Cognition 276 (2026) 106647

Contents lists available at ScienceDirect

Cognition

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

English-hearing toddlers learn words from consistent and variable labels

Kennedy Casey *, Casey Lew-Williams

Department of Psychology, Princeton University, United States of America

A R T I C L E I N F O

Keywords:
Word learning
Child-directed speech
Wordplay
Eye-tracking

A B S T R A C T

Early word learning is traditionally characterized as a process of mapping a single label onto a single referent. Yet, across many languages, child-directed speech (CDS) frequently features multiple labels for the same referent, including diminutivized and reduplicated word forms (e.g., *doggy* or *woof-woof* for *dog*). We examined how consistent versus variable labels impact children's novel word learning. In a within-subjects eye-tracking experiment, we exposed 18- to 36-month-old English-hearing children to novel objects that were named with consistent labels, with typical CDS variants, or with atypical CDS variants. Using looking-time measures of word-recognition accuracy, we found that children reliably learned words in the consistent and typical variation conditions, with weaker performance in the atypical variation condition. Using pupil size synchrony measures of moment-by-moment attention, we found that engagement was higher in both variation conditions compared to the consistent condition, and was highest in the typical variation condition. Our findings suggest (1) that children can successfully contend with multiple types of label variation when learning new words, and (2) that increased attention is a plausible mechanism by which label variation could support—rather than hinder—children's early word learning.

1. Introduction

Across many languages, child-directed speech (CDS) frequently includes "baby-talk" words that are unique, or at least highly specific, to this register (Bynon, 1968; Casagrande, 1948; Ferguson, 1964; Han & Gu, 2023; Mazuka et al., 2008; Ota et al., 2018; Pye, 1986; Sarvasy et al., 2025). Child-directed words are often variants of existing adult word forms, involving phonological and/or morphological changes (e.g., in English: *doggy* for *dog*, Moore & Bergelson, 2024) or replacement by onomatopoeic forms (e.g., *woof-woof* as an alternate label for *dog*: Laing, 2017; Ota et al., 2018), though some share little to no overlap in form or meaning (e.g., *bunny* for *rabbit*: Soderstrom, 2007). For some words, there are a surprising number of common variants (e.g., *pup*, *pooch*, or *bow-wow* for *dog*). While child-directed variants are typically replaced by adult forms later in development (Casey & Casillas, 2022), evidence from a prior naturalistic corpus study indicates that, in the first years of life, multiple labels for the same referent are used seemingly interchangeably in children's input (Moore & Bergelson, 2024). Here, we investigate how such label variation impacts children's word learning.

We first review competing evidence for the effect of label variation on early word learning. Several theoretical accounts predict that label variation should hinder word learning (Section 1.1). However, emerging evidence suggests that children hear substantial label variation in their input and successfully accommodate some multi-label mappings for the same referent (Section 1.2).

1.1. Label variation may hinder word learning

From a traditional one-to-one mapping perspective, children learn to associate individual word forms (i.e., labels) with individual referents (Smith & Yu, 2008; Stevens et al., 2017; but see also Wojcik et al., 2022). Under this view, multiple labels for the same referent could be considered "noise" in the learning signal. Indeed, some evidence from lab-based studies suggests that children often avoid adopting a second label for a referent they already have a name for (i.e., the mutual exclusivity bias: Markman et al., 2003; Markman & Wachtel, 1988; Merriman & Bowman, 1989; see also Mervis et al., 1994; Swingley, 2016). Using multiple variants to label the same referent also violates the pragmatic assumption that unique word forms have unique meanings (i.e., the principle of contrast: Clark, 1988).

Another challenge is that label variation may reduce the consistency or predictability of children's input for word learning. Introducing multiple interchangeable labels decreases the frequency of the standard

* Corresponding author at: Department of Psychology, Princeton University, 40 Woodlands Way, Princeton, NJ 08544.
E-mail address: kcasey@princeton.edu (K. Casey).

https://doi.org/10.1016/j.cognition.2026.106647
Received 28 January 2026; Received in revised form 3 June 2026; Accepted 6 July 2026
Available online 13 July 2026

label (one of the best predictors of earlier word learning: Braginsky et al., 2019). Variable labels may also make utterances less predictable and thus potentially harder to learn new words from (e.g., Reuter et al., 2019). Finally, the presence of multiple interchangeable labels in children's input can create lexical competition (Swingley & Aslin, 2007), increase memory interference (Regier, 2005), or impose greater general memory demands during learning (see Vlach, 2019, for further discussion).

1.2. *Label variation may support word learning*

However, despite theoretical reasons to expect that label variation should impede word learning, children's real-world learning environments contain substantial variation, and young learners seem to accommodate some types of label variation. English-hearing children in North America have been shown to recognize multiple labels for the same referent (e.g., *horse/pony*, *sheep/lamb*: Potter & Lew-Williams, 2024) and produce multiple labels themselves, including child-directed variants (e.g., *doggy/dog, bunny/rabbit*: Casey & Casillas, 2022) and variants describing different features of the same referent (e.g., *cracker/magnet* for a cracker-shaped magnet: Deák et al., 2001).

In fact, children's earliest learned words are associated with the greatest amount of label variation in their input (Moore & Bergelson, 2024). Moreover, the presence of certain types of label variation in their input has been shown to correlate both with children's concurrent vocabulary size (Callanan & Sabbagh, 2004) and future vocabulary growth (Ota et al., 2018) as well as their earlier learning of individual words (Moore & Bergelson, 2024). Thus, young learners appear to successfully contend with—and potentially benefit from—label variation in their input.

