CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Referential transparency of verbs in child-directed input by Japanese and American caregivers

Allison Fitch[1,2] (ID), Amy M. Lieberman[2], Michael C. Frank[3] (ID), Jessica Brough[2,4], Matthew Valleau[2] and Sudha Arunachalam[2,5]

[1]Rochester Institute of Technology, USA; [2]Boston University, USA; [3]Stanford University, USA; [4]University of Edinburgh, UK and [5]New York University, USA
Corresponding author: Allison Fitch; Email: ahfgsh@rit.edu

Abstract

Children acquiring Japanese differ from those acquiring English with regard to the rate at which verbs are learned (Fernald & Morikawa, 1993). One possible explanation is that Japanese caregivers use verbs in referentially transparent contexts, which facilitate the form-meaning link. We examined this hypothesis by assessing differences in verb usage by Japanese and American caregivers during dyadic play with their infants (5-22 months). We annotated verb-containing utterances for elements associated with referential transparency and compared across groups. Contrary to our hypotheses, we found that Japanese caregivers used verbs in fewer referentially transparent contexts than American caregivers, or did not significantly differ from American caregivers, depending on the measure. These findings cast doubt on cross-cultural differences in referential transparency between Japanese and American child-directed input.

Introduction

The first words children acquire vary by individual and reflect both the language the children are acquiring and the context in which that language is acquired. However, it appears to be consistent – at least across languages where this question has been studied in depth – that children generally acquire words for objects (nouns) before words for actions (verbs) (Au et al., 1994; Caselli et al., 1995; Gentner, 1982; Setoh et al., 2021).

While the existence of a noun bias is consistent, its strength (i.e., the relative proportion of nouns to verbs in early vocabulary) seems to vary cross-linguistically (Frank et al., 2021). Children acquiring English have a strong noun bias with relatively few verbs in their early vocabularies, while children acquiring Japanese (along with Korean, and Mandarin) have a more substantial proportion of verbs in their early vocabularies (but are still noun-biased; Tardif, 1996; Choi & Gopnik, 1995; Tardif et al., 1999; Fernald & Morikawa, 1993; C. C. Y. Chan et al., 2011, see also Frank et al., 2021). We refer to these latter languages as verb-friendly.

Explanations for the variation in degree of noun bias center around two separate hypotheses – a structural argument and an input argument. Those who propose structural arguments hypothesize that the grammar of Japanese is more verb-friendly (Tardif et al., 1997). For example, Japanese places verbs in salient positions in utterances (Caselli et al., 1995; Slobin, 1985). In contrast, input accounts argue that the context in which Japanese is acquired is verb-friendly: caregivers use verbs in referentially transparent contexts, facilitating a connection between the verb and its referent action (Choi & Gopnik, 1995; Tardif et al., 1999).

In the current study, we aim to further explore the input hypothesis by comparing two languages that differ in their degree of noun bias: Japanese and American English. Specifically, we assess the referential transparency of verbs in child-directed Japanese and American English. To do this, we capitalize on a corpus of parent-infant play interactions from Japanese and American dyads (Fernald & Morikawa, 1993), and examine the use of referential transparency cues coupled with imageable verbs.

Why is it difficult to learn verbs?

Although the degree of noun bias varies across languages, verbs are almost always harder to learn than nouns. Several non-mutually exclusive accounts have been proposed to explain why this is the case, but many converge on the notion that verb meanings are more difficult to connect with their forms than nouns. For example, some theories argue that humans have innate assumptions that novel words refer to objects (Markman, 1989) or that novel words refer to entire categories of objects (Waxman, 1990), but no such assumption exists for events. Gentner's (1982) Natural Partitions hypothesis argues that it is easier to individuate a concrete object than it is an event. Indeed, while verbs generally refer to dynamic events that unfold over time, nouns often refer to static objects that persist over time. In the context of word learning, the ability to simultaneously label a referent and attend to it (also referred to as joint attention) is considered extremely powerful (Tomasello & Farrar, 1986; Trueswell et al., 2016). Static objects offer ample opportunity to engage in joint attention, but dynamic events are fleeting and finite. Likewise, what qualifies as simultaneous in the context of an unfolding event is unclear. Watching the event of *diving*, for example, could be labeled at the launch from the diving board, the brief time in the air, or the entry into the water.

Verbs may also be ambiguous because events are composed of many parts, of which the verb only refers to some (e.g., Gleitman, 1990). For example, in a sentence like *He's diving into the pool*, the event includes a subject, a method of movement, and a destination. The verb *diving* only refers to the manner in which the subject enters the pool and a prepositional phrase is used to describe the destination. While objects are also composed of many parts (shape, size, color, texture, etc.), languages are more consistent in which part they label (shape). We see differences cross-linguistically in which aspects of events are labeled by verbs. Whereas English verbs typically denote manner, Spanish verbs typically denote path (Talmy, 1975). To a language-acquiring child, it requires experience to learn which event feature a verb in their language denotes, and thus this may be more difficult than identifying nouns (Gentner, 1982).

Verbs thus should be harder to learn than nouns. Why then do languages vary in their verb-friendliness? Cross-linguistic differences in verb acquisition may be correlated with differences between languages and communities in how they reduce the ambiguity associated with verbs. Verb-friendly languages could draw clearer connections between

verb form and meaning either through properties inherent to the languages' structure (structural accounts), or through how input is provided to children acquiring those languages (input accounts). We review each of these potential accounts in turn.

The structural account of differences in verb bias

Given that different languages encode and use verbs differently, perhaps language-inherent properties are responsible for cross-linguistic differences in verb friendliness. For example, word order dictates the positional salience of verbs. In English and other SVO languages, the verb is in the middle of the utterance, while SOV languages like Japanese place the verb in the utterance final position, which is more salient (e.g., Seidl & Johnson, 2006). This salience might offer clearer connections between the linguistic input and the event it references. However, on its own this is not a wholly satisfying explanation. Not every SOV language is verb-friendly, so this account cannot fully explain cross-linguistic differences in verb friendliness (Caselli et al., 1995).

