Published in final edited form as:

Psychiatry Res. 2018 May ; 263: 74–79. doi:10.1016/j.psychres.2018.02.037.

Computational Linguistic Analysis Applied to a Semantic Fluency Task to Measure Derailment and Tangentiality in Schizophrenia

Luca Pauselli[a,*], **Brooke Halpern**[b], **Sean D. Cleary**[c], **Benson Ku**[d], **Michael A. Covington**[e], and **Michael T. Compton**[a]

[a]Department of Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY, USA

[b]Department of Psychiatry, Lenox Hill Hospital, New York, NY, USA

[c]Department of Epidemiology and Biostatistics, Milken Institute School of Public Health, The George Washington University, Washington, DC, USA

[d]Hofstra Northwell School of Medicine, Hempstead, NY, USA

[e]Covington Innovation, Athens, GA, USA

Abstract

Although rating scales to assess formal thought disorder exist, there are no objective, high-reliability instrument that can quantify and track it. This proof-of-concept study shows that CoVec, a new automated tool, is able to differentiate between controls and patients with schizophrenia with derailment and tangentiality. According to ratings from the *derailment* and *tangentiality* items of the *Scale for the Assessment of Positive Symptoms*, we divided the sample into three groups: controls, patients without formal thought disorder, and patients with derailment/tangentiality. Their lists of animals produced during a one-minute semantic fluency task were processed using CoVec, a newly developed software that measures the semantic similarity of words based on vector semantic analysis. CoVec outputs were *Mean Similarity*, *Coherence*, *Coherence-5*, and *Coherence-10*. Patients with schizophrenia produced fewer words than controls. Patients with *derailment* had a significantly lower mean number of words, and lower *Coherence-5* than controls and patients without derailment. Patients with *tangentiality* had significantly lower *Coherence-5* and *Coherence-10* than controls and patients without tangentiality. Despite the small samples of patients with clinically apparent thought disorder, CoVec was able to detect subtle differences between controls and patients with either or both of the two forms of disorganization.

*Corresponding author: Luca Pauselli, M.D., New York State Psychiatric Institute, 1051 Riverside Drive, Unit 100, New York, NY 10032. Tel: 1 (917) 340-8762. pausell@nyspi.columbia.edu.

Keywords

Automatic Data Processing; Formal Thought Disorder; Psychosis; Schizophrenia; Semantics; Semantic Fluency Tasks

1. Introduction

Formal thought disorder is characterized by disorganized and difficult to follow speech, and includes *derailment*, a sudden switching of topic with no obviously apparent logic or segues, and the less severe *tangentiality*, a response pattern that increasingly deviates off topic. These hallmark features of schizophrenia were recognized by Bleuler as "loosening of associations," or disordered thinking so severe that associations among ideas become fragmented and disturbed, and as a result, lacking in logical relationships (Bleuler, 1950). Bleuler's earliest description of patients with schizophrenia illustrated that the primary language impairment is in "context-dependent language understanding" (Bagner et al., 2003; Bazin et al., 2000; Bleuler, 1950; Linscott, 2005). He stated that although patients diagnosed with schizophrenia produce a lot of words, they do not intend to convey anything or to communicate with the environment (Meilijson et al., 2004). Formal thought disorder impairs social relationships, and greatly interferes with educational and vocational performance (Bowie and Harvey, 2008; Harrow et al., 1983a; Kuperberg, 2010; Marengo and Harrow, 1997). Unfortunately, there are few, if any, treatments for disorganization. Furthermore, there has been less research on disorganization than on other symptoms, such as delusions or hallucinations, and some researchers have recently called for more research on this often persistent and disabling domain of symptomatology (Elvevåg et al., 2007; Hart and Lewine, 2017).

Clinicians have few tools at their disposal for measuring disorganization longitudinally. The usual documentation of a mental status examination simply notes whether thought disorder is present or absent, and if present, how it manifests (e.g., loose associations, neologisms), without any numerical ratings. Some clinical documentation relies on qualitative ratings, such as "mild," "moderate," or "severe" formal thought disorder/disorganization—a rating system popularized by the 20-item *Scale for the Assessment of Thought, Language, and Communication* (Andreasen, 1979). Although subjectively evaluating the patient's verbal self-presentation is an essential diagnostic tool (Bleuler, 1950; Kraepelin, 1915; McKenna and Oh, 2008) and assessing discourse is important for prognostication (Andreasen and Grove, 1986; Harrow et al., 1983b), the characterization of incoherent ideas remains vague given the diverse types of disorganization, and the multidimensional nature of the underlying pathology (Cuesta and Peralta, 1999; Harrow et al., 1982; McKenna and Oh, 2008; Sass and Parnass, 2017). Although formal thought disorder might be an overt symptom that is recognizable, there are currently no commonly used measures for a clinician to record severity or follow severity (including improvements or worsening) over time. Our field needs highly reliable, efficient, automated, and finely detailed measures of disorganization severity, which would identify the types of disorganization and their longitudinal severity in an objective and more standardized manner. This potential value of quantifying thought disorder would be useful for prognosis, in assessing treatment responsiveness, and for diverse types of research concerning schizophrenia (Elvevåg et al., 2007). Computational linguistic approaches might advance the field.

