Building Language Resources for Exploring Autism Spectrum Disorders

Julia Parish-Morris¹, Christopher Cieri², Mark Liberman², Leila Bateman¹, Emily Ferguson¹, Robert T. Schultz²

¹Center for Autism Research, Children’s Hospital of Philadelphia
²Linguistic Data Consortium, University of Pennsylvania
Outline

- Autism
- Challenges
- Opportunities
- Prior research
- Current collaboration
- Future projects
Autism Spectrum Disorder

- Brain-based disorder typically identified in early childhood
  1.5% of U.S. children (CDC, 2016)

- Diagnostic criteria:
  - Impairments in social communication
  - Presence of repetitive behaviors or restricted patterns of interests

- “Spectrum” = mild to severe symptoms

- Significant public health cost

- Swift, accurate, early diagnosis is critical to improved outcomes

- Behaviorally defined: no brain scan or blood test

- Significant symptom overlap with other disorders

- Many children diagnosed late
PROBLEM:

sample heterogeneity +
small samples +
poor measurement =

non-reproducible scientific results
Opportunities

◆ Natural language interaction
  • Highly nuanced outward signal of internal brain activity
  • Fundamentally social

◆ Most children with ASD acquire language; nearly all vocalize

◆ Can HLT and Big Data methods help us identify ASD more reliably and understand it better?
Variable vocalization throughout development:
- Differences evident in infancy
- Language delay as toddlers/preschoolers
- Difficulty being understood & understanding humor, sarcasm
- Conversational quirks
  - unusual word use
  - turn-taking
  - synchrony
  - accommodation

Real-life effects of pragmatic language problems:
- Difficulty forming/maintaining friendships
- Increased risk of being bullied
- Difficulty with romantic relationships
- Difficulty maintaining employment
Early vocalization in ASD

- **4 mo**: fewer complex pitch contours during cooing (Brisson et al., 2014)
- **6 mo**: Higher and more variable F₀ in cries, poorer phonation (Orlandi et al., 2012; Sheinkopf et al., 2012)
- **9 mo**: Fewer well-formed babble sounds (Paul et al., 2011)
- **12 mo**: Less waveform modulation and more dysphonation in cries, compared to TD and DD (Esposito & Venuti, 2009)
- **16 mo**: fewer responses to parent vocalizations, especially when directing to people (Cohen et al., 2013)
- **18 mo**: Higher F₀ in cries, compared to TD and DD (Esposito & Venuti, 2010)
ASD speech communication:

- Many small variations accumulate to create an odd impression
- Difficulty to determine what exactly differs
- Difficult to recognize
Characterizations

- Too slow
- Too quiet
- Too fast
- Robotic
- Stilted
- Pedantic
- Disorganized
- "Little Professor"
- Flat
- Too loud
- Sing-songy
The truth?

- The generalizations in the literature are mostly impressions (or stereotypes....)
  - There are few empirical studies
  - Sample sizes are generally very small
- In fact:
  - The ASD phenotype is very diverse in speech communication as in other ways
  - The truth is probably neither a point nor a “spectrum” but a complex multidimensional multimodal distribution in a space that we all live in
- We don’t really know the dimensions of this space and figuring it out will take careful analysis of lots of data
Clinical Computational Linguistics

- Natural language:
  - Nuanced signal (marriage of cognitive and motoric systems)
  - Few practice effects

- Can automatically identify and extract features ("linguistic markers")

- Specific linguistic features associated with:
  - Depression
  - Dementia
  - PTSD
  - Schizophrenia

- ...Autism
Prior Research

On average, individuals with ASD have been found to:

- Produce idiosyncratic or unusual words more often than typically developing peers (Ghaziuddin & Gerstein, 1996; Prud’hommeaux, Roark, Black, & Van Santen, 2011; Rouhizadeh, Prud’Hommeaux, Santen, & Sproat, 2015; Rouhizadeh, Prud’hommeaux, Roark, & van Santen, 2013; Volden & Lord, 1991)

- Repeat words or phrases more often than usual (echolalia; van Santen, Sproat, & Hill, 2013)

- Use filler words “um” and “uh” differently than matched peers (Irvine, Eigsti, & Fein, 2016)

- Wait longer before responding in the course of conversation (Heeman, Lunsford, Selfridge, Black, & Van Santen, 2010)

- Produce speech that differs on pitch variables; these can be used to classify samples as coming from children with ASD or not (Asgari, Bayestehtashk, & Shafran, 2013; Kiss, van Santen, Prud’hommeaux, & Black, 2012; Schuller et al., 2013)
Collaboration

- Center for Autism Research (CAR)
  - autism expertise
  - data samples
- Linguistic Data Consortium (LDC)
  - corpus building methods
  - expertise in linguistics analysis
Process and analyze recorded language samples from Autism Diagnostic Observation Schedule ("ADOS"; Lord et al., 2012)

- Conversation and play-based assessment of autism symptoms
- Recorded for reliability and clinical supervision, coded on a scale, then filed away

- 600+ at CAR alone, thousands more across the U.S. and in Europe; never compiled

- Associated with rich metadata that includes family history, social, cognitive, and behavioral phenotype, genes, and neuroimaging
Pilot

Goals

- Assess feasibility
- Identify and extract linguistic features
- Machine learning classification and/or discovery of relevant dimensions
- Correlate features with clinical phenotype
Transcription