Another structural feature of some verb-friendly languages that may confer support to verb acquisition is pro-drop. Pro-drop allows for the omission of pronominal arguments that can be inferred from context. This in turn allows for bare verbs to comprise entire utterances, stripped of extra information. These bare verb utterances are frequent in child-directed input in Japanese and Korean, two pro-drop languages (e.g., Kim, 2000; Smith & Frank, 2012). Like word order salience, this allows for a clear connection between the input and a new event being introduced, labeling only the new information (the event itself) and omitting what is already known. Indeed, studies in which children are asked to learn novel verbs demonstrate that children acquiring Japanese and Korean, two pro-drop languages, perform better when arguments are dropped than when they are overt (Arunachalam et al., 2013; Imai et al., 2008). However, this factor too cannot fully explain cross-linguistic differences in verb friendliness, as pro-drop is not perfectly correlated with verb friendliness (Tardif et al., 1997).

The input account of differences in verb bias

In contrast to linguistic accounts, input accounts argue that child-directed input in verb-friendly languages highlights verbs in ways that enhance verb learning relative to child-directed input in noun-friendly languages. One piece of evidence in favor of this account comes from bilingual children who acquire one noun-friendly and one verb-friendly language from the same caregiver(s). Prior work has demonstrated a greater number of nouns in the acquisition of Mandarin (compared to monolingual norms) by Mandarin–English bilinguals in an English dominant environment (W. H. Chan & Nicoladis, 2010), as well as a greater number of verbs in the acquisition of English among Mandarin–English bilinguals compared to Malay-English bilinguals (with Mandarin being more verb-friendly than Malay; Chai et al., 2021).

Additional evidence in favor of the input account comes from word frequency, which predicts substantial variation in children's early vocabularies (Frank et al., 2021; Goodman et al., 2008). Caregiver input from verb-friendly languages tends to contain verbs at higher frequencies (Tardif et al., 1997, 1999; Fernald & Morikawa, 1993). This variation could in turn be due to differences in cultural values, such as an emphasis on interactions and routines over object labelling (Fernald & Morikawa, 1993). However, in Italian (a noun-friendly language), caregivers also tend to use more verbs than nouns in their

child-directed input (Camaioni & Longobardi, 2001). Note though, that this could be an argument in favor of the structural hypothesis instead – the pattern may be a by-product of pro-drop, which allows nouns to be omitted from the utterance. That said, the relative proportion of verbs and nouns in caregiver input may vary by context and age of the child (at least between 12 and 24 months), even in verb-friendly languages (Ogura et al., 2006). Thus, verb frequency is unlikely to be the sole driver of verb-friendliness.

Another way variation in input might highlight verbs over nouns is through referential transparency, which is the degree to which a word's meaning can be readily deduced from the extralinguistic context (Gillette et al., 1999). Most studies addressing referential transparency have focused on nouns, and have demonstrated that referential transparency is greatest during moments of joint attention. Referential transparency is especially high when the caregiver follows the child's focus of attention into the episode (follow-in joint attention, Baldwin, 1991, 1993), as well as when there are few other objects in the child's field of view (Pereira et al., 2014) and when there is perfect co-occurrence between the label being uttered and attention being directed to the referent object (Trueswell et al., 2016).

Caregiver and child eye-gaze are particularly good indicators of a noun reference (Frank et al., 2013; using the same English corpus as the current study), but fewer studies have focused on what cues lead to referential transparency for verbs. Joint attention and other eye gaze cues do not neatly map onto fleeting events in the same way as static objects. Studies on caregiver directed input in English suggest that verbs are learned best in impending contexts – that is, when they are labeled BEFORE the event occurs (Tomasello & Kruger, 1992). We currently do not know if this timing is optimal across the range of verb-friendly languages, or if it is limited to English and/or other noun-friendly languages.

While we know little about what leads to referential transparency for verbs, there is evidence that referential transparency in caregiver input varies cross-linguistically and cross-culturally. Work by Snedeker et al. (2003) demonstrated that referential transparency in English child-directed input is much higher for nouns than verbs, but roughly equivalent for nouns and verbs in Mandarin, a verb-friendly language. They used a Human Simulation Paradigm, in which participants were asked to guess the meaning of a target word based on only the video of the extralinguistic context in which that word was used (the sound was off, so there was no linguistic context upon which to base the guess – Gillette et al., 1999). Regardless of whether the participants themselves spoke English or Mandarin, guessing accuracy was higher for nouns for the English videos and similar between nouns and verbs in the Mandarin videos. Fitch et al. (2021) obtained a similar finding in a study where sign-naïve English speakers were significantly better at determining the meanings of child-directed verbs than child-directed nouns in American Sign Language (ASL). Children acquiring ASL typically have more verbs in their early vocabularies than children acquiring English (though ASL is still noun-friendly; Anderson & Reilly, 2002) and child-directed ASL tends to be verb heavy (Fieldsteel et al., 2020).

One final way that the timing of a verb label relative to a referent event might be affected by language or culture is in the kinds of speech acts that are used in the input. Imperatives and interrogative requests are used most often in impending contexts by nature. Utterances such as: *can you touch the doggy's nose?* Or *throw the ball!* are not likely to be labeling an event that has already happened. In contrast, a declarative can describe events in impending, ongoing, or completed contexts equally well. There is evidence to suggest that the use of different speech acts varies by culture. With regard to American vs. Japanese culture, one study noted that Japanese mothers were less likely to use interrogatives and more likely to use imperatives with their two-year-olds than American

mothers (Clancy, 1985). On the other hand, a different study on much younger children (three-month-olds) suggested Japanese mothers used fewer imperatives than American mothers, but was in agreement that American mothers used more questions than Japanese mothers (Toda et al., 1990).

The current study

Taken together, the findings highlighted above suggest substantial gaps in the literature, particularly whether or not verb input in verb-friendly languages is more referentially transparent (as suggested by Snedeker et al. and Fitch et al.), and if so, what cues lead to that referential transparency. In the current study, we seek to address these gaps by examining differences in child-directed input by American caregivers (with their English-acquiring infants) and Japanese caregivers (with their Japanese-acquiring infants). We used an existing corpus in which American caregiver-infant dyads and Japanese caregiver-infant dyads played in their homes with a standard set of objects (Fernald & Morikawa, 1993). By analyzing interactions that are consistent across two cultures, we hope to identify possible factors in maternal input that may contribute to the relative advantage for verb learning in one verb-friendly language, Japanese. Specifically, we assessed differences in verb input that prior literature suggests should contribute to referential transparency.