Studies of thought disordered speech in the 1960s and 1970s focused primarily on predictability and variability of a particular word within the sentence, and have experimented with Cloze procedures (finding missing words), type-token ratios (number of different words, divided by the total number of words, as a measure of lexical variation), and readability indices (measures of word or sentence complexity) (Manschreck et al., 1981). Other than these analyses of the appearance of certain words in speech, there are also patterns of lexical and syntactic errors. Chaika (1974) described many of these errors to be exacerbations of the types of speech errors produced by healthy individuals. Analysis of these errors suggested that speech of patients with schizophrenia is generally more grammatically deviant (Hoffman and Sledge, 1988) and less syntactically complex than that of controls (Fraser et al., 1986; Morice and Ingram, 1982; Sanders et al., 1995). Unlike the relatively simple approaches to statistical linguistic measures, analysis of speech in terms of this lexical and syntactic structure more holistically captures the richness of human discourse while maintaining standardization and objectiveness. Further work is needed to develop a proper linguistically based quantitative method to characterize these deviations and complexities in a more meaningful way (Elvevåg et al., 2007; Elvevåg et al., 2017). However, like other objective linguistic tests for schizophrenia, a problem with manual approaches has been “the hours of parsing and data processing required per patient” (Fraser et al., 1986).

More recent studies of linguistic measures used automated/computational techniques (computer-derived semantic, syntactic, or pragmatic measures), and such measures were then correlated with disorganization severity. Maher (2005) used computational models to characterize the statistical properties of thought-disordered speech by quantifying the frequency of normal associations in utterances of patients with schizophrenia. Their findings that patients produced higher mean totals of associations compared to controls are consistent with models of language disturbance in schizophrenia. Elvevåg and colleagues (2007) used latent semantic analysis (LSA) to examine transcripts of patients’ speech. LSA is used to quantitatively measure “loose associations” among words. It provides a measure of semantic relatedness between text passages with the assumption that words that appear together within the same context usually have stronger associations than words appearing in different contexts. Strous et al. (2009) used machine learning to differentiate between text written by patients with schizophrenia compared to unaffected individuals via lexical and syntactical features. Word graph analysis is chronologically the most recent of the quantitative linguistic methods that has been applied to explore thought disorder using transcription of speech samples (Cabana et al., 2011; Mota et al., 2017). This method derives from developments in network theory and information science. According to this model, each word is a node, and the temporal sequences of consecutive words are directed edges; through this representation it is possible to calculate attributes that characterize graph structure, such as connectedness. In 2012, Mota and her group found that graph analysis of speech produced by psychotic patients can be used to quantitatively sort participants with mania from those with schizophrenia, detecting symptoms such as poor speech, logorrhea, and flight of thoughts even when inter-individual differences in verbosity were accounted for. In 2017, the same group applied word graph analysis in 21 recent-onset psychosis patients undergoing first clinical contact. A Disorganization Index (function of different aspects of connectedness) was built and was able to classify negative symptom severity and predict a diagnosis of schizophrenia at 6 months.

Semantic fluency tasks are a common test of speech production used in assessing neurocognition. The subject is asked to say as many words belonging to a semantic category (e.g., animals, vegetables) as possible in a certain amount of time, usually 60 seconds. Performing this task requires mental flexibility, multitasking, efficient retrieval and recall of words, cognitive self-control, reaction initiation, and inhibition (Henry and Crawford, 2004). Semantic fluency tasks are usually scored simply by counting the number of words produced. Bokart and Goldberg, in their meta-analysis (2003), demonstrated that patients with schizophrenia were consistently impaired on semantic fluency. Troyer and colleagues (1997) described a qualitative method to score fluency tasks that takes into consideration *semantic clusters* (responses are organized into groups of semantically related words) and *switches* (frequency of transitions between these groups) through manual determination of whether or not adjacent words belong to the same category. This manual approach is subjective, time consuming, and difficult to standardize, making it unlikely to be used in everyday clinical psychiatric settings outside of controlled research studies.

As noted above, LSA uses automated computational semantic indices to measure how two different words are related (Elvevåg et al., 2007; Landauer et al., 2011; Landauer and Dumais, 1997). It is one of several matrix-based approaches to comparing the contexts in which words appear, overcoming some limitations of other linguistic indices for semantic analysis (e.g., Linguistic Inquiry Word Count), which lacks the capability of measuring textual coherence as token-based methods (Neil, 2016). Conceptually, one could determine the similarity of two words by comparing all the places those two words occur in a large corpus representing the language as a whole. Doing so directly would produce a large matrix that is sparse (because most words fail to occur in most contexts) that misses indirect similarities (so that if A is similar to B and B is similar to C, no similarity of A to C would be implied). LSA uses singular value decomposition to reduce the rank of the matrix and fill in indirect similarities; CoVec (Covington, 2016) uses a matrix reduced by other methods developed by the Stanford GloVe project (http://nlp.stanford.edu/projects/glove). Either way, the descriptions of the two words being compared are vectors, which can be compared by vector cosines or other standard methods.

