According to Their Plots, Jane Austen's Novels Are Not Comic Romances with Happy Endings

Cynthia Whissell[1]

[1] Psychology Department, Laurentian University, Sudbury, Ontario, Canada

Correspondence: Cynthia Whissell, Psychology Department, Laurentian University, Sudbury, Ontario, Canada.

Abstract

In order to answer two specific questions ("Do the plots of Jane Austen's novels match the plot of *Cinderella*?" and "Do Austen's novels include a comic or happy ending, defined as one where the author employs more pleasant language at the end of the novel than she did at the beginning?"), Jane Austen's six major novels and *Cinderella* were scored for the pleasantness of their language with the Dictionary of Affect (Whissell, 2009). The answer to both questions, based on results of regression analyses and means comparisons, is negative. Austen's novels are not variants of the Cinderella story, nor do they have the type of endings that characterize comic romances. *Cinderella* is very pleasant and has a distinct happy ending. In contrast, *Emma*, *Pride and Prejudice*, and *Northanger Abbey* are less pleasant and have equivocal endings, while *Mansfield Park* and *Sense and Sensibility* have tragic (relatively unpleasant) endings. *Persuasion* employs the least pleasant language overall but has a happy ending.

Keywords: Austen, emotion, plot, comic romance, Dictionary of Affect

1. Introduction

1.1 General Purpose

The research described in this paper employs quantitative analyses of natural language to answer two related questions "Do the plots of Jane Austen's novels match the comic romance plot of the Cinderella story?" and "Do Austen's novels include comic (happy) endings, defined as those where the author employs more fortunate or pleasant language at the end of the novel than she did at the beginning?" The instrument employed to answer these questions is the Dictionary of Affect in Language. The Dictionary was designed to score natural language according to the pleasantness of its words (Whissell, 2009); it allows for comparisons of the pleasantness of text segments to one another, and leads to the establishment of plot structure.

1.2 The Marriage-Success Plot

Jane Austen's six novels (*Northanger Abbey*, *Sense and Sensibility*, *Pride and Prejudice*, *Mansfield Park*, *Emma*, and *Persuasion*) all tell stories of young women who overcome difficulties in order to marry happily. The marriages are always upwardly mobile, although the step up is much higher in some cases (e.g., for Elizabeth, in *Pride and Prejudice*) than others (e.g., for Elinor, in *Sense and Sensibility*). Like many Harlequin® romances of the 20th and 21st centuries (Whissell, 1998), Austen's works could be accused of telling the same story over and over again. This possibility is noted and then dismissed by Hinnant (2006), who points to variability among plots in terms of the difficulty that the heroine must overcome in order to get married. Notwithstanding the diverse challenges that the heroine faces, could it be that all of Austen's novels actually *are* variants of the Cinderella story? *Cinderella* is a prototypical folk tale involving a young woman who overcomes difficulties in order to marry well. Andrew Lang, writing in his introduction to Marian Roalfe Cox's 1892 work *Cinderella: Three Hundred and Forty-Five Variants*, (Note 1) suggests that the Cinderella story has three key elements. All its versions involve a downtrodden heroine of "mean or obscure position," who achieves an advantageous or "good" marriage with the aide of a "supernatural" helper. In folk tales from across the world, supernatural helpers have ranged from fairy godmothers to friendly animals. In romantic novels, the supernatural helper has often been the author who crams couples into wedded bliss whether the plot of the novel has been developing in that direction or not. In this vein, one might recall Marianne and Colonel Brandon's abrupt transition to marriage at the end of *Sense and Sensibility*, or Fanny's very, very, late success in gaining Edmund's attention as a

potential wife in *Mansfield Park*. As Brown (1969) and others have pointed out, Austen's novels are not about the heroine's developing marriage plans, but rather about her development of self-awareness, with the marriage appearing almost as a parody at the end of the novel. Magee (1987) argues that Austen's achievement lies in her ability to take the shell of the marriage plot and work within it to create great literature.

Smith (1980) maintains that a "variant" is not simply a re-telling of some fundamental tale (the Platonic ideal of the tale) because no such thing exists. Instead, she suggests that story-tellers produce "versions" of a tale that are elicited by particular motives and serve varying functions (p. 221). Smith also points out the danger of nominating a central tale (such as *Cinderella*), and then concluding that just about every other success story is a variant of it. Where does one draw the line, and where does a Cinderella-success become different from some other success? What defines a tale as a variant rather than an independent story? An assumed variant should at least match the emotional orientation of the reference tale in order to qualify as a "variant" of it. For example, in order for *Mansfield Park* to be classified as a variant of the Cinderella story (as Hinnant, 2006, suggests it might be), one ought to be able to demonstrate *at least* that *Mansfield Park* matches *Cinderella* in terms of its comic narrative—that it is pleasant and has a happy ending.

1.3 Comic Romance

Austen's novels have been described as comic romances. They are labeled as comic largely because they are seen to have fortunate or happy endings. In contrast to comic works, tragic ones have unfortunate or sad endings. This distinction as to the emotional transition embodied by the narrative (from misfortune to fortune for comic works and in the reverse direction for tragic ones) is more than two thousand years old, and can be traced to Aristotle's *Poetics*. (Note 2) In a work designed to help advanced students of Austen's novels, Gill & Gregory (2003, p. 5) blithely affirm that "all Jane Austen's novels are comic." There are alternative viewpoints. Brown (1969) interprets the novels as parodies of comic romance, noting that Austen is "self-conscious in her courtship success scenes" (p. 1582), and Emsley (2007) identifies *Mansfield Park* as a tragedy. Newman (1983), on the other hand, suggests that Austen's endings are realistic rather than utopian. There is considerable agreement in terms of the classification of Austen's novels as romances. Northrop Frye defines a romance as an "extended and literary development" of the kind of story that is found in folk tales (Frye, 1976, p. 3) and notes that "the conventions of prose romance show little change over the centuries" (p. 4). He classifies Jane Austen as a writer of romances (p. 39).

