



Cognitive Science 48 (2024) e13469

© 2024 The Author(s). *Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society (CSS).

ISSN: 1551-6709 online

DOI: 10.1111/cogs.13469

Vision Verbs Emerge First in English Acquisition but Touch, not Audition, Follows Second

Lila San Roque,^{a,*} Elisabeth Norcliffe,^{b,*} Asifa Majid^b

^a*School of Culture, History, and Language, The Australian National University*

^b*Department of Experimental Psychology, University of Oxford*

Received 16 October 2023; received in revised form 28 March 2024; accepted 15 May 2024

Abstract

Words that describe sensory perception give insight into how language mediates human experience, and the acquisition of these words is one way to examine how we learn to categorize and communicate sensation. We examine the differential predictions of the typological prevalence hypothesis and embodiment hypothesis regarding the acquisition of perception verbs. Studies 1 and 2 examine the acquisition trajectories of perception verbs across 12 languages using parent questionnaire responses, while Study 3 examines their relative frequencies in English corpus data. We find the vision verbs *see* and *look* are acquired first, consistent with the typological prevalence hypothesis. However, for children at 12–23 months, touch—not audition—verbs take precedence in terms of their age of acquisition, frequency in child-produced speech, and frequency in child-directed speech, consistent with the embodiment hypothesis. Later at 24–35 months old, frequency rates are observably different and audition begins to align with what has previously been reported in adult English data. It seems the initial orientation to verbalizing touch over audition in child–caregiver interaction is especially related to the control of physically and socially appropriate behaviors. Taken together, the results indicate children’s acquisition of perception verbs arises from the complex interplay of embodiment, language-specific input, and child-directed socialization routines.

Keywords: Language acquisition; Perception verbs; Sensory language; Sensory dominance; Embodiment

*Lila San Roque and Elisabeth Norcliffe contributed equally to this work.

Correspondence should be sent to Asifa Majid, Department of Psychology, University of Oxford, Oxford OX2 6LE, UK. E-mail: asifa.majid@psy.ox.ac.uk

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Basic perception verbs—such as *look*, *hear*, and *smell* in English—label sensory experiences according to distinct modalities. As the senses provide the basis for our experience of the world, perception verbs have held a particular fascination for language scientists. On the one hand, they encapsulate the universality of our shared human physiology. On the other, cross-linguistic research has revealed substantial diversity across languages in how basic perceptual experiences are lexically represented and referred to (Norcliffe & Majid, 2024; Majid & Levinson, 2011; San Roque et al., 2015; Viberg, 1983).

One theory to emerge from the cross-linguistic study of perception verbs is that a biologically motivated hierarchy of senses constrains perceptual language (Evans & Wilkins, 2000; Viberg, 1983). Viberg (1983, 2001) proposed that despite variability across languages, vision verbs are universally linguistically privileged over other perception verbs, followed by hearing, and then the “lower” senses of touch, then taste, and smell. Verbs at the top of the hierarchy are argued to be more frequent, shorter, and more diachronically stable. While later typological studies of perception verbs have not found support for a universally fixed hierarchy of the five senses (Norcliffe & Majid, 2024; San Roque et al., 2015; see also Majid et al., 2018, for property words denoting sensory categories), research does suggest that vision and audition tend to dominate over the other senses in language. For example, *see* and *look* are the most frequent basic perception verbs in typical English conversation between adults, followed by *hear* and *listen* (Floyd, San Roque, & Majid, 2018; Winter, Perlman, & Majid, 2018), and a similar frequency ranking with respect to vision, audition, and other senses has also been found in a range of other languages (San Roque et al., 2015).

While perception verbs have been the focus of cross-linguistic typological study for many decades, our knowledge of how children acquire and use these key vocabulary items is nascent. A few studies have examined first language acquisition of perception verbs in relation to questions concerning such topics as semantics and polysemy, complementation structures, and pragmatic functions (e.g., Bloom, Rispoli, Gartner, & Hafitz, 1989; Edwards & Goodwin 1985; Johnson, 1999; Landau & Gleitman 1985; San Roque & Schieffelin 2019), and some scholars have suggested the acquisition of perception verbs will follow the sensory hierarchy outlined above (e.g., Bloom et al., 1989; see also Viberg 1994; Gentner & Bowerman 2009): children will learn sight and then hearing verbs before other perception verbs because humans are attuned to the salience of visual and auditory meaning, and supported in this through patterns of language use. This can be referred to as the typological prevalence hypothesis, following Gentner and Bowerman (2009).

However, it is also argued that children’s early embodied experience as a whole is crucial for understanding the order in which verbs emerge in child speech, and senses other than vision and audition may be critical earlier in development. This view is based on the idea that the physicality of the body and how it acts in the world is critical to cognition, and this embodied experience underlies conceptual representation and processing (Barsalou, 2008; Pulvermüller, 1999). Accordingly, the unfolding sensory-motor development of children may have implications for language learning too.

Consistent with this, Maouene, Hidaka, and Smith (2008) compiled a set of 101 early learned verbs in English (based on vocabulary questionnaire data) and asked adult English speakers to associate verbs with body parts (e.g., the verb *kiss* might call to mind the body part *mouth*). They hypothesized that early learned verbs would show consistent, structured associations to specific body parts, as well as a recognizable learning trajectory, reflecting infant reliance on embodied meaning. Indeed, they found that parents assessed children at younger ages as knowing more verbs relating to actions featuring the mouth (such as *kiss*), followed by verbs relating to the hands (e.g., *take*), then eyes (e.g., *cry*), while ear-related verbs such as *hear* trailed behind. They concluded this language acquisition pattern is “tantalizing in its similarity to traditional Piagetian (Piaget, 1953) descriptions of the developmental course of sensory-motor development as infants first explore relations in their world” (Maouene et al., 2008, p. 1212): that is, first exploring objects with the mouth and hands, and later moving to other ways of engaging with the world. Thus, within the domain of basic perception verbs specifically, an embodied approach to word meaning would posit touch and taste as frontrunners relative to other senses. Accordingly, the embodiment hypothesis might predict touch and taste verbs are acquired before sight and hearing.

In this paper, we examine these two possibilities—the typological prevalence hypothesis and the embodiment hypothesis. In Study 1 (Section 2), we used a set of parent-report vocabulary production checklists from the MacArthur–Bates Communicative Development Inventory (CDI; Fenson et al., 2007) where parents are asked, for a large set of individual words, whether their child understands and produces the word at a given age. CDIs are able to measure infant vocabulary acquisition on a large scale and are generally considered reliable indicators of infant word learning (Bates, Bretherton, & Snyder, 1988; Dale, 1991; Dale, Bates, Reznick, & Morisset, 1989; Fenson et al., 1994; Styles & Plunkett, 2009). Our analysis of data for English-speaking children indicates acquisition of vision verbs, followed by touch verbs, and then hearing and taste. In Study 2 (Section 3), we tested the cross-linguistic robustness of this ordering by examining available CDI data in a sample of 11 languages from four language families and found a similar pattern of perception verb acquisition.

While the age-of-acquisition findings are suggestive, the CDI data do not include all verbs of interest, and the method is open to response bias, since adult responses may reflect attitudes and expectations about the senses (e.g., vision is most important), rather than accurately reflecting child language use. We therefore complemented this approach for English by examining the rate of perception verb usage in child speech over time (Study 3; Section 4), using naturalistic corpus data from the North American collection of the Child Language Data Exchange System (Childes; MacWhinney, 2000), a publicly available dataset of interactions involving children. This enabled us to include the missing basic perception verbs *smell* and *feel* (not present in CDI checklists) and to verify whether the order of acquisition of perception verbs, as measured by parental reports, is in agreement with the relative frequency of perception verbs in actual child speech. We found that vision verbs were produced most frequently, but that touch verbs were used earlier and more frequently than other non-visual perception verbs in under 2-year-olds. At older age ranges, productions of touch verbs were no more frequent than hearing verbs. The child frequency patterns in the earliest age bracket

therefore mirror the acquisition ordering found both in English and across languages. In order to gain further understanding of acquisition in context, we also examined verb frequency patterns in child-directed speech in the same dataset and additionally conducted a focused qualitative analysis of perception verb usage in child-caregiver interactions.

