The impact of caregivers' multimodal behaviours on children's word learning: A corpus-based investigation

Ed Donnellan[1]
ed.donnellan@ucl.ac.uk

Antonia Jordan-Barros[1]
antonia.jordan@ucl.ac.uk

Niki Theofilogiannakou[1,2]
niki.theofilogiannakou.20@ucl.ac.uk

Gwen Brekelmans[3]
g.brekelmans@qmul.ac.uk

Margherita Murgiano[1]
marghe.murgiano@gmail.com

Yasamin Motamedi[1]
yasamin.motamedi@gmail.com

Beata Grzyb[1]
b.grzyb@ucl.ac.uk

Yan Gu[1,4]
yan.gu@ucl.ac.uk

Gabriella Vigliocco[1]
g.vigliocco@ucl.ac.uk

[1]Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, UK
[2]School of Psychology, University of East Anglia, Norwich, NR4 7TJ, UK
[3]School of Biological and Behavioural Sciences, Queen Mary University of London, London, E1 4NS, UK
[4]Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK

Abstract

Studies have shown the importance of caregivers' multimodal behaviours (e.g., prosody, gestures, gaze) on children's word learning. However, most studies focus on only one specific behaviour (e.g., only prosody). Here, we investigate which multimodal behaviours used by caregivers best predict children's word learning and vocabulary growth. Using data from the ECOLANG corpus, we analysed caregiver behaviour in semi-naturalistic interactions with their child (3 to 4 years old) in which they talked about known and unknown toys. We analysed caregivers' (*n*=36) use of multimodal cues while labelling the objects, specifically their use of yes/no questions, pitch, representational gestures, pointing, object manipulations and gaze. Caregivers' pitch, use of yes/no questions and pointing predicted children's word learning. In particular, higher pitch when labelling unknown toys predicted immediate word learning. The degree to which caregivers used higher pitch when producing the label for known compared to unknown toys predicted both immediate learning and vocabulary growth. Furthermore, the degree to which caregivers used yes/no questions more for unknown toys predicted immediate learning, while the frequency of yes/no questions when naming unknown toys predicted vocabulary growth. Lastly, caregiver pointing also predicted immediate label learning and vocabulary growth, but in the opposite direction from prosody: the more they pointed towards known toys, the better children's learning of novel toy labels. Other behaviours did not predict word learning. Overall, these results provide evidence for the important role of multimodal caregiver behaviours, particularly prosody, on children's lexical development.

Keywords: multimodal communication; caregiver input; language learning

Introduction

It is clear from individual differences in children's vocabulary size across the early years (Fenson et al., 1994; Rowe et al., 2012) that the rate at which children learn new words varies substantially. A number of interacting factors may contribute to this variability, including caregiver socio-economic status (SES, Fernald et al., 2013), responsiveness (McGillion et al., 2017), quantity of input (Hurtado et al., 2008) and children's early phonetic perception (Tsao et al., 2004). Variability in multimodal behaviours (both verbal and non-verbal) used by caregivers when interacting with their children can also contribute to variability in learning trajectories above and beyond factors such as SES and input quantity (Cartmill et al., 2013). Here, we focus on a number of verbal and non-verbal behaviours (speech pitch, use of yes/no questions, representational gestures, pointing, object manipulation and gaze direction) and assess their role in immediate learning and vocabulary growth.

Multimodal behaviours and word learning

There are a number of verbal and non-verbal cues that can facilitate *immediate* word learning as well as *vocabulary growth*.

Focusing on verbal cues, repeated use of a novel label (i.e., the name of an object) in the presence of an associated object could facilitate slow word-mapping, with learning processes happening over time in the presence of different arrays of potential targets (McMurray et al., 2012). However, the precise relationship between frequency of caregiver word use and immediate word learning is not straightforward. Schroer & Yu (2022) found that the number of caregiver repetitions of a novel word is by itself not enough to predict 1- to 2-year-olds' immediate word learning. Instead they found that word learning was predicted by the frequency with which novel words were produced by caregivers while children attended and manipulated the objects. Additionally, repeated naming of novel objects that occurs specifically in quick bursts predicts immediate word learning (Slone et al., 2023). In terms of vocabulary growth, Bang et al. (2022) found that the frequency that caregivers used object labels with 18-month-olds predicted expressive vocabulary at 25 months.

Furthermore, the type of utterance in which an object name is used in speech to children could be predictive. Dong et al. (2021) found that the number of utterances consisting of yes/no questions used by caregivers to their 3- to 4-year-old children (but not declaratives, imperatives or wh- questions)

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

predicts children's immediate word learning and later vocabulary (see also Rowe, 2008). Though this finding related to all utterances, regardless of whether they contained an object name, it is possible that this could also apply to label-containing utterances.

