Chapter 12

Cindy K. Chung
The University of Texas at Austin, USA

James W. Pennebaker
The University of Texas at Austin, USA

ABSTRACT
Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, & Francis, 2007) is a word counting software program that references a dictionary of grammatical, psychological, and content word categories. LIWC has been used to efficiently classify texts along psychological dimensions and to predict behavioral outcomes, making it a text analysis tool widely used in the social sciences. LIWC can be considered to be a tool for applied natural language processing since, beyond classification, the relative uses of various LIWC categories can reflect the underlying psychology of demographic characteristics, honesty, health, status, relationship quality, group dynamics, or social context. By using a comparison group or longitudinal information, or validation with other psychological measures, LIWC analyses can be informative of a variety of psychological states and behaviors. Combining LIWC categories using new algorithms or using the processor to assess new categories and languages further extend the potential applications of LIWC.

INTRODUCTION
Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2007) is a word counting software tool widely used for quantitative text analysis in the social sciences. Although LIWC is able to quantify features in text that allow for text classification and predictions for a variety of behavioral outcomes, it has primarily been used to identify word features that are informative of the underlying psychological states of an author or speaker or groups. LIWC was originally developed to address content analytic issues in experimental psychology. Today, there are an increasing number
of applications across fields such as computational linguistics, forensics, marketing, and social computing. Together, the widespread applications to various fields, along with the psychological information that can be gleaned about an author or speaker, place LIWC in the realm of applied natural language processing (ANLP) approaches.

In this chapter, we describe the rationale of the LIWC approach, and highlight several applications of LIWC—such as a lie detector, a status decoder, or a social barometer. The main issues of a LIWC approach relative to other natural language processing (NLP) approaches are discussed, along with several solutions and recommendations. Finally, we present novel ways in which LIWC is being applied, and how it might be used in future research. (Note: LIWC is pronounced “Luke”; LIWC may be used as a noun, adjective [e.g., the LIWCed data], and verb [e.g., “Please LIWC the speech”]; adverb usage is under development.).

BACKGROUND

The story of the origin and development of LIWC is instructive for how the use of LIWC has evolved and expanded. In the nineteen-eighties, the second author conducted a series of studies examining health improvements after writing about one’s deepest thoughts and feelings regarding a traumatic or stressful event for 15 minutes a day over three to four days (e.g., Pennebaker & Beall, 1986). As more studies confirmed the health effects of expressive writing, multiple labs started to test what factors could account for the health and psychological changes associated with putting emotional upheavals into words.

In an attempt to determine if some features of the ways people wrote could shed light on the expressive writing phenomenon, Pennebaker initially had large groups of research assistants conduct qualitative content analysis on the essays. The task proved to be too complex, unreliable, and subjective, leading him to seek a computer-based approach to analyzing text (Graybeal, Seagal, & Pennebaker, 2002). With the help of his graduate student Martha E. Francis, he developed a relatively simple precursor of LIWC that counted a small group of emotion-related words. The initial success of the program led to a more concerted effort to capture a wider range of categories that ultimately resulted in the first commercial version of the LIWC program (e.g., Pennebaker & Francis, 1996).

It is important to acknowledge that a small number of computer-based word counting programs preceded LIWC in the social sciences. The most impressive, General Inquirer, had been created in the 1960s (Stone, Dunphy, Smith, & Ogilvie, 1966) for use on a mainframe computer. A handful of word counting approaches soon followed but, like General Inquirer, were not transparent in their operation, unavailable for desktop computers, and were heavily theoretical in their orientation, making them less broadly applicable. Whereas General Inquirer was developed to test need-based motivational theories, others were influenced by Freud and the psychoanalytic movement (Gottschalk, 2000; Martindale, 1990; Mergenthaler, 1996; Weintraub, 1989). All of the computer-based word counting systems at the time, including more recent programs such as Hart’s (2000) DICTION program, relied on dictionaries and complex algorithms not accessible to the users.

THE LIWC DICTIONARY

The heart of the LIWC program is its dictionary system. The current version of LIWC relies on approximately 80 default dictionaries made up of a total of about 4,000 words and word stems. The default dictionaries generally fall into one of four broader language dimensions that are grammatical (e.g., articles, numbers, pronouns), psychological, (e.g., social, emotions, cognitive
Related Content

Systematic Review of Speech Generating Devices for Aphasia
[www.igi-global.com/chapter/systematic-review-speech-generating-devices/40863?camid=4v1a](www.igi-global.com/chapter/systematic-review-speech-generating-devices/40863?camid=4v1a)

They Can't Fix What They Can't Hear: Improving Pre-Service Teachers' Spoken Grammar
[www.igi-global.com/chapter/they-cant-fix-what-they-cant-hear/108780?camid=4v1a](www.igi-global.com/chapter/they-cant-fix-what-they-cant-hear/108780?camid=4v1a)

Machine Audition of Acoustics: Acoustic Channel Modeling and Room Acoustic Parameter Estimation
[www.igi-global.com/chapter/machine-audition-acoustics/45496?camid=4v1a](www.igi-global.com/chapter/machine-audition-acoustics/45496?camid=4v1a)

Demystifying Domain Specific Languages
[www.igi-global.com/chapter/demystifying-domain-specific-languages/108723?camid=4v1a](www.igi-global.com/chapter/demystifying-domain-specific-languages/108723?camid=4v1a)