12

The Psychological Functions of Function Words

CINDY CHUNG and JAMES PENNEBAKER

THE PSYCHOLOGICAL FUNCTIONS OF FUNCTION WORDS

*L*anguage is the currency of most human social processes. We use words to convey our emotions and thoughts, to tell stories, and to understand the world. It is somewhat odd, then, that so few investigations in the social sciences actually focus on natural language use among people in the real world. There are many legitimate reasons for not studying what people say or write. Historically, the analysis of text was slow, complex, and costly. The purpose of this chapter is to suggest that social scientists in general and social psychologists in particular should reconsider the value of language studies. With recent advances in computer text analysis methods, we are now able to explore basic social processes in new and rich ways that could not have been done even a decade ago.

When language has been studied at all within social psychology, it has usually relied on fairly rigorous experimental methods using an assortment of standardized human coding procedures. These works are helping researchers to understand social attribution (Fiedler & Semin, 1992), intercultural communication (Hajek & Giles, 2003), and even how different cultures think about time (Boroditsky, 2001). When verbal samples have been collected, it has often been assumed that the best strategy is to not ask about one's personal states directly. Instead, participants have been asked to describe an ambiguous picture or tell a story, and the deep under-lying meaning in the elicited statements has been interpreted (e.g. Schultheiss & Brunstein, 2001; Winter & McClelland, 1978)

Over the last decade, a small group of researchers have adopted a somewhat different strategy. Their goal has been to understand how the words people use in their daily interactions reflect who they are and what they are doing. As detailed below, this strategy has also been method-driven. With the development of increasingly versatile computer programs and the availability of natural language

text on the internet, we are now standing at the gates of a new age of understanding the links between language and personality. It should be emphasized that this method-driven approach has also forced us to begin investigations by looking at word usage rather than exploring the broader meaning of language within a phrase or sentence (e.g. Semin, Rubini, & Fiedler, 1995), conversational turn (Tannen, 1993), or an entire narrative (McAdams, 2001).

This chapter summarizes much of our own research that attempts to map and understand how word use can reflect basic social, personality, cognitive, and biological processes. Relying on computerized text analysis procedures, we are finding that the examination of often-overlooked "junk words" – more formally known as function words or particles – can provide powerful insight into the human psyche.

RECENT DEVELOPMENTS IN MEASUREMENT

It is beyond the scope of this paper to summarize the many computerized strategies available to researchers (for a more comprehensive review see Pennebaker, Mehl, & Niederhoffer, 2003). Some methods, for example, simply count words related to particular themes (e.g., the DICTION program: Hart, Jarvis, Jennings, & Smith-Howell, 2005), whereas others look for words or phrases that reveal psychoanalytic concerns (Gottschalk, 1997) or themes related to drives or motives (e.g., the General Inquirer: Stone, Dunphy, & Smith, 1966). Various inductive methods have been evolving from the world of artificial intelligence. One such program, called Latent Semantic Analysis (LSA; Foltz, 1996), compares the similarity of any two texts in terms of their content.

In our laboratory, we have been relying on a text analysis program that we developed called Linguistic Inquiry and Word Count, or LIWC (Pennebaker, Francis, & Booth, 2001). LIWC searches for and counts both content and style words within any given text file. LIWC was developed by having groups of judges evaluate the degree to which about 2000 words or word stems were related to each of several dozen categories. The categories include negative emotion words (sad, angry), positive emotion words (happy, laugh), standard function word categories (first, second, and third person pronouns, articles, prepositions), and various content categories (e.g., religion, death, occupation). For each essay, LIWC computes the percentage of total words that these and other linguistic categories represent.

The original intent of this program was to better understand how people used language when writing about emotional upheavals in their lives. Starting in the 1980s, we discovered that when people wrote about traumatic experiences for 3–4 days for as little as 15–30 minutes per day, they subsequently exhibited improvements in physical health (e.g., Lepore & Smyth, 2002; Pennebaker, Kiecolt-Glaser, & Glaser, 1988). LIWC, then, allowed us to see what word types ultimately correlated with health changes.

The development of LIWC resulted in researchers in other laboratories sending us their own text samples from their experiments to analyze. Soon, we

had hundreds, then thousands of essays written by people from all over the English-speaking world in text format. With the rapid development of the Internet, we began to expand our text archive. Although we now have over 400,000 text files in our archive, this article focuses on the analyses of approximately 95,000 text files representing over 80,000 different people. As can be seen in Table 12.1, the data for part of this paper are based on the analysis of 67 million words across seven written and spoken genres.

FUNCTION WORDS VERSUS CONTENT WORDS

Simply counting words is an admittedly crude way to understand what people are saying. Most computer programs do a poor job of appreciating context. They are generally unable to appreciate irony, sarcasm, and the use of metaphors. In English, words often have different meanings in different settings. The LIWC program, for example, counts the word "mad" as an anger and negative emotion word. Phrases such as "I'm mad about my lover" and "he's mad as a hatter" are simply miscoded. Word count programs are ultimately probabilistic.

