The Psychology of Word Use in Depression Forums in English and in Spanish: Testing Two Text Analytic Approaches

Nairan Ramirez-Esparza[1], Cindy K. Chung[2], Ewa Kacewicz[2], and James W. Pennebaker[2]

[1]University of Washington
I-LABS, Box 357988
Seattle, WA 98195
nairan@u.washington.edu

[2] University of Texas at Austin
1 University Station A8000
Austin, TX 78712
cindyk.chung@mail.utexas.edu
ekacewicz@mail.utexas.edu
pennebaker@mail.utexas.edu

Abstract

The present studies demonstrate two computerized approaches to examining the expression of depression on the Internet. Study 1 observed linguistic markers of depression in English and Spanish forums. English and Spanish posts by depressed (N=160) and non-depressed individuals (N=160) were collected from Internet forums using bulletin board systems (bbs). A computer program (LIWC2001) was used to compare linguistic categories across groups within the English and Spanish language. Study 2 analyzed the themes people use when talking about their depression in forums in English and Spanish. Posts by English speaking (N=404) and Spanish speaking depressed people (N=400) were collected from bbs. An automated inductive approach to content analysis was used to compare how the expression of depression varies across cultures. The results for Study 1 showed that linguistic cues previously associated with depression were found to be higher in depressed than in non-depressed posts in English and Spanish. Study 2 showed that depressed people who wrote in Spanish were more likely to mention relational concerns than depressed people who wrote in English, who were more likely to mention medicinal concerns. Implications for informational and support needs and concerns in psychotherapy vs. online support across cultures are addressed and discussed.

Introduction

A growing number of people are turning to Internet forums for self-help, especially for stigmatizing illnesses such as depression (Baker et al. 2003; Berger, Wagner, and Baker 2005; Fox and Fallows 2003). Forums provide an anonymous and inexpensive way for large numbers of people with depression to get information and support, even for those who live in geographically remote areas (Houston, Cooper, and Ford 2002; Stjernsward and Ostman 2006).

Recent advancements in computerized text analyses have allowed researchers to explore naturalistic online behavior by examining the language of users. For example, Owen and colleagues (2003) used a word count and categorization software, Linguistic Inquiry and Word Count (LIWC, Pennebaker, Francis, and Booth 2001) to analyze emotional and cognitive expression in Internet cancer support groups to assess coherence, disclosure, and insight gained from online participation. In a LIWC analysis of emotional expression in online support groups for women with breast cancer, Lieberman and Goldstein (2006) found that women who used more anger words improved in their health and quality of life, whereas women who used more anxiety words experienced increased depression. Lyons, Mehl, and Pennebaker (2006) found that relative to recovering anorexics, pro-anorexics were characterized by more positive emotions, less anxiety, and a lower degree of cognitive reflection and self-focused attention. In sum, using an automated word counting approach has been shown to be an efficient way to characterize the language of online groups in theoretically and psychologically meaningful ways.

Forums can be public or private (where private login access is required), and are often supervised by a moderator for offensive content. A series of replies to a post, akin to a conversation log, is called a thread. Threads are available to readers browsing or participating on the forum. This archival feature makes bbs a great place for writers to refer to information already addressed in the forum, and to determine if it is a forum in which they would like to participate. For researchers, the archival features of bbs make them an ecologically valid source for collecting and

examining naturalistic behaviors. The current study sought to examine the linguistic markers of depression online (Study 1), and to explore the topics that naturally arise on Internet forums for depression (Study 2) across languages.

In Study 1 we compared language use in depressed and non-depressed people from bbs in English and Spanish. In this study we used LIWC (Pennebaker, Francis, and Booth 2001) and its Spanish translated version (Ramirez-Esparza et al. 2007) to analyze linguistic markers of depression.

In Study 2 we analyzed the themes people use when they talk about their depression in bbs in English and in Spanish. In this study we used a recently developed procedure that analyzes language in an inductive way. This approach is called the Meaning Extraction Method (Chung and Pennebaker in press) and it uses computerized text analytic tools to determine the most frequently occurring words in a corpus. First, the most frequently occurring words are determined across texts. Then, just as a factor analysis reduces items on a questionnaire down to a few factors made up of items that go together, the MEM uses factor analysis to group the most frequently occurring words into factors made up of words that co-occur in the same text. The resultant factors tend to be coherent and intuitively comprehensible groups of words.

