Children gesture when speech is slow to come

Daniel Yurovksy, Madeline Meyers, Nicole Burke, Amanda Woodward, and Susan Goldin-Meadow
{yurovsky, mcmeyers, nicoleburke, woodward, sgm}@uchicago.edu
Department of Psychology
University of Chicago

Abstract

We test this prediction in a corpus of videos of parent-child interaction in the home recorded longitudinally from 14- to 34-months. naturalistic parent-child o

Keywords: Add your choice of indexing terms or keywords; kindly use a semi-colon; between each term.

Introduction

Children learn a striking amount of language in their first few years of life–thousands of sounds, words, grammatical categories, and the combinatoric properties that allow them to be combined to produce meaningful utterances (E. V. Clark, 2009). They also come to understand what all of this language *is for*: communicating with other people (Zipf, 1949). Furhter, there is good reason to think that these two problems are deeply interwined. The language that children hear is rarely a running commentary on the world around them–when a child's parents return home work, they are much more likely to say "whatcha been doing all day?" than "I am opening the door" (Gleitman, 1990). Understanding that their parent is not trying to tell them about the door may go a long way to learning what the words they are hearing mean.

The understanding that speakers productions are intended to communicate information is at the core of the kinds of inferences that adults routinely make when processing language. These pragmatic inferences, for instance, are the reason that hearing a speaker say that they ate "some of the cookies," causes us to think that some cookies still remain on the plate (Grice, 1969). Children's ability perform these kinds of complex inferences appears relatively late in language development (Noveck, 2001). However, a growing body of empirical evidence shows that a basic undertanding of the communicative purpose of language is already present in the first year of life (Tomasello, 2000). For instance, children appear to understand that speakers communicate information to other adults, even if they themselves do not understand the words being said (Vouloumanos, Martin, & Onishi, 2014; Vouloumanos, Onishi, & Pogue, 2012). But is this understanding of communicative goals present in children's *language production* as well?

The core of extended communicative interactions is taking turns: participants each contribute to the discourse, but only one at a time (Sacks, Schegloff, & Jefferson, 1974). Turn taking is not only universal among both modern and indigenous cultures, the length of time between turns is highly stereotyped, and predicted by the same factors across cultures (Stivers et al., 2009). Evidence from both early observational studies and more recent experiments suggests that tracking of turn boundaries emerges early in infancy–perhaps in the course of scripted interactions like patty cake (Bruner, 1983; Casillas & Frank, 2017).

The regularity of these turns makes communication inherently time constrained: If you stop talking for too long, you lose your turn. Adults are sensitive to this time pressure, for instance producing filled pauses like "um" when they are having difficulty retrieving the words they want to produce in order to signal their desire to hold onto their turn (H. H. Clark & Fox Tree, 2002). If retrieval is still unsuccessful, lingusitically-proficient adults can opt for an alternative word or even a description that gives their interlocutor enough information to help retrieve the word for them (H. H. Clark & Schaefer, 1989). Children still learning their native language, for whom such strategies are unavailable, might resort to an alternative mode of communcation: pointing.

Children produce deictic gestures early in infancy, and appear to understand that these gestures both direct attention and communicate intentions by the time they are 12-months-old (Liszkowski, Carpenter, & Tomasello, 2007; Tomasello, Carpenter, & Liszkowski, 2007). Around the same time, infants begin producing their first spoken words (Bloom, 2000). Over the next few years, infants will produce many more words, and need to rely less on deictic gesture to communicate. However, while children master some words early, others which are less frequent may remain difficulty to retrieve and produce. If children, like adults, are sensitive to the time pressures of communication, then then they may use gesture even for *known* words if these words are slow to come. # Communication as a race between modalities

When children wish to share their interest in an object with a caregiver, they have two modalities available to them. One possibility is to use spoken language, producing the canonical label for it (e.g. "ball"). Alternatively, they can use a deictic gesture, e.g. a point, to draw the caregiver's attention to it. When should children use each of these modalities?

If the child does not know that the object is called "ball," they have no choice but to point. However, if they do know it's label, time pressure on communication produces a race between modalities. If the child can recall the word quickly, they should prefer to use language–as speech is less effortful than pointing (Zipf, 1949). However, if the recall process of recalling and producing the word is progressing too slowly, the child risks losing their conversational turn and should instead point.

