2018

Detecting Language Impairments in Autism: A Computational Analysis of Semi-structured Conversations with Vector Semantics

Adam Goodkind
Northwestern University, a.goodkind@u.northwestern.edu

Michelle Lee
Northwestern University, michelleannemarie2017@u.northwestern.edu

Gary E. Martin
St. John's University, marting@stjohns.edu

Molly Losh
Northwestern University, m-losh@northwestern.edu

Klinton Bicknell
Northwestern University, kbicknell@northwestern.edu

Detecting language impairments in autism: A computational analysis of semi-structured conversations with vector semantics

Adam Goodkind[1], **Michelle Lee**[2,3], **Gary E. Martin**[4], **Molly Losh**[3], and **Klinton Bicknell**[1]
[1]Dept. of Linguistics, Northwestern Univ., Evanston, IL 60208
[2]Clinical Psychology, Feinberg School of Medicine, Northwestern Univ., Chicago, IL 60611
[3]Dept. of Communication Sciences and Disorders, Northwestern Univ., Evanston, IL 60208
[4]Dept. of Communication Sciences and Disorders, St. John's Univ., Staten Island, NY 10301
{a.goodkind, michelleannemarie2017}@u.northwestern.edu,
marting@stjohns.edu, {m-losh, kbicknell}@northwestern.edu

Abstract

Many of the most significant impairments faced by individuals with autism spectrum disorder (ASD) relate to pragmatic (i.e. social) language. There is also evidence that pragmatic language differences may map to ASD-related genes. Therefore, quantifying the social-linguistic features of ASD has the potential to both improve clinical treatment and help identify gene-behavior relationships in ASD. Here, we apply vector semantics to transcripts of semi-structured interactions with children with both idiopathic and syndromic ASD. We find that children with ASD are less semantically similar to a gold standard derived from typically developing participants, and are more semantically variable. We show that this semantic similarity measure is affected by transcript word length, but that these group differences persist after removing length differences via subsampling. These findings suggest that linguistic signatures of ASD pervade child speech broadly, and can be automatically detected even in less structured interactions.

1 Introduction

From its earliest descriptions (Kanner, 1943), autism spectrum disorder (ASD) has been associated with language impairment, and pragmatic language impairment in particular. Problems with pragmatic language are a key component of current diagnostic criteria for ASD, and both atypical and idiosyncratic language are noted as features of ASD un-der current standards used in both the DSM-IV and DSM-5 (American Psychiatric Association, 2000, 2013). However, current methods for assessing pragmatic language impairment are often subjective, can be very time intensive, and distal from underlying mechanisms. Computational models of language production in ASD thus have the potential to improve diagnostic assessments, contribute to research into the basis of language impairment in ASD, and may also show strong utility in clinical treatment as objective and quantitative measures of response to intervention.

Additionally, evidence that more subtle language differences are evident at elevated rates among relatives of individuals with ASD points towards pragmatic language as a genetically meaningful domain in ASD, with potential for informing molecular genetic studies, which examine more specific ties to component phenotypes in ASD that may segregate independently and relate to distinct genetic underpinnings (Losh, Sullivan, Trembath, & Piven, 2008). The development of computational tools for quantifying language impairment in ASD, such as the present study, may therefore contribute to future studies of ASD genetics as well. This can be accomplished by applying the present study's methods to large-scale datasets, which are appropriate for broad genetic studies. In addition, the methods presented below provide a continuous measure of pragmatic impairment, which can be more readily compared against genetic data.

The pragmatic language impairments in ASD are

evident in a range of linguistic features. For instance, limited frequency and diversity of complex syntax has been shown to significantly impact narrative and conversational quality in ASD (Losh & Capps, 2003; Prud'hommeaux, Roark, Black, & Van Santen, 2011). Individuals with ASD produce more non-contingent discourse in narrative and conversation (Capps, Kehres, & Sigman, 1998; Losh & Capps, 2003, 2006). In a similar vein, people with ASD tend to fixate on a single topic, even though a conversation may have moved away from that single topic in another direction (Nazeer & Ghaziuddin, 2012). Additionally, both inappropriate semantic and pragmatic language has been demonstrated (Tager-Flusberg & Sullivan, 1995). From a semantic standpoint, one way these differences may manifest is when children with ASD use very uncommon words in a context in which a common word suffices. Children with ASD can also fail to make common pragmatic inferences, such as understanding the semantics of a question like *Can you close the door?* but failing to understand its pragmatics, and so responding by saying *Yes, I can.*

Given this evidence that individuals with ASD exhibit such pragmatic impairments, prior work has used computational models to distinguish individuals with ASD from typically developing individuals, using distributional semantic word models (Rouhizadeh, Prud'hommeaux, Roark, & Van Santen, 2013; Losh & Gordon, 2014). For example, both Losh and Gordon (2014) and Lee et al. (2017) used Latent Semantic Analysis (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990) with transcripts from picturebook narratives, a narrative recall task, and a less structured narrative elicitation task. Both studies showed that narratives from individuals with ASD diverged significantly in vector semantic space from a gold standard (either the original narrative, or a narrative derived from the TD group of participants) compared to (non-gold standard) typically developing controls.

