

# Detecting linguistic idiosyncratic interests in autism using distributional semantic models

Masoud Rouhizadeh<sup>†</sup>, Emily Prud'hommeaux<sup>°</sup>, Jan van Santen<sup>†</sup>, Richard Sproat<sup>§</sup>

<sup>†</sup>Center for Spoken Language Understanding, Oregon Health & Science University

<sup>°</sup>Center for Language Sciences, University of Rochester

<sup>§</sup> Google, Inc.

{rouhizad,vansantj}@ohsu.edu, emilypx@gmail.com, rws@xoba.com

## Abstract

Children with autism spectrum disorder often exhibit idiosyncratic patterns of behaviors and interests. In this paper, we focus on measuring the presence of idiosyncratic interests at the linguistic level in children with autism using distributional semantic models. We model the semantic space of children's narratives by calculating pairwise word overlap, and we compare the overlap found within and across diagnostic groups. We find that the words used by children with typical development tend to be used by other children with typical development, while the words used by children with autism overlap less with those used by children with typical development and even less with those used by other children with autism. These findings suggest that children with autism are veering not only away from the topic of the target narrative but also in idiosyncratic semantic directions potentially defined by their individual topics of interest.

## 1 Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by impaired communication and social behavior. One of the core deficits associated with ASD is an intense preoccupation with a restricted set of interests (American Psychiatric Association, 2000; American Psychiatric Association, 2013), which can often be observed in an individual's tendency to perseverate on specific, idiosyncratic topics of conversation. Because this symptom is explicitly mentioned among the diagnostic criteria for ASD

used in the DSM-IV and DSM-5, many diagnostic instruments (Lord et al., 2002; Rutter et al., 2003) require a qualitative assessment of this phenomenon. Instances of perseveration on a particular topic in the spontaneous spoken language of children with ASD, however, are not typically explicitly counted in a clinical setting, making comparisons with typically developing children difficult to quantify.

Expert manual analysis of conversations and narratives of individuals with ASD has shown that children and teenagers with autism include significantly more bizarre and irrelevant content in their narratives (Loveland et al., 1990; Losh and Capps, 2003) and introduce more abrupt topic changes in their conversations (Lam et al., 2012) than their typically developing peers. Automatic detection of poor topic maintenance has also been explored using techniques originally developed for information extraction (Rouhizadeh et al., 2013). There has been little work, however, in annotating the precise direction of the departure from a target topic. Thus, it is not clear whether children with ASD are instigating similar topic changes or pursuing idiosyncratic directions in their narratives and conversations consistent with their restricted interests.

In this paper, we attempt to automatically identify topic changes and idiosyncratic interests expressed in the language of children with ASD by measuring the semantic similarity of narrative retellings produced by children with and without ASD. We first use word overlap measures to calculate the semantic similarity between every possible pair of narratives. We then build three pairwise comparison matrices: one comparing pairs of typically developing (TD) children; one comparing pairs of children with ASD; and a third com-

paring pairs consisting of one child with ASD and one child with TD. We calculate the significance of the differences between the pairs in the three matrices using the Monte Carlo method to shuffle the diagnosis label of each child.

We find that TD children share the greatest word overlap with one another, while children with ASD have significantly less word overlap with TD children and even less word overlap with other ASD children. These results indicate that TD children tend to adhere to the target topic in the narrative retellings, while children with ASD often stray from the target topic. Furthermore, the fact that the word choices of an individual child with ASD seem not to resemble the word choices of other children with ASD suggests that when a child with ASD chooses to abandon the target topic, he or she does so in an idiosyncratic way. Although these results are only indirect indications of the presence of restricted interests, the work presented here highlights the potential of computational language analysis methods for improving our understanding of the social and linguistic deficits associated with the disorder.

## 2 Data

Participants in this study included 39 children with typical development (TD) and 21 children with autism spectrum disorder (ASD). ASD was diagnosed via clinical consensus according to the DSM-IV-TR criteria (American Psychiatric Association, 2000) and the established threshold scores on two diagnostic instruments: the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 2002), a semi-structured series of activities designed to allow an examiner to observe behaviors associated with autism; and the Social Communication Questionnaire (SCQ) (Rutter et al., 2003), a parental questionnaire. None of the children in this study met the criteria for a language impairment, and there were no significant between-group differences in age (mean=6.3) or full-scale IQ (mean=115.5).