With respect to the issue of the happy ending, it would be appropriate to inquire if all of Austen's novels actually do end happily. If one were to assume that being married was the equivalent of being happy then they must, because they all end in marriage. However, this would be a dangerous assumption to make given Austen's critique of the sentimental marriage plot (discussed by Brown, 1969) and her cavalier treatment of it in *Northanger Abbey*, her parody of Gothic romances. Were one able to actually measure the emotional tone of the author's writing, one could test the hypothesis that the novels are comic by examining them for the presence of a happy conclusion. In a comic romance, language should be at its most pleasant when the author is describing the marriage-success at the end of the novel, and emotion should rise in pleasantness across the novel to this high point at the end.

1.4 Design of the Research

This paper offers an example of the usefulness of quantitative techniques in application to literature—particularly when a specific hypothesis can be formed and tested. The research was designed to examine Austen's novels in terms of the similarity of their plots to the plot of *Cinderella*, a comic romance, and also to establish whether the concluding marriage in the novels is actually being treated (by the author) as a fortunate, pleasant, happy event. In order to address these questions, the research employs a technique developed to score the pleasantness of natural language known as the Dictionary of Affect in Language (Whissell, 2009). The Dictionary matches about 90% of the words in most English sources, assigning a pleasantness score to them on the basis of earlier respondents' opinions about the words. The pleasantness of an entire novel, or of portions of it, can be represented by an average across all matched words. The English translation of Perrault's *Cinderella,* (Note 3) analyzed below, illustrates the process of scoring a natural English text for pleasantness. The analysis of *Cinderella* also demonstrates the manner in which plot shape is established, and proposes several statistics to be employed in the evaluation of the ending of *Cinderella* as a happy one. Perrault's *Cinderella* is one of the oldest full European versions of the tale available, dating back to the late 17th century.

1.5 Scoring a Text for Its Emotional Tone: The Cinderella Example

The words in *Cinderella* (N=2564) were divided, without respect to content, into 10 successive half-page segments, each containing an average of 256 words. The Dictionary of Affect provided pleasantness scores for

91% of all words. For example, according to the Dictionary, "beautiful," "winning," "love" and "success" are all words of very high pleasantness falling more than two standard deviations above the pleasantness norm of 50. These words appeared in the concluding segment of *Cinderella* (segment 10 in Figure 1). In comparison "poor," "sadly," "fool," and "nothing," are words of very low pleasantness falling more than two standard deviations below 50, and these words appeared in segment 8, one of the least pleasant portions of the tale. A segment of text including many pleasant words would have a higher pleasantness mean which would indicate that a mood of fortune or happiness prevailed within the segment. On the other hand, a segment of text including many unpleasant words would have a lower pleasantness mean, which would suggest a mood of unhappiness or misfortune.

Table 1. Contents and pleasantness scores for each of the 10 divisions of *Cinderella*

| Division	Summary of Contents	Pleasantness Score
1	Cinderella relegated to sitting among the ashes	51.36
2	The step-sisters anticipate the Prince's ball	51.43
3	Cinderella wishes she could attend the ball	52.02
4	The godmother appears and begins to work magic	50.66
5	A coach, servant, and clothes are magically created	49.03
6	Cinderella is a hit at the (first) ball	52.39
7	Cinderella thanks her godmother	51.34
8	Cinderella goes to the (second) ball, forgets time	48.89
9	Cinderella runs away leaving a glass slipper behind	48.31
10	Cinderella is identified, her sisters treat her well	55.60
Entire Story		50.93
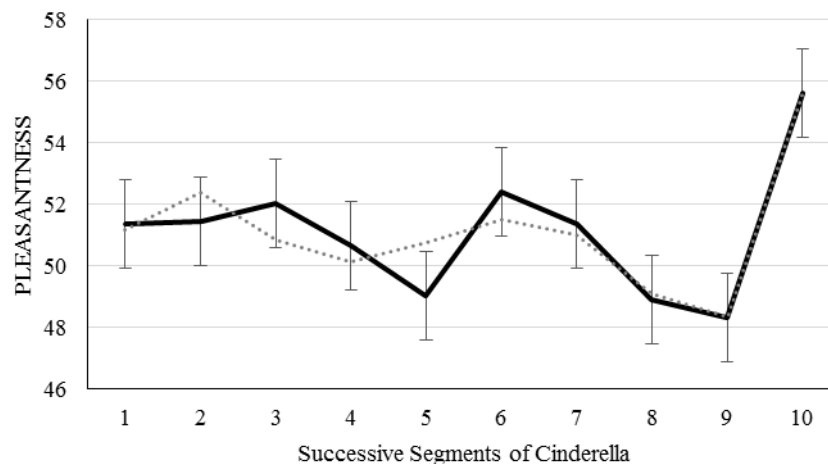
Figure 1. The rise and fall of pleasantness across the segments of Perrault's *Cinderella*

Note. The solid line represents actual means and includes standard error bars, while the dotted line represents means predicted by the polynomial equation. Non-overlapping standard error bars are indicative of a significant difference between means.

In terms of the scale employed for this research (Note 4), the overall average pleasantness for Everyday English, which was established on the basis of samples of broadcasts, texts, essays, and discussions (Whissell, 1998), is 50. The mean for the whole of Perrault's *Cinderella* is 50.93, with a standard deviation of 22.12. A *t* test for a single sample was conducted to compare the pleasantness of *Cinderella* to that of the population mean for Everyday English; results of this test indicated that *Cinderella* was more pleasant than the norm (t_{2333}=2.03, *p*<.05), although the difference was not a large one (Cohen's d=.04). A second comparative population involved the texts of 39 great novels published before 1914 (Note 5); this population represents the early English novel. The novels, which contained some six million words, were scored with the Dictionary of Affect. The mean pleasantness for the early English novel was 48.69, and *Cinderella* was more pleasant than this comparative standard as well (t_{2333}=4.89, *p*<.01, Cohen's d=.10). The results of these two comparisons establish Perrault's *Cinderella* as a pleasant tale both by everyday and by literary standards.

