Personality as Manifest in Word Use: Correlations With Self-Report, Acquaintance Report, and Behavior

Lisa A. Fast and David C. Funder
University of California, Riverside

The use of words is one of the most direct means of expressing thoughts and feelings. However, past studies have had limited success in correlating word use with personality. The purpose of the present study was to identify categories of word use relevant to personality using a broad range of personality data. Using data from 181 participants, the present study correlated word use within a 1-hr life history interview with self-judgments of personality, judgments of personality provided by close acquaintances (who were not exposed to the language sample), and behavioral ratings based on direct observation from a context entirely separate from that from which the language sample was derived. Several categories of word use yielded a large number of correlates with self- and acquaintance personality ratings and behavior. It is suggested that word use is related to personality to a larger degree than previously observed and deserves increased attention as a source of data in personality assessment.

Keywords: personality, language, word use, verbal behavior

Individuals vary in the words they use, and it is reasonable to expect that this variation would reflect psychological differences. A person who speaks in terms that convey certainty and confidence, or who consistently refer to sexual matters, is probably not doing so haphazardly. These word choices may reflect stable psychological processes; that is, they may be reflective of personality. In fact, the very definitions of personality and language imply a relationship between them. Personality is an individual's enduring patterns of thought, emotion, and behavior (Funder, 2007), and language is a systematic means of conveying thoughts and emotions (Merriam-Webster Online Dictionary, 2005). A core concern for personality psychology involves identifying observable, behavioral manifestations of constructs (Funder, 2001), and language may offer one promising possibility.

Longstanding psycholinguistic theory suggests that individuals produce language in four stages: (a) conceptualizing an idea to be expressed, (b) formulating a linguistic plan, (c) articulating the plan, and (d) monitoring articulation (Carroll, 1999). Freud (1916/1964) proposed that personality is encoded in language at the first two stages. He argued that personality involves the particular thoughts and feelings ("ideas") that are most available to an individual and that those primed thoughts and feelings will appear in language. In other words, personality has a top-down influence on the ideas that are conceptualized, and on how these ideas are in turn put into words. Although Freud's theory is certainly interesting, it does not yield specific predictions about how personality might be expressed in language. For example, what aspects of language should be most relevant to personality?

Several methods of systematically and quantitatively coding language have been developed in the social sciences, ranging from contextual human judgments of thematic content (e.g., Gottschalk & Gleser, 1969) to computer-based approaches that are devoid of context (e.g., LIWC; Pennebaker, Francis, & Booth, 2001; for a comprehensive review of methods, see Pennebaker, Mehl, & Niederhoffer, 2003). Deciding between these methods involves a tradeoff. Human coders are able to evaluate language in context because these judgments are based on larger linguistic units (e.g., sentences or phrases) but require extensive time and resources. In contrast, computer-based approaches are more efficient and objective but disregard contextual nuance because they are based on smaller linguistic units (e.g., words). The present focus was on a computer-based approach called the *word count method*.

The word count method involves predefining word categories of interest (e.g., emotion words, certainty words, and achievement words) and counting the proportion of words in a language sample that fall into those categories. Given the loss in context when words are the unit of analysis, one might wonder how useful this approach could be. For example, how often does the word *happy* from the sentence "I am *not* happy" get misclassified as a happy word? However, this may not be a misclassification because there is a subtle, yet important difference between saying, "I am not happy" versus "I am sad." The former suggests that an individual is thinking along a continuum from happy to not happy, whereas the latter suggests that an individual is thinking along a continuum of sadness. Someone who is not happy is not necessarily sad. Moreover, Berry, Pennebaker, Mueller, and Hiller (1997) compared word count classifications using human raters with a

computer-based word count program and found that computer-based word counts do not result in a substantial number of incorrect classifications of words.

There are several reasons to believe that word counts (or word use) might be a promising method for examining the relationship between personality and language. First, word use may offer psychological information that is not obvious to the naked eye (or ear). Pennebaker et al. (2003) suggested that word counts provide linguistic information "from a distance" because in ordinary social situations, people are so busy comprehending and responding to what is being said that they are unable to monitor word use as either speaker or listener. For example, imagine trying to count the number of self-references and positive emotion words a person uses throughout a conversation, while simultaneously comprehending and participating in the conversation. It would not be easy and might not even be possible. Chung and Pennebaker (2007) argued that this unique vantage point allows access into the subtle patterns of language through a technique that is more reliable and efficient than "the most conscientious of human coders" (p. 356).

Second, the usual psychological interpretation of a word count score is congruent with the very definition of personality. Researchers generally have assumed that an individual's score for a given word category reflects the degree to which that category is important to him or her (Stone, Dunphy, Smith, & Ogilvie, 1966). For example, the self may be especially important to individuals who use more self-references (e.g., *I* and *me*) than others, and emotion might be more important to individuals who use more emotion words (e.g., *enjoy* and *sad*). Pennebaker and Stone (2003) found that use of positive emotion words increases with age, whereas use of negative emotion words and self-references decrease with age, suggesting that older individuals are less self-focused and experience more positive and less negative emotion.

A final reason why the word count method may be appropriate for personality research is that growing evidence indicates that individual differences in written and spoken word use are stable across time and context. Pennebaker and King (1999) obtained multiple writing samples per author from a number of sources and found a surprising degree of consistency of word use across a variety of linguistic dimensions. The mean alpha coefficient across 72 dimensions was .59. Similarly, Mehl and Pennebaker (2003) obtained spoken samples of word use over 4 weeks across a variety of contexts. Of the 23 word categories they examined, 16 categories had retest correlations of at least .24, with particularly high correlations for word categories especially relevant to spoken language (e.g., swear words). Considering that the average high school graduate knows about 45,000 words (Nagy & Anderson, 1984), yet individuals pull from their personal lexicons in stable ways, it seems that some form of internal organization is at work, and perhaps personality is part of this internal organization.

Although there are good reasons to believe that the word count approach may be a good candidate for studying the relationship between personality and language, the empirical question is: Do simple word counts tell us something about personality? This question can be evaluated by examining three main types of criteria, including the degree to which word use correlates with self-judgments of personality, others' judgments of personality, and observed behavior. To the extent that these correlations form a psychologically coherent pattern, it can reasonably be assumed that word counts are relevant to personality. A thin spread of research has addressed these relations, mostly focusing on the relationship between self-reports of personality and word use, and the results have been variable. As Pennebaker et al. (2003) observed, "Although self-reports of personality are often associated with word use, the magnitudes of the relationships are surprisingly small" (p. 559).

