

Otilia Pacea¹

CorpTweet: Brands, Language and Identity in Web 2.0²

Abstract: Microblogging represents the largest trend in social media that recently emerged, spearheaded by its most popular service, Twitter. Neither companies nor organizations could have missed the opportunity to spawn their plenipotentiary voices and establish a safe social presence on the web, converging public and corporate interest into social brands. The microblogging phenomenon of Twitter has re-accommodated interpersonal interaction online, from self talk (diaristic posts), direct talk (replies), and indirect talk (talking to someone in front of everyone) to public talk. The currency is a body of words, the most difficult to control being the function or junk words (pronouns, articles, prepositions, conjunctions, auxiliary verbs) as markers of social engagement. This paper applies computerized corpus-based text analysis of style-related words in tweets from the most active social brands. My hypothesis is that universal patterns of social behaviour can be identified, which may further brand microblogging as a form of construction and co-construction of identity online.

Key words: electronic discourse; interpersonal communication; social brands; tweets; corpus-based analysis

1. Introduction

The advent of the social media represents the most significant paradigm shift since the Industrial Revolution – the Social Media Revolution. To the internet novice, social media are the brainchild of the postmillennial digiterati, to the internet savvy, however, social media are easily traceable to CompuServe and other online service providers and pioneers of social networking in the late 1970s. Microblogging is the largest trend in social media since the evolution of Web 2.0 that recently emerged two years ago, spearheaded by its most popular service, Twitter. We cannot ignore, for instance, that Ashton Kutcher and Ellen DeGeneres have more Twitter followers than the entire populations of Ireland, Norway, and Panama; or that eighty percent of Twitter activists use their mobile phones to

¹ Ovidius University of Constanța, Romania

² This work is part of the research project “Translations of American Myths, Icons and Brands in Post-Communist Romanian Culture (TRANSMIT)”, supported by CNCSIS-UEFISCSU, Project number PNII – IDEI – 802/2009.

post updates anywhere anytime, which is no plain sailing for customer satisfaction (Qualman 2009).

Brands are ubiquitous in global popular culture today. Neither companies nor organizations could have missed the opportunity to establish a safe social presence on the web, converging public and corporate interest into social brands. The third wave approach in media studies that brands the popularity of hybrid forms of transmedia storytelling which is exemplified by the stories that unfold across multiple platforms (film, video games, etc.) simultaneously such as the Matrix, Survivor and American Idol has brought forth a new cultural logic. The increasingly active and creative engagement of the consumer/audience/readership in the process of production empowers the individual more than ever before. The story remains the same and yet the voices of the storytellers are becoming subtler in the postmodern arena of cyberspace that in the case of the most popular microblogging service, Twitter, is technically rooted in the countless number of mobile and website plug-ins, which enable users to connect and publish Twitter messages through other websites, applications or mobile devices.

The idea of microblogging and the sense of exceptional customer service are embedded in corporate culture. Corporations in general use microblogging to interact with current or potential customers via the electronic route. It offers an unprecedented way to get involved in the community and to bring companies to life. A growing number of companies are keeping track of what is said about their brands on Twitter: from Comcast, Dell to General Motors and Kodak, they have all acknowledged the fact that the control of the brand has transferred from the hands of its producers to those of its consumers, due to the latest technological developments. Conversely, they use their own microbloggers to build trust and brand engagement. The personality of the microblogger is intrinsically linked to the brand they write about in that individuals can re- and co-construct the identity of the brand. It is no surprise that many argue that microblogging can humanize the brand so that it might appear as if “talking to Zappos is like talking to a friend that happens to sell shoes” (Balwani 2009). Corporate microblogging thus involves individuals tweeting as brands, boosting or undermining brand reputation in the way they choose to use language.

