

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Presentations, Talks, and Seminar Papers --
Department of English

English, Department of

5-1-2014

"More than Custom has Pronounced Necessary": Exploring the Correlation between Gendered Verbs and Character in the 19th Century Novel

Nebraska Literary Lab

University of Nebraska - Lincoln, mjockers@gmail.com

Oliver Baylog

Laura Dimmit

Travis Heller

Gabi Kirilloff

See next page for additional authors

Follow this and additional works at: <http://digitalcommons.unl.edu/englishtalks>



Part of the [Digital Humanities Commons](#), and the [English Language and Literature Commons](#)

Literary Lab, Nebraska; Baylog, Oliver; Dimmit, Laura; Heller, Travis; Kirilloff, Gabi; Smith, Shannon; Thomas, Grace; Warren, Chandler; and Wehrwein, James, "More than Custom has Pronounced Necessary": Exploring the Correlation between Gendered Verbs and Character in the 19th Century Novel" (2014). *Presentations, Talks, and Seminar Papers -- Department of English*. 7. <http://digitalcommons.unl.edu/englishtalks/7>

This Article is brought to you for free and open access by the English, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Presentations, Talks, and Seminar Papers -- Department of English by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Nebraska Literary Lab, Oliver Baylog, Laura Dimmit, Travis Heller, Gabi Kirilloff, Shannon Smith, Grace Thomas, Chandler Warren, and James Wehrwein

“More than Custom has Pronounced Necessary”: Exploring the Correlation between Gendered Verbs and Character in the 19th Century Novel

Oliver Baylog Laura Dimmit Travis Heller Gabi Kirilloff
Shannon Smith Grace Thomas Chandler Warren James Wehrwein

5/1/2014

Abstract

During the 19th century, gender politics played a crucial role in shaping the emergence of the novel as a popular and successful form of literature. Not only were middle class women becoming an important part of the reading public, women were also authoring novels and creating complex heroines that at times pushed against, and at other times bolstered, traditional conceptions of propriety and femininity. Along with a rise in popularity came a rise in the critique of the novel as a valid literary genre; many critics claimed that novels were capable of corrupting their female readership. Authors responded to this claim both by seeking to highlight moral propriety through fiction and by questioning conventional notions of moral female behavior. Our project engaged with the complex relationship between gender and the 19th century novel by exploring the way that female characters are portrayed in a corpus of 3,500 19th century novels. We examined the agency associated with male and female characters by studying the different types of verbs used in conjuncture with male and female pronouns. Through the course of our research we focused on the following questions: Are male and female characters performing different actions in the 19th century novel? Are the female characters created by male authors associated with different actions than those created by female authors? During the course of the 19th century, do the types of actions associated with women change?¹

I. INTRODUCTION

IN both the micro and macro scales, stereotypes are an oft-recurring theme when discussing character. Many of the ways that we refer to characters, in casual conversation and in popular media, are practically stereotypes on their own terms—the result of generalization and constant use. Strong woman. Damsel in distress. Evil genius. Comic relief. In thinking about this, we began to wonder if these stereotypes were indicative of large numbers of characters, or if they were outliers made prominent by their particular qualities. Specifically, our questions led us to the realm of the 19th-century novel. The 1800s were a fertile time

for fiction in both the United States and Great Britain, but, more than that, they were also a period that fostered vocal and prolific discussion about gender in both public and private spheres. This is the time period that gave us both the Victorian ‘Angel in the House’ and the classic American cowboy. We wondered: did characters in 19th-century novels align themselves with prevailing gendered stereotypes? Would we discover passive women and active men? Would we discover that an author’s gender impacted the portrayal of character? To ask these types of questions, we first had to establish a way of tracking character in a large corpus of texts. Before examining the types of male and female characters present,

¹This research began in a class taught by Matthew Jockers, without whose support this work would not have been possible.

we first had to determine a way of locating characters and recording their genders. There were several possible methods to choose from to accomplish this task. Tracking the occurrences of characters' names using named entity recognition would have been one option. However, this approach would only catch references to proper names, leaving us without a significant chunk of data. This approach would have overlooked places where pronouns stand in for character names and would have potentially returned words that were not character names, such as place names. We chose to use personal pronouns as our main indicator of character presence, believing that pronouns would give us the widest swath of data. Focusing on pronouns, such as "she" and "he," allowed us access to data that was inherently gender-specific, which in turn enabled various lines of inquiry about the relationship between character and gender, and between author gender and character. While the decision to track pronouns was an important one that echoed through the entire project, we wanted to go a step further by examining the specific actions that were associated with different characters; once we had a way to track character presence, we were able to track verbs. Gathering the first verb that appeared after every male or female pronoun gave us one way to observe broad trends in characterization. By investigating action specifically, we were able to question the expectations and stereotypes that accompanied popular 19th century attitudes. Specifically, we were curious about questions of classification. The predominant social stereotypes of the 19th century would lead us to believe not only that male and female characters would manifest in different ways, but that male and female authors would approach them differently. Using the metadata of our text corpus, which included author gender, we were able to follow this line of inquiry. We were especially interested in establishing which verbs were most useful in predicting character (pronoun) gender. While our research examines the broad trends in the relationship between gender and character discussed above, our work also takes

a more concentrated approach: in addition to generating quantitative data about a corpus of 3,500 texts, we also worked closely with texts authored by Jane Austen, Charlotte Brontë, and George Eliot. Doing so allowed us to examine how these prolific writers fit into the social and literary trends of the 19th century. Given the way in which critics have singled out and canonized the works of these authors over the last two centuries, it seemed especially productive to situate these eminent female writers within a wider social context. In her seminal work *A Room of One's Own* Virginia Woolf comments on the ways in which all three of these writers form an important part of the history of female authorship:

"...they had no tradition behind them, or one so short and partial that it was of little help. ..there was no common sentence ready for her use...It was a sentence that was unsuited for a woman's use. Charlotte Brontë, with all her splendid gift for prose, stumbled and fell with that clumsy weapon in her hands. George Eliot committed atrocities with it that beggar description. Jane Austen looked at it and laughed at it and devised a perfectly natural, shapely sentence proper for her own use and never departed from it. Thus, with less genius for writing than Charlotte Brontë, she got infinitely more said. Indeed, since freedom and fullness of expression are of the essence of the art, such a lack of tradition, such a scarcity and inadequacy of tools, must have told enormously upon the writing of women. Moreover, a book is not made of sentences laid end to end, but of sentences built, if an image helps, into arcades or domes. And this shape too has been made by men out of their own needs for their own uses." (77)

As Woolf observes, the three authors highlighted here are worthy of individual attention. Charlotte Brontë's novels provide an intimate and deeply psychological portrait of female protagonists, who sometimes inhabit roles and take actions that would—by the going standard—be considered masculine. Jane Eyre pulls a man out of a burning bed; she breaks an engagement; she battles the elements and her

own psyche. Jane Austen has become one of the singular figures in English literature. In addition to exerting vast cultural influence, her works have also inspired scholarly debate. Do her works perpetuate a specific privileged lifestyle only available to a few? Or, by capturing it, do her works instead create a space for subtle critique? George Eliot, also known as Marian Evans, wrote under a male pseudonym to escape what she perceived as limitations on female authors in the 19th century. Writing in 1856, she said: "Silly novels by Lady Novelists are a genus with many species, determined by the particular quality of silliness that predominates in them—the frothy, the prosy, the pious, or the pedantic" (301).

Using computational methodologies allowed us to take a unique approach to studying these renowned female authors; through a combination of text mining and close reading, we hoped to explore how Austen, Brontë, and Eliot's use of male and female characters compare not only to 19th century social stereotypes, but also to the characters being created by their male and female contemporaries.

II. BACKGROUND

In order to question the presence of 19th-century gender stereotypes in a diverse corpus of 19th century fiction, we first had to familiarize ourselves with what those stereotypes might be. It is crucial to note that stereotypes are historically and geographically dependent. The social situation we examined for this project was predominantly American and British, white, and middle-class. The contents of our corpus were, in part, dictated by the texts that were available for us to manipulate. The corpus that we utilized was built from texts that were already out of copyright and digitized. Thus, the raw material that our project utilized reflects a specific slice of 19th century socioeconomic life. However, the type of analysis conducted during our project could easily be transferred to other areas; examining perceived social norms in conversation with literature could be done with any combination

of reading material and reading public.

The 19th century was not only a time of great literature production, but also of prescriptive literature production. Not unlike Cosmo quizzes or "25 Lifehacks for your Home" articles, 19th century moralistic novels and etiquette guides serve as an indication of the social norms of their period. At the same time however, the 19th century was also a time of dramatic social change, especially in regards to gender roles. This period produced many female writers, such as Jane Austen, Charlotte Brontë, and George Eliot, whose long lasting influence has continued to shape literature into the 21st century. While not all female authors defy traditional notions of gender within their works, the very act of publishing a book was a rebellious and transgressive act for female writers. In some ways, the social paradoxes of the Victorian period can be seen in 19th century literary production: on one hand, women were taking greater control in terms of literary production and consumption, on the other hand, conservative gender stereotypes abounded in many of the popular works of the period.

Certain patterns begin to emerge when examining 19th century social attitudes about ideal female behavior. In general, during the Victorian and Antebellum periods women were regarded as emotional creatures, whose main virtues stemmed primarily from their ability to act as compassionate wives and mothers. In the 19th century, an idea called "the Cult of True Womanhood" constructed societal ideals for how women should act. As Barbara Welter notes in "The Cult of True Womanhood: 1820-1860," "Submission was perhaps the most feminine virtue expected of women. Men were supposed to be religious, although they rarely had time for it, and supposed to be pure, although it came awfully hard to them, but men were the movers, the doers, the actors. Women were the passive, submissive responders" (158-159). Welter also explains that any actions performed by women were usually "morally uplifting tasks," such as cleaning the house, cooking, and preparing tea or coffee (164-165).

The 19th century emphasis on women as

compassionate creatures bound to the domestic sphere led to the popular stereotype of the ideal Victorian woman as “the angel of the house.” In their seminal work on 19th century female authors, *The Madwoman in the Attic*, Sandra Gilbert and Susan Gubar focus on the Victorian categorization of women as either “angels,” or “monsters,” noting that, for male authors, “a Victorian angel-woman should become her husband’s holy refuge from the blood and sweat that inevitably accompanies a ‘life of significant action’...” (24) In contrast to the domestic angel stands the stereotype of the woman as a “monster” or a “demon.” Women who failed to conform successfully to notions of feminine domestic propriety were often thought of as unnatural and grotesque. What made these women particularly threatening was the way in which a departure from domestic duties problematized the arrangement of 19th century family life; women who refused to orient their lives around being dutiful wives and mothers not only upset their own positions within society, but also destabilized the positions of their husbands, fathers, and brothers.