The strongest correlations with self-reports have generally been found with pronouns. In particular, self-references (e.g., *I* and *me*) are the most studied type of pronouns, and their use has been positively related to self-reported Machiavellianism, depression, narcissism, and introversion (Ickes, Reidhead, & Patterson, 1986; Weintraub, 1989; Gill & Oberlander, 2002, respectively; Raskin & Shaw, 1988). On the basis of research and practice with clinical samples, Weintraub (1989) proposed that individuals who use a relatively high number of self-references are self-preoccupied, individuals who use a moderate number of self-references are autonomous, and those who use self-references infrequently are relatively detached and impersonal. This suggests that Machiavellianism, depression, narcissism, and introversion involve a high degree of self-focus. In contrast, self-monitoring scores are negatively related to use of self-references but positively related to use of other-references (e.g., *you* and *we*), suggesting that self-monitoring involves a high degree of self-detachment and focus on others (Ickes et al., 1986).

Less research examines correlates between self-reports of personality and word categories beyond pronouns, and the results have not been promising. Gill and Oberlander (2002) found that self-reported extraversion is positively related to using a higher total number of words and fewer references to numbers (indicating lack of concrete or precise language), but the authors were not impressed with the magnitudes of these relationships. Pennebaker and King (1999) analyzed language samples along 15 word categories and, using factor analysis, identified four broad word categories, including Immediacy, Making Distinctions, The Social Past, and Rationalization. There were marginal to no relationships between these categories and self-reports of positive and negative affect, self-esteem, and the Big Five personality traits. Moreover, when these broad word categories were correlated with behavior, only a handful of correlations reached significance. Use of immediacy words (e.g., articles and self-references) was related to classroom participation and final grades, and use of rationalization words (e.g., causation words and negative emotion words) was related to self-reported affiliative behaviors, such as preferences for working on a project within a group rather than alone.

Although attempts to correlate word use with self-reports of personality and behavior have had limited success, examinations of correlates of others' reports of personality have seemed more promising. Mehl, Gosling, and Pennebaker (2006) used the electronically activated recorder (EAR; Mehl, Pennebaker, Crow, Dabbs, & Price, 2001) to capture word use over 2 days and asked judges to rate targets' personalities by listening to their EAR recordings. These stranger ratings of the Big Five personality traits (based solely on the targets' language sample) and self-reports of the Big Five were correlated with 23 categories of word use. Overall, use of pronouns, swear words, negative emotion words, and past tense verbs had the strongest correlations. For example, individuals who were heard to use more swear words were described as lower in Agreeableness, Conscientiousness, and Openness, and higher in Extraversion. This confirms the stereotype that

those who swear a lot are outgoing, yet uncultivated and rebellious. As in past studies, however, self-reports of personality had relatively weak correlations with word use. Also, a distinctive finding was that many of the correlations were gender specific, suggesting that attention should be given to possible gender differences and similarities when examining the relationship between personality and word use.

The Present Study

Many of the previously mentioned studies concluded with the observation that important steps for this research involve examining a broader range of personality data in relation to word use (Pennebaker et al., 2003) and identifying the word categories that are relevant to personality (Mehl et al., 2006; Pennebaker & King, 1999; Pennebaker & Stone, 2003). Moreover, the two studies that examined the broadest range of word categories in relation to the broadest range of personality variables (Mehl et al., 2006; Pennebaker & King, 1999) may have overlooked important word categories because of the criteria used to determine what categories to focus on. For example, both of these studies excluded word categories with alpha reliabilities less than .60. Although the degree to which words are used consistently is an important issue, it may be more fruitful to identify word categories on the basis of validity concerns. In response, we identify categories of word use in the present study that are relevant to personality using criteria that are based on validity in a broad data set gathered by the Riverside Accuracy Project–Phase II (RAP-II).

This data set allows us to expand the range of personality variables that have been examined in relation to word use. First, 181 target participants described themselves and were described by two informants along 100 characteristics of personality. Past studies focused on global personality attributes, such as the Big Five, and because word use is relatively molecular-level phenomenon, it may be useful to match the level of analysis and examine narrower aspects of personality in relation to word use. Second, the informants were people who knew the targets well in everyday life, yet they had no exposure to the targets' language samples. In contrast, most previous studies used other's judgments of personality that were based solely on the language sample (Mehl et al., 2006). Such studies may address whether people infer personality from language, but they cannot address whether word use is associated with personality as judged by acquaintances on the basis of frequent, ordinary, daily interaction. Third, the behavior of each target was coded along 64 dimensions, and we used real behavior as directly observed and independently coded by trained raters in the present study, as opposed to self-reported behavior (Pennebaker & King, 1999). Finally, the range of word use categories examined so far is expanded by assessing language samples along 66 different word categories.

With these data, categories of word use that are relevant to personality are identified in the present study by examining the following questions: (a) What word categories are related to self- and acquaintance ratings of personality? (b) What categories of word use have a similar pattern of personality correlates across gender? Past studies suggest that there may be gender differences in the relationship between word use and personality (Mehl et al., 2006), and word categories that have similar patterns of correlates across gender are identified in the present study. (c) Of the word categories related to self- and/or acquaintance ratings of personality, which of them are also related to behavior? (d) To what degree are the word categories used consistently across the life history interview?

Method

The data for the present study were collected as part of a larger research project, the RAP-II, designed to examine the factors involved in accurate personality judgment. The complete project put participants through several different sessions and procedures, and because some participants missed one or more of these sessions, the sample size for particular analyses varies. Although this is the fifth article to come out of the RAP-II data set, the analyses do not overlap with previous projects (Letzring, Block, & Funder, 2005; Letzring, Wells, & Funder, 2006; Vazire & Funder, 2006; Wagerman & Funder, 2007) or with future planned projects.

Participants

Data from 511 undergraduate students (181 target participants and 330 acquaintances) from the University of California, Riverside, who participated in RAP-II, were included in the present study. One hundred eighty-one target participants (90 women, 91 men) were recruited through the placement of flyers throughout the campus and announcements made in undergraduate psychology courses. Each of them was paid $10/hr to participate and could earn up to $100 if he or she showed up for every session and completed every procedure. The ethnic breakdown for targets was 41% Asian, 22% Hispanic, 14% Caucasian, 12% African American, and 11% "other" or not specified.

Acquaintances

Each target participant was asked to nominate 2 acquaintances who knew him or her well and were available in the area. These acquaintances (182 women, 148 men) were then contacted by project staff and scheduled to visit the lab to provide personality judgments of the target participants with whom they were acquainted. The acquaintances were paid $10/hr for their participation. The average length of acquaintanceship between targets and acquaintances was 34 months (SD = 43 months, range = 1–407 months). Of the acquaintances, 84% were nonromantic friends, 10% were romantic partners, 2% were family members, 2% were friends from work, and 2% "other"/not specified.

Overview of Procedures

Participation for target participants involved four separate laboratory sessions and the completion of several take-home packets. Only the lab visits and materials relevant to the present study are reviewed. During the first session, targets were given a take-home packet consisting of several personality questionnaires. They were also asked for the names and contact information for two local acquaintances, as described above. In the second session, targets were placed in a three-person group, and the group interactions were videotaped so that the behavior of each interactant could be coded. In the final visit, targets participated in a 1-hr life history interview conducted by a clinically trained psychologist.

