January 2016

Language In Schizophrenia: What We Can Learn From Quantitative Text Analysis

Sasha Deutsch-Link
Yale University, sasha.deutsch-link@yale.edu

Language in Schizophrenia: What We Can Learn from Quantitative Text Analysis

A Thesis Submitted to the
Yale University School of Medicine
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Medicine

By

Sasha Deutsch-Link

2016

LANGUAGE IN SCHIZOPHRENIA: WHAT WE CAN LEARN FROM QUANTITATIVE TEXT ANALYSIS

Sasha Deutsch-Link, Sarah Fineberg, Philip R. Corlett. Department of Psychiatry, Yale University School of Medicine, New Haven, CT.

Background: People living with schizophrenia demonstrate broad language and communication deficits. Prior research has focused on qualitative changes in thought, speech production, and language comprehension. In this thesis I expand upon our understanding of language in schizophrenia, by using a novel, quantitative method for language analysis, Linguistic Inquiry Word Count (LIWC).

Methods: I examined essays by four groups of authors: people with schizophrenia (n=77), family history of schizophrenia (n=25), psychiatric controls (mood/anxiety, n=29), and non-psychiatric controls (college students, n=418). Essays were then processed using Linguistic Inquiry Word Count (LIWC), which organizes language into parts of speech and pre-defined content themes and then calculates percent word type used. I performed multiple ANOVAs to examine group differences followed by Tukey post-hoc and FDR correction for multiple variables. Standard of significance was determned at $p < .05$. Afterwards, I performed a cluster analysis using MClust in R to examine whether patients living with schizophrenia tended to cluster with people with a family history of schizophrenia.

Results: Mood disorder essays used more affective language and "I." Essays written by patients with schizophrenia used more external referential language ("humans" and "religion") and less "I" than controls. Family members used less "I" and more religion-related words than controls, but had similar reference to other humans. In function word use, schizophrenia and family member essays used more articles, and fewer pronouns than controls. Schizophrenia and family member essays also contained more perceptual language.

Conclusions: Important differences emerged between schizophrenia essays, family member essays and controls. Decreased "I" and increased external referential language likely reflect loss of agency/power in schizophrenia. Decreased pronoun use in schizophrenia and family members likely reveals a degree of social isolation or withdrawal. The cluster analysis demonstrated that family members clustered with the schizophrenia group, suggesting a possible intermediate language phenotype. These findings can be expanded upon in future studies by analyzing spoken and/or unedited speech, controlling for demographic variables, and by using more standardized essay prompts.

Acknowledgements

I would like to thank my mentors Dr. Philip Corlett and Dr. Sarah Fineberg for their guidance and support throughout this process and for their incredible academic mentorship throughout all four years of medical school, helping me foster my passion and curiosity within the field of psychiatry and medicine.

I would like to thank other members of the BLAM lab for their help, including Emily Finn (for her linguistic expertise) and Adina Bianchi (for her incredible research support in lab).

I would also like to thank Dr. James Pennebaker and Dr. Cindy Chung at the University of Texas at Austin, who are influential researchers in language analysis, and whose research methods (especially Linguistic Inquiry Word Count) have inspired the work for the thesis and beyond.

I appreciate the time Dr. Marc Potenza and Dr. Sarah Yip took to read and provide feedback, helping further strengthen my thesis and particularly my third chapter for potential publication.

Finally, I would like to thank my family and friends for their moral support and unwavering confidence in my abilities.

Table of Contents

Chapter 1: Introduction

Background

Language is one of the primary means by which we engage with our social environment. Nearly all mammals have some form of language, though none exhibit greater complexity or nuance than humans. The incredible breadth and precision of human language offers amazing advantage, and is the reason we are able to live in such an interconnected global community. However, in the complexity of our language, we find unique vulnerability. Phenomena such as a hearing impediment, a congenital defect, brain damage, or mental illness can compromise the fine-tuned system that produces and comprehends fluent language. Of these, one particular mental illness, schizophrenia, can produce global, and heterogeneous deficits in language. These deficits have been studied for several decades, and have been analyzed qualitatively in the work of Andreasen and others (Andreasen, 1979). However, until recent, quantitative analyses of language in schizophrenia have been limited. This thesis uses novel quantitative word-counting software, Linguistic Inquiry Word Count (LIWC) (Pennebaker, 2007), to examine one of the most fundamental components of language – the words themselves. By counting different parts of speech (articles, pronouns, verbs, etc), and theme categories (words relating to negative emotions, religion, social environment, etc), we can gain valuable insight into how language differences reflect the experience and phenomenology of schizophrenia.

Schizophrenia is a common disease with a prevalence of about .7%, or 7 in 1000 (McGrath, Saha, Chant, & Welham, 2008). To put the commonality of schizophrenia in perspective, this rate is almost four times the global cases of Alzheimer's Disease (Hebert, Weuve, Scherr, & Evans, 2013), and six times the prevalence of type 1 diabetes (Shaw,

Sicree, & Zimmet, 2010). In addition to being remarkably common, schizophrenia also has devastating impacts. Schizophrenia is associated with a markedly increased rate of morbidity and mortality (Goff et al., 2005). In Sweden, the life expectancy of someone with a diagnosis of schizophrenia is 62.6 years for men and 70.5 years for women, compared to 77.6 years and 82.5 years in for the general population (Crump, Winkleby, Sundquist, & Sundquist, 2013). Though no formal life expectancy comparison has been published in the United States recently, given a less comprehensive safety net system and greater measures of inequality (Ezzati, Friedman, Kulkarni, & Murray, 2008), we can assume the United States must have even poorer measures. In terms of disability, schizophrenia is the 8th leading cause of disability worldwide, comprising 1.1% of all disability-adjusted life years (DALYs) (Theodoridou & Rössler, 2010). Though these hard-numbers are impactful and quite staggering, they don't begin to explain the qualitative distress for the patients living schizophrenia or for loved ones who are often caregivers (Barrowclough, Tarrier, & Johnston, 1996; Selten, Wiersma, & van den Bosch, 2000).

According to the DSM-V, schizophrenia is diagnosed by a series of symptom clusters based on careful patient population study. Symptoms are generally broken down into positive and negative symptoms. Positive symptoms include delusions, hallucinations, and disorganized behavior or speech. They generally represent more apparent and outward symptoms of psychosis. Negative symptoms include apathy, social withdrawal, cognitive deficits and poverty of speech. As one can see, disorganized speech and poverty of speech are mentioned in the diagnostic thread of schizophrenia in the DSM-V (APA, 2013). This thesis will further explore how language is impacted in schizophrenia, and how it may relate to both negative and positive symptom clusters.

Linguistic Perspectives on Language Production in Schizophrenia:

One of the earliest accounts of language differences in schizophrenia dates back to the 1960s. In 1961, psychiatrist Dr. Maria Lorenz published a paper describing several language differences she had observed working clinically with patients suffering from schizophrenia. As a psychiatrist, and not a linguist, she wrote: "We are faced with the paradox that while we recognize schizophrenic language when we see it, we cannot define it" (Lorenz, 1961). In the ensuing decades, linguists, psychiatrists, psychologists and other scientific disciplines would work towards "defining" language in schizophrenia.

In the formal analysis of language in schizophrenia, we can subdivide differences into a few core features: lexicon, phonology, morphology, syntax, semantics and pragmatics (Covington et al., 2005; McGregor, 2015). Several aspects of language in schizophrenia seem to draw from each of these categories simultaneously, with certain language features straddling two or more different arenas.

Phonology, which represents the sound system of language, is, for the most part, unaffected in schizophrenia (Covington et al., 2005). Phonology includes all individual sounds in words, such as the "p" in the word "pot." These sound units seem to be entirely preserved in schizophrenia (Lecours, 1974), (Covington et al., 2005). This may surprise some readers, as sometimes language produced by patients with schizophrenia can seem quite unintelligible. However, the "unintelligible" gestalt one perceives, is actually rooted in several other language deficits that make this type of speech difficult to follow.

One interesting finding in schizophrenia closely related to phonology, is the concept of clanging. Clanging refers to a pattern of speech driven by sound relationships between words, rather than the actual content. One example from a patient interviewed in a 1979

Andreasen paper demonstrates clanging: “I’m not trying to make noise. I’m trying to make sense. If you can make sense out of nonsense, well, have fun. I’m trying to make sense out of sense. I’m not making sense (cents) anymore. I have to make dollars” (Andreasen, 1979). Here, we see the word “sense,” is confused for “cents,” which has an identical sound. Then, we observe the topic of conversation shift toward money, and suddenly the patient is talking about dollars. Although not an overt deficit in phonology, (“sense” phonologically is the same as “cents” and is pronounced correctly in this example), the concept of clanging demonstrates that phonology is, in fact, an important feature to examine in schizophrenia.

Another language feature in schizophrenia that may fall within the realm of phonology is echolalia. Echolalia, deriving from the word echo, refers to a pattern of speech in which the patient will repeat words or phrases of the interviewer. When present it is often repetitive and persistent (Andreasen, 1979). An example would be an interviewer asking a patient “How are you?” and patient responding with “How are you?” Here the phonology is important, because echolalia is driven by the sounds of words and short phrases rather than structure or content. Interestingly, we see echolalia in other brain disorders, most notably in stroke or traumatic brain injury involving frontal lobe damage (Grattan & Eslinger, 1991).

Deficits in morphology, like basic deficits in phonology, are also uncommon. Morphology refers to the structure of words broken down into morphemes, lexemes and other linguistic units such as root words. Morphemes are the most basic unit of language that have meaning, so for instance “town hall” can be broken down into two morphemes – “town” and “hall.” A lexeme allows for the extension of morphemes to represent a new meaning – so “town hall” would be considered its own lexeme, as both town and hall have separate meanings, but when put together create a new one. (McGregor, 2015) Chaika, a linguist who

studied language in schizophrenia for decades, describes the possibility for morphological errors in schizophrenia, with a patient substituting the word “medicate” for “medicine,” (E. O. Chaika, 1990). However, we cannot necessarily deem this a morphological error. Medicate is a correct word, albeit in a separate context, so this may reflect a word-finding difficulty or perhaps a syntactical error (as a wrong part of speech, using the verb instead of the noun).

In linguistics, syntax refers to the set of rules that dictate sentence structure (McGregor, 2015). In other words, it is how we put together sentences: “The young boy went to the store” is a syntactically correct sentence, however, the “The boy young the to went store,” might convey the correct meaning to most listeners if they take extra time to interpret it, but the syntax is incorrect. Another famous quote by linguist Noam Chomsky offers a syntactically perfect sentence with little meaning: “Colorless green ideas sleep furiously” (Chomsky, 1956). Here, all parts of speech are correctly aligned, but the sentence does not really mean anything – how can green be colorless or ideas sleep furiously? In general, language schizophrenia is not characterized by overt syntactic errors (Covington et al., 2005). However, we do see some syntactical differences that deserve mention.

Through the careful analysis of transcribed interviews and qualitative syntactic coding, Fraser et al demonstrated that patients with schizophrenia exhibit an overall simplified, albeit correct, syntactical structure to their language when compared to controls (Fraser, King, Thomas, & Kendell, 1986). Research conducted by Thomas et al replicated this finding, and conducted additional analysis on syntactical complexity and how it correlates with negative and positive symptoms. They observed strong correlations between negative symptoms and syntactical simplicity, but no association with positive symptoms

(Thomas, King, & Fraser, 1987). Further research has shown that the reduction in syntactical complexity is associated with chronicity of schizophrenia diagnosis (Thomas, King, Fraser, & Kendell, 1990), suggesting a dose response effect: the longer the exposure to schizophrenia, the more that language, as measured by syntactical complexity, is affected.

Lexicon is another component of language affected in schizophrenia. Lexicon refers to the vocabulary, or subset of words, in one's repertoire (McGregor, 2015). Initial lexical analyses of language in schizophrenia have shown an overall reduction in volume of speech. In 1979 Andreasen developed The Scale for Thought, Language and Communication, an 18-item scale to evaluate thought disorder and language in psychosis (Andreasen, 1979). Using this scale, Andreasen & Grove conducted a study looking at language in schizophrenia compared to control groups (Andreasen & Grove, 1986). They observed deficits in the volume of language produced in schizophrenia, or "poverty of speech," as well as diminished actual content of the language that was produced, called "poverty of content of speech" (Andreasen & Grove, 1986). Further quantitative research has shown an overall reduction in the number of distinct words used by people with schizophrenia, even after controlling for total number of words used (Allen, Liddle, & Frith, 1993). Once poverty of speech was established, the next question became whether this could be attributed to a problem in one's pool of known words, or rather to a word-finding difficulty (a word accessing problem). Experiments conducted by Allen et al demonstrated that known vocabulary was normal in schizophrenia, but rather the access to that language and the ability to produce it was impaired (Allen et al., 1993), or in other words, a dysfunction in word fluency.

Word-finding difficulties are not only apparent in the diminished volume of words produced, but can also be exemplified in word approximation. In McKenna's book

Schizophrenia and Related Syndromes, he describes this phenomenon: "*Metonymic distortion* (word approximation)… in which imprecise verbal constructions are substituted for common-place words" (McKenna, 2007). As we discussed in the prior paragraph, patients with schizophrenia exhibit word-finding difficulties. This can be seen in poverty of speech, but is also shown in their tendency to replace those words they have difficulty accessing with words that have similar meaning, but are imprecise, or incorrect. In Andreasen's Thought, Language and Communication scale, she discusses examples of word approximations, such as the word "hand shoes" being used to describe gloves or "time vessel" instead of the word "watch" (Andreasen, 1979). Here, the actual meaning is mostly intact, as we are able to understand what hand-shoes refer to, especially in context. However, the words used are atypical and imprecise.

Neologisms are another way that word-finding difficulties may manifest. Neologisms are novel words not part of a language's appropriate dictionary or common vernacular. We have not yet determined whether neologisms truly derive from word-finding difficulties, or whether they represent a unique phenomena in schizophrenia. Some examples of neologisms from research studies have included a spectrum ranging from words that are somewhat comprehensible such as "crusady," an adjective derived from the word crusade, to completely bizarre such as "tarn-harn" or "geshinker," which we cannot relate to any real word or concept without more context (Andreasen, 1979; McKenna, 2007).

Semantics is another subset of linguistics that deserves mention in the study of language in schizophrenia. Semantics describes the meaning of words and phrases (McGregor, 2015). The initial example of syntactical error "the boy young the to went store," demonstrates an example in which semantics are preserved, despite a sentence filled with

structural (syntactic) error. Dysfunction in semantics has been well-established in schizophrenia (Oh, McCarthy, & McKenna, 2002; Rodriguez-Ferrera, McCarthy, & McKenna, 2001). A case study published in 2002 by Oh et al provides an excellent example of semantic disruption. When the patient in this study was answering a question about what life in the hospital was like, he patient replied: "Oh it was superb, you know, the trains broke, and the pond fell in the front doorway" (Oh et al., 2002). Here, we see clear semantic error – a pond cannot fall in a doorway, and if it could, how would this answer the question of what life was like in a hospital? It appears the sentence does not actually convey anything meaningful to a general audience. This is also reminiscent of Chomsky's quote "colorless green ideas sleep furiously," an example of true syntactic precision, but overt semantic error (Chomsky, 1956).

The most severe form of semantic dysfunction in schizophrenia manifests as "word salad." This phenomenon refers to a patient speaking in a manner that is voluminous and completely incomprehensible. It is rare, but when observed, can be quite remarkable. Andreasen offers one example of this from a patient interview: "Interviewer: 'What do you think about current political issues like the energy crisis?' Patient: 'They're destroying too many cattle and oil just to make soap. If we need soap when you can jump into a pool of water, and then when you go to buy your gasoline, my folks always thought they should get pop" (Andreasen, 1979). Here, the words seem to be strung together with no meaning or relation to one another. It is similar to the quote from Oh et al, but more voluminous and dysfunctional.

Pragmatics is a relatively newer field of linguistic study that also deserves consideration. Pragmatics encompasses how context contributes to meaning (Covington et

al., 2005). Context is interpreted broadly and includes aspects such as the tone, the people having the conversation, the topic, the setting and the flow of language (what comes before and after it, and how). Pragmatics concerns some of the most complex features of human language, so it comes as no surprise that it is the area most disordered in schizophrenia.

Tone, one component of pragmatics, refers to the pitch at which we deliver our speech. Also referred to as prosody, tone can turn "great job" from a compliment to an insult. The ability to deliver and interpret intonation correct is incredibly important in social interactions. Historically, patients with schizophrenia have often been described as having a flat, or even monotone voice (Stein, 1993). In Chen et al's clinical language scale (CLANG), lack of intonation was shown to be one of the key features of language in schizophrenia (Chen, Lam, Kan, & Chan, 1966). The deficits in tone production extend further, as research indicates people with schizophrenia also have more trouble comprehending tone and inflection (Rieber & Vetter, 1994).

