Automatic detection of pragmatic deficits in children with autism

Emily Prud'hommeaux and **Masoud Rouhizadeh**

Center for Spoken Language Understanding, Oregon Health & Science University, Portland, Oregon, USA

Abstract

Autism spectrum disorder (ASD) is characterized by atypical and idiosyncratic language, which often has its roots in pragmatic deficits. Identifying and measuring pragmatic language ability is challenging and requires substantial clinical expertise. In this paper, we present a method for automatically identifying pragmatically inappropriate language in narratives using two features related to relevance and topicality. These features, which are derived using techniques from machine translation and information retrieval, are able to distinguish the narratives from children with ASD from those of their language-matched peers and may prove useful in the development of automated screening tools for autism and neurodevelopmental disorders.

Index Terms

spoken language evaluation; child language; diagnostic tools; discourse analysis

1. Introduction

Atypical or idiosyncratic language has been associated with autism since Kanner first gave a name to the disorder [1]. This atypicality is used as a diagnostic criterion in many of the widely used screening instruments for autism [2, 3, 4]. The current view is that phonological, morphological, and syntactic skills are typically spared in autism spectrum disorders (ASD), while pragmatic expression is impaired [5, 6], which contributes to the perceived atypicality of language in individuals with ASD.

Narrative retelling tasks, in which a subject listens to a brief narrative and must retell the narrative to the examiner, are widely used in neuropsychological assessment instruments in order to test memory and language ability [7, 8, 2]. Since relating a narrative requires pragmatic competence, the study of narrative production in ASD is particularly useful for revealing pragmatically atypical language. A narrator must stay on topic and present the story in a meaningful way in order to guarantee that the listener understands the narrative. Deviations from these expected communication behaviors can be indicative of pragmatic language problems.

In this paper, we present work on identifying features associated with pragmatic language deficits in the spoken narrative retellings of children with ASD. Previous work in the

emilypx@gmail.com, rouhizad@ohsu.edu.

neuropsychology literature on this topic has relied on extensive expert manual annotation. In contrast, the features discussed here, while grounded in observations from the literature, can be derived in an automated fashion using techniques adapted from machine translation and information retrieval.

The evaluation of pragmatic language ability on a very general scale, as is required by many of the standard diagnostic instruments for autism, is inherently subjective and difficult to perform reliably. The automated methods described here provide an entirely objective and efficient way of characterizing some of the specific pragmatic deficits observed in ASD narratives. Although there are many other factors contributing to atypical language in ASD, automated approaches like these could lead to future work in developing diagnostic screening tools for autism and other developmental disorders.

2. Background

Clinically elicited narratives are often analyzed in terms of syntactic or lexical properties, such as sentence length, vocabulary diversity, or grammatical complexity, but such features rarely distinguish children with ASD from their language-matched peers [9, 10]. The more reliable distinctive features tend to be related to pragmatics. Expert manual annotation and analysis of narratives in young people with ASD has revealed at least one consistent trend. The narratives of children and young adults with autism include significantly more bizarre, irrelevant, and off-topic content [11, 12, 13] than the retellings of their language-matched peers. In addition, ASD narratives contain more frequent instances of inappropriate story-telling [10].

Existing work on automated analysis of narratives in children has focused primarily on syntactic and structural features rather than pragmatic or semantic features. Using the CHILDES corpus [14], which includes a narrative task, Sagae et al. [15] extracted grammatical relations in order to calculate the Index of Productive Syntax, a measure of syntactic development in children [16]. Gabani et al. [17] used part-of-speech language models, as well as other features such as utterance length and subject-verb agreement, to characterize agrammaticality in narratives of children with language impairments.

Automated analysis of the content and pragmatics of narratives in the elderly has received more attention. In particular, several researchers have developed techniques for scoring narrative retellings relative to the source narrative. Latent semantic analysis [18] and unigram overlap [19] have both been used to approximate the standard scoring procedures normally applied to narrative retellings. In previous work, we automatically scored narrative retellings produced by elderly subjects, according to the published scoring guidelines, by aligning the retellings to the source narrative [20]. We have also used alignments to explore other narrative features relating to the relevance of content and ordering of events in the narrative retellings of elderly subjects [21]. In the analysis of children's narrative described in Section 4, we will use alignments to derive a measure of topical content. In addition, we will explore measures of word atypicality that can be estimated using word centrality measures used in information retrieval.

