Title

The distributional and embodied contexts of verbs in caregiver-infant interactions.

Authors

Zhang, Vivian Hanwen
Chang, Lucas M
Deák, Gedeon O

Publication Date

2024-01-08

DOI

10.1017/s0305000923000636

Journal of Child Language (2024), 1–15
doi:10.1017/S0305000923000636

CAMBRIDGE UNIVERSITY PRESS

BRIEF RESEARCH REPORT

The distributional and embodied contexts of verbs in caregiver-infant interactions

Vivian Hanwen ZHANG[1,2], Lucas M. CHANG[1] and Gedeon O. DEÁK[1]

[1]Department of Cognitive Science, University of California, San Diego, USA
[2]Department of Psychology, Cornell University, USA
Corresponding author: Gedeon O. Deák; Email: gdeak@ucsd.edu

(Received 06 February 2023; revised 17 September 2023; accepted 27 September 2023)

Abstract

The process by which infants learn verbs through daily social interactions is not well-understood. This study investigated caregivers' use of verbs, which have highly abstract meanings, during unscripted toy-play. We examined how verbs co-occurred with distributional and embodied factors including pronouns, caregivers' manual actions, and infants' locomotion, gaze, and object-touching. Object-action verbs were used significantly more often during caregiver-infant joint attention interactions. Movement and cognition verbs showed distinct co-occurrences with different contexts. Cognition and volition verbs were differentiated by pronouns. These findings provide evidence for how verb acquisition may be supported by the distributional and embodied contexts in caregiver-infant interactions.

Keywords: distributional cues; embodiment; verb learning

Introduction

As infants develop from knowing little about language to comprehending dozens of words during their first year (Frank et al., 2017), they learn aspects of word usage from interactions with caregivers in social contexts (Tamis-LeMonda et al., 2014). To understand this process, previous research has mostly focused on words with relatively concrete referents: object-labeling nouns. However, little is known about contexts that predict word instances in infant-directed speech (IDS) with less concrete meanings (Maguire et al., 2006). For example, how do infants acquire verbs, which cannot be mapped onto physical objects or often even concrete actions? How might caregiver-infant interactions facilitate this – for example, what contextual factors covary with verb usage in naturalistic infant-parent interactions?

Verb acquisition

Verb learning appears to be harder than noun learning (e.g., Golinkoff & Hirsh-Pasek, 2008). Infants show evidence of understanding their first nouns around 6.0 - 7.5 months of age (Bortfeld et al., 2005), but do not recognize their first verbs until 11.0 - 13.5 months

(e.g., Nazzi et al., 2005). Similarly, although infants typically produce their first words around the end of the first year, first verbs are not produced until months later (Fenson et al., 1994). This gap in acquisition has been argued to reflect differential difficulty in concept learning: children induce object categories easier than relational or state/change categories (Rattermann & Gentner, 1998). Thus, they may attend to, and learn, regularities corresponding to noun meanings prior to regularities corresponding to verb meanings (Gentner, 2006).

There is indirect evidence, however, that input factors moderate verb as well as noun learning. Differential word usage in language input and events in caregiver-infant interactions might contribute to the early prevalence of nouns vs. verbs (Chan & Nicoladis, 2010). For example, cross-linguistic differences in the prominence of verbs and nouns in IDS might lead to different proportions of nouns and verbs in early vocabularies (Waxman et al., 2013). Mandarin-speaking mothers end sentences with verbs (which facilitates encoding) more often than English-speaking mothers (Tardif et al., 1997). Conversely, when reading picture books, English-speaking mothers of 20-month-olds produced more nouns than Mandarin-speaking mothers (Tardif et al., 1999). Such differences might contribute to the higher proportion of verbs in Mandarin-speaking toddlers' vocabulary, relative to English-speaking peers (Tardif et al., 1999). Interestingly, English-speaking adults guess verbs from muted videos of caregiver-infant interactions more accurately if the caregiver is Mandarin-speaking than if she is English-speaking (Snedeker et al., 2003), suggesting that non-verbal cues to verb meaning also differ cross-culturally. Evidence therefore suggests that both linguistic *and* non-linguistic input during caregiver-infant interactions support infants' verb acquisition. Thus, it is important to document how and when verbs are used in naturalistic caregiver-infant interactions.

Distributional contexts

The distributional contextual patterns of natural speech are far from random (Redington et al., 1998), and evidence suggests that infants are sensitive to these patterns. Words strongly and successively constrain the types and positions of other words in an utterance. For example, "I eat…" is much more likely to be followed by "apples" than by "walk," "exuberant," or "tigers." Such patterns of lexical co-occurrence could help infants infer word meaning and assign words to syntactic categories. Willits et al. (2014) suggested that the more frequent and consistent distributional contexts of nouns compared to verbs might contribute to the noun precedence in infant vocabularies. They also confirmed that 7.5- and 9.5-month-old infants learn distributional statistics of verbs, because infants looked longer in response to verbs appearing in infrequent linguistic contexts.

With respect to lexical co-occurrences, recent studies showed that verbs co-occur non-randomly with pronouns (Babineau et al., 2020; Laakso & Smith, 2007), object nouns (Yuan et al., 2011), and adverbs (Syrett et al., 2014). However, these studies examined a limited number of specific verbs to establish the possible importance of distributional statistics. A broader survey of the degree to which many verbs from varied verb categories co-occur with multiple lexical AND non-lexical contextual factors would more clearly show how naturalistic input patterns might support verb learning.