One partial explanation for this pattern is that child-directed variants are more learnable than standard adult word forms. Reduplicated variants, which include repetition of full or partial syllables (e.g., *woof-woof*, *bow-wow*), are relatively common in CDS (though estimated rates vary substantially across languages: e.g., Han & Gu, 2023; Mazuka et al., 2008; Ota et al., 2018). Ota and Skarabela (2016) found that (Scottish) English-learning 17- to 19-month-olds learned reduplicated labels for novel objects better than non-reduplicated labels. Diminutivized variants, which include /i/ epenthesis in English (e.g., *doggy*), are also common in CDS (though again with cross-linguistic variation in rates: e.g., Kempe et al., 2001; Ota et al., 2018; Savickiene, 1998). Moore and Bergelson (2024) found that diminutives were the most common type of label variation observed in a corpus of 6- to 17-month-old (US) English-learning infants. Critically, hearing a higher proportion of both reduplicated words and diminutivized words has been linked to faster vocabulary growth from 9 to 21 months for (Scottish) English learners (Ota et al., 2018). Additionally, both of these types of label variants have been shown to facilitate word segmentation (Kempe et al., 2007; Ota & Skarabela, 2018), and in doing so, could support word learning (e.g., Yurovsky et al., 2012).

What prior findings do not address is how the interchangeable use of multiple variants in the same interaction could facilitate—rather than hinder—children's word learning. We consider attention as a plausible mechanism: label variation may heighten engagement across repeated labeling instances, particularly over short timescales. Just as CDS as a register is generally preferred by young listeners over adult-directed speech (ManyBabies Consortium, 2020), hearing multiple CDS variants may help capture and sustain children's attention. Indeed, word repetition does not draw children's attention (Schwab & Lew-Williams, 2016; Segal & Newman, 2015), but label variation may introduce the kind of novelty that promotes engagement (see Tippenhauer, 2026, for related discussion), while still offering sufficient predictability to be learnable (Kidd et al., 2012, 2014). This account aligns with evidence that other types of variability within CDS, such as prosodic contours (Nencheva et al., 2021; Räsänen et al., 2018) and vowel articulations (Miyazawa et al., 2017), engage children's attention and scaffold their word learning.

1.3. *The current study*

Despite competing possibilities, no studies to date have directly tested how consistent versus variable labels impact children's ability to learn new words in a single language (but see also Tsui et al., 2023). The present study addresses this gap by manipulating the presence and type of label variation in a controlled, lab-based novel word learning task. In an eye-tracking study using a looking-while-listening (LWL) design (Fernald et al., 2008; Swingley, 2011), English-hearing toddlers learned novel word–object mappings across three within-subjects conditions. In the **consistent** condition, toddlers heard repeated exposures to the exact same standard label. In two variation conditions, toddlers heard repetitions of the standard label in addition to two label variants. The **typical variation** condition featured variants common in English CDS (diminutivized and reduplicated word forms). The **atypical variation** condition featured phonologically altered word forms that are unattested in English CDS. Looking-time measures were used to assess children's word-recognition accuracy (Fernald et al., 2008; Swingley, 2011), and pupil size synchrony measures (Kang & Wheatley, 2017; Nencheva et al., 2021) were used to examine children's moment-to-moment engagement during learning.

1.3.1. *Hypotheses*

For the effects of condition and age on children's novel word-recognition accuracy, we pre-registered the following predictions:

- In line with the vast majority of prior word learning experiments conducted to date, which overwhelmingly test children's recognition of one-to-one mappings, we hypothesized that children would demonstrate above-chance word-recognition accuracy in the **consistent** condition.
- The key condition of interest was the **typical variation** condition. We did not pre-register a specific directional hypothesis for this condition given competing theoretical predictions and uncertainty around how label variation impacts learning over short timescales. If children can successfully contend with variation attested in English CDS during a short lab task (as they do over longer developmental timescales: e.g., Moore & Bergelson, 2024), then we should expect to see above-chance word-recognition accuracy in this condition. If this type of variation impedes novel word learning (at least in the context of a short lab task), then we should expect chance-level performance in this condition.
- We hypothesized that words would be more difficult to learn in the **atypical variation** condition and that toddlers would show chance-level word-recognition accuracy, on average, for this condition.
- We also predicted that word-recognition accuracy would increase across age.

In addition to analyzing looking behavior during test trials, we pre-registered to examine children's pupillary responses during the learning phase as a complementary index of attentional engagement. Prior studies with adults (Kang & Wheatley, 2017) and toddlers (Nencheva et al., 2021) suggest that pupil size synchrony between participants (i.e., the extent to which participants' pupils dilate and constrict together over time) provides a dynamic trace of shared attention. Here, we asked whether participants attend differently to consistent versus variable labels. Specifically, we tested whether pupil size synchrony during the learning phase (a) differed between conditions and (b) predicted toddlers' word-recognition accuracy during the test phase (consistent with Nencheva et al., 2021).

2. Method

Our study design and analysis plan were pre-registered (https://osf.

io/uqz6d). All stimuli, anonymized data, and analysis scripts are publicly available at https://github.com/kennedycasey/LexVar/.

2.1. Participants

Participants ($N = 58$ total, 28 female, 30 male), aged 18 to 36 months ($M_{age} = 28.08$), were full-term, monolingual English-hearing children (>80% exposure) with no known hearing, vision, language, and/or developmental delays and were recruited from a volunteer database of local families in and around central New Jersey. Caregivers reported race/ethnicity information for their children in an open-ended format at the time of database sign-up. Participants were primarily Non-Latine white ($N = 27$, 46.55%) with some representation of other races/ethnicities: Mixed race ($N = 12$, 20.69%), Asian ($N = 5$, 8.62%), Black ($N = 2$, 3.45%), Latine white ($N = 1$, 1.72%), Not reported ($N = 11$, 18.97%). Caregivers also reported maternal education, though more sparsely: High school degree ($N = 1$, 1.72%), Bachelor's degree ($N = 12$, 20.69%), Master's degree or higher ($N = 11$, 18.97%), Not reported ($N = 34$, 58.62%).

The final sample size was pre-registered and determined using an a priori power analysis conducted in G*Power (Faul et al., 2009) and informed by prior eye-tracking studies with a similar age range (Nencheva et al., 2021; Quam & Swingley, 2023). Thirty-two additional children participated but were excluded from the final analyses due to refusal to wear the EyeLink target sticker or failure to complete calibration ($N = 9$), significant interruption by a family member ($N = 4$), contribution of too few usable trials based on pre-registered criteria due to fussiness or inattention ($N = 18$), or failure to meet the language requirement ($N = 1$). Although our inclusion criteria were pre-registered, the resulting high attrition rate may limit the generalizability of our findings. However, comparable exclusion rates have been reported in infant eye-tracking studies using a looking-while-listening design (e.g., Reuter & Emberson, 2025).

