

NO OTHER THAN A WOMAN'S REASON  
TEXTUAL ANALYSIS OF GENDER IN SHAKESPEARE AND  
ELIZABETHAN DRAMA

---

*Author:*

Katherine ROE

*Supervisor:*

Prof. Bob BERWICK

May 2015

# Contents

# 1 Introduction

William Shakespeare, widely regarded as one of the best playwrights in any language, created hundreds of well-known and timeless characters of both genders. Although many thousands of books and papers have examined the role of gender in Shakespeares plays, very few if any have performed a rigorous quantitative analysis of the texts. For my 6.UAP project, I have performed such an analysis on the lines delivered by men versus the lines delivered by women to determine whether Shakespeare wrote male and female characters in a quantifiably distinct manner. I analyzed the number of unique words, types of speech, sentence length, line length, and sentiment of every line in 36 Shakespeare plays . Additionally, I performed the same analysis on plays by six of Shakespeares contemporaries in order to determine whether Shakespeares treatment of female characters was unique for his time or part of a broader cultural influence. My aim in this project was to write text analysis software that can be used to shed light on subtle aspects of gender treatment in Elizabethan theater and other literature.

## 2 Methods

### 2.1 Text

The first challenge was to represent Shakespeares plays in a way that allowed for data analysis. I started with an HTML version of every play analyzed. I then wrote a script that converted each line into a python object. These objects included the lines play, location in the play, location in the speech, act and scene, speaker, speaker gender, and text.

### 2.2 Analysis

Once the plays were prepared, I wrote a series of Python scripts to analyze the texts along several dimensions. These dimensions were number of words per line, speech length, and number of unique words. Additionally, I used natural language processing Python packages to analyze the parts of speech distribution, subjectivity, and polarity of each line.

COMEDIES	TRAGEDIES	HISTORIES
All's Well That Ends Well	Antony and Cleopatra	King John
As You Like It	Coriolanus	Richard II
Comedy of Errors	Hamlet	Henry IV Part 1
Cymbeline	Julius Caesar	Henry V
Love's Labours Lost	King Lear	Henry VI Part 1
Measure for Measure	Macbeth	Henry VI Part 2
The Merchant of Venice	Othello	Henry VI Part 3
The Merry Wives of Windsor	Romeo and Juliet	Richard III
A Midsummer Night's Dream	Timon of Athens	Henry VIII
Much Ado About Nothing	Titus Andronicus	
Pericles		
Theaming of the Shrew		
The Tempest		
Troilus and Cressida		
Twelfth Night		
Two Gentlemen of Verona		
The Winter's Tale		

Table 1: Shakespeare Play Categorization

## 2.3 A Note on Categorization

For the data analysis, I divided Shakespeare's plays into comedies, histories, and tragedies as shown in Table 1.

# 3 Results

## 3.1 Words Per Line

Shakespeare is well known for writing in iambic pentameter, meaning that, with very few exceptions, each line of his plays include ten syllables. Counting the average number of words per line, therefore, is a good measure for the average

length of a spoken word. I postulated that female characters might have a higher number of average words per line, correlating to shorter words. Contrary to this expectation, men and women speak almost exactly the same number of words per line. In fact, in all the plays, men speak about 7.53 words per line compared to women's 7.41 (about a 1.6% difference). Women consistently spoke fewer words per line than men in all three categories of play, although there were individual plays (such as *Coriolanus*) where the overall trend was reversed. The average number of words per line for men and women in the tragedies are shown in Figure 1.

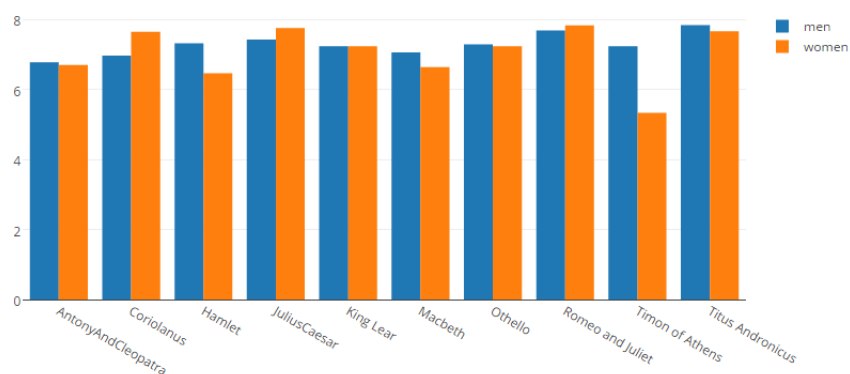


Figure 1: Example image.

## 4 Conclusion

## References

- [Figueredo and Wolf, 2009] Figueredo, A. J. and Wolf, P. S. A. (2009). Assortative pairing and life history strategy - a cross-cultural study. *Human Nature*, 20:317–330.