Narrative recall and picturebook description tasks afforded clear gold standards for comparisons, with a very clear objective semantics to communicate the original narrative. Thus, they were optimal for a computational linguistic approach. However, it is less clear whether such an approach can generalize to other, more variable and naturalistic language

contexts, in which there may be no objective gold standard. Arguably, though, these naturalistic studies are more ecologically valid, and also constitute the discourse context posing the most serious challenges to individuals with ASD.

1.1 Goals

The primary goal of the present work is to investigate whether this computational approach could be applied in a more open-ended conversational setting, in which there is no objective gold standard. The primary contributions of this work are as follows:

- We show that language from individuals with ASD can be distinguished from that of typically developing individuals, by applying vector semantic models, *even on semi-structured conversational data.*

- We demonstrate that semantic similarity metrics are affected by transcript length, raising the question of whether such metrics can yield valid conclusions with very small language samples, such as often occur with children or lower functioning populations.

- We present a method for adapting semantic similarity analyses to accommodate possibly-small language samples from younger or lower functioning populations who have more limited language abilities.

To do this, we analyzed semi-structured conversational interactions, consisting of relatively free ranging conversation in a number of somewhat consistent situations. We construct an approximate gold standard of comparison from the transcripts of a few individuals with typical development who had similar and typical language and cognitive abilities.

We focus on conversational interactions because this language context is among the most challenging for individuals with ASD, since lack of structure and high interpersonal demands pose serious barriers to effective communication (as opposed to picturebook narratives, for example) (Losh & Capps, 2003). Furthermore, prior studies of computational linguistic approaches to characterizing discourse in ASD have focused on more structured contexts. This study is the first to apply this technique to conversational interaction in ASD. If this technique can successfully

13

differentiate neurotypical individuals and individuals with ASD, based on semi-structured conversation, it would suggest that the semantic differences in the language of individuals with ASD are quite widespread, and detectable across a range of everyday tasks. Finally, because the semi-structured interactions we analyze come from a standard ASD diagnosis task (the ADOS, see below), such a computational model has the potential to help with diagnosis.

The remainder of this paper is structured as follows. The following section describes our transcript dataset. Following that, Section 3 describes how we use word embeddings to quantify distances to gold standards. Section 4 describes our first experiment applying this method to conversation data. Section 5 points out that this similarity metric is confounded by transcript length and presents a method to remove this confound. Section 6 concludes.

2 Interaction session transcripts

2.1 Participants

We selected 109 participants, in three groups: (1) (younger) typically developing children used as a control due to comparable cognitive ability to the clinical groups (TD); (2) school aged children with idiopathic ASD, unrelated to any other known genetic disorders (ASD); and (3) children with ASD comorbid with fragile X syndrome (FXS-ASD). For all three groups, children were selected based upon the nonverbal mental age from the Leiter International Performance Scale (Wechsler, 2008). For typically developing children, mental age should on average match chronological age. However, for children with developmental impairments, mental age is often lower than chronological age.

Fragile X syndrome (FXS) is the most common heritable intellectual disability, and has common comorbidity with ASD (Rogers, Wehner, & Hagerman, 2001; Kaufmann et al., 2004; Martin et al., 2017; *inter alia*). Like ASD, fragile X syndrome often shows pragmatic deficits as well. Evidence also exists that language impairment within fragile X syndrome affects males with FXS more than it affects females with FXS (Abbeduto, McDuffie, & Thurman, 2014). For this reason, all selected participants were male. This also eliminates sex as a possible confound.

All participants were selected based on a men-tal age of approximately 5;0. All participants had a mean length of utterance (MLU) of at least three words per utterance and were L1 English speakers. Participants were drawn from a larger longitudinal study reported in Martin et al. (2017). Additionally, individuals with idiopathic ASD were required to have a previous clinical diagnosis, confirmed by administration of the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 2000) and/or the Autism Diagnostic Interview - Revised (Lord, Rutter, & Le Couteur, 1994). Individuals with FXS-ASD were confirmed based only on the ADOS. The average chronological and mental age for each group are provided in Table 1.