More problematic is deciding what words should be counted. Most early content analysis approaches by both humans and computers focused on words that suggested specific themes. By analyzing an open-ended interview, a human or computer can detect theme-related words such as family, health, illness, and money. Generally, these words are nouns and regular verbs. Nouns and regular verbs are "content heavy" in that they define the primary categories and actions dictated by the speaker or writer. It makes sense. To have a conversation, it is important to know what people are talking about.

There is much more to communication than content. Humans are also highly attentive to the ways people convey a message. Allport (1961) emphasized the idea of stylistic behaviors or, more broadly, personality styles. The ways people walk, use gestures, and even peel an orange can reflect their motives, needs, and important dimensions of personality. Just as there is linguistic content, there is also linguistic style – how people put their words together to create a message.

What accounts for "style"? Consider the ways in which three different people might summarize how they feel about ice cream:

Person A: I'd have to say that I like ice cream.
Person B: The experience of eating a scoop of ice cream is certainly quite satisfactory.
Person C: Yummy. Good stuff.

All three are saying essentially the samething, but their ways of expressing themselves are hinting at other issues: Person A is a bit tentative; Person B is overly formal and stiff; Person C more easy-going and uninhibited. The three people differ in their pronoun usage, use of large versus small words, verbosity, and dozens of other dimensions. We can begin to detect linguistic style by paying attention to "junk words" – those words that do not convey much in the way of

TABLE 12.1　Text Archive Characteristics

	Descriptions	Experiments	Internet	Published	Personal	Spoken	Grand Total
Examples	Non-emotional descriptions of an object, event, daily routine	Expressive writing about emotional events	Blogs, bulletin board posts, chat room logs	Novels, lyrics, poems, newspapers	Diaries, stories, personal accounts of emotional events	Natural conversation, TV/radio interviews, speeches	
Total files	11,347	12,975	9537	10,870	34,988	16,782	96,499
Number of words	5,632,475	5,099,444	3,305,468	26,641,920	14,997,848	11,095,099	66,772,254
Different words	53,619	41,285	60,927	132,850	79,963	51,466	
Mean letters/word	4.25	3.97	4.02	4.58	3.97	3.89	4.11

content. These junk words, usually referred to as function words or particles, serve as the cement that holds the content words together.

Function words include pronouns, prepositions, articles, conjunctions, and auxiliary verbs. Whereas the average native English speaker has an impressive vocabulary of well over 100,000 words, fewer than 400 are function words (Baayen, Piepenbrock, & Gulikers, 1995). This deceptively trivial percentage (less than 0.04%) of our vocabulary accounts for over half of the words we use in daily speech (Rochon, Saffran, Berndt, & Schwartz, 2000). Despite the frequency of their use, they are the hardest to master when learning a new language (Weber-Fox & Neville, 2001).

Table 12.2 lists the 20 most commonly used words in our text archive. All are function words and are used at surprisingly high rates. The top ten words alone account for over 20% of the words we use. As can be seen, function words are generally very short (usually 1–4 letters), are spoken quickly (at a speed of 100–300 milliseconds – the rate often used in laboratory studies testing priming or subliminal perception), and glossed over even more quickly when we read (Van Petten & Kutas, 1991).

We have a terrible memory of our own as well as other's use of function words. When composing a letter or making a speech, we might think briefly about these words. In daily conversation, however, we have virtually no control or memory over how and when they are used either by the speaker or by ourselves. As evidence, estimate how frequently you have seen articles (a, an, the) on the last page. Has this paper used more or fewer articles than you would in normal speech? [Hint: the answer is much more: 6.6% in this chapter compared to 4.0% in normal speech.] Despite rarely paying them any conscious attention, function words have a powerful impact on the listener/reader and, at the same time, reflect a great deal about the speaker/writer. Returning back to the three hypothetical people describing ice cream, their different uses of function words mark them in predictable ways. The ways people use function words reflects their linguistic style.

Humans, of course, are highly social animals. If we examine the human brain and compare it with every other mammal, the frontal lobe of the cerebral cortex is disproportionately large. In recent years, researchers have begun to emphasize the frontal lobe in guiding our social behaviors (e.g., Damasio, 1995; Gazzaniga, 2005). Most social emotions, skills in reading others' emotions and intentions, and the ability to connect with others are highly dependent on an intact frontal lobe.

