The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods

Yla R. Tausczik and James W. Pennebaker

Abstract

We are in the midst of a technological revolution whereby, for the first time, researchers can link daily word use to a broad array of real-world behaviors. This article reviews several computerized text analysis methods and describes how Linguistic Inquiry and Word Count (LIWC) was created and validated. LIWC is a transparent text analysis program that counts words in psychologically meaningful categories. Empirical results using LIWC demonstrate its ability to detect meaning in a wide variety of experimental settings, including to show attentional focus, emotionality, social relationships, thinking styles, and individual differences.

Keywords

computerized text analysis, LIWC, relationships, dominance, deception, attention, pronouns

James J. Bradac (1986, 1999) celebrated the many ways that scientists could simultaneously study both language and human communication. He understood the value of highly controlled laboratory studies and, at the same time, the importance of exploring the ways people naturally talk in the real world. Of particular importance to him, however, was that language research replicates its theories and findings across a wide array of methods and samples. This article draws heavily from Bradac’s approach to research by applying a new array of computer-based text analysis tools to the study of everyday language.

1University of Texas at Austin, Austin, TX, USA

Corresponding Author:
James W. Pennebaker, Department of Psychology, University of Texas at Austin, Austin, TX 78712, USA
Email: Pennebaker@mail.utexas.edu
The words we use in daily life reflect who we are and the social relationships we are in. This is neither a new nor surprising insight. Language is the most common and reliable way for people to translate their internal thoughts and emotions into a form that others can understand. Words and language, then, are the very stuff of psychology and communication. They are the medium by which cognitive, personality, clinical, and social psychologists attempt to understand human beings.

The simultaneous development of high-speed personal computers, the Internet, and elegant new statistical strategies have helped usher in a new age of the psychological study of language. By drawing on massive amounts of text, researchers can begin to link everyday language use with behavioral and self-reported measures of personality, social behavior, and cognitive styles. Beginning in the early 1990s, we stumbled on the remarkable potential of computerized text analysis through the development of our own computer program—Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, & Francis, 2007). We are now witnessing new generations of text analysis coming from computer sciences and computational linguistics.

This article is divided into three sections. The first is a brief history of text analysis in psychology. The second focuses on our own efforts to develop LIWC along with some of the basic psychometrics of words. The third explores the links between word usage and basic social and personality processes.

**Computerized Text Analysis: A Brief History**

The roots of modern text analysis go back to the earliest days of psychology. Freud (1901) wrote about slips of the tongue whereby a person’s hidden intentions would reveal themselves in apparent linguistic mistakes. Rorschach and others (e.g., Holtzman, 1950; Rorschach, 1921) developed projective tests to detect people’s thoughts, intentions, and motives from the way they described ambiguous inkblots. McClelland and a generation of thematic apperception test (TAT) researchers (e.g., McClelland, 1979; Winter, 1998) found that the stories people told in response to drawings of people could provide important clues to their needs for affiliation, power, and achievement. In all cases, trained raters read the transcripts of people’s descriptions and tagged words or phrases that represented the dimensions the investigators were studying.

More general and less stimulus-bound approaches began to evolve in the 1950s. Gottschalk and his colleagues (e.g., Gottschalk & Gleser, 1969; Gottschalk, Gleser, Daniels, & Block, 1958) developed a content-analysis method by which to track Freudian themes in text samples. The original Gottschalk method required patients to talk in a stream of consciousness way into a tape recorder for 5 minutes. The language samples were transcribed and broken down into grammatical phrases. Judges, then, evaluated each phrase to determine the degree it might reflect one or more themes related to anxiety (e.g., death, castration), hostility toward self or others, and various interpersonal and psychological topics. The Gottschalk method later was used in the psychiatric diagnoses of cognitive impairments, alcohol abuse, brain damage, and mental disorders. Attempts to translate the original Gottschalk–Gleser scoring scheme
to a computer program have proven difficult with modest correlations to the judge-based “gold standard” (e.g., Gottschalk & Bechtel, 1993).

The first general purpose computerized text analysis program in psychology was developed by Philip Stone and his colleagues (Rosenberg & Tucker, 1978; Stone, Dunphy, Smith, & Ogilvie, 1966). Using a mainframe computer, the authors built a complex program that adapted McClelland’s need-based coding schemes to any open-ended text. The program, called General Inquirer, relied on a series of author-developed algorithms. The General Inquirer and other programs like it (e.g., Hart’s, 1984, DICTION program; Martindale, 1990) have proven valuable in distinguishing mental disorders, assessing personality dimensions, and evaluating speeches. One limitation of these approaches is that they have relied on the manipulation and weighting of language variables that were not visible to the user.

The first truly transparent text analysis method was pioneered by Walter Weintraub (1981, 1989). Weintraub, a physician by training, became fascinated by the everyday words people used—words such as pronouns and articles. Over the span of a decade, he hand-counted people’s words in texts such as political speeches and medical interviews. He noticed that first-person singular pronouns (e.g., I, me, my) were reliably linked to people’s levels of depression. Although his methods were straightforward and his findings consistently related to important outcome measures, his work was largely ignored. His observation that the simple words of everyday speech reflected psychological state nevertheless was prescient. (See also the work of Mergenthaler, 1996, who developed a computer program TAS/C that taps abstraction and emotion in psychotherapy sessions.)

The Development of LIWC and the Psychometrics of Words

In the 1980s, we discovered that when people were asked to write about emotional upheavals in their lives they subsequently evidenced improvements in physical health (e.g., Pennebaker & Beall, 1986). The first group of writing studies generated hundreds of writing samples that revealed deeply moving human stories. Intuitively, the ways the stories were written should have been related to whether people’s health improved or not. In an attempt to link the stories with health outcomes, judges were asked to read the emotional essays and to rate them along multiple dimensions. Some of the categories included the degree to which the stories were organized, coherent, personal, emotional, vivid, optimistic, and evidenced insight.

Relying on judges’ ratings yielded three important findings: (a) even with in-depth training, judges do not agree with each other in rating most dimensions when evaluating a broad range of deeply personal stories; (b) rating essays by multiple judges is extremely slow and expensive; and (c) judges tend to get depressed when reading depressing stories.

To find a more efficient evaluation method, we turned to the promise of computerized text analysis programs to assess the essays. At the time, no simple text analysis program existed. Consequently, Martha Francis and the second author began the task
of developing one. Our goal was to create a program that simply looked for and counted words in psychology-relevant categories across multiple text files. The result has been an ever-changing computer program named Linguistic Inquiry and Word Count, or LIWC (pronounced “Luke”).

