INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Objective Language Feature Analysis in Children with Neurodevelopmental Disorders during Autism Assessment

Manoj Kumar[1], Rahul Gupta[1], Daniel Bone[1], Nikolaos Malandrakis[1], Somer Bishop[2], Shrikanth Narayanan[1]

[1]Signal Analysis and Interpretation Laboratory (SAIL), USC, Los Angeles, USA
[2]Department of Psychiatry, UCSF School of Medicine, San Francisco, USA

Abstract

Lexical planning is an important part of communication and is reflective of a speaker's internal state that includes aspects of affect, mood, as well as mental health. Within the study of developmental disorders such as autism spectrum disorder (ASD), language acquisition and language use have been studied to assess disorder severity and expressive capability as well as to support diagnosis. In this paper, we perform a language analysis of children focusing on word usage, social and cognitive linguistic word counts, and a few recently proposed psycholinguistic norms. We use data from conversational samples of verbally fluent children obtained during Autism Diagnostic Observation Schedule (ADOS) sessions. We extract the aforementioned lexical cues from transcripts of session recordings and demonstrate their role in differentiating children diagnosed with Autism Spectrum Disorder from the rest. Further, we perform a correlation analysis between the lexical norms and ASD symptom severity. The analysis reveals an increased affinity by the interlocutor towards use of words with greater feminine association and negative valence.

Index Terms: Autism spectrum disorder, Lexical norms, Linguistic analysis

1. Introduction

Autism Spectrum Disorder (ASD) refers to a heterogeneous group of complex neurodevelopmental disorders characterized by social-communicative deficits along with restricted, repetitive behaviors. Prevalence of ASD among American children has been rapidly increasing, from about 1 in 150 children in 2002 to 1 in 68 children in 2014 [1]. Language is an integral part of any social interaction, and hence has been extensively studied in relation to ASD [2, 3, 4].

Previous studies have investigated language use in children with ASD under various settings, primarily through manual annotation. Different forms of delayed and impaired language production have been reported, such as meaningless repetition of an interlocutor's words (i.e., echolalia), formation of novel words (i.e., neologism), idiosyncratic phrasing [5], and preference to a more formal style of language [6]. Similarly, deficiencies in language comprehension (verbal and non-verbal) have been observed [7, 8].

Computational researchers continue to create new measures to understand the complexities of language production. For example, psycho-linguistic norms provide a dimensional description of language. Emotion norms have been quite popular for sentiment analysis [9], where low-level valence polarity can be used to infer the intended communication of a text. Norms beyond emotions, including age of acquisition, familiarity and concreteness have also been quite popular and made available publicly in various languages [10], [11]. With manually annotated corpora being quite costly to create, automatic scaling to new vocabulary has been explored in recent works [12, 13, 14]. Recently, Gibson et al. [15] used such psycho-linguistic norms for predicting empathy from a therapist in a motivational interview based psychotherapy targeting addictive behavior.

In this paper, we use automatic computational language analysis to characterize the verbal behavior of children with ASD and non-ASD developmental disorders. We quantify language usage by N-grams, word counts and lexical norms of psycho-linguistic nature. The word counts and lexical norms used are numeric representations of a specific emotional, linguistic or a psychological dimension. Since differences between ASD and non-ASD children with respect to manually annotated psycholinguistic dimensions have been widely researched [16, 17], we hypothesize that the norm representations can be used to quantify these differences. Our study is also a first step towards an automatic assessment of language in developmental disorder, with an overarching goal of aiding language-specific assessment and intervention progress tracking.

We conduct two sets of experiments to analyze the use of language in children. First, we conduct a classification experiment to investigate if there exists a difference in language based cues between ASD and non-ASD groups. Specifically, the experiment is performed using three sets of cues: (i) N-gram frequencies, (ii) psycho-linguistic lexical norms [14], (iii) and categorical word counts from the Linguistic Inquiry and Word Count software (LIWC) [18]. The N-gram count serves as a brute-force baseline that captures gross word usage. The psycho-linguistic norms offer quantifications of the emotional, linguistic and psychological dimensions which are hypothesized to provide complementary information beyond the frequency of word usage. We supplement the classification experiment with a correlation analysis to investigate the relationship between psycho-linguistic norms and autism severity. We aim to investigate the dependency (evaluated through empirical assumptions) between each individual norm and ASD severity, as opposed to the combined discriminative power evaluated in the classification experiments.