Circumstantiality and tangentiality have been well-documented as core features of language in schizophrenia (Andreasen, 1979) and can be attributed to errors in pragmatics. Circumstantiality refers to the tendency of speech to veer away from the main content, with still reaching the ultimate goal of the argument or point being made. It represents a circuitous path to expressing a general idea. Tangential speech similarly refers to the tendency for speech to deviate from the main topic, however the original theme is not maintained. It also represents a circuitous path, but the patient never comes back to the initial topic or concept, and the goal of the initial argument is lost. Circumstantiality and Tangentiality represent errors in pragmatics because the actual meaning of each sentence might still be intact, but they don't necessarily flow logically from one step to the next, and context is lost. These

features have both been demonstrated the speech in schizophrenia patients (Hoffman, Stopek, & Andreasen, 1986). Notably, Hoffman et al created a sentence-by-sentence diagram portraying linkages between one word and the next. They found impaired coherence in the language of schizophrenia and an overall lack of contextual structure.

Derailment, also called "loose associations" or "flight of ideas" is another pattern of speech that falls within the realm of pragmatics. Derailment refers to a type of speech in which a patient will constantly change from one idea to the next – seemingly making associations between these ideas. Sometimes this string of ideas forms a linear pattern, and a listener is able to logically follow why one idea led to the next, however some seem to be completely unrelated. An example from Andreasen's Thought Language and Communication Scale demonstrates what derailment sounds like:

> "—Interviewer: 'What did you think of the whole Watergate affair?' Patient: 'You know I didn't tune in on that, I felt so bad about it. I said, boy, I'm not going to know what's going on in this. But it seemed to get so murky, and everybody's reports were so negative. Huh, I thought, I don't want any part of this, and I was I don't care who was in on it, and all I could figure out was Artie had something to do with it. Artie was trying to flush the bathroom toilet of the White House or something. She was trying to do something fairly simple. The tour guests stuck or something. She got blamed because of the water over¬ flowed, went down in the basement, down, to the kitchen. They had a, they were going to have to repaint and restore the White House room, the enormous living room.'" (Andreasen, 1979)

In this quote we observe a loosely associated string of ideas, with idea seeming to develop from the next, but not following a clear, coherent flow. The error is not in the sentences themselves, but rather in how they fit together (their context). The presence of loose associations, like those above, has been strongly linked to positive symptoms, such as delusions and hallucinations (Kay, Flszbein, & Opfer, 1987). The aberrant associations seen in flight of ideas, may relate to the involvement of abnormal perceptual associations in delusion formation (Corlett, Honey, & Fletcher, 2007).

Two more components of language within schizophrenia deserve mention without falling neatly into any of the aforementioned linguistic categories – blocking and distractible speech. Blocking occurs when a patient interrupts his/her flow of speech when the thought or idea has not yet been completed. This very noticeable speech quality likely reflects the array of cognitive symptoms people with schizophrenia experience. Distractible speech is somewhat similar, but the situation in which topic is lost secondary to a distraction. In distractible speech, a patient will often stop in the middle of a sentence and thought, and change the subject in response to some sort of stimulus in the room.

Within the realm of language production in schizophrenia, we see several deficits and differences in the areas of phonology, lexicon, syntax, semantics, and pragmatics. However, we must also examine language comprehension. Although production and comprehension are discussed here separately for the sake of organization, they are inextricably intertwined - different manifestations of a common underlying language deficit.

Language Comprehension in Schizophrenia:

My discussion of language in schizophrenia thus far has focused on language production, however in the arena of language comprehension, we also notice robust deficits. Most research, to date, has focused on the ability of patients with schizophrenia to understand complex language, and abstract concepts such as metaphor, proverb or idiom.

In the previous section, I mentioned that patients with schizophrenia tend to produce sentences with a simplified syntactical structure. Interestingly, in the study of language comprehension, we see a diminished capacity for *understanding* complex sentence structure. In 2002, Chondray et al analyzed the ability of patients with schizophrenia to understand

sentences with increasing syntactical complexity (Condray, Steinhauer, van Kammen, & Kasparek, 2002). Compared to controls, people diagnosed with schizophrenia had more difficulty interpreting sentences as the syntax became more complex. The discrepancy between the schizophrenia and control groups widened as syntax became more complicated.

The ability to comprehend abstract concepts also seems to be impaired in schizophrenia. Patients with schizophrenia tend to choose incomplete, concrete interpretations of metaphor, irony and proverbs (Kuperberg, 2010). Studies by Brune et al and Kiang et al have assessed the ability of patients with schizophrenia to interpret proverbs, such as: *"you can't judge a book by its cover"* (Brüne & Bodenstein, 2005; Kiang et al., 2007). Both studies found significant deficits in proverb interpretation characterized by an overly concrete, or literal, interpretation of the proverb. Accordingly, deficits are also seen in the ability to interpret irony (Mo, Su, Chan, & Liu, 2008) and metaphor (de Bonis, Epelbaum, Deffez, & Féline, 1997).

Language Analysis: A Quantitative Approach

Much of my discussion thus far has focused on qualitative methods for analyzing language and schizophrenia. In these instances, language is often transcribed and then coded using a linguistic framework (such as Andreasen's Thought, Language and Communication scale, or Chaika's intensive linguistic case study) (Andreasen, 1979; E. Chaika, 1974). The information gathered from these methods has been invaluable in helping us understand and characterize language deficits in schizophrenia, however new methods now exist that can help us delve deeper.

Novel, quantitative approaches to language analysis made possible by computers and their ever-expanding processing capacity, have added another layer to our study of language in schizophrenia. We now have dozens of quantitative software programs in our toolbox, ranging from the analysis of the ways in which words are put together [Crawdad Technologies (Corman & Dooley, 2006)] to actual word counting [programs such as DICTION (Hart, 2000) or Linguistic Inquiry Word Count (J. W. Pennebaker, Francise, M.E., & Booth, J.R.,, 2007)]. In thesis, I employed the use of the word counting software program, Linguistic Inquiry Word Count.

Linguistic Inquiry Word Count or LIWC, was developed by James Pennebaker and colleagues at the University of Texas Austin as an objective approach to analyzing language (J. W. Pennebaker, Francise, M.E., & Booth, J.R.,, 2007). This approach to language analysis reduces language to some of its most fundamental components – the actual words themselves. Using 4,500 words, word stems and emoticons, this program calculates percentage of certain types of words in a given sample. These include common words (I, we, me), parts of speech (articles, nouns, etc), and content categories (sadness, religion, biological processes). (See appendix for word categories)

Analyses using LIWC have been applied across various settings to help us better understand how language is influenced by changes in internal state and the surrounding environment. LIWC can be used to analyze speeches, essays, articles and even internet blogs to characterize language changes in light of personal or shared tragedy, depression and schizophrenia.

For instance, an analysis of former New York City Mayor Guiliani's speeches before and after a period of severe emotional turmoil for him (prostate cancer diagnosis, divorce and

withdrawing from his senate race against Hilary Clinton), demonstrated a clear upswing in the use of personal referents (I, me), going from 2% to 7% of total words (J. W. Pennebaker & Lay, 2002). This aligns with prior researching showing that increased use of the first person singular is associated with negative affect states (Weintraub, 1989).

There is also significant data on language changes after shared upheaval. After the events of 9/11, blogs and other types of personal posting on the internet were extracted and analyzed for language changes. Cohn et al showed that in the immediate aftermath of 9/11, participants expressed more negative emotion content words and more words indicating social engagement (social pronouns such as he/she), indicating a sense of camaraderie and overall feeling of togetherness. After 6 weeks, however, social referencing (she/he) decreased even more than baseline, possibly reflecting a more psychologically distanced state. (Cohn, Mehl, & Pennebaker, 2004).

LIWC analyses have also been used to examine language changes in depression. One analysis performed by Rude et al in 2004, examined college student essays combined with data on depression status (Rude, Gortner, & Pennebaker, 2004). They found that students who currently met criteria for major depressive disorder used more personal referents (I, me) than their non-depressed counterparts. In fact, even a history of depression was associated with this pattern, suggesting depression may have a lasting impact on language. Personal referent use has such a strong relationship with depression, that another analysis by Mehl et al demonstrated that personal referential language was even more predictive of depression that use of negative emotion words (Mehl 2004).

Until very recently, word-counting analyses had not been used in schizophrenia. In the recent years, however, LIWC and other quantitative methods have been used to examine

language in schizophrenia, giving us a new lens for assessment. Studies performed by Buck et al and Minor et al employed LIWC to examine cohorts of patients with schizophrenia (with no comparison group), and determine how certain language markers correlated with overall functioning (Buck, Minor, & Lysaker, 2015; Minor et al., 2015). Buck et al showed that language complexity was associated with cognitive capacity, and that function word use was positively correlated with social cognition. Minor et al showed that anger words were associated with lower functioning, and social words were related to higher levels of cognitive functioning. Together, these studies show us that quantitative language analysis in schizophrenia can give us valuable information about the clinical status and cognitive functioning of patients with schizophrenia.

Thus, the capacity for language analysis in schizophrenia is greater than ever before. We have a rich history of qualitative linguistic analysis that provides a strong foundation and framework for how we approach our study. And with the advent of impressive technology and quantitative software programs, we have an amazing opportunity to better understand how schizophrenia disrupts the human language system that is so fundamental to our being. The purpose of this thesis is to use one of these novel quantitative methods, Linguistic Inquiry Word Count (LIWC), to compare word use in patients with schizophrenia to that of patients with mood disorders, schizophrenia family members and a healthy control group. From this comparison, I hope to glean important distinctions that help us better understand how language is uniquely affected in schizophrenia, and what implications that may bring to a better understanding of pathophysiology and clinical outcomes.

Chapter 2: LIWC Analysis of Language in Schizophrenia Compared to Mood Disorders and Controls*

Background:

As discussed in the introduction, language serves a vital role in the social lives of humans, and can often be disrupted in mental illness, particularly schizophrenia. Until recent, our analysis of language in schizophrenia has been qualitative and linguistic in nature. However, with the advent of greater technology and processing power, we have the capability to analyze language in schizophrenia on a more fundamental level than ever before – by looking at the words themselves.

Linguistic Inquiry Word Count has been used to analyze language across various emotional states, socioeconomic class, gender, and depression (Chung & Pennebaker, 2007). As mentioned above, it has also been used to correlate language with cognitive and social functioning in schizophrenia (Buck et al., 2015; Minor et al., 2015). One of the only papers using LIWC to compare language in schizophrenia to a control group was published by Hong et al in 2015 (Hong et al., 2015). In this study, Hong and colleagues collected short (30-90 second) autobiographical narratives from 120 patients with schizophrenia and 80 controls, and analyzed the text samples using LIWC. The results showed that patients with schizophrenia used a similar number of overall words, but the speech included more self-reference and repetitions, and had an overall reduction in language complexity.

---

* This chapter has been published as a single journal article: Fineberg, S. K., Deutsch-Link, S., Ichinose, M., McGuinness, T., Bessette, A. J., Chung, C. K., & Corlett, P. R. (2015). Word use in first-person accounts of schizophrenia. *The British Journal of Psychiatry*, *206*(1), 32-38.

To complement the findings by Hong et al, I examined written autobiographical narratives, and included a more active control group – mood disorder patients. As Andreasen has shown in her Thought, Language and Communication Scale analysis of psychosis and mania, some features of language in schizophrenia are shared by other mental illnesses (Andreasen, 1979). Therefore, my aim was two-fold. First, I wanted to examine whether certain language features were common to mental illness in general, given that features such as isolation, impaired functioning and suffering are common amongst almost all spectrums of disease. And second, I wanted to separate out the components of language in schizophrenia that were unique to schizophrenia (and not simply a result of mental illness).

Research Hypotheses

H1: I hypothesized that the mood disorder group would use more affect-related language and personal referents compared to schizophrenia and control groups. Words relating to affect are categorized in LIWC into "affect" (happy, cry), "positive emotion" (love, nice), "negative emotion" (hurt, ugly), "anger," "sadness," and "anxiety." Because by definition, mood disorders are disorders of emotion and affect, I thought these words would be represented in higher frequency in their essays. Then, to further validate the sample, I expected to replicate a previous study showing that depression is associated with more personal referents, such as "I" and "me" (Rude et al., 2004).

H2: Schizophrenia is often associated with a disrupted sense of self and agency (Hemsley, 1998; Synofzik, Thier, Leube, Schlotterbeck, & Lindner, 2010). Given this, I expected to see differences in personal referent use (I) in schizophrenia samples compared to control and mood-disorder essays. Given the often increased salience of external forces seen in

schizophrenia, I also expected to see an increase in external referent use (they, them, humans).

H3: I expected to observe shared language features amongst the schizophrenia and mood disorder groups. By definition in the DSM-V, one must display impaired functioning to be diagnosed with depression, anxiety or schizophrenia (APA, 2013). Also, people diagnosed with a mental illness often report higher levels of distress and isolation (Ziskind, 1958). Given these shared experiences across mental illness, I expected to find unifying features in language related to: social processes (family, friends) and personal concerns (work, achievement, leisure).

H4: Because schizophrenia is often characterized by strange perceptual experiences, I expected this to be reflected in language. The most common form of perceptual disturbance is auditory hallucinations, but patients can also experience those that are visual, tactile and olfactory (Mueser, Bellack, & Brady, 1990). I therefore hypothesized that schizophrenia writing would contain more language related to perceptual experiences, with auditory perceptions (words related to hearing) at the forefront.

Methods:

Data Sources*:*

I examined a total of 525 first-person essays written by three groups of authors: 77 by people diagnosed with schizophrenia from the first person accounts series in Schizophrenia Bulletin, 29 mood disorder (our psychiatric "control") from mood disorder journals and advocacy association websites, and 418 control essays written by college students. The essay prompt

for schizophrenia and mood disorder patients was to write about their experience living with mental illness, and the control essays were written by college students about the transition to college. Because schizophrenia and mood disorder essays were published and publicly available, and the college essays had already been processed using LIWC by the Pennebaker Lab at the University of Texas – Austin and had no identifiers, the study was granted exemption by the Humans Investigations Committee (HIC) at Yale.

Language Processioning:

To analyze the essays, I used Linguistic Inquiry Word Count (LIWC), a text analysis software that calculates the degree to which people use distinct words, word categories or parts of speech. LIWC's dictionary includes over 4,500 words, word stems and emoticons. For each dictionary word, LIWC includes a corresponding dictionary entry that defines one or more word categories. For instance, "the word *cried* is part of five word categories: sadness, negative emotion, overall affect, verbs, and past focus. Hence, if the word cried is found in the target text, each of these five subdictionary scale scores will be incremented" (J. W. Pennebaker, Chung, Ireland, Gonzales, & Booth, 2015). The development of the LIWC dictionary involved word collection using common emotion rating scales, Roget's thesaurus and standard English dictionaries. Words were preliminarily placed into categories, and independent reviewers evaluated them, needing a majority agreement for them to be placed in the final LIWC edition. LIWC was then validated using qualitative coding of texts involving different emotional states (J. W. Pennebaker et al., 2015). (See appendix for word categories)

Data Analysis:

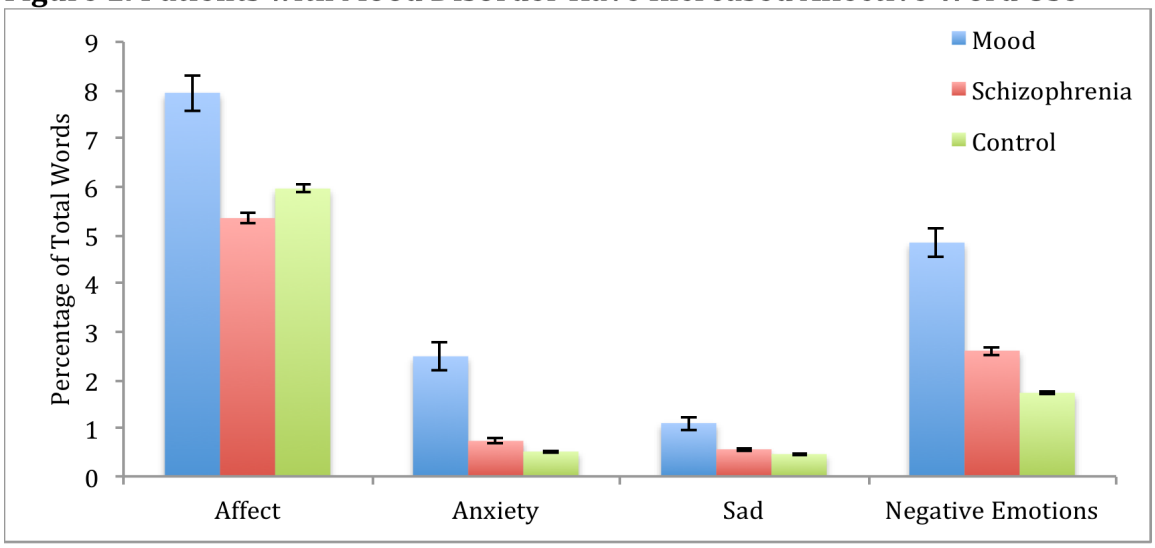
I used One-Way ANOVAs to examine group differences followed by Tukey post-hoc correction for pairwise comparisons. Given the large number of variables, I also used a false discovery rate (FDR) correction to account for multiple variables in between group differences. FDR has been used in prior research to correct for multiple variables (Benjamini & Hochberg, 1995). This correction strengthens the validity of the results. Standard of significance was set for both at $p<.05$.