This kind of race model can be formalized nicely as two competing accumulators (see e.g. S. D. Brown & Heathcote, 2008). Each modality accumulates activation at its own in-

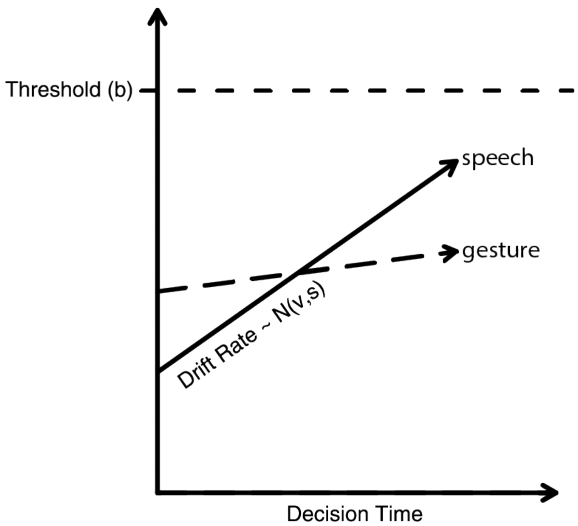

Figure 1: Referential communication as a race between speech and gesture. While the rate of pointing should be independent of referent, the speech accumulator should vary with properties of words, e.g. frequency

dependent rate, and whichever is the first to reach threshold wins the race and is chosen as the referential modality. Although the difficulty of pointing may vary due to issues of proximity of the speakers to each-other, the location of the target referent, etc., the difficulty of pointing should in general be independent of the thing being pointed to. On the other hand, the difficulty of recalling and producing a word varies from word to word. In adults, this difficulty is influenced by many features of the word, including the phonology and orthography of both the word and its neighbors in the lexicon (see e.g. Vitevitch, 2008). Here we focus on just one–contributor: Input frequency (Wingfield, 1968). The more frequently we hear a word, the easier it is for us to retrieve and produce it. Children's *language processing* shows similar effects of frequency–children's speed and accuracy of known words increases as they become more frequent (Swingley, Pinto, & Fernald, 1999). If their *language production* is similarly affected by frequency, than the rate of the speech accumulator should increase as frequency increases, resulting in it winning the race for reference more often.

The utility of this framework is that it makes detailed predictions about the relationship between modality and production time as features of the target referent change. We test three specific predictions of this model in children's spontaneous productions from 14- to 34-months:

1. As the frequency of a referent in children's input increases, they should be relatively more likely to use speech, and less likely to use gesture to communicate about it.

2. As children develop and learn more language, words should be known better and thus be easier to retrieve. Thus, speech should win the race more often–especially for low frequency words.

3. Recent use of a word should make it easier to retrieve, thus children should be relatively more likely to use speech for low frequency referents in if they have occurred previoulsy in the same discourse than at baseline.

Method

The data analyzed here are transcriptions of recordings parent-child interactions in the homes of 10 children from the Chicagoland area. Each recording was ˜90min long, and participants were given no instructions about how to interact–the goal was to observe the natural ecology of language learning. Each child was recorded 6 times at 4-month intervals starting at 14-mo. and ending at 34-mo.

These children's data was drawn from the larger Language Development Project dataset pseudo-randomly to preserve the socio-economic, racial, and gender diversity representative of the broader Chicago community. Of the 10 children, 5 were girls, 3 were Black and 2 were Mixed-Race. Familes spanned a broad range of incomes, with 2 families earning $15,000 to $34,999 and 1 families earning greater than $100,000. The median family income was $50,000 to $74,999.

Data Processing

The original Language Development Project transcripts consist of utterance-by-utterance transcriptions of the 90 minute recordings in CHAT format (MacWhinney, 2000), as well as a transcription of all communicative gestures produced by children and their conversational partners, including conventional gestures (e.g. waving "bye"), representational gestures (e.g. tracing the shape of a square), and deictic gestures (e.g. pointing to a ball).

For each of these communicative acts, we coded all concrete noun referents indicated in either the spoken or gestural modality (see Table 1). As it is difficult both to gesture about, and to code, gestures for abstract entities like "weekend," we focused only on nouns that could be referred to in either gesture or speech. Spoken referents were coded only if a noun label was used (e.g. no pronouns were included), and only deictic gestures were counted as referential to minimize ambiguity in coding. Synonyms, nicknames, and proper nouns were all coded according to a manual that can be found in the linked github repostory.

