Communicative modes in parent—child conversations

Anonymous

Abstract

Keywords: language input, language acquisition, child-directed speech, corpus analysis

From their first utterances, children are not just producing language but *using* it to communicate. A child who can produce only one word at a time 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). Indeed, 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). [. . .]

Describing children's communicative acts computationally, 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 taken their approach at the level of words (i.e., vocabulary learning). The goals and intentions underlying those words are less amenable to such manipulation. In this paper, we put forth one approach to modeling the 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.

[. . .]

Studying children's communicative acts using a computational model also allows us to extract communicative patterns across many children with less a priori specification of what those patterns are. Traditionally, studies of communicative acts among children have brought frameworks from adult communication, such as Speech Act theory (Austin, 1962; Searle & Searle, 1969) and Conversation Analysis (Sacks, Schegloff, & Jefferson, 1974), to bear on children's conversations. While concepts like 'directive', 'expressive', 'accusation' and 'justification' can be useful to characterize children's conversations, young children may not have the same communicative needs as adults—these may not be relevant distinctions in children's communication. An advantage of our approach is that we do not need to specify the types of communicative acts a priori. [. . .]

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 without top-down specification, this model extracts several communicative acts analogous to those observed in close case studies of children's communication. We then show that use of these acts has a developmental trajectory in line with findings from close studies of children's communication. [. . .]

Corpus

We use 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 filtered the transcripts to include only utterances spoken by the target child or their parents. Overall, our training data included 4,043 transcripts from 411 children 6 months to 60 months old.

Model

We use a Hidden Topic Markov Model (Gruber, Weiss, & Rosen-Zvi, 2007) to extract communicative modes from parent—child conversations. 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.

The Hidden Topic Markov Model (HTMM) differs from a

simple topic model in that it 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) develop this model and use 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 acknowledgements. Typically, function words are removed from corpora before training topic models to aid detection of thematic content. Here, we aim to classify communicative modes underlying utterances rather than thematic topics. We expect function words to be diagnostic of these modes, so include them in our training data.

After training, the model produces a set of topics with associated probability distributions over words and a probability distribution over topics for each utterance in the corpus. In Part 1, we will show that the model captures some aspects of communicative acts and explore the static distribution of these utterance types. In Part 2, we will examine trajectories of topic use across development among parents and children. In Part 3, we will examine the temporal dynamics of topic use within discourse.

Part 1: Topics and their static structure

The most distinctive words of each topic in our 15–topic model, as measured by difference between a word's probability within a topic and its average probability across all topics, are shown in Table 1.

One way to characterize the diversity of communicative acts a person engages in is to measure the entropy of their communicative acts. Overall, parents' communicative acts have higher entropy than children's. Children's communicative act entropy increases drastically over development between 6 and around 24 months, and remains relatively stable across the rest of our age range, as shown in 3. Our communicative acts are therefore capturing some capacity that becomes more adult-like across development. [more description / perhaps some stats]

Beyond becoming more complex across development, the ability to engage in a variety of communicative acts might reasonably be expected to correlate with other measures of language ability. Indeed, Snow et al. (1996) find that the number of speech act types children use correlates highly with the number of word types they use, but does not consistently correlate with mean length of utterance (MLU). We replicate these findings here: children's number of communicative act types produced correlate with their word types produced (1) and does not consistently correlate with MLU (2).

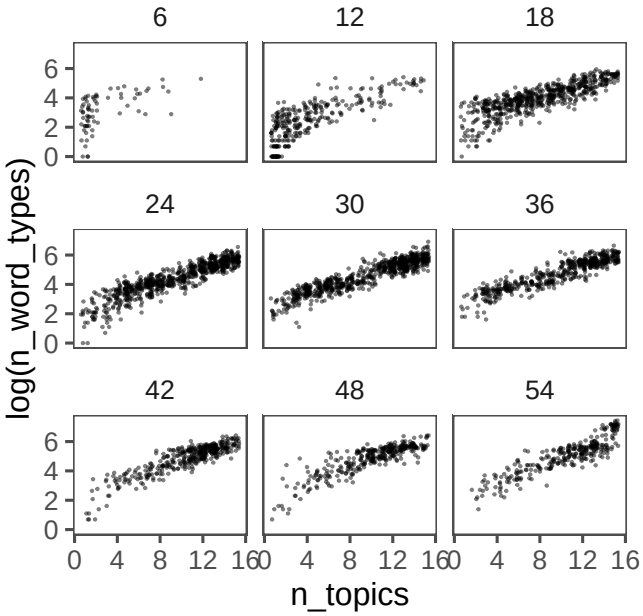

Figure 1: The number of communicative act types children produced plotted against the number of word types they produced in six-month age ranges.