2.2 Procedure

Language samples were derived from the ADOS and/or ADOS-2, gold standard diagnostic tools for ASD. The ADOS includes several structured activities as well as opportunities for naturalistic interaction, in order to probe for social-communication skills and the presence of restricted and repetitive behaviors. Play-based activities included the opportunity to play with action figures and other toys. Non-play based activities included conversation between tasks, describing a picture, or telling a story from a book.

2.3 Transcription

Subsections of entire language samples were transcribed from high quality audio recordings by trained transcribers. The transcripts were based on a subset of the full assessment: specifically, 55 intelligible play based turns and 55 non-play based turns were transcribed (or fewer in the rare case that there were not 55 intelligible turns).

2.4 Processing

All child utterances were extracted from the transcripts, including filled pauses and stop words. Although stop word removal is common practice in distributional semantics and NLP (Levy, Goldberg, & Dagan, 2015), this class of words can be psychologically informative (Chung & Pennebaker, 2007). This also seems especially relevant to ASD, where incorrect pronoun usage is common, e.g. using the second-person *you* when referring to oneself, instead of the correct first-person *I* (Naigles et al., 2016).

Diagnostic Group	n	Chronological Age (SD)	Mental Age Equivalent (SD)
Typically developing (TD)	22	4.7 (1.1)	5.1 (1.2)
Autism spectrum disorder (ASD)	39	8.7 (2.9)	6.9 (3.4)
Fragile X syndrome + ASD (FXS-ASD)	48	10.6 (2.6)	5.0 (0.6)

Table 1: Participant chronological age and mental age equivalent for each diagnostic group.

2.5 Gold standard transcripts

Two of the transcripts from children with typical development were designated gold standard transcripts. This designation was performed by two researchers who were both familiar with the tasks in the interactions. Gold standards were selected based on detailed clinical-behavioral ratings. We selected TD participants who, based on this coding, demonstrated minimal pragmatic language deficits and highly-rated core features of conversation, such as contingency, reciprocity, and initiation. For the purposes of analyses conducted below, these two transcripts were excluded from the TD group, so as not to bias the results.

3 Word embeddings

A number of previous studies have used word embeddings (vector semantics) to study language transcripts of people with autism (Rouhizadeh et al., 2013; Rouhizadeh, 2015). A vector semantic model specifies an embedding, or mapping, from each word in the vocabulary to a point in a continuous vector space. A document in such models typically consists of an unordered collection of words. A vector semantic representation of a document can be obtained by combining the embeddings of the words it contains in some way, such as summing.

For the present study, each word in the transcript was converted to a vector using the `word2vec` model, via the pretrained Google News embeddings (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), which are 400-dimensional. A vector semantic representation of each document was created by summing the vectors for each of its words, and then normalizing to have unit length.

The gold standard vector was calculated as the mean of the two gold standard transcript vectors, and used as the basis of comparison for semantic similarity.[1] The transcripts identified to be the gold standard

[1]Other bases for comparison were also considered, includ-

were excluded from the TD group for all analyses.

Semantic distance of a transcript vector to the gold standard was measured as the cosine distance between them. That is, for a given transcript vector $\vec{v_i}$ and the gold standard vector $\vec{g}$, the distance was then calculated as

$$d(\vec{v_i}, \vec{g}) = \frac{\vec{v_i} \cdot \vec{g}}{\|\vec{v_i}\|_2 \|\vec{g}\|_2} \quad (1)$$

Because the transcript vectors are all normalized to have unit length, this reduces to a simple dot product.

A lower cosine distance means that the vectors being compared have more similar dimensions. Given this, we then defined the semantic similarity of a vector to the gold standard as one minus the cosine distance $(1 - d(\vec{v}, \vec{g}))$, so that a lower distance resulted in a higher semantic similarity score.

The code for converting transcripts to vectors and computing similarities is freely available on an open-source repository[2]. Although our transcripts themselves cannot be shared because of privacy concerns, we used a standard format for transcription, making our tools readily usable by other investigators.

4 Experiment 1

We performed three sets of analyses to compare TD individuals to the two populations with ASD.

4.1 Similarity to gold standard

For each transcript, we calculated the cosine distance between its vector embedding and the gold standard, yielding a single similarity score for each transcript. Figure 1 illustrates the mean semantic

ing using the mean vector from all transcripts as well as the mean vector of all of the typically developing transcripts. Future studies will compare and contrast the utility of selecting different bases for comparison.

