STYLISTIC AND EMOTIONAL ANALYSES OF ELEANOR PORTER'S NOVEL POLLYANNA USING THE NEW DICTIONARY OF AFFECT

Cynthia Whissell
Laurentian University

This article describes the results of stylistic and emotional analyses of the adolescent novel *Pollyanna*. The fictional Pollyanna first appeared on the literary scene as the major character in Eleanor Porter's 1913 novel by the same name. She made such a strong impression on readers that her name soon came to be equated with a particular emotional attitude - that of persistent optimism and good spirits.

Webster's Unabridged International Dictionary (1971) includes three words which have entered the English language as a result of Porter's novel: pollyanna, pollyanaism, and pollyannaish. These words are used to describe people or points of view that are markedly or even blindly optimistic, and those that insist on focusing on the good in everything while ignoring the bad.

When discussing the common theme of many older girls' novels including *Pollyanna*, *Anne of Green Gables,* and *Rebecca of Sunnybrook Farm,* Nodelman (1996) referred to the "warmhearted world of the traditional novel for girls" peopled by heroines whose "magical qualities seem to triumph over every bad circumstance." In these novels, the heroine's "major talent is the ability to restore the past - to return grown-ups to the happiness they felt in their youth," or, in other words, "to awaken dormant joyousness." The Oxford Companion to Children's Literature (1984, page 418) encapsulated the novel by pointing out that Pollyanna's favourite word was "glad."

It should be noted that the sense in which the word "pollyanna" is used is often pejorative. Definitions from Webster's dictionary emphasize the realist's denial of pollyanaism. The words which have evolved from Pollyanna's name can be and have been used to describe an overly sweet or mawkish attitude towards life's vicissitudes.

There were two questions driving the stylistic and emotional analyses performed on this novel: the first led to a search of the emotional language in the novel for an objective and reliable description of the warm, optimistic, and hopeful (or sweetly sentimental) tone perceived by readers, and the second led to an examination of development within the novel itself and an identification of patterns of movement across chapters.

Pollyanna was obtained from Project Gutenberg and scored on a chapter by chapter basis using the program TEXT.NLZ (Whissell, 1994) and the New Dictionary of Affect. The original Dictionary of Affect contained some 4700 words which had been rated by several people along the emotional dimensions of pleasantness and activation. Words had been selected for inclusion on the basis of their likely emotionality. The New Dictionary of Affect, which was developed in an ecologically valid manner, contains some 8700 words which were selected on the basis of frequency of usage in various texts.

Hit rates, which were in the range of 15-25% for the original dictionary have risen to 80-98% for the New Dictionary. The newer dictionary includes ratings for all parts of speech (interjections, prepositions, conjunctions, articles, nouns, verbs, adjectives, and adverbs). For *Pollyanna*, the matching rate was 89%. Nine out of every ten words in the novel were matched and scored by the New Dictionary of Affect.

STYLISTIC AND EMOTIONAL ANALYSES OF POLLYANNA

The computer program TEXT.NLZ (Whissell, 1994) provided scores for each chapter along the dimensions of pleasantness and activation. It also reported on the usage of extreme emotional words as a proportion of all emotionally scored words. The extreme categories evaluated were nasty words (words which were both unpleasant and active), sad words (those which were both unpleasant and passive), nice words (those which were both pleasant and passive), and cheerful or fun words (those which were both pleasant and active). In order to belong to one of the extreme categories, a word had to have a score in the top or bottom quartile for the appropriate dimensions. Samples of scored words from various extreme categories are included in Table 1.

Table 1. Examples of emotionally rated words from chapter 1 of *Pollyanna*, as scored by the New Dictionary of Affect.

Words rated high on Pleasantness: baby, best, glad, good, hope, love, mother, welcome.
Words rated low on Pleasantness: alone, dead, dried, frown, nothing, sick, stern, stranger.
Words rated high on Activation: enthusiasm, forced, strength, surprise, spite, hurried, do.
Words rated low on Activation: attic, cloth, death, envelope, floor, little, old, years.
Sad words (unpleasant *and* passive): alone, frown, dead, last, lonely, old, only.
Fun words (pleasant *and* active): babies, children, friend, glad, great, nice, people, surprise.
Nasty words (unpleasant *and* active): forced, hastily, spite, told, break, explanations.
Nice words (pleasant *and* passive): care, home, hoping, ideals, sleep, sympathy.

Sentence length (in number of words), and number of words per chapter were also scored by the program. A word frequency list previously developed by Whissell (1998) was used to assess mean word frequency per chapter for matched words, and the proportional use of unmatched words.

Scores from Pollyannna were interpreted in comparison to scores obtained for a widely sampled corpus of English texts (Whissell, 1998) containing more than a third of a million words of English. The corpus, which included samples from conversations, books, television shows, and magazines, was taken as a normative sample of English language.

RESULTS & DISCUSSION

The main stylistic findings of the research were completely consistent with the fact that *Pollyanna* was a work of juvenile fiction written close to a century ago. Sentence length was roughly 15 words. This value was significantly lower than that from several sources of adult materials which place sentence length at approximately 20 words ($z>100$, $p<.001$). Word frequency however was somewhat lower than mean word frequency for the broadly sampled corpus (roughly 2300 versus 2500, $z=7.36$, $p<.001$), and words not matched by the frequency list tended to appear somewhat more often than normal for current English texts (12% as opposed to 10%, $z=13.3$, $p<.001$).

