PATTERNS OF LANGUAGE USE IN MANIA

by

BESS SIRMON FJORDBAK

(Under the Direction of William A. Kretzschmar, Jr.)

ABSTRACT

This dissertation investigated the patterns of language use in mania using corpus linguistic methodology. The data were drawn from a collection of journals composed over a twenty eight year period by an individual diagnosed with mania, yielding a Master Corpus of 242,589 words which was subjected to an array of analyses, including comparison to patterns of 'typical' language use as represented in the Freiberg-Brown Corpus of American English (FROWN). The Master Corpus was also compared to a reference corpus composed of journals produced by other writers for a within-genre analysis. Multiple intra-individual analyses were conducted in order to test hypotheses concerning differences in language behavior during medicated versus unmedicated conditions, manic versus non-manic writing, and *more* versus *less* severe episodes of mania. A final analysis investigated changes in patterns of language use relative to variation in content and frequency of entries across early, middle, and late phases of a single manic episode. The data were analyzed with *WordSmith Tools* to determine patterns of word frequency and collocation, and for the derivation of keyness statistics between corpora. Results confirm all hypotheses, which postulated there would be significant differences between language as observed during manic episodes and the respective reference corpora, and likewise intra-

individual differences in all the previously described experimental conditions. Corpus analysis showed sensitivity in detecting language behavior that correlated with the diagnosis of mania, and also the language effects of treatment with medication. According to the *DSM-IV*, clinical diagnosis of mania and other mental illnesses relies substantially upon the presence of anomalous patterns of language use in a constellation of psychological, emotional and behavioral manifestations. Therefore, utility of corpus linguistic methodology in clinical applications was proposed, and a model for conceptualizing the constellation of linguistic criteria for diagnosing mania was delineated, in an effort to clarify the terminology used for identification of linguistic behaviors in mania. Areas of future research involving enhanced diagnostic validity and reliability were proffered, as were other directions for using corpus linguistic analysis in the diagnosis and monitoring of mental illness.

INDEX WORDS: Mania, corpus linguistics, language variation, WordSmith, mental illness, pragmatics, assessment

PATTERNS OF LANGUAGE USE IN MANIA

by

BESS SIRMON FJORDBAK

Master of Science, Georgia College, 1992

Master of Science, Florida State University, 1987

Bachelor of Science, Florida State University, 1984

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2006

PATTERNS OF LANGUAGE USE IN MANIA

by

BESS SIRMON FJORDBAK

Major Professor:	William A. Kretzschmar, Jr.
Committee:	Donald L. Rubin Rebecca J. Shisler Marshall

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2006

DEDICATION

For Tim

This is dedicated to the one I love.

ACKNOWLEDGEMENTS

Fellow travelers always make the journey sweeter, and the trip through a doctoral program is no exception. Thanks to Cati Brown, who was one of the first people I met when I came to UGA, and she told me one day, “I understand that Dr. Kretzschmar is looking for someone to work with him on a very special project”. Thanks to Garrison Bickerstaff, who always has a kind word and a Starburst, and whose love of language I greatly admire. Arigato gozaimasu to Alberto Centeno-Pulido, and muchas gracias a mi querida amiga, Alejandra Arboleda. Many thanks to Jaret Walton and Chris Martin, with whom I learned to do the freak-out dance while learning a whole lot about statistics, and about balancing school and real life. Very special thanks to Robin Bramlett, one of the brightest and Southernest women I’ve ever met, whom I have come to love like one of my sisters. To all of you, I will treasure your friendship always.

At the University of Georgia, I have been fortunate to work with several professors who were not only great teachers in their area of expertise, but who also provided wonderful models for being professors, and have become good friends in the process: Laida Restrepo, Joe Wisenbaker, Marlyse Baptista, Yolanda Keller-Bell, Michael Covington and the CASPR research team, and Al De Chicchis, a research collaborator and mentor, who gave me the opportunity to teach in the CMSD department, an experience that I enjoyed from beginning to end. I am also thankful to the late Dr. William Provost, who welcomed me with open arms into the Linguistics Program.

Thanks to my dissertation committee, Drs. Don Rubin and Rebecca Shisler Marshall, who agreed to serve and never hesitated to ask the hard questions. Very special thanks to Dr. Bill Kretzschmar, who is a stellar exemplar of what a major professor should be. Thank you for always starting with "Sure, you can do that", for directing (and redirecting) my work, for encouraging me to embrace life on the cutting edge of interdisciplinary research and face the challenges of working outside the box, and for setting a standard as a researcher and a teacher to which I will always aspire.

There are several people to whom I will always be grateful, without whose encouragement this process could never have happened, beginning with LMN, whose willingness to share her work and her stories served as the impetus for this entire project. Thank you for your patience and encouragement and friendship. Thanks to Tony and Karen Murray, the dear friends and faithful alumni who initiated me into the Bulldawg Nation, taught me how to bark, and how to tailgate, UGA-style. Thanks also to Dr. Franklin Johnson, who told me that getting a Ph.D. was like eating an elephant; the only way to do it is to take one bite at a time. Dr. J, you were right! To Dr. Nancy Canolty, I look forward to a long and happy collaboration with you. Special gratitude is due to Dr. Ralph Boynton, who helped me to see that I could, indeed, do this, and to the late Bruce Boron, my buddy from aerobics class at Big Canoe who coached me for the GRE. I am also grateful for the memories of Dr. Philip Dagostino and Cliff Herrell, both of whom always encouraged me to reach a little further.

In this journey towards a Ph.D., one must be surrounded by love and support, because even more than chocolate, this is what sustains doctoral students. For me, this came from my Fjordbak and Sirmon families who were always ready with exhortation, prayer, and confirmation. Thank you to my in-laws, Everitt and Ann Fjordbak, and my sisters and brothers,

Julia Waldrep, Lou Miller, John Sirmon, Lynn Hogan, Ed Fjordbak, Steve Fjordbak, and all your spouses and children, who have forgiven me for the missed family gatherings and forgotten birthdays. I love you all.

Although it has taken me a long time to realize this goal, the desire to pursue higher education was planted early in my life, and for that I thank my parents - my darling mother, Betty Ann, who surrounded us with books and read to us every day, and my late father, Wiley Sirmon, who always wanted to be a professional student. You demonstrated for us and instilled in all of us a life-long love of reading and learning, and you always made education a priority. Thank you for that. I have always wanted to make you proud.

I especially want to thank my husband, Dr. Timothy Fjordbak. You are my mainstay, my anchor, my helpmate, my biggest cheerleader, my tango partner, my sweetheart, my best friend, and the love of my life. I couldn't have done this without you.

Finally, and most of all, thank God!

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

Chapter One - Introduction

The purpose of this dissertation is to investigate the role of patterns of language use in the manifestation of mental illness. Historically, linguistic investigation has dealt with the development and use of language from a normal perspective, that is, how language typically occurs, with little accounting for variations on the theme of 'normal'. The study of language is a dynamic one, relative to the changes observed in language use, and the shifting sands of pronunciation, meaning, and intent. Linguists are compelled to monitor such changes as they correlate with geographic, socio-political, and economic variation, and generally tend to regard pathological deviations as anomalous, and not representative of the populations at large. Further, linguistic study has been noted to rely on the observations of a single individual's language use as characteristic of a region, or in some cases, on the intuitions of an individual researcher to represent theoretical models of language for all populations (Kretzschmar, in preparation).

On the other hand, professionals in the clinical realm dealing with those variations in language use as markers of mental illness have reported on patterns that appear to be different in their diagnostic significance for the classifications of pathologies, without necessarily having a complete grasp on what is considered to be 'normal language', seeming to assume that being a successful user or clinical observer of language qualifies one to describe its structure and use. Thus, without established linguistic support, the parameters used to define patterns of mental

disturbance are based on observation of behaviors, many of which are linguistically-based, and are documented for the purpose of initial diagnosis, as well as indicators of progress and response to treatment. At a very fundamental level, the use of descriptive terminology is one that needs refinement; for example, use of terms such as *poverty of speech* and *thought disorder* are incorporated in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV, American Psychiatric Association, 1994) and have become part of the vernacular in mental health, with general agreement on the matching symptomatology (cf. McKenna & Oh, 2005). However, less frequently used terms only add to the confusion and distortion of meaning, as there is neither an apparent quantifiable mechanism nor linguistic basis employed in describing speech which is *gibberish* (E. O. Chaika, 1990) or *wooly* (Reilly, 1975).

The end result is that linguists have not fully embraced the study of language variation outside the range of what is considered normal, and clinical disciplines (most particularly psychology and psychiatry) have not embraced the terminology nor theoretical frameworks offered by years of linguistic research. This project attempts to bridge the gap from the linguistic world of applied language science involving the identification of patterns of language use and generative theories in which the human is removed from formulation, into the clinical world of investigation of patterns of language use in the presence of neurological, cognitive, or mental processes of illness, injury or disorder. In an effort to link the fields of inquiry, this work is written so that the practitioners typically involved with observation, identification, and intervention of disorders based on "abnormal" use of language will understand the underlying linguistic constructs, and conversely, so that linguists will become more familiar with

the diagnostic manifestations of clinical and pathological patterns of language. In this process of discovery, the field of communication sciences and disorders can also benefit. In the United States, the conception of mental illness representing a communication disorder is a novel one, and although it is within the scope of practice for speech-language pathologists (American Speech-Language-Hearing Association, 2001), there are few who practice in this arena. In other parts of the world, intervention in populations with mental illness and in psychiatric facilities is more readily recognized as within the purview and scope of practice of speech-language pathologists, who function as members of interdisciplinary treatment teams (France & Muir, 1997). All of these disciplines have much to learn from each other.

This particular investigation of language use focuses on the expression of linguistic form as observed with a diagnosis of a Major Mood Disorder: Bipolar Disorder – Mania Only. Mania is recognized as one of the polar extremes experienced typically in a Bipolar Disorder, although in some unipolar cases, the contrasting cycles from melancholia to euphoria swing in only one direction. The DSM-IV (1994) describes the pathological linguistic aspects of mania as inclusive of speech which lacks cohesion and is difficult to interpret, with word selection sometimes governed by phonological properties (i.e., alliteration) rather than semantic content; and 'flight of ideas', in which there is an accelerated rate of production with abrupt shift in topic, sometimes to the point of incoherence. Concomitantly, there also may be cognitive-behavioral changes observed in

individuals with mania[1], including an increase in organization and planning behaviors involving spending money, developing grand and expansive ideas, and increased engagement in social interactions, all of which may be noted along a spectrum of severity. These linguistic changes are observed in the context of a distinct period of mood disturbance and neurovegetative changes in which there are atypically and persistently elevated and expansive behaviors that may become severe enough to impair interpersonal or work-related function.

This study involves the analysis of a series of hand written journals prepared across a span of nearly thirty years, unique because of the author's diagnosis of Bipolar Disorder – Mania Only. The serendipitous nature of the acquisition of this data set permitted an opportunity for in-depth linguistic investigation into a phenomenon typically understood as a disorder or mental illness. Analysis of the patterns of language use in mania has been historically conceptualized from a clinical perspective as a nomothetic representation of psychological function based on specific criteria, and this study takes that perspective with quantitative analyses of language production. However, this dissertation also provides an idiographic lens of individual variation through which manifestations of mania in the written expression of this individual may be qualitatively examined as situated in the contexts of her personal identity.

In this investigation, the DSM-IV definition will be examined in comparison to the linguistic behaviors evidenced in written text. Of note, specific patterns of thematic discourse emerge in the journals, most of which have a fundamental relation

[1] The tem "manic" used as a noun to describe a patient population is historically accurate, but in mental health and other clinical professions, person-first terminology is now preferred. Although ostensibly more awkward, it will be used exclusively in this paper.

to the marking of the passage of time. This individual frequently notes ideas taking shape, referencing these to prior journal entries, and across the years, she also describes particular interest in the utility of prime numbers as a mechanism for understanding herself and her environment.

As a follow-up to the nomothetic diagnostic perspective, the way in which the writer situates herself in the contexts of her life - as an academic, a parent, and a citizen of her community, among others – will be examined, comparing her use of language in journals with corpora representative of what would be considered 'normal' language use. The goal of this work is to provide a useful lens for examination of the linguistic and inherent psychosocial manifestations of mania as represented by or correlated with the written expression of this individual during different psychological states along the aforementioned severity spectrum, as well as in comparison with reference corpora representing more typical patterns of language use.

Participant and Data

The participant in this study (hereafter identified as LMN) is a woman who was diagnosed with Bipolar Disorder – Manic Type in 1985, but in 1978 (prior to her formal diagnosis), she began experiencing periodic "episodes" in her life during which she reported feeling considerably more energized and creative and required little sleep. Throughout these early episodes, she began keeping a series of journals to document her thoughts and ideas. Over the past twenty-eight years, this has become a consistent pattern as she has experienced recurrences of her manic state

several times per year, each time delineating detailed plans and germinating concepts, with observations of herself and the disease process, reported with time and date signatures for her stated purpose of contributing to the greater knowledge and understanding of this diagnosis.

LMN is a retired academic researcher, trained in the scientific method, and describes the process of keeping journals as her effort to record her experience for posterity, with the eventual goal of compiling the work into a book or documentary, for the purpose of bringing the nature of this diagnosis into better public awareness with the hope of reducing some of the associated stigma. She reported feeling a need to carefully document her experience with mania as less of an "obsession" or a "commitment", and more toward something she has felt "compelled" to do, for the purpose of contributing to the greater knowledge and understanding of this diagnosis. Further description of the participant, her history, and situation will be elucidated in Chapter Three.

The corpus used in the current analysis was built from selections of a larger data set, comprised of 28 years of personal journals, currently numbering well over 100 volumes. The selections are representative of the entire span in an effort to track any changes evidenced across time. Throughout her experience, LMN's habit has been to set the current journal aside after it is completed, as she typically does not read back over her work or subject it to any sort of editorial process, instead storing it for future reference.

Through a series of personal and professional contacts, the participant offered this set of papers to a well-known university-based corpus linguist for the purpose of

research and analysis, as a means of meeting her goal for increased understanding of mania. For participation in the current project, she helped to draft and subsequently signed the consent for participation form as required and approved by the Institutional Review Board Human Subjects Office at the University of Georgia (included as Appendix A).

The journals she has produced are typically small bound books, with variable patterns of handwriting in blue or black ink. On each page, she has recorded the month, day, and year, typically with references to time in an hour:minute:second pattern at the beginning and ending of each entry. She writes on several different levels, sometimes as reminders to herself, lists and descriptions of daily activities, and sometimes taking a metalinguistic and reflective perspective on her behavior and her contributions to the world at large. There are several consistent themes in her journal entries, including the generation of ideas, which are often recorded, but not fully explored until later in the writing process. This is a behavior she describes as a means of warding-off the tendency to become tangentially focused on one thing to the exclusion of all other thoughts. Over the course of time, one of her particular interests that has emerged is the notation of prime numbers, and her pattern is never to end a thought on a prime number, so she will continue recording the time at the beginning or end of an entry until it gets to a non-prime designation (See Appendix B for examples of prime time notations.).

The journals are also remarkable for a series of non-linguistic markings, which she refers to as “automatic sketches”, described as reflecting changes in her perceived physiological state. These sketches vary in length and quality, but typically involve a

pattern of jagged marks made in a back and forth motion across the pages in a non-overlapping manner, which eventually resolve into smooth and rounded shapes. The participant has never been diagnosed with seizures, but she believes the sketches involve a change in her level of conscious arousal and awareness. There are also examples of "circular writing" in which she will continue writing in sentence form, but will change the direction of the paper, generating writing that moves across the page in a variety of random shapes, frequently involving inward-spiraling concentric circles. Across the course of an episode, which may last from a few days to a few weeks in length, there are also notable changes in the size and shape of her handwriting.

The journals are not the only medium through which she has recorded her experiences with mania. At certain periods, she would record her thoughts onto audio cassettes, or use video recordings (8-mm, analog, and digital video) to archive her process. She also has extensive collections of Polaroid photographs, as well as loose sheets of notes with random observations, newspaper clippings, and other ephemera collected in formats not conducive to organized journals.

The Significance of Mania

Approximately one percent of the US population over the age of eighteen is diagnosed with Bipolar Disorder, an undetermined subset of whom experience solely manic episodes (DSM, 1994; National Institute of Mental Health [NIMH], 2001). Bipolar disorder is considered along a spectrum of expressions; notably, the mild to moderate phase of mania may be experienced as hypomanic states in which energy

levels are raised, productivity is enhanced, and mood is pleasant. Thus, the reported incidence of actual diagnoses of mania may be keyed to the more pronounced phases which are more severe and debilitating in nature as the cycle progresses.

As noted, among the diagnostic criteria for the subset of patients experiencing only the manic symptoms of Bipolar Disorder are several which are primarily linguistically-based. Additionally, people experiencing mania may often be portrayed as irritable, restless, distractible, and aggressive, while denying that anything is wrong or different, as the literature suggests that people with mania may demonstrate decreased insight into their experience (Akiskal et al., 2001; Dell'Osso et al., 2000; Ghaemi & Rosenquist, 2004). Clinically, a manic episode is diagnosed from a constellation of manifestations over a period of time, involving observable signs and reported symptoms.

Mania is most frequently managed pharmacologically, historically with the psychotropic lithium carbonate (cf. Katzung, 1995; J. B. Murray, 1985), a medication which must be carefully maintained at therapeutic levels to avoid subclinical symptom breakthrough or high levels producing lithium toxicity, a potentially fatal overdose. This medication may sometimes be used in concert with anticonvulsant medications, the combination of which has been shown to stabilize extreme swings in mood, normalize patterns of function (including language use), and reduce the frequency of relapse. Additionally, certain antipsychotic medications are sometimes required when treating acute cases of mania. Psychotherapeutic intervention is also effective in the treatment of mania, targeting cognitive-behavioral therapy, psychosocial education, family-centered intervention, and treatment compliance.

In recent years, researchers have paid particular attention to manifestations of bipolar disorder in the aging population, such inquiries having relevance to the study of journals produced across nearly three decades. Depp and Jeste's (2004) review of the literature noted changes across the developmental life-span, reporting longer episodes and shorter intervals of remission as patients age, with the suggestion that medications may become less effective in managing the symptoms across time. There does not appear to be a progressively deteriorating course for this diagnosis relative to cognitive function, but on the other hand, the psychological and emotional symptoms do not appear to dissipate over time. Generally, the manifestations of mania do not appear to be different in the population over the age of fifty as compared to younger people with the same diagnosis. Similarly, there appears to be little clinical difference between the manifestations of people experiencing early onset (defined as less than forty years old) versus late onset of bipolar disorder (Depp et al., 2004). However, Shulman (1997), as well as Depp and his colleagues contend that there is an increased prevalence of neurological involvement in elderly patients experiencing first episodes of mania, particularly those experiencing late onset, possibly secondarily related to changes in right hemisphere function and patterns of disinhibition, suggesting potential implications for linguistic breakdown, most notably in social/pragmatic function.

On a more positive note, the diagnosis of mania has sometimes been correlated with increased creativity, attributed across time to certain poets, writers, composers, artists, as well as to historically significant political, military and religious leaders (Carreno & Goodnick, 1998; Hershman & Lieb, 1998). As an example,

Ludwig van Beethoven is reported to have dealt with several protracted periods of depression during his life. However in 1801, he experienced a change in mood resulting in a cycle of increased productivity during which he wrote, "I live entirely in my music; and hardly have I completed one composition when I have already begun another. At my present rate of composing, I often produce 3 or 4 works at the same time." (Hershman & Lieb, 1998, p. 76). On a less grand scale, but still as a noteworthy example, Andreasen and Powers (1975) compared a sample of patients with mania to creative writers, noting similarities in their use of "behavioral and conceptual overinclusion", but differences in the quality of thought processes. Creativity is variously measured along dimensions of novel creation, complex thought patterns, and motivation, based in theories of Janusian cognitive processes, bi-hemispheric hyperconnectivity, or biological thermodynamics (cf. Carreno & Goodnick, 1998), all of which, it is suggested, may be enhanced by the experience of mania.

This is, of necessity, a very brief review of the significance and impact of mania[2]; the more relevant questions investigated by this dissertation deal specifically with the processes and patterns of language use in the presence of mania.

The Problem: Mania and Language

The research addressing the linguistic changes present in a diagnosis of mania is scarce, even though the criteria for identifying this disorder are based substantially on linguistic and other cognitive-behavioral changes which are ostensibly

[2] For more complete information on mania and other mental illnesses, see NIMH (2001), or visit the National Institute of Mental Health home page at www.nimh.nih.gov, or the National Alliance for Mental Illness website at www.nami.org.

linguistically mediated. The literature dating from the 1950's through the present demonstrates inconsistent results, the outcome of which speaks to the failure systematically and objectively to define how aberrant language truly manifests in mental disorders, even though language has been a primary diagnostic marker for the identification of such psychological disturbances. Among the research questions which emerge are these: what are the measurable changes in patterns of language use in the written discourse of a person diagnosed with mania, and how can these be described and investigated in a systematic and ideally replicable manner?

Subjectively, LMN describes her awareness of the favorable changes she perceives during manic episodes, and she has documented these throughout the process:

> "I've long recognized that I become more 'organized' in one of these 'creative periods'. Seem to clean and straighten more. With research, get more organized in method of storing, ideas about research become more focused. Also, tend to be more interested in wardrobe and personal appearance." (26 March 1984, 4:58:45 a.m.).

As keen an observation as this qualitative description may be, this study also provides a quantitative description of such changes, comparing the individual to larger populations, as well as comparing variations across time and genre in her own work, situating it all in the context of the multiplicity of roles she has filled across her life.

Corpus Linguistic Analysis

The primary method of analysis employed is that of corpus linguistics, as an analytical process involving investigation of meaning as situated in language use, based in large bodies of text. There is debate as to whether corpus linguistics is a theoretical subspecialty within the larger domain of linguistics, or if it is better regarded as a methodological pursuit adaptable to virtually any type of linguistic inquiry (Meyer, 2002). Generativists tend to reject corpus linguistics on the grounds that it is not hypothesis-driven in the manner of experimental science (cf. Chomsky's comments in an interview with Andor, 2004), and would be more appropriately regarded as opportunistic data-mining. However, Neo-Firthian theorists (Kretzschmar, in press; Stubbs, 1996) and functionalists agree that the analysis and discussion of language use is best situated in the context of language actually being used, and have embraced the methods and theories ascribed therein. Whatever the preferred lens through which corpus linguistics is viewed, the theoretical orientations and applications are evolving, and center on the idea that meaning in language is identified through repeated events which can only be recognized through analysis of large bodies of actual occurrences of words in attested and authentic (as opposed to intuitive or invented) text (cf. Hockey, 2000; Stubbs, 1996).

Corpora are defined broadly as collections of any naturally-occurring text, but Meyer (2002) narrows the definition somewhat to "a collection of texts or parts of texts upon which some general linguistic analysis can be conducted." (p. xi). Corpora may be compiled from any number of sources to address a range of research questions. The corpus utilized in the current research project is best classified as an

opportunistic or convenience sample, as the participant came unsolicited to the university and offered the entire data set. There are several sets of reference corpora to which this data is compared, including a monitor corpus, which represents diachronic analysis of within-subject changes over time.

In corpus linguistic analysis, there are levels of analysis, the most basic of which is *collocation*, which involves word frequency counts and patterns of co-occurrence; Stubbs (1996) noted that words are not selected in isolation and that patterns of co-selection are indeed meaningful. The next level of analysis is that of *semantic preferences*, through which conceptual word associations are identified and addressed (e.g., words related to the passage of time, such as *morning*, *late*, *o'clock*, *before*). The most complex level of analysis involves identification of *discourse prosodies*, in which the attitude or evaluation of a collocative node is interpreted as having positive or negative affective connotation in a particular context.

Corpus work allows for examination of samples of language use in context, from which meaning can be derived based on structural variations and interpretation of patterns that emerge through analysis of linguistic substructures. Use of corpus linguistic methodology is appropriate for this study because of the nature of the data - a series of written texts spanning several decades transcribed into a large body of work - as well as the research questions which address potential differences in language use in the presence of an identified diagnosis. It is through repeated patterns that identification and justification of 'unusual' or pathological patterns of language use can be identified, and these patterns will only be evident via corpus study.

Hypotheses

This study proposes a linguistic investigation in consideration of the identified problem of defining language use in mental illness according to the DSM-IV criteria, the significance of mania and other manifestations of disordered language in the spectrum of mental illnesses, and the variations in the expression of language in written texts as produced across a span of almost thirty years as identified through corpus linguistics. The primary dependent variable to be examined in this research is changes in word frequency as a marker of variation in language use. Based on the questions raised, the hypotheses to be considered in this dissertation are these:

1. There are measurable differences in the written texts of this participant as compared to corpora of 'normal' language use across genres using:
    a. the Freiburg-Brown Corpus of American English (FROWN) corpus; and
    b. journals of the same genre produced by other individuals.
2. There are significant within-subject differences in patterns of written language use as measured:
    a. during unmedicated versus medicated periods of time;
    b. between manic versus non-manic writing, incorporating other samples of text generated by the same writer (i.e., personal letters, e-mail correspondence);
    c. during episodes of contrasting amplitude, addressing levels of acuity experienced as more or less significantly manic; and
    d. within a given manic episode, between early, middle and late phases.

Overview of the Dissertation

As this introductory chapter has provided an overview of the project, further chapters will explore the topic in more depth. Chapter Two provides a review of the literature, historical and current, relative to manifestations of language use in mental illness in general, and mania specifically; an exploration of the journal as expressive genre; and establishment of the idiographic/nomothetic paradigm for analyses of these texts. Chapter Three explicates the methodology, first with an in-depth description of the participant and her history, and explication of the analyses involved in the investigation of her patterns of writing, including comparisons to external corpora representing 'normal' language use, to writing samples in other journals, and within-subject comparisons of writing style in unmedicated versus medicated periods, as well as addressing varying levels of acuity, including non-journal writing samples (e.g., personal letters, e-mail). The analyses will take the form of descriptive inferential statistics addressing word frequency counts, comparison of patterning in semantic categories across conditions, and analysis of keyword frequency between corpora.

Chapter Four addresses the results of the study, beginning with statistical analyses and in-depth descriptions and interpretation of the Master Corpus compiled from LMN's journals, and continuing with comparative studies employing external corpora as points of reference to situate her writing in contrast to patterns considered 'typical' in general American English as well as in the genre of journal writing. Chapter Five continues the analyses, targeting intra-individual comparisons of LMN's patterns of language use as identified in changes in word frequency between

conditions she has experienced across the twenty eight year process. Related to these analyses, Appendix D offers a brief treatment of the observer's paradox as manifest in this research, with discussion of the Heisenberg and Hawthorne effects, and comparative samples of how this was demonstrated and accommodated for in this data set. The final chapter is a discussion of the findings, noting the potential for the applied uses of corpus linguistics in diagnostic formulation and in clinical practice. Limitations of the present work and further directions for research are offered, both in terms of this data set, and for this type of analysis in general.

Chapter Two - Literature Review

In an effort to gain a clearer perspective on the research questions posed, it is necessary to consider the seemingly disparate topics in this discussion of mania and journal writing at converging points in the body of current literature and theoretical orientations. The literature review begins with a discussion of patterns of language use in mental illness, first from a fairly broad historical perspective, then specifically referencing the topic of mania and situating the idea of variation in language as a manifestation of mental illness. With this as background, the review then foregrounds a discussion of journal as genre, addressing history of the form, writing practices, construction of audience, and description of a taxonomy for understanding differences within journal-keeping style. As a technical stylistic note, use of the verb form *journaling* or the titling of the person writing as *journalist* will be avoided in favor of the preferred and more descriptive *journal-keeping (-keeper)*. The final section of this chapter will discuss the juxtaposition of language and mania with journal as genre, the constructs of disordered pragmatics and discourse, and the idiographic and nomothetic frames of reference as expressed in the written discourse of this individual.

Language and Mental Illness

In the United States, approximately two percent of the total population is diagnosed with major mental illness of either the bipolar or schizophrenic type

(American Psychiatric Association, 1994). Analyses of patterns of language use have historically been employed as inclusion criteria for identifying such disturbances, providing fertile ground for linguistic investigation, to date most notably in schizophrenia. A recurrent complaint in the literature deals with the difficulty in accurately defining the linguistic phenomena under study (cf. Wykes & Leff, 1982), with poor diagnostic descriptors and irreproducible results. Few of the studies addressing language use in mental illness have been written by linguists, and hence, such investigations and the description of the results obtained leaves much open to interpretation. This has resulted primarily because there has been inconsistent and inventive appropriation of terminology between the linguistic and clinical domains, as exemplified by terms such as 'vague and wooly'(Reilly, 1975); 'extreme fabulizing' (Daniels et al., 1988); 'disturbed' (Rosenberg & Tucker, 1979); or 'verbigeration' (Andreasen, 1979), among others.

Rochester and Martin (1979) provided a seminal analysis of conversational breakdown in schizophrenia, asking at what point the listener stops making allowances and accommodations to unusual patterns of discourse, and becomes unable to comprehend the speaker's message based on cohesion errors produced in a model after Halliday and Hasan (1976). They identified patterns of predominant reliance upon lexical cohesion markers in the language produced by people with schizophrenia, rather than the broader array of cohesive ties including reference, substitution, ellipsis and conjunction. In addressing such patterns of discourse breakdown, these investigators presented a fundamental terminological question regarding the privileging of *thought disorder* as a diagnostic marker, since the noted

changes manifested in linguistic coherence errors were based on deficits in internal cohesion. Chaika and Lambe (1989) were critical of Rochester and Martin's analysis and their descriptive terminology, and replicated their work with modified methodology, yielding differing results, suggesting that communicative breakdown was more evident at the level of speech disorders in schizophrenia, as there were no apparent underlying systemic language disorders. They argued that cohesive ties occurred with equal frequency when produced by people with schizophrenia as they did in 'normal' language, although there was less anaphor noted. Although the use of such linguistic structures does not distinguish impaired versus normal populations, Chaika and Lambe suggest that the differences in narrative formulation observed in patients experiencing psychosis were related to other factors, such as distractibility from external stimuli leading to tangential digressions in conversation.

Alverson and Rosenberg (1990) reviewed Rochester and Martin's work, and were also critical of their findings regarding the decontextualized nature of their linguistic analysis, suggesting that claims of incoherence cannot be made based on language not situated in meaningful context. They describe cohesion and coherence in a manner more closely related to Halliday and Hasan (1976) than that observed by Rochester and Martin, noting coherence as the quality of "hanging together at the level of theme", and that cohesion is situated in the surface structure of sentences in discourse (p. 175). These authors argue that coherence requires the gestalt of language function in use, and this can only be analyzed in the context of larger scale discursive practices since there is, at best, only a tenuous relationship between cohesion and coherence. Fine (1995) responded, noting that cohesion is but one

analytical tool, and that it is likewise only one of the influences on coherence in expressive language, suggesting instead a systemic theoretical approach, as language operates in complex situations, rather than in isolation.

Chaika (1990) offered a broader spectrum look at linguistic behavior in the population diagnosed with schizophrenia, noting "disintegration of linguistic ability" at virtually every level of function, sparing only phonemic processes (p.183). For example, she described the preservation of phonological processes, noting that even in the presence of neologisms or 'gibberish', the rules are still maintained, even though meaning cannot be decoded from the context. At the level of semantic function, glossomania, a process of associational chaining whereby words are selected based on shared meaning which progressively moves further away from the intended meaning, is compounded by the complications of intrusive thoughts that also contribute to sudden and unexpected topic change. Related to this, Chaika's explanation of 'word salad' suggests that normal words are used in incoherent sequences, the combination of these factors converging to yield communication which is pragmatically and functionally inconsistent with the context. Chaika argues that the linguistic breakdowns in schizophrenia represent a failure in subordinating competing stimuli or thoughts from different cognitive and communicative levels at different times. Alternatively, Kramer, Bryan and Frith (1998; 2001) narrowed the focus of investigation to a single domain, suggesting that the breakdown noted in linguistic processes of patients with schizophrenia is more appropriately addressed and remediated from a perspective of discourse frames involving semantic and/or pragmatic language use in context, rather than from the point of syntactic structure.

They developed an intervention program based on increasing self-monitoring of narrative, procedural/expository, and conversational language in a model whereby comprehension preceded production. Although their sample size was small, their outcome data indicated that in a relatively short time, there was measurable improvement in the ability of patients with schizophrenia to provide sufficient information and get to the point without extraneous material, consistent with Grice's (1975) Maxims of Quantity and Relation. Although this population was diagnosed with an illness that will not be cured, the implications of such work suggest potential for improving the communication function of these patients, and thereby improving their quality of life.

Acknowledging the similarities in disordered patterns of language use across diagnostic classifications, Peuser (1987) provided a descriptive comparison of the linguistic output observed in *schizophasia* and *jargonaphasia* [his preferred spelling], offering criteria for a differential diagnosis between the two disorders. The single subject identified with schizophasia produced language which was notable for intermittent changes from normal to deviant, with episodic breakdown in coherent sequencing of words in sentences, and semantic paraphasias or word substitutions. This participant also experienced hallucinations and demonstrated overall decreased psycholinguistic competence related to functional language use. The patient with jargonaphasia produced continuous and unremitting fluent neologistic jargon, with no apparent awareness of his expressive language deficits, and severely impaired comprehension. His functional communication was marked by intact use of phatic speech, but otherwise complete breakdown in linguistic competence. The ability to

identify and quantify patterns of impaired language use in terms of neologisms versus paraphasias, or quality of functional comprehension has implications for differential diagnostic classification and management of such disorders.

Cognitive processes in mental illness

Barr, Bilder, Goldberg, Kaplan and Mukherjee (1989) approached the analysis of schizophrenic speech from the neuropsychological assessment perspective, arguing for the presence of frontal lobe dysfunction, and possibly subcortical disturbances. They noted parallels to speech patterns associated with certain types of aphasia, but presented stronger evidence for executive function breakdown as manifest in linguistic patterns of repetition at the phonemic, syntactic and semantic levels of function. Exemplary of the connection between patterns of language use and frontal lobe function, perseveration was observed in both schizophrenia and aphasia, marked as distractibility and an inability to shift from a previous topic to a new topic. Consistent with other research in the area, the significant breakdown identified by Barr and colleagues was in the area of self-monitoring. McKenna and Oh (2005) also presented perspectives based on recent research into the analysis of disorders at the intersection of cognition, language and communication, and the neuropsychological influence upon the linguistic manifestations of schizophrenia and other patterns of mental illness. They propose a dyssemantic hypothesis of deteriorating patterns of language use involving loss of word meaning, topical derailment secondary to overinclusive thinking and interference from semantic priming, and decreased semantic knowledge. This constellation of linguistic symptoms may manifest as a

semantic memory impairment in schizophrenia, experienced as a degraded store or impaired lexical access. From a linguistic perspective, the neuropsychological deficit approach has utility for identification of patterns of language use relative to executive functions breakdown (e.g., perseveration, word retrieval and substitution errors, loss of topic).

A fairly large body of research has addressed the issue of thought disorder from a cognitive orientation, notably Thomas and Fraser (1994), who describe the multi-level disturbance in thought disorder as abnormalities in thinking, language processing and social cognition, convincingly making a case for the term 'communication disorder' as more representative of the global nature of the deficits. Thomas (1997) further argues that thought disorder in mental illness is a misleading construct for several reasons. He suggests that not all speech conveys thought, but even if that were the case, would disordered speech necessarily equal disordered thought? The introduction of the confounds of intelligence, culture, semantic priming, and contextual variables in individual cases further impact the reliability of assessment of thought disorder. Hence, Thomas contends that thought disorder is not a homogeneous entity, supplying further support for the notion of a broader spectrum communication disorder in mental illness. Use of this term privileges linguistic formulation as equal to cognitive/behavioral function, and necessitates assessment of semantic and pragmatic function, in particular.

Theory of Mind (TOM) offers another cognitive perspective on understanding meaning in language use in mental illness. TOM, also referred to as *mentalizing*, involves an individual's ability to explain or predict the behavior of others secondary

to appreciation of the mental state of the others, and is also closely linked with appreciation of pragmatic function and linguistic presupposition (Langdon, Davies, & Coltheart, 2002). Corcoran and Frith (1996) presented forty nine patients with schizophrenia with a series of stories involving observation of Grice's (1975) maxims, appreciation of implicature, and politeness principles in a forced-choice format. They noted decreased awareness of conversational rules in all patients, but reported that those with primarily negative symptoms (e.g., poverty of speech and content, decreased social function) were most likely not to follow the rules, suggesting a TOM deficit relative to expected conduct. Bazin, Sarfati, Lefrère, Passerieux and Hardy-Baylé (2005) also argue that the communication deficits manifest in schizophrenia, and to a lesser degree in mania, are due to cognitive deficits related to TOM and faulty integration of information from the context of the interaction.

Langdon, Davies and Coltheart (2002) investigated TOM in schizophrenia relative to interpretation of metaphor and irony. They provide further evidence for frontal lobe involvement based on the linguistic manifestations of schizophrenia, observing the difficulty in suppressing prepotent tendencies, a skill required for both TOM processing and interpretation of figurative language. Langdon and colleagues suggest that problems with TOM are directly related to problems with appreciation of irony, as both require a level of abstract interpretation. Related to levels of abstract processing, Tirassa (1999) takes the unusual perspective of suggesting that communication is not language-based, but is instead socially-based, more accurately described as a cognitive competence, breakdown of which is clearly a pathognomic marker of mental illness.

Rosenberg and Tucker (1979) also speak to the subjective nature of the diagnostic criteria of thought disorder, noting that under certain stresses, even 'normals' evidence linguistic patterns of thought disorder. These authors suggest, because the linguistic hypotheses typically used in clinical realms are largely ineffective in differential diagnosis, a move towards analysis of the content of language. Such analyses yielded significant differences between male and female patients in their language use, as compared to stereotyped sex roles; unfortunately, these researchers contribute to the mix of unquantifiable terminology by describing the use of language in female patients with schizophrenia as 'gushier' than that of their male counterparts. Terminology notwithstanding, Rosenberg and Tucker report that as psychotic episodes remit, the semantic content of language approaches more socially acceptable parameters, noting that it is the deviations from the expected that yield the perceptions of linguistic breakdown.

Language variation in mental illness

Situating linguistic function in the context of language variation is another apposite model for analysis of mental illness. In considering the fluid nature of language and the interaction of personality variables, social influences, and cognitive functions, the challenge lies in formulation of diagnostic criteria for defining behavior based on what would be considered within the boundaries of what is usually expected in speech. Those criteria for defining mental illness are historically uneven, with differences reported between disciplines and even between countries, which is not completely unexpected, since 'appropriate' function is defined within each discourse

community (cf. McKenna & Oh, 2005). Hence, the spectrum of linguistic markers of mental illness is variable in nature, without common etiology, and manifesting along a continuum of attributes (Covington, He et al., 2005). Whether such language is incapacitating depends on a number of factors and circumstances.

Frow (2001) makes a strong case for identification of mental illness as a metaphor, by which the constellation of unusual behaviors and aberrant language use symbolize a disease process which is attributed to some biochemical breakdown. He suggests that thoughtful analysis of the linguistic content of patients diagnosed with schizophrenia or other psychoses may reveal meaningful content, but this requires careful listening and interpretation by the treating clinicians. The patient is vulnerable in the identification and application of specific roles in this milieu, with the power differential inherent in such institutions, and rather than unpacking the specific dynamics necessary for understanding the situation, he argues that the tendency is for those holding the power to include the individual in the heterogeneous mix that falls under the larger umbrella of mental illness, rather than appreciating individual variation. Ribeiro (1994) notes that even in the presence of florid hallucinations and verbal incoherence, the subject she studied was incorporating the perspective of individual agency in multiple frames of reference (e.g., assuming the role of child while speaking to her dead grandmother, then assuming the role of parent while speaking to the doctor as child). The patient was noted to observe certain cultural and linguistic variables throughout her psychotic discourse which were judged suitable for the context, however contrived her interaction or out of touch with reality she may have seemed.