Bokart and Goldberg (2003) suggested investigating any potential association between semantic fluency (i.e., linguistic production) and semantic disorganization (i.e., thought disorder). In this proof-of-concept study, we demonstrate that CoVec, a new automated linguistic software, when applied to semantic fluency word lists, is able to detect clinically rated speech disorganization, specifically derailment and tangentiality. This represents the first attempt to detect formal thought disorder with a widely used, very brief cognitive task rather than natural language or free speech. With this initial demonstration, we could potentially develop an automated instrument to measure derailment and tangentiality in a clinical setting with a commonly used 60-second verbal fluency task.

2. Methods

We used for this study a sample of 105 individuals, 58 (55.2%) with a diagnosis, according to the *Structured Clinical Interview for DSM-IV Axis I Disorders* (SCID-I; First and Gibbon, 2004), of schizophrenia or first-episode non-affective psychosis (schizophreniform disorder and psychotic disorder, not otherwise specified), along with 47 (44.8%) unaffected controls (no Axis I diagnoses of psychotic or mood disorders according to the SCID-I). The latter also had no first-degree family history of a psychotic disorder according to their own report. The patients were recruited both in Washington D.C (n=23), and New York City (n=35). In Washington, D.C., patients were enrolled from a Core Service Agency (CSA) that provides outpatient community mental health services in the Georgia-Petworth neighborhood (n=3), another CSA in the northwestern D.C. (n=7), the inpatient psychiatric unit of a private, downtown, university-affiliated teaching hospital (n=7, 12.1%), and the inpatient psychiatric unit of a large community hospital in northwestern D.C. (n=6). In New York, patients were recruited from the inpatient psychiatric unit of a large community hospital in the Upper East Side of Manhattan (n=14), the outpatient mental health clinic of that hospital (n=3), an early intervention for psychosis service also affiliated with that hospital (n=2), an adult inpatient unit of a large psychiatric hospital in Queens (n=5), the outpatient mental health clinic affiliated with that hospital (n=10), and by referral from a social worker at a college who heard about the study (n=1). Data from a total of 47 unaffected controls were used for this analysis. They were recruited through advertisements placed in *AM New York* (n=28), and Craigslist (n=3); by word-of-mouth (n=4); and through flyers posted or handed out in public areas such as houses of worship, grocery stores, the YMCA, and various community centers (n=12). Eligible participants were native English-speaking and aged 18–50. Those with known or suspected intellectual disability or dementia, or a medical condition compromising ability to participate were excluded, potential controls with a SCID-based diagnosis of a psychotic or mood disorder were excluded.

All participants were administered a semantic fluency test (naming as many animals as possible in 60 seconds) as part of the *MATRICS Consensus Cognitive Battery* (Kern et al., 2008; Nuechterlein et al., 2008). Not knowing that the animal list would later be used as primary data once CoVec was developed, reliable transcripts of the animal list were available for only the above-described 105 of the subjects from a larger project involving 199 participants. The samples' sociodemographic characteristics are given in Table 1. Psychotic symptoms were assessed, among the patients, using the *Scale for the Assessment for Positive Symptoms* (SAPS; Andreasen et al., 1995). Derailment and tangentiality are assessed in the SAPS with a 6-point rating scale (0=None, 1=Questionable, 2=Mild, 3=Moderate, 4=Marked, and 5=Severe), which is used to evaluate all of the positive symptoms.

For this analysis, we divided the sample into three groups: (1) controls, (2) patients who received a score of "None" according to the SAPS derailment and tangentiality scores, and (3) patients who received a score of "Moderate," "Marked," or "Severe" on those items. The patients who were rated as "Questionable" or "Mild" were not included to ensure that the analyses took into consideration only patients with and without clear manifestations of derailment or tangentiality. Regarding *derailment*, 46 patients did not have this thought disorder, four did, and eight were excluded due to "Questionable" and "Mild" ratings. Regarding *tangentiality*, 35 patients did not have this form of thought disorder, five did, and 18 were excluded due to "Questionable" and "Mild" ratings. As such, among the seven patients with a formal thought disorder, two had derailment but not tangentiality, three had tangentiality but not derailment, and two had both derailment and tangentiality.

The transcripts of the animal list were converted to plain ASCII text and hand-edited (by a researcher blinded to the subject's status) to enforce standard spelling and punctuation, including combining two words into one where appropriate (e.g., red bird to redbird). It was observed that the samples were generally free of repetitions and of words not denoting animals.