Language, too, has an important link to frontal lobe function. In general, the majority of language functions are housed in the temporal and frontal lobes. Within the left temporal lobe (at least for most people) is Wernicke's area. Wernicke's area is critical for both understanding and generating most advanced speech – including nouns, regular verbs, and most adjectives. Broca's area, on the other hand, is situated in the left frontal lobe. Damage to Broca's area – while Wernicke's area is intact – results in people speaking in a painfully slow, hesitating way, often devoid of function words. People with functioning Broca's area – but with damage to Wernicke's area – exhibit a completely different social style. These people often speak warmly and fluidly while maintaining eye contact with the target person. The only problem is that they primarily use function words with no

TABLE 12.2   Frequency of the 20 Most Commonly-Used Words as a Function of Genre (from our text archive)

	Descriptions	Experiments	Internet	Published	Personal	Spoken	Mean
I	2.63	5.75	2.57	1.04	5.35	4.47	3.64
The	3.99	3.18	3.00	4.93	2.98	2.77	3.48
And	2.48	3.28	1.90	3.14	3.25	3.46	2.92
To	2.99	3.57	2.31	2.54	3.20	2.83	2.91
A	1.99	1.95	1.76	1.84	2.08	2.02	1.94
Of	1.96	1.57	1.33	3.02	1.65	1.47	1.83
That	1.29	1.67	1.06	0.90	1.92	2.06	1.48
In	1.33	1.20	1.09	1.83	1.24	1.06	1.29
It	1.07	1.26	0.97	0.71	1.39	1.75	1.19
My	0.64	2.28	0.65	0.37	1.53	0.99	1.08
Is	1.06	0.91	1.29	0.64	1.30	1.15	1.06
You	1.45	0.32	1.06	0.70	0.84	1.93	1.05
Was	0.72	1.40	0.56	0.67	1.25	1.45	1.01
For	0.87	0.91	0.79	0.89	0.74	0.61	0.80
Have	0.63	0.86	0.70	0.39	0.88	0.77	0.70
With	0.68	0.75	0.55	0.71	0.69	0.63	0.67
He	0.56	0.64	0.36	0.60	0.80	1.03	0.66
Me	0.55	1.03	0.44	0.31	0.82	0.70	0.64
On	0.77	0.67	0.60	0.65	0.55	0.56	0.63
But	0.50	0.71	0.48	0.38	0.83	0.80	0.62
Top 10 words	21.18	25.91	17.37	20.86	24.65	24.21	21.76
Top 20 words	28.16	33.91	23.47	26.26	33.29	32.51	29.60
Top 50 words	39.85	47.26	34.55	34.95	47.71	47.95	41.82

Note: Numbers reflect percentage of total words within any given text. For example, in any given text from the Description archive, 2.64% of all words are the word "I" (this includes I'm, I'd, I'll, I've).

content at all (e.g., Miller, 1995). Even at the brain level, then, function words are linked to social skills.

A closer analysis of function words points to their social functions more clearly. Pronouns, for example, are words that demand a shared understanding of their referent between the speaker and listener. Consider the following sentence:

I can't believe that he gave it to her.

This is a completely normal sentence. We can imagine someone saying this to us and knowing exactly what is meant. This sentence makes absolutely no sense, however, unless you know who the "I", "he", and "her" are, as well as what the "it" is. In a normal conversation, we would know who the various players and objects were based on shared knowledge between the speaker and listener. Some social skills are required here. The speaker assumes that the listener knows who everyone is. The listener must be paying attention and know the speaker to follow the conversation. So the mere ability to understand a simple conversation replete with function words demands social knowledge.

The same is true for articles, prepositions, and all other function words. Consider the slightly altered sentences:

I can't believe that he gave her the ring.
I can't believe that he gave her a ring.

The difference between "the" ring and "a" ring is subtle but significant. These sentences hint to possible differences in the speaker's and audiences' shared knowledge, contexts, and interpersonal relationships. Words such as "before", "over", and "to" similarly require a basic awareness of the speaker's location in time and space. The ability to use function words, then, is a marker of rather sophisticated social skills. Talking about nouns and verbs, however, simply requires the ability to understand culturally shared categories and definitions.

FUNCTION WORDS AND SOCIAL PROCESSES

For the last few years, we have begun to track the usage of function words across multiple settings. Most of these studies have focused on pronouns and, occasionally, on articles and prepositions. Given that function words are so difficult to control, examining the use of these words in natural language samples has provided a non-reactive way to explore social and personality processes. Much like other implicit measures used in experimental laboratory studies in psychology, the authors or speakers we examine often are not aware of the dependent variable under investigation (Fazio & Olson, 2003). In fact, most of the language samples we have analyzed come from sources in which natural language is recorded for purposes other than linguistic analyses, and therefore have the advantage of being more externally valid than the majority of studies involving implicit measures.

It is possible that changing communication goals and contexts may drive function word use. This possibility has yet to be ruled out. However, given the wide range of text corpora examined, it is unlikely that specific external factors drive the reported effects. The links between function words and social processes remain, at present, correlational. But the fact that function words do vary according to psychological states is a novel and important finding. Future research can improve upon the findings by adopting linguistic indices for discriminant validity, or through the rapprochement of other assessment methods for predictive validity. Here, we briefly describe some of our most robust findings. We begin with links between words and biological activity and move across levels of analysis to the ways in which words can reflect cultural differences.

