Workshop on Corpus Collection, (Semi)-Automated Analysis, and Modeling of Large-Scale Naturalistic Language Acquisition Data

Elika Bergelson (elika.bergelson@gmail.com)
Brain and Cognitive Sciences, University of Rochester, Rochester, NY, 14627 USA
Department of Psychology & Neuroscience, Duke University, Durham, NC, 27708 USA

Keywords: language acquisition; automatic speech recognition; computational models; speech corpora

Introduction

The main goal of this full-day workshop is to bring together researchers from several distinct fields: behavioral psychologists studying language acquisition, speech technology researchers, linguists, and computational modelers of cognitive development. These groups are broadly interested in the same questions, i.e. what is the nature of speech and language, and how might a system learn to process it in supervised or unsupervised ways? Since the groups interested in these questions work on different analysis levels, cross-pollination has been sparse.

Recent technological innovations have made collecting long naturalistic recordings of children's home environment far simpler than in the past. However, the raw output of such recordings is not immediately usable for most analyses. Simultaneously, speech technology (ST) and machine learning tools have improved immensely over the past decade, making it feasible to use such tools with increasingly diverse and noise-laden data. Relatedly, cognitively viable computational models have made recent strides in explaining learning and development, but few such models can be applied to novel data-sets without encountering many hurdles about translatability across frameworks. This workshop brings together experts from all of these areas, and seeks to build bridges across them, with insight from other similar interdisciplinary efforts in other areas of cognitive science

The program committee is part of a newly formed group called DARCLE (Daylong Audio Recordings of Children's Language Environment); with the help of an NSF grant, DARCLE has created a repository called HomeBank for raw data, metadata, and analysis/processing tools for long-form recordings of child language. This workshop is an opportunity to network with related efforts in Europe, and for a talk and demo of a related effort, the NSF-funded Speech Recognition Virtual Kitchen.

Workshop Organization

Target Audience

The target audience of this workshop is researchers (from students to PIs) with an interest in collecting, analyzing, and modeling language data. It will be especially useful for participants with background in speech and hearing, computer science, linguistics, or psychology, but will be geared to a non-specialist audience.

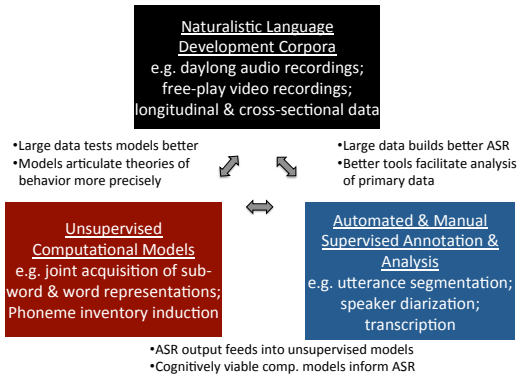

Figure 1: Schematic of Relevant Areas and Questions.

Workshop Format

The workshop comprises ten confirmed speakers in four sets of talks, with interspersed question periods and breaks. Ten institutions from four countries are represented. Recordings will be available on the workshop website. The workshop features experts across a range of disciplines, and provides an opportunity for junior researchers to solicit feedback.

Workshop Organizer and Program Committee

Elika Bergelson is a Research Professor in the University of Rochester's Brain & Cognitive Sciences Department, and an incoming Assistant Professor in Duke's Psychology & Neuroscience Department.

The program committee includes PIs of the HomeBank grant, and the DARCLE board: Elika Bergelson (Rochester/Duke), Alex Cristia (LSCP), Emmanuel Dupoux (LSCP), Brian MacWhinney (CMU), Melanie Soderstrom (Manitoba), Mark VanDam (Washington State), and Anne Warlaumont (UC-Merced).

Summary of Presentation Topics

Speech and Language Development Corpora

Anne Warlaumont After introducing HomeBank, Warlaumont will present a case example of a computational model that uses statistics obtained from daylong home audio recordings of children to provide an account for how growth in speech-related vocalizations varies across groups with differing clinical and socioeconomic status. She will also

discuss how more detailed computational modeling examining specific speech sound production could benefit from daylong home recording data. Finally, she will propose ST challenges that need to be overcome in order to maximize synergy between computational modeling and daylong home recording.

