Linguistic Analysis in Personality Research (including the Linguistic Inquiry and Word Count)

Margaret L. Kerr[1] *and Jessica L. Borelli*[2]

[1] *University of Wisconsin–Madison*
[2] *University of California, Irvine*

Our word choices communicate significantly more than we intend; in fact, researchers analyzing written and spoken language find that word use reveals a great deal about our thoughts, behaviors, and emotions. In an effort to understand the health benefits of expressive writing, a group of researchers developed a computer program known as Linguistic Inquiry and Word Count (LIWC; Pennebaker, Francis, & Booth, 2001) that analyzes the content and style of spoken and written language. Since its initial development 15 years ago, LIWC has become one of the most widely used applications for analyzing linguistic data. Using a computer program to analyze text reduces the resources required for research involving behavioral analysis, enabling investigators to systematically analyze massive quantities of data in very limited amounts of time.

How LIWC Works

The foundation of LIWC is a dictionary composed of approximately 4,500 words and word-stems. Within that dictionary, there are approximately 80 sub-dictionaries that represent word categories. The output contains a score for each category that represents the proportion of words that fall into that category (i.e. positive emotion words, pronouns) compared to the total number of words in the passage. The dictionaries are arranged hierarchically, such that one word can fall in to multiple categories. For example, the word "worry" falls into three word categories: anxiety, negative emotions, and affective processes. LIWC is designed to analyze both spoken and written language and has categories specifically targeted toward spoken language, such as fillers (i.e. "you know"), non-fluencies (i.e. "errr," "um") and swear words.

The 80 word categories represent four main language dimensions. *Language composition* includes descriptive information such as total word count, average words per sentence, and percentage of words longer than six letters, as well as style words such as prepositions,

The Wiley Encyclopedia of Personality and Individual Differences: Measurement and Assessment, Volume II, First Edition. Edited by Bernardo J. Carducci and Christopher S. Nave.

articles, and pronouns. *Psychological processes* include word categories aimed at tapping emotional, cognitive, and social processes and are similar to the categories seen in many other linguistic analysis programs. *Relativity* dimensions encompass time, space, and motion categories. Finally, *current concern* dimensions are common in content analysis and include categories such as money, occupation, and physical states.

The LIWC program underwent rigorous evaluation in order to establish a reliable and valid system for analyzing language. During the development phase, a panel of three judges made independent decisions about the inclusion of each individual word based on lists of potential words created by research assistants. The dictionaries were then re-assessed by a panel of three more judges, resulting in rater agreements ranging from 93% to 100%. In 2007, after over 100 million words were analyzed using the LIWC program it underwent a major revision of its dictionaries, which included expanding commonly used categories and removing some categories that had consistently low base rates or poor reliability (Tausczik & Pennebaker, 2010; Pennebaker et al., 2007). Since its initial development, LIWC has demonstrated strong reliability and validity across multiple studies (Mehl & Pennebaker, 2003).

LIWC researchers distinguish between two broad categories of words that are captured by text analysis. Content words are typically nouns, verbs, and adjectives that convey the substantive content of the text. Style words, also known as function words, are composed of articles, pronouns, prepositions, and auxiliary verbs (e.g. "it," "we," "was") that constitute the way we communicate information through language. Only about .05% of words in the English language are style words, yet we cannot communicate without them. These words make up about 55% of our spoken and written language. Although these words may appear less interesting, they actually reflect subtle but important differences in the speaker's and listener's shared knowledge, context, and relationships. Style words have been linked to personality, stress, depression, social status, and even physiological reactivity (Chung & Pennebaker, 2007). Because these words are so commonly utilized but their production and processing are largely unconscious, they offer a unique and implicit angle from which to examine our thoughts and behavior.

