UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Initiation Asymmetry in the Ontogenesis of Social Routines: In Conversation, Caregivers Scaffold 1-Year Olds to Respond, but 2-year Olds Initiate

Permalink
https://escholarship.org/uc/item/58z60167

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

Authors
Terwilliger, Jack
Rossano, Federico

Publication Date
2025

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Initiation Asymmetry in the Ontogenesis of Social Routines: In Conversation Caregivers Scaffold 1-year-olds to Respond, but 2-year-olds Initiate

Jack Terwilliger (jterwilliger@ucsd.edu)

Federico Rossano (frossano@ucsd.edu)
Department of Cognitive Science
UC San Diego

Abstract

Social routine words – e.g. *yes*, *no*, *hi*, *bye*, *okay*, and *thank you* – are among the first words children learn across different languages and they help constitute a foundational set of actions for conducting social interactions. Despite this, we know little about how these words are acquired. In this paper we begin by showing that social routine words are systematically acquired earlier than statistical models of word acquisition predict. Furthermore, we argue this gap is due to a selective focus of word learning science on reference – how words are mapped onto concepts and events – a relationship which is absent for social routine words. Rather than looking at the properties of words per se, we address this gap by instead looking at how children become conversational partners. Enroute to becoming a conversational partner, a child must orient themselves to others' expectations about the position & composition of their turns relative to their social partner's. We hypothesize that second pair parts (turns which respond, e.g. agreeing, acknowledging, reciprocating a greeting, etc.) afford more scaffolding from caregivers than first pair parts, and therefore words used to compose such responses are more learnable. To support this hypothesis, we sampled 1,442 conversational turns from 5 mother-child dyads (12mo-28mo) and manually labeled the position of these turns within adjacency pairs. We found that 12-month-old's talk is made mainly in response to caregivers who initiate adjacency pairs, but this initiation asymmetry in conversation disappears by 28-months. Because social routine words either stereotypically or frequently are used to compose second position turns, this pattern of initiation asymmetry could explain their early acquisition. More generally, our observation likely reflects a transition from scaffolding to children's active learning.

Keywords: language development; social interaction; conversation analysis; pragmatics

Introduction

Social routine words – e.g. *yes*, *no*, *hi*, *bye*, *okay*, and *thank you* – constitute a foundational set of actions for conducting social interactions. These constituted actions, e.g. greeting, agreeing, ratifying, and leaving, etc, allow us to, among other things, navigate joint commitments (Bangerter & Clark, 2003; Bangerter et al., 2022) – a normative attitude which organizes much of human sociality (Gilbert, 1992) and which begins to develop in 3- and 4-year-olds (Gräfenhain et al., 2009; Kachel et al., 2019; Rossano et al., 2022). Furthermore, these words are among 25% of the first 50 words children learn in English and 25% of the first 10 words across 15 other languages (Casey et al., 2023). While the normative understanding of these words is likely not present when first acquired, they are still used to coordinate with caregivers and likely play an important role in socializing children to become accountable social partners.

Despite this, social routine words have been neglected. As has been remarked, *"Because [social routine words] are performatives, and not referential at all, they are largely ignored in discussions of children's first words"* (Tomasello, 2003). As others have more recently argued, theories of first word learning may have overemphasized the pragmatics of reference, i.e. how a word is mapped onto a concept or event, ignoring alternative functions and contexts of word use (Rohlfing et al., 2016; Wojcik et al., 2022). Furthermore, methods for assessing early knowledge of words do not work for social routine words (Casey et al., 2023) because these methods assume these words operate under the pragmatics of reference. Instead, nearly all work on first words has focused on nouns and verbs (Tomasello, 1992; Waxman et al., 2013) with some notable exceptions (Bates et al., 1975; Bruner, 1975; Gopnik, 1982). There have been good reasons for this focus. Nouns and verbs are the more productive elements of languages, they vastly outnumber social routine words, and investigations into how children correctly infer words' referents despite the innumerable of alternatives (Quine, 1960; MacNamara, 1972) has shed light on children's sophisticated social cognitive abilities (Bloom, 2000; Tomasello et al., 1996; Tomasello & Farrar, 1986). And yet, a theory of word learning ought to account for the learning of all words.

In part 1 of this paper, we demonstrate that the acquisition of social routine words is not only early (Casey et al., 2023), it is poorly accounted for by current predictors of word acquisition. Using word frequency, concreteness, measures of sentence complexity, and other predictors, we show that a statistical model predicting words' age of acquisition (Braginsky et al., 2016), i.e. the age at which 50% of children produce a word, systematically overestimates when they are acquired. It predicts that social routine words ought to be acquired nearly 4 months later than they actually are – a considerable error given these words are acquired between 12-24mo. We believe this gap highlights the need for measurable explanations that do not depend on reference or syntax.

