Parameter Selection for Analyzing Conversations with Autism Spectrum Disorder

Tahiya Chowdhury[1], Veronica Romero[1,2], Amanda Stent[1]

[1]Davis Institute for AI, Colby College, USA
[2]Department of Psychology, Colby College, USA

`tahiya.chowdhury@colby.edu, veronica.romero@colby.edu, ajstent@colby.edu`

Abstract

The diagnosis of autism spectrum disorder (ASD) is a complex, challenging task as it depends on the analysis of interactional behaviors by psychologists rather than the use of biochemical diagnostics. In this paper, we present a modeling approach to ASD diagnosis by analyzing acoustic/prosodic and linguistic features extracted from diagnostic conversations between a psychologist and children who either are typically developing (TD) or have ASD. We compare the contributions of different features across a range of conversation tasks. We focus on finding a minimal set of parameters that characterize conversational behaviors of children with ASD. Because ASD is diagnosed through conversational interaction, in addition to analyzing the behavior of the children, we also investigate whether the psychologist's conversational behaviors vary across diagnostic groups. Our results can facilitate fine-grained analysis of conversation data for children with ASD to support diagnosis and intervention.

Index Terms: developmental disorder, autism, conversation, children's speech.

1. Introduction

Autism spectrum disorder (ASD) refers to a range of developmental disabilities characterized by deficits in social communication and interaction. In the absence of therapeutic intervention, adolescents with ASD exhibit impairments in social interaction throughout their lives [1], including an inability to display or reciprocate verbal aspects of communication appropriately and, in turn, successfully maintain effective social interactions [2, 3, 4, 5]. The diagnosis of ASD is a complex, challenging task as it depends on behavioral symptoms identified by psychologists with no reliable biochemical diagnostic tests available. Existing diagnostic instruments, such as the Autism Diagnostic Observation Schedule (ADOS) [6], rely on qualitative coding by expert assessors for the presence or absence of certain behavioral markers across multiple structured conversational scenarios. The assessor has to simultaneously engage the child in conversation, monitor their own conversational behavior, and make diagnostic notes. AI-based techniques have the potential to reduce the cognitive load on the assessor by providing objective measurements of conversational behaviors and thus to augment the assessor's workflow.

Previous research has explored the efficacy of ML tools using speech and language features for identifying atypical and behavioral signals of ASD in communication from conversation data [7, 8, 9]. Acoustic-prosodic features such as pitch [10], intonation, and rhythm [11] derived from diagnostic conversations have been found useful for identifying ASD individuals. Language features such as word usage [12], social and cognitive linguistic word counts [13], and semantic similarity [14] have also been investigated and found to contain behavioral markers helpful to distinguish between ASD and TD children. While these findings are promising, prior work uses conversation samples from a limited set of diagnostic tasks and may not generalize to different conversational contexts.

Prior research has also explored the use of information from the conversational partner's interactions to predict diagnosis outcome. The conversational partner's acoustic-prosodic cues have been found to be predictive of ASD symptom severity [15] and engagement levels [16]. [13] found a significant correlation between the conversational partner's language use and ASD severity score. [17] used social context in conversations to explore its relationship with symptoms of autism. However, these findings also come from a limited set of diagnostic tasks.

In this work, we aim to answer the following questions:

- What are the most informative acoustic-prosodic indicators of children with ASD in different conversational contexts?
- What are the most informative linguistic indicators of children with ASD in different conversational contexts?
- Can we classify ASD and TD children using a minimal set of characteristics of the child's spoken language?
- Can we classify ASD and TD children using a minimal set of characteristics of the interlocutor's spoken language?

2. Data

We used data collected during sessions of the Autism Diagnostic Observation Schedule - Second Edition (ADOS-2), an assessment tool used to categorize ASD impairment [6]. In this assessment, a child and a certified adult assessor (usually a psychologist) engage in a sequence of semi-structured activities to assess behavioral patterns of the child. Our data includes conversations spanning 14 different subtasks from Module 3 of the ADOS-2, which is designed for verbally fluent children and adolescents (Table 1). Depending on the subtask, the child may be asked to tell a story, play with toys, act out a cartoon, or discuss topics such as emotion, loneliness, and friends. Because we are interested in different conversational contexts, we included all 14 tasks from the module, which are listed in Table 3.

