Learning communicative acts in children's conversations: a Hidden Topic Markov Model analysis of the CHILDES corpus

Claire Bergey[1], Zoe Marshall[2], Simon DeDeo[2,3], & Dan Yurovsky[2]

[1] University of Chicago

[2] Carnegie Mellon University

[3] Santa Fe Institute

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

Correspondence concerning this article should be addressed to Claire Bergey, Postal address. E-mail: cbergey@uchicago.edu

Abstract

Over their first years of life, children learn not just the words of their native languages, but how to use them to communicate. Because manual annotation of communicative intent does not scale to large corpora, our understanding of communicative act development is limited to case studies of a few children at a few time points. We present an approach to automatic identification of communicative acts using a Hidden Topic Markov Model, applying it to the CHILDES database. We first describe qualitative changes in parent-child communication over development, and then use our method to demonstrate two large-scale features of communicative development: (1) children develop a parent-like repertoire of our model's communicative acts rapidly, their learning rate peaking around 14 months of age, and (2) this period of steep repertoire change coincides with the highest predictability between parents' acts and children's, suggesting that structured interactions play a role in learning to communicate.

Keywords: language acquisition, corpus analysis, computational modeling, pragmatics

Word count: X

Learning communicative acts in children's conversations: a Hidden Topic Markov Model analysis of the CHILDES corpus

From their first utterances, children are not just producing language but *using* it to communicate. A child who can produce only one-word utterances can nonetheless convey several communicative intentions: using variations in pitch, she can use the word *mama* to identify a person, question possession of an object, or to call for someone's presence (Dore, 1975). From 14 to 30 months of age, children quickly branch out from communicative acts like *requesting*, *protesting*, and *marking an event* to *agreeing to an action*, *stating intent*, and asking and answering a variety of questions (Snow, Pan, Imbens-Bailey, & Herman, 1996). Close studies of children's conversations, using rich observations to infer intended meaning in context, show that much of the action of language acquisition happens at the level of what children *mean* to say.

Despite the centrality of communicative goals to even infants' comprehension and production of language (Vouloumanos, Onishi, & Pogue, 2012), nearly all of our formal models of language focus on acquisition of words or syntactic categories rather than communicative expressiveness (e.g. Johnson, Griffiths, & Goldwater, 2007; Siskind, 1996; although c.f. Bohn & Frank, 2019). This state of affairs is partly due to the fact that most large-scale data on language acquisition concern the ages at which children comprehend and produce words or correctly inflect those words according to the grammar of their language (Bergmann et al., 2018; Frank, Braginsky, Yurovsky, & Marchman, 2017). We lack a quantitative description of the trajectory of children's communicative capacities, or how they bring these to bear in interactions with their parents and peers.

Describing children's communicative acts on a large scale, however, is a challenging task. Without nuanced, context-sensitive human coding, communicative acts can be hard to identify. Words are amenable to identification, storage, and tabulation using common computational tools; perhaps due to their ease of use, models of language development have often approached language development at the level of words (i.e., vocabulary learning). The goals and intentions underlying those words are less amenable to such manipulation. We present an approach to modeling children's communicative acts, working backwards from the words they produce: we model communicative acts as the latent sources from which words emerge, and characterize children's engagement in these acts across development.

Traditionally, studies of communicative acts among children have brought frameworks from adult communication, such as Speech Act theory (Austin, 1962; Searle, 1969) and Conversation Analysis (Sacks, Schegloff, & Jefferson, 1974), to bear on children's conversations. While these systems can be useful to characterize children's conversations, they may not be the most relevant distinctions in children's communication. Studying children's communicative acts using an unsupervised computational model allows us to extract communicative patterns across many children with less a priori specification of what those patterns should be.

Here, we characterize children's growing repertoire of communicative acts using a Hidden Topic Markov Model. This model observes utterances produced by parents and children and attempts to infer common underlying processes–topics–that produced them. We first show that this model extracts several communicative acts analogous to those observed in close case studies of children's communication without specifying them top-down. We then show that use of these acts has a developmental trajectory in line with those studies: children's act usage quickly proliferates, growing in diversity most quickly at around 14 months and reaching parent-like diversity by 24 months. Further, these acts have distinct developmental trajectories that are in line with prior research. Finally, we show that parents and children engage in these acts contingently within conversations, and that this sequential contingency peaks around the same age that children are expanding their act use most rapidly.

