Toward Objective, Multifaceted Characterization of Psychotic Disorders: Lexical, Structural, and Disfluency Markers of Spoken Language

Alexandria K. Vail
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
avail@cs.cmu.edu

Elizabeth Liebson
Department of Psychiatry
Harvard Medical School
Boston, Massachusetts, USA
eliebson@partners.org

Justin T. Baker
Department of Psychiatry
Harvard Medical School
Boston, Massachusetts, USA
jtbaker@partners.org

Louis-Philippe Morency
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
morency@cs.cmu.edu

ABSTRACT

Psychotic disorders are forms of severe mental illness characterized by abnormal social function and a general sense of disconnect with reality. The evaluation of such disorders is often complex, as their multifaceted nature is often difficult to quantify. Multimodal behavior analysis technologies have the potential to help address this need and supply timelier and more objective decision support tools in clinical settings. While written language and nonverbal behaviors have been previously studied, the present analysis takes the novel approach of examining the rarely-studied modality of *spoken* language of individuals with psychosis as naturally used in social, face-to-face interactions. Our analyses expose a series of language markers associated with psychotic symptom severity, as well as interesting interactions between them. In particular, we examine three facets of spoken language: (1) lexical markers, through a study of the function of words; (2) structural markers, through a study of grammatical fluency; and (3) disfluency markers, through a study of dialogue self-repair. Additionally, we develop predictive models of psychotic symptom severity, which achieve significant predictive power on both positive and negative psychotic symptom scales. These results constitute a significant step toward the design of future multimodal clinical decision support tools for computational phenotyping of mental illness.

ACM Reference Format:
Alexandria K. Vail, Elizabeth Liebson, Justin T. Baker, and Louis-Philippe Morency. 2018. Toward Objective, Multifaceted Characterization of Psychotic Disorders: Lexical, Structural, and Disfluency Markers of Spoken Language. In *2018 International Conference on Multimodal Interaction (ICMI '18), October 16–20, 2018, Boulder, CO, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3242969.3243020

ICMI '18, October 16–20, 2018, Boulder, CO, USA

ACM ISBN 978-1-4503-5692-3/18/10…$15.00
https://doi.org/10.1145/3242969.3243020

1 INTRODUCTION

Psychotic disorders are forms of severe mental illness that cause significant functional impairment and can result in profound lifetime disability and loss of productivity [1]. Assessment of psychotic disorders often relies upon clinical interviews and observation of an individual's day-to-day behaviors, but unfortunately, clinicians put in this role are often bounded by constraints such as time availability, clinician fatigue, or the simple human inability to observe all channels of behavior at once. These difficulties necessitate the development of tools for the computational phenotyping of mental illness, which can offer objective support and data analysis to clinicians to aid in assessment and treatment.

When assessing the psychiatric condition of an individual, clinicians rely upon subjective analysis of atypicalities in the individual's behavior, such as nonverbal cues, social behaviors, and language use. Critically, these behaviors can also be evaluated through multimodal behavior analysis systems. Although a moderate amount of work has focused on nonverbal behaviors through audio-visual information [31, 32], little work has focused on the language use of these individuals with psychotic disorders. Further, to date, almost all work on language use in psychotic disorders has focused on written texts, such as autobiographical narratives and social media interactions [13, 24]. The present work is one of the first studies to examine *spoken* language use in individuals with psychotic disorders from a computational perspective in clinical settings.

Furthermore, most prior work has examined differences between individuals diagnosed with psychotic disorders and those who are not [3, 4, 17, 24], but few studies have examined behaviors within psychotic disorder groups. The primary line of work to date on symptom-specific written language use focuses on anhedonia, a negative symptom of schizophrenia characterized by a reduction in expression of positive affect [2, 3]. This prior work studies only one specific symptom of schizophrenia and does not yet cover the full range of symptoms expressed by psychotic disorders. The present analysis takes the novel approach of examining language use as it

pertains to a broad range of psychotic symptoms and more fully characterizes an individual's manifestation of the disorder.

In this paper, we analyze three facets of spoken language use in individuals with psychotic disorders: (1) lexical markers, through a study of the function of words; (2) structural markers, through a study of grammatical fluency; and (3) disfluency markers, through a study of dialogue self-repair. For each of these three facets, we perform single-facet analyses which will inform our multi-faceted fusion approach. Our multi-faceted interaction analysis is conducted in two parts: a moderation analysis and predictive model building. Our moderation analysis examines how the relationship between an individual's symptom severity and two facets at a time. Our multi-facet predictive models consider the set of features emerging as significant in single-facet analyses as predictors of psychotic symptom severity. We perform our analyses and experiments on a dataset consisting of semi-structured clinical interviews between clinicians and adult individuals with schizophrenia or bipolar disorder recently admitted to an inpatient unit at a major psychiatric facility.

2 PSYCHOSIS AND LANGUAGE

Extended analysis of language use has the potential to influence the understanding of language dysfunction in psychosis as well as the potential further development of clinical assessment tools. We examine features at three levels of a participant's language use: *lexical markers*, *structural markers*, and *disfluency markers*. The following subsections detail previous work in these areas. This prior work will inform our single-faceted research described in Section 4.

2.1 Lexicon

As previously mentioned, any previous studies implementing lexical analysis have (1) focused on written language and (2) compared between psychotic disorder and control groups [3, 4, 17]. However, few studies have examined research within psychotic disorder groups to investigate whether word use is linked with psychotic symptoms themselves. For our lexical marker analyses, we focus on five lexical categories, which we introduce as part of three main groups: affect, power, and reality monitoring.