Billow, Rossman, Lewis, Goldman, and Raps (1997) described an interaction between the writer James Joyce and the psychoanalyst Carl Jung. Jung was treating Joyce's daughter Lucia, for what was most likely schizophrenia. Joyce defended his daughter's rambling, tangential and paraphasic verbal expression as similar to his own patterns of writing. Jung drew the analogy of Joyce and his daughter as heading towards the bottom of a river: "You swim, she falls". The question becomes, then, at what point does creative or idiosyncratic use of language become pathological?

The answer lies in the adaptability of the individual to their circumstance and his or her application of language variation. Psychotic mental illness is defined by impaired reality testing, such as noted in delusions or hallucinations, and is typically manifest, in part, through expressive language. The literature demonstrates that patients with schizophrenia experience breakdown at every level of language: phonological changes, morphological deviations, simplified syntax, semantic substitutions; pragmatic deviations, and discourse impairment (cf., Ribeiro, 1994, Covington, et al., 2005; Peuser, 1987; Chaika, 1990; Thomas & Fraser, 1994 ; Langdon, Davies, & Coltheart, 2002). Parenthetically, in consideration of the basic linguistic constructs of phonology, morphology, syntax and semantics, written production is typically considered superior in form to verbal production due to the added level of self-monitoring required that is usually not seen in extemporaneous verbalization; however, in cases of impaired reality testing, coherence may be equally impaired as that noted in verbal expression.

In the presence of such pervasive deficits of language function, the ability to perform at a level considered acceptable or standard, regardless of the values

established in the community of practice, can be severely limited. Again, mental illness falls along a spectrum, and there is typically enough flexibility within any context for an individual who is considered "odd" or "eccentric" to find a comfortable social niche. In clinical terms, this may involve, for example, schizoptypal personality features or patterns of interaction consistent with Asperger's Syndrome, but when the manifestations of thought disorder (Andreasen, 1979) or language production impairment (Barch & Berenbaum, 1996, 1997) reach a proportion that judgment and safety become impaired, best practice (and common sense) dictates intervention.

Differences in language use, based on cultural, regional, or community practices can only be understood in context, as framed by what is defined as normal in any given setting. Ribeiro (1994) describes the variations in physical proxemics that were acceptable for the patient under observation, noting that in the patient's 'frame' as parent speaking to the doctor as child, it was not necessarily inappropriate for her to touch the doctor's face to gain her attention. Had such a cultural custom not been appreciated in that context, inappropriate touch and violation of personal space would have been identified as a pathological clinical marker. Chaika (1990) makes reference to the pattern of 'glossolalia' in psychotic expression (i.e., chaining of words in connected discourse related phonemically; akin to neologistic paraphasia), noting that in certain charismatic religious frames, such behavior is considered 'speaking in tongues', a transcendental meditative state in which one draws closer to God. She goes on to describe both situations as a suspension of psychoanalytic ego function, although Chaika seemingly appears to attribute intentional behavior in both

circumstances (and presumably, no one would choose to be psychotic). Likewise, other manifestations of psychosis could be variously attributed to religious or political themes (e.g., hearing voices makes one a prophet in some cultures).

In the *zeitgeist* of language variation theory, it is appropriate to consider the individual differences noted in patterns of language use within the context of the individual's community of practice. The ability to shift between communities is necessary for adaptable function, and if such discriminant flexibility is not within one's repertoire, this does represent a pathological circumstance. The *Dictionary of Psychology* (Chaplin, 1985) describes the inability to adapt successfully in an environment as maladjustment, or a mental disorder. Further, *Blakiston's Gould Medical Dictionary* (Gennaro & Gould, 1979) defines maladjustment as:

> "A state of faulty or inadequate conformity to one's environment, due to the inability to adjust one's desires, attitudes or feeling to social requirements."

The degree to which the pathology is identified represents the chasm between what is considered 'acceptable, but unusual' versus what is considered wrong for the circumstance and potentially problematic. Mental illness manifests as a multi-layered spectrum of disorders, and in the nebulous environment of language variation in communities, the ability to incorporate deviations in related language behaviors is best considered in terms of what is 'average', rather than what is 'normal'.

Assessment

Accurate diagnosis of mental illness typically involves implementation of language or linguistically-mediated assessment tools, many of which have been in continuous use since early in the last century. One of the oldest and most well-known tools is the *Rorschach* (Rorschach, 1921). This projective technique involves interpreting the participant's responses (percepts) to a series of ambiguous ink blots, which are explained in degrees of psychopathology. There are many variations of scoring and interpreting this venerable tool, all of which are based on the patient's perceptions and verbal responses as a means of structuring and organizing their world view, both internal and external. The reliability of the Rorschach technique has been determined based on implementation of specific scoring protocols; however, the validity of this tool remains open for debate, as there are questions about quantifying open-ended responses, and thus about the purpose of the technique, the interpretation of single versus multiple responses, and the efficiency of administration and scoring. The Holtzman technique (Holtzman, Thorpe, Swartz, & Herron, 1961) represented an attempt to provide increased structure to the inkblot type of assessment, allowing for only one response per card, and is recognized for its scoring reliability in this type of projective technique. Responses are scored according to semantic content and the inherent associations (e.g., food, movement), patterns of embellishment or elaboration, or discourse themes (e.g., aggression), which are impacted by intellectual, emotional and other factors. Responses are also analyzed in terms of 'part/whole' constructs, and the amount of descriptive detail provided. Notably, the presence or absence of

frequently-occurring verbal responses is used to gauge the continuity of basic reality testing.

The *Thematic Apperception Test* (H. A. Murray, 1943) is a projective technique based on the theoretical approach of interpretation of verbal expression as reflective of psychological, emotional, and cognitive function. This assessment tool involves generation of narratives based on a series of black and white drawings representing various scenarios, through which the patient consciously or unconsciously describes underlying inhibited tendencies and typical approaches to interacting with their environment. The average length of the narratives is about 300 words for adults, and responses are interpreted according to narrative themes (e.g., emotional intensity, novelty of construction) in story-telling for psychodiagnostic purposes, and also for specific language form and content in some research contexts (cf. Covington, Riedel et al., 2005).

The *Whitaker Index of Schizophrenic Thinking* (Whitaker, 1973) was designed for two purposes: to differentially diagnose patients with schizophrenia from other disorders, and to assign a degree or level of severity to the observable thinking impairment. The technique involves multiple choice responses in a word association task, for which the foils represent differing levels of disorder (i.e., correct responses, loose association, ideas of reference, phonological association, nonsense). The response sets are all linguistically-mediated, and repeated patterns of selection filter the diagnostic classifications along a continuum from less to more impaired. One of the criticisms of this technique is the rather high percentage of both false positives

and false negatives, to which the author responded by pointing out the inadequacy of the current psychiatric diagnostic criterion.

A number of assessment instruments have been developed specifically for investigation of linguistic processes in mental illness, one the most well-known of which is the *Thought, Language, and Communication* (TLC) scale (Andreasen, 1986). Andreasen and her colleagues (Andreasen, 1979; Andreasen & Grove, 1986) identified eighteen patterns of disordered language use (re-ordered to cluster similar symptoms after Covington, He, et al., 2005):

- Poverty of speech, poverty of content – vague, non-specific references
- Pressure of speech, blocking – extremes in verbal fluency ranging from accelerated rate to a complete stop
- Distractibility – shifting attention to external stimuli
- Loss of goal, derailment, circumstantiality – digression and gradual loss of topic
- Tangentiality, perseveration – irrelevant or repetitive replies
- Illogicality, incoherence – disruption in coherent structure
- Neologisms, word approximations – made-up words or approximations
- Stilted speech – overly formal style
- Clanging – word selection based on phonemic relatedness, rather than semantic meaning; frequently manifest as rhyming
- Echolalia – 'echoing' words or lengthier discourse, with varying levels of purposeful intent
- Self-reference – egocentric topical persistence

This list of linguistic markers has been found to correlate with the positive and negative symptoms of schizophrenia, and has also been observed in major mood disorders.

The *Thought and Language Index* (Liddle et al., 2002) offered a less complex model of thought and language disorder based on eight symptoms identified through factor analysis, claiming reliability with identification of even subtle linguistic differences. Chen and colleagues (1996) developed the *Clinical Language* (CLANG) scale which approached linguistic analysis from a similar perspective, targeting pathology within a framework of current theories of language organization, including semantics, syntax and production, including formal features of both speech and language use, (e.g., phonemic processing, prosodic variations).

Sanders, Adams, Tager-Flusberg, Shenton and Coleman (1995) addressed assessment of the linguistic deviance in patients with schizophrenia and bipolar disorder-related psychotic features with a multi-factor analysis, utilizing clinical and psycholinguistic measures of language and thought disorder. A standardized clinical language sample was gathered, and post-hoc analysis included the *Thought Disorder Index* (Johnston & Holzman, 1979) and linguistic analyses of mean length of utterance, Type-Token Ratio, fluency (i.e., false starts and incomplete utterances), cohesion, and syntactic complexity, implemented for the purpose of measuring quantitative as well as qualitative differences in language use. Patients with schizophrenia demonstrated more significant breakdown in utterance length and cohesion, with higher thought disorder scores. The most sensitive indicators for differential diagnosis of schizophrenia were related to the discourse patterns, rather

than the grammatical structure of language, and these researchers suggest that quantitative descriptors of language function may be more accurate diagnostic tools than qualitative observational techniques.

Language in Mania

When compared to schizophrenia, the patterns of linguistic breakdown among the portion of the population manifesting mania as a major mood disorder have been less closely examined. The DSM-IV (1994) describes the pathological linguistic aspects of mania as inclusive of speech which lacks cohesion and is difficult to interpret, with word selection sometimes governed by phonological properties (i.e., alliteration) rather than semantic content; and 'flight of ideas', in which there is an accelerated rate of production with abrupt shift in topic, sometimes to the point of incoherence. Concomitantly, there also may be behavioral changes, including an increase in organization and planning behaviors involving spending money, developing grand and expansive ideas, and increased engagement in social interactions, all of which may be noted along a spectrum of severity. Research in this area has historically described manifestations of mania foremost as linguistically-based behaviors notable for mirthful mood (Fenichel, 1945); colorful language (Andreasen & Pfohl, 1976); difficulty creating meaningful relationships between sentences (Durbin & Martin, 1977); talkativeness and grandiosity (Mendhekar, Srivastav, Jiloha, & Awana, 2004); expansive thinking, and playful, flippant responses (Daniels et al., 1988); disinhibition and overinclusion (Khadivi, Wetzler, & Wilson, 1997); and distractibility (Liddle et al., 2002).

Lorenz and Cobb (1952) produced the first systematic analysis of the speech production of ten patients diagnosed with mania, using transcribed samples of verbal interview discourse in 1000 word samples. In their analysis of these patients and normal controls, they observed that the 'flow' of speech (i.e., the temporal rate of production and verbal sequencing) was not increased or 'pressured' in the presence of mania, but was more uneven and variable than seen in normal speech. Although by contemporary standards the samples would be considered inadequate, the researchers counted words and parts of speech, and noted that patients with mania used the same ten most common words as did controls, with equal distributions of elements of speech. The patients with mania used more pronouns and main verbs, and fewer adjectives. However, there were fewer word types noted in the samples collected from patients with mania, and these words were repeated with higher frequency of occurrence consistent with the predictions of Zipf's Law. Although Lorenz and Cobb argued that the basic linguistic structure was still present in mania, as their samples did not appear to be grossly disorganized, there were observations of defective process at higher integrative levels of language formulation. Later researchers (e.g., Durbin & Martin, 1977) discounted some of their findings based on the primitive nature of their linguistic analysis, suggesting that a satisfactory delineation and description of the speech production of persons diagnosed with mania has not yet been provided.

In her detailed description of the language of a patient experiencing an acute exacerbation of psychosis, Ribeiro (1994) also related to the concepts of 'flow' and 'flight' in verbal expression, arguing that 'pressure of speech' is one of the

conclusively pathological linguistic markers of mania, observed as an increased amount of spontaneous speech with an enthusiasm to pursue new ideas. Her single subject case study addressed conversational frames in which the participant communicated, suggesting that there were thematic units which established coherence, even though the patient's reality testing was clearly impaired. In the context of the institutional discourse between patient and doctor, the subject of study was able to perform her social identity in a frame consistent with the asymmetry of the interaction, despite her condition. Ribeiro also argues for interpretation of language in the presence of such acute mental illness as consisting of *metamessages* and referential meaning, through which the patient was making an effort to communicate meaningfully.

As a diagnosis of mania represents one of the polar extremes of a bipolar disorder, Andreasen and Pfohl (1976) compared sixteen patients diagnosed with mania to patients at the other end of the spectrum who were diagnosed with depression (n = 15), and reported that patients with mania had a slightly higher Type-Token Ratio in spoken language, but noted no difference between the two groups in syntactic complexity. Patients with mania tended to use less abstract language, with more nouns, adjectives, and action verbs, suggesting language use that was described as more 'colorful' than the patients with depression. Their results were consistent with those gleaned by Lorenz & Cobb (1952), providing some evidence for an emerging diagnostic pattern relative to lexical density of certain parts of speech.

Durbin and Martin (1977) also investigated speech patterns in patients with mania, addressing an analytically structured protocol of linguistic competence, and a

contextually-based analysis of extemporaneous discourse. Six patients in the acute phase of a manic cycle were included in this study, and the general patterns demonstrated suggested intact linguistic competence, including observation of phonological processes (although one patient demonstrated rule-governed neologisms), and functional lexical retrieval with typical vocabulary content. Coherence in the conversational discourse of this population was found to be faulty, particularly in the use of unreferenced ellipses, despite the presence of flaws that limited the sensitivity of the methodology of this study, including the small *n*. The researchers described analytic confounds, including problems with assessing the rate of speech based on constantly changing conversational topics, and syntactic performance which was judged to be generally intact and as complex as expected, although flawed owing to tangential responses. Regarding coherence of text, Durbin and Martin noted that anaphoric devices were used appropriately by all participants in all structured tasks, and at the level of intrasentential cohesion in connected discourse. However, detailed analysis indicated that while elliptical referents were used appropriately at the level of sentence interpretation, at the more functionally-based discourse level, anaphoric errors were pervasive and consistent with no self-monitoring or self-correction noted, and constraints on such discourse conventions were also violated with obscure and unexplained references. For example:

> In respect to *them* being problems, *they* have already been solved, however, the solution of *them* is a problem, the passing *them* on is a problem and that is why I am here and actually, some of these solutions were acquired in my coming here and that's part of the reason I was here, and, the perfecting of

> *them* came through here and the only problem now is the one we are solving now. (Emphases in original, p. 213)

Fleck, Shear, Zimmerman, Getz, Corey, Jak, et al. (2003) expanded the study of mania to include specific cognitive-linguistic processes involving patterns of verbal memory. These researchers postulated that memory failure is a cognitive process issue related to decreased encoding and consolidation of information, and potentially inefficient retrieval. They suggest that the ability to encode, access and retrieve cognitive information is a linguistically-mediated process, related to subcortical functions involving the striatum, amygdala and hippocampus, and argue that these skills are quantifiably and diagnostically different for people with mania. Fleck and his colleagues conclude that the symptoms of verbal memory deficits experienced by people with mania (typically mild to moderate as measured on first hospitalization) are related to verbal encoding and retrieval.

In association with observed deficits in cohesion and distortion of verbal memory, an area of particular breakdown that *appears* to separate persons with mania from the rest of the population is the tendency to engage in presuppositional behaviors, in which speakers provide insufficient background reference for understanding of a topic, apparently working under the assumption that the hearer is privy to the same referents as the speaker. Presupposition is identified through use of appropriate devices for maintaining cohesion and coherence (cf. Hoffman, Stopek, & Andreasen, 1986; Sanders, Adams, Tager-Flusberg, Shenton & Coleman, 1995). This is what Durbin and Martin (1977) described as breakdown in the complex linguistic

operations involving ellipses and discourse anaphora, whereby redundancy is eliminated, along with the potential for recovery of meaning from the context.

Researchers have also compared the speech patterns evident in acute episodes of mania with those seen in patients experiencing floridly psychotic episodes of schizophrenia. Wykes and Leff (1982) compared corpora of verbal expressions of thought disordered speech in a small sample of both populations (n = 4 patients with mania; n = 8 with schizophrenia), and reported that patients with mania demonstrated more frequent use of cohesive ties (i.e., lexical, conjunction, reference) per sentence unit than did patients with schizophrenia, although this was not necessarily related to how well a sentence could be understood by a listener. They suggest that frequency counts of this type of language use, however, will not enhance diagnostic accuracy between mania and schizophrenia. Similarly, other researchers describe no substantial difference between the linguistic output observed in these schizophrenia and mania (Barch & Berenbaum, 1997; Docherty, DeRosa & Andreasen, 1996 ; Harrow, Grossman, Silverstein, Meltzer, & Kettering, 1986).

Conflicting arguments are presented by other lines of psycholinguistic research. These have addressed patterns of speech and language noted in mania and those typical of schizophrenia, reporting statistically significant differences in the areas described as cohesion (Bartolucci & Fine, 1987; Hoffman, et al., 1986; Sanders, et al., 1995); quality of thinking disturbance (Daniels, Shenton, Holzman, Benowitz, Coleman, Levin et al., 1988); and discourse connectedness (McPherson & Harvey, 1996), although such arguments must be considered in the context of differing levels of acuity for the respective disorders at the time of analysis.

Assessment of language in mania

Clinically, assessment of mania has been approached from the perspective of differential diagnosis from other mental illnesses, but some specific tools have been devised to address identification and track improvement in this population. Altman and his colleagues (Altman, 1998; Altman, Hedeker, Peterson, & Davis, 1997, 2001) developed a linguistically-based self-rating scale specifically for assessing change in patients with mania, which has proven to be a reliable and valid measure, even with patients demonstrating limited insight into their illness. Patients interpret sets of statements, identifying their own experience by endorsing responses rated from 'not present' to 'present in severe degree'. A primary criticism has been that the efficacy of such tools appears limited, as self-appraisal is more useful with mild to moderately involved cases, since more severely impaired patients tended to be either unwilling or unable to complete the self-assessment. Certain items included in earlier iterations of this measure (e.g., distractibility, racing thoughts, grandiosity) were endorsed frequently by patients with mania and those with other diagnoses, and were determined to be less useful in differential diagnosis. Nonetheless, Altman's scale and others have evolved into useful clinical applications for initial screening, patient education and self-monitoring, and documenting treatment outcomes.

Khadivi, Wetzler and Wilson (1997) applied the techniques from the *Thought Disorder Index* (Johnston & Holzman, 1979) to responses on the *Rorschach* (Rorschach, 1921) as generated by patients with mania, in a task designed for differential diagnosis. They noted a distinct pattern of combinatory thinking in the patients with mania, in which unrelated topics were linked in an overinclusive manner,

related to the tendency toward distractibility (i.e., attending to more than one topic at a time). These results suggested specific concerns related to cohesion and coherence, as the connections tended to be faulty and inappropriate, as well as expansive and grandiose, but not necessarily unlike the connections sometimes expressed in creative thinking.

With the understanding that some variations in patterns of language use are typical, but others are recognized as a sensitive indicator of psychopathological function, Hoffman, Stopek and Andreasen (1986) demonstrated that people with mania exhibit unexpected shifts in discourse from one coherent topic to another, a subtle distinction, but nonetheless a measurable one representative of aberrant language use. Using linguistic markers of cohesion, Wykes (1981) demonstrated that psychiatrists could be trained to discriminate differences in the transcripts of speech between patients with mania and patients with schizophrenia. As speech patterns appear quite similar in the acute stages of each illness, the physicians were able to demonstrate a statistically significant improvement in their diagnostic accuracy based on brief training looking for patterns of cohesive ties. Similarly, Ceccherini-Nelli and Crow (2003) compared measures of language function to traditional behavioral observational criteria for differential diagnosis of psychotic disorders, and convincingly argued that language disturbances were diagnostically superior in discriminating schizophrenia from other psychiatric diagnoses.

The historical discussion of language use among mentally ill populations has not adequately addressed the question of whether it is the language that is disordered, or if the language produced is an outward manifestation of a mental disorder of either

psychological or organic etiology. In the presence of mental illness, observations of patterns of language production are used as diagnostic criteria, upon which various psychological, environmental, behavioral, and/or pharmacological interventions are based, and language use often serves as one of the outcome measures of successful treatment (Andreasen, 1986; Gavell & France, 1991; Liddle, et al, 2002; Rieber & Vetter, 1994; Rosenberg & Tucker, 1979; Sanders, et al, 1995; Thomas, 1997). However, the individuals comprising the groups designated as mentally ill actually represent a heterogeneous population for whom there is not always a consistent profile of linguistic function. Among other things, intellectual endowment and realization of cognitive capacity interact with symptoms, interventions, and etiological factors, and the means by which atypical manifestations evidence (complicated, as noted, by the inherent difficulty in classifying disorders based on such an admixture). Thus, a review of the literature addressing patterns of language use in mental illness presents a continuum of results. There are acknowledged similarities between schizophrenia and mania, particularly in the most acute phases of both illnesses, at which time differential diagnosis is almost impossible. Disagreement remains regarding the linguistic profile that could be considered typical in these acutely ill populations, although the most obvious breakdowns appear to be in semantic content and pragmatic function, with some variations in syntax also suggested by Covington, He, Brown, Naci, McClain, Fjordbak, et al., (2005).

Despite the heterogeneity of the populations and confounds introduced by differing acuity, disease processes of varying etiology and diagnosis, and the spectrum of acceptable variations in language use, researchers do seem to agree that

there are typical linguistic breakdowns observed in schizophrenia and bipolar disorder, most obviously in cohesion and coherence, probably based on factors that limit the individual's ability to focus and attend to a stimulus without distraction, yielding pragmatic deficits of functional use of language in context. In order to better situate these issues in the current study, the discussion turns to analysis of patterns of writing specific to the genre of keeping a journal.

Journal as Genre

> "To suggest that people should keep journals is to suggest
> that even their quickest thoughts and feelings caught
> in mid-flight might be valuable to them and to others."
>
> Macrorie, 1987

As noted, the patterns of language analyzed in the present study were produced in the form of journals, texts written by LMN who was adopting a role as scientist/observer of her experience, as a notation book in which she records everyday experiences (e.g., grocery lists), as well as more abstract notions, and her ideas and insights relative to her life. In order to propose a method for adequately deconstructing this data set, the review of the literature addresses the concept of journal as genre.

From an historical perspective, Lowenstein (1987) describes the emergence of the journal as early as 56 AD in China, involving public documentation of communal materials (e.g., household accounts), but notes that the more private and autobiographical use of written discourse did not emerge until the Renaissance.

Variations on the genre have been acknowledged as significant for recording literary and historical events as they occurred, and even more personal writings (e.g., the *journal intime*), have offered important insights into political, economic, and social issues across time. Gannett (1992) has parsed the historical semantic differences between keeping a journal versus keeping a diary, suggesting that the meanings have evolved in such a way that the former requires a measure of profound reflection and deep thought, whereas the latter, typically given feminine attributions, involves the recording of intimate and personal contemplations. These tend to be inconsistent semantic variations, however, and both terms are used synonymously with others such as 'appointment calendar' or 'daily log' in a range of projective or reflective applications as records of personal experiences and information.

Fulwiler (1987) notes that over time, a set of common features of 'good' journals as utilized in the academic setting has emerged, with particular relevance to the research at hand. The content and process of these commonalities are broken down into three main areas. First, in the area of language features, he suggests that they should look like conversational speech in written form, using colloquial diction, informal punctuation, first person references, with the freedom to be experimental and unpredictable. The cognitive activities involved should include observations, questions and speculation about the writer's experiences, increasing self-awareness, digression from the topic at hand, synthesis and revision. The formal features of good journals include frequent and long entries with chronological ordering. Gannett (1992) offers a taxonomy of journal writing, reasonably divided into four types: the common-place book, the travel journal, the spiritual journal, and the professional

journal (e.g., ship captain's log). Each retains a particular purpose with little overlap, although in practical application, the different strands of record-keeping and notation are frequently woven together.

Lowenstein (1987) argues that although journals are written by one person, they are not produced as isolated phenomena, noting that some works are specifically prepared for audiences, but even those which are not should be interpreted by the reader without distortion in the context in which they were created; hence, the writer's perceptions and phenomena can be situated and understood appropriately. Sperling (1993) suggests that "one's past and future conversation partners lurk" in every text, arguing for the influence of the audience, even if it is not explicitly identified (p. 5). Rubin (1988) describes writing as an act of social construction involving communicative contexts which impact stylistic variation in writing. These include the purpose of the communication, participant roles, setting, topic, and discourse structure, and he notes that writers construct variation in these domains to serve different linguistic functions (e.g., persuasion, exposition). Consistent with this line of reasoning, Gannett describes the act of keeping a journal as a form of social construction, acknowledging that "no act of writing… is completely personal or private" (1992, p.2), and that while contemporary journal-keepers may employ a variety of sub-genre (e.g., blogs), the audience for such writing still exists along a continuum ranging from the private-self-as-audience to a completely unknown public audience via the World Wide Web (C. Gannett, personal communication, April 2, 2006). This fact is also acknowledged by Stillman (1987) who, while recognizing the inherently personal nature of keeping a journal, comments on the value of one's

written word as a gift for posterity. Writing also situates the individual within the community, as Harris (1989) notes, arguing that the practices of one's community constrain and reflect what may be said. According to Sperling, writing then, is constructed as a social action with intertextual links to meaning in context, and is fostered and rooted in interaction with others (p. 43).

Rubin also observed that identity of the writer is constructed by the reader (1995) in work related to gender identity. To expand on this theory, a worthy point of consideration addresses whether the reader of the LMN's journals constructs an identity for the writer based on *a priori* assumptions of mental illness in general, and mania specifically. These notions may be based on any of a number of premises: on the DSM-IV (1994) classification criteria; on personal anecdotal experiences; or on the manifestations of these diagnoses as constructed by popular culture and the media. Perhaps the more pressing question would be to ask if the reader did not have these preconceptions before reading the journals produced by LMN, would she necessarily have been identified as manic based on these samples of writing, particularly when viewed through Fulwiler's (1987) criteria?

Putting the pieces together: language and mania and journals

The participant in this study has defined her work as journals which she has written during periods of mania, begun as a means of recording her experience for an unspecified audience for the purpose of closer investigation. As the primary directive of science involves describing, explaining and then predicting a phenomenon, she began her extensive record-keeping as the first step in that process. Based on her

training in the applied scientific method, she recognized a need to record her perceptual changes and thought processes with the onset of her first episode in 1978. The end result of this effort is a product that incorporates multiple media, including audio and video recordings, photographs, computer programs, and most significantly, journals. Within the larger category of journals, however, she has employed sub-genre, as her writing may alternately be common-place, when she writes about her daily activities (e.g., taking her son to a movie); spiritual, as she describes the transcendental pleasures of a sunny day or the insights she gains from exploring prime numbers; and log-keeping, as she records the items she has purchased at the grocery store or the number of grapes she has eaten for a snack; the complex combination of which has shaped her discourse.

Writing occurs as a series of dynamic oppositions, occurring both as an individual cognitive exercise, as well as a social and conventional practice; as an expression of idiosyncratic identity situated within a social matrix; and a process through which a writer's identity is simultaneously reflected and created (Rubin, 1995). LMN's writing clearly fits each of these points, as she has chosen to engage in the socially acceptable and conventional practice of keeping a journal, but is using the genre as a means of exploration of her emerging identity as a person with a diagnosis of mania.

Pragmatics

As described, the fundamental differences in language in cases of mental illness are at the level of functional use in context, or pragmatics, and this study will

attempt to discern such variation in written discourse. Pragmatics is a broad ranging construct that deals fundamentally with the functional purpose of language and the nature of inferences and assumptions inherent therein (Levinson, 1983), involving the maintenance of politeness boundaries (Leech, 1983) and observation of cooperative rules and implicature in language use (H. P. Grice, 1975). Pragmatic language disorders are recognized as pathologies of communication in developmental and acquired etiologies, most consistently evident relative to involvement of the frontal lobes of the brain, particularly in the right hemisphere. As noted, several research groups (cf., Corcoran & Frith, 1996; Langdon, Davies, & Coltheart, 2002) have investigated Theory of Mind deficits in populations with mental illness, arguing for a clear link between the limited ability to appreciate the perspective of others and impairment in the ability to observe the 'rules' for appropriate interaction in a conversational context.

Although the defining variables of pragmatic language are a moving target based on local cultural norms and acceptable discourse practices, observational techniques (cf. Prutting & Kirchner, 1983, 1987) address communication at larger units of function, rather than smaller units of decontextualized form. Pragmatic analysis is then an effective means of delineating communicative breakdown at the level of function and purpose.

Context is described as the foundation upon which effective pragmatic communication is situated, either written or spoken, but it is also ambiguously defined. Levinson (1983) speaks to the cultural and linguistically relevant aspects involved in the production and interpretation of utterances, factoring in role and status,

spatial and temporal location, level of formality, knowledge of the medium and subject matter, and appreciation of the register. Additionally, social and psychological factors shape the context of effective language use, and these can vary at any time (Ferguson, 2000; Ochs & Schieffelin, 1979).

Smith and Leinonen (1992) describe pragmatics as the process and function of language, and closely related to this, they define discourse as the product or structure of communication, inclusive of all modes of expression. Discourse is the interaction between naturally occurring language, its use in social and cultural contexts, and the acquisition, storage and use of knowledge. The analysis of discourse involves principles that constrain the structure and order of language, and the application of social and cognitive knowledge to the generation of coherent and sequential organization, and the breakdown at any of these levels has the potential to impact function at all other levels (Mentis & Thompson, 1991). Appreciating the multiple influences on discourse, Wood and Kroger (2000) argue that the goal of discourse analysis is, "to understand variability and to employ it for analytical purposes, not to eliminate it" (p. 10), so rather than seeking to identify universals in language use, the more useful approach would be to embrace the differences, even intra-individually as observed in LMN's journals, based on the situational context.

In that context of discourse as text, discourse prosody (Stubbs, 2001) is a construct that addresses the attitudinal tone of language in an evaluative and subjective manner, inclusive of semantic meaning as well as pragmatic intent. This carries across sentences, speaking to the affective connotations of entire pieces of text. Discourse prosody contributes to a level of analysis inclusive of positive or negative

attributions, external locus of control, making meaning of circumstance, and individual agency (Brown, Nolan, Crawford, & Lewis, 1996; Casey & Long, 2003; Harden, 2000; McAdams et al., 2004), all of which serve to situate the writer's intent in context based on patterns of language use.

Returning to the model of linguistic oppositions, there would appear to be two seemingly disparate, although not necessarily mutually exclusive ways to formulate an understanding of LMN's patterns of written language. The idiographic manifestations are those involved in establishing her baseline model of interpreting and responding to reality, against which further behavioral functions will be measured (J. W. Grice, 2004; Molenaar & Valsiner, 2005). In LMN's case, these behaviors conform to the DSM-IV diagnostic criteria relative to flight of ideas and grandiose idea generation. In this framework, the specific diagnostic conditions have been identified, and her experiences fit the pattern. Acknowledging this, she has taken mood-stabilizing medication for many years in an effort to gain some control over the variability inherent in these cycles. In that process, she was also attempting to conform to exogenous influences brought to bear (e.g., family pressures). LMN's patterns of word choice, discourse prosody, collocations and idea germination may represent idiographic traits by which she is defined, based on her life experiences, her educational achievement, and her professional endeavors. Her voluntary associations, as well as her involuntary condition have shaped her lexicon, and have likewise influenced the linguistic behaviors of the networks and communities of practice with which she has been affiliated as member and leader. The process of differentiation is realized by multiple layers of language use, as a professional, as a parent, as one who

records personal thoughts in a journal, and it is the sum of her collective parts that defines her. From her perspective, the diagnosis of mania has been a positive force which she has learned to embrace, despite the vagaries and unpredictability of the exacerbations.

On the other hand, a diagnosis of mental illness necessarily places one in a position of comparison to other people, addressing the nomothetic variations of a particular individual situated within the population at large. Chambers (2003) suggests that personality factors as manifest in language carry no social significance, but this perspective misses the point that mental illness carries its own social stigma, especially because many such illnesses are identified based on the observed use of language as different from the expected. Although one would not choose to be diagnosed with a major mood disorder, this participant has chosen to use language as a tool to record her experiences for posterity. Viewing mania and the writing produced during such episodes through a nomothetic lens provides a perspective of the boundaries of what is considered "normal" by the general public. In the heterogeneous society at large, and even in the more circumscribed communities of practice in which LMN operates, she has been able to function as a spouse, a parent, an academic researcher, a teacher, friend, and a productive community member. All those in close contact with her have been aware of her episodes, and have dealt with them as one of the qualities of her character. Grice (2004) argues that the "true study of personality is … necessarily idiographic *and* nomothetic" (p. 205, italics in original); hence it is from both of these perspectives that these data will be addressed.

The journals produced reflect 'good' journal-keeping habits as noted by Fulwiler (1987), as LMN has observed the conventions of chronology by indicating dates and times, writing in a conversational manner, and constructing meaning based on her observations and perceptions. Her stated intention has always been to prepare a data set which will offer insight into the manifestations of mania, and her use of language in these texts provides such a glimpse. The ultimate goal of the present study is to determine if the patterns of language used in the formulation of this collection of journals can resolve some of the paradoxical questions – what in her language is perceived as normal or pathological, pragmatically functional or impaired, individual or socially situated, idiographic or nomothetic?

Chapter 3 - Methodology

Description of general methodology

This chapter includes a description of the methodology involved in this dissertation, beginning with a description of the participant (LMN) and her journals, then an explication of the data set, the reference corpora, and the analyses involved in comparative investigation of the data set.

Corpus linguistic research methodology is based on the theoretical orientation of language use in context, as patterns can be identified and defined only through analysis of large bodies of text. As the data produced by the participant in this study is unique in its construction and scope, reference corpora are organized and implemented for the purpose of comparison to a standard considered to be typical. The analyses involve synchronic and diachronic comparisons across corpora and across time, and although LMN's corpus may not be representative of language use in all individuals diagnosed with mania, based on the longitudinal nature and the volume of data, it is representative of her experience. Prior to initiating the investigation, the researchers and participant drafted and signed an informed consent agreement (Appendix A) as required in the application process of the Human Subjects Office/Institutional Review Board (IRB), Office of the Vice President for Research, University of Georgia; approval for the research was granted based on meeting this and all other IRB guidelines.

The participant

LMN was born in a small town in a rural area of the Midwest in 1942, the second of five children and the only daughter. Shortly after her birth, her parents bought an 80 acre farm from the maternal grandparents, located about four miles from the local school, and about 15 miles from the nearest small town. On the farm, her father had corn, wheat, and soybean crops and raised hogs and dairy cows, while her mother raised broilers (young chicks) and for one summer, turkeys. In addition to farming, her father worked in road construction and was involved in the building of interstate highways. During LMN's childhood years, he would be away from home for several days at a time, and she fondly recalled the pleasure of visiting him on the job site with her mother and brothers during the months of summer vacation, playing around the large mounds of dirt and the earth-moving machinery. Her mother was a homemaker for most of LMN's life, although in later years she owned and operated a restaurant in the small town nearby. Because they lived on a farm, there was always plenty to eat, and the family had a big garden, the tending of which was one of the children's responsibilities. Parenthetically, LMN notes that in her adult life, she has had no inclination to work in a garden, probably as a result of that early experience. The extended family lived nearby, and LMN grew up in a close-knit environment involving frequent interactions with aunts, uncles and cousins. In retrospect, LMN notes that perhaps the area in which she grew up would be considered "poor", but she had no awareness of that at the time, and she considered herself and her family to be fairly normal, noting that "I never felt I didn't have what I needed".

LMN attended a small school with a total enrollment of about 90 students; there were 18 in her graduating class. She describes her early educational experiences as typical, involving active participation in the 4-H Club and being a cheerleader for four years during high school. The school's resources were limited, with a small library that took up less than one entire wall. There were no books in her home, with the exception of a copy of *Huckleberry Finn*, which was given to the children as a gift from a well-meaning friend of the family who felt that children needed books. The family got their first television set in 1953, and around that same time also got their first telephone.

Based on her outstanding academic performance in high school, LMN did not attend her senior year, as she was accepted into an early enrollment program at one of the largest universities in the state. She left the farm at the age of 17 and moved to a the town where the university was located to live in a co-op house on campus, where every resident had specific responsibilities in return for a decrease in their living expenses. Her initial ambition was to be a typing teacher, based on the influence of an aunt who held that position, but instead this role-model encouraged her to pursue a major in Home Economics based on LMN's early involvement with the 4-H Club. She was married in January of 1962 during her junior year of college and had a baby a few months before her graduation in June of 1963. She took a teaching position that fall at a high school that included grades 7-12, a position which she held for four years. She was divorced in 1966.

Her next professional position was at a larger high school where she continued teaching Home Economics for two more years. At that time, her state had a

regulation requiring all high school teachers to hold graduate degrees, so she returned to her alma mater and graduated with a master's degree in 1968. Prior to leaving the university, she interviewed for and accepted a teaching position at a state college on the west coast.

In the summer of 1968, LMN spent three months touring Europe, and upon her return, she was married for the second time, to a man from her hometown. As she began her college teaching position, it became evident that continuing in a post-secondary work setting required another graduate degree, so she decided to pursue her PhD. She was accepted into an extremely rigorous and prestigious institution in that state, and began her studies in January of 1970 with a small cohort of students, some of whom remain close friends.

As she neared completion of her studies, she accepted a position at a branch of the state university system prior to finishing her dissertation. She finished writing in January of 1974 and graduated with her PhD in June of the same year. Her second child was born in November of 1974, and she remained at that university developing her lines of research and fulfilling her teaching responsibilities through June of 1980.

In February of 1978, she experienced the onset of her first manic episode, what she described at the time as a 'creative period', during which she primarily stayed in her office at the university, going for several days without sleep. Based on her unusual behavior, and concern for the care of her young son, her colleagues summoned her husband from his job in a city five hours away. He took her to an emergency room, and after a brief examination in which the resident physician reportedly would only speak to her husband, while not addressing any comments to

her, LMN was medicated and sent home to rest. No diagnosis was made at that time, and no longer-term pharmacological intervention was prescribed. Shortly thereafter, she briefly engaged the services of a psychologist and then a psychiatrist, but there was still no clinical diagnosis assigned. It was during this period of time that she began documenting her phenomenology in a journal, recognizing that from a research perspective her experiences were unusual, which motivated her to leave a record for future investigation. This particular episode lasted for a few days, then remitted after medical attention was sought, at which time she ceased the recording of her experiences and returned to a more normal pace of existence until the next episode occurred.

During a subsequent episode in 1979, LMN had a friend record her using 8-mm video for which there is no accompanying audio track, but on which her rapid changes of position and frenetic motion are easily observed. On this piece of film, she also demonstrates the perception of a 'popping' feeling in her head, which she describes as 'keno pops', similar to a motion she had seen in gambling machine in Las Vegas.

In 1979, she was up for tenure review at the university and was denied. Based on comments in the letters from her department head, there were concerns expressed about her unusual behavior, which although unidentified, had begun to repeat in an unpredictable cycle of episodes. She recalls being confrontational with a visiting professor in her department during a faculty seminar, demanding that he articulate the purpose and value of his research, and in retrospect, she feels that such incidents contributed to the denial of tenure. Ironically, in her performance reviews prior to the

onset of the episodes, her research was acknowledged as positive, but subsequent reviews of her research were unfavorable, perhaps as an artifact of her episodic behavioral changes.

In July of 1980, she started in a new position at a large state university on the east coast, noting that she immediately felt at home, as the rolling hills of the small college town were reminiscent of the area in which she grew up. Across the next few years, she continued to experience periodic episodes, all of which seemed to be manageable from a functional standpoint, as she felt increasingly creative and productive, all the while continuing to document her experiences using a variety of media, including photographs, audiotapes, and writing. When she came up for tenure at her second academic position, some of the same concerns regarding her manic behavior were allegedly discussed within the committee, but her application was successful and tenure was awarded.