Method

Corpus

We used transcripts of conversations from the Child Language Data Exchange System (CHILDES), a database of child conversation corpora (MacWhinney, 2000). These corpora predominantly record spontaneous conversations between children and their family members, often in the home. We trained our model on transcripts from the North American English collection of CHILDES among children under 60 months old, and filtered these transcripts to include only utterances spoken by the target child or their parents. To exclude transcripts with extremely sparse child utterances, our analyses include children 6 to 60 months old, and transcripts with at least 10 child utterances and 10 parent utterances. Overall, our training data included 4043 transcripts from 411 children, and here we analyze 3016 transcripts from 372 children.

1	mental states:	{I, you, what, know, don't, it, do, that, did, think, like}
2	labeling:	{a, that, what's, is, this, that's, what, the, look, who's, baby}
3	counting:	{two, one, dis, three, four, de, dat, five, duh, six, eight}
4	evaluation:	{not, it's, good, i'm, you're, a, that's, very, be, he, he's}
5	proposed actions:	{you, it, me, to, can, want, okay, I, up, get, come}
6	non-present events:	{to, we, go, and, you, going, did, when, at, school, the}
7	miscellaneous:	{of, they, we, have, like, the, i, and, them, are, all}
8	requests:	{some, you, want, i, more, have, eat, milk, juice, your, drink}
9	knowledge questions:	{you, what, are, do, look, see, with, at, play, those, doing}
10	location:	{the, in, there, put, here, it, go, on, right, this, over}
11	social routines:	{mommy, daddy, hi, down, baby, bye, ball, n, tape, where's, where}
12	backchannels:	{yeah, no, oh, okay, hm, uh, huh, ah, mhm, mm, yes}
13	description:	{a, one, this, is, that's, blue, red, green, make, big, yellow}
14	storytelling:	{the, and, he, was, his, said, in, she, of, to, they}
15	body routines:	{your, on, my, put, his, her, hair, off, in, head, mouth}

Figure 1. The most distinctive words in each topic. Distinctiveness is measured by the difference between a word's probability in a given topic and its average probability across all topics.

Model

We used a Hidden Topic Markov Model (HTMM) to extract communicative acts from parent—child conversations (Gruber, Weiss, & Rosen-Zvi, 2007). Topic models represent documents as mixtures of topics, and topics as mixtures of words. For instance, a simple topic model trained on news articles may extract a topic whose distinctive words are *fire*, *flood*, and *aid* and another whose distinctive words are *speech*, *legislation*, and *administration*. Based on its distribution of words, an article about politicians' provision of disaster relief may be correctly inferred to feature these two topics, among others. Intuitively, the goal of a topic model is to recover the underlying sources—topics—from which the words in a document spring.

In contrast to a standard topic model, the Hidden Topic Markov Model takes into account the sequential utterance structure of a document, not just its static distribution of words. The HTMM assumes that words within an utterance are of the same topic, and that sequential utterances may be more likely to be of the same topic. It represents topic transitions between utterances in a coarse-grained way: either switch or stay. Gruber et al. (2007) developed this model and used it to segment machine learning conference papers, showing that the model can distinguish instances of the word "support" in mathematical contexts (describing support vectors) from those in the context of acknowledgments.

We trained the HTMM on all the utterances in our corpus. Some markers for unintelligible or non-word speech were removed; when this resulted in empty utterances, a 'non-word' token was included to preserve the temporal structure of the dialogue. Transcribed babbling (e.g., "awaoo") was included, and words that were transcribed as compounds or contractions due to pronunciation (e.g., "gimme") or idiosyncratic transcription standards (e.g., "thank_you") were retained in the corpus as written. Typically, function words are removed from corpora before training topic models to aid detection of thematic content. Here, we aim to classify communicative acts underlying utterances rather than thematic topics. We expect function words to be highly diagnostic of these acts, so we included them in our training data.