All experimental protocols, including procedures for obtaining informed consent, were approved by the Princeton University Institutional Review Board (Protocol: 0000007117). Families received a $10 digital gift card and a small gift for participating.

2.2. Stimuli

Auditory stimuli were recorded by a female native English speaker using naturalistic child-directed prosody. Each target word (Table 1) appeared in four sentence frames for exposure trials ("Look, ___! That's a ___. Do you see the ___? Yes, that's a ___.") and in two sentence frames for test trials ("Can you find the ___? See the ___?" following the test trial structure of Ota & Skarabela, 2016).[1] Exposure trials were 14,131 ms long, on average (range = 14,094–14,150 ms), including 2,000 ms silent periods between sentences. Test trials were all 4,966 ms long, with target word durations ranging from 613 to 634 ms ($M = 624$ ms), and including 2,000 ms of silence between sentences. All auditory stimuli were normed to a standard mean intensity of 65 dB (Boersma & Weenink, 2025).

The visual stimuli included six isoluminant objects with 12 unique colors (two per object) that were modeled after Nencheva et al. (2021)'s stimuli and hand-drawn using Procreate software (Savage Interactive, 2025).

Table 1
Target word forms by variant type.

Standard	Reduplicated (typical)	Reduplicated (atypical)	Diminutivized (typical)	Diminutivized (atypical)
tanzer (/tæn.zɚ/)	tan-tan (/tæn.tæn/)	zer-zer (/zɚ.zɚ/)	tanny (/tæn.i/)	y-zer (/i.zɚ/)
bogdum (/bɑg.dʌm/)	bog-bog (/bɑg.bɑg/)	dum-dum (/dʌm.dʌm/)	boggy (/bɑg.i/)	y-dum (/i.dʌm/)
kaisol (/kaɪ.sɔl/)	kai-kai (/kaɪ.kaɪ/)	sol-sol (/sɔl.sɔl/)	kaisy (/kaɪ.si/)	y-sol (/i.sɔl/)
mobet (/moʊ.bɛt/)	mo-mo (/moʊ.moʊ/)	bet-bet (/bɛt.bɛt/)	moby (/moʊ.bi/)	y-bet (/i.bɛt/)
woifer (/wɔɪ.fɚ/)	woi-woi (/wɔɪ.wɔɪ/)	fer-fer (/fɚ.fɚ/)	woify (/wɔɪ.fi/)	y-fer (/i.fɚ/)
voopin (/vu.pɪn/)	voo-voo (/vu.vu/)	pin-pin (/pɪn.pɪn/)	voopy (/vu.pi/)	y-pin (/i.pɪn/)

2.3. Procedure

Toddlers sat on their caregiver's lap, approximately 60 cm away from a 17-in. monitor equipped with an EyeLink 1000 Plus eye-tracker, in a dimly-lit testing room. Caregivers wore opaque sunglasses to prevent interference.

The experiment consisted of three blocks (one per condition). Each block consisted of two phases: an exposure phase, during which participants learned novel word–object pairings, and a test phase, during which toddlers were immediately tested on their recognition of the word–object mappings. An isoluminant, non-linguistic, attention-getting video was shown after every block.

During the exposure phase (four trials per block), children learned two novel word–object mappings. Each exposure trial started after a short, variable-length inter-trial interval (ITI; mean = 500 ms, range = 250–750 ms) or when the participant looked at the screen (following Nencheva et al., 2021, to minimize anticipatory fluctuations in pupil size before stimulus onset). During each exposure trial, a single target object appeared silently on the screen for 1,500 ms or until the participant looked at the screen. Then, the exposure audio played. After audio offset, the target object remained on the screen for an additional 1,500 ms.

During the test phase (four trials per block), toddlers saw the two objects from the preceding exposure phase side-by-side on the screen. Each trial started with an ITI (same structure as training phase), followed by a 2,000 ms silent period with both objects on the screen. Then, the target audio played, which included two repetitions of the standard word form. After the offset of the audio, the side-by-side objects remained on the screen for an additional 2,000 ms.

2.4. Conditions

Participants learned words in three conditions (within-subjects manipulation). In all conditions, participants learned the names of two different novel objects across four alternating exposure trials (two per word). In the **consistent** condition, participants heard the standard form of the target word repeated 5× per trial (e.g., "Look, *tanzer*! That's a *tanzer*…*tanzer*. Do you see the *tanzer*? Yes, that's a *tanzer*."). In the **typical variation** condition, participants heard the standard form of the target word repeated 2× per trial in addition to one reduplicated form and one diminutivized form per trial (e.g., "Look, *tanzer*! That's a *tan-tan*. Do you see the *tanny*? Yes, that's a *tanzer*."). Critically, the number of exposures to the first syllable of the target word (i.e., *tan*) was equivalent across these two conditions. In the **atypical variation** condition, participants heard the standard form of the target word repeated 2× per trial in addition to 1× exposure each of reversed versions of the typical

[1] Trials were created in Praat (Boersma & Weenink, 2025) by concatenating each target word onto the final position of each sentence frame. To avoid co-articulation artifacts in trials with articles preceding the target word, the target words were spliced along with the article. Minor acoustic edits were made on some sentences to ensure that all trials sounded naturally produced and had comparable loudness and pitch.

reduplicated and diminutivized forms (e.g., "Look, *tanzer*! That's a *zerzer*. Do you see the *y-zer*? Yes, that's a *tanzer*."). This condition includes phonological variation that is not attested in English CDS but is equated to the typical variation condition in the number of modifications made to the standard label, thereby providing tight experimental control. A full list of exposed word forms can be found in Table 1.

Participants were assigned to one of 18 pseudorandomized orders such that the order of presentation of blocks and the assignment of target labels to target objects was counterbalanced across conditions. Target locations were also counterbalanced such that the target object appeared 50% of time on the left and 50% of the time on the right for each participant.