If *Cinderella* ends happily, then the last segment of the tale should include many pleasant words. This prediction can be visually confirmed from Table 1, where average pleasantness is noted for each segment. The most pleasant segment is the last one. The means from Table 1 are displayed according to segment in Figure 1; joined in sequence, they constitute the plot of the novel. The resulting line rises to represent pleasant segments (it is highest at the end) and falls to represent unpleasant ones (it is low for segments 8 and 9). There is a significant difference among segments ($F_{9,2324}$ =2.27, p=.016, η^2=.01), although the effect size is again weak. (Note 6) Figure 1 includes standard error bars, so it is possible to compare the segments to one another. For *Cinderella*, the last segment is more pleasant than any other segment, as indicated by the non-overlapping error bars. (Note 7) By this criterion, the ending of the tale is a happy one.

A further analytical approach involves the possible presence of a significant linear trend indicating rising pleasantness across the 10 segments. This trend was evaluated with a linear regression predicting the pleasantness of a word from its segment. The linear trend for *Cinderella* was not significant (p=.911). It seems that pleasantness does not rise steadily across the tale, but only at the ending. Between segments 1 and 10 pleasantness both rises and falls (for example, pleasantness rises from segment 5 to segment 6 but falls again between segment 6 and segment 8).

The plot of *Cinderella* which appears in Figure 1 can be modeled as a polynomial function of segment (1-10). This modeling allows for a smoothing and a mathematical description of the plot. A linear regression was employed to predict pleasantness for each word of the tale on the basis of the standard score for segment, and the standard scored raised to the second, third, fourth, and fifth powers. Prediction at the level of individual words was statistically significant (R=.086, p<.004). In an analysis involving entire segments, the correlation between actual and predicted segment means was very strong (r_8=.91, p<.001). The modeled plot for *Cinderella* appears in Figure 1 as a faint dotted line. It is evident, from the figure, that the value predicted by the polynomial equation provides a successful model for the actual pleasantness of various segments of the tale.

Perrault's *Cinderella* qualifies as a comic romance because all analyses (except the one for linear trend, which produced equivocal results) provide evidence in the expected direction and indicate that the tale is a pleasant one with a happy ending. These same analyses are applied to Austen's six novels. Each one is evaluated in terms of its overall pleasantness in comparison to Everyday English and early English novels. The six novels are compared to one another to identify the more and less pleasant among them. To determine if the ending of each work is a happy one, the first and last chapters of each novel are compared, and each novel is examined for the presence of a rising linear trend. Finally, the plot of each novel is modeled with a 5th order polynomial regression, and the shape of the modeled plots is displayed and discussed.

2. Method

The texts of Austen's novels were downloaded from Project Gutenberg. (Note 8) *Emma* (Project Gutenberg file number pg158) was downloaded in June of 2016 and the remaining novels (pp. 141, 121, 105, 1342, 161) in September of 2016. The text of each novel was scored by matching each and every word in it to the Dictionary of Affect. When a match was identified, the pleasantness value for the word was recorded in a computer file. The Dictionary provided pleasantness scores for 91% of the words in the novels (the matching range for individual novels was 90.6% to 91.3%). Together, the novels included 723,457 words of which 658,396 were scored for pleasantness by the Dictionary.

3. Results

3.1 Overall Pleasantness of Austen's Novels

The average pleasantness for each novel is reported in Figure 2, which also includes the average for Perrault's *Cinderella* and the two comparative standards (Everyday English and early English novels). The standard error for Austen's novels, represented by standard error bars in the figure, was .07. By the criterion of a difference greater than two standard errors, none of Austen's novels exceeds the pleasantness of Everyday English. Most of them fall significantly below that standard though *Emma*, the most pleasant novel, is not significantly different from it. On the other hand, all of Austen's novels are significantly more pleasant than the standard for early English novels. Within the set of six novels, *Emma* is more pleasant than *Pride and Prejudice* and *Sense and Sensibility* which do not differ significantly from one another. In turn, these two novels are more pleasant than *Northanger Abbey* and *Mansfield Park*, which do not differ significantly from one another. The least pleasant novel (significantly lower than all the others) is *Persuasion*. All of Austen's novels are significantly less pleasant than *Cinderella*, as evidenced by the total lack of overlap between standard error bars.
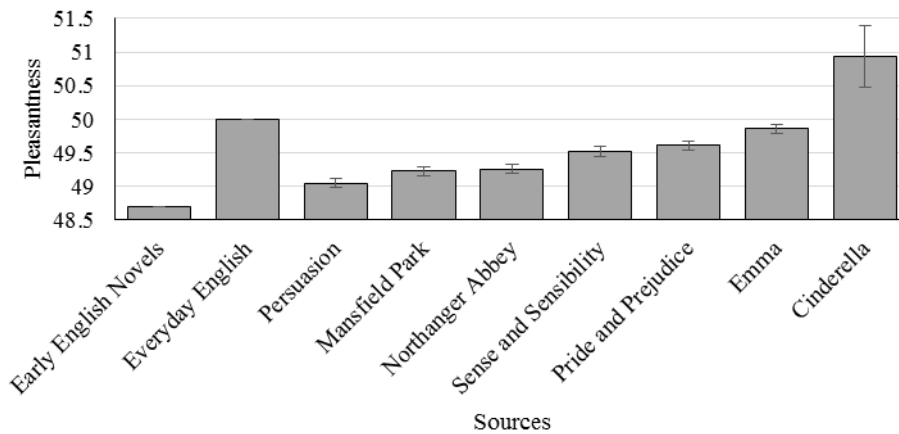
Figure 2. Mean pleasantness for Cinderella, the six Austen novels, and the two comparative populations

Note. Any two samples whose standard error bars do not overlap may be considered significantly different; for comparisons to the populations, a difference of two standard errors from any sample (.14 for the novels, .92 for *Cinderella*) can be considered significant.

