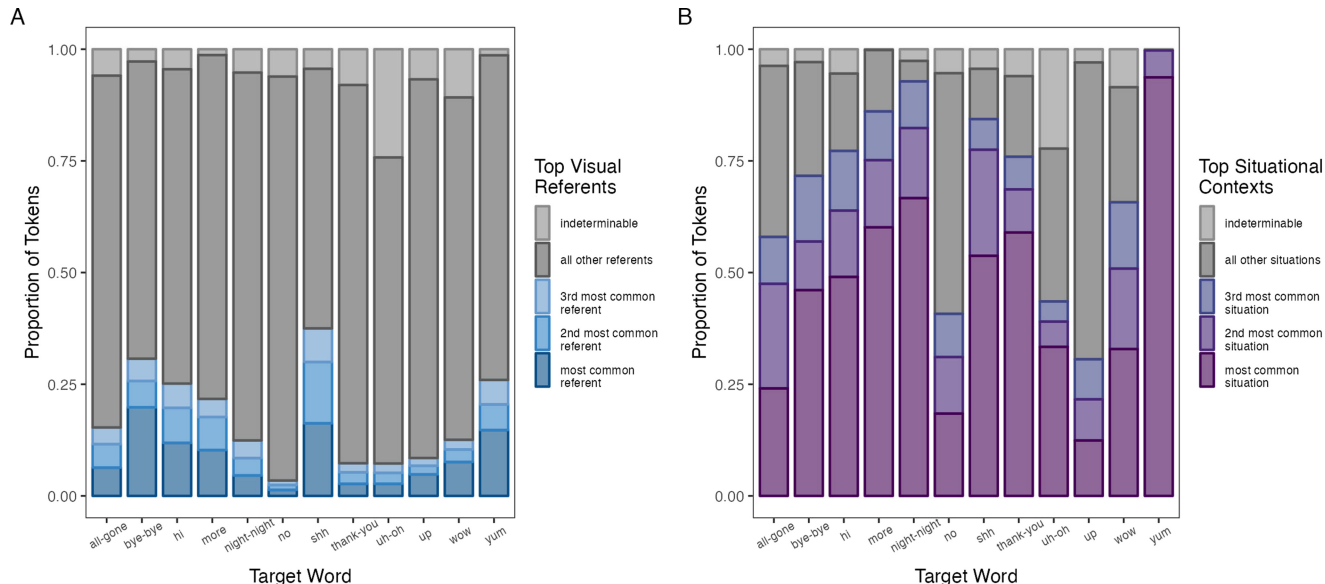
Discussion

Results of Study 2 demonstrated that the visual input surrounding everyday words is highly variable. While dominant theories of word learning tend to rely on referent stability, we found that co-occurring visual referents varied both within and across children. This observation challenges long-held assumptions that imageable, concrete words are the first learned (e.g., Gentner, 1982) because everyday words—despite

³lmer(everyday_word_production_rate ~ visual_stability + situational_stability + (1|word) + (1|child)), where everyday word production rate was calculated as the total number of everyday word tokens produced divided by the total number of all word tokens produced across all recording sessions up to 24 months (scaled, extracted using the *childesr* package, Braginsky et al., 2021). Children produced, on average, 928.2 unique word types across all recording sessions ($Mdn = 587$, range = 237–2,368).

Figure 2

Study 2 (Video Corpus Analysis) Results



Note. (A) Distribution of top visual referents across target words. (B) Distribution of top situational contexts across target words. Dark gray bars refer to tokens not matching the top three visual referents, and light gray bars refer to tokens where intended usage was indeterminable (e.g., due to the speaker and target child being out of frame). The plot for visual referents shows particularly large variability beyond the top three referents, while the plot for situational contexts shows more stability at this level of the input. See the online article for the color version of this figure.

their visual inconsistency—are typically among the first to appear in infants’ vocabularies (Frank et al., 2017, 2021). These findings also call into question the validity of using standard eye-tracking procedures to assess infants’ knowledge of less imageable, routine-based words and therefore help explain the lack of reliable looking to target images in our behavioral experiment.

