

# Using literature to understand authors

## The case for computerized text analysis\*

James W. Pennebaker and Molly E. Ireland

The University of Texas at Austin

Through computerized text analysis, the psychology of literature is on the threshold of becoming a dominant force in psychology and the social sciences. The ways people use words in their writing and in everyday life reflect people's social and psychological states. Whereas most text analysis research has focused on the content of people's writings, the current paper demonstrates that almost-invisible function words can be psychologically relevant as well. Through the analysis of pronouns, prepositions, and other function words used in literature, several studies demonstrate how authors' emotional states, aging processes, theories of mind, and the nature of their romantic and collaborative relationships are revealed through their words. The function word approach provides a glimpse of the rapidly expanding methods available to psychologists interested in tracking the social and psychological worlds of authors. With the upcoming release of data sets such as Google Books, the analysis of literature will likely serve as a foundational method used in the fields of psychology, linguistics, history, and other areas of the behavioral and social sciences.

**Keywords:** text analysis, LIWC, language, literature, function words

The goals of the scientific study of literature vary tremendously depending on the researcher. Several investigators are interested in the essence of literature itself. For example, what features of a narrative help to make a compelling or coherent story? Others examine literature at different historical periods to better understand the ways different cultures thought and felt. Yet another approach — one central to this paper — is to use various types of literature as historical archives that reflect the psychological states of their authors. Novels, poems, song lyrics, or other documents can tell us how the authors were thinking, feeling, and connecting with others at the time their work was published. More broadly, the analysis of literature can shed light on how people think about themselves and their worlds.

The ability to efficiently analyze a large number of authors and their works has been made possible in the last two decades by the revolution in computer technology. In this paper, one particular method will be highlighted that relies on a relatively simple word counting approach. We are standing on the threshold of a new world. The breathtaking new methods coming from computational linguistics, computer sciences, artificial intelligence, and cognitive and social psychology are changing the ways we can study and think about the written and spoken word. Over the next few years, these breakthroughs in computerized text analysis will build a rapidly increasing number of connections between the humanities and sciences, vastly enriching both disciplines in the process.

The current paper is divided into three overlapping sections. The first provides a background to our computer-based approach, with special attention devoted to the meaning of function words versus content words. The second section describes several studies that have used words to reveal the psychological states of various novelists, poets, playwrights, scriptwriters, and songwriters. The final section explores how published writing can also tell us about people's relationships with others.

### Background: Trauma narratives, text analyses, and function words

Our interest in the scientific study of literature came about almost by accident. In the 1980s, our research team discovered that if people are asked to write about emotional upheavals in their lives for as little as 15 minutes a day for three consecutive days, their physical health improved. Later studies found that expressive writing resulted in people visiting the doctor for illness at lower rates, enhancement in immune function, and other health-related outcomes (e.g., Pennebaker & Beall, 1986; Pennebaker & Chung, 2010; ~~for reviews and meta-analysis of almost 150 studies see Frattaroli, 2006~~). The mere ability to write a personal and emotional story has salutary benefits on health and well being.

Why would writing about an emotional upheaval result in better health? The expressive writing studies suggest that the answer is not simple. One approach to answering this question was to examine the stories themselves. Was it possible, for example, that some ways of writing a story yielded greater health improvements than others? Answering this presumably simple question lies at the heart of our research on language and literature.

Beginning in the early 1990s, it became apparent that relying on large groups of human judges to read and evaluate hundreds of depressing emotional essays was costly, slow, and often unreliable. Because no simple computerized text analysis program was available at the time, we wrote one. The underlying idea was that a text analysis program could evaluate individual essays by simply counting the

words that people used. We reasoned that one could assess emotional tone by counting emotion-relevant words. Similarly, cognitive styles, social concerns, and other psychological processes should be revealed through word use in the essays.

