

What do babies hear? Analyses of child- and adult-directed speech

Marisa Casillas¹, Andrei Amatuni², Amanda Seidl³, Melanie Soderstrom⁴, Anne S. Warlaumont⁵,
Elika Bergelson²

¹Max Planck Institute for Psycholinguistics, The Netherlands

²Psychology and Neuroscience, Duke University, USA

³Speech, Language, and Hearing Sciences, Purdue University, USA

⁴Psychology, University of Manitoba, Canada

⁵Cognitive and Information Sciences, University of California, Merced, USA

marisa.casillas@mpi.nl, andrei.amatuni@duke.edu, aseidl@purdue.edu,
m.soderstrom@umanitoba.ca, awarlaumont2@ucmerced.edu, elika.bergelson@duke.edu

Abstract

Child-directed speech is argued to facilitate language development, and is found cross-linguistically and cross-culturally to varying degrees. However, previous research has generally focused on short samples of child-caregiver interaction, often in the lab or with experimenters present. We test the generalizability of this phenomenon with an initial descriptive analysis of the speech heard by young children in a large, unique collection of naturalistic, daylong home recordings. Trained annotators coded automatically-detected adult speech ‘utterances’ from 61 homes across 4 North American cities, gathered from children (age 224 months) wearing audio recorders during a typical day. Coders marked the speaker gender (male/female) and intended addressee (child/adult), yielding 10,886 addressee and gender tags from 2,523 minutes of audio (cf. HB-CHAAC Interspeech ComParE challenge; Schuller et al., in press). Automated speaker-diarization (LENA) incorrectly gender-tagged 30% of male adult utterances, compared to manually-coded consensus. Furthermore, we find effects of SES and gender on child-directed and overall speech, increasing child-directed speech with child age, and interactions of speaker gender, child gender, and child age: female caretakers increased their child-directed speech more with age than male caretakers did, but only for male infants. Implications for language acquisition and existing classification algorithms are discussed.

Index Terms: Addressee, Child Directed Speech, Language Development, Speech Classification, Gender

1. Introduction

(other intro stuff) Here we analyze real-world recordings of conversations during infants’ daily experiences, i.e., conversations involving young children, their caregivers, and other family members in their natural home environment. The data were annotated with an eye towards developing algorithms that successfully tag these data into child- vs. adult-directed speech. Such algorithms, in turn, can be used over thousands of hours of existing data, with great potential for improving our understanding of infant learning and development, and parent-child interaction, for both typically developing kids (such as those in this dataset) and by extension, to children with language delays and deficits. Here we report details on how the corpus was constructed, along with initial analysis based on the manually annotated data.

2. Methods

This corpus was created by sub-sampling daylong audio recordings from four corpora that are part of a much larger corpus of real-world child language recordings known as HomeBank [?] (homebank.talkbank.org): [?], [?], [?], and [?]. All recordings were collected using a LENA recording device, which is a small audio recorder worn by a child in a pocket, in specialized clothing. [?]. All recordings came from four cities in North America, from monolingual English homes, with children who were typically-developing. All families granted permission to share the audio with the research community. Age was sampled as uniformly as possible between 2 and 24 months across sub-corpora, with one recording sampled per child.

The recordings were then analyzed with LENAs proprietary software, which identifies ‘conversational blocks’ (i.e. speech surrounded by 5 s of non-speech), and utterance-boundaries with speaker-tags from a closed set of 12 alternatives (e.g. Female-Adult-Near, Key-Child), among other features. The goal for the present analysis was to subsample evenly over early childhood, and classify all adult speech as (1)child-directed (CDS) or adult-directed (ADS), and (2)female or male.

Within each of the 61 daylong recordings, 20 conversational blocks (as defined above) were selected. Within each block, individual adult speaker utterances (as tagged by LENA’s proprietary diarisation algorithm) were hand-annotated separately by three trained research assistants. The addressee and gender tagging relied on both acoustic-phonetic information and context (see <https://osf.io/d9ac4/> for more detail). Non-speech clips were tagged into a ‘Junk’ category. Annotators achieved high reliability in differentiating CDS/ADS (Fleiss’ $\kappa > .75$, $p < .001$). These annotations are also being used for the ComParE 2017 addressee sub-challenge (Homebank Child/Adult Addressee Sub-Challenge, HB-CHAAC).

3. Results

4. Discussion

5. Conclusions

6. Acknowledgements

This work was funded in part by the NWO Veni Innovational Scheme (269484) (MC; grant to Stephen Levinson), NIH DP5-OD019812 (EB), SSHRC Insight Grant 435-2015-0628 (MS), and NSF SBE-1539129 and NSF BCS-1529127 (AW). The au-

thors further thank the research assistants who provided HB-CHAAC labels and Kelsey Dyck for assistance developing the labelling protocol, Mark VanDam for sharing his corpus, Brian MacWhinney hosting the HomeBank corpora, as well as all families who provided the raw data.