Our intuited visual referents from Study 1 appeared rarely in the naturalistic video corpus (~8% of tokens). However, rather than reflecting invalid adult assumptions about early meanings, the low proportion of exact matches to our experimental stimuli underscores the visual inconsistency associated with everyday words. Even the most common visual referent captured only ~9% of the input and also varied across children, suggesting that the meanings of everyday words are unlikely to be fully represented by a single visual token or event. Notably, there is greater variability in the visual scenes surrounding everyday words compared to the variability between exemplars within a concrete object category (Clerkin et al., 2017). While the distribution of different *cup* instances, for example, is anchored on a frequent prototypical visual token and described by a function, the variability in everyday word visual referents has no obvious anchor or unifying structure. Thus, the characteristics of the visual input surrounding everyday words appear largely incompatible with current approaches to both testing and understanding early word learning.

Everyday words also appeared in diverse situational contexts, though the input was generally skewed toward one or more dominant contexts for each item. The top situation accounted for nearly half of the input, on average across words. That is, *uh-oh* does refer to things falling with some reliability. This finding suggests that even though the visual information associated with everyday words varies considerably, the situational information surrounding utterances appears

more stable and—if children focus on this context level—could support word learning. Still, it is notable that the situational contexts associated with everyday words were far from uniform, particularly compared to the stability of the visual input surrounding concrete nouns. While concrete nouns have been found to co-occur with the same referent over 80% of the time (Bergelson & Swingley, 2013; Custode & Tamis-LeMonda, 2020; Pereira et al., 2014), we found that everyday words co-occurred with the same situation only 48% of the time, indicating that infants must contend with substantial variability even at this broader level of the input to learn everyday words.

General Discussion

Using experimental and descriptive approaches, this investigation examined a common yet generally unexplored category of early-learned words, which we term everyday words. These high-frequency, routine-based words, such as *uh-oh* and *hi*, are well represented in infants’ earliest productive vocabularies, yet their learnability cannot be explained by current word–referent association models of word learning. Study 1 used eye-tracking measures to test whether 1-year-old infants could associate everyday words with experimenter-intuited visual referents but failed to detect evidence of reliable comprehension. Study 2 revealed that the real-world visual input surrounding everyday words is highly unstable but that these words may appear in relatively more consistent situational contexts.

Limitations of Lab-Based Measures of Word Comprehension

Early word knowledge encompasses more than just the ability to map labels onto stable, concrete referents. For example, infants must

generalize words to new contexts and use their limited productive vocabularies to communicate their socio-emotional needs. Yet, lab-based studies generally define word comprehension in a rather restricted sense. By requiring that infants look at, or even point to, something to demonstrate understanding, standard behavioral measures may be insufficient to capture infants' emerging representations of words that may not have direct visual associates (Wojcik et al., 2022), or these measures may reflect behavior that is more expressive than referential (e.g., Dore, 1974). When presenting only two-dimensional static images of intended referents, highly controlled lab-based experiments generally lack the richness and diversity of infants' real-world language environments (Nastase et al., 2020; Reuter et al., 2021; Tamis-LeMonda et al., 2017) and may therefore eliminate connections to socio-emotional, pragmatic, or contextual information and other cues that are useful for word recognition. Dynamic stimuli might better approximate naturalistic referential contexts, especially for everyday words. Using videos rather than static images could improve the odds of detecting comprehension via eye tracking for more routine-based words that may serve as responses to dynamic events in naturalistic contexts (e.g., Bergelson & Swingley, 2013; Syrnk & Meints, 2017). However, this experimental design logic still assumes that there exist prototypical, generalizable, and imageable referents for each word, and therefore remains largely inconsistent with the characteristics of the naturalistic visual input observed in our study.

