The Development and Psychometric Properties of LIWC2007

James W. Pennebaker, Cindy K. Chung, Molly Ireland,
Amy Gonzales, and Roger J. Booth

The University of Texas at Austin
and
The University of Auckland, New Zealand

This article is published by LIWC.net, Austin, Texas 78703 USA in conjunctin with the LIWC2007 software program. Correspondence should be sent to Pennebaker@mail.utexas.edu.

Contents

The Development and Psychometric Properties of LIWC2007 3

The LIWC2007 Framework..... 3

The LIWC2007 Main Text Processing Module..... 3

The Default LIWC2007 Dictionary 4

LIWC2007 Dictionary Development..... 7

Internal Reliability and External Validity..... 8

Base Rates of Word Usage 9

Comparing LIWC2007 with
LIWC2001……13

Helpful References..... 14

The Development and Psychometric Properties of LIWC2007

The ways that individuals talk and write provide windows into their emotional and cognitive worlds. Over the last four decades, researchers have provided evidence to suggest that people's physical and mental health are correlated with the words they use (Gottschalk & Glaser, 1969; Rosenberg & Tucker, 1978; Stiles, 1992). More recently, a large number of studies have found that having individuals write or talk about deeply emotional experiences is associated with improvements in mental and physical health (e.g., Fratteroli, 2007; Lepore & Smyth, 2002; Pennebaker, 1997). Text analyses based on these studies indicate that those individuals who benefit the most from writing tend to use relatively high rates of positive emotion words, a moderate number of negative emotion words, and an increasing number of cognitive words, and switch their use of pronouns from writing session to writing session (e.g., Campbell & Pennebaker, 2002; Pennebaker, Mayne, & Francis, 1997).

In order to provide an efficient and effective method for studying the various emotional, cognitive, and structural components present in individuals' verbal and written speech samples, we originally developed a text analysis application called Linguistic Inquiry and Word Count, or LIWC. The first LIWC application was developed as part of an exploratory study of language and disclosure (Francis, 1993; Pennebaker, 1993). The second version, LIWC2001, updated the original application with an expanded dictionary and a more modern software design (Pennebaker, Francis, & Booth, 2001). The most recent evolution, LIWC2007, has significantly altered both the dictionary and the software options. As with previous versions, however, the program is designed to analyze individual or multiple language files quickly and efficiently. At the same time, the program attempts to be transparent and flexible in its operation, allowing the user to explore word use in multiple ways.

The LIWC2007 Framework

The LIWC2007 application relies on an internal default dictionary that defines which words should be counted in the target text files. Note that the LIWC2007.EXE file is an executable file and cannot be read or opened. To avoid confusion in the subsequent discussion, text words that are read and analyzed by LIWC2007 are referred to as ***target words***. Words in the LIWC2007 dictionary file will be referred to as ***dictionary words***. Groups of dictionary words that tap a particular domain (e.g., negative emotion words) are variously referred to as subdictionaries or word categories.

The LIWC2007 Main Text Processing Module

LIWC2007 is designed to accept written or transcribed verbal text which has been stored as a digital file in one of multiple formats, including raw text, ASCII, unicode, or standard files from Microsoft's Word files. LIWC2007 accesses a single file or group of files and analyses each sequentially, writing the output to a single file. Processing time for a page of single-spaced text is typically a fraction of a second on both PC and Mac computers. LIWC2007 reads each designated text file, one target word at a time. As each target word is processed, the dictionary file is searched, looking for a dictionary match with the current target word. If the target word matches the dictionary word, the appropriate word category scale (or scales) for that word is incremented. As the target text file is being processed, counts for various structural composition elements (e.g., word count and sentence punctuation) are also incremented.

With each text file, approximately 80 output variables are written as one line of data to a designated output file. This data record includes the file name, 4 general descriptor categories (total word count, words per sentence, percentage of words captured by the dictionary, and percent of words longer than six letters), 22 standard linguistic dimensions (e.g., percentage of words in the text that are pronouns, articles, auxiliary verbs, etc.), 32 word categories tapping psychological constructs (e.g., affect, cognition, biological processes), 7 personal concern categories (e.g., work, home, leisure activities), 3 paralinguistic dimensions (assents, fillers, nonfluencies), and 12 punctuation categories (periods, commas, etc). A complete list of the standard LIWC2007 scales is included in Table 1.

The Default LIWC2007 Dictionary

The LIWC2007 Dictionary is the heart of the text analysis strategy. The default LIWC2007 Dictionary is composed of almost 4,500 words and word stems. Each word or word stem defines one or more word categories or subdictionaries. For example, the word *cried* is part of five word categories: sadness, negative emotion, overall affect, verb, and past tense verb. Hence, if it is found in the target text, each of these five subdictionary scale scores will be incremented. As in this example, many of the LIWC2007 categories are arranged hierarchically. All anger words, by definition, will be categorized as negative emotion and overall emotion words. Note too that word stems can be captured by the LIWC2007 system. For example, the LIWC2007 Dictionary includes the stem *hungr** which allows for any target word that matches the first five letters to be counted as an ingestion word (including hungry, hungrier, hungriest). The asterisk, then, denotes the acceptance of all letters, hyphens, or numbers following its appearance.

Each of the default LIWC2007 categories is composed of a list of dictionary words that define that scale. Table 1 provides a comprehensive list of the default LIWC2007 dictionary categories, scales, sample scale words, and relevant scale word counts.

