Characterizing Child-Directed Listening with Corpus and Model-based Analyses

Stephan C. Meylan[1,2], Elika Bergelson[2], Roger P. Levy[1]
[1]Massachussets Institute of Technology [2]Duke University

22nd Bi-Annual International Conference of Infant Studies
“Insights From Outside the Lab” symposium

Extensive research has focused on how we talk to infants and young children (e.g., the properties and benefits of infant-directed speech; Newport and Gleitman, 1977), but what about how we listen to them? Here we propose that parents and other adult caregivers contribute to the language learning process through “child-directed listening.” Just as adults use rich expectations regarding what other people are likely to say in order to overcome ambiguity, noise, and variation in the speech they hear from other adults (Gibson et al. 2013), they employ a related set of abilities to interpret infants’ vocalizations. Critically, child-directed listening allows caregivers to treat children's speech as communicatively significant, felicitous, and actionable—even when it bears little resemblance to the adult language.

We present two analyses as initial steps towards characterizing child-directed listening. First, we analyzed lab-annotated “best-guess” transcriptions from developmental corpora as an approximation for how caregivers interpret children’s vocalizations as words. Specifically, we compared these best-guess transcriptions of child speech in the Providence corpus (Demuth et al., 2006) with matched phonemic (IPA) transcriptions. While phonemic /kæts/ would typically be transcribed as “cats,” we measured the prevalence of “annotator-recovered words” — instances where the best-guess transcription is *not* supported by the phonemic transcription of the child’s speech (e.g. /kæt/ [“cat”] annotated as “cats”). Focusing on a sound pattern that includes plural nouns (e.g., “cats”), possessives (“Mom’s”), and 3rd person singular verbs (“sees”), we looked for words with transcriptions ending in “s” or “z,” but where the matched phonemic transcription does *not* contain a corresponding sound (no word-final /s/, /z/, or approximations /ð/, /θ/, or /ʒ/). Analyzing 47,844 word tokens, we found high rates of annotator-recovered words during infancy, with the proportion of such word tokens per transcript decreasing into childhood (Figure 1).

Next, we evaluated whether substituting these annotator-recovered words yields utterances that are more consistent with adults’ linguistic expectations than the literal interpretation of children’s vocalizations. As a stand-in for adults’ linguistic expectations we used a state-of-the-art neural language model (GPT-2, Radford et al., 2018). We compared the model-estimated probability of two variants for each utterance in a subset from Analysis 1: one with the annotator-recovered word (e.g., “two little kitty cats”), and one with a literal interpretation of the phonemic transcript (“two little kitty cat”). If adults bring their prior linguistic knowledge to the task of interpreting infants' speech, estimated surprisal (negative log probability of an utterance under the model) should be systematically lower among annotator-recovered utterances than literal utterances. Indeed, this is the case (Figure 2; paired t-test, t = -18.531, df = 602, p < .0001), suggesting annotators’ interpretations reflect their prior expectations.

Taken together, these analyses suggest that adult language processing plays a critical role in the language learning process: adults use prior expectations to interpret (and thus act on) infants’ vocalizations. Future research will investigate the implications of child-directed listening for development, including the nature of feedback available to infants in language learning.