Our data included 29 children (14 ASD, 15 TD) whose age ranged between 10-15 years. Each ADOS-2 assessment session lasted on average 40-60 minutes; all sessions were administered by the same psychologist. For this work, we first prepared a separate recording of each subtask using annotations done by a research assistant from the video recordings of the sessions. The average length of these videos is 5 minutes. We used pyannote [18] to perform speaker diarization. We then

Table 1: *Demographic information of the ADOS-2 dataset.*

Category	Statistics
Age (years)	Range: 10-15 (mean: 12.27, std: 1.75)
Gender	21 male, 8 female
Race	White: 20, Afr. Am.: 6, Hisp.: 2, Asian: 1
Diagnosis	ASD: 14, TD: 15

used PyDub [19] to partition the subtask audio into shorter segments based on speaker identity and turn boundaries derived using pyannote. This allowed us to automatically separate the child and psychologist turns into separate audio recordings, resulting in a total of 6440 utterances. After removing the empty utterances, our final set contained 116 utterances per child on average. We then used the Whisper open-source speech recognizer [20] to transcribe each turn.

3. Method

We trained models to analyze the contribution of speech and language features in distinguishing between children with and without ASD during different conversational tasks[1].

3.1. Features

We used openSMILE [21] to extract acoustic-prosodic features from each turn. In this work, we used the ComPARE 2016 feature set [22], which extracts 88 low level features and 6373 features derived by applying statistical functions on the low-level features. This tool and a similar feature set have been used before in research on autism diagnosis [23]. The low-level features include spectral (Mel Frequency Cepstral Coefficients (MFCC), zero crossing rate), voice quality (local shimmer, jitter, harmonic noise ratio), and prosodic features (loudness, pitch). We z-normalized all these features per participant as the mean age of diagnosed children coincided with the age of changing voice.

We used two sets of language features: Linguistic Inquiry and Word Count (LIWC) features and lexical features motivated by prior research. As prior work has shown correlations between the psychologist's language use and ASD severity [13], we calculated these features for each speaker (child and psychologist) for each of the 14 subtasks. To understand the influence of task type and context in predicting the diagnostic group, we also used task as a feature.

LIWC [24] has been widely used for predicting outcomes including psychological [25] and cognitive [26] functioning and personality [27]. We used a total of 119 LIWC features, including percentage of different parts of speech (e.g. pronouns, articles), punctuation categories (e.g. commas, periods), and psychometric measures (e.g. affect, social, politeness).

Separately, we used a set of 12 lexical features inspired by their use in prior autism research [12]. Using the spaCy NLP library [28] and word dictionaries [29], we calculated the relative frequencies of certain parts of speech as well as the number of syllables, hedge words, weasel words, filler words and discourse markers per turn. This resulted in a total of 131 language features. Prior research has shown the importance of appropriate usage of discourse markers (e.g. *and, but, anyway*), hedge

words (e.g. *often, usually*), weasel words (e.g. *may, like, possibly*), and filler words (e.g. *so, ok*) for measuring reciprocity and concreteness in conversations for autism research [30].

3.2. Feature Selection and Classification

First, we performed correlational analyses to estimate the association between our features and ASD diagnosis outcome. As our diagnosis outcome is binary, we performed point-biserial correlations [31]. To control for any false discovery rate arising from multiple tests, we applied Holm's sequential Bonferroni procedure [32].

Second, to understand the role of different features in distinguishing between children diagnosed with ASD and TD, we used a feature fusion strategy where we trained a random forest classifier using different combinations of features from different modalities (speech, text, task) extracted from each speaker in the conversation. To identify a minimal set of characteristic features, we used our correlation analysis as a feature selection strategy where we selected features found to be statistically significant ($p < 0.05$) for each task into the feature subset.

To ensure generalization for out-of-sample testing, we performed cross-validation by using a leave-n-user-out method[2]. We report results averaged over 10 runs, where 80% (23) of children were randomly selected for training and 20% (6) for testing in each run. All classification results are reported as accuracy, precision, recall, F1-score and AUC score. All experiments and models are implemented using the scikit-learn and sciPy libraries with default parameter settings.