Caregivers often use child directed language (CDL; Cox et al., 2022). The higher-pitched speech of CDL may enhance word learning by attracting children's attention (Cristia, 2013) or improving their ability to recognize phonological properties (Trainor & Desjardins, 2002). Indeed, higher-pitched caregiver speech when introducing novel object labels is a robust predictor of children's word learning from 17 up to 48 months of age (Graf-Estes & Hurley, 2012; Grassman & Tomasello, 2007; Shi et al., 2022; Ma et al., 2011), and even adult label learning (Fillipi et al., 2014). Beyond immediate measures of word learning, evidence from infancy of mean pitch effects on vocabulary size is mixed. 3- to 14-month-olds with caregivers who introduce novel object labels using higher-pitched speech (e.g., in the phrase "pet the *gorilla*") had larger expressive vocabularies at 12 to 13 months (Porritt et al., 2014). However, Kalashnikova and Burnham (2018) did not find a relationship between higher pitched caregiver speech to infants from 7 to 19 months and expressive vocabulary at 15 and 19 months. Moreover, beyond infancy, Shi et al. (2022) showed that while higher-pitched speech when introducing unknown object labels to 3- to 4-year-olds did not predict receptive vocabulary one year later, the degree to which caregivers produced higher speech for unknown compared to known object labels did.

Caregivers can also provide nonverbal indexical cues to facilitate fast mapping of novel words to present referents, e.g., pointing at, manipulating or gazing at objects whilst labelling them (Bohn & Frank, 2019; Brooks & Meltzoff, 2008; Radar & Zukow-Goldring, 2010; Vigliocco et al., 2019). Studies have suggested that children between 2 and 5 years use a combination of linguistic and indexical cues (particularly eye gaze and pointing), to map novel labels to objects (Grassman & Tomasello, 2010; Nurmsoo & Bloom, 2008; Yow & Markman, 2014). Evidence suggests that gesturing to novel objects whilst naming them boosts children's immediate word learning. For example, 18- to 30-month-olds learned novel words in a story better when caregivers pointed towards and labelled objects in the story compared to simply labelling them without pointing (Kalagher & Yu, 2006). Booth et al. (2008) showed that 28- to 31-month-olds learned unfamiliar labels for novel objects better when caregivers simultaneously gazed towards objects or gazed while pointing, touching or manipulating the object. Furthermore, there is indirect evidence suggesting caregiver pointing or gazing whilst naming predicts infant vocabulary size. Pan et al. (2005) found that rates of caregiver pointing predicted children's expressive vocabulary growth between 14 and 36 months. Though the study did not focus on pointing accompanying naming, most observed pointing was co-speech, and so this effect could plausibly apply to naming instances. Law et al. (2012) tested 18-month-olds' novel word learning in a gaze-following task where they had to follow an experimenter's gaze to correctly identify the referent. This ability predicted measures of the child's receptive vocabulary at 24 months, and expressive vocabulary at 24 and 30 months. Caregiver naming whilst gazing at objects could therefore facilitate label-object mapping and enhance children's later vocabulary.

Even when objects are absent, mapping may be achieved through caregivers' use of representational gestures (iconically representing object properties whilst labelling; Perniss & Vigliocco, 2014). Representational gestures facilitate verb learning, as representational gestures often depict motion (Aussems, 2020; Sweller et al., 2020; Goodrich & Kam, 2008). However, work has also demonstrated that 4-year-olds learned object labels (i.e., nouns) accompanied by noun-depicting representational gestures (e.g., the shape of a particular animal) better than labels that were provided without such gestures (Vogt & Kaushke, 2017; see also Capone & McGregor, 2005). Caregivers' use of representational gestures is predictive of later vocabulary outcomes too. In a training study, 11-month-olds whose caregivers were encouraged to produce iconic gestures whilst using an associated word (often a label; e.g., scratching underarms while saying "monkey") outperformed control group children (whose caregivers were not instructed in this way) on an array of expressive and receptive vocabulary measures from 15 to 36 months (Goodwyn et al., 2000).

The current study

The current study investigates multimodal caregiver behaviours produced when providing labels for objects to children aged 3 to 4 years old, and how this impacts both the child's learning of previously unknown labels as well as their later vocabulary size. This work addresses four crucial interacting factors that at present preclude a full understanding of which caregivers' multimodal behaviours produced while naming novel objects support children's word learning.

Firstly, much of the evidence surrounding multimodal caregiver behaviours involves task-based experiments in the lab, and doesn't necessarily quantify what behaviours caregivers use in day-to-day interactions with their children. In this study, we investigate multimodal behaviours using data from the ECOLANG corpus (Gu et al., in prep), a dataset focusing on naturalistic interactions between caregivers and their 3- to 4-year-old children taking place in the family home. In this dataset, dyads interact naturalistically, discussing toys that are both familiar and unfamiliar to the child, providing the opportunity to observe caregiver behaviours in a setting closer to its ecological niche.

Secondly, studies on the predictive effects of caregiver multimodal behaviours on word learning typically focus on each behaviour in isolation (e.g., only looking at phonological measures, or focusing only on gestures accompanying labelling). Here, to gain a more complete understanding of what kind of multimodal behaviour predicts learning, we directly contrast relative predictive power of different behaviours in the same sample, thus reducing the possibility that a particular finding could be due to a hidden correlate, i.e., indexing another predictive behaviour (see examples in Bang et al., 2022; Booth et al., 2008). As there are many potential predictors to consider, for speech-properties we focus on mean pitch of labels as a strong phonological predictor of children's word learning (Shi et al., 2022). In terms of semantic properties, we focus on whether labels are produced as part of a yes/no question (e.g., "is this the parsley?"), known to be an important predictor of children's word learning over and above other sentence types (e.g., declaratives, imperatives and wh- questions; Dong et al., 2021). For gestural predictors, we focus on pointing, representational gestures and object manipulations. Finally, we consider whether caregivers are gazing to the objects as they label them.