In early 1985, her experience of the creative periods took a markedly different track, for the first time requiring inpatient hospitalization. At the time of this acute exacerbation, she was working at the university, engaging in ongoing research and teaching, and was also under a particularly stressful extra-curricular workload, as this was the period during which her idea of a national meeting to study mathematical models in her area of interest was brought to fruition. The pressure of organizing and executing an event of this magnitude compounded her accelerated level of energy and motivation to the point of requiring intervention, and she was referred to a large teaching hospital in a nearby city and was admitted to a psychiatric unit. The diagnosis of mania was made shortly after admission, and lithium therapy was begun.

She continued to record her experiences while hospitalized, and although she was initially somewhat resistant, she agreed that the medication was the appropriate course of intervention for her at that time. She was discharged home a few days later, but required hospitalization again three months later as ongoing lab tests indicated a subclinical level of medication.

During the summer of the same year, LMN was hospitalized for a third time, on this occasion due to a dangerously high level of lithium, to the point of toxicity. Her retrospective description of that experience was that she was trying to manage her mania by taking an increased dose of the medication, hoping this would help her to gain clarity on the thoughts and ideas she was having, as well as improve her ability to solve the problems she was processing. She reports that taking the medication in this manner was quite purposeful behavior, and she documented each dose in the journal she was keeping at the time. However, upon admission, her physician documented the possibility of a suicide attempt, although LMN states emphatically that there was no such intent.

Since 1985, there have been no subsequent hospitalizations, and although she has continued to experience manic episodes several times a year, the amplitude of these events has been less critical. From her perspective as a researcher, LMN understood that mania occurred in a cyclical manner with no predictable pattern, and across time, she has made an effort to explore statistical correlations between onset of the episodes with phases of the moon, seasonal variations, menstrual cycles, and biorhythms, all of which have yielded inconclusive results thus far.

The impact of these cycles of acceleration took several forms during the early years of her experience. On one hand, LMN felt that the increased creativity contributed to her research process and her ability to problem solve and think through her scientific investigations, as it was during one of these periods that she had the original idea for a national scientific meeting to address mathematical models in her area of research. On the other hand, as noted, the unpredictability of the episodes as well as the change in her ability to manage the responsibilities of her faculty position contributed to the denial of tenure at that first academic position. She notes that when experiencing a manic "high" there is a tendency to lose the control mechanism or cognitive filter that "…makes you do what you're supposed to do" in terms of regulating verbal behavior or emotional expression (LMN, personal communication, April 11, 2006). Although she describes herself as typically not confrontational, she recalled unusual incidents relative to this loss of control, including the experience at her first academic post involving challenging the visiting professor, as well as shouting at a physician who was perceived as being condescending to her in the hospital emergency room.

The impact of the manic episodes on her family was most evident from the perspective of her husband. They had been married for more than nine years when LMN experienced her first episode, and across time, he became concerned that LMN was becoming more self-centered and private during the episodes, noting that she would pull away from family and friends to spend time alone writing in her journals. Further, he worried that she was ruining her health, as her eating and sleeping patterns would become erratic and inconsistent secondary to the mania. Her younger son, who

was three years old at the time of the initial manic episode, coped fairly well with his mother's episodes. After reaching adulthood, he finally told her at one point that, "…the only thing that bothers me is that you can't focus on one thing." LMN acknowledged that there were probably negative challenges in all of her relationships involving friends, family and colleagues during manic periods, but that she did not necessarily perceive them as negative at the time, because she felt good and was productive in her work. LMN was widowed in September of 2002 after nearly 34 years of marriage.

Although initially agreeing to take the lithium, and continuing to do so for many years primarily at the encouragement of her husband, after nearly 19 years of regular dosages of the mood stabilizing medication, LMN decided to discontinue the lithium therapy in December of 2003, more than a year after her husband's death. She did so under the supervision of a psychiatrist, and has asked selected friends and family to monitor her for extreme changes in behavior. LMN ultimately decided to stop the medication because she continued to experience manic episodes two or three times a year the entire time she was on it.

Since her retirement in June of 2004 (closely coinciding with the discontinuation of the medication), LMN has continued to experience manic episodes. She has continued to write in her journals during each one, while pursuing hobbies and interests including taking and teaching adult continuing education classes, volunteering, traveling, and starting a new business venture. Her most significant interests continue to involve the exploration of the phenomenon of mania as she has experienced it, and seeking to understand how it has impacted her life. She is

continuing to document her experience in her journals, and through the analysis of the contents of these books which now number well over 100 volumes, she hopes scientific investigation of her records can used to further the knowledge and understanding of bipolar disorder.

The journals

The journals in which LMN has recorded her experiences with mania have taken many forms, most consistently in bound books with unlined pages. However, after receiving a variety of blank writing books as gifts, and at other times using what was available, the collection has grown to include an assortment of volumes of different shapes and sizes, including spiral bound notebooks, and three-hole plastic binders with notebook paper pages torn from legal pads. Table 3.1 provides an overview of the types of journals from the three spans of time representative of early, middle and late phases of LMN's experience with mania, corresponding more or less to the years from 1978 to 1984; 1985 to 1994; and 1995 to 2003. Although there is a fairly large body of text that was produced after 2003, journals written after that time were subject to the observer's paradox and inherent variations in style; this will be discussed briefly in Appendix D. As previously described, although the manic periods consistently occurred across all time spans, at certain points, there were fewer journals used for the documentation, with more reliance on other media for the purpose of leaving a record.

LMN notes that one of the reasons she has employed a variety of paper media in keeping a journal is to ward off the devastation of losing a valued volume. In 1984,

she began writing in an expensive leather-bound journal, recording her thoughts as part of a lengthy ongoing series. "Book Thirteen", as this particular one in the sequence was labeled, was left on an airplane, and despite her best efforts, it was not recovered, and represents a loss she continues to lament more than 20 years later.

The contents of the journals are also variable, as some volumes include documentation on every single page, including notes in the margins, sketches and graphs, and coded diagrams of her perception of physiological changes, while others may have brief entries on only a few pages. The temporal scope of the journals is also variable, as some volumes cover a span of time as short as a single day, and others were begun during one episode and continued during another episode, perhaps months or even years later. This has contributed to LMN's challenge of organizing the data set chronologically, but she continues to focus on this goal so that she will be able to more clearly define the timeline of her manic process in an effort to determine any identifiable patterns of occurrence.

Since retirement from her academic post, she has expended a great deal of time and energy in organizing the journals for study. She began with a numbering system, applying numerical identifiers in the order she retrieved a book from storage (i.e., not in chronological order according to the time written). This system has allowed for systematic review and organization of the documents in preparation for the eventual archival process to be undertaken when her papers are donated to the local university library.

Table 3.1

Examples of journals archived by LMN.

Dates	Journal types	Number of volumes
1978 to 1984	Hardbound volumes, composition books, sketchbooks, mini 6 ring loose leaf binder, composition book, leather bound journal (lost)	17
1985 to 1994	spiral notebooks, sketchbooks, composition books, large (3 inch) three ring binders loose leaf binders fabric covered hardbound journals, steno pad	61
1995 to 2003	sketchbooks, laboratory notebook, hardbound journals with decorative fabric covers, small pocket-sized journals with quotes on cover	49

Procedures

Selection. The basis for all comparisons in this study is the Master Corpus, compiled from a selection of LMN's journals into a 242,589 word collection. The journals were organized into a timeline spanning the course of her experience with mania from 1978 to the present, and because of the large quantity of data, spans of time representative of early, middle, and late phases were roughly delineated into

seven to ten year spans, as specified in Table 3.1. The dates of the selected journals prepared as a timeline included as Appendix C. Selection decisions were made jointly with LMN, as a part of the ongoing member-check process, and specific texts were chosen from each time period based upon meeting the following criteria:

1. Texts in which words represented at least 75% of the total content (i.e., not solely automatic sketches or other non-orthographic notations).
2. A number of texts sufficient to represent at least sixty thousand words of the available data from each of the designated time periods (i.e., early, middle, and late as noted in Table 3.1)

During several years of the early period, defined as the time from the initial episode until the time of her hospitalization and diagnosis, LMN stopped keeping a formal journal and utilized other media exclusively for a period of time; rather than writing, she took photographs and made extensive audiotape recordings. Thus, review of the journals in chronological order indicates a gap between the years of 1979 and 1983, but she reports that the episodes were still occurring; her means of expression took a different tack. The data utilized for that portion of the analysis were selected from the years represented by the journal writing.

During the middle period, LMN was at her most prolific, notably in 1985 during which she was hospitalized three times. In July of that year, LMN entered what she describes as likely her most acute exacerbation, an experience which she attempted to regulate through use of medication, resulting in an accidental overdose and lithium toxicity. In the space of twelve days, she filled over twelve hundred

pages. This particular text is included in the middle period sample, but also represents a separate sub-corpus which will be used for further molecular analysis. Other texts from this time period typify her patterns of language use during episodes, particularly as she is involved in a number of work and community-based activities (e.g., participation on the local planning committee for a major civic event).

The late period texts were selected from the years of 1995 through 2003, and document her experiences through the loss of her husband, retirement from her academic career, and the early stages of post-retirement activity, including the decision to discontinue her use of mood-stabilizing medication. Although LMN continues to write, these more recent texts are influenced by the observer's paradox, and with one exception (see method for analyzing medicated versus unmedicated conditions below), these were deemed not suitable for inclusion in this study.

Transcription. The texts selected for inclusion in the Master Corpus were transcribed for the purpose of noting consistent and consecutive patterns in language use. Journals selected for inclusion were transcribed by typists who had been instructed to reproduce the text exactly as written, inclusive of abbreviated sentence structure, spelling errors, mathematical formulae, time notations and marginalia, and omitting non-text drawings and notations, while documenting illegible words as <xxx>. When orthographic convention gave way to sketches and drawings, descriptions were included within brackets, to be excluded during corpus analysis; transcription conventions are included as Appendix E. Transcriptions were converted into text file format (.txt) for data entry into the *WordSmith* program, which is described below. Because LMN's current interest in the research process had begun

in early 2004, she had arranged for transcription of a number of journals prior to the advent of the current project, which were graciously made available for inclusion in this corpus. Although specific typing conventions were outlined, each transcript was reviewed for accuracy and any errors were corrected according to the researcher's interpretation of journal content.

As is typical for journal writing habits, there is not necessarily a clearly delineated language-based beginning and end to a volume. Rather, a journal ended either when she cycled out of a manic episode, or when she ran out of pages in one book and would pick up her train of thought on page one of the next volume; hence, selected texts may not be representative of entire episodes. Her entries frequently reflect notations from all hours of the day or night (e.g., one of her consistent turns of phrase is "It's three o'clock in the morning"), and the journals served as the vehicle by which she was able to ground her process in a temporal and chronological manner.

Relative to the variation in journal stylistics, her individual entries were also variable in length and frequency, as well as purpose. LMN would typically begin and end an entry with an hour:minute:second time stamp, and the text within these boundaries took on a variety of formats. An entry may have been a brief description of a physiological perception (e.g., "I-spot shook to count of 48."); short statements about her current state of mind (e.g., "Drank coffee. Feel good. Happy mode."); or a complete sentence which explored a fulminating idea (e.g., "Software development: Supports the development of new means for solving computational problems unique to biology, i.e., example algorithms for searching databases or understanding nucleotide sequences."). Because of this variability, for the purposes of this study an

entry is defined as the discrete production of a record which begins (and frequently ends) with a time stamp, and is usually separated from the previous and subsequent entries by white space.

Reference corpora

The purpose of the analyses of LMN's journals is to situate the written language of this individual in comparison to what would be considered typical patterns of language use. In order to make such comparisons, points of reference must be fixed; the specific comparisons and the reference corpora utilized in such analyses are described below.

External reference corpora. In order to determine if LMN's corpus is actually representative of patterns of language use in mania, comparisons were required addressing differences from a benchmark of what could be considered to be typical language use. Scott (2006) raised the question of establishing tolerable limits of similarity between two corpora for the purpose of determining differences in language use, noting that different research questions require different reference corpora. He argues that the influence of size and genre of reference corpora are important in determining keyword patterns, and if these two conditions are appropriately satisfied, there are no bad reference corpora.

For the initial analysis, the Freiberg-Brown Corpus of American English (FROWN) was employed. FROWN was compiled in 1991 and consists of 1.47 million words of edited written American English, divided into 2000 word samples of varying genre, and it provides a synchronic comparison of patterns of typical

language use during the late 20th century. FROWN is a replication of the Brown Corpus of American English, which was produced in 1961 as a representation of the language used in texts such as American newspapers, fiction, and government documents (Meyer, 2002). Although this corpus is considered to be somewhat small when compared to others in contemporary use (some exceeding 100 million words), it is representative of the time period in which LMN was writing, and it incorporates *written* language, both being factors which satisfy Scott's (2006) criteria, and hence selection of the FROWN was believed justified as a point of comparison for LMN's work.

Further external comparisons were made using a reference corpus designed to answer the same questions regarding differences in language use from a literary stylistic perspective. The genre of journal-keeping was the basis for selection for samples of written language as generated by other writers. Although the concept of keeping a journal is ancient, the journals which were available for comparative analysis tended to be limited to the contemporary format of a web log, identified colloquially as 'blog', in which personal thoughts and opinions are recorded in an on-line format, available for wide consumption. The blog as genre is not consistent with that of journal-keeping, as the construction of audience in this milieu tends towards addressing the anonymous public end of the spectrum, rather than self-as-audience (cf. Gannett, 1992); therefore the linguistic devices would be less consistent with those produced by LMN. Hence, blogs were not included in this analysis. The texts that were selected reflect a somewhat diverse mix, including works from the late 19th and 20th centuries allowing for both a diachronic and synchronic analysis.

Acknowledging the linguistic confounds of such material based on the temporal differences and the nature of these texts as edited material available in an electronic format, these corpora nonetheless meet Scott's (2006) criteria with regards to consistent genre, and are justified as points of comparison, as they are consistent with the personal and observational tone of LMN's work.

Selections for inclusion in the reference corpus were based on the following criteria:

1. Personal observations of life experiences prepared in sequential chronology.
2. Texts produced by adults.
3. Late 19th and 20th century writers
4. Availability in electronic format.

The first two journals selected for comparison represent diachronic analyses of language use consistent with the genre of journal across time. The first journal compared was written by a woman in 1881 during a six week period in which she left the comforts of home to live with a band of Native Americans of the Sioux nation. Alice Cunningham Fletcher's diaries (published on-line in 2001) tell the story of this 43 year old single woman who was educated in Boston and eventually went to work at a Harvard University museum. Based on her interest in anthropology and ethnographic observation, she accepted an offer to travel west for an unprecedented study of the way of life of Sioux women. She described the conditions of her journey in great detail, the pace and hardships of traveling by horse-drawn wagon, and the racial prejudices and stereotypes she faced upon her arrival, as well as the ones she brought with her. When Fletcher returned to Boston, she became an activist in Native

American affairs, advocating strongly for education, equitable land assignment, and financial support for a nation of people who were rapidly becoming stranded between two worlds. Her contributions to the knowledge base regarding the Sioux, and eventually other Native cultures are recognized as groundbreaking for her time.

The second journal included in this analysis was written by Charles Inman Barnard (1918), an American newspaper correspondent who was in Paris during the political upheaval at the beginning of what would eventually become World War I. He describes his social interactions, his impressions of patriotism, and his ruminations regarding the conversations and conflicts which shaped the role of France in the turmoil of the time. Barnard wrote as an observer of the process, and although his work reads as a memoir, he noted different motivations:

> These notes, jotted down at odd moments in a diary, are published with the idea of recording, day by day, the aspect, temper, mood, and humor of Paris, when the entire manhood of France responds with profound spontaneous patriotism to the call of mobilization in defense of national existence. (Barnard, ¶ 1)

The third journal included in the reference corpus was prepared by Michael Greger (1999), a medical student who was enduring the harsh tribulations of his third year of training, a time during which his resolve to be a doctor wavered, and he questioned the wisdom of his career choice. He describes his incredulity at the insensitivity and political incorrectness of the attending physicians toward their patients and students, as well as the unfairness of the hierarchical structure inherent in such a training paradigm. Greger noted his observations in a journal format that

included descriptions of his disconnection from the regularities of normal life (e.g., sleep patterns), as well as the passage of time, counted not by months, but by completion of clinical rotations. Of the three journals used as reference corpora, this represents more of a synchronic comparison with LMN's work in terms of language used in the late 20th century by professionals trained at the doctoral level.

Internal reference corpora. In an effort to understand the patterns of language use in LMN's experience with mania, it was also necessary to investigate differences that she has manifested at different points in her life. Intra-individual differences in LMN's language use were addressed by systematically parsing the Master Corpus into sub-corpora for the purpose of addressing differences between her medicated versus unmedicated states; writing produced in manic versus non-manic periods; word choices in her *more* versus *less* acute exacerbations; and linguistic patterns observed during the early, middle and late stages of a single episode. For each of these comparisons, LMN served as her own control, consistent with a single subject methodology.

Analysis

WordSmith Tools. Typed texts were saved as .txt files in a format ready for analysis using *WordSmith Tools, v. 4.0* (Scott, 2005), a program designed to analyze patterns of word use in texts. Wordsmith generates frequency counts and relevant statistics, contextual concordances and patterns of collocation, and key word identification. Scott describes the purpose of frequency counts as a simple way to study the type of vocabulary used in a corpus, and more specifically, to allow for

comparisons between files or genres. Concordancing allows for review of the collocates of a specific word in a context in a fixed span to the right and left of the target word, providing specific details for analysis of variation in language use. The KeyWord feature of *WordSmith Tools* is designed to allow for comparison of two corpora, the larger of which serves as a reference corpus, and identifies the words which characterize the text under analysis as different from the reference. Table 3.2 provides an exemplar of the type of analyses to follow, with the top five keywords as identified in a comparison of LMN's Master Corpus and the FROWN corpus. The first column identifies the word which appears more frequently in the Master Corpus, the frequencies of which are noted by the second and third columns reflecting the actual count and proportion. The fourth and fifth columns indicate the frequency count and proportion of the same word as it appears in FROWN. The last column represents the keyness statistic, indicating the degree of difference between the two values for the key word.

Table 3.2

Example of keywords from comparison of Master Corpus and FROWN.

Key	Master Corpus		FROWN		
Word	Frequency	% of total	Frequency	% of total	Keyness*
am	1948	0.8030042	279	0.0189412	6038.435059
prime	1049	0.4324186	75	0.0051091	3579.685303
I	3828	1.5779775	6794	0.4612425	3188.02832
pm	621	0.2559885	2	0.0001357	2404.529297

SPJ [a]	602	0.2481563	0	0.0000000	2356.464600

* $p<0.00000$ for all keyness values.

[a] Abbreviation for *see previous journal*.

Keyness is a non-parametric statistic calculated by *WordSmith* which is similar to a Chi-square statistic, yielding information about significant differences between word frequencies in two corpora. Positive keyness reflects a higher frequency of occurrence of a word in the target corpus, and inversely, negative keyness suggests rare words in the target corpus, with higher frequency of occurrence in the reference corpus.

Standard *WordSmith* conventions are such that alphabetic notations separated by white space are counted as words, and numeral groups are likewise counted. This does not account for differences between notations of numbers as time constructs or as quantity markers. Nonetheless, the conventions are consistent across corpora so comparisons can be made on equal footing; that is, the tokens identified as # in one corpus are identified the same way in the other, so statistics such as word frequency and keyness are calculated based on identical constructs.

Each of the hypotheses outlined in Chapter One involves descriptive and inferential statistics based on the results from the *WordSmith* program, for the purpose of appreciating the scope of the content of the corpus under investigation, as well as understanding relationships between distributions that would be unlikely to occur by chance and are thus empirically meaningful. (Kretzschmar, Meyer, & Ingegneri, 1994). Throughout the analyses, a distinction is made between occurrence

of *content* and *function* words. Function words (e.g., *and, the, it, of, is, to*) are extremely important in tasks such as author identification (cf. Wendelberger, 2006), as these always tend to occupy the positions of highest frequency of occurrence in linguistic analysis, and the LMN corpus is no exception. Because of their ubiquity, however, comparisons of such words are best studied in large samples of the same genre. The rates of occurrence of function words between the target and reference corpora in this study, especially across varying genre, contribute little to understanding the meaning of language content in this context. Therefore, appearance of function words in all analyses will be severely limited, with particular attention being given to the highly productive content words, which more appropriately reflect the intentions, interests and overall meaning of the writer, and will yield evidence for practical applications in clinical settings.

Description of Master Corpus. The first analysis delineated the Master Corpus based upon observations of LMN's linguistic behaviors as referenced to the diagnostic criteria for mania from the DSM-IV (1994), including increased idea generation; idiosyncratic thought patterns, the expression of which are difficult for others to interpret; expansive affect and related mentation; flight of ideas; and associational chaining. This analysis primarily entailed patterns of repeated use noted as word frequency and collocation. Idea generation appears to be a consistent theme for LMN, and because this is considered to be a pathognomic marker of mania, it was considered in terms of both *process* and *content* relative to frequency of occurrence and patterns of collocation. Synonyms for *idea* were also investigated, with an explication of the semantically similar tokens and their contextual meaning in the

Master Corpus. Patterns of idiosyncratic thinking were predicted based on the inordinately high proportional representation of the token *prime*, which holds special meaning for LMN, and occurs primarily in collocation with notations of date and time. Expansive affect was noted in the context of collocates which predicted positive discourse prosodies, as well as in word frequencies which spoke to her increased levels of productivity and elevated mood states. Flight of ideas was analyzed by means of parsing from the master corpus collocated phrases relative to cognitive distractibility, as well as observations of periods in which there were increased numbers of entries in the span of a very short time. Associational chaining, although fairly rare in this corpus, was documented with repeated patterns of alliterative word use. Other references to non-linguistic manifestations of mania were noted by word frequency counts in which LMN was describing physiological perceptions.

Further analysis of the Master Corpus entailed identification of the top 100 content words in lemmatized forms, and division of these words into semantic categories representative of recurrent patterns in LMN's language use. Not unexpectedly in a corpus linguistic study, semantic preferences based on shared features are reflected in word choice (Stubbs, 2002), repeated patterns of which contribute to the construction of meaning. Lists of content words were derived from analysis of either frequency or keyness, and were then grouped according to an *a priori* determination to have about seven semantic categories (plus or minus two). More groupings would have become too specialized, and fewer would have been less effective in ascertaining meaningful differences. Categories emerged from the process of establishing the frequency and distribution of tokens and keyness, and

were refined by the investigator's knowledge of LMN's unique background and circumstance, rather than on popular or cultural consensus alone. Words were sorted into categories based on empirical rate of occurrence and an understanding of her idiosyncratic patterns of language use. Groupings became evident relative to *career and research, home and family, physiological changes, cognitive processes, the writing process*, and of course, *time and quantity* based on recurrent topical themes, and the content words were assigned to the semantic categories accordingly. For example, her use of the word "food" was particularly applicable in this case to her research based on analysis of word collocations, rather than exclusively pertaining to her home life. Likewise, her use of idiosyncratic referents for reporting her physiological perceptions (e.g., *keno pops*) contributed to specific categorical inclusion, as there was only one possible explanation for the use of this term, once the context was understood. In other analyses, the content words changed, and required a reconfiguration of categories based on the topics with which LMN was preoccupied, as identified through word frequency or keyness patterning. Inter-rater reliability coefficients in the *good* to *excellent* range were obtained for all identified categories by three independent coders whose judgments of agreement were evaluated using Fleiss' kappa coefficient.

Comparison to external corpora. The next analyses involved comparisons of the Master Corpus to the external reference corpora for determination of fundamental differences in language use. FROWN and the Master Corpus were compared and 100 content words were identified in terms of both positive and negative keyness variables. That is, in terms of comparing relative rarity of appearance, the top fifty

content keywords which appeared most frequently in the Master Corpus (but not in FROWN), and likewise the top fifty content keywords from FROWN were identified, as measured by positive and negative keyness, respectively.

Based on the semantic themes identified in LMN's Master Corpus, it was anticipated that her patterns of discourse might represent a significant difference from what would be considered 'typical' language use, as compared to FROWN. To appreciate the magnitude of this presumed difference, key content words from the LMN/FROWN comparison corresponding with the Master Corpus list of content words were drawn from an expanded version of the keyword list (>2000 words), and were categorized into thematic groups of ten exemplary tokens each, selected based on frequency of occurrence. In an effort to control for confounds inherent in word identification, only the clearly obvious abbreviations were included in this analysis, eliminating those for which homonymous substitutions could be made as words are taken out of context (e.g., *Feb* was included as this clearly represents *February*, but *May* was not included as it could represent either the month, a polite question formulation, or a modal auxiliary marker). These ten words were ranked within categories according to degree of keyness, comparing the frequency of occurrence and the proportional representation of each word in each corpus. Differences between the percentage of occurrence in groups based upon semantic classification were analyzed using a *z* test for proportions (Hardyck & Petrinovich, 1969). The statistical differences were predicted to be significant. The reliability of semantic classifications was verified by means of inter-coder reliability, the details of which are explicated in the results found in Chapter Four.

To provide another perspective for appreciation of the differences in terms of word frequency, a descriptive table of relative ranking was generated which demonstrates the order in which the key content words appeared in the Master Corpus lemmatized list, their comparative rank from the FROWN corpus, and the statistical significance between the two.

The second comparative analysis involving external corpora addressed the question of the journal-as-genre corpus to the Master Corpus. This involved identification of the top 100 content words which were identified in terms of both positive and negative keyness variables. As with the previous analysis, to compare relative rarity of appearance, the top fifty content keywords which appeared most frequently in the Master Corpus and the top fifty content keywords from the journal-as-genre corpus were identified, as measured by positive and negative keyness, respectively. Likewise, another descriptive table of relative ranking provides insight into the comparative rank, proportional representation, and statistical significance of differences in word frequencies between the two corpora, based on the lemmatized table of top content words from the Master Corpus. These analyses approached the question of difference in patterns of language use from the perspective of genre, and the patterns of word frequency were interpreted relative to their consistency with such a framework, including observation of chronology and the personal perspective on the immediate circumstance of the writer.

Within subject comparisons. The next set of analyses addressed intra-individual comparisons of selected sub-corpora of the Master Corpus, using LMN as her own control. The first of these (and the third hypothesis tested) targeted

differences in patterns of language use between conditions of being unmedicated versus medicated. As noted, LMN did not begin pharmacological intervention until seven years into her process, and after taking the medicine for nearly nineteen years, she made the decision to discontinue the intervention in late 2003. The division of these time periods is noted as *early unmedicated* period from 1978 through 1984, *medicated* period from 1985 to 2003, and *late unmedicated* period from 2003 to 2005. The early and late unmedicated time periods were combined to represent the years in which LMN was not taking the mood-stabilizing medication for the purpose of comparison to her language use during the years in which she was receiving the pharmacological intervention. There are inherent confounds in such comparisons, notably, that the late unmedicated sample is influenced by the introduction of external observers, the result of which involves the researcher's name appearing as one of the keywords (not used in further analysis). There are also variations noted relative to LMN's diachronic interests and preoccupations, as the early texts are notable for commentary on her research and academic activities related to her career; predictably, the later texts involve her interests post-retirement, including increased attention to her journals and the establishment of an archive of her work. The most obvious interpretive confound has to do with the sampling procedures. In order to make definitive statements about differences in the two conditions described, transcriptions of her entire body of work would be necessary to allow for specific analysis of word count per entry, number of entries per day and per episode, and the number of episodes per time periods, none of which can be calculated based on the available corpus. Acknowledging these confounds, the hypothesis nevertheless asks if there

are differences between medicated versus unmedicated conditions, and to address this with the currently available data set, the texts from the early and late unmedicated time periods were compiled into a single batch for comparison to the medicated time period.

The top 100 key content words from this analysis were identified from the positive and negative keyword lists as generated by *WordSmith Tools*. These were sorted into semantic categories, again based on observer judgment and confirmed with inter-coder reliability checks, drawing from both the positive and negative keyness variables, as certain words appeared with more frequency in either the unmedicated or medicated corpora. Frequencies of occurrence were summed and Chi-square statistics were performed to determine if there were significant differences in the patterns of language use (Connor-Linton, 2003) across semantic categories. Further analysis of these findings involved identification of differences between the rates of occurrence of the *expected* word frequencies and the *observed* word frequencies based on the Chi-square results. A proportional representation was calculated using the observed number as numerator, divided by the expected number as denominator, allowing for comparison of relative distributions of semantic representation between the two conditions of unmedicated versus medicated.

The fourth hypothesis tested involved comparison of samples of LMN's writing that was generated during manic episodes with the non-journal writing habits of the same individual as produced during non-manic episodes. LMN offered the use of personal correspondence produced as letters and e-mails to friends and family which was compiled into a corpus totaling 31,815 words. In this more synchronic

milieu, the texts included often preceded or followed episodes of mania by only a few days. Although she was a fairly prolific writer during her career as an academic researcher, samples of texts generated and published in peer-reviewed journals, newspaper columns, or other professional capacities were not included due to the extensive editing and review process intrinsic to such work. Therefore, to retain more consistent and conversational stylistic patterns of language use as utilized in her journal keeping habits, the less formal epistolary genre was employed, and although it is not identical to the stream of consciousness writing employed in the journals, the within subject comparison was considered to be a valid one. Texts provided were generated during non-manic periods in non-journal format as reported by LMN were cross-referenced to the timeline generated by placing the journals in chronological order. Any correspondence for which there was overlap (i.e., anything which may have been written during a manic episode) was eliminated from inclusion in this corpus. Analysis involved formulation of keyword lists, comparing the master corpus texts generated during manic episodes with the smaller corpus of correspondence produced during periods when LMN was not experiencing mania, identifying the top fifty positive and negative keywords which represent highest frequencies of occurrence in the two corpora (i.e., which words appeared in one, but were rare in the other). Notably, texts included in the non-manic writing corpus were of recent vintage, which could accurately be described as subject to the observer influence. However, because these pieces of correspondence were produced during time periods when LMN was not experiencing a manic episode and were written for a specific audience in mind, these texts are not subject to the same exclusionary criteria.

Presumably, all of the casual epistolary writing produced in non-manic periods across the course of her adult life would be fairly consistent in style, with the expected variations in content as life circumstances change. Qualitative observations of pragmatic function of language use in the context of interpersonal communication as compared with the self-as-audience structure inherent in the keeping of journals were also made.

The fifth comparison addressed intra-individual differences in terms of episodic acuity. According to the participant's historical recollection, the manic episode experienced in July, 1985 was a benchmark in terms of written production, a time period in which she produced over a thousand pages of handwritten notes in the space of twelve days immediately prior to hospitalization for accidental self-induced lithium toxicity. In her own words, she described that particular episode as "really bad", observing that no episodes preceding or since that time have manifested that acutely, although the initial experience in 1978 may have been similar.

The texts generated during that episode (henceforth referred to as the July 1985 corpus) were parsed from the master corpus, and used as a point of diachronic reference for comparison with the rest of the master corpus, during which manic episodes evidenced in more typical and expected patterns of behavior, according to LMN's description. At the first level of analysis, keyword comparisons were performed, and the top key content words were identified in positive and negative tables, and then summarized according to semantic categories. In this analysis, as with the previous such categorization, group labels were based on intuitive clusters of words, and inclusion of these keywords (including positive and negative keyness) into

these specific categories was verified by means of inter-rater reliability, the details of which are discussed in the results found in Chapter Five. Statistical differences between the categories were calculated using a Chi-square statistic. As in a previous analysis, further investigation into these findings involved identification of differences between the rates of occurrence of the *expected* word frequencies and the *observed* word frequencies based on the Chi-square results. Again, a proportional representation was calculated using the observed number of target words as numerator, divided by the expected rate of occurrence as denominator, allowing for comparison of relative differences of occurrences of semantic representation between the *more* versus *less* acute manifestations of mania.

The final hypothesis tested approached the analysis from a different perspective, addressing intra-individual variation within the course of a single episode. In terms of episodic acuity, LMN reports that she knows when an episode is beginning because she feels an increased motivation to begin writing: "It seems that another episode has begun." (08 Feb 90 4:22:34). Likewise, her observations of the waning phase of an episode are also reflected in comments in her journals, for example:

> 07 Aug 1983 8:54:12 pm
>
> Don't really know what 'terminates' a period. All I know is that I definitely do KNOW when its over. In Sept 79 it was very obvious - ~ 4 PM. I wrote soon after that 'Hurricane LMN has subsided'. Can't really describe how I know when it's over. One sign (not how I know) is that I sleep through the night. 8:57:13 pm

The research question addressed here deals with changes in her language patterns based on changes in acuity within a single episode. For this molecular synchronic/diachronic analysis, texts generated during a single typical episode were selected (March 10-30, 1984), yielding a considerably smaller corpus, but nonetheless one which was useful for the purpose of descriptive analysis. The identification of onset and resolution of mania was determined by the parameters based on LMN's self-report: she is motivated to begin writing at the advent of an episode, and as her motivation decreases and entries become less frequent and spaced further apart, she becomes aware that an episode has come to a conclusion.

For the purpose of this study, the unit of analysis was an *entry*. Journal entries were defined as beginning with a time stamp, recording a thought as text, and then closing with either a time stamp or white space. The break in discourse was apparent between entries as the next entry began with another time stamp, which preceded the next segment of text. In LMN's journals, the entries varied in length, according to the number of words written as well as the amount of time involved in the actual writing process. The process for identifying entries in this research involved full text coding, rather than part of speech or topical coding, both of which may provide fruitful lines of future research.

For this hypothesis, an entry-based analysis was employed: after separating the single episode into equal early, middle and late phases based on the total number of days noted in the text divided into equal thirds, the number of journal entries per day was manually counted, and divided into two time periods: the six hour span from midnight to 6:00 a.m., and the eighteen hour span from 6:00 a.m. to midnight. These

segments were selected for analysis since individuals experiencing the acute phase of mania often appear energized or prolific during late night/early morning hours when they would typically be asleep, reflecting the neurovegetative changes and increased productivity (in LMN's case, increased writing) which represent another hallmark diagnostic sign of the disorder. Other analyses involved identification of the most frequently occurring key content words across the three time periods, and establishing averages based on number of entries in the early, middle, and late periods, and per time period within each of those segments.

Other Methodological Considerations

Construction of audience. LMN initiated the current investigation, contacting a university-based corpus linguist to offer her collection of texts for analysis as she was retiring from her thirty year career in academia. As noted, she began the collection as a data set, with the intention of conducting her own research, and at some point, making the data available to other researchers. As such, her construction of audience was initially *self* plus an undefined "other", ostensibly researchers who would use her data for specific investigation into the phenomenon of mania. Over the years, her audience construction remained fairly consistent, with a significant shift in perspective occurring at the time of the introduction to the university-based research group. Although an argument can be made that the entire corpus is subject to the confound of the observer's paradox because of her early intentions, it was not until specific investigators were identified that the perceptible shifts in her writing style, word choice, and sentence structure became evident. Likewise, even if she was intent

upon sharing the journals to advance the purpose of research or achieving other personal goal, this mindset would be consistent with the egocentric manifestations that characterize mania. Thus, the perception of early audience construction did not confound the representation of linguistic behaviors associated with the phenomenon under consideration.

Language variation and categorical identification. Acknowledging the interdisciplinary nature of this methodology requires accommodation for the perspective of the continuous nature of language variation, while recognizing the categorical nature of clinical research. The study of meaning in language does not allow for binary distribution of variables as black or white because of the inherent variation and shades of gray interwoven therein by contextual constraints and individual processes in linguistics. When applying language variation theory to the constructs inherent in the categorical world of clinical methodology, however, it is only at the level of extreme variation that labels can be accurately applied. The constellation of signs and symptoms that yield diagnoses of mental illness may percolate at subclinical levels at which an individual may be able to function without consequence. Even at the point of one symptom reaching an extreme, it may not be representative of pathology if another more parsimonious explanation can be made (e.g., evidence of thought disordered speech among 'normals' in cases of fatigue or intoxication). However, if no other explanation is found, and a behavior (linguistic or otherwise) manifests to a degree that the ability to adequately adapt to the environment becomes compromised, then a categorical diagnosis is made. It is at this intersection of categorical and continuous that the philosophical perspectives of

idiographic language use are overlaid onto the nomothetic criteria for inclusion into diagnostic groupings.

Related to this, the diagnostic criteria for Bipolar Disorder – Mania Only as defined in the DSM-IV (1994), as is the case with most diagnoses of mental illness, involves a description of a cluster of linguistic, cognitive, emotional, and behavioral *processes* which manifest differently in individuals based on the *content* of their language, thinking, or actions. For example, in Delusional Paranoid Disorder, the process may involve a theme of persecution with content that may variously include being spied upon by the government, conspired against by family members, harassed by co-workers, or maligned by neighbors. In the case of mania, the processes that can be marked linguistically typically involve grandiose idea generation (with *idea* frequently manifesting as both process and content), being more talkative than usual (increased linguistic production), flight of ideas, distractibility, and increased goal directed activity, all within the context of an elevated or expansive mood state.

The notion of a continuum of linguistic behavior is consistent with that of a constellation of diagnostic markers in mental illness; however, this only increases the variance within which a diagnosis must be made. Including other non-clinical aspects of diagnostic formulation such as educational, emotional, or socioeconomic status underscores the basic fact that the DSM-IV (1994) criteria are difficult to map alongside a linguistic framework because of the inadequate delineation of definitions offered as clinical markers. Further, although certain language-based terms have been incorporated into the diagnostic argot in mental health, there is little accommodation for determining at what point on the continuum of linguistic behavior

the observable manifestations of illness qualify as extreme enough to warrant a diagnosis. It is the aim of this research to inform this discussion.

Chapter 4 - Analysis of Master Corpus and External Corpora

A journal is like a cave. What will we find there?

Macrorie, 1987

This chapter describes the results of analysis of the content and patterns of language use in the Master Corpus, with comparisons made to the diagnostic criteria for mania according to the DSM-IV (1994). Comparisons are then drawn between the Master Corpus and the external reference corpora, offering a nomothetic perspective to this investigation, as LMN's language use is compared to external standards representing patterns of 'typical' language use.

The Master Corpus

Of the 100+ journals produced in the last 28 years, the selections of text transcribed from the original documents were compiled into a corpus totaling 242,589 words. As noted, the original documents include drawings, markings identified as 'automatic sketches', and other non-linguistic markings, which are not pertinent to the current analysis, as the research questions deal with the linguistic content only. This collection, referred to as the Master Corpus, includes journals representative of the spans of time that represent early, middle, and late phases of LMN's process, excluding texts produced after 2003. The analysis of the body of work in the master corpus begins with a description of her language use which demonstrates qualitative congruence with her diagnosis of mania. The fundamental examination of the data

served to situate LMN's patterns of written language use in the spectrum of signs and symptoms of mania, as described in the DSM-IV (American Psychiatric Association, 1994) and other texts (c.f. Kaplan & Sadock, 1998). These include speech patterns which lack cohesion and are difficult to interpret; word selection which may be governed by phonological properties rather than content; flight of ideas involving increased rate of production with rapid topical shifts; and other linguistically-mediated behaviors including grandiose idea generation, increased organization and planning, and increased socialization. The DSM-IV also includes descriptions of concomitant non-linguistic behaviors such as expansive affect, psychomotor agitation, increased energy, and decreased need for sleep, the notation of which takes on specific linguistic patterning. Unfortunately, the inclusion criteria from the DSM-IV used for diagnosing mania and other mental illnesses involving patterns of language provides only cursory definitional statements of general function, without specific linguistic markers or indicators of severity along a discernible continuum. Although there seems to be agreement within the clinical community on how to interpret constructs such as *poverty of speech* or *thought disorder*, accurate measurement remains tentative, and attempts to quantify these concepts in the literature has led to further obfuscation with labels such as *verbigeration* or *extreme fabulizing*.