Results

H1: Mood disorder essays use more affect-related language compared to schizophrenia and control groups. Mood disorder writings also used more self-referential language compared to schizophrenia, but not than controls.

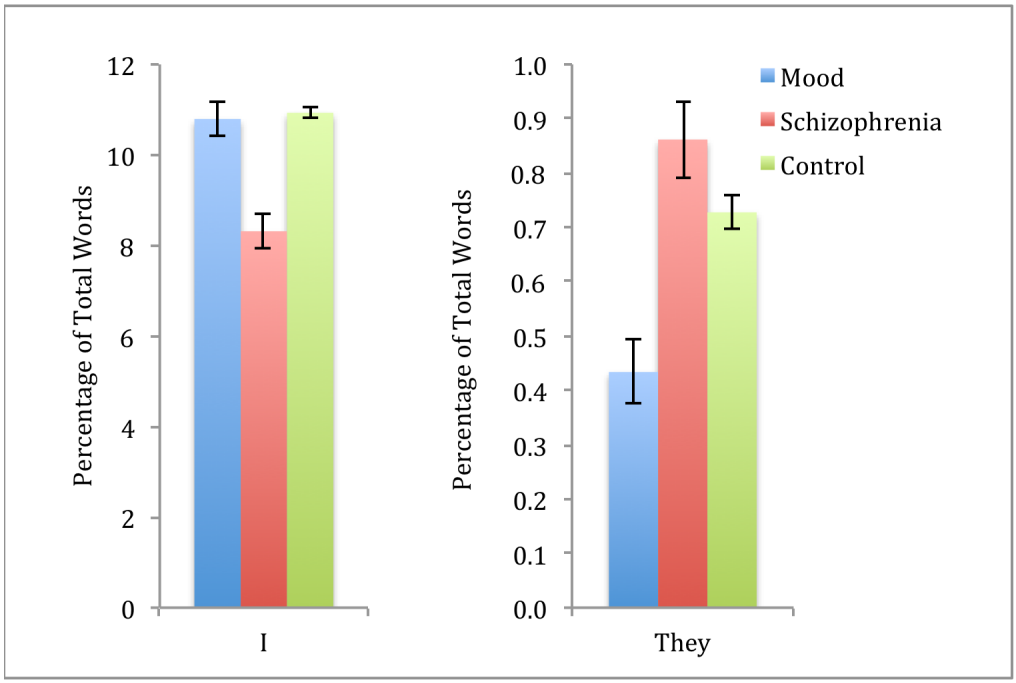
My first hypothesis (H1) predicted that mood disorder essays would use more affect-related words (positive and negative emotions including sadness, anxiety, happiness) compared to schizophrenia and control essays. To test this hypothesis, I measured the frequency of the LIWC category "affective words" and used a one-way ANOVA to compare between groups. Between groups differences were significant ($F=20.08$, FDR-adjusted p-value $<.001$). Specifically, affect-related words were higher amongst mood disorder essays and represented 7.95% of total words used, compared with 5.35% ($p<.001$) in schizophrenia samples and 5.96% ($p<.001$) in control essays. Within the subset of affective words, anxiety, sadness and negative emotion seemed to drive much of the difference between the mood disorder samples and the other groups. **(Figure 1)**

Figure 1: Patients with Mood Disorder Have Increased Affective Word Use



Personal reference as measured by "I" differed significantly between groups (F=61.95, FDR-adjusted p <.001), but pairwise comparisons demonstrated that most of this difference was secondary to a reduction in "I" in schizophrenia essays. Use of I did not differ significantly between mood disorder (10.8% of total words) and control (10.9%) groups, p =.993. "I" usage in schizophrenia was lower at 8.3%, and significantly different from mood disorder and control groups (p<.001). **(Figure 2)**

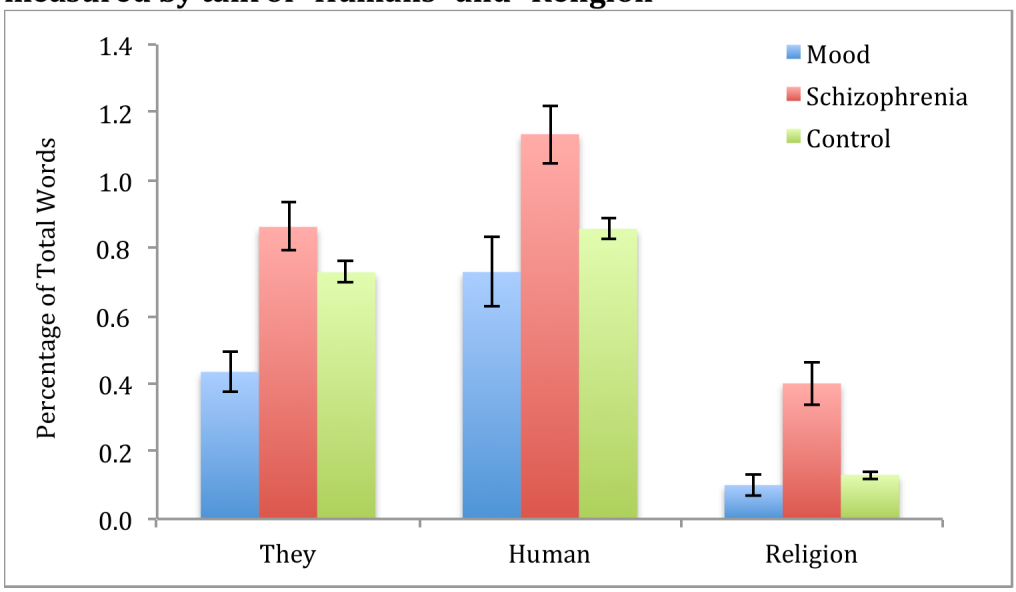
Figure 2: Patients with schizophrenia use more external referents as measured by function word use.



H2: Schizophrenia writers use more external referents and fewer personal referents when compared to mood disorder and control groups.

I also confirmed the second hypothesis (H2), which predicted that patients with schizophrenia would show changes in personal and external referential language compared to mood disorder and control groups. **(Figures 2 & 3)** Patients with schizophrenia used "I" significantly fewer times compared to controls and mood disorder patients, with "I" representing 8.3% of total words in schizophrenia writing samples, and 10.8% and 10.9% in mood disorder and control samples, respectively ($p<.001$). "They" was used in relative higher frequency in schizophrenia samples (.87% of total words) when compared to mood disorder samples (.43%, $p<.001$). The frequency was lower in controls (.72%) but was not significant at ($p=.176$).

Figure 3: Patients with Schizophrenia Use More External Referents as measured by talk of "Humans" and "Religion"

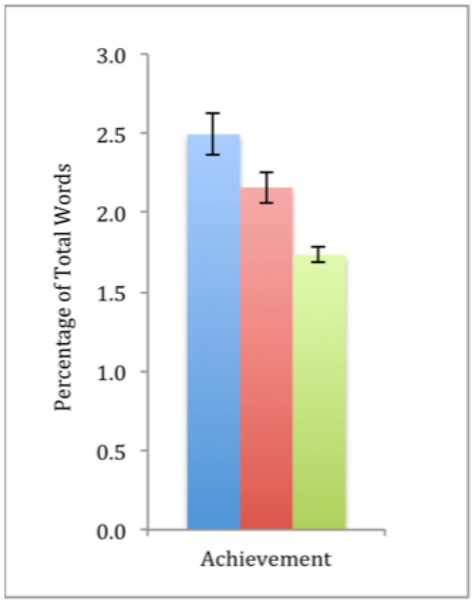


Content words strongly associated with external referents (mention of "humans" and "religion") were also more common in schizophrenia writing samples. (**Figure 3)** The "humans" category in LIWC is a category that includes all words relating to other humans, (adult, man, woman). It excludes relational words referring to humans (sister, friend, buddy, husband), because those words belong to the category "friend" or "family." Thus, the "human" category represents a more distant reference to humans. The "religion" category includes all language referring to religion, (pious, priest, altar, church). Between group differences were significant for humans ($F=5.97$, FDR-adjusted $p<.001$) and religion ($F=18.06$, FDR-adjusted $p<.001$). Specifically, schizophrenia essays used human-related words at a higher frequency, totaling 1.14% of total words used, with controls at .86% and mood disorder patients at .73% ($p=.009$ and $p=.001$, respectively). Schizophrenia essays also had more religion-related words (.40% of total words), with control and mood disorder samples using religion-related words at a frequency of .13% and .10%, respectively, $p<.001$).

H3: Both psychiatric writing samples (schizophrenia and mood disorder) wrote about achievement and the past tense more than controls.

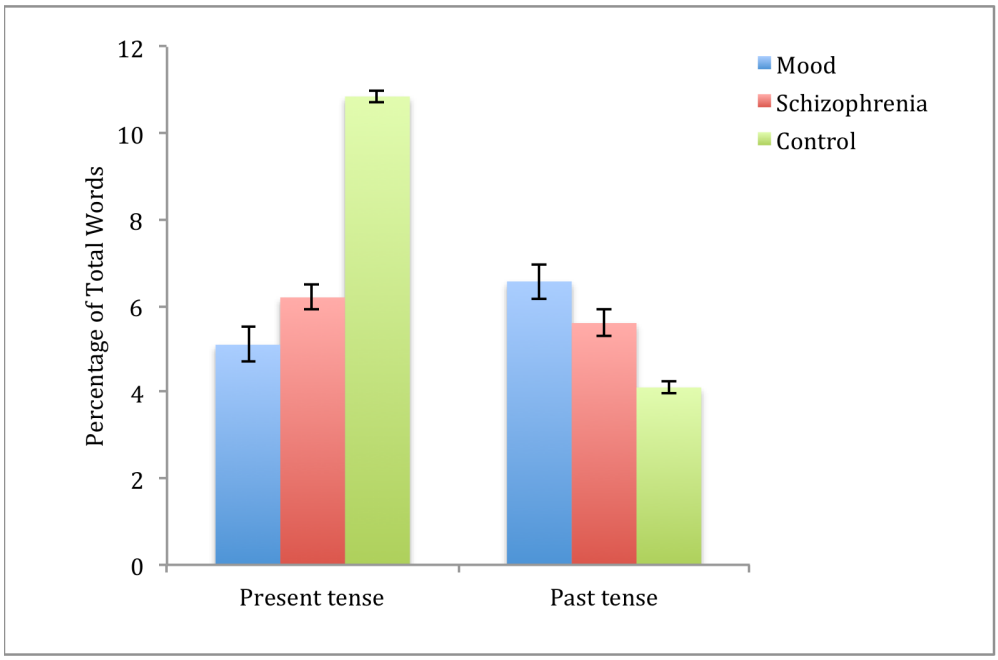
I expected to see a unifying pattern amongst the psychiatric samples (mood disorder and schizophrenia) when compared to controls (H3). The two areas in which schizophrenia and mood disorder samples clustered together, and separately from controls, was in use of achievement-related words and the present and past tense. **(Figures 4 & 5)** The "achievement" category in LIWC is a category that includes 186 words related to personal achievement including "earn," "hero," and "win." Between group differences in talk of achievement were significant (F=11.10, FDR-adjusted p<.001). Talk of achievement was significantly greater in mood disorder samples (2.5%) and schizophrenia samples (2.16%) when compared to controls (1.75% with p<.001, p=.001, respectively). However, mood disorder and schizophrenia samples were not significantly different from one another (p=.223).

Figure 4: Mental Illness Groups (Mood Disorder and Schizophrenia) talk more about achievement than controls.



Between group differences in present and past tense were also significant (F=129.65, FDR-adjusted p<.001, F=30.22, FDR-adjusted p<.001, respectively). The present tense was far more common in control essays (10.85%) compared with schizophrenia (6.20%, p<.001) and mood disorder essays (5.10%, p<.001), and again schizophrenia and mood disorder samples were not significantly different from one another (p=.142). Past tense usage was significantly lower in control essays (4.11%) when compared to schizophrenia (5.60%, p<.001) or mood disorder samples (6.57%, p<.001).

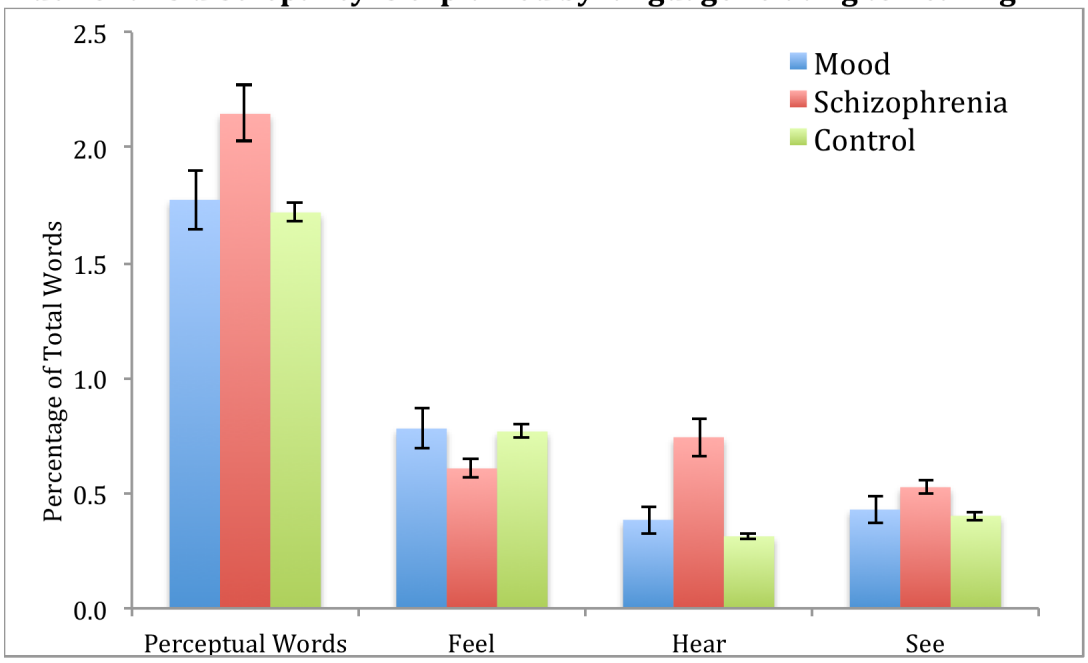
Figure 5: Mental Illness Groups (Mood Disorder and Schizophrenia) have increased use of past tense, and decreased use of present tense.



H4: Schizophrenia samples had greater usage of words related to perceptual experiences, particularly related to hearing.

Finally, as hypothesized (H4), I observed significant differences in language related to perceptual experiences. **(Figure 6)** For all language related to perception, between group differences were significant with F=7.413, FDR-adjusted p<.001. Schizophrenia essays used more perceptual-related words (2.15%) when compared to control essays (1.72%, p<.001). Perceptual related words were not significantly different between schizophrenia (2.15%) and mood disorder essays (1.77% p=.089), but may be trending towards significance with greater power. However, after breaking down perceptual words into subcategories, much of the discrepancy is accounted for by words related to hearing. (.74% in schizophrenia, versus .38% in mood, and .31% in controls, p<.001). In fact, patients with schizophrenia used perceptual words related to feeling (.61%) less than controls (.77%, p=.035), and were not significantly different from their mood disorder counterparts (.78%, p=.276).

Figure 6: Schizophrenia essays include more perceptual-related language. Much of this discrepancy is explained by language relating to hearing.



Discussion

Using Linguistic Inquiry Word Count, I employed a quantitative approach to analyzing language in schizophrenia. Using LIWC software, I was able to compare word usage in schizophrenia to mood disorder and control groups. The inclusion of a mood disorder group allowed for analysis of a more "active" control group, and helped distinguish what language features were unique to schizophrenia, rather than mental illness in general. Several significant differences emerged, some of which supported by my initial hypotheses, and others that replicated previous findings to help validate the sample. The patterns I observed provide an important lens into how language is affected in schizophrenia, and what that might inform about patient experience.

Affective Language

As expected, I observed increased use of affective language in the mood disorder sample. This has been shown in prior research, and aligns with the notion that mood disorders are affective in nature (Bernard, 2012). These particular findings are not considered novel, but serve the important purpose of validating this study's sample.

Personal Reference in Mood Disorders

I expected to see increased use of personal referents ("I") in depression compared to schizophrenia and control groups. Prior research has shown increased use of "I" in depression and negative affect states (Chung & Pennebaker, 2007). We can attribute this to both the tendency of those with depression to turn inward and self-isolate. My results somewhat replicated this finding, as they showed increased use of "I" in the mood disorder group compared to schizophrenia. However, the mood disorder and control groups were not significantly different from one another. This may be a result of this study's convenience control sample. The control essays were written by college students who were describing their transition to college. Research shows that the transition from high school to college is a high-stress period associated with higher levels of depression and anxiety (Wei, Russell, & Zakalik, 2005). Perhaps this reveals that the control group was not a corpus of true "healthy" controls, but instead represented a psychiatrically heterogeneous group with significant levels of anxiety and depression.

