UNCOVERING PSYCHOPATHS: AN AUTOMATED LINGUISTIC APPROACH

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

M.S. in Communications

by

Randy Au

January 2009

ABSTRACT

The identification of psychopaths, especially those who commit violent crimes, is important to society because of their increased risk of criminal recidivism rates. In this paper, we examine communication patterns that may be unique to individuals high in psychopathy. Several linguistic features in the narratives of prisoners convicted of murder identified by the use of automated Natural Language Processing (NLP) suggest that spontaneously produced psychopathic speech differs from non-psychopathic speech. Psychopaths produced more speech disfluencies, past tense verbs, and food/drink related words in their narratives compared to non-psychopaths. Surprisingly, no overall differences in affective word use were detected.

BIOGRAPHICAL SKETCH

Randy Au was born and raised in New York City, eventually graduating from Stuyvesant High School before earning a B.S. while majoring in Business Administration and Philosophy at SUNY at Stony Brook. He then moved to Cornell to study for his M.S in Communication.

ACKNOWLEDGEMENTS

I'd like to thank my advisor, Jeff Hancock, for dealing with the constant stream of drafts, my committee for all their patience, the wonderful department staff for all their support, and all my friends for dealing with the occasional outburst of stress.

TABLE OF CONTENTS

LIST OF TABLES

UNCOVERING PSYCHOPATHS: AN AUTOMATED LINGUISTIC APPROACH

Psychopaths, individuals with serious emotional, ethical, and moral deficiencies, hold an important position within the criminal population. They tend to have higher recidivism rates for violent crimes than non-psychopathic criminals. Indeed, psychopathic characteristics are a better indicator of recidivism than a criminals' prior criminal history data (Serin, 1996). Given that psychopaths are more likely to commit serious crime after release from incarceration, identifying psychopaths is of great importance to society as a whole.

Currently, the primary method of identifying or classifying psychopaths is a psychological diagnostic test called the Psychopathy Checklist-Revised (PCL-R, Hare, 1991) that must be administered by a certified clinician based on a person's case history and an interview conducted under controlled conditions. The ultimate objective of the present study is to develop an alternative classification technique that analyzes a subject's communication patterns to classify the speakers as likely to be psychopathic or not. As such, the two primary goals of the study are to 1) identify the linguistic correlates of psychopathy, and 2) develop a statistical classification model that can distinguish between texts produced by psychopaths and non-psychopaths.

Characterizations of Psychopaths

Psychopaths have been described as having a number of unique traits that set them apart from the general population. Reiber and Vetter (1996, p.2) quote Coleman (1956) in saying that the psychopath is "marked by a lack of ethical or moral development and an inability to follow socially approved codes of

behavior." They also quote Coleman's (1956) description that psychopaths are unable to "profit from mistakes and ordinary life experiences except by learning to exploit people and to escape punishment" (Reiber & Vetter, 1996, p.5) and they describe psychopaths along four major characteristics: 1) thrill-seeking, 2) pathological glibness, 3) antisocial pursuit of power, and 4) the absence of guilt.

The thrill-seeking characteristic of psychopathy is marked by an extreme boredom with the mundane and ordered world. It is also different from the characterizations of impulsiveness that comes from earlier literature, in that the thrill-seeking aspect can require a great amount of planning in order to execute. "True psychopaths prefer an open-ended world... they are looking to create situations of ambiguity and potential danger" (Reiber & Vetter, 1994, p11). The established rules of society are not interesting enough to the true psychopath, and they instead look to break rules or add secret new ones to make things more exciting.

The pathological glibness dimension refers to a psychopath's ability to speak colorfully, but with no relation to facts. Words are tools for manipulating reality, such as avoiding punishment, or tricking others, and since psychopaths use words as tools, they are unmoved by words themselves. In addition, there is some evidence that psychopaths possess a more shallow understanding of the meaning of words, especially emotional words (Blair, Richel, Mitchel, Leonard, Morton, Blair, 2006).

There are two major aspects of the antisocial pursuit of power. The first is an obsession and extreme sensitivity to power relationships. Obtaining maximum power and preventing the loss of any power is of importance to psychopaths. The second aspect is the focus upon using power for destructive ends. Because

power is understood by psychopaths only in the context of victimization, in which there must always be some victim, a weaker person is meant to be exploited, as part of the natural order.

Finally, the true psychopath described by Reiber and Vetter does not feel guilt. They are capable of rationalizing away, or otherwise evading the emotion altogether. They might be able to give the outward appearance of guilt, once again as a tool, but they don't actually feel it, nor does it affect their actions. In addition to guilt, psychopaths are generally impoverished of affective reactions, often being described as "callous."

It has also been suggested that psychopaths are less developed according to Loevinger's theory of ego development (Endres, 2004). This offers some potential explanation for the traits listed earlier. For example, children are less emotionally developed, and are more able to understand power in terms of victimization (e.g. bullying) than more complex forms of power. Abrahamsen (in Reiber & Vetter, 1994 p6.) notes that the symptoms of psychopathy are most generally found among young people and that the psychopath appears to reach the unstable emotional state of the formative years of life sooner and ends it later, than normal people.

Reiber and Vetter (1994) classify psychopaths into various broad groups, ranging from schizoid psychopaths that can become shut-ins and daydreamers, to aggressive psychopaths who display unusual competitive aggression in the world. Out of all of these categories, and even crossing categories, the group that is of most interest to the criminal justice system and society are the psychopaths that ultimately commit violent crimes.

The intersection of psychopathy and violent crimes is of especial interest because it has been noted that violence by psychopaths tend not to be "crimes of passion" but rather, externally motivated (Woodworth & Porter, 2002). That is, crimes by psychopaths tend not to be emotionally driven, such as a fit of jealous rage, but instead are driven by some external reason, such as armed robbery for financial gain. Whereas other prisoners who commit violent crime are not likely to commit another act of violence upon release, violent psychopaths are demonstrably more likely (Serin, 1996; Woodworth & Porter, 2002).

Given the increased likelihood that psychopaths will re-offend after release, accurate classification of psychopaths is extremely important. As noted earlier, the PCL-R is the primary measurement device used to identify psychopaths. It is a psychological test that requires a human administrator and interpreter. The full PCL-R has an estimated administration time of 90-120 minutes for the interview, and 60 minutes for the collateral review (Pearson Assessments, 2007). The actual checklist is divided into two major factors: Factor 1, selfish, callous and remorseless use of others, and Factor 2, chronically unstable and antisocial lifestyle.

Recognizing that the resource demands of the PCL-R are great, a shorter screening version, the PCL-SV of the checklist was developed. This version has an estimated administration time of 45 minutes for the interview, and 30 minutes for the collateral review (Pearson Assessments, 2007). The PCL-SV primarily condenses the items of the PCL-R while still measuring the two factors that are the focus of the PCL-R. The PCL-SV items were also modified, where applicable, so that no access to a formal criminal record is needed. Despite these modifications, the PCL-SV has been demonstrated to have high correla-

tions with the PCL-R (Cooke, Michie, Hart & Hare, 1999).

Language and Psychopathy

Is it possible to use the language produced by criminals to identify their level of psychopathy? Research in social psychology suggests that personality traits are reflected in a person's language use. For instance, there is a significant, though weak, relationship between the Big Five personality traits (i.e., neuroticism, extroversion, openness, agreeableness, and conscientiousness) and linguistic features used, as well as a relationship between some lifestyle variables, such as the amount of smoking and drinking, and linguistic features used (Pennebaker & King, 1999). For example, high-level neurotics tend to use more first person singular (i.e. "I") and inclusive terms (e.g. with, include) than low-level neurotics (see Oberlander & Gill, 2006). These findings suggest that it is possible that language can provide us with a "window" to observe and measure psychopathic traits.