Affect. The foremost line of study of this topic is focused on anhedonia [2, 3], a negative symptom of schizophrenia characterized by a reduction in expression of positive affect. Cohen et al. observed that participants exhibiting high levels of anhedonia used more negative affect words when discussing pleasant topics than those exhibiting low levels of anhedonia [4]. In our analysis we follow this line of work by investigating *affect* words as they relate to the broader spectrum of psychotic symptoms.

Power. The most characteristic symptoms of psychotic disorders revolve around delusions and grandiosity [19]. Individuals that express high levels of delusion tend to hold beliefs which are unfounded, unrealistic or idiosyncratic [19]. Grandiosity, on the other hand, involves an exaggerated self-opinion and unrealistic convictions of superiority, which can include delusions of extraordinary abilities, wealth, knowledge, fame, power, or moral righteousness [19]. Our analysis examines the impact of delusions and grandiosity on the language of individuals with psychotic disorders via words of *power*: words relating to the drive for influence and dominance.

Reality monitoring. Another significant segment of the lexicon examined in the present analysis involves words related to *reality monitoring* [16]. The concept of reality monitoring extends from the idea that people recall information from two primary sources: external sources (such as perceptual processing and contextual information) and internal sources (such as reasoning). Reality monitoring refers to the processes people use to decide whether information was generated from an external source or an internal source.

Numerous studies have observed reality monitoring impairments in individuals with psychotic disorders compared to healthy controls [7, 9, 20], but most work focuses on the neurocognitive aspects of the phenomenon, as opposed to detection in the field. The present analysis takes the novel approach of investigating reality monitoring as it manifests in conversational settings (i.e., spoken language). In particular, it features a focus on the use of words that reflect each of the two potential sources of information: external sources through *perceptual processing* and *relative* (contextual) words, and internal sources through *cognitive processing* words.

2.2 Language Structure

Individuals with speaking disorders or cognitive impairment tend to express themselves atypically compared to control groups [8]. Prior work on written language has used language models to study this phenomenon by estimating the probability of a given utterance being produced, e.g., in studies of language impairment in children [8] and language dominance prediction in multilingual individuals [30]. Hong et al. conducted a study of autobiographical narratives written by individuals with and without schizophrenia; this work suggested that different language models optimally explain part-of-speech tag sequences within the two groups [13].

Few previous studies have examined perplexity itself as a measure of grammatical integrity in schizophrenia and psychosis. A study by Mitchell et al. compared posts by social media users voluntarily self-labeled as experiencing schizophrenia against posts from a control group; a marginal difference between these sets of users suggested that those with schizophrenia generated higher-perplexity posts than the control group [24]. The present study takes the novel approach of investigating perplexity as an indirect measure of psychotic symptom severity, rather than as a distinguishing characteristic between individuals with psychotic disorders and those without.

2.3 Disfluency

Disfluencies, such as self-repairs, pauses, and fillers (such as *er* and *umm*) are pervasive in day-to-day dialogue [28]. These disfluencies are generally regarded as symptomatic of problems in communication, whether caused by production or self-monitoring issues [22]. Disfluencies can also highlight the interactive nature of dialogue — some disfluencies occur as a result of tailoring dialogue to a specific listener, or in response to feedback from interlocutors [11].

What brought you into the hospital?
Has anything in particular been on your mind?
What has the team here been helping you with?
Would you say that they are doing a good job?
What are your goals for the hospitalization?
How are people treating you?
How is the food?
How is your mood? / How are your spirits?
How is your thinking/focus?
How is your energy?
How have you been sleeping?
How is your self-confidence compared to how it usually is?
What changes do you observe since you were hospitalized?

Table 1: A list of interview questions administered during the session.

Individuals with psychotic disorders tend to have difficulties with language and social cognitive skills, and especially with self-monitoring [15] and turn-taking [25], but little research has examined how these problems affect interaction. Work by Leudar et al. found that the less self-repair that an individual with schizophrenia employs, the more verbal hallucinations they tend to experience [21]. Further work by McCabe et al. discovered that other-initiated repairs (clarification of a clinician's dialogue, in particular) are associated with improved adherence to treatment [23]. The present work, therefore, examines the disfluencies and self-repairs present in the dialogue of individuals with psychotic disorders as they relate to symptom severity.

3 DYADIC PSYCHOSIS INTERVIEW DATASET

The dataset examined in the present analysis consists of a series of clinical interviews with adult individuals recently admitted to an inpatient psychotic disorder unit at a major psychiatric facility. Video and audio recordings, as well as transcripts, were collected from 53 sessions (28 unique participants). Each session consisted of a semi-structured clinical interview between the admitted individual and a clinician, lasting approximately 10–15 minutes each. The interview script was modeled upon existing everyday clinical interactions designed to elicit reactions that may be illustrative of the psychiatric condition of the individual[1]. A list of interview questions is presented in Table 1.

Following the conclusion of each interview, each participant was administered a series of clinical scales, including the Positive and Negative Syndrome Scale (PANSS) [19], a scale used for measuring psychotic symptom severity. PANSS involves seven-point ratings of 30 symptoms across three dimensions: *positive symptoms*, involving behaviors in excess or distortion of normal function; *negative symptoms*, involving behaviors diminished or suppressed below normal function; and *general psychiatric symptoms*, involving items that cannot be linked decisively to either syndrome. In this paper, we focus on the symptoms from the positive and negative scales (see descriptions in Table 2). The average positive scale score in the present sample is $\mu = 14.88$ ($\sigma^2 = 7.82$), and negative scale score $\mu = 12.14$ ($\sigma^2 = 4.71$), both in a possible range of 7 to 49 (see Figure 1 for the distribution of the present sample).

[1]Although participants varied with regard to previous exposure to interactions of this type, this diversity is reflective of the larger population, and we believe that this strengthens the applicability of this analysis.