**The Logic and Development of LIWC**

The LIWC program has two central features—the processing component and the dictionaries. The processing feature is the program itself, which opens a series of text files—which can be essays, poems, blogs, novels, and so on—and then goes through each file word by word. Each word in a given text file is compared with the dictionary file.

For example, if LIWC were analyzing the first line of the novel *Paul Clifford* by Edward Bulwer-Lytton (1842):

> It was a dark and stormy night

the program would first look at the word “it” and then see if “it” was in the dictionary.

It is and is coded as a function word, a pronoun, and, more specifically, an impersonal pronoun. All three of these LIWC categories would then be incremented. Next, the word “was” would be checked and would be found to be associated with the categories of verbs, auxiliary verbs, and past tense verbs.

After going through all the words in the novel, LIWC would calculate the percentage of each LIWC category. So, for example, we might discover that 2.34% of all the words in a given book were impersonal pronouns and 3.33% were auxiliary verbs. The LIWC output, then, lists all LIWC categories and the rates that each category was used in the given text.

The dictionaries are the heart of the LIWC program. A dictionary refers to the collection of words that define a particular category. When LIWC was first created, the goal was fairly modest. We simply wanted the computer to calculate the percentage of positive and negative emotion words within a text. To do this, we needed to specify exactly which words to look for. Based on our judges’ ratings, we also wanted to include measures of thinking styles—for example, signs of self-reflection, and causal thinking. Over several weeks, the number of categories we were interested in expanded from the original 2 to more than 80.

Across the 80 categories, several language dimensions are straightforward. For example, the category of articles is made up of three words: “a,” “an,” and “the.” Other dimensions are more subjective. For example, the emotion word categories required human judges to evaluate which words were suited for which categories. For all subjective categories, an initial selection of word candidates was gleaned from dictionaries, thesauruses, questionnaires, and lists made by research assistants. Groups of three judges then independently rated whether each word candidate was appropriate to the overall word category.
All category word lists were updated by the following set of rules: (a) a word remained in the category list if two out of three judges agreed it should be included; (b) a word was deleted from the category list if at least two of the three judges agreed it should be excluded; and (c) a word was added to the category list if two out of three judges agreed it should be included. This entire process was then repeated a final time by a separate group of three judges. The final percentages of judges’ agreement for this second rating phase ranged from 93% to 100% agreement.

The initial LIWC judging took place between 1992 and 1994. A significant LIWC revision was undertaken in 1997 and again in 2007 to streamline the original program and dictionaries. Text files from several dozen studies, totaling more than 100 million words were analyzed. Some low base rate word categories were deleted and others were added. For details of the process and specific findings, see Pennebaker, Chung, Ireland, Gonzales, and Booth (2007).

The Psychometrics of Word Usage

Unlike the typical development of a new measurement instrument, verifying the validity and reliability of word usage is trickier. Consider how psychologists typically develop and test a new measurement instrument. For questionnaires, for example, after specific questions have been generated and initially tested, the investigator computes reliability statistics to be sure that all items are correlated with the sum of the remaining items. Generally, a factor analysis of the items is run to see if the items reflect more than one dimension. Next, the investigator computes the test–retest reliability of the questionnaire. And, finally, there are a series of validation tests to see if the questionnaire correlates with or predicts real-world behaviors that it is supposed to measure.

Word categories are unlike questionnaire items. Words are rarely normally distributed, they generally have low base rates, and standard measures of reliability are not always appropriate. Consider, for example, the category of articles—“a,” “an,” and “the.” All three words serve the same function, which is to signal the upcoming use of a concrete noun. From a classically trained psychometric perspective, for us to consider “articles” to be a coherent, internally consistent category, use of the three words should be highly correlated with each other—with Cronbach’s $\alpha$ of at least .60 or .70, it is hoped. Tragically, words do not adhere to traditional psychometric laws that we see in questionnaires. For example, our lab frequently relies on a random assortment of about 2,800 text files that includes a wide range of text genres, including blogs, experimental essays, poetry, books, science articles, and natural speech transcripts to examine the psychometrics of words. Within this text corpus, articles represent 5.43% of all words used (where “a” = 1.96, “an” = 0.19, “the” = 3.27). The intercorrelation among these words is low but highly significant (“a” with “an” = .13, “a” with “the” = .09, “an” with “the” = .09), resulting in Cronbach’s $\alpha$ of .14 (for a summary of all reliability statistics, see Pennebaker et al., 2007).

Note that assessing the psychometrics of word use is even more complicated than what the above statistics suggest. To get reliability data for a questionnaire, we typically give people the same test of often-redundant questionnaire items on two occasions.
In theory, the questionnaire has exactly the same meaning on the two administrations. Asking people to, say, describe themselves on two occasions will generally evoke different types of responses. For example, within the open-ended response itself, people generally don’t repeat themselves (meaning one rarely gets good split-half reliability). Second, if people tell an experimenter who they are today, they will likely change their stories next time either because they have changed a bit or they want the experimenter to have a fuller sense of who they were from the previous time. Furthermore, saying the same thing as they did to the person on the first occasion would be redundant and, perhaps, a bit rude. In short, the psychometrics of word use pose a new set of problems that questionnaires avoid.

**Content Versus Style Words**

When LIWC was first developed, the goal was to devise an efficient system that could tap both psychological processes and the content of what people were writing or talking about. Within a few years, it became clear that there are two very broad categories of words that have different psychometric and psychological properties. **Content words** are generally nouns, regular verbs, and many adjectives and adverbs. They convey the content of a communication. To go back to the phrase “It was a dark and stormy night” the content words are: “dark,” “stormy,” and “night.” Intertwined through these content words are **style words**, often referred to as function words. Style or function words are made up of pronouns, prepositions, articles, conjunctions, auxiliary verbs, and a few other esoteric categories. In the phrase these words are “it,” “was,” “a,” and “and.”

Although we tend to have almost 100,000 English words in our vocabulary, only about 500 (or 0.05%) are style words. Nevertheless, style words make up about 55% of all the words we speak, hear, and read. Furthermore, content and style words tend to be processed in the brain very differently (Miller, 1995).

From a psychological perspective, style words reflect how people are communicating, whereas content words convey what they are saying. It is not surprising, then, that style words are much more closely linked to measures of people’s social and psychological worlds. Indeed, the ability to use style words requires basic social skills. Consider the sentence, “I will meet you here later.” Although grammatically correct, the sentence has no real meaning unless the reader knows who “I” and “you” refer to. Where is “here” and what is meant by “later”? These are all referents that are shared by two people in a particular conversation taking place at a particular time. To say this implies that the speaker knows that the listener shares the same knowledge of these style words (cf. Chung & Pennebaker, 2007).

