Building Language Resources for Exploring Autism Spectrum Disorders

Julia Parish-Morris[*], Christopher Cieri[°], Mark Liberman[°], Leila Bateman[*], Emily Ferguson[*], Robert T. Schultz[*]

[°]Linguistic Data Consortium. University of Pennsylvania
3600 Market Street, Suite 810, Philadelphia, PA, 19104 USA

[*]Center for Autism Research. Children's Hospital of Philadelphia
3535 Market Street, Suite 860, Philadelphia, PA, 19104 USA

Corresponding Author: julia.parish.morris AT gmail.com

Abstract

Autism spectrum disorder (ASD) is a complex neurodevelopmental condition that would benefit from low-cost and reliable improvements to screening and diagnosis. Human language technologies (HLTs) provide one possible route to automating a series of subjective decisions that currently inform "Gold Standard" diagnosis based on clinical judgment. In this paper, we describe a new resource to support this goal, comprised of 100 20-minute semi-structured English language samples labeled with child age, sex, IQ, autism symptom severity, and diagnostic classification. We assess the feasibility of digitizing and processing sensitive clinical samples for data sharing, and identify areas of difficulty. Using the methods described here, we propose to join forces with researchers and clinicians throughout the world to establish an international repository of annotated language samples from individuals with ASD and related disorders. This project has the potential to improve the lives of individuals with ASD and their families by identifying linguistic features that could improve remote screening, inform personalized intervention, and promote advancements in clinically-oriented HLTs.

Keywords: language resources, collection, annotation, data distribution, quality control, autism spectrum disorder

1. Introduction

Autism Spectrum Disorder (ASD) is a brain-based developmental condition that affects a growing number of individuals across the globe (Baxter et al., 2015; Elsabbagh et al., 2012). In the U.S., approximately 1 in 68 school children are identified having an ASD (Blumberg et al., 2013), with prevalence ranging from 1 in 38 in South Korea (Elsabbagh et al., 2012), to 1 in 100 in Iceland (Saemundsen, Magnússon, Georgsdóttir, Egilsson, & Rafnsson, 2013), 1 in 115 in Mexico (Fombonne et al., 2016), 1 in 124 in Sweden (Gillberg, Cederlund, Lamberg, & Zeijlon, 2006), 1 in 146 in Denmark (Parner et al., 2011), 1 in 196 in Western Australia (Parner et al., 2011), and 1 in 263 in the U.K. (Taylor, Jick, & MacLaughlin, 2013). These relatively high numbers have significant economic consequences. In 2014, the annual public health cost of ASD in the United States was projected to reach into the billions of dollars (Lavelle et al., 2014) and the lifetime per capita incremental societal cost of ASD was estimated to be nearly $2.5 million in both the U.S. and the U.K. (Buescher, Cidav, Knapp, & Mandell, 2014).

Access to swift, accurate, low-cost diagnosis is one of the most significant challenges in autism today, and could be greatly aided by targeted advancements in Human Language Technologies (HLT). Consensus practice parameters recommend multidisciplinary assessment of children suspected of having ASD (Volkmar et al., 2014).

Of course, many children across the world do not have access to a single healthcare provider that is knowledgeable about ASD, much less a highly trained interdisciplinary team of providers (Samms-Vaughan, 2014; Tomlinson et al., 2014). Even in developed countries, access to care is limited for a large proportion of children, resulting in late or missed diagnoses (Daniels & Mandell, 2013). The issue of late or missed diagnosis is not trivial; early, intensive social and behavioral intervention has been repeatedly shown to improve long-term outcomes in children with ASD (Ben Itzchak & Zachor, 2009; Howlin, Magiati, & Charman, 2009; Remington et al., 2007), which likely reduces lifetime cost-of-care. Diagnostic challenges have proved difficult to solve, however, in part because ASD is behaviorally defined and has symptoms that overlap with other disorders (Grzadzinski, Dick, Lord, & Bishop, 2016). There is no blood test or brain scan to facilitate rapid diagnosis in autism. Rather, clinicians must rely on time-intensive, in-person clinical assessments (Wiggins et al., 2015). Moreover, even highly trained experts disagree with one another about whether or not an individual meets criteria (Gabrielsen et al., 2015; Westman Andersson, Miniscalco, & Gillberg, 2013).

There is considerable variability in ASD severity and in the clinical profile of those diagnosed with ASD. The notorious heterogeneity of ASD is leading to a revision in thinking about its nosological status; there is considerable interest in the scientific community in exploring

dimensional approaches to understanding mental disorders, as an alternative to categories (Insel, 2014). Similarly, recent questions about whether the autism spectrum should be viewed as the tail of a Gaussian distribution of human variation have led to important discussions about neurodiversity, including when/whether/how to address independence and functional impairment (Armstrong, 2015; Kenny et al., 2015; Odom, 2016).