So where do referential explanations and the early acquisition of social routine words come apart? Early words like *mommy*, *dog*, *give* are acquired before the most frequent words in a language like articles such as *the* and copulas such as *is*. This is because early words refer to very concrete things and actions, meaning (1) they are also easy to label and (2) preverbal infants may already individuate the

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

concepts of *mommy*, *dog*, and *give* from other things and actions (Gentner & Boroditsky, 2001). Because social routine words are not referential, they are considered abstract according to concreteness norms from adult speakers (Brysbaert et al., 2014), so, we should predict that they are acquired later. However, we suspect that, while social routine words are evidently hard to label referentially (Casey et al., 2023), they are still easily made salient in joint activities and they are easy for infants to cognitively represent. This is to say, social routine words may not actually be so abstract from an interactional perspective.

We suggest social routine words constitute concrete actions within highly predictable & frequently recurring action sequences, and so, infants can readily individuate between the basic occasions of their use. Through what we call *sequence scaffolding*, caregivers make these action slots salient. In other words, they are parts of pragmatic frames (Rohlfing et al., 2016). We build the intuition and details of this hypothesis as follows.

First, conversational turn taking is an attractive phenomena because it plays a pivotal role in language development (Zimmerman et al., 2009). A growing body of evidence suggests that the amount of conversational turn-taking children engage in strongly predicts vocabulary growth as well as brain-related changes that mediate language development (Romeo et al., 2018). Turn-taking appears to explain-away the effects of overheard speech, refining our understanding of the 30-million word gap (Donnelly & Kidd, 2021). What is not evidenced, however, is how exactly conversational turn-taking facilitates this.

But we can glean insight into this from Conversation Analysis. Conversational turn-taking is commonly organized around a structure called an *adjacency pair*. An adjacency pair contains (1) an initiating turn called a *first pair part* (FPP), and (2) a responding turn called a *second pair part* (SPP) (Sacks & Schegloff, 1973). In some contexts, different speakers' words, actions, and cognitive work is asymmetrically distributed within the first and second position of adjacency pairs. Early child-caregiver conversations may be one of these contexts.

Second, social routine words are either stereotypically or frequently used in second position depending on the specific social routine. For example, *yes* and *no* are polar answers to questions (Raymond, 2003; Enfield et al., 2019). *Hi* and *bye* are responses to greetings and degreetings (Sacks & Schegloff, 1973). *Thank you* may often act, initially, as a completive speech act (Bruner, 1975), i.e. acknowledging completion of another's initiating action.

Third, we argue that second position turns afford more sequence scaffolding from caregivers than first pair parts (see Figure 1). Through recipient design (Sacks et al., 1974), caregivers may construct a first pair part so that its action type is more recognizable, e.g., an offer, and design it to mobilize a response (Stivers & Rossano, 2010), e.g. interrogative rising intonation. If this fails to elicit a response, caregivers can pursue a response from a child (Filipi, 2013) by repeating or redesigning their turn (see Table 1 line 4). Pursuits index absent action slots which help children learn the obligatory sequential structure of social routines. If a child succeeds in contributing a second pair part but the turn causes some interactional trouble, e.g. a misunderstanding or is unintelligible, a caregiver can initiate an other-initiated repair (Clark & Andersen, 1979; Clark, 2020). While other-initiated repair can fix trouble in both first and second pair parts, it is likely more effective when the caregiver initiates and the child responds. When caregivers initiate, they have an expected response in mind and therefore are more able to initiate a repair with a *restricted offer* than with less informative *open request*, e.g. *huh?* This means they can bear more cognitive and linguistic labor in the interaction (Dingemanse et al., 2015).

