Discourse Processes, 45:211–236, 2008

ISSN: 0163–853X print/1532–6950 online
DOI: 10.1080/01638530802073712

Gender Differences in Language Use: An Analysis of 14,000 Text Samples

Matthew L. Newman
Department of Social and Behavioral Sciences
Arizona State University

Carla J. Groom*
Department of Psychology
The University of Texas at Austin

Lori D. Handelman
Oxford University Press
New York

James W. Pennebaker
Department of Psychology
The University of Texas at Austin

Differences in the ways that men and women use language have long been of interest in the study of discourse. Despite extensive theorizing, actual empirical investigations have yet to converge on a coherent picture of gender differences in language. A significant reason is the lack of agreement over the best way to analyze language. In this research, gender differences in language use were examined using standardized categories to analyze a database of over 14,000 text files from 70 separate studies. Women used more words related to psychological and social processes. Men referred more to object properties and impersonal topics. Although these effects were largely

*Carla J. Groom is now employed by the UK Department for Work and Pensions.

Correspondence concerning this article should be addressed to Matthew L. Newman, Department of Social and Behavioral Sciences, Arizona State University, P.O. Box 37100, Phoenix, AZ 85069–7100. E-mail: matt.newman@asu.edu

consistent across different contexts, the pattern of variation suggests that gender differences are larger on tasks that place fewer constraints on language use.

GENDER DIFFERENCES IN LANGUAGE USE: AN ANALYSIS OF 14,000 TEXT SAMPLES

The last several decades have seen an explosion of research on the nature and existence of differences between men and women. One particularly popular question has been the extent to which men and women use language differently. This popularity stems, in part, from the fact that language is an inherently social phenomenon and can provide insight into how men and women approach their social worlds. Within the social sciences, an increasing consensus of findings suggests that men, relative to women, tend to use language more for the instrumental purpose of conveying information; women are more likely to use verbal interaction for social purposes with verbal communication serving as an end in itself (e.g., Brownlow, Rosamon, & Parker, 2003; Colley et al., 2004; Herring, 1993).

At the same time, a number of theorists have argued against the existence of *any* meaningful differences in men's and women's language (e.g., Bradley, 1981; Weatherall, 2002). One contributor to this doubt may be the lack of a commonly accepted metric of analysis among empirical studies of language. Multiple studies, for example, have analyzed a small number of text samples and then made broad generalizations about the differences between women and men. In this project, we explored gender differences in language use in a very large data set of written and spoken text samples using a computerized text analysis tool. Through this exploration, we hope to provide some empirical resolution to the questions of whether, how, and why men and women use language differently.

PREVIOUS RESEARCH ON GENDER DIFFERENCES IN LANGUAGE USE

The empirical literature has been thoroughly reviewed elsewhere (e.g., Mulac, Bradac, & Gibbons, 2001). What follows is a brief overview of previous research on men's and women's language use. In addition to the overall message goals mentioned earlier, men and women may also have different semantic goals in mind when they construct sentences. Some researchers (e.g., Mulac, Weimann, Widenmann, & Gibson, 1988) found that questions are more common in women's contributions to dyadic interactions (e.g., "Does anyone want to get some food?"), whereas directives that tell the audience to do something (e.g., "Let's go get some food") are more likely to be found in men's conversational contributions. In a study of 96 schoolchildren taken from the 4th, 8th, and 12th grades, Mulac, Studley, and Blau (1990) found that boys in all three age groups were more likely than girls to offer opinions (e.g., "This idea is Puritanical."). When mean sentence length is calculated, women come out as the wordier gender both in writing (e.g., Mulac & Lundell, 1994; Warshay, 1972) and speaking (Mulac & Lundell, 1986; Mulac et al., 1988; Poole, 1979). However, men use more words overall and take more "turns" in conversation (e.g., Dovidio, Brown, Heltman, Ellyson, & Keating, 1988).

Some recent studies have failed to replicate these findings. Thomson and Murachver's (2001) study of e-mail communication found that men and women were equally likely to ask questions; offer compliments, apologies, and opinions; and to hurl insults at their "net pal." Other studies have reported significant differences in the opposite direction. In a comparison of 36 female and 50 male managers giving professional criticism in a role play, it was the *men* who used significantly more negations and asked more questions, and the women who used more directives (Mulac, Seibold, & Farris, 2000). However, the study did confirm that men used more words overall, whereas women used longer sentences. One possible explanation for these contradictory reports is that the different contexts in which the language samples were generated influenced the size and direction of the gender differences.

Beginning with Robin Lakoff's (1975) pioneering work, gender differences have also been investigated at the level of specific phrases. Lakoff identified in women's language two specific types of phrases—hedges (e.g., "it seems like,") and tag questions (e.g., " … aren't you?")—that can be inserted into a wide variety of sentences. A number of studies have reported greater female use of tag questions (e.g., McMillan, Clifton, McGrath, & Gale, 1977; Mulac & Lundell, 1986), although others have found the opposite (e.g., Dubois & Crouch, 1975). Other researchers have found further evidence that women use phrases that may communicate relative uncertainty. Uncertainty verb phrases, especially those combining first-person singular pronouns with perceptual or cognitive verbs (e.g., "I wonder if"), have been found more often in women's writing (Mulac & Lundell, 1994) and speech (Hartman, 1976; Poole, 1979). A related interpretation of women's use of hedge phrases is that women are more reluctant to force their views on another person. Consistent with this idea, Lakoff claimed that women were more likely than men in the same situation to use extra-polite forms (e.g., "Would you mind … "), a claim that was supported by subsequent empirical work (Holmes, 1995; McMillan et al., 1977).

Gender differences have also been examined by studying the actual words people use. Mirroring phrase-level findings of tentativeness in female language, women have been found to use more intensive adverbs, more conjunctions such as but, and more modal auxiliary verbs such as could that place question marks of some kind over a statement (Biber, Conrad, & Reppen, 1998; McMillan et al., 1977; Mehl & Pennebaker, 2003; Mulac et al., 2001). Men have been found to swear more, use longer words, use more articles, and use more references to location (e.g., Gleser, Gottschalk, & John, 1959; Mehl & Pennebaker, 2003; Mulac & Lundell, 1986).

One striking result reported by Mehl and Pennebaker (2003) was that women were more likely to use first-person singular. This is consistent with repeated findings that depressed people use more first-person singular (e.g., Bucci & Freedman, 1981; Rude, Gortner, & Pennebaker, 2004; Weintraub, 1981), given that depression is more common among women (Diagnostic and Statistical Manual of Mental Disorders [4th ed., text revision], American Psychiatric Association, 2000). However, the word "I" intuitively connotes individualism or selfishness, which fits the male stereotype better than the female stereotype. The result is also at odds with a review by Mulac et al. (2001), which cited findings that men used first-person singular more often. However, their conclusion was based on only two studies: one representing analyses of 32 essays (4th-grader sample; Mulac et al., 1990), one representing 148 essays (Mulac & Lundell, 1994), and both using relatively impersonal writing tasks (essays and descriptions of photographs). Certainly, if the entire category of personal pronouns is considered, women frequently are the higher users (e.g., Gleser et al., 1959; Mulac & Lundell, 1986). Based on the existing data alone, therefore, it is not possible to either confirm or disconfirm the stereotype that men use I more than women.