Diagnosing autism in an accurate, reproducible way throughout the world represents a significant challenge for clinicians. One tool, the Autism Diagnostic Observation Schedule (ADOS), is a widely adopted (Kim et al., 2011), semi-structured behavioral observation used to aid in clinical decision-making (Lord et al., 2012). For younger children, the ADOS evaluation provides an opportunity to play with toys and tell stories that might reveal the social communication impairments and repetitive behaviors indicative of ASD. In older verbal individuals, the ADOS includes a conversation similar in form and content to the interviews that have been the focus of prior HLT research, but focused on social-emotional concerns. The first edition of the ADOS was published in English, Danish, Dutch, Finnish, French, German, Hebrew, Hungarian, Icelandic, Italian, Korean, Norwegian, Romanian, Russian, Spanish, and Swedish; to date, the second has been translated to Czech, Danish, Dutch, Finnish, French, German, Italian, Norwegian, and Swedish (Lord et al., 2012).

2. The Case for Developing a Shared Resource

As part of the diagnostic decision-making toolbox for a complex disorder like ASD, ADOS evaluations are routinely recorded (for training and reliability purposes) (Lord et al., 2012). There are many thousands of recorded evaluations in the U.S. alone, and thousands more across the globe in multiple languages. Many of these recordings are associated with clinical metadata such as age, sex, clinical judgment of ASD status, autism severity metrics, IQ estimates, and social/language questionnaires, as well as genetic panels, brain scans, behavioral experiments, and infrared eye tracking. The quality of these audiovisual recordings is variable, with a multitude of recording methods employed. Importantly, these recordings have never been assembled into a large, shareable resource. **We view this as a massive, untapped opportunity for data sharing and clinically oriented advancements in HLT research.** Indeed, a review of language-related questions and scores in the ADOS revealed a number of subjective decisions that clinicians must make, including some that seem to be susceptible to automation using HLT.

At the Children's Hospital of Philadelphia Center for Autism Research (CAR), we have collected data from more than 1200 toddlers, children, teens, and adults, most of which were ultimately diagnosed with ASD. We conducted deep phenotyping with most of our participants, in the form of interviews and questionnaires, cognitive and behavioral assessments, brain scans, eye

tracking, and genetic tests. Importantly, this richly characterized data set is accompanied by language samples from the ADOS evaluation recordings. In 2013, CAR and the Linguistic Data Consortium (LDC) established a collaboration to leverage this untouched resource. Our initial goal was to determine whether automated analysis of language recorded during the ADOS could predict diagnostic status, although our aims have since expanded to include identifying correlates of phenotypic variability within ASD. This second aim is particularly meaningful in the clinical domain; if we can accurately and objectively quantify the linguistic signal, we have a much better chance of reliably mapping it to real-world effects.

The current paper reports on our work-in-progress. Here, we hope to spur discussion about data and methods in this area, describe inter-annotator agreement, get feedback on our workflow, and describe efforts toward growing and sharing valuable resources like this one.

3. Prior Research in ASD

The search for automated, language-based methods of identifying ASD is gaining momentum. In 2013, *Interspeech* issued a challenge: develop an algorithm to discriminate ~2,500 short (read) language samples from 9- to 18-year-old children in the French-language Child Pathological Speech Database (Schuller et al., 2013). Thirty-five out of 99 children had clinical diagnoses of ASD, specific language impairment, or pervasive developmental disorder – not otherwise specified. The winning proposal used voice quality features including Harmonic-to-Noise ratio, shimmer, and jitter, along with standard features such as energy, cepstral, and spectral features, to classify clinical samples (Asgari, Bayestehtashk, & Shafran, 2013). When the same algorithm was applied to the task of distinguishing ASD from other developmental disorders, however, discrimination power dropped significantly. This highlights the need for diagnostic algorithms that capture features specific to ASD.

A recent series of studies by Van Santen and colleagues approaches this goal by analyzing language produced by 146 4- to 9-year-old children with clinical diagnoses during ADOS evaluations. Samples from children with ASD contained different pitch and language features than samples from children with typical development and, in some cases, than children with specific language impairment (Kiss, van Santen, Prud'hommeaux, & Black, 2012; Prud'hommeaux, Roark, Black, & Van Santen, 2011). Follow-up work using a machine-learning approach on coded speech errors resulted in good discrimination between diagnostic groups (Receiver Operating Characteristic (ROC) area under the curve (AUC) <.75) (Morley, Roark, & Van Santen, 2013).