**Caveats concerning computer text analysis.** Psychologists are always looking for measures that reveal the secret, hidden, or distorted “real” self. Freud’s popularity was partly attributable to his assertion that subconscious thoughts, emotions, and experiences drove our behavior. People continue to be enthralled with his methods of dream analysis, slips of the tongue, and other psychoanalytic claims. This trend continues with a new generation of measures and theories that rely on a host of implicit measures such as the implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998),
priming strategies, and various imaging techniques such as functional MRI that all hold out the promise of discovering the “real” person. Many people consider the analysis of language—especially function or style words—to do the same. And, indeed, they sometimes can reveal social psychological processes that people are not able to easily conceal.

Despite the appeal of computerized language measures, they are still quite crude. Programs such as LIWC ignore context, irony, sarcasm, and idioms. The word “mad,” for example, is currently coded as an anger word. When people say things such as “I’m mad about him,” or “He’s as mad as a hatter” the meaning and intent of their utterances will be miscoded. LIWC, like any computerized text analysis program, is a probabilistic system.

The study of word use as a reflection of psychological state is in its earliest stages. As described below, studies are providing evidence that function words can detect emotional and biological states, status, honesty, and a host of individual differences. Nevertheless, the imprecise measurement of word meaning and psychological states themselves should give pause to anyone who relies too heavily on accurately detecting people’s true selves through their use of words.

**The Social and Psychological Meaning of Words**

The words we use in daily life reflect what we are paying attention to, what we are thinking about, what we are trying to avoid, how we are feeling, and how we are organizing and analyzing our worlds. The 80 language categories in LIWC have been linked in hundreds of studies to interesting psychological processes. In this section, we give a brief discussion of psychological processes and a small set of related of language categories. The section concludes with a comprehensive summary of findings about the correlates of word categories from a large group of studies.

**Attentional Focus: Pronouns and Verb Tense**

Tracking people’s attention reveals information about their priorities, intentions, and thoughts. Infants, for example, focus on objects that display novelty, complexity, and motion (Berlyne, 1960), which shows the extent to which they are focused on learning. Our attention can oscillate from our external worlds to our internal feelings or sensations (e.g., Pennebaker, 1982). If we are playing a game of tennis, we might bruise our arm and not notice because our full attention is on the game itself. Alternatively, if the injury is significant, the pain may be so attention grabbing that we no longer are aware of the game at all.

Tracking language use such as tracking people’s gaze can tell us where they are attending. At the most superficial level, content word categories explicitly reveal where individuals are focusing. Those thinking about death, sex, money, or friends will refer to them in their writing or conversation. Function words, such as personal pronouns, also reflect attentional allocation. People who are experiencing physical or emotional pain tend to have their attention drawn to themselves and subsequently use more first-person singular pronouns (e.g., Rude, Gortner, & Pennebaker, 2004). When
people sit in front of a mirror and complete a questionnaire, they use more words such as “I” and “me” than when the mirror is not present (Davis & Brock, 1975). As we might expect, positive ads focus on the political candidate producing the ad and negative ads focus on their opponent; use of pronouns quickly reveals these differences (Gunsch, Brownlow, Haynes, & Mabe, 2000). Gunsch and colleagues show that more self-references (e.g., “I,” “we”) were present in positive political ads compared with mixed and negative political ads, whereas more other-references (e.g., “he,” “she,” “they”) were present in negative ads compared with positive and mixed ads.

Attention can reveal not just who someone is attending to but how they are processing the situation. Students who wrote about their experiences with teasing varied in the pronouns they used depending on whether they were teasing others or were being teased by others (Kowalski, 2000). Participants used more first-person singular and fewer third-person pronouns (e.g., “he,” “she”) when describing an event when they were being teased compared with when they described an event were they were teasing someone else. In both cases, the focus is on the person who was teased—the victim of the event. There was a significant interaction with sex and use of third-person pronouns; male participants used more third-personal pronouns when describing an event in which they were being teased than female participants. Compared with women, men may focus more on the perpetrator of the event when they are the victim, although it remains unclear why this is the case.

Whereas personal pronouns provide information about the subject of attention, analyses of the tense of common verbs can tell us about the temporal focus of attention. In the same study of political ads, the authors found that positive ads used more present and future tense verbs, and negative ads used more past tense verbs (Gunsch et al., 2000). From the tense of the verbs and the personal pronouns used, we can infer that negative ads focus on past actions of the opponent, and positive ads focus on the present and future acts of the candidate.

Studying attention also gives us a deeper understanding of how people are processing a situation or event. Participants were asked to either recall an event that they had discussed with someone else, or an undisclosed event; there were significant differences in the verb tense used in the two conditions (Pasupathi, 2007). Participants used greater past tense in discussing a disclosed event and greater present tense in discussing an undisclosed event. Verb tense differences could indicate increased psychological distance and a higher degree of resolution for disclosed events compared with undisclosed events.

Pronouns and verb tense are useful linguistic elements that can help identify focus, which, in turn, can show priorities, intentions, and processing. Some care should be taken in evaluating how pronouns and verbs are used. An exception to the pronoun-attention rule concerns first-person plural pronouns—“we,” “us,” and “our.” Sometimes “we” can signal a sense of group identity, such as when couples are asked to evaluate their marriages to an interviewer, the more the participants use “we,” the better their marriage (Simmons, Gordon, & Chambless, 2005). “We” can also be used as the Royal We, such as when the advisor announces to his or her graduate students that “we need to analyze that data.” The use of “we” in this case actually means “you students” rather than “you students and I” (see also use of the Royal We by political figures, such as Rudolph Guiliani in Pennebaker & Lay, 2002).
Emotionality: Positive and Negative Emotions

The degree to which people express emotion, how they express emotion, and the valence of that emotion can tell us how people are experiencing the world. People react in radically different ways to traumatic or important events; how people react may say a lot about how they cope with the event and the extent to which the event plays a role in the future. At the heart of reacting and coping with events is people’s emotional response.

Research suggests that LIWC accurately identifies emotion in language use. For example, positive emotion words (e.g., love, nice, sweet) are used in writing about a positive event, and more negative emotion words (e.g., hurt, ugly, nasty) are used in writing about a negative event (Kahn, Tobin, Massey, & Anderson, 2007). LIWC ratings of positive and negative emotion words correspond with human ratings of the writing excerpts (Alpers et al., 2005).