One type of contextual co-occurrence information is pronouns. Pronouns are closed-class markers for subjects or objects, with limited semantic information (e.g., number or gender), that are typically disambiguated by syntactic and pragmatic context. They typically refer to established or "given" constituents (e.g., "I like it", Messer, 1978).

https://doi.org/10.1017/S0305000923000636 Published online by Cambridge University Press

Moreover, although limited in informativeness and concreteness, pronouns are among the most frequent words in English IDS (Laakso & Smith, 2007). Importantly, because pronouns carry both semantic and syntactic information (e.g., "She eats them" vs. "They eat her"), they should have discernable distributional statistics, and these different distributions might differ among verbs and verb categories. Thus, we investigated whether pronouns alone could predict verb semantics within naturalistic American-English IDS. This extends previous findings on verb-pronoun co-occurrences in natural language (Babineau et al., 2020; Laakso & Smith, 2007), and their potential to support infants' verb learning.

Embodied contexts

Caregiver-infant interactions are multimodal and dynamic (Suarez-Rivera et al., 2022). A sequence of actions accompanying speech, like a caregiver showing their infant a toy while talking, then the infant looking at and grabbing the toy, is common in everyday interactions, and might scaffold word learning. For example, infants learn object nouns more readily during joint attention with caregivers (Tomasello & Farrar, 1986), and when the referent is visually dominant (Yu & Smith, 2012). Caregivers' object naming utterances are also predicted by infants' and mothers' gaze and object-directed manual actions, and by shared attentional focus during the interaction (Chang et al., 2016; Custode & Tamis-LeMonda, 2020; West & Iverson, 2017). Similarly, embodied contexts also facilitate verb learning. A recent home-recording study showed that caregivers often used movement verbs when their 13-month-old infants locomoted, and manual verbs when the infants manipulated objects (West et al., 2022). Also, an eye-tracking study showed that infants (15 to 25 months) paid more attention when caregivers produced verbs corresponding with their actions (Liu et al., 2019). This suggests that joint attention to actions (see Deák et al., 2014) might help infants learn verb as well as noun meanings. However, because actions are more transient than objects, action verbs might co-occur less reliably than object-nouns with their respective referents during bouts of shared attention. In investigating the role of verb-action co-occurrences in verb acquisition, previous studies largely focused on object-related actions and on locomotion (West et al., 2022). Here we consider a wider range of verb categories (e.g., cognition/perception and volition verbs) and study how caregivers use them co-occurrent with object handling, gaze target, and locomotion.

The current study

To understand how verbs are distributionally represented in pronominal and embodied contexts during caregiver-infant play, we video-recorded mother-infant free-play sessions at 12 months of age, and transcribed mothers' utterances. We annotated mothers' and infants' gaze and manual actions, as well as infant locomotion. We classified the most frequent verbs and pronouns in mothers' speech based on common semantic (e.g., mental vs. action verbs) and syntactic (e.g., transitive vs. intransitive) features of interest in previous research. We analyzed co-occurrences of each verb category with pronoun categories as well as embodied contextual factors (i.e., gaze, hands, locomotion). We tested the strength of co-occurrence frequencies using linear mixed effects models. We hypothesized that different verb categories co-occur with distinct combinations of linguistic and embodied factors.

https://doi.org/10.1017/S0305000923000636 Published online by Cambridge University Press

Specifically, our first prediction was that, like object-naming nouns (Tomasello & Farrar, 1986), object-action verbs may also co-occur with episodes of joint attention and object handling. Second, we predicted that movement and mental verbs would be differentiated by correlated pronominal and embodied variables during play. Previous research indicated that motion verbs tend to precede the word "it", whereas psychological attitude verbs tend to precede a clause (Laakso & Smith, 2007). Mental verbs are learned later than movement verbs (Bloom et al., 1975; Shatz et al., 1983), so any differential context of usage might be especially crucial for learning the former. Third, we further predicted that among mental verbs, cognition verbs would be differentiated from volition verbs, similar to Laakso and Smith (2007) who found that epistemic verbs (e.g., "think") were more likely to co-occur with "I", whereas deontic verbs (e.g., "like") co-occurred with "you".

Method

Participants

Forty-two mothers with their infants (20 female) were recruited in San Diego County, for a longitudinal study of infant social development (Deák et al., 2013). An experimenter visited the participants' homes every month while infants were between the ages of 3 to 9 months, and again at 12 months of age. Upon recruitment, mothers' mean age was 32.1 years (range = 21-42), with a mean of 16.1 years of formal education (range = 12-21). Twenty-nine infants were White, two were Asian, five were "other" or multiracial. Four infants were of Hispanic origins. Two parents did not provide information about ethnicity[1]. The current study reports data from the 12-month home session, when infants averaged 371 days old (range: 356-450). This age was chosen based on previous related studies of early verb learning and caregiver verb use (see, e.g., Liu et al., 2019; West et al., 2022).

Procedure

Infants were seated across from their mother in a room of their home where the dyad typically played. Dyads were recorded by two cameras while they played with three sets of infant objects (see Figure 1 for details). Mothers were instructed to "play as they normally would" with their infants for about 15 minutes (M = 14.12 min, SD = 1.73). Intervals when infants were fussy or locomoted outside of view, or when an experimenter was present (e.g., to deliver toy sets), were excluded from coding and analyses. On average, 12.19 min of play per session were coded and analyzed (SD = 2.22).

Coding

Videos were digitized and synchronized for coding. Coders (blind to specific hypotheses) annotated the videos using ELAN (2023)[2] for maternal speech, and Datavyu (2014)[3] for

[1]No infant had known neurological, cognitive, or sensory deficits. Four additional participants were excluded because of equipment failure or speaking to the infant in a language other than English.