Measures

The California Adult Q-set. The California Adult Q-set (CAQ; Block, 1961, as modified for use by nonprofessionals by Bem & Funder, 1978) includes 100 items describing a broad range of personality characteristics (e.g., "is critical, skeptical, not easily impressed"; "is charming"; and "is sympathetic and considerate"). In the present study, the self and acquaintances rated each Q-item separately on a Likert-type scale ranging from 1 (*extremely uncharacteristic*) to 9 (*extremely characteristic*), gathering ratings from 2 acquaintances when possible. Two acquaintances provided ratings for 154 of the target participants, and 22 targets were described by a single acquaintance (5 targets had no acquaintances). For targets with 2 acquaintances, a composite score was created by averaging the two ratings for each CAQ item. The average interrater reliability per item is intraclass $r = .29$ ($SD = .15$).

The Riverside Behavioral Q-sort. The Riverside Behavioral Q-sort (RBQ; Funder, Furr, & Colvin, 2000) consists of 64 items that describe a broad range of socially meaningful behaviors (e.g., "tries to control the interaction," "acts playful," and "expresses insecurity"). RBQ items describe behavior at a midlevel of generality between micro-level behavior (e.g., number of smiles) and macro-level impressions (e.g., is successful). The RBQ was used to code behavior from a three-person social interaction. Each targets' behavior was rated by four trained coders who sorted the items into a nine-category, forced-choice, quasinormal distribution ranging from 1 (*extremely uncharacteristic*) to 9 (*extremely characteristic*). The four scores on each item were averaged to form a composite for each target (mean $\alpha = .61$, $SD = .17$).

Three-Person Interactions

Each of the targets participated in a videotaped laboratory interaction with two other people they had never met before. The gender composition of the interaction groups was counterbalanced so that an equal number of groups were composed of all men (MMM), all women (FFF), 2 women and 1 men (FFM), and 2 men and 1 woman (MMF). Targets were quasirandomly assigned to one of five different interaction conditions, ranging from minimal interaction time to 3 hr (see Letzring et al., 2006). Only the data from four of these conditions were used because the minimal interaction condition (in which participants silently completed personality questionnaires) did not allow a range of behavioral display. Preliminary analyses determined that correlations with behavioral ratings from the RBQ were more similar than different across the four conditions, so all of the conditions were combined for subsequent analyses.[1]

One-Hour Life History Interviews

Targets individually participated in a 1-hr long life history interview conducted by one of four clinically trained (e.g., master of social work, master of arts in counseling, or doctorate in clinical psychology) and licensed psychologists who had experience with clients of the same age as our college student sample. The interviews were videotaped with consent of the participants who were told that they would be interviewed by a "professionally trained interviewer" to diminish demand characteristics from the knowledge that they were seeing a clinical psychologist.

The clinicians conducted a semistructured interview adapted from a protocol used for many years by the Institute of Personality Assessment and Research (IPAR; Craik et al., 2002). The interview topics in the present study were adapted to better apply to college students and to capture a broad range of personality-relevant information without explicitly asking about sensitive topics and risky behaviors. All participants were asked the same questions in the same order. Interviews started by asking the participant, "Tell me something about yourself." The topics covered throughout the interview include college and academic experiences, future plans, interpersonal relationships, and childhood and family history. Finally, the interview concluded by asking the participant to "Describe a defining event in your life that had a significant impact on or changed your life in some way."

Text Analysis Procedure

Each target's life history interview was transcribed and analyzed separately (after deleting everything the interviewer said) using the Linguistic Inquiry and Word Count program (LIWC; Pennebaker et al., 2001). LIWC was originally developed to identify characteristics of linguistic expression that are related to self-disclosure and health (Berry et al., 1997). Over time, this has led to the development of numerous word categories, such as positive and negative emotion, cognitive processes, social processes, achievement, religion, physical processes, leisure, and language composition. A total of 87 categories and linguistic elements (e.g., commas, periods, and questions marks) can be analyzed with the LIWC. Linguistic elements and categories that are difficult to transcribe from spoken language (e.g., commas, periods, and nonfluencies such as "um" and "huh") and word categories dealing with verbal ability (e.g., words longer than six letters) were excluded in the present study, leaving 66 dimensions to be analyzed (for a complete list, see Appendix A).

The LIWC analyzes language on a word-by-word basis. It has a master dictionary that consists of over 2,200 words and word stems, and these words are assigned to several subdictionaries on the basis of word categories such as the ones just mentioned. For each target, the LIWC counts the words in the language sample that match each of the 66 categories. Scores for each category are expressed as percentages, or a proportion of words that match to the total number of words used. For example, if a target used 10 words that fall into the leisure word category and spoke a total of 100 words, then that target's score for leisure would be 10% or .10. On average, 72% ($SD = 6\%$) of the words spoken by participants in the interview fell within the LIWC word categories, which is comparable to the 75% ($SD = 8\%$) found by Pennebaker et al. (2001) across 43 studies.

Results

The first research question examined was: What word categories are related to self- and acquaintance ratings of personality? To

[1] Ratings from the RBQ were de-meaned within condition and correlated with word use, and ratings that were not de-meaned were correlated with word use. Both of these methods yielded approximately the same number of significant correlates, and the pattern of correlates were highly similar. Therefore, reported analyses are based on behavioral ratings with means.

Table 1
Word Categories Satisfying the Three Criteria

Criterion	Self-personality ratings; Word categories satisfying the criterion	Acquaintance composite personality ratings; Word categories satisfying the criterion
1. At least 20 CAQ personality correlates significant at $p \leq .10$ (double the number of correlates nominally expected by chance).	17 Categories (I, self, article, optimism, negative emotion, anxiety, anger, sad, certainty, inclusion, leisure, home, sports, TV, physical, body, sexuality)	21 Categories (I, self, positive feelings, sad, cause, tentative, certainty, inclusion, family, friends, senses, hear, motion, future, metaphysical, religion, occupation, physical, body, sexuality, eating)
2. A male/female vector of $r \geq .25$, $p \leq .01$.	12 Categories (Article, negative emotion, anxiety, sad, certainty, inclusion, leisure, home, sports, TV, body, sexuality)	11 Categories (Positive feelings, tentative, certainty, metaphysical, religion, physical, body, sexuality, eating, friends, future)
3. At least 13 RBQ behavioral correlates significant at $p \leq .10$.	6 Categories (Article, negative emotion, certainty, inclusion, body, sexuality)	5 Categories (Certainty, future, physical, body, sexuality)

Note. CAQ = California Adult Q-set; RBQ = Riverside Behavioral Q-sort.

answer this question, each target's word use score on each of the 66 word categories was correlated with his or her self- and informants' ratings of each of the 100 CAQ items. We found that an overwhelming number of word categories yielded numerous significant and meaningful correlates with both self- and informant ratings of personality. For example, individuals who use more anger words describe themselves as relatively unpredictable, not calm, and ready to express themselves through action. Also, individuals who use more self-references were described by their acquaintances as relatively moody, overreactive, and self-indulgent. Of the 6,600 correlations computed between word use categories and self-ratings of personality, 1,042 (16%) correlations were significant at $p \leq .10$, 653 (10%) correlations were significant at $p \leq .05$, 263 (4%) correlations were significant at $p \leq .01$, and 77 (1%) were significant at $p \leq .001$. Of the 6,600 correlations computed between word use categories and informant ratings of personality, 1,056 (16%) correlations were significant at $p \leq .10$, 662 (10%) correlations were significant at $p \leq .05$, 244 (4%) correlations were significant at $p \leq .01$, and 48 (less than 1%) were significant at $p \leq .001$.