Table 1. LIWC2007 Output Variable Information

Category	Abbrev	Examples	Words in category	Validity (judges)	Alpha: Binary/raw
Linguistic Processes					
Word count	wc				
words/sentence	wps				
Dictionary words	dic				
Words>6 letters	sixltr				
Total function words	funct		464		.97/.40
Total pronouns	pronoun	I, them, itself	116		.91/.38
Personal pronouns	ppron	I, them, her	70		.88/.20
1st pers singular	i	I, me, mine	12	.52	.62/.44
1st pers plural	we	We, us, our	12		.66/.47
2nd person	you	You, your, thou	20		.73/.34
3rd pers singular	shehe	She, her, him	17		.75/.52
3rd pers plural	they	They, their, they'd	10		.50/.36
Impersonal pronouns	ipron	It, it's, those	46		.78/.46
Articles	article	A, an, the	3		.14/.14
[Common verbs][a]	verb	Walk, went, see	383		.97/.42
Auxiliary verbs	auxverb	Am, will, have	144		.91/.23
Past tense [a]	past	Went, ran, had	145	.79	.94/.75
Present tense [a]	present	Is, does, hear	169		.91/.74
Future tense [a]	future	Will, gonna	48		.75/.02
Adverbs	adverb	Very, really, quickly	69		.84/.48
Prepositions	prep	To, with, above	60		.88/.35
Conjunctions	conj	And, but, whereas	28		.70/.21
Negations	negate	No, not, never	57		.80/.28
Quantifiers	quant	Few, many, much	89		.88/.12
Numbers	number	Second, thousand	34		.87/.61
Swear words	swear	Damn, piss, fuck	53		.65/.48
Psychological Processes					
Social processes[b]	social	Mate, talk, they, child	455		.97/.59
Family	family	Daughter, husband, aunt	64	.87	.81/.65
Friends	friend	Buddy, friend, neighbor	37	.70	.53/.12
Humans	human	Adult, baby, boy	61		.86/.26
Affective processes	affect	Happy, cried, abandon	915		.97/.36
Positive emotion	posemo	Love, nice, sweet	406	.41	.97/.40
Negative emotion	negemo	Hurt, ugly, nasty	499	.31	.97/.61
Anxiety	anx	Worried, fearful, nervous	91	.38	.89/.33
Anger	anger	Hate, kill, annoyed	184	.22	.92/.55
Sadness	sad	Crying, grief, sad	101	.07	.91/.45
Cognitive processes	cogmech	cause, know, ought	730		.97/.37
Insight	insight	think, know, consider	195		.94/.51
Causation	cause	because, effect, hence	108	.44	.88/.26
Discrepancy	discrep	should, would, could	76	.21	.80/.28
Tentative	tentat	maybe, perhaps, guess	155		.87/.13
Certainty	certain	always, never	83		.85/.29
Inhibition	inhib	block, constrain, stop	111		.91/.20
Inclusive	incl	And, with, include	18		.66/.32

Category	Abbrev	Examples	Words in category	Validity (judges)	Alpha: Binary/raw
Exclusive	excl	But, without, exclude	17		.67/.47
Perceptual processes[c]	percept	Observing, heard, feeling	273		.96/.43
See	see	View, saw, seen	72		.90/.43
Hear	hear	Listen, hearing	51		.89/.37
Feel	feel	Feels, touch	75		.88/.26
Biological processes	bio	Eat, blood, pain	567	.53	.95/.53
Body	body	Cheek, hands, spit	180		.93/.45
Health	health	Clinic, flu, pill	236		.85/.38
Sexual	sexual	Horny, love, incest	96		.69/.34
Ingestion	ingest	Dish, eat, pizza	111		.86/.68
Relativity	relativ	Area, bend, exit, stop	638		.98/.51
Motion	motion	Arrive, car, go	168		.96/.41
Space	space	Down, in, thin	220		.96/.44
Time	time	End, until, season	239		.94/.58
Personal Concerns					
Work	work	Job, majors, xerox	327		.91/.69
Achievement	achieve	Earn, hero, win	186		.93/.37
Leisure	leisure	Cook, chat, movie	229		.88/.50
Home	home	Apartment, kitchen, family	93		.81/.57
Money	money	Audit, cash, owe	173		.90/.53
Religion	relig	Altar, church, mosque	159		.91/.53
Death	death	Bury, coffin, kill	62		.86/.40
Spoken categories					
Assent	assent	Agree, OK, yes	30		.59/.41
Nonfluencies	nonflu	Er, hm, umm	8		.28/.23
Fillers	filler	Blah, Imean, youknow	9		.63/.18

"Words in category" refers to the number of different dictionary words that make up the variable category; "Validity judges" reflect the simple correlations between judges' ratings of the category with the LIWC variable (from Pennebaker & Francis, 1996). "Alphas" refer to the Cronbach alphas for the internal reliability of the specific words within each category. The binary alphas are computed on the occurrence/non-occurrence of each dictionary word whereas the raw or uncorrected alphas are based on the percentage of use of each of the category words within the texts. All alphas were computed on a sample of 2800 randomly selected text files from our language corpus.

The LIWC dictionary generally arranges categories hierachically. For example, all pronouns are included in the overarching category of function words. The category of pronouns is the sum of personal and impersonal pronouns. There are some exceptions to the hierarchy rules:

[a] Common verbs are not included in the function word category. Similarly, common verbs (as opposed to auxiliary verbs) that are tagged by verb tense are included in the past, present, and future tense categories but not in the overall function word categories.

[b] Social processes include a large group of words (originally used in LIWC2001) that denote social processes, including all non-first-person-singular personal pronouns as well as verbs that suggest human interaction (talking, sharing).

[c] Perceptual processes include the entire dictionary of the Qualia category (which is a separate dictionary), which includes multiple sensory and perceptual dimensions associated with the five senses.

LIWC2007 Dictionary Development.

The selection of words defining the LIWC2007 categories involved multiple steps over several years. The initial idea was to identify a group of words that tapped basic emotional and cognitive dimensions often studied in social, health, and personality psychology. With time, the domain of word categories expanded considerably.

Step 1. Word Collection. In the design and development of the LIWC category scales, sets of words were first generated for each category scale. Within the Psychological Processes category, for example, the emotion or affective subdictionaries were based on words from several sources. We drew on common emotion rating scales, such as the PANAS (Watson, Clark, & Tellegen, 1988), Roget's Thesaurus, and standard English dictionaries. Following the creation of preliminary category word lists, brain-storming sessions among 3-6 judges were held in which words relevant to the various scales were generated and added to the initial scale lists. Similar schemes were used for the other subjective dictionary categories.