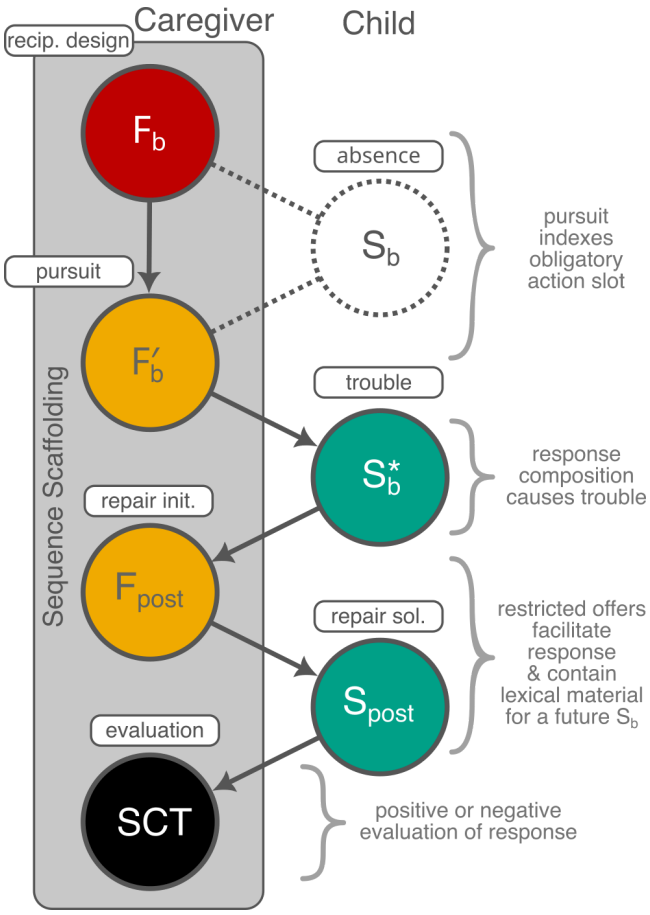

Figure 1: A schematic of sequence scaffolding. Through recipient design, the action type of a first pair part (F_b) is made to be recognizable and mobilize a response (S_b). A pursuit (F'_b) indexes missing action slots where a response (S_b) was expected. A restricted offer (F_{post}) may rework the problematic turn (S^*_b). Sequence closing thirds (SCT) – and post-expansions – may contain feedback on the design and composition of response (S_{post}).

William at 16.4 months | transcript 10412 @ 00:44:45			
	MOT and CHI are next to a swingset in their backyard. CHI runs over to a toddler swing seat.		
1	F_b*	CHI:	swing
2	F_{ins}	MOT:	swing?
3			(1.5 second gap)
4	F'_{ins}	MOT:	you wanna go on the swing?
5	S_{ins}*	CHI:	yeh ((CHI pushes swing))
6	F_{post}	MOT:	d'you wanna push it?
7	S_{post}	CHI:	up ((raises arms up to MOT))
8	F_{post}	MOT:	up?
9	S_b	MOT:	okay
	MOT picks up CHI, puts him on the swing, and praises his use of "up" to ask to be swinged.		

Table 1: A request→accept sequence. See the Appendix for transcript conventions. In this sequence the base pair parts (lines 1 & 9) are separated by 6 turns. Lines 2-5 is an instance of other-initiated repair which clarifies CHI's request on line 1. This repair side-sequence is considered an *insert-expansion* between the F_b and S_b. It is followed by – or *post-expanded* by – a second other initiated repair side-sequence on lines 6-7, likely because of the discord between CHI's action (pushing the swing) and MOT's candidate understanding (Schegloff, 1996) of his request on line 4 (that he wants to be pushed on the swing). Note that while William initiates the base action in this sequence, most of his contributions are responses.

In sequence closings, caregivers often evaluate the quality of second pair parts, as is the case in initiation-response-evaluation (IRE) sequences which are paradigmatic of pedagogical talk (Mehan, 1979).

Furthermore, second position turns are an important aspect of children's interactional competence. Even if a child's turns are linguistically well-formed, these turns may still appear "child-like", as Stivers et al. (2018) found in 4- and 8-year-olds' responses to polar questions. These responses tended to violate conversational norms like preference organization (Pomerantz, 1984), e.g. avoiding offering disaffiliative rejections like *no* by waiting for one's partner to self-reject or mitigating rejections with accounts.

Our hypothesis requires that children's early contributions to conversation occur in second position, but prior work on adjacency pairs in child conversations suggests that the opposite is the case. Bloom et al. (1996) found that child→mother adjacency pairs are more common than mother→child adjacency pairs over the first year of life. However, in this study adjacency pairs were identified by the sequential position of speakers' turns sampled within 2 second windows – similar to the 5 second window commonly used in LENA software (Gilkerson & Richards, 2008). We question the validity of constructing adjacency pairs this way.

On the one hand, gaps between turns are remarkably small: 200ms on average (Stivers et al., 2009; Levinson, 2016). Gaps are meaningful (Sacks et al., 1974; Pomerantz, 1984; Kendrick & Torreira, 2015; Gates et al., 2021). And the timing of children's turns are slower than that of adults, vary considerably between certain response types, and may even slow down during the early language development (Casillas et al., 2016; Hilbrink et al., 2015; Nguyen et al., 2022) with slowdowns found even in 8-year-olds (Stivers et al., 2018). Therefore, because children respond more slowly, it is likely that mother→child adjacency pairs are undersampled using a temporal window method.

On the other hand, gap duration is not a joint along which adjacency pairs and conversations are carved. As originally observed by Schegloff & Sacks, (1973), adjacency pairs are composed of two turns, by different speakers, relatively ordered (so that the first initiates and the second responds), and pair-type[1] related (so that responses relevantly fit with the *social action* of initiating actions) Schegloff, (2007). The most basic *base adjacency pairs*, denoted with subscripts $F_b \rightarrow S_b$, accomplish some main social action. But these two turns need not occur immediately one after another. Adjacency pairs may be nested within each other, e.g. the inserted repair sequence in Table 1. This arrangement of turns may produce sequences such as $[F_b[F_{ins},S_{ins}]S_b]$, which contains two pairs: $F_b \rightarrow S_b$ and $F_{ins} \rightarrow S_{ins}$. A gap based method would instead misrepresent this sequence as containing three pairs: $F_b \rightarrow F_{ins}$, $F_{ins} \rightarrow S_{ins}$, and $S_{ins} \rightarrow S_b$, only one of which is correct. More generally, F_b's and S_b's can be separated by intervening talk. Therefore, temporal window methods, in general, cannot identify adjacency pairs in common forms of conversation.