3. Data

3.1. Study participants

In this paper, we analyze the narrative retellings of 97 children, ranging in age from 4 years to 8 years. All subjects were required to have a full-scale IQ of greater than 70, indicating no intellectual disability, and a mean length of utterance (MLU) of at least 3, indicating an adequate level of verbal fluency. IQ was measured using the WPPSI-III [22] for children under 7, and the WISC-IV [23] for children 7 and older. Exclusion criteria included any brain lesion or neurological condition unrelated to the diagnoses of interest; orofacial abnormalities; bilinguality; extreme unintelligibility; gross sensory or motor impairments; or identified mental retardation. In addition, the exclusion criteria for the typical development (TD) group included a history of psychiatric disturbance and a family history ASD or language impairment.

We categorize the subjects into four diagnostic groups: typical development (TD, $n = 40$); specific language impairment not meeting the criteria for autism (SLI, $n = 17$); ASD not meeting criteria for a language impairment (ALN, $n = 21$); and ASD meeting the criteria for a language impairment (ALI, $n = 21$). A subject received a diagnosis of a language impairment if his CELF Core Language Index [24] was more than one standard deviation below the mean. A diagnosis of ASD was indicated if the child reached the ASD cut-off score on two instruments: the Autism Diagnostic Observation Schedule (ADOS) [2], a semi-structured series of activities designed to allow an examiner to observe behaviors associated with autism; and the Social Communication Questionnaire (SCQ) [4], a parental questionnaire. A team of experienced clinicians generated for each child a consensus diagnosis according the instrumental findings and the diagnostic criteria listed in the DSM-IV-TR [25].

3.2. NEPSY Narrative Memory

The narrative retelling task analyzed here is the NEPSY Narrative Memory task, which was administered to the subjects as part of an earlier study on prosody in autism and language disorders. The NEPSY Narrative Memory (NNM) task is a narrative retelling test in which the subject listens to a brief narrative and then must retell the narrative to the examiner. The examination scoring materials include a list of 17 story elements, each corresponding to a word or phrase expressing a detail or event in the narrative. Figure 1 shows the NNM narrative, with the portions of the text corresponding to the 17 story elements placed in brackets and numbered according to the published scoring guidelines. The score for a retelling reported under standard administration of the task is simply the total number of story elements recalled in the retelling. The elements do not need to be recalled verbatim or in the correct order, and all elements contribute equally to the final score.

The retelling of the NNM narrative for each experimental subject was recorded and later transcribed at the word level, tokenized, and downcased, and all punctuation and pause-fillers were removed. In Figures 2 and 3, we provide two short sample retellings produced by a child with SLI and a child with ALI. The child with SLI struggles to tell the story, but stays on-topic, first setting the stage by describing the characters and then by outlining the

resolution of the story, when Pepper brings the shoe to Anna so that she can rescue Jim. The child with ALI, on the other hand, very quickly strays from the source narrative. The contrast between these two retellings illustrates some of the features of ASD retellings described in Section 2.

4. Measuring topicality and relevance

4.1. Word alignment

The first topicality and relevance feature we will discuss is a measure of how well a retelling can be aligned to the source narrative. In previous work on scoring narrative retellings in the elderly, we generated word-level alignments between a retelling and the source narrative, from which we then extracted several narrative features, including a measure of relevance [21]. These word alignments are similar to those used in machine translation (MT), where each word can be considered a translation of the word to which it is aligned.

Standard word alignment packages for machine translation, such as the Berkeley aligner [26] and Giza++ [27], start with a parallel corpus consisting of sentences in one language, the source language, and translations of those sentences in another language, the target language. These packages then use a pipeline of expectation maximization techniques to determine in an unsupervised fashion which words in each source sentence are aligned to (i.e., are translations of) the words in each corresponding target sentence. The techniques used to generate to word alignments between two languages in MT can also be used to align a collection of narrative retellings to a source narrative. Figure 4 shows a snippet of the word-level alignment between the SLI retelling in Figure 3 in the source narrative shown in Figure 1.

In order to automatically generate word alignments, we need to collect an in-domain parallel corpus and then train a word alignment model on that corpus. In order to ensure a perfect match for the domain, we build a parallel corpus consisting only of data from the NNM narrative retellings. From this corpus, a word alignment model that is tailored to this specific alignment task can be learned. The corpus includes: (1) a source-to-retelling corpus containing the bracketed portions of the source narrative, representing the story elements, and the retellings for all 97 study participants, as well as roughly 160 retellings from other individuals not participating in the study; (2) a full pairwise retelling-to-retelling corpus; (3) a manually derived corpus of the phrases in the non-experimental retellings aligned with their corresponding story elements in the source narrative; (4) a word identity corpus, consisting of each word that appeared in a retelling aligned to itself. Example lines from each of these corpora are provided in Figure 5.