Step 2. Judges' Rating Phases. Once the broad word lists were amassed, words in the Psychological Processes and Personal Concerns and most in the Relativity (excluding verb tense) categories were then rated by three independent judges. In the development of the first LIWC program, the judges were instructed to focus on both the inclusion and exclusion of words in each LIWC dictionary scale list. In the first rating phase, the judges indicated whether each word in the category list should or should not be included on the particular category in question. They were also instructed to include additional words they felt should be included in the category. All category word lists were updated by the following set of rules: 1) a word remained in the category list if two out of three judges agreed it should be included, 2) a word was deleted from the category list if at least two of the three judges agreed it should be excluded, and 3) a word was added to the category list if two out of three judges agreed it should be included. Due to the objective nature of elements in the Standard Language Dimensions category (e.g., articles, pronouns, prepositions), judges' ratings were not collected for the various lists in that category.

The second rating phase involved the discrimination of LIWC category word elements. Judges were given category level alphabetized word lists (e.g., all Cognitive Process words) and asked to indicate whether each word in the list should or should not be included in the high-level category in question. Judges were then instructed to indicate in which, if any, of the mid-level scale lists the word should be included (e.g., Insight, Causation). All category scale word lists were updated by the following rules: 1) a word remained on the scale list if two out of three judges agreed it should be included and 2) a word was deleted from the scale list if at least two of the three judges agreed it should be excluded. The final percentages of judges' agreement for this second rating phase ranged from 93% agreement for Insight to 100% agreement for Ingestion, Death, Religion, Friends, Relatives, and Humans.

Step 3. Psychometric Evalutation. The initial LIWC judging took place in 1992-1994. A significant LIWC revision was undertaken in 1997 to streamline the original program and dictionaries. Text files from several dozen studies, totaling over 8 million words were analyzed using the 1997 version of LIWC as well as WordSmith, a powerful word count program used in discourse analysis. Original LIWC categories that were used at very low rates (less than 0.3 percent of words made up the category) or that suffered from consistently poor reliability or

validity were omitted. Several new categories, including social processes, several personal concern categories, and the relativity dimensions, were added following the same stringent judge-based procedures described above (including both passes). Finally, once the entire new LIWC dictionary was assembled, any words that were not used at least 0.005 percent of the time in our previous text files or were not listed in Francis and Kucera's (1982) *Frequency Analysis of English Usage* were excluded.

Step 4. Updates and Expansions. The most recent version, LIWC2007, involved substantial updating of the dictionaries and modification in the dictionary structure. Drawing on over several hundred thousand text files made up of several hundred million words from both written and spoken language samples, we sought to identify common words and word categories not captured in the earlier LIWC versions. Examining the 2000 most frequently used words, a group of four judges individually and collectively agreed which new words and new word categories were appropriate for inclusion. Based on recent studies suggesting that function words are particularly relevant to psychological processes, we added the categories of Conjunctions, Adverbs, Quantifiers, Auxiliary Verbs, Commonly-used Verbs, Impersonal Pronouns, Total Function Words, and Total Relativity Words. In addition, third person pronouns were divided into 3rd person singular and 3rd person plural. Finally, a large group of punctuation marks have been added as separate categories.

For those who are familiar with LIWC2001, it will be clear that some of the original categories have been removed – primarily because these categories had consistently low base rates and were rarely used: Optimism, Positive Feelings, Communication Verbs, Other References, Metaphysical, Sleeping, Grooming, School, Sports, Television, Up, and Down. The category of Unique Words (also known as Type/Token ratio) has also been removed. This category typically correlates with word count at -.80. Note that an alternative default LIWC2001 dictionary is available.

LIWC2007: Internal Reliability and External Validity

Assessing the reliability and validity of text analysis programs is a tricky business. On the surface, one would think that you could determine the internal reliability of a LIWC scale the same way it is done with a questionnaire. With a questionnaire that taps anger or aggression, for example, participants complete a self-report asking a number of questions about their feelings or behaviors related to anger. Reliability coefficients are computed by correlating people's answers to the various questions. The more highly they correlate, the reasoning goes, the more the questions all measure the same thing. Voila! The scale is deemed internally consistent.

A similar strategy can be used with words. The LIWC Anger scale, for example, is made up of 184 anger-related words. In theory, the more people use one type of anger word in a given text, the more likely they should be to use other anger words in the same text. To test this idea, we can determine the degree to which people use each of the 184 anger words across a select group of text files and then calculate the intercorrelations of the word use. Indeed, in Table 1, we include these internal reliability statistics, including those of Anger where the alpha reliability

ranges between .92 (binary method) and .55 (uncorrected) depending on how it is computed. The internal reliability statistics are based on the correlation between the occurrence of each word in a category with the sum of the other words in the same category. The binary method converts the usage of each of the single words within a given text into either a 0 (not used) or a 1 (used one or more times). The uncorrected method is based on the percentage of total words that each of the category words are used. The binary method has the potential to overestimate reliability based on the length of texts; the uncorrected method tends to underestimate reliability based on the highly variable base rates of word usage within any given category.

But be warned: the psychometrics of natural language use are not as pretty as with questionnaires. The reason is obvious once you think about it. Once you say something, you generally don't need to say it again in the same paragraph or essay. The nature of discourse, then, is we usually say something and then move on to the next topic. Saying the same thing over and over again is generally bad form.

Issues of validity are also a bit tricky. We can have people complete a questionnaire that assesses their general moods and then have them write an essay which we then subject to the LIWC program. We can also have judges evaluate the essay for its emotional content. In other words, we can get self-reported, judged, and LIWC numbers that all reflect a participant's anger.

One of the first tests of the validity of the LIWC scales was undertaken by Pennebaker and Francis (1996) as part of an experiment in which first year college students wrote about the experience of coming to college. During the writing phase of the study, 72 Introductory Psychology students met as a group on three consecutive days to write on their assigned topics. Participants in the experimental condition (n = 35) were instructed to write about their deepest thoughts and feelings concerning the experience of coming to college. Those in the control condition (n = 37) were asked to describe any particular object or event of their choosing in an unemotional way. After the writing phase of the study was completed, four judges rated the participants' essays on various emotional, cognitive, content, and composition dimensions designed to correspond to selected LIWC Dictionary scales.

Using LIWC output and judges' ratings, Pearson correlational analyses were performed to test LIWC's external validity. Results, presented in Table 1, reveal that the LIWC scales and judges' ratings are highly correlated. These findings suggest that LIWC successfully measures positive and negative emotions, a number of cognitive strategies, several types of thematic content, and various language composition elements. The level of agreement between judges' ratings and LIWC's objective word count strategy provides support for LIWC's external validity.