The program, which came to be known as Linguistic Inquiry and Word Count, or LIWC (pronounced “Luke”), took about three years to develop because it relied on multiple judges to evaluate which words were appropriate for our various language dictionaries. For example, in creating an anger dictionary to capture anger or hostility in writing, judges had to agree if each of dozens of words were appropriate. For example, words like *hate*, *furiosus*, and *rage* were obvious candidates for inclusion in the anger dictionary. But what about *mad* (yes), *hurt* (no), or *rude* (yes)? Just looking at these examples, it is clear that LIWC is a rather crude program. Consider the word *mad*. A person can be madly in love or mad as a hatter and not be the slightest bit angry. Nevertheless, LIWC proved to work as a probabilistic system — more times than not, when a person used the word *mad*, they used it in a way that conveyed anger.

The original LIWC program expanded rather quickly to include over 70 different language dictionaries that captured basic linguistic categories (e.g., personal pronouns, articles, prepositions), basic psychological processes (e.g., emotions, cognitive mechanisms, social and biological processes), and a small number of current concerns or content themes (e.g., work, leisure, religion, sex, illness). The current LIWC version has about 4,500 words and word stems in its 80 dictionaries and typically captures 80–90 percent of the words in most works of fiction (Pennebaker, Booth, & Francis, 2007).

Some of the initial LIWC analyses of the expressive writing narratives yielded some promising results. Basically, the more that people used positive emotion words in their writing (such as *love*, *caring*, and *nice*) the more likely they were to exhibit health improvements compared with people who used low rates of positive emotion words. More striking, however, were findings concerning cognitive words (such as *because*, *effect*, *reason*, and *understanding*). The more that people increased in their use of cognitive words from the first day of writing to the last, the more their health improved (Pennebaker, Mayne, & Francis, 1997).

The emotion and cognitive word findings made sense. At the same time, we started to see a confusing number of significant effects for pronouns, prepositions, and other forgettable language categories that had no clear meaning to us. These words were related to key variables such as age, sex, social class, emotional state, and personality measures we were collecting at the time. Fortunately, George Miller’s (1991) book, *The Science of Words*, helped to change our thinking.

According to Miller, words can be broken into two broad categories — content and function words. Content words include nouns, regular verbs, and most adjectives and adverbs. Function words are the small filler words that help contextualize

and connect content words, and include pronouns, prepositions, articles, conjunctions, auxiliary verbs, and a handful of other categories. Researchers and physicians who deal with brain injuries have long known that damage to Broca's area (which is located in the frontal lobe) results in people speaking in slow, hesitant ways using primarily nouns and regular verbs. Damage to Wernicke's area (in the temporal lobe) results in people being able to speak fluidly but using primarily function words. Evidence from brain damaged patients suggests a fact that becomes clear once one looks closely at function word usage in everyday language: Function words are profoundly social. Consider the following sentence:

"He was sitting next to me by the table."

The content words *sitting* and *table* are clearly understood by anyone who speaks English. All the other words, though, are function words. Who is *he*? The word *was* implies in the past — but when? What is *next to* and who is *me*? *The table* — which table would that be? Only the speaker and listener share the knowledge of what these function words refer to. To be able to use language with another person, we must all have the rudimentary social and cognitive skills to be able to interpret what function words mean within any given social context.

And that is the beginning of the story. Although there are fewer than 200 common function words in English, they account for almost 60 percent of the words we use in everyday language. By analyzing them, we can begin to understand people's thinking. For example, people's use of pronouns tells us where they are focusing their attention. Someone who uses *I* at high rates is by definition more self-focused than someone who doesn't. People who use a high rate of articles (*a*, *an*, and *the*) are by definition using more concrete nouns than people who don't use articles. Article users, then, are paying more attention to objects and things. And people who use high rates of conjunctions tend to be thinking more complexly than people who don't (for an overview, see Tausczik & Pennebaker, 2010).

Function words reveal people in ways that they cannot imagine.

### Understanding authors by measuring their function words

People's psychological states color the ways they perceive and interact with the world. Having a terrible fight with a coworker in the morning can make other relationships more difficult later in the day. Either our mind keeps returning to the fight thereby distracting us during other conversations or our negative moods come out in what we say or write. The social and psychological states of authors are no different. By tracking the ways that authors write we can begin to gain insight into issues that may be affecting them.

*Depression, suicide, and self-focus*

Being a published poet is one of the most dangerous jobs in America. According to Kay Jamison (1995), published poets are up to 18 times as likely as the general population to commit suicide. Her research suggests that poets have an extraordinarily high rate of bipolar disorder which is known to be closely linked to suicidal behaviors.