Scale Item	Brief Description of Behavior
Positive Scale	
Delusions	Beliefs which are unfounded, unrealistic, and idiosyncratic.
Conceptual Disorganization	Disorganized process of thinking characterized by disruption of goal-directed sequencing, e.g., circumstantiality, tangentiality, loose associations, non-sequiturs, gross illogicality, or thought block.
Hallucinatory Behavior	Verbal report or behavior indicating perceptions which are not generated by external stimuli. These may occur in the auditory, visual, olfactory, or somatic realms.
Grandiosity	Exaggerated self-opinion and unrealistic convictions of superiority, including delusions of extraordinary abilities, wealth, knowledge, fame, power, and moral righteousness.
Hostility	Verbal and nonverbal expressions of anger and resentment, including sarcasm, passive-aggressive behavior, verbal abuse, and assaultiveness.
Negative Scale	
Blunted Affect	Diminished emotional responsiveness as characterized by a reduction in facial expression, modulation of feelings, and communicative gestures.
Emotional Withdrawal	Lack of interest in, involvement with, and affective commitment to life's events.
Poor Rapport	Lack of interpersonal empathy, openness in conversation, and sense of closeness, interest, or involvement with the interviewer. This is evidenced by interpersonal distancing and reduced verbal and nonverbal communication.
Difficulty in Abstract Thinking	Impairment in the use of the abstract-symbolic mode of thinking, as evidenced by difficulty in classification, forming generalizations, and proceeding beyond concrete or egocentric thinking in problem-solving tasks.
Lack of Spontaneity and Flow of Conversation	Reduction in the normal flow of communication associated with apathy, avolition, defensiveness, or cognitive deficit. This is manifested by diminished fluidity and productivity of the verbal-interactional process.

Table 2: Enumeration and brief description of a selection of symptoms contained in the PANSS positive and negative scales [19].

For the following analyses, the dataset was separated into a training set (43 sessions) and a held-out test set (10 sessions). The single-facet analyses were performed upon the training set, and only the multi-faceted predictive models were tested upon the held-out test set at the conclusion of the analysis.

4 SINGLE-FACET LANGUAGE ANALYSIS

Our first set of analyses examines spoken language use at three levels of a participant's dialogue: lexical markers, structural markers, and disfluency markers. The following subsections detail the computational analyses of these three facets of spoken language. The results of these single-facet analyses will be used during the multi-faceted prediction task.

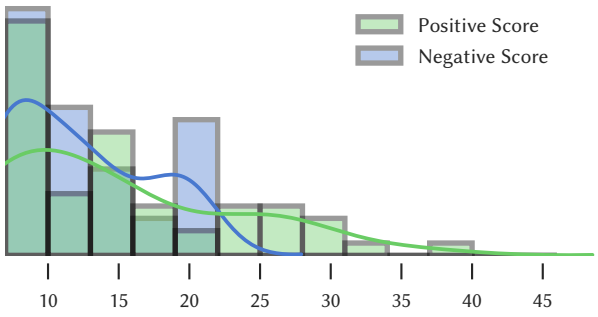


Figure 1: The distribution of PANSS positive and negative scores in the examined sample.

	Positive Score		Negative Score	
	corr(ρ)	p-value	corr(ρ)	p-value
Cognitive Processing	+0.048	0.736	+0.018	0.898
Affect	−0.063	0.655	**+0.287**	**0.037**
Power	**+0.374**	**0.006**	+0.091	0.516
Relative	**−0.302**	**0.028**	**−0.352**	**0.010**
Perceptual Processing	**+0.351**	**0.010**	+0.111	0.429

Table 3: Reported Spearman's rank correlation coefficient between selected LIWC features and PANSS scores. Boldface indicates significant correlations holding under a Benjamini-Hochberg procedure for multiple hypothesis testing; $\alpha = 0.05$.

4.1 Lexicon Analysis

In this study, we focus on five categories of lexical markers: cognitive processing words, affect words, power words, relative words, and perceptual processing words (see Section 2.1 for details). Lexical features of participant speech were extracted using the Linguistic Inquiry and Word Count (LIWC) tool [26], a computerized measure that assesses speech and language content using a dictionary of over 4500 words across over 60 categories. LIWC has demonstrated validity in measuring expression in verbal dialogue [18] and has been used previously to assess word use in schizophrenia for written text [3, 4]. We computed a Spearman's rank correlation coefficient to assess the relationship between each of these categories and two PANSS scales (positive and negative). To account for multiple hypothesis testing, results were filtered within each scale using the Benjamini-Hochberg procedure with a family-wise error rate of $\alpha = 0.05$. All analyses were perfomed upon the training set only. Results are reported in Table 3; significant correlations are discussed below and illustrated in Figure 2.

Affect. Affect words relate to the emotions: for example, *happiness*, *gloomy*, and *sadly*. Previous work has suggested that greater levels of emotion are significantly associated with lower functioning in psychotic disorders [2], and expression of negative affect, in particular, has been linked to anhedonia, a major negative symptom, in the past [4]. There was a significant positive correlation between affect words and negative PANSS score ($\rho(53) = +0.287$, $p = 0.037$). The more negative symptoms expressed by a participant, the more affect words they used.

Power. Power words relate to the drive for dominance: for example, *superiority*, *important*, and *exploit*. Individuals with psychotic disorders often exhibit symptoms of grandiosity and delusions, which are associated with a perception of greater self-power [19]. There was a significant positive correlation between power words and positive PANSS score ($\rho(53) = +0.417$, $p = 0.002$). Overall, the more positive symptoms expressed by a participant, the more power words they used.