Analysis was performed with CoVec version 1.0.5912 (Covington Innovations, www.covingtoninnovations.com/software.html). CoVec measures the semantic similarity of words using the vector methodology of the Stanford GloVe project (http://nlp.stanford.edu/projects/glove). Words are considered similar if they occur in similar contexts in a large set of English texts. The GloVe project's data file, trained on 840 billion words of English text with 300-element vectors, was used as norms. The output of CoVec effectively picks out synonyms and words that are commonly used together for any reason.

Four results were computed on each sample. Mean Similarity is the average similarity of each word to the immediately preceding word. Coherence is the average similarity of each word to each of the other words in the list, regardless of order or proximity. This tends to be lower with longer samples because longer lists are inherently more diverse. Accordingly, Coherence-5 and Coherence-10 are like Coherence, but are computed by moving a 5-word or 10-word window through the text and computing Coherence of the window as if it were the whole text, then averaging the values thus computed for all positions of the window. This produces a measure of local coherence not affected by the length of the sample.

Descriptive statistics and bivariate tests, when appropriate, were performed for sociodemographic variables, derailment and tangentiality scores, and CoVec outputs. Difference in the means between the three groups of subjects (controls, patients without clinically rated thought disorder, and patients with thought disorder) for the CoVec output measures was investigated using analysis of variance (ANOVA) and Tukey's Studentized honest significant difference (HSD) post-hoc analysis for all pairwise comparisons, which controls for Type I experiment-wise error rate and due to unequal size of all groups. When statistical significance was found ($p<0.05$) Cohen's d effect size was also calculated.

3. Results

The sample included 105 subjects: 58 (55.2%) patients with schizophrenia or first-episode non-affective psychotic disorder, and 47 (44.8%) healthy controls. Over half were male (65.7%) and African American (69.5%). The mean age was higher (36.3±9.4) in the control group than in the patient group (30.7±9.6); years of education completed followed the same pattern (Table 1).

The descriptive statistics of the CoVec output measures (Table 2) showed a significant difference in mean number of words (which is the standard outcome of a fluency task and does not require a computational approach to score it), with patients' values lower than controls' (with a large effect size, d=0.95), as well as for *Mean Similarity*, with patients' values lower than controls' (with a small effect size, d=0.22).

Correlation analysis showed that *Mean Similarity* is weakly correlated with number of words (r=−0.09), while *Coherence*, *Coherence*-5, and *Coherence-10* were more strongly correlated with number of words and between each other (range r=−0.22–0.93).

As given in Table 3, patients with derailment had a significantly lower mean number of words (12.25±5.56) than controls (22.13±5.27). Patients with derailment also had a significantly lower *Coherence-5* (0.514±0.047) than patients without derailment (0.552±0.029) and controls (0.557±0.030) with a large effect size (d_1=0.97, d_2=1.09); Table 4 shows the actual list of animals for a control, a patient without derailment, and patient with derailment, selected by the individual-level *Coherence-5* value that most approximated the group mean. There were no significant differences in *Mean Similarity*, *Coherence*, or *Coherence-10*, though means were in the expected direction numerically.

Patients without tangentiality had a significantly lower mean number of words (16.91±6.10) than controls (22.13±5.27), with a large effect size (d=0.91). Patients with tangentiality had a significantly lower *Coherence-5* (0.510±0.039) and *Coherence-10* (0.434±0.035) than patients without tangentiality (respectively, 0.552±0.028, 0.481±0.034; $d_{Coherence\text{-}5}$=1.24, $d_{Coherence\text{-}10}$=1.36), and controls (0.557±0.030, 0.480±0.034; $d_{Coherence\text{-}5}$=1.35, $d_{Coherence\text{-}10}$=1.33). Again, Table 4 shows the actual list of animals for a control, a patient without derailment, and patient with derailment, selected by the individual-level *Coherence-5* value that most approximated the group mean, without duplicating the lists given previously pertaining to derailment.

4. Discussion

This initial demonstration of CoVec, despite the limited sample sizes of patients with moderate to severe clinically rated derailment and tangentiality, shows that a very widely used one-minute cognitive test of verbal fluency may contain information beyond the simple number of words listed. Within this data, when modern computational linguistic methods are applied, there may be evidence that tools could be developed to provide computerized, objective, easy-to-obtain, quantitative measures of formal thought disorder. This new software determines whether words occurring near one another in a semantic fluency task are, in some sense, similar or coherent. CoVec detected signals capable of not only differentiating patients with derailment or tangentiality from healthy controls, but also patients with and without these clinical features. Furthermore, it may detect a lowering of coherence that could be very difficult to detect "manually," using non-computational techniques, as demonstrated in the actual lists of words given for six participants with scores closest to their respective group mean scores (i.e., even patients affected by derailment and tangentiality have a degree of similarity and coherence that is more apparent than their subtle non-coherence).