Our research extends this emergent literature by exploring language produced by school-aged children and teens with and without ASD, during the conversation and reporting section of the ADOS. Within the LDC/CAR collaboration our goals were: 1) create a Pilot corpus for

linguistic analysis of speech from participants with and without ASD, 2) develop analytic methods that correlate linguistic form with clinical features, 3) assuming success in 1 & 2, develop a scalable methodology for corpus creation and extend the corpus, and 4) expand our research goals into new interaction types and analytics and more refined investigations into the ASD phenotype.

Given the unique constraints associated with assembling a corpus from a pediatric clinical population, we carefully describe our data selection and processing, our transcription protocol, quality control procedures, and preliminary results based on a pilot subsample.

4. Data Selection and Processing

We processed two cohorts of data: a *Pilot* corpus and an *Extension* corpus. The Pilot corpus included 46 children/teens: 18 with classic ASD, 14 with symptoms on the less severe end of the spectrum or other diagnoses (a "mixed clinical" group), and 14 with typical development (TD). For this exploratory sample, we matched participants on sex, age, and IQ measured via the Differential Abilities Scales, Second Edition, (Elliott, 2007). The Extension corpus included an additional 47 participants with ASD, 4 with non-ASD mixed clinical diagnoses, and 3 with typical development, selected to have verbal and nonverbal IQ scores above 80 but otherwise unmatched. In the sample as a whole (N=100), 74% of participants were male and 76% identified as Caucasian. Average age was 10.24 years (ASD: 9.96, TD: 10.21, mixed clinical: 11.32) and average IQ was 103 (ASD: 105, TD: 104, mixed clinical: 98). Research reliable PhD level clinical psychologists and/or psychology trainees administered the ADOS module 3 to all participants.

After we obtained consent from participants to use their sessions for research purposes, entire video recordings were copied from their original media onto a shared file system accessible only to project members with current certifications for research on human subjects. Audio was extracted from the video stream and saved in lossless FLAC format. Except for extraction and format conversion, the data was identical to the original recording.

5. Transcription

Transcription teams at LDC and CAR created time aligned, verbatim, orthographic transcripts of the conversation and reporting section of the ADOS evaluation for each participant (mean length of transcribed section ~20 minutes). The LDC transcription team consisted of two junior and two senior transcribers, all college educated native speakers of American English. The junior transcribers performed segmentation of the audio files into pause groups and transcription. The senior transcribers corrected the initial transcripts and occasionally did transcription from scratch.

For this effort, LDC created a new transcription specification that resembles those used for conversational speech. The principal differences are that the current specification requires that participants be labeled only by their role (Interviewer and Participant) and that the boundaries between speech and non-speech be placed rather accurately because (inter-)turn duration is a factor of interest.

Once the Pilot proved successful, CAR developed a team to extend the corpus and also begin evaluating inter-annotator agreement. The CAR team consists of multiple pairs of college educated native speakers of American English that transcribe the conversation and reporting section of the ADOS independently, a third more senior transcriber responsible for comparing and adjudicating the work of the first two, and a fourth transcriber who compares CAR and LDC transcripts when the latter are available, and adjudicates remaining disagreements (Figure 1). In this way, 4 transcribers and 2 adjudicators with complementary goals produce a "gold standard" transcript for analysis and for evaluation/training of future transcriptionists.

6. Quality Control

LDC transcribed 52 files, and CAR transcribed 100 including independent transcriptions of the 52 that LDC transcribed. A simple comparison of word level identity between CAR's adjudicated transcripts and LDC's transcripts revealed 93.22% overlap on average, before a third adjudication resolved differences between the two. In the case of files that were transcribed by CAR only (N=48), pre-adjudication overlap in word-level comparisons between transcribers averaged 92.18%. We are confident that two or three complete transcriptions plus one or two complete adjudications has resulted in a reliable data set.

Figure 1. Transcription and adjudication flow for LDC-transcribed files (Transcriber 1, Transcriber 2) and CAR-transcribed files (Transcriber 3, Transcriber 4), with final adjudication performed on files transcribed by both teams.
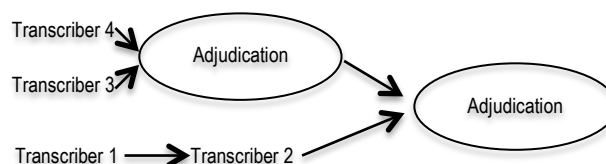
7. Initial Analyses

The brief results reported here are based on data from our Pilot sample. As a proof of concept, we aimed to classify the lowest hanging fruit: clear ASD vs. clear typical development. This was the least challenging discrimination to make (differentiating between language produced by a child with ASD and language produced by the mixed clinical group is likely to be much harder). Beginning with this simple distinction allowed us to explore the basic potential of the conversation and reporting section of the ADOS to discriminate between diagnostic groups, and to examine correlations between

speech-language features and a widely used measure of social skills (the Social Responsiveness Scales) (Constantino et al., 2003).