Base Rates of Word Usage

In evaluating any text analysis program, it is helpful to get a sense of the degree to which language varies across settings. Since 1986, we have been collecting text samples from a variety of studies – both from our own lab as well as from 28 others in the United States, Canada, and New Zealand. For purposes of comparison, six classes of text from 72 separate studies were analyzed and compared. As can be seen in Table 2, these analyses reflect the utterances of over 24,000

writers or speakers totaling over 168 million words. Overall, 29 samples are based on experiments were people were randomly assigned to write either about deeply emotional topics (emotional writing) or about relatively trivial topics such as plans for the day (control writing). Individuals from all walks of life – ranging from college students to psychiatric prisoners to elderly and even elementary-aged individuals – are represented in these studies. A third class of text was based on 113 highly technical articles in the journal Science published in 1997 or 2007. A fourth sample included 714,000 internet web logs, or blogs, from approximately 20,000 individuals who posted either on Blog.com in 2004 or LiveJournal.com in the summer and fall of 2001. The fifth sample was based 209 novels published in English between 1700 and 2004. The American and British novels included best-selling popular books as well as more classic novels. Finally, we analyzed data from seven observational studies in which participants were tape-recorded while engaging in conversations with others. The speech samples ranged from transcripts of people wearing audio recorders over days or weeks, strangers interacting in a waiting room, to couples talking about problems, to open-air tape recordings of people in public spaces.

Table 2. Summary Information for LIWC2007 Statistics

	Emotional writing	Control writing	Science Articles	Blogs	Novels	Talking
Total files	2,931	2,431	113	714,028	209	2,014
Total authors	1,014	841	113	20,146	209	850
Total words	1,299,400	985,698	305,552	149,924,828	14,637,011	1,202,015
Total studies	29	29	1	2	1	10
Total labs	11	11	1	2	1	3

Emotional writing studies require participants to write about their emotions and thoughts about personally relevant topics; Control Writing involves writing about non-emotional topics, such as plans for the day or descriptions of ordinary objects or events; Science articles are published articles in the journal Science in 1997 and 2007. Blogs are from LiveJournal.com which were written in summer and fall, 2001 and from Blogs.com that were downloaded in summer, 2004. Novels refers to either portions or complete works of American and British fiction published between 1800 and 2005; Talking files come from transcripts collected from individuals who are talking in real world unstructured settings.

As can be seen in Table 3, the LIWC2007 version captures, on average, over 86 percent of the words people use in writing and speech. Note that except for total word count and words per sentence, all means in Table 3 are expressed as percentage of total word use in any given speech/text sample. Simple one-way ANOVAs indicated that word usage was significantly different across the four settings for all of the word categories.

Table 3. LIWC2007 Output Variable Information

Category	Emotional writing	Control writing	Science Articles	Blogs	Novels	Talking	Grand Means	Mean SDs
Linguistic Processes								

Word count (mean)	443	405	2,704	7,304	70,033	596	13580	12203
Words/sentence	19.56	19.84	14.61	46.81	22.02	25.87	24.79	67.42
Dictionary words	93.42	88.55	53.66	83.83	83.57	91.49	82.42	4.92
Words>6 letters	13.27	13.87	29.55	14.12	16.33	9.43	16.10	3.71
Total function words	63.93	57.53	34.72	55.29	57.17	60.48	54.85	4.99
Total pronouns	20.23	14.29	3.18	16.07	14.89	21.52	15.03	3.30
Personal pronouns	14.23	10.78	0.82	10.67	10.29	13.63	10.07	2.87
Category	**Emotional writing**	**Control writing**	**Science Articles**	**Blogs**	**Novels**	**Talking**	**Grand Means**	**Mean SDs**
1st pers singular	10.40	8.50	0.12	6.42	2.55	6.30	5.72	2.48
1st pers plural	0.73	0.93	0.37	0.88	0.55	1.09	0.76	0.83
2nd person	0.39	0.20	0.00	1.23	1.29	3.94	1.18	0.93
3rd pers singular	2.01	0.73	0.04	1.48	4.92	1.46	1.77	1.33
3rd pers plural	0.71	0.41	0.28	0.65	0.98	0.84	0.65	0.57
Impersonal pronouns	6.00	3.51	2.36	5.40	4.61	7.89	4.96	1.56
Articles	4.97	6.63	7.67	5.89	8.21	4.42	6.30	1.95
Common verbs[a]	17.44	13.59	4.98	14.61	13.01	19.94	13.93	2.73
Auxiliary verbs	10.65	7.42	3.90	8.81	7.76	12.38	8.49	2.11
Past tense [a]	5.76	4.55	1.45	3.83	6.29	3.98	4.31	2.25
Present tense [a]	9.16	6.74	2.70	8.68	4.57	13.97	7.64	2.73
Future tense [a]	1.12	1.54	0.37	1.06	1.14	0.99	1.04	0.80
Adverbs	6.29	4.48	1.35	5.46	3.76	6.22	4.59	1.44
Prepositions	12.94	16.06	12.87	12.06	14.06	9.33	12.89	2.08
Conjunctions	7.39	7.71	4.30	6.39	6.65	5.67	6.35	1.64
Negations	2.24	0.84	0.40	1.78	1.69	2.92	1.65	0.95
Quantifiers	3.12	2.46	1.93	2.79	2.27	2.23	2.47	0.94
Numbers	1.31	2.73	7.05	1.96	1.17	1.95	2.70	1.60
Swear words	0.11	0.03	0.00	0.33	0.06	0.37	0.15	0.29
Psychological Processes								
Social processes[b]	9.09	5.55	2.61	8.65	12.26	11.75	8.32	2.93
Family	0.99	0.33	0.08	0.38	0.41	0.24	0.41	0.53
Friends	0.50	0.42	0.04	0.25	0.17	0.16	0.26	0.32
Humans	0.84	0.38	0.24	0.79	1.05	0.81	0.69	0.62
Affective processes	6.02	2.57	2.18	5.84	4.89	4.93	4.41	1.59
Positive emotion	3.28	1.83	1.33	3.72	2.86	3.42	2.74	1.27
Negative emotion	2.67	0.71	0.84	2.07	1.98	1.49	1.63	0.91
Anxiety	0.68	0.21	0.16	0.30	0.44	0.18	0.33	0.33
Anger	0.66	0.14	0.13	0.76	0.55	0.58	0.47	0.48
Sadness	0.63	0.14	0.29	0.42	0.57	0.19	0.37	0.37
Cognitive processes	19.66	14.42	11.28	15.97	15.23	15.66	15.37	2.85
Insight	3.25	1.31	1.82	2.17	1.99	2.34	2.15	1.05
Causation	1.85	1.28	2.16	1.42	1.02	1.55	1.55	0.84
Discrepancy	2.13	1.08	0.48	1.54	1.52	1.73	1.41	0.79
Tentative	2.93	2.31	1.33	2.65	2.16	2.36	2.29	1.05
Certainty	1.73	0.80	0.56	1.40	1.43	1.34	1.21	0.64
Inhibition	0.46	0.38	0.63	0.47	0.61	0.37	0.49	0.39
Inclusive	5.09	6.37	4.08	4.66	5.35	3.88	4.91	1.54
Exclusive	3.49	1.71	0.92	2.78	2.22	3.26	2.40	1.06