2. Study 1: Age of acquisition of perception verbs in English

2.1. Data and methods

We accessed the American English CDI using Wordbank (Frank et al., 2016), a structured database which archives raw CDI data across languages and labs. The data were downloaded on December 20, 2022, using the R **wordbank** package (<http://wordbank.stanford.edu/>). In total, the dataset consisted of observations for 7,878 children. We did not exclude any children from the analyses; missing observations were treated as NA.

We estimated the age of acquisition of basic perception verbs using vocabulary data. Viberg (1983) identifies basic perception verbs in English: *see*, *look*, *hear*, *listen*, *feel*, *touch*, *taste* and *smell*.¹ The words used in our analyses are the six verbs included in the American English CDI (Words and Sentences) surveys: *see*, *look*, *hear*, *listen*, *touch*, and *taste*. The verbs *feel* and *smell* are not included in the set of CDI vocabulary and so could not be analyzed. Our analyses were restricted to production data only, for the age range 16–30 months.

2.2. Analyses

All analyses were conducted in R (R Core Team, 2022, version 4.2.1). We used the package “tidyverse” (version 1.3.2; Wickham et al., 2019) for data processing and visualization and “brms” (version 2.17.0; Bürkner, 2017, 2018, 2021) for the regression models. R code for all analyses is available at https://osf.io/zjtmw/?view_only=cefa4dd13a3e424fa08204e9e93ca70a.

We modeled the parental report data of children’s word production using Bayesian mixed effects logistic regression. Children’s acquisition of perception verbs (1 = produced, 0 = did not produce) was modeled as a function of the fixed effects of Age of child (a continuous predictor, partitioned into 1-month intervals), Modality of verb (Sight, Hearing, Touch, Taste), and the higher order interaction between the two. For the fixed effect of Modality, we collapsed over productions of *see* and *look* (such that a child’s production of either or both verbs counted as 1) to create a single-level Sight. We did the same for *hear* and *listen* to create a single-level Hearing. (See Supporting Information, S1, for a complementary analysis that only considers the set of unambiguously agentive perception verbs, i.e., contrasts *look* vs. *listen* vs. *touch*). The fixed effect of Modality was contrast coded, with Touch treated as the reference level. This allowed us to determine whether Sight and Hearing are acquired earlier than Touch, and, secondarily, to test whether there are differences in the timing of acquisition between Touch and Taste.

In the model, random intercepts were included to account for by-subject variation in verb production. The magnitude of the coefficient of each modality contrast gives an estimate of

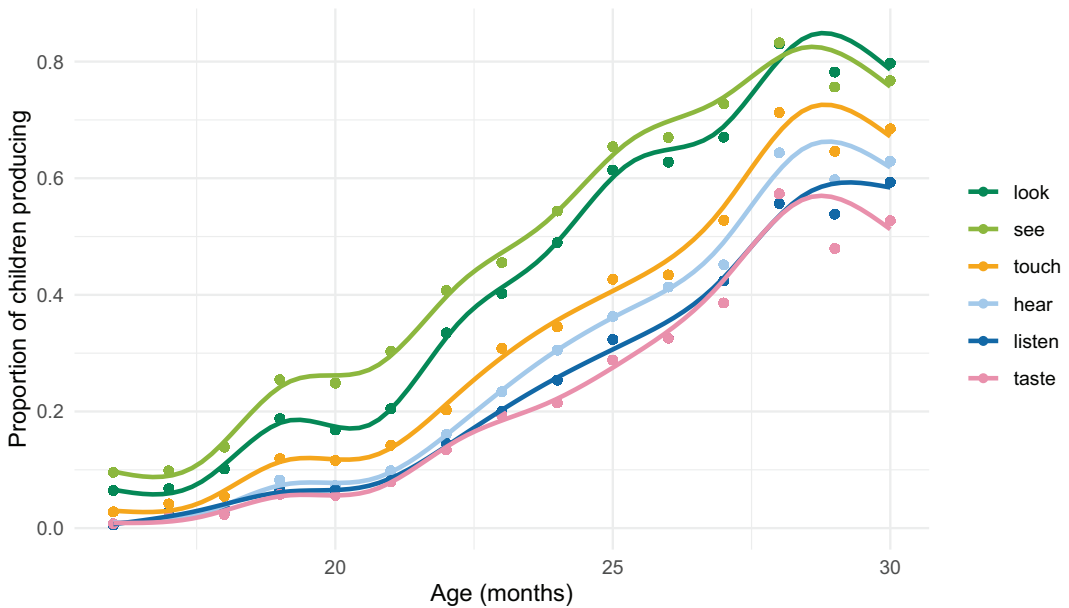


Fig. 1. Proportion of children producing each of the six basic perception verbs over time.

its independent contribution to words being produced by more children. Interactions between Modality and Age indicate how the effect of Modality is modulated for words learned earlier or later. For example, a positive effect of sight (vs. touch) means that verbs referring to sight are acquired by more children than verbs referring to touch; a negative interaction with Age means that this effect results in higher rates of acquisition for younger children.

We used default priors from the package “brms” for intercept and standard deviation and weakly informative priors on fixed slopes (normal distribution centered at 0 with a standard deviation of 1). As our inference criterion, we observed whether the 95% credible intervals of the posterior distributions for each predictor included zero. Credible intervals that did not contain zero were interpreted as providing strong evidence for the effect of each predictor on the dependent variable. We also report the probability of each effect being above zero.

Markov chain Monte Carlo (MCMC) sampling was performed with 4000 iterations each for four chains (2000 iterations discarded after warmup). Model convergence was indicated by visual inspection of the chains and Rhat values of 1. Posterior predictive checks showed a good fit to the data.

2.3. Results

Fig. 1 shows the item trajectories for the six perception verbs. The graph shows that, at every age interval, a higher proportion of children produce *see* compared to all other perception verbs. *Look* ranks second highest, followed by *touch*, and then *hear*, *listen*, and *taste*. To determine if these differences between modalities were reliable, we fitted a logistic regression model to the data.

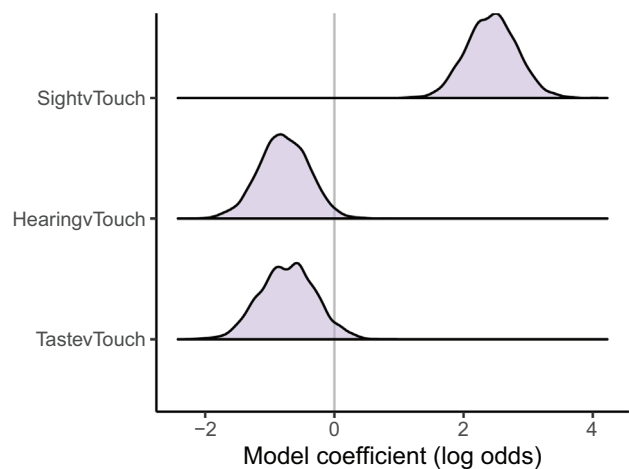


Fig. 2. Posterior distribution of logistic regression coefficients predicting whether English-speaking infants 16–30 months have acquired perception verbs (Touch is the reference level).