In the context of psychopathy, there has been some work showing that psychopaths process language differently than average people. For example, Day and Wong (1996) find that psychopaths show no advantage in processing emotional words in the left field of vision, while non-psychopaths do. Endres (2004) found that verbal reactions of psychopaths to standardized sentence cues also yielded some difference between psychopaths and non-psychopaths. For example, there is a significant correlation between the PCL-R and sentence completions that express power and domination (e.g. lying, possessing people), or statements that concerned with being dominated.

Less has been done, however, on studying the differences in the psycho-

path's ability to generate language. Endres (2004) notes that linguistic behavior has only been recently targeted as an indicator of psychopathy. There have been some references to behavior known as "semantic dementia" in which psychopaths actively play with the meaning of words during conversation. Reiber and Vetter (1994) give an example from The Great McGintey where the lead character says "What's everybody so upset about? What you rob, you spend, so the money goes back to the people anyway, doesn't it?" Other examples exhibit a similar pattern of making some logical sense on the surface, but logically collapse upon closer inspection of the deeper meanings of the words. Additionally, Reiber and Vetter have also identified a number of linguistic behaviors that psychopaths seem to engage in, such as a greater use of negating language (e.g. not, no, nothing). But taken together, very little empirical work has examined linguistic differences between language produced by psychopaths relative to non-psychopaths.

The primary objective of the present study is to identify and extract differences in language produced by convicted murderers. The secondary objective is to use these linguistic differences in an attempt to classify prisoners accurately into psychopathic and non-psychopathic categories. One method to extract linguistic differences for analysis would be hand coding. However, the resources required for human coding narratives would be comparable to, if not worse than, the resource demands of the PCL-R. In light of these practical concerns, we turn to the power of computer automation and the use of tools developed in the field of Natural Language Processing (NLP) as an approach that may be able to identify these linguistic correlates as well as assist in classifying whether a given text belongs to a psychopath or not.

Natural Language Processing

NLP is a rapidly developing field of research, but there are several currently practical types of analysis. Part of speech tagging, the automated process by which all the words of a text are associated with their part of speech (e.g., verb, noun, adjective, etc), is among one of the more well understood areas of NLP and its performance in identifying the part of speech of all words in a given piece of text can match human judgment up to 95-97% of all words tagged (Manning & Shutze, 2003, p.371).

More complex analyses include a broad range of analyses, such as semantic analysis, which focus on analyzing the actual meanings of words, as opposed to analyzing a word's syntactic function (e.g. part of speech) in a sentence. At its simplest level, a semantic analysis could involve "determining if a word or sentence involves food" or determining whether the word "bank" means a "river bank," or a "financial bank." The semantic analysis system, UCREL Semantic Analysis System (USAS), that is used in the study and discussed later has a reported accuracy of around 91-92% (Rayson, Archer, Piao, & McEnery, 2004).

As we try to identify more complex language features beyond part of speech, such as identifying the subject of a sentence, or a paragraph, we progressively lose accuracy. This loss of accuracy cannot be predicted and tends to fall off greatly with even minor amounts of complexity, and so for our efforts, we will stay as close as possible to using the most basic, but more reliable, NLP tools.

Our replacement of human coding with an automated computer program entails some trade-offs, primarily, the exchange of the high costs and time associated with human coders, for the speed and consistency of computers. In turn,

we sacrifice the high-level discernment of humans to identify subtle patterns of language use (e.g. sarcasm, repeated topic shifts, etc.). Although this seems to be a high cost to pay, Pennebaker and King (1999) have shown that computers, even working under such conditions, show significant agreement with human coders and can also identify significant differences between the language created by different individuals.

Pennebaker and King's approach relies upon their Linguistic Inquiry and Word Count (LIWC) system, and uses a dictionary of "over 2,200 words and word stems" that have been validated against human raters (Pennebaker & King, 1999). This lexicon is then used to "bin" words encountered in a text into various categories, such as "negative emotion" or "past-tense verb". It should be noted that a single word may count towards multiple categories at the same time. Once the text has been processed using the system, the resulting output consists of counts of the various linguistic dimensions measured by LIWC and these statistics are then used for analysis.

For our attempts to identify psychopaths using NLP methods, we will avoid relying on simple brute force computing and atheoretical regression analysis. Instead, our approach is to 1) identify various theoretically defined traits in psychopaths, 2) determine the linguistic effects that are correlated with those traits, 3) use NLP techniques that can automatically extract these linguistic features from text, and finally, 4) determine which of these linguistic features are significantly different between psychopaths and non-psychopaths.

Theory-based Linguistic Features of Psychopathic Speech

The characterization that psychopaths are on a lower level of human ego

development (Endres, 2004) suggests some types of language that should be more frequently observed in psychopaths. For instance, psychopaths may be more likely to have ego needs at the lower levels of Maslow's hierarchy (1943), feeling a greater need to fulfill lower level needs such as food, sex, shelter, and safety, whereas non-psychopaths of similar age would feel that their basic needs are more satisfied and thereby focus more on emotional, and social needs.

If such needs were more important to psychopaths than non-psychopaths, then those issues would correspondingly appear in psychopath conversations and narratives more often than non-psychopaths. Therefore, hypothesis 1 states that words related to food, or words related to money, which in modern society allows for the purchase of food, shelter, and many other basic needs, will be produced more often by psychopaths than non-psychopaths.

Also related to the characterization of being developmentally retarded compared to non-psychopaths, there should be a noted difference with respect to affect and emotion in psychopaths. Throughout Reiber and Vetter's (1994) review of the literature and their characterization of psychopaths, a common description is that psychopaths are "callous." Day and Wong (1996) found that psychopaths processed words differently than non-psychopaths, especially affective words, taking significantly longer to complete word tasks involving emotion terms. Therefore, hypothesis 2 states that psychopaths will use words that involve affect and emotion less than non-psychopaths.

As described earlier, psychopaths are sensitive to power relationships, and understand power and relationships in terms of victimization (Reiber & Vetter, 1994). We would expect to find indications of this increased salience of power in their language, with a greater use of words such as "control", "dominate"

or "lead." Also, Pennebaker, Mehl and Niederhoffer (2003) cite several studies where people of higher status tend to use fewer 1st person singular words, especially in Pennebaker and Lay (2002) where they observed Mayor Rudolph Giuliani's shift in his usage of 1st person pronouns. Mayor Giuliani's case is interesting as he was characterized by Pennebaker and Lay as being a very powerful figure and correspondingly used relatively few 1st person pronouns early in his career. Later, after personal crises his 1st person pronoun use increased, making him seem more human. Combined with Reiber and Vetter's description that psychopaths, specifically adaptive psychopaths, who can mask their psychopathic tendencies and function in the world, tend to have attitudes of omnipotence Reiber and Vetter describe as "feelings that they will never get caught" (1996, p.9). Helfgott (2004) also makes similar remarks. Considering these observations, hypothesis 3 states that because psychopaths have a sense of being in a superior position of power they will use fewer 1st person pronouns in certain situations than non-psychopaths.