Thirdly, there are two main ways to conceptualize the relationship between caregiver behaviours and children's language learning. We could consider the variability in caregiver behaviour when labelling objects that they think their child is unfamiliar with. Thus, we could conceptualize variability as simply frequency, or averages, i.e., how much caregivers produce a behaviour whilst labelling, and how this predicts the child's language learning. On the other hand, we could consider variability in the degree to which caregivers *modify* labelling behaviours for words that they think their child is unfamiliar with. Thus, we could conceptualize variability as differences between labelling known and unknown objects, and whether the variation in the magnitude of these differences predicts children's language learning (see Shi et al., 2022). Here, we consider both ways of quantifying caregiver behaviour.

Finally, few studies investigate both immediate word learning (i.e., assessing how labelling behaviours impact on children's recognition of novel words used in the study) and later vocabulary learning (i.e., assessing how differences in labelling behaviours impact on children's vocabulary growth). Here, we make use of data from the ECOLANG corpus to investigate immediate learning (children's recognition of labels for the toys that they were unfamiliar with), as well as assessments of concurrent vocabulary and vocabulary one year later.

Method

Participants

Dyads were taken from a corpus of $N = 38$ dyads involving a caregiver and a child aged 3 to 4 years old (ECOLANG corpus; Gu et al., in prep). Two dyads were not used due to eye-tracker malfunction. Our sample consisted of $n = 36$ caregiver-child dyads (caregivers: F = 35, Age [years] $M = 38.64$, $SD = 3.67$, Range = 29 - 48; children: F = 17, Age [months] $M = 43.06$, $SD = 4.53$, Range = 36 - 52).

A subset of dyads were also used for analysis of vocabulary outcomes one year later. This data was available for $n = 31$ dyads from our sample (caregivers: F = 30, Age [years] $M = 38.77$, $SD = 3.90$, Range = 29 - 48; children: F = 14, Age [months] $M = 42.90$, $SD = 4.61$, Range = 36 - 52).

ECOLANG Corpus

Videoed interaction The corpus involves semi-naturalistic interactions between children and their caregivers in the home. Sessions were video recorded while the caregiver wore a lapel microphone and head-mounted eyetracker (Tobii Pro Glasses 2). Caregivers and children sat at 90° to each other at a table, and talked about 24 toys (drawn from a pool of 98 toys). 12 toys were selected because they were known to the child and 12 were selected as they were unknown (determined by caregiver report). Toys were grouped by category (animals, foods, musical instruments and tools), with 6 toys in each (3 unknown). Dyads spent 6-8 minutes on each set (category), half the time with toys present, and half with toys absent. Category order and whether toys were present or absent first was counterbalanced across dyads.

Caregivers' speech in the corpus was manually annotated using Praat (Boersma & Weenink, 2019). Speech was initially transcribed on an utterance level, defined as a unit that expresses a single event (Berman & Slobin, 1994). Utterances were coded into different types including Yes/No questions (see Dong et al. 2021).

The pitch of object labels was extracted using a Praat script that computed mean F0, and each value was manually checked to correct pitch errors and mistracked points (see Shi et al. 2022, for more detail).

Caregivers' gestures and gaze fixations were annotated using ELAN (Sloetjes & Wittenburg, 2008). Representational gestures were defined as gestures that represent properties of referents, such as the shape or function of an object (e.g., hands forming a circle to represent the round shape of an apple or bringing your hand towards your mouth to represent drinking from a cup). Pointing was defined as gestures that single out a particular referent through deixis (e.g., a canonical index finger point). Object manipulations were any meaningful movement or action performed while touching a toy (e.g., holding a toy to direct the child's attention to it). Actions carried out while incidentally holding objects were not counted. To determine gaze fixations on toys, raw recordings obtained from the eye tracking glasses were processed to mark caregivers' gaze position. Afterwards, the recording was manually annotated by an expert coder on ELAN to mark the specific toy that was the focus of the gaze fixation (coding only those lasting for ≥3 consecutive video frames; see Motamedi et al., 2022 for detail on utterance, gesture and gaze coding in the corpus).

Child object recognition task Immediately after the interaction, children took part in an E-prime recognition task. Children were presented with two pictures of toys side-by-side, while a voice prompted, "Can you help me find the [toy name], where is the [toy name]?". Children pointed to a picture, and an experimenter recorded their response. Each child received 28 trials, consisting of 24 target test trials (2 for each unknown toy used in the interaction), and 4 control trials (featuring known toys). For all target test trials, non-target toys were also unknown toys used in the interaction (i.e., each unknown toy appeared 4 times across test trials: twice as target, twice as non-target).