An alternative explanation for this result could be an age effect. We know that adolescence, including late adolescence that extends into the college ages, is a period of identity preoccupation and formation (Marcia, 1980). Increased self-reference may also be a

result of this particularly identity-focused time period for the study's college-aged control group.

Personal and External Reference in Schizophrenia

Compared to the control and mood disorder groups, the essays written by people with schizophrenia had decreased self-reference ("I") and increased external reference ("humans" and "religion"). These findings align with my initial hypothesis about that language in schizophrenia would show increased external referents and a reduction in referral to self. These changes in referential language likely reflect two themes in schizophrenia. First, schizophrenia is often characterized by a disturbed sense of self (Hemsley, 1998). A disturbed notion of one's personal identity may manifest in a reduction in the use of "I." Second, schizophrenia is often accompanied by a reduced sense of agency – both real and delusional. People diagnosed with schizophrenia often experience hallucinations and delusions, that can co-opt their daily life and routine. As these external forces (voices, paranoid delusions) exert more control over their lives, people with schizophrenia may develop a reduced sense of agency or locus of control (Kaney & Bentall, 1989). In addition to this perceived loss of control, people with schizophrenia often experience a very real loss of control. When severely ill, these patients can lose, temporarily or indefinitely, their very basic human rights. Patients can be locked into psychiatric wards against their will, can be stripped of their belongings if deemed dangerous, told when they can eat, and when they can leave. This would undoubtedly reduce one's sense of agency, and it seems natural that use of "I" might fall under such circumstances.

Increased use of words relating to "humans" (adult, man, woman) in schizophrenia may also reflect a notion of disconnectedness and isolation. The LIWC category "humans" excludes all relational human language. "Friend" and "buddy" are under the category of "friends," while "sister" and "husband" are part of the "family" category. Thus "humans" reflects a more disconnected reference to people. Social isolation and withdrawal are defining features of schizophrenia (APA, 2013). In addition, people diagnosed with schizophrenia are often subject to harsh stigma, and can be excluded and ostracized for their unusual behavior and affect (Link, Cullen, Struening, Shrout, & Dohrenwend, 1989; Link & Phelan, 2014). Increased talk of distant humans in schizophrenia essays may reflect this notion of exclusion and isolation.

Achievement and Past + Present Tense in Mental Illness

I expected to see commonalities between the schizophrenia and mood disorder groups that reflected a shared experiencing of living with mental illness. Two instinct word categories separated the mental illness groups (mood disorder and schizophrenia) from controls: Achievement-related words and the past tense were used at significantly higher rates in the mental illness groups.

Increased talk of achievement (words like "win" and "earn") in schizophrenia and mood disorders likely reveals a common theme in mental illness. By definition, diagnosis with a mental illness requires concomitant "functional impairment" (APA, 2013). Accordingly, mental illness is one of the leading causes of disability in the United States (Murray et al., 2013). The disability associated with mental illness can dramatically limit a patient's ability to achieve. This may be driving the increased used of achievement-related

words in schizophrenia and mood disorder essays. Of note, it is interesting that the control college essays talk less of achievement when the college years are often predicated on the concept of learning and achievement.

Increased use of the past tense in the mental illness groups (schizophrenia and mood disorders) may also reveal a common theme. In people's conception of illness, people tend to focus on the past. This includes various themes such as glorification of their prior health and how it painfully contrasts to their current state, and in conceiving of how the current illness came to be. They may conjecture whether there was a causal or inciting event, or whether there were early signs foreshadowing what was to come. This is often referred to as a retrospective bias, in which patients tend to look back and draw connections between earlier events and their current disease state. (Shafer & Dexter, 2012).

Perceptual Words

Finally, as expected, I observed increased perception-related words in schizophrenia essays when compared to the other groups. This aligns with the notion that schizophrenia is characterized by odd perceptual experiences (Mueser et al., 1990). These experiences are salient and often confusing or disturbing. So naturally, when prompted to talk about their illness, patients mention these unusual perceptual experiences. Hearing-related words accounted for most of this difference, which is consistent with the well-established fact that the majority of odd perceptual experiences in schizophrenia are auditory hallucinations (Mueser et al., 1990). Depression, (aside from psychotic depression, which this sample does not include), and the transition to college are not characterized by odd perceptual experiences, so it follows logically that these groups used less perpetual language.

Chapter 3: LIWC Analysis of Language in Schizophrenia, Family Members and Controls

Background:

Our current understanding of the pathophysiology schizophrenia suggests a complicated origin. Though not entirely elucidated, we have come to realize that the development of schizophrenia can be explained by numerous gene-environment interactions. (Karl & Arnold, 2014) Recent studies have shown strong genetic influences on the development of schizophrenia, (Schizophrenia Working Group of the Psychiatric Genomics, 2014) but epidemiological studies also demonstrate environmental influence. (Wahlberg et al., 1997),(McGrath et al., 2004). Given that family members possess genetic similarities, and often have similar environments, it is possible that family members not diagnosed with schizophrenia may embody an intermediate language phenotype. This concept inspired me to use the quantitative language analysis software Linguistic Inquiry Word Count (LIWC) to examine essays authored by family members of people with schizophrenia and compare them to their affected counterparts.

The evidence for genetic involvement in schizophrenia has been well described for some time. Twin studies dating back to the 1960s have shown a 41-50% concordance rate for schizophrenia amongst identical (monozygotic) twins, compared to 4-11% between dizygotic twins. (Cannon, Kaprio, Lönnqvist, Huttunen, & Koskenvuo, 1998; Cardno et al., 1999; Farmer, McGuffin, & Gottesman, 1987; Kringlen, 2013; Onstad, Skre, Torgersen, & Kringlen, 1991) More recently, advanced genetic analyses using genome-wide association sequencing (GWAS) have demonstrated several specific gene loci that are correlated with schizophrenia diagnosis (Schizophrenia Working Group of the Psychiatric Genomics, 2014).

Although we have compelling evidence for genetic involvement, lack of 100% monozygotic concordance or genetic profiles uniformly associated with a particular outcome suggest that other, environmental forces must be at play.

Some of the earliest conjecture regarding the pathophysiology of schizophrenia involved the concept of the cold, rigid "refrigerator" mother potentially being "schizophrenogenic," and thereby producing children who went on to develop schizophrenia (Kanner, 1968). This theory has long been disproven, but the underlying idea that environment may play a role in the development of schizophrenia has been re-examined with strong quantitative tools and epidemiological studies. For instance, a systematic review by McGrath et al in 2004 examined over 150 epidemiological studies, and showed that immigrant status and urban and environments are associated with increased rates of schizophrenia (McGrath et al., 2004). Additionally, some research indicates that exposure to infection in utero can lead to an increased risk of developing schizophrenia later in life. (Brown, 2006). The en utero environment has also been examined by Susser et al with the Dutch Famine Study Cohort (Susser et al., 1996). This study showed that exposure to famine in the early prenatal period was associated with a two-fold increase in the rate of schizophrenia diagnosis later in life. And finally, there is the questionable role of marijuana in the development of schizophrenia. We observe correlations with excessive marijuana use in adolescence and the development of schizophrenia. (DeLisi, 2008) However, it is difficult to determine whether the early prodrome (the period before the first psychotic episode) is characterized by an increased propensity to use marijuana, or whether in fact high marijuana use results in brain changes that may lead to schizophrenia. Overall, we have a complex

story; we have evidence demonstrating that genetics, environment and the dynamic interplay between the two, influence one's risk of developing schizophrenia.

People who have a strong family history of schizophrenia, but no diagnosis of schizophrenia themselves, provide a unique opportunity for study. They possess an overlapping genetic profile, and often a similar environment to their affected family members. Though these people do not meet diagnostic criteria for schizophrenia, perhaps their gene-environment profile might produce some milder phenotypic changes (in language and beyond) that separate them from the general population.

Research suggests that individuals with a strong family history of schizophrenia are more likely to have a variety of negative psychiatric proclivities including affective disorders (Baron & Gruen, 1991), paranoid personality disorder (Baron et al., 1985), and schizoid and avoidant personality disorder (Kendler et al., 1993). Interestingly, of these associations, schizotypal personality disorder and its associated traits demonstrate the strongest familial relationship to schizophrenia (Kendler et al., 1993). Some researchers conceptualize these higher rates of schizotypy as symptoms occupying a spot on continuum – ranging from healthy (no schizotypy or schizophrenia) to severe schizophrenia (Peters, Joseph, & Garety, 1999). However, currently, the DSM-V categorizes schizophrenia and schizotypal personality disorder as discrete diagnoses. They actually don't even occupy the same cluster of diagnoses; schizophrenia is listed under psychotic disorders and schizotypal personality disorder belongs to the personality disorders (APA, 2013). In the same way that Asperger's is now considered on an autism spectrum, occupying the higher-functioning end of this spectrum (Smith, Reichow, & Volkmar, 2015), one day we might conceptualize schizotypal personality disorder on a schizophrenia spectrum. With the R-DOC initiative recently

declared by the National Institutes of Mental Health to begin to base diagnosis on scientific pathophysiology, this very notion may be realized in the coming years (Glannon, 2015).

As discussed above, people with a strong family history of schizophrenia often demonstrate phenotypic similarities to those with an actual diagnosis of schizophrenia. In this chapter, I sought to examine whether we might also see phenotypic similarities with regards to language use. Interestingly, the DSM-V definition of schizotypyal personality disorder includes, "Odd thinking and speech (e.g., vague, circumstantial, metaphorical, overelaborate, or stereotyped)" (APA, 2013). Clearly, language is an important feature of both schizophrenia and schizotypy.

Prior research on schizotypy and language has been limited. A review by Kiang et al in 2010 discussed two major differences we see in schizotypal speech. When compared to the general population, people with schizotypal personality disorders tend to use more idiosyncratic words and more word-associations (Kiang, 2010). This aligns with the well-described features of schizophrenia language, which include both loose associations and neologisms (Covington et al., 2005). I used similar research methods to those described in Chapter 2 (using LIWC to do a quantitative language comparison) to further characterize language use in people with a family history of schizophrenia.

Research Hypotheses:

H1: Because schizophrenia is often associated with a disturbed sense of self and a feeling of losing agency (Hemsley, 1998), (Synofzik et al., 2010), I expected to see the schizophrenia essays using more external referents (they, them) and fewer personal referents (I) than controls. Family members would occupy an intermediate phenotype.

H2: I hypothesized that family member and schizophrenia writing would demonstrate similarities in function word use. Function words are words with little content-related meaning, but instead serve to express grammatical relationships with other words. Some examples are: the, a, he, she, then, well, thus. Pronouns (he/she) are one class of function words that can be related to theory of mind (one's ability to comprehend others' perspectives) and social engagement (Chung & Pennebaker, 2007). As theory of mind and social engagement are often compromised in schizophrenia (Corcoran, Mercer, & Frith, 1995), I expected to see differences in this type of word use in schizophrenia writing compared to controls, again with family members somewhere in between.

H3: Because schizophrenia is often characterized by strange perceptual experiences, I expected to again see differences in perceptual words. The most common form of perceptual disturbance is auditory hallucinations, but can also be in odd visual, tactile and olfactory experiences (Mueser et al., 1990). I therefore hypothesized that schizophrenia writing would contain more language related to perceptual experiences, particularly auditory. Schizotypal personality disorder can also be associated with perceptual disturbances (A. Raine, 1991), though not as overwhelming, so I expected to observe family member writing somewhere in between.

H4: To further characterize the degree to which family members' language resembles language in schizophrenia, I decided to employ a clustering analysis. I expected that this analysis would show schizophrenia family members either clustering with their affected counterparts, or possibly occupying a separate cluster from both control and schizophrenia writing (an intermediate cluster).

H5: Neologisms are non-dictionary words that are used at higher rates amongst individuals with schizophrenia (Covington et al., 2005). I expected to observe more frequent use of neologisms in schizophrenia writing, again with family members occupying an intermediate phenotype.

Methods:

Data Sources:

I analyzed a total of 520 first person essays: 77 schizophrenia and 25 unaffected family members of schizophrenia patients who wrote first person essays for Schizophrenia Bulletin, and 418 control essays. The essay prompt for schizophrenia patients, as previously stated, was to write about what it was like to have schizophrenia. Family members were prompted to write about their experience of having a family member with schizophrenia. The control essays were composed of 418 college student essays on the experience of coming to college. Because the schizophrenia and family member essays were published and publicly available, and the college student essays had already been processed by the Pennebaker lab at the University of Texas in Austin and were LIWC output with no possible identifiers, the study received exemption from Yale's Human Investigations Committee.

Language Processing:

Essays were processed using Linguistic Inquiry Word Count (LIWC) 2007 (J. W. Pennebaker, Francise, M.E., & Booth, J.R.,, 2007), a text analysis software that calculates the degree to which people use different categories of words. LIWC 2007's dictionary includes over 4,500 words, word stems and emoticons. The development of the LIWC dictionary

involved word collection using common emotion rating scales, Roget's thesaurus and standard English dictionaries. Words were preliminarily placed into categories, and independent reviewers evaluated them, needing a majority agreement for them to be placed in the final LIWC edition. LIWC was then validated using qualitative coding of texts involving different emotional states (J. W. Pennebaker et al., 2015). For each dictionary word, LIWC includes a corresponding dictionary entry that defines one or more word categories. LIWC output is the percentage of each word type in the sample for common distinct words, parts of speech and pre-defined themes. (See appendix for word categories)

Data Analysis:

The first set of analyses involved One-Way ANOVAs to examine group differences followed by Tukey post-hoc correction for pairwise comparisons. I used FDR correction for multiple comparisons, and determined the standard of significance at a level of $p < .05$.

The second set of analyses involved a cluster analysis. I used SPSS to run a principle components analysis on the LIWC output database. I then entered these principle components into an MClust analysis in R, which objectively determines how many clusters to split the cases, and then clusters the cases into the specified number of clusters it determined.

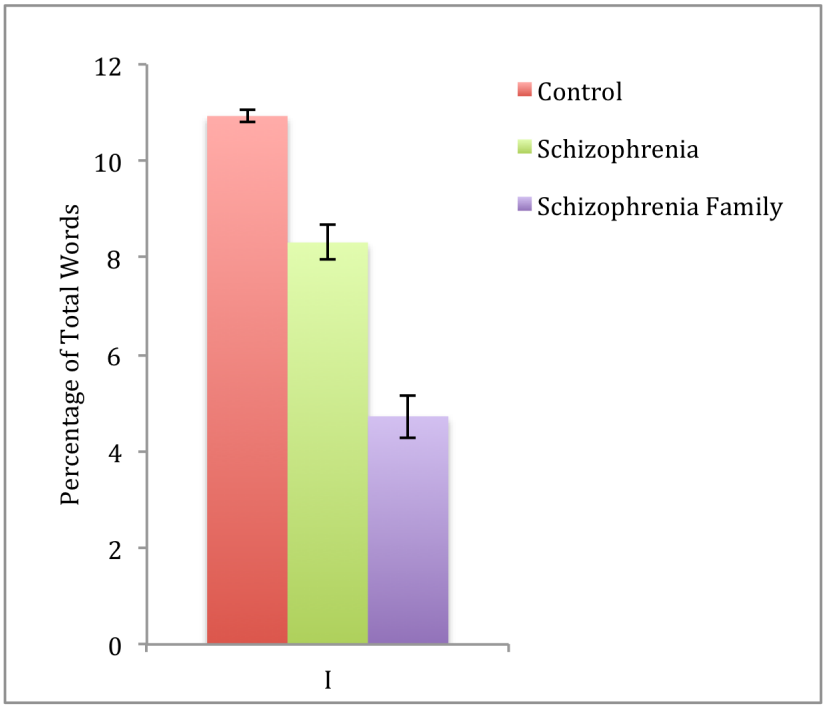
Results:

As expected, the analyses yielded several language patterns that separated schizophrenia and family member essays from controls. These included patterns in use of personal and external referents, function words, perceptual words and punctuation.

H1: Patients with Schizophrenia and family members use fewer personal and more external referential language than controls.