Study 1: Linguistic Markers of Depression

Previous studies using LIWC have found that it is possible to characterize depression through natural language use. LIWC is a computerized word counting tool. LIWC searches for approximately 2000 words and word stems, and categorizes them into linguistic (e.g. articles, numbers, pronouns), psychological (e.g. cognitive mechanisms, emotions, social), or content (e.g. death, home, school) categories. Results are reported as a percentage of words in a given text file, indicating the degree to which a particular word category was used in a given text. The words in LIWC categories have previously been validated by independent judges, and the use of the categories within texts have been shown to be reliable markers for a number of psychologically meaningful constructs (Pennebaker, Mehl, and Niederhoffer 2003).

Using LIWC2001, Stirman and Pennebaker (2001) found that suicidal poets were more likely to use first person pronouns (e.g., I, me, mine) and less first plural pronouns (e.g., we, ours) throughout their writing careers than were non-suicidal poets. These findings supported the social engagement/disengagement model of depression, which states that suicidal individuals have failed to integrate into society in some way, and are therefore detached from social life (Durkheim 1951). Similarly, Rude, Gortner, and Pennebaker (2004) found that currently depressed students used more first person singular pronouns, more negative emotions words, and slightly fewer positive emotion words in their essays about coming to college, relative to students who had never experienced a depressive episode. These results are in line with Pyszcynski and Greenberg's (1987) self-awareness theory of depression, which theorizes that high levels of self-focus lead to negative emotions, self-blame, and, consequently, less attentional resources. Note that clinicians have also reported on the greater use of the pronoun "I" in depressed patients (Bucci and Freedman 1981; Weintraub 1981).

Study 1 Overview

Depressive language has been studied in written text (Rude, Gortner, and Pennebaker 2004, Striman and Pennebaker 2001); however, less is known about the language of depression online. Furthermore, there is even less research exploring the similarity of depression language between English and Spanish. The present study was conducted to explore the linguistic markers used by people who write about their depression in bbs in English and in Spanish. In order to accomplish this, we initially compared linguistic markers used in English internet depression forums with a control group. Breast cancer bbs were selected as our control group. This group was chosen because we should expect that women in these forums are not necessarily depressed, but at the same time, use these forums to express their experience of coping with their illness, as do depressed individuals in depression forums. For instance, women in the breast cancer forums express their emotions and talk about how the cancer has affected their social and psychological lives. This first step allowed us to isolate specific linguistic markers associated with depression. The second step sought to examine language differences between depressed and non-depressed women in Spanish online forums.

Study 1: Method

Data Collection. Posts from depression forums in English and Spanish were collected for text analyses. Breast cancer forum posts in English and Spanish served as a comparison sample for the characterization of depressive language, since breast cancer is a personal topic. All collected posts were from women.

Threads were included only if: a) the posts were from a website in which private login access was not required, b) the posts were at least 100 words long, c) there was enough information on the forum to know that the writer was a women d) the post was written well enough to be judged as written by a native English speaker for the English sample, and a native Spanish speaker for the Spanish sample, e) the post was written by someone who seemed to be interested in the forum because the forum was relevant to the writer's self concerns, and not out of concern for someone they knew, f) only a single post from a screen name was collected to avoid multiple entries from a single user, and g) for the breast cancer group, posts were excluded if the writer explicitly indicated that they were experiencing depression. Each post was spell checked and converted into a single text file.

Text Files. A total of 320 posts were collected. Half of these were in Spanish (160) and half were in English (160). Similarly, half of the posts were from depression forums

(80) and half from the control group (80). The mean word count for forums in English was 229.00 (*SD= 115.54*) for the depression group and 246.88 (*SD = 105.55*) for the breast cancer group. The mean word count for forums in Spanish was 263.78 (*SD= 116.13*) for the depression group and 253.26 (*SD = 121.96*) for the breast cancer group. There were no significant differences for word count across languages.