Child vocabulary Prior to the videoed interaction, the British Picture Vocabulary Scale, 3rd edition (BPVS; Dunn & Dunn, 2009) was used to assess children's concurrent vocabulary. Additionally, a subset of children (*n*=31) completed the BPVS again one year after the interaction.

Measures

For the 36 dyads in the sample, we extracted N = 5794 utterances in which the caregiver used a toy label (n = 2962 unknown to the child, n = 2831 known).

For all utterances, we considered whether it was a *Y/N question*, the *pitch* of the label and whether the utterance containing the label overlapped with a *representational gesture*, *object manipulation*, *pointing* or *gaze* to the toy.

For these measures, we calculated *frequency/mean scores* for each toy that the dyad interacted with. For count behaviours (e.g., pointing), this was the number of times a caregiver pointed at a toy whilst saying the label. For pitch, we calculated the mean pitch across all labelling instances for that toy. We also calculated *difference scores* quantifying differences between caregiver behaviours when naming known and unknown toys. For count behaviours, e.g., pointing, this was the frequency caregivers labelled and pointed at known toys minus the frequency they labelled and pointed at unknown toys. For pitch, we calculated a ratio score, i.e., mean pitch when labelling known toys divided by mean pitch when labelling unknown toys (see Shi et al., 2022).

Children's immediate learning scores For each child, we calculated a binary recognition score for each unknown toy from the recognition task. If they correctly identified the toy in both trials where it was a target, then they were classified as having learned the label (= 1), and if not, they were coded as not having learned the label (= 0).

If a child did not complete both target test trials for a toy in the experiment (due to technical issues with the recognition task), we did not include a score for that toy in the measure of immediate learning. Two participants had scores for <12 unknown toys (having scores for 11 and 8 toys respectively).

Children's vocabulary scores We used BPVS raw scores for concurrent vocabulary and vocabulary one year later.

Analysis

All analyses were conducted in RStudio 9.0.351 (RStudio Team, 2021) running *R* version 4.2.2 using *lme4* (Bates et al., 2015), *lmerTest* (Kuznetsova et al., 2017) and *stats* (R Core Team, 2022) packages.

Immediate learning To analyse the effect of caregiver multimodal behaviours on immediate learning of unknown objects (frequency/mean scores), we constructed logistic mixed-effects models. The dependent variable (DV) was the child's immediate learning score. Independent variables (IVs) were the caregiver's pitch, yes/no questions, pointing, representational gestures, object manipulation, gaze and the total number of labels. We included child's concurrent vocabulary score as a control variable. There was a separate datapoint per toy for each dyad, reflecting the IV and DV relating to that toy. Note, for 10 datapoints (10 toys across *n*=6 participants), the caregiver did not label the toy, so these datapoints were excluded. We included a random intercept of participant, and toy as a random effect on all slopes. In order to ease model convergence, aid interpretability and ensure random effects were identifiable, all IVs were mean-centred and scaled (M = 0, SD = 1). Additionally, we only included data from unknown toys that were used for ≥4 participants. This resulted in 374 datapoints for both DV and IVs, for 37 different toys across the 36 dyads (note that control variables only have 36 datapoints, one per dyad).

To analyse the effect of the difference between multimodal behaviours while producing labels for known and unknown toys on immediate learning outcomes, we again constructed logistic mixed effects models with the same DV, control variable and random effects structure. However, in this model IVs were difference scores, so there were 384 datapoints for the DV (each representing a recognition score) across 37 items, with 1 datapoint per dyad for the IVs and control variables.

Vocabulary growth To analyse the effect of caregiver multimodal behaviours whilst producing novel words on children's later vocabulary, we used simple linear regression (1 datapoint per dyad). The DV was the child's vocabulary score one year later and the control variable was the child's concurrent vocabulary score. For models using frequency/mean scores, IVs represented the frequency/mean of labelling instances across all unknown toys. For models using difference/ratio scores, IVs were identical to the model predicting immediate learning from these differences.

As the sample for vocabulary outcomes was relatively small (n = 31) relative to the number of predictors, we first constructed the full model with all predictors and then used model selection (calculating a model for every possible combination of predictors) and ranking models by AICc, a version of Akaike's Information Criterion that corrects for small sample sizes (Sugiura 1978; Burnham & Anderson, 2002). All models contained the control variable. Across all models, we calculate ΔAICc (AICc for a given model minus the lowest AICc for any model). The model with ΔAICc = 0 is considered the best model, however models with ΔAICc < 2 are considered equivalent (Burnham & Anderson, 2002). To avoid cryptic multiple testing (Forstmeier & Schielzeth, 2011), we present the full model with all predictors, before presenting reduced models.

Multicollinearity Variance Inflation Factors (VIF) were calculated for every model to ensure there was no multicollinearity between predictors. Where VIF scores led to predictor removal, this is noted in the results. All presented models have VIF < 5 for all predictor variables.

Results

Immediate learning outcomes

We first investigated whether the frequency/mean of caregiver labelling of unknown toys (with different multimodal behaviours) predicted immediate learning of unknown toy labels. Of these behaviours, only the pitch of caregiver labels significantly predicted immediate word learning (along with concurrent vocabulary, see Table 1 top). No behaviours relating to caregiver gestures or gaze whilst labelling unknown toys predicted immediate word learning.