To summarise, we theorize that second position facilitates word learning because it is more predictable and affords more caregiver scaffolding than first position. Words which are frequently used (or are expected, by caregivers, to be used by children) to constitute SPPs are more likely to be acquired early. Because social routine words are often used in second position, this could address their early acquisition. A first step toward the theory is showing that children's first contributions to talk primarily occur in second position.

To test our hypothesis we analyze naturalistic interactions between children and their caregivers. Finding evidence of initiation asymmetry would suggest that the sequence scaffolding affordances of second pair parts is a plausible explanation for the early acquisition of social routine words.

Methods

Modeling age of acquisition

In order to examine how well predictors of word learning explain the acquisition of social routine words, we adapted a statistical model from Braginsky et al. (2016). Borrowing predictors from their data, we fit a linear model to predict age of acquisition from frequency, concreteness, valence, babiness, mean length of utterance. However, unlike Braginsky et al., we do not include intercepts or interactions for lexical categories. The reasoning for this is two-fold.

[1] E.g. question→answer, greet→greet, offer→accept

First, in CDIs and Wordbank, social routine words do not appear to belong to a proper lexical category. They have been slotted into the misfit category of *other*, which contains other early acquired but unrelated words like kinship & people terms (mommy, boy, etc.), onomatopoeia (woof woof, vroom, etc.), as well as spatial-temporal items (outside, night, etc.). Second, our intention is to avoid a social routine specific predictor with no theoretical basis, e.g. a special variable which applies only to one lexical category. For example, adding a special social routine word variable would remove any bias, but would offer no cognitive explanation. We aim for a predictor of all words.

Longitudinal corpus

To test the hypothesis that 1-year olds respond rather than initiate in adjacency pairs, we analyzed longitudinal video data from the CHILDES Archive of five North American English speaking children from the Providence Corpus (Demuth et al., 2006). We included Alex, William, Lily, Naima, and Violet but excluded Ethan for whom no video data is available. This corpus contains videos of children between 11 and 48 months collected every few weeks. Recordings are, on average, one hour in length and capture spontaneous caregiver-child interactions at home. Caregivers and children both wore wireless lavalier microphones and were captured using a stationary camera caregivers positioned somewhere in a room when both the caregiver and child were co-present.

Because mothers were by far the most frequently recorded caregiver for all target children in the corpus, we focus our analysis on interactions of mother-child dyads.

To collect a representative sample of conversational sequences, we generated a stratified sample of 1,442 utterances in the corpus. We first grouped the data by 3-month bins (from 12-28), by target child, and by speaker type (child or mother) and then sampled up to 25 utterances from each grouping. 25 utterances were not available for all groupings.

We coded this data using a bespoke browser-based annotation[2]. This tool is compatible with ELAN files. We selected it simply because it allowed us faster transcript navigation and to link directly to sampled turns.

We downloaded the original Providence CHAT files from talkbank.org and converted these to an ELAN format using pympi, a python package for processing ELAN & Praat files. We patched pympi in order to process an old CHAT incompatibility found in the Providence files.

Coding Adjacency Pairs & Sequence Organization

For every turn in our stratified sample of the Providence Corpus, we first identified the entire conversational sequence it belonged to. A sequence could be a simple base adjacency pair containing two turns. It could contain just a single unreciprocated turn. Or it could contain several adjacency pairs. Identifying the boundaries of a conversational sequence is done by first determining the main social action being performed which involves locating the base adjacency pair $F_b \rightarrow S_b$ which could perform a request→accept, offer→reject, inform→acknowledge, etc. But other adjacency pairs or turns could be involved in the performance of this main social action.

These could include adjacency pairs that occur in expansions, namely: (1) pre-expansions, e,g. bids for attention or call→response pairs such as "Mommy"→"Yes honey?", which establishes joint attention in service to a main action, (2) insert-expansions, e.g. other-initiated repair, as in Table 1, or requests for pertinent information needed before the response can be properly made, and (3) post-expansions, e.g. other initiated-repair, closings, or evaluation, which occur after another adjacency pair. For a more complete treatment of expansions, see Kendrick et al. (2020) and Schegloff (2007a).

Then, for every turn within a sequence, whether it was an FPP or SPP. But, sequences can be closed with minimal post-expansion turns called sequence closing thirds (SCT) as in Table 2. Importantly, because these turns, unlike FPPs, do not obligate a response, we did not count SCT as either first or second position for our analysis.

Lily at 23.8 months | transcript 011126 @ 00:09:01			
	MOT & CHI are playing with a kitty cat made of felt		
1	┌F_b	MOT:	yuh want the kitty cat tuh sit
2	│		on the couch?
3	└S_b	CHI:	yeah
4	SCT	MOT:	okay
	MOT puts the felt kitty cat on a felt cutout of a couch		

Table 2: An offer→accept sequence with an SCT. This post-expansion ratifies the activity with "okay".