Bipolar disorder, as with other forms of severe depression, is associated with crippling psychological pain. One theory of depression suggests that emotional pain — like physical pain — forces people to pay attention to themselves (e.g., Pyszczynski & Greenberg, 1987). High degrees of self-focus can be seen in people's inattention to others and preoccupation with the self. Other work on self-focus suggests that states of self-awareness are linked to elevated use of first-person singular pronouns — especially the use of I-words such as *I*, *I'm*, *I'll*, etc. (e.g., Davis & Brock, 1975).

Depressive states have long been linked to higher use of I-words across several genres. In some of the first systematic word count studies, Walter Weintraub (1981) found that stream of consciousness writing produced much higher use of I-words in depressed patients than patients dealing with other medical disorders. In our own research, college students diagnosed with depression used more I-words than non-depressed students when writing essays about their college experiences (Rude, Gortner, & Pennebaker, 2004). In natural speech captured over several days of tape recordings, use of *I* is more frequent among those with high depression scores than those with low depression scores (Mehl, 2006).

Because poets commit suicide at such high rates, we initiated a study to analyze the poetry of 18 poets, half of whom were suicidal. Poets who eventually committed suicide used 1st-person singular pronouns at higher rates than those who did not commit suicide (Stirman and Pennebaker, 2001). Ironically, suicidal poets do not use more negative emotion words than other poets. Overall, suicidal poets' language use showed that they were focused more on the self, and less socially integrated than non-suicidal poets. Consider, for example, the lines of two American poets born about the same time — Sylvia Plath, who committed suicide, and Denise Levertov, who did not:

*Sylvia Plath*

I shut my eyes and all the world drops dead;  
I lift my lids and all is born again.  
(I think I made you up inside my head.)

The stars go waltzing out in blue and red,  
And arbitrary blackness gallops in:  
I shut my eyes and all the world drops dead.

I dreamed that you bewitched me into bed  
 And sung me moon-struck, kissed me quite insane.  
 (I think I made you up inside my head.)

God topples from the sky, hell's fires fade:  
 Exit seraphim and Satan's men:  
 I shut my eyes and all the world drops dead.

I fancied you'd return the way you said,  
 But I grow old and I forget your name.  
 (I think I made you up inside my head.)

I should have loved a thunderbird instead;  
 At least when spring comes they roar back again.  
 I shut my eyes and all the world drops dead.  
 (I think I made you up inside my head.)

— Mad Girl's Lovesong (1953/1996)

Denise Levertov

The ache of marriage:  
 thigh and tongue, beloved,  
 are heavy with it,  
 it throbs in the teeth

We look for communion  
 and are turned away, beloved,  
 each and each

It is leviathan and we  
 in its belly looking for joy, some joy  
 not to be known outside it

two by two in the ark of  
 the ache of it.

— The Ache of Marriage (1962/2002)

What makes the use of 1st-person singular pronouns in these two poems interesting is that you can see how Plath's use of *I* suggests that she is embracing her loss and sadness. Levertov, on the other hand, seems to be holding her aching away at arms' length — almost as if she is looking at it from a more distant (and safer) 3rd-person perspective. Indeed, as one reads the collected works of these two impressive authors, it is apparent how the two differ in owning or embracing their feelings of loss, alienation, and depression. Plath may be the better poet for this reason. With the tool of 1st-person singular pronouns, she takes us closer to the edge so that we can get a feeling of her personal despair.

### *The language of aging*

Age differences in function words are also robust. In analyses of personal writings collected during experiments or surveys from several thousand people, we found that people use fewer personal pronouns — especially first-person singular pronouns — with increasing age (Pennebaker and Stone, 2003). This, along with the greater use of exclusive words, suggests that as people age, they make more distinctions and psychologically distance themselves from their topics. In other words, older people speak with greater cognitive complexity. Interestingly, analysis of the use of emotion-laden words indicates that as people get older, they increase in their use of positive emotion while showing a drop in negative emotion word use. The aging project was interesting because all participants did their writing in the late 1990s or early 2000s.