Use of emotion words has also been used as a measure of the degree of immersion. Holmes et al. (2007) found that among women trying to cope with intimate partner violence, using more positive and negative emotion words to describe the violence led to increased feelings of physical pain over the four writing sessions. The authors conclude that higher use of emotion words showed more immersion in the traumatic event, which led to increased experience of physical pain.

Language emotionality extends beyond the simple expression of more or less emotion; use of emotion words relate to other key language elements. In an examination of the random assortment of around 2,800 texts described earlier, emotion words were negatively correlated with articles (\(r = -0.33\)), prepositions (\(r = -0.38\)), and relativity words (\(r = -0.40\)). These language features as we discuss later, may be important in cognitive complexity and thinking styles. Emotion words were positively correlated with pronoun use (\(r = 0.29\)), auxiliary verb use (\(r = 0.29\)) and negation use (\(r = 0.32\)). All correlations are highly significant, \(p < .001\). The nature of these correlations suggests a deeper importance of the expression of emotion and thinking styles, and social awareness.

Social Relationships

Language at its most basic function is to communicate. Words provide information about social processes—who has more status, whether a group is working well together, if someone is being deceptive, and the quality of a close relationship. Word choice provides information about person perception (Semin & Fiedler, 1988). Certain language clues give away relationships. Pronouns reveal how an individual is referencing those in the interaction and outside of it. Word count explains who is dominating the conversation and how engaged they are in the conversation. Assents and positive emotion words measure levels of agreement. Other language cues are specific to the interaction; here we offer a few situations that have been studied.
Status, Dominance, and Social Hierarchy

Higher-status individuals speak more often and freely make statements that involve others. Lower-status language is more self-focused and tentative. In a study of groups of three crew members, a captain, a first lieutenant, and a second lieutenant engaging in several flight simulations, the use of greater first-person plural correlated with higher rank (Sexton & Helmreich, 2000). The authors found the opposite pattern for question marks: Higher-ranked crew members asked fewer questions compared with lower-ranked crew members. Across five studies in which status was either experimentally manipulated, determined by partner ratings, or based on existing titles, increased use of first-person plural was a good predictor of higher status, and in four of the studies increased use of first-person singular was a good predictor of lower status (Kacewicz, Pennebaker, Davis, Jeon, & Graesser, 2009). Leshed, Hancock, Cosley, McLeod, and Gay (2007) reported that members of small groups are rated as being more involved and task focused by their teammates if they use more words; supporting the assertion that total word count may also indicate status.

Social Coordination and Group Processes

More communication, more unity, and positive feedback may promote better group performance. Word count can act as a proxy for amount of communication; in some circumstances, more first-person plural may show group cohesion; and assents and question marks show how individuals are responding to each other. In the study of flight crews simulating easy and difficult flights, increased group word count, increased use of first-person plural, and increased use of question marks in early simulations predicted better team performance (Sexton & Helmreich, 2000). However, groups of 4 to 6 participants working on a joint task that used less first-person plural rated their group as having more group cohesion, although first-person plural was unrelated to group performance (Gonzales, Hancock, & Pennebaker, in press). The type of first-person plural pronouns may be important, if “we” is being used to promote interdependence as in “we can do this;” it may increase group cohesion if, on the other hand, it is being used to indirectly assign tasks as it may lead to resentment. Increased use of assents (e.g., agree, OK, yes) could signal increased group consensus and agreement; however, the timing of assents is important. Later in a group task, assents may signal consensus, early assents may indicate blind agreement by unmotivated group members (Leshed, Hancock, Cosley, McLeod, & Gay, 2007).

Honesty and Deception

Deceptive statements compared with truthful ones are moderately descriptive, distanced from self, and more negative. Newman, Pennebaker, Berry, and Richards (2003) investigated lying behavior in five experiments; in each experiment, lying was operationalized differently. Across the studies when participants were lying they used more negative emotion, more motion words (e.g., arrive, car, go), fewer exclusion
words, and less first-person singular. More motion words and fewer third-person pronouns were also significant predictors of deception by prisoners instructed to lie or tell the truth about videos they had watched (Bond & Lee, 2005). Hancock, Curry, Goorha, and Woodworth (2008) expanded these findings to study lying within pairs of participants over instant messenger. They found a similar pattern of language use when a participant was lying. They also found that the people being deceived, the partners of the participants lying, also changed their language. When one participant was lying both used a higher total word count, less first-person singular, and more sense words. Motion, exclusion, and sense words all indicate the degree to which an individual elaborated on the description of the scenario. Deceptive statements are balanced in descriptiveness because enough description is required to convince the other person of an untruthful statement but too much information might reveal inaccuracies. Using different linguistic measures, researchers found that non-naïve individuals assigned to be deceptive compared with non-naïve individuals assigned to be truthful or naïve individuals who were truthful used some language features that showed less diversity and complexity (Zhou, Burgoon, Nunamaker, & Twitchell, 2004). Exclusive words are also a marker of complexity. Complexity may be reduced in deceptive speech because of the cognitive load required to maintain a story that is contrary to experience, and the effort taken to try to convince someone else that something false is true.

Close Relationships

Pronoun use is very important in showing the quality of a close relationship, because it shows how individuals are referring to each other. Surprisingly, first-person plural (“we”) has not been found to be related to higher relationship quality, instead use of second person (“you”) is more important in predicting lower-quality relationships. Simmons, Chambless, and Gordon (2008) found that use of second-person pronouns was negatively related to relationship quality. They found in a study of relatives of participants suffering from either obsessive–compulsive disorder or panic attacks with agoraphobia that there were differences in the use of pronouns and that these differences signaled the extent to which they had a poor relationship with the patient. Relatives who used more second person in a taped interview with the patient scored higher on measures of criticism and having an overinvolved emotional reaction to the patient’s condition. In this study, use of second person showed hostility and willingness to confront the patient. In a study of archived instant message conversation between heterosexual romantic partners shows a marginal trend that increased use of second person by the male participant predicted lower ratings of relationship satisfaction (Slatcher, Vazire, & Pennebaker, 2008). Researchers have hypothesized that increased use of first-person plural in conversations between romantic partners should lead to increased ratings of relationship satisfaction and stability. In fact in the study of instant message transcripts of romantic partners shows that increased use of first-person singular by the women leads to higher ratings of satisfaction for both individuals, use of first-person plural is unrelated to the satisfaction. Higher positive emotion words for men lead to increased relationship satisfaction as well.
These are only a few possible interactions and related language categories. Patterns of language use are a rich tool in studying interactions, because so much of the interplay between individuals is carried out through language. However, language use depends on the situational context. For example, in a cooperative coordination context, higher total word count may signal better communication and agreement, whereas in a negotiation context it may signal a breakdown in agreement.