Our strategy is to annotate naturalistic dyadic interactions with each of our features of interest and then to ask which of them vary between the two groups. We first asked if caregivers use verbs to refer to a co-occurring action, reasoning that a minimum criterion for joint attention is the ability to observe the referent event. We then examined the timing of a verb utterance relative to an event. Next, we looked at the agent of the referent event, reasoning that when a caregiver labels an event that the child carries out, that is akin to follow-in joint attention. Finally, we investigated speech act (declarative, imperative, or interrogative), to determine if frequency of imperatives (or perhaps interrogatives) might contribute independent variance in verb friendliness.

We hypothesized that Japanese mothers would provide verb input that differs from American mothers when playing with their infants, in a way that makes the meanings of verbs more referentially transparent. Specifically:

1. Japanese mothers will be more likely to label events that occur and are observed. Such events will be more likely labeled using imperatives.
2. Japanese mothers will differ from American mothers in the timing of when they label an event. If labeling impending events is universally beneficial (as is true for American mothers in prior work; Tomasello & Kruger, 1992), we should see greater use of labels before the referent event occurs in Japanese relative to American mothers. This may be due to increased frequency of modeling the action.
3. Japanese mothers will use more imperatives than American mothers, which are more likely to label impending events.

Methods

Participants

Data were extracted from a corpus of transcribed, dyadic interactions collected by Fernald and Morikawa (1993). Dyads were Japanese mothers (n = 28) with their Japanese-

Table 1. Number of infants within each age bin (see also Fernald & Morikawa, 1993).

	Japanese	American
5–6 months	8	8
11–14 months	9	8
18–21 months	10	8

acquiring infants (5.5-21 months of age, M = 12.67 months, SD = 5.58) and American mothers (n = 24) with their English-acquiring infants (6-21 months of age, M = 12.48 months, SD = 5.43). The age distribution of the infants is in Table 1. This wide age range was motivated by prior literature on cross-cultural maternal input, which encompasses a diverse range of ages from three months (Toda et al., 1990) to two years (Clancy, 1985; Ogura et al., 2006). One Japanese dyad was excluded due to the caregiver not producing any codable utterances (n = 1).

Procedure
Observation
Each interaction was a short, semi-structured play session. In brief, dyads played with a standard set of objects (box, brush, stuffed dog, stuffed pig, car, and truck) for 10-15 minutes. An experimenter recorded these interactions using audio and video recorders, and rotated the toys into and out of the interaction in pairs. Full details on the interaction procedure are available in Fernald and Morikawa (1993).

Coding
Three English–Japanese bilingual coders, who were naïve to study hypotheses, annotated the corpora. They were given access to the recorded interactions and an utterance-by-utterance transcription. They used Praat (Boersma & Weenink, 2013) and a custom coding program designed by one of the authors (MV) to divide each video recording into segments that coincided with each utterance. For each utterance, coders were instructed to determine the following:

1) Whether the utterance contained a concrete, imageable verb that elicited or described an event (i.e., modals, abstract verbs, and displaced speech did not meet this criteria). This first-line of coding determined the final data set for analysis.
2) The speech act, which was coded as declarative, imperative, or interrogative.
3) Whether the event described by the imageable verb in the utterance occurred within 10 seconds of the utterance onset and offset (10 seconds before onset and 10 seconds after offset). On average, utterances were produced every ~3.3 seconds (defined as the number of utterances produced during the session divided by the length of the play session, range = 1.9 – 5.6 seconds), thus we believe this window is more than sufficient to describe the concurrent visual context. However, in the case of repetitive utterances, the event was associated with the FIRST use of the verb (and thus on successive repetitions, the event was counted as occurring even if it fell outside the 10 second window for the latter repetition(s)).

Table 2. Inter-rater reliability values for each coded category.

Category	n double-coded	α
Contains codable verb	18630	0.98
Speech act	3984	.76
Elicits action	5882	0.93
Action occurred	2096	0.64
Who was actor	1835	0.76
Onset of Event	1304	0.99

4) If the event did occur, which participant carried out the event (caregiver or child). This was coded to determine if caregivers were more likely to model the labeled event or not.
5) If the event did occur, when it occurred (in seconds) relative to the verb being uttered.

Inter-rater reliability

Coders were considered reliable by attaining greater than 80% agreement with another trained coder on an initial set of videos. Across the Japanese and English corpora, 18630 utterances were double coded to confirm inter-rater reliability. Disagreements were settled by a third coder or the first author. Krippendorff's alpha was calculated in each category using the IRR package (Gamer et al., 2019); alpha values are listed in Table 2.

Data analysis

Analyses were limited to utterances that elicited or described an imageable verb (n = 1522 Japanese, n = 1180 American). Hypotheses and analyses were pre-registered (https://osf.io/zekcf). De-identified data and reproducible code is available at https://github.com/langcog/fm_verbs.

Results

Event occurrence

Our first analysis focused on event occurrence, i.e., whether or not the event to which the verb referred occurred within 10 seconds of the verb-containing utterance. We first analyzed the role of corpus, child age, and speech act on whether or not a referent event occurred within the 10-second time window on each side of the utterance. We used a generalized mixed-effects linear model (family = binomial) with event occurrence as the dependent variable, and corpus, child age in months, their interaction term, and speech act as fixed effects, with subject as a random factor[1] for the intercept only (Figure 1).