Given the large number of correlations computed and lack of developed theory, it was appropriate to consider that a certain number of correlations would be significant simply by chance. Therefore, a Bonferroni-type criterion was created to address this issue. Each word category had to have at least 20 CAQ personality correlates significant at $p \leq .10$ with either self- or informant ratings in order to be considered in subsequent analyses. The CAQ has 100 items, and 20 correlates is twice the number that would nominally be expected at chance at the .10 level of significance.[2] The .10 level was used because these analyses were exploratory. It should be emphasized that this criterion might result in overlooking categories that have a small number of correlations that are strong in magnitude; however, it will identify word categories that are rich in personality information. As can be seen in Table 1, 17 of the original 66 word categories satisfied this criterion with self-CAQ personality ratings, and 21 of the original 66 categories satisfied this criterion with acquaintance CAQ personality ratings.

Three comments bear mention at this point. First, less than one third of the 66 word categories generated almost half of the correlations with self- and informant personality ratings. The 17 word categories in the self column generated 455 of the 1,042 correlations significant at $p \leq .10$, and the 21 word categories in the acquaintance column generated 542 of the 1,056 correlations significant at $p \leq .10$. Moreover, the word categories in the self and acquaintance columns generated 71% of correlations significant at $p \leq .001$. This suggests that the categories in Table 1 are rich in personality information and generated most of the large correlates with personality, whereas some of the categories excluded by the criterion undoubtedly generate personality correlates as well. Second, our analyses confirm past findings that self-ratings are related to pronouns, but they also seem to be related to many other categories. Table 1 displays a broad variety of word categories, such as articles, negative emotion, certainty, leisure, sports, and sexuality. Moreover, self-personality ratings generated 1,042 significant correlates, so it would seem that self-reports are related to word use to a much larger degree than previously observed. Third, the number of categories satisfying the criterion with acquaintance ratings is striking. This is the first study in which acquaintance ratings were made independent of exposure to the target's language sample, so it was unknown whether acquaintance ratings would be related at all; but indeed they are.

The second research question examined was: What categories of word use have a similar pattern of personality correlates across gender? Using only the categories identified from the previous analysis, this question was addressed by correlating each word category with female targets' CAQ personality ratings, correlating each word category with male targets' CAQ personality ratings, and then correlating the female with the male correlations within each category. This procedure yields an index of degree to which correlations between word categories and personality ratings are similar across genders. As can be seen in Table 1, 12 of the 17 categories in the self column had a similar pattern of correlations

---

[2] The exact number of significant correlations expected by chance cannot be directly calculated because of daunting considerations arising from correlations among correlations; however, past Monte Carlo simulations have suggested that the actual number of correlations expected by chance is lower than what is estimated according to the conventional statistical model (e.g., 20; Block, 1960).

across gender (a male–female vector correlation of .25 or larger), and 11 of the 21 categories in the acquaintance column had a similar pattern of correlations across gender. Consistent with the results of the Mehl et al. (2006) study, we found that use of articles and negative emotion words have similar correlates across gender with self-reports of personality. Unfortunately, no other comparisons can be made because Mehl et al. (2006) focused on different categories.

The third research question examined was: Of the word categories related to self- and/or acquaintance ratings of personality, which of them are also related to behavior? Accordingly, the remaining word categories were correlated with the 64 RBQ ratings of behavior. Of the 768 correlations computed between word use categories and behavior in the self column, 216 (28%) correlations were significant at the $p \leq .10$ level, and of the 704 correlations computed between word use categories and behavior in the acquaintance column, 173 (25%) correlations were significant at the $p \leq .10$ level. Again, a Bonferroni-type criterion was created to address the lack of developed theory and that a certain number of correlations might be significant simply by chance. The criterion was that each category had to have at least 13 RBQ behavioral correlates significant at $p \leq .10$. The RBQ has 64 items, and 13 correlations is twice the number that would nominally be expected at chance; the .10 level of significance was used because analyses were exploratory. As can be seen in Table 1, 6 of the 12 categories remained in the self column, and 5 of the 11 categories remained in the acquaintance column.

The six word categories in the self column generated 81% of the significant correlates with behavior, and the five word categories in the acquaintance column generated 78% of the significant correlates with behavior. This suggests that the categories identified by the criterion generated most of the behavioral correlates. It is important to emphasize that these are correlations with "real" behavior recorded in the laboratory and observed and coded by trained observers, not correlates with self-reported behavior. Furthermore, these behavior samples derive from a context entirely separate from the interview from which the word use scores were obtained. It is impressive that, for example, the number of sexuality words a person used in the life history interview correlated with a large number of behaviors that he or she displayed in a separate three-person interaction. This suggests that word use may be broadly related to behavior to a much larger degree than previously suspected.

The fourth research question examined was: To what degree are the word categories used consistently across the life history interview? Targets were asked a variety of questions throughout the interview, and it is important to examine the degree to which word categories were used consistently regardless of what was being asked. This was examined by splitting each target's interview in half and obtaining word use scores for each half, and then computing a correlation across the halves. The reliabilities of the remaining categories are as follows: Articles ($r = .58$), Negative Emotion ($r = .44$), Certainty ($r = .64$), Inclusion ($r = .63$), Body ($r = .02$), Sexuality ($r = .55$), Physical ($r = .20$), and Future ($r = .41$). Although these reliabilities might be lower than what one might expect (especially the Body and Physical categories), they are actually impressive given that the content of the interview was very different in each half of the interview. In the first half, participants were asked to describe themselves (e.g., "What do you

like most about yourself?"), to answer questions about college and academic experiences (e.g., "What are your study habits like?"), and to discuss their future plans (e.g., "What are your plans after college?"). In the second half, they were asked about interpersonal relationships (e.g., "Are you currently in a romantic relationship?"), childhood and family history (e.g., "What was your mother/father like as a child?"), and to describe a defining event that had an impact on or changed their lives. Also, considering that only two of the eight categories satisfy the inclusion criterion of past studies (e.g., a reliability of at least .60), yet all of these categories generated a large number of personality and behavioral correlates, it seems that word use is related to personality even when reliability is low.