Next we investigated whether the difference in the way caregivers labelled unknown and known toys predicted immediate word learning (Table 1 bottom: total labels removed due to high multicollinearity, VIF = 5.64). The difference between caregivers' pitch during labelling significantly predicted immediate word learning. Children's recognition scores were higher when caregivers used higher pitch for unknown labels compared to known. Furthermore, the difference between the amount caregivers pointed to toys while labelling was predictive. Children's scores were higher when their caregivers pointed more to known toys whilst labelling. The difference between the amount caregivers label as part of a yes/no question was predictive (though non-significant, *p* = .056) in the same effect direction as for pitch (see Figure 1).

Table 1: Immediate learning predicted by frequency/mean of multimodal behaviours (top) and difference in labelling behaviours between known/unknown toys (bottom)

	Est	SE	t	p
(Intercept)	1.264	0.228	5.543	.000***
Pitch	0.482	0.220	2.191	.029*
Y/N	-0.025	0.256	-0.098	.922
RG	0.085	0.238	0.358	.720
OM	0.010	0.330	0.029	.977
Pointing	0.127	0.289	0.440	.660
Gaze	0.088	0.265	0.334	.738
Total Labels	0.125	0.399	0.313	.754
Vocab	0.504	0.184	2.740	.006**
(Intercept)	1.226	0.214	5.730	.000***
Pitch (ratio)	-0.385	0.196	-1.967	.049*
Y/N (diff)	-0.376	0.196	-1.914	.056
RG (diff)	-0.087	0.216	-0.403	.687
OM (diff)	0.356	0.230	1.544	.123
Pointing (diff)	0.442	0.200	2.214	.027*
Gaze (diff)	-0.341	0.243	-1.402	.161
Vocab	0.358	0.168	2.127	.034*

RG=Representational Gesture, OM=Object manipulation. Vocab=concurrent vocabulary. Y/N=Y/N questions. (diff) = the difference in a behaviour between labelling known/unknown toys.

Vocabulary Growth

We investigated whether the frequency/mean of caregiver labelling of unknown toys (with different multimodal behaviours) predicted children's vocabulary scores 1 year after the interaction (Table 2: total labels removed, VIF = 14.85). No IVs significantly predicted vocabulary growth, though concurrent vocabulary score was a significant positive predictor. Model selection identified a single best model including frequency of labels as part of yes/no questions (Table 2).

Table 2: Vocabulary growth predicted by frequency/mean of behaviours for the full (top) and reduced model (bottom)

	Est	SE	t	p
(Intercept)	0.026	0.156	0.166	.870
Pitch	0.078	0.166	0.469	.643
Y/N	-0.304	0.194	-1.568	.131
RG	0.190	0.218	0.868	.395
OM	-0.158	0.257	-0.617	.543
Pointing	-0.166	0.190	-0.873	.392
Gaze	-0.001	0.264	-0.005	.996
Vocab	0.436	0.186	2.349	.028*
ΔAICc=0.00				
(Intercept)	0.036	0.145	0.250	.805
Y/N	-0.412	0.142	-2.906	.007**
Vocab	0.394	0.147	2.681	.012*

Finally, we investigated whether differences in caregivers' multimodal behaviours between naming unknown and known toys predicted children's vocabulary growth (Table 3: total labels removed, VIF = 6.05). The difference between the pitch of caregiver labels significantly predicted vocabulary growth. Children's later vocabulary scores were higher when their caregivers used higher pitch for unknown labels compared to known (Figure 1).

Model selection found no single best model as two models had ΔAICc <2. The pattern was similar across all models: pitch and pointing difference predicted children's vocabulary growth. Children's scores were higher when their caregivers pointed more to known toys whilst labelling (Figure 1). The better fitting model is shown in Table 3. The other model additionally contained gaze difference (as a non-significant predictor ΔAICc = 1.60).

Discussion

Here, we assessed the relative contribution of different multimodal caregiver behaviours whilst producing object labels on children's language learning.

Our results show that verbal behaviours (e.g., prosody, utterance type) play an important role. In particular, we found that caregiver prosody was an important predictor, with higher pitched speech predicting immediate word labelling, and higher pitched speech when naming unknown compared to known toys predicting both immediate word learning and later receptive vocabulary (replicating Shi et al., 2022). This

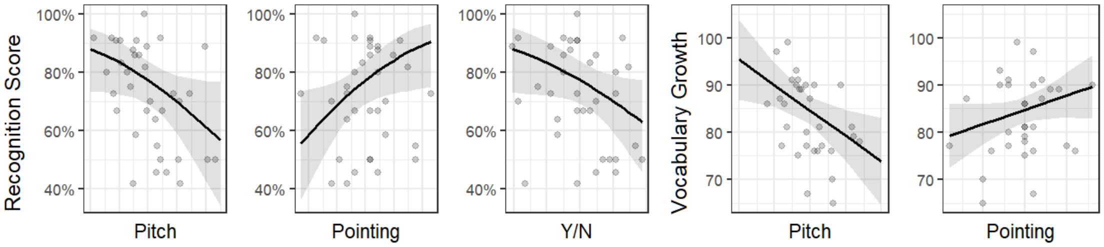

Figure 1: Model predictions (from full models) for recognition scores and vocabulary growth for difference scores relating to pitch, yes/no questions and pointing. The left side of the x-axis represents higher pitch/more pointing or Y/N questions for unknown toys whilst labelling while the right side represents higher pitch/more pointing or Y/N questions for known. Data is collapsed across items per dyad.