The narrative retelling task analyzed here is the Narrative Memory subtest of the NEPSY (Korkman et al., 1998), a large and comprehensive battery of tasks that test neurocognitive functioning in children. The NEPSY Narrative Memory (NNM) subtest is a narrative retelling test in which the subject listens to a brief narrative about a boy and his dog and then must retell the narrative to the ex-

aminer. Under standard administration, the NNM free recall score is calculated by counting how many from a set of 17 story elements were used in a retelling. Following the free recall portion of the test is the cued recall task, in which the examiner then asks the subject to provide answers to questions about all of the story elements that were omitted in the retelling.

The NNM was administered to each participant in the study, and each participant's retelling was recorded and transcribed. The responses for the cued recall portion of the subtest were not included in this work presented here. There was no significant difference between the two diagnostic groups in the standard NNM free recall score.

## 3 Methods

We expect that two different retellings of the same source will lie in the same lexico-semantic space. As a result, they should include high percentage of overlapping words. When a pair of retellings has a low word overlap measure, it could be that one or both retellings include intrusions from unrelated topics. An alternative explanation is that the subjects recalled a non-overlapping set of story elements or simply a small set of story elements. However, since we did not find any significant difference between the TD and ASD groups in the standard narrative recall score, we infer that a low percentage of word overlap indicates a difference in topic between the two retellings.

### 3.1 Word overlap measures

In order to calculate the similarity between a pair of narratives  $i$  and  $j$ , we use type and token overlap measures based on the Jaccard similarity coefficient. Token similarity is defined as the size of intersection of the words (i.e., the actual number of tokens in common) in narratives  $i$  and  $j$  relative to the size of the union of the words in the two narratives (i.e., summing over all tokens in both narratives, the maximum number of instances of that token in either narrative). Type similarity is defined as the size of intersection of the types (i.e., unique words) in narratives  $i$  and  $j$  relative to the size of the union of the types in the two narratives. For instance, for the following set of words  $i$  and  $j$ :

$$i = \{a, b, c, d, c\}$$
$$j = \{a, c, e, c, a, a\},$$

the token intersection is equal to  $\{a, c, c\}$  and

	Group Means		
	TD.TD	TD.ASD	ASD.ASD
Type Overlap	.23	.17	.13
Token Overlap	.19	.14	.11

Table 3: Word overlap pairwise group means

the token union is {a, a, a, c, c, b, e, d}. The token overlap similarity between the two sets  $i$  and  $j$  is therefore  $3/8$ . The type intersection of  $i$  and  $j$  is equal to {a, c} and the type union is {a, c, b, e, d}, yielding a type overlap similarity of  $2/5$ .

### 3.2 Pairwise similarity matrix

We next build a similarity matrix for the type and token overlap measures, comparing every possible pair of children. Every child in the TD and ASD groups is compared to the children in his own group (TD.TD and ASD.ASD), as well as the children in the other group (TD.ASD). The pairwise similarity matrix is diagonally symmetrical, and we thus consider only the top right section of the matrix above the diagonal in our analysis.

### 3.3 Monte Carlo permutation

Since we may not have enough information to make an assumption that the pairwise similarity measures of all children are from a particular distribution, we utilize a non-parametric procedure, the Monte Carlo permutation approach, which is widely used in non-standard significance testing situations.

Given the three sub-matrices in the similarity matrix described above (TD.TD, TD.ASD, and ASD.ASD), we first calculate for each pair of sub-matrices (e.g., TD.TD vs ASD.ASD) three statistics that compare all cells in one submatrix with

the cells in other submatrices: the difference between the means, t-statistics (using the Welch Two Sample t-test), and w-statistics (using the Wilcoxon rank sum test). We label these observed values *observed-mean*, *observed-t*, and *observed-w*. We next take a large random sample with replacement from all possible permutations of the data by shuffling the diagnosis labels of the children 1000 times, and then calculate each of the three above statistics for each shuffle. Finally, we determine the number of times the observed values exceed the values generated by the 1000 shuffles.