3.2 Comparison of First and Last Chapters

The first and last chapter of each of novel were compared with a series of *t* tests for two means. (Note 9) Greater pleasantness in the last chapter would provide evidence in support of a happy ending. No significant difference between first and last chapters was noted for *Pride and Prejudice* (M=51.69, 50.26; *t*=1.28, *p*=.201) or for *Northanger Abbey* (M=49.88, 48.96; *t*=.97, *p*=.333). For *Sense and Sensibility* (50.27, 48.47; *t*=2.39, *p*=.017), *Mansfield Park* (50.124, 40.436; *t*=3.03, *p*=.002), and *Emma* (51.203, 48.708, *t*=3.83, *p*<.001), the last chapter was less pleasant than the first. The only novel including a more pleasant last chapter was *Persuasion* (48.19, 49.88; *t*=-2.12, *p*=.027). These results suggest that a happy ending (defined in terms of a comparison of first and last chapters) is not characteristic of at least three of Austen's six novels (*Sense and Sensibility, Mansfield Park,* and *Emma*), and that results for two novels (*Pride and Prejudice, Northanger Abbey*) are equivocal. According to the criterion of greater pleasantness in the last chapter, only *Persuasion* has a happy ending.

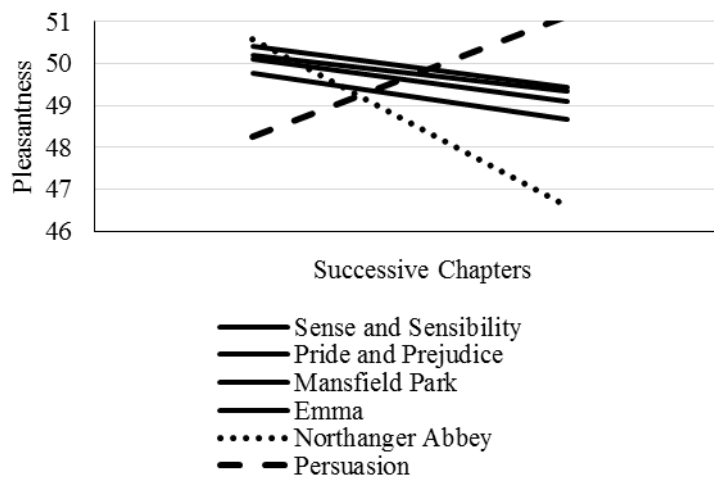### 3.3 The Linear Trend for Pleasantness Across the Novels

Figure 3. Significant linear trends in Austen's six novels

Note. The five solid lines have equal descending slopes and are essentially parallel to one another. Their difference in height represents their overall pleasantness. The lines moving from top to bottom represent *Emma, Pride and Prejudice, Sense and Sensibility,* and *Mansfield Park.*

The simplest way to describe or model the variation of pleasantness in each novel across chapters is in terms of a straight line. Such a line could either rise or fall. A rising line for pleasantness across chapters would indicate a tendency for the pleasantness of the words within a novel to rise more or less steadily throughout. Such a rise

would provide alternative evidence of a happy ending. Linear trend analysis may be more informative than the comparison of first and last chapters because it considers all chapters. Figure 3 displays the straight lines which were the best fit to each of the novels. Prediction of word pleasantness on the basis of chapter was significant in every case (Note 10) (Rs ranging from .013 to .018, p<.001). The slope for the resulting line was negative in all cases except for *Persuasion*, where it was positive (.057, s_e=.012). For *Sense and Sensibility* the slope was -.020 (s_e=.005), for *Pride and Prejudice* it was -.017 (s_e=.004), for *Mansfield Park* it was -.022 (s_e=.004), for *Emma* it was -.020 (s_e=.004), and for *Northanger Abbey* it was -.080 (s_e=.010). Five of the six novels provided evidence of significantly decreasing happiness across chapters. This result is represented, in Figure 3, by the five descending solid lines. The only line that rises is the one for *Persuasion* which is represented by a dashed line in the figure. Comparisons among slopes with the help of values for the standard error of the slope indicate that most novels which gave evidence of declining pleasantness (*Emma, Sense and Sensibility, Pride and Prejudice, Mansfield Park*) were similar in their rate of decline (their beta values were within two standard errors of each other), while one novel (*Northanger Abbey*) was characterized by a significantly steeper overall decline in pleasantness. The steeper decline is evident in Figure 3, where the dotted line representing Northanger Abbey drops precipitously. Not only does the evidence of linear trend analysis not support the presence of a happy ending in most of Austen's novels, it suggests that five of the six novels are best described in terms of a progress towards greater sadness throughout the novel, as indicated by the steadily declining pleasantness across chapters.

3.4 Modeling the Plot of Each Novel

The plot of each of Austen's novels is represented by the pleasantness of sequential chapters. As was the case with *Cinderella* in Figure 1, a rising line in the plot depicts the presence of a more pleasant chapter and a falling line the presence of a less pleasant one. All six plots were modeled in terms of a forced entry polynomial regression with five terms. (Note 11) The prediction was significant in every case (Table 2), and chapter means produced by the model were significantly correlated with actual chapter means (Table 2). Figure 4 depicts the modeled plot for each novel. The downward linear motion mentioned above is evident in every novel except for the last (*Persuasion*). This trend may represent the complication present in all the novels except *Persuasion*—the complication is the difficulty that must be overcome before the happy ending can be achieved. In *Persuasion*, the major complication occurs at the beginning of the novel. There is a recovery phase at the end of three novels (*Northanger Abbey*, *Pride and Prejudice*, and *Emma*) where pleasantness, which has fallen for several chapters, rises again. The interesting fact about the recoveries noted is that none of them exceed the first chapter in pleasantness. Pleasantness falls and then rises again but never rises beyond the level of the early chapters of the novel. Austen's happiest chapters are her early ones. This confirms results obtained from the comparison of first and last chapters. A final recovery of pleasantness does not occur for *Sense and Sensibility* or *Mansfield Park*: these two novels display a tragic downward trend moving from greater pleasantness to lesser pleasantness. The conclusions of these novels are much less pleasant than their introductory chapters.