[2]https://github.com/langcomp/vectoraut
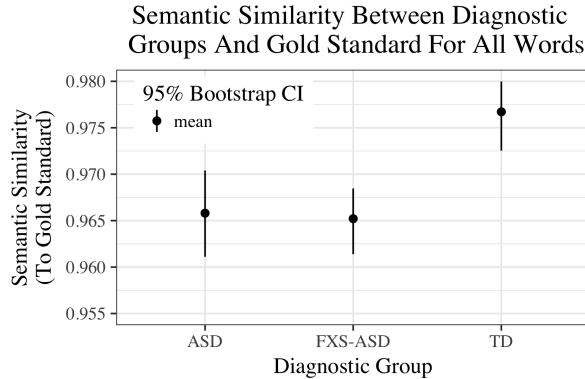
Figure 1: Semantic similarity for each diagnostic group as compared to the semantic content of Gold Standard transcripts

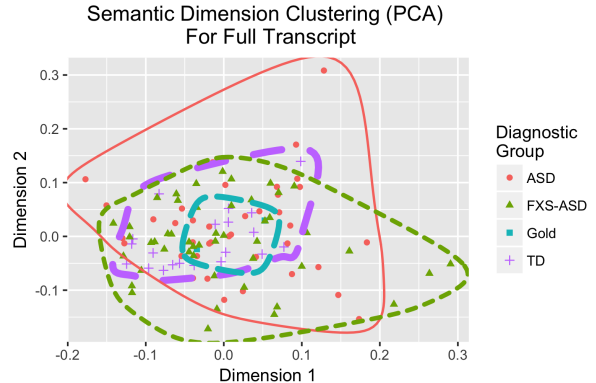Figure 2: Most informative dimensions using full transcripts, as selected by PCA, for different diagnostic groups. Clusters for children with autism are more diffuse.

Comparison	All Words	Word Sampling
TD vs ASD	$p < 0.001$	$p < 0.05$
TD vs FXS-ASD	$p < 0.00001$	$p < 0.001$

Table 2: Significance levels for differences in semantic similarity between diagnostic groups and a gold standard. Both full transcripts as well as random word sampling from transcripts are reported.

Diagnostic Group	Manhattan Distance (mean)
TD	3.44
FXS-ASD	4.42
ASD	4.49

Table 3: Manhattan distance between all pairs of word vectors within each diagnostic group, for full transcripts.

similarity for each group, as well as the 95% bootstrapped confidence intervals. We then ran nonparametric Wilcoxon tests comparing the semantic similarity scores across groups, specifically comparing the TD group to the ASD group, and the TD group to FXS-ASD, the results of which are shown in Table 2, middle column. The results show reliable differences between TD and each of the two other groups, with the TD group being more semantically similar to the gold standard transcripts than the groups with ASD. The two groups with ASD appear very similar to each other.

4.2 Visualizing semantic space

To try to visualize the semantic space these transcripts are embedded in, we used Principal Component Analysis (Jolliffe, 2011) to reduce the 400-dimensional vectors to two dimensions. The results are visualized in Figure 2, where each transcript is identified by its group, and the two transcripts from which the gold standard was constructed have also been added. As can be seen, the ASD and FXS-ASD groups are much more dispersed than the typically developing group, which is relatively tightly clus-

tered around the gold standard. This suggests a possible reason why the two groups with ASD were less similar to the gold standard than was the TD group: because they are more semantically variable.

4.3 Within-group variability

Finally, to test this hypothesis that variability within groups is higher for groups with ASD as opposed to the TD group, we measured the Manhattan distance between all pairs of transcript vectors within each group, using the full 400-dimensional transcript vectors. This distance is an indication of how far apart semantically two transcript are. We elected to use Manhattan distance rather than Euclidean distance because of the former's robustness in high-dimensional space (Aggarwal, Hinneburg, & Keim, 2001). The average distance between each vector pair in each group is reported in Table 3. The results show that the TD group is much more homogeneous, while the two groups with ASD exhibit larger variability. Interestingly, the two groups with ASD again appear very similar to each other.

16

5 Experiment 2

The results of Expt. 1 showed strong evidence that individuals in the ASD and FXS-ASD groups were on average less semantically similar to a gold standard than TD individuals were, and that individuals from both groups with ASD were more semantically variable. However, it is also a feature of this dataset that there is a systematic relationship between the word count of a particular individual's transcript and their semantic similarity. This is shown in Figure 3, where transcripts with a higher word count on average have higher similarity scores to the gold standard, within each of the three groups (although there is some suggestion for the ASD group that this effect may reverse for especially high word counts). There are multiple, non-exclusive hypotheses for where this relationship arises. It may be an objective feature of language production that individuals with more language impairment talk less on average. Alternatively, it may be that the semantic similarity metric becomes noisier and thus lower with smaller language samples.