We observed a significant effect of corpus: verbs used by Japanese mothers were less likely to have an accompanying referent event than verbs used by American mothers

---

[1]When pre-registering this analysis, we also planned to include who carried out the event as a fixed effect in the model. However, this was an error, as who carries out the event is only coded if the event occurs.
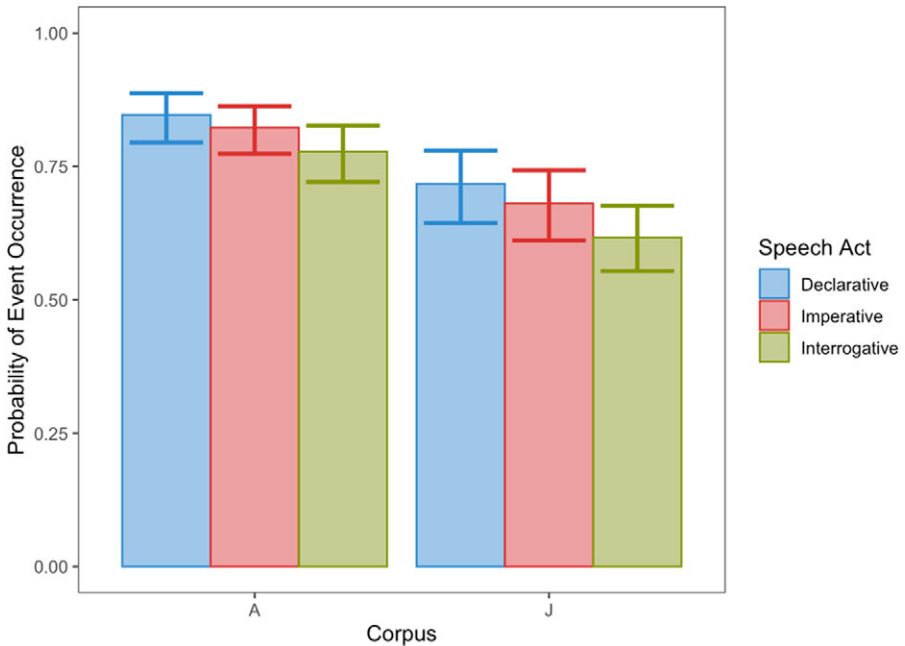
Figure 1. Probability of event occurrence as a function of corpus and speech act.

Table 3. Percentage of verb-containing utterances by corpus and speech act.

	Declarative	Imperative	Interrogative
American	20.25%	50%	28.95%
Japanese	23.17%	16.9%	54.71%

(β = -1.26, *SE* =.6, *Z* = -2.09, *p* =.04). Neither age alone, nor the interaction between age and corpus predicted significant variation in the model (*ps* =.27 and .41 respectively). Speech act also contributed unique variance such that interrogatives led to fewer events occurring relative to declaratives (β = -.45, *SE* = .14, *Z* = -3.21, *p* = .001). A follow-up exploratory (not pre-registered) *t*-test showed that the Japanese caregivers used significantly more interrogatives as a percent of utterances than American caregivers, *t*(44.6) = -4.28, *p* < .001, 95% CI = [-.38, -.13] (see Table 3).

Features of referent events: agent and timing

Our next analysis was designed to identify factors influencing the timing of verbs in parent input relative to its referent event. To determine the timing, we subtracted the timestamp of the target verb onset from the timestamp associated with the onset of the referent event. Thus, events that occurred before the target verb received negative times (e.g., -1 second for an event that started one second before the target verb was uttered) and those that occurred after received positive times. We asked if culture, child age, speech act,

Table 4. Model output: Action onset time relative to verb utterance by corpus, age (continuous), agent of action and speech act

	Estimate	Standard Error	t	p
Intercept	1.16	1.4	0.83	.41
Japanese	1.19	2.34	0.51	.61
Age (months)	−0.06	0.09	−0.62	.54
Parent was agent	1.63	1.4	1.16	.25
Imperative	−0.59	0.64	−0.91	.36
Interrogative	−0.12	0.67	−0.19	.85
Japanese * Age	−0.09	0.15	−.59	.56
Japanese * Parent agent	2.67	2.61	1.02	.31
Age * Parent agent	−0.09	0.1	−0.91	.37
Japanese * Age * Parent	−0.17	0.17	−0.96	.34

Reference levels: American, child, declarative.

and who carried out the action predicted when the verb was uttered relative to the event. For this mixed-effects linear model, we entered relative event onset time as the dependent variable; and corpus, age, whether the parent carried out the event or not, and the interaction term (corpus x age x actor), as well as speech act as fixed factors. Subject was entered as a random factor for the intercept. Findings showed that none of the factors entered into the model predicted significant variation in relative event onset time (see Table 4).

Given this null result, we conducted a follow-up (not pre-registered) Bayesian analysis using the BayesFactor package (Morey et al., 2023), in which we compared Bayes factors for each fixed factor as its own model against the intercept only model, with the constraint that all models must also include the random factor (subject). As shown in Figure 2, simpler models performed better than the full model with two exceptions. The model that included the interaction between age and who carried out the action was slightly preferred to the full model (BF_{10} =.14 ± 8.64%). The best model (which included both age and subject; BF_{10} =.65 ± 10.13%) was preferred to the full model by a factor of ~6.5.

To better understand the range of verb onset time relative to event onset, we plotted the distribution of onset time for each group (Figure 3). This visualization made it clear that there are substantial differences in variability between the two corpora. The distribution of onset times in the Japanese corpus is notably wider and flatter than the American corpus, which has a high and tight peak around 0.