Psychopaths have also been noted for their instrumental perception of the world (Woodworth & Porter, 2002). They look upon the world and their actions as tools to obtain their desires. From this trait, we can expect that psychopaths have conscious, directed goals and plans for their actions, as opposed to acting on pure emotionality. We would then expect them to use explanations when they are describing their past actions more than non-psychopaths. These explanatory statements can be partially identified by counting the number of instances that subordinating conjunctions are used (e.g. though, whereas, because, since). While the function of subordinating conjunctions also handles dependent clauses that don't explain cause-effect relationships in English (e.g. "he ran although he was injured"), they provide a rough measure of the use of

complex sentences. A slightly more refined method would focus on words that are very likely to accompany explanatory statements: "because", and "since." Of course, neither of these methods cover all explanatory statements, (e.g. "I was hungry, so I ate.") nor are they exclusive in measurement, (e.g. "It's been two years since…"). However, this technique may provide some insight into the differences in the instrumental approach to the social world across psychopaths and non-psychopaths. Therefore hypothesis 4 states that psychopaths will use more explanatory statements, as identified by subordinating conjunctions , than non-psychopaths.

A psychopath's instrumental perception of the world should also lead to more "concreteness" in their speech. That is, they are more concerned with material objects and the manipulation of the physical world than with things on a more abstract and emotional level. This is also related to the previous description that psychopaths were developmentally retarded in some way. Measuring concreteness would involve some form of noun counting. That is, we would expect more nouns to be generated in the speech of psychopaths because objects would have higher salience. Related to this would be the number of articles, both definite and indefinite, that are used. The rules for the use of articles in English are complex, but the indefinite articles, a/an, tend to introduce new nouns into discussion, while the definite article, the, tends to refer to previously mentioned nouns, (e.g. "There was a hammer, and I hung a picture up using the hammer and a nail"). It should be noted that concrete nouns, which by definition refer to nouns that can be sensed by the five senses, are objects and often appear with articles, whereas some abstract nouns can appear with articles (e.g. "a/the meeting") and some do not usually appear with articles (e.g. "a/the happiness"). Therefore counting articles is more likely to measure the number

of concrete nouns than simple noun counting, though still only indirectly. Hypothesis 5 states that psychopaths should use more concrete nouns and articles than non-psychopaths.

Finally, in Reiber and Vetter's (1994) description of psychopathic language features they note an increased use of "negators" (i.e. not, no, nothing, etc.) which they attribute to a psychopath's ability to avoid feelings with guilt. A simple tally of the occurrences of the world "not" and the contraction "n't" provides a simple measure of this. Hypothesis 6 states that we expect psychopaths to use negating words "not" and the related contractions "n't" more than non-psychopaths.

With these theorized differences between psychopaths and non-psychopaths, we now move on to doing a study to test whether these linguistic features differ between the two groups, as well as try to identify other linguistic features that may be useful for future theory generation.

Methods

The analysis involves a corpus of interview transcripts of a population of prisoners in a Canadian prison who have been convicted of murder. The prisoners who agreed to participate in the study were interviewed, and asked to describe the events of the murder (Woodworth & Porter, 2002). In total, the number of interview transcripts analyzed was 53. Of those, 14 prisoners have been identified as psychopaths using a PCL-R cutoff score of 25 and higher.

During the interviews, the prisoner was asked to describe the event of the homicide. The interviewer used an interviewing technique based on Yuille's

Step-Wise technique (see Lindberg, et al. 2003, for a more detailed description). The interviewer begins by allowing the prisoner to give a free narrative of their memories, and follows by asking more detailed questions as the interview progresses. The technique, often used in child interrogations, tries to maximize the amount of information given by the interviewee, while minimizing the possibility of contamination and leading on the part of the interviewer.

The responses from the inmates were controlled for demographics and severity of crime committed, they were all convicted of murder. Then, the text transcripts were checked to ensure that they are in a consistent style, as well as a standardized format, to help facilitate later processing. The files were the combined into two large corpora, for psychopaths and non-psychopaths and then entered into the Wmatrix system (Rayson, 2003, 2005). Wmatrix is an online tool for corpus analysis and comparison, and provides access to a number of tools to aid users in analyzing their corpora. However, we make use of only two systems for this study, the Constituent Likelihood Automatic Word-tagging System (CLAWS) part of speech tagger, and the UCREL Semantic Analysis System (USAS). In total, the number of words analyzed by the systems was 44,823 for the psychopaths, and 143,926 for the non-psychopaths.

CLAWS is a part of speech tagger, and its purpose is to tag words, which is done by appending a unique symbol to the word, with their parts of speech according to the surrounding context (e.g. the "fly" in "a house fly" is a noun, while it is a verb in "birds fly"). In general, such systems are "trained" using machine learning techniques on a large set of example text that has been tagged by human coders. The specific algorithm used depends on the specific implementation, and they are also sometimes given special linguistic rules to

help identify other word-tag relationships (e.g. rules in the spirit of "words that come after 'the' are more likely to be nouns than verbs"). Later, the trained system is used to apply tags on new input. The version of CLAWS available on Wmatrix is trained upon a 2-million word sample of the British National Corpus (Garside & Smith, 1997) and the same trained system was later used to tag the entire British National Corpus itself.

What needs to be addressed here is that the CLAWS tagger has been trained using a corpus of data based around British English, while the interview transcripts were in Canadian English. If care is not taken, this could pose a problem with tagging our data set, especially if there is a large number of words that appear in Canadian English that do no appear in the British Corpus. However, we believe that our specific use of the CLAWS output will make this a relatively minor issue. The majority of our measurements rely upon aspects of language that are more functional in nature and common across the dialects of English in question, for example, the 1st person pronouns, the word "not", the explanatory functions of "because" and "since".

The USAS part of the Wmatrix system is a semantic tagger. Instead of applying part of speech tags to a word, it uses a lexicon of about 37,000 words, and 16,000 multi-word units (idioms) and tags them using a set of tags based on McArthur's Longman Lexicon of Contemporary English (Archer, Wilson & Rayson, 2002). The system classifies words with their semantic information, for example the words "cash" and "Dollar" would be tagged with the semantic tag for "money".

After being processed by Wmatrix, the tagged files were further processed using a custom system designed for this study. The custom system first ap-

plies tags according to any predefined dictionaries and lists of words provided, for example, all instances of the word "the" can be marked with a tag such as "DEF-ARTICLE", or similarly, for sequences of words and tags matching a specific pattern, (e.g. "a cat" and "the cat" but not "any cat"). The custom system was designed because existing tools fell short of identifying certain specific features that are of interest in the present analysis of psychopathic language. For example, hypothesis 4 requires looking for explanatory statements, and while Wmatrix has the broad word classification of "subordinating conjunction", we were more interested in the specific subordinating conjunction words that are likely to be associated with explanatory statements. Using a more targeted custom tagging system in addition to the more general Wmatrix allows us to get more specific measurement.

After all tagging is complete, the text includes tags. For example, the phrase "I read the book" would appear as "I_PRN read_VVD the_AT_DEF-ARTICLE book_NN1" in the output. This output is unsuitable for human analysis, and so our custom system then aggregates the data, counts the frequencies of the various linguistic features (which can be single tags, sets of tags, some sequence of tags, etc.) that we specify, and generates output for statistical analysis.

Results

Due to the vast number of linguistic features that were used by CLAWS, USAS, and the custom tagging systems, there were over 370 unique tags and features that were analyzed. Of those, most were not relevant to the hypotheses. For those that were not directly relevant to the hypotheses but may have been of

interest to the general research question concerning psychopathic versus non-psychopathic speech, only linguistic categories that had a minimum number of observations were included in the analysis. The rationale for this minimum criterion is due to the amount of error that's inherent in even state-of-the-art NLP techniques that contributes a certain amount of observational error, and we chose to err on the side of caution. Generally, for the CLAWS parts of speech results, we only accepted entries where at least one corpus (the non-psychopath corpus or psychopath corpus) contained 1% or more observations out of all observed parts of speech. The same standard was held for our custom tags. Similarly, for the USAS results, we retained only entries where at least one corpus had 0.1% or greater. The semantic rejection standard was lower because the amount of semantic variation in a given text of any significant length is generally much larger than the variation in parts of speech; in general, unless speakers are being repetitive, the more they talk, the more semantic ground they cover.

