

Claire Cahoon

Faculty Advisor: Toby Dragon

Whalen Symposium 2017 Extended Abstract

Parsing Personality: Using Data Analysis to Interpret Different Points of View in a Literary Text

Many authors write from different points of view within a text to provide richer context for their storylines. When we read from a different point of view written by a respected author, there is a clear difference between character voices, even in similar situations. But how do we know this? Does the author actually change the way he or she writes?

A computer program can be used to attempt to answer these questions by finding and assessing differences between written points of view. To show that computational statistics and literary analysis can work together to provide an interpretation that surpasses a typical reading, I focused on the work of Toni Morrison. Morrison utilizes narration switches in her novel *Beloved*, writing of the infanticide of Sethe through three different characters. These three chapters are ideal for this project because Morrison chooses to write the same scene with three separate voices, giving direct comparison between the characters. Sethe, the matriarch of the novel and central character in this scene, has no clearly defined role within the larger narrative or characterization. Paul D is an outsider, a stranger in town and a male in a female-centric novel, leaving his role in question as well. The newspaper account is meant to show a more general and stoic view than the two other characters, but its impact is not clear. I propose that the data support existing readings of these widely varied characters and bring additional information to clarify the interpretations. I also hypothesize that Morrison's writing would be tangibly different between the three chapters, giving evidence that quantitative data analysis can supplement or surpass a typical human literary analysis.

In order to test this hypothesis, I wrote an object-oriented program in Python 3 that iterates through each sentence and word in a given text. I started with plain text files for the three chapters from *Beloved* and ran them through this program. To create a control group, I took three passages from the third-person omniscient narrator that provides most of the narration for the book, ran all three passages through the program, and averaged the results for each category of data. The program provides statistics for the number of sentences, number of total words, total

number of different words, number of misspelled words, average word length, average words per sentence, and the ten most commonly used words. To account for the fact that each chapter is a different length, I used four calculations to draw conclusions from these sets of data: average word length, average words per sentence, the percentage of different words, and the percentage of words that were misspelled. These four factors were used in combination with the ten most commonly found words in each chapter to compare the passages.

The results of this analysis supported the characters interpretations, and even added further insight into the nature of each character. This supported my hypothesis that Morrison does vary her writing style when writing from different points of view. For example, Sethe's data is closest to the control, confirming that Sethe is the focus of the narrative. Sethe does use 1.73 times more misspelled words than the control, which shows that Morrison is using speech that is more colloquial in Sethe's voice. Paul D's results differed the most from the control in every category, giving evidence for his status as the "other" in the novel. The newspaper's results showed that it used longer words and sentences with more correctly spelled words. In addition, the common words are mostly directional, like "back," "then," and "when," emphasizing a journalistic narration style, though a racist slur is also included, showing the prejudice of this perspective.

While the nature of literary analysis involves bias, this study shows an example of unbiased data serving as a foundation for further analysis of a text. Future work could include testing for factors like parts of speech or text variations within dialogue to fully assess an author's body of work. In this project, it was shown that Morrison's writing can be analyzed based on a basic reading and a quantitative data analysis, giving more information than either could alone.

Works Cited

Morrison, Toni. *Beloved*. New York: Random House, 1987. Print