supports theories that suggest that CDL characteristics are a key component in children's word learning. As these predict both immediate and later vocabulary, these characteristics are clearly crucial to early language acquisition, perhaps by focusing attention on phonological features of new words. Additionally, we found that the semantic context of labelling was important, in that the degree to which caregivers used labels as part of a yes/no question for known compared to unknown toys predicted immediate word learning. The frequency of labelling as part of a yes/no question was a negative predictor of vocabulary growth, however. This contrasts Dong et al. (2021) who found a positive relationship with vocabulary growth for caregivers' use of yes/no questions, suggesting that the relationship between yes/no questions that specifically include labels impacts differently on vocabulary growth.

Non-verbal behaviours also played a role. However, this was limited to pointing. We found that children's learning of unknown labels was poorer and later vocabulary smaller when caregivers pointed more to unknown rather than known toys. Why might more pointing to unknown toys be negatively related to children's language learning? One possibility is that this relates to following-in behaviour (Carpenter et al., 1998). Research suggests that infants learn labels best when they are already attending to the object when hearing the label. For example, Tomasello & Farrar (1986) found that the amount that caregivers label objects that their 15-month-olds were already attending to predicted later expressive vocabulary. It is possible that pointing and naming unknown objects more indexes a relative lack of sensitivity to the child's attentional state in that moment, and involves redirecting attention to labelled referents instead of labelling objects already being attended to. To test this hypothesis, one would need to take the attentional focus of the child into account or focus only on labels produced in response to children's communicative bids (Olson & Masur, 2015).

Table 3: Vocabulary growth predicted by difference in multimodal behaviours between labelling known/unknown toys for the full (top) and reduced model (bottom)

	Est	SE	t	p
(Intercept)	0.062	0.147	0.423	0.677
Pitch (ratio)	-0.484	0.164	-2.955	0.007**
Y/N (diff)	-0.043	0.180	-0.237	0.815
RG (diff)	-0.230	0.200	-1.151	0.261
OM (diff)	-0.129	0.212	-0.612	0.547
Pointing (diff)	0.324	0.167	1.943	0.064
Gaze (diff)	0.422	0.235	1.798	0.085
Vocab	0.354	0.160	2.213	0.037*
ΔAICc=0.00				
(Intercept)	0.032	0.144	0.225	.824
Pitch (ratio)	-0.381	0.149	-2.558	.017*
Pointing (diff)	0.328	0.153	2.153	.040*
Vocab	0.392	0.152	2.587	.015*

Perhaps the most surprising finding was that most other multimodal behaviours produced in utterances containing a novel object label (e.g., representational gestures, object manipulation, and gaze) did not predict later language. This finding requires context, however, before we disregard the role of other multimodal behaviours. First, we did not investigate combinations of predictors (e.g., labelling whilst pointing + gaze; see Booth et al., 2007; Iverson et al., 1999), or indeed combinations of adult and child behaviours (e.g., mutual gaze). This presents a large analytical challenge as the number of potential predictors, and models under consideration is large compared to the sample size (see Bang et al., 2022; Donnellan et al., 2020). Nevertheless, an assessment of the different combinations of predictors may reveal different predictive relationships between caregiver behaviour and children's word learning. Second, representational gestures may play a role in learning especially when the referents talked about are visually absent (Motamedi et al., 2022). This possibility could not be tested here because of lack of power. Moreover, as representational gestures capture perceptual and action properties of objects (McNeill, 2005), they may support children's learning of conceptual knowledge about the object rather than the mapping between a conceptual representation and its label (e.g., Valenzo et al., 2003; McGregor et al., 2009). Finally, while most of the literature on the relationship between indexical cues and word learning focuses on infants under 2 years of age, the current study tested 3- to 4-year-olds, who are at a relatively advanced stage of linguistic development compared to infants. Thus, while our results suggest that caregivers' pitch modulations may be particularly beneficial for word learning at this developmental stage, it is entirely possible that other multimodal cues influence the lexical development of younger children.

References

Aussems, S. (2020). How seeing iconic gestures facilitates action event memory and verb learning in 3-year-old children. *Language Acquisition, 27*(1), 68–70. https://doi.org/10.1080/10489223.2019.1624759

Bang, J. Y., Bohn, M., Ramírez, J., Marchman, V. A., & Fernald, A. (2022). Spanish-speaking caregivers' use of referential labels with toddlers is a better predictor of later vocabulary than their use of referential gestures. *Developmental Science*. https://doi.org/10.1111/desc.13354

Bates, D., Bolker, B., Machler, M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berman, R. A., & Slobin, D., (1994). *Relating events in narrative: A crosslinguistic developmental study*. Lawrence Erlbaum Associates.