Word frequency in the Master Corpus

From a descriptive statistical perspective, the establishment of the master corpus as baseline involved identification of the top 100 content words as noted in Table 4.1. In compiling this list, familiar function words (e.g., *to, the, of, a, in, is,*

and, it, be) were eliminated so that the content words particular to LMN's corpus could be identified. Since the methodological orientation of corpus linguistics involves analysis of patterning in language use utilizing actual data from a pragmatic/discourse perspective, interpretation of what constituted content words was based on conceptual/semantic relations, rather than syntactic or morphological rules. Hence, lemmatized versions of frequently occurring words were combined to reflect the actual impact of specific word use in the corpus. In two cases, there were abbreviations for words that were also spelled out in other instances (i.e., *lithium* is also noted as *Li*; and the name of the university where LMN was on faculty was referenced on occasion by its initials). In these cases, the abbreviations were expanded into complete word form, as cross-checked by concordancing for context, and included in the count. Likewise, for some content words, variations on spelling and other use were also accommodated as lemmatized forms (e.g., *lithum*; *Lithobid*) as noted in the table to allow for full appreciation of the actual scope of word patterning. In order to protect the confidentiality of individuals specifically referenced in the corpus, family members are identified by their role, rather than by their name, and naturally, *LMN* is used to represent occurrences of the participant's real name. Additionally, the identifying markers of the town she lives in and the university with which she was affiliated are also assigned generic labels. These substitute markers are noted in bold and italic font.

Table 4.1

Top 100 content words from Master Corpus.

Rank	Word	Frequency (242,589 words)	Percentage of total
1	# (representing numerals)	39227	16.170146
2	prime	1049	0.4324186
3	think(ing)/thought(s) [a]	876	0.3611045
4	pm	621	0.2559885
5	good	606	0.2498052
6	SPJ [b]	602	0.2481563
7	one	535	0.2205376
8	smile	528	0.2176520
9	entry(ies) [a]	516	0.2127054
10	January [b]	514	0.2118810
11	now	504	0.2077588
12	idea(s) [a]	503	0.2073465
13	feel [a]	460	0.1896211
14	***LMN***	401	0.1653001
15	time	378	0.1558191
16	November [b]	370	0.1525213
17	data	368	0.1516969
18	when	364	0.1500480
19	tired	355	0.1463380

20	left	349	0.1438647
21	know	315	0.1298492
22	right	309	0.1273759
23	lithium [a, b]	306	0.1261692
24	since	289	0.1191315
25	work	266	0.1096504
26	July [b]	261	0.1075893
27	model [a]	255	0.1051160
28	yes	255	0.1051160
29	first	250	0.1030549
30	August [b]	237	0.0976961
31	shook	237	0.0976961
32	arm(s) [a]	235	0.0968716
33	before	226	0.0931616
34	decide/decision [a]	224	0.0923372
35	***child('s)*** [a]	221	0.0911005
36	February [b]	219	0.0902761
37	leg(s) [a]	216	0.0890394
38	breath/breathe [a]	213	0.0878028
39	morning	213	0.0878028
40	two	213	0.0878028
41	hard	209	0.0861539
42	food	208	0.0857417

43	minute(s) [a, b]	207	0.0853295
44	computer	205	0.0845050
45	count	200	0.0824439
46	coffee	199	0.0820317
47	new	193	0.0795584
48	journal	176	0.0725506
49	deep [a]	174	0.0717262
50	oh	173	0.0713140
51	after	172	0.0709018
52	again	170	0.0700773
53	many	170	0.0700773
54	sleep	166	0.0684284
55	write	166	0.0684284
56	book	164	0.0676040
57	record	164	0.0676040
58	three	164	0.0676040
59	eyes	157	0.0647185
60	mania/manic [a]	156	0.0643062
61	woke	146	0.0601841
62	***town***	145	0.0597718
63	water	143	0.0589474
64	few	142	0.0585352
65	family	137	0.0564741

66	home	136	0.0560619
67	color	129	0.0531763
68	day	128	0.0527641
69	program	127	0.0523519
70	September [b]	127	0.0523519
71	nausea	121	0.0498786
72	December [b]	119	0.0490541
73	April [b]	117	0.0482297
74	paper	116	0.0478175
75	number	115	0.0474052
76	research	114	0.0469930
77	spell [a, b]	113	0.0465808
78	study	112	0.0461686
79	TV	112	0.0461686
80	problem	109	0.0449319
81	earlier	108	0.0445197
82	previous	108	0.0445197
83	slight [b]	108	0.0445197
84	system	108	0.0445197
85	***name of university*** [b]	106	0.0436953
86	brain	105	0.0432830
87	couch	104	0.0428708
88	course	101	0.0416342

89	today	101	0.0416342
90	***husband('s)*** [a]	100	0.0412219
91	group	99	0.0408097
92	important	99	0.0408097
93	office	99	0.0408097
94	proposal	99	0.0408097
95	radio	99	0.0408097
96	nutrition	92	0.0379242
97	project	91	0.037512
98	diet	87	0.0358631
99	class	85	0.0350386
100	student	85	0.0350386

[a] Count includes lemmatized versions of content word.

[b] Count includes abbreviated versions of same word.

LMN's journals are remarkable for repeated themes, as she notes the passage of time and observations about her awareness of and response to her episodes. The list of content words was divided into semantic associational categories for identification of patterns of regularity related to the most significant events and commitments in her life. As noted, during period of mania, these tended to revolve around her fixation with numbers and time, her cognitive and perceptual experiences with the episodic process, her work and home life, and the writing process involved in keeping journals. Selection of items into categories was subjective, grounded in the

theory of the significance of repeated events (Stubbs, 2001) and a working knowledge of the context of LMN's patterns of word use. Selection was verified by means of inter-coder reliability checks, in which three external raters were asked to place the 100 top content words into the same six categories, plus the option of a category labeled *other* if no satisfactory slot could be determined. The raters were informed of the meaning of the idiosyncratic *SPJ* (see previous journal), as well as the token *LMN*, which also appeared among the top 100 words. Reliability was calculated using Fleiss' kappa coefficient of inter-coder concordance, which yielded $\kappa\hat{} = .76$, achieving a level of reliability deemed by convention to be *excellent* (Fleiss, 2003).

There was consistent disagreement on two tokens: s*light* and *LMN*. All external coders placed the token *slight* in the category of *quantity/temporal* words; however, examination of the word in context by means of collocation indicated that it appeared as a modifier of clearly physiological phenomena, as LMN used it this way in over 93% of total appearances of the word in the Master Corpus. For example, SLIGHT* in all its lemmatized forms (i.e., * representing *slight, sl, slightly*) plus <nausea/nauseous> in the R1 position collocated on 54% of the total occurrences, (59 out of 108), with other representations of physical perceptions accounting for 39% of the total (e.g., SLIGHT* collocating with <gag, headache, hungry, tight, numb>). The token *LMN* was placed in a variety of categories by the raters, explained by relation to *family/home* because it was a personal name, *cognitive processes* because she was thinking about herself, and *other* because there appeared to be no better fit. The justification for including it in the *writing process* category was that the token *LMN* served as a signature frequently used when she was noting

the date on a new page, when she had finished an entry, or at the point she was bring a volume to a close. Inclusion of these two tokens in the designated categories was based on knowledge of their specific function and contextual loci, an interpretive parameter that was not available to the external raters. Table 4.2 presents the categorical division of the top 100 content words.

Table 4.2

Semantic and conceptual categorization of top 100 content words.

Quantity/temporal:	# [numerals in text], prime, pm, one, now, January, time, November, when, since, July, first, August, before, February, morning, two, minute(s), count, after, again, many, three, few, day, September, December, April, number, earlier, previous, today
The writing process:	SPJ, entry(ies), LMN, journal, write, book, record, spell
Cognitive experiences:	think(ing)/thought(s), good, smile, idea(s), decide/ decision, feel, know, lithium, yes, new, oh, mania/manic, important
Physiological perception:	tired, left, right, shook, hard, arms, breath/breathe, deep, sleep, eyes, legs, woke, water, color, nausea slight, brain, radio
Family and home life:	*child*, coffee, *town*, family, home, TV, couch, *husband*

Career and research:	data, work, food, computer, program, model, paper, research, study, problem, system, *university*, course, group, office, proposal, nutrition, project, diet, class, student

Analysis 1: Comparison of Master Corpus to Typical Language Use

After the master corpus was compiled, consisting of 242,589 words, the first comparative study involved investigation of these texts as compared to a reference corpus typical of written language use. Hypothesis 1 states: There are measurable differences in the written texts of this participant as compared to corpora of 'normal' language use across genres using the Freiburg-Brown Corpus of American English (FROWN) corpus. The FROWN corpus consists of 1,472,978 words, and as noted in the previous chapter, it was determined to be an appropriate reference corpus based on size and genre. There were some remarkable matches as well, in terms of content words. For example, in both the Master Corpus and FROWN, numerals in the texts (indicated as # in the tables) were the most frequently occurring tokens, accounting for over 15% of the total in each. However, in terms of significance, the Master Corpus included numbers at a rate that was disproportionately larger, producing a keyness statistic (Scott, 2005) associated with a probability (*p*) value so small that it was reported by the program as less than 0.0000. In addition to #, the top ten words ranked according to frequency of occurrence in both corpora included *the, of, to, a, in,*

is, and *for*, all of which represent function words, a predictable finding with corpora of this size.

LMN's journals were remarkable for the consistent linguistic themes noted across the years, as previously noted in Table 4.2. To assess the significant differences in word frequency between LMN's journals and the FROWN corpus, the top 100 key content words are identified, representing both positive and negative keyness, inclusive of all abbreviations and lemma forms as written in the journals (i.e., not collapsed into semantic or orthographic groupings). Table 4.3 presents the top fifty content words from the Master Corpus-FROWN analysis, ranked in terms of positive keyness, that is, words which appear more frequently in the Master Corpus than in FROWN.

Table 4.3.

Top fifty positive key content words from Master Corpus as compared to FROWN.

	Master Corpus		FROWN		
Word	Freq	% of total	Freq	% of total	Keyness
am	1948	0.8030042	279	0.0189412	6038.435059
prime	1049	0.4324186	75	0.0051091	3579.685303
I	3828	1.5779775	6794	0.4612425	3188.028320
pm	621	0.2559885	2	0.0001357	2404.529297
SPJ [a]	602	0.2481563	0	0.0000000	2356.464600
Jan	499	0.2056977	12	0.0008146	1843.000977
etc	470	0.1937433	35	0.0023761	1595.836914

smile	528	0.2176520	93	0.0063137	1570.477661
LMN	401	0.1653001	0	0.0000000	1569.385986
entry	436	0.1797278	48	0.0032587	1408.111694
Nov	359	0.1479869	24	0.0016293	1232.843140
Mar	317	0.1306736	1	0.0000678	1227.326050
tired	355	0.1463380	49	0.0033265	1105.678345
very	627	0.2584618	752	0.0510530	782.826843
lithium	190	0.0783217	2	0.0001357	721.830200
idea	356	0.1467502	206	0.0139853	717.372192
Feb	194	0.0799706	6	0.0004073	707.043029
good	606	0.2498052	788	0.0534971	703.241943
Aug	218	0.0898639	25	0.0016927	699.613403
data	368	0.1516969	257	0.0174476	671.835205
feel	345	0.1422158	221	0.0150036	660.198425
July	245	0.1009938	67	0.0045486	654.538269
I’m	475	0.1958044	508	0.0344879	652.143737
shook	237	0.0976961	62	0.0042091	641.052490
child	150	0.0618329	3	0.0002036	558.301208
room	389	0.1603535	419	0.0284458	530.979675
breathing	175	0.0721384	28	0.0019009	530.402893
town	145	0.0597718	7	0.0004752	512.721252
going	436	0.1797278	566	0.0384256	506.590606
woke	146	0.0601840	10	0.0006788	500.022705

journal	176	0.0725506	40	0.0027155	493.856506
coffee	199	0.0820317	71	0.0048201	489.201019
count	200	0.0824439	79	0.0053632	474.147705
min	138	0.0568863	10	0.0006788	469.805816
nausea	121	0.0498786	1	0.0000678	462.140289
yes	255	0.1051160	195	0.0132385	441.452148
computer	205	0.0845050	103	0.0069926	440.991638
Li [b]	114	0.0469930	2	0.0001357	426.446167
my	793	0.3268903	1956	0.1327922	396.562927
Apr	100	0.0412219	0	0.0000000	391.261444
left	349	0.1438647	472	0.0320439	389.908783
Dec	115	0.0474052	10	0.0006788	383.312255
arms	185	0.0762606	104	0.0070605	377.949462
need	339	0.1397425	467	0.0317045	371.955474
sl [c]	95	0.0391608	0	0.0000000	371.696685
sleep	166	0.0684284	81	0.005499	361.658721
legs	152	0.0626574	62	0.0042091	356.026092
could	633	0.2609351	1472	0.0999336	351.560882
university	89	0.0366875	0	0.0000000	348.219238
morning	213	0.0878028	187	0.0126954	337.621551

Note. Words listed in italics are used to maintain confidentiality.

*p< 0.0000 for all keyness values.

[a] Abbreviation for *see previous journal*.

[b] Abbreviation for *lithium*.

[c] Abbreviation for *slight*.

Analysis of differences between the master corpus and FROWN as reference corpus using the KeyWord feature of WordSmith Tools suggests that *am* is the token for which there is the most significant difference in frequency of occurrence. However, this particular result is not necessarily representative of the semantic variation expected, due to the combination of syntactic and orthographic forms of the word in the master corpus. That is, a disproportionate representation of *a.m.* to indicate the morning hours of the day was expected, but in this case, the count also included the singular present tense form of the copula (i.e., "I *am*"). Analysis of collocates indicate that the unpunctuated form *am* was consistently used to indicate time stamps in the texts, and as noted, this token occurs nearly 2,000 times in the corpus in one form or another. This confound is due to a number of factors, including differences in how LMN may have noted the term indicating the morning hours, as well as inconsistencies in transcription interpretation and typing conventions. Hence the term is reported in this and other tables, but is excluded from further consideration.

Not unexpectedly, the patterns of words that emerged as occurring at a remarkably different rate when compared to a normal corpus is similar to the set of frequently occurring content words identified in the description of the Master Corpus. An obvious exception to this is the token #. Although the representation of numerals in LMN's texts is far and away her most frequently occurring token, it did not appear in the list of top fifty key content words as ranked according to extremes in keyness.

The # token represents a substantial proportion of both corpora, but the keyness statistic of 16.68 (p= 0.000044) indicates that while there is a significant difference in the proportions of numbers in both corpora, it did not to rise to the level observed among the tokens included in Table 4.3, and hence was not included in the top fifty.

Acknowledging that content of language is as conspicuous for what is excluded as it is for what is included, the top fifty negative keywords from the Master Corpus-FROWN analysis were compiled for comparison. Table 4.4 provides such a perspective. As before, the columns portray frequency and percentage of occurrence of key words in each corpus, ranked according to the degree of negative keyness, which denotes the difference between the frequencies of the words in the two corpora, with this set being more representative of the FROWN corpus.

Table 4.4

Top fifty negative key content words from Master Corpus as compared to FROWN.

	Master Corpus		FROWN		
Word	Frequency	% of total	Frequency	% of total	Keyness*
man	5	0.0020610	767	0.0520713	-193.120101
women	5	0.0020610	758	0.0514603	-190.492263
war	4	0.0016488	538	0.0365246	-132.476409
states	4	0.0016488	525	0.0356420	-128.706192
against	9	0.0037099	596	0.0404622	-123.358230
American	14	0.0057710	660	0.0448071	-119.863212
formula	3	0.0012366	462	0.0313650	-116.385208

public	9	0.0037099	555	0.0376787	-112.125595
president	7	0.0028855	468	0.0317723	-97.163368
men	11	0.0045344	527	0.0357778	-96.392601
years	60	0.0247331	1046	0.0710126	-87.337432
young	7	0.0028855	422	0.0286494	-84.570915
father	4	0.0016488	361	0.0245081	-81.669372
city	6	0.0024733	380	0.0257980	-77.477363
year	40	0.0164887	771	0.0523429	-72.861618
old	33	0.0136032	692	0.0469796	-71.736465
woman	11	0.0045344	425	0.0288531	-69.961463
street	3	0.0012366	278	0.0188733	-63.303283
death	7	0.0028855	335	0.0227430	-61.240173
law	8	0.0032977	345	0.0234219	-60.092414
century	4	0.0016488	275	0.0186696	-57.604949
white	29	0.0119543	582	0.0395117	-57.558109
black	36	0.0148399	653	0.0443319	-57.359535
tax	4	0.0016488	264	0.0179228	-54.574546
world	48	0.0197865	750	0.0509172	-53.594638
state	44	0.0181376	706	0.0479301	-52.503452
Clinton	5	0.0020610	262	0.0177870	-49.770267
federal	3	0.0012366	221	0.0150036	-47.289073
court	3	0.0012366	216	0.0146641	-45.900676
act	7	0.0028855	270	0.0183302	-44.402709

God	6	0.0024733	256	0.0173797	-44.355846
late	5	0.0020610	240	0.0162935	-43.929630
general	20	0.0082443	419	0.0284457	-43.387680
national	19	0.0078321	408	0.0276989	-43.335590
under	33	0.0136032	549	0.0372714	-43.004520
view	4	0.0016488	219	0.0148678	-42.334201
natural	3	0.0012366	203	0.0137816	-42.306098
poor	3	0.0012366	202	0.0137137	-42.030559
dead	3	0.0012366	201	0.0136458	-41.755168
local	9	0.0037099	283	0.0192127	-41.155025
wife	3	0.0012366	197	0.0133742	-40.655075
legal	3	0.0012366	196	0.0133063	-40.380432
training	6	0.0024733	239	0.0162256	-39.986324
case	18	0.0074199	367	0.0249155	-36.922435
south	8	0.0032977	251	0.0170403	-36.447155
John	16	0.0065955	340	0.0230824	-35.728065
country	15	0.0061832	328	0.0222678	-35.476722
company	9	0.0037099	259	0.0175834	-35.405261
moral	5	0.0020610	206	0.0139852	-35.071956
money	22	0.0090688	398	0.0270200	-34.840271

*$p < 0.0000$ for all keyness values.

As noted, the FROWN corpus was compiled from written texts of varying genre in 1991, including newspapers, government documents, and fiction writing (Meyer, 2002). Observation of this set of content words suggests the possibility that because of the dearth of representation in the Master Corpus, LMN may have been less attuned to the current events outside her own environment. As an example, the terms *war, law, tax, federal, national,* and *president* in combination account for only .0169% of her total words, compared to the same words accounting for .1344% of FROWN. In addition, other constructs related to daily life as reflected in FROWN were proportionately under-represented in the Master Corpus, as the token *money* was noted only one third as frequently (0.0270% in FROWN and 0.0090% in the Master Corpus); and *local, city,* and *street* collectively accounted for 0.0638% of FROWN, but only 0.0074% of the Master Corpus. From a spiritual/existential perspective, FROWN includes the tokens *death/dead, God,* and *moral* as 0.0677% of the total, with proportional representation in the Master Corpus writing observed as only 0.0086% of the total. The proportional use of the terms *man/men* and *woman/women* was 0.1618% in FROWN, but only 0.0131% in the Master Corpus. These observations may be seen as specious because of variations in genre between the two corpora, but also support the notion of egocentric tendencies and ostensibly limited insight as observed in mania.

Returning to the prominent themes in LMN's writing, the semantic categories identified in the description of the Master Corpus were employed as another point of comparison with the FROWN corpus. Content words drawn from an expanded version of the positive keyword list (>2000 words) were categorized into themes

represented by ten exemplary tokens each, and are presented in separate tables with the corresponding tokens from the FROWN corpus, in an effort to clarify the proportional differences in representation between these two corpora. Once again, inter-coder reliability checks verified selection of items into semantic groupings, as three external raters placed the content words into the same six categories, plus *other* to be used if no satisfactory slot could be determined. As in the previous analysis, the raters were informed of the meaning of the idiosyncratic *SPJ* ("see previous journal"), but no other clarification was provided. Not unexpectedly, the rate of agreement would have been higher had raters been privy to the meaning of certain contextually-dependent tokens (e.g., *diet* as it relates to LMN's research, rather than as it might relate to family/home environment). Fleiss' kappa coefficient of inter-rater reliability was calculated, yielding $\kappa\hat{} = .554$, falling within the range of .40 to .75 deemed by convention to be an indicator of *good* reliability (Fleiss, 2003).

Notably, the significance of the keyness statistic for every term in each of the following six tables was $p < 0.0000$. To control for confounds inherent in word identification, as described in Chapter Three, only the clearly obvious abbreviations of words were included in this level of analysis (e.g., *pm* was included, but not *am*). As expected the proportional representation of the semantic composites rendered significant *z* values in the statistical comparisons, and the degree of difference rendered confirms the hypothesis being tested

Quantity and time. As has been discussed, concepts related to time and numbers hold special significance for LMN, and this is reflected in her word choice in writing. Words in this category accounted for over eighteen percent of her total

lexicon, as noted by just the ten exemplars used in Table 4.5. Although the use of numerals in the text is nearly as well-represented in the FROWN corpus, the difference between LMN's use of other time and quantity constructs and that observed in samples of writing collected across genres reaches the level of statistical significance, and is consistent with the observations noted linking her language use and the diagnostic criteria for mania. The statistical significance between the representative proportions of each word in the respective corpora is indicated by the miniscule *p* value that carries well beyond four decimal places, clearly delineating the difference in discourse topicality as observed in LMN's writing.

Table 4.5

Comparisons of keywords related to quantity and time.

	Master Corpus		FROWN		
Keyword	frequency	% total	frequency	% total	Keyness
prime	1049	0.4324	75	0.0050	3579.69
p.m.	621	0.2559	2	0.0001	2404.53
Jan	499	0.2056	12	0.0008	1843.00
Nov	359	0.1479	24	0.0016	1232.84
Feb	194	0.0799	6	0.0004	707.04
Aug	218	0.0898	25	0.0016	699.61
July	245	0.1009	67	0.0045	654.54
count	200	0.0824	79	0.0053	474.15
now	504	0.2077	1,171	0.0794	280.25

# (numerals)	39,227	16.1701	233,355	15.8423	16.68
*Total %		18.3035		15.9410 (σ_p 0.0743)	

*$z = 31.79$, $p<0.0000$.

Writing process. LMN's involvement in the writing process is notable not only for the sheer volume of words she has generated in the past 28 years, but also for her tendency to situate herself as an observer of her own process. Based on her research training and interests, she approached the keeping of a journal as a process of data collection, and her lexical selection in these texts was replete with references to journals, notebooks, paper, and other methods by which she recorded her experience with mania. Likewise, she also employed the abbreviation *SPJ* (i.e., "see previous journal") as an anaphoric referent, cueing herself that the topic on which she was writing at any given time was one to which she had referred at some point in the past. This unique token is obviously not represented in the FROWN corpus, but accounts for 0.2% of LMN's entire body of work, again speaking to the preoccupations revealed by patterns of language use suggested by the diagnostic criteria for mania. Table 4.6 notes the frequency of occurrence of ten exemplars in this category as compared to the representation of the same words in FROWN, and in this category, the difference in total percentage of the words used reaches astronomical proportions, again reaching a level of $p < 0.0000$.

Table 4.6

Comparisons of keywords related to the writing process.

	Master Corpus		FROWN		
Keyword	frequency	% total	frequency	%total	Keyness
SPJ	602	0.2481	0	0.0000	2356.46
entry	436	0.1797	48	0.0032	1408.11
journal	176	0.0725	40	0.0027	493.85
record	164	0.0676	119	0.0080	292.83
write	166	0.0684	133	0.0090	279.21
paper	116	0.0478	139	0.0094	144.81
journals	47	0.0193	7	0.0004	144.36
creative	67	0.0276	41	0.0027	131.24
notebook	48	0.0197	15	0.0010	123.21
book	164	0.0676	318	0.0215	120.50
**Total %*		0.8183		0.0579 (σ_p 0.0049)	

**z* = 155.65, *p* <0.00000.

Cognitive processes. Second only to her observation of time and numbers, LMN devoted writing space to describing her thought processes and emerging ideas. Her use of the term *prime* as noted in Table 4.7 reflected her observation of the presence of prime numbers in the date and time stamps, and when these occurred, she frequently followed this notation with the word *smile*, to reflect her pleasure at such an observation. She also used *smile* to echo her satisfaction with the gestation of new

ideas. As noted in the description of the Master Corpus, LMN privileged the term *idea* in such a way that it was generally reserved for expression of a fulminating concept which was explored and expanded on the pages of her journals. She frequently evaluated these ideas, and patterns of collocation indicate word pairs such as "good idea", "previous idea", and "new idea" throughout her texts. The extent of her active cognitive processing of her experiences with mania is noted by her use of terms related to *thinking*, which as previously reported in Table 4.1, account for 0.36 % of her total word use. Again, because of familiarity with the context of LMN's repeated patterns of word use, the tokens *lithium* and *manic* were included in this categorization as the use of these was indicative of the thought processes in which she was engaged. Particularly during the period of July, 1985, she commented frequently on her diagnosis and her intake of lithium in attempt to balance and regulate her thinking processes.

The combination of these ten representative tokens and the frequency of occurrence of each in the master and FROWN corpora in Table 4.7 points to the prominence of her cognitive musings and related observations in her writing habits. The statistically significant difference between the total percentages for each corpus is illustrative of the salience assigned to these constructs as outlined, accounting for over one percent of her total words used, and again, to the linguistic behaviors consistent with her diagnosis.

Table 4.7

Comparisons of keywords related to cognitive processes.

	Master Corpus		FROWN		
Keyword	frequency	% total	frequency	% total	Keyness
smile	528	0.2176	93	0.0063	1570.47
idea	356	0.1467	206	0.0139	717.37
lithium	190	0.0783	2	0.0001	721.83
good	606	0.2498	788	0.0534	703.24
feel	345	0.1422	221	0.0150	660.20
yes	255	0.1051	195	0.0132	441.45
think	359	0.1479	587	0.0398	327.79
manic	84	0.0346	7	0.0004	281.43
thoughts	120	0.0494	62	0.0042	254.92
ideas	147	0.0605	116	0.0078	249.60
**Total %*		1.2321		0.1541 (σ_p 0.0080)	

**z* = 135.33, *p* <0.0000

Physiological perceptions. Since LMN was adopting the roles of both observer and participant in her data collection in the journal keeping process, she frequently noted reflections on her process, both of her mentation and perception of physiological phenomena. She described changes in her breathing patterns, and carefully documented the location of particular sensations occurring all over her body. Her lexical selection is remarkable then for frequent references to specific body parts

and specific perceptual and physiological experiences. Table 4.8 indicates the most frequently occurring words in the master corpus related to these observations, several of which deal specifically with unusual or atypical patterns (e.g., *shook, nausea*). The difference between proportional representations of this composite in the master corpus as compared to the typical patterns of American English in the FROWN corpus reaches extreme levels of statistical significance, again, speaking to the relationship between words chosen, and the diagnostic criteria identified.

Table 4.8

Comparisons of keywords related to physiological perceptions.

	Master Corpus		FROWN		
Keyword	frequency	% total	frequency	% total	Keyness
tired	355	0.1463	49	0.0033	1105.67
shook	237	0.0976	62	0.0042	641.05
breathing	175	0.0751	28	0.0019	530.40
woke	146	0.0601	10	0.0006	500.02
nausea	121	0.0498	1	0.0000	462.14
left [side]	349	0.1438	472	0.0320	389.91
arms	185	0.0762	104	0.0070	377.95
sleep	166	0.0684	81	0.0054	361.66
legs	152	0.0626	62	0.0042	356.03
breath	124	0.0511	62	0.0042	267.29

**Total %*	0.8310	0.0628 (σ_p 0.0051)

**z* = 151.42, *p* <0.0000

Family and home. In addition to documenting and conveying curiosity about her experience with mania and associated fascination with time and numbers, and the responsibilities of her career, LMN also expanded upon home and family life. Her journals note participation in routine activities such as going to the grocery store, rearranging the furniture in the house, and taking a shower. LMN described activities of her husband (e.g., traveling on a business trip) and her child (e.g., what time he caught the school bus in the morning), and not unexpectedly, the frequency of occurrence of the proper names of these family members (noted in tables in italics) was significantly different from that observed in FROWN. However as noted in Table 4.9, the proportional representation of all the entries in this semantic category suggests diagnostic significance, particularly in terms of her changes in neurovegetative status and the repeated preparation and consumption of coffee, both of which are non-linguistic hallmarks of mania. Interestingly, she notes such changes as symptomatic of being in an episode. The *z* value measuring the difference in the total number of terms used in this category indicates markedly significant differences in frequency of use.

Table 4.9

Comparisons of keywords related to family and home.

	Master Corpus		FROWN		
Keyword	frequency	% total	frequency	% total	Keyness
child	150	0.0618	3	0.0002	558.30
coffee	199	0.0820	71	0.0048	489.20
couch	104	0.0428	16	0.0010	317.55
shower	83	0.0342	14	0.0009	248.94
photos	62	0.0255	9	0.0006	191.33
husband	91	0.0375	49	0.0033	189.70
TV	112	0.0461	94	0.0063	182.87
clock	72	0.0296	24	0.0016	181.05
Christmas	73	0.0300	42	0.0028	147.45
milk	60	0.0247	23	0.0015	143.79
**Total %*		0.4142		0.0230 (σ_p 0.0031)	

Note. Keywords noted in italics are used to protect confidentiality.

**z* = 126.99, $p < 0.0000$

Career/research. As a person who had pursued advanced graduate education in her particular field of interest, and has since been devoted to participating in the research community, LMN consistently referred to work in her writing. This was most frequently expressed in the generation of ideas, as she used the journals as her medium for documenting her "brainstorms", frequently outlining graphs, sketching

diagrams, or calculating formulae. She would make notes and then comment on her thought process (e.g., "Wait a minute", "This might work"), or would "try out" several sequential variations on themes such as book titles, research proposals, or names of organizations. Predictably, terms specific to her area of research occurred at rates significantly different from the reference corpus, but even words common to the American English vernacular as measured by FROWN were more significantly represented in the master corpus. The *z* value noted at the bottom of Table 4.10 affirmed the tremendous magnitude of difference in the proportion of occurrence between LMN's words within the category and those of the FROWN corpus.

Table 4.10

Comparisons of keywords related to career and research.

	Master Corpus		FROWN		
Keyword	frequency	% total	frequency	% total	Keyness
data	368	0.1516	257	0.0174	671.83
computer	205	0.0845	103	0.0069	440.99
university	189	0.0779	0	0.0000	348.22
food	208	0.0857	208	0.0141	300.58
nutrition	92	0.0379	13	0.0008	285.29
nutrient	67	0.0276	0	0.0000	262.14
proposal	99	0.0408	36	0.0024	241.74
diet	87	0.0358	33	0.0022	209.29
authorware	47	0.0193	0	0.0000	183.88

lab	71	0.0292	23	0.0015	180.19
**Total %*		0.5903		0.0453 (σ_p 0.0042)	

Note. Keywords noted in italics are used to protect confidentiality.

**z* = 126.10, *p* <0.0000

The final analysis of this hypothesis involved gaining another perspective on the differences in word frequency between the Master Corpus and FROWN by means of a descriptive table of relative ranking. This was generated to demonstrate the order in which key content words and their lemma appeared in the Master Corpus, and the comparative rank of each from the FROWN corpus. The first column presents the token, followed by the rank order of the word, the frequency of occurrence, and the relative frequency as represented by the percentage of total in the Master Corpus. The next three columns present the information for the token as it appears in FROWN. For lemmatized tokens, an estimated rank in the FROWN corpus is noted in <angle brackets>, calculated based on combined frequencies of occurrences of all of the target word forms (e.g., *think, thinking, thought, thoughts*), and rank of words which appear with the same frequency in that corpus. For example, the combined frequency of the lemma for *think* = 1365, which yielded an estimated rank of <75>, which would have placed it between *these*, ranked at #74 with 1384 appearances, and *then* which occurred 1320 times and was ranked at #75 in FROWN. Statistical significance (*p*) is reported in the last column, indicating differences between proportional representations in the two corpora.

Table 4.11

Fifty top content words as ranked in the Master Corpus compared to ranking, frequency, and relative proportion in FROWN.

	Master Corpus (242,589 words)			FROWN (1,472,978 words)			
Word	rank	freq	rel freq (%)	rank	freq	rel freq (%)	*p*
#	1	39227	16.1701	1	233355	15.8423	0.0000443
prime	2	1049	0.43218	1517	75	0.00510	0.0000000
think [a]	3	876	0.36110	*<75>*	1365	0.00092	0.0000000
pm	4	621	0.25598	25743	2	0.00013	0.0000000
good	5	606	0.24980	113	788	0.05349	0.0000000
SPJ [b]	6	602	0.24815	--	0	0.00000	0.0000000
one	7	535	0.22053	35	3191	0.21663	0.667200[ns]
smile	8	528	0.21765	1224	93	0.00631	0.0000000
entry [a]	9	516	0.21270	*<2035>*	58	0.00393	0.0000000
Jan [b]	10	514	0.21188	*<1381>*	82	0.00556	0.0000000
now	11	504	0.20775	82	1171	0.07949	0.0000000
idea(s) [a]	12	503	0.20734	*<296>*	322	0.02186	0.0000000
feel [a]	13	460	0.18962	*<191>*	472	0.03204	0.0000000
LMN	14	401	0.16530	--	0	0.00000	0.0000000
time	15	378	0.15581	71	1453	0.09864	0.0000000
Nov [b]	16	370	0.15252	*<1029>*	109	0.00739	0.0000000

data	17	368	0.15169	378	257	0.01744	0.0000000
when	18	364	0.15004	45	2292	0.15560	0.47160 [ns]
tired	19	355	0.14633	2402	49	0.00332	0.0000000
left	20	349	0.14386	191	472	0.03204	0.0000000
know	21	315	0.12984	105	848	0.05757	0.0000000
right	22	309	0.12737	136	665	0.04446	0.0000000
lithium [a, b]	23	306	0.12613	*<16034>*	4	0.00027	0.0000000
since	24	289	0.11913	150	604	0.04100	0.0000000
work	25	266	0.10965	101	906	0.06150	0.0000000
July [b]	26	261	0.10758	1720	67	0.00454	0.0000000
model [a]	27	255	0.10511	*<213>*	424	0.02828	0.0000000
yes	28	255	0.10511	523	195	0.01323	0.0000000
first	29	250	0.10305	77	1289	0.08750	0.0201396
August [b]	30	237	0.09769	*<1315>*	86	0.00583	0.0000000
shook	31	237	0.09769	1907	62	0.00420	0.0000000
arm(s) [a]	32	235	0.09687	*<551>*	187	0.01269	0.0000000
before	33	226	0.09316	107	836	0.05675	0.0000000
decide [a]	34	224	0.09233	*<200>*	449	0.03048	0.0000000
child('s) [a]	35	221	0.09110	20186	3	0.00020	0.0000000
February [b]	36	219	0.09027	*<2418>*	48	0.03357	0.0000000
leg(s) [a]	37	216	0.08903	*<1112>*	101	0.00685	0.0000000
breath [a]	38	213	0.08780	*<932>*	121	0.00821	0.0000000
morning	39	213	0.08780	551	187	0.01269	0.0000000

two	40	213	0.08780	72	1445	0.09810	0.10520 [ns]
hard	41	209	0.08615	309	304	0.02063	0.0000000
food	42	208	0.08574	481	208	0.01412	0.0000000
minute [a, b]	43	207	0.08532	<*428*>	231	0.01568	0.0000000
computer	44	205	0.08450	1078	103	0.00699	0.0000000
count	45	200	0.08244	1427	79	0.00536	0.0000000
coffee	46	199	0.08203	1591	71	0.00482	0.0000000
new*	47	193	0.07955	73	1439	0.09769	0.0059464
journal	48	176	0.07255	2840	40	0.00271	0.0000000
deep [a]	49	174	0.07172	<*348*>	273	0.01853	0.0000000
oh	50	173	0.07131	737	146	0.00991	0.0000000

[ns] not significant.

***** Significant in terms of negative keyness.

[a] Count includes lemmatized versions of content word.

[b] Count includes abbreviated versions of same word.

The proportional representation of # (numerals in text) in LMN's writing is quite similar to that observed in the FROWN corpus. Although her rate of occurrence is less than one half of one percent higher than the reference corpus, it is her pattern of use that speaks to the remarkable differences. There are several points to consider: the FROWN corpus is compiled of writing from different genres, and hence, the use of numbers in text may vary based on the format and context; also, FROWN was compiled by a multitude of different writers, and the LMN corpus is the result of n =1.

Most importantly, however, is the observation of LMN's use of numbers in context. Analysis of collocational patterns of numbers was conducted, examining each digit from zero to nine using the concordance feature of *WordSmith Tools*. In the Master Corpus, tokens indicating something other than time, date, or quantity accounted for less than five percent of the collocates for digits 0-9. The other patterns that did occur were noted in context as lists, mathematical formulae, or proportional representations as fractions or percentages.

Additionally, the combination of # with other representations of temporal or quantity concepts accounts for nearly 20% of all of LMN's words. This is consistent with her stated fascination and preoccupation with prime numbers, as well as marking the passage of time (e.g., estimates of the number of seconds or minutes she experienced a particular phenomenon), observations of her mathematical interests related to her research, or the tendency to record other counting behaviors (e.g., number of grapes eaten for a snack). These manifestations were all consistent with the linguistic indicators of mania.

Notably, in Table 4.11 there are three tokens (i.e., *one, two, when*) for which there is no statistically significant difference in their proportional representations between the two corpora. This can be explained by the high frequency of quantity/temporal themes in LMN's writing, as well as in general writing across the genres included in FROWN. This semantic group of tokens represents a substantial proportion of LMN's content words, but for this small set, it is not necessarily different from the proportions observed in other written texts. Likewise, in this analysis, the token *new* was significantly different in terms of *negative* keyness; that

is, it appears proportionately more frequently in FROWN than it does in the Master Corpus, even though it is one of LMN's most frequently occurring words. Although these are unexpected findings, they do not disconfirm the hypothesis that the content of LMN's patterns of language use are different from those considered to be 'typical'.

Analysis 2 - Comparison of Master Corpus to language use in journal-as-genre.

The second comparative analysis addressed the question of journal-as-genre, and involved a reference corpus built from samples of journals from other writers. Hypotheses 2 states: There are measurable differences in the written texts of this participant as compared to corpora of 'normal' language use noted in journals of the same genre as produced by other individuals. The analyses performed with the reference corpora were consistent with the previous investigation, that is, keyword comparisons of the master corpus with the reference corpora as a batch. *WordSmith* keyness analysis requires that the larger of two corpora is used as the reference corpus, and the journal-as-genre corpus was approximately half the size of the master corpus; hence the master corpus becomes the point of reference. Parenthetically, this would not yield any difference in results when comparing the samples, but the layout of tables was modified so that the Master Corpus would retain the initial position of presentation and positive keyword connotations. Two tables are included in this section outlining the top 50 statistically significant words from each corpus as compared to the other, again using keyness as the statistic of comparison.

Three journals were compiled for use as a point of reference for analysis of language use in journal-keeping as a specific genre. Although the time spans

involved range from the late 19th century to the late 20th century, the genre-specific framework remained the same, as writers made chronological entries which reflected their current circumstance, and their observations of events, places, and personalities in their environment. Review of the top fifty negative keywords in Table 4.12, so noted because of their difference from the master corpus, speaks to the respective writers' situations, as Fletcher (2001) wrote of *buffalo, Indians,* and *Wajapa* (the proper name of a Native American acquaintance); Barnard (1918) used the terms *Paris, war, French,* and *German* repeatedly in his description of his life in Europe at the advent of the first World War; and Greger (1999) regularly incorporated the terms *medical, patient, doctor* in his narrative of his training as a physician. The first two columns indicate LMN's use of the same terms and the percentages represented in her writing. The last column presents the negative keyness statistic, again noting words which are rarer in the Master Corpus.

Table 4.12.

Comparisons of the top fifty negative key content words in Master Corpus and journal-as-genre corpus.