Empirical Evidence

Biologicial Activity   Surprisingly few researchers have examined the possible links between biological activity and function words. Scherwitz, Berton, and Leventhal (1978), for example, found that coronary-prone Type A interviewees who used first person singular pronouns more frequently exhibited higher blood pressure than did those who referred to themselves less frequently. Type B interviewees, who are not prone to coronary heart disease (CHD), did not exhibit a relationship between self-references and any of the measures taken. In a later prospective study, neither density nor frequency of self-references could predict CHD, but the relationship for frequency of self-references and Type A personality remained significant (Graham, Scherwitz, & Brand, 1989).

In our own work, we have recently examined manipulated changes in testosterone with language use. In the study, two adults (one biological male and one biological female) who were undergoing testosterone therapy for different reasons provided us with 1–2 years of their daily text files – personal journal or outgoing emails – as well as a history of their testosterone injections (Pennebaker, Groom, Loew, & Dabbs, 2004). Overall, testosterone had the effect of suppressing the participants' use of non-I pronouns. That is, as testosterone levels dropped in the weeks after the hormone injections, the participants began making more references to other humans. Contrary to stereotypes about the subjective experience of energy, positive affect, heightened sexuality, and aggression thought to be related to this hormone, no consistent mood or other linguistic correlates of testosterone emerged. One function of testosterone, then, may be to steer people's interests away from other people as social beings.

Depression   Across multiple studies, we have found that use of first person singular is associated with negative affective states (see also Weintraub, 1989). When asked to write about coming to college, currently depressed students use more first person singular pronouns than either formerly depressed or never depressed students. In addition, formerly depressed students use more first person singular pronouns than never depressed students (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, 2004). In both studies, pronouns are a better marker of depression than the use of negative emotion words.

In the analysis of the poetry of suicidal versus non-suicidal poets, poets who eventually committed suicide used first person singular pronouns at higher rates than those who did not commit suicide (Stirman & Pennebaker, 2001). Overall, suicidal poets' language use showed that they were focused more on the self and were less socially integrated than non-suicidal poets.

Reactions to Individual Life Stressors

Rudolph Guiliani was mayor of New York City from 1993 to 2001. He held press conferences multiple times per year answering a wide array of questions from the press. In late Spring 2000, a series of events occurred to him within a month: he announced the breakup of his marriage, his affair with another woman was made public, he was diagnosed with prostate cancer, and he withdrew from the senate race against Hillary Clinton. Text analyses of his press conferences in the months surrounding his personal upheavals revealed that his use of first person singular pronouns increased from about 2% of his words to over 7% (Pennebaker & Lay, 2002).

Equally intriguing was his shift in first person plural words. The cultural stereotype is that words such as "we" and "us" reflect the speaker's close emotional ties to others. Sometimes this is true; just as often, it is not. Males especially use "we" in a distancing or royal-we form: "we need to analyze that data" or "we aren't going to put up with higher taxes." In Guiliani's press conferences during his first four years of mayor, he used "we" words at exceptionally high rates – over 2.5% of his total words in press conferences. When his life fell apart, this rate dropped to the more normal rate of 1%. The 9/11 attacks brought Giuliani to the center of the world's stage where he was viewed as heroic in his strength and warmth. During the final phase of being mayor, his use of "I" words was 3% and "we" words was 3.2%. Interestingly, judges who rated his use of "we" words found that his early mayor period was marked by distanced or royal "we" words whereas his post-9/11 "we" words referred to specific individuals or identifiable groups.

Reactions to Socially-Shared Stressors

Whereas first person singular pronouns suggest attention on the self, most other pronouns implicitly or explicitly suggest that the person is attending to other individuals. Congruent with the social support literature, the more that people make reference to others, the healthier they are. Findings concerning the use of third person pronouns (she, he, they) suggest that their use is linked to adaptive coping that leads to physical health benefits.

Using an alternative text analysis method based on latent semantic analysis, it was found that people who alternated in their use of personal pronouns – switching from high rates of "I" to high rates of other personal pronouns when writing about emotional upheavals in their lives – evidenced greater health improvements in the months after writing (cf., Campbell & Pennebaker, 2003). More recently, we have reanalyzed three previous expressive writing studies and found a positive correlation between non-I personal pronoun use and subsequent health: $r = .29$, $p < .01$.

Across every study we have conducted dealing with a cultural and/or community-wide upheaval, people's use of first person plural pronouns increases. These studies include chat room discussions in the wake of Princess Diana's death (Stone & Pennebaker, 2002) and newspaper accounts of the Texas A&M Bonfire tragedy (Gortner & Pennebaker, 2003). Most striking, however, has been the analysis of over 1000 bloggers who were tracked in the months before and after 9/11 (Cohn, Mehl, & Pennebaker, 2004).