## 4 Results

The comparison of the group means of each of the three sub-matrices described in Section 3.2 show that TD children have the greatest overlap with each other; children with ASD have less word overlap with TD children than TD children have with one another and even less word overlap with other ASD children. The group means of both type and token overlap are summarized in Table 3. In addition, examples of overlapping and non-overlapping terms between the groups are provided in Tables 1 and 2 respectively.

The level plot of the pairwise token overlap is shown in figure 1. We see that the TD.TD sub-matrix has the lightest color, indicating higher overlap, followed by TD.ASD. The ASD.ASD submatrix has the darkest color, indicating low word overlap.

In the next step, we determine the significance of the group mean differences. As described in Section 3.3, using the Monte Carlo permutation to test the significance of the following comparisons: TD.TD vs ASD.ASD, TD.TD vs TD.ASD, and TD.ASD vs ASD.ASD. The results of these signif-

Group	Top 10 overlapping words
TD.TD	<i>shoe, tree, climb, ladder, fall, Pepper, Jim, dog, sister, branch</i>
TD.ASD	<i>shoe, tree, Jim, climb, dog, ladder, Pepper, fall, branch, sister</i>
ASD.ASD	<i>shoe, tree, Jim, dog, climb, Pepper, ladder, branch, boy, run</i>

Table 1: Top 10 overlapping words between the groups

Group	Examples of non-overlapping words
TD.TD	<i>coconut, couch, jew, lie, picture, spike, stuff, t-rex, tight, watch</i>
TD.ASD	<i>arm, bottom, cousin, doctor, eat, fruit, giant, meat, push, sense</i>
ASD.ASD	<i>bite, bridge, crunch, donut, gadget, lizard, microphone, sell, table, vision</i>

Table 2: Examples of non-overlapping words between the groups

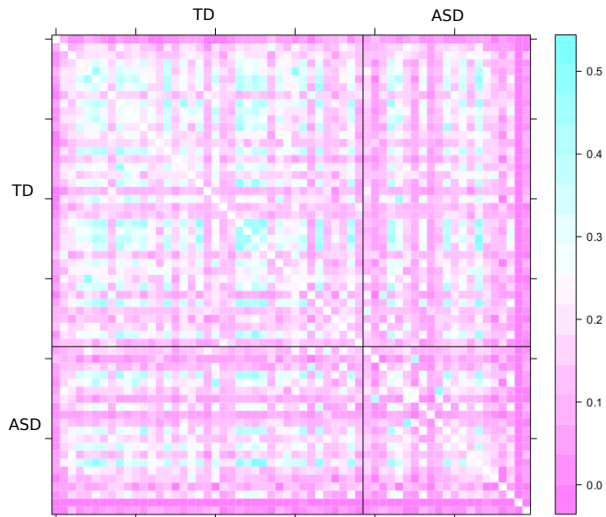


Figure 1: Level plot of the pairwise token overlap (lighter colors indicate higher overlap)

ificance tests are summarized in table 4, and in all cases the differences are significant at  $p < 0.05$ .

## 5 Conclusions and future work

The methods presented for comparing the lexical choices made by children with and without ASD while generating a narrative retelling demonstrate the utility of language analysis for revealing diagnostically interesting information. The low rates of word overlap between retellings produced by children with ASD and those produced by typically developing children suggest that the children with ASD are having difficulty maintaining the target topic. Furthermore, the low overlap between pairs of children with ASD suggests that children with ASD are not straying from the topic in similar ways but are instead exploring topics that are of idiosyncratic interest.

These findings can be potentially used for diagnostic purposes in combinations of other

applications of speech and language processing for automated narrative retelling assessment (Lehr et al., 2013), detection of off-topic words (Rouhizadeh et al., 2013), and pragmatic deficits (Prud’hommeaux and Rouhizadeh, 2012). From a clinical standpoint, diagnostic measures utilizing these methods for automated evaluation of disordered language could be very useful in diagnosis and planning interventions.