We wondered if this distribution was related to whether caregivers were talking about events in the past or that were about to happen. We conducted an exploratory analysis in which verbs were classified as describing an impending action (the verb described an event that has yet to begin) or an ongoing/completed action (the verb described an event that had already begun). For example, if a caregiver asked: "Can you throw the ball?" followed by the child throwing the ball, the verb *throw* would be classified as describing an impending action, whereas a child throwing a ball and a caregiver commenting "You threw it!" would be classified as describing a completed action. We used a mixed-effects linear model where the proportion of impending actions was entered as the dependent variable, corpus and speech act and their interaction were entered as fixed factors, and
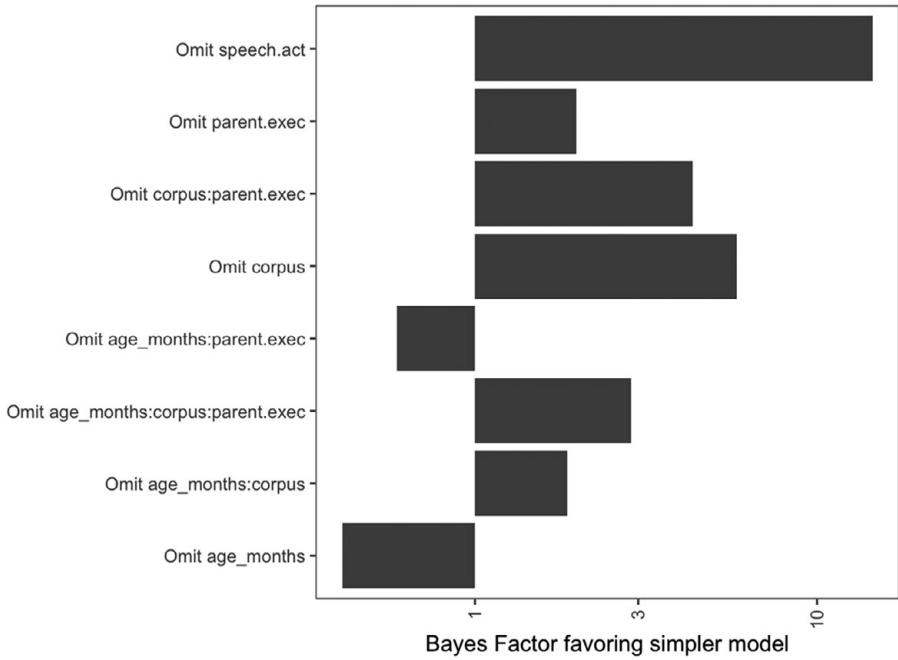
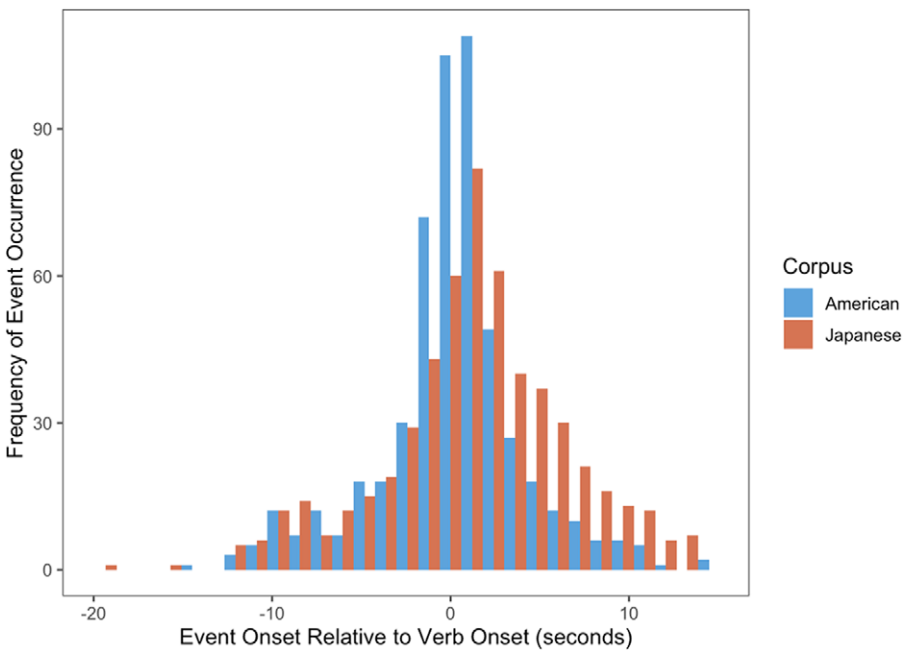**Figure 2.** Bayes factors associated with each fixed factor model.

Figure 3. Distribution of action verb referent occurrence by onset time relative to its label.

Table 5. Model output: Likelihood of impending action by corpus and speech act

	Estimate	Standard Error	z	p
Intercept	0.08	0.38	0.211	0.83
Japanese	1.53	0.79	1.95	0.052
Imperative	−0.05	0.37	−0.15	0.88
Interrogative	0.34	0.4	0.84	0.4
Japanese * Imperative	−0.6	0.78	−0.77	0.44
Japanese * Interrogative	−1.64	0.79	−2.09	0.037

American corpus and declarative utterances were used as reference.
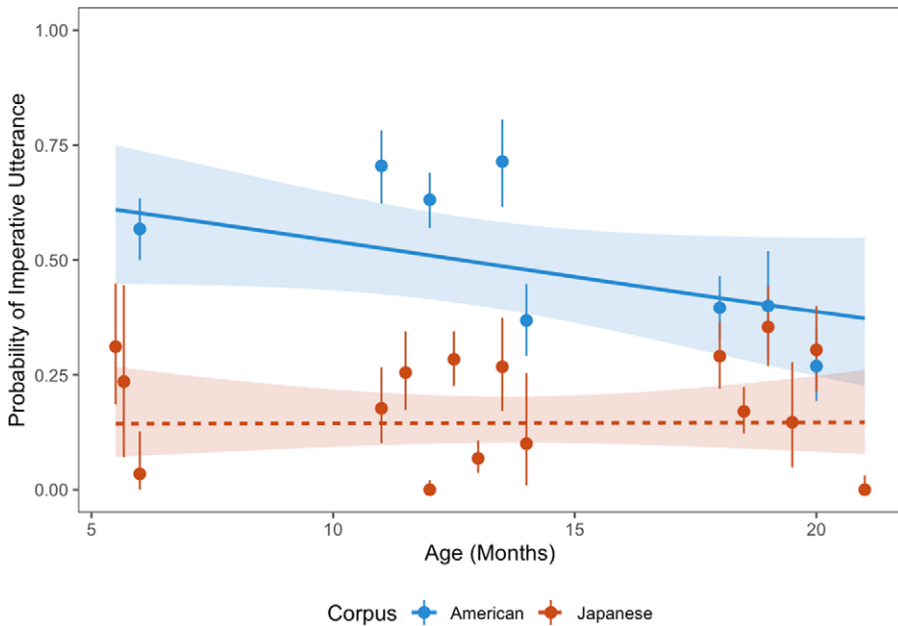

Figure 4. Probability that an utterance will be imperative as a function of corpus and child age.

subject was entered as a random factor (intercept grouped only). We observed only a significant interaction between corpus and speech act, where relative to the reference level (American, declarative utterances), interrogative utterances in the Japanese corpus were more likely to describe an impending action (see Table 5).