	Master Corpus (242,589 words)		Journal-as-genre Corpus (124,933 words)		
Keyword	Frequency	% of total	Frequency	% of total	Keyness*
Paris	1	0.000800	340	0.272145	-721.5034
Mr.	7	0.005603	254	0.203308	-489.8110
war	4	0.003201	220	0.176094	-438.2059

French	7	0.005603	201	0.160886	-378.5375
buffalo	0	0.000000	170	0.136072	-367.0146
German	2	0.001600	167	0.133671	-340.4718
Wajapa	0	0.000000	155	0.124066	-334.6187
medical	13	0.010405	190	0.152081	-324.3967
women	5	0.004002	168	0.134472	-321.5520
man	5	0.002061	142	0.113661	-267.0566
tent	0	0.000000	119	0.095251	-256.8782
came	44	0.018138	210	0.168090	-255.7525
patient	4	0.001648	130	0.104056	-247.9795
day	128	0.052764	299	0.239328	-230.3703
chip	4	0.001648	120	0.096051	-227.0167
little	61	0.025145	217	0.173693	-226.5959
doctor	2	0.000824	110	0.088047	-219.0402
American	14	0.005771	136	0.108858	-212.1539
France	2	0.000824	105	0.084045	-208.4289
Mrs.	19	0.007832	142	0.113661	-205.4471
over	83	0.034214	225	0.180097	-195.7182
Indians	0	0.000000	90	0.072039	-194.2638
five	14	0.005771	124	0.099253	-188.6985
one	535	0.220538	619	0.495466	-187.2640
camp	2	0.000824	89	0.071238	-174.5389
hospital	14	0.005771	114	0.091248	-169.3322

military	0	0.000000	78	0.062433	-168.3570
wife	3	0.001236	88	0.070437	-166.0634
appendix	1	0.000412	80	0.064034	-162.7286
horses	3	0.001236	84	0.067236	-157.7022
patients	13	0.005358	105	0.084045	-155.5818
ambassador	0	0.000000	68	0.054429	-146.7691
hundred	1	0.000412	71	0.056830	-143.5364
fire	3	0.001236	77	0.061633	-143.1028
medicine	4	0.001648	77	0.061633	-137.6554
white	29	0.011954	116	0.092849	-129.3606
government	0	0.000000	59	0.047225	-127.3410
wounded	0	0.000000	59	0.047225	-127.3410
horse	2	0.000824	66	0.052828	-126.0666
killed	0	0.000000	55	0.044023	-118.7065
told	40	0.016488	124	0.099253	-118.6670
general	20	0.008244	95	0.076040	-115.3970
British	1	0.000412	56	0.044824	-111.6272
soldiers	1	0.000412	56	0.044824	-111.6272
away	32	0.013191	108	0.086446	-109.1835
army	9	0.003709	73	0.058431	-108.2921
several	0	0.000000	50	0.040021	-107.9137
bear	1	0.000412	54	0.043223	-107.3821
hill	2	0.000824	57	0.045624	-107.2158

miles	6	0.002473	65	0.052027	-104.1468

*p< 0.0000000000 for all values of negative keyness.

From a different perspective, analysis of the words which appeared in the master corpus but not in the journal-as-genre corpus revealed a familiar pattern of word use, as indicated in Table 4.13. As before, the differences in word choice reflect idiosyncratic patterns of language use, based on the life experiences and the salient demands of the journal keeper. For example, although the writers of the three journals utilized as comparative examples used numbers in their texts, LMN's persistent attention to the date and time in her entries disproportionately outweighed the frequency of occurrence of this very common token. Similarly, her continuous observation of prime numbers, as well as the other repeated patterns of keyword use in her journals all occurred to a degree that was statistically significant (p <0.0000).

LMN's interests and preoccupations as recorded in her journals have been clearly delineated, and this table indicated that even though the genre was ostensibly the same, the content continues to be remarkably different. A finding of particular interest is the number of personal pronoun referents (e.g., *I'm, I've, I'll*) evidenced in the Master Corpus, and the significance with which these keywords exceed the proportions noted in the patterns of other journal writers. LMN is clearly observing the conventions of keeping a journal, but based on the comparisons of content, her notations tend to reflect a more egocentric perspective, rather than the level of commentary on the external environment as reflected in the genre-based reference corpus.

Table 4.13.

Comparisons of top fifty positive key content words from the Master Corpus and journal-as-genre corpus.

	Master Corpus (242,589 words)		Journal-as-genre Corpus (124,933 words)		
Keyword	Frequency	% of total	Frequency	% of total	Keyness
#	39227	16.170146	1051	0.841250	27260.1934
prime	1049	0.432418	3	0.002401	838.3918
smile	528	0.217652	16	0.012806	329.1841
data	368	0.151696	4	0.003201	270.3414
idea	356	0.146750	14	0.011206	206.9909
tired	355	0.146338	23	0.018409	171.3699
July	245	0.100993	6	0.004802	159.9177
shook	237	0.097696	7	0.005603	148.5677
need	339	0.139742	32	0.025613	132.8281
breathing	175	0.072138	3	0.002401	121.4610
I’m	475	0.195804	75	0.060032	118.5090
feel	345	0.142215	41	0.032817	113.8653
good	606	0.249805	124	0.099253	105.9863
something	328	0.135208	41	0.032817	103.6388
could	633	0.260935	136	0.108858	101.9831
record	164	0.067604	6	0.004802	97.3224
count	200	0.082443	13	0.010405	96.3679

room	389	0.160353	63	0.050427	94.1779
since	289	0.119131	35	0.028015	93.8556
what	732	0.301744	180	0.144077	90.7741
both	278	0.114597	34	0.027214	89.5217
program	127	0.052351	5	0.004002	73.7828
legs	152	0.062657	10	0.008004	72.8209
coffee	199	0.082031	23	0.018409	67.1819
thoughts	120	0.049466	6	0.004802	64.4200
write	166	0.068428	17	0.013607	61.4658
think	359	0.147986	75	0.060032	60.6283
color	129	0.053176	9	0.007203	60.0761
going	436	0.179727	102	0.081643	59.9007
yes	255	0.105116	43	0.034418	58.7370
arms	185	0.076260	23	0.018409	58.7137
can't	228	0.093986	36	0.028815	56.8463
thought	255	0.105116	46	0.036819	53.7674
brain	105	0.043283	6	0.004802	53.5144
I’ve	182	0.075024	25	0.020010	52.6427
thinking	142	0.058535	15	0.012006	51.4009
research	114	0.046993	9	0.007203	49.7556
maybe	162	0.066779	21	0.016809	49.5123
shower	83	0.034214	3	0.002401	49.4126
bit	122	0.050290	11	0.008804	49.2154

finished	112	0.046168	9	0.007203	48.3999
nutrition	92	0.037924	5	0.004002	47.8453
don't	358	0.147574	85	0.068036	47.7294
mode	85	0.035038	4	0.003201	46.6246
I'll	142	0.058535	17	0.013607	46.5531
problem	109	0.044932	9	0.007203	46.3765
now	504	0.207758	141	0.112860	45.6660
journal	176	0.072550	27	0.021611	45.3375
book	164	0.067604	25	0.020010	42.5445
type	88	0.036275	6	0.004802	41.4421

*p< 0.0000000000 for all values of positive keyness.

This analysis yielded information about the differences between LMN's journal-keeping habits and those of people who apparently have no diagnosis of a mental illness, but who were dealing with their own set of life challenges. Because of the incongruent nature of the personal circumstances of the writers of the four sets of journals under investigation (including LMN's), there were few consistent content words. Returning to the example of her use of numerals in text, noted with the symbol #, the journal-as-genre corpus was remarkable for utilizing that particular token at a proportion representing only 1/20th of LMN's total use. The significantly disproportionate use of numerals in her text in comparison to the other writers, particularly in combination with other words relating to quantity and time (e.g., *prime, July, count*), speaks to LMN's patterns of preoccupation consistent with her diagnosis,

as does the prevalence of other tokens which hold a place of privilege in her writing (e.g., *smile, idea, breathing, good*).

As was the case with testing the previous hypothesis, the final piece of this analysis involved differences in word frequency between the Master Corpus and the journal-as-genre corpus as addressed by relative ranking and proportional representation. Based on the key content words and their lemma in the Master Corpus, the same words were comparatively ranked as they appeared in the genre-based corpus. Again, the first column presents the token, followed by the rank order of the word, the frequency of occurrence, and the relative frequency as represented by the percentage of total in the Master Corpus. The next three columns present the information for the token as it appears in the journal-as-genre corpus, with estimated rank of lemmatized tokens appearing in <angle brackets>. Statistical significance (*p*) is reported in the last column, indicating differences in proportional representations between the two corpora.

Table 4.14.

Fifty top content words as ranked in the Master Corpus compared to ranking, frequency, and relative proportion in the journal-as-genre corpus.

	Master Corpus (242,589 words)			Journal-as-Genre (124,933 words)			
Word	rank	freq	rel freq (%)	rank	freq	rel freq (%)	*p*
#	1	39227	16.1701	9	1051	0.84125	0.0000000

prime	2	1049	0.43218	4097	3	0.00240	0.0000000
think [a]	3	876	0.36110	*<108>*	142	0.11366	0.0000000
pm	4	621	0.25598	--	0	0.00000	0.0000000
good	5	606	0.24980	115	124	0.09925	0.0000000
SPJ [b]	6	602	0.24815	--	0	0.00000	0.0000000
one	7	535	0.22053	21	619	0.49546	0.0000000
smile	8	528	0.21765	903	16	0.01280	0.0000000
entry[a]	9	516	0.21270	*<3734>*	3	0.00240	0.0000000
January [b]	10	514	0.21188	*<5421>*	2	0.00160	0.0000000
now	11	504	0.20775	108	141	0.11286	0.0000000
idea(s) [a]	12	503	0.20734	1016	14	0.01120	0.0000000
feel [a]	13	460	0.18962	*<187>*	78	0.06243	0.0000000
LMN	14	401	0.16530	--	0	0.00000	0.0000000
time	15	378	0.15581	84	180	0.14407	0.12860 [ns]
Nov [b]	16	370	0.15252	*<5654>*	2	0.00160	0.0000000
data	17	368	0.15169	2926	4	0.00320	0.0000000
when	18	364	0.15004	42	343	0.27454	0.0000000
tired	19	355	0.14633	646	23	0.01840	0.0000000
left	20	349	0.14386	122	120	0.09605	0.0000000
know	21	315	0.12984	165	89	0.07123	0.0000000
right	22	309	0.12737	191	76	0.06083	0.0000000
lithium [a, b]	23	306	0.12613	--	0	0.00000	0.0000000
since	24	289	0.11913	424	35	0.02801	0.0000000

work	25	266	0.10965	177	83	0.06643	0.0000000
July [b]	26	261	0.10758	2182	6	0.00480	0.0000000
model [a]	27	255	0.10511	*<3976>*	3	0.00240	0.0000000
yes	28	255	0.10511	333	43	0.03441	0.0000000
first	29	250	0.10305	99	153	0.12246	0.0064000
August [b]	30	237	0.09769	*<281>*	49	0.03922	0.0000000
shook	31	237	0.09769	1969	7	0.00560	0.0000000
arm(s) [a]	32	235	0.09687	*<456>*	32	0.02561	0.0000000
before	33	226	0.09316	152	97	0.07764	0.0060000
decide [a]	34	224	0.09233	*<408>*	36	0.02881	0.0000000
child('s) [a]	35	221	0.09110	--	0	0.00000	0.0000000
February [b]	36	219	0.09027	*<3772>*	3	0.00240	0.0000000
leg(s) [a]	37	216	0.08903	*<1022>*	14	0.01120	0.0000000
breath [a]	38	213	0.08780	*<677>*	21	0.01680	0.0000000
morning	39	213	0.08780	138	109	0.08724	0.92820 [ns]
two	40	213	0.08780	70	216	0.17289	0.0000000
hard	41	209	0.08615	338	42	0.03361	0.0000000
food	42	208	0.08574	360	40	0.03201	0.0000000
minute [a, b]	43	207	0.08532	*<362>*	40	0.03201	0.0000000
computer	44	205	0.08450	--	0	0.00000	0.0000000
count	45	200	0.08244	1063	13	0.01040	0.0000000
coffee	46	199	0.08203	625	23	0.01840	0.0000000
new	47	193	0.07955	119	121	0.09685	0.0062000

journal	48	176	0.07255	536	27	0.02161	0.0000000
deep [a]	49	174	0.07172	*<444>*	33	0.02641	0.0000000
oh	50	173	0.07131	282	49	0.03922	0.0000000

[ns] not significant.

[a] Count includes lemmatized versions of content word.

[b] Count includes abbreviated versions of same word.

Analysis of these patterns of frequencies and proportions indicates that two of the tokens in this sample, *time* and *morning*, are not significantly different in their representations between the Master Corpus and the journal-as-genre corpus. While there is clearly different content based on the personal reflective nature of the journal genre, these similarities speak to the consistent observation of the conventions by all writers in this comparison. As Fulwiler (1987) observed, 'good' journal keeping habits involve production of conversational speech in written form using colloquial diction, informal punctuation, and first person references, with observations, questions and speculation about the writer's experiences and increasing self-awareness, in the context of frequent and long entries with chronological ordering. The tokens *time* and *morning* represent a consistent acknowledgement of the time/quantity content inherent in the chronological themes (process) of the genre, and the consistency with which these are noted in these two corpora speaks LMN's adherence to the genre register. These observations suggest that in terms of qualitative stylistic production in journal-keeping, LMN's corpus is not substantially different. However, from the perspective of specific word use and patterns of lexical

selection, noted in these analyses particularly with the use of #, the differences in content are remarkable and this hypothesis is also confirmed.

The Master Corpus and Mania

Analysis of the Master Corpus of LMN's writing has demonstrated significant differences in patterns of use when compared to two external reference corpora. This investigation has also clearly delineated behaviors which are consistent with the DSM-IV criteria for the linguistic processes observed in mania.

Grandiose ideas. One of the pathognomic indicators of mania is the tendency toward grandiose idea generation, and the work produced by LMN is no exception. This is noted by her relatively disproportionate use of the term *idea*, which in its lemmatized form represents 0.2 percent of her total words (503 occurrences), as compared to the more typical rate of use as suggested by the FROWN corpus (0.02 %). LMN's use of this token represents a higher rate than even her expression of such common function words as *are* (0.19% in the Master Corpus; 0.30% in FROWN) or *an* (0.16% in the Master Corpus; 0.20 % in FROWN). There are a number of words related to *idea*, as retrieved from two on-line thesauri (Merriam-Webster, 2006; Roget, 2006), including *thought, picture, feeling, plan, image,* and *guess*. Of these, *thought* was the only one to reach the level of significance of occurring at least once per 1000 words written. When combined with lemma including *thoughts, think, and thinking*, this semantic construct accounted for 0.36 percent (876 occurrences) of the total words in the master corpus, speaking directly to

the introspective purpose of keeping the journals. Importantly however, LMN notes that

> "…for me *idea* was very specific; it recognized (recorded) that I had thought of something new-to-me, something I personally had not thought of before. So, *idea* categorized the result of a thought process as being unique to me. The idea did not have to be earth-shattering or monumentous [sic], just something I had not thought of before…When I used *thought*, I probably meant I was thinking about something, recalling something, pondering something. That is I was processing information, not synthesizing something new-to-me." (LMN, personal communication, May 20, 2006).

Collocates for *idea* were analyzed, and typically occurred in patterns relative to LMN's interpretation of the quality or feasibility of the thought. She goes on to note

> "I did evaluate the idea, indicating what I thought of it, using phrases such as "this is an excellent idea", "this is an okay idea", "this is an outstanding idea" (LMN, p.c.).

Such patterns of collocation were measured in a span of five words to the left and right (5:5) of the node IDEA. Not surprisingly, <good> occurred immediately to the left (position L1) of IDEA 124 times, eighteen of those being preceded by <very> in position L2 (i.e., "very good idea"). Parenthetically, LMN used the term *good* 607 times in the Master Corpus, over twenty percent of which were in the context of a qualitative statement in collocation with *idea*. Other frequent collocates noted in the L1 + *idea* position as represented in both the Master Corpus and FROWN are noted in Table 4.15.

Table 4.15

Frequency of occurrence for selected collocates of IDEA in Master Corpus and FROWN.

	Master Corpus	FROWN
good idea	124	9
very good idea	18	0
new idea	15	0
excellent idea	7	1
brilliant idea	4	0
previous idea	4	0

The striking frequency of this recurrent grouping of words in the Master Corpus as compared to FROWN demonstrates a pattern of fixation on *idea* in LMN's language use. The frequent collocations with positive discourse prosody qualify this in the direction of grandiose thinking as described in the DSM-IV. However, "grandiose" may not be the best descriptive qualifier for the idea generation seen in mania, as LMN's ideas ranged from the profound to the mundane. A more accurate descriptive criterion would perhaps refer instead to the *quantity* of ideas generated, rather than the *quality*, as this was clearly the pattern noted in the Master Corpus.

Idiosyncratic thinking. In the DSM-IV (1994), one of the indicators of mania is production of speech which is difficult to interpret. Consistent with this construct, another frequently occurring term in LMN's lexicon is the use of *prime*. This term represents a particular fixation of hers, frequently causing her to be distracted from

the topic of her writing. She uses this term exclusively in reference to prime numbers, and usually in the context of observing the occurrence of such a number in the recording of the hour:minute:second time stamp at the beginning or end of an entry. Her pattern is never to end a thought on a prime number, so she will continue recording the time until it gets to a non-prime designation. For LMN, the observation of prime numbers holds special significance; that is, recording primes is a qualitative acknowledgement of the intrinsic simplicity and austere beauty of the numbers. Prime numbers are those which are indivisible by anything except themselves and the number one, and symbolically represent for LMN an idea that is indivisible by other concepts that can only be appreciated when reduced to its very uncluttered essence. In the master corpus, the node PRIME collocated with either numerals indicating date and/or time, words referring to time concepts (i.e., <July, zero, a.m., minute>), or the word <smile>, used to indicate the pleasure at having experienced a prime number; some combination of which occurred in over 95% of occurrences. Related to this pattern, the node SMILE (85 occurrences) collocated with <prime> 46% of the time, again, speaking to the evident gratification of the experience.

Her preoccupation with time is consistent with this observation of prime numbers, as across the course of the years, she has moved from recording only hours and minutes in a twelve hour format (i.e., noting a.m. and p.m.), to keeping atomic clocks and watches within reach, recording hours:minute:seconds in a 24-hour format. As she has been noted to record up to three entries within the space of a single minute, early in her process she noted that:

"The recording of seconds became important because when ideas are

rapidly evolving, the minutes don't separate." 9:13:09 a.m. 07 Aug, 1983. She has developed a coding system for noting repeated occurrences of prime numbers, finding particular delight in noting these within a single entry (e.g., 11:07:13). In recent years, this has become more elaborate and a different set of collocated tokens has emerged (e.g., *twin primes*).

Relative to the concept of language use which is difficult for the reader to interpret, corpus linguistic analysis identifies recurrent patterns which may rise to the level of topical fixation or perseveration, if no other more parsimonious interpretation can be made for such high rates of occurrence of particular content words. In the Master Corpus, the repeated notation of *prime* serves as an exemplary model of idiosyncratic thinking, as the meaning of this token holds unique significance for LMN, based on the proportional percentage of occurrence and the lack of motivation based on any other relevant context (e.g., *prime time, prime minister,* or *prime rate* as observed in FROWN).

Expansive affect and related mentation. Since "repeated events are significant" (Stubbs, 2001, p. 221)**,** discourse prosodies were investigated based on the collocations of the content words in a given context, allowing for the attitudes and emotional content of the context to be addressed. LMN made frequent use of the phrase, "It's three o'clock in the morning!", which for her, represented a positive observation, consistent with the mirthful mood often noted in mania (cf. Fenichel, 1945). Notations of this time in LMN's journals were typically related to her increased productivity and creative impulses, increased animation and energy, and decreased need for sleep, and usually finished with an exclamation point at the end of

the sentence. Repeated use of this comment (75 occurrences) speaks to her continued observation and documentation of the passage of time. There were no characteristically negative contextual connotations noted to indicate less than favorable circumstances at that hour, such as insomnia or loneliness. Neither did she complain about unusual physical discomfort or ruminative agitation, describing her experiences more from the perspective of observer rather than sufferer.

The DSM-IV (1994) offers less specific criteria in this area, instead allowing for more interpretive flexibility of this diagnostic marker. This construct is understood to include behaviors such as increased animation and energy, and decreased need for sleep, all of which may be mediated linguistically, certainly in the narrative format of journal-keeping. As noted, the positive discourse prosodies and neurovegetative changes reported in the Master Corpus are consistent with expansive affect, but these may be more appropriately classified in terms of separate cognitive, linguistic, and behavioral processes, rather than falling under one nebulously defined umbrella criteria.

Flight of ideas. Another of the linguistic markers of mania described in the DSM-IV relates to the frequency with which an individual shifts topical referent and demonstrates a tendency to be distracted by both internal and external stimuli. Among other ways, the flight of idea marker was manifested in LMN's writing by her repeated use of the phrase "can't retrieve". The occurrence of *idea* has been discussed, but in the context of documenting ideas as they occurred to her, LMN sometimes lost her train of thought before she was able to commit it to paper. Use of this phrase was noted on 51 occasions in the master corpus. In addition to the loss of

train of thought, flight of ideas was observed particularly relative to the number of entries within a given time period. Although this varied widely across the corpus, during episodes of acute exacerbation, she was noted to demonstrate an increased rate of production of up to three entries per minute, during which focus on the topic about which she was writing would be interrupted secondary to sudden shifts of attention to time notations. For example, from the journal dated July 12, 1985:

11:39:50 Pita bread - whole wheat. Excellent!

11:40:15 Other food was grapes.

11:40:47 Nothing else [loke] looked good

11:41:20 Stopped pita bread mid-bite

11:42:57 Took Lithium. Drank Buttermilk with it. Not as good as when need it.

11:43:29 Going to eat rest of pita bread.

Although this is a venerated construct in the DSM-IV, the meaning of which is relatively well-defined and agreed-upon, the actual meaning of *flight of ideas* involves several processes, none of which are adequately delineated. Rapid shifting of conversational topic appears to be the more pathognomic indicator of mania, but that particular linguistic behavior is fundamentally influenced by a failure to filter distractions from both internal and external sources. This breakdown in attention to a specific topic results in the separation of discourse into unlinked parts, rather than allowing for production or appreciation of a longer, more coherent unit.

Associational chaining. The DSM-IV describes selection of linguistic tokens governed by sounds rather than content as a pattern of language use typical of persons with mania, in which words are selected based on their alliterative quality, rather than their relevance to the topic at hand. While this is the type of behavior often attributed to increased creativity and poetic writing, it is recognized as pathological in clinical contexts. Although this is not one of her more overt expressions of mania, the journals written during LMN's most acute episodes do reflect instances of this pattern, noted in use of phrases such as "sneezes and breezes", "Oh no. Oh my.", "Quantum jump, quantum rate, jump start", repeated references to colors, or as in this series of entries:

11:11:12 PRIME. But I don't yet know it

KP - clearing of air ways

Breezes

Breezes

SEA BREEZES

SALT Regulator

Breezes

Breezes

Breezes

Breezes

WHERE

Purple Blue Yellow

11:13:16 Nearly prime

No regulation

Membranes KP

Left Breezes Breezes Breezes

Breezes - Gale

Storm

eyes

Storm

Storm

toronado [sic]

Head forward

Ca++

Breezes, Breezes Breezes.

So Storm in cellular seawater causes (breezes) (14 July, 1985)

Although the DSM-IV is more specific in the description of this particular linguistic behavior, it does not provide clear guidance for delineation of the point at which this behavior rises to the level of diagnostic relevance. Again, such repetitive phonological linking may be indicative of poetic creativity (e.g., Bob Dylan), but as with all the other diagnostic markers, it is only in the observation of use in functional context that such determinations can be made. The advantage of corpus linguistic analysis in the investigation of associational chaining behavior is from the perspective of large quantities of text which are available for review. In LMN's case, it was during the course of her more severe exacerbations that such patterns emerged, as

there were considerably fewer notations relative to such formulations noted throughout the rest of her twenty eight year span of writing.

Other references to non-linguistic manifestations of mania

In addition to the linguistic markers of mania observed in LMN's writing, she also documented other manifestations in what has become the self-reporting mechanism of her symptomatology. *Psychomotor agitation* is described as excessive motor activity that is non-productive and repetitious (DSM, 1994), and LMN reports such movements as shaking of the extremities or 'dancing' during her episodes; for example, within a span of about twenty minutes in the early morning hours of December 10, 1994:

> 1:01:24 Both arms & legs shook very hard, but small 'arc' to count of 200. Breathing quite <XXX>. 1:02:18
>
> 1:08:59 Jump-start - both arms & legs shook very hard to count of 120. Breathing hard. 1:09:32
>
> 1:20:28 'Dance' - no, 'mime' - quit, move in random fashion, exagar.[exaggerated] motion. Don't know about time, probably approximately 3 min. 1:22:07

Her use of the descriptive term *shook* to describe her perception of physiological irregularities is remarkable for the frequency of occurrence, as it is noted 237 times (0.097% of total). Similarly, LMN reports auditory phenomena which represent perceptual changes which may be either hyperacusia, illusions, or hallucinations in

which she reports hearing crickets (42 occurrences) or hearing a radio, although none is playing (29 occurrences).

Summary

The analyses of the content of the Master Corpus, and subsequent comparison to external reference corpora as benchmarks for typical language use indicated distinct patterns of linguistic variation which were consistent with the diagnostic criteria for mania, according to the DSM-IV. LMN demonstrated recurrent patterns relative to specific themes through which she explored and recorded her experience with mania. The FROWN corpus provided a synchronous point of reference for analysis of the Master Corpus, allowing for specific conclusions to be drawn relative to the differences in word frequency observed in each. Likewise, the journal-as-genre corpus comparison yielded data that was significant not only for variation in content, but for also for thematic differences (e.g., writing about internal preoccupations, rather than observing circumstances of the environment), although LMN adhered to basic genre-specific conventions of noting chronological entries and using the first person perspective in her journals.

Examples of efficacious use of linguistic information for diagnostic purposes were extracted from the Master Corpus, the interpretation of which requires consideration of the balance between *process* and *content* in the context of functional language use along a continuum from typical to pathological. The DSM-IV criteria provide a very general framework within which specific contextual references may be placed. Regarding idea generation, the process and content markers may be

equivalent, but in identification of other criteria, the case is made for clarifying particular diagnostic points of reference, parsing out specific behaviors to be measured, and applying corpus linguistic methodology to further elucidate the process.

Chapter 5 – Analysis of Intra-individual Differences

This chapter describes the results of comparative analyses between the Master Corpus and internal reference corpora; that is, using LMN as her own control, various manifestations of her patterns of language use were investigated to address intra-individual differences in her experience with mania. The hypotheses being tested provide evidence for an idiographic perspective into her process, as certain texts are being established as benchmarks against which other texts can be compared, all of which were produced by the same person.

Analysis 3 - Comparison of Language Use during Medicated versus Unmedicated periods

Because these journals were written over an extended number of years, there is inherent variation in the content, based on the variation inherent to the passage of time. One of the most significant distinguishing variables in this collection is the introduction of pharmacopoeia seven years after the writing process began. As noted, LMN began experiencing episodes of mania in early 1978, but was not diagnosed until early in 1985 after experiencing an acute exacerbation that led to hospitalization. For the first seven years of her experience with mania, she was unmedicated, with no pharmacological intervention until starting lithium therapy in February of 1985. She continued taking lithium and continued writing in her journals for nearly nineteen years, but in late 2003 she decided to discontinue the medication because she was

unable to discern the benefits, as she continued to feel as if she was experiencing episodes of mania two to three times a year, even while medicated. Hypotheses 3 states: there are significant within-subject differences in patterns of written language use as measured during unmedicated versus medicated period of time. The analysis of this hypothesis takes on a variation of the classic A-B-A research design as the master corpus was divided into time segments based on administration of medication, as noted in Table 5.1.

Table 5.1.

Number of words per sub-corpus comparing medicated versus unmedicated language use.

Time span	Status	Word Count
Feb, 1978 to Jan, 1984 (71 months)	Unmedicated	74,782
Feb, 1985 to Dec, 2003 (226 months)	Medicated	167,807
Jan, 2004 to June, 2005 (17 months)	Unmedicated	29,845

This diachronic analysis of intra-individual variation (Molenaar & Valsiner, 2005) allows for examination of changes in patterns of language use based on differences associated with the effects of the medication. Therefore, the relevant question and hypothesis being tested is this: are there differences in LMN's patterns of language use when she is medicated, as compared to when she is unmedicated? The original intent of this hypothesis was to address conditions in an A-B-A design to

study the effects of medication on language behavior. The *A* conditions (unmedicated) were separated by a period of nearly nineteen years during which LMN was taking medication (*B*). The impact of the lengthy medicated period had to be considered, as potential differences in these periods reflect diachronic variation as well as the effects of medication.

As noted in Table 5.2, comparison of the early unmedicated condition A_1 with medicated condition B indicated that there were profound differences in the two samples, with substantially more tokens devoted to the topics involving *career and research* during periods when LMN was not taking the prescribed mood-stabilizing medication. After medication was implemented, her primary focus clearly shifted, and commentary in her journals became centered largely on *family and home*.

Table 5.2

A_1 versus B: Proportions of words by condition and semantic category reported as percentages of total.

	Time/quantity	Career/research	Physiological perception	Home/family
A_1 Unmedicated Total = 74,782	16.772	0.5990	0.2366	0.0855
B Medicated	17.9569	0.2347	0.6549	0.4797

Total =	
167,807	

$\chi^2 = 582.524$, $df = 3$, $p < 0.0000000$

Comparison of the late unmedicated A_2 period with the B medicated condition also indicated differences, although in a different direction, as noted in Table 5.3. LMN's interests appeared to be remarkably different in her retirement (A_2) when compared with A_1, coinciding with the beginning of her career twenty five years earlier. Thus, the analysis indicated that there was significantly less interest in *career and research* during the second unmedicated period than that noted during the medicated time span, reflecting her shift in interest to post-retirement (i.e., non-career) activities. Notably, her interest in time and quantity concepts remained consistently high across all conditions.

An acknowledged confound of the A_2 corpus deals with the observer's paradox, because with the introduction of external investigators, there were observable differences noted in LMN's language use, particularly in the style with which she wrote. Prior to 2004, the journals were primarily written as stream of consciousness, but subsequent to that time, LMN became more conscious of the fact that her words were being analyzed, and her format shifted, as she began to write in more complete sentences, and frequently took an epistolary register (e.g., beginning entries with "Dear Bess…"). As noted, this external influence will be addressed in more detail in Appendix D. The corpus compiled from texts generated during her late

phase with no medication (i.e., 2004-05) includes journals which are not included in the Master Corpus because of this influence.

Table 5.3.

A_2 versus B: Proportions of words by condition and semantic category reported as percentages of total.

	Time/quantity	Career/research	Physiological perception	Home/family
A_2 Unmedicated Total = 29,845	17.77	0.07	0.7069	0.4355
B Medicated Total = 167,807	17.9569	0.2347	0.6549	0.4797

$\chi^2 = 33.573$, $df = 3$, $p < 0.0000$

Because of the differences in the language patterning in A_1 and A_2, the A-B-A research design was abandoned in favor of the simpler *unmedicated versus medicated* formulation. Although the two sub-corpora from the unmedicated conditions are different in content as indicated in the preceding tables, they are also both different from the medicated condition. These two samples were combined to represent a

single unmedicated condition, allowing for a larger sample size as recommended by Scott (2006) and more efficacious comparisons according to corpus linguistic methodology. The inclusion of both segments provided a means by which extreme outliers represented in either corpus would be somewhat mitigated (i.e., increasing the word count would ostensibly produce a more representative distribution of what would be considered 'average' language use).

Table 5.4 notes the top fifty content words ranked in order of positive keyness; that is, these words appear more frequently in the unmedicated corpus, and are considered to be uncommon in the medicated corpus. The semantic content of this list is notable for specific references to LMN's research methodology and equipment from the early unmedicated time period (e.g., observations of *wt* [weight], *ROG* [rate of gain], *biological, data, variables, study, CompuJournal, recording, quantity, absolute*). In addition to providing a means for recording her experiences with mania, the early journals also served as lab notebooks in which germinating ideas were explored and framed prior to implementation in her work. Consistent with the diagnosis of mania, there were idiosyncratic patterns of language use observed which would be difficult for an outside reader to interpret. Two of LMN's individually-defined identifiers were noted in this list: *FKS* (also noted as *FK*) which stands for "function key side" and *CKS* for "control key side", both of which were used as notations of spatial orientation to situate perceptual sensations. These markers were employed briefly in the mid-1980's in LMN's journals, and were eventually abandoned for the more conventional use of *right* and *left* to note the same spatial orientation information.

This list is also remarkable for referents specific to the late unmedicated period during which LMN used particular phrases to describe her perception of physiological changes during an episode, noted as *pc* and "*puppet* on a string". She also referred repeatedly to her pet by name in her later writing, as well as to a particular piece of music with the term *Tibet* in the title, hence the appearance of those unusual tokens in this set. Not unexpectedly, use of certain terms related to time (i.e., *Mar, Aug, Jan, February*) were noted as prevalent in the unmedicated corpus; perhaps the more surprising implication is that these were used less frequently during the medicated period.

Table 5.4.

Top fifty positive key content words from unmedicated corpus as compared to medicated corpus.

	Unmedicated (104,627 words)		Medicated (167,807 words)		
Keyword	frequency	% total	frequency	% total	Keyness*
pm	627	0.599271	190	0.113225	499.5419
Mar	245	0.234165	73	0.043502	197.2698
pc	140	0.133808	17	0.010130	176.8609
data	253	0.241811	119	0.070914	133.4277
KWT	51	0.048744	0	0.000000	97.6287
study	89	0.085064	33	0.019665	59.9206
ROG	35	0.033452	1	0.000595	58.8267

biologists	29	0.027717	0	0.000000	55.5106
mast	29	0.027717	0	0.000000	55.5106
proc	28	0.026761	0	0.000000	53.5963
mixture	34	0.032496	3	0.001787	47.1654
biological	63	0.060213	22	0.013110	44.7073
pet's name [a]	30	0.028673	2	0.001191	44.4002
dice	32	0.030584	3	0.001787	43.6849
Aug	132	0.126162	86	0.051249	43.5984
FK	22	0.021027	0	0.000000	42.1106
pillows	31	0.029629	3	0.001787	41.9525
plot	31	0.029629	3	0.001787	41.9525
response	48	0.045877	13	0.007746	41.2773
affective	21	0.020071	0	0.000000	40.1963
levels	32	0.030584	4	0.002383	40.0135
protein	59	0.056390	22	0.013110	39.5135
twin	24	0.022938	1	0.000595	38.5109
CKS	20	0.019115	0	0.000000	38.2821
FKS	20	0.019115	0	0.000000	38.2821
propped	19	0.018159	0	0.000000	36.3679
basis	25	0.023894	2	0.001191	35.5325
Jan	265	0.253280	253	0.150768	34.6510
puppet	18	0.017203	0	0.000000	34.4537
line	70	0.066904	35	0.020857	34.2456

friend	24	0.022938	2	0.001191	33.7752
base	35	0.033452	8	0.004767	33.4304
CompuJournal	26	0.024850	3	0.001787	33.3841
February	26	0.024850	3	0.001787	33.3841
numbers	36	0.034407	9	0.005363	32.5953
comparable	17	0.016248	0	0.000000	32.5395
scientist	17	0.016248	0	0.000000	32.5395
wavy	17	0.016248	0	0.000000	32.5395
variables	37	0.035363	10	0.005959	31.8605
gain	24	0.022938	3	0.001787	30.0091
kcal	24	0.022938	3	0.001787	30.0091
composition	31	0.029629	7	0.004171	29.8153
fat	38	0.036319	12	0.007151	29.2590
Tibet	15	0.014336	0	0.000000	28.7111
wt	23	0.021982	3	0.001787	28.3352
immobile	26	0.024850	5	0.002979	27.2208
recording	35	0.033452	11	0.006555	27.0481
quantity	14	0.013380	0	0.000000	26.7970
level	46	0.043965	21	0.012514	25.0784
absolute	13	0.012425	0	0.000000	24.8828

*p< 0.00000 for all positive keywords.

[a] Tokens noted in italics are used to protect confidentiality.

Analysis of the words which were more frequently occurring in the medicated corpus (and more rare in the unmedicated corpus) is provided in Table 5.5. The medicated corpus represents a substantially longer period of time, and is therefore a larger corpus, but the keyness statistic is based on proportional representation, noted as *% total* columns in this table. The patterns appear to take on the now familiar groupings of particular interest and idiosyncratic use for LMN, including prevalence of *prime, lithium, smile,* and # represented as occurring more consistently during the medicated segment. Her use of # in this analysis is interesting because of the near symmetry in proportions (a difference of only 1.07%), and observational review of the journals suggests that her utilization was fairly consistent with notations of date and time stamps. However, this habit was less ingrained during the earliest stages of her experience, and has been refined and expanded as time has passed, resulting in elaborated notations of hour:minute:second and more frequent use of numbers in text. The difference in frequency of occurrence for # between unmedicated and medicated conditions reached a level of statistical significance, as that token appeared more consistently during the medicated condition. An advantageous framework for interpreting this table may be that use of these familiar terms as noted by negative keyness actually emerged late during the medicated period. These terms are very likely still represented in the late unmedicated segment, but the influence of combining the late unmedicated period with the diachronically dissimilar early medicated segment served to average out or moderate the variation, and hence the terms appear more frequently in the medicated period.

Table 5.5.

Top fifty negative key content words from unmedicated corpus as compared to medicated corpus.

	Unmedicated (104,627 words)		Medicated (167,807 words)		
Keyword	frequency	% total	frequency	% total	Keyness*
prime	270	0.258059	1018	0.606649	-181.4930
July	8	0.007646	239	0.142425	-176.4420
lithium	3	0.002867	189	0.112629	-158.0854
count	5	0.004778	199	0.118588	-155.5538
arms	3	0.002867	184	0.109649	-153.3950
room	45	0.043009	358	0.213340	-151.1957
legs	4	0.003823	148	0.088196	-114.1436
hard	17	0.016248	200	0.119184	-107.2196
left	53	0.050656	316	0.188311	-104.1315
SPJ	168	0.160570	602	0.358745	-97.3610
tired	59	0.056390	313	0.186523	-90.9984
smile	134	0.128074	503	0.299749	-88.7645
both	44	0.042054	247	0.147192	-76.4426
entry	107	0.102268	408	0.243136	-74.0312
nausea	9	0.008601	121	0.072106	-69.0931
TV	7	0.006690	110	0.065551	-67.0279
radio	7	0.006690	98	0.058400	-56.9579

#	16224	15.506513	27807	16.570821	-54.1064
good	167	0.159614	501	0.298557	-54.0322
sl	8	0.007646	95	0.056612	-51.1514
Christmas	3	0.002867	72	0.042906	-50.3404
model	11	0.010513	105	0.062571	-50.0859
right	68	0.064992	265	0.157919	-49.9223
last	84	0.080285	303	0.180564	-49.5752
yes	51	0.048744	218	0.129911	-47.6649
min	20	0.019115	129	0.076874	-45.8083
red	6	0.005734	79	0.047077	-44.6820
family	19	0.018159	124	0.073894	-44.5070
Dec	13	0.012425	103	0.061380	-43.3311
web	3	0.002867	62	0.036947	-41.5236
floor	3	0.002867	62	0.036947	-41.5236
models	7	0.006690	79	0.047077	-41.4401
town [a]	26	0.024850	136	0.081045	-38.8722
child's	5	0.004778	67	0.039926	-38.1951
display	3	0.002867	55	0.032775	-35.4386
water	24	0.022938	124	0.073894	-34.9303
leg	4	0.003823	56	0.033371	-32.5431
since	77	0.073594	248	0.147788	-31.8856
dry	8	0.007646	69	0.041118	-30.8231
cleaning	3	0.002867	49	0.029200	-30.2961

video	13	0.012425	84	0.050057	-29.8727
breath	25	0.023894	118	0.070318	-29.6767
university	13	0.012425	83	0.049461	-29.1926
mode	13	0.012425	82	0.048865	-28.5158
child	30	0.028673	128	0.076278	-27.8988
fell	5	0.004778	53	0.031583	-26.8741
taste	3	0.002867	44	0.026220	-26.0751
head	70	0.066904	215	0.128123	-24.6184
committee	5	0.004778	50	0.029796	-24.5221
couch	18	0.017203	90	0.053633	-24.3638

*p< 0.000000 for all negative keywords.

