

TOWSON UNIVERSITY  
OFFICE OF GRADUATE STUDIES

PREDICTING PERSONALITY TRAITS AND PSYCHOLATHOLGY FROM LINGUISTIC  
STYLE

By

Monika Conrad

A thesis  
in partial fulfillment  
of the requirements for the degree  
MASTER OF ARTS

Department of Psychology

Towson University  
8000 York Road,  
Towson, Maryland 21252

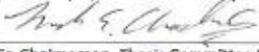
May 2014

TOWSON UNIVERSITY  
OFFICE OF GRADUATE STUDIES  
THESIS APPROVAL PAGE

This is to certify that the thesis prepared by:  
Monika Conrad

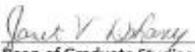
Entitled: Prediction of Personality Traits and Psychopathology from Linguistic Style

has been approved by the thesis committee as satisfactorily completing the thesis requirements for the degree Master of Art, Experimental Psychology  
(for example, Master of Science)

 Mark E. Chachich, Ph.D. 5/9/14  
Co-Chairperson, Thesis Committee Signature Date

 Frederick Parente, Ph.D. 5/9/14  
Co-Chairperson, Thesis Committee Signature Date

 Bryan Devan, Ph.D. 5/9/14  
Committee Member Signature Date

 Janet V. Delany, D. Ed. 5-14-14  
Dean of Graduate Studies Date

## **Acknowledgements**

It would not have been possible to write this thesis without the help, support, and sounding board of many kind people around me. From family members, friends, fellow class cohorts and committee members; to only some of whom it is possible to give particular mention here.

The excellent advice, caring patience, guidance, encouragement and friendship have been invaluable from my advisor and mentor, Dr. Mark Chachich. I am forever grateful for his knowledge, as well as his belief and faith in me. His ability to talk me down before I had to “chew through the restraints” has been instrumental.

Dr. Frederick Parente, as always, my mere expression of thanks, does not suffice. His unsurpassed knowledge, guidance, corrections of data analysis and writing has been priceless. His immense knowledge, calming demure and unlimited amount of guidance has been much needed and beyond appreciated.

Last, and certainly not least, I have to thank my wonderful husband of 25 years; without whom I would never have even attempted to venture into this adventure. Rick has given me the courage and confidence to take this step in the first place. He has provided unwavering strength and support, especially when I did not believe I had the strength or ability to continue.

These caring, thoughtful, and incredibly patient people are the ones responsible for this entire process. There are not enough words to express my extreme gratitude for them being a part of my life.

## **Abstract**

### Predicting Personality Traits and Psychopathology from Linguistic Style

Monika Conrad

This study assessed the extent to which characteristics of written language predicted various personality traits and measures of psychopathology. Linguistic text analysis of written essays were used to predict self-assessment scores derived from the HEXACO six-dimensional personality scale and the SCL-90 psychopathology scale. The results indicated that each of the six HEXACO dimensions could be predicted significantly from the text analysis variables. Likewise, each of three psychopathology variables was predicted significantly by the same set of variables. These data suggest that it is possible to accurately assess personality and psychopathology from written communication.

## Table of Contents

	Page
List of Tables _____	vi
Introduction _____	1
Early Linguistic Style Research _____	3
Personality, Psychopathology and LIWC _____	8
Linguistic Style and Associated Brain Structures _____	11
Current Study _____	12
Participants _____	12
Assessments and Measures _____	13
Proposed Analyses _____	14
Results _____	19
Discussion _____	22
Appendices _____	26
Appendix (A) IRB Approval Form _____	26
Appendix (B) Informational Letter _____	27
References _____	29
Curriculum Vitae _____	33

## List of Tables

Table	Page
1 Function Word Type and Examples_____	6
2 Percentage of Function Word Usage_____	7
3 LIWC2007_____	14
4 Multiple Regression Rsq. of HEXACO Outcome with LIWC Predictors____	19
5 Multiple Regression Rsq. Of Dissociative Outcomes with LIWC Predictors_	21

## **Predicting Personality Traits and Psychopathology from Linguistic Style**

Freud (1901) theorized that a person's speech and written expression revealed their unconscious motivation which, in turn, determined their behavior. He believed a person's choice of words was a window to their unconscious and that "Freudian slips" often revealed the hidden and true self (Groom & Pennebaker 2002; Pennebaker, Mehl & Niederhoffer, 2003). The formatting of written expression, combination of linguistic elements, choice of tense, synthesis of verbal concepts, and the use of specific words types, uniquely identifies each author and allows a vision of the person that permeates the camouflage of their outer mask (Pennebaker & King 1999). Each of us comments, requests, describes, criticizes, and inquires every day; even though our verbiage refers to the same objects, our use of different word combinations makes each statement unique.

Individual differences among humans extends well beyond spoken language. Physical appearance and makeup, gait and station while walking, hand gestures and facial expressions while carrying on a conversation, are as distinct as that individual. Subtle social exchanges such as avoiding eye contact, looking at one's watch while conversing, or pauses in the conversation can provide information about the person's emotional state. It is therefore possible to abstract a rough estimate of a person's personality based on their outward and public display of behavior. It is equally possible to assess personality by the subtle differences that occur in their written communications.

Personality is defined by the American Psychology Association as "the unique psychological qualities of an individual that influence a variety of characteristic behavior patterns, overt and covert, across different situations and over time". The word

“personality” originated from the Latin word “persona”, meaning “mask”. Personality traits have been shown to emerge early in life, are partially predetermined genetically, and are fairly consistent throughout the lifespan (Anderson & Bienvenu, 2011).

Personality is not just displayed in our external behavior, but also our thoughts, feelings and decision processes. Personality includes not only our conscious, but also unconscious tendencies and characteristics.