Reality monitoring. Relative words relate to situations regarding time and space: for example, *yesterday*, *lately*, and *nearby*. These words relate to the phenomenon of reality monitoring, and particularly to the attachment of information to external stimuli [16]. There was a significant negative correlation between relative words and negative PANSS score ($\rho(53) = -0.381$, $p = 0.005$), as well as a significant negative correlation between positive PANSS score ($\rho(53) = -0.302$, $p = 0.028$). We can infer from this result that the more positive or negative symptoms expressed by a participant, the fewer relative words they used.

Perceptual processing words relate to the senses: for example, *feeling*, *see*, and *listened*. Like relative words, these words also tend to relate to reality monitoring, and these words are also linked to the perception of external stimuli [16]. There was a significant positive correlation between perceptual processing words and positive PANSS score ($\rho(53) = +0.434$, $p = 0.001$). Overall, the more positive symptoms expressed by a participant, the more perceptual processing words they used.

4.2 Language Structure Analysis

The structure of the language — including vocabulary and syntactic constructions — expressed by a participant can be measured via *perplexity*, a measurement based on entropy, and can be interpreted to roughly estimate how predictable is a sequence of words. The present work trains a trigram backoff language model on the Switchboard corpus [10], a sizable multispeaker corpus of conversational speech and text through telephone conversations about varying topics. This corpus can be viewed as an approximation of non-psychotic disorder spoken dialogue. The model is then tested on the transcript of each session, and the overall perplexity is calculated. A Spearman's rank correlation coefficient is computed to assess the relationship between perplexity and each of the PANSS scales. All analyses were performed upon the training set only.

Results. The results suggest no significant correlation between negative PANSS score and perplexity ($\rho(53) = -0.046$, $p = 0.746$), but a significant positive correlation between positive PANSS score and perplexity ($\rho(53) = +0.313$, $p = 0.022$). The more positive symptoms an individual expresses, the higher the perplexity of their utterances. Individuals high in positive scale symptoms tend to express symptoms such as excitement and conceptual disorganization, which may interfere with sentential construction [19].

4.3 Disfluency Analysis

Disfluencies in the form of speech repair are typically assumed to have a tripartite *reparandum-interregnum-repair* structure [29], as illustrated in the following example.

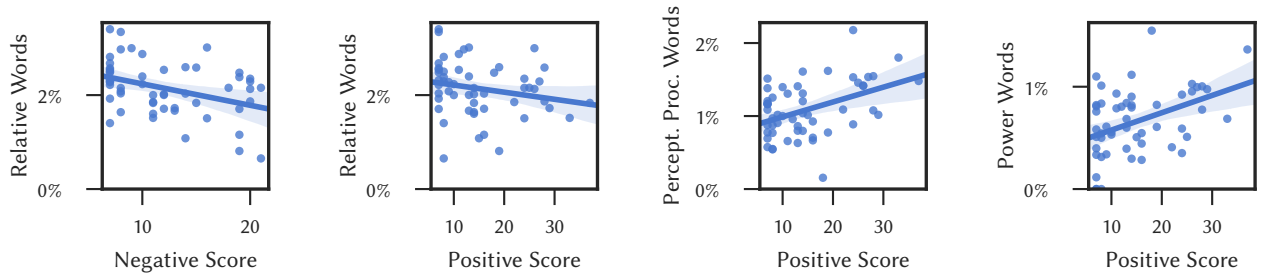


Figure 2: Regression plots of four of the significant correlations between LIWC features and PANSS scores.

"John [likes] [uh] [loves] Mary"
reparandum interregnum repair

A *reparandum* is an error in speech that is subsequently corrected by the speaker; a *repair* term is the corrected speech. An *interregnum* term is a filler token or a cue phrase between the reparandum and repair terms, often a stalling measure while the speaker generates the repair term.

We examine three forms of disfluencies: edits, repeats, and restarts. If the reparandum and the repair terms are absent, the disfluency is considered to be reduced to an isolated *edit* term. In this canonical example, the interregnum is a pause filler token ("uh"), but more phrasal terms such as "I mean" and "you know" are also often used.

The other two forms of repair we examine in the present analysis are *repeat* terms and *restart* terms. The occurrence of a *repeat* term is reasonably straightforward — this is when an individual repeats a word or a short phrase. A *restart* term occurs when an individual changes a partially-complete spoken utterance, as in the example above.

Self-repairs were annotated automatically using a deep-learning-driven incremental disfluency detection model developed by Hough et al. [14]. This model consists of deep learning sequence models that consume incoming words and use word embeddings, part-of-speech tags, and other features to predict disfluency labels for each word in a strictly left-to-right, word-by-word fashion.

Similar to the lexicon analysis, a Spearman's rank correlation coefficient was computed to assess the relationship between each type of self-repair and each PANSS scale (positive and negative). To control for multiple hypothesis testing, results were filtered within each scale using the Benjamini-Hochberg procedure with a family-wise error rate of $\alpha = 0.05$. All analyses were performed upon the training set only.

Results. Results are reported in Table 4; significant correlations are discussed below and illustrated in Figure 3. Both significant correlation results are related to negative PANSS score. The negative PANSS score is characterized by symptoms such as poor rapport, difficulty in abstract thinking, and lack of spontaneity and awkward flow of conversation [19]. There was a significant positive correlation between the negative PANSS score and edit terms ($\rho(53) = +0.309$, $p = 0.024$) as well as a significant positive correlation between the negative PANSS score and restarts ($\rho(53) = +0.334$, $p = 0.014$). The more negative symptoms expressed by an individual, the more edit terms and restarts they express.

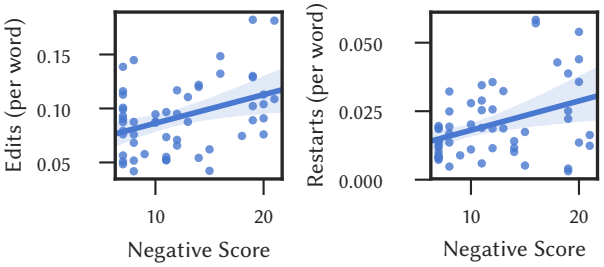


Figure 3: Regression plots of the significant correlations between self-repair features and PANSS scores.

