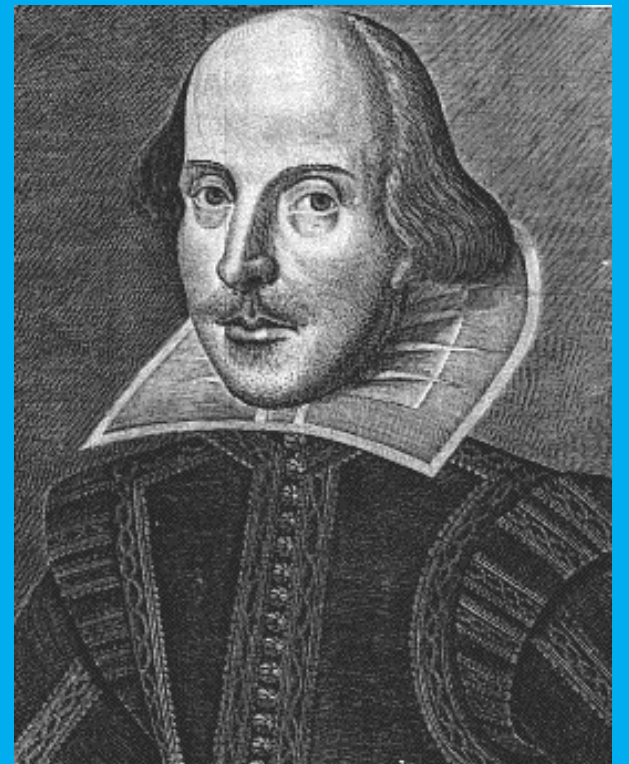


A Computational Stylometric Analysis Of Characterization by Playwrights

Computational Linguistics Group
Intelligent Systems Laboratory
School Of Computer Science and Statistics

Gerard Lynch
gplynch@cs.tcd.ie

Supervisor: Carl Vogel
vogel@cs.tcd.ie



William Shakespeare (1564 -1616)

"The play's the thing, wherein
I'll catch the conscience of
the King"

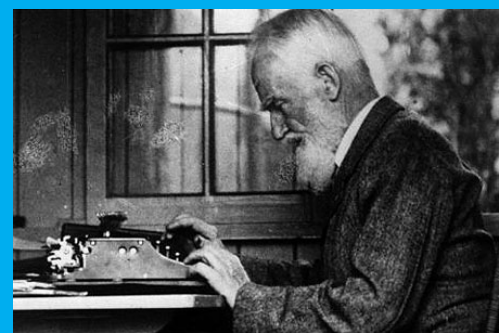
Hamlet

Introduction

Is it possible for an author to create characters truly independent of their own voice? Do characters from the same play exhibit similar characteristics in style? Are some playwrights arguably better at creating unique characters that transcend their own personal style and become individual entities in their own right? Are playwrights who adhere to iambic pentameter or other metrical constraint thereby confined to the construction of characters whose language cannot be differentiated on the basis of the syntax of their language? This study seeks to address these questions and provide more information about what constitutes characterization in plays. Methods from the field of authorship attribution, which are normally used to determine the provenance of anonymous texts are used on the individual text of characters from a number of playwrights.

"That proves it's not
by Shaw, because all
Shaw's characters are
himself: mere puppets
stuck up to spout Shaw"

from *Fanny's First Play*



George Bernard Shaw (1856-1950)

Background and Methods

Previous research in this field focused on finding individual voices in the poetry of Irish poet Brendan Kennelly [2]. The results of this were that individual voices were identifiable across poems and collections, indicating good control over characterization for the author in question. Texts are compared based on the distribution of a selected feature, either word frequencies, letter frequencies, parts of speech(nouns,verbs etc) frequencies or consecutive groups of either letters, words or parts of speech(known as n-grams).

The statistical test known as χ^2 (chi squared) is used to compare the similarity of of two texts, by comparing the distributions of the features. To compare groups of texts with other groups of texts, the Mann-Whitney rank sum test is used.

Plays are divided into the contributions of individual characters using *PlayParser* [1].

Corpus

The playwrights consulted were William Shakespeare, Ben Jonson(a contemporary of Shakespeare), Oscar Wilde and George Bernard Shaw. Due to the fact that more of his work was available, a larger number of Shakespearean works were used in the analysis. 32 Shakespearean dramas were used, with 5 from Oscar Wilde and 4 each from Shaw and Jonson which were all obtained from Project Gutenberg, an online repository of free machine readable literary texts. The main characters from each play were split into five parts and compared against all of the other characters from all playwrights to ascertain whether they were more similar to themselves than any other characters. The same experiments were repeated, this time the individual plays were used as categories. The third experiment involved only four categories, the 4 playwrights examined in the study.



Oscar Wilde (1854-1900)

rather than with the utterances of other characters either in their play or in other plays. In particular, some characters by Jonson and Shaw are found to be distinct, with Shaw's characters also being the least likely to find him as author; another indication of diversity and originality in authorship. Interesting to note is that the famous Shakespearean characters such as *Hamlet* do not show up as strong under this analysis, as they may in studies of dramatic effect, whereas other less famous characters(*Cleopatra* from *Antony and Cleopatra* are found to be strong. This work has been supported by Science Foundation Ireland (RFP 05/RF/CMS002)

References

- [1] Gerard Lynch and Carl Vogel. Automatic Character Assignment. In Frans Coenen Max Bramer and Miltos Petridis, editors, *Research and Development in Intelligent Systems XXIV*, volume 5, pages 335–348, 2007.
- [2] Carl Vogel and Sandrine Brisset. Hearing Voices in the Poetry of Brendan Kennelly. *Belgian Journal of English Language Literature*, pages 1–16, 2007.

Gerard Lynch

Computational Linguistic Group, Intelligent Systems Laboratory,
School of Computer Science and Statistics.