There are also systematic differences between groups in transcript length, visualized in Figure 4. On the hypothesis that differences in transcript length are an artifact of the measure, not necessarily related to language proficiency, this raises the possibility that the reason the TD group had more semantic similarity to the gold standard on average was that it had longer transcripts on average. To rule out this possibility, and gain some insight into the relationship between semantic similarity and transcript length, we developed a simple method to remove the variance in transcript length.

5.1 Word sampling

To remove this variance in transcript length, we performed a random sampling algorithm. Specifically, we selected a target transcript length of 300 words, which was lower than the majority of transcript lengths in every diagnostic group. Then, we sampled, without replacement, 300 words from every transcript. For the 2 transcripts that fell below the 300-word threshold, we selected the entire transcript. With this new set of transcripts of a uniform length, we performed the same analyses as in Expt. 1. To leave in tact full information about the
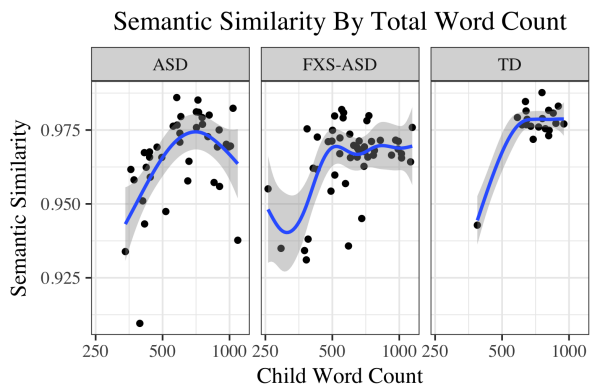
Figure 3: Child word counts versus semantic similarity to gold standard. As children produce more words, similarity to the gold standard also increases. Smoothing lines were calculated using a generalized additive model (GAM).
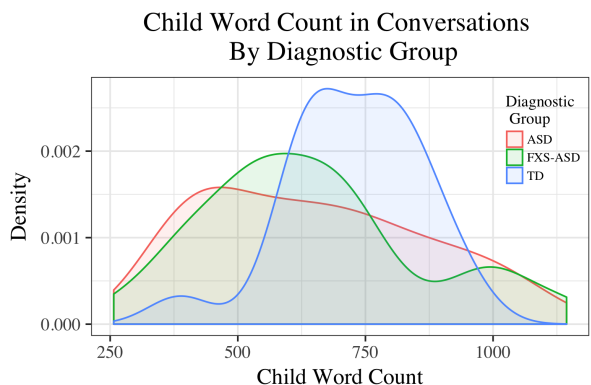

Figure 4: Child words counts for different diagnostic groups. Different diagnostic groups have markedly divergent distributions.

gold standard, we left it unchanged from Expt. 1.

5.2 Similarity to gold standard

The similarity of these uniform-length transcripts to the gold standard are shown in Figure 5. As can be seen, even without transcript length differences, semantic similarity is still lower for the ASD and FXS-ASD groups than for the TD group. The results of a Wilcoxon test are given in the rightmost column of Table 2, showing that these differences are still significant, suggesting that the group differences obtained in Expt. 1 were not merely an artifact of shorter transcript lengths.

Comparing the values in Figure 5 to those in Figure 1 from Expt. 1, however, we can see that the similarity values here are a bit lower, especially those
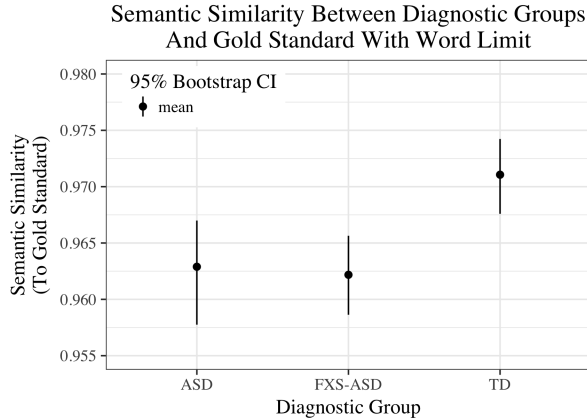
Figure 5: Semantic similarity for each diagnostic group as compared to the semantic content of Gold Standard transcripts, using a 300 word random sample
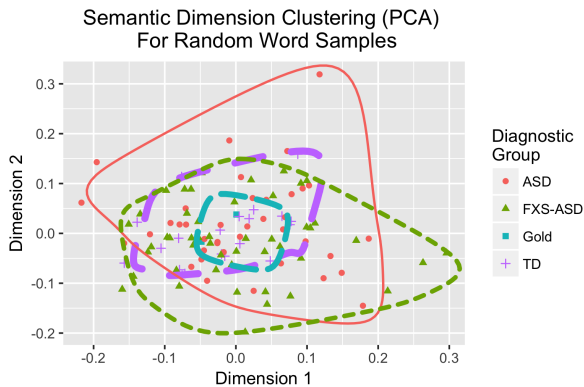
Figure 6: Most informative dimensions using random word sampling, as selected by PCA, for different diagnostic groups. Clusters for children with autism are more diffuse.