In the last decade, millions of Americans have discovered online bulletin boards or web logs (blogs). One such blog is LiveJournal.com. At the time of this writing, LiveJournal receives over 40,000 posts per hour from its 2–3 million active members. Working with LiveJournal, we downloaded the postings of over 1000 people who wrote at relatively high rates in the two months before and after 9/11. Analyses of these 71,800 text files revealed startling changes in pronoun use over time. First, people dropped in their use of first person singular pronouns in the hours after the 9/11 attacks from a baseline of 7.1% to 5.9%. Within about a week, their usage was still significantly below baseline (6.7%) where it remained for the next two months of monitoring. Interestingly, a corresponding increase in first person plural pronouns occurred, that is, people switched from attending to themselves to focusing on friends, family, and others within their group.

Linguistic and acoustic data from people who happened to be wearing an electronically activated recording device (called the EAR) during and immediately following the 9/11 attacks provided further support for the relation between non-I pronouns and belongingness (Mehl & Pennebaker, 2003). The elevated use of non-I personal pronouns in natural speech after the 9/11 attacks occurred at the same time that people changed in their patterns of social interactions. Overall, there was a reduction in the amount of time that people spent in groups of three or more whereas a corresponding increase in dyadic interactions occurred. In other words, in the 5–6 days after the attacks, people spent more time at home with one other person rather than congregating in large or moderate-sized groups. Interestingly, the more that people deviated from this social profile, the less well-adjusted they appeared to be two weeks later.

Based on the above findings, what does the use of first person singular reflect? At its most basic level, the use of the word "I" suggests that the speaker is briefly paying attention to the self. Too much attention to the self is associated with highly negative emotional states such as depression. Interestingly, relatively healthy people facing the upheavals of 9/11 actually evidenced a drop in "I" words rather than an increase. Feeling sad is quite different from being depressed. To the degree that an emotional upheaval results in people feeling closer to others, it may actually be associated with adaptive coping. Indeed, in a study of Texas A&M students dealing with the tragic death of 12 fellow students, we discovered that the student body used elevated rates of "we" and reduced use of "I" in newspaper articles and letters. All indications are that the students were extremely saddened by the events. However, over the next 6 months, students went to the student health center for illness at much lower rates than they had the year before or in comparison with students at other universities at the time (Gortner & Pennebaker,

2003). Pronouns, then, are powerful markers of affiliation, with implications for predicting health outcomes.

Deception   Pronouns and other function words also provide hints about the truthfulness of statements. Conjunctions, negations, and certain prepositions are used to make important distinctions about categories. A particularly interesting class of words is exclusive words. These include words like "but", "except", "without", "exclude". Factor analytically, these words typically load with negations (no, not, never), and are associated with greater cognitive complexity (Pennebaker & King, 1999). Across multiple experiments where people have been induced to describe or explain something honestly or deceptively, the combined use of first person singular pronouns and exclusive words predicts honesty (Newman, Pennebaker, Berry, & Richards, 2003). In other words, when people are telling the truth (as opposed to lying), they are more likely to "own" it by making it more personal and, at the same time, are more likely to describe their story in a more cognitively complex way.

Status   Of all the function words, the relative use of first person singular pronouns is a particularly robust marker of the status of two people in an interaction. Within dyads, we have found that the person whose use of "I" words is lower tends to be the higher status participant. In the analysis of the incoming and outgoing emails of 11 undergraduates, graduate students, and faculty, the rated status of the interactant was correlated −.40 with the relative use of "I" words (Pennebaker & Davis, 2006).

Similarly, our analyses of the Watergate tapes involving dyadic interactions between President Nixon and H.R. Haldeman, John Erlichmann, and John Dean indicated that Nixon had very different relationships with the three men. In their conversations, Nixon's use of first person singular was significantly lower when talking to Erlichman (Nixon = 3.0%, Erlichman = 5.7%) and Dean (3.9 vs. 5.3) than in his interactions with Haldeman (5.1 vs. 5.0). Indeed, John Dean (personal communication, August 30, 2002) noted that Nixon and Haldeman were true partners in running the White House – although they were not close personal friends. Dean's relationship with Nixon was formal and respectful. Interestingly, Dean characterized Erlichman as arrogant yet insecure and was often "over his head" with respect to Washington politics. In listening to the Watergate tapes himself, Dean was impressed with the degree to which Erlichman was making a power play in the hopes of getting Haldeman's job. In his interactions with Nixon, Erlichman was overly solicitous, almost groveling. Nixon's reaction was that of even greater psychological distance than with Dean, with whom he had a more formal distant relationship. The analysis of "I" words, then, can help to uncover the subtle differences in relationships among historical figures.

Demographics: Sex and Age   There are sex differences in the use of virtually all function words: pronouns, prepositions, articles, and auxiliary verbs. In a study of over 10,000 text files, Newman et al. (2003) found that females tend to use first person singular pronouns at a consistently higher rate than do males.

Possible reasons for this difference could be that females are generally more self-focused than men, are more prone to depression than men, or that women have traditionally held lower status positions relative to men. Another large sex difference is that males' natural speech and writing contain higher rates of article and noun use, which characterizes categorization, or concrete thinking. On the other hand, females use more verbs (especially auxiliary verbs), which highlights females' relational orientations.