	Positive Score		Negative Score	
	corr(ρ)	p-value	corr(ρ)	p-value
Edits	−0.089	0.525	**+0.309**	**0.024**
Restarts	+0.173	0.217	**+0.334**	**0.014**
Repeats	+0.028	0.844	+0.215	0.123

Table 4: Reported Spearman's rank correlation coefficient between selected self-repair features and PANSS scores. Boldface indicates significant correlations holding under a Benjamini-Hochberg procedure for multiple hypothesis testing; $\alpha = 0.05$.

4.4 Discussion

In this section, we summarize our observations for all three facets of spoken language: lexical markers, structural markers, and disfluency markers. For lexical markers, we group our observations following the three lexical category groups introduced in Section 2.1.

Affect. Our analyses investigated a series of lexicon categories as used by individuals with psychotic disorders (Section 4.1). There existed a positive correlation between affect words and negative symptoms: the more affect words an individual used, the more severe their negative symptoms. Interestingly, this counters the intuition regarding the negative symptom of emotional withdrawal

and blunted affect [19]; one might believe that an individual with severe negative symptoms may not be very forthcoming about their emotions. This result relates to prior work on anhedonia, which suggested that individuals with this negative symptom do not use significantly fewer affect words than those without, but instead use affect words with a more negative valence [2].

Power. Another result involves power words: the more power words an individual expresses, the higher the severity of their positive symptoms. Some of the characteristic positive symptoms include delusions and grandiosity, which involve holding beliefs that are unfounded, unrealistic, or idiosyncratic, exaggerated self-opinion, and unrealistic conventions of superiority [19]. Considering that these symptoms are central to the positive symptom scale, this finding represents a useful contribution toward computational phenotyping of psychotic disorders.

Reality monitoring. Two lexicon categories emerged that are related to reality monitoring: relative words and perceptual processing words, both of which are related to information recall from external sources [16]. Relative word use is negatively associated with both negative and positive symptoms: that is, the more severe the psychotic symptoms an individual expresses, the less they speak in relative terms. It is interesting to see that this correlation holds for both symptom scales; this may be an indication of a general difficulty in psychotic disorders, rather than dependent on its manifestation. This result reinforces the findings from previous studies that suggested that reality monitoring impairments are generally characteristic of psychotic disorders [7, 20]. There was also a positive association between positive symptoms and perceptual processing: the more perceptual processing words an individual used, the more severe their positive symptoms. Unlike relative word use, perceptual processing word use appears to be dependent upon the particular manifestation of the disorder: one of the characteristic positive symptoms is hallucinatory experiences, which may lead to an individual being more aware of their surroundings, real or imagined, which in turn leads to more discussion about what they feel, see, and hear.

Structure. A correlation was discovered between positive symptom severity and language perplexity (Section 4.2). Positive symptoms entail higher-activity behaviors in excess of typical function, so individuals expressing these symptoms acutely may experience difficulty in constructing sentences; this follows from previous work suggesting that individuals with cognitive impairment may express themselves atypically compared to control groups [8].

Disfluency. There were two results regarding self-repairs during dialogue (Section 4.3). In particular, negative symptom severity was positively correlated with both edit terms and restarts. Disfluencies are generally regarded as symptomatic of problems in communication [22]. Individuals with high negative psychotic symptom severity characteristically experience problems in communication through poor rapport and flow of conversation [19]; it follows logically that this may be expressed linguistically through dialogue disfluencies.

5 MULTI-FACETED LANGUAGE ANALYSIS

Building from the results of the single-facet computational analyses, we are interested in examining the interactions between the different facets of spoken language. In this section, we leverage these results in two multi-facet analyses: an analysis of moderation and predictive modeling. The moderation analysis will focus on two facets at a time, while the predictive modeling will integrate all three facets.

5.1 Moderation Analysis

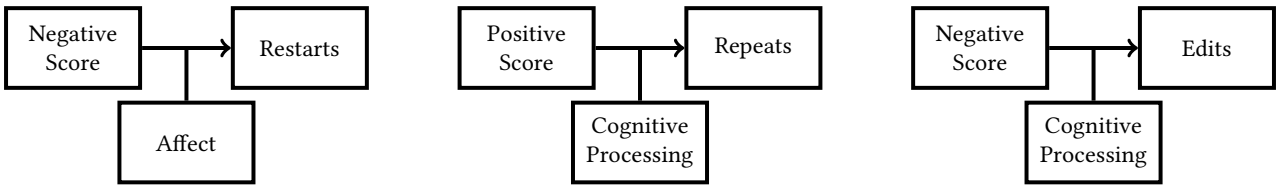
Each of the two PANSS scales (positive and negative) were examined as a moderator of the relation between each of the lexicon features and each form of self-repair. In other terms, the analysis focused on how individuals expressing high positive or negative symptoms might self-repair more frequently when speaking on particular topics (see Figure 4 for an illustration). This work is conducted as a form of regression analysis [5]. Given a PANSS score X_S and a lexicon feature X_L, we predict a given dependent variable (i.e., a self-repair feature) Y_R with the model

$$Y_R = \beta_S X_S + \beta_L X_L + \beta_{SL} X_S X_L,$$

such that β_S, β_L, and β_{SL} are learned parameters via ordinary least squares on the training set [27]. For example, Y_R could indicate self-repair repeats, while X_S and X_L indicate positive PANSS score and affect words, respectively. We describe below three moderation models with significant interactions.