from the TD group. That suggests that these semantic similarity metrics are indeed somewhat biased lower by the smaller samples. For work such as this applying vector semantic similarity metrics to transcripts from younger and/or lower functioning individuals, analyses like this may be a useful tool.

5.3 Visualizing semantic space

As in Expt. 1, we reduced all of the transcript vectors to two dimensions using PCA. Even when randomly selecting words, the same spatial relationships seen in Expt. 1 still hold, with greater dispersion for the ASD and FXS-ASD groups, and more concentration for the TD and gold standard groups.

Diagnostic Group	Manhattan Distance (mean)
TD	3.76
FXS-ASD	4.54
ASD	4.67

Table 4: Manhattan distance between all pairs of word vectors within each diagnostic group, for random word samples.

5.4 Within-group variability

We aimed to understand whether the transcript vectors from randomly sampled words still had a very short distance between them, or whether the random word sampling obfuscated the similarities seen in Expt. 1. As seen in Table 4, the uniform-length transcripts still had the same qualitative distance relationships, with the vectors in the TD group being closer together than those within either of the two groups with ASD.

Comparing Table 4 to its Expt. 1 equivalent, Table 3, we see that the mean distance between TD group transcripts is about 0.3 units higher here, whereas the ASD and FXS-ASD group distances are only 0.1–0.2 units higher. This result suggests again, however, that there is some evidence that the variability within groups seen in Expt. 1 was biased to be somewhat higher for the groups with shorter transcript, and that this method can correct for that bias.

6   Discussion

In this study, we used vector semantics to show that semi-structured conversations produced by individuals with ASD were semantically further from a gold standard conversational sample by children with typical development, and that there was more variability within the groups with ASD than within the typically developing group. We also presented evidence that these semantic similarity and distance measures were moderately biased by transcript length, with low transcript lengths yielding larger semantic distances from a gold standard as well as yielding more within-group variability. Finally, we showed that this bias could not explain the differences in semantic similarity and distance across groups.

Many previous studies applying vector semantics to the language of autism relied upon narratives (Losh & Capps, 2003; Prud'hommeaux et al.,

2011; Rouhizadeh, Prud'hommeaux, van Santen, & Sproat, 2014; Losh & Gordon, 2014; Lee et al., 2017; *inter alia*), for which there is an objective gold standard of semantic evaluation, i.e., the original narrative. The present study demonstrates that such methods can be extended to the more naturalistic context of semi-structured conversation. This is important because narrative retellings are expected to use a more constrained vocabulary; thus, unexpected words are even more surprising when the vocabulary is expected to be more minimal. On the other hand, it is easy to assume that a naturalistic conversation would have a large variety of vocabulary, making unexpected words less surprising and less telling, since a conversation can have a more wide array of topic areas to discuss. Nonetheless, even when participating in a more unconstrained conversation, our study still picked up semantic differences for the ASD groups, despite the more freewheeling dialogue.

The results of this study of semi-structured conversation parallel those of Losh and Gordon (2014) and Lee et al. (2017) on narratives. All studies showed that individuals with ASD produced language semantically further from a gold standard of typical development. All studies also showed that language produced by typically developing children clustered together more closely in semantic space, whereas that from children with ASD was more variable and diffuse in semantic space (cf. Rouhizadeh et al., 2014). These findings suggest that greater semantic variability may be a general property of the language of ASD not confined to narrative retellings.

To our knowledge, this is the first study to present a method for analyzing semantic similarities in the presence of strong differences in word length of transcripts across individuals and groups, which may be a common occurrence in work analyzing populations that differ in conversation length. Since MLU has been commonly found to be strongly correlated to (chronological) age (Miller & Chapman, 1981), and would lead to more words in a conversation, this may be a common issue for younger populations, as well as those that are lower functioning.