Part 2: Use of communicative acts over development

Part 3: Dynamics of communicative acts in conversation

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.

Dore, J. (1975). Holophrases, speech acts and language universals*. *Journal of Child Language*, *2*(1), 21–40. http://doi.org/10.1017/S0305000900000878

Gruber, A., Weiss, Y., & Rosen-Zvi, M. (2007). Hidden Topic Markov Models. In *Artificial Intelligence and Statistics* (pp. 163–170). PMLR. Retrieved from http://proceedings.mlr.press/v2/gruber07a.html

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

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, *50*(4), 696–735. http://doi.org/10.2307/412243

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

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

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
i	a	two	not	you	to	of	some	you	the		yeah	a	the	your
you	that	one	it's	it	we	they	you	what	in	mommy	no		and	on
what	what's	dis	good	me	go	we	want	are	there	daddy	oh	one	he	my
know	is	three	i'm	to	and	have	i	do	put	hi	okay	this	was	put
don't	this	four	you're	can	you	like	more	look	here	down	hm	is	his	his
it	that's	de	a	want	going	the	have	see	it	baby	uh	that's	said	her
do	what	dat	that's	okay	did	i	eat	with	go	bye	huh	blue	in	hair
that	the	five	very	i	when	and	milk	at	on	ball	ah	red	she	off
did	look	duh	be	up	at	them	juice	play	right	n	mhm	green	of	in
think	who's	six	he	get	school	are	your	those	this	tape	mm	make	to	head

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

act_type	gloss_23	speaker_23	gloss_39	speaker_39	gloss_53
5	put back	Target_Child	i'm going to turn off the tape recorder	Father	hey can i take this to
12	oh thank_you	Father	no no	Target_Child	yeah
4	that is so nice	Father	why not	Father	that's neat
5	thank_you for putting that back	Father	cause turn it on so i could reach	Target_Child	daddy you could ope

Table 2: Four examples of the same communicative act sequence (5, 12, 4, 5) in conversations from different children 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.

Speech Act Use*. *Social Development*, *5*(1), 56–84. http://doi.org/https://doi.org/10.1111/j.1467-9507.1996.tb00072.x

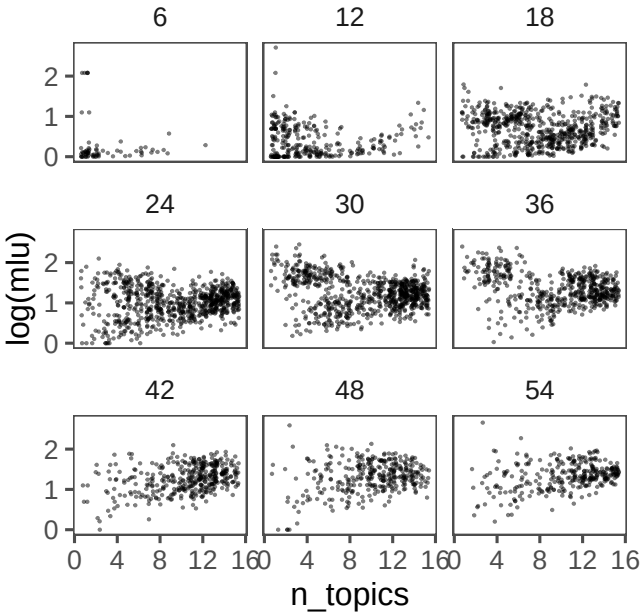

Figure 2: The number of communicative act types children produced plotted against their mean length of utterance.

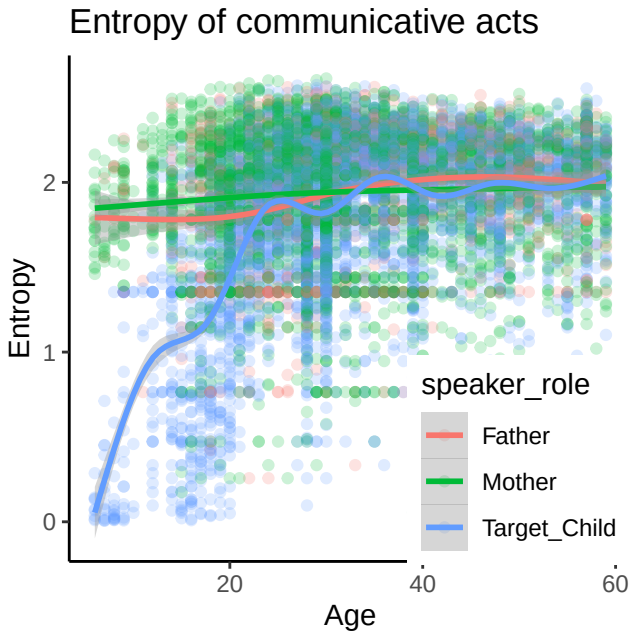

Figure 3: Entropy of topics produced by children, mothers and fathers over development.

