Lexical Analysis in Schizophrenia: How Emotion and Social Word Use Informs Our Understanding of Clinical Presentation

Kyle S. Minor\textsuperscript{a*}, Kelsey A. Bonfils\textsuperscript{a}, Lauren Luther\textsuperscript{a}, Ruth L. Firmin\textsuperscript{a}, Marina Kukla\textsuperscript{b}, Victoria R. MacLain\textsuperscript{a}, Benjamin Buck\textsuperscript{c}, Paul H. Lysaker\textsuperscript{b,d}, Michelle P. Salyers\textsuperscript{a}

\textsuperscript{a}Department of Psychology, Indiana University- Purdue University Indianapolis, Indianapolis, IN, United States
\textsuperscript{b}Roudebush VA Medical Center, Indianapolis, IN, United States
\textsuperscript{c}Department of Psychology, University of North Carolina, Chapel Hill, NC, United States
\textsuperscript{d}Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, United States

*Corresponding Author: Kyle S. Minor; IUPUI School of Science, Department of Psychology, LD 124, 402 N. Blackford St., Indianapolis, IN, 46202; Phone: (317) 274-2933; Fax: (317) 274-6756; email: ksminor@iupui.edu

This is the author's manuscript of the article published in final edited form as:
Abstract

Background: The words people use convey important information about internal states, feelings, and views of the world around them. Lexical analysis is a fast, reliable method of assessing word use that has shown promise for linking speech content, particularly in emotion and social categories, with psychopathological symptoms. However, few studies have utilized lexical analysis instruments to assess speech in schizophrenia. In this exploratory study, we investigated whether positive emotion, negative emotion, and social word use was associated with schizophrenia symptoms, metacognition, and general functioning in a schizophrenia cohort.

Methods: Forty-six participants generated speech during a semi-structured interview, and word use categories were assessed using a validated lexical analysis measure. Trained research staff completed symptom, metacognition, and functioning ratings using semi-structured interviews.

Results: Word use categories significantly predicted all variables of interest, accounting for 28% of the variance in symptoms and 16% of the variance in metacognition and general functioning. Anger words, a subcategory of negative emotion, significantly predicted greater symptoms and lower functioning. Social words significantly predicted greater metacognition.

Conclusions: These findings indicate that lexical analysis instruments have the potential to play a vital role in psychosocial assessments of schizophrenia. Future research should replicate these findings and examine the relationship between word use and additional clinical variables across the schizophrenia-spectrum.

Keywords: Schizophrenia; Speech; Negative Emotion; Social Functioning; Metacognition
I. Introduction

The words that people choose offer windows into internal states, feelings, and views of the world around them. Computerized lexical analysis is a promising tool for examining word use, as it provides a fast, reliable method for scanning narratives by grouping words and word stems into thematic categories. Researchers have observed that analyzing word use, particularly in emotion and social categories, can yield information on a range of constructs, including cortical activity (Saxbe et al., 2013), childhood behavior problems (Slatcher and Trentacosta, 2012), and psychopathological symptoms (Rude et al., 2004). Employing lexical analysis tools, either as stand-alone or supplemental methods, offers the potential to increase accuracy in behavioral assessments, as these objective measures are free of some limitations inherent in other types of measures (e.g., social desirability bias). Lexical analysis would appear to be tailor-made for investigating word use in schizophrenia, which is characterized by disorganized speech that is highly reactive to phenomenological state (Burbridge and Barch, 2002; Docherty and Hebert, 1997). These tools are unique compared to most speech instruments used in schizophrenia research, as lexical analysis focuses solely on speech content rather than syntax or word order. However, few studies have utilized lexical analysis instruments in this population.

Previous studies implementing lexical analysis have focused on comparing word use in schizophrenia and control groups, and have observed clinically significant differences between groups (Buck et al., In Press; Cohen et al., 2009; Hong et al., 2013; Junghaenel et al., 2008; but also see St- Hilaire et al., 2008). There has been little research within patient groups to investigate whether word use is linked with schizophrenia symptoms. Cohen and colleagues (2009) published the lone study on this topic, observing that patients with schizophrenia high in
anhedonia, a negative symptom characterized by reduced positive affect expression, used significantly more negative emotion words when discussing pleasant topics than those low in anhedonia or controls. Cohen and colleagues focused specifically on anhedonia; to our knowledge, no previous study has investigated whether word use is associated with the full spectrum of schizophrenia symptoms.