2.5. Data pre-processing

Gaze behavior was used to determine toddlers' recognition of novel word–object pairings during two windows of analysis (300–2,000 ms following each target word onset, Fernald et al., 2008; Swingley & Aslin, 2000). We included two instances of the target word during test trials to ensure that children had sufficient opportunity to demonstrate recognition (consistent with Ota & Skarabela, 2016). We calculated participants' proportion of looking to target objects during test trials (relative to their total time looking at either the target or distractor object) as the measure of word-recognition accuracy. For any given test trial to be included in analyses, we required at least one fixation to either the target or distractor object during the pre-naming baseline period and looking to either object for at least 1/3 of each critical post-labeling window. In order for a block to be included, children must have contributed usable data for at least two of four test trials. Finally, in order for a participant to be included, they must have contributed at least two of three blocks. On average, participants contributed 9.98 trials (median = 10.50, range = 4–12) across 2.81 conditions (median = 3).

For pupillometry analyses, following Nencheva et al. (2021), we first converted toddlers' raw pupil size values during exposure sentences to percent change from the average pupil size for each trial. We removed potential artifacts due to partial blinks by excluding any 50 ms intervals with more than 15% change in pupil size (Merritt et al., 1994). We used Stineman interpolation to fill in gaps shorter than 100 ms (Johannesson et al., 2009). Finally, we excluded exposure sentences where more than 50% of a participant's pupil size data was missing (e.g., due to the child looking away from the screen or technical error).

3. Results

3.1. Word recognition measured by gaze behavior

Our primary pre-registered analysis examined whether toddlers successfully learned novel word–object mappings across conditions. Using the *lme4* package (Bates et al., 2015), we fit a mixed-effects logistic regression model to test the effects of condition and age on children's novel word-recognition accuracy (i.e., proportion looking to the target object). The model included fixed effects for condition (atypical variation vs. consistent vs. typical variation) and mean-centered age,[2] along with the maximal random effects structure that allowed for model convergence (by-participant and by-item random intercepts). We also replicated this analysis using zero-one-inflated beta regression under a Bayesian framework (Liu & Eugenio, 2018), which yielded similar results (see Supplementary Materials).

To evaluate whether children's performance differed from chance within each condition, we first used the fitted model to estimate marginal means for each condition and then compared each estimated marginal mean (EMM) to chance performance (0.50). As predicted, toddlers' performance was reliably above chance in the consistent condition ($EMM = 0.65$, $SE = 0.04$, 95% CI [0.56, 0.73]). In addition, as predicted, toddlers' performance was at chance in the atypical variation condition ($EMM = 0.54$, $SE = 0.04$, 95% CI [0.45, 0.62]). The main condition of theoretical interest was the typical variation condition, in which we found that toddlers performed reliably above chance ($EMM = 0.60$, $SE = 0.04$, 95% CI [0.51, 0.68]).

We next examined between-condition differences, comparing performance in the consistent and typical variation conditions to the atypical variation condition (the reference level). Model parameter estimates revealed a significant positive effect of the consistent condition ($\beta = 0.47, SE = 0.22, p = 0.030$), and a positive but non-significant effect of the typical variation condition ($\beta = 0.26$, $SE = 0.21$, $p = 0.231$) on overall accuracy. Thus, although toddlers' predicted performance in the typical variation condition was reliably above chance, it was not significantly higher than performance in the atypical variation condition. Toddlers' proportion looking to target also increased significantly with age, as hypothesized ($\beta = 0.26$, $SE = 0.10$, $p = 0.010$).

Exploratory analyses suggested descriptively stronger condition differences in the first post-labeling window, which is the typical window of analysis used in LWL studies to examine word recognition (Fernald et al., 2008). In the second post-labeling window, which is less frequently implemented in LWL studies, children's performance was reliably above chance in all conditions (all $ps < 0.05$), indicating some evidence of learning across all types of exposure (see Supplementary Materials for further details). Adding post-labeling window to the base model revealed no effect of window ($\beta = 0.10$, $SE = 0.14$, $p = 0.482$). Adding an interaction between window and condition did not significantly improve the model fit ($\chi^2(2) = 0.54$, $p = 0.765$). Fig. 1 shows model-predicted accuracy across conditions, overall and separately for the two post-labeling windows.

3.2. Moment-by-moment attention measured by pupil size synchrony

Following Nencheva et al. (2021), we used dynamic time-warping distance (DTW, Tormene et al., 2009) to quantify the temporal alignment of participants' pupil responses when hearing individual exposure sentences (normalized for duration). This DTW measure quantifies the dissimilarity between two time series of pupil fluctuations. To transform these values into synchrony scores, we subtracted DTW distance scores from the maximum distance between two participants on each trial. This transformation rescales DTW distances such that higher values reflect greater synchrony (i.e., greater shared engagement between toddlers). To compare synchrony across conditions, we computed pairwise synchrony scores for all possible pairs of participants (for each exposure sentence per condition). Higher synchrony values indicate that participants' pupils dilated and constricted more similarly over time, suggesting more aligned attentional engagement during learning. Synchrony values were z-scored prior to analysis, so positive values indicate greater-than-average synchrony in our sample, and negative values indicate lower-than-average synchrony.

We compared synchrony during exposure sentences with typical variants versus atypical variants versus consistent labels (see also Supplementary Materials for word-level effects and analyses for all exposure sentences). A linear mixed-effects model (Bates et al., 2015) predicting average pupil size synchrony (z-scored) from condition, with by-participant random intercepts, revealed a significant main effect of condition on pupil size synchrony [$F(2, 246) = 26.66$, $p < 0.001$]. Follow-up pairwise comparisons using Tukey's HSD correction indicated that there were statistically significant differences between all condition pairs. Pupil size synchrony was higher in the atypical variation condition ($\Delta M = 0.20$, $p = 0.006$, Cohen's $d = 0.45$) and the typical variation

[2] Following our pre-registered analysis plan, we tested whether including an interaction between condition and age improved model fit—we compared the model with the interaction term to the simpler additive model using a likelihood ratio test. This analysis suggested no reliable difference in predictive accuracy, indicating that the interaction term did not improve model fit ($\chi^2(2) = 1.01$, $p = 0.602$).