Twitter is a pioneer of microblogging whose popularity lies in its simplicity. Unlike traditional blogs, microblogs limit the number of characters per post producing more of a message rather than an article. The word count for a message on Twitter which is known as a *tweet* is set to 140 characters, often including links to webpages, photos or videos. One can follow updates in others’ Twitter streams, retweet or republish updates from others’ stream in one’s own stream, converse with others, either in public Twitter streams or through private direct messaging. In other words, microblogging can be a form of *Instant Messaging*, adding a social aspect.

With limited linguistic interaction, in size but not at all in effect, Twitter re-accommodates interpersonal interaction online, from self talk (diaristic posts), direct talk (replies), indirect talk (talking to someone in front of everyone) to public talk. Among other internet genres, the internet linguistic variety of microblogging shares more speech-like characteristics, thus not at all spoken language written down but rather written language substantially drawing from speech – an unexpected linguistic gold mine for the researcher interested in *naked* and pure forms of writing that became impossible to recover after standardizing processes and the institutionalization of the standard language.

In recent years, more specifically since the ninth decade of the twentieth century, there has been a surge of interest in empirical, corpus-based linguistics. Due to the large availability of technological tools and statistical methods that were adopted for linguistic research, it has become commonplace for linguists of all kinds and tastes to classify words not only on morphological, syntactic or semantic criteria, but also on the grounds of their co-occurrence with other words. With the advent of the internet, a pragmatist approach is inevitable in the context of the recent emphasis on numerical evaluation and concrete deliverables. Natural language use in the real world has become widely accessible and collectable on the internet, otherwise cumbersome to process. It is after all the same converging argument that is transferred from technology to language research, whereby social sciences and linguistics have crossed paths recently, in an effort to better understand human psyche. Words people use in everyday interactions can offer windows into their own mind and action and instruct us more efficiently on whom they are and what they do. Computer-based text analysis methods and natural language data accessibility help determine the association between language and personality. It is for the first time in the history of language studies that in this way we can link daily word use to social behaviour.

2. Research Hypothesis

Previous research in the psychology of word use confirmed that words people use in routine exchanges represent triggers of psychological processes (Pennebaker 2002; Pennebaker, Lay 2002; Stone, Pennebaker 2002; Campbell, Pennebaker 2003; Slatcher, Pennebaker 2006; Slatcher et al. 2007; Chung, Pennebaker 2007; Tausczik, Pennebaker 2010). Word choice offers a key into their personalities, emotional states, social connections and thinking styles.

With the advent of modern computerized tools, more efficient and successful methods were developed that allow us to examine discourse at a microscopic level. Function words or particles a.k.a. *junk words* fall into this category. They include pronouns, articles, prepositions, conjunctions or auxiliary verbs, in general small words that typically hold together more content-heavy nouns or verbs. Consequently, they reflect the style rather than the content of the communication, separating what people are communicating in tweeting from how they are communicating. In general, linguistic content is conveyed through the use of nouns, verbs, adjectives, adverbs. Language style can be recovered from the close analysis of such function, style-related or simply junk words, which although very few in comparison to content words, record higher frequencies of use (50-60% of the words we use are style words). Their social function is well-grounded in that they require a shared understanding between the addressor and the addressee.

Based on the Linguistic Inquiry and Word Count application and the meaning extraction method that was recently built to investigate Twitter feeds (Pennebaker, Wilson and Booth 2009), I hypothesized that the personality of the individual who tweets in the name of the brand can influence the social construction of the brand. The three language styles that were identified from the analysis of function words in the tweets of the most active social brands – the emotional style, the social style and the thinking style – can project various patterns of

social behaviour – the emotional bird, the social realist and the thinking machine – which may further *brand* corporate microblogging as a new form of construction and co-construction of identity that can be used as a marketing tool.

3. Data Collection

In collecting language data for the corpus of social brand tweets, I took Chafe's advice who back in the 1990s looked forward to the day when linguists of all types (introspective, experimental or corpus linguists) would be more versatile in their approaches and as a result, use a variety of methodologies and techniques (Chafe 96). In classifying the corpus, at least three criteria can be found relevant to the corpus-oriented analysis: in relation to the medium of communication, the corpus of social brand tweets is an electronic corpus, neither spoken nor written, *spoken language written down* – a fourth category that adds to the three categories included, viz. spoken, written and mixed. Equally relevant is the criterion of genre that in this case introduced the internet genre on the agenda of the language researcher. In what concerns the open-endedness criterion, the corpus of social brand tweets remains a closed, unalterable corpus, although I cannot ignore, in fact I encourage, its potential for continuity.

