Psychiatry Research 259 (2018) 63–67

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Semantic coherence in psychometric schizotypy: An investigation using Latent Semantic Analysis

Matthew P. Marggraf[a,*], Alex S. Cohen[b], Beshaun J. Davis[a], Paula DeCrescenzo[a], Natasha Bair[a], Kyle S. Minor[a]

[a] Department of Psychology, Indiana University-Purdue University Indianapolis, United States
[b] Department of Psychology, Louisiana State University, United States

ARTICLE INFO

Keywords:
Schizotypy
Latent semantic analysis
Semantic coherence

ABSTRACT

Technological advancements have led to the development of automated methods for assessing semantic coherence in psychiatric populations. Latent Semantic Analysis (LSA) is an automated method that has been used to quantify semantic coherence in schizophrenia-spectrum disorders. The current study examined whether: 1) Semantic coherence reductions extended to psychometrically-defined schizotypy and 2) Greater cognitive load further reduces semantic coherence. LSA was applied to responses generated during category fluency tasks in baseline and cognitive load conditions. Significant differences between schizotypy and non-schizotypy groups were not observed. Findings suggest that semantic coherence may be relatively preserved at this point on the schizophrenia-spectrum.

1. Introduction

People with schizophrenia often exhibit gross reductions in semantic coherence (Breier and Berg, 1999). Following recent technological advances, computational assessments of semantic coherence have emerged that are potentially more objective, faster, and require less training compared to traditional symptom-rating measures (Cohen and Elvevag, 2014; Elvevag et al., 2007, 2010). Latent Semantic Analysis (LSA; Landauer and Dumais, 1997; Landauer et al., 1998) is one computational method that has shown promise for differentiating semantic coherence in schizophrenia-spectrum samples from healthy controls (Bedi et al., 2015; Davis et al., in preparation; Elvevag et al., 2007; Elvevag et al., 2010; Nicodemus et al., 2014). LSA is a statistical technique that enables quantification of semantic coherence in transcribed speech passages; it is based on the principle that when examined across large corpora, semantically-related words or groups of words occur together more frequently compared to words that are not semantically related (see Landauer and Dumais, 1997 and Landauer et al., 2007 for more details). LSA can be used to generate variables that assess different aspects of semantic coherence. In the schizophrenia literature, average cosine and vector length are commonly used metrics of semantic coherence (Elvevag et al., 2007; Elvevag et al., 2010; Holshausen et al., 2014; Nicodemus et al., 2014). Average cosine is a measure of the degree of semantic relatedness between sequences of words or sentences; whereas vector length is a quantification of the typicality or unusualness of each word.

While LSA has typically been applied to free speech samples, several studies have used LSA to examine semantic coherence of sequential responses on structured speech tasks, namely, category fluency tasks (Elvevag et al., 2007 [Experiment 2]; Holshausen et al., 2014; Nicodemus et al., 2014; Davis et al., in preparation). Category fluency tasks are likely to provide a fertile context for examining semantic coherence as the nature of the task requires participants to quickly generate exemplars from specific semantic category. Schwartz et al. (2003) describe an application of the "spreading activation" theory in category fluency tasks, noting that the structure and organization of memory networks is such that words are conjointly activated depending on their semantic proximity. In addition to semantic information processing, there is an executive functioning component - inhibiting and monitoring responses –believed to be involved in this task (Crawford and Henry, 2005). Semantic information processing and executive function impairments have been found to be associated with formal thought disorder in schizophrenia patients (Kerns and Berenbaum, 2002) indicating semantic coherence and verbal fluency performance may be driven by the same underlying cognitive processes.

Using a category fluency task, Elvevag et al. (2007) observed reductions in semantic coherence in schizophrenia patients by applying LSA to responses on a category fluency task. Davis et al. (in preparation) found a similar pattern of results in an early psychosis sample. Nicodemus et al. (2014) differentiated schizophrenia patients from

* Correspondence to: IUPUI School of Science, Department of Psychology, LD 124, 402 N. Blackford St., Indianapolis, IN 46202, United States.
E-mail address: mpmarggr@iupui.edu (M.P. Marggraf).

http://dx.doi.org/10.1016/j.psychres.2017.09.078
Received 26 June 2016; Received in revised form 23 May 2017; Accepted 25 September 2017
Available online 30 September 2017
0165-1781/

Table 1
Demographic information.

	Schizotypy (n = 42) Mean (SD)	Non-schizotypy (n = 38) Mean (SD)	
Age	19.48 (1.61)	18.94 (1.09)	$p = 0.09$
Gender	79% female	71% female	$p = 0.44$
Ethnicity	89% Caucasian	83% Caucasian	$p = 0.60$
Verbal Intelligence (WRAT 4)	105.15 (10.66)	104.19 (11.75)	$p = 0.71$

Note: WRAT 4 = Wide Range Achievement Test, 4th edition (Wilkinson and Robertson, 2006); SD = standard deviation.

unaffected first degree family members and healthy controls based on LSA-derived measures of semantic coherence using responses generated on a category fluency task. Although LSA has been used to demonstrate reduced semantic coherence in early and chronic stages of psychosis, it is unknown if semantic coherence is also reduced in individuals with schizotypy—those who endorse attenuated schizophrenia-like traits and are at increased risk for developing psychotic and other psychiatric disorders (Gooding et al., 2005; Meehl, 1962). Reductions in semantic coherence could be particularly important to the development of psychotic disorders, as Bedi et al. (2015) found that LSA-derived indices of semantic coherence predicted conversion to psychosis in clinically high risk (CHR) youths.