Women were not alone in bearing the burden of constraining stereotypes; men, especially those of the working class, were also expected to conform to certain ideals. In a time when the division of classes was increasing greatly, there was a strong presence of the new oppressed class in literature. In *The Struggle for the Breeches: Gender and the Making of the British Working Class*, Anna Clark describes some of the troubles that workingmen often went through in terms of identifying themselves: “Working men could not live up to the ideal of a patriarch, supporting his family and running his own workshop, but they did not know whether they should instead be belligerent artisans whose first loyalty was to their mates rather than their families, or self-improving Methodists who shut themselves off from neighborhood companions. A man might go to chapel where the elders would rebuke him for drinking and fighting, yet if he mended his ways he would find his workmates teasing him as a ‘molly’ for refusing to go to the pub.”

(265-266)

Several interesting questions are raised by Clark’s work: does the typical male character in 19th century novels fall into a traditional workingman stereotype? How do 19th century male authors, who tended to be middle class and not working class, portray male characters? Were the authors in touch with the struggles and life styles of workingmen, or are characters in 19th century novels mostly limited to the perspective of middle class writers? If working class characters are abundant in 19th century novels, than male characters might be associated with particularly types of behaviors, such as those linked with violence, poverty, drinking, and religion.

Female and male 19th century stereotypes were as unrealistic as they were confining. Men were encouraged to be the single providers for their family even though the workforce was quickly diversifying, while women, still taught to be meek and emotional, faced increasingly taxing physical demands. Thus, 19th century life in many ways embodied a growing tension between the lives of real people and the popular, but unrealistic, visions of ideal femininity and masculinity. These stereotypes and the social tensions they created seeped into 19th century fiction in the form of male and female characters who epitomize, transgress, and at times parody gender conventions.

One way to explore the relationship between gender and character in the 19th century novel is to focus on the degree of agency afforded to male and female characters. As the above discussion of gender ideals indicates, many 19th century gender stereotypes focus on the behaviors considered ideal for men and women. In light of this, our study attempted to examine what exactly male and female characters were “doing” in 19th century novels. Different types of actions and behaviors are closely linked to the allocation of agency, which can be analyzed through the type of verbs associated with different male and female characters. Were male characters, “riding,” “walking,” “forcing,” and “doing” more than their female counterparts? Were certain verbs pertaining

to domestic duties or emotional states used primarily in relation to female characters?

Using verbs to study the relationship between agency and gender is not unprecedented; in a 1997 study on the psychological perception of agency, Marianne LaFrance, Hiram Brownell and Eugene Hahn observe that different types of verbs imply different degrees of power. This study is particularly useful to our enterprise because of the way that it is focused on the relationship between text and perception. Hiram, Brownell, and Hahn comment on the unique way that grammar and word choice can impact a reader's, or listener's, perception of power relations. The authors note that "verb type and gender stereotype combine to affect people's perception about who is perceived to bring about interpersonal events..."(1) The authors argue that when action verbs are used, such as "to walk" or "to ride," it is assumed that the subject of the sentence causes the action. However, when a verb describing an emotional state is used, such as "to love," it is assumed that the object of the sentence somehow elicited the emotion. Thus, the subject of an action verb, and the recipient of an emotion verb, are both perceived as causal and assumed to have a high degree of agency. Even though this study was conducted recently, when combined with notions of Victorian gender stereotypes, this study points towards an interesting line of reasoning: if the ideal Victorian woman was associated with acts of emotion, such as loving, feeling, and worrying, than even though she may be associated with a large number of verbs, these verbs do not imply the same degree of agency as verbs related to physical action.

While studying a large corpus of texts allowed us to examine broad patterns in the types of verbs used in 19th century novels, our methodologies also enabled us to hone in on individual works; in finding trends in our data, we were able to determine which works deviated from these trends. In her book, *The Proper Lady and the Woman Writer*, Mary Poovey observes that many 19th century novels written by women "betray their author's unmis-

takable inhibition or hesitation before the cultural definition of femininity, but others reveal considerable ability to expand stereotypical images of the female..."(xi) Our research was particularly interested in examining whether the works of noted women writers, such as Austen, Brontë, and Eliot, exhibit these types of "inhibitions," or "expand stereotypical images" by creating characters that behave subversively. Many of Austen's original reviewers, including the 19th century rhetorician Richard Whately, praised her work precisely because they viewed it as appropriate rather than rebellious: "Miss Austen has the merit (in our judgment most essential) of being evidently a Christian writer:...The moral lessons also of this lady's novels, though clearly and impressively conveyed, are not offensively put forward..."(232) It is because her novels were initially perceived as "proper" that feminist readers have at times taken issue with Austen's works, noting for example the ways in which virtuous female characters are rewarded with marriage. At the same time however, many recent critics have argued that under a shallow veneer of convention, Austen's works are actually quite subversive.

Unlike Austen, Charlotte Brontë has traditionally been invoked as an author whose works challenge Victorian notions of femininity, certainly in part because she and her sisters (Anne and Emily) chose originally to write under male pseudonyms. This decision alone would seem to indicate that Brontë believed being known as a female writer would serve as a disadvantage, and her own writing confirms this. In a new 1850 forward to *Wuthering Heights*, Brontë wrote that "We did not like to declare ourselves women...because—without at that time suspecting that our mode of writing and thinking was not what is called 'feminine'—we had a vague impression that authoresses are liable to be looked on with prejudice" (Levine 276). This not only speaks to Brontë's feelings about the cultural reception of female authors, but also to something intangible that she sensed about her own writing, something she did not see reflected in other female-authored

works.