Emotion words appear to be another area of conflicting findings, despite the existence of a fairly clear stereotype. Several studies have reported that women refer to emotion more often than do men (Mulac et al., 1990; Thomas & Murachver, 2001). Yet, Mulac et al.'s (2000) study of managers providing criticism in a role play found precisely the reverse. Mehl and Pennebaker (2003) offered a potential reconciliation: Women used more references to positive emotion, but men referred more to anger—a finding that is perfectly consistent with gender stereotypes.

Limitations of Previous Research

A frustration of studying natural language is that people use words in a variety of ways that change as a function of context. To draw broad conclusions about how men and women differ in their language use across settings, nontraditional methods with large samples are often required. In previous studies examining age and individual differences, mean effects sizes for gender have typically ranged between .08 and .20. For example, Pennebaker and Stone (2003) reported a significant effect size (*d*) of .12 between gender and sentence length with a sample of over 2,000 people. However, the mean words-per-sentence was 23.4 for men and 19.1 for women, with a standard deviation of 35.1, across written and spoken text. A difference of 4 words compared to a standard deviation of 35 words may not be apparent in a small sample with low statistical power. Thus, large samples may be required to detect the smaller differences between men's and women's language.

Unfortunately, many previous studies have had fewer than 50 participants per cell. Larger samples are often difficult to collect when each sample must be hand coded. The need to conserve coder time also reduces the number of features that can be coded in a single study. This reality has focused attention toward features of language that can be easily related to gender stereotypes (e.g., hedges), potentially missing differences in less obvious language categories (e.g., pronouns). Thus, a strategy that allowed for the efficient analysis of large samples of text could help to create a more complete picture of gender differences in language use.

A related limitation is that coding schemes are not always consistent across studies. Even where the name of the language category is shared by two or more studies, the actual features coded for may be different. One researcher's uncertainty verb phrase is another's hedge. This problem is exacerbated by multivariate approaches that compare men and women on a set of language features, rather than reporting mean differences on individual features. The simplest form of this approach was used by Crosby and Nyquist (1977) in which they created a composite "female register" index. The more complex multivariate approaches (e.g., Mulac & Lundell, 1986) use multivariate analyses of variance (MANOVAs) in which language features are weighted differently to achieve maximum discrimination between the genders. Mulac and his associates (Mulac et al., 1986; Mulac et al., 1990; Mulac et al., 2000) argue that language is produced and comprehended as a *gestalt,* and should be analyzed accordingly. However, such an approach makes it difficult to compare results of one study with results of a second study that uses different combinations of features. A standardized set of language categories, composed of a standardized set of features used in coding for that category, would shed new light on the ways in which men and women communicate differently.

Function Words

Because of these limitations, empirical studies of language itself have yet to provide a coherent picture of gender differences in language use. Perhaps the greatest stumbling block has been in deciding how to analyze language as it relates to women and men. Language is inherently complex, and can be analyzed at several levels of analysis. As discussed earlier, explorations of gender differences in language have ranged from the overall structure of men's and women's narratives (e.g., Herring, 1993; Tannen, 1990), down to the level of specific phrases (e.g., Holmes, 1995; McMillan et al., 1977; Thomas & Murachver, 2001) and words (e.g., Biber et al., 1998; Danner, Snowdon, & Friesen, 2001; Gleser et al., 1959; Mehl & Pennebaker, 2003; Mulac et al., 2001). Which dimensions of language should we examine to capture differences in how men and women approach the world?

A growing body of research suggests that we can learn a great deal about people's underlying thoughts, emotions, and motives by counting and categorizing the words they use to communicate (Pennebaker & King, 1999; Pennebaker, Mehl, & Niederhoffer, 2003; cf. Shapiro, 1989). This approach has proved particularly fruitful with respect to "function words," which include pronouns, articles, prepositions, conjunctions, and auxiliary verbs. These words are distinct from content words (nouns, verbs, and adjectives), and are used to "glue" other words together. In the English language, there are fewer than 200 commonly used function words, yet they account for over one half of the words we use.

Differences in the use of function words reflect differences in the ways that individuals think about and relate to the world. For example, using "you and I" instead of "we" reflects a different perspective on the relationship between the speaker and the referent. Using more pronouns in general (rather than nouns) refers to a shared reality, in that both parties have to understand who "he" is. Empirically, the use of first-person singular has been associated with age; depression; illness; and, more broadly, self-focus (Pennebaker et al., 2003). First-person plural can variously be a marker of group identity and, on occasion, a sign of emotional distancing (Brewer & Gardner, 1996; Pennebaker & Lay, 2002). Function words can also reflect psychological state independent of content. For example, people telling the truth use more first-person singular and more qualifying conjunctions (e.g., but) than those instructed to lie—although they are discussing the same topics (Newman, Pennebaker, Berry, & Richards, 2003). This approach to language suggests that differences in how individuals communicate can sometimes be as meaningful as what they communicate.

An examination of gender differences in function word use might shed new light on the psychology of men and women. Apart from personal pronouns, however, function words have been relatively neglected in previous gender difference work. In one notable exception, Koppel, Argamon, and Shimoni (2003) discriminated between male and female authors in a sample of fiction and nonfiction from the British National Corpus. Their goal in that study was to predict author gender without regard to the psychological meaning of the words. These authors used a set of training documents to create a prediction equation, which was used to classify writing by gender, at an accuracy rate near 80%. Empirically, those words that best discriminated between men and women were function words. In a second notable exception, Biber et al. (1998) used parts of speech to create an index of whether a text sample was "involved" (e.g., more pronouns, present-tense verbs) or "informative" (e.g., more nouns, long words). Consistent with prior research, females' language was more involved than males' language.

PRESENT STUDY

In this study, we conducted a comprehensive survey of gender differences in language. This survey was brought within reach by a combination of two methodological developments. The first was our text analysis program, Linguistic Inquiry and Word Count (LIWC; Pennebaker, Francis, & Booth, 2001), which allowed us to perform an extensive linguistic analysis on each individual text in our archive. LIWC analyzes text samples on a word-by-word basis and compares each to a dictionary of over 2,000 words divided into 74 linguistic categories. Output is expressed as a percentage of the total words in the text sample. Some of the categories are defined purely grammatically. For example, the "articles" category searches for instances of a, an, and the. Other categories, such as positive emotion words, were formed initially by having independent judges decide which words should go into each category. Thus, an element of qualitative human judgment is incorporated into an automated and perfectly consistent coding system. LIWC usually recognizes about 80% of the words in a given text sample—proper nouns and very low-frequency words comprise the other 20%. Once the text samples are assembled, thousands of samples can be analyzed on dozens of dimensions in a matter of seconds.

A word count strategy such as LIWC is an admittedly crude way by which to study language use. It cannot detect the context or underlying meaning of words. It fails to appreciate sarcasm or irony. Because of these problems, words are often incorrectly categorized. For example, the word "mad" is currently categorized as an anger word. Phrases such as "I'm mad about you" (suggesting positive emotion) or "mad as a hatter" (indicative of mental health problems) will be miscoded by the computer. It is best to think of word count strategies as probabilistic ways of studying language use. Statistically, we have found that mad is correctly coded about 90% of the time. Fortunately, in any given text, someone who is angry will use several other anger-relevant words. In those cases where the person uses mad and is actually happy, other positive emotions words will surface. In short, word count approaches are prone to errors; but, with large data sets, the likely error rate is extremely low.