In our final analysis, we looked to determine the relation between whether or not an utterance was imperative, and the age and culture of the child. To do this, we again conducted a mixed-effects generalized linear model (family = binomial) using corpus, child age in months, and the interaction term as fixed factors, and subject as a random factor (intercept only), with imperative status as the dependent variable (Figure 4). Findings demonstrated a significant effect of corpus, with Japanese caregivers producing

fewer imperatives than American caregivers (β = -2.58, *SE* = .79, *Z* = -3.26, *p* = .001). Neither age nor the interaction term yielded significant effects (*ps* = .09 and .25 respectively).

Discussion

The goal of our study was to assess one aspect of the hypothesis that Japanese child-directed input is more verb-friendly than American English input. Specifically, we hypothesized that Japanese caregivers use verbs in more referentially transparent contexts by labeling events that occur and by timing their utterances differently from American caregivers perhaps, by using more imperative utterances. Across all hypotheses, we found no supporting evidence for these preregistered hypotheses. Instead, we found that Japanese caregivers were LESS likely to label events that occurred and were LESS likely to use imperatives than American mothers; no significant differences were identified in the relative timing of label and event onset. In sum, we did not see strong evidence that Japanese child-directed input uses more referentially transparent contexts for verbs than American English input.

The findings presented here are surprising in light of the conclusions from Snedeker et al.' (2003) Human Simulation Paradigm which showed that the extralinguistic context of a verb-friendly language (Mandarin) provided increased referential transparency for verbs relative to a noun-friendly language (English). However, the differences between their study and the one presented here may explain the disparity in findings. Notably, this study was not a Human Simulation Paradigm (a necessary limitation of the study due to data privacy concerns). Rather than determining the referential transparency of each verb through the extralinguistic context, we predetermined a set of cues associated with referential transparency and assessed the extralinguistic context surrounding each verb for those cues. If we had conducted a Human Simulation Paradigm on this corpus, we do not know if we would have found the same results as Snedeker et al. (2003). Moreover, we saw some disagreement among trained coders as to whether or not those cues to referential transparency were available (Table 2), particularly for whether or not the event occurred. Anecdotally, disagreements occurred most often in manner-focused events where the intensity of an action can distinguish it from similar actions. For example, *giving* and *throwing* a ball to someone can be difficult to disentangle, particularly for 12-month-olds infants who may "throw" a ball from just inches away. Likewise *throwing* a ball into a basket may more closely resemble *putting* or *dropping* the ball when an infant is the agent of the action. Regardless, there was much greater agreement for who carried out the action and when it occurred – both features of the context that cannot be coded if the action did not actually occur. As such, we felt emboldened to move forward with the analysis despite the lower alpha value.

A further difference between Snedeker et al. (2003) and the current study is that the language studied here was Japanese as opposed to Mandarin. While both languages are verb-friendly, it's not clear that the mechanisms that support verb-friendliness are the same in each. In other words, while Mandarin may lead to greater referential transparency for verbs, Japanese input may promote verb acquisition through other culture-specific input mechanisms or language-specific structural mechanisms. For example, Mandarin is an SVO language while Japanese is SOV. While SOV word order alone seems to not explain verb-friendliness, perhaps SOV word order and pro-drop together are enough to highlight verbs for new language learners (but see Frank et al., 2021 re: Korean). This does not rule out the input account though. The cues we assessed here were previously

associated with referential transparency in English. It's possible that these cues do not generalize to referential transparency in Japanese or other verb-friendly languages.

It is also possible that cues to referential transparency may interact with verb meaning in a way we have not addressed here. Based on prior findings from English (Tomasello & Kruger, 1992) we assumed impending contexts would provide the greatest referential transparency for verbs. However, manner verbs (e.g., jump or crawl) may be better learned in impending contexts, while result verbs (e.g., fall or break) may be better learned in completed contexts (when the event is labeled AFTER it has finished; Ambalu et al., 1997). Additionally, it may be the case that familiar verbs and novel verbs differ in referential transparency. Unlike prior experiments, we were unable to control for verb familiarity, and it is likely that the children in this study were familiar with at least some of coded verbs. We did not see any age effects where older children (and thus those more familiar with words) received different cues than younger children, which casts some doubt on this hypothesis, but does not eliminate it. Further research into these potential differences is warranted.

In addition to cross-linguistic differences, it is important to consider cross-cultural differences in determining the factors supporting early verb acquisition. Prior work has clearly demonstrated that cultural values affect the content of caregiver input (e.g., Bornstein et al., 1992; Caselli et al., 1995; Choi & Gopnik, 1995; Morikawa et al., 1988), and this might extend to the kinds of words that are used and emphasized. For example, individualistic cultures may be more likely to emphasize the names for individuals and objects (i.e., nouns) while collectivist cultures may be more likely to emphasize relations between individuals and objects (which are often activities, and therefore verbs; Lavin et al., 2006). Indeed, Japanese caregivers are more likely to emphasize social routines and scripts (such as greeting and modeling appropriate behavior) in their child-directed input compared to American mothers, who are more likely to label objects (Clancy, 1985; Fernald & Morikawa, 1993). This emphasis on social routines may support the acquisition of particular kinds of verbs (e.g., waving, bowing, eating) while object emphasis may support the acquisition of object-focused verbs (e.g., cutting, sweeping). Cultural differences may also influence the degree to which caregivers gesture along with their speech. Gesture, which has been shown to support the acquisition of new verbs (Mumford & Kita, 2014), varies in both kind and frequency across cultures including between Japanese and Americans (e.g., Kita, 2009). It may be the case that Japanese caregivers provide more supportive gesture along with their verb productions than American caregivers. Lastly, studies of bilingual children have shown that noun- and verb-friendliness can be influenced by their caregiver's culture and/or environment (Chai et al., 2021; W. H. Chan & Nicoladis, 2010). Thus, culture-specific indicators to verb-friendliness are a promising area for future research.