Reliability

In order to ensure the integrity of the coded data for further analyses, we first assessed inter-rater reliability, and then assessed whether the coded referents were present in the scene.

Inter-Rater Reliability To assess the reliability of referent coding, 25% of the transcripts were coded by a second independent coder. Reliability between coders was good (Cohen's $\kappa = 0.76$). Issues and discrepancies in coding decisions were

person	utterance	gesture	spoken	gestured
parent	do you want to read a book quick with mom		book;mom	
child	no			
child	mommy		mom	
parent	no			
parent	oh you want to wear your necklaces		necklace	
parent	uhoh			
parent	I think it's stuck			
parent	you need some help			
child		hold		necklace
parent	why don't you just say help instead of yelling			
parent	can you say help mommy		mom	

Table 1: An example of the output of data processing

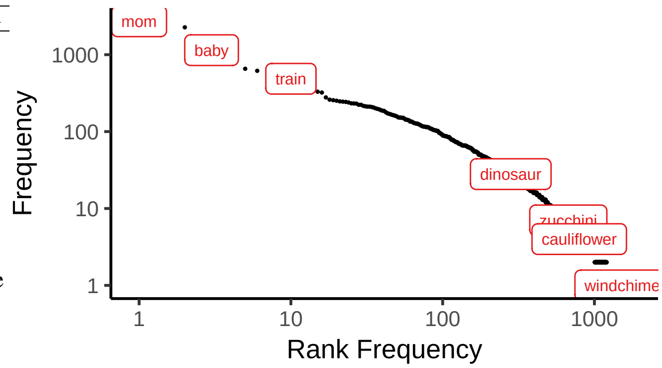

Figure 2: One column image

dicussed and resolved during the formation of a coding manual.

Referent Presentness Although we coded for concrete referents that had the potential to be produced either in speech or in gesture, we found that these referents were not always physically present in the environment. If a referent was not present, it could not be referred to in the gestural modality–potentially biasing our analyses. After coding all referents from the transcripts, the primary coder judged whether each was likely to be present in the scene according to a list of criteria described in the coding manual. Across the 60 transcripts, 90% of referents were judged to be present. Absent referents were included in estimates of input frequency, but excluded from analysis of production modality.

Results

The key predictions of our race model of reference all connect the ease of recall of a referent's spoken label. Although ease of recall is likely related to a number of factors (e.g. phonotactic probability, neighborhood density, etc), we focus here on one easily quantifiable and well-attested predictor: input frequency (Wingfield, 1968).

Estimating Frequency

To estimate the input frequency of each referent in the corpus, we summed its frequency of use across all children and parents and across both the speech and gestural modalities. This estimator is of course imperfect– It assumes, for instance, that every child receives the same input, and that input frequency is stationary across development. Nonetheless, because of the difficulty of estimating these frequencies well especially from a corpus of this size, we felt that a more complex estimator would introduce more error than it was worth.

Figure 1 shows the frequency distribution of the 1533 individual referents in this corpus across all recordings. As with many other frequency distributions in language, referential frequencies were approximately Zipfian, appearing approximately linear on a log-log scale (Piantadosi, 2014). These frequency estimates were used to test the first key prediction of the race model of communcation: Labels for referents that are easier to retrieve from memory are more likely to be produced in speech.

Testing the Race Model

Figure 1 shows the distribution of

primary independent variable in the predictions made by the race model of reference is

All artwork must be very dark for purposes of reproduction and should not be hand drawn. Number figures sequentially, placing the figure number and caption, in 10 point, after the figure with one line space above the caption and one line space below it. If necessary, leave extra white space at the bottom of the page to avoid splitting the figure and figure caption. You may float figures to the top or bottom of a column, or set wide figures across both columns.

Tables

Number tables consecutively; place the table number and title (in 10 point) above the table with one line space above the caption and one line space below it, as in Table 1. You may float tables to the top or bottom of a column, set wide tables across both columns.

You can use the xtable function in the xtable package.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.15	0.10	1.5	0.15
x	1.99	0.10	19.1	0.00

Table 2: This table prints across one column.

Frequency of use

One-column images

Single column is the default option, but if you want set it explicitly, set `fig.env` to `figure`. Notice that the `num.cols` option for the caption width is set to `1`.

Figure 3: This image spans both columns. And the caption text is limited to 0.8 of the width of the document.