The second methodological development has been the creation of a text archive itself. Over the last decade, we have collected a large corpus of over 500,000 text files in the development of LIWC. Labs from all over the world have provided us with language samples based on written and transcribed spoken language. In addition, we have accrued samples of books, poems, song lyrics, and other art forms—many of which had never been subjected to linguistic coding. As a result, we have the opportunity to observe gender differences on a much larger scale than has been attempted in the past (e.g., Biber et al., 1998). This corpus is described in more detail in the Method section.

This research strategy shares certain features with a traditional meta-analysis, but is also distinct in several important respects. As in the case of a meta-analysis, we drew on data collected in a variety of labs on a variety of populations, enabling us to take advantage of the statistical power, increased external validity, and opportunities to examine the effects of study-level moderators that are yielded by such pooling of data. However, we did not pool analyses conducted by other researchers; instead, we used LIWC to code and categorize all of the original raw data, and performed our own primary analyses on the resulting measures. A traditional meta-analytic strategy would have been to seek out all the language studies that had ever coded for a particular language feature (e.g., use of first-person singular); use reported means for men and women to obtain study-level effect sizes; and calculate an overall, weighted, meta-analytic effect size for this language feature. Such a traditional strategy would yield no more and no less than a quantitative synthesis of existing research, constrained by the limited number of studies that had coded for each feature using varying definitions.

Our first goal in this study was to ask a rather simple question: Do men and women use language differently? To answer this question, a large corpus of text samples was subjected to LIWC analysis, and these linguistic data were analyzed for main effects of gender. In analyzing this large body of primary data, our goal was to resolve some of the discrepancies in previous studies. We expected results consistent with the overall picture from previous research—that is, men's language should focus relatively more on conveying information, and women's language should focus relatively more on social connections. However, because of the issues with previous research mentioned earlier, and because of the large and diverse corpus of text at our disposal, we refrained from making predictions about specific language categories. Instead, we took an exploratory, bottom-up approach to men's and women's language use (see Oberlander & Gill, 2006, for a discussion about the merits of a bottom-up approach). We expected the largest differences to be found on function words because these words appear to be particularly good markers of how individuals relate to the world. However, we also examined a range of social and psychological process words, including references to friends, family, and emotions, to better understand how men and women differ in their language use (see Table 1 for a list of language categories).

Our second goal in this study was to examine whether the context in which text samples were produced affected gender differences in language use. Few previous attempts have been made to systematically study how context influences the size and direction of gender differences in language use. As described later, the corpus contained text samples from seven different context categories (see Table 2). We expected that language would differ dramatically across these contexts. However, our primary interest was in the interactions between gender and communication task. We predicted that the overall picture of gender differences would persist across context; but, in keeping with our bottom-up approach to this project, we refrained from making predictions about specific language categories. In addition, previous research has identified differences between speaking and writing, such that the latter is thought to involve more planning and complexity (e.g., Biber,

TABLE 1
Main Effects of Gender on Language Use

		Female		*Male*		
LIWC Dimension	*Examples*	*M*	*SD*	*M*	*SD*	*Effect Size (d)*
Linguistic dimensions						
Word count		1,420	5,403	1,314	6,016	*ns*
Words per sentence		21.26	31.22	23.90	48.12	–0.07
Question marks		3.21	7.33	3.07	7.86	ns
Words ≥ six letters		13.99	4.42	15.25	5.91	–0.24
Numbers		1.37	1.31	1.59	1.55	–0.15
Negations	no, never, not	1.85	1.10	1.72	1.17	0.11
Articles	a, an, the	6.00	2.73	6.70	2.94	–0.24
Prepositions	on, to, from	12.46	2.44	12.88	2.64	–0.17
Inclusive words	with, and, include	6.42	1.88	6.34	2.03	*ns*
Exclusive words	but, except, without	3.82	1.54	3.77	1.64	*ns*
Psychological processes						
Emotions		4.57	1.99	4.35	2.07	0.11
Positive emotions	happy, pretty, good	2.49	1.34	2.41	1.40	*ns*
Optimism	certainty, pride, win	0.56	0.58	0.58	0.61	*ns*
Positive feelings	happy, joy	0.61	0.61	0.51	0.65	0.15
Negative emotions		2.05	1.65	1.89	1.56	0.10
Anxiety	nervous, afraid, tense	0.48	0.68	0.38	0.64	0.16
Sadness	grief, cry, sad	0.55	0.76	0.47	0.70	0.10
Anger	hate, kill	0.61	0.81	0.65	0.92	*ns*
Swear words	damn, ass, bitch	0.09	0.25	0.17	0.44	–0.22
Sensations		2.22	1.27	2.06	1.30	0.12
Feeling	touch, hold, feel	0.58	0.67	0.47	0.66	0.17
Hearing	heard, listen, sound	0.78	0.74	0.71	0.72	0.10
Seeing	view, saw, look	0.72	0.78	0.74	0.83	*ns*
Cognitive processes		7.35	2.57	7.17	2.82	0.07
Causation	effect, hence	1.02	0.76	1.02	0.88	*ns*
Insight	think, know	2.40	1.28	2.28	1.38	0.09
Discrepancy	should, would, could	2.32	1.31	2.23	1.46	0.07
Tentative	perhaps, guess	2.54	1.43	2.54	1.57	*ns*
Certaintyns	always, never	1.35	0.94	1.21	0.96	0.14
Hedge verb phrases	I + guess, I + reckon	0.57	0.67	0.50	0.67	0.11
Social processes						
Social words		9.54	4.92	8.51	4.72	0.21
Communication	talk, share, converse	1.26	0.95	1.20	0.95	*ns*
Friends	pal, buddy, coworker	0.37	0.51	0.33	0.53	0.09
Family	mom, brother, cousin	0.77	1.04	0.64	1.01	0.12
Humans	boy, woman, group	1.22	1.33	1.15	1.33	*ns*

(continued)

TABLE 1 (*Continued*)

		Female		*Male*		
LIWC Dimension	*Examples*	*M*	*SD*	*M*	*SD*	*Effect Size (d)*
Pronouns		14.24	4.06	12.69	4.63	0.36
First-person singular	I, me, my	7.15	4.66	6.37	4.66	0.17
First-person plural	we, us, our	1.17	2.15	1.07	2.12	*ns*
Second person	you, you're	0.59	1.05	0.65	1.15	–0.06
Third person	she, their, them	3.41	3.45	2.74	3.01	0.20
Time and space						
Time	till, started, hour	4.09	1.94	4.03	2.14	*ns*
Past-tense verb	walked, were, had	4.36	2.97	4.02	2.84	0.12
Present-tense verb	walk, is, be	11.71	4.00	10.98	4.10	0.18
Future-tense verb	will, might, shall	1.27	1.03	1.33	1.18	*ns*
Space	here, up, around	2.40	1.18	2.47	1.31	*ns*
Motion verbs	walk, move, go	1.22	0.89	1.15	0.93	0.07
Current concerns						
Occupation	work, class, boss	2.34	1.88	2.59	2.10	–0.12
Money	cash, taxes, income	0.25	0.39	0.29	0.49	–0.10
Metaphysical	death, god	0.41	0.88	0.47	0.97	–0.06
Leisure	house, TV, music	1.11	0.92	1.07	1.08	*ns*
Home	house, kitchen, lawn	0.80	0.76	0.68	0.79	0.15
Sports	football, play, game	0.13	0.30	0.19	0.50	–0.15
Physical functions	ache, breast, sleep	1.33	1.22	1.28	1.29	*ns*
Sex	lust, pregnant, gay	0.30	0.53	0.27	0.60	*ns*

Note. Means (except for word count and words per sentence) refer to percentages of the total words in a sample. Effect size (Cohen's *d*) was calculated by dividing the mean difference by the pooled standard deviation. Positive effect sizes mean women used the category more; negative effect sizes mean men used it more. All mean differences except those labeled "*ns*" were significant at $p < .001$, based on univariate statistics from a multivariate analysis of variance.