While today we find Brontë's works under her own name, rather than her pseudonym, to think about George Eliot is still to think about "a woman living within a man's name" (Wolff 25). Though Marian Evans eventually claimed ownership of her works (after someone else falsely admitted to being the real George Eliot), hardly anyone today would associate the name Marian Evans with *Middlemarch*, *The Mill on the Floss*, or *Daniel Deronda*. Of our three authors of interest, Eliot's novels and biography stand most obviously in opposition to the "Cult of True Womanhood" and the morally restrictive ideas that surrounded it. Eliot's novels are both philosophical and political; Anna Katona notes that, in an ever-growing and divergent 19th-century reading public, Eliot stood as an alternative for those who found Dickens to be "vulgar" (48). In life, Eliot publicly carried on a decades-long affair with a married man. (What was unusual was not the affair, but that it came without any attempts at stealth or secrecy.)

Given the ways in which all three of these writers have come to symbolize the emergence of the 19th century female author, it seems pertinent to examine how their approaches to creating character might differ from, or correspond with, the work done by their male and female contemporaries. In addition, much of the scholarship surrounding these women is concerned with the complex way that gender roles function within their work. Consequently, these three authors offered the opportunity for a unique case study that pits the socio-political trends of a burgeoning literary marketplace against the timeless works of three significant female writers.

III. METHODS

Digital copies of the novels used in this study were created through Optical Character Recognition (OCR), and were then Part-of-Speech (POS) tagged to enable the examination of pronouns and verbs. The compounding error rate of OCR and POS tagging should be taken into account when interpreting our results. Our

corpus of 3,516 texts included the distribution of works found in the following two tables.

Table 1: *Distribution of Texts (Nationality)*

Texts By Nationality	
Nation	Number of Texts
British	2351
American	628
Irish	310
Scottish	182
Welsh	15
French	10
Canadian	6
German	5
Dutch	4
Polish	1
Spanish	1
South American	1
Italian	1

Table 2: *Distribution of Texts (Author Gender)*

Texts By Author Gender	
Gender	Number of Texts
Male	1893
Female	1391
Unknown	232

Table 3: *Distribution of Texts (Date)*

Texts By Publication Date	
Date	Number of Texts
<=1799	68
1800-1833	625
1834-1866	1340
1867-1899	1430

In order to examine only gendered pronouns, two vectors were created: a male vector containing "male," "him," "his," "he," and "himself," and a female vector comprised of "female," "she," "her," "hers," and "herself."

While similar vectors were created for “catching” neutral, first person, second person, and non-gendered pronouns, only the male and female pronouns were entered into our data frame for testing. For each novel, an R programming script noted the first pronoun that it encountered, along with the first verb that followed this pronoun, and entered each into a data frame. Originally, the algorithm examined every verb between two gendered pronouns, which may have included multiple sentences and dialogue. To stay true to context, we decided the first verb would be most closely associated with the previous pronoun.

The relative frequencies of both gendered pronouns were then entered into a larger data frame; one row for each gender. Relative frequencies were calculated as a frequency of verbs following pronouns, as opposed to entire words in the text. The first nine columns of the large data frame contain the meta data and the pronoun data, followed by 72,708 columns of verb relative frequencies. This data frame was then winnowed by removing any word (column) that was only associated with either male or female pronouns; this resulted in our final data frame—consisting of shared verbs spanning 25,173 columns—that was used for testing. Removing these columns eliminated a number of OCR and POS errors that may have interfered with the overall results, as well as removing any verbs that might heavily skew accuracy towards one gender. Though the goal was to observe large differences in verb usage between gendered pronouns, it was also essential to avoid those verbs that may have skewed the data because of its effect on the pamr, or predictive analysis, tests used later in our methodology. The predictive analysis tests were run using the pamr package of functions in the R programming language, which uses the Kaplan-Meier Method for predictive analysis. These verbs may have lowered pamr error rates because of the ability to easily associate strongly related verbs to their respective pronouns.

Our focus and metric for observation was the error rate from tests run with the

pamr.confusion () function used to fit nearest shrunken centroids. Train sizes of data frames randomly selected 9/10 of the data, with the remaining 1/10 acting as the test. In some cases, test texts were specifically chosen (such as in the case with Austen, Brontë, and Elliot) in order to observe the tested data frame’s accuracy against an individual author and the gendered pronouns within their texts. By using the gendered pronoun column as our signal column, tests were executed to note the error rate with which a trained set of rows could correctly predict the gender of the pronoun tested against. For instance, when observing overall error rate on the entire corpus, a typical result would look like this:

```
> pamr.confusion(pamr.cv.out, threshold=2)
      1      2 Class Error rate
1 1857   69    0.03582555
2  158 1751    0.08276585
Overall error rate= 0.059
```

Figure 1: Sample of a pamr Test

In all tests, the “1” row represents male gendered pronouns, the “2” row represents female gendered pronouns, and the pamr.confusion() function displays the error rate for each class and an overall error rate (as an average of the two). Though the overall error rate is a telling metric, attention was also paid to the difference between pronoun accuracy, or the difference between error rates of 1 and 2. With the overall accuracy as the base error rate, tests were run on new data frames created by selecting certain variables including publication date, author gender and nationality (i.e. keeping the same number of columns but selected different combinations of rows). By comparing the error rates resulting from testing variations of these data frames—for example, the pronoun accuracy among texts published between 1800 and 1833 versus those between 1834 and 1866—claims can be made about the ability of computational learning to distinguish between genders and, as a result, the degree of difference and of adherence to gender archetypes within each corpus tested.

Additionally, more qualitative understanding of these difference were observed by comparing the top features used in computer learning to distinguish between the pronoun groups. The top ten verbs used overall and their importance for each gendered pronoun:

	feature	male	female
1	countenance	-0.3678	0.372
2	gaze	-0.2327	0.2354
3	sobbed	-0.2104	0.2128
4	bless	0.2069	-0.2092
5	face	-0.1993	0.2015
6	strode	0.1969	-0.1991
7	trembling	-0.1954	0.1976
8	embrace	-0.191	0.1932
9	knows	0.1801	-0.1822
10	took	0.178	-0.18

Figure 2: Top Ten Most Distinguishing Features

From this, one can glean the most important verbs in classifying these pronouns. It is telling that not only is there little variation in these listed features despite date, author gender, or other observable differences in metadata, but the features also strongly adhere to gender stereotypes often associated with the 19th century.