During the past decade, statistical language processing and machine learning have been increasingly used in the study of speech in people with serious mental illness (Cohen and Elvevåg, 2014). Different approaches have aimed at finding significant differences between patients with schizophrenia and controls. Elvevåg and colleagues (2010) analyzed natural speech samples of patients with schizophrenia, family members, and controls. Using their modeling approach, they demonstrated that it is possible to obtain an accurate discrimination of the three groups based on three types of measures; namely, measures of *statistical language features,* measures based on *the semantic similarity* of a discourse sample to patient or control discourse sample, and *surface features of the discourse* (such as sentence length or variability as measured by numbers of words or syllables). Semantic features analyzed with LSA played the most important role in discriminating between groups, confirming previous findings from the same group (Elvevåg et al., 2007). Bedi and his group (2015) took into consideration transcripts of interviews with youths at clinical high-risk (CHR) for psychosis. Semantic and syntactic features predicted later psychosis onset. Carrillo and colleagues (2016) combined discrete mathematics algorithms for graph characterization, with natural language processing techniques to train classifiers that can distinguish interviews from individuals with schizophrenia and controls. Holshausen's team (2014) focused their attention on formal thought disorders in older inpatients suffering from schizophrenia using LSA to process fluency tasks. For each word uttered in the semantic fluency task, they computed its vector length, and for every pair of sequential words, the cosine between the vectors for those words was computed. The average for the first set generated the *average vector length* (word unusualness measure) and the average of the cosines generated the *average cosine* (coherence measure) for each participant. Their findings suggest that measures of LSA of speech are associated with disorganized speech, performance on verbal fluency tasks, and adaptive functioning. Our approach, a vector-based method different from LSA, found that CoVec, processing a 60-second semantic fluency task transcript, can detect statistically significant differences not only between patients with formal thought disorder and healthy control s, but also between patients with and without formal thought disorder.

Several methodological limitations and caveats in interpretation are noteworthy. First is the small sample sizes in the groups with derailment and tangentiality. Despite this, significant signals were observed that merit further exploration. Second, even though we used the SAPS —a widely recognized and utilized instrument to measure positive symptoms—there is no reason to think that it is a completely accurate or "gold standard" way of evaluating formal thought disorder because the scoring is based on a clinical interview and subjective rating. In fact, future CoVec-type measures will probably be the "gold standard" as they are completely objective and perfectly reliable. Third, although in this proof-of-concept analysis we wanted to examine the four different CoVec output measures, we acknowledge that they are moderately to strongly inter-correlated, meaning that the main findings are to some extent redundant. Fourth, because we had initially had no *a priori* intent to use the lists of words as primary data (as they were collected to get a semantic fluency score for the *MATRICS Consensus Cognitive Battery*), there was a fair amount of missing data in terms of reliable and usable transcripts, and this appeared to be a greater problem among the controls (54%, compared to 40% among patients), perhaps because they tended to speak more fluently and thus give more words, making it hard to record all responses by writing (reverting instead to just counting). For this reason, future studies should audio-record the listing of words and implement a computerized transcription to keep the process as automated as possible. Fifth, consideration should be given in future studies to the fact that semantic variables are influenced by culture, experience, and geography; this could lead to biased results when the corpora chosen to derive the vectors for the analysis are not representative of the background characteristics of the population from which the sample is drawn.

This proof-of-concept study needs to be followed with larger sample sizes, and longitudinal studies would allow a test of whether measures such as those produced by CoVec could meaningfully track symptom change and response to any potential treatments.

Acknowledgments

Research reported in this publication was supported by National Institute of Mental Health grant R21 MH097999-02 (“Applying Computational Linguistics to Fundamental Components of Schizophrenia”) to the last author. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or National Institute of Mental Health. The authors report no financial relationships with commercial interests.

References

Andreasen NC. Thought, language, and communication disorders. II Diagnostic significance. Arch Gen Psychiatry. 1979; 36:1325–1330. DOI: 10.1001/archpsyc.1979.01780120055007 [PubMed: 496552]

Andreasen NC, Arndt S, Del Miller D, Flaum M, Nopoulos P. Correlational studies of the Scale for the Assessment of Negative Symptoms and the Scale for the Assessment of Positive Symptoms: an overview and update. Psychopathology. 1995; 28:7–17. DOI: 10.1159/000284894 [PubMed: 7871123]

Andreasen NC, Grove WM. Thought, language, and communication in schizophrenia: diagnosis and prognosis. Schizophr Bull. 1986; 12:348–359. DOI: 10.1093/schbul/12.3.348 [PubMed: 3764356]

Bagner DM, Melinder MR, Barch DM. Language comprehension and working memory deficits in patients with schizophrenia. Schizophr Res. 2003; 60:299–309. DOI: 10.1016/S0920-9964(02)00280-3 [PubMed: 12591591]

Bazin N, Perruchet P, Hardy-Bayle MC, Feline A. Context-dependent information processing in patients with schizophrenia. Schizophr Res. 2000; 45:93–101. DOI: 10.1016/S0920-9964(99)00167-X [PubMed: 10978877]

Bedi G, Carrillo F, Cecchi G, Slezak DF, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. Schizophrenia. 2015; 1:1–7. DOI: 10.1038/npjschz.2015.30