1991). Based on this literature, we also predicted that the size of the gender difference would be largest in spoken language (i.e., the conversation category) because it is more natural and spontaneous (e.g., Biber, 1991).

Method

Text corpus.[1] Our archive of electronic text samples represented 70 studies from 22 laboratories. These laboratories included 14 universities in the United States (63 studies), 1 university in New Zealand (4 studies), and 3 universities in England (3 studies). Forty-four of the studies (63%) were conducted by at least one of the authors. The studies were conducted over a 22-year period (1980–2002), and included samples of fiction going back as far as the 17th century. All the files

[1]We are happy to make this corpus available to other researchers. Interested parties can contact James W. Pennebaker for more information at pennebaker@mail.utexas.edu

TABLE 2
Characteristics of Text Files, by Context and Participant Gender

Context	*No. Text Files*	*Mean Word Count (SD)*	*% Aged 18–22*	*% Written*
Emotion				
Men	3,603	1,566 (5,159)	62.7	96.7
Women	5,263	1,574 (4,724)	57.9	93.0
Time management				
Men	520	299 (213)	82.3	96.9
Women	723	295 (201)	86.4	97.2
Stream of consciousness				
Men	793	481 (246)	89.8	95.2
Women	1,033	561 (260)	88.5	96.2
Fiction				
Men	37	21,593 (52,366)	0.0	100.0
Women	29	24,302 (55,593)	0.0	100.0
TAT–inkblot				
Men	680	266 (175)	98.9	100.0
Women	996	311 (211)	97.3	100.0
Exam				
Men	170	631 (636)	100	96.5
Women	90	706 (455)	93.3	97.8
Conversation				
Men	168	3,466 (4,370)	54.1	0.0
Women	219	7,808 (7,575)	39.3	0.0

Note. N = 14,324 text files (5,971 men; 8,353 women). Age percentage refers only to the subset of the sample (70.7%) for which age information was available. TAT = Thematic Apperception Test.

contained primary data from individual participants, either written (93%) or transcribed from speech (7%). It is noteworthy that only about two thirds of the text files were from college-age participants, in contrast to most psychological studies of language to date. There was also a good mix of spoken and written samples, with the entirely spoken conversation category being balanced by predominantly written samples in the other categories.

After excluding files for which no gender information was available, and studies including only men or women, there remained text samples from 11,609 participants, consisting of approximately 45,700,000 words. In many of the studies, participants had provided multiple samples within a particular context, to aid in the reliability of linguistic style (M = 3.5 samples per person, SD = 2.4). When this was the case, samples were aggregated such that there was only 1 text file per person, per context. Of the 11,609 participants, 2,130 contributed text samples in two or more contexts. The aggregation process yielded 14,324 final text files, with 5,971 written by men and 8,353 written by women.

The corpus contained text samples from seven different context categories: *emotion, time management, stream of consciousness, fiction, Thematic Apperception Test (TAT)–inkblot, exams,* and *conversation.* Table 2 summarizes for each category the number of text files, the mean word count, the percentage of college-age participants, and the percentage of written (vs. spoken) samples. The *emotion* category contained language samples in which participants addressed emotional (usually traumatic) life events. These included writing studies conducted in a traditional laboratory setting and interviews in which participants discussed such topics as their family histories or their thoughts and feelings about traumatic events. The *time management* category contained language samples from the control conditions in writing studies in which participants wrote or spoke about time management. The *stream of consciousness* category contained language samples from college student participants who were asked simply to track their thoughts and feelings as they occurred as part of class assignments in at least 10 Introductory Psychology classes. The *fiction* category contained the full text of fictional novels, mostly consisting of the top-selling fiction books from the year 1996 (e.g., *The Alienist* by Caleb Carr, and *K is for Killer* by Sue Grafton). The *TAT–inkblot* category consisted of participants' free responses to describing drawings of specific scenes (TAT) or standard inkblots typically used in the Rorschach test. The *exam* category consisted of essays written for class exams in psychology courses. Finally, the *conversation* category contained spoken natural language samples, talk-show transcripts, and other nontraumatic face-to-face interviews.

Text analysis. By default, LIWC analyzes each text on 74 language dimensions. Two categories unique to conversation, *nonfluencies* (e.g., er, hmm, um) and *fillers* (e.g., y'know), were excluded. We also excluded categories with very low base rates (< 0.2%), such as *TV* and *sleep,* although we made an exception for two low base rate categories, *swear words* and *sports,* because of their close relations to gender stereotypes.

In addition, we created a new dictionary category of "hedges." Previous research (e.g., Mulac et al., 2000) suggested that women might be more likely to use verb phrases such as "I figure" and "I guess" to qualify the impact of a sentence. To capture hedges with LIWC, 12 graduate student raters were asked to decide whether each of 43 candidate phrases was commonly used as a hedge, intended to qualify a statement and reduce its force. Only those items with high interrater reliability (agreement by at least 10 of the 12 raters) were selected for the final dictionary. Our final data set represented 14,324 text files, analyzed along 54 language dimensions.

Results

Main effects of gender on language. To answer the most basic question of whether men and women in our sample used language differently, we performed a MANOVA on all of the LIWC categories that met our inclusion criteria, with gender as the independent variable. As expected, this analysis yielded a significant multivariate effect, $F(53, 14{,}270) = 30.66, p < .001$. Individual univariate analyses of variance allowed us to identify the language dimensions that men and women used differently. Table 1 presents the results of each univariate test, including effect sizes (Cohen's d) for significant analyses. Due to the large number of analyses being conducted, as well as the exceptionally large sample size, only highly significant p values < .001 were considered significant. In addition, many of the significant effect sizes are in or below the range generally considered small; Cohen (1992) recommended a cutoff of .20 for a "small" effect. Given the bottom-up nature of this project, we took a more liberal approach and confined our interpretations to effect sizes of d = .10 or greater (see the Discussion section for more on this issue).

The table is organized in generally the same way as LIWC itself was constructed. The overarching structure distinguishes between (a) linguistic properties, (b) words connoting psychological processes, (c) words connoting social processes, (d) words that clarify temporal and spatial context, and (e) current concerns that are wholly content rather than style oriented. The overall picture painted by Table 2 is of a multitude of differences combined with a good deal of overlap between the language of men and women.