**Thinking Styles: Conjunctions, Nouns, Verbs, and Cognitive Mechanisms**

Language can track what information people are selecting from their environment by monitoring attentional focus. By the same token, natural language use provides important clues as to how people process that information and interpret it to make sense of their environment. Thinking can vary in depth and complexity; this is reflected in the words people use to connect thoughts. Language changes when people are actively reevaluating a past event. It can also differ depending on the extent to which an event has already been evaluated.

Depth of thinking can vary between people and situations; certain words can reveal these differences. Cognitive complexity can be thought of as a richness of two components of reasoning: the extent to which someone differentiates between multiple competing solutions and the extent to which someone integrates among solutions (Tetlock, 1981). These two processes are captured by two LIWC categories—exclusion words and conjunctions. Exclusive words (e.g., but, without, exclude) are helpful in making distinctions. Indeed, people use exclusion words when they are attempting to make a distinction between what is in a category and what is not in a category. Exclusive words are used at higher rates among people telling the truth (Newman et al., 2003) and by Gore compared with Kerry and Edwards (Pennebaker, Slatcher, & Chung, 2005). Conjunctions (e.g., and, also, although) join multiple thoughts together and are important for creating a coherent narrative (Graesser, McNamara, Louwerse, & Cai, 2004).

Prepositions (e.g., to, with, above), cognitive mechanisms (e.g., cause, know, ought), and words greater than six letters are all also indicative of more complex language. Prepositions, for example, signal that the speaker is providing more complex and, often, concrete information about a topic. “The keys are in the box by the lamp under the painting.” Within published journal articles, authors use more prepositions in the discussion than the introduction or abstract. Discussions are often the most complex part of an article because results must be integrated and differentiated from past findings (Hartley, Pennebaker, & Fox, 2003).

The use of causal words (e.g., because, effect, hence) and insight words (e.g., think, know, consider), two subcategories of cognitive mechanisms, in describing a past event can suggest the active process of reappraisal. In a reanalysis of six expressive writing studies, Pennebaker, Mayne, and Francis (1997) found that increasing use of causal and insight words led to greater health improvements. This finding suggests that changing from not processing to actively processing an event in combination of emotional writing leads to better outcomes. In these experiments, increasing use of
casual and insight words may be analogous to making reconstrual statements. In other work, use of reconstrual in combination with discussion of a traumatic events has shown to have the best health outcomes (Kross & Ayduk, 2008). Participants in describing a painful relationship breakup used more cognitive mechanisms, particularly causal words, in describing the breakup and postbreakup compared with the prebreakup (Boals & Klein, 2005). The authors argue that causal words are used in the most traumatic parts, the breakup and postbreakup, because they are being used to create causal explanations to organize the participant’s thoughts.

The language that people use to discuss an event can reveal something about the extent to which a story may have been established or is still being formed. When people are uncertain or insecure about their topic, they use tentative language (e.g., maybe, perhaps, guess) and more filler words (e.g., blah, I mean, you know). Participants who recounted an event that they had already disclosed to someone else used fewer words from the tentative category than participants who recounted an undisclosed event (Pasupathi, 2007). Possibly, higher use of tentative words suggests that a participant has not yet processed an event and formed it into a story. Similarly, Beaudreau, Storandt, and Strube (2006) found that in recounting a personal story younger participants used more filler words compared with older participants. However, there was no difference in filler words when the two groups described a story based on a picture. In this experiment, use of filler words may suggest the degree to which the story was well formed, presumably older participants had more perspective on the personal life events and may have recounted them many more times than the younger participants.

**Individual Differences**

The self-focus, cognitive complexity, social references, and emotional tone inherent in language use can help identify individual differences. These linguistic characteristics differ with age, sex, personality, and mental health. Language use, like any behavioral manifestation, can reflect individual differences. These language features can be used to make predictions about individuals and also may underlie causal processes that create some individual differences.

As people age, they become less self-focused, refer more to the moment, and do not decline in verbal complexity. Pennebaker and Stone (2003) examined the writing of participants of varying ages in emotional writing studies. In a second experiment, the authors examined the text of published authors from the span of their writing career. Across these two studies, first-person singular decreased with time, whereas insight words, future tense verbs, and exclusive words increased. The authors observe these patterns of language use both in studies of different individuals at different points in their lives, and of authors over the course of their life. From the results, they reason that there are shifts in self-focus as people age and, counter to expectations, attention to time is more present and future oriented, and verbal complexity may increase or at least stay the same as people age, evidenced by insight words and exclusive words.

Sex differences in language use show that women use more social words and references to others, and men use more complex language. A meta-analysis of the texts
from many studies shows that the largest language differences between males and females are in the complexity of the language used and the degree of social references (Newman, Groom, Handelman, & Pennebaker, 2008). Males had higher use of large words, articles, and prepositions. Females had higher use of social words, and pronouns, including first-person singular and third-person pronouns. There were also large effect sizes for use of swear words, feeling words, and present tense verbs. The fact that there are predictable differences in language used between sexes makes it possible to predict the sex of the user without knowledge of the true sex. An open research question remains what it means if a participant uses sex atypical language.

Studies measuring personality in participants through writing samples (Pennebaker & King, 1999) and spoken dialogue (Mehl, Gosling, & Pennebaker, 2006) have shown that some LIWC categories correspond with big-five personality traits. For example, Mehl and colleagues found that for both males and females higher word count and fewer large words predicted extraversion. Pennebaker and King showed that other LIWC categories showing complexity of language (such as articles, exclusive words, causal words, and negations) were less frequent in the writing of people who scored high on extraversion. Social and emotional language also differed with respect to extraversion; people who scored high on extraversion used more social words, more positive emotion, and less negative emotion. The findings from these two studies partially support traditional personality models. Models of extraversion would predict that extraverts engage in more social interaction, and have a more positive response to that engagement. Also, these models would predict that people high in extroversion would be less inhibited in their language production, possibly leading to less complex language.