Topic models require pre-specification of the number of topics. To determine the right number of topics, we trained the model several times with different numbers of topics—5 to 30 topics, in intervals of 5—with Dirichlet parameters of $\alpha = 1/k$, where k is the number of topics, and $\beta = 0.01$. Each model produced a sequence of the most likely topic assigned to each utterance. Our selection metric was the proportion of other-topic transitions in this sequence: since we aimed to characterize the temporal structure of topic transitions, we wanted to choose a model that had many transitions between topics rather than long stretches of utterances all assigned to the same topic. However, increasing the number of topics almost necessarily increases the number of other-topic transitions, and may make results harder to interpret as topics proliferate; therefore, we balanced the proportion of other-topic transitions against the number of topics. Plotting this proportion across number of topics, we judged 15 topics as an inflection point after which increasing the number of topics had diminishing effect on other-topic transitions; we therefore chose the 15-topic model.

To check that the 15-topic model was not capturing drastically different topic distinctions from other runs of the model, we calculated the mutual information between utterance-level topic classifications from the 15-topic model against the 10-topic and 20-topic models. These mutual information values were relatively high: between the 15- and 10-topic models, mutual information was 1.53 (upper bound: entropy of 10-topic model, 2.25) and between the 15- and 20-topic models, mutual information was 1.84 (upper bound: entropy of 15-topic model, 2.63). Thus, the model we chose captures similar information to alternative runs of the model, and is unlikely to represent a one-off set of distinctions.

Results and Discussion

We begin by showing that the model captures some aspects of communicative acts and exploring the static structure of these utterance types. We then examine trajectories of topic use across development among parents and children. Finally, we examine the temporal dynamics of topic use within discourse.

Part 1: The structure of communicative acts

After training, the model produces a set of topics with associated probability distributions over words. One can conceive of these topics as bags of words, in which some words will be highly likely to be produced and others will be unlikely to be produced. Using this probability distribution of words within topics, the model also produces a probability distribution over topics for each utterance in the corpus. Since we aim to characterize communicative acts and not thematic content or conversational topics, the label 'topic' for these types can be misleading; from here on, we will refer to these types as communicative acts rather than topics.

The most distinctive words of each communicative act in the model, as measured by the difference between a word's probability within a type of act and its average probability across all act types, are shown in Figure 1. Assigning labels to these acts is subjective. We examined these most distinctive words and looked at samples of utterances in each act type to produce labels that capture the gist of each automatically identified topic, for clarity of exposition.

Not all communicative act categories are equally coherent, but even these top distinctive words are illuminating. These distinctive words are largely function words and other very common words, not content-related words. This suggests that the model is picking up on distinctions between utterances that are not just conversational topics or themes. Further, several act types have fairly coherent sets of words, such as backchannels and interjections (e.g., “okay”, “mm”, “no”, “huh”), requests (e.g., “you”, “want”, “have”, “some”), and counting (e.g., “two”, “one”, “three”, “dis”). On the other hand, these types clearly do capture some aspects of thematic content: the request type has several food-related words, and while the body routines type (e.g., “on”, “put”, “head”, “off”) contains many commands and statements about current activities, it is united by words about the body. These types are likely capturing a mixture of semantic content and communicative function, which are likely quite correlated in children’s language environments.

Some of the act types align nicely with speech act or interchange types classified in prior research (Ninio, Snow, Pan, & Rollins, 1994). For instance, the labeling and attention type resembles the *directing hearer’s attention* interchange type, the non-present events type resembles *discussing the non-present*, and the proposed actions type resembles *negotiating the immediate activity*; further, the evaluations type resembles the *evaluations* speech act, the proposed actions and requests types pull out *directives*, and the social routines type contains many *markings* and *performances*. While we do not aim to align these types perfectly with speech acts or interchange types, this resemblance suggests we are capturing some of the same patterns as these classification systems.

To further validate our method, we considered sets of exchanges that share the same communicative act sequence but vary in their participants and content. Here and in Part 2, we consider patterns of usage when utterances are classified as their most probable act type. Table **??** shows three exchanges with the same act sequence between parents and their 23-, 39-, and 57-month-olds. Each exchange begins with a proposed action which receives a backchannel, the action is evaluated, and a new action is proposed. Across these three examples, the words in the utterances are different, and sometimes even of opposite valence (e.g., “okay” vs. “no no”). The role of the speaker also varies–the same act can be produced by both parents and children. Nonetheless, there is a structural resemblance across them.