In an attempt to bridge the divide between macro and micro scale analysis-and to allow for an analysis of canonical texts and associated literary criticism-we also ran tests against works by specific authors: Jane Austen, Charlotte Brontë, and George Eliot. To prepare for these tests, the works written by the specific author were held out of the data frame for each test, and the training was done on all of the rest of the texts in the corpus.

IV. OBSERVATIONS

There were a number of significant observations made when running the various tests by publication date. To begin with, there was no observable difference between the overall error

rate between the different thirds of the 19th century texts.

Test	Male	Female	Overall
Overall	0.044081	0.0879365	0.066
1800-1833	0.026446	0.0967213	0.062
1834-1866	0.044628	0.0840266	0.064
1867-1899	0.03028	0.0885781	0.059
1800 - 1833 (Male)	0.052632	0.1463415	0.1
1800-1833 (Female)	0.027237	0.0864198	0.056
1834-1866 (Male)	0.040595	0.1093117	0.075
1834-1866 (Female)	0.02451	0.0875	0.056
1867-1899 (Male)	0.036474	0.0843558	0.06
1867-1899 (Female)	0.026882	0.0976909	0.062
1800-1833 (British)	0.017949	0.0989848	0.059
1834-1866 (British)	0.052186	0.08	0.066
1867-1899 (British)	0.024742	0.1010309	0.063
American Overall	0.061728	0.133452	0.097
Irish Overall	0.061594	0.0918728	0.077

Figure 3: Error Rates for Various Tests

As the above table demonstrates, the error rates for our tests were surprisingly linear over dates, author gender, and nationality (For a graphical representation of this data, please see Figure 5 in Appendix A). Additionally, there is a consistently higher error rate for female tests than for males, usually falling within a range of about 4-7%. One of the outliers to this trend, which also tests with the highest overall error rate, is the test done on texts written between 1800 and 1833. As shown in Table 3 above, this may be a result of a smaller number of texts being used for training compared to the other two thirds of the corpus. It might also be noted that, upon rerunning some of our tests-particularly with specific author tests-the results of the classifications differed slightly. This is a result of the random sampling done during the training process of the classification tests. For better and more reliable results in future work, tests should be run a large number of times while recording the results of each test. Doing so would result in a clear mean and standard deviation for each test, as well as a percentage of failure occurrences for pronouns within specific author texts.

V. RESULTS

Our observations both support and contradict our initial expectations. Perhaps one of our most basic, but none the less significant, conclusions is that in the 19th century novel certain verbs were strongly associated with male characters while different verbs were strongly associated with female characters. Based on the high degree of accuracy obtained from the test we used to predict the gender of a pronoun based on the verb associated with it, we can conclude that within our corpus, authors chose to portray male and female characters differently by associating them with divergent groups of verbs. This result is not surprising, especially given the way in which ideas about proper behavior differed for males and females within 19th century society. However, this result still has several far reaching implications, one of which is that “actions,” or verbs, are in fact an important part of creating and determining character. The fact that there was a general literary trend in the use of verbs in relation to gender suggests that verbs play a significant role in the creation of literary archetypes and stereotypes. What characters are “doing” is linked to who they are and how they are perceived. While many aspects of a story and its characters may change, such as appearance, style, genre, and plot, our research suggests that there is something fundamental about the links among character, gender, and behavior.

1	countenance	-0.351	0.3538	1
2	gaze	-0.2234	0.2253	2
3	sobbed	-0.2107	0.2124	3.1
4	bless	0.2038	-0.2055	4.5
5	trembling	-0.1975	0.1991	5.9
6	face	-0.1973	0.1989	6.3
7	strode	0.1969	-0.1985	6.9
8	took	0.1967	-0.1983	6.8
9	embrace	-0.1856	0.1871	8.9
10	knows	0.1804	-0.1819	9.9
11	muttered	0.177	-0.1784	10.7
12	manded	0.163	-0.1643	12.7
13	shouted	0.1619	-0.1632	13.1
14	wept	-0.1536	0.1549	15
15	blushed	-0.1532	0.1545	15
16	weeping	-0.1505	0.1517	16.8
17	walked	0.1486	-0.1498	17.2
18	loved	-0.1471	0.1483	18.5
19	marry	-0.144	0.1452	20.1
20	rode	0.1421	-0.1433	21
21	were	-0.1408	0.1419	21.6
22	tined	0.1396	-0.1407	22.2
23	terminated	0.1356	-0.1367	25.3
24	accept	-0.1346	0.1357	25.1
25	burst	-0.1345	0.1356	25.1
26	clung	-0.1324	0.1335	26.5
27	betrothed	-0.1303	0.1314	28.3
28	tripped	-0.1296	0.1307	29
29	shook	0.1285	-0.1296	30.1
30	clared	0.1277	-0.1288	30.8

Figure 4: Top Thirty Most Distinguishing Features

A close look at the verbs that authors used in conjunction with male and female characters also confirmed our initial expectations; based on a preliminary survey of the verbs associated with each gender, the results appear to support both our initial hypothesis and existing literature: action verbs and verbs associated with “dominant” behavior are associated with men while domestic, emotional, and submissive verbs are more typically associated with women. The verbs listed above are the 30 verbs that were most essential in determining gender; all of the verbs were used in conjunction with both male and female characters, but had a large difference in the frequency with which they were used for male or female characters. A positive number in the third column indicates that the verb was more frequently aligned with male pronouns, while a positive number in the fourth column indicates that the verb was more frequently aligned with female pronouns.