Boersma, P. & Weenink, D. (2023). Praat: doing phonetics by computer. Version 6.3.05. [Computer program]. http://www.praat.org

Bohn, M., & Frank, M. C. (2019). The Pervasive Role of Pragmatics in Early Language. *Annual Review of Developmental Psychology, 1*(1), 223–249. https://doi.org/10.1146/annurev-devpsych-121318-085037

Booth, A. E., McGregor, K. K., & Rohlfing, K. J. (2008). Socio-Pragmatics and Attention: Contributions to Gesturally Guided Word Learning in Toddlers. *Language Learning and Development, 4*(3), 179–202. https://doi.org/10.1080/15475440802143091

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language, 35*(1), 207–220. https://doi.org/10.1017/S030500090700829X

Burnham, K. P., & Anderson, D. R. (Eds.). (2004). *Model Selection and Multimodel Inference*. Springer New York. https://doi.org/10.1007/b97636

Capone, N. C., & McGregor, K. K. (2005). The Effect of Semantic Representation on Toddlers' Word Retrieval. *Journal of Speech, Language, and Hearing Research, 48*(6), 1468–1480. https://doi.org/10.1044/1092-4388(2005/102)

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development, 63*(4). https://doi.org/10.2307/1166214

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences, 110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

Cox, C., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G., & Fusaroli R. (2022). A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech. *Nature Human Behaviour, 7*, 114-113. https://doi.org/10.1038/s41562-022-01452-1

Cristia, A. (2013). Input to Language: The Phonetics and Perception of Infant-Directed Speech: The Phonetics and Perception of Infant-Directed Speech. *Language and Linguistics Compass, 7*(3), 157–170. https://doi.org/10.1111/lnc3.12015

Dong, S., Gu, Y., & Vigliocco, G. (2021). The impact of child-directed language on children's lexical development. *Proceedings of the Annual Meeting of the Cognitive Science Society, 43*, 1444–1450. Retrieved from https://escholarship.org/uc/item/38X9h9h4.

Donnellan, E., Bannard, C., McGillion, M. L., Slocombe, K. E., & Matthews, D. (2020). Infants' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language: A longitudinal investigation of infants' vocalizations, gestures and word production. *Developmental Science, 23*(1). https://doi.org/10.1111/desc.12843

Dunn, D. M., Dunn, L. M. (2009). National Foundation for Educational Research in England and Wales, & GL Assessment (Firm). The British picture vocabulary scale. GL Assessment.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development, 59*(5). https://doi.org/10.2307/1166093

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science, 16*(2), 234–248. https://doi.org/10.1111/desc.12019

Filippi, P., Gingras, B., & Fitch, W. T. (2014). Pitch enhancement facilitates word learning across visual contexts. *Frontiers in Psychology, 5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01468

Flynn, V., Masur, E. F., & Eichorst, D. L. (2004). Opportunity versus disposition as predictors of infants' and mothers' verbal and action imitation. *Infant Behavior and Development, 27*(3), 303–314. https://doi.org/10.1016/j.infbeh.2003.12.003

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology, 65*(1), 47–55. https://doi.org/10.1007/s00265-010-1038-5

Goodrich, W., & Hudson Kam, C. L. (2009). Co-speech gesture as input in verb learning. *Developmental Science, 12*(1), 81–87. https://doi.org/10.1111/j.1467-7687.2008.00735.x

Goodwyn, S. W., Acredolo, L. P., & Brown, C. A. (2000). Impact of Symbolic Gesturing on Early Language

Development. *Journal of Nonverbal Behavior, 24*(2), 81–103. https://doi.org/10.1023/A:1006653828895

Graf Estes, K., & Hurley, K. (2013). Infant-Directed Prosody Helps Infants Map Sounds to Meanings. *Infancy, 18*(5), 797–824. https://doi.org/10.1111/infa.12006

Grassmann, S., & Tomasello, M. (2007). Two-year-olds use primary sentence accent to learn new words*. *Journal of Child Language, 34*(3), 677–687. https://doi.org/10.1017/S0305000907008021

Grassmann, S., & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science, 13*(1), 252–263. https://doi.org/10.1111/j.1467-7687.2009.00871.x

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science, 11*(6), F31–F39. https://doi.org/10.1111/j.1467-7687.2008.00768.x

Iverson, J. M., Capirci, O., Longobardi, E., & Cristina Caselli, M. (1999). Gesturing in mother-child interactions. *Cognitive Development, 14*(1), 57–75. https://doi.org/10.1016/S0885-2014(99)80018-5

Kalagher, H., & Yu, C. (2006). The effects of deictic pointing in word learning. *Proceedings of the 5th International Conference of Development and Learning*, Bloomington, IN.

Kalashnikova, M., & Burnham, D. (2018). Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. *Journal of Child Language, 45*(5), 1035–1053. https://doi.org/10.1017/S0305000917000629

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13). https://doi.org/10.18637/jss.v082.i13

Law, B., Houston-Price, C., & Loucas, T. (2012) Using Gaze Direction to Learn Words at 18 Months: Relationships with Later Vocabulary. *Language Studies Working Papers 4*, 3-14.