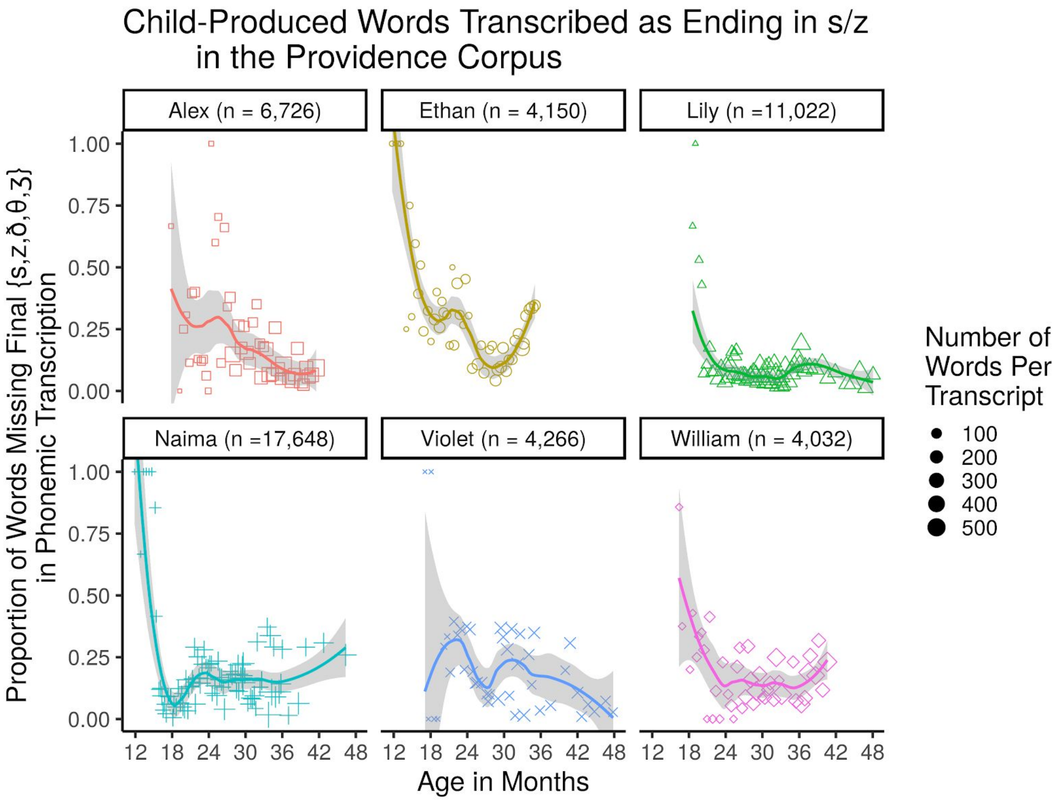

Figure 1: Prevalence of "annotator-recovered" word tokens per child age in the Providence corpus: Among children's words transcribed as ending in /s/ or /z/, how often are these sounds or their approximate realizations actually produced in the corresponding phonemic transcription? Higher y-axis values indicate adult listeners (annotators) are *more* reliant on prior expectations—other words in the utterance or discourse context—to interpret children's utterances.

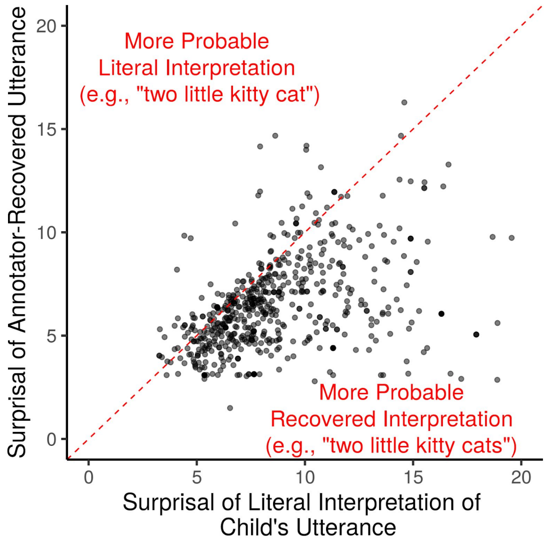

Figure 2: Model-based surprisal (negative log probability) of the literal interpretation of a child's utterance (x-axis) vs. the annotator-recovered utterance (y-axis) for *n* = 603 child utterances.
For example, given the phonemic transcription /ˈtuː ˈlɪtəl ˈkɪti ˈkæt/, we evaluate the probabilities of the literal interpretation ("two little kitty cat") and the matched annotator-recovered transcription ("two little kitty cats"). Points below the diagonal indicate that annotator-recovered utterance is more probable (less surprising) than the corresponding literal interpretations under a generative language model used in adult psycholinguistics.