Classification. Our preliminary effort to classify participants using the ASD and TDC groups from the Pilot corpus used a well-matched sample [ASD: N=18, Mean age=11.14 years (SD=2.25, range=6-14), Mean IQ=106 (SD=14.44, range=78-131); TDC: N=14, Mean age=11.14 years (SD=1.62; range=8-14), Mean IQ=108 (SD=13.19, range=82-123)]. Using weighted log-odds calculations with leave-one-out-cross-validation, we found that Naïve Bayes classification on the basis of word choice alone correctly classified 78% of ASD patients and 100% of typical participants. Receiver Operating Characteristic (ROC) analyses revealed high sensitivity and specificity using this classification metric, with AUC=92%, CIs 82%-100%, *p*<.001 (Figure 2).

Figure 2. Receiver operating characteristic on word choice separates diagnostic groups.
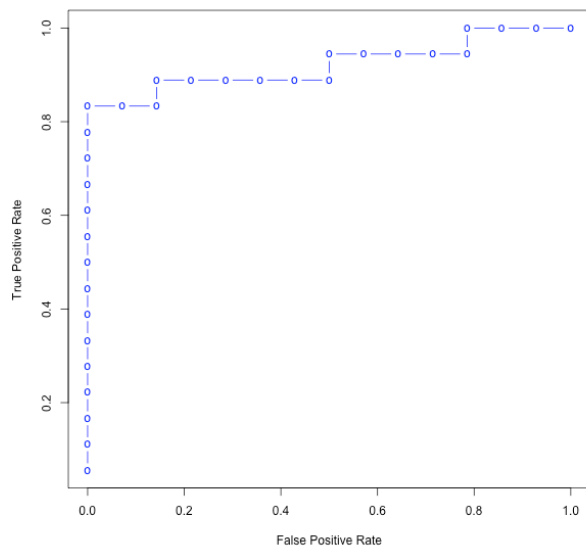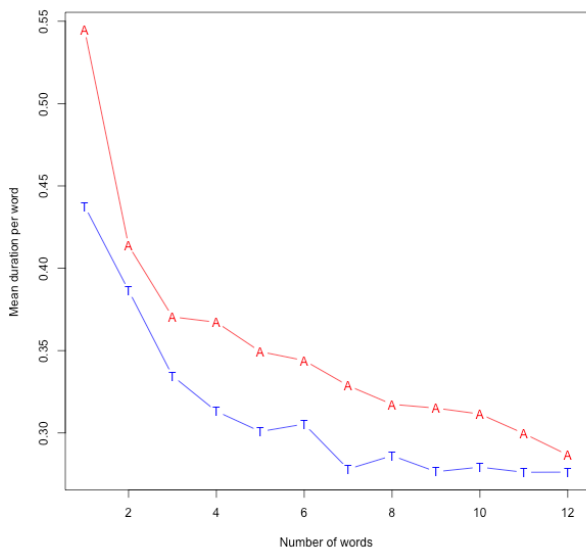

Figure 3. Speaking rate: turn length (in words) versus mean turn duration.

Correlations between linguistic features and clinical phenotype. We observed differences in speaking rate (Figure 3) and inter-turn gaps between the ASD and typical samples (Figure 4). Pearson tests revealed significant negative correlations between speaking rate, rate of conversational turn-taking, and overall speaking time with autism symptom severity as measured by the Social Responsiveness Scales (*r*s range from -.34 to -.44, all *p*s<.01 (Constantino et al., 2003)) but did not reveal significant correlations with IQ or age (the Pilot corpus was matched on these variables). In light of our small initial sample, these results are highly encouraging.

Figure 4. Inter-turn pauses during conversation (Blue=TD, Peach=ASD, Red=Overlap).
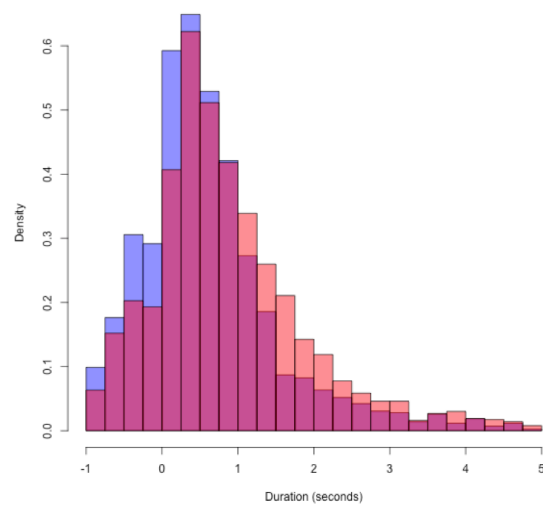
8. Discussion

In this paper, we described a new resource for developing HLTs aimed at children with ASD, and presented data showing its potential clinical utility. Using a small pilot sample, we demonstrated that language produced during the conversation and reporting section of the ADOS can:

1. Classify children as being on the autism spectrum (or not) with high sensitivity and specificity; and
2. Relate meaningfully to parent and clinician ratings of social phenotype.