Perceptual processes[c]	2.08	1.91	1.15	2.27	3.28	2.27	2.16	1.16
See	0.53	0.83	0.65	0.87	1.26	0.99	0.86	0.79
Hear	0.44	0.35	0.06	0.65	1.15	0.69	0.56	0.47
Feel	0.96	0.62	0.24	0.60	0.74	0.48	0.61	0.50
Biological processes	1.95	2.97	1.02	2.05	2.13	1.52	1.94	1.44
Body	0.51	1.05	0.28	0.75	1.21	0.59	0.73	0.85
Health	0.93	0.49	0.57	0.54	0.44	0.31	0.55	0.65
Category	**Emotional writing**	**Control writing**	**Science Articles**	**Blogs**	**Novels**	**Talking**	**Grand Means**	**Mean SDs**
Sexual	0.34	0.05	0.06	0.41	0.18	0.32	0.23	0.39
Ingestion	0.26	1.44	0.15	0.44	0.36	0.37	0.50	0.65
Relativity	13.77	20.13	10.19	13.75	13.92	12.77	14.09	3.21
Motion	2.07	3.57	1.21	2.06	2.18	2.69	2.30	1.15
Space	5.38	7.92	6.08	5.61	6.83	5.46	6.21	1.82
Time	6.03	8.20	2.65	5.72	4.65	4.34	5.27	1.84
Current Concerns								
Work	2.14	3.74	1.74	1.71	1.01	1.67	2.00	1.40
Achievement	1.63	1.47	1.60	1.45	1.13	0.95	1.37	0.84
Leisure	0.78	1.86	0.41	1.60	0.69	1.04	1.06	0.84
Home	0.64	1.86	0.14	0.52	0.63	0.36	0.69	0.62
Money	0.34	0.56	0.36	0.59	0.51	0.60	0.49	0.54
Religion	0.17	0.17	0.06	0.34	0.39	0.19	0.22	0.45
Death	0.18	0.03	0.06	0.15	0.23	0.07	0.12	0.20
Spoken categories								
Assent	0.11	0.07	0.08	0.64	0.19	3.61	0.78	0.76
Nonfluencies	0.19	0.13	0.06	0.32	0.14	0.73	0.26	0.35
Fillers	0.03	0.01	0.00	0.02	0.00	1.20	0.21	0.35
Punctuation								
Total Punctuation	12.19	12.85	33.94	23.80	22.05	49.37	25.70	10.48
Periods	6.12	6.60	11.73	10.66	5.51	9.81	8.41	4.16
Commas	2.90	3.24	7.63	4.09	7.36	5.05	5.05	2.16
Colons	0.05	0.58	0.21	0.73	0.16	0.07	0.30	0.74
Semicolons	0.04	0.03	0.38	0.11	0.63	0.05	0.21	0.41
Question marks	0.17	0.04	0.05	0.60	0.57	2.33	0.63	1.03
Exclamation marks	0.12	0.07	0.00	1.27	0.46	0.21	0.36	0.70
Dashes	0.32	0.45	2.54	1.11	1.60	0.75	1.13	1.65
Quotation marks	0.27	0.21	0.18	0.71	3.39	0.17	0.82	0.82
Apostrophes	1.69	0.95	0.16	2.37	2.11	3.82	1.85	1.50
Parentheses	0.15	0.20	4.87	0.50	0.05	0.01	0.96	0.56
Other punctuation	0.20	0.29	1.32	1.08	0.14	27.11	5.02	4.87

Note: Grand Means are the unweighted means of the six genres; Mean SDs refer to the unweighted mean of the standard deviations across the six genre categories. The LIWC dictionary generally arranges categories hierachically. For example, all pronouns are included in the overarching category of function words. The cateory of pronouns is the sum of personal and impersonal pronouns. There are some exceptions to the hierarchy rules:
[a] Common verbs are not included in the function word category. Similarly, common verbs (as opposed to auxiliary verbs) that are tagged by verb tense are included in the past, present, and future tense categories but not in the overall function word categories.

[b] Social processes include a large group of words (originally used in LIWC2001) that denote social processes, including all non-first-person-singular personal pronouns as well as verbs that suggest human interaction (e.g., talking, sharing).
[c] Perceptual processes include the entire dictionary of the Qualia category (which is a separate dictionary), which includes multiple sensory and perceptual dimensions associated with the five senses.

In many ways, Table 3 points to the important role that context plays in people's use of language. Not surprisingly, the topics of writing – as reflected in the current concerns category – vary substantially as a function of genre. More striking, however, are the large differences in people's use of function words as well as punctuation from genre to genre (cf., Biber, 1988).

Comparing LIWC2007 with LIWC2001

For users of LIWC2001, a new edition of LIWC that uses a different dictionary can be an unsettling experience. Many of the older dictionaries have been slightly changed, a few have been substantially updated (e.g., exclusive words, cognitive mechanisms), and others have been removed or added. To help older users, we include Table 4 which lists the means, standard deviations, and correlations between the two dictionary versions. These analyses are based on a comparison of over 2800 randomly selected texts from each of the genres listed in Tables 3 and 4.

