COMPUTATIONAL ANALYSIS OF TRAJECTORIES OF LINGUISTIC DEVELOPMENT IN AUTISM

Emily Prud'hommeaux, Eric Morley†, Masoud Rouhizadeh†, Laura Silverman°*
Jan van Santen†, Brian Roark‡, Richard Sproat‡, Sarah Kauper°, Rachel DeLaHunta°

*Center for Language Sciences, University of Rochester
†Center for Spoken Language Understanding, Oregon Health & Science University
°Department of Pediatrics, University of Rochester Medical Center
‡Google Research

ABSTRACT

Deficits in semantic and pragmatic expression are among the hallmark linguistic features of autism. Recent work in deriving computational correlates of clinical spoken language measures has demonstrated the utility of automated linguistic analysis for characterizing the language of children with autism. Most of this research, however, has focused either on young children still acquiring language or on small populations covering a wide age range. In this paper, we extract numerous linguistic features from narratives produced by two groups of children with and without autism from two narrow age ranges. We find that although many differences between diagnostic groups remain constant with age, certain pragmatic measures, particularly the ability to remain on topic and avoid digressions, seem to improve. These results confirm findings reported in the psychology literature while underscoring the need for careful consideration of the age range of the population under investigation when performing clinically oriented computational analysis of spoken language.

Index Terms— natural language processing, clinical spoken language analysis, automated neurological assessment

1. INTRODUCTION

Atypical language has been associated with autism spectrum disorder (ASD) since the condition was first identified over 70 years ago [1, 2]. In verbal individuals without a comorbid language disorder, ASD language is characterized primarily by deficits in semantic and pragmatic expression, with syntax, morphology, and phonology left relatively unimpaired [3, 4]. In general, the sort of linguistic assessment that can reveal these trends is performed either through the administration of standardized instruments by clinicians or via manual annotation of spoken language transcripts by trained annotators, requiring substantial time and expertise. Automated methods of language analysis have the potential to offer an efficient and objective supplement to the standard manual procedures.

One common method for eliciting spontaneous language data is the narrative retelling paradigm, in which a subject listens to a story and must retell the story to an examiner, who then evaluates the content of the story in real time or transcribes the story for more detailed manual analysis. Previous such work has found that children with ASD, when generating narrative retellings, produce more off-topic content [5, 6], unusual word use [7], and different distributions of disfluencies [8, 9, 7] than their age- and language-matched typically developing peers. These results, however, have sometimes been contradictory, leading to questions about the ease of reliably performing the required annotations, as well as the relative significance of the between-group differences as language development progresses.

The goals of this paper are thus twofold. First, we attempt to capture the linguistic characteristics associated with autism as reported in the psychology literature using automated techniques described in the computational linguistics literature. Second, we try to determine whether any between-group differences that are observed persist with age. We find that many of the automated measures we propose seem to tap into previously identified linguistic features characteristic of ASD narratives. By applying these measures to two distinct populations of subjects from disjoint age groups, we find that although some differences between ASD narratives and those produced by children with typical development (TD) persist with increased age, others seem to disappear, particularly those related to pragmatic expression. The results presented here suggest that some difficulties experienced by individuals with ASD when generating narratives can persist, while others resolve either with remediation or simply with advancing age and development.

2. BACKGROUND

There is a substantial body of work involving manual analysis of narratives in children, including children with developmental disorders such as ASD. A number of interesting differences between children with ASD and typical development (TD) in the language each group uses to express narratives retellings have been observed, and when the diagnostic groups are matched for IQ and language ability, most of these differences seem to be found the areas of semantics and pragmatics. Differences in syntax, morphology, and raw word and sentence count are typically not observed when the language measures of the ASD cohort are within typical range [6, 10].

One way in which pragmatic expression can be disrupted is in the use of disfluencies, including revisions, repetitions, and false starts, as well as the use of filled pauses. Certain types of disfluencies have been reported to have different distributions in ASD speech, but results have been mixed, with repetitions reported to be more common in ASD and revisions and false starts more common in TD [9, 7]. Previous work in analyzing filled pauses in ASD spoken language has produced somewhat contradictory results, with some finding that ASD language contains fewer filled pauses than TD language [9], some finding no difference [7], and still others finding differences in only one class of filled pauses [8]. It has been argued that only a subset of disfluencies, including self-corrections and filled pauses, serve a communicative purpose, making them perhaps more difficult to use for people with pragmatic impairments [9, 11]. Notably, these differences seem to be most significant in younger children, with ASD adults displaying only higher rates of repetitions [12].