The classification experiment reveals that the proposed use of psycho-linguistic norms provides additional information beyond the baseline features, while the correlation analysis offers interesting insights into the verbal interaction, such as increased presence of negative valence and feminine association in the psychologist's speech in their interaction with children with higher ASD severity scores.

http://dx.doi.org/10.21437/Interspeech.2016-563

Table 1: Demographic Details

Category	Statistics
Age (years)	Range: 3.58-13.17. (Mean,std) = (8.61,2.49)
Gender	123 Male, 42 Female
Non-verbal IQ	Range: 47-141. (Mean,std) = (96.01,18.79)
Best-estimate Clinical Diagnosis	86 ASD, 42 ADHD 14 Mood/Anxiety Disorder 12 Language Disorder 10 Intellectual Disability, 1 No Diagnosis

2. Conversational Data

The Autism Diagnostic Observational Schedule [19] consists of a series of semi-structured activities between the subject and an examiner (a trained psychologist) used to evaluate behaviors associated with ASD. A session, which lasts 30-60 minutes, is broken down into different subtasks that are intended to elicit responses from the subject in different social and communicative settings. There are five modules (including the toddler module) in ADOS, and the psychologist chooses between them based on the subject's verbal fluency: Module 1 for least fluent to Module 4 for most fluent. Scoring in an ADOS session takes place as follows: The psychologist rates the subject's behavior in accordance to module-specific codes, wherein a code is categorical. The codes are combined in order to obtain a final ADOS classification score and transformed into an ASD severity score [20]. The psychologist is also required to give a best-estimate clinical diagnosis based on all the information gathered during the assessment.

We chose to work on the Emotions and Social Difficulties & Annoyance subtasks from Module 3, which are expected to impose high social and cognitive demand on the subject. Both subtasks consist of specific questions[1] posed to the child. In the *Emotions* subtask, the child is asked to identify things that trigger various emotions in him/her, and describe how he/she is affected by them. In the *Social Difficulties & Annoyance* subtask, the questions explore the child's notion of various social problems (at home and school), whether the child has understood the nature of those problems, and how he/she has tried to adapt to them.

The dataset used in this paper consists of 165 children (86 ASD, 79 Non-ASD). The demographic details are provided in Table 1. The audio was extracted from the video recordings for each session. The conversations were manually transcribed using a modified version of the SALT transcription guidelines [22]. All speech disfluencies along with temporal markings of utterance boundaries were removed, including laughter, fillers, sound effects and false starts; while disfluencies may be relevant to autism and will be investigated in future work, we do not consider them in the present study since we focus on language usage. Additional details on this corpus can be found in [23]. Each session consisted of 718 words on average.

3. Experimental Methodology

We conduct two experiments in this paper: (i) prediction of ASD clinical best-estimate diagnosis based on lexical features and, (ii) a correlation analysis between the lexical norms and ADOS severity. The prediction experiment is performed to estimate the joint discriminative power of the lexical features and the correlation analysis is performed to investigate the correlation of each lexical norm with the ASD severity. We discuss these two experiments in detail below.

[1] The exact set of questions is available in the ADOS manual [21]

3.1. Prediction of ASD diagnosis based on lexical features

To investigate the differences of lexical and psycho-linguistic features between ASD and non-ASD children, we perform a classification experiment to predict the diagnosis. We would like to emphasize that the goal of this experiment is solely to understand the discriminative power carried by the lexical and psycho-linguistic cues instead of developing a diagnostic mechanism. The latter objective is far more complicated and would require incorporation of several other factors apart from the language based cues. For the purpose of evaluation, we perform a 10 fold cross-validation with a unique set of children in each fold. 8 splits are used as training set, 1 as development set and 1 for testing. We use three sets of language based features in this paper for the prediction of ASD diagnosis: (i) a baseline N-gram feature set, (ii) Linguistic Inquiry and Word Count (LIWC) based features and, (iii) psycho-linguistic norms. We describe each of these feature sets in detail.