The research was first conducted on the Corpus of Social Brand Tweets (CorpTweet) which was collected in August 2010 and consists of almost 600 English-language tweets originating from global brands on Twitter. I sampled naturally occurring language data from the most active social brands on Twitter, which I identified based on various classifications that have been proposed starting with 2008³. I arranged them in five classes according to the functional criterion of the field they represent such as car industry, food and beverages, media and technology, travel and entertainment.

Companies that tweet are thus described qualitatively, in what regards their established social media popularity on Twitter and quantitatively, in what regards the number of words in the tweets that were investigated. An immediately efficient way to assess the level of popularity on Twitter is to consider the number of followers which represent the number of people who follow, i.e. read, your tweets – Tweeter followers, and the number of people you follow – Twitter following. I also identified the individuals behind the brand in each category based on the information that is widely accessible on their Twitter account.

In the car industry category, I included tweets from Ford and Honda. Scott Monty is the official tweeter for Ford, with 42,980 Twitter followers and 31, 561 following. Analyzed were 1073 most recent words. Alicia Jones is the official tweeter for Honda, with 7,525 Twitter followers and other 5,740 Twitter following. Investigated were 259 most recent words that were tweeted.

³ Both classifications included were completed by experienced internet contributors who chose their own blogs as the publishing platform. Cf. "Twitter Brand Index" in *Fluent Simplicity: Thoughts on Strategy an Operations*. 8 August 2008. 9 August 2010: <http://blog.fluentsimplicity.com/twitter-brand-index/>; Paul Dunay, "Brands That Tweet". *Buzz Marketing for Technology*. 15 December 2008. 9 August 2010: <http://pauldunay.com/brands-that-tweet/>.

In the food and beverage industry category, the most active social brands selected are Starbucks and Pepsi. Starbucks is represented by a large community of Twitter followers of 988,609 and other 81,444 Twitter following. Pepsi includes 32,906 Twitter followers and other 33,823 following. Analyzed were 260 words from tweets sampled from Starbucks and 834 words from tweets sampled from Pepsi.

In the media and technology industry category, I included CNN, Fox News, CBC News and New York Times, on one hand, IBM, and Comcast, on the other hand. CNN is represented by a very large community of followers of 1,212,272 and a small number Twitter following – 278. A similar distribution can be noticed in the number of Twitter followers for Fox News – 267,588 as opposed to the only 42 Twitter following; also CBC News with 34,198 followers vs. 89 following; New York Times – 2,524,571 followers and 198 – following. Conversely, the official tweeter for IBM is Steve Lazarus and counted were 1,762 followers and 1,626 following. The official tweeter for Comcast is Bill Gerth with 44,870 followers and 42,646 following. Investigated were 574 words from CNN tweets, 583 words from Fox News tweets, 1390 words from CBC News tweets, 1419 words from NY Times tweets and 93 words from IBM tweets and 1243 words from Comcast tweets.

In the travel industry category, included are two airline companies, Virgin America and JetBlue, the former with 105,824 followers and 18,602 following, and the latter with 1,595,064 followers and 115,163 following. Counted were 716 words from tweets sampled from the former and 574 words from tweets sampled from the latter.

In the entertainment industry category, sampled were tweets from Marvel Entertainment, with 71,201 Twitter followers and 310 following. Counted were 816 words from Marvel tweets.