My initial hypothesis (H1) was that family members would occupy an intermediate phenotype between patients and controls in their use of personal (I) and external (humans, religion) referents. I found significant between group differences in the use of I (F= 61.95, FDR-adjusted p <.001). Schizophrenia family members and patients with schizophrenia used "I" less frequently (4.72% and 8.31%, respectively) than controls (10.80%, p<.001). Family member essays also used significantly less I than participants in the schizophrenia group (p<.001), and therefore did not occupy an intermediate phenotype here. **(Figure 7)**

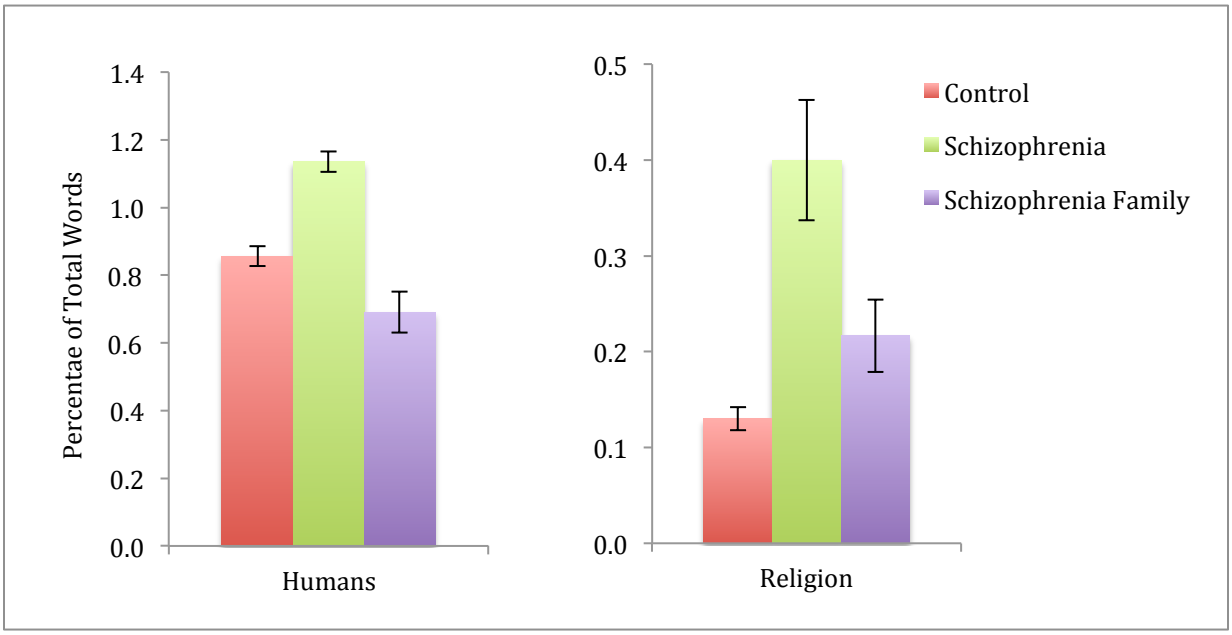
Figure 7: Schizophrenia and Schizophrenia Family Member Groups use less "I"



Usage of words related to "humans" and "religion," which both represent external/outside forces and influence, demonstrated family members with an intermediate pattern. The "humans" category includes words referring to other people (adult, man,

woman), but excludes closely related people. Words like "friend" and "buddy," are in the separate "friend" category, and "husband" and "sister" are in the "family" category. Between group differences were significant for "humans" (F=5.97, FDR-adjusted p<.001) and "religion" (F=18.06, FDR-adjusted p<.001). Talk of religion was higher in schizophrenia (.40%) and schizophrenia family member (.22%) samples when compared to controls (.13%, p<.001). However, human-related language did not show a unifying pattern amongst schizophrenia patients and family members. Schizophrenia samples used more language related to humans (1.14%) than both schizophrenia family members (.69%) and controls (.86%), p < .001. (**Figure 8)**

Figure 8: Schizophrenia and family member essays have increased talk of religion. Only schizophrenia essays have increased talk of other humans.



H2: Function word use, especially words related to social engagement and theory of mind, are higher in control samples than schizophrenia and family member essays.

Function word use (non-content words such as the, a, he/she, it) demonstrated strong similarities between schizophrenia and family member samples. Total function word use

showed significant between-subjects effects (F=126.07, FDR-adjusted p<.001). Total function word use was higher in control essays (64.08%) than either schizophrenia (58.42%) or schizophrenia family member writing (56.60%). Examining sub-categories of function words shows further nuance. Use of articles was also significantly different between groups (F=54.99, FDR-adjusted p<.001). Article use was higher in schizophrenia and family member writing (6.90% and 6.65%, respectively), than controls (4.96%, p<.001). However, the reverse was true for pronoun usage. Pronoun usage demonstrated significant between groups effects (F=40.60, FDR-adjusted p<.001). However, pronoun use was far higher in control samples (19.22%) than schizophrenia (15.84%, p<.001) and schizophrenia family member writings (16.45%, p<.001). Interestingly, in use of personal pronouns, writers with schizophrenia used significantly fewer (10.73%) than family members (12.68%, p=.002) and controls (13.14%, p<001), with between groups differences significant (F=22.50, FDR-adjusted p<.001). Family members and controls were not significantly different from one another (p=.782), suggesting a possible intermediate phenotype for family members. **(Figures 9,10,11)**

Figure 9: Schizophrenia and Schizophrenia Family Members Used Fewer Total Function Words

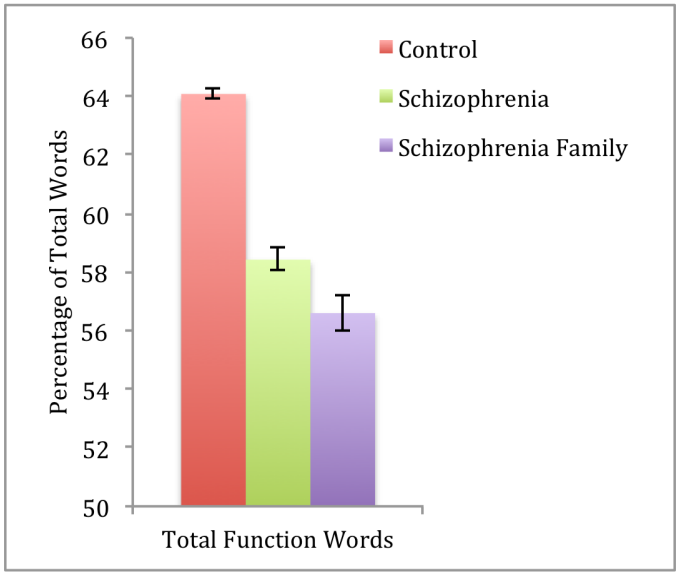


Figure 10: Schizophrenia and Family Member Essays Use More Articles and Fewer Pronouns

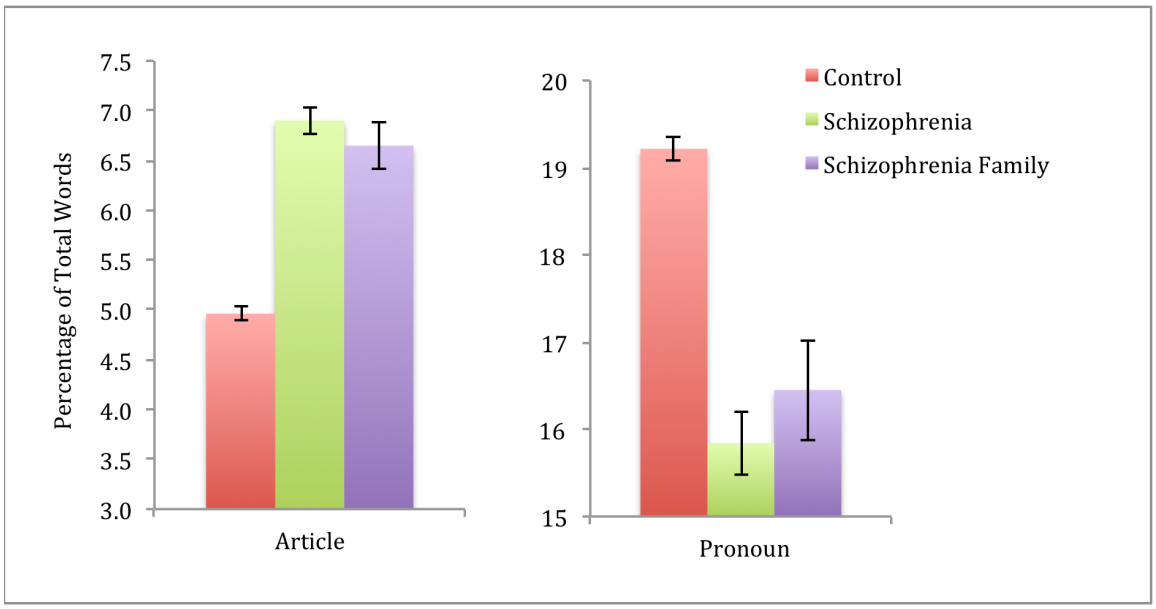
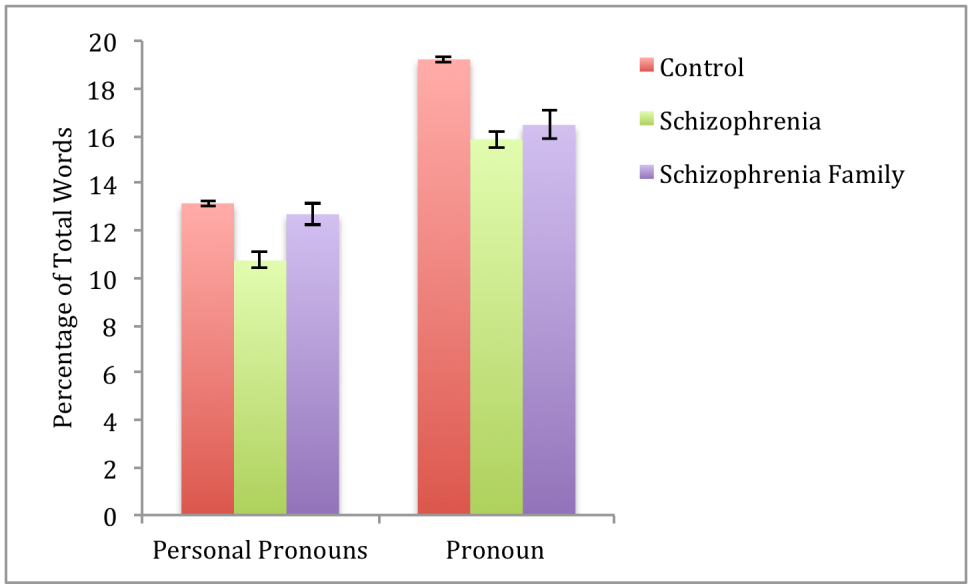


Figure 11: Schizophrenia and family member essays used fewer pronouns. Only Schizophrenia essays used fewer personal pronouns

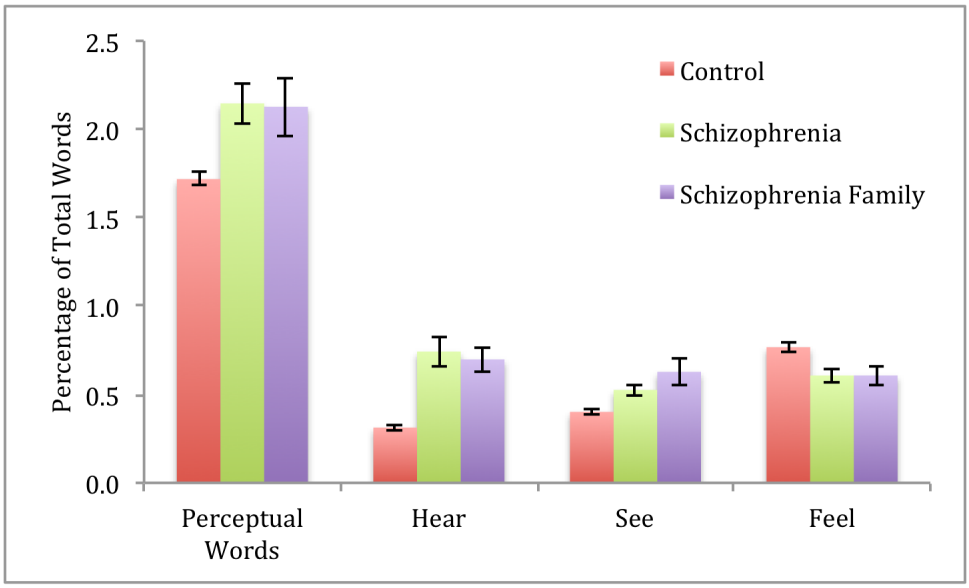


H3: Patients with schizophrenia and their family members use more perceptual language than controls; most of this can be attributed to talk of hearing.

Significant between group differences emerged in perceptual word use (F=7.41, FDR-adjusted p<.001). Perceptual words were higher in schizophrenia essays (2.15%) compared to controls (1.72%, p<.001). In family member essays, perceptual words were higher (2.13%) compared to controls (1.72%), with differences trending towards significance (p=.071). **(Figure 12)**

When broken down into perceptual sub-categories, we see that much of the difference can be explained through discussion of hearing/sound-related perceptions. Words related to hearing showed significant between group differences (F=28.52, FDR-adjusted p<.001) Talk of hearing was higher in schizophrenia (.74%) and family member (.70%) writings when compared to controls (.31%, p<.001). **(Figure 12)**

Figure 12: Schizophrenia and Schizophrenia Essays Use more Perceptual related words, driven largely by words related to hearing.

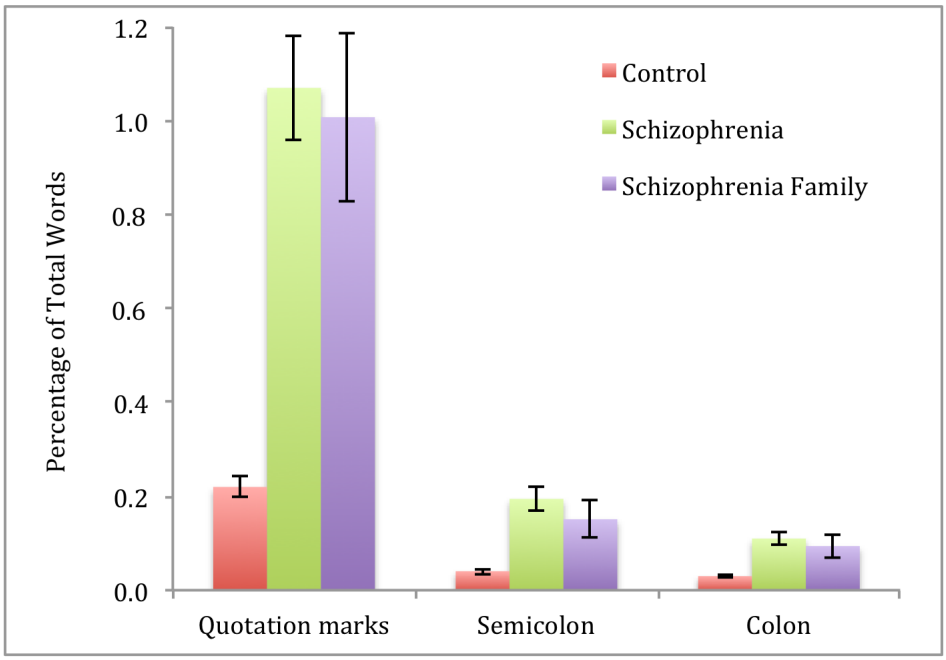


Unexpected Differences: Schizophrenia and family members used certain types of punctuation (including quotation marks, semicolons and colons) more frequently than controls.

The use of quotation marks, semicolons and colons differed significantly between controls, and schizophrenia and family member essays. Quotation mark use was significantly different between groups ($F=56.54$, FDR-adjusted $p<.001$) and were used at much higher rates by schizophrenia (1.07%) and schizophrenia family member essays (1.01%) than controls (.22%, $p>.001$). Semicolon use also differed significantly between groups ($F=25.55$, FDR-adjusted $p<.001$). They were also used more frequently by schizophrenia (.20%) and family members (.15%) in writing than controls (.04%, $p<.001$). Finally, colon use differed significantly ($F=28.33$, $p<.001$). Colons were also used more often by patients with schizophrenia (.11%) and family members (.09%), than controls (.03%, $p<.001$ and $p=.024$

respectively). However, schizophrenia and family member essays did not differ significantly from one another on any of these factors. **(Figure 13)**

Figure 13: Quotation, Semicolon and Colon Use Were Higher in Schizophrenia and Schizophrenia Family Members



H4: Cluster Analysis: schizophrenia and family members tended to cluster with one another based on a principle set of language components, with controls occupying a separate cluster.

To further analyze similarities between patients with schizophrenia and schizophrenia family members, I performed a cluster analysis to independently evaluate how writing samples tended to cluster with one another.

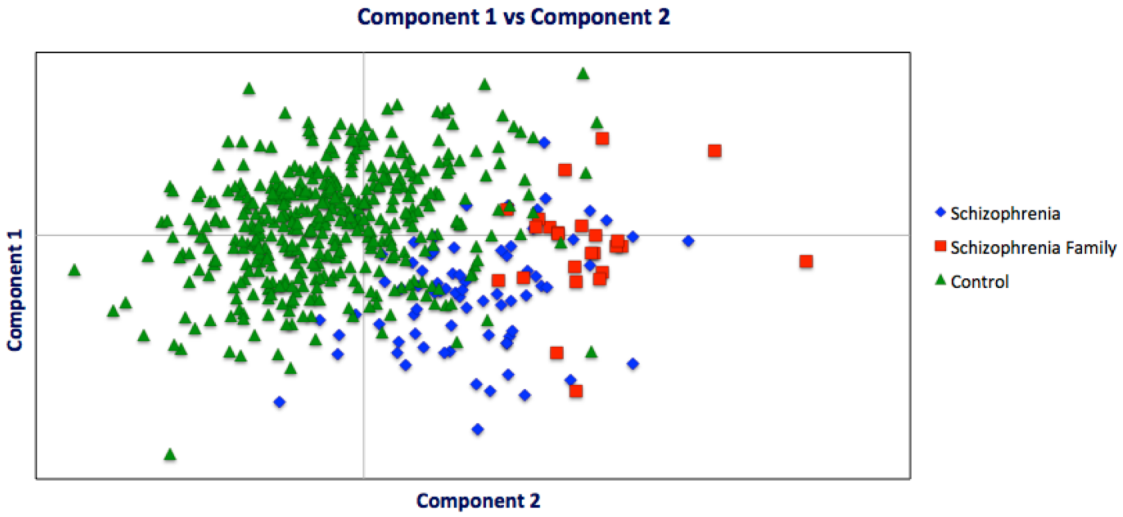
A principle components analysis demonstrated 5 main components within the language items database. **(Figure 14)** Kaiser-Meyer-Olkin Measure of Sampling Adequacy was .530, Bartlett's test of sphericity p=.000, with all Eigen values > 2.0. Adequacy was also confirmed by parallel analysis.