Data Analysis. The text files were analyzed with Linguistic Inquiry and Word Count (LIWC, Pennebaker, Francis, and Booth 2001). LIWC is a computerized tool that uses a dictionary to count the frequency of words and word stems in a category. LIWC computes the degree to which a category was mentioned in a text file, as a percentage of the total number of words in that text file. In this study, for example, LIWC category percentages indicate the degree to which a particular word category was used in a given post. Recently the English dictionary of LIWC was translated into Spanish (Ramirez-Esparza et al. 2007). The Spanish version of LIWC is equivalent to the English LIWC; however, there are some categories that differ across languages. For example, the English LIWC counts more first person singular words than the Spanish LIWC. This is because in Spanish the pronoun is dropped most of the time. Thus, even though both the English and Spanish LIWC versions have the same categories and subcategories of words, there are subtle differences across versions caused by differences in the language and translation biases. Due to these biases, this study first compared linguistic differences within the English language and then examined whether the same pattern of differences occurred within the Spanish language. No differences were analyzed across languages.

In sum, the English depression and breast cancer posts were processed using the English LIWC2001 and the Spanish depression and breast cancer posts were processed with the Spanish version. Mean levels of linguistic categories previously found to be associated with depression were compared between groups in both languages. The linguistic categories examined were: first person pronouns, first person plural, negative emotion words, and positive emotion words.

Study 1: Results

Table 1 shows the mean and standard deviation for both depression forums and breast cancer forums in English and Spanish. The last two columns show the two-tailed t-values for the main effects. In order to avoid Type 1 error, .01 was used as the significance level.

As found in previous research on depression in written language, online depressed writers used significantly more 1st person singular pronouns, less first person plural pronouns in both the English and Spanish forums. These findings supported the social engagement/disengagement model of depression (Durkheim 1951), which states that depressed individuals tend to focus in themselves and detach from others. In addition, women from depressed forums used less positive emotion words and more negative emotion words than women from breast cancer forums in English and Spanish. These findings were in line with Pyszcynski and Greenberg's (1987) self-awareness theory of depression, which states that depressed individuals tend to focus more on negative aspects of their life.

	Forums				Main Effects	
	English		Spanish		Eng.	Spa.
Categories	Dep.	Breast Cancer	Dep.	Breast Cancer		
	N=80 Mean (SD)	N=80 Mean (SD)	N=80 Mean (SD)	N=80 Mean (SD)	t-value	t-value
First person singular	12.24 (2.97)	4.03 (3.01)	9.30 (2.34)	5.03 (2.76)	-17.39*	-10.54*
First person plural	.18 (.33)	.72 (1.06)	.22 (.39)	1.02 (1.28)	4.36*	5.32*
Positive Emotions	1.72 (1.14)	2.54 (1.72)	2.99 (1.36)	3.53 (1.93)	3.56*	2.04+
Negative Emotions	3.34 (1.48)	1.18 (1.06)	3.64 (1.56)	1.93 (1.23)	-10.59*	-7.69*

Table 1. Differences in Linguistic Categories between Depression and Breast Cancer Forums in English and in Spanish. Note: Dep. = Depression; Eng. = English, Spa = Spanish; t-tests are based in 158 degrees of freedom; + p< .05, * p < .001

Study 1: Conclusion

Online depression groups used similar language to what has previously been found to correlate with depression in lab- and clinical-based studies. Interestingly, Spanish-speakers also used language in similar ways as English-speakers. One of the limitations of this study is that it was not possible to compare differences across languages. This is because the Spanish LIWC is based on translation and is not exempt from cross-language biases. Other limitations of this study are that the sample was based only on women and other demographic information was not collected (e.g., age). Finally, the sample was relatively small. Study 2 attempted to address each of these issues. Specifically, we introduce an inductive method that extracts the themes people use when they talk about their depression. Furthermore, this method allows us to observe qualitative differences in themes across languages. Finally, in Study 2, significantly more posts of depressed individuals were collected and other demographic information was coded.

Study 2: Themes Depressed People Post About

Previous studies on language use in depressed people have focused on the linguistic style of depressive language, as opposed to analyzing *what* people talk about. While the linguistic style approach has many promising features that

can be generalized across topics, discovering the actual topics of conversation in a particular sample has potentially far-reaching implications. This can clue us into the informational and support needs that people naturally seek online from different groups.

