Traditional and Emotional Stylometric Analysis of the Songs of Beatles Paul McCartney and John Lennon

Cynthia Whissell
Psychology Department, Laurentian University of Sudbury, Sudbury, Ontario P3E 2C6, Canada
Email: cwhissel@Nickel.Laurentian.Ca

Key words: stylometric analysis, emotional stylometric analysis, music, Beatles

Abstract

Traditional stylometric measures such as word usage, word length, and word repetition were paired with six new measures that described word emotionality in terms of a word's pleasantness, its activation level, and the combination of these factors. All measurements were applied to the songs composed by Beatles Paul McCartney and John Lennon between 1962 and 1970. Stylistic and emotional differences between composers and across years were found to be in agreement with observations made by critics and analysts of Beatles' songs, suggesting that emotional stylometry is a valid instrument for the analysis of text. Lennon was the less pleasant and the sadder lyricist, and the Lennon-McCartney lyrics became less pleasant, less active, and less cheerful over time. Several other differences were observed and reported. A technique for summary emotional description (the emotion clock) was also introduced.

Computer-driven stylometric analysis capitalizes on the computer's ability to count large numbers of words speedily and accurately. A question frequently addressed by early stylometrists was that of authorship or of differences among authors. Discussions taking place in the Shakespeare Newsletter of 1984 (4, 5) suggest that this question was of interest a dozen years ago. It continues to be of interest to researchers such as Smith ("Sir Thomas Moore, Pericles, and Stylometry" was published in 1994). A second common line of inquiry has addressed the ontogenesis or development of style. Development comes into question when works of a single author (or those in a single category) are spread across a considerable span of time, during which they might have been expected to change. In his note on ""Edmund Ironside' and Stylometry," Sams (470) recognizes the importance of changing styles, though he criticizes the assumptions that Smith used to describe Shakespeare's different modes of writing.

Stylometry has been criticized for being a cold and impoverished method of textual analysis – one which studies words without studying their meanings (Sams, 471). On the other hand, stylometry has been lauded for its relative objectivity (Merriam, 5). This report describes a new stylometric technique – one which adds some degree of meaning to word-counting analyses – and it applies this new technique, in conjunction with some standard stylometry, to the lyrics of songs written by Beatles' composers Paul McCartney and John Lennon.

In his 1969 article, Charles Osgood (194–199) distinguished between the two types of meaning that may be carried by a word. The first is denotative or descriptive meaning, and the second connotative or emotional meaning. Denotative meaning is both complex and multidimensional. It is therefore very difficult to quantify. Connotative meaning, however, may be quantified in terms of two or three straightforward dimensions. The quantification of connotative meaning was pioneered by Osgood with the help of the semantic differential rating technique. According to Osgood, three emotional dimensions (evaluation, potency, activation, or E, P, and A) were sufficient to
explain most of the variability in semantic differential ratings. Other researchers have reported that two dimensions (evaluation, activation) were sufficient to explain some 80% of differences among words rated on scales such as good-bad, fast-slow, blue-yellow, nice-nasty, and so on (Whissell, “Pleasure and Activation Revisited” 871–874; Russell 1152–1168).

In terms of denotative meaning, there is little similarity between an ice cream cone and a smile yet as researchers who describe connotative meaning have noted, there is considerable similarity in the emotional meaning of these two words because both are pleasant (positively evaluated) and moderately active. Similarly, nightmares and broken bones are both unpleasant and active, while sleeping and rocking a rocking-chair are both pleasant and passive, and failure and loneliness are both unpleasant and passive. Whissell and her colleagues (Whissell, Fournier, Pelland, Weir, & Makarec, 877) have established a dictionary of connotative meaning by asking people to rate several thousand words in terms of the two main dimensions of emotion-activation and pleasantness (or evaluation). Where standard dictionaries provide a description of the denotative meaning of individual words, the Dictionary of Affect provides two ratings that represent connotative meaning. For example, the Dictionary of Affect reports ratings of 4.2 (activation) and 6.4 (pleasantness) for the word “delighted” and ratings of 2.4 (activation) and 3.2 (pleasantness) for the word “gloomy.” Given that both ratings were made with reference to a seven-point scale which had a standard deviation of one unit, it becomes evident that “delighted” carries a positive and moderately active connotation, while “gloomy” carries an unpleasant, inactive one. The two dimensions are unrelated or orthogonal: a high score on one will in no way predict a high or low score on the other.

In her 1994 article dealing with the writing of Hemingway, Galsworthy, and Faulkner, Whissell (“A Computer Program” 818) has suggested that whole passages may be quantified in terms of the emotional or connotative meaning of their component words, and has further demonstrated that a combination of stylometric measures with emotional measures provides an improved method of text description which comes closer to representing the complexity of critical commentaries that describe authors’ styles than do techniques which do not quantify emotion. Whissell has also used a combined stylometry (with both emotional and traditional measures) to examine writing styles in popular fiction (“Objective Analysis of Text” 1567–1570). Bestgen, Hogenraad, and their colleagues, as well as Anderson and Masters have also addressed the issue of emotional stylometry, and have provided convincing evidence for the contribution which it can make to the analysis of texts such as stories, titles, and sentences (Bestgen, 21–36; Hogenraad, Bestgen, & Durieux 455–478; Anderson & McMaster 3–9).

Although emotional textual analysis suffers from some of the same limitations as traditional stylometry (the principle limitation being the paucity of meaning carried by single words) it brings the analyst one step closer to the complex judgments of text made by readers on an ongoing basis, and it accomplishes this while utilizing relatively objective and perfectly replicable assessment techniques.

Any new measurement technique must be vetted in terms of its reliability and its validity (its consistency of measurement, and its ability to measure what it claims to measure) before it can pass into general use (Anastasi describes the requirements of acceptable measurement instruments in chapters 5 and 6 of her book on psychological testing). Where emotional stylometry is concerned, reliability is not problematic. The same passage or text will generate exactly the same emotional measures each time it is assessed because measures depend on ratings of the emotional dimensions as these apply to the words in the passage. A computer matches the passage to the Dictionary of Affect, and summarizes the resulting data in terms of passage means or percentage usage of various categories of emotional word. Reliability across passages but within author has also been demonstrated by Whissell (“A Computer Program” 820–21). Passages written by different authors can be correctly classified with considerable success on the basis of a combination of traditional and emotional stylometric measures.

A reliable or consistent measure is sure to be measuring something, but there is no guarantee that it is measuring what it claims to measure. The question which remains to be answered addresses the validity of what is being measured, and this question is much more difficult to answer in absolute terms. One species of validity discussed by Anastasi is convergent validity (156–158). Emotional stylometric measures could be evaluated against a criterion of critical commentary in a search for convergence or agreement among methodologies. For example, Whissell has demonstrated a species of validity for emotional and traditional stylometry by pointing out the consistency between