Additionally, we coded whether turns were pursuits, e.g. if a speaker contributed a turn which repeated the same action type as their previous turn with a one second gap (which could be shortened with due judgement). These were counted as distinct turns.

Occasionally, existing utterance boundaries in the Providence Corpus were misaligned with conversational turn boundaries. This occurred two different ways. A single Providence utterance could span multiple turns from an individual. In this case, we manually cut the utterance into its constituent turns. Or multiple Providence utterances could, together, constitute a single turn. In this case, we preserved the original Providence utterances, but marked these as constituting one turn in order to avoid overcounting them.

While transcripts in the Providence Corpus do not include gestures, we tried to account for them in our coding of conversational sequences. For co-speech gestures, e.g. when a child vocalizes while pointing, the gesture was treated as part of a turn's design. For gestures without accompanying speech, we accounted for their position in the surrounding verbal turns, but did not add gestures to the transcripts. So,

[2] https://github.com/jackft/beholder-tool

for example, if a child pointed at a cup and a mother labeled it, the mother's labeling was counted as a second pair part, although the child's silent point was not explicitly labeled.

We discarded utterances if they were not a part of a mother-child interaction. These types of utterances were not obviously directed at one's social partner, e.g. a mother talking on the phone to another adult or solo-babbling. We also discarded utterances if they turned out to be non-conversational. These types of utterances were either from non-turn-taking formats of interaction or did not constitute social actions, e.g. an unbroken monologue while reading a book or an involuntary burp from a child.

We assessed reliability for coding adjacency pairs using Fleiss's kappa. Three coders (JT and two research assistants) achieved substantial agreement (κ=.73).

Part 1: The puzzle of social routine words: evidence of a selective gap in science of Word learning

To see whether statistical models overestimate the age of acquisition (AoA) of social routine words, we investigated the residual between actual age of acquisition and model predictions. As shown in the actual-vs.-predicted AoA plot in Figure 2, social routine words are acquired earlier than the model predicts. On average they were acquired 3.72 months earlier than predicted.

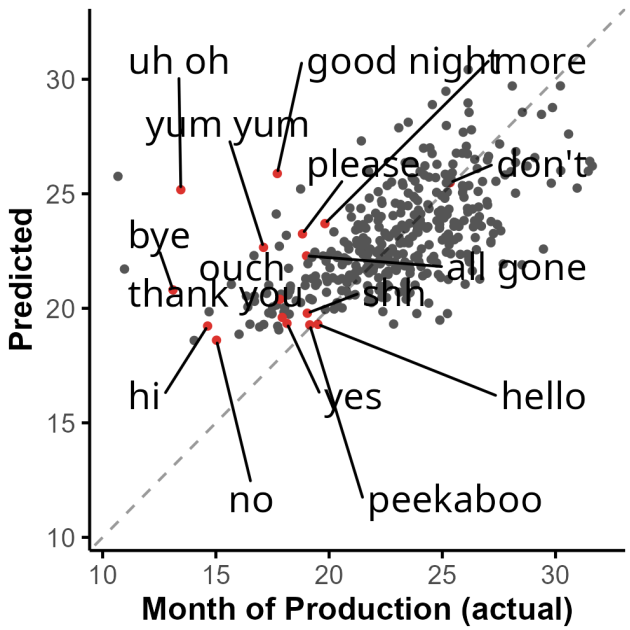

Figure 2: A scatter plot of actual vs. model-predicted age of acquisition. Each point represents a word. Social routine words are colored red whereas others are colored black. Words acquired earlier than the model predicts lie above the dashed line. Words acquired later than the model predicts lie below the dashed line.

By modeling residuals with a linear model and testing whether the marginal mean residual was non-zero, we confirmed that the residual for social routine words was significantly biased ($e_{routine}$=3.72±1.71, p=.031). But, residuals were not biased for verbs and nouns (e_{noun}=0.04±0.45, p=0.91; e_{verb}0.02±0.66, p=.97). Among words acquired before 24 months, social routine words are less concrete than other early words (t=-8.132, p<.001).

Part 2: Initiation asymmetry in caregivers & 1-year old's conversational contributions

To see whether conversations between caregivers and children exhibit initiation asymmetry, we measured the probability a speaker's turn was an FPP or SPP in an adjacency pair (see Figure 3). To test for initiation asymmetry, we jointly modeled the probability that children and parents contribute turns in second position by fitting a generalized linear mixed effects model (gLMM) with a Bernoulli link function (where FPP=0, SPP=1) with fixed effects for age, speaker type (CHI or MOT), and an interaction between age & speaker type. Our model accounts for two hierarchical structures of our data. First, it includes by-dyad random intercepts and random slopes (for