Psychopathology exists at the extremes of personality. The study of psychopathology includes the study of mood fluctuation, cognitive dysfunction, perceptual and behavioral disturbances (Andersen & Bienvenu 2011). These include extremes of anxiety, depression, hostility, interpersonal sensitivity, obsessions/compulsions, paranoid ideation and somatization. Correlation between personality and various psychopathologies has been theorized since the beginning of psychology (Kotov, Gamez, Schmidt, & Watson, 2010). Mechanisms and causal relationships are still elusive; however, some recent research shows strong evidence of interactions between personality traits and various psychopathologies (Kotov, et al., 2010).

The purpose of the present study is to investigate the linguistic features of written expression as predictors of personality traits and psychopathology. This relationship is becoming more important with the increased use of writing via an expanding electronic world. The study begins with a brief overview of various linguistic programs used to study emotional health, cognition, physical health and personality traits. The discussion continues with an overview of Linguistic Inquiry and Word Count Application (LIWC) and a closer look at previous linguistic studies of personality and psychopathology that

have focused on these traits as they relate to linguistic style. The discussion will also review brain structures associated with verbal language, written language, and personality.

### **Early Linguistic Style Research**

Ways to comprehend text organization have focused on choices of words, assembly of thought and phrase composition; a concept that began taking shape in the 1950's (Pennebaker & King, 1999). One of the first text analysis scales was developed in the early 1960's by a physician, Louis Gottschalk, named Gottschalk Gleser Content Analysis Method of Measures (Pennebaker & King, 1999). This scale required participants to speak for a non-stop stream of thought for five minutes and required highly trained raters to evaluate sentence clauses. The results were used to compare linguistic styles of people with varying medical diagnoses. Another early approach, also developed by a physician, Walter Weintraub, involved categorizing words and phrases into 15 categories; also used to compare others with varying medical diagnoses (Pennebaker & King, 1999).

The General Inquirer technique was the first word-based, versus phrases-based, text parsing programs. Stone, Dumphy, Smith, and Ogilvy (1966), combined specific words linked to various psychosociological phrases as defined in the Harvard dictionary to analyze linguistic styles (Pennebaker & King, 1999). This procedure was trailed by several other word-based analysis programs, many of which focused only on word count. All of these techniques needed human raters and did not take into account the varieties of writing that described the same things.

The Linguistic Inquiry and Word Count (LIWC) derived from the need to provide a wide range word-based program capable of analyzing various psychological dimensions in a person's written and verbal language, Pennebaker et al., (1993). The original program has since been updated and enhanced with two major modifications, Pennebaker (2011). The program originally focused on emotional writing and which aspects of writing, if any, could forecast health improvements following emotional upheavals. These authors asserted that the words people used would reflect their feelings (Pennebaker & King 1999).

The LIWC2001 began with the creation of word dictionaries designed to capture different psychological concepts, linguistic processes, biological, personal concerns, and filler words. Each category of the LIWC dictionary is a collection of words that fall into a particular category. The dictionary covers 80 specific categories, containing most words an average person uses daily. The classifications of which words should fall into unique or multiple dictionary categories took several years and had to be unanimously agreed upon by a conglomerate of independent student judges (2011). The LIWC2001 begins analyzing the text by first counting the words in a phrase; it then looks at each word independently and categorizes it into multiple dictionaries. The final step calculates the total percent of words that correspond to each applicable dictionary category.

One of the first studies by Wiltsey, Stirman, & Pennebaker (2001) using the LIWC2001, analyzed poems written by poets who committed suicide and compared them to a control group of poets whom did not commit suicide. This study included three separate time frames of written poetry: early work, middle, within the two suicide designations. Nine suicide poets' work was compared to nine non-suicide poets' works

of similar age and era. Analysis of these data showed the suicidal vs. non-suicidal poets' differed significantly in their use of integration vs. hopelessness words (Wiltsey, Stirman, Pennebaker, 2001). Post-hoc analysis, indicated that the suicide poets used significantly more first person singular references throughout their careers relative to the non-suicide control group. This finding lead to the conclusion that some of the people were already predisposed to suicide and this tendency did not increase because of added stress created from their newly found fame. Further, first person singular usage was also directly related to depression (Bucci & Freedman, 1981); Weintraub, 1981).

Rude, Gortner, & Pennebaker (2004) also reported a significance relationship between the use of pronouns and depression. These researchers had a group of undergraduate students write for 20 minutes about their emotions and state of mind regarding starting college. Thereafter, the students filled out the Beck Depression Inventory (BDI) and the Inventory to Diagnose-Depression-Lifetime (IDD-L), both of which measured current or previous depression symptoms. The writing samples were then categorized as those that derived from participants who were currently depressed, previously depressed, or never depressed. The samples included 31 currently depressed students, 26 formerly depressed students and 67 never depressed students. The study focused on the linguistics pertaining to word usage of first person singular, first person plural, social references, negative and positive emotional nouns. There were no significant differences between the ages of the participants. There were, however, significant difference between current, formerly, and never depressed students in all linguistic categories except social references. Depressed students used significantly more first person singular words than never depressed students. The most sensitive measure

was the use of first person singular “I”, but no differences in the uses of the first person singular words “me” or “my”. Also, depressed students used more negative words and fewer positive words than did never depressed students. Lastly, the formerly depressed group used “I” more often than the never depressed group (Rude, Gortner, & Pennebaker, 2004). These studies indicate that even simply spoken words can also serve as an index of personality.

The first person “I” is an example of a “function word” which is often used to connect the other words in our vocabulary. Pronouns such as the first person “I” have been shown to discriminate depression and it is therefore possible that other categories of function words can also identify different emotional states. The following shows major categories and examples of function words, according to Pennebaker, (2011):

Table 1: Function Word Type and Examples

<u>Category</u>	<u>Examples</u>
Pronouns	I, she, it
Articles	a, an, the
Prepositions	up, with, in, for
Auxiliary verbs	is, don’t, have
Negations	no, not, never
Conjunctions	but, and, because
Quantifiers	few, some, most
Common adverbs	very, really

About 0.04% of the average person’s vocabulary of 100,000 words are function words, however, these are used in over half (55%) of all the words we use in daily conversations

or writings (Chung & Pennebaker, 2007). They are difficult to use correctly when learning a second language. They are also difficult to detect when listening to spoken language and they are easily overlooked when reading (Chung & Pennebaker, 2007). The top 20 function words and the percentage of usage amongst written words and spoken speeches, according to (Chung & Pennebaker, 2007; Pennebaker, 2011) are summarized below.