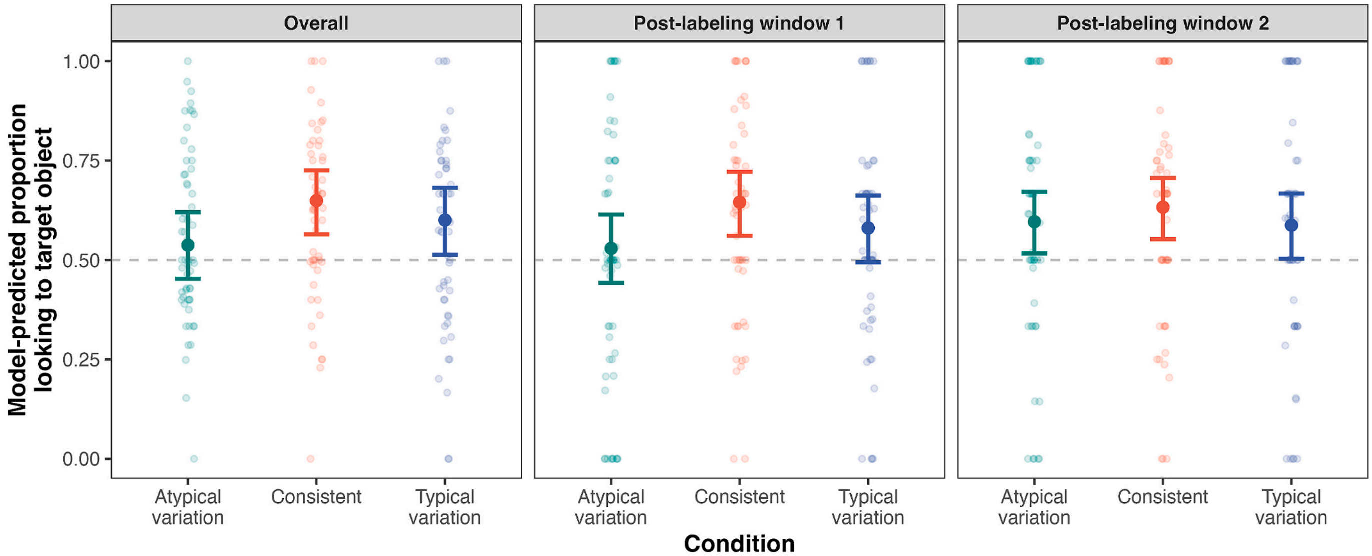

Fig. 1. Model-predicted accuracy (i.e., proportion looking to target object) by condition, overall (averaged between windows) and across both post-labeling windows (300–2,000 ms following target label onset). Point ranges reflect estimated marginal means with 95% confidence intervals. Jittered dots show raw means for individual participants. The dashed line at 0.50 indicates chance performance.

condition ($\Delta M = 0.47$, $p < 0.001$, Cohen's $d = 1.07$) relative to the consistent condition. Pupil size synchrony was also higher in the typical variation condition relative to the atypical variation condition ($\Delta M = 0.27$, $p < 0.001$, Cohen's $d = 0.62$; Fig. 2).

To understand if toddlers' moment-to-moment engagement with consistent versus variable labels predicted their novel word learning, we tested whether or not a given participant's proportion looking to each target object during test trials could be reliably predicted from their average pupil size synchrony for that target word during training trials. We added mean pupil size synchrony (z-scored) as an additional fixed effect, along with a condition-by-synchrony interaction term, to our main mixed-effects logistic regression model from Section 3.1; however, the addition of these terms did not significantly improve model fit ($\chi^2(3) = 2.41$, $p = 0.491$).

4. Discussion

We tested how consistent versus variable labels impact English-hearing toddlers' learning of novel word–object mappings. Specifically, we examined whether 18- to 36-month-olds showed better novel word-recognition accuracy when they were exposed to consistent labels versus typical label variants (reduplicated/diminutivized word forms attested in English CDS) versus atypical label variants (word forms unattested in English CDS). This is, to our knowledge, the first experiment to directly manipulate children's exposure to label variation during a word learning task. Building on prior naturalistic corpus analyses of CDS variant use (e.g., Casey & Casillas, 2022; Moore & Bergelson, 2024; Ota et al., 2018), this study provides a next step toward understanding the causal impact of hearing multiple label variants on children's learning of new words.

As hypothesized, children performed above chance in the consistent condition, replicating decades of findings showing that children can

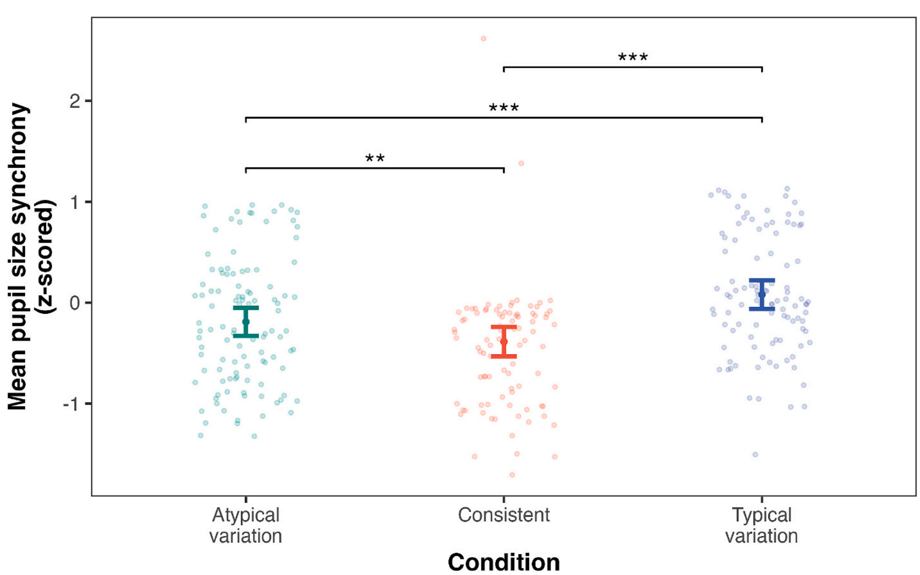

Fig. 2. Model-predicted mean pupil size synchrony (z-scored) across conditions with 95% confidence intervals. Pupil size synchrony is a proxy for shared attention between participants, so greater values represent more shared attention. Jittered points represent participant means. $^{**}p < .01$, $^{***}p < .001$.