These preliminary results have a variety of implications. First, although automated classification is not designed to replace expert clinical judgment, it could serve as a valuable and objective additional guide to inform the complex set of decisions made by diagnosticians. Many clinical teams are forced to maintain long evaluation wait lists, and families often wait months or even a year before experts can assess their children. In later stages of HLT development, automated classification using easily acquired natural language data could be a first step toward giving answers to concerned families.

Remote screening is a second possible application for technologies trained on this shared resource. Clinical teams often employ some form of initial screening to determine whether full diagnostic testing is necessary for

a given child (e.g., a parent history questionnaire or ASD symptom checklist). In the future, remote language-based screening that produces "high risk" or "low risk" classifications could help identify children that would benefit most from follow-up evaluations with humans. Ideally, an automated algorithm could result in a multi-level risk score indicating high-priority cases with especially concerning patterns. This could be especially valuable for geographically isolated or underserved demographic groups that do not have ready access to expert in-person screening. Future downward extensions of this approach could lead to remotely ascertained vocalization patterns tracked from birth, that could produce automated risk indices set to alert pediatricians on the lookout for abnormal developmental patterns. Early identification could lower the age of treatment onset, thus optimizing long-term outcome and enabling more individuals with ASD to achieve their full potential.

Third, linguistic data gathered from natural language samples could help researchers in their quest to understand the biological basis of ASD. ASD is multi-determined and multi-faceted, with some individuals affected only mildly and others profoundly impaired. Applying computational linguistics approaches to the problem of characterization will result in much finer-grained behavioral indicators of underlying brain and genetic differences than say, a 5-point Likert scale that collapses across a variable skill set or heterogeneous feature cluster. The development of a collaborative, multi-site, international repository of ADOS evaluations has significant potential to advance this goal, as many research efforts included biosample collection and brain imaging along with gold standard diagnostic assessments.

A fourth implication of our preliminary data lies in the potential for making personalized treatment recommendations. The fine-grained behavioral data associated with careful quantification could be used to create individualized profiles of linguistic strengths and weaknesses. Speech-language pathologists and other practitioners could use these profiles when setting treatment goals and personalizing plans for intervention. These profiles of strengths and weaknesses might then serve as benchmarks to measure treatment efficacy. In addition, since the input for these profiles would be natural language samples rather than standardized tests, the usual test-retest complications and practice effects will be eliminated, vastly improving our chances of detecting real effects of treatment.

Finally, using advancements in HLT to reduce the human burden associated with complicated and time-consuming transcription and coding could improve our ability to track individual development over time. Sensitive periods of growth (e.g., 0-5 years, puberty, the transition to adulthood) are times of rapid change that, when development goes awry, result in cascading long-term effects. These time periods could be densely sampled using automated approaches, and assessed for trajectories of change that might differ by diagnostic status, suggesting unique paths to intervention, and allowing

individuals to track their own developmental courses. This represents an important potential step toward putting the science of autism research back in the hands of individuals on the spectrum on their families.

9. Current Limitations

Our first attempts to analyze this data were necessarily limited. For example, due to small sample sizes, we matched our participants on age, sex, and IQ. This attempt to control for as many "extraneous" features as possible allowed us to isolate specific effects of ASD on language produced during the ADOS. Although this approach has many advantages with a small data set, it also limits our ability to generalize our findings to samples that are not comparable to our own. In addition to limiting generalization, this approach may obscure important contributions of variables such as sex (which we know is related to speech/language differences) and IQ (which may covary importantly with autism severity and functional outcome). Age is another critical unknown; typical children's language grows and changes over time in ways that are well understood from a cognitive and linguistic perspective, but less well characterized from the perspective of HLTs. Children with ASD may develop subtle linguistic features in a different order or with different qualities than their typically developing counterparts.

A myriad other outstanding questions have been sparked by our preliminary research. For example, we still know very little about interlocutor effects on child language during the ADOS; how does a clinician's level of ASD expertise affect child language features? Can untrained conversational partners ask ADOS questions and still elicit language that differs by diagnostic category? Can we select the most predictive questions from the conversation and reporting section, using those as a "short form" to classify participants with comparable accuracy, or is the entire section necessary? Might it be possible to collect the ADOS conversation and reporting sections via internet or telephone? Is there an added benefit to including full-day, naturalistic recordings from wearable technology in our algorithms? Exploring these exciting questions demands a larger and more varied data set than we currently have.