Table 4. Comparisons Between LIWC2007 and LIWC2001: Means, Standard Deviations, and Correlations

	LIWC2007		LIWC2001		
	mean	**sd**	**mean**	**sd**	**correlation**
Word count	1687.84	7697.27	1687.84	7697.27	1.00
Words per sentence	22.38	44.38	22.38	44.38	1.00
Dictionary words	86.31	10.13	75.32	10.64	0.97
Words>6 letters	13.26	4.56	13.26	4.56	1.00
Pronouns	12.14	4.09	14.16	4.52	0.97
1st person singular	7.82	3.68	7.78	3.67	1.00
1st person plural	0.78	0.90	0.78	0.90	1.00
2nd person	1.08	1.57	1.09	1.60	1.00
Articles	5.36	1.94	5.33	1.94	1.00
Past tense verbs	4.62	3.09	4.74	3.14	1.00
Present tense verbs	8.77	3.80	10.46	4.07	0.96
Future tense verbs	1.14	1.07	1.28	1.22	0.88
Prepositions	12.24	2.85	12.23	2.82	0.99
Negations	1.91	1.11	1.85	1.11	0.97
Numbers	2.52	2.15	2.51	2.15	1.00
Swear words	0.31	0.64	0.30	0.63	0.99
Social words	8.63	3.97	7.92	3.82	0.98
Family	0.53	0.85	0.51	0.84	0.99
Friends	0.33	0.46	0.32	0.46	0.99
Humans	0.73	0.66	0.67	0.61	0.95
Affect	5.12	2.25	4.04	1.91	0.93
Positive emotions	3.02	1.62	2.26	1.33	0.89
Negative emotions	2.04	1.43	1.76	1.31	0.97
Anxiety	0.39	0.46	0.28	0.39	0.91

Anger	0.69	0.86	0.59	0.79	0.97
Sadness	0.41	0.50	0.37	0.47	0.97
Cognitive mechanisms	16.34	4.02	6.41	2.50	0.75
Insight	2.20	1.26	1.86	1.05	0.86
Causal	1.44	0.80	0.90	0.61	0.83
Discrepancy	1.63	0.98	2.14	1.13	0.87
Tentative	2.60	1.30	2.45	1.27	0.84
Certainty	1.31	0.80	1.08	0.71	0.81
Inhibition	0.43	0.39	0.30	0.30	0.73
	LIWC2007		**LIWC2001**		
	mean	**sd**	**mean**	**sd**	**correlation**
Inclusive	4.96	1.90	5.80	1.62	0.72
Exclusive	2.89	1.49	3.56	1.35	0.61
Seeing	0.79	0.72	0.68	0.53	0.61
Hearing	0.56	0.56	0.96	0.77	0.60
Feeling	0.69	0.63	0.44	0.53	0.68
Body	0.77	0.86	0.69	0.81	0.79
Sexual	0.36	0.66	0.33	0.59	0.91
Motion	2.33	1.34	1.54	1.07	0.86
Space	5.86	2.02	3.41	1.41	0.76
Time	5.75	2.40	4.60	2.10	0.93
Occupation	1.87	1.63	2.12	1.55	0.89
Achievement	1.27	0.87	0.78	0.59	0.80
Leisure	1.20	1.05	1.25	1.11	0.67
Home	0.77	0.90	0.73	0.80	0.89
Money	0.49	0.60	0.35	0.46	0.91
Religion	0.23	0.47	0.20	0.43	0.79
Death	0.14	0.32	0.12	0.30	0.96
Assent	0.73	1.28	0.45	0.87	0.92
Nonfluencies	0.30	0.49	0.10	0.38	0.82
Fillers	0.22	0.80	0.21	0.79	0.99

LIWC Dictionary Translations

The LIWC dictionaries have been translated into several languages, including Spanish, German, Dutch, Norwegian, Italian, Portuguese. Several other language translations are underway, including Arabic, Korean, Turkish, and Chinese. To date, these translations have relied on the LIWC2001 dictionary rather than LIWC2007.

LIWC2007 comes with the Spanish and German translations. All others must be received from the original authors (contact Pennebaker@mail.utexas.edu for more information). The Spanish translation (Ramirez-Esparza, Pennebaker, Garcia, & Suria, 2007) was overseen by a native speaker of Mexican Spanish with close help by a Columbian Spanish speaker. The final version involved the collaboration of a native Spanish speaker from Spain. The German LIWC version (Wolf, Horn, Mehl, Pennebaker, & Kordy, 2008) was developed by all native speaking Germans using high German rather than local dialects.

Additional languages will be added to the LIWC dictionary options as they become available.

Helpful References

Argamon, S., Koppel, M., Fine, J., and Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text*, **23**(*3*).

Argamon, S., Koppel, M., Pennebaker, J.W., & Schler, J. (in press). Automatically profiling the author of an anonymous text. *Communications of the Association for Computing Machinery (CACM)*.

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

Biber, D. (1988). *Variation across speech and writing.* Cambridge: Cambridge University Press.

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

Bosson, J.K., Swann, W.B., Jr., & Pennebaker, J.W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79,* 631-643.

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

Brown, R. (1968). *Words and Things: An Introduction to Language*. New York: Free Press

Bruner, J. S. (1973). *Beyond the Information Given: Studies in the Psychology of Knowing.* Oxford: W. W. Norton; 1973.

Bucci, W. (1995). The power of the narrative: a multiple code account. In J.W. Pennebaker (Ed.), *Emotion, Disclosure, and Health* (pp. 93-122). Washington, DC: American Psychological Association

Buchanan, L., Westbury, C., & Burgess, C. (in press). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomics Bulletin & Review*.

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

Chambers, J. K., Trudgill, P., and Schilling-Estes, N., eds. (2004). *The Handbook Of Language Variation And Change* (London: Blackwell).

Chung, C.K., & Pennebaker, J.W. (2005). Assessing quality of life through natural language use: Implications of computerized text analysis. In W.R. Lenderking and D.A. Revicki (eds.), *Advancing health outcomes research methods and clinical applications* (pp 79-94). Washington, DC: Degnon Associates.