A parallel project analyzed the collected works of a number of well-established poets, playwrights, and novelists over the course of their lives (for summary of results, see Pennebaker and Stone, 2003). The results were similar to the previous findings. For example, the older Jane Austen was, the less she tended to use first-person singular ( $r = -.59$ ); Shakespeare ( $-.22$ ) and Yeats ( $-.13$ ) followed the same pattern. Considering the general pattern of these correlations, one can see that aging is generally associated with works of literature subtly changing to reflect the authors' psychological state. The later plays, poems, and books by these authors tended to use fewer personal pronouns, more words related to positive emotion, and more markers of cognitive complexity.

### *Theories of mind*

Literature can also reveal differences in the ways that authors think about other people. In a recent study, we examined how scriptwriters of both sexes portrayed the dialog of men and women in their scripts (Ireland & Pennebaker, 2010). The goal was to assess the degrees to which authors are able to accurately imagine what the opposite sex might think, feel, and say by analyzing dialog from plays or movies.

Imagining *what* the opposite sex thinks about turns out to be simple given knowledge of traditional gender norms and common stereotypes. For example, men talk about money, work, and sports more than women do, and women focus more on home, family, and friends (Newman, Groom, Handelman, & Pennebaker, 2008). On the other hand, imagining *how* the opposite sex thinks is more challenging. Men and women reliably use function words differently, reflecting basic differences in the ways that men and women think. Compared with men, women focus more on themselves and others (greater use of first-person singular

and third-person singular pronouns), qualify their descriptions more often (more high frequency adverbs such as *very* and *just*), and use longer, more complex sentences (more conjunctions like *but* and *because*). Compared with women, men talk more about objects (greater use of articles) and locations (more prepositions such as *around* and *over*) (Ireland & Pennebaker, 2010; Mulac, Bradac, & Gibbons, 2001; Newman et al., 2008).

An analysis of dialog from more than 100 plays and movies showed that men replicate these stylistic sex differences in their fictional male and female characters' dialog more accurately than do women. For example, sex differences between male and female characters in Thornton Wilder's play *Our Town* are among the most accurate in our sample. The play tracks the lives of several characters, including Emily Webb. We see Emily as an eager high school student who later falls in love with Joe. After dying in childbirth, Emily appears one final time in the graveyard after she has just gone back in time and watched a scene from her childhood. When she returns to her seat in the cemetery in the final scene, she talks with some of the other dead souls about the shock of watching her early family life.

EMILY: I didn't realize. So all that was going on and we never noticed. Take me back — up the hill — to my grave. But first: Wait! One more look... Do any human beings ever realize life while they live it? — every, every minute?

STAGE MANAGER: No. The saints and poets, maybe — they do some. ...

EMILY: Oh, Mr. Stimson, I should have listened to them.

SIMON STIMSON: Yes, now you know. Now you know! That's what it was to be alive. To move about in a cloud of ignorance; to go up and down trampling on the feelings of those ... of those about you. To spend and waste time as though you had a million years. To be always at the mercy of one self-centered passion, or another. Now you know — that's the happy existence you wanted to go back to. Ignorance and blindness....

EMILY: They don't understand, do they?

MRS. JULIA GIBBS: No, dear. They don't understand.

Even in this brief scene, the women — Emily and her mother-in-law Julia Gibbs — focus on their own feelings and those of others. The men, Simon Stimson and the Stage Manager, use virtually no pronouns and describe the world in nicely objective, manly ways. Wilder's men do indeed talk like prototypical men, and his women talk like prototypical women.

Compare Wilder's writing with Callie Khouri's screenplay, *Thelma and Louise*. Through a series of haunting misadventures, the two main characters, Thelma and Louise, find themselves running from the law after shooting a would-be rapist. The primary investigator, Hal, talks with Louise by phone and pleads for her to give herself up:



HAL: How are things goin' out there?

LOUISE: Weird. Got some kind of snowball effect goin' here or somethin'.

HAL: You're still with us though. You're somewhere on the face of the earth?

LOUISE: Well, we're not in the middle of nowhere, but we can see it from here.

HAL: I swear. Louise, I almost feel like I know you.

LOUISE: Well. You don't.