Bolstered by the coherence of communicative act categories and the resemblance of

Table 1

Three examples of the same communicative act sequence (5, 12, 4, 5) in conversations between different parents (P) and children (C) at different ages. Though they involve different topical content, they follow a similar communicative pattern: a suggestion or request regarding action and location; an affirmative or negative response; an evaluative statement or question; and another suggestion regarding action and location.

act	23 months	39 months	57 months
5	C: put back	P: i'm going to turn off the tape recorder	C: i'm gonna fold this
12	P: oh thank_you	C: no no	P: okay
4	P: that is so nice	P: why not	P: why are you folding it
5	P: thank_you for putting that back	C: cause turn it on so i could reach	C: so it will fit me

their sequences, we next provide first qualitative and then quantitative descriptions of parents' and children's production of these acts over development.

Part 2: Use of communicative acts over development

Children's use of communicative acts changes drastically as they grow (Figure 2). Early on, children's utterances are predominantly backchannels, interjections, and affirmations or negations (e.g., "yeah", "no", "uh huh"). By around 20 months of age, social routines such as greetings and naming family members (e.g., "hi", "bye", "mommy") displace backchannels to become a substantial part of children's repertoire. Around the same age, an increasing proportion of children's utterances direct attention to or label things in the environment (e.g., "what's that?", "there's a. . . "). Parents also increasingly use these labeling utterances when their children are around 15-30 months old; after about 30 months, these utterances diminish in both parent and child speech.

Across our observed age range, both parents and children consistently increase the