Traditional content analyses have explored corpuses of text by coding for particular topics with modest rates of inter-reliability, much time and effort involved in coding, and with subjective decisions about the level of specificity desired in coding for topics. While a full range of labor-intensive topic sampling can be beneficial, what is ultimately useful to the researcher is a summary of topics that spread across a wide portion of the corpus. With recent advancements in computerized text analyses, it is now possible to to summarize the main topics in a corpus quickly and reliably.

Our lab has recently developed the Meaning Extraction Method (Chung and Pennebaker in press). The MEM uses automated text analytic tools to identify the most commonly used content words in written text, and then to determine how these words co-occur in a given text. To appreciate how the MEM works, imagine that 500 American students are asked to describe who they are. Let's assume that each personality self-description takes about 15 minutes, and the average person generates about 300 words. Although this task will generate 150,000 different words, there will probably only be about 3,000 different words used. Once the standard function words are removed (e.g., articles, auxiliary verbs, prepositions, pronouns, etc.) as well as words not used by at least 3% of the participants, perhaps only 100 different words will remain. Imagine that we now go back through each of the 500 essays and determine if each essay either did (coded 1) or did not (coded 0) use each of the 100 words using a binary approach. The resultant matrix is made up of 500 essays in rows and 100 words in columns.

The 500 x 100 matrix of 1's and 0's can now be subjected to a traditional factor analysis. Although Chung and Pennebaker (in press) used principal components analysis with varimax rotation, most factor analytic methods produce comparable results. The number of optimal factors that emerge is determined by the scree plot method. Based on previous work, the number of dimensions will probably range between 5 and 9. This number is much smaller than the hundreds or so dimensions that are typically extracted through similar data reduction methods for text, such as LSA (Foltz 1966). Rather than using the word factors to precisely disambiguate word meaning, identify authorship, or to classify documents (Griffiths, Steyver, and Tennenbaum 2007), the primary aim of the MEM is to describe a few broad topics in a corpus.

By relying on factor analyses, one can determine the degree to which any group of words tends to co-occur. Unlike a simple co-occurrence matrix, however, the factor analysis tells us which words make up coherent clusters. Each word cluster, then, is essentially a meaning or theme-based cluster. For example, a person who uses the word *smart* may also use words like *quick, intelligent*, and *thoughtful*. Another person may use some of the same terms or other synonyms. The strength of this approach is that all of these terms will likely yield a factor that we would define as a theme for intelligence.

There are two powerful advantages of this method that are particularly well-suited for cross-cultural research. First, the emergent dimensions are purely inductive. The words are not dependent on judges' ratings, or pre-defined by domains created by culturally biased investigators. The method, then, avoids imposing constructs, theories or list of terms onto texts. By analyzing open-ended depression texts, this method is able to capture natural connections among words found in the posts, rather than grouping words into some pre-selected theme or dimension. Second, the method does not rely on translation. The MEM simply looks at words – a group of letters separated by blank spaces or punctuation marks. The MEM analyzes how groups of words are statistically clustered together. Translation issues are only relevant at the end of the process (Ramirez-Esparza 2007).

Study 2: Method

Data Collection. This study employed the same procedure for data collection as Study 1. However, this time we collected postings that provided estimations of the participants' age category. Each post was spell checked and converted into a single text file. Five Depression forums were used to collect the English posts, and two depression forums were used to collect the Spanish posts.

Text Files. Our target sample included 404 English posts and 400 Spanish posts from depression forums. Table 2 shows the demographic information collected from the posts.

	Sex		Age		
Forum	Female	Male	<25	25 to 45	>45
English	290	114	168	214	22
Spanish	252	148	166	207	26

Table 2. Demographic Information for English and Spanish Depression Forum Samples Note. Cell numbers indicate frequencies. Age, <25 = younger than 25 years, 25 to 45 = 25 to 45 years old, >45 = older than 45 years.

Data Analysis. The MEM can be thought of as a factor analysis of words across a large number of texts. As with all factor analyses, the technique provides a way of determining how groups of items (in this case, words) naturally covary. Each language sample was analyzed separately.