To enhance the clinical utility of lexical analysis instruments, it is imperative to examine whether word use can inform our understanding of commonly observed deficits in schizophrenia, such as metacognition and general functioning, where affective and social processes play critical roles. Metacognition, which is often defined as thinking about thinking (Flavel, 1979; Frith, 1992), involves a range of activities from discrete acts that require recognizing specific thoughts and feelings to synthetic acts that necessitate combining an array of intentions, thoughts, feelings, and connections between events into complex representations of others (see Gumley, 2011; Lysaker and Dimaggio, 2014; Lysaker et al., 2011). To demonstrate metacognition, one must exhibit an ability to integrate cognitive and emotional experiences in the moment and in memory, as well as recognize emotions in oneself and others. These abilities are often examined in schizophrenia by assessing the speech narratives of patients, as those with metacognitive deficits are likely to demonstrate greater impoverishment through language. Lexical analysis instruments offer the potential to examine speech narratives more deeply in order to identify moments when core psychological processes, such as emotional or social processes, are engaged. Despite the importance of emotional and social processes to metacognition and the common use of speech narratives during assessment, no previous study has investigated whether significant overlap exists between metacognition and emotion or social word use.
Regarding functioning, greater levels of negative emotion are significantly associated with lower functioning in schizophrenia (Blanchard et al., 1998). Gauging social relationships is also a key component of functional assessments. Examining associations between functioning and emotion and social word use, compared to other categories, could inform our understanding of how lexical analysis can be utilized to assess general functioning in schizophrenia. If significant relationships are observed, this would offer preliminary evidence for the utility of lexical analysis instruments as an objective screening measure of general functioning in patients.

This is the first study to implement lexical analysis to explore relationships between word use and overall symptoms, metacognition, and general functioning in individuals with schizophrenia. These relationships were examined across fifteen word use categories, spanning themes related to psychological processes and personal concerns. We expected word use to significantly predict all three clinical variables, and that emotion and social word categories would be the strongest individual predictors measured in this study.

2. Materials and Method

2.1. Participants

Participants were outpatients from a Midwestern VA Medical Center ($n = 17$) and a community mental health clinic ($n = 29$) with confirmed DSM-IV diagnoses of schizophrenia ($n = 17$) or schizoaffective disorder ($n = 29$). Exclusion criteria included an age < 18, presence of severe cognitive impairments as measured by a six-item cognitive screener (see Callahan et al., 2002), and incomplete baseline data for speech, symptoms, metacognition, or functioning. This data was part of a randomized controlled trial examining the impact of Illness Management and Recovery over 18 months (Salyers et al., 2013). For this project, we focused solely on baseline
scores (i.e., prior to any intervention). The final sample consisted of 46 participants; most were male ($n = 35, 76\%$), African-American ($n = 26, 57\%$), not currently married ($n = 31, 67\%$), completed high school or their GED ($n = 30, 65\%$), and were earning below $20,000$ annually ($n = 40, 90\%$). All study procedures were approved by the university Institutional Review Board and informed consent was obtained for all participants prior to the onset of the study.

2.2. Measures

Participants generated speech in response to open-ended questions on the Indiana Psychiatric Illness Interview (IPII; Lysaker et al., 2002), a semi-structured interview assessing perceptions of one’s life and illness. The IPII is divided into five parts, where participants are asked: 1) to tell the story of their life, beginning with their earliest memory; 2) if they think they have a mental illness and, if so, whether this mental illness has affected different facets of their lives; 3) if their mental illness controls their life and how they seek to control their mental illness; 4) how their condition affects other people; and 5) what they expect to remain the same and change for them in the future. All IPII’s were conducted by trained research assistants and typically lasted 30 to 60 minutes. The IPII was chosen for this study based on its open-ended nature, which granted participants a considerable degree of freedom when discussing topics. It also provided researchers with large segments of speech for lexical analysis (Total spoken words: $M = 3,588; SD = 2,087$).