4. Methodology

The methodology used in this paper relies on the computer-aided text analysis – Linguistic Inquiry and Word Count (LIWC) – that helps determine the rate at which the authors/speakers use positive or negative emotion words, self-references, big words, or words that refer to sex, eating, or religion within emails, speeches, poems, or transcribed daily speech (Pennebaker, Booth, Francis 2007). The original program was designed by its developers, in essence social psychologists, to calculate the degree to which people use different categories of words across a wide array of texts. The application was developed by having groups of judges evaluate the degree to which about 2000 words or word stems were related to each of several dozen categories. The categories include negative emotion words (sad, angry), positive emotion words (happy, laugh), standard function word categories (first, second, and third person pronouns, articles, prepositions), and various content categories (e.g., religion, death, occupation). For each text analyzed, LIWC computes the percentage of total words that these and other linguistic categories represent. The original intent of this program was to better understand how people used language when writing about emotional upheavals in their lives.

Function words in general were neglected insofar as they do not convey much in terms of content. Previous research attested an impressive vocabulary of over 100,000 words for the average native English speaker, among which fewer than 400 are function words (cf. Baayen et al. 1995). Yet over half of the words used in everyday exchanges are represented by function words (cf. Rochon et al. 2000). Among the most commonly used function words, we can find pronouns (first person such as *I, my, me*; second person such as *you*; third person such as *it, he*; demonstrative pronouns such as *that*), articles (*a, the*), auxiliary verbs (*is, was, have*), conjunctions (*and, but*), prepositions (*to, of, in, for, with, on*) (Chung and Pennebaker 348).

To illustrate the social function of *junk words*, I drew on Chung and Pennebaker (2007) who chose a commonly used word sequence in a casual conversation: “I can’t believe that he gave it to her” (Chung, Pennebaker 349). Our daily exchanges are similar in the abundance of function words such as the use of personal pronouns in the example proposed, which, given their deictic nature, take reference from the communicative situation they are attached to. As Chung and Pennebaker suggested, the sequence makes absolutely no sense, *unless you know who the “I”, “he”, and “her” are, as well as what the “it” is* (Chung, Pennebaker 349). They anchor communicative situations and serve as the linguistic glue that builds context.

The latest text analysis tool proposed by the original developer of LIWC (cf. Pennebaker et al. 2009) – *Analyze Words* – takes a Twitter user’s most recent updates and runs them through the program, closely looking into the frequencies of such style-related words. The output determines how emotional and social a user is, as well as what type of thinking style they employ when they post tweets (*Analyze Words News*, 2 November 2009). Scores are obtained for each of the categories proposed, which range from 0 to 100 percent. *Analyze Words* can be used for any Twitter user who posts fairly regularly and has an open access account. My investigation was completed only for tweets collected on August 11, 2010 from active Twitter accounts that were opened in the name of various popular brands.

The first language style was identified as the emotional bird and included are positive words and “we” talk (“upbeat”), anxious language (“worried”), hostile words and overuse of capitalization which is used for heightened emotion in the internet linguistic variety (“angry”), self-references and frequent use of depressive words (“depressed”).

The second language style is represented by the social realist and accounted for in the use of social words (“plugged in”), positive emotions words, questions, reference to others (“personable”), big words and fewer self-references (“arrogant/distant”), higher use of present tense verbs and punctuation (“spacy/valley girl”).

The third language style proposed the thinking machine that can be identified on the basis of ample and large words and phrases that include complex thinking styles (“analytic”), relating feelings to the environment (“sensory”), reference to the current events (“in-the-moment”).

A functional association among situational and linguistic parameters is also proposed by a quantitative measurement which lay the foundation for empirical investigation in exploring issues of language variation in general, in the computer-mediated discourse of microblogging in particular. Two situational parameters were established in the analysis of tweets from the most active social brands: the addressor (established as the official tweeter or non-official tweeter) and the company profile (product-oriented, service-oriented and

mixed: products and services), which can be further developed to included quantitative scales – ordinal and dichotomous.

Following Biber’s procedures (1994), I also considered and recoded tweets produced by the most active social brands for the two interrelated situational parameters: the addressor and the company profile. The former relates to the identity of the person who tweets in the name of the brand that can be recovered from the Twitter account. The latter refers to the main activities of the company behind the brand. The addressor was coded as dichotomous scales as in Table 1 and the company profile as continuous scales as in Table 2. Tweets were coded separately for each social brand with respect to the proposed situational parameters, as illustrated in Table 3.