In comparing the two corpora for significant differences, a log-likelihood calculation has been used. The calculation is comparable to a Chi-square test for frequency differences between two corpora (see Rayson, Berridge & Francis, 2004 for a detailed description). The calculation takes differing corpora sizes into account when calculating whether the proportions of tags observed in a corpus are likely to be different or not. Log-likelihood values of 6.63 corresponds to $p < 0.01$, 10.83 to $p < 0.001$, and 15.13 to $p < 0.0001$ levels of significance. It should be noted that it is the size of the corpora analyzed, 44,823 for psychopaths, and 143,926 for non-psychopaths, is what is used to calculate the log-likelihoods. See the appendix for a more detailed discussion of the log-likelihood calculation.

Table 1. Linguistic findings related to lower levels on Maslow's Hierarchy of Needs by psychopath and non-psychopath (Hypothesis 1)

		Psychopaths		Non-psychopaths		
Language Feature	Tag	Obs.	Pct. of corpus	Obs.	Pct. of corpus	Log-likeli-hood
Clothes and personal belongings	B5 (USAS)	125	0.36	294	0.26	8.37*
Food	F1 (USAS)	85	0.24	120	0.11	31.63***
Drinks	F2 (USAS)	226	0.64	413	0.36	44.16***
Money, generally	I1 (USAS)	91	0.26	173	0.15	15.64***
Business: Selling	I2.2 (USAS)	62	0.18	168	0.15	1.34
Weather	W4 (USAS)	38	0.11	53	0.05	14.49***

Note: * P<0.01, ** P<0.001, *** P<0.0001
Note: USAS denotes a tag associated with UCREL Semantic Analysis System (USAS) semantic category

Hierarchy of Needs

Hypothesis 1 stated that psychopaths would use words related to concerns associated with lower levels of Maslow's hierarchy of needs, including clothing, food, drink, and money, more than non-psychopaths. As described in Table 1, hypothesis 1 was supported for each of these categories. Language produced by psychopaths involved more terms related to drinks (psychopaths:

0.64%, non-psychopaths 0.36%), food (psychopaths: 0.24%, non-psychopaths: 0.11%), money (psychopaths: 0.26%, non-psychopaths: 0.15%), and clothes and personal items (psychopaths: 0.36%, non-psychopaths: 0.26%). Related to the theme of basic needs in food and money but not directly hypothesized, words referring to the weather were also used by psychopaths more than non-psychopaths (psychopaths: 0.11%, non-psychopaths 0.05%).

Table 2. Emotional language production across psychopaths and non-psychopaths (Hypothesis 2)

		Psychopaths		Non-psychopaths		
Language Feature	Tag	Obs.	Pct. of corpus	Obs.	Pct. of corpus	Log-likelihood
Liking	E2+ (USAS)	33	0.09	148	0.13	3.14
Calm/Violent/Angry	E3- (USAS)	166	0.47	449	0.40	3.66
Fear/bravery/shock	E5- (USAS)	23	0.07	108	0.10	2.89

Note: * P<0.01, ** P<0.001, *** P<0.0001
Note: USAS denotes a tag associated with UCREL Semantic Analysis System (USAS) semantic category

Emotion terms

Hypothesis 2, which stated that psychopaths would use fewer emotional and affect terms than non-psychopaths, was not supported. As described in Table 2, the category of "emotional/affective terms" doesn't only include words simply indicating emotional states, such as "happy", "sad" and "angry", but also can include words with positive or negative emotional content, such as

"crying" and "relief". None of the linguistic features involving emotion and affect showed significant differences between psychopaths and non-psychopaths. There were 18 different tags of emotional words that could be generated. Of those, the tag for general emotional states and processes (tag E1) was non-significant, but psychopaths used them 0.01% of the time, while non-psychopaths used them 0.03%.

Table 3. Pronoun production by psychopaths and non-psychopaths (Hypothesis 3)

		Psychopaths		Non-psychopaths		
Language Feature	Tag	Obs.	Pct. of corpus	Obs.	Pct. of corpus	Log-likelihood
1st person sing. objective personal pronoun (me)	PPIO1 (CLAWS)	340	0.97	1177	1.04	1.33
1st person sing. subjective personal pronoun (I)	PPIS1 (CLAWS)	2014	5.73	6476	5.71	0.02

Note: * P<0.01, ** P<0.001, *** P<0.0001
Note: CLAWS denotes a tag associated with the CLAWS Part of Speech Tagger

Parts of Speech Analyses

The last set of hypotheses were concerned with parts of speech usage, including pronoun patterns (Table 3), explanatory devices (Table 4), noun concreteness (Table 5), and negation (Table 6). Hypothesis 3 stated that psychopaths would use 1st person less often than non-psychopaths because people who feel that they are in a position of power use fewer 1st person pronouns was not supported in any measurements of 1st person pronoun use.

Table 4. Use of subordinating conjunction by psychopaths and non-psychopaths (Hypothesis 4)

		Psychopaths		Non-psychopaths		
Language Feature	Tag	Obs.	Pct. of corpus	Obs.	Pct. of corpus	Log-likelihood
Subordinating conjunction (e.g. if, because, unless, so, for)	CS (CLAWS)	618	1.76	1734	1.53	8.66*
Because/Since	QUAL (CUST)	166	0.37	367	0.25	15.12**

Note: * P<0.01, ** P<0.001, *** P<0.0001
Note: CLAWS denotes a tag associated with the CLAWS Part of Speech Tagger
Note: CUST denotes a tag associated with tagging from a customized tagger

Table 5. The concreteness of language (use of articles, nouns) by psychopaths and non-psychopaths (Hypothesis 5)

		Psychopaths		Non-psychopaths		
Language Feature	Tag	Obs.	Pct. of corpus	Obs.	Pct. of corpus	Log-likeli-hood
Article (e.g. the, an)	AT (CLAWS)	1529	4.35	4489	3.96	9.95*
Singular article (e.g. a, an, every)	AT1 (CLAWS)	588	1.67	1864	1.64	0.13
Singular common noun (e.g. book, girl)	NN1 (CLAWS)	3269	9.30	10196	8.99	2.79
Plural common noun (e.g. books, girls)	NN2 (CLAWS)	541	1.54	1946	1.72	5.13

Note: * P<0.01, ** P<0.001, *** P<0.0001
Note: CLAWS denotes a tag associated with the CLAWS Part of Speech Tagger

Table 6. Use of negation by psychopaths and non-psychopaths (Hypothesis 6)

		Psychopaths		Non-psychopaths		
Language Feature	Tag	Obs.	Pct. of corpus	Obs.	Pct. of corpus	Log-likeli-hood
Not, n't	XX (CLAWS)	445	1.27	1700	1.50	10.42*

Note: * P<0.01, ** P<0.001, *** P<0.0001
Note: CLAWS denotes a tag associated with the CLAWS Part of Speech Tagger

Hypothesis 4 stated that psychopaths would use more explanatory statements than non-psychopath, as indicated by subordinating conjunctions ("because", "since", and "however"), and this hypothesis was supported. Using the Wmatrix category for subordinating conjunctions (tag CS), psychopaths used those words 1.76% of the time, while non-psychopaths used them 1.53% of the time. Using a more constrained count of just the words "however", "since", and "because", psychopaths also used these terms significantly more (0.37%) than non-psychopaths used them (0.25%)

Hypothesis 5 stated that psychopaths should use more concrete nouns than non-psychopaths, where concrete nouns were measured by counting the number of articles used. Consistent with this hypothesis, psychopaths used more articles than non-psychopaths (psychopaths 4.35%, non-psychopaths 3.96%)

Hypothesis 6 stated that psychopaths would use more negating words in the form of "not" and the contraction "n't" more than non-psychopaths. Surprisingly, the observed behavior was opposite from what was expected, with psychopaths using fewer negators than non-psychopaths (psychopaths 1.27%, non-psychopaths 1.5%).