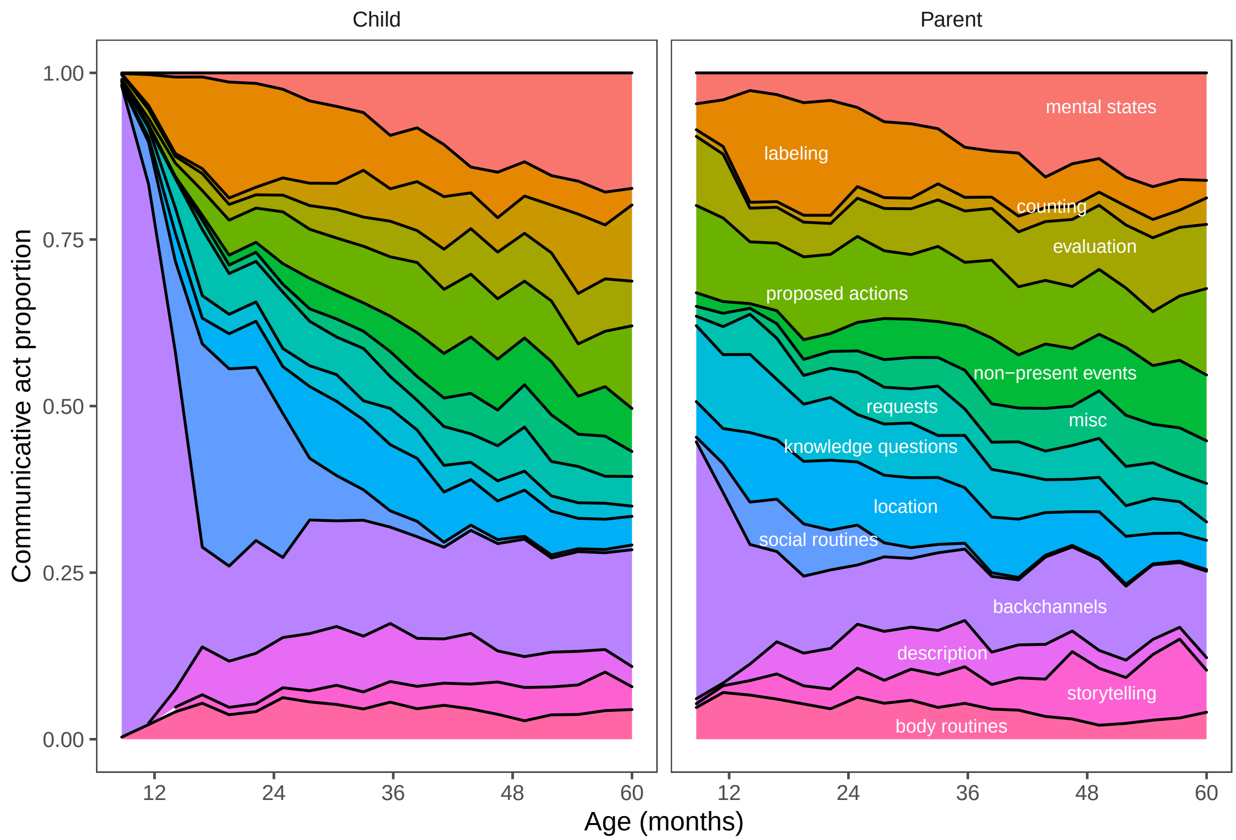

Figure 2. The proportion of utterances classified as each communicative act type produced by children and parents across development.

proportion of their statements about mental states and testimony (e.g., "know", "think", "said"). This is consistent with prior work showing an increase in mental state talk over the 3- to 5-year-old age range, which is correlated with theory of mind task performance (Hughes & Dunn, 1998; Villiers, 2000). It may also reflect growing mastery of the complex syntax required to produce embedded constructions about mental states (e.g., "She thinks it will rain").

Notably, children at the lowest end of our age range start with at least some representation of several communicative act types, and expand their repertoire rapidly. One way to characterize the diversity of communicative acts a person engages in is to measure the entropy of their communicative acts. Children's communicative act entropy increases drastically over development until around 24 months, at which point it matches parents' entropy and remains relatively stable across the rest of our age range, as shown in Figure 3. Our automatically-identified communicative acts are therefore capturing some capacity that becomes more adult-like across development.

To characterize the learning process for these acts and compare it to word learning, we plot parents' and children's act entropies and the entropies of the words they produced in each transcript and fit a sigmoid function to children's trajectories (Figure 3). We find that for both words and acts, children reach 90% of their final repertoire at roughly the same age, 24 months. Children's act entropy is growing most steeply at around 14 months, and their word entropy is growing most steeply at around 12 months. Meanwhile, parents are remarkably stable across all ages: parents provide a constant, high level of diversity of words and communicative acts. Prior work on speech acts has found that productive vocabulary relates to diversity of speech acts: Snow et al. (1996) found that the number of speech act types children use correlates highly with the number of word types they use. We replicate that finding here: children's number of communicative act types produced correlates with their word types produced, even within one-month age intervals (lowest correlation, $r = 0.512$, $p < 0.001$).

Within these trajectories there is variability between individuals in their communicative act diversity. Parents' and children's individual act entropies are correlated ($r = 0.583$, $p < 0.001$), and when divided into one-month age intervals remain correlated in all groups 12 months and older (lowest correlation, $r = 0.321$, $p = 0.010$): parents who use more diverse communicative acts are talking to children who use more diverse communicative acts.

Having characterized qualitative and quantitative changes in children's act repertoire across development, we now ask how contingently parents and children use these acts in conversation.