Figure 14: Results of the Principle Components Analysis

1		2		3		4		5	
article	-0.688	work	-0.68	family	-0.482	home	0.322	insight	-0.499
health	-0.413	quant	-0.432	preps	-0.417	preps	0.332	anx	-0.387
bio	-0.356	posemo	-0.391	achieve	-0.417	past	0.414	negemo	-0.308
preps	-0.315	present	-0.352	health	-0.393	number	0.457	shehe	0.315
sad	0.303	auxverb	-0.301	incl	-0.366	motion	0.482	family	0.338
discrep	0.346	see	0.321	past	-0.343	space	0.546	friend	0.355
adverb	0.35	ingest	0.34	article	-0.311	time	0.592	incl	0.381
negate	0.387	past	0.396	bio	-0.305	relativ	0.881	home	0.407
present	0.417	sad	0.403	quant	0.349				0.407
auxverb	0.493	health	0.462	certain	0.353				0.431
verb	0.734	hear	0.485	assent	0.397				0.453
i	0.755	death	0.495	tentat	0.404				0.506
pronoun	0.841	body	0.522	verb	0.421				
ppron	0.881	shehe	0.568	auxverb	0.423				
		bio	0.612	swear	0.454				
		anger	0.613	negate	0.474				
		negemo	0.649	adverb	0.553				
				present	0.572				
				ipron	0.574				
				excl	0.603				

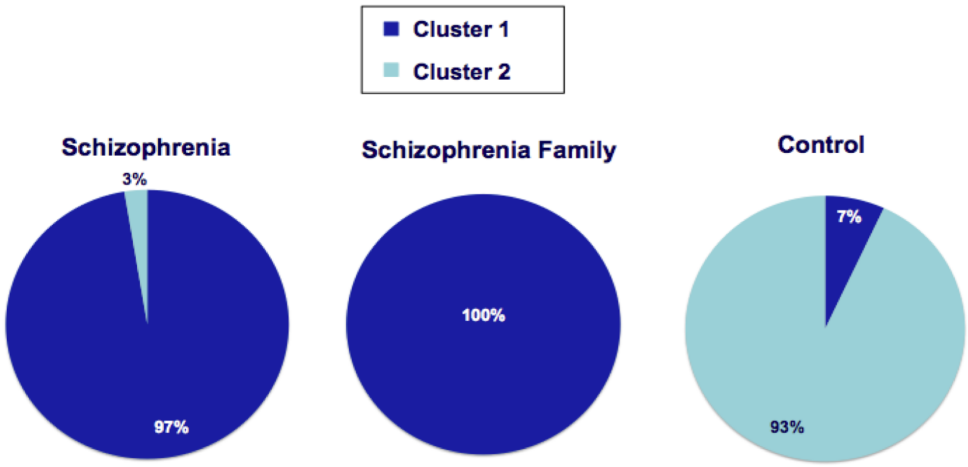
Of these principle components, components 1and 2 most strongly separated control writing samples from schizophrenia and family members. **(Figure 15)** Component 1 is characterized by increased use of pronouns, proper nouns, I, verbs, and the present tense. It is negatively associated with talk of health and biological processes, and with use of articles and prepositions. Schizophrenia and family members rate lower on component 1 when compared to controls. Component 2 is positively associated with perceptual language, the past tense, negative emotions death, and is negatively associated with talk of work, quantitative langue present tense and positive emotions. Schizophrenia and family members rate high in component 2.

Figure 15: Components 1 and 2 Most Strongly Separate Schizophrenia, Family Member and Control Groups



A cluster analysis using mclust in R objectively determined that two clusters existed in the essay corpus. Results showed that schizophrenia and schizophrenia family members tended to belong to cluster 1 (100% and 97%, respectively), whereas the controls were mostly in cluster 2 (93%). **(Figure 16)**

Figure 16: Schizophrenia and Schizophrenia Family Members Cluster Separately into Cluster 1



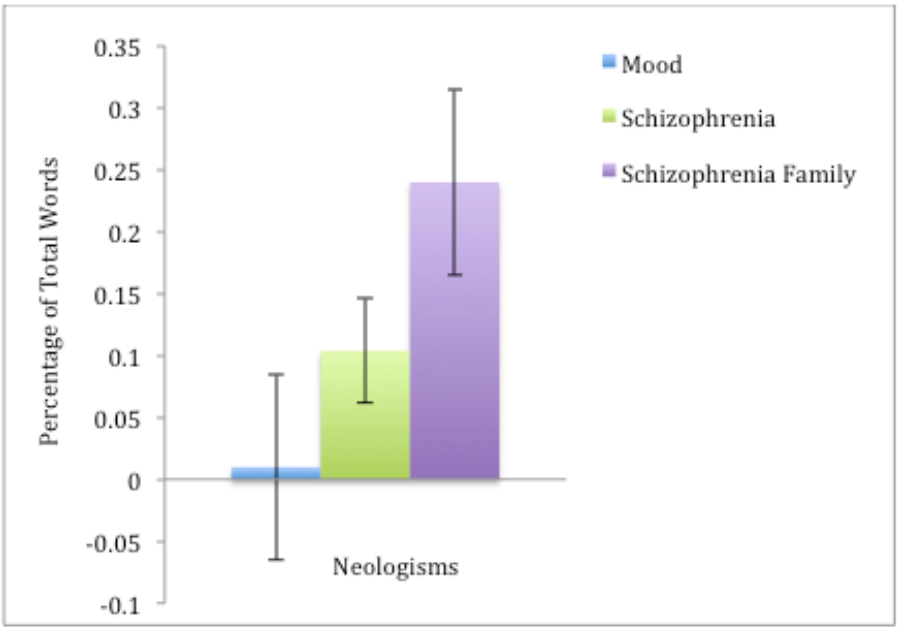
H5: Neologisms were qualitatively more "odd" in schizophrenia and family member essays.

Qualitative observation of neologisms (non-dictionary words invented by the writer/speaker) demonstrated the interesting finding that patients with schizophrenia as well as schizophrenia family members tended to use more neologisms, and neologisms that were more unusual when compared with a non-psychosis group. **(Figure 17)** I did not have access to possible neologisms in the control language corpus, so instead looked at neologisms in the mood disorder essays for comparison. Figure 17 indicates that patients with schizophrenia used unusual neologisms such as "transvestophobe" "fumbly" and "transmutate." Family members used words like "colickiness" "shadowsides" and "opalized." Mood disorder essays included some neologisms, but they seemed to be less unusual, including "perfectionistic" "ping-ponging" and "googling." An ANCOVA examining neologism counts in the schizophrenia, family member and mood disorder essays, controlling for word count, did not show significant between-groups effects ($F=1.03$, $p = .361$), with means of .01 for mood, .10 for schizophrenia and .24 for schizophrenia family. **(Figure 18)**

Figure 17: Neologisms in Schizophrenia and Schizophrenia Family Member Essays Were More Unusual

Schizophrenia	Schizophrenia Family	Mood Disorder
Transvestophobe	Colickiness	Perfectionistic
Fumbly	Shadowsides	Ping-ponging
Pschometrists	Scraggle	Googling
Transmutate	Opalized	
Antisexuals	Nondelusionary	

Figure 18: Neologism frequency in Mood Disorder, Schizophrenia and Schizophrenia Family Member Essays



Discussion:

Overall, this research supports this chapter's initial hypothesis that writers with a strong family history of schizophrenia would separate from controls and demonstrate some similarity to the writing authored by individuals with a diagnosis of schizophrenia. These results coincide with prior research showing that people with a strong family history of schizophrenia tend to have intermediate phenotypes on the schizophrenia spectrum. Although their language use did not neatly occupy an intermediate phenotype in each of the areas hypothesized, I did observe several important similarities.

Personal Referents

In the analysis of personal referents (mainly in the use of "I"), schizophrenia and family member essays demonstrated diminished use when compared to controls. This

concept coincides with the notion of a disturbed sense of self often seen in schizophrenia (Hemsley, 1998). People living with schizophrenia frequently describe delusions and auditory hallucinations as exerting strong control on their daily life, which results in a reduced locus of control, or a sense of being a master of one's destiny (Kaney & Bentall, 1989). Instead of saying, "I went to the park, so I could run along the beautiful lake," they might instead say, "The voices ordered me to go to the park and run away from those who were following me." The voice becomes passive, reducing the use of "I".

Decreased use of "I" may also communicate an important experience quite common in severe psychotic mental illness. Patients with severe psychosis are often subject to the loss of very basic human rights. They can be locked inside wards against their will, told when to eat, when they can speak to their family members, and when they are allowed to leave. This undoubtedly will contribute to a reduced sense of agency, and may be another reason for reduced use of "I" in those diagnosed with schizophrenia.

In schizotypal personality disorder, odd beliefs may also decrease one's sense of agency. These may include overwhelming influence of religion or belief in other unusual external forces influencing daily life. For instance, higher scores in schizotypy tend to be associated with higher religiosity and even participation in novel religious movements (Day & Peters, 1999). Other paranormal experiences are also associated with schizotypy (Hergovich, Schott, & Arendasy, 2008). We must also consider the experience of living with and taking care of a family member with schizophrenia as reducing one's sense of control, as this could be a potential confounder. Having a family member with schizophrenia may make someone feel like their daily routine is disrupted and even co-opted by his/her family member's illness. This too could result in diminished use of "I."

We must also recognize a limitation in this sample. Family members were asked to write about their experience as a family member of someone with schizophrenia. By definition, their prompt involves writing about another person. Depending on the essay, the reduction in "I," may be in part due to increased focus on their family members.

Increased External Referents (religions, humans)

Schizophrenia essays used more words relating to other "humans" and "religion" than controls. Family members did not clearly align with the schizophrenia group; they clustered with schizophrenia essays in talk of religion, but clustered with controls in talk of humans.

Increased reference to other humans in schizophrenia essays, but not in family member or controls, likely reflects the loss of agency people with schizophrenia often experience as other external forces (both real and imagined) take control of their lives. Talk of "humans" may increase as patients ascribe power to outside forces, such as voices in the form of auditory hallucinations, delusions or in real-life experiences with doctors, law enforcement, and support staff. The "human" category in LIWC reflects a more distanced and disconnected reference to humans (adult, man, woman), because referring to humans relationally close (friend, husband, sister) is categorized in different categories (friends, family). Distant references made to humans in the schizophrenia essays may reflect the social isolation and harsh stigma that patients with schizophrenia are often subject to.

In talk of religion, schizophrenia and family member essays clustered together, using more religion-related words than controls. Prior research indicates that religiosity is often higher in people with schizophrenia and schizotypal personality disorder (Day & Peters, 1999). People with schizophrenia often have delusions that are religious in nature, and people

with schizotypal personality disorder are more likely to engender odd beliefs with religious undertones. This may explain the differences compared to controls. However, it is also important to consider the effect of family environment. If a patient with schizophrenia engenders odd, religious beliefs, their family members may discuss and write about religion more than a control group.

Function Word Use:

As a reminder, function words are words with little content-related meaning, but are instead used to connect words and describe their relationship to one another (the, a, he, she, an, then, than). The results showed lower overall function word use in schizophrenia and family members. A paper by Chung & Pennebaker (2007) gives some interesting insight into the meaning of this. According to them, function word use requires a shared understanding in communication and the ability understand a listener's perspective. For instance, when discussing a story about a dog, after the dog is mentioned, someone who understands his/her audience's contextual perspective will use function words to refer to the dog, (it, he, him). People with schizophrenia, often exhibit deficits in the ability to understand other people's perspectives (theory of mind), and thus may have difficulty making this transition from "dog" to "him." According to Chung and Pennebaker, "The ability to use function words, then, is a marker of rather sophisticated social skills" (Pg. 349, Chung & Pennebaker, 2007).

When examined further, we can see that much of this discrepancy in function word frequency is actually driven by pronoun (he/she) use. This makes sense because out of the function words, pronouns reflect the most social engagement, and people with schizophrenia often experience social disengagement. This likely occurs through multiple different

mechanisms. First, one of the basic defining negative symptoms of schizophrenia is social withdrawal (APA, 2013). In addition to that, people with schizophrenia often experience negative repercussions of severe stigma. People with schizophrenia quite reasonably anticipate rejection from others, and evidence shows that the more they do, the more they are likely to withdraw and have smaller social support networks (Link et al., 1989; Link & Phelan, 2014).

Perceptual Words:

Odd perceptual experiences, usually in the form of auditory hallucinations, are a hallmark feature of schizophrenia (APA, 2013); schizotypal personality disorder is also defined by odd perceptual experiences (Adrian Raine et al., 1994). The LIWC results coincide with this, as schizophrenia and family member essays demonstrated increased perceptual words, particularly words related to hearing.

Some of this effect may be secondary to the topic of family member essays. Their prompt was to write about their personal experience living with someone with schizophrenia. Some language related to hearing may be a result of discussing their family member's odd perceptual experiences.

Punctuation:

Punctuation also showed some interesting patterns, with increased quotation marks, colons and semicolons in schizophrenia and family member writing. Increased quotation mark use likely reflects discussion of auditory hallucinations in schizophrenia. When discussing these hallucinations, writers would often directly quote the voices using

quotations within the text. In family member writing, much of the quotation mark use was driven by discussion of their family members, as they directly quoted the speech of their family members.

Colons and semicolons were also higher in schizophrenia and family member writing, which may derive from the explanation of unusual and different experiences. Colons and semicolons are used in sentences to extend a thought or idea. Because unusual and odd experiences are defining themes in schizophrenia and schizotypy, writers with these diagnoses may use more colons and semicolons to extend their explanation of experiences that are difficult to explain to a general audience.

Cluster Analysis:

Given the similarities I saw between schizophrenia and family member essays, I decided to perform a cluster analysis to determine whether family members tended to cluster with controls or patients. The unbiased mclust cluster analysis in R showed that family member essays tended to cluster more with schizophrenia essays than controls. This coincides with the various similarities observed in the preliminary ANOVAs.

Although the principle components analysis yielded five major factors, components one and two most strongly separated schizophrenia and family member essays from controls, with schizophrenia and family members rating lower in component one and higher in component two.

Component one, which was high in pronouns, proper nouns, I and verbs, and low in articles, biological processes and prepositions, aligns well with some of the findings from the aforementioned ANOVAs. Specifically, this reinforces prior results demonstrating

significantly reduced pronoun, proper noun and "I" use in schizophrenia and family member essays. As previously stated, lower use of "I" in schizophrenia samples may represent a theme of decreased personal reference as this mental illness is often characterized by a disturbed sense of self, and external and often overwhelming outside voices (delusions, auditory hallucinations). Family members, if high in schizotypy, may also be focused on external forces drawing from odd beliefs, though it should be mentioned that lower use of "I" may also reflect family members' essays focusing primarily on their affected family members rather than themselves. Schizophrenia and family member essays again showed lower pronoun and proper noun use, which may represent deficits in the ability to understand their audience's perspective and social disengagement/withdrawal.

Although use of verbs did not reach statistical significance in the ANOVAs, it is interesting that it was a strong factor within component one. This indicates that verb usage was generally lower in schizophrenia and family members. Perhaps decreased verb use may also reflect decreased agency and locus of control.

Low loadings of articles in component one suggest that schizophrenia and family members are using more articles – which is opposite of their pronoun use pattern. Increased articles and decreased pronouns may reflect social disengagement (more "it" than "he/she"). It is important to note that although family members are often talking about a family member, they too use fewer pronouns than controls. One would expect exactly the opposite given their subject matter, so perhaps they too are somewhat socially disengaged.

Component one also shows decreased talk of biological processes in schizophrenia and family members, which may again relate back to decreased personal reference and a preoccupation on external forces (delusions and auditory hallucinations).

Another factor positively associated with component one is the category of discrepancy. Increased talk of discrepancy, again not significantly different amongst the groups but higher in the control group, may reflect a couple of different ideas. Discrepancy words (than, or, compare) indicate the writer is comparing ideas, objects, people or, which indicates an engagement with one's environment. This may reflect the notion that people with schizophrenia can become socially withdrawn and isolated, and may have deficits in abstract and relational thinking (Mohamed, Paulsen, O'Leary, Arndt, & Andreasen, 1999).

