

Shedding Light on Dickens' Style through Independent Component Analysis and Representativeness and Distinctiveness

Carmen Klaussner

May 11, 2014

Abstract

The present work explores the style of Charles Dickens opposed to both fellow writer Wilkie Collins and a set of contemporary writers using the statistical techniques of Independent Component Analysis and Representativeness & Distinctiveness to find characteristic deviations of word frequencies. The work in part replicates a task treated in a study by Tomoji Tabata in 2012 using Random Forest Classification. Considerable overlap in results of the different methods strongly indicates the presence of individual stylistic elements of the author.

Additionally, the thesis explores different means of evaluation to ensure quality and consistency of the features selected as well as their ability to separate documents by different authors.

Keywords:

Non-traditional Authorship Attribution, Dickens' Style Analysis, Text Classification