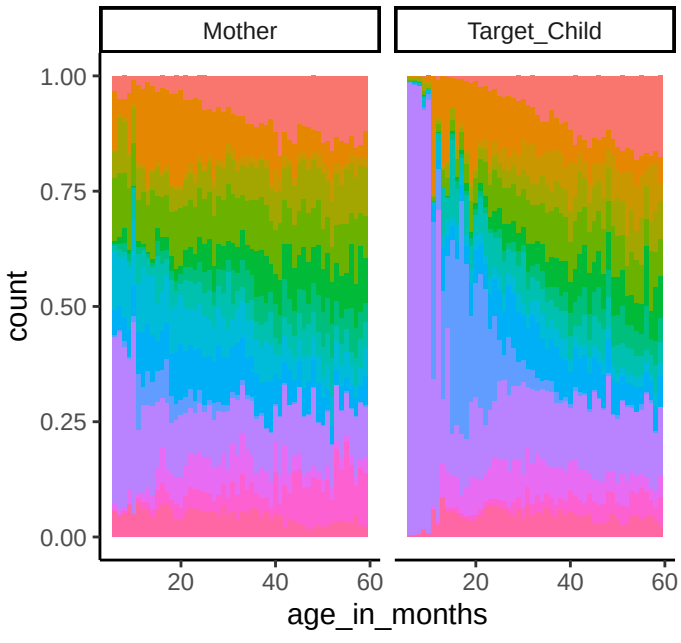

Figure 4: Plot showing the distribution of topics produced by children and mothers across development.

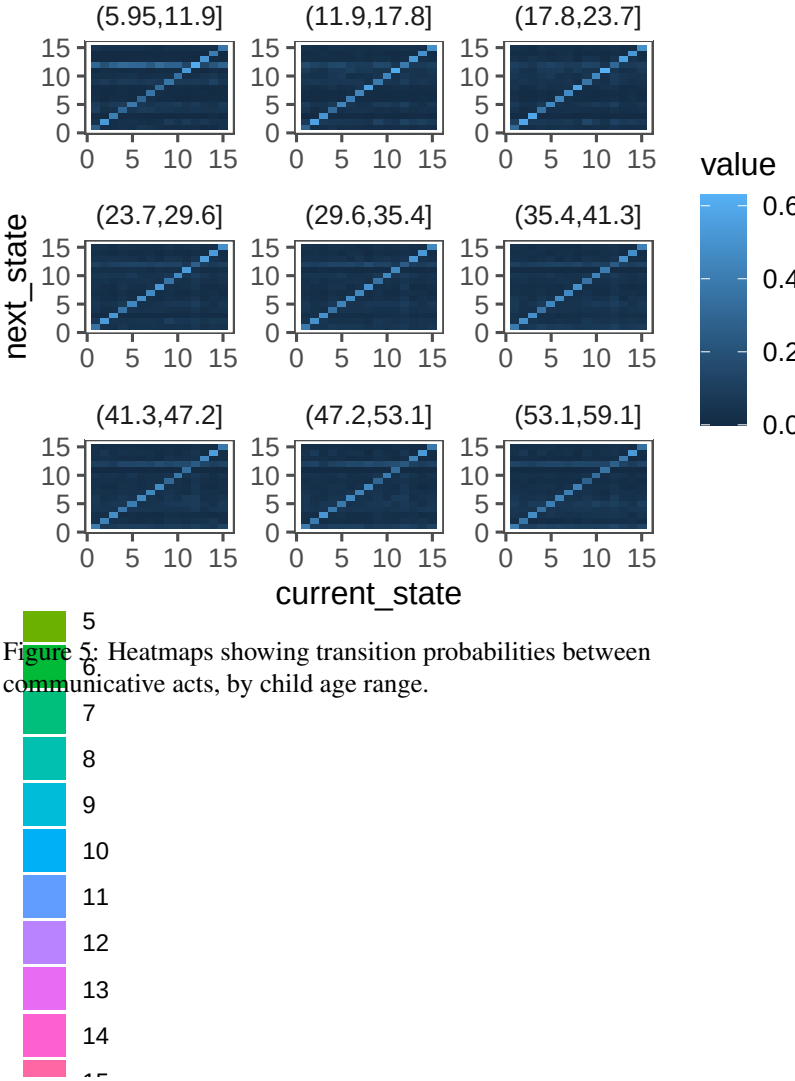

Figure 5: Heatmaps showing transition probabilities between communicative acts, by child age range.