Despite these limitations, our findings are promising; they replicate and extend prior research showing that computational approaches to assessing language produced during ADOS evaluations can result in a clinically useful correlation. In fact, the initial analyses reported here likely underestimate the characterization power of complex speech/language signals, as we have yet to include in our algorithm aspects of voice quality that have been found to discriminate diagnostic groups in the past (Schuller et al., 2013). We hope that our preliminary findings spark renewed interest in assembling, processing, and disseminating de-identified language samples from ADOS evaluations into a joint international repository that can be disseminated widely, for the benefit of all.

10. Future Directions

At LDC and CAR, we envision a future where language samples from individuals with neurodevelopmental disorders are processed using automated algorithms developed in tight collaboration between linguists and clinicians. These algorithms will produce personalized profiles of linguistic strengths and weaknesses, and will inform targeted intervention. Careful clinical phenotyping combined with linguistic analysis on large, diverse samples may ultimately provide information about the treatment options that are most likely to benefit a given child, and may shed light on the underlying biology of autism. Understanding longitudinal trajectories and the effects of developmental change on linguistic output will be key to identifying and treating a complex, lifelong disorder like ASD.

We are in the process of extending this Pilot work along three dimensions: First, sample size. The preliminary sample described here is small, especially given that ADOS evaluations are collected every day throughout the world. Currently, transcription of a third CAR cohort is underway (consent is being obtained from N=234 participants with ASD or other clinical diagnoses, phenotypic data, and previously recorded ADOS evaluations), and a new collaborative sample is in the works (N=150 typical participants and N=150 with ASD or language impairment). Our goal is to recruit an ever-growing number of researchers and clinicians to join our effort, as we strive to establish a joint repository of de-identified ADOS recordings and high-quality annotations.

Our second dimension is sample composition: We recognize that classifying children as having ASD or typical development is not as difficult as distinguishing between ASD and other disorders. Future samples will include a mixed clinical population to train our classifier, including participants with anxiety, ADHD, depression, specific language impairment, intellectual disability, and brain trauma, all of which are available through the Children's Hospital of Philadelphia and collaborators. Our ultimate aim is to develop a pediatric clinical language sample repository that includes all of these challenging cases and more.

Third, algorithm composition: our first machine learning classification algorithm relied on word choice alone to determine group membership, and was surprisingly successful. However, our working hypotheses about ASD include linguistic features such as turn-taking and the distribution of inter-turn intervals, speaking rate, the distribution of speech and silence intervals, intonational prosody, co-articulation, n-gram frequency, syntactic structure, and contraction use, among other variables. We are now annotating our data for all of these variables, in most cases by methods that are fully automatic (though human-checked) given an accurate time-aligned transcript. Thus future research, based on larger data sets as well as larger sets of behavioral variables, will offer a better basis for theoretical investigation as well as more effective methods for diagnostic classification.

Finally, we want this effort to be scalable, which means that we eventually need to obtain our primary data outside the lab. We are currently designing remotely deployable solutions to collecting ADOS conversations and comparable language samples using inexpensive confederates and computer programs.

11. De-Identification and Distribution

The ultimate goal of this project is dissemination of a valuable resource to the HLT community. However, working with clinically sensitive pediatric populations makes data sharing slightly more complicated than usual. As with most linguistic data collected in a clinical context, ADOS session recordings and their transcripts are constrained not only by our university's Institutional Review Board and by the informed consent documents that patients or their parent(s) sign prior to visiting CAR, but also by the U.S. Health Insurance Portability and Accountability Act (HIPAA). Given the collaborative goals of this project, CAR staff is contacting families to request their consent to share data beyond the current study. To satisfy HIPAA requirements, our team is developing a protocol to remove all personally identifying and sensitive information from audio files and associated transcripts. This information includes: proper names (people, places, pets, etc.), ages, elements of date including date of birth and date of visit, age, geographic subdivisions (e.g., address, city, county, zip code), telephone numbers, e-mail addresses, and any other information that could be used alone or in combination to identify a unique individual. Once these identifiers are removed, transcripts and audio files will be shared as part of the LDC database (anticipated Fall 2016).

In the future, we hope to join with like-minded researchers at other institutions in creating a large multi-national collection of ADOS recordings that can be shared for research purposes (with appropriate safeguards), following the model of the Alzheimer's Disease NeuroImaging Initiative ("ADNI Home," n.d.). This repository will require genuine informed consent and careful removal of personally-identifying information, as well as agreement by recipients not to attempt de-anonymization or other inappropriate use. Our joint efforts will encourage resource sharing as a way to improve reproducibility and raise the overall performance of relevant technologies, with the goal of improving outcomes for individuals with developmental challenges. Every year, thousands of ADOS interviews are carried out around the world, by trained interviewers in controlled settings. A sample of this material, carefully transcribed and consistently annotated, will be an extraordinary force for research progress.