The low dimensionality of common lab-based measures may be especially problematic for everyday words, which by their routine-based nature may be tied to less-visible situational factors, such as affective state (e.g., *uh-oh* is likely to be produced under more dramatic circumstances than many concrete nouns, Ponari et al., 2018) or event timing (*uh-oh* is likely to coincide with salient event transitions, such as when the child falls, *hi* is likely to occur when a new interaction begins, which may create opportunities for learning during moments of heightened attention, Kosie & Baldwin, 2019). Furthermore, early word knowledge may be highly idiosyncratic, as evidenced by the cross-household variability observed in our study, making it difficult to capture early comprehension with one standard set of stimuli.

The Richness and Complexity of Learning Contexts for Everyday Words

Naturalistic contexts for word learning are visually complex and introduce challenges for young learners (Medina et al., 2011), but these real-world contexts may also provide opportunities to detect previously ignored regularities in the input. Existing research explains infants' ability to learn nouns, in particular, by exploiting consistent referent-linked visual cues to word meaning. However, this narrow conception of word knowledge as the ability to link labels with their concrete, imageable, consistent visual referents cannot fully account for infants' early learning of everyday words, for which other cues may better predict learning.

Our video corpus analysis identified situational stability as one such potential cue. Even if *uh-oh* does not reliably co-occur with the visual referent of a fallen cup, for example, this word may appear consistently when a variety of objects fall, perhaps providing enough information for infants to detect this broader meaning-relevant pattern. The skewed nature of the situational-level input could support learning by facilitating efficient recognition of words that appear in their dominant

situational contexts, while also promoting successful generalization to novel situations, given that meanings are not tightly constrained to a single context. Not unique to this set of everyday words, skewed distributions are a common feature of children's real-world linguistic (Goldberg et al., 2004; Montag et al., 2018) and visual input (Clerkin et al., 2017; Jayaraman et al., 2015) and have been identified as supportive of learning and generalization (Boyd & Goldberg, 2009; Carvalho et al., 2021; Maguire et al., 2008). Consistent with this view, we uncovered a preliminary link between regularity of the situational-level input and infants' rate of everyday word production (i.e., infants who heard a given target word in more consistent situational contexts may have been more likely to produce the word, relative to infants who saw the same word used in more variable situations).

It is worth noting, however, that while we aimed to avoid top-down assumptions, our categorization of situations for everyday words required adult decisions that may not map onto infants' parsing of complex, naturalistic scenes. Moreover, our decisions about categories may not map into infants' learning processes. This question of where to divide meaningful "units" of information (i.e., between situational contexts that involve many dimensions of information, such as objects, actions, people, emotions, etc.) is not specific to the present study. Even when operating within a single domain, like vision (Johnson et al., 2009; Slone & Johnson, 2015), speech (Frank et al., 2010; Mattys et al., 2005), or event processing (Baldwin & Kosie, 2021), there are similar questions surrounding the level at which learners segment and encode information. Still, it is promising that our coded context-level features were able to capture potentially meaningful variation in infants' production (consistent with Roy et al., 2015). A key contribution of this work is the recognition that there may be learnable patterns in broader aspects of the input surrounding words, visual or otherwise, which remain largely overlooked in conventional studies of noun-to-object mapping. Future work is needed to determine how infants attend to and encode information in a given situation, and how this information is integrated across multiple episodes.