We use the Berkeley aligner [26] to build an alignment model on this four-part parallel corpus. The Berkeley aligner estimates a set of initial word-to-word alignment probabilities using a simple mixture model relying on co-occurrence statistics. These initial probability estimates serve as the input probabilities for an HMM, in which expectation maximization is used not only to refine the word-to-word alignment probabilities but also to estimate probabilities modeling the reordering of the words in a retelling relative to the source narrative. We use the alignment model generated by the Berkeley aligner to align every

retelling to the bracketed portions of the source narrative corresponding from the story elements.

Recall that previous work has found that the retellings of children with ASD contain more off-topic and bizarre content and more instances of inappropriate story telling. The feature we extract from the alignments captures how well the child stays on topic during his retelling by measuring the amount of a retelling's content that is related to the content of the source narrative. First we determine which words in a retelling transcript are content words rather than function words. We then count the number of these content words that are aligned to some word in the source narrative. The percentage of content words in a child's retelling that cannot be aligned to a word in the source narrative is an estimate of how much of the retelling is unrelated to the source narrative, which in turn is a measure of how far off-topic the child has veered.

4.2. Log odds score

Another way to evaluate the relevance and topicality of a retelling is to measure the degree to which the retelling contains unexpected or unlikely words given the context of the source narrative. We expect that different retellings of the same source narrative will share a common vocabulary. The presence of words outside of this vocabulary in a retelling may indicate that the speaker has strayed from the story or is failing to select the appropriate words, both of which are signs of pragmatic language difficulties.

Identifying these outliers is similar to the information retrieval (IR) task of clustering documents together based on a common topic. Documents with a similar vocabulary are likely to be about the same topic; documents containing numerous words that are not found in the other documents are likely to be about a different topic. There are several document lexical similarity measures used in IR to cluster documents. Here we will use one of those similarity measures, the log odds score, to characterize the relevance and topicality of a retelling relative to the other retellings.

In order to establish the topic of the NNM narrative relative to other topics, we include transcribed conversational data from four interactive activities of the Autism Diagnostic Observation Schedule (ADOS) [2], an instrument used in the diagnosis of ASD. Each child's retelling is compared to a collection of documents that includes every other retelling as well as each of the 4 ADOS activities for every other child. The log odds score for each word in a child's retelling will be 0 or close to 0 if the word is as likely to appear in the retelling as in any other document. As the log odds score increases, so does the probability that the word is specific to the retelling and unlikely to appear in another retelling or any of the ADOS activities.

In order to generate a single log odds score for a complete retelling, we calculate the log odds score for each word in the retelling and take the mean of that set of scores. A retelling with a high mean log odds score likely contains many words that are unusual or unexpected given the context. We predict that this will indicate that the child has strayed off-topic or is failing to communicate the story in a typical or appropriate way.

5. Diagnostic classification

We now proceed to use these two features of topicality and relevance in order to perform diagnostic classification. Recall that many of the features of narratives that purport to distinguish typical development from ASD fail to distinguish the groups when they are matched for language ability. For this reason, we will perform three pairwise classification tasks: 1) typical development (TD) vs. ASD with no language impairment (ALN); 2) specific language impairment (SLI) vs. ASD with a language impairment (ALI); and 3) Non-ASD (TD and SLI) vs. ASD (ALN and ALI).

We perform classification using the support vector machine (SVM) implemented in LibSVM [28] within the Weka API [29]. Classification accuracy is evaluated using a leave-pair-out cross-validation scheme, in which every pairwise combination of a positive and negative example is tested against an SVM built on the remaining examples. We report accuracy as the area under the receiver operating characteristic curve (AUC) [30].

In Table 1, we see that although both features individually yield reasonable classification accuracy for the two smaller group comparisons, the feature measuring the amount of unaligned content produces better results than the log odds score. There seems to be no real cumulative effect of combining the features when distinguishing the two smaller groups, but we find that when comparing the non-ASD subjects to the ASD subjects, we achieve much higher classification accuracy when using both features together. We note that neither of these features was able to distinguish the two ASD groups from each other, which suggests that we have identified two features of narrative topicality that are specific to ASD and independent of language ability.

6. Conclusions and future work

The work presented here is a promising first step in automatically identifying pragmatic language impairments in children with autism spectrum disorder, a task that has long relied on extensive manual annotation and clinical expertise. The features we developed are grounded in clinical observations of language in ASD but were extracted from transcripts in an unsupervised and automatic fashion. The results indicate that searching for additional features of topicality and relevance that can be derived using NLP techniques may further improve diagnostic classification accuracy.