[2]https://archive.mpi.nl/tla/elan

[3]http://datavyu.org

https://doi.org/10.1017/S0305000923000636 Published online by Cambridge University Press

Figure 1. An Example of Play Session Recording
Note. A frame from one camera from a home session recording illustrating the interaction configuration (faces blurred to de-identify participants). Two cameras on tripods on the floor pointed to the mother and to the infant respectively; the recordings were later synchronized for coding. The dyad interacted with one another using three sets of toys. The first set contained three blocks (yellow, red, green), two bugs (green, red), and rings; the second set contained multi-colored nesting cups, a ball, and a duck; the third set contained a turtle, two dolls, a bird, and a boat. These toy sets were chosen to incorporate various shapes, colors, and nameable categories, and to afford different actions.

nonverbal actions. All utterances were transcribed, and actions classified, with their start and stop times specified (frame-wise, 10 Hz precision), using coding protocols developed within our lab (available at https://osf.io/bnyhk). Utterances were defined as bouts of meaningful speech separated by pauses >200ms[4] (e.g., Chang et al., 2016). Different coders independently coded behaviors including: infants' gaze target (object or mother's face), infants' object-touches (defined as any deliberate contact of infant's hand, arm, or mouth to an object), infant locomotion (by self or mother), and mothers' manual actions (to infant-visually-attended or -unattended object; see Table 1). To assess reliability, a second coder independently annotated 20% of files (randomly selected). Cohen's kappas (Cohen, 1968) were .76 for infant gaze, .81 for infant touches, .81 for infant locomotion, and .88 for mother manual actions.

We selected the 67 most frequent verbs and 17 most frequent pronouns from the dataset. These occurred at least eight times total and were used by at least five mothers. Frequencies of specific verbs and pronouns are shown in Supplementary Tables S1 and S2. Verbs were classified semantically as *Movement* (e.g., swim, go; Laakso & Smith, 2007;

[4]To assess the robustness of this operationalization, we randomly sampled 10% of all utterances. In this sample, 56.6% contained one grammatically complete sentence (e.g., "where's the duck?"), 30.0% were one-word utterances (e.g., "look!"), 7.6% contained two or three grammatically complete sentences with a gap < 200ms (e.g., "what's that? is that your turtle?"), and 5.8% contained grammatically incomplete sentences (e.g., "just a little person"). Thus, this definition mapped most (86.6%) utterances onto grammatical single (or one-word) sentences.

https://doi.org/10.1017/S0305000923000636 Published online by Cambridge University Press

Table 1. Categories of Embodied Behavior Variables and Specific Categories Coded

Embodied variable category	Categories coded
Infants' gaze target	Object
	Mother's face
Number of objects infants touching	None
	One
	Two
Infant locomotion	Crawling
	Walking
	Mother moving infant
Mothers' manual actions target match	Matching infant gaze
	Not matching

Note. Mothers and infants' behaviors were coded with Datavyu (http://datavyu.org), a free open-source software application.

West et al., 2022), *Object-action* (e.g., squeeze, hold; West et al., 2022), *Cognition/Perception* (e.g., think, see; Davis & Landau, 2021; Laakso & Smith, 2007), or *Volition* (e.g., want, like; Laakso & Smith, 2007) verbs. In addition, verbs were tagged for the syntactic categories *Transitive* (e.g., want, eat; Kline et al., 2017), *Intransitive* (e.g., swim, look) and *Auxiliary* (e.g., do, can; Tincoff et al., 2000). Pronouns were categorized as *First person* (e.g., I, me), *Second person* (e.g., you, your), *Third person* (e.g., she, his), or *Deictic* (e.g., that, this; Strauss, 2002). Verbs could belong to multiple categories, whereas pronouns only belonged to one.

Data analysis

To test whether verb categories were differentiated by linguistic and embodied contexts in mothers' naturalistic IDS, we constructed binomial Generalized Linear Mixed Models (binomial GLMM) for each verb category (glmer R function[5]; Bates et al., 2015). Data were separated by utterance, with binary columns for each verb category and for each contextual variable. Co-occurrence with a pronominal context was defined as the verb and pronoun occurring in the same utterance. Co-occurrence with any embodied context was defined as the verb utterance overlapping temporally with the embodied context.

Each verb category was entered as a predicted variable, and pronominal and embodied contextual variables were entered as predictors. Dyad was included as a random effect. Predictors were considered significant if $p < 0.05$. We calculated predictor effect sizes as Odds Ratios (OR), which are independent of variable base rates (Spitznagel & Helzer, 1985). A predictor more likely than chance to co-occur with a verb type has OR > 1.0; a predictor less likely than chance to co-occur has OR < 1.0[6].

[5]www.rdocumentation.org/packages/lme4/versions/1.1-28/topics/glmer

[6]ORs above 1.68 or below 0.6 are considered small effects; those above 3.47 or below 0.29 are medium effects, and above 6.71 or below 0.15 are large effects (Chen et al., 2010).

https://doi.org/10.1017/S0305000923000636 Published online by Cambridge University Press

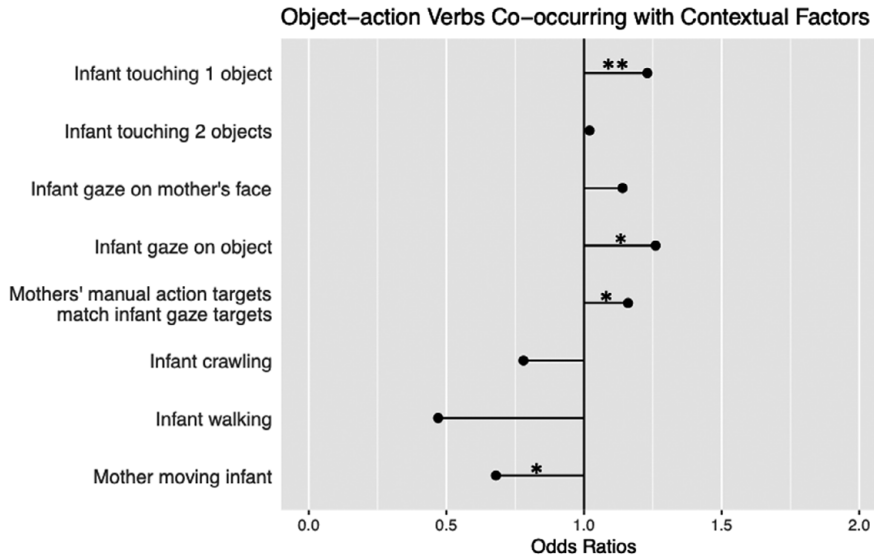

Figure 2. Odds Ratios of Object-action Verbs Co-occurring with Different Contextual Factors
Note. Odds ratios indicating how many times more (OR > 1) or less (OR < 1) likely than chance *Object-action Verbs* co-occurred with various contextual features (listed along Y axis). Significance levels were obtained from binomial linear mixed effects models. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Results