Bleuler E. Dementia praecox or the group of schizophrenias International Universities Press; Oxford, England: 1950

Bokat CE, Goldberg TE. Letter and category fluency in schizophrenic patients: a meta-analysis. Schizophr Res. 2003; 64:73–78. DOI: 10.1016/S0920-9964(02)00282-7 [PubMed: 14511803]

Bowie CR, Harvey PD. Communication abnormalities predict functional outcomes in chronic schizophrenia: Differential associations with social and adaptive functions. Schizophr Res. 2008; 103:240–247. DOI: 10.1016/j.schres.2008.05.006 [PubMed: 18571378]

Cabana Á, Valle-Lisboa JC, Elvevåg B, Mizraji E. Detecting order disorder transitions in discourse: Implications for schizophrenia. Schizophrenia research. 2011; 131:157–164. DOI: 10.1016/j.schres.2011.04.026 [PubMed: 21640558]

Carrillo F, Mota N, Copelli M, Ribeiro S, Sigman M, Cecchi G, et al. Automated Speech Analysis for Psychosis Evaluation. Springer, Cham. 2016; :31–39. DOI: 10.1007/978-3-319-45174-9_4

Chaika E. A linguist looks at “schizophrenic” language. Brain Lang. 1974; 1:257–276. DOI: 10.1016/0093-934X(74)90040-6

Cohen AS, Elvevåg B. Automated computerized analysis of speech in psychiatric disorders. Curr Opin Psychiatry. 2014; 27:203–9. DOI: 10.1097/YCO.0000000000000056 [PubMed: 24613984]

Covington MA. [accessed 2.3.17] Covington Innovations Software [WWW Document] 2016 URL http://www.covingtoninnovations.com/software.html

Cuesta MJ, Peralta V. Thought disorder in schizophrenia. Testing models through confirmatory factor analysis. Eur Arch Psychiatry Clin Neurosci. 1999; 249:55–61. DOI: 10.1007/s004060050066 [PubMed: 10369150]

Elvevåg B, Foltz PW, Rosenstein M, Delisi LE. An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. J Neurolinguistics. 2010; 23:270–284. DOI: 10.1016/j.jneuroling.2009.05.002 [PubMed: 20383310]

Elvevåg B, Foltz PW, Rosenstein M, Ferrer-i-Cancho R, De Deyne S, Mizraji E, Cohen A. Thoughts about disordered thinking: measuring and quantifying the laws of order and disorder. Schizophrenia bulletin. 2017; 43:509–513. DOI: 10.1093/schbul/sbx040 [PubMed: 28402507]

Elvevåg B, Foltz PW, Weinberger DR, Goldberg TE. Quantifying incoherence in speech: An automated methodology and novel application to schizophrenia. Schizophr Res. 2007; 93:304–316. DOI: 10.1016/j.schres.2007.03.001 [PubMed: 17433866]

First MB, , Gibbon M. The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) and the Structured Clinical Interview for DSM-IV Axis II Disorders (SCID-II) John Wiley & Sons Inc; Washinghton, D.C: 2004

Fraser WI, King KM, Thomas P, Kendell RE. The diagnosis of schizophrenia by language analysis. Br J Psychiatry. 1986; 148:275–278. DOI: 10.1192/bjp.148.3.275 [PubMed: 3719219]

Harrow M, Grossman L, Silverstein M. Thought pathology in manic and schizophrenic patients: Its occurrence at hospital admission and seven weeks later. Arch Gen Psychiatry. 1982; 39:665–671. DOI: 10.1001/archpsyc.1982.04290060027006 [PubMed: 6124224]

Harrow M, Silverstein M, Marengo J. Disordered thinking: Does it identify nuclear schizophrenia? Arch Gen Psychiatry. 1983; 40:765–771. DOI: 10.1001/archpsyc.1983.01790060063008 [PubMed: 6860077]

Hart M, Lewine RR. Rethinking thought disorder. Schizophrenia Bulletin. 2017; 43:514–522. DOI: 10.1093/schbul/sbx003

Henry JD, Crawford JR. A meta-analytic review of verbal fluency performance following focal cortical lesions. Neuropsychology. 2004; 18:284–295. DOI: 10.1037/0894-4105.18.2.284 [PubMed: 15099151]

Hoffman RE, Sledge W. An analysis of grammatical deviance occurring in spontaneous schizophrenic speech. J Neurolinguistics. 1988; 3:89–101. DOI: 10.1016/0911-6044(88)90008-5

Holshausen K, Harvey PD, Foltz PW, Bowie CR. Latent semantic variables are associated with formal thought disorder and adaptive behavior in older inpatients with schizophrenia. Cortex. 2014; 55:88–96. DOI: 10.1016/j.cortex.2013.02.006 [PubMed: 23510635]

Kern RS, Nuechterlein KH, Green MF, Baade LE, Fenton WS, et al. The MATRICS Consensus Cognitive Battery, Part 2: Co-norming and standardization. Am J Psychiatry. 2008; 165:214–220. DOI: 10.1176/appi.ajp.2007.07010043 [PubMed: 18172018]