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

Chung, C.K., & Pennebaker, J.W. (in press). Revealing people's thinking in natural language: Using an automated meaning extraction method in open-ended self-descriptions. *Journal of Research in Personality*.

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science, 15,* 687-93.

Crammer, K. and Singer, Y. (2003). Ultraconservative Online Algorithms for Multiclass Problems. *Journal of Machine Learning Research*, **3**:951—991.

Damasio, A. R. (1995). *Descartes' Error: Emotion, Reason and the Human Brain.* New York: Harper Collins.

Davison, K.P, Pennebaker, J.W., & Dickerson, S.S. (2000). Who talks? The social psychology of illness support groups. *American Psychologist*, *55*, 205-217.

Feixas, G., Geldschlager, H., & Neimeyer, R. A. (2002). Content analysis of personal constructs. *Journal of Constructivist Psychology, 15,* 1-19.

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, 5,* 547-557,

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

Francis, W.N., & Kucera, H. (1982). *Frequency analyses of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

Gazzaniga, M. S. (2005). *The Ethical Brain.* New York: Dana Press.

Genkin, A., Lewis, D. D., and Madigan, D. (2006). Large-scale Bayesian logistic regression for text categorization. *Technometrics* (to appear).

Gill, A. (2003). *Personality and language: The projection and perception of personality in computer-mediated communication.* Unpublished doctoral dissertation. University of Edinburgh, Edinburgh, Scotland.

Gill, A. J., Oberlander, J., & Austin, E. (2006). The perception of e-mail personality at zero-acquaintace. *Personality and Individual Differences, 40,* 497-507.

Gortner, E.M., & Pennebaker, J.W. (2003). The 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.

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

Graesser, A. C., Gernsbacher, M. A., & Goldman, S. R. (2003). Introduction to the Handbook of Discourse Processes. In A. C. Graesser, M. A. Gernsbacher, and S. R. Goldman, *Handbook of Discourse Processes* (pp. 1-23). Mahwah, NJ: Lawrence Erlbaum Associates.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers, 36,* 180-193.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments & Computers, 36,* 193-202.

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

Graybeal, A., Seagal, J.D., & Pennebaker, J.W. (2002). The role of story-making in disclosure writing: The psychometrics of narrative. *Psychology and Health, 17*, 571-581.

Groom, C.J., & Pennebaker, J.W. (2005). The language of love: Sex, sexual orientation, and language use in online personal advertisements. *Sex Roles, 52*, 447-461.

Groom, C.J., & Pennebaker, J.W. (2003). Words. *Journal of Research in Personality, 36*, 615-621.

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 Associates, Publishers.

Halliday, M. A. K., and Matthiessen, C. (2004). *An Introduction To Functional Grammar* (3rd ed.) (London: Arnold).

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.

Hartley, J., Pennebaker, J.W., & Fox, C. (2003). Using new technology to assess the academic writing styles of male and female pairs and individuals. *Journal of Technical Writing and Communication, 33*, 243-261.

Hartley, J., Sotto E., & Pennebaker, J. W. (2003). Speaking versus typing: A case-study of the effects of using voice-recognition software on academic correspondence. *British Journal of Educational Technology, 34*, 5-16.

Hartley, J., Sotto, E. and Pennebaker, J. W. (2002). Style and substance in psychology: Are influential articles more readable than less influential ones. *Social Studies of Science, 32*, 321-334.

Heberlein, A.S., Adolphs, R., Pennebaker, J.W., & Tranel, D. (2003). Effects of damage to right-hemisphere brain structures on spontaneous emotional and social judgments. *Political Psychology, 24,* 705-726.

Kanagawa, C., Cross, S. E., & Markus, H. R. (2001). "Who am I?" The cultural psychology of the conceptual self. *Personality & 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.

Koppel, M., Schler, J., and Zigdon, K. (2005), Determining an Author's Native Language by Mining a Text for Errors (short paper), *Proceedings of KDD*, Chicago IL, August 2005.

Koppel, M., Schler, J., Argamon, S., and Pennebaker, J. W. (2006). Effects of age and gender on blogging. Presented at *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*, Stanford, CA, March 2006.

Lee, Chang H., Nam, K., & Pennebaker, J.W. (2004). Is writing as much phonological as speaking? Homophone usage across speaking and writing. *Psychologia: An International Journal of Psychology in the Orient, 47*, 1-9.

Lepore, S. J., & Smyth, J. M. (2002). *The Writing Cure: How Expressive Writing Promotes Health and Emotional Well-Being.* Washington, DC: American Psychological Association.

Li, J., Zheng, R., and Chen, H. (2006). From fingerprint to writeprint. *Communications of the ACM* 49:4 (Apr. 2006), pp. 76-82.

Liehr, P., Mehl, M.R., Summers, L.C., & Pennebaker, J.W. (2004). Connecting with others in the midst of a stressful upheaval on September 11, 2001. *Applied Nursing Research, 17*, 2-9.

Liehr, P., Takahashi, R., Nishimura, C., Frazier, L., Kuwajima, I. & Pennebaker, J.W. (2002). Embodied language: Comparison of the cardiac and stroke health experience for Japanese elders. *Journal of Nursing Scholarship, 34*, 27-32

Lyons, E. J., Mehl, M. R., & Pennebaker, J. W. (2006). Linguistic self-presentation in anorexia: Differences between pro-anorexia and recovering anorexia internet language use. *Journal of Psychosomatic Research, 60,* 253-256.

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., Pennebaker, J. W. (2003). The social dynamics of a cultural upheaval: Social interactions surrounding September 11, 2001. *Psychological Science, 14,* 579-85.

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

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

Mitchell, T. (1999). *Machine Learning*. (New York: McGraw-Hill)

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.

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

Niederhoffer, K.G. & Pennebaker, J.W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology, 21,* 337-360.

Nisbett, R. E. (2003). *The Geography of Thought: How Asians and Westerners Think Differently.* New York, NY: Free Press.

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

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

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological Science, 8*, 162-166.

Pennebaker, J. W. (2002). What our words can say about us: Towards a broader language psychology. *Psychological Science Agenda, 15*, 8-9.