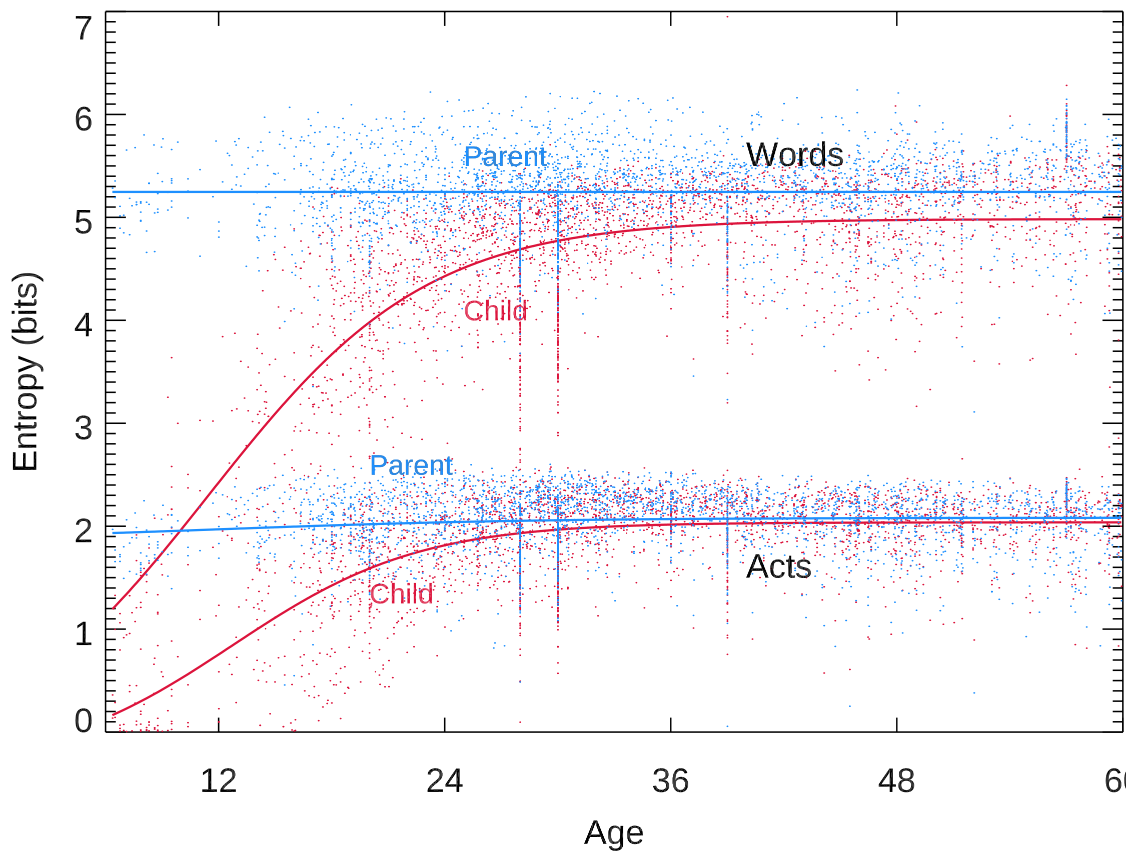

Figure 3. Entropy of communicative acts and words produced by children and parents over development.

Part 3: Dynamics of communicative acts in conversation

Over development, children gain the ability to engage in more structured dialogue, eventually becoming able to respond appropriately and contingently to their conversational partner (Bruner, 1985). To characterize the temporal dynamics of communicative acts in conversation, we turn our attention to the transitions between utterances in parent-child conversations. One way to measure how contingent parents' and children's utterances are on each other is to calculate their mutual information, a measure of how much knowing one piece of information (a prior utterance) reduces uncertainty about another (the current utterance). However, discourse with adult-like structure may not maximize mutual information between sequential utterances: because discourse has recursive structure, raising issues and closing them in an embedded way, it may not be highly regular at the level of utterance-to-utterance transitions (Sacks et al., 1974). Mutual information between sequential utterances is maximized when utterances are highly predictable based only on the utterance prior, and can be dampened either because contingency is weak or because higher-order structure is at play.

To calculate the mutual information between parents' and children's utterances, we use the full distribution of communicative act probabilities for each utterance produced by the model. These probabilities are averaged within turns; that is, if a child produced four utterances in a row, the probabilities are averaged to produce one probability per act type per turn. We then compute conditional probabilities for each current act type given each prior act type. These conditional probabilities are used to calculate mutual information. We calculate mutual information values for the parent's utterance given the child's prior utterance, the child's utterance given the parent's prior utterance, the child's utterance given their own prior utterance (their own most recent turn), and the parent's utterance given their own prior utterance.

When children are young, mutual information is highest between the child's current utterance and their own last utterance (Figure 4). One way to interpret this is that the child is on their own conversational track, steered less by what their conversational partner just said than what they themself said prior. By around 24 months, this tendency lessens considerably, but children remain more predictable based on their own prior utterances than based on their parents' utterances across our age range.