Interestingly, even when reducing the lengths of the transcripts by a substantial amount (often more than 50%) via random sampling, the transcripts of individuals with ASD were still significantly se-

mantically further from the gold standard than were those of typically developing individuals. This suggests that the language of ASD may be pervasive in speech and detectable from even shorter samples.

The question remains, though *What exactly are we detecting through different mean semantic similarities for each group?* Are we picking up different *styles* of language use, or are we picking up the same type of language, just describing different topics? Either of those differences could affect semantic similarity in a similar fashion.

To test this, we looked at the transcripts' connections to type-token ratio, which measures lexical diversity. We ran both Pearson and Spearman correlations between semantic similarity and type-token ratio (TTR). TTR is the ratio of the number of word types to the number of word tokens If lexical diversity was significantly different between groups in the same way that semantic similarity was different between groups, this would be a strong indicator that the difference in group semantic similarity was due to the use of different types of language, or at least different levels of lexical diversity. However, this correlation was very low and neither individual groups nor overall correlations approached statistical significance.

This results suggests that linguistic style (at least as indexed by the type-token ratio) is not a driving factor. Instead, we are left with the most glaring differences being due to using the same type of language while discussing different topics, and thereby using different words to describe different topics.

Perhaps this underscores the idiosyncratic nature of utterances from both the ASD and FXS-ASD groups. In other words, even though conversational language is more unconstrained than narrative retellings, Figures 2 and 6 still seem to illustrate that TD participants use a more confined vocabulary. This may seem counter-intuitive, as a more natural conversation can be assumed to be diverse, with a less constrained vocabulary. This seems to point to the robustness of word vectors, since they can still capture the diversity of the uncommon words often selected by children with ASD and FXS-ASD, while downplaying the natural diversity that is expected from conversations with typically-developing children.

6.1 Conclusion

The primary contribution of this study is to show that vector semantics distinguishes language of individuals with ASD from language of individuals who are typically developing, even when that language was produced in a semi-structured conversational setting with no objective semantic standard – the language context where individuals with ASD exhibit most severe impairments. It also presented evidence that such semantic metrics can be applied to populations who yield smaller word counts (such as younger, lower functioning children with intellectual disability) – despite short transcripts biasing the semantic metrics – and presented a simple method to help quantify and control this bias that can be implemented across age and ability levels.

This work represents a step toward developing a metric of language impairment in ASD that is empirically quantifiable, objective, and automatically generated, which has the potential to improve clinical assessments, offer objective quantitative indices of language impairment that could be used to stratify groups in biological studies, and possibly serve as sensitive measures of response to clinical interventions. The results here were based on conversational data from the ADOS assessment, a standard assessment for ASD including semi-structured conversational samples. This is a more generalized form of dialogue than the narrative retellings used in many previous studies. This finding, then, opens the door to investigate further diverse discourse settings where semantic similarity tests might be effectively implemented. For instance, if a classroom or family dinnertime setting could be used as a reliable source, then a child would not need to be removed from their typical daily activities in order to perform testing. Such assessments would also capture pragmatic impairments in more naturalistic settings, affording more generalizable findings.

Future investigations will investigate the seeming disparity between extremely uncommon word choice and word choice that is *too* identical to that employed by a conversation partner. Perhaps these phenomena cancel each other out when an entire conversation is considered, leaving both of these idiosyncrasies diluted by averaging. We plan on investigating this alongside semantic similarity as a means to predict, not merely quantify, ASD.

We also have data collected from the same participants at subsequent time points, which will allow us to test the rate of change for individuals. For example, if language production did not advance at an expected rate over a period of time, this could also be a sign of developmental deficits.

Finally, by quantifying the language of ASD in a continuous way, this method has possible applications to genetic studies of ASD, where quantitative as opposed to categorical phenotypic measures afford greater power to detect molecular genetic associations for complex traits and disorders such as ASD. For example, it could be used to quantify the extent to which family members of individuals with ASD, who are themselves typically developing, nevertheless exhibit values of this continuous measure that more closely resemble the language of ASD. More generally, these findings suggest that the linguistic signatures of ASD pervade child speech, and may be automatically detectable under wider conditions than previously demonstrated.

References

Abbeduto, L., McDuffie, A., & Thurman, A. J. (2014). The fragile x syndrome–autism comorbidity: what do we really know? *Frontiers in genetics*, *5*.

Aggarwal, C. C., Hinneburg, A., & Keim, D. A. (2001). On the surprising behavior of distance metrics in high dimensional spaces. In *Icdt* (Vol. 1, pp. 420–434).

American Psychiatric Association. (2000). Task force on DSM-IV. diagnostic and statistical manual of mental disorders: DSM-IV-TR. *Washington, DC: American Psychiatric Association*, *4*.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders: DSM-5. *Washington, DC: American Psychiatric Association*.