3.1.1. Baseline features: N-gram features

We use a set of unigrams and bigrams extracted using the SRILM toolkit [24] from the conversational transcripts as baseline features. These features capture the gross information in transcripts regarding the diagnosis, however do not offer quantification of abstract concepts such as affect and language sophistication (captured using other features discussed later). The N-grams are further appended with the corresponding speaker role (child or psychologist) to carry the information regarding the source of the N-gram.

The N-grams are extracted separately on the transcripts from the Emotions subtask and the Social Difficulties & Annoyance subtask and are compiled into a single feature vector. Since the dimensionality of the N-gram based feature vectors is very high, we prune the N-grams based on a minimum N-gram count and entropy of distribution criteria. An N-gram is retained only if it is observed more than a threshold count in the training set. Furthermore, the entropy of distribution of the N-grams between the ASD and non-ASD classes should also be lower than an entropy threshold for selection. We tune the count and entropy thresholds on the development set.

3.1.2. Linguistic Inquiry and Word Count based features

Researchers have investigated the effect of autism on a child's ability to understand and reciprocate the other person's emotions and feelings [25]. Further, impairment in the social use of language has been widely associated with autism and Asperger syndrome [26], and autism has been observed to impair and even arrest social development [27]. Hence, we hypothesize that there exists a dependency between the social and affective content in the vocabulary and ASD severity. We make use of the LIWC computation tool [18] to estimate *Social* and *Affect* features at session level. LIWC gives categorical word counts for each socio-behavioral aspect. We estimate the value of these features separately for each subtask as well as each speaker (child and the psychologist). Therefore, we get a set of four values for each child-psychologist pair, which is later used for the purpose of classification.

3.1.3. Psycho-linguistic Norm based feature

Next, we use a set of three psycho-linguistic norms introduced by [14] within the classification setup. These norms are termed *Age of acquisition*, *Concreteness* and *Gender Ladennes*. We describe the motivations for these features below.

Age of acquisition: Language development is commonly delayed in people with ASD. [28]. It has also been used as an early indicator for ASD, amongst other development disorders [29]. Hence, a child with a high ASD severity score could be hypothesized to possess impaired language development, which would be reflected by the choice of words among other characteristics such as pronunciation and prosody. The *Age of Acquisition* norm quantifies this language sophistication by associating it with an age at which the elicited vocabulary is expected to be obtained.

Concreteness: Children with autism have been known to prefer *concrete* thinking over abstraction. [30]. Further, decreased synchronization between language and spatial centers was observed in the brain for children with autism, implying that the *mental image* of words spoken (and listened to) may not be accurately/adequately formed. Finally, a task of selecting books based on abstract or concrete characteristics [31] showed that children with autism preferred *concrete* characteristics over abstract ones.

Gender Ladennes: This feature is motivated from the "Male Brain Theory" [32] in autism. In this theory, autism is seen as an extreme form of the male brain, in the sense that children with autism exhibit average or above average levels for systemizing, but significantly lower levels of empathizing than even typical male children (*Systemizing* and *Empathizing* have been proposed as two components which are prevalent in the male and female brain respectively). A number of experiments have been conducted (summarized in [32]) that emphasize the differences between these components. In this work, we check whether there exists any possible relation between the male brain theory and preference to use words with more masculine association.

We normalize all the norms between -1 and 1, with a lower norm value for a word indicating reduced age-of-acquisition, reduced concreteness and reduced feminine association. In the next section, we describe the classification setup over the baseline, LIWC and psycho-linguistic norm based features.