Pennebaker, J. W. (2003). The social, linguistic, and health consequences of emotional disclosure. In J. Suls and K.A. Wallston (Eds.), *Social psychological foundations of health and illness* (pp 288-313). Malden, MA: Blackwell Publishing.

Pennebaker, J. W. & Campbell, R.S. (2000). The effects of writing about traumatic experience. *Clinical Quarterly, 9*, 17-21.

Pennebaker, J. W. & Chung, C.K. (2005). Tracking the social dynamics of responses to terrorism: Language, behavior, and the Internet. In S. Wessely and V.N. Krasnov (Eds.), *Psychological responses to the new terrorism: A NATO-Russia dialogue*. Amsterdam: ISO Press.

Pennebaker, J. W. & Graybeal, A. (2001). Patterns of natural language use: Disclosure, personality, and social integration. *Current Directions in Psychological Science, 10,* 90-93.

Pennebaker, J. W. & Lee, Chang H. (2002). The power of words in social, clinical, and personality psychology. *The Korean Journal of Thinking and Problem Solving, 12*, 35-43.

Pennebaker, J. W., & Chung, C.K. (in press). Computerized text analysis of Al-Qaeda transcripts. In K. Krippendorff & M. Bock (Eds.), *A content analysis reader*. Thousand Oaks, CA: Sage.

Pennebaker, J. W., & Francis, M.E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition and Emotion, 10*, 601-626.

Pennebaker, J. W., Francis ME, Booth RJ. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC2001.* Mahwah: Lawrence Erlbaum Associates.

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., & Ireland, M. (in press). Analyzing words to understand literature. In W. van Peer and J. Auracher (Eds.), *New beginnings for the study of literature*. Cambridge, UK: Cambridge Scholars Publishing.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality & 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,* 271-82.

Pennebaker, J. W., Mayne, T., & Francis, M. E. (1997). Linguistic predictors of adaptive bereavement. *Journal of Personality and Social Psychology, 72*, 863-871.

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., 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 lifespan. *Journal of Personality and Social Psychology, 85,* 291-301.

Pennebaker, J. W., & Stone, L.D. (2004). Translating traumatic experiences into language: Implications for child abuse and long-term health. In L.J. Koenig, L.S. Doll, A. O'Leary, and W. Pequegnat (Eds.), *From child sexual abuse to adult sexual risk: Trauma, revictimization, and intervention* (pp 201-216). Washington, DC: American Psychological Association

Pennebaker, J. W., Slatcher, R.B., & Chung, C.K. (2005). Linguistic markers of psychological state through media interviews: John Kerry and John Edwards in 2004, Al Gore in 2000. *Analysis of Social and Public Policy, 5*, 1-9.

Ramirez-Esparza, N., & Pennebaker, J.W. (2006). Do good stories produce good health? Exploring words, language, and culture. *Narrative Inquiry, 16*, 211-219.

Ramirez-Esparza, N., Pennebaker, J.W., Garcia, F.A., & Suria, R. (2007). La psychología del uso de las palabras: Un programa de comutadora que analiza textos en Español (The psychology of word use: A computer program that analyzes texts in Spanish). *Revista Mexicana de Psicología, 24*, 85-99.

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.

Rosenberg, S.D. & Tucker, G.J. (1978). Verbal behavior and schizophrenia: The semantic dimension. *Archives of General Psychiatry, 36*, 1331-1337.

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

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

Schiller, R., Tellegen, A., & Evens, J. (1995). An idiogrpahic and nomothetic study of personality description. In J. N. Butcher and C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 10, pp. 1-23). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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, Special issue: More data on the current Rorschach controversy,* 71-86.

Scott, M. (1996). *WordSmith.* New York, NY: Oxford University Press.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, **34**(1).

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 & Social Psychology Bulletin, 21,* 834-841.

Slatcher, R.B. & Pennebaker, J.W. (2006). How do I love thee? Let me count the words: The social effects of expressive writing. *Psychological Science, 17*, 660-664.

Slatcher, R.B., Chung, C.K., Pennebaker, J.W., & Stone, L.D. (2007). Winning words: Individual differences in linguistic style among U.S. presidential and vice presidential candidates. *Journal of Research in Personality, 41*, 63-75.

Slobin, D. (1996). From "thought" and "language" to "thinking" for "speaking". From J. J. Gumperz and S. J. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70-96). New York, NY: Cambridge University Press.

Stiles, W.B. (1992). *Describing talk: A taxonomy of verbal response modes*. Newbury Park, CA: Sage.

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 & Applied Social Psychology, 24,* 172-182.

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 & Cognition, 19,* 95-112.

Väyrynen, J.J., & Honkela, T. (2005). Comparison of independent component analysis and singular value decomposition in word context analysis. In T. Honkela, V. Könönen, M. Pöllä, and O. Simula (Eds.), *Proceedings of AKRR'05, International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning* (pp. 135-140). Espoo, Finland.

Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063-1070.

Weber-Fox, & Neville (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.* NY: 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.

Wolf, M., Horn, A., Mehl, M., Haug, S., Pennebaker, J. W., & Kordy, H. (2008, in press). Computergestützte quantitative Textanalyse: Äquivalenz und Robustheit der deutschen Version des Linguistic Inquiry and Word Count [Computer-aided quantitative text analysis: Equivalence and robustness of the German adaption of the Linguistic Inquiry and Word Count]. *Diagnostica*.

Zijlstra, H., van Meerveld, T., van Middendorp, H., Pennebaker, J.W., & Geenen R. (2004). De Nederlandse versie van de Linguistic Inquiry and Word Count (LIWC), een gecomputeriseerd tekstanalyseprogramma [Dutch version of the Linguistic Inquiry and Word Count (LIWC), a computerized text analysis program]. *Gedrag & Gezondheid, 32,* 273-283.

Portions of the research reported in this manual were made possible by grants from the National Institutes of Health (MH52391). We are deeply indebted to a number of people who helped with different phases of this project: Laura King, Cheryl Hughes, Becky Smith, Kathy Davison, Janie Keller, Mary Sue Hayward, Brooke Novales, Anne Vano, Michael Crow, Sally Dickerson, and Bernard Rimé.