Table 2: Percentage of Function Word Usage

I	3.64%	Of	1.83%	Is	1.06%	With	.67%
The	3.48%	That	1.48%	You	1.05%	He	.66%
And	2.92%	In	1.29%	Was	1.01%	Me	.64%
To	2.91%	It	1.19%	For	.80%	On	.63%
A	1.94%	My	1.08%	Have	.70%	But	.62%

Function words have been shown to predict present and past behavior of political candidates. A study by Gunsch et al., (2000) showed that political ads that used more present and future tense verbs, combined with the pronouns focused on present actions of the candidate whereas negative ads using pronouns along with past tense verbs focused more on past actions of the opponent (Tauscik & Pennbaker 2010).

Linguistic style may also be correlated with life stressors. For example, a study by Pennebaker & Lay (2002), analyzed the many press conferences held by Rudolph Guiliani from 1993 to 2001. After experiencing several life altering events in a period of one month in 2000, the analysis of his speeches showed that he increased his use of first person singular pronouns from 2% to approximately 7% following that particular month. His use of words such as “we” or “us”, often associated with close emotional ties to other groups, dropped from 2.5% to 1% and raised to over 3% following the attacks of 9/11 (Chung & Pennebaker, 2007).

Newman, Pennebaker, Berry, & Richards (2003) conducted five separate research studies to measure honesty and deception using linguistic styles. Lying used more negative emotional and motion words, while using fewer first person singular words (Bond & Lee 2005; Hancock, Curry, Goorha, & Woodworth, 2008; Tausezik & Pennebaker, 2010).

### **Personality, Psychopathology and LIWC**

The study of a relationship between personality traits and linguistic style both old and new. Several studies described above have shown that linguistic style can predict personality traits. The correlation is based on the assumption that personality derives from thoughts, emotions, whereas language use is a primary mechanism for expressing these thoughts, emotions (Fast & Funder, 2008).

The majority of these studies have incorporated the Big Five factor model of personality traits, developed by Costa & McCrae (1992). The Big five, also referred to as the NEO personality scale, breaks down personality traits into five categories: Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness. Those who score high on agreeableness tend to behave in a socially acceptable, forgiving, cooperative and helpful manner, while lower scores suggest a person with more abrupt and offensive behaviors. Scoring high on the conscientiousness category predicts higher levels of reliability, organization, and a task-oriented individual; whereas scoring lower suggests messier and disorganized task completion and daily activities. A high score on Extraversion predicts a more social and outgoing personality, ambition, task orientation, and a strong-will. A lower score is associated with more avoidance of social activities. Those who score high on the neuroticism trait category tend to be self-conscious,

anxious, distrustful, and fearful in novel social venues. Lower scorers are more “go with the flow”, and less unnerved. A high score on the openness scale suggests more unconventional behavior and openness to new ideas; a lower score suggests ability to stay with routine and concrete matters (Gill, 2011; Hirsh & Peterson, 2009).

Several studies have reported significant correlations with regard to linguistic style and personality traits using the NEO. Hirsh & Peterson (2009) asked 94 undergraduate students to write about a past experience and how it had changed their lives or future goals. The students also completed the big five personality scale. Results indicated a significant average correlation of  $r = .23$  between word usage and the five personality traits measured. Pennebaker & King (1999) found neuroticism correlated with negative emotional word use and use of first person singular pronouns. Scherwitz, Graham, & Ornish (1985) and Pennebaker & King (1999) found extraversion traits were correlated with the use of positive emotion words, as well as social processes words. Agreeableness correlates with the use of negative and positive emotion words (Tausezik & Pennebaker, 2010). Conscientious traits correlated with use of more positive and less negative words (Groom & Pennebaker, 2002). While openness traits correlated with use more article words, but fewer self-reference words.

Yarkoni (2010) invited over 5,000 bloggers to participate in a large online personality and linguistic style correlation study in which each of them completed the NEO personality scale. The results showed similar correlations of word use to personality traits of the big five as previous studies (Fast & Funder, 2008; Hirsh & Peterson, 2009; Pennebaker & King, 1999). Openness also correlated with category-level linguistic as well as word-level linguistics, for example, words such as “sang”,

restaurant”, “crowd”, and cultural related words, such as “art”, “literature” “poet” (Yarkoni, 2010). Openness has also been shown to have the lowest correlation to any form of psychopathology (Andersen & Bienvenu, 2011). While the other four traits correlated with less than 30 specific words, these same words show stronger correlation to measures of pathology (Andersen & Bienvenu, 2011).

LIWC program has been used in a number of research studies. Various categories of word-use has been shown to correlate with education, social class, social interest, attentional focus, aggression, emotion, goal orientation, inhibition, aggression, verbal skills, cognitive complexity, honesty, and even agreement (Tausczik & Pennebaker, 2010), as shown in an appendix attached to Tausczik & Pennebaker (2010).

Although most of the studies outlined above demonstrate a link between linguistic style and personality, a few studies suggest a link between linguistic style and psychiatric disorders. (Oxman, Rosenberg, Schnurr, and Tucker, 1985). For example, patients with somatization disorder used more negative words than patients suffering from depression. Writing samples have been used to classify patients into various psychiatric disorder categories; such as, depression, paranoia, or schizophrenia. This technique may more accurately classify patients in to diagnostic categories relative to clinical judgement (Pennebaker, Mehl & Niederhoffer, 2003).

Junghaenel, Smyth, & Santner, (2008) recruited 27 psychiatric outpatient participants, suffering from varying degrees of schizophrenia, delusional disorders, or bipolar disorder, and compared their linguistic style with a control group of 17 volunteers, not diagnosed with any psychiatric disorders. They instructed the participants to write an essay about their lives, family, school, or hobbies. The specific topic was left

up to the participant. Their primary focus was on use of pronouns, negative words, positive words, time reference words and cognitive processing. The results indicated that psychiatric patients used significantly fewer positive words, fewer cognitive processing words reflecting inhibition and certainty, and fewer time reference category words. Pronoun use and negative emotion word use however, were not significantly different among the groups (Junghaenel, Smyth, & Santner, 2008).