Females' language was more likely than men's to include pronouns and social words, a wide variety of other psychological process references, and verbs. Negations and references to the home were also features of the female profile. Men exceeded women on a number of linguistic dimensions including word length, numbers, articles, and prepositions. Men also discussed various current concerns more frequently, and swore more often. It is worth noting that some of these word frequencies are inextricably linked. For example, using more prepositions requires using more articles ("to the store," makes sense; "to store" does not). Consistent with this, the Pearson correlation between gender and article use slightly decreased when we controlled for prepositions (from $r = .12$ to $r = .10$).

Null effects were found for word count, question marks, first-person plural, and a variety of relatively narrow categories including anger, gender, time, and space. Table 1 highlights the fact that men and women did not merely focus on different word categories (e.g., anxiety vs. anger, or emotion words vs. cognitive words), but actually used different superordinate language dimensions. Female language emphasized psychological processes, social processes, and verbs. Male language emphasized current concerns. Thus, the results are consistent with the idea that men and women employ language for different reasons.

To illustrate the main effects of gender on language, we selected representative examples from the "stream of consciousness" communication context. This particular context was selected for two reasons: (a) The cell sizes were reasonably large (793 men; 1,033 women), and (b) it imposed the fewest constraints on the content of participants' language. Excerpts from 4 representative women and 4 representative men are presented below.

Examples of females' language. In the following excerpts, female college students tracked their thoughts as they occurred. Free from constraints, these women made references to psychological processes (e.g., mad, uneasy, remember, nervous), social processes (e.g., sister, friends), verbs (e.g., watching, taking, talking, thinking), and negations (e.g., can't, not):

> Female #1: Okay, well, I am watching this movie. I'm not really watching it because I'm typing, but I'm listening to it. I really can't type that well, so there are probably going to be a few misspelled words. My sister made me mad a while ago because I asked her to call me when her husband got home and she didn't.

> Female #2: Palms are sweaty, my stomach is uneasy, and my head just feels in pain. I'm sick, I'm not supposed to get sick. I'm pre-med I'm supposed to be taking good care, promoting health. I need to get better, it is essential that I get better.

> Female #3: okay, so I'm sitting here talking to my friends. I miss them so much. they live back home in houston. I wish I could see them just like old times. I remember when we would all hang out at school together. it was great.

> Female #4: Right now, I am thinking about my chemistry homework and test. I am very nervous about it and I am worried that I may not succeed to my fullest potential.

Examples of males' language. In the following excerpts, male college students tracked their thoughts as they occurred. Free from constraints, these men used longer words and more articles (e.g., "a bit," "the music," "a journal"). Men also made references to current concerns (e.g., assignment, apartment), including multiple references to the "stream of consciousness" instructions that are absent from females' language:

> Male #1: Sorry for any grammar mistakes in this timed writing. There's a bit of pressure writing every thought you have within 20 minutes and try and make it completely coherent. The music in the back ground plays that of falling falling falling.

> Male #2: I find it amusing that in writing a stream of consciousness about what I am thinking, my mind is completely focused on what I am going to write in the stream of consciousness paper. Thus, my stream of consciousness is about my stream of consciousness about my stream of consciousness, etc.

> Male #3: Stream of Consciousness? How do you start something so vague. I keep a journal which I write in occasionally, but I can not remember the last time an assignment consisted solely of write your thoughts.

> Male #4: Cool. I'm currently sitting in the campus library completing this assignment because the Time Warner Cable people consistently refuse to show up to our apartment to install our internet. The two guys next to me keep talking about web sites and how they can improve the overall aesthetic beauty of the page by writing some of the code in java script.

Age as a Moderator Recent research has suggested that language use also varies according to an individual's age (Pennebaker & Stone, 2003), and that gender differences vary across children of different ages (Mulac et al., 1990). Therefore, we repeated the previous analysis using linguistic categories that had been adjusted for age. Age information was available for 70.7% of the sample, either specific to each individual or, in relatively homogenous populations where the group mean was known (primarily Introductory Psychology classes), by using the mean to replace missing values. This resulted in a total of 10,131 text files with associated age information.

A significant multivariate effect of gender was again found for the age-adjusted means, $F(51, 10{,}079) = 17.73$, $p < .001$. The pattern of univariate results was nearly identical when the effects of age were controlled. Not a single effect switched directions from female to male advantage or vice versa. Six previously significant effects now failed to meet our alpha level ($p < .001$): second-person pronouns, total cognitive words, discrepancies, hedge verb phrases, motion verbs, and metaphysical references. This may be partly attributable to the loss of 29.3% of the sample for which age information was unavailable. Nevertheless, the overall picture was of gender differences in language use that remained unchanged when age was controlled for.

Function Words Versus Other Categories We hypothesized that the largest differences between males' and females' language would be found with function words. Function words, which tend to be processed in the brain differently from more traditionally studied nouns and regular verbs, have been found to be linked to emotional state such as depression (Rude et al., 2004), social connections with others (Chung & Pennebaker, 2007), and how individuals think about the world around them (Pennebaker et al., 2003). As a beginning way to examine the nature of function words and gender differences, we compared the average effect size for the function word categories to the average effect size for the content categories of nouns and verbs.

Because our interest was in the size rather than the direction of this analysis, we used the absolute value of effect sizes that fit into each category (see Table 1). We also limited this comparison to effect sizes that met our alpha criterion ($p < .001$). The average effect size for verbs (feeling, hearing, insight, past-tense verbs, present-tense verbs, and future-tense verbs) was $d = .10$, and the average effect size for nouns (friends, family, occupation, money, metaphysical, home, and sports) was d

= .11. The average effect size for function words (articles, prepositions, discrepancy words, and total pronouns) was $d = .21$. Consistent with our predictions, the size of the gender difference was roughly twice as large for function words as for either nouns or verbs. However, it is also clear from looking at Table 1 that men and women differed in their use of both content and function words.

Communication Context To examine the effects of communication context, a 2 (Gender) × 7 (Context) MANOVA was performed on the set of 54 linguistic categories. A significant multivariate main effect of context was found, $F(432, 85{,}464) = 119.84$, $p < .001$; in addition to a main effect for gender, $F(72, 14{,}239) = 5.72$, $p < .001$. These were qualified by a significant Gender × Context interaction, $F(432, 85{,}464) = 4.64$, $p < .001$. Table 3 lists the 37 categories for which significant interactions were found, together with effect sizes for the gender differences and the form of the interaction.

In about one half of the cases (19 out of 37), the interactions were simply additive—that is, the gender differences differed in magnitude across the conditions, but were in the direction reported in Table 1. The remaining 18 interactions were crossovers—that is, the overall gender difference was *reversed* in at least one condition; or, in the cases where no overall effect had been found, different conditions showed effects in opposite directions. In nearly all of the crossover cases, the effect was driven by reversals in conversation contexts. In conversation, men and women still differed quite dramatically from each other, but not in the same ways that they did in writing or more formal speech. In conversations, for example, men used many more negations, negative emotion words, present-tense verbs, and references to leisure activities, whereas women predominated in their references to numbers. A closer examination of these findings suggests that men's speech is characterized by more negative emotion and references to the past relative to men's writing.

The bottom section of Table 3 presents average effect sizes for each condition, computed using the absolute value of each language dimension. The average effect sizes for the five superordinate categories show a complex pattern. Men's and women's use of both linguistic categories and time and space references differed most in the conversation context. In contrast, the use of psychological and social processes both showed the largest gender differences in the fiction context. Finally, gender differences in reference to current concerns were equally large in both the fiction and conversation contexts.