Mothers on average produced 17.06 utterances/min ($SD = 7.94$), each containing a mean of 3.23 words ($SD = 1.63$). Mothers held objects on average 10.90 times/min ($SD = 5.03$). Infants on average locomoted 0.50 times ($SD = 0.38$), were moved by mothers 0.62 times ($SD = 0.41$), produced 13.10 gaze fixations to an object or mother's face ($SD = 4.64$), and touched objects 7.56 times ($SD = 3.48$), all per min. There were no significant gender differences in any of these rates (two-tailed, all $ps > 0.05$).

GLMMs revealed that verbs were differentiated by their linguistic and embodied contexts. As we predicted, co-occurrence patterns differentiated *Object-action* verb use (Figure 2; Table 2), differentiated *Movement* vs. *Cognition/Perception* verbs (Figure 3; Table 3), and differentiated *Cognition/Perception* vs. *Volition*-based mental verbs (Figure 4; Table 4). Our analyses focused on the three predictions above. However, the full model of every verb category in relation to all contextual factors is provided in Table S4.

Co-occurrence of object-action verbs with joint attention

Mothers used *Object-action* verbs more often than chance when focusing on the same object as their infants (i.e., joint-attention; Table 2, Figure 2). *Object-action* verbs also co-occurred above chance when infants touched one object ($OR = 1.23$, $p < 0.01$) or gazed at one object ($OR = 1.26$, $p < 0.05$). However, they occurred BELOW chance when mothers moved the infant ($OR = 0.68$, $p < 0.05$). Using object-action verbs during joint-attention to objects might be optimal input for infants to learn associations between actions and related verbs (e.g., rolling a ball and "roll").

https://doi.org/10.1017/S0305000923000636 Published online by Cambridge University Press

Table 2. Co-occurrences of Object-action Verbs With Different Contextual Factors

	Object-action verbs
Infant touching 1 object	1.23** "What can we **do** with these?"
Infant touching 2 objects	n.s.
Infant gaze on mother's face	n.s.
Infant gaze on object	1.26* "let's **put** her in the boat"
Mothers' manual action targets match infant gaze targets	1.16* "I will **put** a duck in your cup"
Infant crawling	n.s.
Infant walking	n.s.
Mother moving infant	0.68*

Note. Numbers reported are odds ratios indicating how many times more or less likely than chance for a verb type and a context to co-occur. Significance levels were obtained from binomial linear mixed effects models. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Examples of utterances for each significant co-occurrence pattern are provided. The bolded verb was the most frequent *Object-action* verb for each co-occurrence scenario.

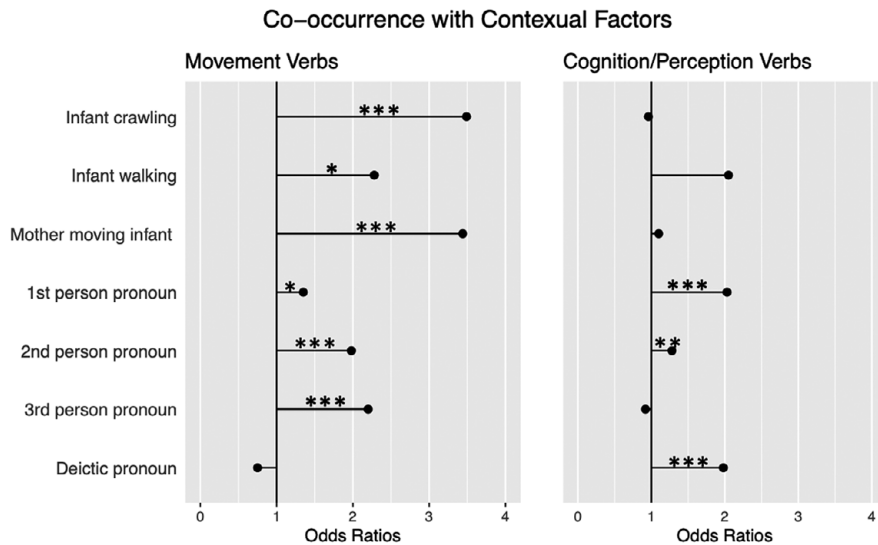

Figure 3. Odds Ratios of Movement and Cognition/Perception Verbs Co-occurring with Different Contextual Factors

Note. Odds ratios indicating how many times more (OR > 1) or less (OR < 1) likely than chance for *Movement* (left panel) and *Cognition/Perception Verbs* (right panel) co-occurred with various contextual features (listed along Y axis). Significance levels were obtained from binomial linear mixed effects models. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Differentiation of movement and cognition/perception verbs

Movement and *Cognition/Perception* verbs co-occurred with distinctly different embodied and linguistic contexts (Table 3, Figure 3). Regarding embodied factors, *Movement* verbs co-occurred above chance with infant crawling ($OR = 3.49$, $p < 0.001$), infant walking ($OR = 2.28$, $p < 0.05$), and mother moving the infant ($OR = 3.44$,

https://doi.org/10.1017/S0305000923000636 Published online by Cambridge University Press

Table 3. Co-occurrences of Movement and Cognition/Perception Verbs With Different Contextual Factors

	Movement verbs	Cognition/Perception verbs
Infant crawling	3.49 *** "Ok **come** back here sweetie pie."	n.s.
Infant walking	2.28 * "Where you **going**?"	n.s.
Mother moving infant	3.44 *** "**Come** here, relax."	n.s.
1st person pronoun	1.35* "There we **go.**"	2.03 *** "I **see** you."
2nd person pronoun	1.98 *** "Where are you **going** Mr.?"	1.28 ** "You **see** the bears?"
3rd person pronoun	2.20 *** "Where does a boat **go**? Does it float in water?"	n.s.
Deictic pronoun	n.s.	1.98 *** "**Look** at this!"