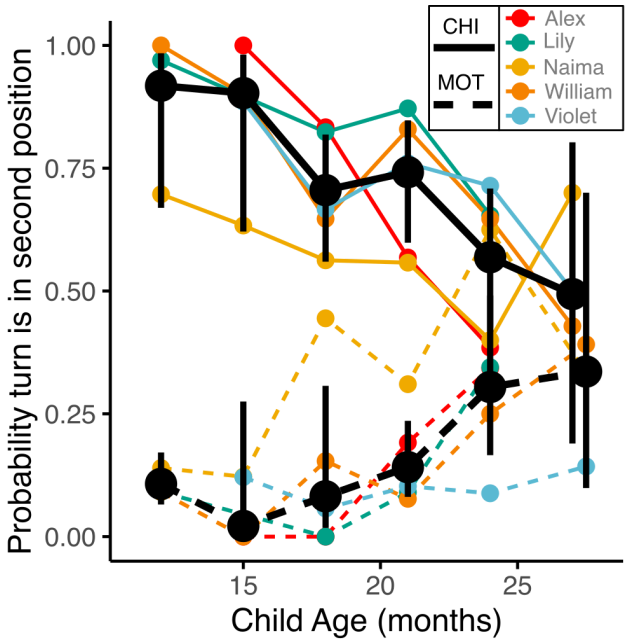

Figure 3: The distribution of children's and mothers' turns within adjacency pairs over development. Solid lines are children; dotted lines are mothers; black dots & lines are estimated marginal means with 95% CIs; and colored dots are per-dyad empirical means. At 12 months, children tend to respond (second pair part turns) to their mother's initiating turns (first pair part turns). Over the second year of life, children's and mother's contributions become more similarly distributed within adjacency pairs. EMMs at 27mo are offset to prevent overlapping confidence intervals.

age, speaker type, and their interaction). For example, this allows the Naima dyad to be more symmetric than others and to be more stable over time. Second, it includes by-sequence random intercepts which affords variation between sequences, e.g. sequences where a mother contributes many FPPs in pursuit of an absent SPP or where a mother contributes many FPPs repairing a child's FPP.

Over the second year of life, children's but not mothers' contributions to talk are most likely to occur in second position. We verified this by testing whether the estimated marginal mean (EMM) probability of SPP turns was greater than that of FPPs (EMM_{CHI}=0.76±0.05, p<.001; EMM_{MOT}=0.13±0.04, p=1.000). Their contributions are particularly asymmetric when children are 12 months, when children begin to utter their first words, (EMM_{CHI}=0.92±0.66, p<0.001; EMM_{MOT}=0.11±0.03, p=1.000). We note however, that there is considerable variation between dyads. For example, there is much less initiation asymmetry in Naima's dyad than in others.

Fitting a model that treats age as a continuous variable, we find a significant effect of age on turn position (β_{age}=-0.11±0.04, p=.002) and a significant interaction between age and speaker type ($\beta_{age\&MOT}$=0.24±0.06, p<.001). A post-hoc estimated marginal mean analysis of this interaction revealed that mothers' turns were more likely to be uttered in second position over child age (β_{age}=0.13±0.04, p=.001). Together these tests verify that children's and mothers' contributions become more symmetric over the second year of life.

Discussion

Our results provide evidence of initiation asymmetry in caregiver-child dyads, in the second year of life, contrary to the claim that they are the primary initiators of adjacency pairs in conversation (Bloom et al., 1996). We argue that measuring adjacency pairs in terms of the conditional relevance of the constituent turns is a promising method a) to explain why turn taking is so relevant for language acquisition broadly and b) for application to pragmatic frame theory. Taken together, our results suggest it is plausible that second position turns are an important site in children's early word learning via sequence scaffolding, which we argued is especially important for the acquisition of social routine words.

More broadly, our results are compatible with both scaffolding and active learning theories of word learning as children begin to initiate adjacency pairs more with age. We found the degree of initiation asymmetry decreases substantially from 12 to 28 months as children's conversational turns shift from primarily made in response to caregivers to a balance of initiation and response.

Here, we note that there are case studies of the shift from responding to initiating in specific routines. In a longitudinal study of peekaboo[3], in one mother-child dyad, Bruner & Sherwood (1975) found that prior to 15 months, reappearances in the game were controlled by the mother, but at 15 months, the child invented their own variant of the game in which she initiated the reappearance.

A homologous shift in initiation asymmetry appears to occur in other great apes despite relying on a different social learning mechanism, i.e. ontogenetic ritualization. One recurring social routine in bonobos (Pan paniscus) is the mother-infant carry, whereby an infant comes to be attached to the other for the purpose of joint travel. Both mother and infant can initiate a carry and use learned idiosyncratic repertoires of gestures to do so. Halina et al. (2013) found that, between 10mo-25mo, carries shifted from being mainly initiated by mothers to being mainly initiated by infants. Taken together, these studies may be evidence of a general ontogenetic principle for how signaling systems emerge between infants and caregivers through social feedback for the purpose of coordination.