Discussion

In our attempt to identify psychopaths through their use of language, we hypothesized a number of possible linguistic behaviors that could be measured using automated computer coding techniques. The linguistic features hypothesized were varied in nature, and drew from different aspects of existing work on psychopaths. Our efforts were met with mixed results, with hypotheses that

were supported, and some that weren't.

There was general support for the hypotheses involving food, drink, shelter, money, and other basic needs. This lends support to considering the possibility that psychopaths may be in a more basic state of mental development compared to non-psychopaths as proposed by Endres (2004). Under Maslow's hierarchy of needs, the need to fulfill a lower level need takes precedence over a higher need — avoiding starvation is more than seeking social acceptance. Accordingly, while most non-psychopaths would not have to worry about basic necessities of food, drink, or shelter, psychopaths may feel a much stronger need and may feel unfulfilled enough for it to influence their thinking.

The hypothesis that psychopaths feel they are in a position of power and would use fewer 1st person pronouns as a result was not supported despite both Helfgott's (2004) Reiber and Vetter's (1994) description of feelings of omnipotence and "would never be caught" which would suggest that psychopaths do feel themselves to be in a position of power. Instead, psychopaths used words that involved power relationships (e.g. "led", "ordered", "power") significantly more than non-psychopaths (psychopaths 0.18%, non-psychopaths 0.09%). This seems to indicate that the notion of power relationships is salient in their minds while they might not feel to be in a position of power.

Why didn't "I" differentiate psychopaths and non-psychopaths? Pennebaker, Mehl and Neiberhoffer's (2003) analysis points to other situations that can affect use of the 1st person pronoun in either direction, and this situation of describing a memory of murder has no strong parallel with any of the described situations of social interaction, emotional upheaval, deception, and formal/informal situations and therefore none may apply. Meanwhile, Helfgott's (2004) measure-

ments of omnipotence involve coding statements, such as "I am big enough to kick and shake holes in the ground" (p.13) which are beyond the power of the current analysis.

The greatest puzzle in the results is that emotional content, such an important and often commented on difference between psychopaths and non-psychopaths, yielded no differences. However, this may be a result of the method of measuring emotional content within the text. By only measuring the simple quantity of such terms, we ignore one of the defining aspects of a psychopath's relationship to emotion and affect, that they have a shallow understanding (Blair, Richel, Mitchel, Leonard, Morton, Blair, 2006), as well as taking more time to process such words (Day & Wong, 1996).

"Shallow understanding" is a qualitative statement about how psychopaths understand the meanings behind the emotional words that they use, and such an analysis can prove difficult for a human coder, much less an automated system to detect. It may be that psychopaths utter a similar quantity of emotional terms as non-psychopaths, but they use them in different ways. For example, they may be able to understand and describe emotions that they feel themselves, while they may have more difficulty attributing emotion to others. Because the current analysis is unable to make that distinction, they are currently being lumped together, masking the difference.

Perhaps related to emotionality, was the finding that psychopaths mentioned kin, wife, mother, etc., very rarely in their narrations compared to non-psychopaths. This may also be related to the situation surrounding the event of the memory, in that non-psychopaths are likely to have involved their kin in the act of murder while psychopaths, acting more instrumentally, may involve

strangers. The results also indicate that psychopaths use significantly fewer proper nouns (tag NP1) in their narratives compared to non-psychopaths. Why might this be the case? One possibility is that it related to previous observations that psychopaths tend to commit crimes for instrumental reasons (see Woodworth & Porter, 2002). Woodworth and Porter found that non-psychopaths were almost evenly divided between murders that are primarily reactive (unplanned) and instrumental (planned), while psychopaths were overwhelmingly instrumental.

A non-psychopath who commits murder out of emotion is likely to be among people they know; it seems unlikely that strangers would elicit enough emotion to kill. This may give a non-psychopath more names to use in their narratives, whereas a psychopath who may have killed during a robbery for money, would have relatively fewer names to mention. However, this reasoning is speculative and requires a human coder focusing on the nature of proper nouns used (e.g. whether they were place names, or names of people) before drawing any firm conclusions.

It is surprising that the expected difference in the use of negating words, "not" and the contraction "n't", was in the opposite direction, with psychopaths using less than non-psychopaths. Reiber and Vetter (1996) cites a study by Eichler (1965) where psychopaths were found to use more negating words, in an attempt to deny feelings. The main difference between Eichler's study is that the subjects were allowed to speak about whatever topic they chose, whereas for this study, the subjects were specifically told to discuss their memory of a murder they have committed. This difference is significant in that a past memory has undeniably happened. If this is the case, non-psychopath might be more

inclined to deny their involvement, deny their feelings, perhaps out of guilt. Psychopaths however, generally described to be callous and remorseless in the literature, may feel no guilt about their past actions, and they may not feel the need to deny the events of the past. In the case of unrestricted speech the topic may involve the future or the present, the denial of a psychopath's instrumental intentions for things that have not yet happened may be important to the psychopath, and so the reverse is observed.

Unexpected differences observed

Reviewing the full results from the linguistic analysis (Appendices A, B and C) revealed a number of statistical differences between psychopathic and non-psychopathic language that were not predicted. Although not hypothesized, these differences may be important to report for future research. For example, one unexpected difference observed was the use of different verb tenses (tags starting with a "V" all relate to verbs, dividing into various tenses, with separate tags for special verbs such as "to be" and "do/did/done"). Psychopaths appeared to use more past tense verbs than non-psychopaths. Here, the term "past tense" encompasses both the simple past tense (e.g. "have done"), as well as the complex past tenses (e.g. pluperfect, "had done", future perfect "will have done"). The complement to that set of past tense verbs, we call the "non-past tense" verbs, which covers everything else, including infinitives, participles, present, future tenses, etc. When individual verb tags were aggregated together into non-past or past verbs, psychopaths produced significantly more past than non-past tense verbs. Again, this may be the result of the specific context of the transcripts used, interacting with a psychopath's inability to feel guilt. When recounting a memory, describing events in the past and pluper-

fect tenses would emphasize the completion of an event, and consequently the event's separation from the speaker, who is "in the present." Of course, tenses exist to describe the order of events in a narration, so someone would have trouble using the same tense for all verbs used, but just as writers can choose to cast a narration in a general tense overall, psychopaths on some level made a similar choice.

Another unexpected result is the amount of speech disfluencies that psychopaths use compared to non-psychopaths. Disfluencies under the CLAWS tagger is included in the category for words used in verbal discourses, which include words such as "uh", "um", "yeah", and "whoa". Because the category of words under the CLAWS system was so large, a more focused custom tagger was also used to count only instances of "uh" and "um". Both of these show very significant differences between the populations with psychopaths using "uh" and "um" for 2.25% of their corpus and non-psychopaths using 1.62%.