Additionally, it is worth noting that even within the set of early-learned everyday words investigated in the present study, there is considerable individual variation in the input. That is, while the general patterns of visual variability and situational stability appear for the five children and 12 words in our sample, idiosyncrasies at the child and item level cannot be ignored. Combined with evidence of idiosyncrasies in adult word representations (Wang & Bi, 2021), our findings suggest that early word meanings may be specific to infants' unique experiences. Word learning theories must account for individual trajectories (Samuelson, 2021) and acknowledge that word representations need only be similar enough from person to person to successfully communicate. Furthermore, theories must address how infants may attend to a diverse array of cues in the environment, beyond object-specific visual characteristics. As identified in several previous studies, useful characteristics include overall label frequency (Goodman et al., 2008), frequency of occurrence in isolation (Brent & Siskind, 2001; Lew-Williams et al., 2011) or in the sentence-final position (Braginsky et al., 2019), strong association with infancy (i.e., "babiness": Perry et al., 2015), short mean length of utterance and high contextual distinctiveness (Roy et al., 2015), relatedness of word meanings (Floyd & Goldberg, 2021), and social cues such as eye gaze and referential gestures (Yu & Ballard, 2007). However, these features still cannot scale to a comprehensive theory of early word learning, and it remains unclear how multiple features

are encoded and integrated within and across episodes. Together, recognized key predictors of word learning only account for up to 29% of the variance in age of acquisition across 386 English words on the MacArthur-Bates CDI (Fenson et al., 1994) and notably less for words that do not fit into established lexical categories (Braginsky et al., 2016).

This article takes a first step toward understanding infants' early learning of “everyday words.” While grouped together here to contrast with commonly studied words from other word classes (nouns, verbs, adjectives, sounds), routine-based words within this set are diverse. Each everyday word could make its own case study, and specific theories about each word's acquisition could be informed by prior research documenting children's use of these words over their first several years of life (e.g., *uh-oh* emerging as a response to the failure of an attempted action: Gopnik & Meltzoff, 1984; *more* first serving as a request and then becoming a quantifier, Gathercole, 1985; Weiner, 1974; *no* as the earliest-produced form of negation: Szabó & Kovács, 2022; *thank-you* as an explicitly taught and repeatedly prompted politeness routine: Gleason et al., 1984). Future research can continue to investigate how these words are first learned and produced, in addition to how children's semantic representations may change over development.

Looking forward, continued rigorous manual coding efforts (Mendoza & Fausey, 2021) and advances in machine learning that allow for automatic processing of even larger amounts of naturalistic data (Orhan et al., 2020; Tsutsui et al., 2020) will position the field to be able to uncover novel features of infants' real-world input that may be supportive of word learning. Such work will greatly benefit from the increasing availability of modern video corpora that capture the first-person perspective of the learner across time (Bergelson et al., 2019; Sullivan et al., 2021). By embracing bottom-up approaches to describing the distributions of cues in the input, including multisensory, social, and contextual information, we can better understand how infants build early word representations.