Collapsing across all LIWC categories, the average overall effect sizes were largest for fiction writing ($d = .31$), natural conversation ($d = .26$), and exam essays ($d = .22$). Overall effect sizes were considerably smaller for stream of consciousness ($d = .11$), TAT–inkblot ($d = .09$), emotional writing ($d = .08$), and time management ($d = .08$). It is also noteworthy that these latter two conditions,

TABLE 3
Significant Gender × Context Interactions

		Effect Sizes						
LIWC Category	*Crossover Effect?*	*Emotion*	*Time Management*	*Stream of Consciousness*	*Fiction*	*TAT–Inkblot*	*Exams*	*Conversation*
Linguistic								
Word Count		.00	–.01	.23*	.04	.16*	.10*	.50
Words per sentence	Yes	–.03	.03	–.12*	–.14*	–.01	.28*	.00
Question marks	Yes	–.02	–.03	.04	–.04	.04	–.07*	.49*
Words ≥ six letters		–.16*	–.04	–.09*	–.32*	–.17*	–.45*	–.44*
Numbers	Yes	–.09	–.15*	–.13*	–.37*	–.10*	–.09*	.14*
Negations	Yes	.08	.04	.17*	.58*	.05	.13*	–.22*
Articles		–.21*	–.07	–.33*	–.70*	–.22*	–.05	–.77*
Prepositions		–.11*	–.09*	–.12*	–.26*	–.09*	–.11*	–.74*
Inclusive words		.04	.05	.03	–.38*	.07	–.17*	–.35*
Exclusive words	Yes	.03	.05	.03	.06	–.04	.13*	–.51*
Psychological								
Positive emotion		.03	.03	–.03	.31*	.15	–.06	.12*
Negative emotion	Yes	.09*	.05	.05	.01	.03	–.26*	–.30*
Anxiety	Yes	.13*	.05	.09*	.04	.12*	–.30*	.16*
Anger		–.02	–.05	–.09*	–.10*	.04	–.18*	–.43*
Swear words		–.14*	–.10*	–.24*	–.28*	–.12*	.06	–.43*
Senses	Yes	.10*	.09*	.12*	.56*	.03	.25*	–.12*
See	Yes	.00	.02	–.08	.24*	–.03	.14*	–.32*
Hear		.05	.21*	.13*	.45*	.02	.22*	.00
Discrepancies	Yes	.02	.03	.05	.57*	.16*	.34*	–.13*
Tentative		.01	.03	–.05	.25*	–.06	.35*	–.19*
Social								
Social words		.11*	.20*	.21*	.74*	.28*	.56*	.14*
Communication		.02	.20*	.13*	.55*	.02	–.08	.05
Humans	Yes	.06	.05	–.06	–.01	.01	.47*	–.17*

continued)

TABLE 3 (*Continued*)

	Effect Sizes							
LIWC Category	*Crossover Effect?*	*Emotion*	*Time Management*	*Stream of Consciousness*	*Fiction*	*TAT–Inkblot*	*Exams*	*Conversation*
Total pronouns		.25*	.20*	.33*	.79*	.30*	.46*	.03
First-person singular		.16*	.00	.17*	.27*	–.01	.29*	.10*
Second person		–.07	.00	–.05	.32*	–.08	.24*	.33*
Third person		.16*	.18*	.20*	.66*	.32*	.51*	–.08
Time and space								
Space	Yes	–.04	.09*	–.07	–.19*	–.04	–.06	–.70*
Motion verbs	Yes	.06	.00	.10*	.11*	.05	.13*	–.27*
Time	Yes	.05	–.06	–.04	.23*	–.03	.21*	–.23*
Past	Yes	.09*	.04	.01	.43*	.11*	.05	–.10*
Present		.14*	.04	.17*	.28*	.09*	.41*	–.21*
Future	Yes	–.05	–.12*	.02	.10*	–.03	.32*	–.16*
Current concerns								
Money	Yes	–.07	–.07	–.04	–.18*	–.05	.11*	–.39*
Leisure	Yes	.06	–.14*	–.03	.44*	.09*	.16*	–.32*
Home		.12*	–.11*	.20*	.46*	.15*	.25*	.02
Sports		–.08	–.13*	–.19*	–.05	–.06	–.10*	–.39*
Overall average:		0.08	0.08	0.11*	0.31*	0.09*	0.22*	0.26*
(absolute value)								
Linguistic		0.08	0.06	0.13*	0.29*	0.10*	0.16*	0.42*
Psychological		0.06	0.07	0.09*	0.28*	0.08	0.22*	0.22*
Social		0.12*	0.12*	0.16*	0.48*	0.15*	0.37*	0.13*
Time and space		0.07	0.06	0.07	0.22*	0.06	0.20*	0.28*
Current concerns		0.08	0.11*	0.12*	0.28*	0.09*	0.20*	0.28*

Note. All interactions reported in this table are significant at $p < .001$ based on univariate analyses of variance. Effect size (Cohen's d) was calculated by dividing the mean difference by the pooled standard deviation. Positive effect sizes mean women used the category more; negative effect sizes mean men used it more. "Crossover effect" means that significant effect sizes were found in opposite directions across conditions. "Average" effect sizes at the bottom of the table represent the mean *absolute value* across all language dimensions. LIWC = Linguistic Inquiry and Word Count.

$^*p < .001$.

both from studies of emotional writing, showed the largest differences in the use of social process words.

DISCUSSION

Comparing the language of men and women in a large, heterogeneous sample of written and spoken text reveals small but consistent gender differences in language use. For the women who contributed 8,353 text files to the study, the English language was more likely to be used for discussing people and what they were doing, as well as communicating internal processes to others, including doubts. Thoughts, emotions, senses, other peoples, negations, and verbs in present and past tense figured high on the list of words that women used more than men. For the men who contributed 5,970 files, language was more likely to serve as a repository of labels for external events, objects, and processes. Along with discussion of occupation, money, and sports were technical linguistic features such as numbers, articles, prepositions, and long words. Swear words added emphasis to male language. Contrary to popular stereotypes, men and women were indistinguishable in their references to sexuality, anger, time, their use of first-person plural, the number of words and question marks employed, and the insertion of qualifiers in the form of exclusion words (e.g., but, although).

It mattered little whether age was controlled for, although the map of gender differences did vary somewhat across communication context. The primary difference was that men's speech was characterized by more negative emotion and more references to the past relative to men's writing. It is interesting to note that the effect sizes were almost all larger in natural conversation, perhaps because spoken language is more "natural" than writing (e.g., Biber, 1991). Similarly, natural language, by its very nature, tends to be more informal and less constrained. Function words, such as pronouns, are used at much higher rates in conversation—especially by women. In addition, when given the freedom to talk about any topic, men (but not women) elected to talk about concrete objects, which require nouns and, of course, articles. Taken together, the general pattern of variation across contexts suggests that gender differences are larger on tasks that place fewer constraints on language use.

The effect sizes on all language dimensions were in the range generally considered small. In fact, only five dimensions met Cohen's (1992) criterion for a small effect when we collapsed across communication context—long words, articles, swear words, social words, and pronouns. Although larger effects were seen in the conversation and stream of consciousness contexts, these five dimensions appear to be the most meaningful differences in males' and females' language. In interpreting the size of these effects, two considerations are worth noting.