One use of speech disfluency in a conversation is to maintain a speaking turn, by making an utterance, while giving the speaker time to think of more to say. As such, it seems that psychopaths are groping for words, or memories of events, more than non-psychopaths. This could be for various reasons. A psychopath's more instrumental view of the world could be a cause, since if one had reasons for acting a certain way, one might want to convey that in a narration, but need extra time to recall the reasoning. Also, recall that Day and Wong's (1996) results suggest that psychopaths take longer in processing emotional terms, so it is reasonable to hypothesize that generating emotional terms may also take more effort, causing a psychopath to hesitate before speaking. Moreover, instead of processing a word given to them, psychopaths would

be forced to consider a number of words before finding one that describes an affective state, so the time required may even be longer. However, with just a simple text transcript, it is impossible to directly measure the length of time a speaker is taking to think of words to say. Speech disfluencies would play a critical indirect indicator of this phenomenon.

However, there were more results that were unexpected. Most surprising was that there was no measured difference in the number of emotion terms used by psychopaths and non-pscyhopaths despite it being a defining feature. Similarly puzzling are the observation that psychopaths used more pluperfect tense than simple past perfect and how psychopaths are more disfluent in their speech. Moreover, psychopaths used fewer negating words than non-psychopaths, the opposite of what was expected.

Limitations

One thing to note is that this study has a small sample size of 14 psychopath and 39 non-psychopaths and all the transcripts for each group was combined into a single group corpus for analysis, the total number of words analyzed was 44,823 for the psychopaths, and 143,926 for the non-psychopaths. The size of the corpora gives the log-likelihood calculation the power to find a number of significant differences, but the fact that a small number of participants doesn't change. Meanwhile, comparing individual transcripts to other transcripts would not have enough statistical power to detect anything, due to the large variances inherent in a single transcript.

Effectively, this means that we must be careful about drawing any conclusions from small but significant differences. In the case of looking at words

involving basic needs, the differences in usage is mostly within 0.2%, the difference of two words in a thousand. A single observation with a 0.1% difference does seem small to make a judgement upon, however for that hypothesis, there had been five obsevations that all supported the hypothesis in the anticipated direction. This triangulation offers us more confidence in saying that there is something there that is worth further investigation.

The literature on psychopathy frequently makes note of differences in language use by psychopaths, however the majority of the literature notes differences on the level of the meanings of words and sentences. That is, they mention statements that embody notions such as "omnipotence" (Helfgott, 2004, Reiber and Vetter, 1994) or a "shallow understanding of affective terms" (Reiber and Vetter, 1994). However, this study looks primarily the level of words and parts of speech, only going so high as to consider some multi-word idioms that are conceptually a single word unit. Despite that, it has found some features of psychopath speech that is consistent with existing theory on what defines a psychopath from a non-psychopath. However, these results are like shadows compared to looking at the higher level linguistic features of psychopaths that are out of reach.

All the mixed results observed indicate the need for further clarification of the methods used to identifying features in this paper. The strongest difference between this study and other studies on psychopaths is that the situation in the study is very specifically restricted to the psychopath's memory of an act of murder they committed. This stands in contrast to other studies in the literature which varied widely in context and content. It is likely that a number of the findings in this paper are correlated with the fact that the subjects were re-

calling memories. However, this tight control over the potential variance of the data provides a nearly ideal testing ground for the proposed method of using automated computer systems to aid in the coding task involved in separating psychopaths and non-psychopaths through their language differences. Moreover, the discrepancies with the established theory of psychopaths that arose in this study indicates that more work needs to be done in refining the measurement and operationalizations of this technique.

Conclusion

Initially, this paper set the goal of developing a system that could detect psychopaths by analyzing their language. Now looking at the results, that possibly is still out of reach. At the level of NLP technology that was applied in this paper, the possibility is remote. Recall that a number of the differences found were significant, but differed only a few tenths of a percentage. For a typical interview, this would mean an increase in use of a given linguistic feature would be only slightly more than a non-psychopath, and the effect could easily be overshadowed by individual linguistic variation.

As natural language processing methods improve, allowing for reliable analysis of language on the sentence, paragraph, and higher levels, then an automated detection system may be viable. However, ethically, the most that these systems should ever do is add one more indicator of psychopathy. Since such a designation holds many serious consequences, only qualified people doing careful analysis should ever be able to make that judgment. This effort to create an automated system is more suited toward raising awareness, much like police can become suspicious and look for probable cause to act, that may be used to screen for potential candidates to give the PCL-R.

APPENDIX A

FULL CLAWS OUTPUT

Appendix A. Full CLAWS analysis output

Appendix A. Full CLAWS analysis output

		Psychopaths		Non-psychopaths		
Language Feature	Tag	Obs.	Pct. of corpus	Obs.	Pct. of corpus	Log-likeli-hood
possessive pronoun, pre-nominal (e.g. my, your, our)	APPGE	552	1.57	2043	1.80	8.43*
article (e.g. the, no)	AT	1529	4.35	4489	3.96	9.95*
singular article (e.g. a, an, every)	AT1	588	1.67	1864	1.64	0.13
coordinating con-junction (e.g. and, or)	CC	1755	4.99	5661	4.99	0
subordinating con-junction (e.g. if, because, unless, so, for)	CS	618	1.76	1734	1.53	8.66*
singular determiner (e.g. this, that, an-other)	DD1	569	1.62	1877	1.65	0.22
general preposition	II	1721	4.89	5456	4.81	0.39
general adjective	JJ	937	2.66	3213	2.83	2.74
singular cardinal number (one)	MC1	464	1.32	1452	1.28	0.32

Appendix A. *(Continued)*

Language Feature	Tag	Psychopaths		Non-psychopaths		Log-likelihood
		Obs.	Pct. of corpus	Obs.	Pct. of corpus	
singular common noun (e.g. book, girl)	NN1	3269	9.30	10196	8.99	2.79
plural common noun (e.g. books, girls)	NN2	541	1.54	1946	1.72	5.13
3rd person sing. neuter personal pronoun (it)	PPH1	608	1.73	1776	1.57	4.39
3rd person sing. objective personal pronoun (him, her)	PPHO1	447	1.27	1286	1.13	4.26
3rd person sing. subjective personal pronoun (he, she)	PPHS1	682	1.94	2765	2.44	29.91***
1st person sing. objective personal pronoun (me)	PPIO1	340	0.97	1177	1.04	1.33
1st person sing. subjective personal pronoun (I)	PPIS1	2014	5.73	6476	5.71	0.02

Appendix A. *(Continued)*

Language Feature	Tag	Psychopaths		Non-psychopaths		Log-likelihood
		Obs.	Pct. of corpus	Obs.	Pct. of corpus	
2nd person personal pronoun (you)	PPY	676	1.92	2109	1.86	0.57
general adverb	RR	1335	3.80	4213	3.71	0.48
infinitive marker (to)	TO	399	1.13	1262	1.11	0.12
interjection (e.g. oh, yes, um)	UH	1307	3.72	3300	2.91	54.15***
was	VBDZ	918	2.61	3290	2.90	8.11*
modal auxiliary (can, will, would, etc.)	VM	426	1.21	1319	1.16	0.54
base form of lexical verb (e.g. give, work)	VV0	798	2.27	3087	2.72	21.68***
past tense of lexical verb (e.g. gave, worked)	VVD	2020	5.74	6012	5.30	9.67*
-ing participle of lexical verb (e.g. giving, working)	VVG	647	1.84	2364	2.08	8.07*
infinitive (e.g. to give... It will work...)	VVI	1116	3.17	3607	3.18	0

Appendix A. *(Continued)*

Language Feature	Tag	Psychopaths		Non-psychopaths		Log-likelihood
		Obs.	Pct. of corpus	Obs.	Pct. of corpus	
past participle of lexical verb (e.g. given, worked)	VVN	409	1.16	1378	1.21	0.6
not, n't	XX	445	1.27	1700	1.50	10.42*
singular letter of the alphabet (e.g. A,b)	ZZ1	436	1.24	1370	1.21	0.23