Unusual word use or idiosyncratic language is another characteristic of the pragmatic deficits long observed in ASD language [2]. Such words can be semantically appropriate but simply rare, overly formal, or unexpected [7]. Alternatively, the unusual language can be unusual precisely because of its inappropriate, digressive, and off-topic nature [6, 5]. We note that the work reporting increased topic excursions focused on pre-adolescent populations, and similar trends were not observed in older adolescents [7].

Until recently, there has been little work on automated analysis of ASD language for the purposes of identifying pragmatic deficits. Using techniques derived from information extraction and distributional semantics, Rouhizadeh and colleagues [13] found that young children with ASD produced more off-topic words in their narratives, which was confirmed via manual annotation. In a separate study, Rouhizadeh and colleagues [14] found that the rate of word overlap between the narratives of young children with ASD and those of children with TD is lower than that between children with TD and one another, and still lower than that with other children with ASD, suggesting that children with ASD are veering off-topic in idiosyncratic ways. Using latent semantic analysis (LSA) to compare children's narrative retellings with the source narrative, Losh and Gordon [15] found that the ASD narratives had lower LSA scores than TD narratives. All three of these previous studies using automated measures were performed on narratives from young school-aged children and pre-adolescents.

Dx	n	Mean age	Mean NVIQ	Mean VIQ
TD	39	6.2	117	118
ASD	21	6.5	113	105

Table 1. Demographics of younger children's (YC) corpus.

Dx	n	Mean age	Mean NVIQ	Mean VIQ
TD	24	15.4	104	105
ASD	25	15.8	103	107

Table 2. Demographics of older children's (OC) corpus.

3. DATA

We analyze two separate corpora of narratives retellings for this paper: one produced by early school-aged children between the ages of 4 and 9, and the other produced by adolescents between the ages of 13 and 17. To our knowledge, this is the first attempt to apply multiple computational measures to two sets of narratives produced by children from two distinct age groups.

The first corpus, which we will refer to as the younger children's (YC) corpus, consists of narrative retellings elicited via the the Narrative Memory subtask of the NEPSY, an instrument used to evaluate neurospychological development [16]. In the Narrative Memory subtask, a child listens to and retells a story about a boy who gets stuck in a tree and is rescued by his dog. The NEPSY retellings were produced by 39 children with typical development and 21 children with ASD. Demographic information about the participants is shown in Table 1. There were no significant between-group differences in age or non-verbal IQ. There was a significant between-group difference in verbal IQ ($p < 0.001$), but none of the children in either group met criteria for a language disorder. The children generating these narratives ranged in age from 4.1 to 9.0, representing the younger of the two populations investigated here.

The second corpus, the older children's (OC) corpus, consists of narrative retellings of a story designed to study of the impact of gesture on narrative memory. The narrative used in this study, which tells the story of a girl named Claire who traveled to Paris, was divided into three parts, two of which were presented in video-and-audio format and one of which was presented as audio only. The subjects were asked to retell each part immediately upon hearing or viewing that part. Twenty-four children with typical development and 25 children with ASD participated in this study. Demographic information about the participants is shown in Table 2. There were no significant differences in age, verbal IQ, or non-verbal IQ. The children generating these narratives ranged in age from 13.2 to 17.9, making this group the older of the two populations investigated here.

In both corpora, the narratives were recorded and then transcribed by research assistants at the two data collection sites. Each transcript was then manually reviewed by the first author in order to ensure consistency across collection sites in spelling, punctuation, utterance boundary insertion, and the use of contractions, abbreviations, and pause-fillers. For some of the analyses performed, further normalization was performed, as described below.

4. FEATURE EXTRACTION

Each of the features described below was extracted from each retelling transcript in both corpora. Recall that both groups of children investigated here have unimpaired language. Thus, we expect that there were will be few, if any, between-group differences in the syntactic features. Given previous results, however, we do anticipate that some of the disfluency features and semantic and pragmatic features will distinguish the narratives of children with ASD from those with TD in one or both of the age groups.