First, Eagly (1995) pointed out that a sizable portion of other gender differences have effect sizes ranging from small to moderate. These results are particularly compelling because of the diverse content of the text samples (i.e., some of the samples came from experimental studies—ranging from writing about trauma to describing a picture—whereas others came from fiction writing and natural conversations). Despite this, men and women used language in reliably and systematically different ways. Writing about a traumatic experience is very different from writing a class exam, but men and women wrote differently across both contexts. This mirrors the substantial intraindividual consistency in language use reported in earlier work (Pennebaker & King, 1999).

Second, it is important to note the context in which samples were collected. On the surface, the difference between using 14% pronouns and using 12% pronouns seems rather subtle. However, these differences are based on an average of 15 min of communication, comprising an average of 1,000 words. This means that women used 140 pronouns compared to men's 120 pronouns. These numbers translate into a difference of roughly 2 to 3 pronouns every minute. Thus, gender differences in written and spoken language appear to be subtle, but reliable. The fact that we are confronted with these differences every day yet fail to notice them highlights the degree to which they are a part of everyday life. At the same time, it is important to keep in mind that these differences are averages at the *population* level. The implication of this fact is that predictions about language use by *individuals* should be made cautiously, if at all.

Comparison With Previous Research

Many of the findings of our analyses echo those reported by earlier studies using more traditional text analysis methods; other findings clarify disputed questions in the literature. Table 4 summarizes the relations between previous data and these results. Past research has identified differences at four main levels of analysis—words, phrases, sentences, and overall messages. Because LIWC is a word-based text analysis system, the most direct correspondence to previous literature is at the word level, and these results are presented first. However, many phrase-level, sentence-level, and message-level features are associated with particular word choices; and several of LIWC's word categories serve as effective proxies. Human coders and LIWC may have used slightly different coding strategies, but the basic conclusions overlap.

Different words. Women's greater use of pronouns mirrored previous work, and the finding that women used more certainty words parallels earlier discoveries that women used more intensive adverbs (Biber et al., 1998; McMillan et al., 1977; Mehl & Pennebaker, 2003; Mulac et al., 2000). Successful replications for men's language included substantially greater use of numbers, articles, long words, and

TABLE 4
Comparisons Between Previous Findings and Present Results

Previous Findings	*Relevant LIWC Category*	*Present Results*
Different words		
F > M		
Pronouns	Total pronouns	✓ Replication
Intensive adverbs	Certainty words	✓ Replication
M > F		
Numbers	Numbers	✓ Replication
Articles	Articles	✓ Replication
Long words	Words ≥ six letters	✓ Replication
Swearing	Swear words	✓ Replication
Mixed		
First-person singular	First-person singular	F > M
Emotion words	Total emotion	F > M
	Negative emotion	F > M
	Positive feelings	F > M
	Anger	F = M
Different phrases		
F > M		
Polite forms	Discrepancies (e.g., could)	✓ Replication, small effect
Hedging phrases	Tentative (e.g., maybe)	✗ F = M
	Hedges (e.g., suppose)	✓ Replication
M > F		
Locatives	Space (e.g., above)	✗ M = F
	Prepositions	✓ Replication
Mixed		
Oppositions	Exclusive (e.g., but)	F = M
Justifiers	Causation (e.g., because)	M = F
Different sentences		
F > M		
Long sentences	Words per sentence	M > F, small effect
Negations	Negations (e.g., never)	✓Replication
M > F		
Word count	Word count	✗ F = M
Mixed		
Directives	Second-person pronoun	M > F, small effect
Questions	Question marks	M = F
Different messages		
F > M		
Personal concerns and interpersonal queries	Psychological words	✓ Replication
	Social process words	✓ Replication
M > F		
Information exchange	Numbers	✓ Replication
	Prepositions	✓ Replication
	Articles	✓ Replication
	Current concerns	✓ Replication

Note. "Small effect" refers to effect sizes of $d < .10$. LIWC = Linguistic Inquiry and Word Count; M = male; F = female.

swearing (e.g., Gleser et al., 1959; Mehl & Pennebaker, 2003; Mulac & Lundell, 1986). Reflecting the mixed bag of earlier work on emotional references, women use more affect words, but this was not restricted to positive emotions, as one earlier study suggested (Danner et al., 2001). Women were more likely than men to refer both to positive feelings and to negative emotions—specifically, sadness and anxiety (cf. Thomas & Murachver, 2001; Mulac et al., 1990). The previous finding of a male advantage in anger words was not replicated (Mehl & Pennebaker, 2003). The most striking discovery was that women, not men, were the more prolific users of first-person singular pronouns (i.e., I, me, and my).

Different phrases. LIWC's discrepancy category approximates the category of polite forms (e.g., "Would you mind if … ," "Should I get the door?"). Fitting with earlier work, we found a small but reliable tendency for such words to appear more often in women's texts (e.g., Holmes, 1995; McMillan et al., 1977). LIWC has two categories relevant to the critical linguistic area of hedging, but these yielded surprisingly different results. LIWC's new hedging category, combining "I" with a variety of verbs such as "guess," found the expected gender difference: Women were more likely to hedge (cf. Hartman, 1976; Mulac & Lundell, 1994; Poole, 1979). However, women were no more likely to use words from the tentative category (e.g., maybe, perhaps). The use of phrases, such as "I guess," may reflect previous findings that women use more polite forms (Holmes, 1995; McMillan et al., 1977), and are reluctant to force their views on other people.

Different sentences. It is difficult to construct a negation—a sentence of what something is not—without one of the words from LIWC's negations category (e.g., no, not, never). Fitting with the prior observation that women were more apt to make such statements (Mulac & Lundell, 1986; Mulac, Lundell, & Bradac, 1986), we found a small difference favoring women in this domain. Previous literature has found that in terms of words used, it was actually men who consumed more "airtime" (e.g., Dovidio et al., 1988). Our data found no evidence of any differences in overall word count. Finally, we failed to find any tendency for women to use question marks, contrasting with earlier reports that women asked more questions and inserted more tag questions into their sentence (e.g., McMillan et al., 1977; Mulac & Lundell, 1986).

Different messages. This is the most difficult level of analysis to relate to LIWC's word-count output because, by definition, it concerns what is implicit in language rather than what is contained in language's manifest features. However, it is informative to consider the types of topics that men and women use their words to talk about. This study provides strong evidence that women seem to have more of a "rapport" style, discussing social topics and expressing internal thoughts and feelings more often, whereas men "report" more often, describing the quantity and location of objects (e.g., Herring, 1993; Tannen, 1990). The absence of a difference in first-person plural may indicate that the word "we" is not a simple marker of a communal, interdependent mindset (cf. Brewer & Gardner, 1996), rather than indicating doubts about whether women really are rapport oriented.

CONCLUSION

Text analyses based on word count cannot, by their very nature, capture the context in which words are used. Interpreting the gender differences is clearly a nuanced matter. Part of our aim was to use LIWC technology to get a broader sample than any hand-coded study could ever manage. A qualitative investigation of the gender differences we have reported would be one useful avenue for future research. Such an investigation would allow for a more complete explanation of the ways in which social roles and relationships between speakers contribute to differences in language use. It must also be acknowledged that our data came from a pre-existing archive of texts that had either been collected in our own laboratory or had been volunteered by outside labs. However, the size and diversity of the dataset suggest that a more extensive sample would not have altered the overall findings.