Applications of LIWC

Empirical work using linguistic analysis is expansive and covers a wide array of psychological constructs; a few of the major topics will be reviewed below. Language style matching (LSM) refers to the extent to which individuals mirror each other's word use and language style. Language matching occurs in all types of verbal interchanges between two or more people and is typically undetectable by both the speakers and outside observers. In fact, LSM levels are unrelated to self-reported quality of the conversation (Niederhoffer & Pennebaker, 2002). LSM is focused on the use of function words because these words are used frequently, processed quickly, and require shared social knowledge between conversation partners. LSM is thought to capture the extent to which individuals attempt to engage the other person as well as the degree to which those efforts are reciprocated. Higher LSM has been found to predict group cohesiveness and peaceful conflict resolution as well as romantic interest and relationship stability (Gonzales, Hancock, & Pennebaker, 2010; Ireland et al., 2011).

Verbal immediacy, a construct that represents degree of psychological engagement, is marked by high rates of first-person singular pronoun use and less frequent use of articles, long words, and discrepancy words. People using high verbal immediacy appear emotionally immersed in the topic being discussed while less verbal immediacy indicates a more distant or unengaged tone. For example, a statement such as, "I am mad about that" would score high on verbal immediacy whereas "That conversation angered me to the depths of my being" displays low verbal immediacy. Researchers have examined verbal immediacy as a predictor of formality, coping with trauma or loss and attachment security (Borelli et al., 2013; Cohn et al., 2004).

LIWC is commonly employed as a measure of verbal expression of emotional experiences. Validity studies reveal that more positive emotion words are used when writing about amusing events compared to sad or neutral events, and more negative emotion words are used when writing about sad events (Kahn et al., 2007). Emotion word use also signifies level of immersion in emotional experiences; those who use more negative emotion words when writing about trauma experiences also felt more physical pain (Holmes et al., 2007). In studies on savoring, participants who were instructed to write about positive relational memories used more positive emotion words than those asked to discuss a neutral event (i.e. morning routine; Burkhart et al., 2015).

Research Using LIWC

Age differences in language use have been found across longitudinal and cross-sectional studies. The use of first-person pronouns decreases over time while insight words, future tense verbs, and exclusion word use increase with age (Pennebaker & Stone, 2003). Other studies have found that younger people use more filler words (i.e. "you know") than older people (Laserna et al., 2014). An analysis of 14,000 text samples from 70 different studies revealed multiple gender differences in language use. Men use more complex language, such as longer words, more numbers, and more articles and prepositions, while women use more pronouns and social words, as well as less swear words than men (Newman et al., 2008).

Several studies examining thousands of text samples have found consistent patterns of word use across Big Five personality dimensions. Higher word count, fewer large words, and less verbal complexity has been associated with extraversion. People high in extraversion also tend to use more positive emotion words, less negative emotion words, and more social words. Conscientiousness and Agreeableness are consistently related to more positive emotion words, fewer negative emotion words, and fewer swear words, while Neuroticism is related to less positive emotion word use and more negative emotion words. Verbal immediacy is positively related to Openness to Experience and inversely related to Agreeableness (Mehl, Gosling, & Pennebaker, 2006; Pennebaker & King, 1999; Yarkoni, 2010).

The utility of LIWC has allowed researchers to examine other psychological processes that are difficult to detect via self-report. In studies on deception, Newman and colleagues (2003) concluded that when discussing the same topic, people instructed to tell a lie used fewer markers of cognitive complexity, more negative emotion words, fewer first-person pronouns, and fewer qualifiers (e.g. "but") than those telling the truth. In the context of groups and social relationships, first-person plural pronoun use has been linked to

higher social status and more group cohesion while first-person singular pronouns indicate lower social status (Kacewicz et al., 2014).

Assessing word use among depressed people and victims of suicide has illuminated the inner thoughts of those suffering from mental disorders. One study found that depressed students use more first-person pronouns and negative emotion words than students who have never been depressed (Rude et al., 2004). An examination of 300 poems from suicidal and non-suicidal poets revealed that suicidal poets used more first-person pronouns and more death-related words than non-suicidal poets (Stirman & Pennebaker, 2001). In a study of 40 suicide notes, researchers found that notes from completed suicides included more future tense verbs, more social references, and more positive emotion words than notes from attempted suicides (Handelman & Lester, 2007).