Kraepelin E. Lectures on clinical Psychiatry. Am J Med Sci. 1915; doi: 10.1037/10789-000

Kuperberg GR. Language in Schizophrenia Part 1: An Introduction. Lang Linguist Compass. 2010; 4:576–589. DOI: 10.1111/j.1749-818X.2010.00216.x [PubMed: 20936080]

Landauer TK, Dumais ST. A solution to Plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychol Rev. 1997; 104:211–240. DOI: 10.1037/0033-295X.104.2.211

Landauer TK, , Mcnamara DS, , Dennis S, , Kintsch W. Handbook of Latent Semantic Analysis Routledge Taylor & Francis Group; New York: 2011

Linscott RJ. Thought disorder, pragmatic language impairment, and generalized cognitive decline in schizophrenia. Schizophr Res. 2005; 75:225–232. DOI: 10.1016/j.schres.2004.10.007 [PubMed: 15885514]

Maher BA, Manschreck TC, Linnet J, Candela S. Quantitative assessment of the frequency of normal associations in the utterances of schizophrenia patients and healthy controls. Schizophr Res. 2005; 78:219–224. DOI: 10.1016/j.schres.2005.05.017 [PubMed: 16005190]

Manschreck TC, Maher BA, Ader DN. Formal thought disorder, the type-token ratio and disturbed voluntary motor movement in schizophrenia. Br J Psychiatry. 1981; 139:7–15. [PubMed: 6117348]

Marengo JT, Harrow M. Longitudinal Courses of Thought Disorder in Schizophrenia and Schizoaffective Disorder. Schizophr Bull. 1997; 23:273–285. DOI: 10.1093/schbul/23.2.273 [PubMed: 9165637]

McKenna PJ, , Oh TM. Schizophrenic Speech: Making Sense of Bathroots and Ponds That Fall in Doorways Cambridge University Press; Cambridge: 2005

Meilijson SR, Kasher A, Elizur A. Language Performance in Chronic Schizophrenia. J Speech Lang Hear Res. 2004; 47:695.doi: 10.1044/1092-4388(2004/053) [PubMed: 15212578]

Morice RD, Ingram JCL. Language Analysis in Schizophrenia: Diagnostic Implications. Aust New Zeal J Psychiatry. 1982; 16:11–21. DOI: 10.3109/00048678209161186

Mota NB, Copelli M, Ribeiro S. Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. npj Schizophrenia. 2017; 3:18.doi: 10.1038/s41537-017-0019-3 [PubMed: 28560264]

Mota NB, Vasconcelos NA, Lemos N, Pieretti AC, Kinouchi O, Cecchi GA, et al. Speech graphs provide a quantitative measure of thought disorder in psychosis. PloS one. 2012; 7:e34928.doi: 10.1371/journal.pone.0034928 [PubMed: 22506057]

Neal A. Count or Context: Investigating Methods of Text Analysis 2016 https://cdr.lib.unc.edu/indexablecontent/uuid:da60faf1-dc20-4a82-82db-5b12713d607e

Nuechterlein KH, Green MF, Kern RS, Baade LE, Barch DM, Cohen JD, et al. The MATRICS Consensus Cognitive Battery, Part 1: test selection, reliability, and validity. Am J Psychiatry. 2008; 165:203–213. DOI: 10.1176/appi.ajp.2007.07010042 [PubMed: 18172019]

Sanders LM, Adams J, Tager-Flusberg H, Shenton ME, Coleman M, Andreasen NC, et al. A comparison of clinical and linguistic indices of deviance in the verbal discourse of schizophrenics. Appl Psycholinguist. 1995; 16:325–338. DOI: 10.1017/S0142716400065942

Sass L, Parnas J. Thought Disorder. Subjectivity, and Self Schizophrenia Billetin. 2017; 43:497–502. DOI: 10.1093/schbul/sbx032

Strous RD, Koppel M, Fine J, Nachliel S, Shaked G, Zivotofsky AZ. Automated Characterization and Identification of Schizophrenia in Writing. J Nerv Ment Dis. 2009; 197:585–588. DOI: 10.1097/NMD.0b013e3181b09068 [PubMed: 19684495]

Troyer AK, Moscovitch M, Winocur G. Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. Neuropsychology. 1997; 11:138–146. DOI: 10.1037/0894-4105.11.1.138 [PubMed: 9055277]

Highlights

- Semantic fluency tasks might contain hidden data about formal thought disorder.
- Animal lists during a 1-minute semantic fluency task were processed using a new software measuring word similarity.
- CoVec is a new tool that may be able to detect formal thought disorder in semantic fluency tasks.