4. Analysis and Results

4.1. Correlation Analysis

Based on the correlation analyses, we found a subset of 28 (out of 88) acoustic-prosodic features, 3 (out of 131) lexical features and 21 (out of 219 total) combination features, extracted from the child conversational participants' turns, to have statistically significant correlations with an ASD diagnosis outcome. In the case of the experimenter, we have 19, 7, and 17 features respectively that have statistically significant correlations with an ASD diagnosis outcome. We present in Table 2 the 7 features we find significant most frequently across different tasks.

In our correlation analysis, we also explored the top features based on significant correlation and strength in each task. To find shared features for both child and psychologist, we compared the top features for the two speakers in the conversational context, which we present in Table 3.

4.2. Observations

Child. Here, we include some additional observations regarding informative features for different task contexts. When using features from the child conversational participant, in conversations about social difficulties children with ASD used fewer discourse markers ($r = -0.2401$) and fewer words. Children with ASD used fewer pronouns ($r = -0.3703$) and discourse markers ($r = -0.3465$), and their turns were shorter ($r = -0.3299$). For the make-believe play task, children with ASD used more hedge words ($r = +0.2150$). Interestingly, when talking about friends, children with ASD used fewer words from the *tech* category than TD children ($r = -0.1093$).

---

[1]These experiments and the original data collection were reviewed and approved by the Institutional Review Board at the participating institutions.

[2]We chose this over 10-fold cross validation, as this ensures our training set does not include information from a child who is also present in our test set.

Table 2: *Correlation coefficients of statistically significant indicators of ASD in ADOS-2 tasks as assessed using the conversation turns of child conversation participants. We include only the tasks for which at least one feature was found significant after Holm's sequential Bonferroni correction.*

Tasks	MFCC	Spect. Harmonicity	Log HNR	Shimmer	Spectral Energy	Pitch	Disc. Markers
Description	0.2409	-	-	-	-	-	-
Emotion	-0.2504	0.2633	0.2641	0.2439	0.2993	0.2193	-
Social	-	-	-	-	-	-	-0.2401
Friends	-0.2167	-	-	-	-	-	-
Loneliness	-0.5402	-	-	-	-	-	-
Interactive	0.3424	-	-	-	-	-0.2601	-
Telling	0.2242	-	-	-	-	-	-

When talking about emotion, describing a picture, and during the break, the count of words, words per sentence, nouns, syllables and turn duration were all negatively correlated with an ASD diagnosis outcome, which suggest shorter responses from ASD children. Acoustic-prosodic features such as loudness and pitch were positively correlated with an ASD diagnosis outcome for the cartoons subtask, but negatively correlated for silent play during break.

Psychologist. When talking about emotions or friends, the psychologist used fewer discourse markers ($r = -0.2091$), weasel words, words, and syllables, and produced shorter turns, when talking with children with ASD. When talking about loneliness, use of the word *you* ($r = 0.3173$) is increased and *i* decreased. The counts of words, words per sentence and pronoun usage (*he/she, we, they*) were all negatively correlated with an ASD diagnosis outcome in conversation, break and emotion subtasks. Overall, language features were found to be more informative in the psychologist data for predicting the diagnosis outcome, whereas only MFCC features among all acoustic-prosodic features were found to be significant. This observation is reflected in our classification results discussed next.

4.3. Classification

We use the findings of our correlation analyses to inform our feature selection strategy when training classifiers to distinguish between children with ASD and TD children. We first trained a classifier with all features from each speaker turn, and then separately trained classifiers with selected features from different combinations of feature sets. We used random forests because their results are highly interpretable by psychologists. Our goal was to find a minimal feature set for the purpose of aiding diagnosis.

Our classification results for child turn data are presented in Table 4. For each feature type, we trained with full and selected feature sets (the number of features, N, is indicated in parentheses). We found language features were better indicators for the diagnosis outcome than acoustic-prosodic features. However, the combination of features from the two modalities gave improved accuracy ($F1 = 0.61$). When we trained with features selected using point-biserial correlation analysis, the number of features used in the model was reduced by nearly 80% for the A, A+L and A+L+T settings, while classification accuracy improved or remained unchanged. After including the subtask category as a categorical feature, we obtained the best classification results with both full and selected feature sets.