Negative symptoms, affect, restarts. The first model involves negative PANSS score, affect words, and restarts (see Figure 4a). In the first step of the regression analysis, negative PANSS score and affect words are entered as predictors of restarts; this model significantly predicted restarts ($F(50, 2) = 4.797$, $p = 0.012$, $r = +0.401$). In the second step of the analysis, the interaction term (the product of the negative PANSS score and affect word use) was introduced; this model also significantly predicted restarts ($F(49, 3) = 4.733$, $p = 0.006$, $r = +0.474$). This difference was statistically significant ($\Delta r = +0.073$, $p = 0.050$). See Table 5 for the final interaction model; β is the coefficient for each term, and t and p refer to a t-test value and p-value indicating its significance. From these results we can observe that the higher an individual's negative PANSS score and the more affect words they used, the more they restarted their sentences, but when high-negative-score individuals spoke about affective utterances, they expressed *fewer* restarts than in general.

Positive symptoms, cognitive processing, repeats. The second model involves positive PANSS score, cognitive processing words, and repeats (see Figure 4b). In the first step of the regression analysis, positive PANSS score and cognitive processing words are entered as predictors of repeats; this model marginally predicted repeats ($F(50, 2) = 1.952$, $p = 0.153$, $r = +0.269$). In the second step of the analysis, the interaction term (the product of the positive PANSS score and cognitive processing word use) was introduced; this model did significantly predict repeats ($F(49, 3) = 2.754$, $p = 0.052$, $r = +0.380$). This difference was statistically significant ($\Delta r = +0.111, p = 0.048$). See Table 5 for the final interaction model; β is the coefficient for each term, and t and p refer to a t-test value and p-value indicating its significance. From these results we can observe that the higher an individual's positive PANSS score, and the more cognitive processing words they used, the more repeats in their dialogue, but when high-positive-score individuals spoke



(a) (b) (c)

Figure 4: An illustration of the structure of the moderation analyses with significant interaction effects described in Section 5.1.

Restarts =	β	t	p
Affect Words	+0.468	+1.443	0.155
Negative PANSS Score	+1.107	+2.947	0.005
Interaction Term	−1.020	−2.006	0.050

Repeats =	β	t	p
Cognitive Processing Words	+0.335	+1.116	0.270
Positive PANSS Score	+2.255	+2.168	0.035
Interaction Term	−2.171	−2.028	0.048

Edits =	β	t	p
Cognitive Processing Words	−1.278	−1.568	0.123
Negative PANSS Score	−0.572	−1.716	0.092
Interaction Term	+1.788	+2.070	0.044

Table 5: The regression models examining the moderation between PANSS scores and lexical categories as predictors of self-repairs.

about cognitive processing terms, they expressed *fewer* repeats than in general.

Negative symptoms, cognitive processing, edits. The third model involves negative PANSS score, cognitive processing words, and edits (see Figure 4c). In the first step of the regression analysis, negative PANSS score and cognitive processing words are entered as predictors of edits; this model significantly predicted edits ($F(50, 2) = 4.559$, $p = 0.015$, $r = +0.393$). In the second step of the analysis, the interaction term (the product of negative PANSS score and cognitive processing word use) was introduced; this model also significantly predicted edits ($F(49, 3) = 4.667$, $p = 0.006$, $r = +0.471$). This difference was statistically significant ($\Delta r = +0.078$, $p = 0.044$). See Table 5 for the final interaction model; β is the coefficient for each term, and t and p refer to a t-test value and p-value indicating its significance. From these results we can observe that the higher an individual's negative PANSS score, and the more cognitive processing words they used, the fewer edits in their dialogue, but when high-negative-score individuals spoke about cognitive processing terms, they expressed *more* edits than in general.

Discussion. There were three significant results observed during our moderation analysis. In particular, as individuals speak of specific topics, individuals with more severe symptoms tend to repair their language more or less often than in general. For example, individuals with high levels of negative symptoms were much less likely to restart their sentences when speaking about affective topics than in general, which may be explained by the blunted affect symptoms; it may be more straightforward for these individuals to speak about their emotions if they are not experiencing many of them. In another case, individuals with more severe positive symptoms were less likely to repeat themselves when speaking with cognitive processing terms, and individuals with more severe negative symptoms were more likely to edit themselves when speaking with cognitive processing terms. These three results are hinting to the fact that there are multi-faceted interactions in spoken language of individuals with psychotic disorders. Following these intuitions, we next learn multi-faceted prediction models.

5.2 Predictive Modeling

The final multi-faceted analysis consisted of the development of two sets of predictive models, one for each of the PANSS scales: positive and negative. Each model includes features that appeared as significant in the single-faceted analyses (see Section 4). For the positive PANSS scale, the features are the lexicon categories of power words and perceptual processing words, as well as perplexity. For the negative PANSS scale, the features are lexicon category of time words and the self-repair features of edits and restarts. As previously mentioned, all the single-facet analyses were performed on the training set, allowing for a fair evaluation of the prediction models on the test set (with new participants not in the training set).

Prediction experiments. We compare both ϵ-support vector machines [6] and multi-layer perceptron models [12] for prediction of PANSS scales. These models were trained using ten-fold cross-validation for hyperparameter tuning on the training set, optimizing upon the Pearson's r correlation coefficient. Hyperparameters included the kernel (linear or radial basis function), $C = \{10^{-5}, 10^{-4}, \ldots, 10^{4}\}$, $\epsilon = \{10^{-5}, 10^{-4}, \ldots, 10^{-1}\}$, and $\gamma = \{0.00, 0.05, \ldots, 1.00\}$ (in the case of the RBF kernel) for the support vector machines, and the number of hidden units ($\{1, 5, 10, 50, 100, 500\}$) and activation function (logistic, hyperbolic tangent, or rectified linear unit) in the multi-layer perceptron. Test set results are summarized in Table 6. The multilayer perceptron significantly outperformed the SVM in both cases ($p < 0.01$ in both cases according to a one-way ANOVA).