Word use has also been studied in the context of romantic relationships. In addition to the LSM research discussed above, women's use of first-person pronouns in instant message conversations predicts higher relationship quality among dating couples (Slatcher, Vazire, & Pennebaker, 2008). Alternatively, "you talk," or second-person pronoun use predicts lower relationship quality and more negative relationship interactions, while "we talk" is associated with more positive problem solving (Simmons, Gordon, & Chambless, 2005). "We talk" in couples is also predictive of better health outcomes in heart failure patients, and even more than self-reported marital quality (Rohrbaugh et al., 2008).

Conclusion

This brief review of a large and robust body of research reveals the importance of studying word use and its implications for understanding our personalities, relationships, and psychological processes. While we are beginning to gain an understanding of the dynamics of our language use, there is still a great deal left to discover. The development of Linguistic Analysis and Word Count has opened the doors to a greater understanding of human behavior through the way we speak, write, and converse.

See Also

Individual Differences in Coping with Stress
Personality and Language

References

Borelli, J. L., David, D. H., Rifkin-Graboi, A., Sbarra, D. A., Mehl, M. R., & Mayes, L. C. (2013). Language use in the Adult Attachment Interview: Evidence for attachment-specific emotion regulation. *Personal Relationships*, *20*, 23–37.

Burkhart, M. L., Borelli, J. L., Rasmussen, H. F., & Sbarra, D. (2015). Cherish the good times: Relational savoring in parents of infants and toddlers. *Personal Relationships*, *22*, 692–711. http://doi.org/10.1111/pere.12104

Chung, C., & Pennebaker, J. W. (2007). The psychological functions of function words. *Social Communication*, *1*, 343–359.

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

Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2010). Language style matching as a predictor of social dynamics in small groups. *Communication Research*, *37*(1), 3–19.

Handelman, L. D., & Lester, D. (2007). The content of suicide notes from attempters and completers. *Crisis*, *28*(2), 102–104.

Holmes, D., Alpers, G. W., Ismailji, T., Classen, C., Wales, T., Cheasty, V., ... & Koopman, C. (2007). Cognitive and emotional processing in narratives of women abused by intimate partners. *Violence Against Women*, *13*, 1192–1205.

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, *22*(1), 39–44.

Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count. *American Journal of Psychology*, *120*, 263–286.

Laserna, C. M., Seih, Y. T., & Pennebaker, J. W. (2014). Um ... who like says you know: Filler word use as a function of age, gender, and personality. *Journal of Language and Social Psychology*, *33*(3), 328–338.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, *90*, 862–877.

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 and Social Psychology*, *84*, 857–870.

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, *45*, 211–236.

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

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Sharing one's story: On the benefits of writing or talking about emotional experience. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 573–583). New York: Oxford.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The development and psychometric properties of LIWC2007. Retrieved from http://homepage.psy.utexas.edu/HomePage/faculty/Pennebaker/reprints/LIWC2007_LanguageManual.pdf

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

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

Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the life span. *Journal of Personality and Social Psychology*, *85*(2), 291–301.

Rohrbaugh, M. J., Mehl, M. R., Shoham, V., Reilly, E. S., & Ewy, G. A. (2008). Prognostic significance of spouse we talk in couples coping with heart failure. *Journal of Consulting and Clinical Psychology*, *76*, 781–789.

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

Simmons, R. A., Gordon, P. C., & Chambless, D. L. (2005). Pronouns in marital interaction: What do "you" and "I" say about marital health? *Psychological Science*, *16*(12), 932–936.

Slatcher, R. B., Vazire, S., & Pennebaker, J. W. (2008). Am "I" more important than "we"? Couples' word use in instant messages. *Personal Relationships*, *15*, 407–424.

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

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54.

Yarkoni, T. (2010). Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363–373.

Further Reading

Pennebaker, J. W. (2012). *The secret life of pronouns: How our words reflect who we are*. New York: Bloomsbury.