We observed similar findings when using the psychologist's turn data which we present in Table 5. Using only acoustic-prosodic features (A) and only text (T), the model trained on the psychologist's turns provided results similar to the one trained on the childs' turns. We observe that the classification results remained unchanged or improved for all settings (A, L, A+L, A+L+T) when training with nearly 80% fewer features selected through the correlation analysis feature selection procedure, an observation similar to the results from child's turns.

We also applied dimensionality reduction (PCA) as an alternate feature selection strategy and compared this with the correlation-based selection strategy. The number of principal components was chosen to preserve 95% variance in the data. We found that this did not improve performance compared to using the full feature set. One reason could be that we fit a PCA model to the training data and transformed the testing data using the fitted model. Since our training and testing data come from different children, applying PCA did not help in this case.

Ethical discussion and limitations. Our goal is to provide AI decision tools to help assessors more effectively use ADOS for ASD diagnosis. It would be unethical to deploy the machine learning approach outlined here on its own (without human judgment). Also, it would be unethical to deploy such an approach in a non-clinical setting. In order to ensure no misuse of our findings, we do not plan to release the trained models or raw data to the public.

For our experiments, we used entirely automatic preprocessing. No stage of preprocessing is perfect. In particular, there are many cases where pyannote missed speaker changes. Our experimental results could be improved with manual diarization and transcription, but automatic diarization and transcription are more practical in our target setting.

Another limitation of this work is that our data includes only two diagnostic groups (ASD and TD) and the language of the conversations is English. Whether the informative features for ASD are also applicable for other co-occurring neurodevelopmental disorders (e.g. ADHD) and other languages is a direction for future work. Finally, the current data set only included one psychologist, so we do not know how our findings would generalize to other practitioners.

5. Conclusions and Future Work

Early diagnosis and interventions are critical to positive outcomes for autism. There is a growing need for identifying behavioral indicators that can help clinicians to diagnose ASD. Speech and language features have been extensively studied in prior works to distinguish between children with ASD and TD children. In this work, we explored the diagnostic informativeness of different acoustic-prosodic and language features across multiple diagnostic conversational contexts, separately examining data from child and psychologist turns. We used acoustic-prosodic indicators and language indicators to iden-

Table 3: *Most predictive features for different subtasks with task descriptions. We add the indicator's correlation coefficient with an ASD diagnosis outcome for features extracted from the experimenter's turns when found significant. (C = Child, P = Psychologist)*

Subtask Name	Task Description	Feature	Corr. Coefficient (C | P)
Description	Description of a picture	MFCC [13]	+0.2409 | −0.2064
Conversation	Conversation about topics of interest	Duration	| −0.2675
Emotions	Interview about their feelings	Spectral energy	+0.2993
Social Difficulties	Interview about their social lives	Discourse Markers	−0.2401
Friends	Interview about their relationships	MFCC [5]	−0.2167
Loneliness	Interview about loneliness	MFCC [7]	−0.5402
Construction	Work on a puzzle	MFCC [14]	+0.2223 | −0.2671
Make-believe Play	Child plays with toys alone	MFCC [13]	+0.2973 | +0.4467
Interactive Play	As previous (with psychologist)	MFCC [6]	+0.3424 | +0.2603
Demonstration	Demonstrates how to do a task	Spectral energy	+0.2335
Telling	Makes up story based on picture book	MFCC [13]	+0.2242
Cartoons	Act out a cartoon story	Duration	+0.3044
Imaginative Story	Creating a story with small prop items	MFCC [1]	+0.1786 | +0.1765
Break	Silent play and unstructured conversation	MFCC[6]	+0.1997 | +0.2014

Table 4: *ASD diagnosis classification from child turns using a random forest classifier trained with different combinations of full and selected features. The number of features used in each model is indicated by N. Best results are boldfaced. (A: Acoustic-prosodic, L: Lexical, T: Task category)*

Features	Acc.	Prec.	Rec.	F1	AUC
A (N = 88)	0.48	0.67	0.48	0.50	0.55
A (N = 28)	0.51	0.65	0.51	0.55	0.56
L (N = 131)	0.60	0.72	0.60	0.57	0.63
L (N = 3)	0.61	0.76	0.61	0.55	0.62
A + L (N =219)	0.66	0.73	0.66	0.61	0.62
A + L (N = 21)	0.62	0.81	0.62	0.58	0.66
A+L+T (N = 220)	0.74	**0.85**	0.74	0.72	**0.76**
A+L+T (N = 22)	**0.76**	0.84	**0.76**	**0.75**	0.71