Note. Numbers reported are odds ratios indicating how many times more or less likely than chance for a verb type and a context to co-occur. Significance levels were obtained from binomial linear mixed effects models. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Examples of utterances for each significant co-occurrence pattern are provided. The bolded verb was the most frequent *Movement* or *Cognition/Perception* verb for each co-occurrence scenario.

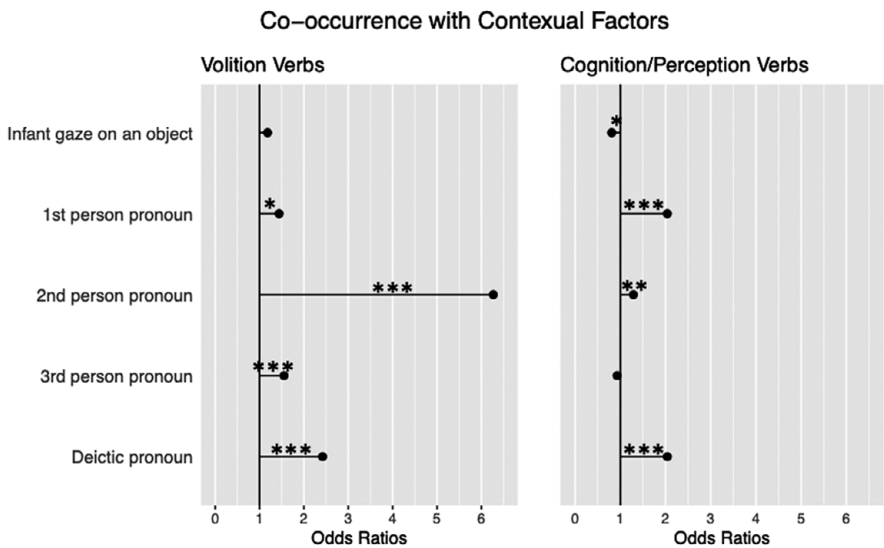

Figure 4. Odds Ratios of Volition and Cognition/Perception Verbs Co-occurring with Different Contextual Factors
Note. Odds ratios indicating how many times more (OR > 1) or less (OR < 1) likely than chance *Volition* and *Cognition/Perception Verbs* co-occurred with various contextual features (listed along Y axis). Significance levels were obtained from binomial linear mixed effects models. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

$p < 0.001$), whereas *Cognition/Perception* verbs co-occurred near chance levels with these contexts. Regarding linguistic context, *Movement* verbs co-occurred above chance with 3rd-person pronouns ($OR = 2.20$, $p < 0.001$) and near chance with deictic pronouns, whereas *Cognition/Perception* verbs co-occurred above chance with deictic pronouns ($OR = 1.98$, $p < 0.001$), and near chance with 3rd-person pronouns.

https://doi.org/10.1017/S0305000923000636 Published online by Cambridge University Press

Table 4. Co-occurrences of Volition and Cognition/Perception Verbs With Different Contextual Factors

	Volition verbs	Cognition/Perception verbs
Infant gaze on an object	n.s.	0.81 *
1st person pronoun	1.44 * "**Let**'s see if we can play a game."	2.04 *** "Can I **see** the ball?"
2nd person pronoun	6.27 *** "You **like** that one?"	1.29 ** "Did you **see** all the colors?"
3rd person pronoun	1.55 *** "Do you **wanna** play with it?"	n.s.
Deictic pronoun	2.42 *** "Do you **like** that one?"	2.04 *** "Hey **look** at this."

Note. Numbers reported are odds ratios indicating how many times more or less likely than chance for a verb type and a context to co-occur. Significance levels were obtained from binomial linear mixed effects models. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Examples of utterances for each significant co-occurrence pattern are provided. The bolded verb was the most frequent *Volition* or *Cognition/Perception* verb under each co-occurrence scenario.

Differentiation of cognition/perception and volition verbs

Within the class of mental verbs, *Volition* and *Cognition/Perception* verbs shared some contextual distribution patterns but differed in others (Table 4, Figure 4). Both verb types co-occurred above chance with 1st-person (*Volition*: $OR = 1.44$, $p < 0.05$; *Cognition/Perception*: $OR = 2.04$, $p < 0.001$), 2nd-person (*Volition*: $OR = 6.27$, $p < 0.001$; *Cognition/Perception*: $OR = 1.29$, $p < 0.01$), and deictic (*Volition*: $OR = 2.42$, $p < 0.001$; *Cognition/Perception*: $OR = 2.04$, $p < 0.001$) pronouns. However, *Volition* verbs co-occurred above chance with 3rd-person pronouns ($OR = 1.55$, $p < 0.001$) and near chance level with infant gazing at an object. In contrast, *Cognition/Perception* verbs co-occurred near chance with 3rd-person pronouns, and below chance with infant gazing at an object ($OR = 0.81$, $p < 0.05$). Note that *Volition* verbs were more likely to co-occur with 2nd- than 1st-person pronouns, whereas *Cognition/Perception* verbs showed the opposite pattern. Laakso and Smith (2007) has reported similar pronominal co-occurrence differences between *Volition* and *Cognition/Perception* verbs.