The results suggest that there are several word categories relevant to personality. Although the categories identified by our analyses are related to multiple sources of personality data, the criteria are extremely conservative, so many of the categories excluded are likely rich in personality information as well. Still, it could be argued that the present results are based on a strictly empirical, nontheoretical method and that perhaps they are arbitrary. In order to address this possibility, two word categories were identified for which to present the personality and behavioral correlates. These two categories (Certainty and Sexuality[3]) were chosen because they satisfied all the criteria in both the self and acquaintance columns.[4] To the extent that correlations between word use and self-judgments of personality, acquaintance judgments of personality, and observed behavior form a psychologically coherent pattern, it can reasonably be assumed that the results are not arbitrary.

The correlation between Certainty word count scores and Sexuality word count scores is .32. To determine whether these categories should be reported independently or combined, Certainty word use was correlated with self- and acquaintance CAQ ratings and RBQ ratings, partialing for Sexuality word use, and Sexuality word use was correlated with self- and acquaintance CAQ ratings and RBQ ratings, partialing for Certainty word use. For both categories, this did not appreciably change the pattern of correlates or the magnitude of the correlates. Subsequently, the correlates of these categories, at the $p \leq .05$ level of significance, are reported separately.[5]

---

[3] The internal reliability of these word categories has not been reported in past research. In the present data, we found that the alpha internal reliabilities for the Sexuality and Certainty word categories were .46 and .34, respectively. In an attempt to enhance reliability, revised categories were created by discarding items (i.e., words) with negative item-total correlations; however, this did not substantially improve reliability, change the pattern of personality or behavioral correlates, or change the number of significant correlates of the word categories. Subsequently, all analysis are reported using the original categories.

[4] Although the Body word category also satisfied all criteria in both the self and acquaintance columns, it had a reliability of only $r = .02$, which is unacceptably low according to any standards.

[5] Tables including the complete personality and behavioral correlates up to the .10 level of significance can be viewed online at http://rap.ucr.edu/Lisa/WebsiteTables.pdf

Table 2
Correlations Between Certainty Word Use and CAQ Personality Ratings

No.	Q-item description	Self	Acquaintances
Positive correlations			
43	Is facially and/or gesturally expressive.	.31**	.17*
98	Is verbally fluent; can express ideas well.	.29**	.25**
16	Is introspective and concerned with self as an object.	.27**	.18*
66	Enjoys esthetic impressions.	.24**	.17*
8	Appears to have a high degree of intellectual capacity.	.23**	.26**
44	Evaluates the motivation of others.	.22**	.12
52	Behaves in an assertive fashion.	.21**	.20**
47	Has a readiness to feel guilt.	.21**	.01
15	Is skilled in social techniques of play/humor.	.20**	.13
56	Responds to humor.	.20**	−.04
64	Socially perceptive of a wide range of interpersonal cues.	.20**	.00
71	Has a high aspiration level for self.	.19**	.21**
58	Enjoys sensuous experiences.	.19**	.04
46	Engages in personal fantasy and daydreams.	.18*	.07
3	Has a wide range of interests.	.17*	.11
82	Has fluctuating moods.	.16*	.08
90	Is concerned with philosophical problems.	.16*	.17*
1	Is critical, skeptical, not easily impressed.	.16*	.08
60	Has insight into own behavior.	.15*	.02
51	Genuinely values intellectual matters.	.15*	.19**
68	Is basically anxious.	.15*	.11
59	Concerned w/ own body and physiological functioning.	.14	.21**
29	Is turned to for advice and reassurance.	.14	.22**
77	Appears straightforward, candid in dealing with others.	.13	.19**
35	Has warmth; capacity for close relationships.	.12	.16*
96	Values own independence and autonomy.	.11	.15*
91	Is power oriented; values power in self and others.	.10	.15*
2	Genuinely dependable and responsible.	.06	.18*
5	Behaves in a giving manner toward others.	.05	.20**
80	Interested in members of the opposite sex.	.03	−.18*
Negative correlations			
86	Has repressive and dissociative tendencies.	−.25**	−.19**
97	Is emotionally bland; has flattened affect.	−.23**	−.23**
61	Creates and exploits dependency in people.	−.19**	−.21**
45	Has a brittle ego-defense system.	−.16*	−.03
21	Arouses nurturant feelings in others.	−.15*	.03

Note. Examples of certainty words are *absolutely, exact, guarantee, sure,* and *truly.* CAQ (California Adult Q-set) item content is abbreviated. No. = number.
* $p \leq .05$.   ** $p \leq .01$.

Certainty

Examples of words that fall into this category are *absolutely, clearly, definite, exactly, guarantee, precisely,* and *sure* (for a complete list, see Appendix B). On average, participants used 78 certainty words throughout the interview ($SD = 28$), and results indicate that women ($M = 86$, $SD = 36$) used significantly more certainty words than men ($M = 72$, $SD = 36$), $t(182) = 2.71$, $p = .007$. The following is a direct quote from a participant who scored high in use of certainty words: "I have *always* enjoyed computers and the Internet and that is *certainly* where I want to focus my attention to in graduate school and it is *certainly* an amalgamation of everything I have learned and enjoy."

The personality and behavioral correlates can be seen in Tables 2–3, and the degree of convergence across data sources is remarkable.[6] According to self-ratings, acquaintance ratings, and behavioral ratings, individuals who are confident and aspiring (e.g., behaves in an assertive fashion, has a high aspiration level, and does not express insecurity), likable (e.g., seems likable, exhibits social skills and says

or does interesting things), smart and thoughtful (e.g., is verbally fluent, appears to have a high degree of intellectual capacity, and is introspective), yet not emotionally flat (e.g., not emotionally bland and not repressive), use more certainty words.

The most revealing aspects of certainty word use may be the association with smart/thoughtfulness and emotionality. After analyzing cases of brain damage (e.g., Phineas Gage and "Elliott"), Antonio Damasio (1994) concluded that emotions are just as important as cognition in good judgment because emotions are powerful motivators and help people to identify what is important and what is not. In line with this observation, it seems that certainty word use taps a construct of good judgment or wisdom. These results are particularly interesting given that early language theorists assumed that people who have linguistic styles characterized by "certainty" would be paranoid, rigid, and irascible

---

[6] The vector correlation between self and acquaintance CAQ personality correlates is .67 ($p < .0001$).