Figure 4: One column image.

R Plots

You can use R chunks directly to plot graphs. And you can use latex floats in the fig.pos chunk option to have more control over the location of your plot on the page. For more information on latex placement specifiers see **here**

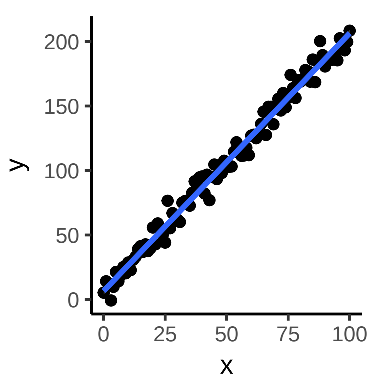

Figure 5: R plot

Discussion

Young children are inundated with language, hearing on the order of 30 million words by the time they are four years old (Hart & Risley, 1995). From the statistical relationships within and among these words, children must discover the latent structures that allow them to become fluent speakers of their native language. Some of these words will be overheard, addressed by one parent to another or a sibling to a friend. However, some will be directed to the child, and these child-directed words maybe be particularly supportive of learning (Weisleder & Fernald, 2013). Child-directed speech differs from adult-directed speech along a number of dimensionsChild-directed speech is not merely a corpus of well-formed language; It is contingent on the child's own attention, interests, and prior knowledge, and thus can be directed *by the child themself*.

Young children are notorious for asking questions "why?" In her analysis of 5 children from the Brown and Kuczaj corpora in CHILDES, Chouinard (2007) reports children asking over 100 questions per hour they interaction with adults over the 2-5 year range. These questions are powerful because they allow children to learn about two important things simultaneously: The causal relationships in the world around them, and also about the structure of language itself. By driving the discourse into predictable areas of content, they can reduce referential ambiguity in learning new language for this content.

Long before they can explicitly direct their input with wh-questions, children can sometimes acchieve a similar outcome simply by referring to objects in their environment. Having observed a referential event, parents will often follow-in with expansions and additional information about the child's target of interest (H. H. Clark, 2014)

Acknowledgements

This research was funded by a James S. McDonnell Foundation Scholar Award to DY and National Institutes of Health P01HD40605 to SGM.

References

Bloom, P. (2000). *How children learn the meanings of words*. MIT press: Cambridge, MA.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178.

Bruner, J. (1983). *Child's talk: Learning to use language* (pp. 111–114). Norton.

Casillas, Marisa, & Frank, M. C. (2017). The development of children's ability to track and predict turn structure in conversation. *Journal of Memory and Language*, *92*(C), 234–253.

Chouinard, M. M. (2007). Children's questions: A mechanism for cognitive development. *Monographs of the Society for Research in Child Development*, *72*(1), 1–112.

Clark, E. V. (2009). *First language acquisition*. Cambridge University Press.

Clark, H. H. (2014). How to talk with children. In I. Arnon, M. Casillas, C. Kurumada, & B. Estigarribia (Eds.), *Language in interaction: Studies in honor of eve v. clark* (pp. 333–352). John Benjamins.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to Discourse. *Cognitive Science*, *13*(2), 259–294.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.

Grice, H. P. (1969). Utterer's meaning and intention. *The Philosophical Review*, *78*, 147–177.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young american children.* Paul H Brookes Publishing.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). Reference and attitude in infant pointing. *Journal of Child Language*, *34*, 1–20.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188.

Piantadosi, S. T. (2014). Zipfs word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, *21*(5), 1112–1130.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. *Language*, *50*, 696–735.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*(2), 73–108.

Tomasello, M. (2000). The social-pragmatic theory of word learning. *Pragmatics*, *10*, 401–413.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*(3), 705–722.

Vitevitch, M. S. (2008). What Can Graph Theory Tell Us About Word Learning and Lexical Retrieval? *Journal of Speech, Language, and Hearing Research*, *51*(2), 408–422.

Vouloumanos, A., Martin, A., & Onishi, K. H. (2014). Do 6-month-olds understand that speech can communicate? *Developmental Science*, *17*(6), 872–879.

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*, 12933–12937.

Weisleder, A., & Fernald, A. (2013). Talking to children matters early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152.

Wingfield, A. (1968). Effects of frequency on identification and naming of objects. *The American Journal of Psychology*, *81*(2), 226–234.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley Press.