3.1.4. Classification setup

Since our baselines features are a set of N-grams, we use the Maximum Entropy (MaxEnt) classifier as the baseline classifier. The LIWC and psycho-linguistic norm based features are numeric and we employ a Support Vector Machine (SVM) classifier for these features. We perform a forward feature selection on the combined set of LIWC and psycho-linguistic norm based features to reduce the feature dimensionality on which the SVM classifier is trained. The final class assignment is performed based on a weighted fusion of outputs from the MaxEnt and SVM classifiers. We represent the output probabilities for the ASD class from the MaxEnt classifier as $p^{\text{ASD}}_{\text{MaxEnt}}$ and the output probability from the SVM classifier (computed by fitting logistic models to distance from hyperplane boundaries [33]) as $p^{\text{ASD}}_{\text{SVM}}$. The final class assignment is determined based on (1). The weighting parameter α is tuned on the development set.

$$\alpha \times p^{\text{ASD}}_{\text{MaxEnt}} + (1-\alpha) \times p^{\text{ASD}}_{\text{SVM}} \underset{\text{Non-ASD}}{\overset{\text{ASD}}{\gtrless}} 0.5 \qquad (1)$$

We report the classification accuracies using the baseline and additional features under four different settings: (i) using baseline features only (α in equation (1) is set to 1), (ii) using only the LIWC features in the SVM classifier, (iii) using only the psycho-linguistic norm based features in the SVM classifier and, (iv) using both, LIWC and psycho-linguistic norm based features in the SVM classifier. The results are stated in Table 2.

Table 2: Classification accuracies using the baseline and additional features. Chance accuracy is the proportion of majority class.

Feature set	Accuracy
Chance	52.1%
Baseline (N-gram)	66.7%
Baseline + LIWC	69.1%
Baseline + psycho-linguistic norm	69.7%
All features	69.1%

3.1.5. Results

From the results in Table 2, we observe that every set of feature is significantly better than chance (Mc-Nemar's test, p-value $<$0.005). We also observe improvements over the baseline system by individually appending the LIWC and psycho-linguistic norm based features. However, these improvements are not significant over the baseline system. Due to a limited sample of data, achieving a significant boost over baseline requires higher improvements. We believe that it is not possible to obtain a high diagnosis accuracy in this sample using a single modality (language in our case), therefore limiting the improvements below significant. Although there is no improvement in training with all features, it is encouraging that the additional features do provide a boost over the baseline and we perform analysis of these features by listing the top few baseline and additional features, as discussed next.

3.1.6. Discussion

We perform a feature analysis over the baseline and additional features by listing the most frequently selected features during cross validation from both the categories. We separately discuss the baseline n-gram and additional features below.

N-gram features: Top N-gram features from the baseline system are the ones that highly favor one class over the other, as determined by the MaxEnt classifier. That is, if the output MaxEnt probability for the ASD (/Non-ASD) class is very high solely based on the N-gram under consideration, the N-gram is considered to be a strong indicator of the ASD (/Non-ASD) class. The N-grams are further sorted by their frequency of occurrence in the data to remove low-frequency N-grams that are spuriously determined to be highly relevant to a class. Table 3 shows the top few N-grams, categorized by the speaker group (child, psychologist). They are further arranged by the class they favor.

The results suggest a frequent usage of the phrase *I don't know* by an ASD child. This is expected, since children with autism have been known to possess difficulties comprehending language in general [34]. Nothing significant can be inferred in the case of a Non-ASD child except that there are more references to other people. However, characteristic words from the psychologist's speech suggests that social questions (for instance, *Have you ever had problems getting along with other people at school?*) are given more emphasis for ASD children over Non-ASD children.

LIWC and psycho-linguistic norm based features: We identify the top LIWC and psycho-linguistic norm based features as the ones that are picked during the feature selection in ev-