The presence of additional cognitive demand may further reduce semantic coherence in schizophrenia-spectrum populations. According to cognitive load theory (Sweller, 1983), greater cognitive load reduces the cognitive resources in working memory that are available to other functions such as producing semantically coherent speech; in turn, performance on both tasks suffer. Although the effect of increased cognitive load on semantic coherence has not been explicitly examined in schizotypy, a recent meta-analysis of neurocognitive performance in psychometric schizotypy (Chun et al., 2013) revealed that working memory deficits ($d = -0.27$) were the most affected domain in this population, suggesting a particular vulnerability to increased cognitive load. Further, Kerns and Becker (2008) found that working memory performance significantly predicted reductions on a behaviorally-based measure of semantic coherence, after accounting for schizotypy status. Taken together, these results indicate that working memory may play a critical role in semantic coherence in schizotypy. If diminished semantic coherence – a core deficit of schizophrenia – and cognitive reactivity are present in individuals at-risk for, but not yet experiencing, overt psychosis symptoms, this would suggest that subtle discourse deviations may be identifiable risk factors for schizophrenia. In this study, when compared to a non-schizotypy group, we expected those with schizotypy to exhibit: 1) Reduced semantic coherence, as measured by average cosine and vector length values; and 2) A steeper decline in semantic coherence from baseline to cognitive load conditions. An exploratory aim of this study was to examine whether semantic coherence variables would be associated with positive, negative, or disorganized schizotypy traits.

2. Method

2.1. Sample

Participants were recruited from a public university in the Southeastern United States. Participants were recruited via e-mail to complete a schizotypy questionnaire on-line and were compensated by receiving course credit if offered by the instructor or were entered into a drawing to win a $25 gift card. A total of 1296 participants completed the schizotypy questionnaire. Schizotypy and non-schizotypy groups were determined from questionnaire responses using gender- and ethnicity- derived means. Individuals included in the schizotypy group obtained a *z*-score > 1.65 above the mean (above the 95th percentile) on positive, negative or disorganized subscales; while those in the non-schizotypy group scored obtained z-scores < mean on *each* of the three subscales. Individuals who met criteria for either group (non-schizotypy or schizotypy) were invited to the laboratory for further testing. Eighty participants were included in the final sample (schizotypy $n = 42$; non-schizotypy $n = 38$). All study procedures were approved by the university's institutional review board. Groups did not significantly differ on age, gender, ethnicity, or verbal intelligence (see Table 1).

2.2. Measures

The Schizotypy Personality Questionnaire – Brief Revised (SPQ-BR; Cohen et al., 2010) consists of 34 Likert-scale items (1 = *Strongly Disagree*, 5 = *Strongly Agree*) and was used here to assess schizotypy. It has shown good internal consistency across all three subscales (positive, negative, and disorganized). Previous research suggests individuals with elevations on schizotypy measures to be at an increased risk for developing psychotic and other psychiatric disorders (Chapman et al., 1994).

Category fluency tests were administered across three conditions (two baseline and one cognitive load) and each participant completed all three conditions. For each condition, participants had 60 s to provide as many examples of a category as possible. In cognitive load conditions, participants provided examples while simultaneously completing a computerized '1-back' task which consisted of 30 geometric shapes presented consecutively on a computer screen. Participants were instructed to press the "S" key if the shape on-screen was the same as the previous shape or the "L" key if the shape was not the same. Two categories (fruit, vegetable) were counterbalanced across participants, so that approximately half of the participants completed the fruit test in baseline condition and the vegetable test in the cognitive load condition, and vice versa. The order in which participants completed the cognitive load and baseline conditions was counterbalanced. Additionally, all participants completed a second baseline test (animals) so that results could be compared directly to previous studies (Davis et al., in preparation; Elvevag et al., 2007). The second baseline test was always completed last.

Two LSA-derived variables were generated to assess semantic coherence: average cosine and vector length. Average cosine scores were calculated for each participant on the three category fluency tests following steps outlined in Elvevag and colleagues (2007). The semantic space consisted of a large text corpus composed of the type and amount of reading to which an average first-year college student would be exposed. This corpus consists of 37,561 documents and 92,409 unique words (http://lsa.colorado.edu/) Analyses were conducted using 300 dimensions. This corpus and dimensional representation is consistent with Elvevag' et al. (2007) analytic approach. Word-to-word comparisons were used, with each value represented by a cosine ranging from −1 to +1. For each test, cosines were averaged for the participant by summing cosine values for each sequential comparison and dividing by the number of word-to-word comparisons, resulting in a mean coherence score adjusted for number of words generated. Greater positive cosine values indicated greater semantic coherence (see Elvevag et al., 2007; Foltz et al., 1998 for more detail). For example, if a participant generated: 'carrot, radish, cucumber' on the vegetable-naming fluency