As can be seen from this list, verbs that one would associate with submissive, emotional behavior, such as “sobbing,” “blushing,” and “trembling,” are all associated with female characters. Based on the distribution of verbs, the average female character in our corpus seems to behave in a manner similar to “the angel of the house.” Upon close examination, one trend which emerges from our data is that verbs associated with motion, such as “strode,” “walked,” and “rode” are associated with male characters. In order to accomplish these types of actions, it is often necessary to travel - moving between multiple locations or spheres. The fact that it is men who move about in this way, seems to support the Victorian notion that men belonged out in the world, while women belonged in the home.

While overall our results seem to support our initial hypotheses, our findings on the relationship between author gender and verb choices defied our initial predictions. Based on our study, it appears that the association of different verbs with male and female characters varies somewhat, but not radically, based on the gender of the author. Our results indicate that male and female authors were using the same, or similar verbs, to portray the actions of their male and female characters, respectively. This is a somewhat unexpected result given the way in which many scholars have stressed the rebellious nature of 19th century female authorship. We initially would have assumed that female writers would have been more likely to break with conventional gender stereotypes by creating active heroines who at times adopted behaviors deemed masculine by Victorian society. However, it is important to not that this result does not mean that female writers were not creating strong female characters. As some scholars have noted, women writers at times used stereotypes to their advantage, creating conventional female characters but portraying them as ironic or humorous.

Given the ways in which gender roles changed over the course of the 19th century, we were also expecting to find that the distinction between verbs associated with male characters

and verbs associated with female characters began to break down over time. However, our research found that there was not a significant change in the relationships among gender, verbs, and character over the course of the 19th century. By the end of the 19th century, writers were still using different sets of verbs to define the actions of their male and female characters. In fact, using our predictive model to examine verb classification in the first and second half of the 19th century showed no significant difference between the two. Our model for determining the gender of a pronoun based on the verb associated with it was just as successful for the novels from 1800-1833 and 1834-1866 as it is for 1867-1899, with error rates ranging from .062 and .064 to .059 respectively. Because of the way in which gender roles gradually became less rigid in society over the course of the century, one would expect the roles in literature to change as well. This observation raises more questions than it answers. One possible explanation for the results is that society simply took longer to catch up to real life changes. Nostalgia and yearnings for more conventional representations of gender could have pervaded the literary market.

Not all of the works in our corpus exhibited the same pattern in regards to gender and verb use. The observation of outliers during certain tests revealed a number of avenues for future research. Classification using Irish and Scottish literature as a whole resulted in a drastic disparity between the accuracy for each gender. Accuracy dropped even more when using just male authors. While the overall accuracy rate (the average) resulted in an error rate of anywhere from 20-30%, the breakdown between error rate of classifying male pronouns was between 10-15%; the female error rate between 30-40%. In other words, training with Irish and Scottish novels results in only slightly better than average accuracy when testing female characters.

Table 4: Author Classification Results

Results for Specific Authors	
Date and Author	Pronouns Incorrect
Overall Austen	None incorrect
1800-1833 Austen	<i>Emma</i> male <i>Sense</i> male
1834-1866 Austen	<i>Emma</i> male
1867-1899 Austen	None incorrect
Overall Brontë	<i>JaneEyre</i> female <i>Agnes</i> female
1800-1833 Brontë	<i>JaneEyre</i> female
1834-1866 Brontë	<i>JaneEyre</i> female <i>Agnes</i> female
1867-1899 Brontë	<i>JaneEyre</i> female
Overall Eliot	None incorrect
1800-1833 Eliot	<i>SilasMarnor</i> male
1834-1866 Eliot	None incorrect
1867-1899 Eliot	None incorrect

Our corpus contained all seven of Austen's finished, published novels: *Sense and Sensibility* (1811), *Pride and Prejudice* (1813), *Mansfield Park* (1814), *Emma* (1814), *Northanger Abbey* (1815), and *Persuasion* (1818). In addition, our corpus included three of her shorter works, two of which are unfinished: *Lady Susan* (1871), *The Watsons* (1804), and *Sanditon* (1817). We tested all of these novels against the rest of our corpus of 3,500 texts to determine how Austen's works compared to broader literary trends.² In each of her works, Austen's use of verbs describing the behavior of female characters corresponded with overall trends in the corpus: Austen defined her female characters using the same types of verbs her contemporaries used to define their female characters.

Though Austen has recently been championed by many feminist critics, the fact that her portrayal of female action does not radically de-

part from other authors is perhaps not entirely surprising. In *Jane Austen: Irony as Defense and Discovery* Marvin Mudrick observes that Jane Austen used irony to "expose the incongruities between form and fact, all the delusions intrinsic to convention art and conventional society."⁽¹⁾ It stands to reason that as a female author Austen's use of irony was at least partially motivated by the pressure society put on female authors to conform to conventions. Given that Austen used irony and parody as a way to circumvent conventions and establish her own authority as a writer, it stands to reason that her female characters would not necessarily be associated with highly unconventional verbs. While our data suggests that the women in Austen's books were doing the same things as women in other 19th century books, this is not to say that they were doing things in the same way or to the same effect.

Austen's portrayal of male characters however, was not as typical as her portrayal of female characters. In two of her novels, *Emma* and *Northanger Abbey*, male characters do not follow the trends we observed in our larger corpus. We ran our predictive model to test these novels against the corpus as a whole. It should be noted that, because of the random sampling done by the classification tests, these results vary to some degree. In some cases, gendered pronouns in Austen's other works—such as male pronouns in *Sense and Sensibility* or male pronouns in *Lady Susan*—also fail the prediction tests.