All words in each language sample were ranked according to frequency using WordSmith (Scott 1996). Function words (i.e. pronouns, prepositions, articles, conjunctions, and auxiliary verbs) were not considered for analyses. A dictionary including approximately 200 of the

most frequently used words and their root forms were created for each language. With the English dictionary, LIWC was used to process the occurrence of the 197 words in the English depression posts. With the Spanish dictionary, LIWC was used to process the occurrence of the 195 words in the Spanish depression posts. The outputs were a 197 word by 404 text file matrix for English, and a 195 by 400 text file matrix for Spanish. Each matrix was made up of 1's and 0's, representing whether or not a post used that particular word.

Principal components factor analyses were run using varimax rotation for each matrix. A scree of the Eigenvalues (Cattell 1966, Stevens 1992) suggested that about 5 factors be extracted for each sample. Each factor represented a group of content words that tended to co-occur within each language sample.

Study 2: Results

Tables 3 and 4 show the groups of words that clustered together to form coherent themes. In both languages 5 factors were found. From these three were similar across cultures: Treatment, Family, and School. However, slight differences appeared in how the themes were expressed. For example, Treatment for the English sample included more specific terms relating to side effects and medication (e.g. *effect, experience, side*) than in the Spanish sample. The School factor for the English sample was associated with relationships, while the School factor for the Spanish sample was associated with feelings of insecurity (e.g. *timid, incapable, foolishness*). Family for the English sample was largely made up of specific roles (e.g. *mom, daughter, child*), whereas Family for the Spanish sample was tied to specific issues, as indicated by the inclusion of terms such as *money, studies,* and *suicide.*

Themes that were unique to English were Disclosure and Symptoms. Disclosure was associated with sharing feelings with others (e.g. *feel, happy, people, tell*). The Symptoms factor detailed sleep (e.g. *night, sleep, wake*) and eating habits (e.g. *eat, food, weight*), both of which are known to change during depressive episodes (DSM-IV, 1994).

Themes unique to Spanish were Relationship History and Emotional Event. The Relationship factor was associated with the history of a specific relationship, as indicated by relationship and temporal words such as *boyfriend, end, relationship, love,* and *met.* The Hopelessness factor described an event (e.g. *eyes, night, second, die*) with feelings (e.g. *heart, hope, peace*).

Overall, these results suggest that depression forums in English were more concerned with medicinal questions and concerns, whereas the Spanish depression forums were used for sharing and disclosing information about relational concerns. Note that each of the Spanish themes were tied to emotions or relationships, whereas a couple of the English factors were more concrete and descriptive (i.e. Family, Symptoms, and School).

FACTOR 1: Treatment		**FACTOR 1: Disclosure**		**FACTOR 3: Family**	
Medication	.62	Tell	.43	Mom	.49
Effect	.46	People	.41	Daughter	.48
Depression	.43	Know	.39	Child	.48
Side	.35	Happy	.35	Family	.48
Week	.34	Talk	.35	Brother	.43
Therapy	.34	Feel	.34	Sister	.43
Suffer	.34	Want	.34	Dad	.41
Disorder	.33	Suppose	.33	Son	.33
Doctor	.33	Read	.32	Love	.33
Antidepressant	.32	Hurt	.32	Girl	.33
Experience	.32	Wrong	.32	Young	.32
Major	.32	Emotional	.31	Parent	.32
Mental	.31	Mind	.31	House	.31
Psychiatrist	.31	Sad	.31	Husband	.30
		Make	.31	Crazy	.30
		Hate	.31		

FACTOR 4: Symptoms		**FACTOR 5: School**	
Sleep	.51	Constant	.46
Hour	.48	Relationship	.45
Food	.44	School	.40
Wake	.44	High	.40
Morning	.44	Lack	.39
Night	.41	University	.38
Bed	.39	Social	.38
Stay	.38	College	.36
Weight	.37	Move	.32
Eat	.36	Friend	.32
Place	.32	Girlfriend	.32
		Class	.31

Table 3. Factor Word Loadings in English Depression Posts. Note: Loadings greater than .30 are reported.