References

- Baldwin, D. A., & Kosie, J. E. (2021). How does the mind render streaming experience as events? *Topics in Cognitive Science*, 13(1), 79–105. <https://doi.org/10.1111/tops.12502>
- Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., Reilly, J., & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, 21(1), 85–123. <https://doi.org/10.1017/s0305000900008680>
- Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, 22(1), Article e12715. <https://doi.org/10.1111/desc.12715>
- Bergelson, E., & Swingle, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, 109(9), 3253–3258. <https://doi.org/10.1073/pnas.1113380109>
- Bergelson, E., & Swingle, 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., & Swingle, 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>
- Bloom, L., Tinker, E., & Margulis, C. (1993). The words children learn: Evidence against a noun bias in early vocabularies. *Cognitive Development*, 8(4), 431–450. [https://doi.org/10.1016/S0885-2014\(05\)80003-6](https://doi.org/10.1016/S0885-2014(05)80003-6)
- Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer* (Version 6.0.20) [Computer software]. <http://www.praat.org/>
- Bowerman, M. (1978). The acquisition of word meaning: An investigation into some current conflicts. In N. Waterson & C. Snow (Eds.), *The development of communication* (pp. 263–287). Wiley.
- Boyd, J. K., & Goldberg, A. E. (2009). Input effects within a constructionist framework. *The Modern Language Journal*, 93(3), 418–429. <https://doi.org/10.1111/j.1540-4781.2009.00899.x>
- Braginsky, M., Sanchez, A., & Yurovsky, D. (2021). *childesr: Accessing the 'CHILDES' database* [Computer software]. <https://github.com/langcog/childesr>
- Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. (2016). From *uh-oh* to tomorrow: Predicting age of acquisition for early words across languages. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1691–1696). Cognitive Science Society.
- Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, 3, 52–67. https://doi.org/10.1162/opmi_a_00026
- Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81(2), B33–B44. [https://doi.org/10.1016/s0010-0277\(01\)00122-6](https://doi.org/10.1016/s0010-0277(01)00122-6)
- Carvalho, P. F., Chen, C., & Yu, C. (2021). The distributional properties of exemplars affect category learning and generalization. *Scientific Reports*, 11(1), Article 11263. <https://doi.org/10.1038/s41598-021-90743-0>
- Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, 10(2), 159–199. [https://doi.org/10.1016/0885-2014\(95\)90008-X](https://doi.org/10.1016/0885-2014(95)90008-X)
- Çetinçelik, M., Rowland, C. F., & Snijders, T. M. (2021). Do the eyes have it? A systematic review on the role of eye gaze in infant language development. *Frontiers in Psychology*, 11, Article 3627. <https://doi.org/10.3389/fpsyg.2020.589096>
- Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), Article 20160055. <https://doi.org/10.1098/rstb.2016.0055>
- Custode, S. A., & Tamis-LeMonda, C. (2020). Cracking the code: Social and contextual cues to language input in the home environment. *Infancy*, 25(6), 809–826. <https://doi.org/10.1111/infa.12361>
- Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech*, 49(2), 137–173. <https://doi.org/10.1177/00238309060490020201>
- Dore, J. (1974). A pragmatic description of early language development. *Journal of Psycholinguistic Research*, 3(4), 343–350. <https://doi.org/10.1007/BF01068169>
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5), 1–185. <https://doi.org/10.2307/1166093>
- Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, 42(1), 98–116. <https://doi.org/10.1037/0012-1649.42.1.98>
- Fernald, A., Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective–noun phrases. *Cognitive Psychology*, 60(3), 190–217. <https://doi.org/10.1016/j.cogpsych.2009.12.002>
- Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. In I. A. Sekerina, E. M. Fernández, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 97–135). John Benjamins.

