
Characters in 19th Century Novels Display Distinctive Voices as Seen by Stylometric Analysis

Paul J. Fields
pjfphd@byu.net
Brigham Young University, United States of America

Larry Bassist
larrybassist@comcast.net
Brigham Young University, United States of America

Matt Roper
matt_roper@byu.edu
Brigham Young University, United States of America

Abstract

Can novelists create characters with distinct voices as reflected in each character's wordprint? We explored the literary skills of four prominent nineteenth century authors who are generally considered to have been experts in creating characters within their novels:

- Jane Austen – *Pride and Prejudice* and *Sense and Sensibility*.
- Charles Dickens – *Oliver Twist* and *Great Expectations*,
- James Fennimore Cooper – *The Last of the Mohicans* and *The Deerslayer*,
- Mark Twain – *The Adventures of Tom Sawyer* and *The Adventures of Huckleberry Finn*.

Applying stylometric analysis using non-contextual words and principle components analysis, we found:

The voice of the narrator in each novel was different than the characters in the novel, and the narrator's voice did not match the author's own wordprint voice.

- Each author's characters were distinctively different amongst themselves and also different from other authors' characters.
- The authors displayed varying ability to create distinctive characters, with Dickens' characters being the most distinctive followed by Twain's, Austens' and Cooper's.

We conclude that talented authors can create characters with distinctive voices and that authors have differing ability to do so.

Introduction

Since it is widely accepted that authors tend to have their own unique writing style, and since it is also widely accepted that authors are not able to disguise their overall writing style, it is reasonable to ask if novelists can create different voices for the characters in their books. The appropriate null and alternative hypotheses for this research question are:

- Null Hypothesis: A novelist's characters do not have distinctive voices.
- Alternative Hypothesis: At least some of a novelist's characters have distinctive voices.

Tim Hiatt and John Hilton (1990 and 1993) considered the question of character voice for William Faulkner, James Joyce, Mark Twain and Robert Heinlein and concluded that although an author could create character voices, of the author they tested Faulkner alone was uniquely able to create characters with differing voices. However, we noted that their analyses were simplistic and did not use the multivariate statistical techniques commonly used current in stylometric analysis. Therefore, we chose to test the null hypothesis of non-distinctive character voices based on non-contextual word frequencies and using principle components analysis (PCA).

Method

We applied PCA to novels written by Jane Austen, Charles Dickens, James Fennimore Cooper and Mark Twain (Samuel Clements). We selected two novels from each author and separated the words quoted by each character. We only included characters whose quoted words exceeded 500 words. Austen created sixteen characters in *Pride and Prejudice* and fourteen characters in *Sense and Sensibility* who met the minimum number of quoted words, while Dickens created twenty-three characters in *Oliver Twist* and fourteen in *Great Expectations*. Similarly, Cooper created twelve characters in *The Last of the Mohicans* and ten in *The Deerslayer*, and Twain created nine characters *The Adventures of Tom Sawyer* and fourteen in *The Adventures of Huckleberry Finn*. We then split each character's quoted words into 200 word blocks for analysis. Since each book also had a narrator, we also split the narrator's words into 2000-word blocks.

Results and Discussion

Figure 1 shows plots of the first two principle components for the four authors where each star dot represents a 2000-word block for the narrator and each circle dot represents a 200-word block for each character. Each novel's narrator clearly stands out from the characters in the book. It can also be seen in Figure 1 that the narrator's voices are similar for Austen's, Dickens', and Cooper's novels, while the narrator voices in Twain's novels are distinguishable.

Although we do not show the plots here, a similar analysis as in Figure 1 comparing the author's own wordprint voice based on non-contextual word frequencies in the author's non-novel writing to the narrator's voice in each novel showed that the narrators' wordprints did not match the author's own wordprint for all four novelists

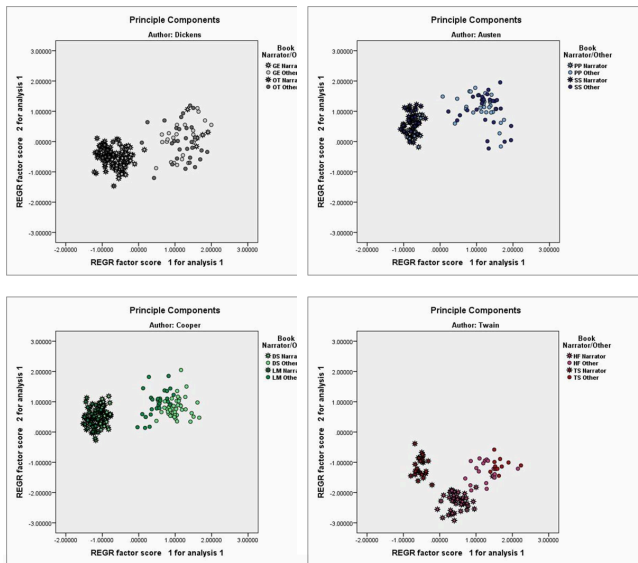


Figure 1: Principle components plots for the 19th century authors. The narrator's voice is clearly distinguishable from the characters' voices within each novel, and there is obvious diversity of voices among each author's characters.