Although we were able to use these automatically derived features to distinguish children with ASD from their language-matched peers, we have not yet manually determined whether these features correlate with human judgements of relevance, topicality, and coherence, or whether such judgements are able to distinguish the subjects in our data set. We plan to manually code the transcripts according to the guidelines described in previous work in the neuropsychology literature on narrative coherence in ASD. In addition, we would like to explore other narrative coherence features, such as element ordering, and other document similarity measures that are widely used in information retrieval, such as those derived using graph-based methods, to more accurately identify unusual and unexpected words in the retellings of children with ASD. Finally, we plan to experiment with using ASR

to generate the transcripts in order to produce a fully automated narrative-based screening tool.

Acknowledgments

References

1. Kanner L. Autistic disturbances of affective content. Nervous Child. 1943; 2:217–250.

2. Lord, C., Rutter, M., DiLavore, P., Risi, S. Autism Diagnostic Observation Schedule (ADOS). Western Psychological Services; 2002.

3. Lord C, Rutter M, LeCouteur A. Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of Autism and Developmental Disorders. 1994; 24:659–685. [PubMed: 7814313]

4. Rutter, M., Bailey, A., Lord, C. Social Communication Questionnaire (SCQ). Los Angeles: Western Psychological Services; 2003.

5. Tager-Flusberg H. Understanding the language and communicative impairments in autism. International Review of Research in Mental Retardation. 2001; 23:185–205.

6. Lord, C., Paul, R. Language and communication in autism. In: Cohen, D., Volkmar, F., editors. Handbook of autism and pervasive developmental disorders. Wiley; 1997. p. 195-225.

7. Wechsler, D. Wechsler Memory Scale – Third Edition. The Psychological Corporation; 1997.

8. Korkman, M., Kirk, U., Kemp, S. NEPSY: A developmental neuropsychological assessment. The Psychological Corporation; 1998.

9. Tager-Flusberg H. Once upon a ribbit: Stories narrated by autistic children. British journal of developmental psychology. 1995; 13(1):45–59.

10. Diehl JJ, Bennetto L, Young EC. Story recall and narrative coherence of high-functioning children with autism spectrum disorders. Journal of Abnormal Child Psychology. 2006; 34(1):87–102. [PubMed: 16485176]

11. Loveland K, McEvoy R, Tunali B. Narrative story telling in autism and down's syndrome. British Journal of Developmental Psychology. 1990; 8(1):9–23.

12. Losh M, Capps L. Narrative ability in high-functioning children withautism oraspergers syndrome. Journal of Autism and Developmental Disorders. 2003; 33(3)

13. Goldman S. Brief report: Narratives of personal events in children with autism and developmental language disorders: Unshared memories. Journal of Autism and Developmental Disorders. 2008; 38

14. MacWhinney, B. The CHILDES Project: Tools for Analyzing Talk. Lawrence Erlbaum Associates; 2000.

15. Sagae K, Lavie A, MacWhinney B. Automatic measurement of syntactic development in child language. Proceedings of ACL. 2005:197–204.

16. Scarborough HS. Index of productive syntax. Applied Psycholinguistics. 1990; 11:1–22.

17. Gabani K, Sherman M, Solorio T, Liu Y. A corpus-based approach for the prediction of language impairment in monolingual English and Spanish-English bilingual children. Proceedings of NAACL-HLT. 2009:46–55.

18. Dunn JC, Almeida OP, Barclay L, Waterreus A, Flicker L. Latent semantic analysis: A new method to measure prose recall. Journal of Clinical and Experimental Neuropsychology. 2002; 24(1):26–35. [PubMed: 11935421]

19. Hakkani-Tur D, Vergyri D, Tur G. Speech-based automated cognitive status assessment. Proceedings of Interspeech. 2010

20. Prud'hommeaux E, Roark B. Graph-based alignment of narratives for automated neuropsychological assessment. Proceedings of the NAACL 2012 Workshop on Biomedical Natural Language Processing (BioNLP). 2012

21. Prud'hommeaux E, Roark B. Extraction of narrative recall patterns for neuropsychological assessment. Proceedings of Interspeech. 2011

22. Wechsler, D. Wechsler Primary and Preschool Scale of Intelligence – Third edition (WPPSI-III). Harcourt Assessment; 2002.

23. Wechsler, D. Wechsler Primary, Wechsler Intelligence Scales for Children – Fourth Edition (WISC-IV). The Psychological Corporation; 2003.

24. Paslawski T. The clinical evaluation of language fundamentals, fourth edition (celf-4): A review. Canadian Journal of School Psychology. 2005; 20(1–2):129–134.

25. American Psychiatric Association. DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders. 4th. Washington, DC: American Psychiatric Publishing; 2000.