Coates and Johnson (2001) pointed out that the study of language provides a uniquely "social" perspective on the study of gender differences. Given that our understanding of other human beings is heavily dependent on language, the average differences in communication style that we report are likely to play a central role in the maintenance of gender stereotypes and may perpetuate the perception of a "kernel of truth" that underlies those stereotypes. However, it is important to note that our analyses merely identify *how* men and women communicate differently, without addressing the issue of *why* these differences exist. Gender differences in language use likely reflect a complex combination of social goals, situational demands, and socialization—just to name a few—but these data do not identify these origins. Rather, our goal was to provide a clear map of the differences in men's and women's language, and to offer a starting point for future research into the nature and origin of gender differences.

Our analyses demonstrate small but systematic differences in the way that men and women use language, both in terms of what they say and how they choose to say it. Although our focus was more on function words than content words, it is clear that both types offer numerous opportunities for future research. By using a very large and diverse data corpus combined with a computerized text analysis program we were able to put the controversial topic of language-based gender differences on firmer empirical ground. Furthermore, our data support and clarify, rather than contradict, previous research, suggesting that word-count strategies are a viable, highly efficient alternative to linguistic analysis based on human coders. Computerized text analysis offers the statistical power and coding consistency that are ultimately essential for a complete answer to the questions that have captured the imagination of laypeople and scientists alike: when, where, why, and how do men and women talk differently?

ACKNOWLEDGMENTS

Preparation of this project was supported by grants from the National Institutes of Health (MH52391) and the Army Research Institute (W91WAW-07-C-0029). We thank Matthias Mehl and Barbara Luka for their helpful comments.

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.

Biber, D. (1991). *Variation across speech and writing*. Cambridge, England: Cambridge University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge, England: Cambridge University Press.

Bradley, P. H. (1981). The folk linguistics of women's speech: An empirical examination. *Communication Monographs, 48,* 73–90.

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.

Brownlow, S., Rosamon, J. A., & Parker, J. A. (2003). Gender-linked linguistic behavior in television interviews. *Sex Roles, 49,* 121–132.

Bucci, W., & Freedman, N. (1981). The language of depression. *Bulletin of the Menninger Clinic, 45,* 334–358.

Carr, C. (1996). *The Alienist*. New York: Random House.

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.

Coates, L., & Johnson, T. (2001). Towards a social theory of gender. In W. P. Robinson & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 451–464). New York: Wiley.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Colley, A., Todd, Z., Bland, M., Holmes, M., Khanom, M., & Pike, H. (2004). Style and content in emails and letters to male and female friends. *Journal of Language and Social Psychology, 23,* 369–378.

Crosby, F., & Nyquist, L. (1977). The female register: An empirical study of Lakoff's hypothesis. *Language in Society, 6,* 313–322.

Danner, D. D., Snowdon, D. A., & Friesen, W. V. (2001). Positive emotions in early life and longevity: Findings from the nun study. *Journal of Personality & Social Psychology, 80,* 804–813.

Dovidio, J. F., Brown, C. E., Heltman, K., Ellyson, S. L., & Keating, C. F. (1988). Power displays between women and men in discussions of gender-linked tasks: A multichannel study. *Journal of Personality & Social Psychology, 55,* 580–587.

Dubois, B. L., & Crouch, I. (1975). The question of tag questions in women's speech: They don't really use more of them, do they? *Language in Society, 4*(3), 289–294.

Eagly, A.H. (1995). The science and politics of comparing women and men. *American Psychologist, 50*, 145–158.

Gleser, G. C., Gottschalk, L. A., & John, W. (1959). The relationship of sex and intelligence to choice of words: A normative study of verbal behavior. *Journal of Clinical Psychology, 15,* 183–191.

Grafton, S. (1996). *K is for Killer.* New York: Ballantine Books.

Hartman, M. (1976). A descriptive study of the language of men and women born in Maine around 1900 as it reflects the Lakoff hypotheses in language and woman's place. In B. L. Dubois & I. Crouch (Eds.), *The sociology of the languages of American women* (pp. 81–90). San Antonio, TX: Trinity University Press.

Herring, S. C. (1993). Gender and democracy in computer-mediated communication. *Electronic Journal of Communication*, *3*(2). Retrieved June 3, 2003, from http://www.cios.org/getfile/ HERRING_V3N293

Holmes, J. (1995). *Women, men and politeness.* Harlow: Longman.

Koppel, M., Argamon, S., & Shimoni, A. (2003). Automatically catagorizing written texts by author gender.. *Literary and Linguistic Computing, 17,* 101–108.

Lakoff, R. (1975). *Language and woman's place.* New York: Harper Colophon Books.

McMillan, J. R., Clifton, A. K., McGrath, D., & Gale, W. S. (1977). Women's language: Uncertainty or interpersonal sensitivity and emotionality? *Sex Roles, 3,* 545–559.

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

Mulac, A., Bradac, J. J., & Gibbons, P. (2001). Empirical support for the gender-as-culture hypothesis: An intercultural analysis of male/female language differences. *Human Communication Research, 27,* 121–152.

Mulac, A., & Lundell, T. L. (1986). Linguistic contributors to the gender-linked language effect. *Journal of Language & Social Psychology, 5*, 81–101.

Mulac, A., & Lundell, T. L. (1994). Effects of gender-linked language differences in adults' written discourse: Multivariate tests of language effects. *Language & Communication, 14,* 299–309.

Mulac, A., Lundell, T. L., & Bradac, J. J. (1986). Male/female language differences and attributional consequences in a public speaking situation: Toward an explanation of the gender-linked language effect. *Communication Monographs, 53,* 115–129.

Mulac, A., Seibold, D. R., & Farris, J. L. (2000). Female and male managers' and professionals' criticism giving: Differences in language use and effects. *Journal of Language & Social Psychology, 19*(4), 389–415.

Mulac, A., Studley, L. B., & Blau, S. (1990). The gender-linked effect in primary and secondary students' impromptu essays. *Sex Roles, 23,* 439–469.

Mulac, A., Wiemann, J. M., Widenmann, S. J., & Gibson, T. W. (1988). Male/female language differences and effects in same-sex and mixed-sex dyads: The gender-linked language effect. *Communication Monographs, 55,* 315–335.

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.

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

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

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality & Social Psychology, 6,* 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*(3), 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 lifespan. *Journal of Personality & Social Psychology, 85,* 291–301.
Poole, M. E. (1979). Social class, sex, and linguistic coding. *Language and Speech, 22,* 49–67.
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.
Shapiro, D. (1989). *Psychotherapy of neurotic character.* New York: Basic Books.
Strainchamps, E. (1971). Our sexist language. In V. Gornick & B. K. Moran (Eds.), *Woman in sexist society* (pp. 240–250). New York: Basic Books.
Thomson, R., & Murachver, T. (2001). Predicting gender from electronic discourse. *British Journal of Social Psychology, 40,* 193–208.
Weatherall, A. (2002). *Gender, language, and discourse*. London: Routledge.
Weintraub, W. (1981). *Verbal behavior: Adaptation and psychopathology.* New York: Springer.