4.1. Syntactic features

Numerous features designed to capture the degree of syntactic complexity of a retelling were extracted from the transcripts after tokenization and removal of punctuation. These include three basic word-level features: total number of words per retelling, total number of utterances per retelling, and average utterance length, which roughly approximates the widely used developmental language measure of mean length of utterance (MLU). We also extracted features from part-of-speech information, constituency parses, and dependency parses, all of which were generated using the Charniak parser [17], trained on Switchboard [18], a corpus of transcribed telephone conversations. These features include mean words per clause, clauses per utterance, and dependency length per word, as well as two measures designed to measure syntactic complexity: Yngve score per word and Frazier score per word. The Yngve score of a word is the size of the stack of a shift-reduce parser after that word, while the Frazier score counts how many intermediate nodes exist in the tree between the word and its lowest ancestor that is either the root or has a left sibling in the tree. All of these features have been used to measure syntactic complexity in previous research in analyzing language in populations with neurodevelopmental and neurodegenerative disorders [19, 20].

4.2. Disfluency features

We next extracted features related to disfluencies. Using an existing disfluency detection algorithm [21], which considers revisions, repetitions, false starts, and filled pauses to be a single category of disfluency, we identified disfluencies in the tokenized transcripts. In previous work, this algorithm yields highly accurate disfluency tagging and bracketing, with F1 measures as high as .90 and .84 respectively, depending on the test and training corpus used. From these annotations we extracted the number of disfluencies normalized over the total number of utterances. Previous work on disfluency use indicates that the English pause-fillers *um* and *uh* serve different purposes and have different distributions [11, 22]. Following Heeman and colleagues [8], who found that children with autism produce fewer instances of *um* than children with TD but a comparable number of instances of *uh*, we also calculated the per-utterance rate of use of these two pause-fillers.

4.3. Lexical and semantic features

Finally, we consider features that we propose are directly related to semantic and pragmatic expression, the area of language most widely reported to be impaired in ASD. Two of these measures are designed to capture commonalities between retellings. First, using tokenized transcripts with punctuation removed, we identify every word in every retelling that never appears in another retelling and calculate for each child the rate of use of unique tokens and types. Secondly, following Rouhizadeh and colleagues [14], after stemming and lemmatizing the transcripts and removing stop words, we calculate the percent of word overlap between each pair of children and derive the average word overlap between each child and all other children. Following the work of Losh and colleagues [15], we use the University of Colorado latent semantic analysis (LSA) online tool (http://lsa.colorado.edu) to calculate the semantic similarity between each retelling and the text of the source narrative. For the YC corpus, we selected the 3rd-grade-level topic space, and for the OC corpus, we use the 12th-grade-level topic space.

5. RESULTS

To determine whether the features extracted here are significantly different between diagnostic groups within each corpus, we performed two sample, one-tailed t-tests. As expected, none of the syntactic complexity features were significantly different in either corpus. The lack of differences in any of the measures of syntactic complexity is consistent with previous findings that syntax is relatively unimpacted in individuals with ASD without language impairment.

The results of the analysis of the disfluency features are found in Table 3. In both corpora, we see significant between-group differences in the number of disfluencies per utterance. In the younger group, it is the children with TD who produce more disfluencies per utterance, while in the older group it is the ASD children who produce more disfluencies per utterance. Because the algorithm we used does not distinguish the different subtypes of disfluencies, it remains to be determined whether the specific differences observed in previous work, such as the presence more disfluencies overall among young TD children and more repetitions in particular in individuals of all ages with ASD, obtain. The rate of use of the pause-fillers *uh* and *um*, however, reveals an interesting change over time. Among the younger children, the TD group uses *um* more frequently than the ASD group, as predicted by [8].

Feature	YC			OC		
	TD mean	ASD mean	*p*	TD mean	ASD mean	*p*
Disfluencies per utterance	0.55	0.39	$<$0.05	0.78	0.92	$<$0.05
uh per utterance	0.07	0.09	n.s.	0.16	0.12	n.s.
um per utterance	0.14	0.07	$<$0.05	0.19	0.20	n.s.

Table 3. Comparison of disfluency features in younger children (YC) and older children (OC).