Table 3: Most characteristic N-grams of different diagnostic groups

	Child	Psychologist
ASD	I_DON'T	FEEL_WHEN
	DON'T_KNOW	IT_FEEL
	AND_I	OTHER_PEOPLE
	UM_I	MAKES_YOU
	BUT_I	DO_YOU
Non ASD	MY_BROTHER	YOU_FEEL
	IN_THE	WHEN_YOU'RE
	I_GET	HOW_DOES
	LIKE_I	CAN_YOU
	I_JUST	FEEL_INSIDE

ery iteration of the 10 fold cross-validation. Feature selection on the LIWC feature set always returned the *Affect* feature from the negative valence questions (sadness, anger, fear) for both child and psychologist. On the other hand, we obtained the *Gender Ladennes* norm as a selected feature (from the set of psycholinguistic norms) during feature selection. This suggests that there exists variability between the two diagnostic groups in the conduct of negative valence questions. The child's response and the psychologist's follow up, and this variability is manifested in their language. Further, the gender association of the vocabulary is the most significant discriminating feature, over the perceived language delay (age-of-acquisition) and affinity to concreteness. In the next section, we continue the analysis on these features by correlating with the ASD severity.

3.2. Norm-Severity Correlation analysis

In this section, we investigate whether lexical norms can be used to make any potential inferences about language use patterns in ASD using a correlation analysis. From the LIWC feature set, we consider the *Social*, *Affect* and *Cogmech* dimensions. Transcripts from Social Difficulties & Annoyance subtask are used to correlate the LIWC features. From the psycho-linguistic features, we consider emotion norms *Arousal*, *Valence* and *Dominance* from the Emotions subtask, and *Gender Ladennes*, *Age of Acquisition* and *Concreteness* features from transcriptions of both subtasks taken together.

For each feature, we use the Stanford Part-Of-Speech Tagger [35] to compute the mean score across a session by considering only content words (nouns, adjectives, verbs and adverbs) and remove filler words like *mm-hm*, *hmm*, *mm* and *um*. We correlate these scores with the Calibrated Severity Scores (CSS) [20], which are more robust to demographic variations. Further, all correlation results are normalized over the child's age, gender and IQ (non-verbal) levels. The correlation results are presented in Table 4.

3.2.1. Results

After adjusting the p-values of the multiple correlation coefficients, we found a significant negative association between children's linguistic valence and ASD severity. Our results suggest significant correlation between the child's severity scores and the psychologist's linguistic features. The psychologist's affect and valence correlates significantly with ASD severity, indicating high occurrence of affective words with neutral valence. We further observe a significant relation between the psychologist's expressed Gender Ladennes and the children's severity scores, i.e. the psychologist expresses more feminine oriented words (i.e. of higher empathy) towards children with more severe

Table 4: Correlation of lexical norms with Calibrated Severity Scores. **Significant** at ($p < 0.05$) using student t-distribution [36]

Norm	Child	Psychologist
Social	-0.11	0.10
Affect	0.08	**0.30**
Cogmech	-0.01	–
Arousal	0.01	-0.04
Valence	**-0.15**	**-0.20**
Dominance	-0.07	-0.05
Age of Acquisition	0.09	-0.05
Gender	-0.07	**0.32**
Concreteness	0.09	-0.10

ASD diagnosis. These are consistent with previous findings indicating the association of ASD severity to the psychologist's prosodic patterns [37]. The results are also consistent with the classification setup, since most of the features returned by the feature-selection algorithm correlate significantly. Surprisingly, *Age of Acquisition* and *Concreteness* do not correlate significantly although one would expect them to, based on existing hypotheses. One possible reason for *Age of Acquisition* could be that Module 3 of ADOS is intended only for high functioning group of children with ASD. It is not immediately clear whether the norms do not accurately reflect the psycho-linguistic dimension, or whether the lexical information alone is not enough to deduce the underlying behavioral characteristics.

4. Conclusion

Lexical use has been studied by psychologists to understand and characterize developmental disorders such as ASD. In the context of autism research, extracting knowledge-driven features from acoustic data in a conversational model has been shown to correlate closely with clinically diagnosed severity measures. In this paper, we propose different objective lexical analyses on ADOS transcripts involving children with various neurodevelopmental disorders. We use N-gram models, linguistic word counts and recently proposed computational psycholinguistic norms and show using classification and correlation experiments that such objective measures assist in discriminating between the two groups, although they are prone to noise and misinterpretation.