The posterior distributions of the Bayesian logistic regression model strongly support an overall effect of sensory modality (Sight, Hearing, Touch, Taste) on children’s perception verb acquisition (see Fig. 2). Children were more likely to acquire sight verbs compared to touch verbs (odds = 11.5 to 1, log odds = 2.44, $SE = 0.39$, 95% $CI = [1.70, 3.19]$, $\text{post} > 0 = 100$). Children were less likely to acquire hearing verbs compared to touch verbs (odds = 0.4 to 1, log odds = -0.78 , $SE = 0.38$, 95% $CI = [-1.55, -0.05]$, $\text{post} < 0 = 98.1$) and less likely to acquire taste verbs compared to touch verbs (odds = 0.5 to 1, log odds = -0.73 , 95% $CI = [-1.54, 0.11]$, $\text{post} < 0 = 95.6$). There was, unsurprisingly, also a positive main effect of Age: perception verb acquisition increased with age (odds = 2.1 to 1, log odds = 0.74, $SE = 0.02$, 95% $CI = [0.70, 0.78]$, $\text{post} > 0 = 100$). The model did not yield any noteworthy interactions between Age and Modality. For Sight (vs. Touch)*Age: odds = 1 to 1, log odds = 0.02, $SE = 0.02$, 95% $CI = [-0.01, 0.05]$, $\text{post} > 0 = 90.3$; for Hearing (vs. Touch)*Age: odds = 1 to 1, log odds = 0.03, $SE = 0.02$, 95% $CI = [0.00, 0.06]$, $\text{post} > 0 = 97.6$; for Taste (vs. Touch)*Age: odds = 0.97 to 1, log odds = -0.03 , $SE = 0.02$, 95% $CI = [-0.07, 0.00]$, $\text{post} < 0 = 97.6$.

2.4. Summary

Children were more likely to acquire touch verbs compared to hearing verbs, consistent with the embodiment hypothesis which holds that early child development privileges touch over hearing. However, they were also more likely to acquire vision verbs compared to touch verbs, consistent with the typological prevalence hypothesis which suggests vision dominates over the non-visual senses overall. Thus, with respect to order of acquisition, there is evidence that vision verbs are learned earliest, but touch precedes hearing and taste.

3. Study 2: Age of acquisition of perception verbs across languages

3.1. Data and methods

We next looked beyond English to ascertain whether the same pattern is found cross-linguistically in child language acquisition, again drawing from Wordbank. Other than English varieties, Wordbank currently contains vocabulary data from 29 languages, though these do not have equal vocabulary coverage. To identify languages with the target perception verbs, we used Wordbank's "unilemma" ("universal lemma") coding. Unilemmas are "cross-linguistic mappings from lexical items to single (English) forms that stand for a particular conceptual abstraction" (Frank et al., 2016). We began by selecting languages that had word forms for any of the unilemmas *see*, *look*, *hear*, *listen*, *touch*, and *taste* ($n = 25$). Of these, 11 languages were missing word forms that corresponded to either of the unilemmas *touch* or *taste* and one language was missing any verb for hearing (i.e., both "hear" and "listen" were absent). We excluded these languages, together with one language (Finnish) that was undersampled (only four children in total) and one language (Swedish) that was only sampled every 3 months. We also excluded one variety of French (France), because the language was already represented by Quebecois French (which had a greater number of observations). This left a final sample of 11 languages that had lexical coverage of the four sense modalities targeted by CDI, sight, hearing, touch and taste (see Table 1). Some of these languages do not have separate verb forms for agentive and non-agentive sight or hearing, at least with respect to the concepts sampled in Wordbank; and some have multiple forms listed (Table 1).

Sample size varied across languages. In total, the cross-linguistic dataset consisted of observations for 8,692 children: Croatian ($n = 377$), Danish ($n = 2863$), Dutch ($n = 208$), French (Quebecois) ($n = 887$), German ($n = 1,181$), Italian ($n = 639$), Kigiriama ($n = 88$), Kiswahili ($n = 89$), Korean ($n = 931$), Latvian ($n = 365$), Mandarin (Beijing) ($n = 1,056$). We did not exclude any children from the analyses; missing observations were treated as NA. As in Study 1, analyses were restricted to production data for the age range 16–30 months.

3.2. Analyses

We modeled the data using Bayesian mixed effects logistic regression, following the same model specifications as Study 1 but now including language and subject as nested random effects to capture by-language and by-subject variability (nested within language). MCMC sampling was performed with 4,000 iterations each for four chains (2,000 iterations discarded after warmup). Visual inspection of the chains and Rhat values of 1 indicated the model converged and posterior predictive checks showed the model fit the data well.

3.3. Results

The posterior distributions of the model strongly support an overall effect of sensory modality (Sight, Hearing, Touch, Taste) on children's perception verb acquisition across languages. As with the English-only dataset, children were more likely to acquire sight verbs compared to touch verbs (odds = 5.21 to 1, log odds = 1.65, $SE = 0.29$, 95% $CI = [1.09, 2.24]$, post

Table 1
List of the word forms corresponding to the target unilemmas in each language

Language Family	Language	Unilemmas					
		Sight		Hearing		Touch	Taste
		See	Look	Hear	Listen		
Indo-European	Croatian	<i>vidjeti</i>	<i>gledati</i>	<i>čuti</i>	<i>slušati</i>	<i>dirati</i>	<i>kušati</i>
	Danish	<i>se</i>	<i>kigge</i>	<i>høre</i>	<i>lytte</i>	<i>røre</i>	<i>smage</i>
	Dutch	<i>zien</i>	<i>kijken</i>	<i>hoor(t)!</i> <i>horen</i>	<i>luisteren</i>	<i>aanraken</i> <i>aankomen</i>	<i>proeven</i>
	French (Quebecois)	<i>voir</i>	<i>regarder</i>	<i>entendre</i>	<i>écouter</i>	<i>toucher</i>	<i>goûter</i>
	German	<i>sehen</i>		<i>hören</i>	<i>zuhören</i>	<i>anfassen</i>	<i>schmecken</i>
	Italian	<i>vedere</i>	<i>guardare</i>	<i>sentire</i>		<i>toccare</i>	<i>provare</i>
Atlantic-Congo	Latvian	<i>redzēt</i>	<i>skatīties</i>	<i>dzirdēt</i>	<i>klausities</i>	<i>skart pieskarties</i>	<i>garšot</i>
	Kigiriana	<i>ona</i>	<i>lola</i>	<i>sikira</i>	<i>sirikiza</i>	<i>guta</i>	<i>thatha</i>
	Kiswahili	<i>ona</i>	<i>tazama</i> <i>angalia</i>	<i>sikia</i>	<i>sikiliza</i>	<i>gusa</i>	<i>onja</i>
Koreanic	Korean	<i>boyeo</i> <i>boda</i>	<i>bwa</i>		<i>(sorireul) deureo</i>	<i>sondae</i>	<i>masbwa</i>
Sino-Tibetan	Mandarin (Beijing)		<i>kàn</i>	<i>tīng</i>	<i>tīnghuà</i>	<i>mō</i>	<i>cháng</i>
			<i>qiáo</i> <i>chǒu</i>	<i>tīngjiàn</i>			

> 0 = 100) and were less likely to acquire hearing verbs compared to touch verbs (odds = 0.52 to 1, log odds = -0.66, SE = 0.30, 95% CI = [-1.24, -0.07], post < 0 = 98.8). There was no reliable difference between the acquisition of taste and touch verbs (odds = 1.25 to 1, log odds = 0.22, SE = 0.31, 95% CI = [-0.37, 0.82, post > 0 = 76.2]. There was again a positive main effect of Age: perception verb acquisition increased with age (odds = 1.69 to 1, log odds = 0.52, SE = 0.01, 95% CI = [0.49, 0.54], post > 0 = 100). There were no noteworthy interactions between Age and Modality: Sight (vs. Touch)*Age: odds = 1.02 to 1, log odds = 0.02, SE = 0.01, 95% CI = [-0.01, 0.4]; Hearing (vs. Touch)*Age: odds = 1.05 to 1, log odds = 0.05, SE = 0.01, 95% CI = [0.03, 0.07]; Taste (vs. Touch)*Age: odds = .99 to 1, log odds = -0.01, SE = 0.01, 95% CI = [-0.03, 0.02] (Fig. 3).

3.4. Summary

The cross-linguistic analysis revealed a pattern of acquisition similar to that of English children from Study 1. Across the 11 languages in Study 2, children were more likely to acquire touch verbs compared to hearing verbs and were also more likely to acquire vision verbs compared to touch verbs. We did not find any reliable differences in the timing of acquisition of touch and taste, in contrast to the results from English-speaking children.