Feature	YC			OC		
	TD mean	ASD mean	*p*	TD mean	ASD mean	*p*
Unique type rate	0.04	0.07	$<$0.01	0.03	0.05	$<$0.01
Unique token rate	0.03	0.05	$<$0.01	0.02	0.03	$<$0.01
Mean percent overlap with other children	0.18	0.14	$<$0.01	0.26	0.24	n.s.
LSA with source narrative	0.70	0.61	$<$0.05	0.82	0.80	n.s.

Table 4. Comparison of lexical and semantic features in younger children (YC) and older children (OC).

This difference, however, fades in the older children. If the pause-filler, *um*, does indeed serve a subtle communicative purpose, these results suggest that the children may eventually grasp the pragmatic importance of using *um*. We note that the higher rate of use of pause-fillers in general in the older group is to be expected given the reported positive correlation between disfluency frequency, utterance length, and grammatical complexity [23].

Finally, we present the between-group differences observed in the lexical and semantic features in Table 4. The rate of use of both unique tokens and types is significantly different between the two diagnostic groups in both corpora. We interpret this to mean that children with ASD are using idiosyncratic language in the form of unusual or unexpected words to describe the events of the narrative. Interestingly, the overlap measure and the LSA score are significantly different between groups in the younger children but not in the older children. In other words, the retellings of younger children with ASD are tend to have less in common with other children's retellings and the source narrative itself, both lexically and semantically. This suggests that the younger children are veering away from the target topic of the narrative, indicating problems with pragmatic competence. These differences are not observed in the retellings of the older children. Again, it seems that some of the impairments in pragmatic expression that have been observed in ASD, such as poor topic maintenance, improve with age, while others, particularly idiosyncratic language, persist.

6. CONCLUSIONS AND FUTURE WORK

Despite significant work dedicated to analyzing spoken narratives in order to detect the pragmatic deficits often associated with autism spectrum disorder, the precise nature of these deficits in a narrative context is not entirely clear. We propose that this may partly be due to the focus of such work on either a very wide age range or a very narrow range of relatively young subjects, particularly in the research using automated computational techniques for language analysis. In this paper, we offer evidence that some pragmatic abilities seem to change and even improve with age, underscoring the need to carefully consider the study population demographics when applying these analyses, whether manual or automated, to make generalizations about language in ASD.

One of the main weaknesses of the work presented here is that we rely on previous work to inform the relationships between features observed by human annotators, such as degree of topic maintenance, and features that can be extracted automatically from a transcript of a spoken narrative. In future work, we plan to manually annotate instances of digressions and idiosyncratic word use in order to determine whether our automated measures truly capture these phenomena. In addition, given the somewhat mixed disfluency results, we would like to train models not only to identify disfluencies but to separate them into their respective subclasses, which will require further manual annotation. Finally, we would like to explore other widely used measures of text similarity and topic modeling, which might provide further computationally derived evidence for the pragmatic deficits reported here and elsewhere in the literature.

Although we have focused on ASD in this research, most of the techniques and features explored here can be easily applied to spoken language data elicited from subjects with other neurological disorders, including specific language impairment and dementia. Seniors in the early stages of dementia in particular display many of the same pragmatic and semantic problems in their narratives as children with ASD, although the etiology of these symptoms is very different. In future work, we would like to apply these same methods to narrative retellings of the elderly in order to identify the digressions and instances of unusual word use often observed in the early stages of cognitive impairment.

7. ACKNOWLEDGMENTS

This work was supported in part by NSF Grant #BCS-0826654, NIH NIDCD grants #1R01DC012033-01 and R01DC007129. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the NSF or the NIH.

8. REFERENCES

[1] Leo Kanner, “Autistic disturbances of affective content,” *Nervous Child*, vol. 2, pp. 217–250, 1943.

[2] Leo Kanner, “Irrelevant and metaphorical language,” *American Journal of Psychiatry*, vol. 103, pp. 242–246, 1946.

[3] Helen Tager-Flusberg, “Understanding the language and communicative impairments in autism,” *International Review of Research in Mental Retardation*, vol. 23, pp. 185–205, 2001.

[4] Catherine Lord and Rhea Paul, “Language and communication in autism,” in *Handbook of Autism and Pervasive Developmental Disorders*, D. Cohen and F. Volkmar, Eds., pp. 195–225. Wiley, New York, 1997.