Interestingly, there is an increase in mutual information in all of our measures–parent given child, child given parent, child given child, and parent given parent–from the beginning of our age range peaking at about 12-18 months, after which mutual information declines or stays steady. Though this is an exploratory finding, it points to the possibility that parents and children interact most contingently when children are still quite young. That is, at the

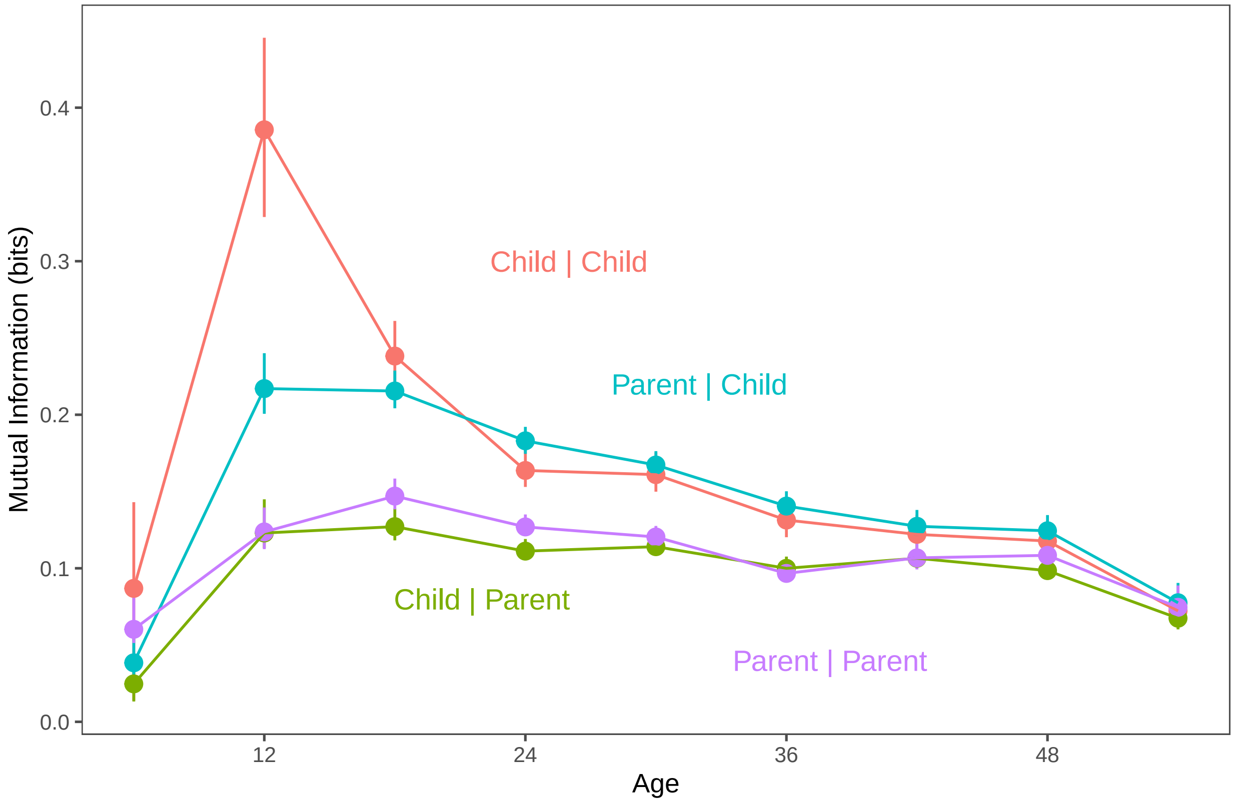

Figure 4. Mutual information between parents' and children's communicative acts over development. We show mutual information for parents' utterances given a child's prior utterance (Parent | Child), children's utterances given a parent's prior utterance (Child | Parent), children's utterances given their own most recent utterance (Child | Child), and parents' utterances given their own most recent utterance (Parent | Parent).

youngest ages, there may be low mutual information between parents' and children's utterances because of disorganized interaction; mutual information may then increase as sequences become more orderly, but discourse is not necessarily structured on longer timescales; and mutual information may fall again as discourse structure emerges that is not captured in sequential utterance transitions.

The point of highest mutual information coincides with the point of steepest growth in communicative act entropy, between 12 and 18 months, suggesting that learning of communicative acts is most rapid when interactions have the most sequential structure.

During this period, children and parents engage in high rates of social routine utterances and labeling and attention-directing utterances. This finding is in line with theories of communicative development that emphasize highly routinized interactions as crucial for learning to engage in structured discourse (Bruner, 1985). Future work may further explore how sequentially predictable interactions give children a foothold in discourse, and how this sequential structure gives way to longer-range discourse structure over communicative development.