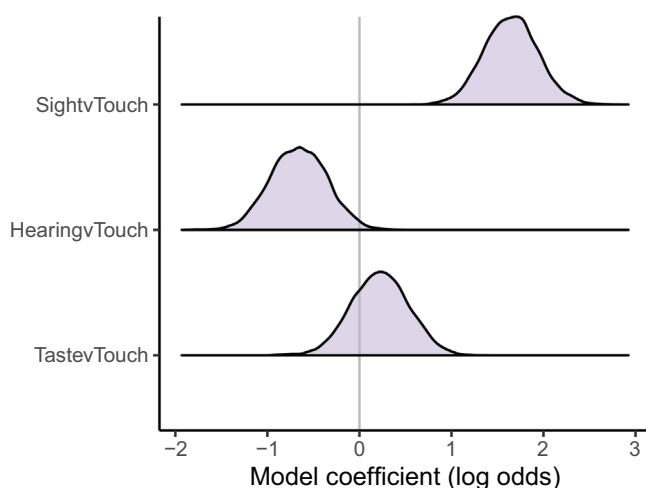


Fig. 3. Posterior distribution of logistic regression coefficients predicting whether infants across languages between 16 and 30 months have acquired perception verbs (Touch is the reference level).

4. Study 3: Perception verbs in child and child-directed speech

4.1. Data and methods

For frequency analyses of child and child-directed speech, we used corpus data from the North American collection of the Child Language Data Exchange System, CHILDES (MacWhinney, 2000). This collection includes interactional data from 48 subcorpora, recorded for a wide range of research purposes over a variety of times and contexts (e.g., conversations in the home, lab visits with toy-based play, narrative-focused recordings, etc.). The data were accessed using *chldes-db* (version 2021.1), a database-formatted mirror of CHILDES, using the R **chldr** package (Sanchez et al., 2018). The collection contained a total of 10,455,042 word tokens. We excluded tokens coded as unintelligible ($n = 241,100$), or for which information about the child's age was missing ($n = 1,023,125$), and that fell outside the age range of 1–5 years old ($n = 1,472,728$). This resulted in a final dataset of 7,718,089 tokens from 641 children and their parents.²

Each token was tagged for speaker role, “child” or “parent” (collapsed over mother and father), age of the child when the word was produced, and whether the word was a perception verb or not. Perception verb tokens were coded for their lemma, which consisted of the following categories: *see*, *look*, *hear*, *listen*, *touch*, *feel*, *taste*, and *smell*. Any inflected form counted as an instance of the lemma type (e.g., *see*, *seen*, *saw*, *seeing*, *sees* are all treated as instances of *see*).³ We were also able to include two perception verbs that were not in the CDI, *feel* and *smell*. To complement the quantitative analyses, we include a few illustrative examples of conversation between children and caregivers.

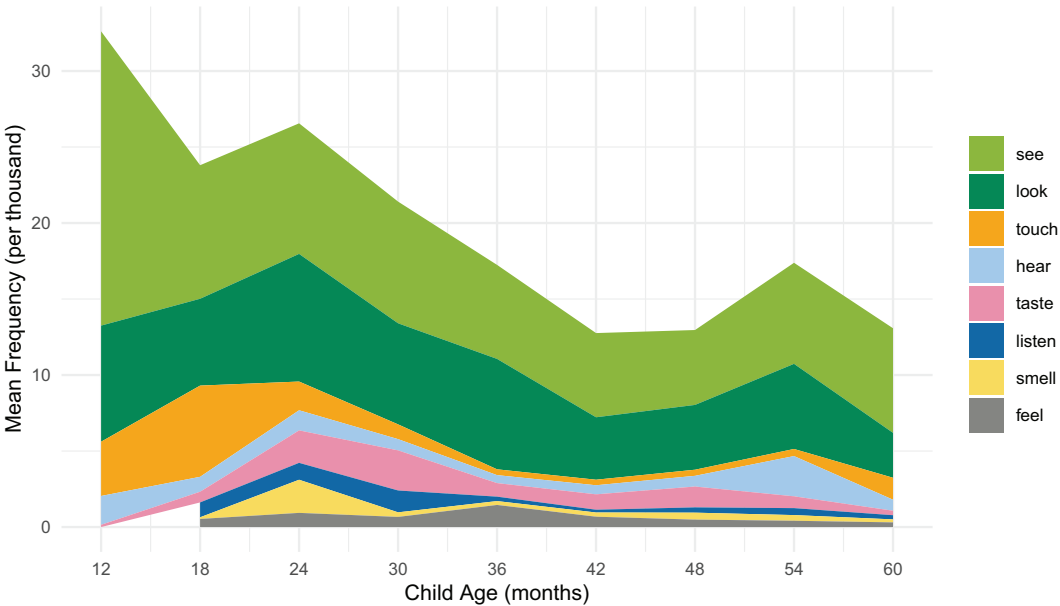


Fig. 4. Mean normalized perception verb frequencies in child speech as a function of age (category values are stacked).

4.2. Results

4.2.1. Child speech

For an initial overview of the trajectory of perception verb production in child speech, we counted the number of times an instance of a perception verb lemma was produced by each child, with counts grouped into 6-month age bins. These counts were then divided by the total words produced by each child within each age bin and divided by 1,000 to yield normalized frequencies (Fig. 4). Overall, children’s production of *see* and *look* is more frequent than the non-visual perception verbs for all ages. Not all perception verbs are produced in the earliest age bin (12–18 months): *listen*, *smell*, and *feel* only appear at 18 months.

Visual inspection of Fig. 4 suggests in the first 24 months, *touch* is the most frequently produced of the non-visual verbs; after 24 months *touch* productions decrease in frequency. This is clearer in Fig. 5, which zooms in on the mean normalized frequencies of non-visual perception verbs. Within the earliest age band (12–23 months), children produced *touch* tokens over five times per thousand words on average, whereas *hear* was produced around one time per thousand words (Fig. 5, top left). This is also evident when collapsed across agentive and non-agentive uses of the verb (Fig. 5, top right). This pattern shifts at 24–35 months, however. The difference between *touch* and the other non-visual senses is reduced—all non-visual perception verbs are now produced around two times per thousand words or fewer (Fig. 5, bottom left and right).

We fit two separate negative binomial regression models to the frequency data, one for each age band. We specified a negative binomial distribution because the dependent variable