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development, 7*(3), 185–201. https://doi.org/10.1080/15475441.2011. 579839

McGillion, M., Pine, J. M., Herbert, J. S., & Matthews, D. (2017). A randomised controlled trial to test the effect of promoting caregiver contingent talk on language development in infants from diverse socioeconomic status backgrounds. *Journal of Child Psychology and Psychiatry, 58*(10), 1122–1131. https://doi.org/10.1111/jcpp.12725

McGregor, K. K., Rohlfing, K. J., Bean, A., & Marschner, E. (2009). Gesture as a support for word learning: The case of under. *Journal of Child Language, 36*(4), 807–828. https://doi.org/10.1017/S0305000908009173

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review, 119*(4), 831–877. https://doi.org/10.1037/a0029872

McNeill, D. (2005). *Gesture and Thought.* Chicago, IL: University of Chicago press.

Motamedi, Y., Murgiano, M., Grzyb, B., Gu, Y., Kewenig, V., Brieke, R., Marshall, C., Wonnacott, E., Perniss, P., & Vigliocco, G. (2022). Language development beyond the here-and-now: Iconicity and displacement in child-directed communication [Preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/8rdmj

Nurmsoo, E., & Bloom, P. (2008). Preschoolers' Perspective Taking in Word Learning: Do They Blindly Follow Eye Gaze? *Psychological Science, 19*(3), 211–215. https://doi.org/10.1111/j.1467-9280.2008.02069.x

Olson, J., & Masur, E. F. (2015). Mothers' labeling responses to infants' gestures predict vocabulary outcomes. *Journal of Child Language, 42*(6), 1289–1311. https://doi.org/10.1017/S0305000914000828

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal Correlates of Growth in Toddler Vocabulary Production in Low-Income Families. *Child Development, 76*(4), 763–782. https://doi.org/10.1111/1467-8624.00498-i1

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651). https://doi.org/10.1098/rstb.2013.0300

Porritt, L. L., Zinser, M. C., Bachorowski, J.-A., & Kaplan, P. S. (2014). Depression diagnoses and fundamental frequency-based acoustic cues in maternal infant-directed speech. *Language Learning and Development, 10*(1), 51–67. https://doi.org/10.1080/15475441.2013.802962

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing. http://www.R-project.org/

Rader, N. de V., & Zukow-Goldring, P. (2010). How the hands control attention during early word learning. *Gesture, 10*(2–3), 202–221. https://doi.org/10.1075/gest.10.2-3.05rad

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. Jo*urnal of Child Language, 35*(1), 185–205. https://doi.org/10.1017/S0305000907008343

Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The Pace of Vocabulary Growth Helps Predict Later Vocabulary Skill: Pace of Vocabulary Growth. *Child Development, 83*(2), 508–525. https://doi.org/10.1111/j.1467-8624.2011.01710.x

RStudio Team (2021). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/.

Schroer, S. E., & Yu, C. (2022). Looking is not enough: Multimodal attention supports the real-time learning of new words. *Developmental Science*. https://doi.org/10.1111/desc.13290

Shi, J., Gu, Y., & Vigliocco, G. (2022). Prosodic modulations in child-directed language and their impact on word learning. *Developmental Science*. https://doi.org/10.1111/desc.13357

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category - ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation* (LREC 2008).

Slone, L. K., Abney, D. H., Smith, L. B., & Yu, C. (2023). The temporal structure of parent talk to toddlers about objects. *Cognition*, *230*. https://doi.org/10.1016/j.cognition.2022.105266

Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics - Theory and Methods*, *7*(1), 13–26. https://doi.org/10.1080/03610927808827599

Sweller, N., Shinooka-Phelan, A., & Austin, E. (2020). The effects of observing and producing gestures on Japanese word learning. *Acta Psychologica*, *207*, 103079. https://doi.org/10.1016/j.actpsy.2020.103079

Tomasello, M., & Farrar, M. J. (1986). Joint Attention and Early Language. *Child Development*, *57*(6), 1454. https://doi.org/10.2307/1130423

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, *9*(2), 335–340. https://doi.org/10.3758/BF03196290

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech Perception in Infancy Predicts Language Development in the Second Year of Life: A Longitudinal Study. *Child Development*, *75*(4), 1067–1084. https://doi.org/10.1111/j.1467-8624.2004.00726.x

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, *28*(2), 187–204. https://doi.org/10.1016/S0361-476X(02)00007-3

Vigliocco, G., Motamedi, Y., Murgiano, M., Wonnacott, E., Marshall, C., Milán-Maillo, I., & Perniss, P. (2019). Onomatopoeia, gestures, actions and words: *How do caregivers use multimodal cues in their communication to children?* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/v263k

Vogt, S., & Kauschke, C. (2017). Observing iconic gestures enhances word learning in typically developing children and children with specific language impairment*. *Journal of Child Language*, *44*(6), 1458–1484. https://doi.org/10.1017/S0305000916000647

Yow, W. Q., & Markman, E. M. (2015). A bilingual advantage in how children integrate multiple cues to understand a speaker's referential intent*. *Bilingualism: Language and Cognition*, *18*(3), 391–399. https://doi.org/10.1017/S1366728914000133