In the future, we would like to increase automation in the analysis process by repeating the investigation on ASR (Automatic Speech Recognition) decoded transcripts. ASR systems for children speech in general, are yet to scale up in performance to adults' speech recognition systems. One would expect more variability in the speech of children with ASD, thereby presenting opportunities and challenges related to noise suppression into ASR based lexical analysis. We also aim to integrate the lexical analysis with other modalities for a holistic understanding of communication in children with ASD. Finally, in collaboration with ASD specialists, we also aim to impact the diagnosis and intervention in developmental disorders based on the proposed language and other multi-modal cues.

5. Acknowledgement

This work was supported by National Science Foundation (NSF), National Institute of Child Health and Human Development (NICHD) and National Institute of Mental Health (NIMH).

6. References

[1] J. Baio, "Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, united states, 2010." *Morbidity and mortality weekly report. Surveillance summaries (Washington, DC: 2002)*, vol. 63, no. 2, p. 1, 2014.

[2] E. M. Ornitz, D. Guthrie, and A. H. Farley, "The early development of autistic children," *Journal of autism and childhood schizophrenia*, vol. 7, no. 3, pp. 207–229, 1977.

[3] M. Bucholtz and K. Hall, "Language and identity," *A companion to linguistic anthropology*, vol. 1, pp. 369–394, 2004.

[4] L. Bloom, *Language development*. University of Chicago Press Chicago, 1975.

[5] J. Volden and C. Lord, "Neologisms and idiosyncratic language in autistic speakers," *Journal of Autism and Developmental Disorders*, vol. 21, no. 2, pp. 109–130.

[6] C. Lord, "Language in high-functioning adolescents with autism: Questions about deviance and delay," in *Rochester Symposium on Developmental Psychopathology*, vol. 7. UNIVERSITY OF ROCHESTER PRESS, 1996, pp. 149–166.

[7] S. V. Huemer and V. Mann, "A comprehensive profile of decoding and comprehension in autism spectrum disorders," *Journal of Autism and Developmental Disorders*, vol. 40, no. 4, pp. 485–493, 2009.

[8] J. Ricketts, C. R. G. Jones, F. Happé, and T. Charman, "Reading comprehension in autism spectrum disorders: The role of oral language and social functioning," *Journal of Autism and Developmental Disorders*, vol. 43, no. 4, pp. 807–816, 2012.

[9] M. M. Bradley and P. J. Lang, "Affective norms for english words (anew): Instruction manual and affective ratings," Technical report C-1, the center for research in psychophysiology, University of Florida, Tech. Rep., 1999.

[10] M. Coltheart, "The mrc psycholinguistic database," *The Quarterly Journal of Experimental Psychology Section A*, vol. 33, no. 4, pp. 497–505, 1981.

[11] O. Lahl, A. S. Göritz, R. Pietrowsky, and J. Rosenberg, "Using the world-wide web to obtain large-scale word norms: 190,212 ratings on a set of 2,654 german nouns," *Behavior Research Methods*, vol. 41, no. 1, pp. 13–19, 2009.

[12] M. Mohler, M. T. Tomlinson, D. B. Bracewell, and B. Rink, "Semi-supervised methods for expanding psycholinguistics norms by integrating distributional similarity with the structure of wordnet." Citeseer.

[13] T. Liu, K. Cho, G. A. Broadwell, S. Shaikh, T. Strzalkowski, J. Lien, S. M. Taylor, L. Feldman, B. Yamrom, N. Webb *et al.*, "Automatic expansion of the mrc psycholinguistic database imageability ratings."

[14] N. Malandrakis and S. S. Narayanan, "Therapy language analysis using automatically generated psycholinguistic norms," in *INTERSPEECH*, 2015.

[15] J. Gibson, N. Malandrakis, F. Romero, D. C. Atkins, and S. Narayanan, "Predicting therapist empathy in motivational interviews using language features inspired by psycholinguistic norms," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[16] A. M. Wetherby and C. A. Prutting, "Profiles of communicative and cognitive-social abilities in autistic children," *Journal of Speech, Language, and Hearing Research*, vol. 27, no. 3, pp. 364–377, 1984.