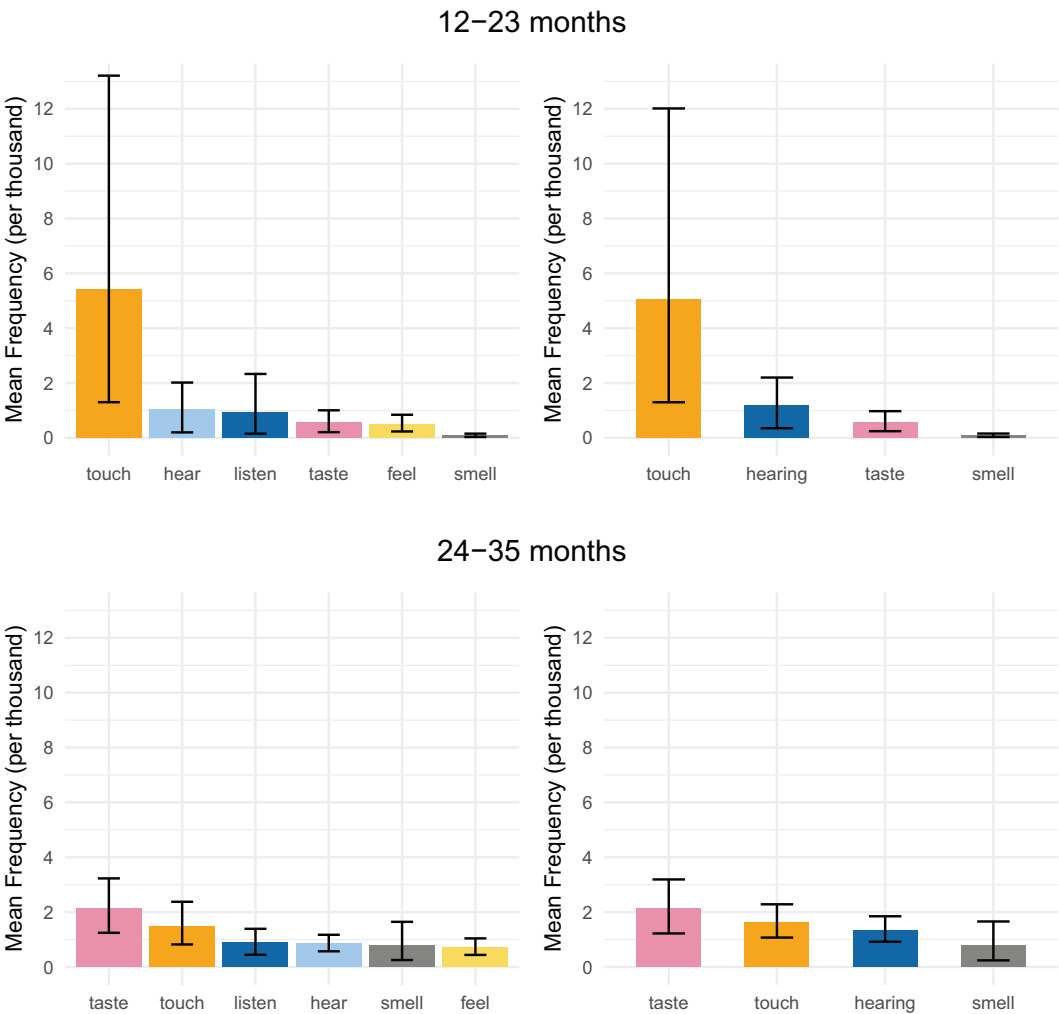


Fig. 5. Mean normalized frequencies of non-visual perception verbs by lemma (left) and collapsed across sensory modality (right) in child speech between the ages of 12–23 months (top) and 24–35 months (bottom). Error bars indicate bootstrapped 95% confidence intervals across words in each category.

was count data (Winter & Buerkner, 2021). The models included an offset term to allow us to model perception verb counts over units of exposure (log of the child’s total word count within the age range). This controls for the number of total words and gives a rate at which the target verbs are produced per word in the children’s data over the specified age range. Children’s rate of perception verb usage was modeled as a function of Modality, collapsed over lemmas, to create a single level for each sensory modality and contrast coded with Touch as the reference level (Supporting Information, S2, considers agentive perception verb lemmas only). In the model, random intercepts were included to account for by-speaker variation in usage rates of perception verbs.

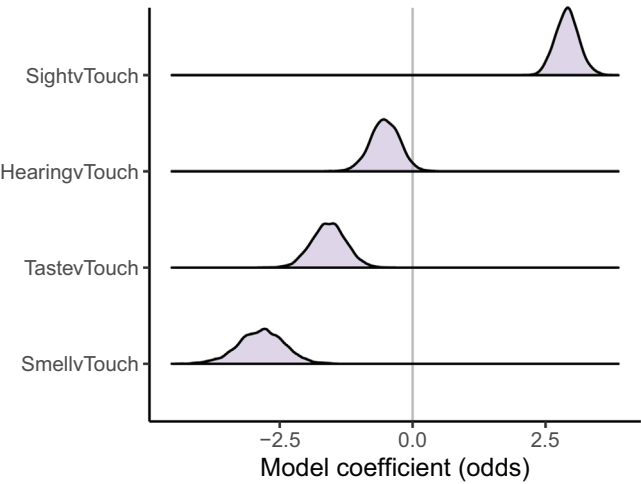


Fig. 6. Posterior distribution of negative binomial regression coefficients predicting differences in the rates of perception verb usage in child speech between 12 and 23 months (Touch is the reference level).

Children at 12–23 months produced reliably more sight verbs compared to touch verbs (log estimate 2.90, $SE = 0.22$, 95% $CI = [2.48, 3.33]$, $\text{post} > 0 = 100$), but fewer hearing verbs compared to touch verbs (log estimate -0.53 , $SE = 0.26$, 95% $CI = [-1.03, -0.02]$, $\text{post} < 0 = 98$). They also produced reliably fewer taste (log estimate -1.58 , $SE = 0.31$, $CI = [-2.20, -0.98]$, $\text{post} < 0 = 100$) and smell verbs compared to touch verbs (log estimate -2.84 , $SE = 0.42$, $CI = [-3.68, -2.01]$, $\text{post} < 0 = 100$); see Fig. 6.

Within the 24–35 month age range, children continued to produce reliably more sight verbs compared to touch verbs (log estimate 3.20, $SE = 0.11$, 95% $CI = [2.99, 3.42]$, $\text{post} > 0 = 100$). However, there was no reliable difference between hearing and touch verbs any longer (log estimate -0.01 , $SE = 0.13$, 95% $CI = [-0.27, 0.24]$). Children did produce fewer taste (log estimate -0.67 , $SE = 0.15$, $CI = [-0.96, -0.38]$, $\text{post} < 0 = 100$) and smell verbs compared to touch verbs (log estimate -1.57 , $SE = 0.16$, $CI = [-1.88, -1.26]$, $\text{post} < 0 = 100$); Fig. 7.

To summarize, vision verbs were produced most frequently by children of all ages, while smell and taste verbs were produced less frequently across the board, consistent with the typological prevalence hypothesis. Critically, however, in the earliest productions of children at 12–23 months, reference to touch appeared more frequent than hearing; by 24–35 months this difference disappeared and touch and hearing verbs were produced on par. This could be taken as tentative support for the embodiment hypothesis since touch appears before hearing in children’s early language production.

The discrepancy between the child frequency data and adult frequency data reported elsewhere (e.g., Floyd et al., 2018; San Roque et al., 2015; Winter et al., 2018) raises the question of whether English child-directed speech recapitulates perception verb production as has been reported in adult speech. Alternatively, it is possible caregivers speak differently in the

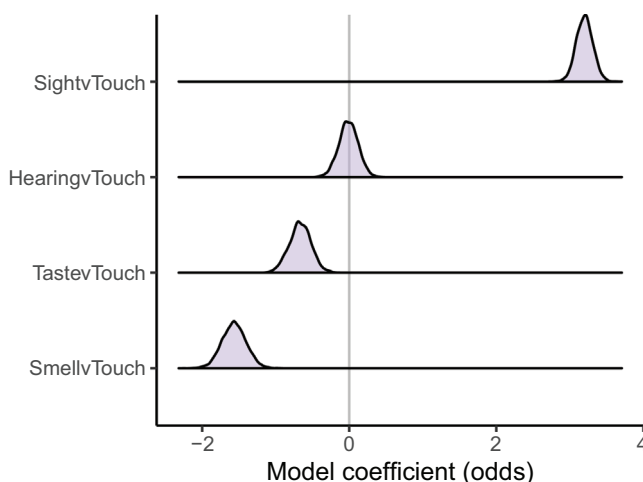


Fig. 7. Posterior distribution of negative binomial regression coefficients predicting differences in the rates of usage of perception verbs between 24 and 35 months (Touch is the reference level).

context of child interaction and refer to touch more frequently than hearing. To disentangle these possibilities, we examined child-directed speech within the same corpus.

4.2.2. Child-directed speech

Following the same procedure as for child speakers (Section 4.2.1), we counted the number of times a parent produced an instance of each of the perception verb lemmas, with counts grouped into 6-month child age bins. These counts were converted to normalized frequencies by dividing the total words produced by each child's parents within each age bin.

Parents' production of *see* and *look* far outstripped non-visual perception verbs at all ages, paralleling the child production data. Focusing on the non-visual perception verbs, at 12–23 months parents produced *touch* most frequently, with *touch* tokens appearing just under twice per thousand words on average. While this mirrors children's production, the actual rate of *touch* was higher in the child data (5 times per thousand words). *Taste* had the next highest mean frequency in the parent speech. Notably, *listen* and *hear* were less frequent—children heard *listen* at a quarter of the rate of *touch* on average, and when collapsed across agentive and non-agentive uses, children heard *listen/hear* verbs at around half the rate of *touch/feel* verbs. By 24–35 months the situation was different. Parents produced hearing verbs most frequently after vision. Notably, parents' production of *touch* dropped to half that of the earliest age range.