IPII Interviews were recorded, transcribed, and processed to include only participant speech for lexical analysis. Linguistic Inquiry Word Count (LIWC; Pennebaker et al., 2007) is a computerized measure that assesses speech content using a dictionary of over 4500 words/word stems across 68 categories. In this study, we focused on the seven primary psychological process
categories (positive emotion, negative emotion, social, cognitive mechanisms, perception, biological, and relativity) on the LIWC and two personal concern categories (work, achievement) chosen based on their potential link with functioning. Subcategories for negative emotion (anxiety, anger, sadness) and social words (family, friend, humans) were also examined. LIWC does not contain subcategories for positive emotion. For each category, LIWC calculates percentage scores to account for total words spoken. Higher percentages indicate more frequent word use (see Table 1 for raw data and examples of categories/subcategories). LIWC has demonstrated good validity for measuring verbal emotional expression (Kahn et al., 2007), and has been used previously to assess word use in schizophrenia (Buck et al., In Press; Cohen et al., 2009; Hong et al., 2013; Junghaenel et al., 2008; St- Hilaire et al., 2008).

Schizophrenia symptoms, metacognition, and functioning were measured using validated instruments. The Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987) is a 30-item symptom scale that has been used extensively in schizophrenia research. The overall scale has demonstrated good internal consistency (Kay et al., 1987), interrater reliability (Bell et al., 1992; Lysaker et al., 2013), and predictive validity (Bell et al., 1992). PANSS items range from 1 (absent) to 7 (extreme) and have a three factor structure (reality distortion, negative, disorganized; Bell et al., 1994). In this study, we focused on the overall PANSS score ($M = 70.27, SD = 16.66$), as well as reality distortion ($M = 15.04, SD = 5.21$), negative ($M = 17.50, SD = 5.18$), and disorganized ($M = 14.35, SD = 4.57$) symptoms.
The Metacognition Assessment Scale (MAS-A; Lysaker et al., 2005) measures the ability to use implicit and explicit information to form and reflect upon integrated representations of self and others, and to use that knowledge to respond to psychosocial challenges. The MAS-A has exhibited good reliability and validity in psychotic populations (Lysaker et al., 2012). It consists of a total score and four domain scores: self-reflectivity, awareness of the mind of the other, decentration, and mastery (see Lysaker et al., 2013 for more information). Higher scores indicate better metacognitive abilities. For this study, only total scores (which can range from 1 to 28) were examined ($M = 12.08, SD = 3.43$). All MAS-A raters underwent extensive training and had to demonstrate good inter-rater reliability ($> .80$) with expert raters prior to rating the IPII interviews used in this study.

The Quality of Life Scale- abbreviated version (QLS; Bilker et al., 2003) was used to measure functioning. The abbreviated scale consists of seven items rated from 0 to 6, with higher scores representing better functioning. It has demonstrated high convergent validity with the original instrument. In this study, total score was examined ($M = 20.91, SD = 7.31$). Ratings for all instruments in the current study were completed by trained research staff who had demonstrated good interrater reliability on all measures. MAS-A scores were based on responses from the IPII; PANSS and QLS scores were rated using separate interviews.

2.3. Data Analysis

Data analysis was conducted in three parts. First, to control for potential differences in word use as a function of diagnosis, independent $t$-tests were conducted to examine whether LIWC categories varied in schizophrenia and schizoaffective subgroups. Second, correlations between word use categories, overall symptoms, metacognition, and functioning were analyzed
to explore expected relationships between speech content and clinical variables. Second, three linear regressions were conducted to investigate whether word use categories predicted overall symptoms, total metacognition, or general functioning. Predictor variables for each regression were selected based on a trend level relationship ($p < .10$) in correlation analyses.

3. Results

When word use categories were compared in schizophrenia and schizoaffective subgroups, we observed that biological words were used more frequently by participants with schizophrenia, $t(45) = 2.34, p = 0.02$. A trend level difference was observed in sadness words, $t(45) = 1.96, p = 0.06$, with the schizoaffective subgroup using these words at a higher rate. In subsequent analyses, controlling for these two results did not affect the interpretation of any finding.

As shown in Table 2, greater negative emotion and anger word use were significantly related to overall symptoms, while greater use of friend, cognitive mechanism, and work words showed trend level associations with symptoms. When specific symptoms were examined, three significant findings were observed. Reality distortion symptoms were significantly associated with increased use of anger words, $r = 0.33, p = 0.02$, disorganized symptoms were associated with fewer work related words, $r = - 0.35, p = 0.01$, and negative symptoms were linked with fewer social words, $r = - 0.30, p = 0.04$. Social word use was positively associated with greater metacognition, and perception words exhibited a trend level association with metacognition. A significant inverse association for anger word use and functioning was observed, while social word use demonstrated a trend level association with functioning. These analyses provided initial
support that emotion and social words were the most strongly associated LIWC categories with the clinical variables investigated here.