Table 1

Sociodemographic characteristics of the study sample

	Total Sample (n=105)	Controls (n=47)	Patients (n=58)	Test statistic, df, p
Age, mean±SD	33.2±9.9	36.3±9.4	30.7±9.6	t=2.99, df=103, p=0.003
Gender, N (%):				χ^2=0.61, df=1, p=0.436
Male	69 (65.7%)	29 (61.0%)	40 (69.0%)	
Race, N (%)				χ^2=3.98, df=1, p=0.137
African American	73 (69.5%)	28 (59.6%)	45 (77.6%)	
Caucasian	17 (16.2%)	10 (21.3%)	7 (12.1%)	
Other	15 (14.3%)	9 (10.1%)	6 (10.3%)	
Marital status, N (%):				χ^2=0.012, df=1, p=0.914
Single and never married	92 (87.6%)	41 (87.2%)	51 (87.9%)	
Years of education, mean±SD	12.7±2.6	13.4±2.4	12.1±2.6	t=2.723 df=102 p=0.008

Table 2

Descriptive statistics of CoVec output measures (mean±SD) for the total sample, and comparisons between controls and patients

	Total Sample (n=105)	Controls (n=47)	Patients (n=58)	Test statistic, df, p
Number of Words	19.1±6.3	22.1±5.3	16.7±6.1	t=4.839, df=103, p<0.001
Mean Similarity	0.445±0.047	0.494±0.040	0.454±0.052	t=2.054, df=103, p=0.043
Coherence	0.488±0.044	0.435±0.038	0.482±0.045	t=1.257, df=103, p=0.212
Coherence-5	0.552±0.031	0.557±0.030	0.549±0.031	t=1.359, df=102, p=0.177
Coherence-10	0.478±0.033	0.480±0.034	0.477±0.033	t=0.532, df=96, p=0.596

Table 3

Group values (mean±SD) and significance from ANOVA post-hoc tests (Tukey Studentized honest significant difference (HSD), for derailment and tangentiality

	Number of Words	**Mean Similarity**	**Coherence**	**Coherence-5**	**Coherence-10**
Derailment					
A. Patients with moderate to severe derailment (*n*=4)	12.25±5.56	0.451±0.067	0.442±0.037	0.514±0.047	0.451±0.039
B. Patients without derailment (*n*=46)	17.20±6.27	0.486±0.045	0.456±0.055	0.552±0.029	0.479±0.033
C. Controls (*n*=47)	22.13±5.27	0.494±0.040	0.435±0.038	0.557±0.030	0.480±0.034
Global F test (*p*-value)	11.64 (<0.0001)	1.93 (0.15)	2.24 (0.11)	3.82 (<0.05)	0.76 (0.47)
Significance ($p < 0.05$)	C>A C>B	n.s.	n.s.	C>A B>A	n.s.
Tangentiality					
A. Patients with moderate to severe tangentiality (*n*=5)	18.00±6.36	0.453±0.041	0.407±0.030	0.510±0.039	0.434±0.035
B. Patients without tangentiality (*n*=46)	16.91±6.10	0.487±0.043	0.458±0.059	0.552±0.028	0.481±0.034
C. Controls (*n*=47)	22.13±5.27	0.494±0.040	0.435±0.038	0.557±0.030	0.480±0.034
Global F test (*p*-value)	8.73 (p<.0001)	2.26 (0.11)	3.88 (<0.05)	5.60 (<0.01)	3.63 (<0.05)
Significance ($p < 0.05$)	C>B	n.s.	n.s.	C>A B>A	C>A B>A

n.s. = no significant differences in means

Levene's test for homogeneity of variances was not significant for all comparisons.

Table 4

Actual lists of animals from six participants, illustrating differences across groups

	Analysis Pertaining to Derailment and Coherence-5			Analysis Pertaining to Tangentiality and Coherence-5		
	Control	**Patient without Derailment**	**Patient with Derailment**	**Control**	**Patient without Tangentiality**	**Patient with Tangentiality**
Group Mean	0.557	0.552	0.514	0.557	0.552	0.510
Illustrative Individual Subject's Score	0.557	0.553	0.506	0.555	0.548	0.486
Animal List	Dog Cat Lion Tiger Bear Cow Horse Bird Lizard Fish Dinosaur Guinea Pig Rat Snake Whale Shark Hippopotamus Rooster Chicken Pig Eagle	Dog Cat Fish Cow Horse Duck Lamb Shark Whale Dolphin Chicken Bird Snake Lama Flamingo	Horse Cat Elephant Lion Bear Tiger Dog Rat Bat Squirrel Mosquito Sloth Orangutan Monkey Bamboo Moth Butterfly	Dog Cat Hamster Tiger Lion Bear Koala Fish Shrimp Lobster Crab Horse Pony Donkey Snake Bird Fly Worm Rabbit Monkey Ape Gorilla Jaguar	Cat Dog Mouse Lion Tiger Bear Snake Moose Mongoose Butterfly Bee Spider	Dog Cat Killer Whale Seal Piranha Stingray Catfish Clam Crab Tyrannosaurus Rex Flounder Horse Tiger Lion Giraffe Hippopotamus Canary Parakeet Snake Gerbil Hamster Ferret