"Quite Right, Dear and Interesting": Seeking the Sentimental in Nineteenth Century American Fiction

Tom HORTON
Kristen TAYLOR
University of Virginia

Bei YU
Xin XIANG
University of Illinois at Urbana-Champaign

This paper connects research by the nora Project (http://noraproject.org), a study on text mining and humanities databases that includes four sites and scholars from many areas, with current critical interests in nineteenth century American sentimental literature.

The term sentimental has been claimed and disparagingly applied (sometimes simultaneously) to popular fiction in this time period since its publication; academic study of sentimental fiction has enjoyed widespread acceptance in literature departments only in the past few decades. Academic disagreement persists about what constitutes sentimentality, how to include sentimental texts on nineteenth century American syllabi, which sentimental texts to include, and how to examine sentimental texts in serious criticism. Most of the well-known and widely-taught novels of the time period exist in XML format in the University of Virginia's Etext Center, one of the libraries in partnership with the nora Project; the original XML data for three texts discussed below was taken from this source.

The term sentimental novel is first applied to eighteenth century texts such as Henry Mackenzie's Man of Feeling, Samuel Richardson's Pamela, and Lawrence Sterne's Sentimental Journey and Tristam Shandy. Usually included in courses on the theory of the novel or eighteenth century literature, these works illustrate the solidification of the novel form. Sentimental novels emphasize, like Mackenzie's title, (men and women of) feeling. Feeling is valued over reason and sentimental is used with the term sensibility (recall Jane Austen's title Sense and Sensibility.) Although definitions of sentimentality range widely, and are complicated by the derogatory deployment of the term by contemporary and current critics, the group of texts loosely joined as being in the mid-nineteenth century sentimental period is a crucial link for humanities scholars that work on the novel and British and American texts in the nineteenth century connecting Victorian texts with their predecessors.

Sentimental texts are a particularly good place to look at how a group of texts may exhibit certain recognizable features; sentimental fiction uses conventional plot development, stock characters, and didactic authorial interventions. The emphasis is the exposure of how a text works to induce specific responses in the reader (these include psychophysiological responses such as crying and a resolve to do cultural work for nineteenth century causes such as temperance, anti-slavery, female education, and labor rights); readers do not expect to be surprised. Instead, readers encounter certain keywords in a certain order for a sentimental text to build the expected response. Although the term sentimental is used to represent an area of study and title literature courses, there is no set canon of sentimental texts because scholars do not agree on what constitutes textual sentimentality. Using text-mining on texts generally considered to exhibit sentimental features may help visualize levels of textual sentimentality in these texts and ultimately measure sentimentality in any text.

Two groups of humanist scholars scored three chapters in Harriet Beecher Stowe's text, Uncle Tom's Cabin, the most well-known and critically-acknowledged text in the group often considered sentimental. Although chapters in this text may be quite long and contain varying levels of sentimentality, the chapter as a unit was preferred as the original division structure of the text and the fact that humanist scholars expect this division and assign class reading and research by chapter units. UTC was later adapted into theatrical productions, and the idea of scenes (within chapters) may be a fruitful place to begin studying the sentimental fluctuations with a chapter unit in later phases of this project.

For the initial rubric, though, chapters were scored on a scale of 1 to 10: low is 1-3, medium 4-6, high 7-10. 10 is considered a "perfect"ly sentimental score, and as such, is only to be used when the peak of sentimental

conventions is exhibited: a character nears death and expires in a room usually full of flowers and mourners who often "swoon." The training set for this experiment includes two other texts that were scored on the same sentimentality scale, Susanna Rowson's 1794 novel Charlotte: A Tale of Truth and Harriet Jacobs's Incidents in the Life of a Slave Girl.

Since these texts were considered sentimental, most chapters were scored in the medium or high range, so the categories were changed to "highly sentimental" and "not highly sentimental." With D2K, the Naive Bayes method was used to extract features from these texts, which we might call markers of sentimentality. Looking at the top 100 of these features, some interesting patterns have emerged, including the privileging of proper names of minor characters in chapters that ranked as highly sentimental. Also interesting are blocks of markers that appear equally prevalent, or equally sentimental, we might say: numbers 70-74 are "wet," lamentations," "cheerfulness," "slave-trade," and "author." The line of critical argument that goes that the sentimental works focus on motherhood is borne out by "mother" at number 16 and "father" not in the top 100.

As we move into the next three phases of the project, we will include stemming as an area of interest in classifying the results. Phase two will use two more novels by the same authors as those in the training set; phrase three may include ephemera, broadsides, and other materials collected in the EAF collection at the UVa Etext Center. Phase four will run the software on texts considered non-sentimental in the nineteenth century and other phases might include twentieth and twenty-first century novels that are or are not considered sentimental. We hope to discover markers that can identify elements of the sentimental in any text.

Performing Gender: Automatic Stylistic Analysis of Shakespeare's Characters

Sobhan HOTA
Shlomo ARGAMON
Department of Computer Science,
Illinois Institute of Technology

Moshe KOPPEL
Iris ZIGDON
Department of Computer Science,
Bar-Ilan University

1. Introduction

A recent development in the study of language and gender is the use of automated text classification methods to examine how men and women might use language differently. Such work on classifying texts by gender has achieved accuracy rates of 70-80% for texts of different types (e-mail, novels, non-fiction articles), indicating that noticeable differences exist (de Vel et al. 2002; Argamon et al. 2003).

More to the point, though, is the fact that the distinguishing language features that emerge from these studies are consistent, both with each other, as well as with other studies on language and gender. De Vel et al. (2002) point out that men prefer 'report talk', which signifies more independence and proactivity, while women tend to prefer 'rapport talk' which means agreeing, understanding and supporting attitudes in situations. Work on more formal texts from the British National Corpus (Argamon et al. 03) similarly shows that the male indicators are mainly noun specifiers (determiners, numbers, adjectives, prepositions, and post-modifiers) indicating an 'informational style', while female indicators are a variety of features indicating an 'involved' style (explicit negation, first- and second-person pronouns, present tense verbs, and the prepositions "for" and "with").

Our goal is to extend this research for analyzing the relation of language use and gender for literary characters. To the best of our knowledge, there has been little work on understanding how novelists and playwrights