Depressed and suicidal individuals are more self-focused, express more negative emotion and sometime use more death-related words. Studies on depression and suicide show that language features can be markers of mental health. Depressed patients are more likely to use more first-person singular and more negative emotion words than participants who have never been depressed in emotional writings (Rude et al., 2004). Suicidal poets in their published works compared with matched nonsuicidal poets use more first-person singular and more death-related words (Stirman & Pennebaker, 2001). This individual difference may show an attentional difference, that is, more self-focus in response to emotional pain, or it may indicate a thinking pattern that is a predilection for experiencing depression (see also work by Wolf, Sedway, Bulik, & Kordy, 2007, dealing with the language of anorexia).

Conclusion

The function and emotion words people use provide important psychological cues to their thought processes, emotional states, intentions, and motivations. We have summarized some of the LIWC dimensions that reflect language correlates of attentional focus, emotional state, social relationships, thinking styles, and individual differences. This review is, by definition, brief and selective. Word use is highly contextual and many of the findings may not hold with different groups of people or across a wide range of settings. More of the research results have come from labs in the United
States working with college-aged students, often in highly contrived settings. Very little work has explored the differences between spoken and written language.

As can be seen in the appendix, an increasing number of studies are beginning to link daily word use to broader social and psychological processes. What is most striking has been the relatively fast growth of the language–behavior research endeavor.

The connections between language and social psychology are changing at an accelerating rate. When journals such as the *Journal of Language and Social Psychology* were founded, most research was based on written text or transcriptions of spoken text, all of which were hand-typed, hand-scored, and stored in a filing cabinet for later analyses. Researchers interested in language and social processes have historically been trained in laboratory methods whereby participants were run, one at a time, in highly controlled settings to best capture the links between language use, cognitive processing, and communication dynamics.

Innovations in word analysis—as exemplified by Google and Yahoo—are challenging the social psychological methodologies most of us have grown up with. In the amount of time it takes to run a single participant in a social psychology language study, we can now download thousands of personal writings, interaction transcripts, or other forms of text that can be analyzed in seconds. The Internet world provides a far more diverse population from which to draw as well as access to a wide range of languages.

The availability of natural language use and our computational resources are transforming language analysis and modern social science. LIWC represents only a transitional text analysis program in the shift from traditional language analysis to a new era of language analysis. Newer text analysis will be able to analyze more complex language structure while retaining LIWC’s transparency. Studies have begun to look at n-grams, groups of two or more words together in the same way we have used LIWC to look at frequencies of single words (Oberlander & Gill, 2006). Text analysis methods should also increase in flexibility, allowing the researcher to examine language categories specific to his or her research program. New techniques to automatically extract conceptually related words should be expanded to incorporate related patterns of language style with related content words. From research using LIWC, it has become clear that language style information is critical to understanding a person’s state of mind.

Research using these new text analysis methods will also be expanded to capture cultural differences mirrored in language use. Language style conveys subtle information about social relations. The relevant social information can vary greatly between language and cultures (cf. Maass, Karasawa, Politi, & Suga, 2006). Indeed, some of the most striking cultural differences in language—such as markers of politeness, formality, and social closeness—are inherent in function words rather than content words (Boroditsky, Schmidt, & Phillips, 2003).

We are standing on the threshold of a new era of language analysis. One can easily imagine how Jim Bradac would have celebrated the possibilities of tracking natural language across hundreds of millions of people and an unknown number of contexts. The expanding galaxy of computer-based text analysis methods have the potential to add to our current ways of thinking about language and, in Bradac’s (1999) words, “burn ever brighter and illuminate the universe increasingly from their different places” (p. 11).
### Appendix

**Summary Table Linking LIWC Word Categories to Published Research Studies**

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
<th>Words in Category</th>
<th>Psychological Correlates</th>
<th>Published Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linguistic processes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word count</td>
<td></td>
<td></td>
<td>Talkativeness, verbal fluency, Verbal fluency, cognitive complexity</td>
<td>2, 9, 18, 19, 20, 24, 32, 35, 36, 39, 40, 48, 53, 54, 57, 60, 66, 70, 72, 73, 74, 86, 89, 103, 115</td>
</tr>
<tr>
<td>Words/sentence</td>
<td></td>
<td></td>
<td>Verbal fluency, cognitive complexity</td>
<td>3, 7, 39, 43</td>
</tr>
<tr>
<td>Dictionary words</td>
<td>(Percentage of all words captured by the program)</td>
<td></td>
<td>Informal, non-technical language</td>
<td>19, 42, 43, 65, 66, 85, 89</td>
</tr>
<tr>
<td>Words &gt;6 letters</td>
<td>(Percentage of all words longer than 6 letters)</td>
<td></td>
<td>Education, social class</td>
<td>3, 19, 20, 27, 35, 36, 42, 43, 73, 74, 79, 89, 90, 93, 103, 115</td>
</tr>
<tr>
<td>Total function words</td>
<td></td>
<td>464</td>
<td>Informal, personal</td>
<td>1, 19, 36, 43, 55, 89, 90, 119</td>
</tr>
<tr>
<td>Total pronouns</td>
<td>I, them, oneself</td>
<td>116</td>
<td>Personal, social</td>
<td>58, 79</td>
</tr>
<tr>
<td>Personal pronouns</td>
<td>I, them, her</td>
<td>70</td>
<td>Honest, depressed, low status, personal, emotional, informal</td>
<td>1, 3, 4, 5, 11, 13, 18, 27, 35, 36, 46, 55, 56, 64, 65, 66, 68, 69, 72, 73, 74, 78, 80, 81, 87, 90, 93, 94, 100, 101, 105, 108, 109, 112, 113, 115</td>
</tr>
<tr>
<td>First-person singular</td>
<td>I, me, mine</td>
<td>12</td>
<td>Detached, high status, socially connected to group (sometimes)</td>
<td>1, 4, 13, 18, 35, 46, 55, 64, 65, 74, 78, 81, 87, 90, 93, 94, 100, 103, 104, 105, 106, 113</td>
</tr>
<tr>
<td>First-person plural</td>
<td>We, us, our</td>
<td>12</td>
<td>Social, elevated status</td>
<td>1, 18, 27, 41, 55, 90, 100, 105, 106</td>
</tr>
<tr>
<td>Second person</td>
<td>You, your; thou</td>
<td>20</td>
<td>Social interests, social support</td>
<td>1, 3, 14, 36, 39, 55, 64, 66, 80, 87, 88, 90, 95</td>
</tr>
<tr>
<td>Third-person singular</td>
<td>She, her, him</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Third-person plural</td>
<td>They, their; they’d</td>
<td>10</td>
<td>Social interests, out-group awareness (sometimes)</td>
<td>1, 3, 14, 39, 55, 64, 80, 87, 88, 95</td>
</tr>
</tbody>
</table>