Capps, L., Kehres, J., & Sigman, M. (1998). Conversational abilities among children with autism and children with developmental delays. *Autism*, *2*(4), 325–344.

Chung, C., & Pennebaker, J. W. (2007). The psychological functions of function words. *Social communication*, 343–359.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, *41*(6), 391.

Jolliffe, I. T. (2011). Principal component analysis. In *International encyclopedia of statistical science.*

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous child*, *2*(3), 217–250.

Kaufmann, W. E., Cortell, R., Kau, A. S., Bukelis, I., Tierney, E., Gray, R. M., ... Stanard, P. (2004). Autism spectrum disorder in fragile x syndrome: communication, social interaction, and specific behaviors. *American Journal of Medical Genetics Part A*, *129*(3), 225–234.

Lee, M., Martin, G. E., Hogan, A., Hano, D., Gordon, P. C., & Losh, M. (2017). Whats the story? A computational analysis of narrative competence in autism. *Autism*, 1362361316677957.

Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving distributional similarity with lessons learned from word embeddings. *TACL*, *3*, 211-225.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., ... Rutter, M. (2000). 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*, *30*(3), 205–223.

Lord, C., Rutter, M., & Le Couteur, A. (1994). 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*, *24*(5), 659–685.

Losh, M., & Capps, L. (2003). Narrative ability in high-functioning children with autism or asperger's syndrome. *Journal of autism and developmental disorders*, *33*(3), 239–251.

Losh, M., & Capps, L. (2006). Understanding of emotional experience in autism: insights from the personal accounts of high-functioning children with autism. *Developmental psychology*, *42*(5), 809.

Losh, M., & Gordon, P. C. (2014). Quantifying narrative ability in autism spectrum disorder: A computational linguistic analysis of narrative coherence. *Journal of autism and developmental disorders*, *44*(12), 3016–3025.

Losh, M., Sullivan, P. F., Trembath, D., & Piven, J. (2008). Current developments in the genetics of autism: from phenome to genome. *Journal of Neuropathology & Experimental Neurology*, *67*(9), 829–837.

Martin, G. E., Barstein, J., Hornickel, J., Matherly, S., Durante, G., & Losh, M. (2017). Signaling of noncomprehension in communication breakdowns in fragile x syndrome, down syndrome, and autism spectrum disorder. *Journal of Communication Disorders*, *65*, 22–34.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Miller, J. F., & Chapman, R. S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech, Language, and Hearing Research*, *24*(2), 154–161.

Naigles, L. R., Cheng, M., Rattanasone, N. X., Tek, S., Khetrapal, N., Fein, D., & Demuth, K. (2016). "You're telling me!" The prevalence and predictors of pronoun reversals in children with autism spectrum disorders and typical development. *Research in autism spectrum disorders*, *27*, 11–20.

Nazeer, A., & Ghaziuddin, M. (2012). Autism spectrum disorders: clinical features and diagnosis. *Pediatric Clinics of North America*, *59*(1), 19–25.

Prud'hommeaux, E. T., Roark, B., Black, L. M., & Van Santen, J. (2011). Classification of atypical language in autism. In *Proceedings of the 2nd workshop on cognitive modeling and*

computational linguistics (pp. 88–96).

Rogers, S. J., Wehner, E. A., & Hagerman, R. (2001). The behavioral phenotype in fragile x: symptoms of autism in very young children with fragile x syndrome, idiopathic autism, and other developmental disorders. *Journal of developmental & behavioral pediatrics*, *22*(6), 409–417.

Rouhizadeh, M. (2015). Computational analysis of language use in autism. *Oregon Health & Science University*.

Rouhizadeh, M., Prud'hommeaux, E., Roark, B., & Van Santen, J. (2013). Distributional semantic models for the evaluation of disordered language. In *Proceedings of the conference. association for computational linguistics. north american chapter. meeting* (Vol. 2013, p. 709).

Rouhizadeh, M., Prud'hommeaux, E., van Santen, J., & Sproat, R. (2014). Detecting linguistic idiosyncratic interests in autism using distributional semantic models. In *The workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality2014* (pp. 46–50).

Tager-Flusberg, H., & Sullivan, K. (1995). Attributing mental states to story characters: A comparison of narratives produced by autistic and mentally retarded individuals. *Applied Psycholinguistics*, *16*(3), 241–256.

Wechsler, D. (2008). Wechsler adult intelligence scale–fourth edition (WAIS–IV). *San Antonio, TX: NCS Pearson*, *22*, 498.