General Discussion

With the growing use of the Internet as a resource for help with depression, it is important to examine the naturalistic behavior of online depression groups. Study 1 demonstrated that linguistic markers differ across depressed and non-depressed individuals in English forums. The same pattern of differences was found within Spanish-speaking groups. Depressed people used more 1st person singular pronouns and negative emotion words, and less 1st person plural pronouns and positive emotions than non-depressed people. In Study 2, using text analytic tools, we were able to identify the key words and themes used by depressed people across cultures. The MEM revealed that depressed people who wrote in Spanish were more likely to mention relational concerns than depressed people who wrote in

English, who were more likely to mention medicinal concerns. Possible moderators for depression forums or more controlled online satellite sites for outpatients might include pharmacists for English speaking samples, and relationship counselors for Spanish speaking samples. With analyses of ecologically valid contexts that depressed people have selected, we were able to broaden our understanding of the expression of depression in daily life.

FACTOR 1: Family		**FACTOR 2: Relationship History**	
MADRE/mother	.48	RELACION/relationship	.51
HERMANO/brother	.46	ENAMORADO/love	.48
ABUELA/grandmother	.44	CONOCI/met	.45
PADRES/parents	.44	HABLAR/talk	.45
PAPA/father	.43	CHICO/guy	.44
HORRIBLE/horrible	.39	AMIGOS/friends	.44
CASA/house	.39	NOVIO/boyfriend	.43
SUICIDIO/suicide	.36	JUNTOS/together	.37
DINERO/money	.34	ESPECIAL/special	.36
ESTUDIOS/studies	.33	EMPECE/start	.33
CLASE/class	.32	TIEMPO/time	.33
FEA/ugly	.32	DEJAR/leave	.33
ASCO/disgust	.32	FINAL/end	.33
COMER/eat	.31	HISTORIA/history	.31
PEQUENIOO/small	.31	MESES/months	.31
FAMILIA/family	.31	LLEGAR/arrive	.31
HIJOS/sons	.30		
FACTOR 3: Hopelessness		**FACTOR 4: School**	
NOCHE/night	.45	TIMIDA/timid	.41
SEGUNDO/second	.44	CARRERA/career	.41
MORIR/die	.42	COLEGIO/school	.40
PAZ/peace	.39	CONFIANZA/trust	.37
OJOS/eyes	.39	ESTUDIOS/studies	.37
ESPERO/hope	.38	INCAPAZ/incapable	.33
TERRIBLE/terrible	.36	UNIVERSI./university	.33
FUERTE/strong	.34	TONTERIA/foolishness	.32
CORAZON/heart	.31		
SUENIOS/dreams	.31		
FACTOR 5: Treatment			
PSICOLOGO/psychologist			.40
ANSIEDAD/anxiety			.40
EMPRESA/company			.39
ANTIDEP/antidepressants			.39
SINTOMAS/symptoms			.34
MEDICAMENTO/medicines			.30

Table 4. Factor Word Loadings in Spanish Depression Post. Note: Loadings greater than .30 are reported.

Limitations

This non-reactive and cross-sectional sampling procedure enabled us to bypass some of the response rate and selection biases that result when forum participants are invited to participate in a survey. However, the tradeoff is that those studies allowed for the potential assessment of actual levels of depression, diagnostic information, and comorbidity with other disorders or illnesses.

Although significant differences were found between depression and breast cancer in terms of various word categories, especially with those previously found to indicate depression, the breast cancer sample may not have served as the best comparison sample. It is possible that some people in the breast cancer sample were also depressed. It is difficult to know from the current study, whether or not depressed people only use depressive language with a particular audience, or if they could be identified when not addressing other people suffering from depression. In addition, it is not known how the features of depressive language might be different on public vs. private (logon and password required) forums.

Future Directions

The goal of future research might be to analyze the language use of those who experience short term or long term relief of depression through online forum participation. Examining the natural behavior and topics discussed in online forums might be the first step in understanding informational and support needs across various patient populations and cultures.

Conclusions

The Internet serves as an excellent source for studying the naturalistic daily behaviors and concerns of populations that we normally only get to observe in small numbers in practice or in the laboratory. Interestingly, our study found that the features of language that characterize depression online are similar in social media and nonsocial media text. In addition, both the writing style and content were similar to those found in lab and clinical based studies. The text analytic strategies used in the present paper allowed us to characterize depressive language online, and to examine themes naturally mentioned across cultures. The examination of language markers and topics of interest to other illness populations and cultures are recommended.

References

Baker, L., Wagner, T. H., Singer, S., and Bundorf, M. K. 2003. Use of The Internet and E-Mail for Health Care Information. *Journal of the American Medical Association* 289: 2400-2406.