Both *Emma* and *Northanger Abbey* include male characters who do not comply with 19th century attitudes about ideal masculinity. Mr. Philip Elton, an unsuccessful suitor in the novel *Emma*, is arrogant, resentful, and pompous. Similarly, John Thorpe, a failed suitor from *Northanger Abbey*, is portrayed as vain and arrogant. It is possible that in creating these characters Austen turned towards verbs convention-

²In addition to the largest data frame created from our pronoun detecting script and the smaller winnowed data frame made by removing gender unique columns, a third data frame was produced by finding the standard deviation of column relative frequencies, and then removing columns that were under the standard deviation of all column standard deviations. Table 4 displays the results from using the second data frame, but very closely matches the results from the largest data frame. The explanation of results is based of off tests done on the smallest data frame. Because there are disparities between the data frames used, part of our future work will be to investigate why these differences in author tests occur.

ally associated with women in order to portray these men as shallow and ineffectual. However, it is worth observing that Austen's other novels also contain examples of ridiculous and arrogant men, such as the memorable William Collins from *Pride and Prejudice*. What then, is particularly unique about Austen's characterization of men in *Emma* and *Northanger Abbey*? Does Austen's use of verbs have something to do with the specific qualities of the characters she is creating? Since we did not look specifically at which individual characters were associated with which verbs there is no way to know whether old men or young men, rich men or poor men, or heroes or villains are in fact being associated with unconventional actions. This initial result however, does point towards several productive avenues for future inquiry.

Our overall corpus also included all four finished novels written by Charlotte Brontë: *Jane Eyre* (1847), *Shirley* (1849), *Villette* (1853), and *The Professor's House* (published posthumously in 1857). What the results testing Brontë's corpus against the classifier indicate is that Brontë's writing is period typical, and does not register as an outlier. None of her novels were misclassified, which, given the existing scholarship, is an unexpected result. From the initial date of publication, readers and reviewers alike noticed that Brontë's characters did not act in the expected ways. Examining the reception of *Jane Eyre* is helpful in highlighting the ways in which Brontë's writing runs up against prevailing opinions of how women ought to have acted. An 1848 review read, in part: "...we think we see 'a beard under the muffler.' Jane Eyre, rare and excellent though she may be, is not to our mind a genuine woman" (LeFavour 117). Her actions, as perceived by readers, make the novel a prime example of what Maureen Corrigan calls "the woman's extreme adventure story":

"The precarious situations described in these female extreme-adventure stories—childbirth, unwanted pregnancies, abortions (legal and illegal), abusive relationships, fatiguing caregiving—are ones that are faced almost

exclusively by women. Their physical ordeals are augmented or even outweighed by heavy emotional burdens...above all, it's the quotidian quality of their pain that separates the women from the boys" (9).

This reconception of action and adventure is useful in that it provides a possible reevaluation of traditional submissive stereotypes. By validating emotional "action," Corrigan's idea paves the way to re-read overarching patterns. Sangeeta Dutta also observed that, in Brontë's fiction as a whole, the central female characters tend to function independently in a "male-defined world" (2312), even if sometimes this independence was mental rather than physical. Her novels can be read as an attempt to legitimize a woman's right to complex, emotionally powerful lives.

Our data, which seems to question these assertions, could lead in several directions. The first possibility is that what strikes readers as remarkable or unusual about Brontë's works is contained in a textual aspect that we were not measuring. Perhaps it is not in their actions that Brontë's characters are distinct, but rather in their descriptions. Perhaps it is not the verbs themselves that are the most indicative of character, but the context which surrounds the verb. At this stage in our research, we are confident that we can track the relationship between verb usage and gender, but what is still required is to determine how fundamental that relationship is for reader reaction and interpretation.

When considering Brontë's results, the other possibility is that singular characters, such as Jane Eyre, are swallowed or normalized by the sheer volume of female pronouns that are associated with more stereotypically anticipated verbs. Though she is the primary female character, Jane is not the only woman in her story; the same goes for Brontë's other novels. Taking a macro view of these books demonstrates that the characters scholars have traditionally been drawn to study do not exist in isolation. Their actions may stand out in a reading of a single text, but not necessarily when taken as a part of a much larger whole.

Our corpus also contained eight novels writ-

ten by George Eliot: *Adam Bede* (1859), *The Mill on the Floss* (1860), *Silas Marner* (1861), *Romola* (1863), *Felix Holt, the Radical* (1866), *Middlemarch* (1871–72), and *Daniel Deronda* (1876). As with our Brontë results, the data for Eliot's novels was somewhat surprising; *Romola* and *Silas Marner* appear as outliers, which opens up additional future avenues for exploration.

Many of the questions produced by our study of Eliot are similar to the questions we have raised about Brontë. Like Brontë, Eliot has been identified as an author interested in pushing on the boundaries of how women appeared in literature. Specifically, Michael Wolff describes the "...tension between the debilitating otherness which was stereotypically attached to women and some form of heroism which [Eliot] thought possible both for herself in life and for the female protagonists of her writing" (28). It is particularly interesting that he invokes the idea of heroism when, by many traditional Victorian accounts, a female heroine would have been an anathema. Combined with our results, this opens up yet another question: does a female writer working towards the feminine heroic approach her male characters any differently? Since the anomalies are in the classification of male characters, it is possible that the answer is yes. Only future examination will tell us if the misclassified characters are more similar in verb use to her female characters, but, given her standing in the criticism, it is an examination worth making.