Elika Bergelson Bergelson will discuss aspects of data collection and analysis of daylong audio and hour long video recordings from her longitudinal corpus, SEEDLingS. Specifically focusing on object word acquisition from 6-18 months, Bergelson will discuss links between eyetracking data and home environment, and discuss which aspects of the variability in the input map onto variance in word comprehension and production in the first two years of life.

Kim Oller Oller will articulate his approach to large scale analysis of vocal data from infants. He will highlight cautions about coding infant vocalizations in terms of adult sound categories, emphasizing the fact that in the vast majority of cases in the first year infant utterances are neither phonemic nor lexical. Oller will thus underscore the challenge for computational analysis of infant vocal data: finding ways to monitor infant categories as they are progressively transformed from primitive or discoordinated phonatory/articulatory sequences into mature speech.

Annotation and Automatic Speech Recognition

Reiko Mazuka Mazuka will introduce a corpus of Japanese infant-directed speech (Riken-Japanese Mother-Infant Conversation Corpus), which is fully annotated, and time-aligned both for segmental and intonational details. She will discuss the challenges inherent in attaining expert phonetic annotation for a sizable naturalistic speech corpus, and the necessity of doing so for certain research questions.

Florian Metze Metze will introduce the Speech Recognition Virtual Kitchen, which is dedicated to improving community research and education infrastructure in speech technology. The goal of this work is to allow non-experts to access state of the art ST and Automatic Speech Recognition tools. The preconfigured virtual-machine-based "kitchen" provides the infrastructure for "appliances" (e.g., speech recognition toolkits), "recipes" (scripts for creating state-of-the art systems), and "ingredients" (language data).

Metze will introduce these resources, including demonstrations on users' machines. He will highlight how the "kitchen" can facilitate the development of audio related algorithms in research on child language acquisition using large corpora.

Computational Modelers of Cognitive Development

Okko Räsänen Räsänen will discuss challenges and opportunities in using large-scale audio/audio-visual recordings for computational models of speech perception and early language acquisition (LA). He will focus on how such data could enable a transition from the study of isolated aspects of LA to the development of integrated models (e.g., joint acquisition of sub-word and word representations, including semantic grounding of words to referential contexts). Räsänen will also discuss computational model evaluation, the related requirements this imposes on the metadata/annotation of recordings, and the challenges associated with the use of linguistic representations, such as phonemes or words, as proxies for learning targets in early stages of development.

Tove Gerholm Gerholm will discuss her group's longitudinal study of parent-child interaction from 3 months to 3 years of life. This video-corpus based work seeks to look at modalities across language acquisition, build a reusable corpus, and model interaction during development. Her talk will focus on an expansion of models that map sound strings to objects by adding in tactile and gestural information, and caregiver feedback. Gerholm will discuss model assumptions and highlight her groups modeling efforts that have met both failure and success, helping others with similar research goals.

Synthesis

Early Stage Work There will be a group discussion allowing early-career participants to describe ongoing or planned work, and solicit audience feedback.

Caitlin Fausey Fausey will briefly review findings from her work about the distributions of activities, contexts, and visual instances over the first two years of everyday life at home, addressing relevant inferential issues for learners building multimodal links among language, vision, and action. Fausey will raise discussion points about how coarse- and fine-grained daily rhythms interact in language learning, and how these activity-based rhythms change over the first two years. Finally, she will discuss potential links between distributions of daily events and ST performance.

Mark Liberman Liberman will discuss the present state and future prospects of corpus-based methods in speech and language science. Key questions include infrastructure for sharing data (including annotations of various types), the inventory of tools (and what's missing), and the skills researchers will need in order to participate fully in this area's trans-disciplinary future. Finally, he will discuss some examples, obvious and not so obvious, of the research opportunities created by new social and technological developments.

Acknowledgments

This work was supported by an NIH grant to Bergelson (DP5-OD019812-01) and by a collaborative NSF grant awarded to Anne Warlaumont (1539129), Mark VanDam (1539133) and Brian MacWhinney (1539010).