Table 3
Correlations Between Certainty Word Use and RBQ Behavioral Ratings

No.	RBQ-item description	RBQ
Positive correlations		
54	Speaks fluently.	.34**
8	Exhibits social skills.	.32**
44	Says or does interesting things.	.29**
3	Volunteers information about self.	.29**
16	Shows high enthusiasm and energy.	.28**
21	Is talkative.	.28**
17	Shows a wide range of interests.	.25**
46	Displays ambition.	.23**
57	Speaks in a loud voice.	.23**
50	Behaves in a cheerful manner.	.19*
38	Is expressive in face, voice, or gestures.	.19*
2	Interviews partner.	.18*
42	Shows interest in intellectual matters.	.18*
29	Seems likable.	.18*
33	Expresses warmth.	.17*
Negative correlations		
51	Gives up when faced with obstacles.	−.34**
14	Exhibits an awkward interpersonal style.	−.33**
22	Expresses insecurity.	−.31**
41	Keeps partner at a distance.	−.31**
18	Talks at rather than with partner.	−.30**
37	Behaves in a fearful manner.	−.30**
45	Says negative things about the self.	−.29**
23	Shows signs of tension and anxiety.	−.28**
61	Seems detached from the interaction.	−.27**
9	Is reserved and unexpressive.	−.26**
48	Expresses self-pity or victimization.	−.25**
27	Seeks reassurance from partner.	−.25**
36	Is unusual or unconventional in appearance.	−.24**
52	Behaves in a gender-stereotyped manner.	−.23**
31	Appears to regard self as physically attractive.	−.18*
30	Seeks advice from partner.	−.18*

Note. Examples of certainty words are *absolutely, exact, guarantee, sure,* and *truly.* RBQ (Riverside Behavioral Q-sort) item content is abbreviated. No. = number.
* $p \leq .05$.   ** $p \leq .01$.

(Hayakawa, 1940; Korzybski, 1933). Our data suggests that the opposite is more likely. Certainty word users are smart, thoughtful, confident, and well liked.

Sexuality

Examples of words that fall into this category are *breast*, *butt*, *erection*, *horny*, *love*, *nude*, *orgasm*, and other "dirty words" (for a complete list, see Appendix C). Participants used an average of six sexuality words across the interview ($SD = 6$), and results indicate that women ($M = 8$, $SD = 4$) used significantly more sexuality words than men ($M = 5$, $SD = 4$), $t(182) = 2.21$, $p = .03$. The following is a direct quote from a participant who scored high in use of sexuality words:

> I lost my *virginity* when I was 13 . . . and um, I think she [a friend] just got pissed. . . that she went to my mom. . . my parents flipped the heck out. Of course I got back at [the friend] and I said, well. . . you're the one that taught me the certain positions with *sex* and all things we need to do.

Tables 4–5 display the personality and behavioral correlates, and again, there is a remarkable degree of convergence across

data sources;[7] yet, where there is divergence, it is revealing. According to self-ratings of personality, acquaintance ratings of personality, and behavioral ratings, individuals who are highly extraverted (e.g., talkative and speak in a loud voice), dramatic (e.g., is self-dramatizing/histrionic and is physically animated and expressive in face, voice, and gestures), and unconventional (e.g., rebellious and nonconforming) use more sexuality words. However, according to the acquaintance ratings and behavioral ratings, frequent sexuality word users are also neurotic (e.g., have fluctuating moods and anxiety and tensions find outlet in bodily symptoms) and self-centered (e.g., dominate the interaction, do not express sympathy toward partners, and are not interested in what partners have to say). The pattern of extraversion combined with neurotic self-centeredness and self-dramatizing implies that individuals who use more sexuality words are characteristically high in the need for attention. To our knowledge, sexuality word use has not been studied in previous research, but it seems to tap an interesting combination

---

[7] The vector correlation between self and acquaintance CAQ personality correlates is .57 ($p < .0001$).

Table 4
Correlations Between Sexuality Word Use and CAQ Personality Ratings

No.	Q-item description	Self	Acquaintances
Positive correlations			
43	Is facially and/or gesturally expressive.	.25**	.16*
4	Is talkative.	.22**	.18*
99	Is self-dramatizing; histrionic.	.20**	.24**
15	Is skilled in social techniques of play/humor.	.20**	.15*
52	Behaves in an assertive fashion.	.18*	.24**
83	Able to see the heart of important problems.	.17*	−.05
18	Initiates humor.	.16*	.03
62	Rebellious and nonconforming.	.16*	.19**
77	Appears straightforward, candid in dealing with others.	.16*	.07
16	Is introspective and concerned with self as an object.	.16*	.22**
59	Concerned with own body and physiological functioning.	.16*	.13
98	Is verbally fluent; can express ideas well.	.15*	.18*
82	Has fluctuating moods.	.13	.21**
69	Sensitive to demands.	.13	.16*
10	Anxiety and tension find outlet in bodily symptoms.	.12	.18*
89	Compares self to others.	.11	.20**
66	Enjoys esthetic impressions.	.07	.23**
70	Behaves in an ethically consistent manner.	.06	−.16*
39	Thinks and associates ideas in unusual ways.	.06	.17*
90	Is concerned with philosophical problems.	.05	.18*
36	Is subtly negativistic; undermines.	.00	.18*
Negative correlations			
33	Is calm, relaxed in manner.	−.20**	−.25**
97	Is emotionally bland; has flattened affect.	−.16*	−.24**
2	Genuinely dependable and responsible.	−.16*	−.10
27	Shows condescending behavior with others.	−.16*	.13
75	Has a clear-cut, internally consistent personality.	−.07	−.15*
74	Subjectively unaware of self-concern.	−.05	−.19**
100	Does not vary roles.	−.05	−.20**
7	Favors conservative values.	−.01	−.17*
53	Needs tend toward direct and uncontrolled expression.	−.01	.16*

Note. Examples of sexuality words are *boobs, butt, horny, kiss,* and *love, and nude.* CAQ (California Adult Q-set) item content is abbreviated. No. = number.
* $p \leq .05$. ** $p \leq .01$.

of characteristics. Use of sexuality words appears to be one of several ways in which individuals who have a high need for attention seek to get noticed.

Discussion

The present data suggest that many categories of word use are relevant to personality, and some of the categories that we identified have been overlooked in past research. Of all the categories remaining in the self and acquaintance columns of Table 1, past studies have not examined the personality correlates of Certainty, Physical, Body, or Sexuality word use. These word categories yielded a high number of personality correlates that were similar in pattern across genders, a high number of behavioral correlates, and were used with a surprising degree of consistency across the interview context. What is more, the correlations across data sources formed a coherent pattern of results for Certainty and Sexuality word use. Past studies (Mehl et al., 2006; Pennebaker & King, 1999) excluded word categories that had a mean usage level below 1%, such as Sexuality, and our analyses suggest that this may lead to excluding important word categories. We believe that all of the word categories in Table 1 are relevant to personality because many of the categories that did not satisfy every criterion are still rich with personality information. For example, the Sports cate-

gory had slightly fewer than 13 RBQ correlates (11), but it generated several personality correlates. Also, categories of word use that have gender-specific personality correlates might still have meaningful correlates within gender; they simply need to be analyzed separately for men and women.

Future research might particularly benefit from examining the use of sad words, self-references, and articles.[8] We found that those who used relatively more sad words and self-references were rated by themselves and their acquaintances as higher along a variety of depressive personality characteristics; however, the behavioral correlates of sad word usage were few in number and socially undesirable (e.g., expresses hostility and does not smile frequently), whereas use of self-references was positively correlated with a wide variety of socially desirable behaviors (e.g., enthusiastic, cheerful, and warm). It would be interesting to explore why frequent usage of both types of words seems related to depression, yet high-self-referencing individuals display positive social behaviors, and individuals using more sad words do not.