[INSERT TABLE TWO HERE]

In the first regression model (see Table 3), five categories (negative emotion, anger, friend, cognitive mechanisms, work) were entered simultaneously to predict overall symptoms. This model was significant, $F(5, 40) = 3.05, p = .02$, and the five categories accounted for 28% of the variance in overall symptoms (adjusted $R^2 = .19$). Anger words were the lone significant individual predictor of symptoms. Two categories (social, perception) were entered into the second model, and accounted for 16% of the variance in metacognition (adjusted $R^2 = .12$). The overall model was significant, $F(2, 43) = 4.20, p = .02$, and greater social word use significantly predicted greater metacognition. Functioning was the outcome variable in the third model, and two categories (anger, social) were entered as predictors. The overall model was significant, $F(2, 43) = 4.12, p = .02$, with predictors accounting for 16% of the variance in functioning (adjusted $R^2 = .12$). Greater anger word use significantly predicted lower functioning. To determine if anger word use accounted for significant variance in functioning beyond that of symptoms, we conducted a final regression with anger word use and symptoms entered as independent variables and functioning entered as the dependent variable. The overall model was significant, $F(2, 43) = 15.06, p < .01$, with symptoms and anger word use accounting for 41% of the variance. Greater symptoms significantly predicted lower functioning in this model, $\beta = -0.59, t(45) = -4.78, p < 0.001$. Anger words did not significantly contribute to functioning when entered with symptoms, $\beta = -0.11, t(45) = -0.92, p = 0.36$, accounting for approximately 2% of the variance. These
findings supported our hypothesis that emotion and social words would be the strongest predictors of symptoms, metacognition, and functioning. With functioning, emotion word use no longer accounted for significant variance once symptoms were also entered into the model.

[INSERT TABLE THREE HERE]

4. Discussion

Our findings indicate that lexical analysis of speech can yield important information regarding schizophrenia symptoms, metacognitive abilities, and general functioning in individuals with schizophrenia. Each of these clinical variables was significantly predicted by word use categories, with negative emotion and social word categories demonstrating the strongest relationships across variables. The finding that word use categories explained over a quarter of the variance in overall symptoms is particularly compelling. Although a previous study found that individuals with schizophrenia high in anhedonia produced more negative emotion words in pleasant conditions than those low in anhedonia or healthy controls (Cohen et al., 2009), the current study is the first to use lexical analysis to demonstrate how word use is related to the full range of schizophrenia symptoms. Our observation that word use categories significantly predicted symptoms cannot be accounted for by shared method variance, as word use and symptoms were assessed using separate interviews.

A second key finding was that anger words individually predicted greater symptoms (both overall and positive symptoms) and lower functioning (although it should be noted anger word use no longer significantly predicted functioning when entered simultaneously with symptoms). Anger is a LIWC subcategory of negative emotion containing 184 different
words/word stems that includes both overt (e.g., enrag*, hate, liar*) and more muted (e.g., contradict*, prejudic*, petty*) expressions. Additional work should be conducted to replicate this finding and determine whether both overt and muted expressions of anger predict a diminished clinical presentation. If so, lexical analysis instruments may hold promise as a fast, reliable method of scanning speech narratives to identify more subtle expressions of anger, including phrases that may go undetected by clinicians during initial assessments.

The observed link between social word use and metacognition also underscores the utility of implementing lexical analysis in schizophrenia. Being able to accurately identify metacognitive processes in spoken narratives is time-consuming and necessitates intensive training. If findings from this study are replicated, lexical instruments may have value as a supplemental form of analysis. This could reduce coding time by quickly identifying speech segments where social words are used, allowing raters to evaluate these segments for the presence of metacognition.