Note: * P<0.01, ** P<0.001, *** P<0.0001

APPENDIX B

FULL USAS OUTPUT

Appendix B. Full USAS output

Appendix B. Full USAS output

Language Feature	Tag	Psychopaths		Non-psychopaths		Log-likelihood
		Obs.	Pct. of corpus	Obs.	Pct. of corpus	
General actions, making etc.	A1.1.1	396	1.13	1191	1.05	1.44
Damaging and destroying	A1.1.2	49	0.14	183	0.16	0.85
Open/closed; Hiding/Hidden; Finding; Showing	A10-	20	0.06	116	0.10	6.72*
Open/closed; Hiding/Hidden; Finding; Showing	A10+	68	0.19	234	0.21	0.22
Degree: Non-specific	A13.1	42	0.12	162	0.14	1.1
Degree: Boosters	A13.3	129	0.37	388	0.34	0.47
Degree: Approximators	A13.4	82	0.23	253	0.22	0.12
Exclusivizers / particularizers	A14	231	0.66	825	0.73	1.91
Affect: Modify, change	A2.1+	143	0.41	390	0.34	2.87
Affect: Cause/Connected	A2.2	93	0.26	288	0.25	0.12
Being	A3+	1135	3.23	4050	3.57	9.2*

Appendix B. *(Continued)*

Language Feature	Tag	Psychopaths		Non-psychopaths		Log-likelihood
		Obs.	Pct. of corpus	Obs.	Pct. of corpus	
Generally kinds, groups, examples	A4.1	79	0.22	208	0.18	2.29
Particular/general; detail	A4.2+	51	0.15	129	0.11	2.08
Evaluation: Good/bad	A5.1+	251	0.71	715	0.63	2.81
Evaluation: Authenticity	A5.4+	64	0.18	159	0.14	2.99
Comparing: Similar/different	A6.1-	144	0.41	426	0.38	0.79
Definite (+ modals)	A7+	427	1.21	1333	1.18	0.35
Getting and giving; possession	A9-	54	0.15	149	0.13	0.94
Getting and giving; possession	A9+	601	1.71	1755	1.55	4.36
Anatomy and physiology	B1	232	0.66	995	0.88	16.17***
Health and disease	B2-	27	0.08	199	0.18	19.93***
medicines and medical treatment	B3	49	0.14	206	0.18	2.93

Appendix B. *(Continued)*

Language Feature	Tag	Psychopaths		Non-psychopaths		Log-likelihood
		Obs.	Pct. of corpus	Obs.	Pct. of corpus	
Clothes and personal belongings	B5	125	0.36	294	0.26	8.37*
Liking	E2+	33	0.09	148	0.13	3.14
Calm/Violent/Angry	E3-	166	0.47	449	0.40	3.66
Fear/bravery/shock	E5-	23	0.07	108	0.10	2.89
Food	F1	85	0.24	120	0.11	31.63***
Drinks	F2	226	0.64	413	0.36	44.16***
Crime, law and order: Law and order	G2.1	146	0.42	444	0.39	0.38
Crime, law and order: Law and order	G2.1-	80	0.23	153	0.13	13.48**
Warfare, defence and the army; weapons	G3	109	0.31	341	0.30	0.08
Architecture and kinds of houses and buildings	H1	109	0.31	374	0.33	0.33
Parts of buildings	H2	107	0.30	581	0.51	27.53***
Residence	H4	60	0.17	238	0.21	2.13

Appendix B. *(Continued)*

Language Feature	Tag	Psychopaths		Non-psychopaths		Log-likelihood
		Obs.	Pct. of corpus	Obs.	Pct. of corpus	
Furniture and household fittings	H5	70	0.20	167	0.15	4.3
Money generally	I1	91	0.26	173	0.15	15.64***
Business: Selling	I2.2	62	0.18	168	0.15	1.34
Work and employment: Generally	I3.1	51	0.15	118	0.10	3.75
Life and living things	L1-	140	0.40	459	0.40	0.03
Moving, coming and going	M1	710	2.02	2524	2.23	5.33
Putting, taking, pulling, pushing, transporting etc.	M2	257	0.73	772	0.68	0.97
Vehicles and transport on land	M3	178	0.51	597	0.53	0.21
Location and direction	M6	707	2.01	2168	1.91	1.35
Places	M7	97	0.28	229	0.20	6.35
Remaining/stationary	M8	76	0.22	202	0.18	2.01
Numbers	N1	316	0.90	912	0.80	2.85

Appendix B. *(Continued)*

Language Feature	Tag	Psychopaths Obs.	Psychopaths Pct. of corpus	Non-psychopaths Obs.	Non-psychopaths Pct. of corpus	Log-likelihood
Measurement: Size	N3.2-	35	0.10	80	0.07	2.75
Measurement: Speed	N3.8+	34	0.10	68	0.06	4.88
Linear order	N4	356	1.01	1031	0.91	3.02
Quantities	N5	148	0.42	427	0.38	1.34
Quantities	N5-	49	0.14	137	0.12	0.72
Entirety; maximum	N5.1+	151	0.43	519	0.46	0.48
Quantities	N5+	77	0.22	269	0.24	0.39
Quantities	N5++	57	0.16	189	0.17	0.03
Frequency etc.	N6+	46	0.13	160	0.14	0.21
Objects generally	O2	253	0.72	707	0.62	3.75
Shape	O4.4	36	0.10	53	0.05	12.28**
Linguistic actions, states and processes; communication	Q1.1	26	0.07	164	0.14	11.82**
Paper documents and writing	Q1.2	16	0.05	126	0.11	14.18**
Speech etc: Communicative	Q2.1	502	1.43	1868	1.65	8.3*
Speech acts	Q2.2	295	0.84	855	0.75	2.47
People	S2	64	0.18	254	0.22	2.29

Appendix B. *(Continued)*

Language Feature	Tag	Psychopaths		Non-psychopaths		Log-likelihood
		Obs.	Pct. of corpus	Obs.	Pct. of corpus	
People: Female	S2.1	39	0.11	102	0.09	1.2
People: Male	S2.2	100	0.28	423	0.37	6.27
Relationship: General	S3.1	64	0.18	255	0.22	2.38
Relationship: Intimate/sexual	S3.2	56	0.16	188	0.17	0.07
Kin	S4	87	0.25	604	0.53	53.84***
Obligation and necessity	S6+	99	0.28	256	0.23	3.37
Power, organizing	S7.1+	65	0.18	107	0.09	17.06***
Permission	S7.4+	35	0.10	106	0.09	0.10
Helping/hindering	S8+	49	0.14	139	0.12	0.59
Religion and the supernatural	S9	36	0.10	208	0.18	11.91**
Time	T1	116	0.33	355	0.31	0.24
Time: General: Past	T1.1.1	79	0.22	267	0.24	0.13
Time: General: Present; simultaneous	T1.1.2	97	0.28	260	0.23	2.35

Appendix B. *(Continued)*

Language Feature	Tag	Psychopaths		Non-psychopaths		Log-likelihood
		Obs.	Pct. of corpus	Obs.	Pct. of corpus	
Time: General: Future	T1.1.3	154	0.44	436	0.38	1.9
Time: Momentary	T1.2	49	0.14	108	0.10	4.64
Time: Period	T1.3	335	0.95	905	0.80	7.48*
Time: Beginning and ending	T2-	102	0.29	239	0.21	6.98*
Time: Beginning and ending	T2+	88	0.25	287	0.25	0.01
Time: Beginning and ending	T2++	51	0.15	207	0.18	2.26
Time: Old, new and young; age	T3	34	0.10	105	0.09	0.05
Weather	W4	38	0.11	53	0.05	14.49**
Mental actions and processes	X2	31	0.09	119	0.10	0.77
Thought, belief	X2.1	222	0.63	736	0.65	0.13
Knowledge	X2.2+	392	1.11	1425	1.26	4.49
Investigate, examine, test, search	X2.4	51	0.15	113	0.10	4.71
Sensory: Sound	X3.2	53	0.15	157	0.14	0.28