[a] Tokens noted in italics are used to protect confidentiality.

Because of the meaning inherent in repeated patterns of use (Stubbs, 2002), balanced samples of equal numbers of keywords from the two preceding analyses were sorted into semantic categories of *time/quantity, career and research, physiological perceptions,* and *home and family,* consistent with the divisions observed from the Master Corpus (Table 4.2). Inter-rater reliability was established based on Fleiss' kappa coefficient, yielding $\kappa\hat{} = .544$, in the range rated as *good*. Frequencies of occurrence were summed and Chi-square analyses were performed to determine if there were significant differences between the two conditions across the semantic categories, and Table 5.6 presents these results.

Notably, two of the semantic categories previously identified, *cognitive process,* and *writing process,* were not included in this analysis, as there were not sufficient keywords in these categories identified as occurring at a statistically more significant rate during the unmedicated conditions; therefore these samples could not be balanced with positive and negative keywords in the same manner as the other four categories. The tokens *lithium, smile, SPJ, good,* and *yes* occurred at a statistically significantly higher rate in the medicated condition, but again, there were no semantically-related tokens appearing at a similar rate in the unmedicated condition. This list is also conspicuous for the absence of the token *idea*, which suggests that it has been a consistently recurring theme in LMN's writing across time observed during both conditions, and hence it did not appear as a keyword. Again, caution is indicated in interpreting these observations, as the differences reported may have been a result of LMN's lexical choices being subject to diachronic topical variation, rather than a significant cognitive shift in focus.

Table 5.6.

Proportions of words by condition and semantic category reported as percentages of total.

	Time and quantity	Career and research	Physiological perception	Home and family
Unmedicated Total = 104,627	17.0548	0.4310	0.3708	0.1863

Medicated	17.9569	0.2347	0.6549	0.4797
Total = 167,807				

$\chi^2 = 305.178$, $df = 3$, $p < 0.0000000$

The results were overwhelmingly significant indicating differences between the two conditions for the categories of words from the keyness rankings. Given the magnitude of the difference, the results were further parsed to understand how the *observed* frequencies deviated from the *expected* frequencies as determined by the Chi-square tabulations for each of the semantic categories. These numbers reflect the percentage above or below the expected rate of occurrence as calculated by dividing the *observed* number by the *expected* number as denominator. Table 5.7 provides an explication of the relative representation of each.

Table 5.7.

Percent difference between observed and expected word frequencies by category and condition.

	Unmedicated	Medicated
Time and quantity	+ 1.00%	- 0.007%
Career and research	+ 45.00%	- 26.00%
Physiological perception	- 29.10%	+ 16.95%
Home and family	- 47.02%	+ 27.38%

These results suggest that if the relative contributions of each category were analyzed according to the observed versus expected elements comprising the total Chi-square value, that *time/quantity* differed to a negligible extent. This is a construct in which LMN was apparently equally invested, regardless of her medication state. However, further analysis of the patterns of language use indicate that when LMN was unmedicated, she appeared to be more involved with and preoccupied by her work as the rate of occurrence exceeded the expected by 45%, and she was less focused on her domestic concerns which had a rate of occurrence that dropped 47.02% below the expected. An inverse relationship was observed when she was medicated, as her attention appeared to shift toward her home and family, and away from her work-related interests. Another interesting finding was that her commentary on perceived physiological changes increased substantially when she was on medication (+16.95%), suggesting that she may have been dealing with side-effects of lithium, which may include tremors and nausea, in addition to the increased psychomotor involvement previously described (Katzung, 1995).

Despite the aforementioned confounding factors, Hypothesis 3 appeared to be confirmed as there were differences between patterns of language use between conditions of medicated versus unmedicated. In the analyses conducted, it appears that the initiation of pharmacotherapy served the intended purpose for LMN, that is, to regulate the extremes in her experience of mania to a degree that she was able to devote her energy and attention in a more balanced manner than she had during the years in which she was unmedicated. This is underscored by the observation of the remarkable shift in the topical fixations noted in lexical selection between conditions.

The patterning of words identified the topics to which she was devoting the most of her time and attention, and during the unmedicated periods, the evidence pointed to her career interests with disproportionately fewer notations regarding family and home. With the advent of lithium therapy, her patterns shifted significantly as averaged over nearly nineteen years, and her writing focused more specifically on domestic concerns with remarkably less attention devoted to her career and research pursuits. Review of lexical selection suggests that during the time period in which she was taking the medication, LMN made significantly more notations about her body parts and perceptions in such a way that the lithium appears to have had the unintended consequence of shaping her patterns of linguistic behavior. This observation is consistent with the DSM-IV marker regarding psychomotor agitation, as well as the potential confirmation of medication side-effects noted in the description of her physiological state.

Analysis 4 – Intra-individual Comparison of Manic versus Non-Manic Writing Patterns

Another intra-individual analysis was conducted using samples of LMN's writing that were generated during periods of time when she was and was not experiencing mania. Her habit has been to write in journals only during episodes, and that increased motivation to do so has become one of the signs that she is cycling into a period of mania; conversely, as her motivation to write wanes, so does her episode. Hypothesis 4 states: there are significant within-subject differences in patterns of written language use as measured between manic versus non-manic writing,

incorporating other samples of text generated by the same writer. In order to appreciate the scope of her language use in other contexts, LMN volunteered a selection of personal correspondence in the form of letters and e-mails to family and friends which was compiled into a corpus totaling 31,815 words. Although she had been a fairly prolific writer during her academic tenure, none of her journal articles, newspaper columns, or other professional publications was included in this compilation due to the extensive editing and review process inherent in such writing. This collection of personal correspondence more closely resembled the stream of consciousness production that is the hallmark of the journals, although the epistolary register is necessarily somewhat more structured. Careful reading of the texts indicated that in her letters and e-mails, LMN devoted more careful attention to sentence structure, sequential thought processes and themes, and demonstrated language use consistent with the pragmatic expectations of interpersonal communication. Additionally, she seemed less susceptible to interruptions secondary to internal or external distractions (e.g., no notations of prime numbers), although she reports that the personal correspondences may not have been written from start to finish in a single sitting. As noted in the Methodology section and demonstrated in Table 5.8, there are texts in the non-manic corpus that were written after the introduction of external observers; however, because these pieces of correspondence were produced during time periods when LMN was not experiencing a manic episode and were written for a specific audience in mind, these texts are not subject to the same exclusionary criteria. Although the comparison between the two corpora is not

exact because of the variation in genre and stylistic frame, it is the intra-individual variation in language use that warrants investigation.

Table 5.8.

Number of days per month with writing samples from non-manic periods.

	2000	2001	2002	2003	2004	2005	2006
January	1	2	1		3	6	5
February		2		2	2	16	
March		2	1		8		10
April		2	1		5	5	12
May		1	1			2	
June			2		3		
July		1	2	1	2	6	
August		1		1	2		
September				1	1	3	
October		2	1	1	2	6	
November			2		4	13	
December	1			6		8	

Table 5.9 presents the top fifty keywords from the non-manic corpus as compared to the patterns noted in the master corpus. The differences in frequency of occurrence were all significant to a level of $p < 0.000000$. Not unexpectedly, given

the genre, the content as indicated by words which occurred more frequently in the non-manic writing sample was primarily related to interpersonal interactions in terms of relating family news, expressing gratitude, and discussing upcoming plans. In this set of texts, LMN consistently employed personal pronouns (i.e., *I, you, my, your, we,* me), as well as third person referents (i.e., *he, she, his, they, her,* him) to a significantly greater degree than observed in her journals, as she engaged in interactional dialogue and descriptive narrative (e.g., the "newsy" letter to a relative).

Table 5.9.

Top fifty key content words comparing samples of non-manic writing to Master Corpus.

	Non-manic writing (31,815 words)		Master Corpus (242,589 words)		
Word	frequency	% total	frequency	% total	Keyness
I	1228	3.859814	3828	1.577977	644.4766
you	403	1.266698	575	0.237026	555.9887
he	158	0.496621	165	0.068016	274.4370
my	318	0.999528	793	0.326890	236.6924
email	54	0.169731	0	0.000000	232.7847
house	107	0.336319	83	0.034214	221.4555
your	148	0.465189	210	0.086566	204.4407
family [a]	43	0.135156	1	0.000412	176.0531
thanks	48	0.150872	7	0.002885	166.7059

weekend	49	0.154015	10	0.004122	159.9855
fun	70	0.220022	47	0.019374	155.7046
hope	71	0.223165	60	0.024733	140.1812
April	44	0.138299	17	0.007007	121.6601
walk	54	0.169731	38	0.015664	117.3938
we	166	0.521766	450	0.185498	108.6393
call	57	0.179160	54	0.022259	105.2125
year	51	0.160301	40	0.016488	104.8726
phone	32	0.100581	7	0.002885	102.9426
have	298	0.936665	1120	0.461686	102.6957
movie	44	0.138299	27	0.011129	101.9938
she	80	0.251453	127	0.052351	99.9611
his	64	0.201162	81	0.033389	96.8249
they	96	0.301744	192	0.079146	94.5242
December	27	0.084865	4	0.001648	93.5155
November	32	0.100581	11	0.004534	91.7331
condo	31	0.097438	10	0.004122	90.5242
years	54	0.169731	60	0.024733	89.8291
had	133	0.418041	358	0.147574	87.9678
I’m	158	0.496621	475	0.195804	86.8337
friend	31	0.097438	13	0.005358	83.4042
her	67	0.210592	108	0.044519	82.5149
family	79	0.248310	150	0.061832	82.4233

me	143	0.449473	419	0.172720	82.2202
friend	25	0.078579	5	0.002061	81.9483
pet	21	0.066006	2	0.008244	77.4105
nice	39	0.122583	33	0.013603	76.9179
friend	23	0.072292	5	0.002061	74.0843
February	34	0.106867	25	0.010305	72.2918
January	29	0.091151	15	0.006183	72.2237
October	23	0.072292	7	0.002885	68.2569
enjoyed	27	0.084865	15	0.006183	65.3186
cat	15	0.047147	0	0.000000	64.6461
glad	29	0.091151	22	0.009068	60.6745
car	26	0.081722	16	0.006595	60.1814
him	43	0.135156	60	0.024733	60.1537
nearby city	29	0.091151	23	0.009481	59.2645
day	62	0.194876	128	0.052764	58.7980
retirement	16	0.050290	2	0.000824	56.8910
congrats	13	0.040861	0	0.000000	56.0259
afternoon	26	0.081722	20	0.008244	54.0023

* $p < 0.0000000000$ for all positive keywords.

[a] Words noted in italics are used to protect confidentiality.

Negative keyness is noted in Table 5.10, presenting *all* of the words which occurred more frequently in the Master Corpus than in the non-manic corpus. This

list is obviously shorter, as the *WordSmith Tools* keyword analysis indicated that there were fewer words which were more prominent in the Master Corpus. Therefore, caution is suggested in interpreting this finding based on the disparity in corpus size, as well as the restricted range of time which the non-manic writing represents. Nonetheless, the content of this list is notable several things: the proportional use of # was overwhelmingly smaller, but this class of tokens still accounted for over 2% of her total words used; similarly, she also used the term *idea*, most consistently in her description of post-retirement activities in correspondence with an old friend; and she did include observations of time constructs (i.e., *Jan, Nov, p.m.*), although not to the degree noted in the Master Corpus. More obviously, LMN restricted her use of the typical preoccupations and idiosyncratic notations so prevalent in her journal writing (e.g., *prime, SPJ, pc*), which points directly to the differences in linguistic manifestations between writing composed while in a manic episode as compared to written documents from periods of remission. The level of significance (*p*) for these statistics was also less than 0.000000.

Table 5.10.

Negative key words which appeared more consistently in the Master Corpus when compared to the non-manic writing samples.

	Non-manic writing (31,815 words)		Master Corpus (242,589 words)		
Keyword	frequency	% total	frequency	% total	Keyword*
#	872	2.740845	39227	16.170147	-5596.0600

Jan	3	0.009429	499	0.205697	-99.3209
am	115	0.361464	1948	0.803004	-88.7663
etc	4	0.012572	470	0.193743	-87.0031
Nov	3	0.009429	359	0.147986	-66.7299
pm	21	0.066006	621	0.255988	-58.7045
tired	5	0.015715	355	0.146338	-56.3931
could	29	0.091151	633	0.260935	-42.9316
room	11	0.034574	389	0.160353	-42.5691
idea	10	0.031431	356	0.146750	-39.1473
head	3	0.009429	228	0.093986	-37.1185
food	3	0.009429	208	0.085741	-32.7326
left	12	0.037718	349	0.143864	-32.4682
arms	3	0.009429	185	0.076260	-27.7578
feel	14	0.044004	345	0.142215	-27.1021

*$p < 0.000000$ for all negative key words.

Although there is a relatively small proportion of keywords identified as remarkably different when comparing these two corpora, results of this analysis clearly indicate that there are differences in LMN's patterns of language use in these different conditions. The statistical summary demonstrated a difference in the content of the language used in writing produced during manic episodes as compared to that produced when she was reportedly not experiencing mania. Likewise, qualitative observations of register gleaned from review of the texts support the hypothesis of

differences related to her appreciation of the genre of personal correspondence, as she demonstrated shifts in syntactic structure as well as pragmatic function when engaging in interactive written communication. From that perspective, LMN produced linguistic pragmatic markers inherent to the genre of personal letters and e-mails, as she inquired into the health and activities of her correspondent, responded to comments from previous interactions, and shared pieces of personal information related to activities that might have been of interest to the other party (e.g., upcoming travel plans). As these texts representing interpersonal interaction were produced during periods when LMN was not experiencing mania, no differences from the typical personal correspondence genre were expected or observed.

Confirmation of Hypothesis 4 has potential diagnostic implications for application of corpus linguistic techniques in a clinical setting. The shift in patterns of language use between manic and non-manic periods was clearly defined when viewed through of large bodies of text. Observations of behaviors which match the DSM-IV criteria for flight of ideas, grandiose idea generation, associational chaining, or idiosyncratic patterns of use may be derived from analysis of discourse. Although the patterns would be more overt during florid exacerbations of mania, even less dramatic hypomanic manifestations may provide evidence of emerging or potentially prodromal episodes based on repeated use of particular words and phrases.

Analysis 5 – Intra-individual Comparison of Degrees of Mania

Based on the significant differences observed in the previous analyses, the next investigation involved examination of intra-individual variation based on texts

produced during the most acute experiences with mania, as compared with those produced during more typical episodes. Hypothesis 5 states: there are significant within-subject differences in patterns of written language use as measured during episodes of contrasting amplitude, addressing levels of acuity experiences as more or less significantly manic. LMN estimates that the episode which occurred in July of 1985 was the most severe exacerbation of her entire experience, corresponding with a subsequent hospitalization due to lithium toxicity. Journals produced during that time (the July 1985 corpus, n = 25,940 words) were separated from the Master Corpus (hereafter referred to as the reference corpus for this analysis) for the purpose of describing differences between levels of acuity in LMN's experience, based on her report. Positive keyword comparisons are presented in Table 5.11, noting the occurrences of the top fifty content words which occurred with greater frequency in the July, 1985 corpus, and were relatively more rare in the reference corpus.

Table 5.11.

Top fifty positive key content words comparing the July 1985 and reference corpora.

	July 1985 (25,940 words)		Reference Corpus (216,649 words)		
Key word	frequency	% of total	frequency	% of total	Keyness
left	239	0.921357	110	0.050773	660.2657
Li [a]	112	0.431766	2	0.000923	481.5166
eyes	119	0.458751	38	0.017539	367.3420
lithium	129	0.497301	61	0.028156	352.5772

right	163	0.628373	146	0.067390	335.1241
Ca [a]	71	0.273709	1	0.000461	307.3127
water	104	0.400925	39	0.018001	306.5799
swallows	51	0.196608	0	0.000000	228.1190
nausea	81	0.312259	40	0.018463	217.8416
regulation	52	0.200463	2	0.000923	215.9369
forehead	60	0.231303	11	0.005077	209.6510
dry	57	0.219738	17	0.007846	179.0410
ear	62	0.239013	25	0.011539	178.6272
power	48	0.185042	6	0.002769	178.3762
taste	42	0.161912	4	0.001846	161.5726
canal	50	0.192753	15	0.006923	156.8045
cool	44	0.169622	10	0.004615	147.3063
metallic	32	0.123362	0	0.000000	143.1125
model	64	0.246723	52	0.024002	138.4644
frequency	40	0.154202	8	0.003692	137.4547
clear	46	0.177332	19	0.008769	131.4902
nose	44	0.169622	16	0.007385	130.8213
breezes	29	0.111796	0	0.000000	129.6927
trembly	28	0.107941	0	0.000000	125.2196
crickets	35	0.134927	7	0.003231	120.2670
mouth	37	0.142637	11	0.005077	116.2912
tight	37	0.142637	11	0.005077	116.2912

red	48	0.185042	31	0.014309	115.8648
advisory	32	0.123362	6	0.002769	111.3197
air	37	0.142637	14	0.006462	108.6979
bite	24	0.092521	0	0.000000	107.3278
nasal	37	0.142637	16	0.007385	104.1742
models	47	0.181187	39	0.018001	100.5514
grapes	24	0.092521	1	0.000461	99.1565
regulator	22	0.084811	0	0.000000	98.3822
keno	34	0.131072	15	0.006923	95.0836
freq	30	0.115652	10	0.004615	91.4385
back	98	0.377795	233	0.107547	88.8788
oxygen	29	0.111796	10	0.004615	87.5492
mid	25	0.096376	5	0.002307	85.8968
complementary	22	0.084811	2	0.000923	85.0661
pita	19	0.073246	0	0.000000	84.9645
KP [b]	31	0.119507	15	0.006923	83.9419
breeze	26	0.100231	7	0.003231	83.7496
synthesis	25	0.096376	6	0.002769	82.6942
CHER [c]	20	0.077101	1	0.000461	81.6224
Pisces	20	0.077101	1	0.000461	81.6224
casino	18	0.069391	0	0.000000	80.4921
squint	18	0.069391	0	0.000000	80.4921
tongue	18	0.069391	0	0.000000	80.4921

*p< 0.0000000000 for all positive keyword values.

[a] Notation of chemical element.

[b] Abbreviation for *keno pop*, a physiological perception.

[c] Acronym for an organization described by LMN.

Review of the content words in the preceding table suggests some variation from the usual themes observed in LMN's writing patterns. As noted, the episode in which these texts were generated was particularly acute, and in her effort to control the manifestations of mania during that time, LMN exceeded the recommended dosages of her medication, resulting unfortunately in an accidental and potentially lethal overdose of lithium. LMN used the journals to document her process in detail, including recording each dose of lithium, as well as the physiological complications she was experiencing, including nausea, tremors in her extremities, dryness in her mouth and eyes, and a metallic taste on her tongue. She was evidently attempting to establish a homeostatic balance based on her knowledge of elemental chemistry and food composition, and her journal provides specific insights into that effort, including how many bites of a given food she would eat (i.e., *grapes, pita bread, buttermilk*), and the location and duration of tremors, as well as working through the chemical formulae for regulating and balancing *Li, Ca, helium* and *oxygen*. The idiosyncratic patterns of language use and idea generation evidenced during this episode were consistent with her diagnosis, as observed in her application of the terms *keno pop* (also noted as *KP*) to describe a particular repetitive physiological sensation; *Pisces* as referring to a research-related construct; *crickets* which referred to an unusual

auditory perception; and *casino* and *CHER*, used as she was formulating an idea involving a research organization and the related models of study.

Table 5.12 provides a converse view of the same period with analysis of the negative keyword occurrences, again reflecting the words which occurred with greater proportional frequency in the reference corpus than in the July 1985 sample. As was the case in the previous intra-individual analysis, there were very few tokens which appeared in this list due to the overwhelming shift in her immediate preoccupation and topical fixation. Predictably, she continued to apply # in her writing, although at a proportionately smaller rate than usual. Likewise, this table indicates the expected references to time constructs, as well as the prevalence of function word, the combination of which yielded a remarkably small set. The acuity of the July 1985 episode is notable for the dearth of her usual terms, as there was no evident notation in the keyword analysis of *prime, smile,* or *SPJ*, which speaks to the intensity of and attention devoted to her immediate phenomenon.

Table 5.12.

Negative keywords comparing the July 1985 and reference corpora.

	July 1985 (25,940 words)		Reference Corpus (216,649 words)		
Word	frequency	% of total	frequency	% of total	Keyness
am	7	0.026985	1941	0.895919	-379.2973
#	3197	12.3246	36030	16.630586	-337.3584
the	221	0.851966	4688	2.163868	-250.1247

I	217	0.836546	3611	1.666751	-121.4718
Jan	3	0.011565	496	0.228941	-89.0426
entry	5	0.019275	431	0.198939	-65.2899
a	185	0.713184	2668	1.231485	-61.2723
room	4	0.01542	385	0.177706	-60.4481
and	71	0.273709	1327	0.612511	-56.3528
my	39	0.150347	754	0.348028	-33.9999
since	6	0.02313	283	0.130626	-32.4922
had	10	0.03855	348	0.160628	-32.1799
morning	3	0.011565	210	0.096930	-29.3946
July	5	0.019275	240	0.110778	-27.8430
been	10	0.03855	317	0.146319	-26.9991
an	15	0.057826	373	0.172167	-24.4618

*p< 0.000000 for all negative keyword values.

According to these data, there were obvious differences in the patterning of words between the July 1985 sample and the remainder of the master corpus. LMN's topical referents changed, and she devoted a disproportionate amount of her writing to notes on her perception of the physical changes she was experiencing (e.g., nausea, trembling), as well as her attempt to understand and regulate the chemical interactions of the medication (i.e., lithium) and other elements (e.g., calcium, helium) which she believed might have an impact upon her physiology. As previously noted, this episode culminated in hospitalization secondary to the effects of lithium toxicity.

Positive and negative key content words were sorted into semantic categories based subjective ratings as well as familiarity with the intent based on the context of her writing. Inter-coder reliability was established by three independent raters, again with Fleiss' kappa coefficient of $\kappa\hat{} = .554$ which falls well within the *good* range, according to convention (Fleiss, 2003). Among other things, these findings revealed repeated references to chemical elements in the form of therapeutic medication; perceptual, spatial and physiological phenomena; specific colors; and a small set of food items. Chi-square analysis of word frequency in patterns of discernible categories achieved a statistical significance level (p <0.0000) indicating that based upon language used to describe her phenomenology, LMN's experience during the acute period in July 1985 was remarkably different than patterns which emerged within the combined manic episodes comprising the reference corpus. Table 5.13 delineates the differences in word frequencies represented as proportional percentages between the two corpora across semantic categories.

Table 5.13.

Percentages of occurrence by semantic categories.

	July, 1985 (25,940 words) %	Reference Corpus (216,649 words) %
anatomy	2.6638	0.2266
elements/therapeutics	2.4286	0.0715
research/professional	1.2336	0.0590

perceptual experiences	1.8080	0.1075
spatial orientation	1.7386	0.1269
physiology/motor activity	1.3569	0.0360
physical dynamics	0.8057	0.0263
colors	0.3739	0.0313
food	0.3084	0.00004

$\chi^2 = 194.72$, $df = 8$, $p < 0.0000$

Another lens through which this information can be viewed involves analysis of the difference between the *observed* frequencies of the respective semantic categories in accordance with the *expected* frequencies as indicated by the Chi-square tabulation. Calculating an expected frequency of responses for each category across corpora allows for such predictions, and Table 5.14 presents percentages of difference as formulated using the expected frequency as the denominator and the observed frequency as the numerator. Thus, it appears that during the July 1985 episode LMN wrote more about specific therapeutic and regulatory concerns (e.g., references to lithium, calcium, helium, thorazine); her specific physio-motor behaviors, ostensibly in response to the medication; and the attempt to establish systemic homeostasis (e.g., reduced nausea) with particular foods such as grapes, pita bread, buttermilk. Comparatively, in the reference corpus, LMN tended to write more about specific points on her anatomy (e.g., head, arms, legs) and the perceptual changes she was experiencing (e.g., “keno pops”).

Table 5.14.

Percent difference between observed and expected word frequency in the July 1985 sample versus the reference corpus by category.

	July, 1985	Reference Corpus
anatomy	- 15.04%	+ 33.21%
elements/therapeutics	+ 16.62%	- 36.68%
research/professional	+ 3.79%	- 8.37%
perceptual experiences	- 2.91%	+ 6.43%
spatial orientation	- 9.72%	+ 21.46%
physiology/motor activity	+ 18.95%	- 41.83%
physical dynamics	+ 14.18%	- 31.28%
colors	- 14.54%	+ 32.17%
food	+ 29.18%	- 64.36%

Although it is possible to infer trends from such data, caution is suggested in interpreting such categorical information based on the small numbers in some cells; for example, in the categories of *color* and *food*, LMN produced a fairly large number of occurrences of a small number of words. It does appear reasonable, however, to conclude that her usual focus was altered toward descriptions of the anomalous experience documented in the July 1985 corpus, with concerns about therapeutics or physiology increasing, and issues of anatomical structure or spatial orientation decreasing. Likewise, the volume of her writing during this period speaks to the

accuracy of the DSM-IV diagnostic criteria for identifying florid episodes of mania, in terms of expansive affect and tangential thought processes, increased productivity, decreased need for sleep, and flight of ideas. Based on the evidence presented, Hypothesis 5 was also confirmed.

Although the DSM-IV criteria may be less precise in the case of less overt episodes, their utility in combination with corpus linguistic analysis may present another level of sensitivity and confirmation for diagnostic consideration. Analysis of patterns of language use may offer a more sensitive barometer of variations in mood state, with potential implications for treatment option decisions, including level of service delivery (e.g., outpatient versus inpatient treatment) or regulation of medication.

Analysis 6 - Within-subject within-episode comparison

Differences in severity between episodes have been demonstrated; therefore the last analysis addressed variations in patterns of language use across the span of time as noted within a single episode. Hypothesis 6 states: there are significant within-subject differences in patterns of written language use as measured within a given manic episode, between early, middle, and late phases. The time period of March 10-30, 1984 was selected based on the availability of transcribed texts and the chronological representation of a complete episode. This was prior to LMN's formal diagnosis, and pharmacological intervention had not yet been implemented. The 21-day episode was divided into three equal segments of seven days each, and was divided again into periods reflecting the first six hours of the day (midnight to 6:00

a.m.) and the second 18 hours of the day (6:00 a.m. to midnight). As indicated in Chapter Three, journal entries were defined as beginning with a time stamp, recording a thought as text, and then closing with either a time stamp or white space. The number of entries per day per time period is presented Table 5.15.

Table 5.15.

Number of journal entries per day in early, middle and late phases of a single episode.

Date	12:00 midnight to 6:00 a.m. (6 hour period)	6:00 a.m. to 12:00 midnight (18 hour period)	Total (24 hours)
Early Period			
10 March 84	17	15	32
11 March 84	14	17	31
12 March 84	3	5	8
13 March 84	7	16	23
14 March 84	27	15	42
15 March 84	27	6	33
16 March 84	2	6	8
Early Period Total	*97*	*80*	*177*
Middle Period			
17 March 84	4	4	8
18 March 84	1	34	35
19 March 84	13	8	21

20 March 84	18	40	58
21 March 84	13	4	17
22 March 84	6	0	6
23 March 84	1	3	4
Middle Period Total	*56*	*93*	*149*
Late Period			
24 March 84	1	51	52
25 March 84	29	13	42
26 March 84	27	12	39
27 March 84	4	5	9
28 March 84	21	7	28
29 March 84	0	0	0
30 March 84	5	0	5
Late Period Total	*87*	*88*	*175*
Grand Total	***240***	***261***	***501***

Analysis of the patterns of language use revealed differences which were delineated in Table 5.16. LMN averaged nearly as many entries during the six hour period of the very early morning as she did during the rest of the day, and as she moved toward the end of the episode, there was a slight increase in the number of words per entry, although no remarkable change in the average number of entries. Although reflected in this data set, on March 29th, the day prior to the last entry, LMN

made zero entries in her journal, and on the last day of the episode, there were only five entries. Although results from this single episode cannot be interpreted to indicate global trends in LMN's writing habits, the differences are clear: consistent with the diagnosis of mania, LMN demonstrated increased activity during the early morning hours, and as her episode began to wane, her entries became slightly longer and less frequent, a predictor variable LMN reported as an indication that an episode was nearly over.

Table 5.16.

Descriptive statistics for calculating number of entries per time period.

	Early	Middle	Late	
Word totals:	6802	5380	7571	
Number of entries per period:				
Midnight to 6:00 a.m. (6 hrs)	97	56	87	Mean = 80 (49.71% total)
6:00 a.m. to midnight (18 hrs)	80	93	88	Mean = 87 (50.29% total)
TOTAL	177	149	175	
Mean # words per entry	38.42	36.10	43.26	

Keyword analysis was conducted to assess the interaction between the three time segments, as noted in Table 5.17. These comparisons yielded almost no

differences, with the exception of LMN's use of two individually-defined identifiers: *FK* which stands for "function key side", and *CK* or "control key side", used as notations of spatial orientation to situate perceptual sensations. As mentioned previously, these idiosyncratic markers were used for a fairly brief time during the mid-1980's in LMN's journals, and were eventually abandoned for the more conventional use of *right* and *left*. In this particular data set, she used the terms primarily in the early segment, with significantly fewer occurring in the middle segment, and none at all in the third segment. The only other token which occurred at a significantly different rate was the use of *PM* as part of a time tag. Use of this term occurred across the three samples, but at a rate which was statistically different only between the middle and late periods.

Table 5.17.

Keywords from early, middle, and late sample comparisons within a single episode.

Word	frequency	% total	frequency	% total	Keyness*
	Early Period		Middle Period		
FK	22	0.32	0	0.00	25.67
	Middle Period		Late Period		
PM	95	1.77	48	0.63	36.36
	Early Period		Late Period		
FK	22	0.32	0	0.00	32.96
CK	17	0.25	0	0.00	25.46

*$p < 0.000000$ for all keyness values.

Table 5.18 presents an analysis of the highest frequency content words with proportional representations across the three time periods. Not unexpectedly, given the sequential and proximal nature of the data, the patterns are fairly similar in terms of word order. Numbers (#) continue to be the most frequently occurring token for LMN, and consistently represent about 19% of all her words in all three samples in this analysis. Other familiar constructs are represented consistently as well, with time references (i.e., *Mar*[ch], and *pm*); use of the modifier *good*; and *FK(S)* and *CK(S)*, as previously noted.

Table 5.18.

Proportional representation of top fifty content words of within-episode time periods.

	Early		Middle		Late	
	March 10-16		March 17-23		March 24-30	
	(6802 words)		(5380 words)		(7571 words)	
keyword	freq	%	freq	%	freq	%
#	1273	18.71	1018	18.92	1452	19.18
Mar	88	1.29	69	1.28	79	1.04
I	83	1.22	39	0.72	65	0.86
pm	75	1.10	95	1.77	48	0.63
my	29	0.42	13	0.24	26	0.34
I'm	24	0.35	20	0.37	27	0.36
FK(S)	22	0.32	17	0.32	3	0.04
good	22	0.32	10	0.19	2	0.03

data	20	0.29	9	0.16	22	0.29
CK(S)	17	0.24	13	0.24	7	0.09
computer	17	0.24	4	0.07	9	0.11
very	17	0.24	20	0.37	18	0.23
idea	16	0.23	6	0.11	4	0.05
cold	13	0.19	10	0.18	3	0.04
food	13	0.19	19	0.35	7	0.09
activity	12	0.17	1	0.02	2	0.03
before	12	0.17	2	0.04	8	0.11
manual	12	0.17	1	0.02	6	0.08
brain	11	0.16	0	0.00	6	0.08
feel	11	0.16	9	0.16	8	0.11
LMN [a]	11	0.16	1	0.02	3	0.04
sleep	11	0.16	13	0.24	6	0.08
earlier	10	0.14	0	0.00	1	0.01
menu(s)	10	0.14	1	0.02	0	0.00
record	10	0.14	8	0.14	4	0.05
side	10	0.14	1	0.02	1	0.01
sound	10	0.14	0	0.00	0	0.00
time	10	0.14	3	0.06	2	0.03
write	10	0.14	2	0.04	4	0.05
breathing	9	0.13	8	0.14	10	0.13
color	9	0.13	2	0.04	2	0.03

KP	9	0.13	6	0.11	0	0.00
video	9	0.13	0	0.00	0	0.00
lab(oratory)	8	0.11	1	0.02	5	0.07
intense	6	0.09	13	0.24	2	0.03
affective	0	0.00	11	0.20	10	0.13
disorder(s)	0	0.00	9	0.16	4	0.05
library	0	0.00	9	0.16	1	0.01
journal	1	0.01	8	0.14	12	0.15
mania*	0	0.00	8	0.14	14	0.18
canal	4	0.06	7	0.13	1	0.01
ear	6	0.09	7	0.13	2	0.03
forehead	2	0.03	7	0.13	0	0.00
o'clock	1	0.01	7	0.13	1	0.01
depress*	0	0.00	9	0.16	12	0.15
relax*	3	0.04	6	0.11	12	0.15
child [a]	3	0.04	6	0.11	2	0.03
pc	3	0.04	19	0.35	17	0.22
nutrient	7	0.10	6	0.11	9	0.11
coffee	3	0.04	3	0.06	8	0.10

* Including lemmatized forms.

[a] Words noted in italics are used to protect confidentiality.

The evidence presented confirms Hypothesis 6, demonstrating differences in patterns of language use within the course of a single episode. In addition to the variations in writing observed across the designated periods, one of the most important findings from this data set is the changing patterns in lexical selection across time. One of LMN's most consistent manifestations of mania, the use of *idea*, changes across the early, middle, and late segments of time in this data set, as there is a diminishing proportion of occurrence of this token from early (0.23%) to middle (0.11 %) to late (0.05%). Likewise, her use of the abbreviation *KP* to indicate the physiological sensation she describes as *keno pops* declined from 0.13% to 0.00% during the same time period. Conversely, as the episode progressed, she showed an increased use of the tokens *mania* (0.00% to 0.18%) and *journal* (0.01% to 0.15%) across the same three-week span of time, suggesting an evolving pattern of self-reflection and condition-related insight as her episode waned.

In consideration of diagnostic formulation according to the DSM-IV, however, the more significant finding deals with the variation in the basic count of the number of entries as differing manifestations of mania across the episode. Consistent with the criteria describing expansive affect, decreased need for sleep, and the related increase in energy and activity, LMN's rate of journal entries clearly indicated a manic state at the beginning of the episode, particularly during the six hour period from midnight to 6:00 a.m. As her episode waned, her writing behaviors declined. This variability provides another potential clinical point of reference, as investigation of linguistic behaviors such as frequency of journal entries has implications for monitoring the course of a manic episode.

Summary

The hypotheses tested in this chapter have dealt with variations in patterns of language use in LMN's experience with mania. These analyses were structured to address differences from every perspective available, including conditions of being medicated versus unmedicated; samples of manic versus non-manic writing; episodes gauged as more versus less severe; and early, middle, and late phase changes within a single episode. The results of these analyses were convincing, indeed confirming that there are differences in patterns of LMN's language use as measured at different points in her process. The linguistic behaviors identified were generally consistent with the diagnostic criteria for mania as described in the DSM-IV, but corpus linguistic methodology demonstrated a remarkably subtle sensitivity to the variations observed, which has significant implications for clinical application.

Chapter 6 – Conclusion

This study has investigated the patterns of language use as produced in the written texts of an individual, identified as LMN, diagnosed with Bipolar Disorder – Mania Only. LMN kept a series of journals for a period of twenty eight years, produced only during the time when she was experiencing a cyclic manic episode. Selected samples of the 100+ journals were transcribed and compiled into a corpus totaling 242,589 words, identified as the Master Corpus, which served as the basis for testing six hypotheses using *WordSmith Tools* (Scott, 2005), a software program designed for corpus linguistic investigation. The analyses involved comparison of LMN's patterns of use in an attempt to determine differences between the language used by the individual and reference corpora designed to represent 'typical' language and genre-specific use, as well as intra-individual differences as measured at different points in LMN's experience. This study has provided a series of insights into the variation in language use in mania, and the utility of applied corpus linguistic methodology into diagnostic formulation. This chapter summarizes the overall results of the research, and discusses the potential for the applied uses of corpus linguistics in diagnostic formulation and in clinical practice, the limitations of the current study, and proposals for future directions of further research, and then situates language variation theory in the context of mental illness.

Review of findings

The first analysis conducted confirmed that LMN's linguistic behavior differed significantly from typical language as represented in FROWN, a reference corpus of 1.47 million words of written American English. LMN's patterns of word frequency were notably different from FROWN and among other things revealed her preoccupations with prime numbers, notations of time, commentary on perceived physiological changes secondary to her manic experience, and intense focus on her research interests. Although the criteria tend to be underspecified and ambiguous, these linguistic manifestations were consistent with the DSM-IV (1994) classification of mania as related to an increase in patterns of idiosyncratic speech, as evidenced by particular fixations and accompanied by "flight of ideas," in which accelerated rate of production occurred with abrupt topical shifts. The entries in LMN's journals were notable for rapid shifts from theme to notation of time and prime number references, sometimes occurring at a rate of several per minute. There was also occasional associational chaining noted in which words were selected based on alliterative or assonant qualities, rather than semantic content. Additionally, LMN's repeated use of the token *idea* was characteristic of the tendency toward idea generation in mania, notable for the increased *quantity* of these references in her journals, rather than just *qualitative* aspects of such potentially grandiose manifestations. Likewise, increased organization and planning behaviors were noted as her journals included descriptions involving founding organizations, proposals for grant funding, and other civic and professional activities for which she would take a leadership role. Her patterns of language use were also consistent with other more descriptive aspects of mania as

described in the research literature, as her discourse prosodies included emotional referents such as *smile* in a manner consistent with what has been termed mirthful mood (Fenichel, 1945; Stubbs, 2002). Likewise, there were breakdowns in the maintenance of meaningful relationships between sentences (Durbin & Martin, 1977). LMN's journals also revealed a remarkable rate of productions incorporating grandiose themes (Mendhekar, Srivastav, Jiloha, & Awana, 2004); indicators of expansive thinking, and playful, flippant responses (Daniels et al., 1988); signs of disinhibition and overinclusion (Khadivi, Wetzler, & Wilson, 1997); and repeated instances of distractibility (Liddle et al., 2002).

The second comparison revealed some similarities between LMN's journals and a journal-as-genre reference corpus, as both sets of texts demonstrated adherence to certain basic conventions of journal writing as described by Fulwiler (1987), such as notation of chronological entries, and personal observations and experiences. Thus, in a manner consistent with other journal-keepers, across time LMN used the journals to satisfy a number of traditional conventions (Gannett, 1992), such as common-place observations of her daily experience, commenting on her spiritual or existential reflections, and elaborating on descriptions of her research process. However, there were significant differences between LMN's writing and the patterns of language use produced by writers of the other journals. As expected, the journal-as-genre corpus showed consistent use of specific topical indicators in word selection that reflected the unique circumstance of the respective authors. Hence, word frequencies and themes produced by LMN and the other writers were fundamentally consistent with relevant life experience. However, LMN's notations were observed to reflect a more

egocentric perspective as her use of personal pronoun referents exceeded that observed in the combined reference corpus using the other three journals for comparison (e.g., her use of the tokens *I'm, I've, I'll* equaled 0.3293% of the total for the Master Corpus, and the same tokens represented only 0.0936% of the reference corpus). Additionally, her remarkably frequent use of date and time notations represented a significant difference in patterning as compared to the reference corpus in which there was usually a single indication of date and/or time with an entry. These differences in LMN's patterns as compared to the other writers were noteworthy, and provided cross-validation of the temporal preoccupation and self-centered perspective as identified in the previous comparison with the FROWN corpus.