HAL: You're gettin' in deeper every moment you're gone.

LOUISE: Would you believe me if I told you this whole thing is an accident?

HAL: I do believe you. That's what I want everybody to believe. Trouble is, it doesn't look like an accident and you're not here to tell me about it... I need you to help me here. ... You want to come on in?

LOUISE: I don't think so.

Khouri's men are interesting because their language use is actually more feminine than the women's. Hal's frequent use of pronouns and low use of articles points to his deep interest in other people as opposed to concrete objects. Louise's relatively impersonal sentences are more linguistically androgynous. In fact, all of Khouri's male characters — including the role played by Brad Pitt — have a feminine speaking style. At the same time, none would be considered remotely effeminate in their actions.

Sex differences in the accuracy of fictional dialog appear to reflect basic differences in men's and women's perspective-taking tendencies. Neuroimaging studies suggest that women simulate others' psychological states to a greater degree than do men (Derntl et al., 2010; Schulte-Rüther, Markowitsch, Shah, Fink, & Piefke, 2008), while men are more apt than are women to use a cognitive, rule-based route to infer others' thoughts and feelings (Saxe & Wexler, 2005). While women are normally better than are men at decoding and recalling nonverbal behavior (Hall, Murphy, & Mast, 2006; Rosip & Hall, 2004), women become worse than men at identifying emotional facial expressions when they are prevented from mimicking them (Stel & van Knippenberg, 2008). Consistent with these findings, our dialog analysis suggest that women are less able than are men to imagine others' minds in the abstract, without the benefit of face-to-face contact — perhaps because they practice abstract perspective-taking less in everyday life. Using literature, we are beginning to better understand some of the potential real-life consequences of these basic differences in the ways that men and women think about others' minds in different contexts.

### *The psychology of groups: The Beatles*

In recent years, we have expanded our research into the ways groups of people collaborate in writing songs, articles, or other works. By analyzing the ways individuals collaborate over time, we can begin to dissect the group's thinking styles

and relationships with each other. A good example comes from our research on the singing group, The Beatles (Petrie, Pennebaker, & Sivertsen, 2008).

During their 10 years together, the Beatles wrote and sang about 200 songs, a large number of which became world-wide hits. Three of the four members were particularly prolific: John Lennon, Paul McCartney, and George Harrison. Among lyrics that had a minimum of 50 words, we analyzed 78 by Lennon, 67 by McCartney, 25 by Harrison, as well as 15 by the Lennon-McCartney collaboration. In addition to using LIWC, we also worked with another computer program called Latent Semantic Analysis, or LSA, which is useful in analyzing word patterns. Specifically, LSA can tell us the degree to which any two songs' lyrics are mathematically similar in content or style (Landauer & Dumais, 1997).

Over their 10-year career, the Beatles' lyrics became darker, with dramatic drops in positive emotion words and increasing rates of negative emotion language. They also became more psychologically distant, using 1st-person pronouns and present tense verbs less frequently over time. Most striking, however, were the analyses of the different composers. Paul McCartney's lyrical style proved to be more variable and broad ranging than either Lennon's or Harrison's. Using LSA to assess lyric similarity, Harrison appears to have been more influenced by Lennon than he was by McCartney. Finally, the lyrics jointly written by Lennon and McCartney were mathematically more similar to Lennon's linguistic styles than McCartney's.

### Relationships among authors

Most of the original research on function words focused on individuals. That is, we were interested in exploring how the words produced by an individual reflected that particular individual's psychological state. Although valid, this assumption fails to appreciate the degree to which language is inherently social. After all, language evolved to make communication among people more efficient and comprehensive. In recent years, we have been exploring how the language of ongoing interactions can reveal important clues to the thinking of groups of people.

#### *Couples' relationships over time*

Literature is shaped not only by solitary authors but also by the people in the authors' lives, such as the people with whom they have close relationships. Sometimes, the two coincide. To investigate how authors influence and are influenced by people they are close with, we analyzed the life's work of married poets Elizabeth Barrett and Robert Browning and Sylvia Plath and Ted Hughes (Ireland &

Pennebaker, 2010). Using a measure of the function word similarity between pairs of texts called language style matching (LSM), we found that both couples' poems were more stylistically in sync during happier and more peaceful periods of their lives together. Notably, LSM was lowest when each pair was emotionally and psychologically out of sync: When Elizabeth Barrett Browning welcomed death and her husband dreaded it, and when Ted Hughes was in love with his mistress while Sylvia Plath remained in love with her husband.

