

Dr Andrea Nini
University of Manchester

Who wrote the Jack the Ripper letters? A stylometric analysis

authorship analysis, forensic linguistics, stylometry

In 1888, a number of prostitutes were murdered in Whitechapel, London and the perpetrator(s) were never caught. However, in the 209 letters that were received before and after the events, the murderer allegedly identified himself as 'Jack the Ripper', a name that is to this day associated with the case. These letters were key in providing an evocative name to the press and in creating the persona of Jack the Ripper, still alive today in the form of books, movies, plays, and tours. Interestingly, however, historical evidence suggests that the two most important of these letters responsible for the creation of 'Jack the Ripper' were fabricated by journalists with the aim of selling more newspapers, with the later ones probably written by hoaxers after the police decided to make these two public. The present paper reports on a stylometric analysis aimed at identifying which of the 209 letters allegedly attributed to Jack the Ripper were written by the same person. This task has recently been referred to as 'authorship clustering', a special case of 'authorship verification'. Due to the brevity of the Jack the Ripper letters, frequency methods could not be applied and a novel clustering method based on the presence/absence of word 2-grams using the Jaccard distance was applied instead. The results support the hypothesis that the two most important earliest letters were written by the same person and that a third letter, the 'Moab and Midian' letter, can be connected to these two. This letter is controversial as some historical evidence suggests it was fabricated by journalists at the Central News Agency. The implications for these results for both the Jack the Ripper case, its socio-cultural dimension, and modern authorship analysis will be discussed.