learn novel one-to-one mappings from a brief exposure phase. The primary goal of this study was to determine how children contend with two different types of label variation. As predicted, performance was at chance in the atypical variation condition. However, performance was above chance in the typical variation condition, indicating that toddlers can successfully learn from variable input, particularly when variant labels match the type that they might hear at home. Overall, the key finding was evidence of learning in the consistent and typical variation conditions, but not in the atypical variation condition, based on proportion looking to the target object overall and during the first post-labeling window.

Between-condition effects were generally weak and non-significant, likely driven by the short length of the in-lab task relative to the longer developmental timescales over which children hear multiple interchangeable labels in everyday life (Casey & Casillas, 2022; Moore & Bergelson, 2024; Ota et al., 2018). Accordingly, these results warrant caution when interpreting the magnitude of differences between hearing consistent versus variable labels, particularly over short timescales. Additionally, exploratory analyses of the two post-labeling windows suggested that condition differences were most apparent in toddlers' looking behavior following the first instance of the label, which is the window used most often to index word recognition in looking-while-listening paradigms (Fernald et al., 2008). By the second instance of hearing the target label (cf. Ota & Skarabela, 2016), children reliably recognized the novel word–object mappings in all conditions, suggesting that toddlers eventually demonstrated some recognition of words learned during consistent and variable labeling conditions.

Our current findings cannot fully tease apart different explanations for why children successfully learned from both consistent and (typically) variable labels. That is, we do not know if learning was anchored in toddlers' exposures to the standard word form, or in forming associations between each variant and its referent. It is possible that exposure to repetitions of the standard label was sufficient for children to learn label–object mappings, regardless of the surrounding variants. Children may have "filtered out" (cf. Potter & Lew-Williams, 2019) atypical or unfamiliar forms, focusing on the consistent standard repetitions. This explanation is plausible in the context of this short in-lab experiment but contrasts with naturalistic evidence that CDS variants are common in children's early productive vocabularies (e.g., Frank et al., 2021; Laing, 2019; Tardif et al., 2008) and are used more often than standard adult word forms in the first years of life (Casey & Casillas, 2022). Given the length of the experiment, it was not feasible to also test children's recognition of typical and atypical variants themselves (in addition to standard word forms), thus we cannot determine exactly what children encoded about the specific label variants. Future work could further test the effects of exposure to label variants, for instance, by probing whether toddlers show improved long-term retention of word–object mappings (cf. Vlach, 2019) or better generalization to new label variants (cf. Regier, 2005; Swingley, 2016) after hearing variable rather than consistent input.

Another open question is why toddlers showed more evidence of learning in the typical variation condition than in the atypical variation condition. One possibility is that English-hearing toddlers come to expect certain types of variants based on their experience with label variation in naturalistic speech. Unattested, atypical variants may have been less predictable, potentially disrupting learning. Another possibility is that the typical variants in our study preserved phonological information that supports early word recognition, including word onsets and stressed syllables. Young children make use of word onsets during real-time language processing (e.g., Swingley et al., 1999) and are sensitive to strong-weak disyllabic stress patterns (e.g., Jusczyk et al., 1993), so preserving these cues may have supported learning in the typical variation condition. Future work systematically manipulating variant familiarity, phonological overlap, and stress preservation will be important for identifying the mechanisms underlying children's real-time processing of and learning from label variation. Additional work will also be needed to account for children's ability to accommodate the full scope label variation found in naturalistic input (e.g., *dog*, *bow-wow*, *hound*, *pup*, etc.).

In the current study, pupillometry analyses provide complementary insight into children's moment-to-moment engagement during learning. Attention, indexed by pupil size synchrony across participants (Kang & Wheatley, 2017; Nencheva et al., 2021), was higher in both variation conditions relative to the consistent condition, and was highest in the typical variation condition. These results suggest reliable differences in children's engagement with the labeling sentences across conditions. Although pupil size synchrony during exposure did not reliably predict word-recognition accuracy during test, prior findings suggest a link between synchrony and learning (Nencheva et al., 2021), which might have been observed if children were also tested on their recognition of variant labels. Our findings also fit with prior accounts of a "goldilocks" effect in children's attention (Kidd et al., 2012, 2014), where a moderate amount of variability—here, variation of the type common in children's everyday language input—is associated with the highest engagement. Overall, the current pupillometry findings point to attentional engagement as a plausible, though not yet definitively established, mechanism by which label variation, particularly of the type that children encounter naturalistically, could facilitate novel word learning over short timescales.

In sum, this study provides the first experimental evidence that English-hearing toddlers can successfully contend with label variants when learning new words. Typical CDS variants were associated with both greater attentional engagement during learning and above-chance word-recognition at test. These findings suggest that naturalistic label variation in children's input is not merely noise and may play a facilitative role in early word learning. Future work that directly tests children's learning of label variants, tracks longer-term retention of multiple labels, or examines generalization across novel variants will be critical for understanding how variability may support vocabulary growth across early development.

Author note

The study design and analysis plan was pre-registered (https://osf.io/uqz6d). Anonymized data and fully reproducible manuscript code are available at https://github.com/kennedycasey/LexVar. The authors declare no conflicts of interest. This research was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01R01HD095912-05 awarded to CLW) and the National Science Foundation (Graduate Research Fellowship awarded to KC). We are grateful to the participating families as well as the members of the Princeton Baby Lab and Psychology of Language Lab for their feedback on this work.