General Discussion

In this paper, we present a method for characterizing children's communicative acts on a large scale. In doing so, we gain the ability to examine the communication of more children in more contexts and across a wider age range than afforded by hand coding, and to examine patterns of usage that only become clear across such a wide range of data. Using this method, we find that children start off with a few act types at the beginning of our age range, and quickly branch out as they grow. Parents and children produce these acts contingently, depending on the prior utterance, and we find a period of heightened contingency when children are most rapidly expanding their act repertoire. These findings prompt further exploration of individual differences in communicative act development, longitudinal prediction of language measures, and conversations' discourse structure. Examining these patterns of broad developmental change and contingency would be prohibitive without an automated approach.

Of course, this approach loses the nuance and specificity captured by close observations of children's interactions. Children can achieve communicative goals even before they can use language to do so: they can use gestures and vocalizations both to request a desired object and to call a person's attention to something in the environment (Bates, Camaioni, & Volterra, 1975). Studies of children's early one-word utterances demonstrate that they can use the same single word to fulfill multiple communicative goals (Dore, 1975). These contextual and non-verbal factors are glossed over when we examine communicative acts through words alone. Further, the communicative acts we describe do not solely pick out communicative intent: their distinctions likely reflect a mixture of communicative goals, semantic content, and common syntactic patterns, factors which are likely correlated in children's language environments (Cameron-Faulkner, 2014; Shatz, 1979). Thus, we see this model as a complementary approach to hand-coded observations of children's conversations, picking up on some well-characterized distinctions while allowing us to characterize new patterns in children's communication.

Understanding how children learn to *use* language to communicate is a puzzle at the heart of language development. Our model provides one way to take advantage of corpora of child conversations in the study of children's communicative development, offering a new angle from which to tackle this puzzle.

References

Austin, J. L. (1962). *How to do things with words.* Cambridge: Harvard University Press.

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.

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development*, *89*(6), 1996–2009.

Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language. *Annual Review of Developmental Psychology*, *1*, 223–249.

Bruner, J. (1985). Child's talk: Learning to use language. *Child Language Teaching and Therapy*, *1*(1), 111–114.

Cameron-Faulkner, T. (2014). The development of speech acts. In D. Matthews (Ed.), *Pragmatic development in first language acquisition*.

Dore, J. (1975). Holophrases, speech acts and language universals. *Journal of Child Language*, *2*(1), 21–40.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677.

Gruber, A., Weiss, Y., & Rosen-Zvi, M. (2007). Hidden Topic Markov Models. In *Artificial Intelligence and Statistics* (pp. 163–170). PMLR.

Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: Longitudinal associations with mental-state talk between young friends. *Developmental Psychology*, *34*(5), 1026.

Johnson, M., Griffiths, T. L., & Goldwater, S. (2007). Adaptor grammars: A framework for specifying compositional nonparametric bayesian models. *Advances in Neural Information Processing Systems*, *19*, 641.

MacWhinney, B. (2000). *The CHILDES Project: The database*. Psychology Press.

Ninio, A., Snow, C. E., Pan, B. A., & Rollins, P. R. (1994). Classifying communicative acts in children's interactions. *Journal of Communication Disorders*, *27*(2), 157–187.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, *50*(4), 696–735.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Shatz, M. (1979). How to Do Things by Asking: Form-Function Pairings in Mothers' Questions and Their Relation to Children's Responses. *Child Development*, *50*(4), 1093–1099.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1-2), 39–91.

Snow, C. E., Pan, B. A., Imbens-Bailey, A., & Herman, J. (1996). Learning How to Say What One Means: A Longitudinal Study of Children's Speech Act Use. *Social Development*, *5*(1), 56–84.

Villiers, J. de. (2000). Language and theory of mind: What are the developmental relationships? In *Understanding other minds: Perspectives from developmental cognitive neuroscience, 2nd ed* (pp. 83–123). New York, NY, US: Oxford University Press.

Vouloumanos, A., Onishi, K. H., & Pogue, A. (2012). Twelve-month-old infants recognize that speech can communicate unobservable intentions. *Proceedings of the National Academy of Sciences*, *109*(32), 12933–12937.