In addition to analyzing the co-occurrences between verb types and various contextual variables, the dendrogram of verbs based on hierarchical clustering (hclust function in R[7]; R Core Team, 2021) is shown in Figure 5. The input to hclust was the probability of each verb co-occurring with each of 25 contextual factors (17 pronouns and 8 embodied variables). This visualization shows how the similarity of verbs' context predicts various verbs' common semantic or syntactic categories.

Discussion

Infants acquire word knowledge as observers and participants in social interactions. Historically, investigations of infants' word learning have largely focused on object nouns; however, researchers have argued that verbs are harder for infants and children to learn (e.g., Golinkoff & Hirsh-Pasek, 2008). However, little research has examined how infants eventually learn verbs from naturalistic interactions, and in particular what verbal and nonverbal information is available for infants to disambiguate verb usage and meanings.

[7] www.rdocumentation.org/packages/stats/versions/3.6.2/topics/hclust

https://doi.org/10.1017/S0305000923000636 Published online by Cambridge University Press

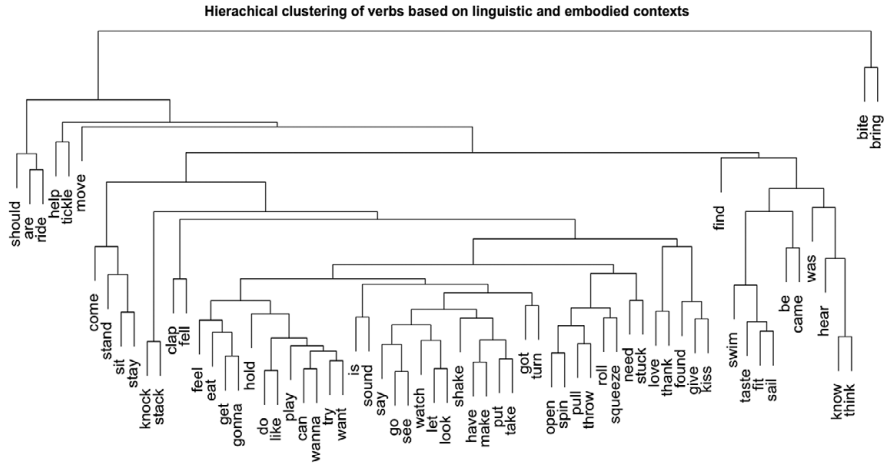

Figure 5. Clustering of Verbs

Note. Hierarchical clustering showing the proximity relationships among verbs, based on each verb's co-occurrences with pronominal and embodied contexts. The clusters were generated with pairwise complete-linkage clustering of Euclidean distances between verbs.

The current study provides new evidence that pronouns, infant manual actions, gaze, and locomotion, as well as parent object actions, differ across verb types.

Object-action verbs

Infants prefer looking at caregivers manipulating objects more than caregivers' faces or isolated objects (Deák et al., 2014). Additionally, caregiver/infant object attention-sharing is positively correlated with caregivers' utterance rates and infants' object-noun vocabularies (Tomasello & Farrar, 1986). During free-play, caregivers tend to use object-action verbs when either they or their infants manipulate objects (West et al., 2022). Moreover, infants attend to caregivers' actions around times when the caregiver uses an action verb (Liu et al., 2019). Our results complement these findings: mothers produced object-action verbs (e.g., spin, turn) while they manipulated an object and infants watched, or when infants touched or looked at an object.

Like object-naming nouns, then, object-action verbs were frequently used when caregivers and infants jointly attended to an object. However, despite this contextual similarity, object-action verbs are learned and produced by English-learning infants later than object nouns (Fenson et al., 1994). Can usage co-occurrence statistics explain this gap in learning object-related verbs? One possible explanation is that caregivers more regularly produce object nouns than action verbs while the infant is watching the action (or, relatedly, that verb forms are more variable than object labels during actions). Another possibility is that actions are more transient, so infants are less likely to attend to both a particular action and its co-occurring action verb (Liu et al., 2019). To test these hypotheses, finer-grained, higher-power future studies should explore verb-noun, verb-action, and noun-action co-occurrences within naturalistic infant-caregiver interactions, ideally across diverse languages and cultures. Such studies could reveal generalized distributional patterns that support infants' verb learning.

https://doi.org/10.1017/S0305000923000636 Published online by Cambridge University Press

Mental versus movement verbs

Mental verbs are learned late. English-learning children produce and understand some action or movement verbs in the second year (Bloom et al., 1975), but do not regularly produce or comprehend multiple mental verbs until the third year (Shatz et al., 1983). However, children's production of mental verbs can be facilitated by syntactic and observational cues in a scene description task (Papafragou et al., 2007). Thus, it is natural to wonder how mental verbs are used by caregivers, and with what contextual cues. Our results indicate that some mental verbs are used frequently by English-speaking caregivers of 12-month-olds, and co-occur with specific lexical and nonverbal contextual elements. For example, cognition/perception verbs were used significantly more when caregivers handled the object infants gazed at, and used deictic pronouns. Cognition/Perception verbs also co-occurred with first person pronouns, whereas movement verbs co-occurred with second and third person pronouns, replicating and expanding findings from Laakso and Smith (2007). Movement verbs were used often when infants locomoted or when caregivers moved infants (replicating West et al., 2022), whereas cognition/perception verbs were used near chance frequency during those times. Co-occurrences between movement verbs and infant locomotion might build infants' semantic associations between specific verbs and actions. In addition, caregivers often used the same movement verbs with first, second or third person pronouns in different utterances, to describe the movement of the dyads (e.g., "Let's go and check that out!"), infants (e.g., "where are you going mister?") or objects (e.g., "where did the boat go?"). These cross-situational usages of the verb "go" might bootstrap infant learning of how "go" references the movements of different entities. Plausibly, movement verbs were deliberately chosen by mothers to narrate salient events when the infant locomoted or was relocated by the mother. By contrast, mothers used mental verbs to comment on infants' mental states while they were looking at and handling objects. However, these events also co-occurred with other types of verbs, notably object-action verbs. This non-specificity of verbs to context, and of context to verbs, might partly explain toddlers' late acquisition of mental verbs (Bloom et al., 1975; Shatz et al., 1983).