Table 2. Standardized polynomial regressions predicting pleasantness from all terms in a 5th order polynomial regression

Novel	Ch	Ch^2	Ch^3	Ch^4	Ch^5	R	r with chapter
Northanger Abbey	-.08	-.01	-.08	.02	-.03	.04	.54
Sense and Sensibility	-.03	.03	.09	-.05	-.10	.02	.42
Pride and Prejudice	-.07	.03	.12	.00	.07	.03	.54
Mansfield Park	.05	.00	-.12	.00	.05	.02	.57
Emma	.04	-.05	-.15	.05	.10	.02	.42
Persuation	.00	.03	.03	-.03	-.02	.02	.46

Note. Ch stands for chapter; all regressions significant at p<.001; the last column represents the correlation of chapter means with predicted chapter means.
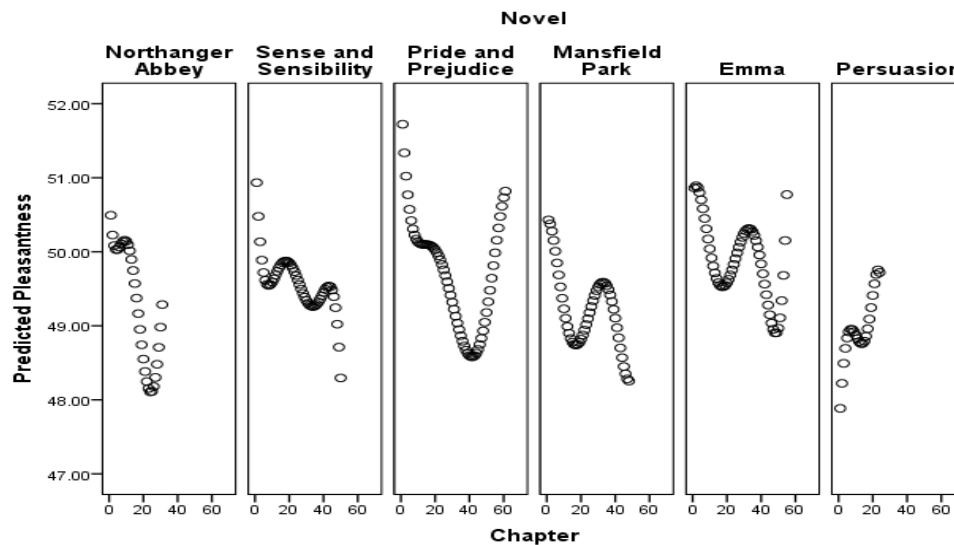
Figure 4. Modeled plots for Austen's novels based on 5th order polynomial equations in Table 2

The plots depicted in Figure 4 change direction at various points. If these plots are correctly describing the novels they represent, actions described in the chapters at the point of change should make sense in terms of the level of pleasantness indicated. For example, there is clearly a low point in pleasantness for *Northanger Abbey* in chapters 23-26 and for *Pride and Prejudice* in chapters 40-43. Do the events described in these chapters merit the unpleasantness of the language employed within them? What follows is a brief overview of the plot of each novel with special attention being payed to its turning points and its ending (Figure 4).

The plot for *Northanger Abbey* descends somewhat in pleasantness and then temporarily levels off at chapters 8-10 where love interests (Isabella-James, Catherine-Henry) are being developed, and Catherine makes friends with Eleanor. It then continues to descend steadily all the way to chapters 23-26: these are the chapters where Catherine is giving full sway to her Gothic imaginings about the evils of the Abbey (and General Tilney) and where she succeeds in visiting the room of the late Mrs. Tilney. Within the same set of chapters, Isabella breaks off her engagement to Catherine's brother, James. Chapters 23-26 are the low point of the novel for pleasantness. Even chapter 28, where Catherine is abruptly told (by Eleanor) that she must leave the Abbey, is more pleasant than chapters 23-26. From here, pleasantness rises to the more positive conclusion, with the fortuitous marriage of the main characters (because Eleanor married money, Henry could marry Catherine) being discussed in the last chapter.

The plot for *Sense and Sensibility* is tragic in shape as it displays no recovery of pleasantness in the final chapters. Pleasantness falls until chapter 8, where the Dashwoods occupy Barton Lodge and meet their neighbours, then rises somewhat to chapter 18, where Edward considers his future profession (as a vicar) and hears about Willoughby. After another short descent to the neighbourhood of chapter 34 (where Mrs. John Dashwood snubs her sisters-in-law), the plot displays another gentle rise to chapters 43-44 where Willoughby attempts to excuse himself to Elinor for his betrayals and Mrs. Dashwood arrives to tend the sick Marianne. In its last several chapters, the plot descends in pleasantness to the conclusion in chapter 50 where the successful weddings of Marianne and Elinor are announced. Again, a lucky happenstance (a gift of a small amount of money from Mrs. Ferrars) allows a couple (Edward and Elinor) to marry.