Also employed is Pearson’s correlation coefficient, which originates in social sciences, but was recently proposed for modeling language data across disciplines (cf. Johnson 2008).

Value	Addressor
0	no official tweeter
1	official tweeter

Table 1. Dichotomous Coding for Addressor

Value	Profile
1	Products
2	Services
3	Mixed

Table 2. Ordinal Scales for Company Profile

Social Brand	Addressor	Profile
Honda	1	1
Ford	1	1
Starbucks	0	1
Pepsi	0	1
CNN	0	2
Fox News	0	2
CBC News	0	2
NY Times	0	2
IBM	1	3
Comcast	1	3
Virgin	0	2
JetBlue	0	2
Marvel	0	3

Table 3. Values with Respect to Two Situational Parameters for 13 Social Brand Tweets on 11 August 2010

5. Results

Given the proposal for coding per corpus, situational parameters as ordinal parameters can be correlated with language styles as continuous parameters, to measure the degree of association between the two variables. As shown in Table 4, Pearson's correlation coefficient was suggested to measure the relationship among the three language styles (emotional, social and thinking) and the two situational parameters previously coded (addressor and the company profile). The r value represents the correlation coefficient and the value r^2 or "R-squared" the coefficient of determination which establishes a direct relationship between the two variables in the sense that the percentage of variation in one variable can be predicted on the basis of the other.

	Addressor		Company Profile	
	r	r^2	r	r^2
Correlation Coefficient				
Language Style				
Emotional Bird	.57	32%	.81	66%
Social Realist	.51	26%	.83	68%
Thinking Machine	.61	37%	.85	73%

Table 4. Correlations between Language Styles and Situational Parameters

The parameter of addressor shows a weak correlation to the language styles proposed, with 26% to 37% percent of the variability in the distribution of the language styles in tweets predicted by the appointment of an official tweeter who microblogs in the name of the brand. The fact that the identity of the official tweeter may or may not be easily recovered has no consequence on the language style of the tweets they produce, which can be accounted on the shared technological knowledge on how Twitter actually works. It is obvious to all, followers or non-followers, that individuals tweet in the name of the companies whose brand they boost in the electronic environment, regardless of their public status as official or non-official tweeters.

Conversely, language styles can be influenced by the company profile that supports the brand identity in the electronic environment, which is shown by the strong association between language styles and the situational parameter of the company profile: correlations of 66 percent ($r = .81$), 68 percent ($r = .83$) or 73 percent ($r = .85$) indicate that the main activities of the company may have an impact on the language styles of the tweets that are published in the name of the brand.

Such findings establish the direct relationship between extra-linguistic criteria and language styles (emotional, social and thinking) in the investigation of social brand tweets. Of the situational parameters selected for this analysis, the company profile, in what concerns the main activities of the companies in relation to their brands, is predominantly correlated with the language styles.

In relation to the use of style words, I found that product-oriented company brands such as Honda, Ford, Starbucks and Pepsi are in general maximally "upbeat" (Honda 89%, Ford 65 % and Starbucks and Pepsi 100%), less "depressed" (Honda 27%, Ford 44 %, Starbucks 17% and Pepsi 4%) and more "plugged in" (Honda 85%, Ford 88%, Starbucks

79% and Pepsi 96%) than other social brands, in particular Starbucks and Pepsi are highly “personable” (Starbucks 83% and Pepsi 91%). They are all quite “analytic” (Honda 49%, Ford 55%, Starbucks 33% and Pepsi 46%) and “in-the-moment” (Honda 41%, Ford 44%, Starbucks 54% and Pepsi 52%).

Service-oriented company brands such as Virgin America and JetBlue are also maximally “upbeat” (Virgin 88% and JetBlue 100%), with high scores in the “plugged in” category (Virgin 62% and JetBlue 85%) and quite “analytic” as well (Virgin 49% and JetBlue 61%).