[17] L. Wing and J. Gould, "Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification," *Journal of Autism and Developmental Disorders*, vol. 9, no. 1, pp. 11–29, 1979.

[18] J. W. Pennebaker, R. J. Booth, and M. E. Francis, "Linguistic inquiry and word count: Liwc [computer software]," *Austin, TX: liwc. net*, 2007.

[19] C. Lord, S. Risi, L. Lambrecht, E. H. Cook Jr, B. L. Leventhal, P. C. DiLavore, A. Pickles, and M. Rutter, "The autism diagnostic observation schedulegeneric: A standard measure of social and communication deficits associated with the spectrum of autism," *Journal of autism and developmental disorders*, vol. 30, no. 3, pp. 205–223, 2000.

[20] K. Gotham, A. Pickles, and C. Lord, "Standardizing ados scores for a measure of severity in autism spectrum disorders," *Journal of autism and developmental disorders*, vol. 39, no. 5, pp. 693–705, 2009.

[21] C. Lord, M. Rutter, P. DiLavore, S. Risi, K. Gotham, and S. Bishop, *Autism diagnostic observation schedule: ADOS: Manual*. Western Psychological Services Los Angeles, CA, 2002.

[22] J. Miller and A. Iglesias, "Systematic analysis of language transcripts (SALT), Research Version [Computer Software]," 2012.

[23] D. Bone, S. Bishop, R. Gupta, S. Lee, and S. Narayanan, "Acoustic-prosodic and turn-taking features in interactions with children with neurodevelopmental disorders," in *INTERSPEECH*, 2016.

[24] A. Stolcke, J. Zheng, W. Wang, and V. Abrash, "Srilm at sixteen: Update and outlook," in *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop*, 2011, p. 5.

[25] S. Baron-Cohen, *Mindblindness: An essay on autism and theory of mind*, 1997.

[26] ——, "Social and pragmatic deficits in autism: cognitive or affective?" *Journal of autism and developmental disorders*, vol. 18, no. 3, pp. 379–402, 1988.

[27] R. J. Landa, K. C. Holman, and E. Garrett-Mayer, "Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders," *Archives of general psychiatry*, vol. 64, no. 7, pp. 853–864, 2007.

[28] L. Wing and J. Gould, "Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification," *Journal of Autism and Developmental Disorders*, vol. 9, no. 1, pp. 11–29.

[29] S. Mitchell, J. Brian, L. Zwaigenbaum, W. Roberts, P. Szatmari, I. Smith, and S. Bryson, "Early language and communication development of infants later diagnosed with autism spectrum disorder," *Journal of Developmental & Behavioral Pediatrics*, vol. 27, no. 2, pp. S69–S78, 2006.

[30] L. Kanner, *Autistic disturbances of affective contact*, 1943.

[31] D. Ropar and D. Peebles, "Sorting preference in children with autism: the dominance of concrete features," *Journal of autism and developmental disorders*, vol. 37, no. 2, pp. 270–280, 2007.

[32] S. Baron-Cohen, "The extreme male brain theory of autism," *Trends in cognitive sciences*, vol. 6, no. 6, pp. 248–254, 2002.

[33] J. Platt *et al.*, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in large margin classifiers*, vol. 10, no. 3, pp. 61–74, 1999.

[34] H. Tager-Flusberg, R. Paul, C. Lord *et al.*, "Language and communication in autism," *Handbook of autism and pervasive developmental disorders*, vol. 1, pp. 335–364, 2005.

[35] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. Association for Computational Linguistics, 2003, pp. 173–180.

[36] H. Soper, A. Young, B. Cave, A. Lee, and K. Pearson, "On the distribution of the correlation coefficient in small samples. appendix ii to the papers of "student" and r. a. fisher," *Biometrika*, vol. 11, no. 4, pp. 328–413, 1917.

[37] D. Bone, C.-C. Lee, M. P. Black, M. E. Williams, S. Lee, P. Levitt, and S. Narayanan, "The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody," *Journal of Speech, Language, and Hearing Research*, vol. 57, no. 4, pp. 1162–1177, 2014.