Age differences in function words are also robust. Pennebaker and Stone (2003) found that people use fewer first person singular words and greater first person plural words with age. 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, the analysis of their auxiliary verbs indicates that people use more future tense and less past tense the older they get, suggesting a shift in focus through the aging process.

Culture   Along with the stereotypes that "we" and "us" represent strong social bonds, one might surmise that the pronoun "we" would be more common in collectivist cultures, and the pronoun "I" more frequent in individualistic cultures. Investigating these very questions, we have compared translations of Japanese newspapers, poems, and novels to comparable American texts. Judges' ratings of the first person plural pronouns showed that both countries used first person plural pronouns in a close, personal way at the same rates. However, American authors used first person plural pronouns in a distant, royal-we way at double the rate that was found in the Japanese texts. This accounted for the overall greater rate of first person plural pronouns in American than in Japanese texts. Also counter to stereotypes, the Japanese texts used first person singular pronouns at a higher rate than did American texts. Indeed, American texts were higher in their use of first person plural pronouns (Chung & Pennebaker, 2005).

What could account for these counterstereotypical findings? Recall that the work reviewed in this chapter found that, overall, "I" use reflects self-focus. Given that focus on the self is required to achieve collectivistic values such as harmony, empathy, and self-criticism to please the ingroup (e.g. Kanagawa, Cross, & Markus, 2001; Markus & Kitayama, 1991), this finding is perhaps not so surprising. Similarly, the use of "we" has been shown to engender feelings of closeness, similarity, and of sharing a common fate with another more than the use of "Other and I" (Fitzsimmons & Kay, 2004), "they", or "it" (Brewer & Gardner, 1996). In a hierarchically modeled social system as in Japan, it would be rather insulting or debasing to imply that one is closer, similar, and shares a common fate with one's superior or subordinate. In these cases, grammatical constructions such as "other and I" would be more appropriate than using "we". However, the presumptive, distant, royal-we would more frequently be used where sharp distinctions in social status are not as salient. These data support this.

The phenomenon of pronoun-drop in some languages suggests that speakers from these cultures may be more collectivistic in their thinking (Kashima & Kashima, 1998; see also Chapters 2 and 4 in the present volume). However,

comparisons in a common language (including the use of translations) point to how pronouns are more than just ostensible markers of self-focus and collective-focus; pronoun use across cultures can point to other cultural values such as uncertainty avoidance (Kashima & Kashima, 2005), and convey status similarities and differences. Indeed, in several languages of high-power distance cultures, it is not even possible to use a pronoun without first having established the relative social status between speaker and addressee. Comparisons in a common language suggest that these differences in cultural patterns in status are maintained, to some degree, in translations.

Cultural researchers have also been concerned with the nature of thinking across cultures. Peng and Nisbett (1999) argue that Western thought from the time of the early Greeks has been highly categorical. Categorization is an essential process by which we are able to generalize or to reason "beyond the information given" (Bruner, 1973). Having categories allows us to think about the world in an ordered way, and to make inferences regarding a particular class of objects, ideas, or events based on category membership. Of course, East Asians also naturally categorize, but Peng and Nisbett argue that Eastern thinking and philosophy are less guided by categorization and more by movement and process.

Function words that indicate categorization include articles (a, an, the) which are used with nouns. In our own work, we are finding that translations of Japanese texts have significantly fewer articles and nouns than comparable American works (Chung & Pennebaker, 2005). These findings provide linguistic evidence for the Eastern and Western ways of thinking found in social cognitive tasks (Nisbett, 2003). These cross-cultural comparisons using translations provide convergent evidence for structural differences existing in the English language and some Asian languages (e.g. Japanese and Korean). Further research examining why linguistic differences emerge in translations may yield valuable insights into their respective cultures.

CONCLUSIONS

Our findings to date suggest that the words we use in natural language reflect our thoughts and feelings in often unpredictable ways. They also reveal a tremendous amount of information about our social interactions and personality. Function words, in particular, carry an array of psychological meanings and set the tone for social interactions. Before discussing the possible implications of these findings, two important concerns must be addressed.

How can we say that the various effects that we have discussed reflect function word differences and not differences in content or context? Perhaps these effects are merely reflections of differences in syntax – some people simply put sentences together in different ways. We placidly concede that the content and context of language use may vary across levels of stress, age, culture, or honesty.

However, it is important to consider that linguistic content and the contexts in which people speak are not randomly assigned. Humans choose where to talk and write and what to talk or write about. That function words and not traditional

content words consistently vary as a function of psychological state is important by itself. We can begin to measure these words in order to get rough proxies of people's psychological worlds.