Differences in LSM between the couples reflected well-known disparities between the couples. Even at its highest point, LSM for Plath and Hughes' poetry was much lower than the Brownings' average degree of matching. Taking into consideration the relative happiness of each couple, this did not come as a surprise. Although Hughes and Plath's courtship was passionate, Sylvia suffered from depression intermittently throughout her adult life and suspected Ted of infidelity from nearly the beginning of their relationship. The Brownings, on the other hand, were famously happy together. They overcame monolithic obstacles in order to elope and reportedly never spent a day of their married life apart. Even at the end when their synchrony waned, Elizabeth died peacefully in Robert's arms.

### *The psychology of collaborations*

Most of the history of literature has focused on the individual writer. Novels, song lyrics, and poems are generally penned by a single person. Within the academic world, however, articles are rarely solo authored. Indeed, there is a growing trend for popular novels to be jointly authored. For example, in December, 2010, 3 out of the top 25 novels on the New York Times had multiple authors. (It is also likely that others may have been ghost written with no acknowledgement given to the primary or supplemental authors.)

This raises a fascinating question. Does a joint collaboration between two authors result in a work that is the average of the two authors or does it yield a product unrelated to either's previous work? From a language perspective, we can easily address this question. Reanalyses of two case studies hint that two heads are very different than the sum of their, hmmm, parts (from Pennebaker, 2011). Recall that 15 of John Lennon and Paul McCartney's songs were true collaborations whereby both men agreed that it was impossible to disentangle who wrote what. Alexander Hamilton and James Madison jointly wrote three Federalist papers. In both sets of these collaborations, we are able to compare the language of jointly written work with that of songs and texts written alone. Across the various dimensions of language and even punctuation, we can calculate what percentage of the time the collaboration produces an effect that is the average of the two collaborators working on their own. There are three clear hypotheses:

1. *Just-like-another-member-of-the-team hypothesis*. Collaborative writing projects produce language that is similar to that produced by a single member of the team writing alone. Sometimes the work will use words like one author and other times like the other author.
2. *The average person hypothesis*. More interesting is the possibility that collaborations produce language that is the average of the two writers. If Lennon uses a low rate of *we*-words and McCartney uses a high rate, it would follow that their collaboration would produce a moderate number of *we*-words.
3. *The synergy hypothesis*. Even more interesting is the idea that when two people work closely together, they create a product unlike either of them would produce on their own. Their collaboration's language style will be distinctive in such a way that most people would not be able to recognize who the authors are.

And the winner is the synergy hypothesis. When Lennon and McCartney and Hamilton and Madison were working together, they produced works that were strikingly different from works produced by the individual writers themselves. When collaborating, the Lennon-McCartney team produced lyrics that were much more positive and used more *I*-words, fewer *we*-words, and much shorter words than either artist normally used alone. Similarly, when Hamilton and Madison worked together they used much bigger words, more past tense, and fewer auxiliary verbs than either did on their own. In fact, across about 75 dimensions of language and punctuation, more than 90 percent of the time collaborations resulted in language that differed significantly from the language of the two writers on their own.

Collaborations produce quite different language patterns than what the individuals would naturally do on their own. What's not yet known is if collaborative work is generally better than individual products. This is a research question that is begging to be answered.

### The promise of computerized text analysis for the psychology of literature

The relatively obscure discipline of psychology and literature is about to change. As the current paper suggests, the analysis of function words in books, lyrics, and poems can reveal a great deal about individuals and groups. Note that the analysis of function words is only one way by which to examine the psychology of authors. More traditional strategies such as content analysis which tends to focus on nouns and regular verbs can provide insight into authors' motives and the general themes that affect people (cf., Winter, 1991). Literature can provide a powerful window into the psychology of authors.