Applying negative binomial regression to the child-directed speech following the same model specifications as the models of child speech (Section 4.2.1; see Supporting Information, S3, for analysis of agentive perception verbs) revealed the posterior distributions strongly support an overall effect of sensory modality (Sight, Hearing, Touch, Taste, Smell) on rates of perception verb usage talking to children aged 12–23 months. Parents produced reliably more sight verbs compared to touch verbs (log estimate 2.67, $SE = 0.09$, 95% $CI = [2.50, 2.86]$, $post > 0 = 100$). On the other hand, caregivers produced reliably fewer hearing verbs

compared to touch verbs (log estimate -0.79 , $SE = 0.11$, 95% $CI = [-0.99, -0.58]$, $post < 0 = 100$), fewer taste verbs (log estimate -2.08 , $SE = 0.13$, $CI = [-2.35, -1.82]$, $post < 0 = 100$), and fewer smell verbs (log estimate -1.76 , $SE = 0.13$, $CI = [-2.00, -1.51]$, $post < 0 = 100$).

With children aged 24–35 months, parents again produced reliably more sight verbs than touch verbs (log estimate 3.38 , $SE = 0.07$, 95% $CI = [3.23, 3.91]$, $post > 0 = 100$). In contrast to the younger age range, however, there was no reliable difference in the rates of hearing verbs and touch verbs (log estimate 0.07 , $SE = 0.08$, 95% $CI = [-0.10, 0.23]$, $post > 0 = 80.2$). Still, caregivers produced reliably fewer taste verbs (log estimate -0.98 , $SE = 0.10$, $CI = [-1.18, -0.78]$, $post < 0 = 100$) and fewer smell verbs compared to touch verbs (log estimate -1.80 , $SE = 0.12$, $CI = [-2.03, -1.57]$, $post < 0 = 100$).

4.3. Summary

Children's production data showed tentative support for the embodiment hypothesis with young children aged 12–23 months referring to touch more than hearing; this difference disappeared later at 24–35 months. On the other hand, consistent with the typological perspective, vision dominated, while taste and smell appeared infrequently in the child data. The caregiver data were enlightening in this context. Unlike previous studies of English adult-to-adult speech (e.g., Floyd et al., 2018; San Roque et al., 2015; Winter et al., 2018), adults produced more touch verbs than hearing verbs when speaking to children aged between 12 and 23 months, mirroring (or leading) child production patterns as well as age of acquisition data. The dominance of touch relative to hearing in the child-directed speech shifted over time: in the 24–35 month age range, the difference between hearing and touch evened out. This appears to be driven both by a reduction in the frequency of touch as well as an increase in the frequency of hearing. To understand this shift better, we qualitatively examined instances of spontaneous conversation; the primary context of first language acquisition.

4.4. Use of perception verbs in context

To understand the differential use of touch and hearing verbs in child–caregiver interactions, we examined contextualized examples of the active perception verbs from the Providence Corpus within CHILDES (Demuth, Culbertson, & Alter, 2006). This subcorpus features six children recorded monthly or bimonthly in naturalistic interactions (usually indoors at home, engaging in a range of activities such as toy-play and make believe, reading texts and examining picture books, food preparation, eating and drinking, looking at photographs, chatting, having altercations, singing or chanting rhymes, among other activities) from around 12 months of age until into their third or fourth years (Demuth et al., 2006). These examples illustrate that perception verb usage is quite varied, suggesting a more nuanced interpretation of the child language data is required.

We have already seen that vision verbs were overwhelmingly the most frequent in both child and child-directed speech. Examples of their usage in context help to illustrate their multiple functions. As we might imagine, adults and children in the recordings commonly use *look* in imperative or jussive sentences to direct each other's attention in the shared environment (Examples 1–3) (see also Edwards & Goodwin, 1985).

Example 1

Mother: *Okay. Daddy's gonna wave. Ready? **Look** out here. He's gonna wave.* [Violet 010907]⁴

Example 2

Mother: *Alex. Put your blocks in the truck, **look. Look.** Fill it up.* [Alex 010427]

Example 3

Alex: ***Look** at that.*

Mother: *Okay which one?*

Alex: *This. That.* [Alex 011103]

As well as commands or encouragements to *look* (likely a dominant function of this verb with young children), a range of other functions are represented, such as talking about somebody else's visual engagement, requesting that one be allowed to look, and discussing how something or someone looks to the speaker or addressee. In their conversations, adults and children appear to orient to looking as a desirable and prosocial behavior, for example, in the sense that looking facilitates joint attention, shared appraisals, and coordinated projects (such as filling a truck with blocks, admiring a picture, etc.).

Turning to audition, for both parents and children, *listening* is often verbalized as a pleasurable and interesting activity, for example, in relation to listening to music and stories. In some of the recordings, caregivers and children also treat listening as a sense to be actively discussed and explored. In Example 4, Ethan's mother even explains what listening "is" (here, in relation to the recording/playback device).

Example 4

Mother: *Yeah, that's **listening**. When you put it into your ear, it's **listening**.*

Ethan: ***Listening**.*

Mother: ***Listening**. But you talk into the microphone.* [Ethan 010918]

Both parents and children also use *listen* in relation to listening to human speech; the dominant context for audition verbs in adult–adult conversation more generally (San Roque et al., 2015). One type of this usage is "discourse-oriented" in that *listen* directs the addressee to attend to upcoming talk (Examples 5, 6).

Example 5

Mother: ***Listen** did you tell Manuela where we're gonna go later?*

Lily: *The zoo.* [Lily 020108]

Example 6

Naima: *Listen Mommy.*

Mother: *What?*

Naima: *Nursie Daddy.*

Mother: *Nursie Daddy, that's another joke!* [Naima 010604]

For parents in particular, the recordings show a further emphasis on listening to speech as being “good” and behaving well (cf. a common extension of hearing verbs is to mean “obey”; Sweetser, 1990). Listening can be suggested as a behavior that facilitates the resolution of distressing situations, and *not* listening is identified as “naughty” or non-cooperative (e.g., when Alex, aged 1 year and 9 months, is told *you're being naughty, you're not listening*). Functions of *listen* thus suggest that the growing frequency of audition verbs in adult speech may result from an increased focus on verbal instruction and interaction as children become more competent and metalinguistically aware conversational partners.

In contrast to the plentiful verbal encouragement and positive parent attitude toward looking and listening, acts of touch are overwhelmingly verbalized by adults in contexts of prohibition. Directives of one kind or another *against* touching in fact account for around 200 of the roughly 330 parent-produced examples of *touch* in the Providence corpus. Especially for younger children, these instructions are often simply constructed using negatives like *don't* or *no*, including expressions that are not standard in adult speech (e.g., *no touch*), and often involve repetition (Example 7).

Example 7

Parent: *It's a camera sweetie. No no no no touch. No touch.*
You can look at it later, okay? [Ethan 001104]

Simple negative commands are interspersed with more complex prohibitions (e.g., *I'd rather you didn't touch...*; *didn't I say not to touch...*) as well as those that give reasons or specify consequences. Caregivers typically only invite or instruct children to touch within specific contexts, such as action rhymes (e.g., touching one's toes) or with toys and books that have been designed with tactile engagement and textural focus in mind.

Children do, at times, themselves embrace the use of touch prohibitions, such as *Don't touch it* (e.g., as spoken by Ethan at 1 year and 8 months). In Example 8, a nearly 3-year-old Naima highlights the right to touch her blocks as a privilege that her toy bear, Sleepy (voiced by Naima's mother), is not allowed to share.

Example 8

Child: *These blocks are, are part of the sofa, Sleepy.*
 Parent [as Sleepy]: *Oh can I jump on them?*
 Child: *No.*
 Parent: *Oh okay. Sleepy you better come here.*
 Child: *Only I get to **touch** them.* [Naima 021123]

Overall, children's utterances seem to be less dominated by prohibitions than those of their parents. Only around one fifth of the 57 *touch* tokens produced by children were clear directives against touching, while the rest (barring five unclear instances) exemplified a variety of contexts such as narrating; requesting to touch—potentially the “flipside” of a prohibition; or

just contemplating touch, for example, as a possible form of engagement (Example 9). Here, Lily and her mother are discussing snail slime.