Cognition/perception versus volition verbs

Toddlers' acquisition of mental verbs might be facilitated by co-occurrence statistics that differ among semantic sub-types. Although both cognition/perception and volition verbs co-occurred weakly (i.e., small effect) with 1st-person pronouns, the co-occurrence with 2nd-person pronouns was much greater (i.e., medium effect) for volition than cognition/perception verbs. These effect size differences could hypothetically contribute to infants' differentiation of pronouns that describe their own and others' mental states. For example, when infants indicate a desire for one specific toy, they might hear caregivers say, "you like that one?" Their affective state, and prior learning that a caregiver can satisfy their desire for objects, might increase their attentiveness to the caregiver's utterance. Such co-occurrences, if regular, could bootstrap infants' mapping of the words "you" and "like" to their own volitional states. Comparatively, caregiver use of cognition verbs to describe their own mental activities (e.g., "I think that's good") might be less salient to infants, due to a less intense and/or focused co-occurring affective state. This would predict that infants comprehend volition verbs earlier than cognition verbs, because volition verbs are more often associated with their own experience of salient emotional states. Future studies should

https://doi.org/10.1017/S0305000923000636 Published online by Cambridge University Press

compare the age of acquisition volition and cognition verbs, relative to pronoun context as well as social-emotional context, and investigate whether these co-occurrences might better explain acquisition differences.

Summary

This study examined how different verb categories co-occurred with distributional and embodied contexts during caregiver-infant free-play. We found that caregiver linguistic input and play contexts provided statistical regularities that could facilitate infant verb acquisition. However, several limitations should be addressed by future studies. First, we segmented caregiver utterances with a temporal cut-off, which might alter the complexity of utterance content that infants might further differentiate. Future studies could consider additional utterance boundary cues like terminal pitch contours or grammatical units. Also, our linguistic context only considered pronouns, but nouns and adverbs also co-occur non-randomly with verbs. We focused on pronouns because they are frequent, and their co-occurrence patterns have seldom been examined as a distributional cue for infants. Future studies with larger datasets should investigate co-occurrences between verbs and nouns, adverbs, and other closed-class elements as well as pronouns. Lastly our results only sampled one language and subculture, and are therefore limited in generalizability.

The present study is among several that document the co-occurrence patterns of both language AND embodied contextual variables across a range of verb categories within naturalistic caregiver-infant interactions. The results suggest that infant acquisition of verbs could be supported by learning mechanisms and environment statistics parallel to those that support acquisition of object nouns – that is, a capacity to learn contextual regularities, rather than a specific "verb-learning module". Our results suggest potential explanations for the later acquisition of verbs, and specifically mental verbs: notably, contextual factors co-occurring with mental verbs were least specific. The approach exemplified here should be applied to datasets from diverse populations, for comparisons that will broaden our understanding of how infant-caregiver contextual statistics support verb learning.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0305000923000636.

Acknowledgements. This research was supported by grants from the National Science Foundation (SES-0527756) and from the UC - San Diego Academic Senate. We thank student members of the Cognitive Development Lab for their assistance with data collection and coding, and we thank the families who participated in this research.

Competing interest. The authors declare none.

References

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

Bates, D., **Maechler, M.**, **Bolker, B.**, & **Walker, S.** (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48.

Bloom, L., **Lightbown, P.**, & **Hood, L.** (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development*, 1–79.

Bortfeld, H., **Morgan, J. L.**, **Golinkoff, R.**, & **Rathbun, K.** (2005). Mommy and me: Familiar names help launch babies into speech stream segmentation. *Psychological Science*, **16**, 298–304.

Chan, W. H., & **Nicoladis, E.** (2010). Predicting two Mandarin-English bilingual children's first 50 words: Effects of frequency and relative exposure in the input. *International Journal of Bilingualism*, **14**(2), 237–270.

Chang, L., **de Barbaro, K.**, & **Deák, G.** (2016). Contingencies between infants' gaze, vocal, and manual actions and mothers' object-naming: Longitudinal changes from 4 to 9 months. *Developmental Neuropsychology*, **41**(5-8), 342–361.

Chen, H., **Cohen, P.**, & **Chen, S.** (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—simulation and Computation®*, **39**(4), 860–864.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, **70**(4), 213.

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.

Datavyu Team (2014). Datavyu: A Video Coding Tool. Databrary Project, New York University. URL http://datavyu.org.

Davis, E. E., & **Landau, B.** (2021). Seeing and believing: the relationship between perception and mental verbs in acquisition. *Language Learning and Development*, **17**(1), 26–47.

Deák, G. O., **Krasno, A.**, **Triesch, J.**, **Lewis, J.**, & **Sepeda, L.** (2014). Watch the hands: Human infants can learn gaze-following by watching their parents handle objects. *Developmental Science*, **17**, 270–281.

Deák, G. O., **Triesch, J.**, **Krasno, A.**, **de Barbaro, K.**, & **Robledo, M.** (2013). Learning to share: The emergence of joint attention in human infancy. *Cognition and brain development: Converging evidence from various methodologies p.* 173–210. American Psychological Association.

ELAN (Version 6.7) [*Computer software*]. (2023). Nijmegen: Max Planck Institute for Psycholinguistics, The *Language* Archive. Retrieved from https://archive.mpi.nl/tla/elan.

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*, i-185.

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.