### **Linguistic Style and Associated Brain Structures**

Frontal lobe functioning is often associated with emotions, social appropriateness, social skills, and judgment (Pennebaker, 2011). Language is also mediated by the temporal and frontal lobe. Wernicke's area, located in the left temporal area, is vital for understanding and comprehending, language. Brocca's area, situated in the left frontal, is vital for speech. Brocca's area, if damaged, produces speech lacking function words. Damage to these areas can affect a person's ability to communicate which may, therefore, produce a measureable linguistic profile that correlates with pathology (Pennebaker, 2011).

The study of linguistic style, personality, and brain physiology is in its infancy. DeYoung, et al., (2010) compared brain region volume with the big five personality traits. DeYoung and associates recruited 116 participants to complete the big five personality scale and return on a separate day to be scanned in an MRI machine. The researchers found correlations between personality traits and regional brain volume in four of the big five personality traits (DeYoung, et al., 2010). The Extraversion trait correlated with more volume in medial orbitofrontal cortex; area involved in reward feedback. The Neuroticism trait correlated with reduced volume in dorsomedial pre-

frontal cortex, as well as posterior hippocampus and increased volume in mid-cingulate gyrus. Previous evidence has shown that functioning of the hippocampus correlates with stress and depression, while function in the mid-cingulate gyrus has been associated with physical and emotional pain (Anderson & Bienvenu, 2011). Dorsomedial pre-frontal cortex has been correlated to self and emotional control; whereas, the Agreeableness trait correlated with reduced posterior left superior temporal sulcus. Posterior cingulate cortex functioning correlates with altruistic behavior and conscientiousness correlates with left lateral pre-frontal functioning. Openness was the only trait which did not show a significant difference in brain volume regions (DeYoung, et al., 2010).

To summarize, the studies outlined above suggest that different personality traits correlate with linguistic style and functioning in different areas of the brain. To the extent that there is damage to any of these brain areas, then personality would fluctuate. Linguistic style may therefore be one way of assessing these fluctuations thereby serving as a diagnostic tool. There is, however, relatively little research that relates linguistic style to measures of psychopathology. The underlying logic of these investigations is, however, the same. If psychopathology represents extremes of personality, and because personality has been shown to correlate with linguistic style, then it is reasonable to suggest that linguistic style would also correlate with measures of psychopathology.

### **Current Study**

#### **Participants**

Participants were student volunteers from Towson University who received course credit for their participation. The participants were administered two separate self-reporting inventory questionnaires (SRI). Upon completion of the SRI's, they

compose two essays. The first essay described the current weather conditions in as much detail as possible; in order to establish an unemotional and neutral linguistic style sample. The second essay focuses on self-narrative linguistic style on which the participant composed a written reflection on their current and anticipated “end of semester” mood, plans, or outlook. Upon completion, in order to ensure anonymity, they were provided envelopes to enclose their essays and measurements.

### **Assessments and Measures.**

The administered SRI's included the SCL-90-R symptom inventory. The SCL-90-R consists of 90 questions on a five-point Likert scale used to measure the possible levels of Somatization, Obsessive-Compulsive, Interpersonal sensitivity, Depression, Anxiety, Hostility, Phobic anxiety, Paranoid ideation, and Psychoticism. The personality component of the SRI was assessed using the recent HEXACO mode of personality scale. The HEXACO acronym stands for six broad personality dimensions; Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to Experience. The sixth dimension, not found in the widely administered NEO, Honesty-Humility is characterized by facets of sincerity, fairness, greed-avoidance, generous and modest. The HEXACO scale consists of 100 questions, also encompassing the five-point Likert level scale. The essays were analyzed with the LIWC2007 to identify the participants' unique linguistic style. This analysis provided numerical indices of function words, psychological processes, and affective process.

## Proposed Analyses

These data will be organized in those items that measure personality (HEXACO) and those that measure psychopathology (SCL90-R). Analysis of the writing samples was organized into several sections that are described by the table in the LIWC manual that is reproduced below:

Table 3: LIWC2007

Category	Abbrev	Examples
<b>Linguistic Processes</b>		
Word count	wc	
words/sentence	wps	
Dictionary words	dic	
Words>6 letters	sixltr	
Total function words	funct	
Total pronouns	pronoun	I, them, itself
Personal pronouns	ppron	I, them, her
1 <sup>st</sup> pers singular	i	I, me, mine
1 <sup>st</sup> pers plural	we	We, us, our
2 <sup>nd</sup> person	you	You, your, thou
3 <sup>rd</sup> pers singular	shehe	She, her, him
3 <sup>rd</sup> pers plural	they	They, their, they'd
Impersonal pronouns	ipron	It, it's, those
Articles	article	A, an, the
[Common verbs] <sup>a</sup>	verb	Walk, went, see
Auxiliary verbs	auxverb	Am, will, have
Past tense <sup>a</sup>	past	Went, ran, had
Present tense <sup>a</sup>	present	Is, does, hear

Future tense <sup>a</sup>	future	Will, gonna
Adverbs	adverb	Very, really, quickly
Prepositions	prep	To, with, above
Conjunctions	conj	And, but, whereas
Negations	negate	No, not, never
Quantifiers	quant	Few, many, much
Numbers	number	Second, thousand
Swear words	swear	Damn, piss, fuck
<b>Psychological Processes</b>		
Social processes <sup>b</sup>	social	Mate, talk, they, child
Family	family	Daughter, husband, aunt
Friends	friend	Buddy, friend, neighbor
Humans	human	Adult, baby, boy
Affective processes	affect	Happy, cried, abandon
Positive emotion	posemo	Love, nice, sweet
Negative emotion	negemo	Hurt, ugly, nasty
Anxiety	anx	Worried, fearful, nervous
Anger	anger	Hate, kill, annoyed
Sadness	sad	Crying, grief, sad
Cognitive processes	cogmech	cause, know, ought
Insight	insight	think, know, consider
Causation	cause	because, effect, hence
Discrepancy	discrep	should, would, could
Tentative	tentat	maybe, perhaps, guess
Certainty	certain	always, never
Inhibition	inhib	block, constrain, stop
Inclusive	incl	And, with, include