Pride and Prejudice has some claim to a happy ending because of the recovery of pleasantness at the end of the plot. The pleasantness of the novel falls briefly and then stabilizes for chapters 11-15 which describe Elizabeth's visit to Netherfield to help a sick Jane, and Mr. Collins' unwelcome marriage proposal. Pleasantness continues to fall to a low point in chapters 40-43. These chapters describe Elizabeth's visit to Pemberley (during which she experiences some regrets for her treatment of Darcy), and her description of Wickham's true character to Jane. Chapters 40-43 are a turning point and pleasantness recovers for the remaining 18 chapters, with Lydia being married and both Elizabeth and Jane being happily united with their loves by the last chapter.

Mansfield Park contains a potential early recovery of pleasantness, but this recovery is thwarted. After chapter 34 pleasantness declines steadily until the novel has ended. The problems of Fanny's life make for an

increasingly unhappy tale as pleasantness drops all the way to chapters 15-19 which describe the crisis of the play *Lover's Vows* that the young people of Mansfield Park and their friends want to present. Fanny disapproves of the play on moral grounds, and her uncle Sir Thomas eventually puts an end to the play-acting aspirations of the group. Pleasantness then rises to chapters 31-34 where Fanny learns of her brother's commission in the navy and persists in refusing the proposal of Henry Crawford (which has been supported by both Edmund and her uncle). After these chapters, Edmund becomes disillusioned with Mary Crawford and many different relationships (including one marriage) break down. Pleasantness falls until the end of the novel where Fanny is suddenly united with Edmund. Fanny's happiest points are her beginnings (where the pleasantness of the plot is highest) and her virtuous refusal of Henry Crawford's proposal (where pleasantness is moderately high). By contrast, Fanny's union with Edmund, crammed into the very last chapter, is discussed in dry terms, despite of the author's assurance that "Fanny… must have been very happy in spite of everything" (Austen, p. 750).

The plot of *Emma* is shaped like a W, with two low points in it. The pleasantness of Austen's language descends until chapters 16-19 where Emma receives Mr. Elton's presumptuous proposal (she had intended Mr. Elton to marry Harriet Smith). Pleasantness then rises until chapters 31-35 where Emma makes efforts to be fair to the Eltons (Mr. Elton has married a woman whom Emma considers to be vulgar), and Frank Churchill announces another visit to the area. Pleasantness falls again until chapters 47-50, where Emma finds out that Frank is marrying Jane (Emma eventually decides that she is happy for them), and where she realizes her own love for Mr. Knightley, and supposes that Harriet is her rival (Mr. Knightly disabuses her of this notion). Harriet marries elsewhere and the pleasantness of Austen's language rises towards a happy ending in chapter 55.

Persuasion was Austen's last-written novel and it was published after her death. This novel shows a rise in pleasantness until chapters 7-8 where Anne and Captain Wentworth meet again and interact. The novel enters a holding pattern and pleasantness remains relatively flat until chapters 13-15 where Louise, who had been injured in Lyme, recovers and there is gossip about a potential relationship between Anne and Captain Benwick. After this, pleasantness rises steadily towards the end of the novel, with Anne and Captain Wentworth becoming progressively closer to one another and eventually deciding to marry.

4. Discussion

Table 3. Meta-analysis: an overview of Austen's novels in terms of the five criteria of comic romance with matching scores for each novel to the comic criterion (Tot1) and to *Cinderella* (Tot2)

SCORES	FIVE CRITERIA OF COMIC ROMANCE						
Novel/Tale	More pleasant than Everyday English	More pleasant than Early Novels	More Pleasant at End than Beginning	Rising Linear Trend	Model Plot Recovers at the End	Tot1	Tot2
Cinderella	+	+	+	0	+	4	NA
Northanger Abbey	-	+	0	-	+	0	2
Sense and Sensibility	-	+	-	-	-	-3	1
Pride and Prejudice	-	+	0	-	+	0	2
Mansfield Park	-	+	-	-	-	-3	1
Emma	0	+	-	-	+	0	2
Persuasion	-	+	+	+	+	3	3

Notes. A plus sign indicates confirmation of the criterion for a comic romance (a significant result in the expected direction); a minus indicates disconfirmation (a result opposite to the one expected), and a zero indicates an equivocal result. Tot1 is the total match score for each novel to the standard for a comic romance (pleasant text, happy ending); all + add one point and all – subtract one point from the total; the maximum total is 5/5; negative scores indicate a tragic pattern. Tot2 is the total match score for each novel with *Cinderella* (only similar signs of +, -, or 0 contribute points); the maximum match is 5/5; scores closer to zero indicate greater differences from *Cinderella*.

4.1 Answering the Research Questions (In the Negative)

Are Austen's novels comic romances? Table 3, which will help answer this question, summarizes the results of the analyses conducted above. There are five criteria for a comic romance. Such a romance must be more pleasant in tone than early English novels; it must also be more pleasant than Everyday English; its last chapter (ending) must be more pleasant than its first (beginning); it must be described by a rising linear trend for pleasantness and its modeled plot must show some signs of a recovery—a rise in pleasantness at the end of the novel. Positive evidence in terms of several of these criteria, especially in the absence of contradictory evidence, would qualify a work for the label of "comic romance." Entries in Table 3 are in the form of a plus sign