Feature analysis. To examine the relative importance of the included features in the multi-layer perceptron model, a greedy

PANSS Scale	SVM	MLP
Positive Scale	+0.570	+0.879
Negative Scale	+0.566	+0.710

Table 6: Average Pearson's r correlation coefficient achieved over ten-fold cross-validation, hold-out testing on prediction of positive and negative PANSS scores.

Positive Scale

	Top Predictive Features	Δr
1	power words	+0.406
2	perceptual processing words	+0.336
3	perplexity	+0.046

Negative Scale

	Top Predictive Features	Δr
1	self-repair edits	+0.330
2	time words	+0.262
3	self-repair restarts	+0.239

Table 7: A tabulation of the most significant features in each of the multi-faceted predictive models.

step-wise feature selection process was performed, using a ten-fold cross-validation procedure over the entire set[2]. At each iteration, candidate features were evaluated, and the single best feature to be added was selected via the highest average change in Pearson's r (Δr). Results are summarized in Table 7.

Discussion. In our predictive modeling analysis, we compared the performance of support vector machines (SVMs) and multi-layer perceptrons on a prediction task for positive and negative symptom severity. Although SVMs performed reasonably on both tasks, they were outperformed by multi-layer perceptrons in both cases. A higher performance was observed in predicting positive symptom severity, which may suggest that an individual's language use is more reflective of positive symptoms than negative symptoms in general. While positive scores were significantly predicted by lexical categories, negative scores were more significantly predicted by self-repairs. This may suggest that individuals with high negative scores have more difficulty in communication, while individuals with high positive scores are more characterized by what they speak about.

6 DISCUSSION AND CONCLUSIONS

Most psychiatric disorders are diagnosed with the aid of significant clinical evaluation of an individual's abnormalities in behavior patterns, but the complexity of the many different ways these disorders can manifest can limit this evaluation. Multimodal behavior analysis systems have the potential to fill this gap, but limited work has focused on the computational analysis of spoken language, despite psychological evidence for its pertinence. The present analysis approached language in three facets — through lexical, structural, and disfluency perspectives — and exposed a series of exciting results within each category as well as within interactions between them.

Words of power are heavily associated with positive symptom severity. Power words, such as *superiority*, *important*, and *exploit*, emerged as significantly predictive of positive symptom severity. The most characteristic symptoms of the positive scale involve delusions and grandiosity, which are defined by unfounded and exaggerated self-opinion and convictions of superiority, so the capability to detect these symptoms through language use is critical. Furthermore, the proportion of words of power used by an individual was the feature providing the most influence in a predictive model for positive symptom severity, above all other features.

Lack of relative language is highly indicative of more severe psychotic symptoms. Although much work has identified reality monitoring as a particular difficulty for individuals with psychotic disorders, little to no work has examined how this difficulty might be reflected in language use. Our analyses revealed that a lack of contextual language — relative words such as *yesterday*, *lately*, and *nearby* — is highly predictive of both positive and negative symptom severity. The fewer of these words an individual uses, the more severe their psychotic symptoms in general.

Linguistic difficulty during cognitive processing can be related to negative symptom severity. Although speaking in cognitive processing terms does not strictly indicate negative symptom severity, the higher an individual's negative symptom score, the more they will self-repair (and specifically edit their language) while speaking in cognitive processing terms. This behavior is often indicative of hesitation while constructing the sentences, so it may be representative of the cognitive difficulties characteristic of the negative psychotic symptom scale.

Future work will delve into more symptom-specific analyses, as each of the positive and negative scales are subdivided into measures of seven different symptom items. Augmenting these analyses with those of audio-visual modalities also holds great promise for improving the explanatory power of these models. Through these analyses we can achieve an even more nuanced characterization of psychotic disorders, which will constitute a significant step toward the design of future multimodal clinical decision support tools for computational phenotyping of mental illness.

[2]The full dataset was used in this step as a post-hoc analysis for feature importance.

ACKNOWLEDGMENTS

This material is based upon work partially supported by National Science Foundation Award #1722822 and the National Science Foundation Graduate Research Fellowship Program. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, and no official endorsement should be inferred.

REFERENCES

[1] 2018. RAISE Questions and Answers. https://www.nimh.nih.gov/health/topics/schizophrenia/raise/raise-questions-and-answers.shtml.
[2] J. J. Blanchard, K. T. Mueser, and A. S. Bellack. 1998. Anhedonia, positive and negative affect, and social functioning in schizophrenia. *Schizophrenia Bulletin* 24 (1998), 413–424. Issue 3.
[3] B. Buck, K. S. Minor, and P. H. Lysaker. 2015. Lexical Characteristics of Anticipatory and Consummatory Anhedonia in Schizophrenia: A Study of Language in

Spontaneous Life Narratives. *Journal of Clinical Psychology* 71 (2015), 696–706.

[4] A. S. Cohen, A. St-Hilaire, J. M. Aakre, and N. M. Docherty. 2009. Understanding anhedonia in schizophrenia through lexical analysis of natural speech. *Cognition and Emotion* 23, 3 (2009), 569–586.

[5] J. Cohen, P. Cohen, L. S. Aiken, and S. H. West. 2003. *Applied multiple regression/correlation analysis for the behavioral sciences.*

[6] Harris Drucker, Christopher J. C. Burges, Linda Kaufman, Alex J. Smola, and Vladimir Vapnik. 1997. Support vector regressor machines. In *Proceedings of the Ninth International Conference on Neural Information Processing Systems.*

[7] M. Fisher, K. McCoy, J. H. Poole, and S. Vinogradov. 2008. Self and other in schizophrenia: a cognitive neuroscience perspective. *American Journal of Psychiatry* 165 (2008), 1465–1472. Issue 11.