Table 5: *ASD diagnosis classification from psychologist's turns using a random forest classifier trained with different combinations of full and selected features. The number of features used in each model is indicated by N. Best results are boldfaced. (A: Acoustic-prosodic, L: Lexical, T: Task category)*

Features	Acc.	Prec.	Rec.	F1	AUC
A (N = 88)	0.49	0.60	0.49	0.50	0.52
A (N = 19)	0.51	0.63	0.51	0.51	0.56
L (N = 131)	0.56	0.79	0.56	0.54	0.63
L (N = 7)	0.63	0.82	0.63	0.60	0.66
A + L (N = 219)	0.55	0.70	0.55	0.47	0.59
A + L (N = 17)	0.61	0.74	0.61	0.53	0.56
A+L+T (N = 220)	0.59	0.80	0.59	0.56	0.64
A+L+T (N = 18)	**0.68**	**0.83**	**0.68**	**0.65**	**0.67**

tify the most informative features for different conversational tasks administered during the ADOS-2. Our results from correlational and classification experiments suggest the presence of informative indicators that are shared across different contexts and co-appear in conversational turns from both speakers.

We are currently working to confirm these results via manually corrected speaker diarization and transcription. In future work, we would like to extend this analysis to multimodal conversational data. We also plan to examine modeling approaches that take into account more of the conversational context, for example, whether the child conversation participant constructs contextually appropriate dialog acts, or the ways in which child and assessor align in acoustic-prosodic or linguistic ways during conversation.

6. Acknowledgements

7. References

[1] American Psychiatric Organization, *Diagnostic and statistical manual of mental disorders (DSM-5)*. Washington, DC: American Psychiatric Publishing, 2013.

[2] H. Tager-Flusberg and E. Caronna, "Language disorders: Autism and other pervasive developmental disorders," *Pediatric Clinics of North America*, vol. 54, no. 3, Jun. 2007.

[3] H. Tager-Flusberg, "A psychological approach to understanding the social and language impairments in autism," *International Review of Psychiatry*, vol. 11, no. 4, pp. 325–334, Jan. 1999.

[4] P. Mundy and J. Markus, "On the nature of communication and language impairment in autism," *Mental Retardation and Developmental Disabilities Research Reviews*, vol. 3, no. 4, pp. 343–349, 1997.

[5] R. Landa, "Social language use in Asperger syndrome and high-functioning autism," in *Asperger Syndrome*. New York, NY, US: The Guilford Press, 2000, pp. 125–155.

[6] C. Lord, S. Risi, L. Lambrecht, E. H. Cook, B. L. Leventhal, P. C. DiLavore, A. Pickles, and M. Rutter, "The Autism diagnostic observation schedule—generic: 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, Jun. 2000.

[7] R. Fusaroli, E. Weed, D. Fein, and L. Naigles, "Hearing me hearing you: Reciprocal effects between child and parent language in

autism and typical development," *Cognition*, vol. 183, pp. 1–18, Feb. 2019.

[8] R. Fusaroli, A. Lambrechts, D. Bang, D. M. Bowler, and S. B. Gaigg, "Is voice a marker for autism spectrum disorder? a systematic review and meta-analysis," *Autism Research: Official Journal of the International Society for Autism Research*, vol. 10, no. 3, pp. 384–407, 2017.

[9] R. Fusaroli, R. Grossman, N. Bilenberg, C. Cantio, J. R. M. Jepsen, and E. Weed, "Toward a cumulative science of vocal markers of autism: A cross-linguistic meta-analysis-based investigation of acoustic markers in American and Danish autistic children," *Autism Research: Official Journal of the International Society for Autism Research*, vol. 15, no. 4, pp. 653–664, 2022.

[10] G. Kiss, J. P. v. Santen, E. Prud'Hommeaux, and L. M. Black, "Quantitative analysis of pitch in speech of children with neurodevelopmental disorders," in *Proceedings of INTERSPEECH*, 2012.

[11] D. Bone, M. P. Black, A. Ramakrishna, R. B. Grossman, and S. S. Narayanan, "Acoustic-prosodic correlates of 'awkward' prosody in story retellings from adolescents with autism," in *Proceedings of INTERSPEECH*, 2015, pp. 1616–1620.