<b>Category</b>	<b>Abbrev</b>	<b>Examples</b>
Exclusive	excl	But, without, exclude
Perceptual processes <sup>c</sup>	percept	Observing, heard, feeling
See	see	View, saw, seen
Hear	hear	Listen, hearing
Feel	feel	Feels, touch
Biological processes	bio	Eat, blood, pain
Body	body	Cheek, hands, spit
Health	health	Clinic, flu, pill
Sexual	sexual	Horny, love, incest
Ingestion	ingest	Dish, eat, pizza
Relativity	relativ	Area, bend, exit, stop
Motion	motion	Arrive, car, go
Space	space	Down, in, thin
Time	time	End, until, season
<b>Personal Concerns</b>		
Work	work	Job, majors, xerox
Achievement	achieve	Earn, hero, win
Leisure	leisure	Cook, chat, movie
Home	home	Apartment, kitchen, family
Money	money	Audit, cash, owe
Religion	relig	Altar, church, mosque
Death	death	Bury, coffin, kill
<b>Spoken categories</b>		
Assent	assent	Agree, OK, yes
Nonfluencies	nonflu	Er, hm, umm
Fillers	filler	Blah, I mean, you know

The LIWC computer program provides numerical values for each of the individual entries under each bolded category in the table. Multiple regression analyses were used to correlate the set of personality and psychopathology variables with the set of writing variables provided by the LIWC. The HEXACO provides a set of six personality characteristics and the SCL-90 yields a set of nine psychopathology dimensions. The LIWC yields a set of 48 measures of written communication.

The first set of analyzes explored the relationship between the personality and psychopathology measures and the various categories from the LIWC. All of the analyses included, both the personality or psychopathology dimensions and the entire set of linguistic processes from the LIWC. Regarding the HEXACO scale, six multiple regression analyses correlated the LIWC measures with each of the HEXACO subscales. For the SCL-90 scales, the multiple regressions correlated the LIWC measures with three symptom dimensions measuring anxiety/depression, paranoia/psychoticism, and panic/phobia (Steinberg, Barry, Sholomsky & Hall, 2005) with the various LIWC categories. Several forward solution multiple regression analyses assessed the relationship between the HEXACO personality subscale, the three SCL90 psychopathology dimensions, and the various linguistic scores from the LIWC. These analyses tested the three global hypotheses listed below by assessing the significance of the multiple  $R^2$  between the set of LIWC linguistic variables, the HEXACO personality subscales, and the SCL90-R psychopathology outcomes. Beyond these tests however, the analyses were exploratory. That is, the forward solution was selected because it provided a multiple  $R^2$  that was used to test the hypothesis of a relationship between the linguistic variables and the various outcomes. The assessment of the individual

linguistic predictors was, however, post-hoc because there were no specific predictions regarding which would be significant.

Apriori Hypotheses:

1. The various studies summarized above predict that there should be an overall significant relationship between the personality variables and measures of expressive writing. Specifically, one or more of the measures of personality, defined by the HEXACO, will correlate significantly with the set of LIWC linguistic variables (Groom & Pennebaker, 2002; Hirsh & Peterson, 2009; Pennebaker & King, 1999; Tausezik & Pennebaker, 2010). This hypothesis will be assessed using multiple regression analysis where each of the HEXACO personality variables will be predicted from the set of LIWC measures.
2. To the extent that psychopathology represents the extremes of personality, then it is reasonable to suggest that one or more of the psychopathology measures will also correlate with the set of writing variables. Therefore, the LIWC measures will predict one or more of the SCL90 psychopathy subscale scores.
3. The studies referenced above also predict that measures of linguistic style, e.g., total pronouns, personal pronouns, 1<sup>st</sup> person singular, 1<sup>st</sup> person plural, Affective processes, Positive emotion, Negative emotion, Anxiety, Anger, Sadness, and Assent, will also correlated significantly with one or more of the HEXACO and SCL-90 dimensions.

## Results

### Predicting Personality (HEXACO subscales)

An analysis of the 69 participants' data began by screening the variables for model fit. The analysis revealed that regression residuals were distributed in bell-shaped and symmetric fashion. The multiple Rsquare from the forward solution regression models assessed the overall relationship between the LIWC linguistic variables and the HEXACO trait subscales (Table 4). These analyses tested hypothesis 1. Each of the Rsquares in the table was significant ( $p < .05$ ) which provided support for hypothesis 1. Specifically, each of the HEXACO subscales correlated significantly with a subset of the LIWC measures.

Table 4: Multiple Regression Rsq. of HEXACO Outcome with LIWC Predictors

<b>HEXACO</b>	<b>R SQ</b>	<b>LIWC</b>	<b>EXAMPLE</b>	<b>CATEGORY</b>
Openness	.423	Family	Daughter, Husband, Aunt	Psychological-Social
		Anger	Hate, Kill, Annoyed	Psychological-Social
		Include	And, With, Include	Psychological-Cognitive
		Quant	Few, Many, Much	Linguistic-verb
Agreeableness	.390	Filler	Blah, I mean, You know	Spoken
		Exclude	But, Without, Exclude	Psychological-cognitive
		See	View, Saw, Seen	Psychological-perceptual
		Quote	Quotation marks	Punctuation
		Exclam	Exclamation mark	Punctuation
Conscientiousness	.345	You	You, Your, Thou	Linguistic-2 <sup>nd</sup> person
		Home	Family	Personal concerns
		Insight	Think, Know, Consider	Psychological-cognitive
		OtherP	She, Their, Them	Linguistic-3 <sup>rd</sup> person
		Feel	Feels, Touch	Psychological-Perceptual
Extraversion	.342	Quant	Few, Many, Much	Linguistic-Common verb
		Allprent		Punctualtion
		Dash		Punctualtion
Honesty	.141	See	View, Saw, Seen	Psychological-Perceptual
		Auxverb	Am, Will	Linguistic-Common verb