- Time aligned, verbatim, orthographic transcripts
  (~20 minutes of conversation per interview, from ADOS Q&A segment)
- New transcription specification developed by LDC,
  (adapted from previous conversational transcription specifications)
- 4 transcribers and 2 adjudicators from LDC and CAR produced a “gold standard” transcript for analysis and for evaluation/training of future transcriptionists

  ![Diagram of transcription process]

- Simple comparison of word level identity between CAR’s adjudicated transcripts and LDC’s transcripts: 93.22% overlap on average, before a third adjudication resolved differences between the two
- Forced alignment of transcripts with audio
Participants

- Pilot sample
- N=100
- Mean age=10-11 years
- Primarily male
- 65 ASD, 18 TD, 17 Non-ASD mixed clinical
- Average full scale IQ, verbal IQ, nonverbal IQ
Bag-of-words classification:

- Correctly classified 68% of ASD participants and 100% of TD participants
- Naïve Bayes, leave-one-out cross validation and weighted log-odds-ratios calculated using the “informative Dirichlet prior” algorithm (Monroe et al., 2008)
- Receiver Operating Characteristic (ROC) analysis revealed good sensitivity and specificity; AUC=85%
Word Choice

- 20 most “ASD-like” words:
  - \{nsv\}, know, he, a, now, no, uh, well, is, actually, mhm, w-, years, eh, right, first, year, once, saw, was
  - \{nsv\} stands for “non-speech vocalization”, meaning sounds that have no lexical counterpart, such as imitative or expressive noise
  - “uh” appears in this list, as does “w-”, a stuttering-like disfluency.

- 20 least “ASD-like” words:
  - like, um, and, hundred, so, basketball, something, dishes, go, york, or, if, them, \{laugh\}, wrong, be, pay, when, friends.
  - “um” appears, as does the word friends and laughter
Fluency

- Rates of um production across the ASD and TD groups (um/(um+uh))
- ASD group produced UM during 61% of their filled pauses (CI: 54%-68%)
- TD group produced UM as 82% of their filled pauses (CI: 75%-88%)
- Minimum value for the TD group was 58.1%, and 23 of 65 participants in the ASD group fell below that value.
Mean word duration as a function of phrase length

TD participants spoke the fastest (overall mean word duration of 376 ms, CI 369-382, calculated from 6891 phrases)

Followed by the non-ASD mixed clinical group (mean=395 ms; CI 388-401, calculated from 6640 phrases)

Followed by the ASD group with the slowest speaking rate (mean=402 ms; CI: 398-405, calculated from 24276 phrases)
Latency to Respond

- Characterizes gap between speaker turns
- Too short = interrupting or speaking over a conversational partner
- Too long (awkward silences) interrupts smooth exchanges
- ASD somewhat slower than TD
Fundamental Frequency

- Mean absolute deviation from the median (MAD)
  - Outlier-robust measure of dispersion in F0 distribution
  - Calculated in semitones relative to speaker’s 5th percentile

- MAD values are both higher and more variable within the ASD and non-ASD mixed clinical group than the TD group
  - ASD: median: 1.99, IQR: 0.95
  - Non-ASD: median: 1.95, IQR: 0.80
  - TD: median: 1.47, IQR: 0.26
Next Steps

❖ Expand sample sizes
  ● Improve classification metric
    ■ Focus on specificity (differentiate ASD from its cousins)
    ■ Identify relevant dimensions of variation
  ● Hone HLT for pediatric clinical population

❖ Emerging collaborations include more ADOS evals with phenotypic data, neuroimaging, and genetics
  ● Large body of shared data
  ● Goal: gene-brain-behavior mapping

❖ Enlarge age range
  ● Goal: downward extension to infancy
  ● Identify clusters of acoustic markers
  ● Chart growth to pinpoint critical points of divergence (targets for intervention)
We have subject consent and IRB clearance for publication of anonymized transcripts and audio.

Larger ADOS sample from CAR in process.

Possible multi-site project (like ADNI) to pool very large collection of existing ADOS interviews processed and analyzed to the same standard.

**BUT**

- New ADOS interviews require expensive, time-consuming in-person collection.
- **NEED**: Scalable, inexpensive methods to collect natural language from large, diverse samples.
Future Directions

- **Phone bank**
  - Inexpensive student worker asks ADOS questions
  - Child and parent language samples, questionnaires, online IQ
  - Nationally representative cohort
  - Longitudinal samples

- **Computerized Social Affective Language Task** (C-SALT)
  - Self-contained laptop-based audio/video collection
  - Records language and social affect in schools, clinics and homes
  - Controlled recording is conducive to automated approaches (reduces need for transcription)

- **Combine data sources to improve predictive power:**
  - Motor, language, medical records, parent/teacher report, clinical judgment, performance tasks, imaging, genetics
CAR and LDC are eager to collaborate:

looking for novel analytic approaches

and outside-the-box ideas!
Applications

- **Support clinical decision-making and improve access**
  - Low-cost, remote screening
  - Direct behavioral observation: record in clinics, integrate into EHR
  - Inform identification efforts and assist in differential diagnosis

- **Identify behavioral markers of underlying (treatable) pathobiology**
  - Profiles of individual strengths and weaknesses link to biology = personalized treatment planning and improved outcomes
  - Granular assessment of response to intervention – dense sampling

- **Give participants and families more information about themselves**
  - Online feedback
  - Monitor growth trajectories
Acknowledgements

- Participants and Families
- Clinicians, research, staff from CAR and LDC
- Funding sources
  - Autism Science Foundation
  - McMorriss Autism Program
  - NIH K12