Example 9

Parent: ...it's a little yucky but it's neat too.

Child: Yeah.

Parent: Yeah.

Child: You could **touch** some.

Parent: Well you could **touch** some that's true I guess. [Lily 020728]

In a recording made a few months later, looking together at pictures, Lily (at 2 years, 11 months, 25 days) responds to pictures of an urchin, lobster, crab, and dolphin by immediately identifying whether or not these are things that she likes to touch, suggesting this is a salient property for her and that, at least for a brief period, she may have routinized touch as a “go-to” sense in certain contexts. Lily also identifies a characteristic of touch that is rarely spontaneously verbalized by adults in the recordings: its importance to comfort, as where she comments on needing to *touch blankie* after hurting herself (age 2 years, 11 months, 6 days). Touch prohibitions themselves may also be the topic of extended discussion and explanations. For example, in one conversation Lily asks why one is not allowed to touch a fragile sculpture and her mother explains, *we can only look at it but not touch it* [Lily 030121].

Summing up, the ways children and adults use the verbs *look* and *listen* suggest an overall positive and prosocial attitude toward these activities. In addition, *listen* is used in relation to attending to speech, showing cooperative behavior, and participating in a conversation. At least for parents, however, *touch* is strongly identified as an activity that must be carefully regulated and is fraught with risk. It may be encouraged in controlled contexts but is most commonly verbalized in prohibitions such as *don't touch*. Children also use touch prohibitions and orient toward touch as an acceptable or unacceptable behavior, but in this corpus their contextual range for *touch* appears potentially broader and more exploratory than that of their caregiving adults.

5. Discussion

At the outset, we asked if young children's perception verb acquisition is swayed by early experience of touch and taste, as suggested by the embodiment hypothesis, or more closely aligns with the typological prevalence hypothesis that has identified vision and hearing as dominant cross-linguistically. We used parent questionnaire data to analyze perception verb learning in English and other languages, and naturalistic corpus data to further examine frequencies and functions of English perception verbs in children's and child-directed speech.

In Study 1, we found that adult assessments of children's production vocabulary in English put *see* and *look* as clear front-runners in the acquisition of basic perception verbs, with *touch* next, well ahead of *hear* and *taste* (note, *feel* and *smell* were not in this dataset). Study 2 established a similar pattern in 11 other languages: assessments of children's production

vocabulary indicated vision verbs preceded all other sensory verbs, and touch verbs preceded audition. However, no reliable difference was found between the relative ordering of touch and taste terms. Study 3 focused on perception verbs in child speech and child-directed speech in naturalistic conversations. The usage patterns painted a similar picture: at the earliest ages, 12–23 months, vision verbs were dominant in both child and child-directed speech, and touch verbs were more frequent than hearing, taste, and smell. By 24 months, the frequency of touch verbs diminished in both children's speech and child-directed speech.

The dominance of vision observed across all three studies echoes a robust finding in the literature: across languages, vision verbs are linguistically privileged over the other senses. This has been shown with respect to their usage frequencies (Floyd et al., 2018; San Roque et al., 2015; Winter et al., 2018), as well as other dimensions of linguistic expression, for example, their relative morphological simplicity and their tendency to be lexically differentiated from other senses (Norcliffe & Majid, 2024, Viberg, 1983). For some, vision sits at the top of a biologically motivated hierarchy of senses, with audition accorded second place, followed by touch, taste, and smell (Viberg, 1983). While later work does not find support for a fixed hierarchy of senses across languages (Majid et al, 2018; Norcliffe & Majid, 2024; San Roque et al., 2015), the available evidence does suggest that verbs of hearing tend to dominate over non-visual perception verbs in usage in English and typically (although not universally) across languages (Floyd et al., 2018; San Roque et al., 2015). Strikingly, our study reveals the dominance of audition over other non-visual senses in adult usage frequencies is not mirrored in the relative order of acquisition of perception verbs or in speech by or to children: in English (Study 1) and across other languages (Study 2), we found that touch verbs are learned earlier than hearing verbs. Touch verbs were also produced more than hearing verbs by children and by parents speaking to children in the first 24 months (Study 3).

This robust touch-before-hearing finding thus potentially supports Maouene and colleagues' (2008) identification of tactile experience as highly relevant to early acquired verbs (cf. also Freeborn, 2022), consistent with embodied theories of word meaning (Pulvermüller, 1999; Barsalou, 2008). However, the same perspective would also predict the comparatively early learning of taste verbs, due to the high salience of “mouthy” experiences in early development. The English data do not support this, although the cross-linguistic questionnaire data (Study 2) do keep open the possibility that taste verbs may, like touch verbs, be acquired early in relation to audition verbs in some languages. This possibility requires further in-depth study.


Moreover, a deeper fact about the use of *touch* in child-directed speech suggests an alternative motivation for the use of touch verbs. Adults' increased use of the word *touch* appears likely to be reflective of its importance in behavior regulation, especially in prohibitions against touching, rather than a focus on tactile sensation per se. Within our society, 1-year-olds are likely to be moving around independently in the home (e.g., as opposed to being carried on the body for hours) but are viewed as vulnerable, unpredictable, and in frequent need of explicit and repeated direction concerning their physical behavior. Looking is safe, but touching is dangerous. There is less evidence that the comparatively high frequency of touch verbs in children's speech (as opposed to child-directed speech) also stems primarily from prohibitions. This disparity across adults and children makes sense from our expectation

of social roles (e.g., adults monitor their children's safety and tell them what to do or not do) but leaves open for further exploration what motivates children to use a verb like *touch* themselves. Parent concerns about tactile contact as unsafe or as a social infraction may contribute to early learning of the word *touch* (via frequency effects) but do not necessarily determine how children apply it. Further cross-cultural research on child and child-directed speech patterns could help to disentangle motivating factors.

As English-speaking children move into their third year, they and their caregivers reduce the relative frequency of touch terms. For this age range, adults simultaneously increase their use of audition terms in child-directed speech, and thus start to align more closely with patterns of adult discourse. Reference to audition typically relates to the perception of speech across languages (cf. Buck, 1949; Evans & Wilkins 2000; San Roque et al. 2015; Sweetser, 1990). Children and caregivers also exhibit these speech-oriented uses of *listen*, and adults in particular tend to present *listening* to others as cooperative and prosocial behavior. It seems reasonable, then, that the observed shift in child-directed speech comes about through adults' declining focus on prohibiting physical touch and increasing emphasis on promoting attention to aural stimuli, specifically, spoken language. This possibility could be explored by more fine-grained examination of perceptual language in context.

To conclude, numerous studies of adult language use across cultures consistently find that *see* and *look* are the most frequent basic perception verbs followed by *hear* and *listen*, with touch, taste, and smell verbs in contrast appearing relatively infrequently (Floyd et al., 2018; San Roque et al., 2015; Winter et al., 2018). Accordingly, the typological prevalence hypothesis would predict that children would acquire and produce these verbs in the same order. Analysis of English CDI (Study 1) and corpus data (Study 3), as well as CDI data from eleven additional languages (Study 2), shows important discrepancies from the typological prevalence hypothesis. Although vision verbs do appear first and more frequently in young children's language production, touch appears before hearing. This is aligned with embodied accounts of meaning which predict language use is related to specific bodily experiences.

Open Materials Badge

 This article has earned Open Materials badges. Materials is available at <https://osf.io/mdtb4/>.