Emotionality	.106	We	We, Us, Are	Linguistic-1 <sup>st</sup> person plural
--------------	------	----	-------------	--

The multiple regression analysis computed on the Openness to Experience HEXACO subscale yielded an adjusted Rsquare of .423,  $F(4, 68)=11.734$ ,  $p<.05$ , indicating that about 43% of the variance in a person's openness to experience could be accounted for by linguistic style. The LIWC variables predicted about 39% the Agreeableness trait,  $F(5, 68)=8.054$ ,  $p<.05$ . About 35% of the conscientiousness trait,  $F(5,68)=9.717$ ,  $p<.05$  and 34% of Extraversion,  $F(3,68)=11.242$ ,  $p<.05$ , could be predicted by linguistic style. Although the Rsquares were significant, the Honesty (Rsqr at .141),  $F(2,68)=5.428$ ,  $p<.05$ , and Emotionality (Rsqr = .10),  $F(1,68)=7.939$ ,  $p<.05$ , variables showed the least amount of predictability.

Table 4 indicates that words that reflect social and cognitive psychological processes related to family, anger, inclusion, and quantification predicted Openness to experience. The use of punctuation, filler words, words that expressed exclusion, and perceptual words expressing vision, predicted Agreeableness. Conscientiousness was related to second and third person usage, personal concerns, insightfulness, family, and home. Total word usage, quantitative word usage, and use of punctuation predicted Extroversion. Use of words reflecting first person plural words predicted emotionality traits.

#### Predicting Psychopathology (SCL90-R subscales)

Analysis of these data also began with a screening of the outcome variables to determine if they met the assumptions of the regression model. This analysis also indicated that the regression residuals were bell-shaped and symmetrically. The multiple

regression model showed an overall significance of the following Dissociative outcomes from the SCL-90 with the LIWC linguistic styles showing R square in table 5.

Table 5: Multiple Regression Rsq. Of Dissociative Outcomes with LIWC Predictors

<b>DISSOCIATIVE</b>	<b>R SQ</b>	<b>LIWC</b>	<b>EXAMPLE</b>	<b>CATEGORY</b>
Anxiety/Depression	.421	Quant	Few, Many, Much	Linguistic-Quantifiers
		Verb	Walk, Went, See	Linguistic-Verbs
		OtherP	Punctuation	
		Motion	Arrive, Car, Go	Psychological-Relativity
Paranoia/Psychotism	.405	Past	Went, Ran, Had	Linguistic-Common verb
		Relative	Area, Bend, Exit, Stop	Psychological-Relativity
		WC		Linguistic-Word count
		Sad	Crying, Grief, Sad	Psychological-Affective
Panic/Phobia	.359	Anger	Hate, Killed, Annoyed	Psychological-Affective
		Past	Went, Ran, Had	Linguistic-Common verb
		Space	Down, In, Thin	Psychological-Relativity
		Parenth	Punctuation	Punctuation
		Future	Will, Gonna	Linguistic-Common verb
		WC	Word Count	Linguistic-Word count

The regression analysis indicated that approximately 42% of the variance in a person's psychopathology traits could be accounted for by linguistic style,  $F(4,67)=11.442, p<.05$ . The LIWC measures predicted Paranoia/Psychoticism with about 41% accuracy,  $F(5,67)=8.456, p<.05$ , and the Panic/Phobia dimension with about 36% accuracy,  $F(5,67)=6.945, p<.05$ . Use of linguistic quantifiers, verb usage, increased punctuation, and relative words that implied motion characterized anxiety and depression. Paranoia and psychoticism were characterized by increased verb usage with reference to the past, use of relative words that denoted boundary, increased word usage, and use of affective words that displayed anger and sadness. Common verbs related to the past and

future, relative words related to space, and increased word use and punctuation in general, characterized the panic and phobic dimension.

### **Discussion**

There are three major findings in these data that relate to the hypotheses outlined above. The first is that one or more of the LIWC linguistic variables predicted each of the personality variables which is consistent with the published literature. This finding supports the first hypothesis. Second, one or more of the LIWC variables predicted each of the psychopathology variables which were confirmed hypothesis 2. The third finding was that several of the LIWC variables that were significant predictors were also identified as significant covariates of personality in the published literature.

Each of the HEXACO subscale scores were predicted significantly by one or more of the LIWC text variables. This finding indicated that linguistic style can predict various personality traits as defined by the HEXACO personality scale. This finding is consistent with the literature in which linguistic analyses statistics are also correlated with various subtests on the NEO (Big 5) personality scale. One subtest in particular, the Extroversion scale, consistently correlates with linguistic analysis variables across studies. Pennebaker et al. (2003) reported that Extraversion correlated significantly with the use of social words, positive affect words, and total word count. This study also showed significant correlations between Extroversion and the total percentage of words used. However the use of verbs and punctuation also predicted Extroversion. Correlations between Extroversion and the use of social words and positive affect words were not apparent in these data (Pennebaker et al., 2003). Taken together, the most consistent predictor of Extroversion may be the person's verbal fluency.

Honesty was another trait which was predicted significantly by the LICW variables but the pattern of predictors did not correspond to those Pennebaker et al. (2003) reported. Pennebaker indicated that pronoun use and use of cognitive words predicted Honesty. This study identified visual-perceptual word use and the use of verbs as the best predictors of Honesty. Pennebaker et al. (2003) concluded depression showed a higher use of first person pronoun, with a lack of second and third person pronouns and Psychological negative affect. Although the HEXACO does not measure depression, these results did reveal a significant correlation between the use of the first person plural use and Emotionality.