[8] K. Gabani, M. Sherman, T. Solorio, Y. Liu, L. M. Bedore, and E. D. Peña. 2009. A Corpus-Based Approach for the Prediction of Language Impairment in Monolingual English and Spanish-English Bilingual Children. In *Proceedings of the 2009 Annual Conference of the North American Chapter of the ACL: Human Language Technologies.*

[9] J. R. Garrison, E. Fernandez-Egea, R. Zaman, M. Agius, and J. S. Simons. 2017. Reality monitoring impairment in schizophrenia reflects specific prefrontal cortex dysfunction. *NeuroImage: Clinical* 14 (2017), 260–268.

[10] J. J. Godfrey, E. C. Holliman, and McDaniel J. 1992. SWITCHBOARD: telephone speech corpus for research and development. In *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*. 517–520.

[11] C. Goodwin. 1979. The interactive construction of a sentence in natural conversation. In *Everyday Language: Studies in Ethnomethodology*. 97–121.

[12] T. Hastie, R. Tibshirani, and J. Friedman. 2009. *Elements of Statistical Learning: Data Mining, Inference, and Prediction.*

[13] K. Hong, C. G. Kohler, M. E. March, A. A. Parker, and A. Nenkova. 2012. Lexical Differences in Autobiographical Narratives from Schizophrenic Patients and Healthy Controls. In *Proceedings of the 2012 Joint Conference on Emperical Methods in Natural Language Processing and Computational Natural Language Learning.*

[14] J. Hough and D. Schlangen. 2017. Joint, Incremental Disfluency Detection and Utterance Segmentation from Speech. In *Proceedings of the 15th International Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers.*

[15] L. C. Johns, S. Rossell, C. Frith, F. Ahmad, D. Hemsley, E. Kuipers, and P. McGuire. 2001. Verbal self-monitoring and auditory verbal hallucinations in patients with schizophrenia. *Psychological Medicine* 31 (2001), 705–715.

[16] M. K. Johnson and C. L. Raye. 1981. Reality monitoring. *Psychological Bulletin* 88 (1981), 67–85.

[17] D. U. Junghaenel, J. M. Smyth, and L. Santner. 2008. Linguistic Dimensions of Pscyhopathology: A Quantitative Analysis. *Journal of Social & Clinical Psychology* 27 (2008), 36–55. Issue 1.

[18] J. H. Kahn, R. M. Tobin, A. E. Massey, and J. A. Anderson. 2007. Measuring emotional expression with the Linguistic Inquiry and Word Count. *American Journal of Psychology* 120 (2007), 263–286. Issue 2.

[19] S. R. Kay, A. Fiszbein, and L. A. Opler. 1987. The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophrenia Bulletin* 134 (1987), 261–276. Issue 4.

[20] R. S. E. Keefe, M. C. Arnold, Bayen U. J., J. P. McEvoy, and W. H. Wilson. 2002. Source-monitoring deficits for self-generated stimuli in schizophrenia: multinomial modeling of data from three sources. *Schizophrenia Research* 57 (2002), 51–67. Issue 1.

[21] I. Leudar, P. Thomas, and M. Johnston. 1992. Self-repair in dialogues of schizophrenics: Effects of hallucinations and negative symptoms. *Brain and Language* 43 (1992), 487–511. Issue 3.

[22] W. Levelt. 1983. Monitoring and self-repair in speech. *Cognition* 14 (1983), 41–104. Issue 1.

[23] R. McCabe, P. G. T. Healey, S. Priebe, M. Lavelle, D. Dodwell, R. Laugharn, A. Snell, and S. Bremner. 2013. Shared understanding in psychiatrist-patient communication: Association with treatment adherence in schizophrenia. *Patient Education and Counseling* 93 (2013), 73–79. Issue 1.

[24] M. Mitchell, K. Hollingshead, and G. Coppersmith. 2015. Quantifying the language of schizophrenia in social media. In *Proceedings of the Second Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality.*

[25] K. T. Mueser, A. S. Bellack, M. S. Douglas, and R. L. Morrison. 1991. Prevalence and stability of social skill deficits in schizophrenia. *Schizophrenia Research* 5 (1991), 167–176. Issue 2.

[26] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn. 2015. *The development and psychometric properties of LIWC2015.*

[27] C. R. Rao. 1973. *Linear statistical inference and its applications.*

[28] E. Schegloff, G. Jefferson, and H. Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language* 53 (1977), 361–382. Issue 2.

[29] E. Shriberg. 1994. *Preliminaries to a Theory of Speech Disfluencies.* Ph.D. Dissertation. University of California, Berkeley.

[30] T. Solorio, M. Sherman, Y. Liu, L. M. Bedore, E. D. Peña, and A. Iglesias. 2011. Analyzing language samples of Spanish-English bilingual children for the automated prediction of language dominance. *Natural Language Engineering* 17 (2011), 367–395. Issue 3.

[31] A. K. Vail, T. Baltrušaitis, L. Pennant, E. Liebson, J. Baker, and L.-P. Morency. 2017. Visual Attention in Schizophrenia: Eye Contact and Gaze Aversion during Clinical Interactions. In *Proceedings of the 7th International Conference on Affective Computing and Intelligent Interaction.* 490–497.

[32] T. Wörtwein, T. Baltrušaitis, E. Laksana, L. Pennant, E. S. Liebson, D. Öngür, J. T. Baker, and L.-P. Morency. 2017. Computational Analysis of Acoustic Descriptors in Psychotic Patients. In *Proceedings of the 2017 Annual Conference of the International Speech Communication Association.* 3256–3260.