Notes

- 1 Viberg (1983) also includes *sound* as a basic audition verb (e.g., *that sounds scratchy*), the only verb in the group that is purely “copulative,” that is, the subject is almost invariably the percept itself. We exclude this from our study and focus only on those that can commonly occur as “activity” or “experiencer” verbs (i.e., the subject of the verb is potentially the perceiver).
- 2 We treated all instances produced by parents in the corpus as “child-directed” in relation to the target child, who is usually present throughout the entire recording session. For

- some sessions, tokens may be directed to others who are present (e.g., a sibling, other adult family members, researcher) as well as or instead of the target child.
- 3 We did not distinguish between verbal and nominal uses of perception terms (e.g., I *smell* the orange vs. the orange has a nice *smell*). This difference cannot be consistently assessed in all cases (e.g., in a child's single word utterance there may be no grammatical evidence to favor one interpretation). The feature of nominal-verbal polysemy is present across all five modalities, and we did not have an a priori expectation of how it would impact the results. Similarly, we did not distinguish between different senses of perception verbs (e.g. *see* "understand" as opposed to *see* "visually apprehend") because the analysis is based solely on lexical form. Fine-grained and systematic semantic analyses would be valuable to gain additional insight into the acquisition of individual verbs (see, e.g., Johnson, 1999) as well as the acquisition of polysemous forms more generally.
 - 4 The child's name and their age in years, months, and days is shown in square brackets as the example source.

References

- Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, 59(1): 617–645. <https://doi.org/10.1146/annurev.psych.59.103006.093639>
- Bates, E., Bretherton, I., & Snyder, L.S. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge, England: Cambridge University Press.
- Bloom, L., Rispoli, M., Gartner, B., & Hafitz, J. (1989). Acquisition of complementation. *Journal of Child Language*, 16, 101–120.
- Buck, C. D. (1949). *A dictionary of selected synonyms in the principal Indo-European languages: A contribution to the history of ideas*. Chicago, IL: University of Chicago Press.
- Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1), 395. <https://doi.org/10.32614/RJ-2018-017>
- Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- Dale, P. S. (1991). The validity of a parent report measure of vocabulary and Syntax at 24 months. *Journal of Speech, Language and Hearing Research*, 34, 565–571. <https://doi.org/10.1044/jshr.3403.565>
- Dale, P. S., Bates, E., Reznick, J. S., & Morisset, C. (1989). The validity of a parent report instrument of child language at twenty months. *Journal of Child Language*, 16, 239–249. <https://doi.org/10.1017/S0305000900010394>
- Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis, and coda licensing in the acquisition of English. *Language & Speech*, 49, 137–174. <https://doi.org/10.1177/00238309060490020201>
- Edwards, D., & Goodwin, R. (1985). The language of shared attention and visual experience. A functional study of early nomination. *Journal of Pragmatics*, 9, 475–493. [https://doi.org/10.1016/0378-2166\(85\)90017-7](https://doi.org/10.1016/0378-2166(85)90017-7)
- Evans, N., & Wilkins, D. (2000). In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language*, 76(3), 546–592. <https://doi.org/10.2307/417135>
- Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., Pethick, S., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5), v–173.

- Fenson, L., Marchman, V. A., Thal, D., Dale, P., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates communicative development inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Brookes Publishing Company.
- Frank, M. C., Barginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, 44, 677–694. <https://doi.org/10.1017/S0305000916000209>
- Freeborn, S. (2022). *What we talk about when we talk to babies: The sensorimotor associations of speech to Scottish infants* [Unpublished Honours thesis]. University of Sydney, Sydney, Australia.
- Floyd, S., San Roque, L., & Majid, A. (2018). Smell is coded in grammar and frequent in discourse: Cha'palaa olfactory language in cross-linguistic perspective. *Journal of Linguistic Anthropology*, 28(2), 175–196. <https://doi.org/10.1111/jola.12190>
- Gentner, D., & Bowerman, M. (2009). Why some spatial semantic categories are harder to learn than others: The typological prevalence hypothesis. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & Ş. Özçalışkan (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin*, 465–480. New York: Lawrence Erlbaum Associates.
- Howes, D. (Ed.). (1991). *The varieties of sensory experience: A sourcebook in the anthropology of the senses*. Toronto, Canada: University of Toronto Press.
- Johnson, C. (1999). Metaphor vs. conflation in the acquisition of polysemy: The case of *see*. In M. K. Hiraga, C. Sinha, & S. Wilcox (Eds.), *Cultural, psychological and typological issues in cognitive linguistics: Selected papers of the bi-annual ICLA meeting in Albuquerque, July 1995* (155–170). Amsterdam, the Netherlands: John Benjamins.
- Landau, B., & Gleitman, L. (1985). *Language and experience: Evidence from the blind child*. Cambridge, MA: Harvard University Press.
- Little, E. E., Carver, L. J., & Legare, C. H. (2016). Cultural variation in triadic infant-caregiver object exploration. *Child Development*, 87(4), 1130–1145. <https://doi.org/10.1111/cdev.12513>
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Majid, A., & Levinson, S. C. (2011). The senses in language and culture. *Senses and Society*, 6(1), 5–18. <https://doi.org/10.2752/174589311X12893982233551>
- Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., Woll, B., LeLan, B., de Sousa, H., Cansler, B. L., Shayan, S., de Vos, C., Senft, G., Enfield, N. J., Razak, R. A., Fedden, S., Tufvesson, S., Dingemanse, M., Ozturk, O., Brown, P., Hill, C., Le Guen, O., Hirtzel, V., van Gijn, R., Sicoli, M. A., & Levinson, S. C. (2018). Differential coding of perception in the world's languages. *PNAS*, 115, 11369–11376. <https://doi.org/10.1073/pnas.1720419115>
- Maouene, J., Hidaka, S., & Smith, L. B. (2008). Bodyparts and early-learned verbs. *Cognitive Science*, 32, 1200–1216. <https://doi.org/10.1080/03640210802019997>
- Norcliffe, E., & Majid, A. (2024). Verbs of Perception: A Quantitative Typological Study. *Language* 100(1). 81–123. <https://doi.org/10.1353/lan.2024.a922000>
- Piaget, J. (1953). *The origins of intelligence in children*. London, England: Routledge & Kegan Paul.
- Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences* 22(2), 253–336. <https://doi.org/10.1017/50140525X9900182X>
- R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Sanchez, A., Meylan, S., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank, M. C. (2018). *Chilides-db: A flexible and reproducible interface to the Child Language Data Exchange System*. <https://doi.org/10.31234/osf.io/93mwx>
- San Roque, L., Kendrick, K. H., Norcliffe, E., Brown, P., Defina, R., Dingemanse, M., Dirksmeyer, T., Enfield, N. J., Floyd, S., Hammond, J., Rossi, G., Tufvesson, S., Van Putten, S., & Majid, A. (2015). Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive Linguistics*, 26, 31–60. <https://doi.org/10.1515/cog-2014-0089>

- San Roque, L., & Schieffelin, B. B. (2019). Perception verbs in context: Perspectives from Kaluli (Bosavi) child-caregiver interaction. In L. Speed, C. O'Meara, L. San Roque, & A. Majid (Eds.), *Perception metaphors* (pp. 347–368). Amsterdam. The Netherlands: Benjamins.
- Styles, S., & Plunkett, K. (2009). What is “word understanding” for the parent of a one-year-old? Matching the difficulty of a lexical comprehension task to parental CDI report. *Journal of Child Language*, 36, 895–908. <https://doi.org/10.1017/S0305000908009264>
- Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge, England: Cambridge University Press.
- Viberg, Å. (1983). The verbs of perception: A typological study. *Linguistics*, 21, 123–162. <https://doi.org/10.1515/ling.1983.21.1.123>
- Viberg, Å. (1994). Crosslinguistic perspectives on lexical organization and lexical progression. In K. Hyldenstam & A. Viberg (Eds.), *Progression and Regression in language: Sociocultural, neuropsychological and linguistic perspectives*, 340–383. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511627781>
- Viberg, Å. (2001). The verbs of perception. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals. An international handbook*, vol. 2, 1294–1309. Berlin: De Gruyter.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Golemund, G. et al. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Winter, B., Perlman, M., & Majid, A. (2018). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition*, 179, 213–220. <https://doi.org/10.1016/j.cognition.2018.05.008>
- Winter, B., & Bürkner, P. C. (2021). Poisson regression for linguists: A tutorial introduction to modelling count data with brms. *Language and Linguistics Compass*, 15(11), e12439. <https://doi.org/10.1111/lnc3.12439>

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary Materials