A second major finding in this study was that linguistic style measures on the LIWC predicted three measures of psychopathology. The analysis of this multivariate relationship has been largely overlooked or ignored in the published literature. Therefore, the results reported herein do not replicate any specific body of published research. They do, however, generate several interesting hypotheses that should be replicated. For example, measures of anxiety and depression are related to the person's use of linguistic quantifiers, verbs, and motion. Paranoia and psychosis are related to use of common verbs, words that suggest boundary, total word usage, and words that reflect sadness and anger. Feelings of panic and phobia are reflected in more use of common verbs, words reflecting space, and preoccupation with the past and future. These findings were never intended as tests of literature based hypotheses. They are offered simply as generated hypotheses that should be formulated into more specific testable hypotheses in future studies.

A third finding was that several of the linguistic style variables that reportedly were significant correlates of personality in the published research were replicated in these results. For example, total word usage, use of first person, and words that depict negative emotions such as anger and sadness were all significant predictors of the personality and psychopathology subscale scores. This finding only partially replicates the earlier studies because the same personality scale (NEO) was not used in the present study. It is, however, significant to note that several of the same linguistic style variables predicted the personality dimensions regardless of which scale was used. By way of speculation, total word use may reflect hyperverbality resulting from anxiety. Use of first person may indicate narcissism. Words that reflect negative emotions are often associated with unhappiness, irritability, and dismay.

Although this study did test some apriori hypotheses, the study also generated as many questions as it answered. These results generally replicate the major findings from a limited body of published research that shows that personality and psychopathology can be predicted from a person's written communication. The same body of literature however, provides only limited consensus regarding what aspects of writing predict the various aspects of personality. It will therefore be necessary to replicate the consistent findings from this and other published studies before proposing any meaningful theory that can be tested. It will also be necessary to extend these findings to other cultural populations and to determine if the results are consistent with writing samples derived from different styles (e.g., prose, technical, memos, text messages etc.). Development of a personality scale based on the unobtrusive analysis of a person's writing may be especially useful to clinicians. Although the application of these findings is in its

infancy, the results of this and other published studies documents the promise and potential of the technology.

Appendix A  
IRB approval form



**EXEMPTION NUMBER: 14-X133**

To: Monika Conrad  
From: Institutional Review Board for the Protection of Human  
Subjects Elizabeth Katz, Member   
Date: Friday, April 25, 2014  
RE: Application for Approval of Research Involving the Use of  
Human Participants

Office of Sponsored Programs  
& Research

Towson University  
8000 York Road  
Towson, MD 21252-0001  
t. 410 704-2236  
f. 410 704-4494  
[www.towson.edu/ospr](http://www.towson.edu/ospr)

Thank you for submitting an application for approval of the research titled,  
*Composed Linguistic Correlation with Personality and  
Psychopathological Traits*

to the Institutional Review Board for the Protection of Human Participants  
(IRB) at Towson University.

Your research is exempt from general Human Participants requirements  
according to 45 CFR 46.101(b)(2). No further review of this project is  
required from year to year provided it does not deviate from the submitted  
research design.

If you substantially change your research project or your survey  
instrument, please notify the Board immediately.

We wish you every success in your research project.

CC: M. Chachich  
File

## Appendix B

### Informational Letter

Dear Student,

My name is Monika Conrad and I am a graduate student in the Experimental Psychology program at Towson University. As part of my research for my master's thesis, I am conducting a research experiment to determine if any personality traits correlate with writing style, as well as if varying forms of distressed thoughts or behaviors correlate with writing style and/or personality traits. The purpose of this study is to see if linguistic style, combined with self-reported surveys, can be another significant tool used to help identify individuals suffering from various forms of emotional, behavioral or psychological distress. This, in turn, could help contribute to speedier and more precise treatment. While you will not directly benefit from participating in this research, findings from this study may help to determine whether written work can be used, in combination with other measures, to identify individuals who are experiencing distress. You will not be compensated for your participation.

Your decision to participate is strictly voluntary. There will be no compensation if you chose to participate. And, if you chose not to, your standing in class will not be affected. You also do not need to answer any questions you do not want to answer. You may discontinue participation in this study at any time without penalty. Please do not put your name or any other identifying marks on the forms. You will notice there is also an envelope attached to the answer form; this is to further ensure anonymity that no one inadvertently sees any responses. Please only insert the answer page into the envelope and insert the envelope into the sealed box with the slotted top at the front of the classroom. This box will be opened after all forms have been collected from all classes participating and are ready to be analyzed.

This study consists of two questionnaires to be filled out. One includes 100 questions and the other 90 questions. Both have a 5-point Likert scale answer. These should take you 15-20 minutes to answer. The next section asks you to write two separate short essays. Just write the first thoughts that come to your mind in as many words as possible. You need not worry about grammar or spelling, and this section should take you 10-15 minutes.

Given the personal nature of these questionnaires, they may cause you some distress. If you are now or at any time in the future experiencing distress for any reason, you are encouraged to seek services at the Towson University Counseling Center. The counseling Center is there to help you and offers 12 free counseling sessions to students. Please do not hesitate to contact them for a confidential and free appointment at 410-704-2512. The attached bookmark also has their information.

Please keep this information form in case you find you have some questions or concerns after completing this study. If you have any other questions about this project, you may contact me at 443-643-5268 or mconra5@student.towson.edu, my thesis advisor, Dr. Mark Chachich at 410-704-3211, or the Chairperson of Towson University's Institutional Review Board, Dr. Debi Gartland at 410-704-2236.

This project has been reviewed by the Institutional Review Board for the protection of human participants at Towson University. By completing the questionnaires and essays, I am indicating that I understand the purpose of the study and its potential risks and benefits, and I am indicating my willingness to participate in this research.

Thank you so much for your help in this project. Remember to keep this information form for later use if you have any questions or concerns and do not hesitate to contact the Counseling Center, they are there to help you.