CRediT authorship contribution statement

Kennedy Casey: Writing – original draft, Investigation, Formal analysis, Conceptualization. **Casey Lew-Williams:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Appendix A. Supplementary data

Supplementary materials for this article can be found online at https://doi.org/10.1016/j.cognition.2026.106647.

Data availability

Link shared in manuscript

References

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

Boersma, P., & Weenink, D. (2025). Praat: Doing phonetics by computer [computer software]. https://praat.org.

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

Bynon, J. (1968). Berber nursery language. *Transactions of the Philological Society, 67*(1), 107–161. https://doi.org/10.1111/j.1467-968X.1968.tb01132.x

Callanan, M. A., & Sabbagh, M. A. (2004). Multiple labels for objects in conversations with young children: Parents' language and children's developing expectations about word meanings. *Developmental Psychology, 40*(5), 746. https://doi.org/10.1037/0012-1649.40.5.746

Casagrande, J. B. (1948). Comanche baby language. *International Journal of American Linguistics, 14*(1), 11–14. https://doi.org/10.1086/463971

Casey, K., & Casillas, M. (2022). From doggy to dog: Developmental shifts in children's use of register-specific words. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenozi (Eds.), *Proceedings of the 44th Annual Meeting of the Cognitive Science Society* (pp. 49–56). https://escholarship.org/uc/item/7hb9113g.

Clark, E. V. (1988). On the logic of contrast. *Journal of Child Language, 15*(2), 317–335. https://doi.org/10.1017/S0305000900012393

Deák, G. O., Yen, L., & Pettit, J. (2001). By any other name: When will preschoolers produce several labels for a referent? *Journal of Child Language, 28*(3), 787–804. https://doi.org/10.1017/s0305000901004858

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist, 66*(6), 103–114. https://www.jstor.org/stable/pdf/668164.pdf.

Fernald, A. E., 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). John Benjamins Publishing Company. https://doi.org/10.1075/lald.44.06fer.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The Wordbank project*. MIT Press. https://doi.org/10.7551/mitpress/11577.001.0001

Han, M., & Gu, Y. (2023). Che-che ("car-car") and chi-chi ("eat-eat"): Reduplication in Mandarin Chinese child-directed speech. In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Meeting of the Cognitive Science Society* (pp. 920–927). https://escholarship.org/uc/item/8ks9p299.

Johannesson, T., Bjornsson, H., Grothendieck, G., & Johannesson, M. T. (2009). *Package "stinepack"*. Icelandic Meteorological Office. https://rdrr.io/cran/stinepack/.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*(3), 675–687. https://doi.org/10.1111/j.1467-8624.1993.tb02935.x

Kang, O., & Wheatley, T. (2017). Pupil dilation patterns spontaneously synchronize across individuals during shared attention. *Journal of Experimental Psychology: General, 146*(4), 569. https://doi.org/10.1037/xge0000271

Kempe, V., Brooks, P. J., Gillis, S., & Samson, G. (2007). Diminutives facilitate word segmentation in natural speech: Cross-linguistic evidence. *Memory & Cognition, 35*(4), 762–773. https://doi.org/10.3758/BF03193313

Kempe, V., Brooks, P. J., & Pirott, L. (2001). How can child-directed speech facilitate the acquisition of morphology?. In *Research on Child Language Acquisition: Proceedings of the 8th Conference of the International Association for the Study of Child Language* (pp. 1237–1247).

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One, 7*(5), Article e36399. https://doi.org/10.1371/journal.pone.0036399

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The goldilocks effect in infant auditory attention. *Child Development, 85*(5), 1795–1804. https://doi.org/10.1111/cdev.12263

Laing, C. E. (2017). A perceptual advantage for onomatopoeia in early word learning: Evidence from eye-tracking. *Journal of Experimental Child Psychology, 161*, 32–45. https://doi.org/10.1016/j.jecp.2017.03.017

Laing, C. E. (2019). A role for onomatopoeia in early language: Evidence from phonological development. *Language and Cognition, 11*(2), 173–187. https://doi.org/10.1017/langcog.2018.23

Liu, F., & Eugenio, E. C. (2018). A review and comparison of Bayesian and likelihood-based inferences in beta regression and zero-or-one-inflated beta regression. *Statistical Methods in Medical Research, 27*(4), 1024–1044. https://doi.org/10.1177/0962280216650699

ManyBabies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science, 3*(1), 24–52. https://doi.org/10.1177/2515245919900809

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20*(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology, 47*(3), 241–275. https://doi.org/10.1016/s0010-0285(03)00034-3

Mazuka, R., Kondo, T., & Hayashi, A. (2008). Japanese mothers' use of specialized vocabulary in infant-directed speech: Infant-directed vocabulary in Japanese. In *The Origins of Language: Unraveling Evolutionary Forces* (pp. 39–58). Springer. https://doi.org/10.1007/978-4-431-79102-7_4.

Merriman, W. E., & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development, 54*. https://doi.org/10.2307/1166130

Merritt, S. L., Keegan, A. P., & Mercer, P. W. (1994). Artifact management in pupillometry. *Nursing Research, 43*(1), 56–59. https://doi.org/10.1097/00006199-199401000-00012

Mervis, C. B., Golinkoff, R. M., & Bertrand, J. (1994). Two-year-olds readily learn multiple labels for the same basic-level category. *Child Development, 65*(4), 1163–1177. https://doi.org/10.2307/1131312

Miyazawa, K., Shinya, T., Martin, A., Kikuchi, H., & Mazuka, R. (2017). Vowels in infant-directed speech: More breathy and more variable, but not clearer. *Cognition, 166*, 84–93. https://doi.org/10.1016/j.cognition.2017.05.003

Moore, C., & Bergelson, E. (2024). Wordform variability in infants' language environment and its effects on early word learning. *Cognition, 245*, Article 105694. https://doi.org/10.1016/j.cognition.2023.105694

Nencheva, M. L., Piazza, E. A., & Lew-Williams, C. (2021). The moment-to-moment pitch dynamics of child-directed speech shape toddlers' attention and learning. *Developmental Science, 24*(1), Article e12997. https://doi.org/10.1111/desc.12997