Component two also yielded some interesting findings. It was positively associated with perceptual language, the past tense, death and negative emotions, and was negatively associated with talk of work, quantitative language and positive emotions. It more strongly separated schizophrenia and family members from controls than component one, with schizophrenia and family members rating higher in component two than controls. Increased perceptual language was also found in the initial ANOVAs and is likely related to the odd perceptual experiences (mainly auditory) that people with schizophrenia encounter. Within family members it is difficult to elucidate whether this reflects their family members odd experiences or their own. Death and negative emotions were also higher in schizophrenia and family members. Negative emotions may coincide with the emotional turmoil that often accompanies schizophrenia for both the patient and the family. Following from this, positive emotions were decreased. Talk of death may reflect a couple of things. Sadly, the rate of suicide is far higher in patients with schizophrenia than the general population (Palmer, Pankratz, & Bostwick, 2005), so increased talk of death may reflect this phenomenon. However, death and the afterlife is often a strong component of religion, so this discussion may also relate to religiosity.

Decreased talk of work and quantitative language in schizophrenia and family members likely correlates with diminishing cognitive ability (what we see with the negative symptoms of schizophrenia). This in turn results in lower rates of employment (Rosenheck et al., 2006). Family members, may intermediate symptoms (possibly somewhere in the schizotypal personality dimension) that may also interfere with their own productivity (Skodol et al., 2005).

Neologisms:

Although a quantitative analysis of neologisms did not reach statistical significance, the volume and oddity of the neologisms used by family members of people living with schizophrenia should be noted. Given the low number of overall neologisms in all essays, perhaps significance could have been reached with a greater number of participants in each group. It was interesting to see that family members not diagnosed with schizophrenia or any other psychotic disorder used words like “shadowsides,” “colickiness,” and “opalized.” Although maybe not as odd as the words used by patients with schizophrenia (“transvestophone” “fumbly” and “transmutate”), they were definitely stranger than those used by patients with mood disorders (“perfectionistic” “ping-ponging” and “googling,”), the latter of which is essentially common vernacular. This may be an independent reflection of schizotypy, but may also reflect a familial effect. People learn language primarily in the home, from their caregivers and close family members. There is a possibility that family members may use more unusual words because they hear their family members with schizophrenia using them on a daily basis.

Chapter 4: Limitations and Conclusions

Language differences and deficits in schizophrenia are broad and heterogeneous. For many decades, psychiatrists and linguists have studied and written about the unique language patterns we see in schizophrenia. Their research provides us with a rich repertoire of knowledge laying the foundation for novel quantitative methods to take this study deeper. This thesis used one of these novel methods, the quantitative software program Linguistic Inquiry Word Count, to examine one of language's most fundamental building blocks – the words themselves.

I analyzed essays written by four unique groups: people with schizophrenia, mood disorders or a family history of schizophrenia, and controls. With this method, I examined several novel concepts. Firstly, I looked at what made language in schizophrenia unique from mental illness in general by comparing schizophrenia essays to a psychiatric "control" (mood disorders). The major findings here were increased external reference and decreased self-reference in the schizophrenia essays. This particular analysis also allowed for the assessment of shared language features in mental illness by comparing both psychiatric groups to the control group. Results showed that the mental illness groups demonstrated increased talk of achievement and the past tense.

I was also able to analyze language in people with family history of schizophrenia, and observe whether language could be affected in a non-ill, but genetically and environmentally similar cohort to the schizophrenia. In many instances, family members occupied an intermediate phenotype, but not all of them.

Overall, these results add key concepts to our understanding of how language is altered in schizophrenia. While this research involves some inherent limitations, the results are significant and carry many important implications.

Limitations and Future Improvement:

This study involved a few limitations, which could be improved upon in further study. First, sample selection was somewhat flawed. The mood disorder, schizophrenia and family member groups all represented participants with those diagnoses, however the control group may have been more heterogeneous, and may not have contained only "healthy" controls. The control group was comprised of college students writing about their transition to college. This cohort was not screened for mental illness, so may have included some authors who actually met criteria for a wide range of psychiatric disorders. As described above, transition to college is often associated with increased rates of anxiety and depression, which may be one of the reasons this group resembled the mood disorder group in some ways (particularly in self-reference).

Following from this, the sample selection was also flawed in that I could not gather or control for any potentially confounding demographic variables. The essays were not published including age, race, gender, or socioeconomic status. The college group likely reflected a younger, and possibly more affluent, patient population, as the average age for beginning college is around 18 years of age, and socioeconomically advantaged people are more likely to go to college. A prospective study that gathered these variables would allow one to analyze, control and match for these differences, which would help remove any confounders that might exist.

Essay content represents another limitation of this study. Each group was given a slightly different prompt based on their circumstances, so the subject matter became inherently different. Schizophrenia and mood disorder groups were told to write about their experience with their respective mental illness, so the essays were about different mental illnesses. Family members were asked to write about their experience living with and taking care of someone with schizophrenia – so their essays likely focused on another person more than the others. Finally, college students were asked to write about coming to college. These writers did not need to focus on illness at all. The diversity of these prompts could lead to inherent differences in word use just based on content. A standardized prompt might allow future research to pick up differences in language irrespective of content. However, language is inherently affected by context, and this context affects content, which is important (Chung & Pennebaker, 2007). Thus, it is important to study language under both standardized and contextualized conditions, with this study's strength being the latter of those.

Another limitation of this study was that the samples consisted of edited written language. These samples represent a more "filtered" form of language, and some of the mistakes we see in every day language are undoubtedly missed. These errors might be important to study as they often reflect real-time language problems.

Finally, the schizophrenia essay writers should be considered. The Schizophrenia Bulletin "First Person Accounts" series calls for people diagnosed with schizophrenia to write about their illness. This sample likely consists of a generally higher-functioning cohort within schizophrenia and one with more insight. Being able to write a coherent essay requires a level of organization and insight that many people diagnosed with schizophrenia do not have. Further study of a more diverse group of patients with schizophrenia, and those in

acute psychotic episodes, could be more representative of the range of language differences that exist within schizophrenia.

Implications

Studying language in schizophrenia using novel, quantitative methods has several important implications. Text analysis software programs, like LIWC, carry the potential of quick data collection and analysis that may change the way we diagnose, risk-stratify and treat patients in the coming decades.

This thesis highlights some in important language changes we see in schizophrenia. In some cases, language changes and deficits are incredibly apparent. For instance, clanging, word salad and poverty of speech are easily recognized by most practitioners and even lay people. However, certain language changes in schizophrenia are subtler (such as changes in function word use or syntactic complexity). Subtle changes in language may become apparent before the disease itself does. In these cases, quantitative language analysis software could be used to detect early psychosis before florid functional impairment develops. For instance, research by Bedi et al in 2015, looked at language in youths considered to be at high-risk for developing schizophrenia. By analyzing language samples, they were able to predict conversion to psychosis based on syntactic complexity and semantic coherence (Bedi et al., 2015). This could be particularly important because early detection and treatment of psychosis is associated with better long-term outcomes (Bertolote & McGorry, 2005; Yap, 2010). Therefore, if enough standardized data are published regarding language changes in schizophrenia, text analysis of patient speech may allow for earlier detection and treatment of psychosis.

The Bedi article presents another possibility for intervention in early adolescence, before one converts to a diagnosis of schizophrenia. First, we know that high rates of familial schizophrenia confer greater risk of developing schizophrenia in youth. Recently there has been some promising data to show that targeting high-risk adolescents with particular interventions may allow us to reduce the number of full conversion rates to schizophrenia. For instance, research has shown marijuana use in adolescence is correlated with the development of schizophrenia (DeLisi, 2008). Interventions aimed at reducing marijuana use, and educating children at particularly high risk for developing schizophrenia about the potential dangers of this drug could be helpful. Also, teaching stress management and coping strategies using the principles of cognitive behavioral therapy can also be helpful in the prevention of conversion to schizophrenia (Morrison et al., 2004). Very recently, research has also suggested that omega-3 fatty acids can be helpful in preventing conversion to schizophrenia in the early prodome, as many have hypothesized that conversion to schizophrenia involves an inflammatory process leading to increased synaptic pruning, among other things (Amminger et al., 2010).

We have three potential interventions to reduce the risk of conversion to schizophrenia – marijuana use reduction, stress management and coping skills, and omega-3 fatty acid supplementation. The results of this study along with further quantitative language study could help us target these interventions. Text analysis involving large groups of at-risk youth could allow us to detect subtle features that indicate even higher risk, allowing us to better target resources and interventions to help reduce conversion rates. Reducing the total number of children converting to schizophrenia would have profound effects on the lives of

patients and their families, not to mention societal benefits from increased productivity and reduced healthcare system burden.

Text analysis in schizophrenia may also have implications in diagnostic precision. Because the diagnosis of mental illness is currently based on symptom clusters, and many symptoms bridge across diagnosis group, diagnosis can sometimes be unclear and difficult**.** For instance, some symptoms of schizophrenia such as disorganization or paranoia, may also be present in bipolar I mania, or even psychotic depression (APA, 2013). It is often difficult at first episode to determine what the core diagnosis actually is. Although one could argue that treatment overlaps to some degree (antipsychotics are used in all), there are particular treatment nuances that make precise diagnosis useful. For instance in mania, mood stabilizers such as lithium or valproic acid, can be helpful adjuncts to antipsychotics. In psychotic depression, one would also think about starting an antidepressant like a selective serotonin reuptake inhibitor (SSRI). In the cases where it is difficult to distinguish between these diagnoses, text analysis may be able to provide additional clues to aid in precise diagnosis and appropriate treatment.

Finally, quantitative language analysis may offer another tool for monitoring patients' disease course and severity of illness. In psychiatry, we often rely on patient and family member self-reports, which are inherently biased and may not always be reliable. A strong body of quantitative language markers could potentially be used to track patients' illness and perhaps clue us in to moments when intervention is needed. For instance, research shows that in depressed poets, "I" is increased in the period right before suicide (Stirman & Pennebaker, 2001). With more study of language patterns and how they fluctuate with symptom severity,

we might be able to identify important moments for clinical intervention, ultimately allowing us to provide better, high value care to patients with schizophrenia.

Final Remarks

This thesis employed a novel quantitative approach (Linguistic Inquiry Word Count) to analyzing language patterns in schizophrenia. Results demonstrated unique changes in content and function word use, and how these changes fit into a spectrum with psychiatric controls and family members. These results, combined with further research on language patterns in the prodrome and across different levels of disease severity, will provide new opportunities to enhance preventive actions and advance clinical care. These efforts will hopefully help improve the lives of people living with or affected by schizophrenia.

References

Allen, H. A., Liddle, P. F., & Frith, C. D. (1993). Negative features, retrieval processes and verbal fluency in schizophrenia. *Br J Psychiatry, 163*, 769-775.

Amminger, G Paul, Schäfer, Miriam R, Papageorgiou, Konstantinos, Klier, Claudia M, Cotton, Sue M, Harrigan, Susan M, . . . Berger, Gregor E. (2010). Long-chain ω-3 fatty acids for indicated prevention of psychotic disorders: a randomized, placebo-controlled trial. *Archives of General Psychiatry, 67*(2), 146-154.

Andreasen, N. C. (1979). Thought, language, and communication disorders. I. Clinical assessment, definition of terms, and evaluation of their reliability. *Arch Gen Psychiatry, 36*(12), 1315-1321.

Andreasen, N. C., & Grove, W. M. (1986). Thought, language, and communication in schizophrenia: diagnosis and prognosis. *Schizophr Bull, 12*(3), 348-359.

APA. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5.* . Washington, D.C.: American Psychiatric Association.

Baron, M., Gruen, R., Rainer, J. D., Kane, J., Asnis, L., & Lord, S. (1985). A family study of schizophrenic and normal control probands: implications for the spectrum concept of schizophrenia. *Am J Psychiatry, 142*(4), 447-455.

Baron, M., & Gruen, R. S. (1991). Schizophrenia and affective disorder: are they genetically linked? *Br J Psychiatry, 159*, 267-270.

Barrowclough, Christine, Tarrier, Nicholas, & Johnston, Marie. (1996). Distress, expressed emotion, and attributions in relatives of schizophrenia patients. *Schizophr Bull, 22*(4), 691.

Bedi, Gillinder, Carrillo, Facundo, Cecchi, Guillermo A, Slezak, Diego Fernández, Sigman, Mariano, Mota, Natália B, . . . Corcoran, Cheryl M. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. *npj Schizophrenia, 1*.

Benjamini, Yoav, & Hochberg, Yosef. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.

Bernard, Jared D. (2012). *Depression, language, and affect: An examination of the influence of baseline depression and affect induction on language*: SOUTHERN ILLINOIS UNIVERSITY AT CARBONDALE.

Bertolote, J, & McGorry, P. (2005). Early intervention and recovery for young people with early psychosis: consensus statement. *The British Journal of Psychiatry, 187*(48), s116-s119.

Brown, Alan S. (2006). Prenatal infection as a risk factor for schizophrenia. *Schizophrenia bulletin, 32*(2), 200-202.

Brüne, Martin, & Bodenstein, Luise. (2005). Proverb comprehension reconsidered—'theory of mind'and the pragmatic use of language in schizophrenia. *Schizophrenia Research, 75*(2), 233-239.

Buck, Benjamin, Minor, Kyle S, & Lysaker, Paul H. (2015). Differential lexical correlates of social cognition and metacognition in schizophrenia; a study of spontaneously-generated life narratives. *Comprehensive psychiatry, 58*, 138-145.

Cannon, Tyrone D, Kaprio, Jaakko, Lönnqvist, Jouko, Huttunen, Matti, & Koskenvuo, Markku. (1998). The genetic epidemiology of schizophrenia in a Finnish twin cohort: a population-based modeling study. *Archives of general psychiatry, 55*(1), 67-74.

Cardno, Alastair G, Marshall, E Jane, Coid, Bina, Macdonald, Alison M, Ribchester, Tracy R, Davies, Nadia J, . . . Sham, Pak C. (1999). Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. *Archives of general psychiatry, 56*(2), 162-168.

Chaika, Elaine. (1974). A linguist looks at "schizophrenic" language. *Brain and language, 1*(3), 257-276.

Chaika, Elaine Ostrach. (1990). *Understanding psychotic speech: Beyond Freud and Chomsky*: Charles C Thomas, Publisher.

Chen, Eric YH, Lam, Linda CW, Kan, CS, & Chan, Charlotte KY. (1966). LANGUAGE DISORGANISATION IN SCHIZOPHRENIA: VALIDATION AND ASSESSMENT VVITH A NEW CLINICAL RATING INSTRUMENT. *Hong Kong Journal of Psychiatry (I966), 6*(1), 1-13.

Chomsky, Noam. (1956). Three models for the description of language. *Information Theory, IRE Transactions on, 2*(3), 113-124.

Chung, C., & Pennebaker, J. (2007). The Psychological Functions of Function Words. *Social Communication*, 343-359.

Cohn, Michael A, Mehl, Matthias R, & Pennebaker, James W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological science, 15*(10), 687-693.

Condray, R., Steinhauer, S. R., van Kammen, D. P., & Kasparek, A. (2002). The language system in schizophrenia: effects of capacity and linguistic structure. *Schizophr Bull, 28*(3), 475-490.

Corcoran, R., Mercer, G., & Frith, C. D. (1995). Schizophrenia, symptomatology and social inference: investigating "theory of mind" in people with schizophrenia. *Schizophr Res, 17*(1), 5-13.

Corlett, PR, Honey, GD, & Fletcher, Paul C. (2007). From prediction error to psychosis: ketamine as a pharmacological model of delusions. *Journal of Psychopharmacology, 21*(3), 238-252.

Corman, Steven, & Dooley, Kevin. (2006). Crawdad text analysis system 2.0. *Chandler, AZ: Crawdad Technologies, LLC*.

Covington, M. A., He, C., Brown, C., Naci, L., McClain, J. T., Fjordbak, B. S., . . . Brown, J. (2005). Schizophrenia and the structure of language: the linguist's view. *Schizophr Res, 77*(1), 85-98. doi: 10.1016/j.schres.2005.01.016

Crump, C., Winkleby, M. A., Sundquist, K., & Sundquist, J. (2013). Comorbidities and mortality in persons with schizophrenia: a Swedish national cohort study. *Am J Psychiatry, 170*(3), 324-333. doi: 10.1176/appi.ajp.2012.12050599

Day, Samantha, & Peters, Emmanuelle. (1999). The incidence of schizotypy in new religious movements. *Personality and Individual Differences, 27*(1), 55-67.

de Bonis, Monique, Epelbaum, Catherine, Deffez, Valérie, & Féline, André. (1997). The comprehension of metaphors in schizophrenia. *Psychopathology, 30*(3), 149-154.

DeLisi, L. E. (2008). The effect of cannabis on the brain: can it cause brain anomalies that lead to increased risk for schizophrenia? *Curr Opin Psychiatry, 21*(2), 140-150. doi: 10.1097/YCO.0b013e3282f51266

Ezzati, Majid, Friedman, Ari B, Kulkarni, Sandeep C, & Murray, Christopher JL. (2008). The reversal of fortunes: trends in county mortality and cross-county mortality disparities in the United States. *PLoS Med, 5*(4), e66.