Sincerely,

Monika Conrad, student

Experimental Psychology, Master's Program

## References

- Andersen, A. M. & Bienvenu, O. J. (2011). Personality and psychopathology. *International Review of Psychiatry*, 23, 234-247. Doi: 10.3109/09540261.2011.588692
- Argaman, O. (2010). Linguistic markers and emotional intensity. *Journal of Psycholinguist Research* 39, 89-99. Doi: 10.1007/s10936-009-9127-1
- Augustine, A. A., Mehl, M. R., & Larsen, R. J. (2011). A positivity bias in written and spoken English and its moderation by personality and gender. *Social Psychological and Personality Science* 2(5), 508-515. Doi: 10.1177/10.1948550611399154
- Chung, C., & Pennebaker, J. (2007). The psychological function of function words. In K. Fiedler (Ed.), *Social Communication* (pp. 343-359). Retrieved from
- Chung, C., & Pennebaker, J. (2008). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality* 42, 96-132. Doi: 10.1016/j.jrp.2007.04.006
- DeYoung, C. G., Hirsh, J. B., Shane, M. S., Xenophon, P., Rajeevan, N., & Gray, J. R. (2010). Testing predictions from personality neuroscience: Brain structure and the big five. *Psychological Science* 21(6), 820-828. Doi: 10.1177/0956797610370159
- Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: Correlation with self-report, acquaintance report, and behavior. *Journal of Personality and Social Psychology* 94(2), 334-346. Doi: 10.1037/0022-3514.94.2.334

- Gill, A. J. (2011). Personality and language in computer-mediated communication. *Information Design Journal* 19(3), 250-257. Doi: 10.1075/idj.19.3.06gil
- Groom, C. J. & Pennebaker, J. W. (2002). Words. *Journal of Research in Personality* 36(2002), 615-621. Retrieved from [www.sciencedirect.com](http://www.sciencedirect.com)
- Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of Research and Personality* 43(2009), 524-527. Doi: 10.1016/j.jrp.2009.01.006
- Holleran, S. E., & Mehl, M. R. (2007). Let me read your mind: Personality judgments based on a person's natural stream of thought. *Journal of Research and Personality* 42(2008), 747-754. Doi: 10.1016/j.jrp.2007.07.011
- Junghaenel, D. U., Smyth, J. M., & Santner, L. (2008). Linguistic dimensions of psychopathology: A quantitative analysis. *Journal of Social and Clinical Psychology* 27(1), 36-55. Retrieved from [www.sciencedirect.com](http://www.sciencedirect.com)
- Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" five personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin* 136(5), 768-821. Doi: 10.1037/a0020327
- Li, C., & Chignell, M. (2010). Birds of a feather: How personality influences blog writing and reading. *International Journal of Human-Computer Studies* 68, 589-602. Doi: 10.1016/j.ihcs.2010.04.001
- Pennebaker, J. W. (2011) *The Secret Life of Pronouns: What our Words say about us*. {Kindle Edition}. Retrieved from <http://www.amazon.com/Secret-Life-Pronouns-Words-About->

ebook/dp/B005E8AB16/ref=sr\_1\_1\_title\_0\_main?s=books&ie=UTF8&qid=1399552565&sr=1-1&keywords=pennebaker

Pennebaker, J. W., & Chung, C. K. (in press). Expressive writing connections to physical and mental health. H. S. Friedman (Ed.). *Oxford Handbook of Health Psychology*. New York, N.Y.: Oxford University Press.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology* 77(6), 1296-1312.

Retrieved from

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review Psychology* 54, 547-577. Doi: 10.1146/annurev.psych.54.101601.145041

Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the life span. *Journal of Personality and Social Psychology* 85(2), 291-301. Doi: 10.1037/0022-3514.85.2.291

Roivainen, E. (2013). Frequency of the use of English personality adjectives: Implications for personality theory. *Journal of Research in Personality* 47,417-420. Doi: 10.1016/j.jrp.2013.04.004

Rude, S. S., Gortner, E., & Pennebaker, J. W. (2004). Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, 18(8), 1121-1133. Doi: 10.1080/02699930441000030

Seih, Y., Chung, C. K., & Pennebaker, J. W. (2011). Experimental manipulations of perspective taking and perspective switching in expressive writing. *Cognition and Emotion* 25(5), 926-938. Doi: 10.1080/02699931.2010.512123

- Steinberg, M., Barry, D. T., Sholomskas, D., & Hall, P. (2005). SCL-90 symptom patterns: Indicators of dissociative disorders. *Bulletin of the Menninger Clinic*, 69(3), 237-249. Retrieved from <http://www.strangerinthemirror.com/articles/Steinberg-scl90symptoms-BullMennClinic05.pdf>.
- Stirman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine* 63, 517-522. Retrieved from
- Tanaka, E., Sakamoto, S., Kijima, N., & Kitamura, T. (1998). Different personalities between depression and anxiety. *Journal of Clinical Psychology* 54(8), 1043-1051. Retrieved from
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29(1), 24-54. Doi: 10.1177/0261927X09351676
- Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis personality and word use among bloggers. *Journal of Research in Personality* 44, 363-373. Doi: 10.1016/j.jrp.2010.04.001

## Curriculum vita

Monika Conrad  
[REDACTED]  
[REDACTED]  
[REDACTED]

### Education

Master of Arts, Experimental Psychology  
Expected Graduation May 2014  
Towson University  
Towson, Md.

Bachelors of Science, Psychology 2012  
Towson University  
Towson, Md.

Associates of Arts, Psychology 2010  
Harford Community College  
Bel Air, Md.

### Teaching Assistantships

Psychology of Aging (2014)  
Towson University  
Dr. Mark Chachich

Physiology (2014)  
Towson University  
Dr. Mark Chachich

### Employment

Wellspring Village Program Director  
Brightview Senior Living (2007-2013)  
Bel Air, Md.

Substitute Teacher  
Harford Co. School District (2007-2009)  
Bel Air, Md.

Realtor  
Keller Williams Realty (2003-2006)  
Remax Elite (2001-2003)  
Coldwell Banker (1998-2001)

**Organizations**

PsiChi (2012-present)

Golden Key (2011-present)

Phi Theta Kappa (2009-present)