Since the narrators for each author are obviously different in function word frequencies than the characters in each book, we only used the non-narrator words to compare the distinctiveness of the voices each author created for his or her characters. Again using PCA we applied multivariate analysis of variance (MANOVA) and constructed 95% confidence ellipsoids around the characters in each author's books. Figure 2 shows the ellipsoids using the first two principle components. Each dot in the plots represents one

character. Each ellipsoid shows the diversity of character voices for one novel.

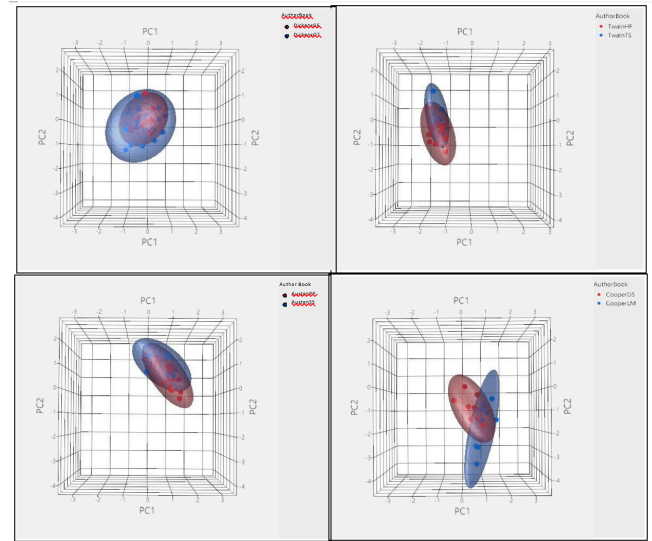


Figure 2: Principle components plots with 95% confidence ellipsoids for each author's characters by novel. The character voices are distinguishable between each novel for all four authors.

It can be seen in Figure 2 that the character voices in one novel are distinguishable from the character voices in the other novel for each author, even though there is some overlap of the ellipsoids.

The volumes of the ellipsoids for each author provide a measure of that author's character diversity – the larger the volume, the greater the diversity of character voices within a novel. Figure 3 shows a comparison across authors of the total volume of a 95% confidence ellipsoid encompassing both of each author's novels.

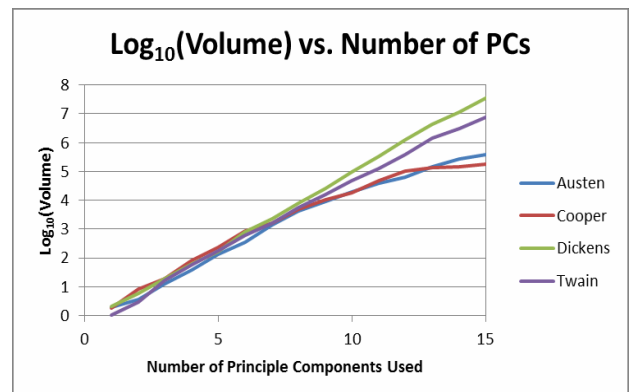


Figure 3: The volume of a 95% confidence ellipsoid encompassing the character voices of both novels for each author compared across all four authors. The log₁₀ is shown when the ellipsoid includes from two to fifteen principle components. For eight principle components and

above, the diversity of character voices for Dickens is clearly greater than for the other three authors.

It can be seen in Figure 3 that Dickens and Twain show greater character diversity than Austen and Cooper. For example, Dickens' character diversity (measured by ellipsoid volume with fifteen principle components) is about one hundred times bigger than Austen's. This shows that even among great authors, there is a spectrum of ability in creating character voices.

Conclusions

We examined the diversity of character voices created by four of the most respected nineteenth century novelists as shown in their characters' functional word frequencies. Using stylometric analyses we considered whether or not an author can create characters with different wordprint voices and found persuasive evidence that they can. We also found that the narrator in each novel speaks with a different voice than the characters and a different voice than the author's own non-novel voice. Further, not only can a talented author create characters with differing voices among characters within a novel, the characters are different among novels. In addition, we found that authors exhibit varying ability to create character voice diversity. Although all four were great novelists and could create distinctive characters, Dickens' and Twain's characters showed even greater voice diversity than Austen and Cooper.

Bibliography

Hiatt, T. and Hilton, J. (1990). "Can authors alter their wordprints? Faulkner's narrators in *As I Lay Dying*." *Deseret Language and Linguistic Society Symposium*: Vol. 16, Iss. 1, Article 6.

Hiatt, T. (1993). "Can authors alter their wordprints? James Joyce's *Ulysses*." Master's Thesis, Brigham Young University.