12. Acknowledgements

13. References

ADNI Home. (n.d.). Retrieved October 14, 2015, from http://www.adni-info.org/

Armstrong, T. (2015). The Myth of the Normal Brain: Embracing Neurodiversity. *AMA Journal of Ethics*, *17*(4), 348. http://doi.org/10.1001/journalofethics.2015.17.4. msoc1-1504

Asgari, M., Bayestehtashk, A., & Shafran, I. (2013). Robust and accurate features for detecting and diagnosing autism spectrum disorders. In *INTERSPEECH* (pp. 191–194). Retrieved from http://www.cslu.ogi.edu/~zak/is13-autism.pdf

Baxter, A. J., Brugha, T. S., Erskine, H. E., Scheurer, R. W., Vos, T., & Scott, J. G. (2015). The epidemiology and global burden of autism spectrum disorders. *Psychological Medicine*, *45*(03), 601–613. http://doi.org/10.1017/S003329171400172X

Ben Itzchak, E., & Zachor, D. A. (2009). Change in autism classification with early intervention: Predictors and outcomes. *Research in Autism Spectrum Disorders*, *3*(4), 967–976. http://doi.org/10.1016/j.rasd.2009.05.001

Blumberg, S. J., Bramlett, M. D., Kogan, M. D., Schieve, L. A., Jones, J. R., & Lu, M. C. (2013). Changes in prevalence of parent-reported autism spectrum disorder in school-aged US children: 2007 to 2011–2012. *National Health Statistics Reports*, *65*(20), 1–7.

Buescher, A. V. S., Cidav, Z., Knapp, M., & Mandell, D. S. (2014). Costs of Autism Spectrum Disorders in the United Kingdom and the United States. *JAMA Pediatrics*, *168*(8), 721. http://doi.org/10.1001/jamapediatrics.2014.210

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., … Reich, W. (2003). Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised. *Journal of Autism and Developmental Disorders*, *33*(4), 427–433.

Daniels, A. M., & Mandell, D. S. (2013). Explaining differences in age at autism spectrum disorder diagnosis: A critical review. *Autism*, 1362361313480277.

Elliott, C. D. (2007). Differential Ability Scales®-II - DAS-II. San Antonio, TX: Harcourt Assessment. Retrieved from http://www.pearsonclinical.com/education/produ cts/100000468/differential-ability-scales-ii-das-ii.html

Elsabbagh, M., Divan, G., Koh, Y.-J., Kim, Y. S., Kauchali, S., Marcín, C., … Fombonne, E. (2012). Global Prevalence of Autism and Other Pervasive Developmental Disorders: Global epidemiology of autism. *Autism Research*, *5*(3), 160–179. http://doi.org/10.1002/aur.239

Fombonne, E., Marcin, C., Manero, A. C., Bruno, R., Diaz, C., Villalobos, M., … Nealy, B. (2016). Prevalence of Autism Spectrum Disorders in Guanajuato, Mexico: The Leon survey. *Journal of Autism and Developmental Disorders*. http://doi.org/10.1007/s10803-016-2696-6

Gabrielsen, T. P., Farley, M., Speer, L., Villalobos, M., Baker, C. N., & Miller, J. (2015). Identifying Autism in a Brief Observation. *Pediatrics*, *135*(2), e330–e338. http://doi.org/10.1542/peds.2014-1428

Gillberg, C., Cederlund, M., Lamberg, K., & Zeijlon, L. (2006). Brief Report: "The Autism Epidemic". The Registered Prevalence of Autism in a Swedish Urban Area. *Journal of Autism and Developmental Disorders*, *36*(3), 429–435. http://doi.org/10.1007/s10803-006-0081-6

Grzadzinski, R., Dick, C., Lord, C., & Bishop, S. (2016). Parent-reported and clinician-observed autism spectrum disorder (ASD) symptoms in children with attention deficit/hyperactivity disorder (ADHD): implications for practice under DSM-5. *Molecular Autism*, *7*(1). http://doi.org/10.1186/s13229-016-0072-1

Howlin, P., Magiati, I., & Charman, T. (2009). Systematic Review of Early Intensive Behavioral Interventions for Children With Autism. *American Journal on Intellectual and Developmental Disabilities*, *114*(1), 23–41. http://doi.org/10.1352/2009.114:23-41

Insel, T. R. (2014). The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry. *American Journal of Psychiatry*, *171*(4), 395–397.