Although not the main focus of the research question, it is important to address our findings on speech acts. Prior studies were mixed with regard to the number of imperative utterances used by Japanese relative to American caregivers. However, those findings agree that Japanese caregivers use fewer interrogatives than American caregivers (Bornstein et al., 1992; Clancy, 1985; cf. Toda et al., 1990). Our findings are in partial agreement with the prior literature. We find Japanese caregivers use fewer imperatives as a proportion of utterances, consistent with the idea they are less directive. Our findings diverge from the prior literature on interrogatives, however: we found greater interrogative use in the Japanese relative to the American caregivers. This disparity may reflect a difference in methodology. Here, we only counted utterances with an imageable verb, which excluded questions like *What is it?* or *Is it a Doggy?*, which are more likely to

comprise input to American infants than Japanese infants (Fernald & Morikawa, 1993). This method also excluded utterances without a verb, such as onomatopoeia, which is more frequent in Japanese than American English (Fernald & Morikawa, 1993), and reduces the proportion denominator relative to prior studies. In our findings, speech act and corpus interacted such that Japanese interrogative utterances were more likely to use verbs that labeled impending actions. This suggests that those interrogatives may have been requests to carry out an action, which is a different kind of question than has been previously associated with American child-directed input (e.g., Toda et al., 1990). Another reason our findings may have differed from prior work is that the trained coders in our study had lower reliability scores for speech act than we anticipated. Many of these cases are attributable to the ambiguity of child-directed speech and/or the exact play scenario. For example, one caregiver said "open the door" and "shut the door" while her child opened and shut the doors to a toy car. Due to the camera angle, it is difficult to ascertain if she was narrating/labeling the actions her child was engaging in (a declarative act) or if she was instructing her child to open and shut the doors (an imperative act). Likewise, a similar statement such as "Can you open the door?" may have been interpreted as either interrogative or imperative.

Taken together, our findings do not support the hypothesis that Japanese child-directed input is more referentially transparent for verbs than American English child-directed input, although these findings may be limited by how referential transparency was operationalized in the current study. Future work on verb-friendliness may need to identify what referential transparency means in the context of Japanese (and perhaps other verb-friendly languages), as well as examine the intersections of child-directed input and culture. This may be best carried out using large scale datasets that include a larger variety of languages and cultures. Nevertheless, the current results add to our understanding of the complex relationship between features of both language and input that contribute to verb learning in young children.

Competing interest. The authors declare none.

References

Ambalu, D., **Chiat, S.**, & **Pring, T.** (1997). When is it best to hear a verb? The effects of the timing and focus of verb models on children's learning of verbs. *Journal of Child Language*, **24**(1), 25–34. https://doi.org/10.1017/S0305000996002978

Anderson, D., & **Reilly, J.** (2002). The MacArthur Communicative Development Inventory: Normative Data for American Sign Language. *The Journal of Deaf Studies and Deaf Education*, **7**(2), 83–106. https://doi.org/10.1093/deafed/7.2.83

Arunachalam, S., **Leddon, E. M.**, **Song, H.**, **Lee, Y.**, & **Waxman, S. R.** (2013). Doing More With Less: Verb Learning in Korean-Acquiring 24-Month-Olds. *Language Acquisition*, **20**(4), 292–304. https://doi.org/10.1080/10489223.2013.828059

Au, T. K. F., **Dapretto, M.**, & **Song, Y. K.** (1994). Input Vs Constraints: Early Word Acquisition in Korean and English. *Journal of Memory and Language*, **33**(5), 567–582. https://doi.org/10.1006/jmla.1994.1027

Baldwin, D. A. (1991). Infants' Contribution to the Achievement of Joint Reference. *Child Development*, **62**(5), 875–890. https://doi.org/10.1111/j.1467-8624.1991.tb01577.x

Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology*, **29**(5), 832–843. https://doi.org/10.1037/0012-1649.29.5.832

Boersma, P. W., & **Weenink, D.** (2013). *Praat: Doing phonetics by computer*. [Computer software]. https://www.praat.org

Bornstein, M. H., **Tamis-LeMonda, C. S.**, **Tal, J.**, **Ludemann, P.**, **Toda, S.**, **Rahn, C. W.**, **Pêcheux, M.-G.**, **Azuma, H.**, & **Vardi, D.** (1992). Maternal Responsiveness to Infants in Three Societies: The United States, France, and Japan. *Child Development*, **63**(4), 808–821. https://doi.org/10.1111/j.1467-8624.1992.tb01663.x

Camaioni, L., & **Longobardi, E.** (2001). Noun versus verb emphasis in Italian mother-to-child speech. *Journal of Child Language*, **28**(3), 773–785. https://doi.org/10.1017/S0305000901004846

Caselli, M. C., **Bates, E.**, **Casadio, P.**, **Fenson, J.**, **Fenson, L.**, **Sanderl, L.**, & **Weir, J.** (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, **10**(2), 159–199.

Chai, J. H., **Low, H. M.**, **Wong, T. P.**, **Onnis, L.**, & **Mayor, J.** (2021). Extra-linguistic modulation of the English noun-bias: Evidence from Malaysian bilingual infants and toddlers. *Journal of Cultural Cognitive Science*, **5**(1), 49–64. https://doi.org/10.1007/s41809-021-00078-5

Chan, C. C. Y., **Tardif, T.**, **Chen, J.**, **Pulverman, R. B.**, **Zhu, L.**, & **Meng, X.** (2011). English- and Chinese-learning infants map novel labels to objects and actions differently. *Developmental Psychology*, **47**(5), 1459–1471. https://doi.org/10.1037/a0024049

Chan, W. H., & **Nicoladis, E.** (2010). Predicting two Mandarin–English bilingual children's first 50 words: Effects of frequency and relative exposure in the input. *International Journal of Bilingualism*, **14**(2), 237–270. https://doi.org/10.1177/1367006910363059

Choi, S., & **Gopnik, A.** (1995). Early acquisition of verbs in Korean: A cross-linguistic study. *Journal of Child Language*, **22**(3), 497–529. https://doi.org/10.1017/S0305000900009934

Clancy, P. M. (1985). The Acquisition of Japanese. In *The Crosslinguistic Study of Language Acquisition*. Psychology Press.

Fernald, A., & **Morikawa, H.** (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development*, **64**(3), 637–656.

Fieldsteel, Z., **Bottoms, A.**, & **Lieberman, A. M.** (2020). Nouns and Verbs in Parent Input in American Sign Language during Interaction among Deaf Dyads. *Language Learning and Development*, **16**(4), 351–363. https://doi.org/10.1080/15475441.2020.1784737

Fitch, A., **Arunachalam, S.**, & **Lieberman, A. M.** (2021). Mapping Word to World in ASL: Evidence from a Human Simulation Paradigm. *Cognitive Science*, **45**(12), e13061. https://doi.org/10.1111/cogs.13061

Frank, M. C., **Braginsky, M.**, **Yurovsky, D.**, & **Marchman, V. A.** (2021). *Variability and Consistency in Early Language Learning: The Wordbank Project*. MIT Press.