(continued)
### Appendix (continued)

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
<th>Words in Category</th>
<th>Psychological Correlates</th>
<th>Published Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indefinite pronouns</td>
<td>It, it’s, those</td>
<td>46</td>
<td>Use of concrete nouns, interest in objects and things</td>
<td>19, 36, 43, 74, 79, 80, 89, 92, 115</td>
</tr>
<tr>
<td>Articles</td>
<td>A, an, the</td>
<td>3</td>
<td></td>
<td>58, 79</td>
</tr>
<tr>
<td>Common verbs</td>
<td>Walk, went, see</td>
<td>383</td>
<td>Informal, passive voice</td>
<td>1, 13, 37, 62, 68, 73, 79, 87, 89, 91, 93, 115</td>
</tr>
<tr>
<td>Auxiliary verbs</td>
<td>Am, will, have</td>
<td>144</td>
<td>Focus on the past</td>
<td>1, 13, 37, 62, 68, 73, 79, 87, 89, 91, 93, 115</td>
</tr>
<tr>
<td>Past tense</td>
<td>Went, ran, had</td>
<td>145</td>
<td></td>
<td>1, 13, 37, 42, 62, 68, 73, 87, 89, 90, 93, 115</td>
</tr>
<tr>
<td>Present tense</td>
<td>Is, does, hear</td>
<td>169</td>
<td>Living in the past</td>
<td>13, 36, 37, 42, 62, 68, 73, 87, 89, 90, 93, 115</td>
</tr>
<tr>
<td>Future tense</td>
<td>Will, gonna</td>
<td>48</td>
<td>Future and goal oriented</td>
<td>13, 26, 37, 41, 62, 64, 76, 90, 93, 114</td>
</tr>
<tr>
<td>Adverbs</td>
<td>Very, really, quickly</td>
<td>69</td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>Prepositions</td>
<td>To, with, above</td>
<td>60</td>
<td>Education, concern with precision</td>
<td>43, 79, 89, 92, 115</td>
</tr>
<tr>
<td>Conjunctions</td>
<td>And, but, whereas</td>
<td>28</td>
<td></td>
<td>24, 39, 40, 48, 79, 89, 90, 114, 115</td>
</tr>
<tr>
<td>Negations</td>
<td>No, not, never</td>
<td>57</td>
<td>Inhibition</td>
<td>19, 79</td>
</tr>
<tr>
<td>Quantifiers</td>
<td>Few, many, much</td>
<td>89</td>
<td></td>
<td>24, 39, 40, 48, 79, 89, 90, 114, 115</td>
</tr>
<tr>
<td>Numbers</td>
<td>Second, thousand</td>
<td>34</td>
<td></td>
<td>58, 73, 74, 81, 98</td>
</tr>
<tr>
<td>Swear words</td>
<td>Damn, piss, fuck</td>
<td>53</td>
<td>Informal, aggression,</td>
<td>58, 73, 74, 81, 98</td>
</tr>
<tr>
<td>Psychological processes</td>
<td>Mate, talk, they, child</td>
<td>455</td>
<td>Social concerns, social support</td>
<td>1, 18, 23, 27, 32, 35, 41, 55, 78, 79, 85, 88, 89, 90, 93, 95, 97, 115, 116</td>
</tr>
<tr>
<td>Social processes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td>Daughter, husband</td>
<td>64</td>
<td></td>
<td>18, 95</td>
</tr>
<tr>
<td>Friends</td>
<td>Buddy, friend, neighbor</td>
<td>37</td>
<td></td>
<td>18, 95</td>
</tr>
<tr>
<td>Humans</td>
<td>Adult, baby, boy</td>
<td>61</td>
<td></td>
<td>1, 11</td>
</tr>
<tr>
<td>Affective processes</td>
<td>Happy, cried, abandon</td>
<td>915</td>
<td>Emotionality</td>
<td>12, 27, 28, 32, 33, 34, 40, 44, 50, 54, 57, 58, 60, 62, 69, 77, 85, 86, 119</td>
</tr>
</tbody>
</table>

(continued)
### Appendix (continued)

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
<th>Words in Category</th>
<th>Psychological Correlates</th>
<th>Published Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive emotion</td>
<td>Love, nice, sweet</td>
<td>406</td>
<td></td>
<td>2, 3, 4, 5, 6, 8, 10, 12, 15, 17, 21, 22, 23, 25, 28, 30, 31, 33, 36, 37, 38, 41, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 57, 59, 60, 61, 62, 64, 66, 67, 68, 69, 70, 71, 73, 74, 75, 76, 77, 81, 82, 85, 89, 91, 93, 94, 96, 99, 107, 108, 109, 110, 113, 115, 117, 118</td>
</tr>
<tr>
<td>Negative emotion</td>
<td>Hurt, ugly, nasty</td>
<td>499</td>
<td></td>
<td>2, 3, 4, 6, 10, 12, 13, 16, 17, 20, 21, 22, 25, 28, 29, 30, 33, 35, 37, 40, 44, 45, 46, 47, 48, 50, 51, 52, 53, 55, 57, 59, 61, 62, 63, 64, 66, 67, 70, 71, 72, 73, 74, 76, 79, 80, 81, 82, 84, 85, 89, 91, 92, 93, 94, 96, 99, 102, 107, 113, 115, 117, 119, 121</td>
</tr>
<tr>
<td>Anxiety</td>
<td>Worried, nervous</td>
<td>91</td>
<td></td>
<td>6, 28, 30, 66, 68, 77, 84, 85, 92</td>
</tr>
<tr>
<td>Anger</td>
<td>Hate, kill, annoyed</td>
<td>184</td>
<td></td>
<td>6, 28, 33, 50, 58, 66, 72, 74, 92</td>
</tr>
<tr>
<td>Sadness</td>
<td>Crying, grief, sad</td>
<td>101</td>
<td></td>
<td>6, 28, 33, 50, 58, 66, 72, 74, 92</td>
</tr>
<tr>
<td>Cognitive processes</td>
<td>Cause, know, ought</td>
<td>730</td>
<td></td>
<td>2, 3, 5, 8, 13, 18, 21, 23, 31, 32, 34, 46, 47, 49, 55, 58, 61, 68, 69, 71, 75, 83, 84, 85, 86, 89, 92, 93, 102, 104, 119, 120</td>
</tr>
<tr>
<td>Insight</td>
<td>Think, know, consider</td>
<td>195</td>
<td></td>
<td>1, 4, 18, 19, 25, 35, 37, 45, 53, 59, 68, 73, 76, 89, 90, 91, 92, 93, 97, 99, 111, 113, 115, 118, 119, 121, 122</td>
</tr>
<tr>
<td>Causation</td>
<td>Because, effect, hence</td>
<td>108</td>
<td></td>
<td>10, 13, 16, 20, 35, 37, 39, 45, 53, 72, 76, 89, 90, 91, 93, 97, 99, 115, 121, 122</td>
</tr>
<tr>
<td>Discrepancy</td>
<td>Should, would, could</td>
<td>76</td>
<td></td>
<td>10, 16, 18, 19, 49, 63, 74, 89, 115</td>
</tr>
<tr>
<td>Tentative</td>
<td>Maybe, perhaps, guess</td>
<td>155</td>
<td></td>
<td>18, 19, 24, 37, 38, 49, 73, 87, 89, 98, 115</td>
</tr>
<tr>
<td>Certainty</td>
<td>Always, never</td>
<td>83</td>
<td></td>
<td>38</td>
</tr>
<tr>
<td>Inhibition</td>
<td>Block, constrain, stop</td>
<td>111</td>
<td>Social/verbal skills, emotional stability</td>
<td>1, 16, 18, 19, 49, 90, 111</td>
</tr>
<tr>
<td>Inclusive</td>
<td>And, with, include</td>
<td>18</td>
<td></td>
<td>41, 60, 73, 74, 89, 115</td>
</tr>
<tr>
<td>Exclusive</td>
<td>But, without, exclude</td>
<td>17</td>
<td></td>
<td>24, 49, 73, 80, 89, 92, 93, 115</td>
</tr>
</tbody>
</table>