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2015). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism*, 1362361315588200. http://doi.org/10.1177/1362361315588200

Kim, Y. S., Leventhal, B. L., Koh, Y.-J., Fombonne, E., Laska, E., Lim, E.-C., … others. (2011). Prevalence of autism spectrum disorders in a total population sample. *American Journal of Psychiatry*. Retrieved from http://ajp.psychiatryonline.org/doi/abs/10.1176/a ppi.ajp.2011.10101532

Kiss, G., van Santen, J. P., Prud'hommeaux, E. T., & Black, L. M. (2012). Quantitative Analysis of Pitch in Speech of Children with Neurodevelopmental Disorders. In *INTERSPEECH*. Retrieved from http://20.210-193-52.unknown.qala.com.sg/archive/archive_papers /interspeech_2012/i12_1343.pdf

Lavelle, T. A., Weinstein, M. C., Newhouse, J. P., Munir, K., Kuhlthau, K. A., & Prosser, L. A. (2014). Economic burden of childhood autism spectrum disorders. *Pediatrics*, *133*(3), e520–e529.

Lord, C., Rutter, M., DiLavore, P. S., Risi, S., Gotham, K., & Bishop, S. L. (2012). Autism diagnostic observation schedule, second edition (ADOS-2). Torrance, CA: Western Psychological Services.

Morley, E., Roark, B., & Van Santen, J. (2013). The utility of manual and automatic linguistic error codes for identifying neurodevelopmental disorders. In *Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 1–10). Retrieved from http://www.aclweb.org/old_anthology/W/W13/W13-17.pdf#page=17

Odom, S. L. (2016). Steve Silberman: NeuroTribes: The Legacy of Autism and the Future of Neurodiversity. *Journal of Autism and Developmental Disorders*. http://doi.org/10.1007/s10803-016-2726-4

Parner, E. T., Thorsen, P., Dixon, G., de Klerk, N., Leonard, H., Nassar, N., … Glasson, E. J. (2011). A Comparison of Autism Prevalence Trends in Denmark and Western Australia. *Journal of Autism and Developmental Disorders*, *41*(12), 1601–1608. http://doi.org/10.1007/s10803-011-1186-0

Prud'hommeaux, E. T., Roark, B., Black, L. M., & Van Santen, J. (2011). Classification of atypical language in autism. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics* (pp. 88–96). Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=2021106

Remington, B., Hastings, R. P., Kovshoff, H., Degli Espinosa, F., Jahr, E., Brown, T., … MacLean, W. E. (2007). Early intensive behavioral intervention: outcomes for children with autism and their parents after two years. *American Journal on Mental Retardation*, *112*(6), 418–438.

Saemundsen, E., Magnússon, P., Georgsdóttir, I., Egilsson, E., & Rafnsson, V. (2013). Prevalence of autism spectrum disorders in an Icelandic birth cohort. *BMJ Open*, *3*(6), e002748.

Samms-Vaughan, M. E. (2014). The status of early identification and early intervention in autism spectrum disorders in lower- and middle-income countries. *International Journal of Speech-Language Pathology*, *16*(1), 30–35. http://doi.org/10.3109/17549507.2013.866271

Schuller, B., Steidl, S., Batliner, A., Vinciarelli, A., Scherer, K., Ringeval, F., … others. (2013). The INTERSPEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. Retrieved from http://eprints.gla.ac.uk/93665/

Taylor, B., Jick, H., & MacLaughlin, D. (2013). Prevalence and incidence rates of autism in the UK: time trend from 2004–2010 in children aged 8 years. *BMJ Open*, *3*(10), e003219.

Tomlinson, M., Yasamy, M. T., Emerson, E., Officer, A., Richler, D., & Saxena, S. (2014). Setting global research priorities for developmental disabilities, including intellectual disabilities and autism. *Journal of Intellectual Disability Research: JIDR*, *58*(12), 1121–1130. http://doi.org/10.1111/jir.12106

Volkmar, F., Siegel, M., Woodbury-Smith, M., King, B., McCracken, J., & State, M. (2014). Practice Parameter for the Assessment and Treatment of Children and Adolescents With Autism Spectrum Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(2), 237–257. http://doi.org/10.1016/j.jaac.2013.10.013

Westman Andersson, G., Miniscalco, C., & Gillberg, C. (2013). Autism in Preschoolers: Does Individual Clinician's First Visit Diagnosis Agree with Final Comprehensive Diagnosis? *The Scientific World Journal*, *2013*, 1–7. http://doi.org/10.1155/2013/716267

Wiggins, L. D., Reynolds, A., Rice, C. E., Moody, E. J., Bernal, P., Blaskey, L., … Levy, S. E. (2015). Using Standardized Diagnostic Instruments to Classify Children with Autism in the Study to Explore Early Development. *Journal of Autism and Developmental Disorders*, *45*(5), 1271–1280. http://doi.org/10.1007/s10803-014-2287-3