---

[8] The personality and behavioral correlates of these word categories can be viewed online at http://rap.ucr.edu/Lisa/MoreWebsiteTables.pdf

Table 5
Correlations Between Sexuality Word Use and RBQ Behavioral Ratings

No.	Q-item description	RBQ
Positive correlations		
21	Is talkative.	.33**
6	Dominates the interaction.	.31**
12	Is physically animated.	.29**
38	Is expressive in face, voice, or gestures.	.28**
3	Volunteers information about self.	.27**
16	Shows high enthusiasm and energy.	.27**
44	Says or does interesting things.	.24**
57	Speaks in a loud voice.	.23**
33	Expresses warmth.	.18*
Negative correlations		
25	Expresses sympathy toward partner.	−.38**
9	Is reserved and unexpressive.	−.31**
4	Seems interested in what partner says.	−.28**
41	Keeps partner at a distance.	−.27**
23	Shows signs of tension and anxiety.	−.26**
61	Seems detached from the interaction.	−.25**
14	Exhibits an awkward interpersonal style.	−.24**
52	Behaves in a gender-stereotyped manner.	−.24**
31	Appears to regard self as physically attractive.	−.23**
22	Expresses insecurity.	−.21*
32	Acts irritated.	−.19*
37	Behaves in a fearful manner.	−.19*
30	Seeks advice from partner.	−.19*
19	Expresses agreement frequently.	−.18*

Note. Examples of sexuality words are *boobs, butt, horny, kiss, love,* and *nude.* RBQ (Riverside Behavioral Q-sort) item content is abbreviated. No. = number.
* $p \le .05$. ** $p \le .01$.

Another interesting category might be use of articles. In line with G. Allport's (1961) distinction between adaptive versus stylistic behaviors, the content word categories discussed thus far measure what a person talks about (adaptive behavior), whereas articles deal with how a person talks (stylistic behavior). For example, "Some girl asked about restrooms" versus "The girl asked where the restroom is." Although both sentences share the same basic meaning, the first sentence includes no articles and conveys a sense of dismissiveness toward the subject, whereas the second sentence includes two articles and seems more polite. G. Allport (1961) argued that stylistic behaviors should be more highly related to personality than adaptive behaviors, and Pennebaker and colleagues have similarly argued that stylistic linguistic categories might be more promising for personality research than content categories (Chung & Pennebaker, 2007; Groom & Pennebaker, 2002). Indeed, we found several personality and behavioral correlates of article usage, indicating that those who use more articles are highly intellectual (philosophical, verbally fluent, and skeptical) and open to experience (wide-ranging interests and aesthetic). Future research might explore why use of articles seems to be related to intelligence and openness.

A possible implication of the present data is that there may be a bandwidth–fidelity trade-off in the relationship between word use and personality. Word use is a relatively specific behavior (high fidelity), and it may not be especially useful to correlate word use with high-bandwidth personality characteristics. For example, Pennebaker and King (1999) found only modest correlations between word use and self-reported Big Five traits; however, in the present study, we found an overwhelming number of correlates between word use and

self-reported CAQ items. Perhaps the relationship between personality and word use is maximized with low-bandwidth personality variables. Another possible implication is that word use may have a stronger relationship with observed behavior than with self-reported behavior. Pennebaker and King (1999) found only a handful of correlations between word use and self-reported behaviors, whereas in the present study, we found a large number of correlates between word use and directly observed behavior.

Although word use seems promising for personality research, some potential limitations need to be addressed. First, the life history interview is a special kind of context, and it is uncertain to what degree word use in this context would generalize to other contexts. However, previous studies have demonstrated that word use is reliable across time and context, and it seems implausible that participants could have completely changed their life-long patterns of word use for the interview. Moreover, the patterns of word use in the interview were found to yield correlates with behavior in at least two independent contexts: the laboratory setting, in which behavior was directly observed, and the daily environment, in which informants interacted with our participants and upon which they based their personality ratings. Second, it might be beneficial to explore methods of language analysis beyond the word count approach. Oberlander and Gill (2006) are exploring a bottom-up technique (rather than a top-down approach as in the present study) in which word sequences can be analyzed (rather than single words). Although it is unclear whether this technique can identify linguistic categories beyond what has been previously established using top-down approaches, it seems that this new technique may be more sensitive to context compared with word counts.

Another limitation is the lack of developed theory guiding research on word use and personality. However, Haig (2005) observed that for the past 150 years, science has been dominated by the idea that theory should precede empirical observation (the "hypothetico-deductive method"). He argued that this is a rather narrow of view of the scientific process and proposed an alternative called the "abductive theory of scientific method." Haig suggested that theory can be constructed by, first, defining and detecting the phenomenon of interest through empirical observation and use of multiple methods. After establishing the "robust and stubborn" existence of the phenomenon, one can then proceed to construct theories on the basis of empirical regularities (Haig, 2005, p. 374). In line with Haig's model, we believe that the relationship between personality and language needs to be explored despite the lack of theory. For almost 100 years, researchers have argued that personality is manifested in language (F. Allport, Walker, & Lathers, 1934; Freud, 1916/1964; Furnham, 1990; Sanford, 1942), and we propose that a theory can be developed only after the phenomenon is more fully explored. The present study was a step toward this possibility by identifying word categories that may be useful in this research and demonstrating that word use is related to personality to a much larger degree than previously observed.

References

Allport, F., Walker, L., & Lathers, E. (1934). Written composition and characteristics of personality. *Archives of Psychology, 173,* 1–82.

Allport, G. (1961). *Pattern and growth in personality*. New York: Holt, Rinehart, & Winston.

Bem, D., & Funder, D. (1978). Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review, 85,* 485–501.

Berry, D. S., Pennebaker, J. W., Mueller, J. S., & Hiller, W. (1997). Linguistic bases of social perception. *Personality and Social Psychology Bulletin, 23,* 526–538.

Block, J. (1960). On the number of significant findings to be expected by chance. *Psychometrika, 25,* 369–380.

Block, J. (1961). *The Q-sort method in personality assessment and psychiatric research*. Palo Alto, CA: Consulting Psychologists Press.

Carroll, D. W. (1999). *Psychology of language* (3rd ed.). New York: Brooks/Cole.

Chung, C., & Pennebaker, J. (2007). The psychological functions of function words. In K. Fielder (Ed.), *Frontiers in social psychology* (pp. 343–359). New York: Psychology Press.

Craik, K. H., Ware, A. P., Kamp, J., O'Reilly, C., III, Staw, B., & Zedeck, S. (2002). Explorations of construct validity in a combined managerial and personality assessment programme. *Journal of Occupational and Organizational Psychology, 75,* 171–193.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.