Frank, M. C., **Tenenbaum, J. B.**, & **Fernald, A.** (2013). Social and Discourse Contributions to the Determination of Reference in Cross-Situational Word Learning. *Language Learning and Development*, **9**(1), 1–24. https://doi.org/10.1080/15475441.2012.707101

Gamer, M., **Lemon, J.**, & **Singh, I. F. P.** (2019). *irr: Various Coefficients of Interrater Reliability and Agreement* (0.84.1) [Computer software]. https://cran.r-project.org/web/packages/irr/index.html

Gentner, D. (1982). *Why Nouns Are Learned before Verbs: Linguistic Relativity Versus Natural Partitioning. Technical Report No. 257*. https://eric.ed.gov/?id=ED219724

Gillette, J., **Gleitman, H.**, **Gleitman, L.**, & **Lederer, A.** (1999). Human simulations of vocabulary learning. *Cognition*, **73**(2), 135–176. https://doi.org/10.1016/S0010-0277(99)00036-0

Gleitman, L. (1990). The Structural Sources of Verb Meanings. *Language Acquisition*, **1**(1), 3–55. https://doi.org/10.1207/s15327817la0101_2

Goodman, J. C., **Dale, P. S.**, & **Li, P.** (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, **35**(3), 515–531. https://doi.org/10.1017/S0305000907008641

Imai, M., **Kita, S.**, **Nagumo, M.**, & **Okada, H.** (2008). Sound symbolism facilitates early verb learning. *Cognition*, **109**(1), 54–65. https://doi.org/10.1016/j.cognition.2008.07.015

Kim, Y.-J. (2000). Subject/Object Drop in the Acquisition of Korean: A Cross-Linguistic Comparison. *Journal of East Asian Linguistics*, **9**(4), 325–351. https://doi.org/10.1023/A:1008304903779

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes*, **24**(2), 145–167. https://doi.org/10.1080/01690960802586188

Lavin, T. A., **Hall, D. G.**, & **Waxman, S. R.** (2006). East and West: A Role for Culture in the Acquisition of Nouns and Verbs. In *Action meets word: How children learn verbs* (pp. 525–543). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195170009.003.0021

Markman, E. M. (1989). *Categorization and Naming in Children: Problems of Induction*. MIT Press.

Morey, R. D., **Rouder, J. N.**, **Jamil, T.**, **Urbanek, S.**, **Forner, K.**, & **Ly, A.** (2023). *BayesFactor: Computation of Bayes Factors for Common Designs* (0.9.12-4.5) [Computer software]. https://cran.r-project.org/web/packages/BayesFactor/index.html

Morikawa, H., **Shand, N.**, & **Kosawa, Y.** (1988). Maternal speech to prelingual infants in Japan and the United States: Relationships among functions, forms and referents. *Journal of Child Language*, **15**(2), 237–256. https://doi.org/10.1017/S0305000900012356

Mumford, K. H., & **Kita, S.** (2014). Children use gesture to interpret novel verb meanings. *Child Development*, **85**(3), 1181–1189.

Ogura, T., **Dale, P. S.**, **Yamashita, Y.**, **Murase, T.**, & **Mahieu, A.** (2006). The use of nouns and verbs by Japanese children and their caregivers in book-reading and toy-playing contexts. *Journal of Child Language*, **33**(1), 1–29. https://doi.org/10.1017/S0305000905007270

Pereira, A. F., **Smith, L. B.**, & **Yu, C.** (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, **21**(1), 178–185. https://doi.org/10.3758/s13423-013-0466-4

Seidl, A., & **Johnson, E. K.** (2006). Infant word segmentation revisited: edge alignment facilitates target extraction. *Developmental Science*, **9**(6), 565–573.

Setoh, P., **Cheng, M.**, **Bornstein, M. H.**, & **Esposito, G.** (2021). Contrasting lexical biases in bilingual English–Mandarin speech: Verb-biased mothers, but noun-biased toddlers. *Journal of Child Language*, **48**(6), 1185–1208. https://doi.org/10.1017/S0305000920000720

Slobin, D. I. (1985). *The crosslinguistic study of language acquisition*. Hillsdale, NJ: Earlbaum.

Smith, C., & **Frank, M.** (2012). Zero anaphora and object reference in Japanese child-directed speech. *Proceedings of the Annual Meeting of the Cognitive Science Society*, **34**(34).

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

Talmy, L. (1975). Semantics and syntax of motion. In *Syntax and Semantics volume* **4** (pp. 181–238). Brill. https://brill.com/downloadpdf/book/edcoll/9789004368828/BP000008.pdf

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology*, **32**(3), 492.

Tardif, T., **Gelman, S. A.**, & **Xu, F.** (1999). Putting the "Noun Bias" in Context: A Comparison of English and Mandarin. *Child Development*, **70**(3), 620–635. https://doi.org/10.1111/1467-8624.00045

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

Toda, S., **Fogel, A.**, & **Kawai, M.** (1990). Maternal speech to three-month-old infants in the United States and Japan. *Journal of Child Language*, **17**(2), 279–294.

Tomasello, M., & **Farrar, M. J.** (1986). Joint Attention and Early Language. *Child Development*, **57**(6), 1454–1463. https://doi.org/10.2307/1130423

Tomasello, M., & **Kruger, A. C.** (1992). Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. *Journal of Child Language*, **19**(2), 311–333. https://doi.org/10.1017/S0305000900011430

Trueswell, J. C., **Lin, Y.**, **Armstrong, B.**, **Cartmill, E. A.**, **Goldin-Meadow, S.**, & **Gleitman, L. R.** (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition*, **148**, 117–135. https://doi.org/10.1016/j.cognition.2015.11.002

Waxman, S. R. (1990). Linguistic biases and the establishment of conceptual hierarchies: Evidence from preschool children. *Cognitive Development*, **5**(2), 123–150. https://doi.org/10.1016/0885-2014(90)90023-M