The mix type of company brands such as IBM or Marvel is less “upbeat” than other social brands (IBM 20% and Marvel 43%). Conversely, Comcast recorded the highest score in this category (91%). They are also highly “plugged in” (IBM 67 %, Comcast 95% and Marvel 60%) and “personable” (IBM 53% and Comcast 80%). In the case of IBM, the highest score was recorded in the “analytic” category (88%). Marvel recorded quite low scores in the “personable” category (19%) and in the “analytic” category (25%).

In general, product- and service-oriented company brands are particularly “upbeat” (average 90%), “plugged in” (average 83%) and “personable” (average 63%), whilst the mix type of company brands are more “analytic”(average 63%) and “in-the-moment” (average 58%). The frequencies of use in the functional or style-related words that proposed the three language styles (the emotional, the social and the thinking style) project various types of social behaviour, which based on these findings, can be reinterpreted as an emotional and social bird in the product- and service oriented company brands and a hybrid bird in the mixed type of company brands. In general, in what regards the company profile, what I noted was that the more popular the brand, the warmer and likable the tweet and the more human and easier to relate to. The tweets that are published by companies with an established popularity in the real world are more “personable” and socially engaged online than less popular brands: Starbucks, Pepsi, IBM, Comcast and JetBlue vs. the other social brands included in the investigation.

Unexpected were indeed the low to average scores obtained in all categories in the investigation of tweets originating from traditional media companies such as CNN, Fox News, CBC News or New York Times. They indicate that they did not adapt to the new media technologies; in fact they use the social platform of Twitter as an extension of the conventional arena to boost their brand reputation. If we correlated the language styles with extra-linguistic characteristics such as the company profile (service-oriented, media industry), also significant is the disproportion in their account of two Twitter features which were used as classification criterion for the classes of social brands selected: twitter followers and twitter following. If the number of Twitter followers is by far the highest if compared to other social brands, with millions of followers, they have only a handful of other brands or individuals they themselves follow, which may suggest a slower adaptability to the new technological requirements.

In all social brand tweets analyzed, the scores obtained in the “arrogant” category is homogeneously higher than expected, with an average of 60%, ranging from 51% to 78%. Such heightened values in the “arrogant” category indicate that discussed is action instead of emotion and scores in the self-reference category are quite low, which is natural in social brand tweeting that focus on the company’s action instead. Yet to appeal to more customers, a more emotional approach ought to be considered. What is surprising is that the social brands that scored the highest in the “upbeat”, “plugged in” and “personable”

categories, also scored high in the “arrogant” category. In the case of social brands, the category needs re-evaluation as “distant” instead of “arrogant” in the sense that the social interaction involves a communicative pattern of many-one-many as the company (“many”) represented by the tweeter (“one”) who addresses the customer (“many”).

6. Conclusion

In this paper, correlated were the situational characteristics and the language styles as continuous parameters to find the degree of functional association between the two. A strong correlation between the language styles and the situational parameter of company profile was thus found. Such findings support the interpretation that the company profile influences the construction of the social brand in the electronic environment. No direct consequence on the brand interaction with current and potential customers on Twitter was found in the corporate decision to use people or banners on the corporate Twitter handles, which indicates that the choice to have a customized image, hence the company’s logo, or a name, therefore the individual who tweets in the name of the brand, has no direct impact on the language styles that were built on the distribution of function words in tweets. In previous work, interpretations of the association between language use and the specific characteristics of the newly emerged communicative situation were based on the qualitative assessment of the underlying communicative functions. The present study lays the foundation for the empirical investigation in that a direct measurement of the association between communicative functions and linguistic features is proposed and validated.