Freud, S. (1964). Introductory lectures on psycho-analysis. In J. Strachey (Trans.), *The standard edition of the complete psychological works of Sigmund Freud* New York: Norton. (Original work published 1916)

Funder, D. C. (2001). Personality. *Annual Review of Psychology, 52,* 197–221.

Funder, D. C. (2007). *The personality puzzle* (4th ed.). New York: Norton.

Funder, D. C., Furr, R. M., & Colvin, R. (2000). The Riverside Behavioral Q-sort: A tool for the description of social behavior. *Journal of Personality, 68,* 450–489.

Furnham, A. (1990). Language and personality. In H. Giles & W. P. Robinson (Eds.), *Handbook of language and social psychology* (pp. 73–95). New York: Wiley.

Gill, A., & Oberlander, J. (2002, August). Taking care of the linguistic features of extraversion. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. Fairfax, VA.

Gottschalk, L., & Gleser, G. (1969). *The measurement of psychological states through the content analysis of verbal behavior*. Berkeley: University of California Press.

Groom, C., & Pennebaker, J. (2002). Brief report: Words. *Journal of Research in Personality, 36,* 615–621.

Haig, B. (2005). An abductive theory of scientific method. *Psychological Methods, 10,* 371–388.

Hayakawa, S. I. (1940). *Language in action*. Chicago: Semantics Institute.

Ickes, W., Reidhead, S., & Patterson, M. (1986). Machiavellianism and self-monitoring: As different as "me" and "you." *Social Cognition, 4,* 58–74.

Korzybski, A. (1933). *Science and sanity: An introduction to non-Aristotelian systems and general semantics*. Lancaster, PA: Science Press.

Letzring, T., Block, J., & Funder, D. (2005). Ego-control and ego-resiliency: Generalization of self-report scales based on personality descriptions from acquaintances, clinicians, and the self. *Journal of Research in Personality, 39,* 395–422.

Letzring, T., Wells, S., & Funder, D. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology, 91,* 111–123.

Mehl, M., Gosling, S., & Pennebaker, J. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology, 90,* 862–877.

Mehl, M., & Pennebaker, J. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology, 84,* 857–870.

Mehl, M., Pennebaker, J., Crow, M., Dabbs, J., & Price, J. (2001). The electronically activated recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers, 33,* 517–523.

Merriam-Webster Online Dictionary. (2005). Springfield, MA: Merriam-Webster. Retrieved October 20, 2006, from http://www.m-w.com/dictionary/language

Nagy, W., & Anderson, R. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19,* 304–330.

Oberlander, J., & Gill, A. J. (2006). Language with character: A corpus-based study of individual differences in e-mail communication. *Discourse Processes, 42*(3), 239–270.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Mahwah, NJ: Erlbaum.

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

Pennebaker, J., Mehl, M., & Niederhoffer, K. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54,* 547–577.

Pennebaker, J., & Stone, L. (2003). Words of wisdom: Language use over the life span. *Journal of Personality and Social Psychology, 85,* 291–301.

Raskin, R., & Shaw, R. (1988). Narcissism and the use of personal pronouns. *Journal of Personality, 56, 2,* 393–404.

Sanford, F. (1942). Speech and personality. *Psychological Bulletin, 39,* 811–845.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *The general inquirer: A computer approach to content analysis*. Cambridge, MA: MIT Press.

Vazire, S., & Funder, D. (2006). Impulsivity and the self-defeating behavior of narcissists. *Personality and Social Psychology Review, 10,* 154–165.

Wagerman, S., & Funder, D. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality, 41,* 221–229.

Weintraub, W. (1989). *Verbal behavior in everyday life*. New York: Springer.

Appendix A

LIWC Word Categories Analyzed

1.	pronouns		34.		friends
2.	I		35.		family
3.	we		36.		humans
4.	self		37.		time
5.	you		38.		past
6.	other		39.		present
7.	negate		40.		future
8.	assent		41.		space
9.	article		42.		up
10.	prepositions		43.		down
11.	number		44.		inclusion
12.	affect		45.		exclusion
13.	positive emotion		46.		motion
14.	positive feelings		47.		occupation
15.	optimism		48.		school
16.	negative emotion		49.		job
17.	anxiety		50.		achieve
18.	anger		51.		leisure
19.	sad		52.		home
20.	cognitive mechanisms		53.		sports
21.	cause		54.		tv
22.	insight		55.		music
23.	discrepancy		56.		money
24.	inhibition		57.		metaphysical
25.	tentative		58.		religion
26.	certain		59.		death
27.	senses		60.		physical
28.	see		61.		body
29.	hear		62.		sexual
30.	feel		63.		eating
31.	social		64.		sleep
32.	communication		65.		groom
33.	other references		66.		swear

Note. LIWC = Linguistic Inquiry and Word Count program.

Appendix B

Words in the Certainty Category

1.	*absolut*[a]		16.		*faith*[a]
2.	*all*		17.		*firm*[a]
3.	*altogether*		18.		*forever*
4.	*always*		19.		*guarantee*[a]
5.	*assur*[a]		20.		*indeed*
6.	*certain*[a]		21.		*inevitab*[a]
7.	*clear*		22.		*never*[a]
8.	*clearly*		23.		*positiv*[a]
9.	*completely*		24.		*precisely*
10.	*confidence*[a]		25.		*secure*
11.	*confidently*		26.		*sure*[a]
12.	*definite*[a]		27.		*totally*
13.	*distinctly*		28.		*truly*
14.	*exact*[a]		29.		*undoubt*[a]
15.	*fact*[a]		30.		*very*

[a] Any form of the word contributes to the score for the category.

(*Appendixes continue*)

Appendix C

Words in the Sexuality Category

1.	*abort*[a]	26.	*loves*
2.	*aids*	27.	*lust*
3.	*bi*	28.	*naked*
4.	*boobs*	29.	*nipple*[a]
5.	*breast*[a]	30.	*nude*[a]
6.	*butt*	31.	*orgasm*[a]
7.	*cock*	32.	*ovar*[a]
8.	*condom*[a]	33.	*passion*[a]
9.	*dick*	34.	*penis*[a]
10.	*erection*[a]	35.	*pervert*[a]
11.	*fairy*	36.	*porn*[a]
12.	*fuck*[a]	37.	*pregnan*[a]
13.	*gay*	38.	*prick*[a]
14.	*gays*	39.	*prostat*[a]
15.	*genital*[a]	40.	*prostitu*[a]
16.	*homosexual*[a]	41.	*pussy*[a]
17.	*horny*	42.	*queer*[a]
18.	*hug*[a]	43.	*rape*[a]
19.	*incest*[a]	44.	*screw*[a]
20.	*kiss*[a]	45.	*sex*[a]
21.	*lesbian*[a]	46.	*stud*
22.	*libid*[a]	47.	*vagina*[a]
23.	*love*	48.	*virgin*[a]
24.	*loved*	49.	*womb*[a]
25.	*lover*[a]		

[a] Any form of the word contributes to the score for the category.