Gentner, D. (2006). Why verbs are hard to learn. In K. Hirsh-Pasek, & R. Golinkoff, (Eds.) *Action meets word: How children learn verbs* (pp. 544–564). Oxford University Press

Golinkoff, R. M., & **Hirsh-Pasek, K.** (2008). How toddlers begin to learn verbs. *Trends in Cognitive Sciences*, **12**(10), 397–403.

Kline, M., **Snedeker, J.**, & **Schulz, L.** (2017). Linking language and events: Spatiotemporal cues drive children's expectations about the meanings of novel transitive verbs. *Language Learning and Development*, **13**(1), 1–23.

Laakso, A., & **Smith, L. B.** (2007). Pronouns and verbs in adult speech to children: A corpus analysis. *Journal of Child Language*, **34**(4), 725–763.

Liu, S., **Zhang, Y.**, & **Yu, C.** (2019, July-August). Why some verbs are harder to learn than others: A micro-level analysis of everyday learning contexts for early verb learning [Paper presentation]. Proceedings of the 42nd Annual Meeting of the Cognitive Science Society in Toronto, Canada.

Maguire, M., **Hirsh-Pasek, K.**, & **Golinkoff, R.** (2006). A unified theory of word learning: Putting verb acquisition in context. In K.Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp.364– 391). New York: Oxford University Press.

Messer, D. J. (1978). The integration of mothers' referential speech with joint play. *Child Development*, 781–787.

Nazzi, T., **Dilley, L. C.**, **Jusczyk, A. M.**, **Shattuck-Hufnagel, S.**, & **Jusczyk, P. W.** (2005). English-learning infants' segmentation of verbs from fluent speech. *Language and Speech*, **48**(3), 279–298.

Papafragou, A., **Cassidy, K.**, & **Gleitman, L.** (2007). When we think about thinking: The acquisition of belief verbs. *Cognition*, **105**(1), 125–165.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rattermann, M. J., & **Gentner, D.** (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development*, **13**(4), 453–478.

Redington, M., **Chater, N.**, & **Finch, S.** (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, **22**(4), 425–469.

Shatz, M., **Wellman, H. M.**, & **Silber, S.** (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition*, **14**(3), 301–321.

Snedeker, J., **Li, P.**, & **Yuan, S.** (2003). Cross-Cultural Differences in the Input to Early Word Learning. In *Proceedings Annual Meeting of the Cognitive Science Society* (Vol. **25**).

Spitznagel, E. L., & **Helzer, J. E.** (1985). A proposed solution to the base rate problem in the kappa statistic. *Archives of General Psychiatry*, **42**(7), 725–728.

Strauss, S. (2002). This, that, and it in spoken American English: a demonstrative system of gradient focus. *Language Sciences*, **24**(2), 131–152.

Suarez-Rivera, C., **Schatz, J. L.**, **Herzberg, O.**, & **Tamis-LeMonda, C. S.** (2022). Joint engagement in the home environment is frequent, multimodal, timely, and structured. *Infancy*, **27**(2), 232–254.

Syrett, K., **Arunachalam, S.**, & **Waxman, S. R.** (2014). Slowly but surely: Adverbs support verb learning in 2-year-olds. *Language Learning and Development*, **10**(3), 263–278.

Tamis-LeMonda, C. S., **Kuchirko, Y.**, & **Song, L.** (2014). Why is infant language learning facilitated by parental responsiveness?. *Current Directions in Psychological Science*, **23**(2), 121–126.

Tardif, T., **Gelman, S. A.**, & **Xu, F.** (1999). Putting the "noun bias" in context: A comparison of English and Mandarin. *Child Development*, **70**(3), 620–635.

Tardif, T., **Shatz, M.**, & **Naigles, L.** (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language*, **24**(3), 535–565.

Tincoff, R., **Santelmann, L.**, & **Jusczyk, P.** (2000). Auxiliary verb learning and 18-month-olds' acquisition of morphological relationships. In *Proceedings of the 24th annual Boston University conference on language development* (Vol. **2**, pp. 726–737). Somerville, MA: Cascadilla Press.

Tomasello, M., & **Farrar, M. J.** (1986). Joint attention and early language. *Child Development*, **57**(6), 1454–1463.

Waxman, S., **Fu, X.**, **Arunachalam, S.**, **Leddon, E.**, **Geraghty, K.**, & **Song, H. J.** (2013). Are nouns learned before verbs? Infants provide insight into a long-standing debate. *Child Development Perspectives*, **7**(3), 155–159.

West, K. L., **Fletcher, K. K.**, **Adolph, K. E.**, & **Tamis-LeMonda, C. S.** (2022). Mothers talk about infants' actions: How verbs correspond to infants' real-time behavior. *Developmental Psychology*, **58**(3), 405.

West, K. L., & **Iverson, J. M.** (2017). Language learning is hands-on: Exploring links between infants' object manipulation and verbal input. *Cognitive Development*, **43**, 190–200.

Willits, J. A., **Seidenberg, M. S.**, & **Saffran, J. R.** (2014). Distributional structure in language: Contributions to noun–verb difficulty differences in infant word recognition. *Cognition*, **132**(3), 429–436.

Yu, C., & **Smith, L. B.** (2012). Embodied attention and word learning by toddlers. *Cognition*, **125**(2), 244–262.

Yuan, S., **Fisher, C.**, **Kandhadai, P.**, & **Fernald, A.** (2011). You can stipe the pig and nerk the fork: Learning to use verbs to predict nouns. In *Proceedings of the 35th Annual Boston University Conference on Language Development* (pp. 665–677). Boston, MA: Cascadilla Press.

Cite this article: Zhang, V.H., Chang, L.M., & **Deák, G.O.** (2024). The distributional and embodied contexts of verbs in caregiver-infant interactions. *Journal of Child Language* 1–15, https://doi.org/10.1017/S0305000923000636