The investigation of function or style-related words such as pronouns, articles, prepositions or conjunctions, almost invisible and neglected in previous text analysis research, proved fruitful in tapping the psychological processing underlying tweeting in the case of the most active social brands. Identified were three language styles that project patterns of social behaviour: the product- and the service-oriented company brands as *emotional and social birds* and the mix type of company brands as *hybrid birds*. The former includes the car industry, the food and beverage industry, the travel (airline) industry. The latter refers to two company brand types: technology and entertainment company brands that are better represented socially than emotionally but also media company brands that obtained low to average scores in each category investigated.

Based on such findings, what I observed was that positive, social tweets are in general associated with brands that have already established a reputation in the real world, such as Starbucks, Pepsi, IBM, Comcast and JetBlue vs. the other social brands included in the investigation. I concluded that there are indeed brands that have adapted their corporate discourse to social platforms as Honda, Ford, Starbucks, Pepsi, IBM, Comcast, Virgin America, JetBlue or relatively adapted such as Marvel to brands such as news networks and traditional media companies as CNN, CBC News, Fox News or New York Times that, although they have adopted social networking, have not yet adapted their corporate discourse and re-constructed their social identity to meet the conditions of the new medium.

The computer-aided text analysis of function word frequency use in tweets throws a fresh perspective on how language use can empower individuals who tweet in the name of

companies to engage more current and potential customers into interpersonal interaction online, which plays a significant role in building brand trust today. With the advent of the multilingual internet, the methodology proposed by Pennebaker et al. (2009) can be extended to other languages.

References

- Baayen, Harald R. et al. *The CELEX Lexical Database [CD ROM]*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania, 1995.
- Balwani, Samir. "10 of the Smartest Big Brands in Social Media." *Mashable*. 11 June 2010: <<http://mashable.com/2009/02/06/social-media-smartest-brands/>>
- Biber, Douglas. "An Analytical Framework for Register Studies." *Sociolinguistic Perspectives on Register*. Ed. Douglas Biber et al. New York: Oxford University Press, 1994. 31-56.
- Campbell, R. Sherlock and James W. Pennebaker. "The Secret Life of Pronouns: Flexibility in Writing Style and Physical Health." *Psychological Science* 14 (2003): 60-65.
- Chung, Cindy and James W. Pennebaker. "The Psychological Function of Function Words". *Social Communication*. Ed. Klaus Fiedler. New York: Psychology Press, 2007. 343-59.
- Johnson, Keith. *Quantitative Methods in Linguistics*. San Francisco: Wiley-Blackwell, 2008.
- Qualman, Erik. "Social Media: Fad or Fashion?" *SearchEngineWatch.com*. 10 August 2009. 11 August 2010: <<http://searchenginewatch.com/3634651>>
- Pennebaker, James W. "What Our Words Can Say about Us: Toward a Broader Language Psychology". *Psychological Science Agenda* 15 (2002): 8-9.
- Pennebaker, James W. and Thomas C. Lay. "Language Use and Personality during Crises: Analyses of Mayor Rudolph Giuliani's Press Conferences." *Journal of Research in Personality* 36 (2002): 271-282.
- Pennebaker, James W., Roger J. Booth, and Martha E. Francis. *Linguistic Inquiry and Word Count: LIWC 2007*. Austin, TX: LIWC: <<http://www.liwc.net>>
- Pennebaker, James W., Chris Wilson and Roger J. Booth. *Analyze Words*, 2009. <<http://analyzewords.com/>>
- Rochon, Elizabeth et al. "Quantitative Analysis of Aphasic Sentence Production: Further Development and New Data." *Brain and Language* 72 (2000): 193-218.
- Slatcher, Richard B. and James W. Pennebaker, 2006. "How Do I Love Thee? Let me Count the Words: The Social Effects of Expressive Writing." *Psychological Science* 17 (2006): 660-664.
- Slatcher, Richard B. et al. "Winning Words: Individual Differences in Linguistic Style among U.S. Presidential and Vice Presidential Candidates." *Journal of Research in Personality* 41 (2007): 63-75.
- Tausczik, Yla R. and James W. Pennebaker. "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods". *Journal of Language and Social Psychology* 29 (2010), 24-54.