(representing a significant confirmation of the presence of the expected comic effect), a minus sign (representing a significant result opposite to the expected effect, and suggesting tragedy rather than comedy) and a zero (an equivocal result with no associated statistical significance). For example, the entry for *Northanger Abbey* indicates that this novel is less pleasant than Everyday English (-) but more pleasant than early English novels (+). For *Northanger Abbey*, there is no difference between first and last chapters (0), but the modeled plot in Figure 4 shows a recovery (rise) at the end (+). When these indicators are summed, with single points being added for plus signs, single points being subtracted for minus signs, and zero signs being ignored, *Northanger Abbey* earns a total match score (Tot1 in Table 3) of 0/5. Every piece of evidence in support of a comic structure is matched and canceled by evidence of a contradictory nature. *Pride and Prejudice* and *Emma* produce similar conflicting results and earn match scores of zero as well: each novel is associated some significant results in support of a comic structure and an equal number of contradictory results. The analysis of *Persuasion* produces the most clearly confirmatory results, earning this novel a score of 3/5. Two novels (S*ense and Sensibility* and *Mansfield Park*) produce so many contradictory results that they each earn a Tot1 score of -3/5 and qualify as tragic rather than comic, final marriages notwithstanding. Most of the negative signs in Table 3 are associated with the lack of a happy ending for various novels. This lack is manifested in finales that are not more pleasant than introductions, in linear trends that fall in pleasantness rather than rising, and in the absence of a recovery for pleasantness in the modeled curve. With the exception of *Persuasion*, Austen's novels break the mold of comic romance in their conclusions. *Northanger Abbey*, *Pride and Prejudice*, and *Emma* have some comic elements, but not enough such elements to classify them as comic romances. *Mansfield Park* and *Sense and Sensibility* are tragedies.

The second research question addressed in this report concerns the similarity of each Austen novel to Cinderella and its status as a possible variant of this tale. A matching score called Tot2 (Table 3) was calculated to indicate the number of times that the result for each of Austen's novels matched the results for *Cinderella*. Not surprisingly, the highest matching score (3/5) is noted for *Persuasion*, which is the only definitive comic romance in the set. Some degree of similarity to *Cinderella* (2/5) is noted for *Northanger Abbey, Pride and Prejudice*, and *Emma*, all of which produce partial evidence in support of a happy ending. A minimal match to *Cinderella* (1/5) is evident for *Sense and Sensibility* and *Mansfield Park*. These two novels are more pleasant than early English novels, and this is the only thing that they have in common with *Cinderella*. Six more variants of the Cinderella story? Hardly!

To see if the conclusions described above apply to Austen's early work, *Lady Susan*, her brief epistolary novel (Project Gutenberg file pg946), was also analyzed. The novel was divided into 22 segments of roughly 1000 words. The Dictionary had a matching rate of 90% for this novel, which had a high mean pleasantness of 51.21 (it was as pleasant as Perrault's *Cinderella*). In terms of the codes employed in Table 3, *Lady Susan* was significantly more pleasant than early English novels (+) and than everyday English (+). However, because it did not have a happy ending, its last segment was significantly less pleasant than its first segment (-) and there was a strong downward linear trend across the whole novel, equivalent to that of *Northanger Abbey* (-; see Figure 2). There was no recovery of pleasantness at the end of the modeled plot (-). The total comic score (Tot1) for *Lady Susan* was -1/5, and its total matching score to *Cinderella* (Tot2) was 2/5. *Lady Susan* does not qualify as a comic romance, or as a variant of Cinderella, although it shares its overall pleasantness with this tale.

Austen's rather pessimistic endings cannot be attributed to a general preference for such endings in English literature. Of the 39 great premodern English novels mentioned above, one third gave evidence of a rising linear trend for pleasantness, one third of a falling linear trend, and one third of no trend at all (as was the case with *Cinderella*). Austen's proportion of tragically patterned downwards trends in the six major novels (.83) is significantly higher than that for the population of great novels (.33, z=2.61, p<.01), so less happy endings can be considered a distinctive characteristic of her work. *Persuasion*, published after Austen's death, might have been evidence of a change in the author's pattern.

4.2 The Special Case of Persuasion

Of all Austen's novels, *Persuasion* is the most comic in character and it is the best emotional match to the Cinderella story. The novel opens at a surprisingly low level of pleasantness (Figure 4): its early chapters are the least pleasant among all chapters in all novels. Yet the plot of the novel is characterized by a steady rise in pleasantness across chapters, and a happy ending. *Persuasion* (83,572 words) is considerably shorter *than Pride and Prejudice* (the Austen novel with which it is most commonly compared, 121,706 words). In some ways *Persuasion* seems like half (the latter half) of *Pride and Prejudice*: the reader enters the narration in media res, after a romance has already been initiated, a marriage proposal has been refused, and time has passed. The novel begins at a complication or crisis point and pleasantness improves thereafter. Tarlson (2006) suggests that

Persuasion is the result of a change in Austen's philosophy in the direction of increased Romanticism, which might explain the increase in pleasantness across the novel. Elizabeth Bennet responded to reason and logic, while Anne Eliot, appearing in the author's last novel, responded to her emotions. Thomas (1987) notes that *Persuasion* has long been regarded as the most Romantic of Austen's novels, and that the novel contains near-lyrical language, especially in its later chapters. Thomas suggests the Austen's Romanticism is the result either of a direct influence from the work of Wordsworth and Coleridge or of a "dominance of a particular epistemological ideology" (Romanticism) that influenced all three (1987, p. 920). Austen's "lyrical structure and strategies" (p. 895) would contribute to the pleasant language of the novel.

4.3 Limitations

In this research, the generally positive tone of comic novels was defined in terms of the pleasantness of the author's overall language, and the expected happy ending was defined in terms of the pleasantness of the language in different portions of the novel. This may be a limiting definition, and one can always return to the fact that the novels all end in a marriage for the main female character. This implies a positive ending because marriage, for a woman in Austen's culture, was not just *a* success story, it was *the* success story. But calling marriage a success does not change the tone of the language that Austen employed in her final chapters. If marriage is evidence of success, why is success being described in relatively unpleasant language? Greater understanding of novels is achieved, not by reliance on qualitative or quantitative assessments alone but by a close study of the interaction of the two approaches. Roger Fowler, a proponent of linguistic approaches to literature, defended such approaches on the basis of "the vitality and centrality of the verbal dimension of literature, the continuity of language in literature with language outside literature … [and the] power [of language] to organize concepts, to make fictions," (1971, p. viii) but he also stated that he "never believed in the possibility of a coherent "linguistic criticism"" (p. vii) which would replace other approaches to literature. The quantative view of Austen employed here, which is based on the pleasantness of words in her novels, is not meant to replace other forms of criticism but to be employed in conjunction with them.