Farmer, Anne E, McGuffin, Peter, & Gottesman, Irving I. (1987). Twin concordance for DSM-III schizophrenia: scrutinizing the validity of the definition. *Archives of general psychiatry, 44*(7), 634-641.

Fraser, W. I., King, K. M., Thomas, P., & Kendell, R. E. (1986). The diagnosis of schizophrenia by language analysis. *Br J Psychiatry, 148*, 275-278.

Glannon, W. (2015). Research domain criteria: a final paradigm for psychiatry? *Front Hum Neurosci, 9*, 488. doi: 10.3389/fnhum.2015.00488

Goff, D. C., Cather, C., Evins, A. E., Henderson, D. C., Freudenreich, O., Copeland, P. M., . . . Sacks, F. M. (2005). Medical morbidity and mortality in schizophrenia: guidelines for psychiatrists. *J Clin Psychiatry, 66*(2), 183-194; quiz 147, 273-184.

Grattan, Lynn M, & Eslinger, Paul J. (1991). Frontal lobe damage in children and adults: A comparative review. *Developmental Neuropsychology, 7*(3), 283-326.

Hart, Rodrick. (2000). Diction 5.0, the text-analysis program user's manual, Scolari Software: Sage Press. http://www. dictionsoftware. com.

Hebert, Liesi E, Weuve, Jennifer, Scherr, Paul A, & Evans, Denis A. (2013). Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. *Neurology, 80*(19), 1778-1783.

Hemsley, D. R. (1998). The disruption of the 'sense of self' in schizophrenia: potential links with disturbances of information processing. *Br J Med Psychol, 71 (Pt 2)*, 115-124.

Hergovich, Andreas, Schott, Reinhard, & Arendasy, Martin. (2008). On the relationship between paranormal belief and schizotypy among adolescents. *Personality and Individual Differences, 45*(2), 119-125.

Hoffman, R. E., Stopek, S., & Andreasen, N. C. (1986). A comparative study of manic vs schizophrenic speech disorganization. *Arch Gen Psychiatry, 43*(9), 831-838.

Hong, Kai, Nenkova, Ani, March, Mary E, Parker, Amber P, Verma, Ragini, & Kohler, Christian G. (2015). Lexical use in emotional autobiographical narratives of persons with schizophrenia and healthy controls. *Psychiatry Research, 225*(1-2), 40-49.

Kaney, S., & Bentall, R. P. (1989). Persecutory delusions and attributional style. *Br J Med Psychol, 62 (Pt 2)*, 191-198.

Kanner, L. (1968). Autistic disturbances of affective contact. *Acta Paedopsychiatr, 35*(4), 100-136.

Karl, T., & Arnold, J. C. (2014). Schizophrenia: a consequence of gene-environment interactions? *Front Behav Neurosci, 8*, 435. doi: 10.3389/fnbeh.2014.00435

Kay, Stanley R, Flszbein, Abraham, & Opfer, Lewis A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia bulletin, 13*(2), 261.

Kendler, K. S., McGuire, M., Gruenberg, A. M., O'Hare, A., Spellman, M., & Walsh, D. (1993). The Roscommon Family Study. III. Schizophrenia-related personality disorders in relatives. *Arch Gen Psychiatry, 50*(10), 781-788.

Kiang, Michael. (2010). Schizotypy and language: A review. *Journal of Neurolinguistics, 23*(3), 193-203. doi: http://dx.doi.org/10.1016/j.jneuroling.2009.03.002

Kiang, Michael, Light, Gregory A, Prugh, Jocelyn, Coulson, Seana, Braff, David L, & Kutas, Marta. (2007). Cognitive, neurophysiological, and functional correlates of proverb interpretation abnormalities in schizophrenia. *Journal of the International Neuropsychological Society, 13*(04), 653-663.

Kringlen, Einar. (2013). *Heredity and environment in the functional psychoses: An epidemiological–clinical twin study*: Elsevier.

Kuperberg, Gina R. (2010). Language in schizophrenia part 1: an introduction. *Language and linguistics compass, 4*(8), 576-589.

Lecours, AR., Vanier-Clement, M. (1974). Schizophasia and jargon-aphasi. *Brain Language, 3*(516-565).

Link, Bruce G, Cullen, Francis T, Struening, Elmer, Shrout, Patrick E, & Dohrenwend, Bruce P. (1989). A modified labeling theory approach to mental disorders: An empirical assessment. *American Sociological Review*, 400-423.

Link, Bruce G, & Phelan, Jo. (2014). Stigma power. *Social Science & Medicine, 103*, 24-32.

Lorenz, Maria. (1961). Problems posed by schizophrenic language. *Archives of General Psychiatry, 4*(6), 603-610.

Marcia, James E. (1980). Identity in adolescence. *Handbook of adolescent psychology, 9*, 159-187.

McGrath, J., Saha, S., Chant, D., & Welham, J. (2008). Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiol Rev, 30*, 67-76. doi: 10.1093/epirev/mxn001

McGrath, J., Saha, S., Welham, J., El Saadi, O., MacCauley, C., & Chant, D. (2004). A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Med, 2*, 13. doi: 10.1186/1741-7015-2-13

McGregor, William B. (2015). *Linguistics: an introduction*: Bloomsbury Publishing.

McKenna, Peter J. (2007). *Schizophrenia and related syndromes*: Routledge.

Minor, Kyle S, Bonfils, Kelsey A, Luther, Lauren, Firmin, Ruth L, Kukla, Marina, MacLain, Victoria R, . . . Salyers, Michelle P. (2015). Lexical analysis in schizophrenia: How emotion and social word use informs our understanding of clinical presentation. *Journal of psychiatric research, 64*, 74-78.

Mo, Shuliang, Su, Yanjie, Chan, Raymond CK, & Liu, Jianxin. (2008). Comprehension of metaphor and irony in schizophrenia during remission: the role of theory of mind and IQ. *Psychiatry Research, 157*(1), 21-29.

Mohamed, Somaia, Paulsen, Jane S, O'Leary, Daniel, Arndt, Stephen, & Andreasen, Nancy. (1999). Generalized cognitive deficits in schizophrenia: a study of first-episode patients. *Archives of General Psychiatry, 56*(8), 749-754.

Morrison, Anthony P, French, Paul, Walford, Lara, Lewis, Shôn W, Kilcommons, Aoiffe, Green, Joanne, . . . Bentall, Richard P. (2004). Cognitive therapy for the prevention of psychosis in people at ultra-high risk Randomised controlled trial. *The British Journal of Psychiatry, 185*(4), 291-297.

Mueser, K. T., Bellack, A. S., & Brady, E. U. (1990). Hallucinations in schizophrenia. *Acta Psychiatr Scand, 82*(1), 26-29.

Murray, Christopher JL, Abraham, Jerry, Ali, Mohammed K, Alvarado, Miriam, Atkinson, Charles, Baddour, Larry M, . . . Birbeck, Gretchen. (2013). The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. *Jama, 310*(6), 591-606.

Oh, TM, McCarthy, RA, & McKenna, PJ. (2002). Is there a schizophasia? A study applying the single case approach to formal thought disorder in schizophrenia. *Neurocase, 8*(2), 233-244.

Onstad, Sidsel, Skre, I, Torgersen, S, & Kringlen, E. (1991). Twin concordance for DSM-III-R schizophrenia. *Acta Psychiatrica Scandinavica, 83*(5), 395-401.

Palmer, Brian A, Pankratz, V Shane, & Bostwick, John Michael. (2005). The lifetime risk of suicide in schizophrenia: a reexamination. *Archives of General Psychiatry, 62*(3), 247-253.

Pennebaker, J.W., Francise, M.E., & Booth, J.R.,. (2007). Linguistic Inquiry Word Count. *LIWC2007, Mahwa, NJ: Ebrham Publishers*.

Pennebaker, James W, Chung, CK, Ireland, M, Gonzales, A, & Booth, RJ. (2015). The Development and Psychometric Properties of LIWC 2007 (2007). *Austin, TX, LIWC. Net*.

Pennebaker, James W, & Lay, Thomas C. (2002). Language use and personality during crises: Analyses of Mayor Rudolph Giuliani's press conferences. *Journal of Research in Personality, 36*(3), 271-282.

Peters, Emmanuelle R, Joseph, Stephen A, & Garety, Philippa A. (1999). Measurement of delusional ideation in the normal population: introducing the PDI (Peters et al. Delusions Inventory). *Schizophrenia bulletin, 25*(3), 553-576.

Raine, A. (1991). The SPQ: a scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophr Bull, 17*(4), 555-564.

Raine, Adrian, Reynolds, Chandra, Lencz, Todd, Scerbo, Angela, Triphon, Nelly, & Kim, Deborah. (1994). Cognitive-perceptual, interpersonal, and disorganized features of schizotypal personality. *Schizophr Bull, 20*(1), 191.

Rieber, RW, & Vetter, Harold. (1994). The problem of language and thought in schizophrenia: a review. *Journal of psycholinguistic Research, 23*(2), 149-195.

Rodriguez-Ferrera, S, McCarthy, RA, & McKenna, PJ. (2001). Language in schizophrenia and its relationship to formal thought disorder. *Psychological Medicine, 31*(02), 197-205.

Rosenheck, Robert, Leslie, Douglas, Keefe, Richard, McEvoy, Joseph, Swartz, Marvin, Perkins, Diana, . . . Lieberman, Jeffrey. (2006). Barriers to employment for people with schizophrenia. *American Journal of Psychiatry, 163*(3), 411-417.

Rude, Stephanie, Gortner, Eva-Maria, & Pennebaker, James. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion, 18*(8), 1121-1133.

Schizophrenia Working Group of the Psychiatric Genomics, Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature, 511*(7510), 421-427. doi: 10.1038/nature13595

Selten, Jean-Paul, Wiersma, Durk, & van den Bosch, Robert J. (2000). Distress attributed to negative symptoms in schizophrenia. *Schizophr Bull, 26*(3), 737.

Shafer, Steven L, & Dexter, Franklin. (2012). Publication bias, retrospective bias, and reproducibility of significant results in observational studies. *Anesthesia & Analgesia, 114*(5), 931-932.

Shaw, Jonathan E, Sicree, Richard A, & Zimmet, Paul Z. (2010). Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes research and clinical practice, 87*(1), 4-14.

Skodol, Andrew E, Pagano, Maria E, Bender, Donna S, Shea, M Tracie, Gunderson, John G, Yen, Shirley, . . . Grilo, Carlos M. (2005). Stability of functional impairment in patients with schizotypal, borderline, avoidant, or obsessive–compulsive personality disorder over two years. *Psychological Medicine, 35*(03), 443-451.

Smith, I. C., Reichow, B., & Volkmar, F. R. (2015). The Effects of DSM-5 Criteria on Number of Individuals Diagnosed with Autism Spectrum Disorder: A Systematic Review. *J Autism Dev Disord, 45*(8), 2541-2552. doi: 10.1007/s10803-015-2423-8

Stein, Johanna. (1993). Vocal alterations in schizophrenic speech. *The Journal of nervous and mental disease, 181*(1), 59-61.

Stirman, Shannon Wiltsey, & Pennebaker, James W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine, 63*(4), 517-522.

Susser, E., Neugebauer, R., Hoek, H. W., Brown, A. S., Lin, S., Labovitz, D., & Gorman, J. M. (1996). Schizophrenia after prenatal famine. Further evidence. *Arch Gen Psychiatry, 53*(1), 25-31.

Synofzik, M., Thier, P., Leube, D. T., Schlotterbeck, P., & Lindner, A. (2010). Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. *Brain, 133*(Pt 1), 262-271. doi: 10.1093/brain/awp291

Theodoridou, A, & Rössler, W. (2010). Disease Burden and Disability-Adjusted Life Years Due to Schizophrenia and Psychotic Disorders *Handbook of Disease Burdens and Quality of Life Measures* (pp. 1493-1507): Springer.

Thomas, P., King, K., & Fraser, W. I. (1987). Positive and negative symptoms of schizophrenia and linguistic performance. *Acta Psychiatr Scand, 76*(2), 144-151.

Thomas, P., King, K., Fraser, W. I., & Kendell, R. E. (1990). Linguistic performance in schizophrenia: a comparison of acute and chronic patients. *Br J Psychiatry, 156*, 204-210, 214-205.

Wahlberg, K. E., Wynne, L. C., Oja, H., Keskitalo, P., Pykalainen, L., Lahti, I., . . . Tienari, P. (1997). Gene-environment interaction in vulnerability to schizophrenia: findings from the Finnish Adoptive Family Study of Schizophrenia. *Am J Psychiatry, 154*(3), 355-362.

Wei, Meifen, Russell, Daniel W, & Zakalik, Robyn A. (2005). Adult attachment, social self-efficacy, self-disclosure, loneliness, and subsequent depression for freshman college students: A longitudinal study. *Journal of Counseling Psychology, 52*(4), 602.

Weintraub, Walter. (1989). *Verbal behavior in everyday life*: Springer Publishing Co.

Yap, HL. (2010). Early psychosis intervention. *Singapore medical journal, 51*(9), 689-693.

Ziskind, Eugene. (1958). Isolation stress in medical and mental illness. *Journal of the American Medical Association, 168*(11), 1427-1431.

Appendix: LIWC Word Categories[†]

Category	Abbreviation	Examples	Words in Category
Linguistic Processes			
Word count	wc		
Words/sentence	wps		
Dictionary words	dic		
Words > 6 letters	sixltr		
Total function words	funct		464
Total pronouns	pronoun	I, them, itself	116
Personal pronouns	ppron	I, them, her	70
1st pers singular	I	I, me, mine	12
1st pers plural	we	We, us, our	12
2nd person	you	You, your, thou	20
3rd pers singular	shehe	She, her, him	17
3rd pers plural	they	They, their they'd	10
Impersonal pronouns	ipron	It, it's, those	46
Articles	article	A, an, the	3
Common verbs[a]	verb	Walk, went, see	383
Auxiliary verbs	auxverb	Am, will, have	144
Past tense[a]	past	Went, ran, had	145
Present tense[a]	present	Is, does, hear	169
Future tense[a]	future	Will, gonna	48
Adverbs	adverb	Very, really, quickly	69
Prepositions	prep	To, with, above	60
Conjunctions	conj	And, but, whereas	28
Negations	negate	No, not, never	57
Quantifiers	quant	Few, many, much	89
Numbers	number	Second, thousand	34
Swear words	swear	Damn, piss	53
Psychological Processes			
Social processes[b]	social	Mate, talk, they, child	455
Family	family	Daughter, husband, aunt	64
Friends	friend	Buddy, friend, neighbor	37
Humans	human	Adult, baby, boy	61
Affective processes	affect	Happy, cried, abandon	915
Positive emotion	posemo	Love, nice, sweet	406
Negative emotion	negemo	Hurt, ugly, nasty	499
Anxiety	anx	Worried, fearful, nervous	91

[†] Adapted from: Pennebaker, James W., et al. "The Development and Psychometric Properties of LIWC 2007 (2007)." *Austin, TX, LIWC. Net* (2015).

Anger	anger	Hate, kill, annoyed	184
Sadness	sad	Crying, grief, sad	101
Cognitive processes	cogmech	Cause know out	730
Insight	insight	Think, know, consider	195
Causation	cause	Because, effect, hence	108
Discrepancy	discrep	Should, would, could	76
Tentative	tentat	Maybe, perhaps, guess	155
Certainty	certain	Always, never	83
Inhibition	inhib	Block, constrain, stop	111
Inclusive	incl	And, with, include	18
Exclusive	excl	But, without, exclude	17
Perceptual Processes[c]	percept	Observing, heard, feeling	273
See	see	View, saw, seen	72
Hear	hear	Listen, hearing	51
Feel	feel	Feels, touch	75
Biological processes	bio	Eat, blood, pain	567
Body	body	Check, hands, spit	180
Health	health	Clinic, flu, pill	236
Sexual	sexual	Horny, love, incest	96
Ingestion	ingest	Dish, eat, pizza	111
Relativity	relative	Area, bend, exit, stop	638
Motion	motion	Arrive, car, go	168
Space	space	Down, in, thin	220
Time	time	End, until, season	239
Personal Concerns			
Work	work	Job, majors, Xerox	327
Achievement	achieve	Earn, hero, win	186
Leisure	leisure	Cook, chat, movie	229
Home	home	Apartment, kitchen, family	93
Money	money	Audit, cash, owe	173
Religion	relig	Altar, church, mosque	159
Death	death	Bury, coffin, kill	62
Spoken Categories			
Assent	assent	Agree, OK, yes	30
Nonfluencies	nonflu	Er, hm, umm	8
Fillers	filler	Blah, Imean, youknow	9

[a]Common verbs are not included in the function word category. Similarly, common verbs (as opposed to auxiliary verbs) that are tagged by verb tense are included in the past, present, and future tense categories but not in the overall function word categories.

[b]Social processes include a large group of words (originally used in LIWC2001) that denote social processes, including all non-first-person-singular personal pronouns as well as verbs that suggest human interaction (talking, sharing).

[c]Perceptual processes include the entire dictionary of the Qualia category (which is a separate dictionary), which includes multiple sensory and perceptual dimensions associated with the five senses.