A Digital Analysis of Mystery Novels
Phoebe Bumsted, Class of 2017

This summer, I examined two works of detective fiction using digital techniques. The first, *The Name of the Rose* by Umberto Eco, is a postmodern murder mystery set in a medieval Italian monastery. The second, *The Mysterious Affairs at Styles* by Agatha Christie, centers around the poisoning of a wealthy woman. Both novels involve decoy suspects that distract the reader from the perpetrator. During my research, I used digital techniques to analyze trends in *The Name of the Rose*, then attempted to predict the outcome of *The Mysterious Affairs at Styles* using the same techniques. I decided to focus on character-specific techniques in order to analyze the ways the authors portray their suspects and victims.

I used two primary methods in this research, topic modeling and dispersion charts. In both cases, I wrote programs in the R programming language in order to extract pieces of text that mention a given character’s name. Using Google’s topic modeling tool, I analyzed these pieces of text for words commonly used in conjunction with the character’s name. In order to create a dispersion plot, I pieced together all the passages that include a character’s name to create that character’s individual timeline within the text. Then, I plotted mentions of a second character within that timeline.

In my analysis of *The Name of the Rose*, I noticed several trends related to the murderer Jorge. In my topic modeling of Jorge’s name, I found that words related to the method of murder appeared in conjunction with his name. Jorge poisons people through a forbidden book that leaves traces of poison on the hands, and “hands” and “fingers” appeared in relation to Jorge’s name. Using dispersion plots, I realized that the victims of the novel appeared in Jorge’s timeline around their times of death. When a character was about to die or had just died, Eco would mention the character’s name in proximity to Jorge’s name.

With these trends in mind, I attempted to predict the ending of *The Mysterious Affairs at Styles* using the same techniques. The data I produced seemed to indicate the guilt of Mary Cavendish and Lawrence Cavendish, both of whom were equally suspicious. However, my prediction was incorrect. I suspect that this was primarily due to the naming of characters. The program I used required the input of a one word name for each character. Thus, Emily Inglethorp became “Emily” and Mary Cavendish became “Mary.” Having not yet read the novel, however, my choices for names were not quite accurate. For example, Emily Inglethorp was more frequently referred to as “Mrs. Inglethorp.” Evelyn Howard was often nicknamed “Evie.” These naming disparities may have contributed to my inaccurate prediction.

In the future, it would be interesting to redo an analysis of *Styles* using a different naming schematic. I would expect this analysis to produce a more accurate prediction, but another incorrect prediction may reveal disparities in the ways the authors construct the two mysteries. Additionally, repeating the process with more mystery novels would provide further illumination.

**Faculty Mentor:** Crystal Hall  
**Funded by the Mellon Humanities Initiative**