[5] Molly Losh and Lisa Capps, “Narrative ability in high-functioning children with autism or asperger’s syndrome,” *Journal of Autism and Developmental Disorders*, vol. 33, no. 3, pp. 239–251, 2003.

[6] Joshua J. Diehl, Loisa Bennetto, and Edna Carter Young, “Story recall and narrative coherence of high-functioning children with autism spectrum disorders,” *Journal of Abnormal Child Psychology*, vol. 34, no. 1, pp. 87–102, 2006.

[7] Joyce Suh, Inge-Marie Eigsti, Letitia Naigles, Marianne Barton, Elizabeth Kelley, and Deborah Fein, “Narrative performance of optimal outcome children and adolescents with a history of an autism spectrum disorder (asd),” *Journal of Autism and Developmental Disorders*, vol. 44, pp. 1681–1694, 2014.

[8] Peter A. Heeman, Rebecca Lunsford, Ethan Selfridge, Lois Black, and Jan van Santen, “Autism and interactional aspects of dialogue,” in *11th Annual SIGdial Meeting on Discourse and Dialogue*, 2010.

[9] ?Johanna K. Lake, Karin R. Humphreys, and Shannon Cardy, “Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders,” *Psychonomic Bulletin and Review*, vol. 18, pp. 135–140, 2011.

[10] Livia Colle, Simon Baron-Cohen, Sally Wheelwright, and Heather K. J. van der Lely, “Narrative discourse in adults with high-functioning autism or asperger syndrome,” *Journal of Autism and Developmental Disorders*, vol. 38, pp. 28–40, 2008.

[11] John Fox Tree, “Listeners uses of um and uh in speech comprehension,” *Memory and Cognition*, vol. 29, no. 2, pp. 320–326, 2001.

[12] K.M. Belardi and D.L. Williams, “Disruptions in spontaneous language in hfa: Indicators of linguistic processing challenges,” in *Presented at the International Meeting for Autism Research, Chicago*, 2009.

[13] Masoud Rouhizadeh, Emily Prud’hommeaux, Brian Roark, and Jan van Santen, “Distributional semantic models for the evaluation of disordered language,” in *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, 2013.

[14] Masoud Rouhizadeh, Emily Prud’hommeaux, Jan van Santen, and Richard Sproat, “Detecting linguistic restricted interests in autism using distributional semantic models,” in *Proceedings of the ACL 2014 Workshop on Computational Linguistics and Clinical Psychology*, 2014.

[15] Molly Losh and Peter C. Gordon, “Quantifying narrative ability in autism spectrum disorder: A computational linguistic analysis of narrative coherence,” *Journal of Autism and Developmental Disorders*, 2014.

[16] Marit Korkman, Ursula Kirk, and Sally Kemp, *NEPSY: A developmental neuropsychological assessment*, The Psychological Corporation, San Antonio, 1998.

[17] Eugene Charniak, “A maximum-entropy-inspired parser,” in *Proceedings of the 1st Conference of the North American Chapter of the ACL*, 2000, pp. 132–139.

[18] John J. Godfrey, Edward Holliman, and Jane McDaniel, “SWITCHBOARD: telephone speech corpus for research and development,” in *Proceedings of ICASSP*, 1992, vol. 1, pp. 517–520.

[19] Brian Roark, Margaret Mitchell, John-Paul Hosom, Kristina Hollingshead, and Jeffrey Kaye, “Spoken language derived measures for detecting mild cognitive impairment,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 7, pp. 2081–2090, 2011.

[20] Emily Prud'hommeaux, Brian Roark, Lois Black, and Jan van Santen, "Classification of atypical language in autism," in *Proceedings of the 2nd ACL Workshop on Cognitive Modeling and Computational Linguistics (CMCL)*, 2011.

[21] Eric Morley, Anna Hallin, and Brian Roark, "Challenges in automating maze detection," in *Proceedings of the ACL 2014 Workshop on Computational Linguistics and Clinical Psychology*, 2014.

[22] Herbert Clark and John Fox Tree, "Using uh and um in spontaneous speaking," *Cognition*, vol. 84, pp. 73111, 2002.

[23] Barbara G. MacLachlan and Robin S. Chapman, "Communication breakdowns in normal and language learning-disabled children's conversation and narration," *Journal of Speech and Hearing Disorders*, vol. 53, pp. 2–7, 1988.