One of the most promising aspects of this study is the potential use of lexical analysis to examine language in schizophrenia for clues regarding deterioration in psychological processes. Based on initial findings, overall symptoms, reality distortion symptoms, and functioning appear to be best characterized by negative emotions and metacognition and negative symptoms are best predicted by social word use, which likely relies on richer processing. The relationships between word categories and clinical variables may suggest different psychological or phenomenological processes that characterize impairments in each area. Namely, that reality distortion symptoms are associated with increases in negative affect, whereas metacognition is characterized by psychological processes linked with complex thinking (e.g. social words). Future research should examine the extent that clinical characteristics can be differentiated according to affective and
cognitive characteristics. Lexical analysis could be a useful tool in this pursuit. Such differential processes (distinguishing cognitive and affective characteristics) are mirrored in the study of social cognition (Mancuso et al., 2011).

A strength of this study is the use of the IPII, an open-ended clinical interview that typically lasts between 30-60 minutes. This allowed participants to produce large speech segments for lexical analyses. However, the length of these interviews also limited the number of individuals willing to participate. The small sample hindered our ability to examine demographic differences in word categories, which would have been particularly useful for investigating whether gender played a role in observed relationships between anger word use, symptoms, and functioning. The small sample also increased the possibility of chance findings. A second limitation related to the use of the IPII centers on emotional reactivity. One of the reasons we chose the IPII for this study is that participants control what they discuss and how much they expound on the subject in response to prompts about their life and illness. However, discussing these subjects may prompt increased negative emotional reactions in some participants. Future studies using the IPII and LIWC should obtain pre-post ratings of positive and negative affect in order to determine whether emotional reactivity to these prompts may account for some of the variation in emotion word use. A third limitation of using the IPII is that both LIWC ratings and metacognition ratings were derived from this interview. Although these ratings are quite different (i.e., metacognition is driven by context and meaning, whereas LIWC ratings only focus on how often a word category is used), future studies should also examine whether similar relationships are present when assessing word use from separate interview. A fourth limitation of this study is that participants tended to have been ill for long periods, and findings may not reflect word use in individuals with shorter durations of psychosis. To address these limitations,
future studies should analyze how word use categories differ according to demographic variables and stage of illness.

In the future, researchers should also seek to integrate lexical analysis with other types of speech instruments, such as latent semantic analysis and discourse incoherence measures, both of which have shown promise for investigating speech across the schizophrenia-spectrum (Docherty, 2012; Elvevag et al., 2007; Minor and Cohen, 2010; Minor and Cohen, 2012). Developing novel ways to combine different speech instruments may increase the effectiveness of these tools. Additional research avenues to pursue include examining whether relationships observed here are stable over time, testing how word use categories are related to other clinical variables or different types of psychopathology to enhance specificity, and investigating if lexical analysis can be used to help predict response to specific interventions. Although findings from this study must be replicated, these results represent early steps in each of these directions.
References


Blanchard JJ, Mueser KT, Bellack AS. Anhedonia, positive and negative affect, and social functioning in schizophrenia. Schizophrenia Bulletin 1998;24,413-424.


Docherty NM, Hebert AS. Comparative affective reactivity of different types of communication disturbances in schizophrenia. Journal of Abnormal Psychology 1997;106,325-330.


Gumley A. Metacognition, affect regulation and symptom expression: A transdiagnostic perspective. Psychiatry Research 2011;190,72-78.


Lysaker PH, Carcione A, Dimaggio G, Johannesen JK, Nicolo G, Procacci M, Semerari A. Metacognition amidst narratives of self and illness in schizophrenia: Associations with

Lysaker PH, Clements CA, Plasck-Hallberg CD, Knipscheer SJ, Wright DE. Insight and personal narratives of illness in schizophrenia. Psychiatry 2002;65,197-206.


Lysaker PH, Gumley A, Brune M, Vanheule S, Buck KD, Dimaggio G. Deficits in the ability to recognize one’s own affects and those of others: Associations with neurocognition, symptoms and sexual trauma among persons with schizophrenia spectrum disorders. Consciousness and Cognition 2011;20,1183-1192.


Lysaker PH, Ringer JM, Buck KD, Grant M, Olesek K, Leudtke BL, Dimaggio G.

Metacognitive and social cognition deficits in patients with significant psychiatric and medical adversity: A comparison between participants with schizophrenia and a sample of participants who are HIV-Positive. Journal of Nervous and Mental Disease 2012;200,130-134.


Minor KS, Cohen AS. Affective reactivity of speech disturbances in schizotypy. Journal of