The first intra-individual analysis revealed demonstrable differences in patterns of language use between LMN's conditions of being medicated versus unmedicated. Beginning in 1978 and continuing until 1985, LMN's condition was undiagnosed and she did not receive treatment. In 2004, she elected to discontinue nearly nineteen years of ongoing pharmacological intervention with lithium carbonate which had been prescribed to stabilize her mood states (Katzung, 1995; Murray, 1985). Combining the unmedicated periods for comparison with the medicated time frame provided the opportunity to evaluate the effects of treatment on her patterns of writing. There were notable differences observed during these conditions, including remarkable shifts in topical fixation. For example, *career and research* was a central focus during the unmedicated state, while references to *family and home* became higher priorities during the medicated condition. Semantic categorization of

frequently occurring words clearly illustrated this contrast as an inverse relationship when calculated based on the proportional differences between *observed* and *expected* rates of occurrence. Proportional percentages of *career and research* went from 45.00% above the expected when unmedicated to 26.00% below the expected in the medicated condition. Conversely, the proportional representation of *family and home* was 47.02% below and 27.38% above the expected under the respective unmedicated and medicated conditions. These findings were supported by LMN's report of her husband's insistence that she remain on the medication in order to sustain involvement in family life. The results appear to reflect the therapeutic impact of the medication which moderated the tendency toward what LMN observed as excessive goal-directed pursuits in the workplace, as often characterizes mania. Her changes in patterns of language use according to treatment condition attest to the utility of lithium carbonate in normalizing function, as described by Katzung (1995) and Murray (1985). Interestingly, her observation of time and number constructs was uninterrupted across both conditions, reflecting the predominance of this preoccupation for LMN.

The comparison of writing produced during manic versus non-manic periods was accomplished by contrasting written language in the Master Corpus with a reference corpus compiled from LMN's personal correspondence in the form of letters and e-mails to friends and family. The analysis found significant differences in language use with regard to the predicted patterns of word frequency, and not unexpectedly, a tendency for journal writing during manic episodes to take a more self-centered focus, whereas the personal correspondence revealed a more interactive

and social register. The inherently private and introspective nature of the journals provided the opportunity to observe the extent of LMN's awareness of and insight into the manifestations of mania, while the personal correspondence allowed a perspective on her ability to exert executive control over aspects of her condition that might impair interpersonal communication. LMN successfully shifted from a predominantly internal (self-oriented) to external (other) focus between the genres, and when writing to others, she accommodated to the syntactic and pragmatic conventions of the register. This analysis provided an example of LMN's ability to manage mania in a way that has been essential to her history of success in personal and professional endeavors. For individuals with mania, this type of linguistic analysis has important diagnostic and prognostic implications reflecting the shifts of insight as referred to by Akiskal et al. (2001), Dell'Osso et al.(2000) and Ghaemi and Rosenquist (2004).

The impact of the manifest severity of a particular manic episode upon language behavior was assessed by comparing a single set of journal entries produced when LMN was experiencing a particularly acute exacerbation to the remainder of the Master Corpus. The language used during this particularly acute time period which required hospitalization was found to be significantly different from the patterns evidenced in the rest of the Master Corpus, as LMN's word frequency patterns shifted dramatically to reflect her observations of the changes in her physical and cognitive state, her preoccupation with the medication, and her attempts to establish and regulate physiological homeostasis. Much of her writing was readily categorized into semantic groupings that represented these changing interests, which was illustrated by

an observed 16.62% increase over the expected rate of occurrence for words associated with *elements and therapeutics*. The hospitalization was necessitated by an inadvertent overdose resulting in lithium toxicity, a serious condition with potentially fatal side effects (Katzung, 1995). The language analysis revealed problems potentially related with toxicity, as LMN documented each dose of medicine, and demonstrated an increasing preoccupation with ways to reduce what may have been side-effects of the overdose. Her language incorporated the aforementioned elements (e.g., lithium, calcium), and also referred to possible benefits of combinations of certain foods. Corpus linguistic analysis demonstrated clear sensitivity to her changing clinical status, as well as observations of the apparent side-effects of the medication.

The final hypothesis addressed possible variations in language use within the context of a single manic episode, examining the changes in patterns of language use during the early, middle, and late phases. Notably, an initial analysis without regard for these designated phases revealed a very high rate of late night/early morning activity, as LMN produced an equal number of entries in the six hours of midnight to 6:00 a.m. as she did during the eighteen hour span from 6:00 a.m. to midnight. Such an excessive rate of occurrence during the time typically designated for sleep was consistent with one of the hallmarks of mania as described in the DSM-IV (1994) and other texts (c.f. Kaplan & Sadock, 1998). The twenty-one day manic episode was divided into three phases (seven days each) noted as early, middle and late. Review of the patterns of language found that the use of two tokens in particular, *idea* and *keno pops*, markedly declined across time, suggesting that focus on idea generation,

flight of ideas, and heightened physiological activity progressively diminished with resolution of the episode. Additionally, across the same time span there was a notable increase in the use of the tokens *mania* and *journal*, suggesting an evolving pattern of insight and self-reflection as the symptoms of mania waned. If corroborated by further study, this correlation could have utility for tracking the course of a manic episode and modifying treatment, especially if considered in the context of variation in late night/early morning activity. A related observation concerning the early, middle, and late phases was that there were fewer but longer entries as the episode waned, consistent with the comments by Depp and Jeste (2004).

Diagnostic Implications

The findings provided support for all six hypotheses, and the data were not inconsistent with the underspecified terminology regarding language and mania as described in the DSM-IV (1994) and other literature. LMN thus did fit the established diagnostic criteria, but analysis of her language revealed a great deal more than the imprecise language-related criteria indicated in the DSM-IV. Her journals provided a prototype for more appropriate operational definitions of language use in mania, and such clear-cut identification of the language patterns through corpus analysis underscored the potential diagnostic utility of the procedure. As is the case with all classifications of illness, a diagnosis of mania is made based on a constellation of clinical signs and symptoms that manifest in particular patterns over a designated period of time. Before rendering a categorical diagnosis, a practitioner is required to incorporate all perceptibly relevant factors and influences, including but

not limited to age, family and personal history, cognitive functioning, physical health status, socioeconomic status, environmental stress, and potential impact of any comorbid psychological or medical conditions. Within this context, the nature and amplitude of presenting manifestations of signs and symptoms must be rated on a continuum from mild to severe to finalize the diagnostic formulation. Historically, the ability to consistently rate the linguistically-based manifestations of mania and other mental illnesses has been hampered by subjective diagnostic terminology and inadequate assessment procedures typically limited to brief clinical observation. The current status of the terminology and technology in the field seems only minimally adequate and far from ideal regarding the language behaviors seen in mental illness. This study presents an effort to remedy the situation as it currently exists using results from the corpus linguistic analysis of patterns of language in LMN's journals as a means for offering constructive suggestions.

In this instance, corpus linguistic methodology was clearly demonstrated to be a sensitive and accurate tool for detecting the presence of language-related manifestations of mania. Rather than reliance on subjective observations of decontextualized behaviors in a clinical setting, this methodology could be implemented as a diagnostic tool to analyze large bodies of authentic text. This has potential diagnostic importance in revealing patterns of repeated behaviors in word frequency and collocations that could not otherwise be quantitatively verified without the advantage of such technology. The need for such instrumentation and the value of a condition-specific reference corpus that would derive from this application could help overcome the weaknesses of the poor operational definitions of the DSM-IV and

related literature, and greatly improve upon the small observational samples that characterize clinical interviews or anecdotal reports. A primary flaw in the diagnostic process that has been openly acknowledged by the DSM-IV is the lack of a "consistent operational definition that covers all situations" (1994, p. xxi). The introduction of coined terminology (e.g., *verbigeration*) provides an example of this failure to enhance diagnostic reliability by further obfuscating the operational definitions of meaningful diagnostic constructs.

Such terminology is mostly unquantifiable and representative of many of the ostensibly defining characteristics of illnesses described in the DSM-IV (c.f. Houts, 2002 for further criticism on flaws in the DSM-IV classification system). Perhaps the immediate demands of clinical practice and the evolution of common jargon among practitioners have resulted in tacit acceptance of nebulous terms and non-specific standards for assessing the linguistic parameters of certain mental illnesses. Consideration of how to properly evaluate rarely occurring manifestations of language further contributes to the difficulties produced by the vagueness portrayed in terms such as *flight of ideas, word salad* or *press of speech*. The diagnostic significance of such language variation could also be complicated by other cognitive, linguistic, cultural, or emotional factors which may occur concomitantly. The literature is replete with descriptive terms, many of which have been invented to describe certain phenomena, but which lack clear operational definition and provide little insight into the processes observed. Exemplary of such vagueness are *extreme fabulizing* (Daniels, et al., 1988)*, gibberish* (Chaika, 1990), or *wooly* (Reilly, 1975). A more utilitarian approach would be to discard terminological classifications that

lack either diagnostic validity or operational definitions sufficient to demonstrate adequate inter-examiner reliability. The most useful model would provide appropriate descriptive terminology with demonstrated sensitivity and specificity, in addition to good inter-rater reliability required to provide a useful description of signs and symptoms. These could be combined with techniques that render quantifiable findings which can be referenced to normative parameters reflecting the actual pathology. The addition of corpus linguistic methodology to the assessment process could offer greatly enhanced objectivity and sensitivity to the measurement of language function, increasing the chances of detecting diagnostically relevant subtle differences that may not be observable under other circumstances.

From a linguistic perspective, mania manifests as cyclical shifts in typical patterns of language use, specifically related to the repeated patterns of unusual tokens and collocational patterning, as well as differences in the use of specific linguistic behaviors. These linguistic phenomena present in a context of rapidly shifting attention, and are simultaneously influenced by non-linguistic factors such as expansive affect and increased energy, decreased need for sleep, and psychomotor agitation. In an effort to streamline and group the diagnostic markers into a model that is more cohesive and consistent with linguistic theory, the following reconfiguration is offered based on the patterns explicated in the literature as well as the results of analyses of the LMN corpus. The array of linguistic and linguistically-mediated indicators of mania is outlined in Table 6.1, and briefly described below.

Table 6.1

Proposed model for linguistic patterning in mania.

I. Attentional shifts and related cognitive/linguistic breakdown
    A. Intra-individual manifestations
        1. Distractibility
        2. Memory impairment
        3. Overinclusive thinking
    B. Extra-individual manifestations
        1. Loss of topic
        2. Limited insight
        3. Language use which is difficult to interpret
        4. Presupposition
        5. Accelerated rate

II. Recurrent patterns
    A. Idea generation
    B. Perseveration

III. Specific linguistic indicators
    A. Lexical fixations
    B. Impaired coherence and cohesion

IV. Other linguistically-mediated processes
    A. Elevated mood
    B. Increased energy

I. Attentional shifts and related cognitive/linguistic breakdown

The inability to maintain attention to a task or topic without being distracted by extraneous stimuli is widely recognized as an indicator of mania, and is discussed in the literature as involving a variety of manifestations. This behavior is recognized as *flight of ideas* in the DSM-IV, but rather than providing a precise diagnostic description, the term tends to serve as a utilitarian catch-all into which many observed linguistic patterns are tossed without further examination. In an effort to present this construct in a more operational and quantifiable manner, the functional components of attentional breakdown are discussed as relating to intra-individual versus extra-individual manifestations, recognizing that such *a priori* delineations are subject to the influences which work in both directions.

A. Intra-individual manifestations

1. Distractibility. The tendency to become distracted from a task or topic of conversation is not necessarily a clinical sign, as this occurs in healthy persons due to any number of influences. However, the inability to appropriately discriminate and regulate one's response to such distractions is suggestive of pathology, and is observed in the clinical formulations of a number of diagnoses (e.g., Attention Deficit Disorder). In mania, this is recognized as contributing to the ubiquitous *flight of ideas*, although as noted, the term does not account for the complexity of the internal processes involved. Distractibility has been defined as a dysregulation of thought, a process that has been identified as more prevalent during acute exacerbations of mania (c.f. Liddle, et al., 2002). For LMN, distractibility was a common occurrence of which she expressed awareness with her notations of "can't retrieve", suggesting

that ideas had occurred to her but were lost before they could be recorded because of interference from other stimuli, as observed in the following example:

2 Nov 91 19:56:27

Very productive idea - generating shower. The first in a long time. The thoughts/ideas were in random order - the randomly generated ideas in random order.

19:57:53 prime 19:58:04

19:58:28 Shower - thoughts (con't)

I don't worry as much about recording every thought because I know it will resurface (but as I write this I'm thinking I really can't remember ideas from shower)

19:39:58

2. Memory impairment. Although this may be a more subtle indicator, Fleck and his colleagues (2003) suggest that there is a cognitive memory component to the deficits observed in mania involving breakdown in the ability to encode and consolidate new learning. This breakdown is a product of distractibility and is associated with inefficient processing of information, secondary to the inability to deal with shifts in attention to other incoming stimuli. Related to this breakdown, there are also deficits in the efficient retrieval of old-learned information. Consideration of a working memory component in this configuration is worthy of future attention. For LMN, this appeared to be related to her distractibility, but she did express insight into this, as noted:

9 July 95 1:38:13

In the midst of the last entry I remembered I told *a friend* I'd mail her son a copy of my grant proposal. That was on July [June] 24 - this was the first time it crossed my mind. Bothers me that I forget so much these days. I have to devise a 'system' to remind me of things I have to do.

1:40:30

3. Overinclusive thinking. Khadivi, et al. (1997) described language in individuals with mania as having tendencies for "connect[ing] things that are normally kept apart" (p. 372), noting that shifting attention from one stimulus to another resulted in attempts to link dissimilar topics, yielding flawed and unrealistic commentary. Daniels et al. (1988) made similar observations of patterns of overly combinatory thinking in people with mania, and Docherty and colleagues (1996) noted problems with figure/ground confusion, in which focus on the target was diffused to include ostensibly irrelevant details from the context. In the LMN corpus, such linkages were notable partly as a consequence of the rapid shifts in attention, as noted:

6 Aug 83 8:54:20 am

Comments on another 9:05:00 am characteristic: I seem to get the urge to 'get my wardrobe together' and to 'clean the house' neither do I normally give very much thought. After coming in from the walk, *husband* was going to take *friend* to *city* to get carpet, he put oil and antifreeze in my car. I started throwing things 'junk (in the car)' away. Came in 'straightened' a little in the kitchen. Washed the sheets off both

beds. Then started to go through my clothes. Want to give to the Goodwill things I don't use.

9:08:58 am

B. Extra-individual manifestations

1. Loss of topic. The inability to maintain a topic of conversation consequent to distractibility and digression is often described as *derailment* in the literature (Andreasen, 1979, 1986; Harrow, et al., 1986). McPherson and Harvey (1996) argue that this behavior represents an interfering process in the linguistic expression of people with mania; they suggest that it involves an attempt to maintain concurrent discussions of multiple coherent topics with figure/ground confounds which require increased effort by the listener/reader to interpret. Loss of topic may present itself through interruption in the flow of a topic, as by the insertion of frequent time references or other disconnections in the continuity of the discourse. The following example of LMN's writing demonstrates the type of disjointed stream-of-consciousness process that is consistent with loss of topic:

9 Aug 83 5:37:00 am

Just moved into the living room. As I walked through dining room, decided I wanted to draw the 'corner with our plants'. Am not going to do it now, but hope to sometime soon.

5:38:32 am

5:39:03am

There is SOMETHING about drawing. When I'm not in 'one of these periods' I do not draw. It seems that there is 'something inside of me that needs to be conveyed' and the means of conveying is drawing. I'd like to learn how!

5:40:54 am

5:41:20 am

I moved in here to write about something and I 'can't retrieve it'.

5:41:56 am

5:45:55 am

I turned out the light, sat in the rocking chair, drinking my coffee. I believe this is what I was going to write.

5:46:48 am

2. Limited insight. The lack of self-awareness of impairment is considered to be a state-dependent function of the cyclical nature of mania (Dell'Osso et al., 2000; Ghaemi & Rosenquist, 2004), with improvements in insight noted as episodes remit. Linguistically, Docherty et al. (1996) describe insight-related communication disturbances as a breakdown in the speaker or writer's ability to appreciate that the meaning is not shared by the receiver of the information. This is suggested to be a cognitive-linguistic deficit notable for vague references, ambiguous word meanings, and unclear structures for which no clarification is provided because the speaker is unaware that clarification is needed (Durbin & Martin, 1977). Impaired insight is recognized as one aspect of the profile of mania (Altman, 1998, Altman et al., 1997, 2001), and is consistent with the positive or inflated self-esteem that is also observed

with this diagnosis (DSM-IV, 1994). Limited insight is grounded in the rapid shift in attention because the speaker or writer works under the assumption that the listener or reader is following their idiosyncratic variation in stream of consciousness patterns, which perpetuates the cycle of decreased awareness as the topic changes.

LMN has acknowledged that in retrospective review of her experience with mania, there were probably times when she was not aware of the impact of her behavior on the feelings or perceptions of others. She went on to note that if she was aware, it was of no consequence to her at that moment because of her elevated mood state and level of energy, although she expressed regret at the thought of having hurt the feelings of close friends and family.

3. Language use which is difficult to interpret. One of the basic constructs described in the DSM-IV (1994) as pathognomic of mania is that of idiosyncratic patterns of speech production, for which referents are not clear to the listener. Barch and Berenbaum (1996, 1997) supported this notion, arguing for recognition of language production deficits in mental illness. They described patterns notable for discourse planning deficits, including incompetent referents that made speech more difficult to interpret due to poorly integrated constructs. In LMN's data, the primary and recurrent pattern of idiosyncratic speech observed related to her observation of prime numbers. Chronological review of her journals allows an appreciation of this pattern in context which reveals that this is a touchstone concept to which LMN repeatedly returns, as it provides a level of structure and meaning which is apparent and significant only for her.

4. Accelerated rate. Ribeiro (1994) described a remarkably increased rate of verbal expression in her study of an acutely psychotic patient in an inpatient hospital setting, using the term *press of speech* (also noted in the literature as *pressured speech* or *push for speech*). This is recognized as involving an increased rate of production of words per minute with resulting uninterruptability (Durbin & Martin, 1977), but also a rapid shifting in topical reference, and may include increased volume and pitch. Mendhekar et al. (2004) and other researchers have also described this pattern as increased talkativeness, which speaks to the increased tendency to initiate conversation, as well as decreased awareness of the pragmatic rules governing conversational turn-taking, and may result in language which is expressive, but not communicative (Lorenz & Cobb, 1952). LMN's corpus is based on language produced in written form, and although the rate of production is ostensibly different in writing as compared to speaking, she still demonstrated unusual acceleration on occasion, most notably during her extreme exacerbations, in which she was noted to record up to three separate entries within sixty seconds.

5. Presupposition. This construct deals with the assumptions that speakers or writers make about the current knowledge set of the receiver of information, in such a manner that there is a common point of reference for both parties. Durbin & Martin (1977) address this in terms of errors within sentences and across discourse, by which the listener or reader is unable to establish intertextual linkage between comments. As an example, in LMN's journals, her links from a current comment with one from the past will frequently be noted with the idiosyncratic abbreviation *SPJ*, the

comprehension of which requires familiarity with her system of notation. As the journals were ostensibly prepared for review and research purposes, her lack of consistent gloss for such terms suggests a presupposition of understanding by the reader. Parenthetically, the proposed use of this construct might be incompatible with Alverson and Rosenberg (1990) who argue that the specific breakdown in establishing such linkage may not rest solely with the individual diagnosed with mental illness, but hypothesize that discourse failure may instead lie in the hearer's interpretation of language rather than in the speaker's formulation.

II. Recurrent patterns noted in repeated use of specific words and phrases

Based on the fundamental principles of corpus linguistic methodology and language variation theory, language is best analyzed in large quantities of authentic text through which repeated patterns of lexical selection and collocation can be identified. Application of this model in clinical practice as an assessment tool would allow for examination of large corpora of spoken or written text, increasing the sensitivity of differential diagnostic formulation through identification of recurrent patterns that can be appropriately appreciated.

A. Idea generation. The DSM-IV (1994) notes the increased occurrence of grandiose idea formulation as one of the historically recognized pathognomic processes in mania. In a state of increased generativity, these ideas may involve small scale projects for personal change, but are typically observed to involve grander designs which would engender favorable recognition or cast the individual into the spotlight. A more appropriate formulation of this indicator in the diagnosis of mania,

however, would acknowledge the *qualitative* aspects evidenced in the process of grandiose idea generation, while attaching equal importance to the overall rate of idea generation, noting the significance of the *quantity* of such expressions. LMN's frequent use of *idea,* and analogous productions by other individuals with mania, would best be identified through corpus linguistic methodology.

The recurrent theme of idea generation is observed as a consistent linguistic token in LMN's writing, appearing as one of her most frequently occurring words, and frequently collocating with qualifiers which speak to the positive discourse prosody of her journals (e.g., "a very good idea", "a brilliant idea"). For LMN, the ideas were recurrent, although the scale of the ideas generated was variable. She wrote of everyday ideas (e.g., painting the living room), as well as expressing confidence of the success of research ideas (e.g., "a Nobel-prize worthy idea"). Notably, her journals served as the medium through which she could express the gestational process of ideas, many of which were discarded along the way. Some of her ideas, however, did bear fruit including founding an organized series of national meetings on mathematical models in her particular area of scientific research.

B. Perseveration. In addition to the fixation on ideas, the process of perseveration on other topics is also evident in the patterns of language use of persons with mania. This typically involves repeated expression of a single theme or incorporation of a single word, the patterning of which resembles the idiosyncratic language and mental preoccupations noted in the DSM-IV (1994). Perseveration is a clinical symptom that is consistent with many diagnostic classifications, including behavioral (e.g., obsessive-compulsive disorder) as well as organic (e.g., frontal lobe

injury) pathologies. Although it has not historically been included in the criteria for mania, the manifestations of repeated use of particular words or phrases, or the continued observation of certain linguistically-mediated topics of interest is consistent with the overall diagnostic picture, and is hence worthy of inclusion. For LMN, this repeated return to topic centered around recognition and appreciation of prime numbers as they occurred in her notations of time as she recorded hour:minute:second tags with each journal entry.

III. Specific linguistic indicators of mania

Several lines of research have addressed the prevalence of particular aspects of language production as assessment tools, primarily in terms of either describing specific diagnostic groups or differentiating similar disorders. An inherent problem with this method of analysis deals with the context in which language is expressed, particularly noted in terms of individual variation in intelligence, comorbid communication disorders (e.g., hearing loss), or the impact of exposure to a second language. Although none of these analyses were conducted on the LMN corpus in the current study, these tools provide another perspective for fine-grained analysis of language use in mania.

A. Lexical fixations. Several studies have addressed specific use of linguistic forms in an attempt to identify patterns which may be consistent with particular diagnoses. The first study of significance into language in mania (Lorenz & Cobb, 1952) approached linguistic analysis from this perspective, and although their samples were inadequate by current standards, specific patterns were noted including

increased use of pronouns and main verbs, with fewer adjectives and prepositions. Andreasen and Pfohl (1976) employed the same investigative criteria, and reported increased use of nouns, verbs, and adjectives, an outcome that differed from the previous research, but resulted in what they described as more "colorful" language, a term that contributes nothing quantifiable to the process. Wykes and Leff (1982) noted that patients diagnosed with mania tended to produce more cohesive ties than did patients with other diagnoses. Other researchers, however, have noted no substantial linguistic differences between groups of patients experiencing more severe manifestations of disturbance (Barch & Berenbaum, 1997; Docherty et al., 1996; Harrow, et al., 1986). This study revealed no specific syntactic markers or other formalist linguistic constructs indicative of mania, but instead, demonstrated patterns of lexical fixation (i.e., *prime, smile*) which may be useful in diagnostic formulation.

B. Impaired discourse cohesion and coherence. The ability to establish linkage within context is fundamental to functional discourse, and this has long been recognized as a specific area of breakdown in the patterns of language use in a number of mental illnesses. Durbin & Martin (1977) identified particular patterns of breakdown in elliptical cohesion in sentences, and in the appropriate use of anaphor at the level of discourse, noted as participants with mania had difficulty creating meaningful relationships between topics, with remarkable digression from topic at hand. Bartolucci and Fine (1987) reported that the conversational cohesive weakness observed in mental illness was not related to verbal intelligence, nor was it representative of a cognitive deficit. They argued that such a breakdown was instead a defective signaling process, in which continuity of meaning was not made plain to

listeners, again lending further support to the construct of speech which is difficult to interpret. The difficulty in application of this construct is based on the decontextualized nature of language assessment (Alverson & Rosenberg, 1990).

One of the more parsimonious models for addressing the constructs of cohesion and coherence in corpus linguistic study is that of Stubbs (2002). He describes cohesion as the linguistic devices formally linking utterances, which involves repetition of words, phrases, and their synonyms across a short span of text, for which connections are built by repetition of the same conceptual structure. On a more abstract level, coherence is the ability to infer meaning of a text from the context of common-sense knowledge. Such background knowledge is based on patterns of collocation and the inherent discourse prosodies that provide a framework within which a positive or negative evaluation of words across texts is established. The patterns of interrupted discourse as manifest in LMN's journals do not provide for the continuity necessary for repetition of words and collocations that would be otherwise interpreted as coherence and cohesion, although this model has potential for future research.

IV. Other linguistically-mediated processes in mania

A. Elevated mood. One of the clinical signs of mania involves elevated or expansive mood states, for which linguistic representations may be observed as expressions in happiness or mirth (Fenichel, 1945); euphoria (Durbin & Martin, 1977); and playful or flippant responses (Daniels, et al., 1988). LMN referred to this

in her journals as her “happy mode”, consistent with her positive discourse prosody, as noted:

25 Apr 97

3:51:43 SMILE PRIME! Yes, it is prime.

I have FELT GOOD FOR A WEEK OR SO. Happy. Content. I smile "to myself" a lot. Again, I feel good. Repeat. I feel good.

3:53:17 smile prime 3:53:26

B. Increased energy. For individuals with mania, another of the hallmarks of the advent of an episode cycle is a decreased requirement for sleep, often noted by awakening in the middle of the night, and reports of increased stamina and energy. The linguistically-mediated expressions of this clinical sign may involve increased generativity in written expression, increased volume of writing, evidence of writing during periods of time typically designated for sleep, or verbal descriptions of increasing amounts of energy and initiation. The literature also notes this as hyperactivity (Durbin & Martin, 1977), and increased potential for engagement in goal-directed or pleasurable activity (DSM-IV, 1994). For LMN, this level of increased energy was noted repeatedly in her use of the phrase, “It’s three o’clock in the morning”, a time period remarkable for decreased need for sleep and increased written productivity.

The literature on patterns of language use in mania commented on two other linguistic analyses which are not included in this model: phonological variation and type/token ratio. Regarding phonology, observation of the rules governing sound

patterning in language typically is not impaired with a diagnosis of mania (Durbin & Martin, 1977). This preserved appreciation of the rules is noted even in the presence of associational chaining in which words are selected based on sounds, rather than appropriate semantic content, or in the presence of neologisms in glossomania as noted in diagnoses of other mental illnesses (Chaika, 1990). Type/token ratio has been implemented in analysis of language use in mania and other mental illnesses, with results indicating fewer word types (Lorenz & Cobb, 1952) and a higher type/token ratio in patients with mania as compared to patients with depression (Andreasen & Pfohl, 1976). The clinical utility of this measure is limited in corpus linguistic studies, however, as increased corpus size tends to neutralize the significance of differences yielded by this ratio, and thus it was not incorporated into the current analyses.

Applied corpus linguistics

As noted, corpus linguistic methodology provides an empirical approach to the analysis of patterns of language use, and in combination with other diagnostic tools and descriptors could serve as the basis for better clinical observation of function. Collection and analysis of texts allows for a larger-scope perspective than would be available in a single diagnostic session, and incorporation of such text analysis in the assessment process would significantly enhance formulation of accurate diagnoses. In practical application, this may involve transcription and analysis of extant texts if an individual under evaluation already keeps journals, or directing the person to record thoughts and impressions across a span of time (e.g.,

writing at least one page per day for two weeks) for inclusion in a formal linguistic analysis. Comparison to a reference corpus (e.g., FROWN) would allow for analysis of differences as measured against a benchmark of language use considered to be 'typical'. Further, if the texts of a sufficient number of patients with a specific diagnosis were compiled and subjected to corpus analysis, these findings could theoretically comprise a representative reference corpus to which other patients could be compared to assist with categorical diagnosis.

In addition to clinical assessment, corpus linguistics has applications for documenting improvement in the therapeutic intervention milieu, as well. This study has demonstrated the utility of corpus linguistic methodology in the analysis of patterns of language use in a personal journal, providing evidence of detectible changes in response to specific treatment (i.e., medication). This suggests the potential for assisting with clinical monitoring, titration of treatment, and determining level of care based on serial re-analysis.

Recording thoughts in journals is a device with a long history in psychotherapy (Schab, 2005). The use of unstructured writing produced as stream of consciousness has been accepted as beneficial and efficacious, but there is also an increasing reliance on more structured uses of journal-keeping to target specific emotional and behavioral intervention (e.g., formulation of letters to self, exploration of other points of view). Such interventions are designed in a hierarchical manner to allow for self-exploration of critical issues in a format that allows for creative expression. The application of corpus linguistic analysis to changing patterns of language use across time in therapeutic journal keeping would provide increased

sensitivity to monitoring of critical changes in mental state as reflected in changes in linguistic behavior.

Limitations of the current study

The limitations of this study are those inherent to any work involving $n = 1$: the results cannot necessarily be extrapolated to the population experiencing mania, nor to the population at large. However, there is a lengthy and respected record of case studies in research (e.g., linguistics, medicine, psychology), and the unique opportunity afforded by the post-hoc analysis gleaned from 28 years of data likely will not be replicated in the literature. Using a single participant like LMN means that the full range of manifestations of mania were not observed, and although she has experienced some severe exacerbations, she has been extremely functional and successful in her life, which is unfortunately not the case for most individuals who are diagnosed with this particular disorder. Other manifestations of mania in less well-adjusted individuals result in extremes of behavior, reckless spending, wanton disregard of rules, stimulus seeking, poor insight, and impaired judgment which may result in injury, broken relationships, incarceration, or death.

Another limitation to the study is that of the observer's paradox, in which the experience of writing in journals has changed because it is being observed. Although LMN was ostensibly preparing all of her journals for investigation, the tenor of her writing changed around the time when the researchers became involved, as she became more acutely aware of the potential for analysis, and included thoughts on

research, the writing process, and even comments to the researcher, all of which were different from the texts produced prior to 2003.

At the level of specific analyses included in this study, the limitations and opportunities for improvement are many. First, the research would benefit from increasing the size of the Master Corpus to include more of the available texts for analysis. The use of FROWN as a reference corpus was appropriate based on size, genre and synchronicity, but inclusion of other corpora which involve larger data sets and more divergent time constructs would provide a different perspective on this comparative analysis. The flaws in the journal-as-genre comparison were evident, as the reference sample was small, and was not representative of the same time period, although the observation of the genre made this a point of comparison worth addressing. Further such analysis should include a larger sample of texts from time periods consistent with those during which LMN was writing.

The intra-individual analyses could also benefit from further re-examination, as the comparison of medicated versus unmedicated conditions revealed remarkable differences as structured, but would profit from both increased corpus size in each condition, and review of potential confounding differences between the early and late unmedicated conditions. Analysis of variance between the unmedicated and medicated conditions may be instructive, revisiting the A_1-B-A_2 model. The analysis of samples of writing produced during manic episodes versus non-manic periods compared different genres of texts, which potentially confounded the investigation, as the stylistic variation was different enough between the two corpora that finding may be perceived as specious. However, the same individual produced both sets of texts,

and the results indicated that she was observant of the expected shifts in register as genre changed, which in fact may have been a difference related to her cycle of mania.

Examination of variation in language based on severity of exacerbation was limited by the sampling technique and dissimilar sizes of the corpora, and would benefit from increased corpus size including transcription of additional texts from the time period of the more acute experience. The final investigation of variations as observed within a single episode would have gained precision with a count of words per episode so that closer analysis could be made of the changes in patterns of writing as the episode progressed.

Future directions

In consideration of further directions for this line of research, from a fundamental linguistic perspective, a deeper analysis of the structure of the language used in the journals is indicated, as there is variation in the syntactic structure of the texts between conditions as explored. There is also potential for investigation of the stylistic and specific linguistic changes noted relative to the introduction of observers to the research process. In addition, tagging of the data for parts of speech and other linguistic markers would allow for quantitative analysis of topical shifts in her writing, predicted to manifest as language-based behaviors specific to the diagnosis of mania. Brown (2004) and He (2006) present useful models for rating idea density in corpus linguistics analysis.

Although the transcripts do not reflect these, analysis of the extralinguistic features LMN included in her work may provide some insight into her process, as she

included a series of sketches and drawings that appear to conform to a closed set of patterns. There are also curious instances of "circular" writing, in which LMN would continue writing in sentence form but would change the direction of the paper, generating writing that moved across the page in inward-spiraling concentric circles. She has indicated that this pattern seems to be related to her cognitive processing of an idea. There are also notable changes in the size and shape of her handwriting across the course of an episode.

From the perspective of gaining a deeper understanding of mania through patterns of writing, continued analyses of the journal-as-genre research has been proposed, involving collection of the journals of a cohort of women of similar age and background as LMN who have not been diagnosed with mania. Comparisons could be made of patterns of language used, observation of the structure inherent to the genre, cohesion markers and frequency of topical shifts, in an effort to delineate patterns specific to the diagnosis and genre. Another reference corpus for such comparative analyses would be the diaries and journals of people who have been diagnosed with other mental illnesses, looking for patterns which would offer insights into differential diagnosis. Additionally, careful review of the timelines of the journals, in combination with analysis of frequency counts for certain words suggests patterns in the episodic nature of LMN's experience with mania related to certain months of the year which appear in her writing with greater frequency. It may be that occurrences are related to seasonal variables such as avoiding extremes of weather in summer or winter, but preliminary review indicates a potential bimodal distribution to her patterns, as the summer months of June, July and August and winter months of

December and January seem to appear in greater frequency. Although such observations are inconclusive at this point, and are admittedly subject to the vagaries of outliers (e.g., the extremely productive episodes in 1985), seasonal patterns of variation are not unexpected concomitant with this diagnosis.

From the corpus linguistics angle, this research has introduced the potential confound of the observer's paradox, in which the introduction of an observer changes the subject being observed, and investigation of this phenomenon and the influence and effects on corpus studies in general is needed. The identification of a specific audience, particularly for the purpose of scientific investigation, has implications for the interpretation of contextual data in transcriptions of spoken and written texts, particularly the fundamental requirement that the language under analysis should be authentic and naturally-occurring.

Finally, a definitive direction for further investigation as instigated by this research deals with the need for re-examination and refining of the linguistic terminology used to differentiate diagnoses of mental illness. The accepted argot for such classifications in the mental health professions is imprecise and open for interpretation, and the danger inherent in such a system is that clinical diagnoses and treatment protocols are based on these interpretations.

Conclusions

For LMN, demystifying the stigma of mental illness has been one of the fundamental driving motivations for the habit of keeping a journal. In the court of public opinion, a diagnosis of mental illness, no matter how debilitating, is never

accorded the same level of understanding or sympathy as physical illness. The perception seems to be that the behavior or language use as manifest in mental illness is a matter of individual choice or character flaw, and given the appropriate boot-straps or other such motivation, one could just straighten up.

Careful chronological review of LMN's record of her experience provides a detailed analysis of the progression of her thought processes. In the earliest writings, she was exploring ideas related to mathematical models of her research, with copious and detailed notes, some of which she refines, and most of which she discards. It is apparent that she was using the journals as the medium for recording her thoughts as they occurred; a process that is undertaken by most people in a less structured manner (i.e., a thought occurs, and if it is important, it is written down as a cue to remember it). Some of the ideas cultivated in the journals were brought to fruition; notably, the first of a series of national meetings on mathematical models in her area of research, a symposium that is now in its tenth iteration (as of 2006). The texts are replete with ideas for research projects, obtaining grant funding, and establishment of organizations for various purposes, all of which were situated in positive discourse prosodies, with the intention of making her corner of the world a better place.

A chronological review of the texts also reveals her personal process of evolving insight in understanding her cycles of energy and creativity. Over time, she appreciated that there are repeated patterns of behavior which began to afford her some predictability in her process (e.g., noting that she drinks more coffee, spends more money, experiences neurovegetative disturbances, and writes in journals when she is entering a manic episode). LMN also documented the changes in her

relationships and interactions with others, including her struggle to be accepted as a professional in her field, and not just someone who could be summarily categorized as "insane". As Frow (2001) noted, it may be that the diagnostic classifications of mental illness are merely convenient metaphors applied to persons who do not fit the profile of what is considered to be typical in a community of practice (also cf. Ribeiro, 1992).

From the vantage point of a chronological review of nearly thirty years of data, it can be concluded that people with mania are people without customary trepidation. The habits, beliefs and social norms that constrain most people typically disallow exploration of all the possibilities of a circumstance. LMN's writings suggest that she was unencumbered by the internal filters that for most people foretell defeat or judge an idea as too far-fetched. At points in her life, LMN explored ideas involving contacting the President of the United States, appearing on television, establishing research institutes, publishing books, founding organizations, and making movies, and as noted, her idea generation has not been without some success.

Not unexpectedly, however, there is another side of the coin. People with mania are frequently recognized as being stimulus seekers, without insight into the consequences of their actions, and without appreciation of the violation of social pragmatics. These individuals may appear unrestrained and fearless in all aspects of their lives, and may engage in behaviors which may have more significant repercussions; for example impulsively spending money or engaging in risky sexual encounters. The challenge for the individual with mania and for the healthcare providers and significant others with whom they are involved is to strike a balance.

In many cases this is found in medication, psychotherapy, environmental modification, family involvement, or some combination thereof. For LMN, it has also taken the form of recording her experiences, documenting her process, and making meaning of her experience with mania, a process through which she is soothed and satisfied. The consistently positive approach with which she faces the challenges of mania is evident in her over-arching goal of preparing the journals and donating them for further investigation.

In addition to the exploration of LMN's patterns of language use in mania, this dissertation has also addressed practical applications of linguistic theory in a clinical model. The study of linguistics has historically addressed language as a system outside the individual, governed by universals which allow for only minor variations on the theme of acquisition and use. In such a model there is no accommodation for intra-individual variation, as has been observed in the writing produced by LMN. This research has achieved a different perspective, examining language that falls within individual variation, in this case representative of mania. Investigation into patterns of language use allowed for identification of changes in a single subject based on differing treatment conditions and exacerbations of symptoms, as well as comparisons to standards of what are considered 'typical' language use, as measured by a reference corpus of typical American English writing, a genre-consistent journal corpus, and an intra-individual corpus of language produced during non-manic periods. The results supported the notion that language is not monolithic, but that, for purposes of diagnosis of particular behavior by particular people, it is best understood at the level of the individual.

Although the DSM-IV is widely accepted as the benchmark for diagnosing mental illness, a better delineated and more nuanced observation of the patterns of language use would enhance diagnostic accuracy at the level of individual function in context. In order to gain this sensitivity, a perspective differing in nature and scope from the more traditional approaches to language is required. The most typical approach to language analysis involves parsing parts of speech and identification of specific markers which can be counted and measured. However, these cannot account for the more difficult-to-measure constructs of functional language use and the appropriacy of contextual circumstance. Such analysis, the study of linguistic pragmatics, deals with the occurrence of functional linguistic acts, rather than the syntactic, morphological, or other structural variations. In mania, it is the language behavior or the use in context that is the indicator of pathology, rather than the particular words used or the structure of the sentences. For LMN, this was observed through recurrent configurations of word in collocation, content analysis, and discourse prosodies which allowed for identification of particular patterns of use that were consistent with the diagnosis.

LMN's language use unequivocally meets the criteria for the diagnosis of mania as defined by the DSM-IV, but those defining features have been demonstrated to be too general and inclusive. The linguistic behaviors described in the DSM-IV are underspecified for identification of exact patterns that serve the purpose of differential diagnosis, particularly when considering the spectrum of severity along which mania may manifest. As noted, pragmatic language variation is based on functional use in context, and when such patterns of use reach extremes that violate

conventional boundaries of behavior considered to be appropriate or acceptable, a diagnosis of pathology is made.