26. Liang P, Taskar B, Klein D. Alignment by agreement. Proceedings of the Human Language Technology Conference of the NAACL. 2006

27. Och FJ, Ney H. A systematic comparison of various statistical alignment models. Computational Linguistics. 2003; 29(1):19–51.

28. Chang C-C, Lin C-J. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology. 2011; 2(27):1–27.

29. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: An update. SIGKDD Explorations. 2009; 11(1)

30. Egan, J. Signal Detection Theory and ROC Analysis. Academic Press; 1975.

31. Hanley J, McNeil B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982; 143:29–36. [PubMed: 7063747]

[Jim]₁ was a boy whose best friend was [Pepper]₂. Pepper was a [big]₃ [black]₄ dog. Jim [liked to walk in the woods and climb the trees]₅. Near Jim's house was a very tall [oak tree with branches so high that he couldn't reach them]₆. Jim always wanted to climb that tree, so one day he [took a ladder from home and carried it to the oak tree]₈. He [climbed up]₇, sat on a branch, and [looked out over his neighborhood]₉. When he started to get down, [his foot slipped, his shoe fell off, and the ladder fell to the ground]₁₀. Jim held onto a branch so he didn't fall, but he couldn't get down. Pepper sat below the tree and barked. Suddenly [Pepper took Jim's shoe in his mouth and ran away]₁₁. [Jim felt sad. Didn't his friend want to stay with him when he was in trouble?]₁₂ [Pepper took the shoe to]₁₃ [Anna]₁₄, [Jim's sister]₁₅. He [barked and barked]₁₆. Finally, Anna understood that Jim was in trouble. She followed Pepper to the tree where Jim was stuck. [Anna put the ladder up and rescued Jim]₁₇. Wasn't Pepper a smart dog?

Figure 1.
Text of NNM narrative with story elements bracketed and numbered.

He had a friend named Pepper. Pepper was a black dog. Pepper. I forgot. Pepper got his shoe. I don't know. Jim was a little boy. Pepper was his friend. Pepper was a black dog, and Pepper rescued his shoe when he brought it to Anna. That's all I know. And then they, then Anna rescued him.

Figure 2.
Sample NNM retelling from a child with specific language impairment (SLI) (score=5).

The way he go down and hurt himself. His shoe fell off. And the ladder go down to the ground. The boy took a picture of the girl. And he stopped taking a picture. And he was about to walk to the best thing. He went off to the zoo, and the girl went on with us, too because she went to the zoo. She sold lots of animals, and the boy sells lots of animals, too.

Figure 3.
Sample NNM retelling from a child with ASD meeting the criteria for a language impairment (ALI) (score=1).

Figure 4.
Snippet of the word-level alignment between the SLI sample retelling and the source narrative.

Story element source	Retelling
jim pepper big black liked to walk in the woods and climb the trees oak tree with branches so high that he could n't reach them took a ladder from home and carried it to the oak tree climbed up looked out over his neighborhood his foot slipped his shoe fell off and the ladder fell to the ground pepper took jim 's shoe in his mouth and ran away jim felt sad did n't his friend want to stay with him when he was in trouble pepper took the shoe to anna jim 's sister barked and barked anna put the ladder up and rescued jim	he had a friend named pepper pepper was a black dog pepper i forgot pepper got his shoe i don't know jim was a little boy pepper was his friend pepper was a black dog and pepper rescued his shoe when he brought it to anna that 's all i know and then they then anna rescued him
Retelling	**Retelling**
he had a friend named pepper pepper was a black dog pepper i forgot pepper got his shoe i don't know jim was a little boy pepper was his friend pepper was a black dog and pepper rescued his shoe when he brought it to anna that 's all i know and then they then anna rescued him	the way he go down and hurt himself his shoe fell off and the ladder go down to the ground the boy took a picture of the girl and he stopped taking a picture and he was about to walk to the best thing he went off to the zoo and the girl went on with us too because she went to the zoo she sold lots of animals and the boy sells lots of animals too
Source phrase	**Retelling phrase**
oak tree with branches too high for jim to reach pepper went to get help	he couldn't reach the branches pepper was really going for help
Word identity	**Word identity**
jim oak little	jim oak little

Figure 5.
Excerpt of parallel corpus used to train the word alignment model.

Table 1

Classification accuracy for relevance and topicality features: AUC.

Features	TD v. ALN	SLI v. ALI	Non-ASD v. ASD
Log odds	0.66	0.61	0.68
Unaligned content	0.71	0.63	0.69
Log odds + unaligned content	0.71	0.63	0.72