Ota, M., Davies-Jenkins, N., & Skarabela, B. (2018). Why choo-choo is better than train: The role of register-specific words in early vocabulary growth. *Cognitive Science, 42*(6), 1974–1999. https://doi.org/10.1111/cogs.12628

Ota, M., & Skarabela, B. (2016). Reduplicated words are easier to learn. *Language Learning and Development, 12*(4), 380–397. https://doi.org/10.1080/15475441.2016.1165100

Ota, M., & Skarabela, B. (2018). Reduplication facilitates early word segmentation. *Journal of Child Language, 45*(1), 204–218. https://doi.org/10.1017/S0305000916000660

Potter, C. E., & Lew-Williams, C. (2019). Infants' selective use of reliable cues in multidimensional language input. *Developmental Psychology, 55*(1), 1. https://doi.org/10.1037/dev0000610

Potter, C. E., & Lew-Williams, C. (2024). Frequent vs. infrequent words shape toddlers' real-time sentence comprehension. *Journal of Child Language, 51*(6), 1478–1488. https://doi.org/10.1017/S0305000923000387

Pye, C. (1986). Quiché Mayan speech to children. *Journal of Child Language, 13*(1), 85–100. https://doi.org/10.1017/S0305000900000313

Quam, C., & Swingley, D. (2023). A protracted developmental trajectory for English-learning children's detection of consonant mispronunciations in newly learned words. *Language Acquisition, 30*(3–4), 256–276. https://doi.org/10.1080/10489223.2022.2069026

Räsänen, O., Kakouros, S., & Soderstrom, M. (2018). Is infant-directed speech interesting because it is surprising?–Linking properties of IDS to statistical learning and attention at the prosodic level. *Cognition, 178*, 193–206. https://doi.org/10.1016/j.cognition.2018.05.015

Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science, 29*(6), 819–865. https://doi.org/10.1207/s15516709cog0000_31

Reuter, T., Borovsky, A., & Lew-Williams, C. (2019). Predict and redirect: Prediction errors support children's word learning. *Developmental Psychology, 55*(8), 1656. https://doi.org/10.1037/dev0000754

Reuter, T. E., & Emberson, L. L. (2025). Relative contributions of predictive vs. associative processes to infant looking behavior during language comprehension. *Journal of Child Language, 52*(6), 1225–1248. https://doi.org/10.1017/S0305000924000102

Sarvasy, H., Rumsey, A., Dahmen, J., Onga, J., & Yam, S. (2025). Child-directed speech in Ku Waru and Nungon (Papua New Guinea). *Language Development Research, 5*(3), 191–244. https://doi.org/10.34842/ldr2025-857

Savage Interactive. (2025). *Procreate (version 5.3.15) [computer software]*. Savage Interactive Pty Ltd. https://procreate.com/.

Savickiene, I. (1998). The acquisition of diminutives in Lithuanian. In *, Vol. 95. Antwerp Papers in Linguistics: Studies in the Acquisition of Number and Diminutive Marking* (pp. 129–153). Antwerpen: Universiteit Antwerpen.

Schwab, J. F., & Lew-Williams, C. (2016). Repetition across successive sentences facilitates young children's word learning. *Developmental Psychology, 52*(6), 879. https://doi.org/10.1037/dev0000125

Segal, J., & Newman, R. S. (2015). Infant preferences for structural and prosodic properties of infant-directed speech in the second year of life. *Infancy, 20*(3), 339–351. https://doi.org/10.1111/infa.12077

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

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review, 27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Stevens, J. S., Gleitman, L. R., Trueswell, J. C., & Yang, C. (2017). The pursuit of word meanings. *Cognitive Science, 41*, 638–676. https://doi.org/10.1111/cogs.12416

Swingley, D. (2011). The looking-while-listening procedure. In L. Wei, & E. Hoff (Eds.), *Methods in Child Language: A Practical Guide* (pp. 29–42). https://doi.org/10.1002/9781444344035.ch3

Swingley, D. (2016). Two-year-olds interpret novel phonological neighbors as familiar words. *Developmental Psychology, 52*(7), 1011. https://doi.org/10.1037/dev0000114

Swingley, D., & Aslin, R. N. (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. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology, 54*(2), 99–132. https://doi.org/10.1016/j.cogpsych.2006.05.001

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71*(2), 73–108. https://doi.org/10.1016/S0010-0277(99)00021-9

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology, 44*(4), 929. https://doi.org/10.1037/0012-1649.44.4.929

Tippenhauer, N. (2026). Is variety the spice of life? The role of variability in early word learning. *Cognitive Development, 78*, 101713. https://doi.org/10.1016/j.cogdev.2026.101713

Tormene, P., Giorgino, T., Quaglini, S., & Stefanelli, M. (2009). Matching incomplete time series with dynamic time warping: An algorithm and an application to post-stroke rehabilitation. *Artificial Intelligence in Medicine, 45*(1), 11–34. https://doi.org/10.1016/j.artmed.2008.11.007

Tsui, R. K.-Y., Kosie, J. E., Fibla, L., Lew-Williams, C., & Byers-Heinlein, K. (2023). Patterns of language switching and bilingual children's word learning: An experiment across two communities. *Translational Issues in Psychological Science, 9*(4), 323. https://doi.org/10.1037/tps0000353

Vlach, H. A. (2019). Learning to remember words: Memory constraints as double-edged sword mechanisms of language development. *Child Development Perspectives, 13*(3), 159–165. https://doi.org/10.1111/cdep.12337

Wojcik, E. H., Zettersten, M., & Benitez, V. L. (2022). The map trap: Why and how word learning research should move beyond mapping. *Wiley Interdisciplinary Reviews: Cognitive Science, 13*(4), Article e1596. https://doi.org/10.1002/wcs.1596

Yurovsky, D., Yu, C., & Smith, L. B. (2012). Statistical speech segmentation and word learning in parallel: Scaffolding from child-directed speech. *Frontiers in Psychology, 3*, 374. https://doi.org/10.3389/fpsyg.2012.00374