Do function words reflect or influence psychological state? A related issue surrounds the causal links between the use of function words and psychological state. Are function words merely reflecting the cognitive architecture of the speaker or is it possible that the ways people use words affect their thinking styles? In all likelihood, function words are mere reflections of underlying cognitive activity. We have conducted multiple unsuccessful studies where we have induced people to use pronouns (e.g., I versus we) in an attempt to make them feel more or less group-oriented. We have also attempted to change the ways people write about emotional upheavals by altering their use of pronouns. Forcing people to talk or write differently has not affected any of our markers of cognitive or other psychological functioning. In short, our work is supporting the cognitive reflection model rather than a more Whorfian causal model.

Implications for Social Psychology

Social psychologists all know that self-reports suffer from multiple shortcomings. Surveys are susceptible to an assortment of response biases that question the validity of these measures. What people say about themselves often reflects their self-theories rather than serve as objective markers of their true thoughts and feelings. Despite the awareness of these problems, researchers remain seduced by their most attractive features; self-reports are cheap, fast, and easy.

Because of these problems, there has also been a push toward more naturalistic and non-obtrusive assessment methods. Language and content analysis has been one alternative. Previous studies have laid the groundwork for understanding how key content words relate to social and cognitive processes. Researchers have interpreted these key content words in their respective contexts. However, this work has required painstaking, laborious coding efforts, thereby restricting both the size and number of linguistic samples in any given study.

Computerized language analyses have brought us to a new frontier in social psychology. We are now able to examine and assess natural language free from the bounds of sampling, coding, and cost, and safe from the pitfalls of self-reports. Computerized tools provide efficient and reliable measurement beyond even the most conscientious of human coders. Instead of focusing on the specific meaning of words in a narrow context, we can widen our lens to the subtle patterns in language that have profound social effects.

Language has evolved to be one of the most effective means by which we communicate our past and current thoughts and feelings. New nouns, verbs, and adjectives (e.g. iPod, googled, cool) are added to our vocabulary with new inventions, fads, or roles, but our function words have remained the same. Until recent computerized linguistic analyses, very few social psychologists ever attended to these words. What we can learn from function words is not to be glossed over as easily as they are in written or spoken language. With the right tools, we now know that function words have real and important social psychological functions.

Streams of text are available wherever natural language occurs: on the Internet, in books, diaries, musical lyrics, during natural conversations, shows, press conferences, court trials, or therapy sessions. With computerized linguistic analyses, we can examine talk in real-time, or analyze words from any historical record. Indeed, several of our analyses have enabled us to examine the psychology of historical figures. From the presumed Word of God (e.g., the Bible, the Koran), the inaugural speeches of our nation's presidents, or ancestral diaries, we are able to know the influential writers or speakers of our past. Serendipitously, we can also start to answer the burning social psychological questions we have in our everyday lives. We can gain access into how our online dating prospects view us, distinguish which rap artists are honest about being true gangsters, diagnose if our therapists are just as depressed as we are, or expose which of our colleagues secretly think they are higher in status than us. What linguistic analyses are telling us is that, in all likelihood, an answer will lie in their use of function words.

ACKNOWLEDGMENTS

REFERENCES

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

Baayen, R. H., Piepenbrock, R., & Bulickers, L. (1995). *The CELEX Lexical Database [CD ROM]*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conception of time. *Cognitive Psychology*, *43*, 1–22.

Brewer, M. B., & Gardner, W. (1996). Who is this "We"? Levels of collective identity and self representations. *Journal of Personality and Social Psychology*, *71*, 83–93.

Bruner, J. S. (1973). *Beyond the information given: Studies in the psychology of knowing*. Oxford: W. W. Norton.

Campbell, R. S., & Pennebaker, J. W. (2003). The secret life of pronouns: Flexibility in writing style and physical health. *Psychological Science*, *14*, 60–65.

Chung, C. K., & Pennebaker, J. W. (2005). The language of East and West: Distinguishing cognitive, emotional, and social processes between Japan and the US through word use. Unpublished data.

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, *15*, 687–693.

Damasio, A. R. (1995). *Descartes' error: Emotion, reason and the human brain*. New York: Harper Collins.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*, 297–327.

Fiedler, K., & Semin, G. R. (1992). Attribution and language as a socio-cognitive environment. In G. R. Semin, and K. Fiedler (Eds.), *Language, interaction, and social cognition*, pp. 58–78. Thousand Oaks, CA: Sage Publications, Inc.

Fitzsimmons, G. M., & Kay, A. C. (2004). Language and interpersonal cognition: Causal effects of variations in pronoun usage on perceptions of closeness. *Personality and Social Psychology Bulletin*, *30*, 547–557,

Foltz, P. W. (1996). Latent semantic analysis for text-based research. *Behavior Research Methods, Instruments and Computers*, *28*, 197–202.

Gazzaniga, M. S. (2005). *The ethical brain*. New York: Dana Press.

Gortner, E. M., & Pennebaker, J. W. (2003). The archival anatomy of a disaster: Media coverage and community-wide health effects of the Texas A&M Bonfire Tragedy. *Journal of Social and Clinical Psychology*, *22*, 580–603.