It is interesting to note that the misclassifications both occurred against the early-19th-century half of the corpus. *Silas Marner* was published in 1861, and *Romola* was published in 1863, so while they do fall in the latter half of the century, it is not by much. It is also important to note that those two novels have wildly different settings. While *Silas Marner* is set in 19th-century northern England, *Romola* is a 15th-century story set in Florence. While it is strange that these two novels, so different on the surface, would return the same type of misclassification, it is also a good confirmation that what we are seeing is some indication of Eliot's style and not just a reflection of outlying

settings.

The other noteworthy feature of the two outliers is that they both prominently feature male clergymen. Since it is the male characters that are not correctly being identified, it could be possible that there is something distinct about the way Eliot writes her male religious figures, even those as different as a Dominican monk and a Calvinist minister. In *Silas Marner*, the religious figures are primarily a function of the background—at the novel's opening, the title character is caring for a sick deacon when he is accused of stealing from his church. That accusation sets in motion the main action of the plot. In *Romola*, the religious men are more central to the plot and some of them are historical figures. Girolamo Savonarola, a key figure in the religious activity of Renaissance Florence, is also key in the novel. In many ways, *Romola* seems to be the obvious outlier in George Eliot's corpus, while *Silas Marner* is a surprising one.

VI. FINAL THOUGHTS

The central tension in our results seems to be between critical reception and what our classifier has indicated is "normal" or "average" for gendered verbs. If traditional critical reception and reaction are all micro-scale activities, then can they be applied in the same manner to a macro project as they can to an additional micro scale project? One of the principal advantages of working with such a large corpus is that we are able to produce results that, left to solely human study, would take years to replicate. The flip side of this, of course, is that no critic is reviewing Eliot, or Brontë, or Dickens, or Twain, with the entire swath of 19th-century novels as a foil. That isn't within the aim or scope of reviewers, or of scholars interested in a particular character, scene, or detail. If we are starting to ask questions about corpuses that are only manageable through machine intervention, then do we also have to change the way we read and incorporate traditional criticism into our research? So far, our results indicate that the answer ought to be a

resounding 'yes.'

REFERENCES

- [1] Clark, Anna. *The Struggle for the Breeches: gender and the making of the British working class*. Berkeley, CA: University of California Press, 1995.
- [2] Corrigan, Maureen. *Leave Me Alone, I'm Reading: Finding and Losing Myself in Books*. New York: Random House, 2005.
- [3] Dutta, Sangeeta. "Charlotte Brontë and the Woman Question." *Economic and Political Weekly* 26.40 (1991): 2311-2313, 2315-2316.
- [4] Eliot, George. "Silly Novels by Lady Novelists" *Westminster Review* 66.1 (1856): 442-461.
- [5] Gilbert, Sandra M. and Susan Gubar. *The madwoman in the attic: the woman writer and the nineteenth-century literary imagination*. New Haven, CT: Yale University Press, 1979.
- [6] Johnson, R. Brimley. *Famous Reviews*. Freeport, NY: For Libraries, 1967. Print.
- [7] Katona, Anna. "The Changing Image of George Eliot." *Hungarian Studies in English* 2.1 (1965): 47-60.
- [8] LaFrance, Marianne, Hiram Brownell, and Eugene Hahn. "Interpersonal Verbs, Gender, and Implicit Causality." *Social Psychology Quarterly* 60.2 (1997): 138-152.
- [9] LeFavour, Cree. "'Jane Eyre' Fever: Deciphering the Astonishing Popular Success of Charlotte Brontë in Antebellum America." *Book History* 7.1 (2004): 113-141.
- [10] Levine, Caroline. "'Harmless Pleasure': Gender, Suspense, and 'Jane Eyre.'" *Victorian Literature and Culture* 28.2 (2000): 275-286.
- [11] Mudrick, Marvin. *Jane Austen: Irony as Defense and Discovery*. Princeton: Princeton UP, 1952. Print.
- [12] Poovey, Mary. *The Proper Lady and the Woman Writer: Ideology as Style in the Works of Mary Wollstonecraft, Mary Shelley, and Jane Austen*. Chicago: U of Chicago, 1984. Print.
- [13] Welter, Barbara. "The Cult of True Womanhood: 1820-1860." *American Quarterly* 18.2 (1966): 151. Print.
- [14] Woolf, Virginia. *A Room of One's Own*. San Diego: Harcourt Brace Jovanovich, 1989. Print.

A. APPENDIX

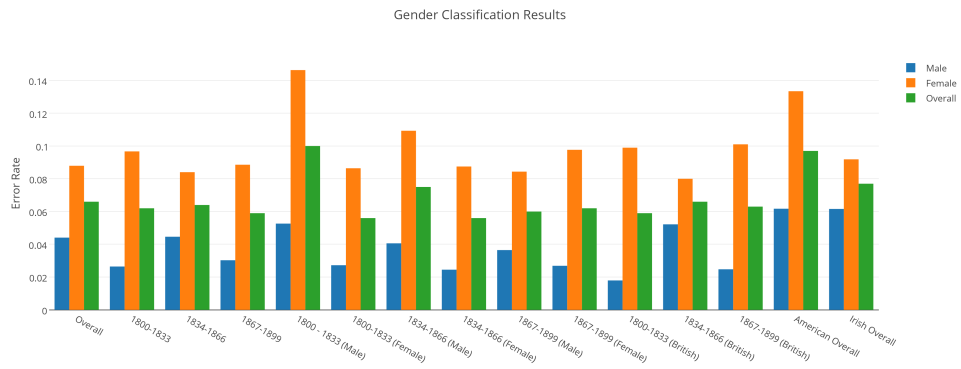


Figure 5: Graph of Error Rates for Various Tests