Our ability to understand historical figures and ancient civilizations is highly dependent on their surviving words. Very few cognitive, personality, or social psychologists were around to study Plutarch, Shakespeare, or even Franklin D. Roosevelt. None of these people completed reams of questionnaires or submitted to polygraph exams. Nevertheless, their words are still available for us to study along hundreds of dimensions. With increasing sophistication, we will be able to get a better sense about events and people that influenced their lives, their emotional upheavals, and the ways they thought about their worlds. Text analysis methods offer an exciting new way to track the psychology of individual authors and the societies that shaped them.

The most striking examples of how text analysis methods can reveal entire cultures through the analysis of literature are coming from the analyses of the Google Books project. Through their collections of several million published books in English and other languages over the last thousand years, the Google Books project will be able to track the evolution of language and thinking with increasing precision (Parry, 2010). We will get a sense of how cultural events such as war, disease, and famine shape the thinking of large groups of people over time. Changes in self-awareness, depression levels, and social behaviors will all be accessible through the written word. And the written word is overwhelmingly literature.

In all likelihood, literature will become the vehicle by which to understand culture and history. The analysis of literary texts will result in new fields of psychology that examine the history of thinking and emotions over large swaths of time. Social psychologists will be able to track daily social behaviors, mating rituals, and gender roles in ways no one could have imagined. Cognitive psychologists will be able to track the ways people categorize and organize concepts across time and culture. Personality researchers will be able to explore the structure of personality as a function of the authors' ages, language, and society.

## Note

\* Preparation of this paper was funded in part by Army Research Institute (W91WAW-07-C-0029) and the National Science Foundation (NSCC-0904822).

## References

- Chung, C.K., & Pennebaker, J.W. (2007). The psychological functions of function words. In K. Fiedler (Ed.), *Social communication* (pp. 343–359). New York, NY: Psychology Press.

(Michel, et al., 2010)

- Davis, D., & Brock, T.C. (1975). Use of first person pronouns as a function of increased objective self-awareness and performance feedback. *Journal of Experimental Social Psychology*, 11, 381-388.
- Derntl, B., Finkelmeyer, A., Eickhoff, S., Kellermann, T., Falkenberg, D.I., Schneider, F., & Habel, U. (2010). Multidimensional assessment of empathic abilities: Neural correlates and gender differences. *Psychoneuroendocrinology*, 35, 67-82.
- Frattaroli, J. (2006). Experimental disclosure and its moderators: A meta-analysis. *Psychological Bulletin*, 132, 823-865.
- Hall, J.A., Murphy, N.A., & Mast, M. (2006). Recall of nonverbal cues: Exploring a new definition of interpersonal sensitivity. *Journal of Nonverbal Behavior*, 30, 141-155.
- Ireland, M.E., & Pennebaker, J.W. (2010). Language style matching in writing: Synchrony in essays, correspondence, and poetry. *Journal of Personality and Social Psychology*, 99, 549-571.
- Ireland, M.E. & Pennebaker, J.W. (2011). Men imitate life, women underestimate it: Sex differences in scriptwriters' portrayal of naturalistic dialog. Manuscript submitted for publication, The University of Texas at Austin.
- Jamison, K.R. (1995). Manic-depressive illness and creativity. *Scientific American*, 272, 62.
- Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- Levertov, D. (1962/2002). The ache of marriage. In D. Levertov (Auth.) & P. A. Lacey (Ed.), *Selected poems* (p. 30). New York, NY: New Directions.
- Lieberman, E., Michel, J-B, Jackson, J., Tang, Y., & Nowak, M.A. (2007). Quantifying the evolutionary dynamics of language. *Nature*, 449, 713-16.
- Mehl, M.R. (2006). The lay assessment of subclinical depression in daily life. *Psychological Assessment*, 18, 340-345.
- Miller, G. (1991). *The science of words*. New York, NY: Scientific American Library.
- Mulac, A., Bradac, J., & Gibbons, P. (2001). Empirical support for the gender-as-culture hypothesis: An intercultural analysis of male/female language differences. *Human Communication Research*, 27, 121-152.
- Newman, M.L., Groom, C.J., Handelman, L.D., & Pennebaker, J.W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, 45, 211-236.
- Parry, M. (May 28, 2010). The humanities go Google. *Chronicle of Higher Education*. Available from: <http://chronicle.com/article/The-Humanities-Go-Google/65713/>
- Pennebaker, J.W. (2011). *The secret life of pronouns: How our words reflect who we are*. New York, NY: Bloomsbury Press.
- Pennebaker, J.W., & Beall, S.K. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology*, 95, 274-281
- Pennebaker, J.W., Booth, R.J., & Francis, M.E. (2007). Linguistic Inquiry and Word Count (LIWC2007): A computer-based text analysis program [Computer software]. Austin, TX: LIWC.net.
- Pennebaker, J.W. & Chung, C.K. (2010). Expressive writing and its links to mental and physical health. In H. Friedman (Ed.), *Oxford handbook of health psychology*. New York, NY : Oxford.
- Pennebaker, J.W., Mayne, T.J., & Francis, M.E. (1997). Linguistic predictors of adaptive bereavement. *Journal of Personality and Social Psychology*, 72, 863-871.
- Pennebaker, J.W. & Stone, L.D. (2003). Words of wisdom: Language use over the lifespan. *Journal of Personality and Social Psychology*, 85, 291-301.