One reason for initiation asymmetry may arise in the first place is as a distributed cognitive strategy. Using such a strategy, cognitive and linguistic work can be offloaded from children onto caregivers and the interaction itself. For example, van Arkel et al. (2020) showed that other-initiated repair can alleviate the need for costly pragmatic inferences. In restricted offers, the speaker in second position only needs to utter *yes*[4] or *uh huh* (Dingemanse et al., 2015) (see Table 1 lines 4 & 5 for an example).

However, our current study is limited only to the distribution of talk by different speakers in adjacency pairs. To further support our sequence scaffolding theory, future work should look for direct evidence of sequence scaffolding strategies, e.g. pursuits and other initiated repair. We would also need to find empirical evidence that early social routine word uses tend to occur in second position. Such future work would help elucidate not just how non-referential words are acquired but might also further our understanding of the role of turn-taking in language acquisition.

Appendix

Sequence Organization Symbols

Convention	Description
F/S	First/second pair part
F_b/S_b	Base first/second pair part
F_{ins}/S_{ins}	First/second pair part of an insert-expansion
F_{post}/S_{post}	First/second pair part of a post-expansion
SCT	Sequence closing third
F′	First pair-part in pursuit of response
*	Turn contains trouble
⊏	Membership in the same adjacency pair

[3] Unlike other social routine words, peekaboo initiates, but the game broadly follows the attested pattern of initiation asymmetry.

[4] See Goodwin (2006) on communicative strategies an aphasic speaker used with only the words *yes*, *no*, and *and*.

Acknowledgments

We thank Samantha Lo and Carmen He for their assistance annotating data. We also thank Julia Di Silvestri, Josh Clarke, and Hunter Flores piloting our coding scheme.

References

Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. Cognitive Science, 27(2), 195–225.

Bangerter, A., Genty, E., Heesen, R., Rossano, F., & Zuberbühler, K. (2022). Every product needs a process: Unpacking joint commitment as a process across species. Philosophical Transactions of the Royal Society B: Biological Sciences, 377(1859), 20210095.

Bates, E., Camaioni, L., & Volterra, V. (1975). The Acquisition of Performatives Prior to Speech. Merrill-Palmer Quarterly of Behavior and Development, 21(3), 205–226.

Bloom, L., Margulis, C., Tinker, E., & Fujita, N. (1996). Early Conversations and Word Learning: Contributions from Child and Adult. Child Development, 67(6), 3154–3175.

Bloom, P. (2000). How children learn the meanings of words (pp. xii, 300). The MIT Press.

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2016). From uh-oh to tomorrow Predicting age of acquisition for early words across languages. CogSci, 6.

Bruner, J. S. (1975). The ontogenesis of speech acts*. Journal of Child Language, 2(1), 1–19.

Bruner, J. S., & Sherwood, V. (1975). Peekaboo and the learning of rule structures. Play : Its Role in Development and Evolution, 277–285.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. Behavior Research Methods, 46(3), 904–911.

Casey, K., Potter, C. E., Lew-Williams, C., & Wojcik, E. H. (2023). Moving beyond "nouns in the lab": Using naturalistic data to understand why infants' first words include uh-oh and hi. Developmental Psychology, 59(11), 2162–2173.

Casillas, M., Bobb, S. C., & Clark, E. V. (2016). Turn-taking, timing, and planning in early language acquisition. Journal of Child Language, 43(6), 1310–1337.

Clark, E. V. (2020). Conversational Repair and the Acquisition of Language. Discourse Processes.

Clark, E. V., & Andersen, E. S. (1979). Spontaneous Repairs: Awareness in the Process of Acquiring Language. Papers and Reports on Child Language Development, No. 16. PRCLD, Department of Linguistics, Stanford University, Stanford, CA 94305.

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, Epenthesis and Coda Licensing in the Early Acquisition of English. Language and Speech, 49(2), 137–173.

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladottir, R. S., Kendrick, K. H., Levinson, S. C., Manrique, E., Rossi, G., & Enfield, N. J. (2015). Universal Principles in the Repair of Communication Problems. PLOS ONE, 10(9), e0136100.

Donnelly, S., & Kidd, E. (2021). The Longitudinal Relationship Between Conversational Turn-Taking and Vocabulary Growth in Early Language Development. Child Development, 92(2), 609–625.

Enfield, N. J., Stivers, T., Brown, P., Englert, C., Harjunpää, K., Hayashi, M., Heinemann, T., Hoymann, G., Keisanen, T., Rauniomaa, M., Raymond, C. W., Rossano, F., Yoon, K.-E., Zwitserlood, I., & Levinson, S. C. (2019). Polar answers. Journal of Linguistics, 55(2), 277–304.

Filipi, A. (2013). Withholding and pursuit in the development of skills in interaction and language. Interaction Studies, 14(2), 139–159.

Gates, V., Callaway, F., Ho, M. K., & Griffiths, T. L. (2021). A rational model of people's inferences about others' preferences based on response times. Cognition, 217, 104885.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word. Language, Culture and Cognition, 3, 215–256.

Gilbert, M. (1992). On Social Facts. Princeton University Press.