The use of sentences much sorter than those in the normative corpus indicates a relatively simple writing style. This simplicity was qualified by the appearance of more - though not many more - infrequent words than normal. The appearance of such words may be attributed to the age of the text or to the inclusion of dialect in it (the characters Old Tom and Nancy often spoke in dialect). Chapters from the novel tended to include 1771 words (sd=753,

range 309-3837), and they were quite variable in length though chapter length did not change systematically as a function of chapter number.

Overall, *Pollyanna* was slightly above average in pleasantness ($z=5$, $p<.01$) and slightly below average in activation ($z=-5$, $p<.01$). The most revealing emotional findings were not associated with the dimensional scores themselves but with the use extreme emotional words. **There were proportionally fewer sad, fun, and nasty words in *Pollyanna* than in the comparative population of English texts (z=-3 or less, p<.01). However, there were many more nice emotional words in *Pollyanna* than expected on the basis of the comparative corpus (z=32, p<.001).** In fact there were almost twice as many (180%) nice words as expected -- 7.2% rather than 4%. Unlike fun words, which are both pleasant and active, nice words are pleasant and passive. The emotional stance which predominated in *Pollyanna* was one of pleasant passivity.

Overall findings are mirrored in the chapter by chapter analyses. Several chapters were significantly more pleasant than normal, and a few significantly less pleasant. Many chapters were of normal activation, but an equal number were of below normal activation. For the proportional use of fun and sad words, there was a scattering of significant values below normal, while for the use of nasty words there were a large number of chapters with below-normal usage.

Not unexpectedly, the most consistent finding had to do with nice words (those which are both pleasant and passive): **every chapter but two (chapters 29 and 32) had more nice words in it than normal.** In spite of the fact that the novel is something of a tearjerker, not a single chapter in it contained more sad words than normal.

In strictly emotional terms, as these are defined by the New Dictionary of Affect, the alchemy of *Pollyanna* transmuted sadness into niceness or softness, changing the valence of the emotion (from negative to positive) but leaving its passivity intact. This is the essence of Pollyanna's game of "finding something about everything to be glad about." The gladness referred to is a passive rather than an active one. It does not match the current interpretation of the word "glad" which is both more pleasant and more active than 80% of the words in the New Dictionary. The spunkiness brought by young actress Hayley Mills to Walt Disney's film version of *Pollyanna* was much appreciated by viewers, but the character described in the novel was in fact a "soppier" (equally pleasant but more passive) version of Mills's Pollyanna.

When emotional and stylistic variables were correlated with chapter number, it became clear than progressive chapters grew more active ($r=.57$, $p<.05$: they rose from below-normal levels of activation to normal levels) and contained relatively fewer sad words (although they never contained many, $r=-.63$, $p<.05$). **As the novel unfolded, chapters became more active and less sad, culminating in a happy ending in chapter 32.**

Sample size creates some interesting problems with textual analyses of the kind described in this report. The actual difference between the mean pleasantness of the novel and the mean pleasantness of the comparative population was .01 (1.86 as opposed to 1.85). A similar finding occurred for activation (1.66 instead of 1.67). These differences seem trivial and they are, indeed, small. The small size of these differences is almost certainly due to regression to the mean found in large samples and to facts of probability described by the Central Limits Theorem. When samples of words are drawn from a population of words, the closeness of the sample mean to the population mean is an inverse function of the square root of sample size. Large samples produce means closer to the population mean than small ones. The population or normative mean for pleasantness for a sample of more than one third of a million words was 1.85 with a standard deviation of .36. The normative mean for activation was 1.67 with a standard

deviation again of .36. For samples with roughly 52,000 words in them, such as *Pollyanna*, the standard error of the mean becomes a minuscule .0015. Significant differences in the second decimal place make sense in view of such small standard errors. The d' values which distinguish the signal (*Pollyanna*) from noise (the comparative population) in this case are of the order of .03, and the variation associated with differences between means is small - but it is still significant.

In view of small effect sizes, the correct interpretation of the findings described in this report would refer to an emotional colouring of language rather than an emotional dominance of language. **It is true beyond levels of chance that *Pollyanna* included more "nice" words than normal, and that it was passive and pleasant in emotional tone.**

References

Carpenter, H. & Prichard, M. The Oxford Companion of Children's Literature. Oxford, Oxford University Press, 1984.

Nodelman. P. "Progressive utopia, or how to grow up without growing up." In Egoff, Stubbs, Ashley, & Sutton, Eds, Only Connect: Readings on Children's Literature. Oxford, Oxford University Press, 74-82, 1996.

Project Gutenberg may be accessed at: http://www.promo.net/pg/

Whissell, C. A computer program for the objective analysis of style and emotional connotations of prose: Hemingway, Galsworthy, and Faulkner compared. Perceptual and Motor Skills, 1994, 79, 815-824.

Whissell, C. A parsimonious technique for the analysis of word-use patterns in English texts and transcripts. Perceptual and Motor Skills, 1998, 86, 595-613.