→ Michel, J-B, Shen, Y.K., Aiden, A.P., Veres, A., Gray, M.K., The Google Books Team, Pickett, J.P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinket, S., Nowak, M.A., and Aiden, E.L. (2010). Quantitative analysis of culture using millions of digitized books. *Science*. DOI: 10.1126/science.1179644.

- Petrie, K.J., Pennebaker, J.W., & Sivertsen, B. (2008). Things we said today: A linguistic analysis of the Beatles. *Psychology of Aesthetics, Creativity, and the Arts*, 2, 197–202.
- Plath, S. (1953/1996). Mad girl's lovesong. In S. Plath (Auth.) & F. McCullough (Ed.), *The Bell Jar*. New York, NY: Harper Collins.
- Pyszczynski, T., & Greenberg, J. (1987). Self-regulatory perseveration and the depressive self-focusing style: A self-awareness theory of reactive depression. *Psychological Bulletin*, 102, 122–138.
- Rosip, J.C., & Hall, J.A. (2004). Knowledge of nonverbal cues, gender, and nonverbal decoding accuracy. *Journal of Nonverbal Behavior*, 28(4), 267–286.
- Rude, S.S., Gortner, E.M., & Pennebaker, J.W. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18, 1121–1133.
- ✓ Saxe, R. & Wexler, A. (2005) Making sense of another mind: The role of the right temporoparietal junction. *Neuropsychologia*, 43, 1391–1399.
- Schulte-Rüther, M., Markowitsch, H.J., Shah, N.J., Fink, G.R. & Piefke, M. (2008). Gender differences in brain networks supporting empathy, *NeuroImage*, 42, 393–403.
- Segalowitz, S.J. & Lane, K. (2004). Perceptual fluency and lexical access for lexical versus content words. *Behavioral and Brain Sciences*, 27, 307–308.
- Stel, M., & van Knippenberg, A. (2008). The role of facial mimicry in the recognition of affect. *Psychological Science*, 19, 984–985
- Stirman, S.W., & Pennebaker, J.W. (2001). Word use in the poetry of suicidal and non-suicidal poets. *Psychosomatic Medicine*, 63, 517–522.
- Tausczik, Y.R., & Pennebaker, J.W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language & Social Psychology*, 29(1), 24–54.
- Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory and Cognition*, 19, 95–112.
- Weintraub, W. (1981). *Verbal behavior: Adaptation and psychopathology*. New York, NY: Springer.
- Winter, D.G. (1991). Measuring personality at a distance: Development of an integrated system for scoring motives in running text. In A.J. Stewart, J.M. Healy, Jr. & D.J. Ozer (Eds.), *Perspectives in personality: Approaches to understanding lives* (pp. 59–89), London: Jessica Kingsley.

*Author's address*

James W. Pennebaker  
 Department of Psychology  
 University of Texas at Austin  
 Austin, TX 78712  
 USA

Pennebaker@mail.utexas.edu

Y<sup>#</sup>

✓ cut (1)