Berger, M., Wagner, T. H., and Baker, L. C. 2005. Internet Use and Stigmatized Illness. *Social Science and Medicine* 61: 1821-1827.

Bucci, W., and Freedman, N. 1981. The Language of Depression. *Bulletin of the Menninger Clinic* 45: 34-358.

Catell, R. B. 1966. The Scree Test for the Number of Factors. *Multivariate Behavioral Research* 1: 245-276.

Chung, C. K., and Pennebaker, J. W. In Press. Revealing Dimensions of Thinking In Open-Ended Self-Descriptions: An Automated Meaning Extraction Method for Natural Language. *Journal of Research in Personality*.

Diagnostic and statistical manual of mental disorders (DSM-IV). 1994. Washington, D.C.: American Psychiatric Association.

Durkheim, E. 1951. *Suicide.* New York, NY: Free Press.

Foltz, P. W. 1996. Latent Semantic Analysis For Text-Based Research. *Behavior Research Methods, Instruments, & Computers , 28*, 197-202.

Fox, S., and Fallows, D. 2003, July 16. Internet health resources. Retrieved June 03, 2007, from http://www.pewinternet.org/pdfs/PIP_Health_Report_July_2003.pdf

Griffiths, T. L., Steyver, M., and Tennenbaum, J. B. 2007. Topics in Semantic Representation. *Psychological Review*, 114: 211-244.

Houston, T. K., Cooper, L. A., and Ford, D. E. 2002. Internet Support Groups for Depression: A 1-year prospective cohort study. *American Journal of Psychiatry* 159: 2062-2068.

Lieberman, M. A., and Goldstein, B. A. 2006. Not All Negative Emotions are Equal: The Role for Emotional Expression in Online Support Groups for Women with Breast Cancer. *Psycho-Oncology* 15: 160-168.

Lyons, E. J., Mehl, M. R., and Pennebaker, J. W. 2006. Pro-Anorexics and Recovering Anorexics Differ in their Linguistic Internet Self-Presentation. *Journal of Psychosomatic Research* 60: 253-256.

Owen, J. E., Yarbrough, E. J., Vaga, A., and Tucker, D. A. 2003. Investigation of the Effects of Gender and Preparation on Quality of Communication in Internet Support Groups. *Computers in Human Behavior*, 19: 259-275.

Pennebaker, J. W., Francis, M. E., and Booth, R. J. 2001. *Linguistic Inquiry and Word Count: LIWC 2001*. Mahwah, NJ: Erlbaum Publishers (www.erlbaum.com).

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

Pyszcynski, T., and Greenberg, J. (1987). Self-Regulatory Perseveration and the Depressive Self-Focusing Style: A Self-Awareness Theory of Depression. *Psychological Bulletin* 102: 122-138.

Ramirez-Esparza, N. 2007. Cross-Cultural Constructions of Self: Americans and Mexicans. Ph.D. diss., Dept of Psychology, University of Texas, Austin, TX.

Ramirez-Esparza, N., Pennebaker, J. W., Garcia, A. F., and Suria, R. 2007. La Psicología del Uso de las Palabras: Un Programa de Computadora que Analiza Textos en Espaniol [The Psychology of Word Use: A Computer Program that Analyzes Texts in Spanish]. *Revista Mexicana de Psicologia* 24: 85-99.

Rude, S. S., Gortner, E. M., and Pennebaker, J. W. 2004. Language Use of Depressed and Depression-Vulnerable College Students. *Cognition & Emotion* 18: 1121-1133.

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

Stevens, J. P. 1992. *Applied Multivariate Statistics for the Social Sciences (2nd ed.).* Hillsdale, NJ: Erlbaum.

Stjernsward, S., and Ostman, M. 2006. Potential of E-Health in Relation to Depression: Short Survey of Previous Research. *Journal of Psychiatric and Mental Health Nursing* 13: 698-703.

Stirman, S. W., and Pennebaker, J. W. 2001. Word Use in the Poetry of Suicidal and Non-Suicidal Poets. *Psychosomatic Medicine* 63: 517-522.

Weintraub, W. 1981. *Verbal Behavior: Adaptation and Psychopathology*. New York, NY: Springer.