Appendix B. *(Continued)*

Language Feature	Tag	Psychopaths		Non-psychopaths		
		Obs.	Pct. of corpus	Obs.	Pct. of corpus	Log-likelihood
Sensory: Sound	X3.2-	77	0.22	273	0.24	0.55
Sensory: Sight	X3.4	143	0.41	467	0.41	0.02
Wanting; planning; choosing	X7+	125	0.36	412	0.36	0.04
Trying	X8+	81	0.23	264	0.23	0.01
Geographical names	Z2	39	0.11	139	0.12	0.31
Discourse Bin	Z4	1611	4.58	4308	3.80	40.05***
Grammatical bin	Z5	9495	27.00	30229	26.65	1.24
Negative	Z6	502	1.43	1840	1.62	6.59
If	Z7	99	0.28	323	0.28	0.01
Pronouns etc.	Z8	7510	21.36	24852	21.91	3.78
Unmatched	Z99	657	1.87	2495	2.20	14.29**

Note: * P<0.01, ** P<0.001, *** P<0.0001

APPENDIX C

CUSTOM TAGS

Appendix C. Full custom tag analysis output

		Psychopaths		Non-psychopaths		
Language Feature	Tag	Obs.	Pct. of corpus	Obs.	Pct. of corpus	Log-likeli-hood
"Uh" and "um"	DIS-FLUENCIES	1010	2.25	2329	1.62	73.699***
Verbs that are not past tense	NONPAST-VERBS	2,414	5.39	8,400	5.84	12.29**
Past tense verbs (incl. complex past tenses)	PASTVERBS	3149	7.03	9486	6.59	9.53*
"Because" and "since"	QUAL	166	0.37	367	0.25	15.13***

Note: * P<0.01, ** P<0.001, *** P<0.0001

APPENDIX D

LOG-LIKELIHOOD CALCULATION

The calculation of the log-likelihood as described by Rayson and Garside (2000) is as follows, and revolves around building the following contingency table.

	Corpus one	Corpus two	Total
Freq. of word	a	b	$a + b$
Freq of other words	$c - a$	$d - b$	$c + d - a - b$
Total	c	d	$c + d$

In this table, a and b are the observed frequencies (O) for a given word in their respective corpora. In order to calculate the expected values (E) of a and b, we use the following formula:

$$E_i = \frac{N_i \sum_i O_i}{\sum_i N_i}$$

Where $N_1 = c$ and $N_2 = d$.

Therefore $E_1 = c * (a + b)/(c + d)$ and $E_2 = d * (a + b)/(c + d)$

The log-likelihood is the calculated as:

$$-2ln\lambda = 2 \sum_i O_i ln\left(\frac{O_i}{E_i}\right)$$

which is equivalent to: $LL = -2 * ((a * log\,(a/E_1)) + (b * log\,(b/E_2)))$

REFERENCES

Archer, D., Wilson, A. & Rayson P. (2002). Introduction to the USAS category system. Benedict project report, October 2002, Retrieved June 25, 2007 from http://www.comp.lancs.ac.uk/ucrel/usas/

Cooke, D. J., Michie, C., Hart, S.D. & Hare, R. D. (1999). Evaluating the screening version of the hare psychopathy checklist-revised (PCL:SV): An item response theory analysis. *Psychological Assessment, 11,* 3-13.

Day, R., & Wong, S. (1996). Anomalous perceptual asymmetries for negative emotional stimuli in the psychopath. *Journal of Abnormal Psychology, 105,* 648-652.

Endres, J. (2004). The language of the psychopath: characteristics of prisoners performance in a sentence completion test. *Criminal Behaviour and Mental Health, 14,* 214-226.

Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4, in Garside, R., Leech, G., and McEnery, A. (eds.) *Corpus Annotation: Linguistic Information from Computer Text Corpora.* Longman, London, 102-121.

Hare, R. D. (1991). Manual for the Psychopathy Checklist-Revised. Toronto: Multi-Health Systems. Helfgott, J. B. (2004). Primitive defenses in the language of the psychopath: Considerations for forensic science. J*ournal of Forensic Psychology Practice, 4,* 1-29.

Lindberg, M. A., Chapman, M. T., Samsock, D., Thomas, S. W. & Lindberg, A.W. (2003) Comparisons of three different investigative interview techniques with young children. *The Journal of Genetic Psychology, 164,* 5-28.

Maslow, A. (1943). A theory of human motivation. *Psychological Review, 50,* 370-396.

Oberlander, J. and Gill, A.J. (2006) Language with character: A corpus-based study of individual differences in e-mail communication. *Discourse Processes, 42,* 239-270.

Pennebaker, J.W., Mehl, M.R., & Niederhoffer, K.G. (2003). Psychological Aspects of Natural Language Use: Our words, our selves. *Annual Review of Psychology 2003, 54,* 547-577.

Pennebaker. J.W. & Lay, T.C., (2002). Language Use and Personality during Crises: Analyses of mayer Rudolph Giulianis press conferences. *Journal of Research in Personality, 36,* 271-282.

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

Pearson Education, Inc. (n.d.). Hare Psychopathy Tests: Pearson Assessments, PCL-R, PCL:SV, PCL:YV, P-SCAN, Retrieved December 20, 2006, from http://www.pearsonassessments.com/tests/hare.htm

Rayson, P. (2005) Wmatrix: a web-based corpus processing environment, Computing Department, Lancaster University.
http://www.comp.lancs.ac.uk/ucrel/wmatrix/

Rayson, P. (2003). Matrix: A statistical method and software tool for linguistic analysis through corpus comparison. Ph.D. thesis, Lancaster University.

Rayson, P., Archer, D., Piao, S. L. & McEnery, T. (2004). The UCREL semantic analysis system. In proceedings of the workshop on *Beyond Named Entity Recognition Semantic labelling for NLP tasks* in association with *4th International Conference on Language Resources and Evaluation (LREC 2004),* 25th May 2004, Lisbon, Portugal, 7-12. Retrieved June 21, 2008 from http://ucrel.lancs.ac.uk/usas/

Rayson, P. & Garside, R. (2000). Comparing corpora using frequency profiling. In proceedings of the workshop on Comparing Corpora, held in conjunction with *the 38th annual meeting of the Association for Computational Linguistics (ACL 2000)*. 1-8 October 2000, Hong Kong, 1-6.

Rayson, P., Berridge, D. & Francis, B. (2004). Extending the Cochran rule for the comparison of word frequencies between corpora. In Purnelle G., Fairon C., Dister A. (eds.) Le poids des mots: Proceedings of the 7th International Conference on Statistical analysis of textual data (JADT 2004): Vol.2 (pp. 926-936.). Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.

Rieber, R.W. & Vetter, H. (1994). The language of the psychopath. *Journal of Psycholinguistic Research, 23,* 1-28.

Serin, R. C. (1996) Violent recidivism in criminal psychopaths. *Law and Human Behavior, 20,* 207-217.

Woodworth, M., & Porter, S. (2002). In cold blood: Characteristics of criminal homicides as a function of psychopathy. *Journal of Abnormal Psychology, 111,* 436-445.