One major focus of our future work will be to manually annotate the narrative retellings used in this study to determine the frequency of topic departures and the nature of these departures. Given the vocabulary differences seen here, we expect to find not only that children with ASD are abandoning the topic of the source narrative more frequently than children with typical development but also that the topics they choose to pursue are related to their own individual specific interests.

A second area we hope to explore is the use of external resources, such as WordNet, to expand the set of terms used to calculate word overlap. It is perfectly reasonable to expect that people will use synonyms and paraphrases in their narrative retellings. It is therefore possible that children with autism are discussing the appropriate topic but choosing unusual words within that topic space in their retellings, which could be consistent with the type of atypical language often observed in children with ASD. By considering semantic overlap rather than simple word overlap, we may be able to distinguish instances of atypical language from true examples of poor topic maintenance.

Third, we are also interested in applying the analysis described above to a set of retellings from seniors with and without mild cognitive impairment, a frequent precursor to dementia. Like children with ASD, seniors with dementia are also more likely to include irrelevant information in

overlap	statistic	p-values		
		TD.TD vs ASD.ASD	TD.TD vs TD.ASD	TD.ASD vs ASD.ASD
Type Overlap	Means	.004	.042	.008
	t.test	.009	.012	.008
	Wilcoxon test	.004	.002	.002
Token Overlap	Means	.012	.034	.028
	t.test	.014	.022	.022
	Wilcoxon test	.012	.002	.002

Table 4: Monte Carlo significance test results

their narrative retellings. These intrusions, however, are often informed by real-world knowledge, and thus may not result in a decrease in measures of word overlap with narratives produced by unimpaired individuals.

Finally, we plan to apply our methods to the output of an automatic speech recognition (ASR) system rather than manual transcripts. Although the ASR output is likely to contain many errors, the fact that our methods focus on content words may make them robust to the sorts of function word recognition errors typically produced by ASR systems.

### Acknowledgments

This work was supported in part by NSF grant #BCS-0826654, and NIH NIDCD grants #R01-DC007129 and #1R01DC012033-01. 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.

### References

- American Psychiatric Association. 2000. *DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Publishing, Washington, DC.
- American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders (5th ed.)*. American Psychiatric Publishing, Washington, DC.
- Marit Korkman, Ursula Kirk, and Sally Kemp. 1998. *NEPSY: A developmental neuropsychological assessment*. The Psychological Corporation, San Antonio.
- Yan Grace Lam, Siu Sze, and Susanna Yeung. 2012. Towards a convergent account of pragmatic language deficits in children with high-functioning autism: Depicting the phenotype using the pragmatic rating scale. *Research in Autism Spectrum Disorders*, 6(2):792–797.
- Maider Lehr, Izhak Shafran, Emily Prud’hommeaux, and Brian Roark. 2013. Discriminative joint modeling of lexical variation and acoustic confusion for automated narrative retelling assessment. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
- Catherine Lord, Michael Rutter, Pamela DiLavore, and Susan Risi. 2002. *Autism Diagnostic Observation Schedule (ADOS)*. Western Psychological Services, Los Angeles.
- Molly Losh and Lisa Capps. 2003. Narrative ability in high-functioning children with autism or asperger’s syndrome. *Journal of Autism and Developmental Disorders*, 33(3):239–251.
- Katherine Loveland, Robin McEvoy, and Belgin Tunali. 1990. Narrative story telling in autism and down’s syndrome. *British Journal of Developmental Psychology*, 8(1):9–23.
- Emily Prud’hommeaux and Masoud Rouhizadeh. 2012. Automatic detection of pragmatic deficits in children with autism. In *Proceedings of the 3rd Workshop on Child, Computer and Interaction (WOCCI)*.
- Masoud Rouhizadeh, Emily Prud’hommeaux, Brian Roark, and Jan van Santen. 2013. Distributional semantic models for the evaluation of disordered language. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
- Michael Rutter, Anthony Bailey, and Catherine Lord. 2003. *Social Communication Questionnaire (SCQ)*. Western Psychological Services, Los Angeles.