Another issue comes from the fact that Austen's novels were written a good two centuries before the Dictionary of Affect was created. Because the Dictionary was constructed on the basis of people's current responses to words, we can correctly assume that the emotional fluctuations picked up by the Dictionary and modeled in Figure 4 are valid for today's readers of the novels. Unfortunately, we cannot be 100% certain that they also apply to the original readers of the works. Here and there a word will likely have changed in its emotional colouring over the centuries. The very high matching rate of the Dictionary (91% of all words) offers some protection against this limitation. The values used in the various analyses of the research employ thousands of words so that a single word with slightly altered meaning would have a minimal influence on results.

5. Conclusions

Jane Austen's six major novels employ pleasant language (in comparison to other novels, although not in comparison to everyday sources of natural English). Strikingly, and with only one exception (*Persuasion*), these novels show a tendency to decline in pleasantness across chapters. With the same exception, their endings are no happier than their beginnings. Some partial evidence in support of a happy ending is noted for *Emma*, *Pride and Prejudice*, and *Northanger Abbey*, but *Persuasion*, the least pleasant of all the novels in overall terms, has the strongest claim to the classification of "comic romance." On the basis of their emotional characteristics, none of Austen's novels qualify as variations of *Cinderella*, and two of them (*Sense and Sensibility* and *Mansfield Park*) actually qualify as tragedies because their emotional plots are characterized by a steady decline in pleasantness across chapters. These quantitative, language-based conclusions are offered to readers of the novels who are invited to consider them in juxtaposition with additional insights offered by other methods of their choice.

References

Austen, J. (n. d.). *The Complete Novels of Jane Austen*. New York, NY: Random House.

Brown, L. W. (1969). The comic conclusion in Jane Austen's Novels. *Publications of the Modern Language Association of America, 84,* 1582-1587. https://doi.org/10.2307/1261504

Emsley, S. (2007). The tragic action of Mansfield Park. *Persuasions On-Line, 28*(1). Retrieved from http://www.jasna.org/persuasions/on-line/vol28no1/emsley.htm

Fowler, R. (1971). *The languages of literature: some linguistic contributions to criticism.* London, UK: Routledge.

Frye, N. (1976). *The secular scripture: A study of the structure of romance* (Vol. 127). Harvard, MA: Harvard University Press.

Gill, R., & Gregory, S. (2003). *Mastering the novels of Jane Austen*. New York, NY: Palgrave Macmillan.

Hinnant, C. H. (2006). Jane Austen's "wild imagination": Romance and the courtship plot in the six canonical novels. *Narrative, 14*(3), 294-310. https://doi.org/10.1353/nar.2006.0014

Magee, W. H. (1987). Instrument of growth: The courtship and marriage plot in Jane Austen's novels. *The Journal of Narrative Technique*, *17*(2), 198-208.

Newman, K. (1983). Can this marriage be saved: Jane Austen makes sense of an ending. *ELH*, *50*(4), 693-710. https://doi.org/10.2307/2872923

Smith, B. H. (1980). Narrative versions, narrative theories. *Critical Inquiry*, *7*(1), 213-236.

Tarlson, C. E. (2006). Jane Austen, *Persuasion*, and the pursuit of happiness. *Lethbridge Undergraduate Research Journal, 1*.

Thomas, K. G. (1987). Jane Austen and the Romantic Lyric: *Persuasion* and Coleridge's Conversation Poems. *ELH*, *54*(4), 893-924. https://doi.org/10.2307/2873102

Whissell, C. (1998). A parsimonious technique for the analysis of word-use patterns in English texts and transcripts. *Perceptual and Motor Skills*, *86*(2), 595-613. https://doi.org/10.2466/pms.1998.86.2.595

Whissell, C. (1998). The formula behind women's romantic formula fiction (Statistical survey of 50 Harlequin-Presents novels). *ARACHNE*, *5*(1), 89-119.

Whissell, C. (2009). Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural language. *Psychological Reports*, *105*(2), 509-521. https://doi.org/10.2466/PR0.105.2.509-521

Whissell, C. (2017). Northanger Abbey: Jane Austen's Connexion with Radcliffe and the Gothic. *International Journal on Studies in English and Literature*, *5*, 51-59.

Notes

Note 1. http://www.surlalunefairytales.com/cinderella/marianroalfecox/index.html

Note 2. http://classics.mit.edu/Aristotle/poetics.html, S. H. Butcher translation, Section 2, XIII

Note 3. http://www.pitt.edu/~dash/perrault06.html

Note 4. The scale employed here is a linear transformation of the original Dictionary scale with a mean of 50 and a standard deviation of 22; it was adopted to spread scores over a wider range (roughly 0 to 100) and to provide easily interpreted scores.

Note 5. https://www.theguardian.com/books/2015/aug/17/the-100-best-novels-written-in-english-the-full-list. These novels were analyzed by the author in conjunction with her book entitled *Emotion and Plot in the Premodern English Novel*, in press.

Note 6. Weak effect sizes are typical of comparisons involving natural language because such language is extremely variable in terms of pleasantness.

Note 7. Non-overlap of whiskers implies a distance of more than two standard errors between means; a difference of this magnitude is statistically significant according to a *t* test comparing two sample means.

Note 8. http://www.gutenberg.org/browse/authors/a

Note 9. Degrees of freedom for the t test exceeded 1900 in every case.

Note 10. For all linear regression analyses in this section, degrees of freedom for the denominator were >70,000.

Note 11. Fifth order polynomials were chosen because they were able to predict a substantial proportion of between-chapter variation in pleasantness without including too many insignificant terms in the equation.

Copyrights