Gilkerson, J., & Richards, J. A. (2008). The LENA natural language study. Boulder, CO: LENA Foundation. Retrieved March, 3, 2009.

Goodwin, C. (2006). Human Sociality as Mutual Orientation in a Rich Interactive Environment: Multimodal Utterances and Pointing in Aphasia. In Roots of Human Sociality. Routledge.

Gopnik, A. (1982). Words and plans: Early language and the development of intelligent action. Journal of Child Language, 9(2), 303–318.

Gräfenhain, M., Behne, T., Carpenter, M., & Tomasello, M. (2009). Young children's understanding of joint commitments. Developmental Psychology, 45(5), 1430–1443.

Kachel, U., Svetlova, M., & Tomasello, M. (2019). Three- and 5-year-old children's understanding of how to dissolve a joint commitment. Journal of Experimental Child Psychology, 184, 34–47.

Kendrick, K. H., Brown, P., Dingemanse, M., Floyd, S., Gipper, S., Hayano, K., Hoey, E., Hoymann, G., Manrique, E., Rossi, G., & Levinson, S. C. (2020). Sequence organization: A universal infrastructure for social action. Journal of Pragmatics, 168, 119–138.

Kendrick, K. H., & Torreira, F. (2015). The Timing and Construction of Preference: A Quantitative Study. Discourse Processes, 52(4), 255–289.

Levinson, S. C. (2016). Turn-taking in Human Communication – Origins and Implications for Language Processing. Trends in Cognitive Sciences, 20(1), 6–14.

MacNamara, J. (1972). Cognitive basis of language learning in infants. Psychological Review, 79, 1–13.

Mehan, H. (1979). “What Time Is It, Denise?”: Asking Known Information Questions in Classroom Discourse. Theory Into Practice, 18(4), 285–294.

Nguyen, V., Versyp, O., Cox, C., & Fusaroli, R. (2022). A systematic review and Bayesian meta-analysis of the development of turn taking in adult–child vocal interactions. Child Development, 93(4), 1181–1200.

Pomerantz, A. (1985). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In J. M. Atkinson (Ed.), Structures of Social Action (pp. 57–101). Cambridge University Press.

Quine, W. V. O. (1960). Word and Object. MIT Press.

Rohlfing, K. J., Wrede, B., Vollmer, A.-L., & Oudeyer, P.-Y. (2016). An Alternative to Mapping a Word onto a Concept in Language Acquisition: Pragmatic Frames. Frontiers in Psychology, 7.

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. E. (2018). Beyond the 30-Million-Word Gap: Children’s Conversational Exposure Is Associated With Language-Related Brain Function. Psychological Science, 29(5), 700–710.

Rossano, F., Terwilliger, J., Bangerter, A., Genty, E., Heesen, R., & Zuberbühler, K. (2022). How 2- and 4-year-old children coordinate social interactions with peers. Philosophical Transactions of the Royal Society B: Biological Sciences, 377(1859), 20210100.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. In Studies in the organization of conversational interaction (pp. 7–55). Elsevier.

Schegloff, E. A. (2007a). Sequence Organization in Interaction. Cambridge University Press.

Schegloff, E. A. (Ed.). (2007b). The adjacency pair as the unit for sequence construction. In Sequence Organization in Interaction: A Primer in Conversation Analysis (Vol. 1, pp. 13–21). Cambridge University Press.

Schegloff, E. A., & Sacks, H. (1973). Opening up Closings. 8(4), 289–327.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. Proceedings of the National Academy of Sciences, 106(26), 10587–10592.

Stivers, T., & Rossano, F. (2010). Mobilizing response. Research on Language and Social Interaction, 43, 3–31.

Tomasello, M. (1992). First Verbs: A Case Study of Early Grammatical Development. Cambridge University Press.

Tomasello, M., & Farrar, M. J. (1986). Joint Attention and Early Language. Child Development, 57(6), 1454–1463.

Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. Journal of Child Language, 23(1), 157–176.

van Arkel, J., Woensdregt, M., Dingemanse, M., & Blokpoel, M. (2020). A simple repair mechanism can alleviate computational demands of pragmatic reasoning: Simulations and complexity analysis. In R. Fernández & T. Linzen (Eds.), Proceedings of the 24th Conference on Computational Natural Language Learning (pp. 177–194). Association for Computational Linguistics.

Waxman, S., Fu, X., Arunachalam, S., Leddon, E., Geraghty, K., & Song, H. (2013). Are Nouns Learned Before Verbs? Infants Provide Insight into a Longstanding Debate. Child Development Perspectives, 7(3), 10.1111/cdep.12032.

Wojcik, E. H., Zettersten, M., & Benitez, V. L. (2022). The map trap: Why and how word learning research should move beyond mapping. WIREs Cognitive Science, 13(4), e1596.

Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S., & Yapanel, U. (2009). Teaching by Listening: The Importance of Adult-Child Conversations to Language Development. Pediatrics, 124(1), 342–349.