- Floyd, S., & Goldberg, A. E. (2021). Children make use of relationships across meanings in word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(1), 29–44. <https://doi.org/10.1037/xlm0000821>
- Forbes, S. H., & Plunkett, K. (2019). Infants show early comprehension of basic color words. *Developmental Psychology*, 55(2), 240–249. <https://doi.org/10.1037/dev0000609>
- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, 44(3), 677–694. <https://doi.org/10.1017/S0305000916000209>
- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The wordbank project*. MIT Press.
- Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, 117(2), 107–125. <https://doi.org/10.1016/j.cognition.2010.07.005>
- Gathercole, V. C. (1985). More and more and more about *more*. *Journal of Experimental Child Psychology*, 40(1), 73–104. [https://doi.org/10.1016/0022-0965\(85\)90066-9](https://doi.org/10.1016/0022-0965(85)90066-9)
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.), *Language development: Language, thought, and culture* (Vol. 2, pp. 301–334). Lawrence Erlbaum Associates.
- Gleason, J. B., Perlmann, R. Y., & Greif, E. B. (1984). What's the magic word: Learning language through politeness routines. *Discourse Processes*, 7(4), 493–502. <https://doi.org/10.1080/01638538409544603>
- Gleason, J. B., & Weintraub, S. (1976). The acquisition of routines in child language. *Language in Society*, 5(2), 129–136. <https://doi.org/10.1017/S0047404500006977>
- Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, 15(3), 289–316. <https://doi.org/10.1515/cogl.2004.011>
- Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, 14(1), 23–45. <https://doi.org/10.1017/S030500090001271X>
- 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>
- Gopnik, A., & Meltzoff, A. N. (1984). Semantic and cognitive development in 15- to 21-month-old children. *Journal of Child Language*, 11(3), 495–513. <https://doi.org/10.1017/S0305000900005912>
- Jayaraman, S., Fausey, C. M., & Smith, L. B. (2015). The faces in infant-perspective scenes change over the first year of life. *PLoS ONE*, 10(5), Article e0123780. <https://doi.org/10.1371/journal.pone.0123780>
- Johnson, S. P., Fernandes, K. J., Frank, M. C., Kirkham, N., Marcus, G., Rabagliati, H., & Slemmer, J. A. (2009). Abstract rule learning for visual sequences in 8- and 11-month-olds. *Infancy*, 14(1), 2–18. <https://doi.org/10.1080/15250000802569611>
- Kartushina, N., & Mayor, J. (2019). Word knowledge in six- to nine-month-old Norwegian infants? Not without additional frequency cues. *Royal Society Open Science*, 6(9), Article 180711. <https://doi.org/10.1098/rsos.180711>
- Kosie, J. E., & Baldwin, D. (2019). Attention rapidly reorganizes to naturally occurring structure in a novel activity sequence. *Cognition*, 182, 31–44. <https://doi.org/10.1016/j.cognition.2018.09.004>
- Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science*, 14(6), 1323–1329. <https://doi.org/10.1111/j.1467-7687.2011.01079.x>
- Lieven, E. V., Pine, J. M., & Barnes, H. D. (1992). Individual differences in early vocabulary development: Redefining the referential-expressive distinction. *Journal of Child Language*, 19(2), 287–310. <https://doi.org/10.1017/S0305000900011429>
- MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.
- Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, 11(4), 628–634. <https://doi.org/10.1111/j.1467-7687.2008.00707.x>
- Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134(4), 477–500. <https://doi.org/10.1037/0096-3445.134.4.477>
- Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America*, 108(22), 9014–9019. <https://doi.org/10.1073/pnas.1105040108>
- Mendoza, J. K., & Fausey, C. M. (2021). Quantifying everyday ecologies: Principles for manual annotation of many hours of infants' lives. *Frontiers in Psychology*, 12, Article 710636. <https://doi.org/10.3389/fpsyg.2021.710636>
- Mirman, D. (2014). *Growth curve analysis and visualization using R*. Chapman & Hall/CRC.
- Montag, J. L., Jones, M. N., & Smith, L. B. (2018). Quantity and diversity: Simulating early word learning environments. *Cognitive Science*, 42(S2), 375–412. <https://doi.org/10.1111/cogs.12592>
- Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: Rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, 222, Article 117254. <https://doi.org/10.1016/j.neuroimage.2020.117254>
- Nelson, K. (1981). Individual differences in language development: Implications for development and language. *Developmental Psychology*, 17(2), 170–187. <https://doi.org/10.1037/0012-1649.17.2.170>
- Orhan, A. E., Gupta, V. V., & Lake, B. M. (2020). Self-supervised learning through the eyes of a child. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Proceedings of the 34th International Conference on Neural Information Processing Systems* (pp. 9960–9971). Curran Associates Inc.
- Parise, E., & Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science*, 23(7), 728–733. <https://doi.org/10.1177/0956797612438734>
- Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, 21(1), 178–185. <https://doi.org/10.3758/s13423-013-0466-4>
- Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS ONE*, 10(9), Article e0137147. <https://doi.org/10.1371/journal.pone.0137147>
- Ponari, M., Norbury, C. F., & Vigliocco, G. (2018). Acquisition of abstract concepts is influenced by emotional valence. *Developmental Science*, 21(2), Article e12549. <https://doi.org/10.1111/desc.12549>
- Reuter, T., Dalawella, K., & Lew-Williams, C. (2021). Adults and children predict in complex and variable referential contexts. *Language, Cognition and Neuroscience*, 36(4), 474–490. <https://doi.org/10.1080/23273798.2020.1839665>
- Rose, Y., & MacWhinney, B. (2014). The PhonBank project: Data and software-assisted methods for the study of phonology and phonological development. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford handbook of corpus phonology* (pp. 380–401). Oxford University Press.
- Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences of the United States of America*, 112(41), 12663–12668. <https://doi.org/10.1073/pnas.1419773112>
- Samuelson, L. K. (2021). Toward a precision science of word learning: Understanding individual vocabulary pathways. *Child Development Perspectives*, 15(2), 117–124. <https://doi.org/10.1111/cdep.12408>
- Slone, L. K., & Johnson, S. P. (2015). Infants' statistical learning: 2- and 5-month-olds' segmentation of continuous visual sequences. *Journal of Experimental Child Psychology*, 133, 47–56. <https://doi.org/10.1016/j.jecp.2015.01.007>

- Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. <https://doi.org/10.1016/j.cognition.2007.06.010>
- Stevens, J. S., Gleitman, L. R., Trueswell, J. C., & Yang, C. (2017). The pursuit of word meanings. *Cognitive Science*, *41*(S4), 638–676. <https://doi.org/10.1111/cogs.12416>
- Sullivan, J., Mei, M., Perfors, A., Wojcik, E., & Frank, M. C. (2021). SAYCam: A large, longitudinal audiovisual dataset recorded from the infant’s perspective. *Open Mind*, *5*, 20–29. https://doi.org/10.1162/opmi_a_00039
- Syrnyk, C., & Meints, K. (2017). Bye-bye mummy—word comprehension in 9-month-old infants. *British Journal of Developmental Psychology*, *35*(2), 202–217. <https://doi.org/10.1111/bjdp.12157>
- Szabó, E., & Kovács, Á. (2022). *Infants’ early understanding of different forms of negation*. PsyArXiv. <https://doi.org/10.31234/osf.io/nfv4j>
- Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental Science*, *20*(6), Article e12456. <https://doi.org/10.1111/desc.12456>
- Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby’s first 10 words. *Developmental Psychology*, *44*(4), 929–938. <https://doi.org/10.1037/0012-1649.44.4.929>
- 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>
- Tomasello, M. (1987). Learning to use prepositions: A case study. *Journal of Child Language*, *14*(1), 79–98. <https://doi.org/10.1017/S0305000900012745>
- Tsutsui, S., Chandrasekaran, A., Reza, M. A., Crandall, D., & Yu, C. (2020). A computational model of early word learning from the infant’s point of view. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society* (pp. 1029–1035). Cognitive Science Society.
- 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*, *61*(12), 2917–2933. https://doi.org/10.1044/2018_JSLHR-L-17-0363
- Wang, X., & Bi, Y. (2021). Idiosyncratic Tower of Babel: Individual differences in word-meaning representation increase as word abstractness increases. *Psychological Science*, *32*(10), 1617–1635. <https://doi.org/10.1177/09567976211003877>
- Weiner, S. L. (1974). On the development of *more* and *less*. *Journal of Experimental Child Psychology*, *17*(2), 271–287. [https://doi.org/10.1016/0022-0965\(74\)90072-1](https://doi.org/10.1016/0022-0965(74)90072-1)
- Wojcik, E. H., Zettersten, M., & Benitez, V. L. (2022). The map trap: Why and how word learning research should move beyond mapping. *WIREs Cognitive Science*, *13*(4), Article e1596. <https://doi.org/10.1002/wcs.1596>
- Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*(13–15), 2149–2165. <https://doi.org/10.1016/j.neucom.2006.01.034>

Received November 11, 2022

Revision received August 1, 2023

Accepted August 2, 2023 ■