Gottschalk, L. A. (1997). The unobtrusive measurement of psychological states and traits. In C. W. Roberts (Ed.) *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts* (pp. 117–129). Mahwah, NJ: Erlbaum.

Graham, L. E. II, Scherwitz, L., & Brand, R. (1989). Self reference and coronary heart disease incidence n the Western Collaborative Group Study. *Psychosomatic Medicine*, *51*, 137–144.

Hajek, C., & Giles, H. (2003). New directions in intercultural communication competence. In J. O. Greene and B. R. Burleson (Eds.), *Handbook of communication and social interaction skills* (pp.935–957). Mahwah, NJ: Lawrence Erlbaum.

Hart, R. P., Jarvis, S. E., Jennings, W. P., & Smith-Howell, D. (2005). *Political keywords: Using language that uses us*. New York: Oxford University Press.

Kanagawa, C., Cross, S. E., & Markus, H. R. (2001). "Who am I?" The cultural psychology of the conceptual self. *Personality and Social Psychology Bulletin*, *27*, 90–103.

Kashima, E. S., & Kashima, Y. (1998). Culture and language: The case of cultural dimensions and personal pronoun use. *Journal of Cross-Cultural Psychology*, *29*, 461–486.

Kashima, E. S., & Kashima, Y. (2005). Erratum to Kashima and Kashima (1998) and reiteration. *Journal of Cross-Cultural Psychology*, *36*, 396–400.

Lepore, S. J., & Smyth, J. M. (2002). *The writing cure: How expressive writing promotes health and emotional well-being*. Washington, DC: American Psychological Association.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*, 224–253.

McAdams, D. P. (2001). The psychology of life stories. *Review of General Psychology*, *5*, 100–122.

Mehl, M. R. (2004). The sounds of social life: Exploring students' daily social environments and natural conversations. Unpublished Doctoral Dissertation.

Mehl, M. R., & Pennebaker, J. W. (2003). The social dynamics of a cultural upheaval: Social interactions surrounding September 11, 2001. *Psychological Science*, *14*, 579–585.

Miller, G. A. (1995). *The science of words*. New York: Scientific American Library.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic style. *Personality and Social Psychology Bulletin*, *29*, 665–675.

Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently*. New York, NY: Free Press.

Peng, K., & Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. *American Psychologist*, *54*, 741–754.

Pennebaker, J. W., & Davis, M. (2006). Pronoun use and dominance. Unpublished data. Department of Psychology, University of Texas at Austin, Austin, TX.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count (LIWC): LIWC2001*. Mahwah: Lawrence Erlbaum.

Pennebaker, J. W., Groom, C. J., Loew, D., & Dabbs, J. M. (2004). Testosterone as a social inhibitor: Two case studies of the effect of testosterone treatment on language. *Journal of Abnormal Psychology*, *113*, 172–175.

Pennebaker, J. W., Kiecolt-Glaser, J., & Glaser, R. (1988). Disclosure of traumas and immune function: Health implications for psychotherapy. *Journal of Consulting and Clinical Psychology*, *56*, 239–245.

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

Pennebaker, J. W., & Lay, T. C. (2002). Language use and personality during crises: Analyses of Mayor Rudolph Giuliani's press conferences. *Journal of Research in Personality*, *36*, 271–282.

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

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

Rochon, E., Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (2000). Quantitative analysis of aphasic sentence production: Further development and new data. *Brain and Language*, *72*, 193–218.

Rude, S. S., Gortner, E. M., & Pennebaker, J. W. (2004). Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, *18*, 1121–1133.

Scherwitz, L., Berton, K., & Leventhal, H. (1978). Type A behavior, self-involvement, and cardiovascular response. *Psychosomatic Medicine*, *40*, 593–609.

Schultheiss, O. C., & Brunstein, J. C. (2001). Assessment of implicit motives with a research version of the TAT: Picture profiles, gender differences, and relations to other personality measures. *Journal of Personality Assessment*, *77*, 71–86.

Semin, G. R., Rubini, M., & Fiedler, K. (1995). The answer is in the question: The effect of verb causality on the locus of explanation. *Personality and Social Psychology Bulletin*, *21*, 834–841.

Stirman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and non-suicidal poets. *Psychosomatic Medicine*, *63*, 517–522.

Stone, L. D., & Pennebaker, J. W. (2002). Trauma in real time: Talking and avoiding online conversations about the death of Princess Diana. *Basic and Applied Social Psychology*, *24*, 172–182.

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

Tannen, D. (1993). *Framing in discourse*. London: Oxford University Press.

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory and Cognition*, *19*, 95–112.

Weber-Fox, C., & Neville, H. J. (2001). Sensitive periods differentiate processing of open- and closed-class words: An event-related brain potential study of bilinguals. *Journal of Speech, Language, and Hearing Research*, *44*, 1338–1353.

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

Winter, D. G., & McClelland, D. C. (1978). Thematic analysis: An empirically derived measure of the effects of liberal arts education. *Journal of Educational Psychology*, *70*, 8–16.