(continued)
## Appendix (continued)

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
<th>Words in Category</th>
<th>Psychological Correlates</th>
<th>Published Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptual processes</td>
<td>Observing, heard, feeling</td>
<td>273</td>
<td></td>
<td>14, 37, 120</td>
</tr>
<tr>
<td>See</td>
<td>View, saw, seen</td>
<td>72</td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>Hear</td>
<td>Listen, hearing</td>
<td>51</td>
<td></td>
<td>13, 41</td>
</tr>
<tr>
<td>Feel</td>
<td>Feels, touch</td>
<td>75</td>
<td></td>
<td>13, 88</td>
</tr>
<tr>
<td>Biological processes</td>
<td>Eat, blood, pain</td>
<td>567</td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>Body</td>
<td>Cheek, hands, spit</td>
<td>180</td>
<td></td>
<td>34, 36, 37, 49, 116</td>
</tr>
<tr>
<td>Health</td>
<td>Clinic, flu, pill</td>
<td>236</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sexual</td>
<td>Horny, love, incest</td>
<td>96</td>
<td></td>
<td>36, 94, 96, 112</td>
</tr>
<tr>
<td>Ingestion</td>
<td>Dish, eat, pizza</td>
<td>111</td>
<td></td>
<td>68, 94</td>
</tr>
<tr>
<td>Relativity</td>
<td>Area, bend, go</td>
<td>638</td>
<td></td>
<td>49, 110</td>
</tr>
<tr>
<td>Motion</td>
<td>Arrive, car, go</td>
<td>168</td>
<td></td>
<td>14, 37, 80</td>
</tr>
<tr>
<td>Space</td>
<td>Down, in, thin</td>
<td>220</td>
<td></td>
<td>14, 120</td>
</tr>
<tr>
<td>Time</td>
<td>End, until, season</td>
<td>239</td>
<td></td>
<td>1, 13, 41, 64, 93, 119, 120</td>
</tr>
<tr>
<td>Personal concerns</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work</td>
<td>Job, majors, xerox</td>
<td>327</td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>Achievement</td>
<td>Earn, hero, win</td>
<td>186</td>
<td></td>
<td>36, 60, 103</td>
</tr>
<tr>
<td>Leisure</td>
<td>Cook, chat, movie</td>
<td>229</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home</td>
<td>Apartment, kitchen, family</td>
<td>93</td>
<td></td>
<td>79</td>
</tr>
<tr>
<td>Money</td>
<td>Audit, cash, owe</td>
<td>173</td>
<td></td>
<td>41, 94</td>
</tr>
<tr>
<td>Religion</td>
<td>Altar, church, mosque</td>
<td>159</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death</td>
<td>Bury, coffin, kill</td>
<td>62</td>
<td></td>
<td>1, 2, 4, 35, 64, 68, 91, 94</td>
</tr>
<tr>
<td>Spoken categories</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assent</td>
<td>Agree, OK, yes</td>
<td>30</td>
<td>Agreement, passivity</td>
<td>48, 60, 81</td>
</tr>
<tr>
<td>Nonfluencies</td>
<td>Er, hm, umm</td>
<td>8</td>
<td></td>
<td>74</td>
</tr>
<tr>
<td>Fillers</td>
<td>Blah, I mean, yaknow</td>
<td>9</td>
<td>Informal, Unprepared speech</td>
<td>9, 74</td>
</tr>
</tbody>
</table>
Appendix (continued)

References Cited in the Table


(continued)
Appendix (continued)


(continued)
Tausczik and Pennebaker

Appendix (continued)


(continued)
Appendix (continued)


(continued)
Appendix (continued)


(continued)
Appendix (continued)


(continued)
Appendix (continued)

Appendix (continued)

Authors’ Note

The original version of this article was presented as part of the James J. Bradac Memorial Lecture at the University of California at Santa Barbara in 2008.

Declaration of Conflicting Interests

The text analysis program, LIWC, is a commercial product co-owned by Pennebaker. Proceeds from his share of the profits are all donated to the University of Texas at Austin. The authors declared no other conflicts of interests with respect to authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research and/or authorship of this article:

Army Research Institute (W91WAW-07-C002), DOD-CIFA (H9C104-07-C0019), and Sandia National Laboratories (26-3963-70).

References


**Bios**

Yla R. Tausczik is a doctoral student in the Department of Psychology at the University of Texas at Austin. She received her BA at the University of California at Berkeley in 2005. Her research interests include using language to understand group dynamics and natural language use in the workplace.

James W. Pennebaker (PhD, University of Texas, Austin) is a professor and chair of the Department of Psychology at the University of Texas at Austin. He is the author of multiple books, including *Opening Up: The Healing Power of Expressing Emotions* (1997). He has recently published in *Science, Psychological Science*, and *Journal of Personality and Social Psychology*. 