[12] A. Song, M. Cola, S. Plate, V. Petrulla, L. Yankowitz, J. Pandey, R. T. Schultz, and J. Parish-Morris, "Natural language markers of social phenotype in girls with autism," *Journal of Child Psychology and Psychiatry*, vol. 62, no. 8, pp. 949–960, 2021.

[13] M. Kumar, R. Gupta, D. Bone, N. Malandrakis, S. Bishop, and S. S. Narayanan, "Objective language feature analysis in children with neurodevelopmental disorders during autism assessment," in *Proceedings of INTERSPEECH*, 2016, pp. 2721–2725.

[14] A. Goodkind, M. Lee, G. E. Martin, M. Losh, and K. Bicknell, "Detecting language impairments in autism: A computational analysis of semi-structured conversations with vector semantics," in *Proceedings of the Society for Computation in Linguistics (SCiL) 2018*, 2018, pp. 12–22.

[15] 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.

[16] R. Gupta, D. Bone, S. Lee, and S. Narayanan, "Analysis of engagement behavior in children during dyadic interactions using prosodic cues," *Comput. Speech Lang.*, vol. 37, no. C, p. 47–66, may 2016.

[17] T. Muskett, M. Perkins, J. Clegg, and R. Body, "Inflexibility as an interactional phenomenon: Using conversation analysis to re-examine a symptom of autism," *Clinical linguistics & phonetics*, vol. 24, no. 1, pp. 1–16, 2010.

[18] H. Bredin, R. Yin, J. M. Coria, G. Gelly, P. Korshunov, M. Lavechin, D. Fustes, H. Titeux, W. Bouaziz, and M.-P. Gill, "pyannote.audio: neural building blocks for speaker diarization," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2020.

[19] (2022) Pydub. [Online]. Available: https://github.com/jiaaro/pydub

[20] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," 2022.

[21] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: The munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM International Conference on Multimedia*. New York, NY, USA: Association for Computing Machinery, 2010, p. 1459–1462.

[22] B. Schuller, S. Steidl, A. Batliner, J. Hirschberg, J. K. Burgoon, A. Baird, A. Elkins, Y. Zhang, E. Coutinho, and K. Evanini, "The Interspeech 2016 computational paralinguistics challenge: Deception, sincerity & native language," in *Proceedings of INTERSPEECH*, vol. 8, 2016, pp. 2001–2005.

[23] Y.-K. Kim, R. Lahiri, M. Nasir, S. H. Kim, S. Bishop, C. Lord, and S. S. Narayanan, "Analyzing short term dynamic speech features for understanding behavioral traits of children with autism spectrum disorder." in *Interspeech*, 2021, pp. 2916–2920.

[24] J. W. Pennebaker and L. A. King, "Linguistic styles: language use as an individual difference." *Journal of Personality and Social Psychology*, vol. 77, no. 6, p. 1296, 1999.

[25] E.-M. Rathner, J. Djamali, Y. Terhorst, B. Schuller, N. Cummins, G. Salamon, C. Hunger-Schoppe, and H. Baumeister, "How did you like 2017? Detection of language markers of depression and narcissism in personal narratives," in *Proceedings of INTERSPEECH*, 2018, pp. 3388–3392.

[26] M. Asgari, J. Kaye, and H. Dodge, "Predicting mild cognitive impairment from spontaneous spoken utterances," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 3, no. 2, pp. 219–228, 2017.

[27] F. Mairesse and M. Walker, "Automatic recognition of personality in conversation," in *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, 2006, pp. 85–88.

[28] M. Honnibal and I. Montani. (2022) spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. [Online]. Available: https://github.com/explosion/spaCy

[29] (2022) List of (possible) English words. [Online]. Available: https://github.com/words

[30] C. Yang, D. Liu, Q. Yang, Z. Liu, and E. Prud'hommeaux, "Predicting pragmatic discourse features in the language of adults with autism spectrum disorder," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Student Research Workshop*, 2021, pp. 284–291.

[31] J. D. Brown, "Point-biserial correlation coefficients," *Statistics*, vol. 5, no. 3, pp. 12–6, 2001.

[32] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 12 2018.