Because it is within the patterns of individual variation that a diagnosis of mania is made, there is no 'manic language' among the general population that can be isolated and identified. Mania is a property of the individual, and it is through observation of the language behaviors that such a diagnosis is applied, within the context of other non-linguistic manifestations. The patterns of language use correspond with the criteria involving both the global linguistic *processes* as noted in the DSM-IV, but also an accommodation for the variations in the *content* of language as identified through individual analysis. One of the goals of the present study is to provide practitioners with a refined set of tools to assist in reaching a diagnosis of mania as derived from patterns of language use. These tools would also provide a means for ongoing monitoring of clinical status and the effects of treatment based on empirical measures of outcome and progress.

The study of language variation involves dynamic examination of changes relative to a number of external influences, including those effected by geographic, political and socio-economic forces. The addition of clinical pathologies to that mix represents a different variation, but one which fits the current methodological theories and practices, including corpus linguistic analysis. The knowledge and expertise from linguistics combined with that of the fields of psychology, speech-language pathology, and communication science can result in construction of a bridge between these specialized disciplines, allowing for more accurate descriptions of linguistic phenomena, greater sensitivity and insight into the cognitive and psychological

manifestations of disorders, more efficacious diagnosis and intervention strategies for individuals experiencing mental illness, and on a more global scale, better understanding and appreciation of the parameters of mental illness.

References

Akiskal, H. S., Hantouche, E. G., Bourgeois, M. L., Azorin, J. M., Sechter, D., Allilaire, J. F., et al. (2001). Toward a refined phenomenology of mania: Combining clinician-assessment and self-report in the French EPIMAN study. *Journal of Affective Disorders, 67*(1), 89-96.

Altman, E. G. (1998). Rating scales for mania: Is self-rating reliable? *Journal of Affective Disorders, 50*(2), 283-286.

Altman, E. G., Hedeker, D., Peterson, J. L., & Davis, J. M. (1997). The Altman Self-Rating Mania Scale. *Biological Psychiatry, 42*(10), 948-955.

Altman, E. G., Hedeker, D., Peterson, J. L., & Davis, J. M. (2001). A comparative evaluation of three self-rating scales for acute mania. *Biological Psychiatry, 50*(6), 468-471.

Alverson, H., & Rosenberg, S. (1990). Discourse analysis of schizophrenic speech: A critique and proposal. *Applied Psycholinguistics, 11*(2), 167-184.

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: Author.

American Speech-Language-Hearing Association. (2001). *Scope of practice in speech-language pathology*. Rockville, MD: Author.

Andor, J. (2004). The master and his performance: An interview with Noam Chomsky. *Intercultural Pragmatics, 1*, 93-112.

Andreasen, N. C. (1979). Thought, language, and communication disorders: II. Diagnostic significance. *Archives of General Psychiatry, 36*(12), 1325-1330.

Andreasen, N. C. (1986). Scale for the assessment of thought, language, and communication (TLC). *Schizophrenia Bulletin, 12*(3), 473-482.

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

Andreasen, N. C., & Pfohl, B. (1976). Linguistic analysis of speech in affective disorders. *Archives of General Psychiatry, 33*(11), 1361-1367.

Andreasen, N. C., & Powers, P. S. (1975). Creativity and psychosis: An examination of conceptual style. *Archives of General Psychiatry, 32*(1), 70-73.Altman, E. G. (1998). Rating scales for mania: Is self-rating reliable? *Journal of Affective Disorders, 50*(2), 283-286.

Barch, D. M., & Berenbaum, H. (1996). Language production and thought disorder in schizophrenia. *Journal of Abnormal Psychology, 105*, 81-86.

Barch, D. M., & Berenbaum, H. (1997). Language generation in schizophrenia and mania: The relationships among verbosity, syntactic complexity, and pausing. *Journal of Psycholinguistic Research, 26*(4), 401-412.

Barnard, C. I. (1918). *Paris war days: diary of an American*. Retrieved 04-06-2006, from http://www.gutenberg.org/dirs/etext06/7prwr10.txt.

Barr, W. B., Bilder, R. M., Goldberg, E., Kaplan, E., & Mukherjee, S. (1989). The neuropsychology of schizophrenic speech. *Journal of Communication Disorders, 22*(5), 327-349.

Bartolucci, G., & Fine, J. (1987). The frequency of cohesion weakness in psychiatric syndromes. *Applied Psycholinguistics, 8*(1), 67-74.

Bazin, N., Sarfati, Y., Lefrere, F., Passerieux, C., & Hardy-Bayle, M.-C. (2005). Scale for the evaluation of communication disorders in patients with schizophrenia: A validation study. *Schizophrenia Research, 77*(1), 75-84.

Billow, R. M., Rossman, J., Lewis, N., Goldman, D., & Raps, C. (1997). Observing expressive and deviant language in schizophrenia. *Metaphor and Symbol, 12*, 205-216.

Brown, B., Nolan, P., Crawford, P., & Lewis, A. (1996). Interaction, language and the "narrative turn" in psychotherapy and psychiatry. *Social Science & Medicine (1982), 43*(11 (Print)), 1569-1578.

Brown, C. (2004). *The computational idea density rater.* Unpublished manuscript.

Carreno, T., & Goodnick, P. J. (1998). Creativity and Mood Disorder. In P. J. Goodnick (Ed.), *Mania: Clinical and Research Perspectives*. Washington, DC: American Psychiatric Press, Inc.

Casey, B., & Long, A. (2003). Meanings of madness: a literature review. *Journal of Psychiatric and Mental Health Nursing, 10*(1 (Print)), 89-99.

Ceccherini-Nelli, A., & Crow, T. J. (2003). Disintegration of the components of language as the path to a revision of Bleuler's and Schneider's concepts of schizophrenia. *British Journal of Psychiatry, 182*(233-240).

Chaika, E. O. (1990). *Understanding psychotic speech: Beyond Freud and Chomsky*. Springfield, IL: Charles C. Thomas, Publisher.

Chaika, E., & Lambe, R. A. (1989). Cohesion in schizophrenic narratives, revisited. *Journal of Communication Disorders, 22*(6), 407-421.

Chambers, J. K. (2003). *Sociolinguistic theory: Linguistic variation and its social significance* (2nd ed. Vol. 22). Oxford: Blackwell Publishers.

Chaplin, J. P. (1985). *Dictionary of Psychology* (2nd Revised ed.). New York: Laurel/Dell Publishing.

Chen, E. Y. H., Lam, L. C. W., Kan, C. S., Chan, C. K. Y., Kwok, C. L., Nguyen, D. G. H., et al. (1996). Language disorganisation in schizophrenia: Validation and assessment with a new clinical rating instrument. *Hong Kong Journal of Psychiatry, 6*(1), 4-13.

Connor-Linton, J. (2003). Georgetown Linguistics Chi Square Tutorial. Retrieved April 17, 2006, from http://www.georgetown.edu/faculty/ballc/webtools/web_chi_tut.html

Corcoran, R., & Frith, C. D. (1996). Conversational conduct and the symptoms of schizophrenia. *Cognitive Neuropsychiatry, 1*(4), 305-318.

Covington, M. A., He, C., Brown, C., Naci, L., McClain, J. T., Fjordbak, B. S., et al. (2005). Schizophrenia and the structure of language: The linguist's view. *Schizophrenia Research, 77*(1), 85-98.

Covington, M. A., Riedel, W. J., Brown, C., He, C., Morris, E., Weinstein, S., et al. (2005). Does ketamine mimic aspects of schizophrenic speech? *Manuscript submitted for publication.*

Daniels, E. K., Shenton, M. E., Holzman, P. S., Benowitz, L. I., Coleman, M., Levin, S., et al. (1988). Patterns of thought disorder associated with right cortical

damage, schizophrenia, and mania. *American Journal of Psychiatry, 145*(8), 944-949.

Dell'Osso, L., Pini, S., Tundo, A., Sarno, N., Musetti, L., & Cassano, G. B. (2000). Clinical characteristics of mania, mixed mania, and bipolar depression with psychotic features. *Comprehensive Psychiatry, 41*(4), 242-247.

Depp, C. A., & Jeste, D. V. (2004). Bipolar disorder in older adults: A critical review. *Bipolar Disorders, 6*(5), 343-367.

Depp, C. A., Jin, H., Mohamed, S., Kaskow, J., Moore, D. J., & Jeste, D. V. (2004). Bipolar Disorder in Middle-Aged and Elderly Adults: Is Age of Onset Important? *Journal of Nervous and Mental Disease, 192*(11), 796-799.

Docherty, N. M., DeRosa, M., & Andreasen, N. C. (1996). Communication disturbances in schizophrenia and mania. *Archives of General Psychiatry, 53*(4), 358-364.

Durbin, M., & Martin, R. L. (1977). Speech in mania: Syntactic aspects. *Brain and Language, 4*(2), 208-218.

Fenichel, O. (1945). *The psychoanalytic theory of neurosis*. New York: WW Norton & Company.

Ferguson, A. (2000). Maximising communication effectiveness. In N. Muller (Ed.), *Pragmatics in speech and language pathology: Studies in clinical applications* (pp. 53-88). Amsterdam: John Benjamin Publishing Co.

Fine, J. (1995). Towards understanding and studying cohesion in schizophrenic speech. *Applied Psycholinguistics, 16*, 25-41.

Fleck, D. E., Shear, P. K., Zimmerman, M. E., Getz, G. E., Corey, K. B., Jak, A., et al. (2003). Verbal memory in mania: Effects of clinical state and task requirements. *Bipolar Disorders, 5*(5), 375-380.

Fleiss, J. L. (2003). *The measurement of interrater agreement: Statistical methods for rates and proportions* (Third ed.). New York, NY: John Wiley & Sons, Inc.

Fletcher, A. C. (2001). *Camping with the Sioux: Fieldwork diary of Alice Cunningham Fletcher*. Retrieved 04-06-2006, from http://www.nmnh.si.edu/naa/fletcher/fletcher.htm.

France, J., & Muir, N. (1997). Speech and language therapy and the mentally ill patient. In J. France & N. Muir (Eds.), *Communication and the Mentally Ill Patient* (pp. 87-97). London: Jessica Kingsley Publishers, Ltd.

Frow, J. (2001). Metaphor and metacommunication in schizophrenic language. *Social Semiotics, 11*, 275-287.

Fulwiler, T. (1987). Introduction. In T. Fulwiler (Ed.), *The Journal Book* (pp. 1 - 8). Portsmouth, NH: Boynton/Cook Publishers.

Gannett, C. (1992). *Gender and the Journal: Diaries and Academic Discourse*. Albany: State University of New York Press.

Gavell, R., & France, J. (1991). *Speech and communication problems in psychiatry*. London: Chapman & Hall.

Gennaro, A. R., & Gould, G. M. (Eds.). (1979). *Blakiston's Gould Medical Dictionary* (4th ed.). New York: McGraw-Hill Book Company.

Ghaemi, S. N., & Rosenquist, K. J. (2004). Is Insight in Mania State-Dependent? A Meta-Analysis. *Journal of Nervous and Mental Disease, 192*(11), 771-775.

Goncalves, O. F., Machado, P. P. P., Korman, Y., & Angus, L. (2002). Assessing psychopathology: A narrative approach. In L. E. Beutler & M. L. Malik (Eds.), *Rethinking the DSM: A psychological perspective*. Washington, DC: American Psychological Association.

Greger, M. (1999). *Heart failure: Diary of a third year medical student*. Retrieved 04-06-2006, from http://upalumni.org/medschool/.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3, pp. 41-59). New York: Academic Press.

Grice, J. W. (2004). Bridging the Idiographic-Nomothetic Divide in Ratings of Self and Others on the Big Five. *Journal of Personality, 72*(2), 203-241.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Harden, J. (2000). Language, discourse and the chronotope: applying literary theory to the narratives in health care. *Journal Of Advanced Nursing, 31*(3 (Print)), 506-512.

Hardyck, C. D., & Petrinovich, L. F. (1969). *Introduction to statistics for the behavioral sciences*. Philadelphia: W.B. Saunders Company.

Harris, J. (1989). The idea of community in the study of writing. *College Composition and Communication, 40*(1), 11-22.

Harrow, M., Grossman, L. S., Silverstein, M. L., Meltzer, H. Y., & Kettering, R. L. (1986). A longitudinal study of thought disorder in manic patients. *Archives of General Psychiatry, 43*, 781-785.

He, C. (2006). *Using text analysis software in schizophrenia research.* Unpublished doctoral dissertation, University of Georgia, Athens.

Hershman, D. J., & Lieb, J. (1998). *Manic Depression and Creativity*. Amherst, NY: Prometheus Books.

Hockey, S. (2000). *Electronic Texts in the Humanities: Principles and Practice*. New York: Oxford University Press.

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

Holtzman, W. H., Thorpe, J. S., Swartz, J. D., & Herron, J. W. (1961). *Inkblot perception and personality*. Austin, TX: University of Texas Press.

Houts, A. C. (2002). Discovery, invention, and the expansion of the modern *Diagnostic and statistical manuals of mental disorders*. In L. E. Beutler & M. L. Malik (Eds.), *Rethinking the DSM: A psychological perspective*. Washington, DC: American Psychological Association.

Johnston, M. H., & Holzman, P. S. (1979). *Assessing schizophrenic thinking*. San Francisco: Jossey-Bass.

Kaplan, H. I., & Sadock, B. J. (1998). *Synopsis of psychiatry* (8th ed.). Philadelphia: Lippincott Williams & Wilkins.

Katzung, B. G. (Ed.). (1995). *Basic and clinical pharmacology* (6th ed.). Norwalk, CT: Appleton & Lange.

Khadivi, A., Wetzler, S., & Wilson, A. (1997). Manic Indices on the Rorschach. *Journal of Personality Assessment, 69*(2), 365.

Kramer, S., Bryan, K. L., & Frith, C. D. (1998). A framework for assessing communication in mental illness. *International Journal of Language & Communication Disorders, 33*, 164-169.

Kramer, S., Bryan, K. L., & Frith, C. D. (2001). Mental illness and communication. *International Journal of Language & Communication Disorders, 36*, 132-137.

Kretzschmar, W.A., Jr. (in preparation). *On parole: The linguistics of speech.*

Kretzschmar, W. A., Jr. (in press). [Review of the book by Sampson (2001) *Empirical Linguistics*]. *Journal of English Linguistics*.

Kretzschmar, W. A., Jr., Meyer, C. F., & Ingegneri, D. (1994). Uses of inferential statistics in corpus studies. In M. Ljung (Ed.), Corpus-based studies in English (pp. 167 - 177). Amsterdam: Rodopi.

Langdon, R., Davies, M., & Coltheart, M. (2002). Understanding minds and understanding communicated meanings in schizophrenia. *Mind & Language, 1 & 2*, 68-104.

Leech, G. (1983). *Principles of pragmatics*. New York: Longman.

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Liddle, P. F., Ngan, E. T. C., Caissie, S. L., Anderson, C. M., Bates, A. T., Quested, D. J., et al. (2002). Thought and Language Index: An instrument for assessing thought and language in schizophrenia. *British Journal of Psychiatry, 181*(4), 326-330.

Lorenz, M., & Cobb, S. (1952). Language behavior in manic patients. *Archives of Neurology & Psychiatry (Chicago), 67*, 763-770.

Lowenstein, S. (1987). A brief history of journal keeping. In T. Fulwiler (Ed.), *The Journal Book* (pp. 87 - 97). Portsmouth, NH: Boynton/Cook Publishers.

Macrorie, K. (1987). Foreword. In T. Fulwiler (Ed.), *The Journal Book* (pp. i - iii). Portsmouth, NH: Boynton/Cook Publishers.

McAdams, D. P., Anyidoho, N. A., Brown, C., Huang, Y. T., Kaplan, B., & Machado, M. A. (2004). Traits and stories: links between dispositional and narrative features of personality. *Journal Of Personality, 72*(4 (Print)), 761-784.

McKenna, P., & Oh, T. (2005). *Schizophrenic speech*. Cambridge: Cambridge University Press.

McPherson, L., & Harvey, P. D. (1996). Discourse connectedness in manic and schizophrenic patients: Associations with derailment and other clinical thought disorders. *Cognitive Neuropsychiatry, 1*(1), 41-53.

Mendhekar, D. N., Srivastav, P. K., Jiloha, R. C., & Awana, S. (2004). Chronic but not resistant mania: A case report. *Acta Psychiatrica Scandinavica, 109*(2), 147-149.

Mentis, M., & Thompson, S. A. (1991). Discourse: A means for understanding normal and disordered language. In T. M. Gallagher (Ed.), *Pragmatics of language* (pp. 199-227). San Diego: Singular Publishing Group.

Meyer, C. F. (2002). English Corpus Linguistics. Cambridge: Cambridge University Press.

Molenaar, P. C. M., & Valsiner, J. (2005). How generalization works through the single case: A simple idiographic process analysis of an individual psychotherapy [Electronic Version]. International Journal of Idiographic

Science, 1. Retrieved February 14, 2006 from www.valsiner.com/articles/molenvals.htm.

Murray, H. A. (1943). *Thematic Apperception Test manual*. Cambridge, MA: Harvard University Press.

Murray, J. B. (1985). Lithium therapy for mania and depression. *Journal of General Psychology, 112*(1), 5-33.

Ochs, E., & Schieffelin, B. B. (Eds.). (1979). *Developmental pragmatics*. New York: Academic Press.

Peuser, G. (1987). Jargonaphasia and schizophasia: An essay in contrastive patholinguistics. In R. Wodak & P. vandeCraen (Eds.), *Neurotic and psychotic language behaviors*. Cleveland, OH: Multilingual Matters, Ltd.

Prutting, C. A., & Kirchner, D. M. (1983). Applied pragmatics. In T. M. Gallagher & C. A. Prutting (Eds.), *Pragmatic assessment and intervention issues in language* (pp. 29-64). San Diego: College-Hill Press.

Prutting, C. A., & Kirchner, D. M. (1987). A clinical appraisal of the pragmatic aspects of language. *Journal of Speech and Hearing Disorders, 52*, 105-119.

Reilly, F. (1975). Looseness of associations in acute schizophrenia. *British Journal of Psychiatry, 127*, 240-246.

Ribeiro, B. T. (1994). *Coherence in psychotic discourse*. New York: Oxford University Press.

Rieber, R. W., & Vetter, H. J. (1994). *The psychopathology of language and cognition*. New York: Plenum Press.

Rochester, S., & Martin, J. R. (1979). *Crazy talk*. New York: Plenum Press.

Rorschach, H. (1921). *Rorschach*. Kirkland, WA: Hogrefe & Huber Publishers.

Rosenberg, S. D., & Tucker, G. J. (1979). Verbal behavior and schizophrenia. *Archives of General Psychiatry, 36*, 1331-1337.

Rubin, D. L. (1988). Introduction: Four dimensions of social construction in written communication. In B. A. Rafoth & D. L. Rubin (Eds.), *The social construction of written communication* (pp. 1-33). Norwood, NJ: Ablex Publishing Corporation.

Rubin, D. L. (1995). Introduction: Composing social identity. In D. L. Rubin (Ed.), *Composing social identity in written language* (pp. 1-30). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Sanders, L. M., Adams, J., Tager-Flusberg, H., & Shenton, M. E. (1995). A comparison of clinical and linguistic indices of deviance in the verbal discourse of schizophrenics. *Applied Psycholinguistics, 16*(3), 325-338.

Schab, L. M. (2005). Journaling II: Directed exercises in journaling. Retrieved April 19, 2006, from www.pdresources.org

Scott, M. (2005). Oxford Wordsmith Tools (Version 4.0): Oxford.

Scott, M. (2006). *Word frequency and keyword extraction: In search of a bad reference corpus.* Paper presented at the AHRC ICT Methods Network Expert Seminar on Linguistics, Lancaster University, UK.

Shulman, K. I. (1997). Disinhibition syndromes, secondary mania and bipolar disorder in old age. *Journal of Affective Disorders, 46*(3), 175-182.

Smith, B. R., & Leinonen, E. (1992). *Clinical pragmatics: unraveling the complexities of communicative failure.* London: Chapman & Hall.

Sperling, M. (1993). *The social nature of written text: a research-based review and summary of conceptual issues in the teaching of writing.* Urbana, IL: National Council of Teachers of English.

Stillman, P. (1987). "Of Myself, for Myself". In T. Fulwiler (Ed.), *The Journal Book* (pp. 77 - 86). Portsmouth, NH: Boynton/Cook Publishers.

Stubbs, M. (1996). *Text and Corpus Analysis: Computer-assisted Studies of Language and Culture* (Vol. 23). Oxford: Blackwell Publishers.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell Publishing.

Thomas, P. (1997). What can linguistics tell us about thought disorder? In J. France & N. Muir (Eds.), *Communication and the mentally ill patient: Developmental and linguistic approaches to schizophrenia.* (pp. 30-42). London: Jessica Kingsley Publishers, Ltd.

Thomas, P., & Fraser, W. (1994). Linguistics, human communication and psychiatry. *British Journal of Psychiatry, 165*, 585-592.

Tirassa, M. (1999). Communicative competence and the architecture of the mind/brain. *Brain and Language, 68*(3), 419-441.

Wendelberger, H. (2006). *Hide and seek: Quantitative authorship identification in language concealment.* Unpublished doctoral dissertation, University of Georgia, Athens.

Whitaker, L. C. (1973). *Whitaker Index of Schizophrenic Thinking*. Los Angeles, CA: Western Psychological Services.

Wood, L. A., & Kroger, R. O. (2000). *Doing discourse analysis: Methods for studying action in talk*. Thousand Oaks, CA: Sage Publications.

Wykes, T. (1981). Can the psychiatrist learn from the psycholinguist? Detecting coherence in the disordered speech of manics and schizophrenics. *Psychological Medicine, 11*, 641-642.

Wykes, T., & Leff, J. P. (1982). Disordered speech: Differences between manics and schizophrenics. *Brain and Language, 15*(1), 117-124.

Appendix A

Consent for Participation

I agree to take part in a research study titled "Individual Variation in Patterns of Language Use in a Case of Major Mood Disorder", which is being conducted by Bess Sirmon Fjordbak, Linguistics Program, University of Georgia (404-593-7298), under the direction of William A. Kretzschmar, Jr., Linguistics Program, University of Georgia (706-542-2246). My participation is voluntary; I can stop taking part at any time without giving any reason, and without penalty. I can ask to have information related to me returned to me, removed from the research records, or destroyed.

The purpose of the study is to analyze the patterns of language use as identified in personal texts I have produced. I acknowledge that I have a diagnosis of Bipolar Disorder – Manic Type, and because of this, the journals which I have kept since 1978 may yield some insights into the processes that accompany such a diagnosis. The benefits that I may expect from this study include preparation of my journals into a form for further investigational analysis by future graduate students as well as my own potential creative endeavors which may include, at my discretion, public acknowledgement of my body of work. The greater benefit of the current research project will be to expand the understanding of Bipolar Disorder.

I acknowledge that I initiated the contact with Dr. Kretzschmar and that I have volunteered to make my journals available for investigation. The current research project will involve transcription and investigation of the content of my journals, which I agree to make available in a timely manner. I understand that the transcription of the raw data into a form ready for text and corpus analysis will be completed by paid employees or research assistants who demonstrate an understanding of the sensitive nature of this work

in its original form, and the need for confidentiality. With this in mind, I understand that no undue discomforts, stresses or risks are expected. After the transcription of the journals is completed, the data are compiled into a corpus for analysis, and the current research project (BSF's dissertation) is completed, there may be future opportunities for further exploration into different aspects of the corpus which are not covered by this consent form.

There will be no deception involved in this research process, and I understand that as documents are prepared, including abstracts of papers, manuscript submissions to journals, and drafts of the dissertation, a member check including review and discussion will be done to insure my understanding of and comfort with the process of analysis and description.

I understand that my confidentiality will be protected throughout this research project, and only those actively involved in the research process will be able to identify me. No individually identifying information about me or provided by me during this research will be shared with others, except if necessary to protect my rights or welfare, or if required by law. I agree that, as a part of the scholarly process, descriptions of data and analysis from this research as well as excerpts from the text will be incorporated into scholarly research, including articles for publication in peer-reviewed journals and presentations at professional meetings, understanding that my identity will be kept confidential. I reserve the right to maintain ownership of my original work, and am free to publish this work in any form I choose beyond the scope of this dissertation.

The researcher will answer any further questions about the research, now or during the course of the project, and can be reached by telephone at 404-593-7298 or via e-mail at fjordbak@uga.edu.

My signature below indicated that the researchers have answered all of my questions to my satisfaction, and that I consent to participate in this study. I have been given a copy of this form.

_____	_____	_____
Name of researcher	Signature	Date
e-mail: fjordbak@uga.edu	phone: (404-593-7298)	

_____	_____	_____
Name of researcher	Signature	Date
e-mail: kretzsch@uga.edu	phone: (706-542-2246)	

_____	_____	_____
Name of participant	Signature	Date

Additional questions or problems regarding your rights as a research participant should be addressed to: The Chairperson, Institutional Review Board, University of Georgia, 612 Boyd Graduate Studies Research Center, Athens, Georgia 30602-7411; Telephone (706) 542-3199; E-Mail Address **IRB@uga.edu**

Appendix B

Samples of texts from journals.

From 14 Nov 89:

2:53:53 Smile Prime

Software development: Supports the development of software to facilitate specialized research in biology. (i.e. image analysis).

Computational Biology: Supports the development of new means for solving computational problems unique to biology (i.e., example algorithms for searching databases or understanding nucleotide sequences).

2:57:14

From 12 July 91

2:29:58

I-spot (SPJ) shook to count of 48.

2:30:20

2:37:47 prime. SPJ

Break since last entry. Drank coffee. Feel good. Happy mode.

2:38:25

2:39:52 Another break.

2:40:05

2:45:49

Both arms & legs shook to count of 79. New mode. Slowly. More "mechanical". Happy mode.

2:46:38

2:52:14 am

Relaxed. Fell asleep, almost. Felt very relaxed. Hot. SPJ.

2:53:07 prime

2:53:14

From 31 Jan 96:

2:15 Heat pump awakened me. Asked about Tag Ensemble … from Switzerland

2:17 smile prime 2:17

2:18

Quite tired. I need to sleep.

2:19 prime 2:19

3:03 prime

"It's 3 o'clock in the morning". Storm outside.

3:03

3:23 prime

Storm again. A bit stronger. Sounds good.

3:23 prime 3:23

Appendix C

Journals included in Master Corpus.

Journal Number	Dates	Word Count
188	10 Feb 78 - 14 Feb 78	12784
192	25 Feb 78 - 11 May 79	5652
122	4 Aug 83 - 9 Aug 83	12093
121	20 Nov 83 - 3 Dec 83	11021
154	10 Mar 84 - 19 Mar 84	9327
155	19 Mar 84 - 30 Mar 84	10426
	Total – 1st Time Period	*61,303*
152	24 Jan 85 - 25 Jan 85	4997
179	25 Jan 85 - 27 Jan 85	8482
120	1 Feb 85 - 7 Feb 85	1296
226	3 Jul 85 - 11 Jul 85	7560
227	12 Jul 85 - 15 Jul 85	18380
185	6 Jan 89 – 31 Jan 89	18043
205	13 Nov 89 – 24 Jul 90	11976
161	12 Jul 91 – 3 Nov 91	12570
170	3 July 93 – 23 Nov 93	12352
166	23 Nov 93 - 28 Nov 93	4009
128	19 May 94 - 19 May 94	679
	Total – 2nd Time Period	*100,344*

167	1 Mar 95 – 12 Jul 95	13612
181	12 Jul 95 – 4 Feb 96	16784
146	31 Jan 96 – 7 Jun 96	14782
169	18 July 96 – 29 Nov 96	10609
175	30 Nov 96 – 20 Jul 97	10928
129	3 Apr 97 – 14 Dec 97	10460
137	17 Jan 03 - 3 Feb 03	1386
118	8 Aug 03-20 Jan 04	2381
	Total – 3rd Time Period	*80,942*
	Grand Total	***242,589***

Appendix D

The Observer's Paradox

Patterns of language use vary according to the situational context in which the language is produced, noted in conventional conversational interaction, in writing across genres, as well as in the presence of mental illnesses which are partially identified based on observations of such linguistic variation. Relative to the context, construction of an audience by the speaker or writer requires accommodation for illocutionary intent as well as perlocutionary force, based on the situation and the roles and dynamics between the involved parties. In written texts, however, the habit of keeping a journal is traditionally considered to be a means of recording personal thoughts and feelings, with audience and interaction primarily limited to *self*. The participant at the center of this investigation, LMN, produced a series of journals for which the constructed audience was not obviously defined, but was clearly always comprised of self, and initially, a nebulously defined research audience. This shifted, however, as formal investigation of her work commenced, and her construction of audience took a more specific direction.

Although the texts were written primarily as personal reflections in a journal, these documents represent a complex interaction involving four conditions:

1. LMN as the individual experiencing the phenomenon of mania, and all the emotional, psychological, social and physical manifestations of it.
2. LMN as the writer adopting a more objective and scientific perspective, with the effort being to maintain objectivity in reporting her experience.

3. Bess as the involved researcher, about whom certain aspects of the text were written and to whom some of the comments are specifically addressed.
4. The general audience (including Bess) as non-specific and uninvolved readers.

The question addressed herein is this: Is there a difference in the patterns of language use for a person with mania as manifested in written text between journals which were prepared without a specific external audience in mind, and those for which a target audience was identified?

Literature Review

The phenomenon of change in a behavior occurring as a result of being observed is variously described as the Hawthorne Effect, the Heisenberg Principle, or simply the Observer's Paradox. The Hawthorne Effect is ambiguously defined as behavior elicited by measurement not being equivalent to that noted in the absence of measurement, based on a subject's knowledge of such measurement (Diaper, 1990). More specifically, the Hawthorne Effect has been described as behavior which is altered by a participant's appreciation of the context of an experiment (Jones, 1992), perhaps as a result of response-consequence operant conditioning and reinforcement (Parsons, 1974). Adair (1984) argues that it is in the context of any clearly identified situation that the individual defines his or her behavior and responds accordingly, consistent with construction of audience (Gannett, 1992).

Similarly, the Heisenberg Uncertainty Principle, borrowed from physics, suggests that the observation of a phenomenon will always distort the object of observation (Tranel, 1981). Portrayed more succinctly as two sides of the same coin,

Stubbs (1996, 65-6) argues for the Observer's Paradox: "You cannot observe people when they are not being observed", and by extension, the Describer's Paradox: "You cannot describe people and their behavior without the description changing the behavior". In clinical practice, the presence of an observer has been shown to affect task performance both positively (cf., Borden & Walker, 1978) and negatively (cf., Yantz & McCaffrey, 2005). Kazdin (1982) describes the effect in terms of the observed participant's behaviors accommodating to more socially desirable responses, different patterns of feedback and regulation, and awareness of the potential for evaluation.

Method

The data selected for analysis of the observer's paradox were drawn from two periods of time: the first is from 1983-4 (44,536 words), written at a time prior to LMN's diagnosis with mania (hereafter identified as Without Observer), and the second is from 2004-5 (13,305 words), after initiating the involvement of university researchers (With Observer). These two small corpora were compared for patterns of language use based upon her vocabulary and semantic selection relative to patterns of word frequency and collocation as measured by *Oxford WordSmith Tools* (Scott, 2005), a program designed to analyze patterns of word use in texts. A reference corpus was also employed for the analysis of vocabulary, comparing LMN's patterns under observation with those judged to be representative of *typical* American English language use in written texts from the Freiberg-Brown Corpus of American English (FROWN).

The specific words used for analysis of vocabulary were established *a priori* based on the following linguistic pathognomic markers of mania:

1. References to *idea*, based on the tendency toward grandiose idea generation (DSM-IV, 1994).
2. *SPJ* ("See previous journal") as an idiosyncratic device marking discourse coherence (Durbin & Martin, 1977), through which she would refer to material previously referenced.
3. Numbers and time references, based on the participant's predisposition to document the hour:minute:second at the beginning and end of each journal entry; related to this, her use of *prime* to note the occurrence of prime numbers in the time reference (cf., Khadivi, Wetzler, & Wilson, 1997 for comment on overinclusion).
4. *Research* as it occurs in the context of her own work and in the current investigation, as well as references to *mania* and the name of the researcher involved in the analysis of her journals.

Results/Discussion

Table D.1 presents analysis of the identified target words, comparing proportional percentages of occurrence across the three corpora. Of note, rather than using actual frequency of occurrence, scores are reported in percentages of total (100%) because of the disparity in corpus size.

Table D.1

Percentage of occurrence of target words from samples and reference corpus

	With Observer (08/2004 – 06/2005)	Without Observer (08/1983 – 03/1984)	FROWN Corpus (1991)
# [numerals]	16.97	17.63	15.84
prime	0.98	0.00	<0.01
SPJ	0.72	0.00	0.00
mania*	0.20	0.05	<0.01
Bess	0.17	0.00	<0.01
idea	0.15	0.19	0.02
research	<0.01	0.07	0.02

*Including lemma

Comparisons of the use of specific terms assigned to the diagnosis of mania across corpora indicated interesting trends as noted in Table D.1, as the use of numbers in written texts was not remarkably different in LMN's writing between the early and late periods, nor was either remarkably different from that observed in typical language use. LMN's use of *idea* remained fairly consistent from early to late writing, but as might be expected based on her diagnosis, occurred at a rate significantly disproportionate to that seen in the population at large. The content of the text samples was notably different in the later sample of writing, as the use of the term *mania* occurred at a significantly higher rate, suggestive of the impact of the

observer and the shift in LMN's perspective based on the formalized involvement of the research. The rate of occurrence noted in the use of *SPJ* and *prime* was unexpected, as these did not appear in the earliest sample, perhaps because LMN had not integrated these constructs completely into her process at that time. Nonetheless, these tokens represented a significant proportion of the total words for the sample in which an observer was involved, and as expected were significantly higher than that that observed in the FROWN corpus. Not unexpectedly, the name of the observer did not appear in LMN's texts until the latest time period in this example. That particular token occurs once in the FROWN corpus, and it is included in this analysis for illustrative purposes.

Conclusion

There were definitive differences observed in the patterns of language use when LMN was observed compared to when she was not. Her intent appeared to change across time as the texts prepared after formal observation began were notable for use of the cohesive referent *SPJ*, as well as addressing the researcher specifically, and making more specific references to mania. Likewise, the frame was altered from a casual stream-of-consciousness style to more formal sentences with appropriate grammatical structure, and frequently involved an epistolary register as her perception of audience shifted from general to specific. The audience was initially undefined, but in very general terms was expected to involve researchers (including LMN) who would be able to use the observations recorded in the journals as the basis for investigation of the phenomena of mania; however, as the paradigm shifted, so did

LMN's writings. Interestingly, under direct observation, LMN's patterns of word selection for vocabulary not specifically included in this analysis and the register with which she recorded her thoughts appeared to approximate the proportional base-rates of occurrence as represented in the FROWN corpus, suggesting a potential "normalizing" effect.

Returning to the initial issue of identification of the participants in this interaction, LMN holds two roles, both as the individual experiencing the phenomenon of mania, and as the scientist/observer who is making careful recordings of her perceptual experiences and her generation of ideas, within the framework of date/time organization. Across time, she shifted roles in her writing, as well as shifting frames; for example, from notation of mathematical formulae in problem-solving a technical aspect of a fulminating idea, to internal perceptions of a more mundane nature (e.g., "I'm hungry."). Apparently, her awareness of the audience was never far from reach, as throughout the course of 28 years she kept a second smaller set of private journals in which she wrote about more personal experiences and perceptions (e.g., recording her menstrual cycle, notes on her anger with colleagues), which were more clearly defined for an audience of self.

Regarding construction of audience, the participation of a specific researcher is explicit and evident in the writing from the time post-2004, evidenced as specific audience in the epistolary format of some of the writing (e.g., beginning an entry with "Dear Bess"), as well as ongoing comments throughout the text with specific suggestions for research or archival purposes. Implicit involvement of the unspecified researcher is also evident, as LMN cast a broad net in her construction of

audience across time. The documents from the earliest days of her manic process indicate her objective to utilize these documents as research data, although her initial intent was to explore the phenomena herself.

Her construction of researcher as audience is determined by the role she has assigned in her writing, also along a continuum from explicit to implicit. Such construction also shapes, and is shaped by, her own role in the writing process, either when adopting the objective scientist/observer or subjectively writing as the person experiencing mania. The roles have become more clearly delineated in her journals with the identification of the observer paradox.

The effects of the observer on the use of language in LMN's journals is evident, and this preliminary analysis clearly indicates the need for further examination of this phenomenon in this body of work, with larger scale implications for the field of corpus linguistic analysis.

References

Adair, J. G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology, 69*(2), 334-345.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Borden, R. J., & Walker, J. W. (1978). Influence of self-observation versus other-observation on immediate and delayed recall. *Journal of General Psychology, 99*, 293-298.

Diaper, G. (1990). The Hawthorne effect: A fresh examination. *Educational Studies, 16*(3), 261-268.

Durbin, M., & Martin, R. L. (1977). Speech in mania: Syntactic aspects. *Brain and Language, 4*(2), 208-218.

Gannett, C. (1992). *Gender and the Journal: Diaries and Academic Discourse*. Albany: State University of New York Press.

Jones, S. R. G. (1992). Was there a Hawthorne effect? *American Journal of Sociology, 98*(3), 451-468.

Kazdin, A. E. (1982). Observer effects: Reactivity of direct observation. *New Directions for Methodology of Social & Behavioral Science, 14*, 5-19.

Khadivi, A., Wetzler, S., & Wilson, A. (1997). Manic Indices on the Rorschach. *Journal of Personality Assessment, 69*(2), 365.

Parsons, H. M. (1974). What happened at Hawthorne? *Science, 183*(4128), 922-932.

Scott, M. (2005). Oxford Wordsmith Tools (Version 4.0): Oxford.

Stubbs, M. (1996). *Text and Corpus Analysis: Computer-assisted Studies of Language and Culture* (Vol. 23). Oxford: Blackwell Publishers.

Tranel, D. D. (1981). A lesson from the physicists. *Personnel & Guidance Journal, 59*(7), 425-429.

Yantz, C. L., & McCaffrey, R. J. (2005). Effects of a supervisor's observation on memory test performance of the examinee: Third party observer effect confirmed. *Journal of Forensic Neuropsychology, 4*(2), 27-38.

Appendix E

Transcription Conventions

These journals are the personal thoughts of an individual, and are mostly written as stream of consciousness. Frequently, her sentences are not complete, or there may just be lists of words. The primary goal is to replicate the content, so type exactly the words she has written, exactly how she wrote them.

1. This is an example of how the files should be labeled: TJ129 1997Apr3. The first part stands for "Transcribed Journal", then the number (inside the front cover), the year, month, and date of the first entry.

2. Begin the first page with the information that is on the first page of the journal – usually the date, some other text, and her name and address.

3. Although her pages are short, you can type them in one long sequence (i.e., you don't have to start a new page every time she does). She does typically begin each new page with the date.

4. Left margin – although her writing is occasionally all over the page, line up her text at the left margin. (Easier to type!)

5. Date and Time – where she has indicated the date and time, please note those exactly as written. This will happen multiple times on a single page, frequently at the beginning and end of an entry.

6. Sketches/drawings - There are numerous non-text sketches. When these occur, please note them in angle brackets like this: <sketch>. I won't be using this in my current analysis, but will need this information for future reference.

7. Graphs/diagrams – when these are included, please type the words that were used, and note in brackets that there was a <graph>.

8. If there are illegible words, please note them in angle brackets as follows : <XXX>.

9. If she has mis-spelled a word, type it just like she spelled it. These will frequently be followed by 'sp', as she catches her error.

10. She uses abbreviations and mathematical formulas in her text. Please type as much of these as you can, but don't worry about the obscure symbols if you can't type them.

11. IMPORTANT – Please save the file as a .txt document in ASCII; you can do this from either Word or Word Perfect. Save each journal on a separate CD,

labeling the CD with the file name, ***and also please e-mail it to me*** at the address below.

12. When you are finished, please write inside the front cover: "Transcribed by… (your initials)" and the date.

If you have **any** questions, please call me or e-mail me **anytime**.