

## The Jack the Ripper case and the evidence for idiolectal lexical bundles

Dr Andrea Nini, University of Manchester

The Whitechapel murders case involved the brutal homicide of a number of prostitutes in Whitechapel in London in the autumn of 1888. The perpetrator, commonly known as ‘Jack the Ripper’ was not caught and this mystery has led to several speculations about his identity. During the case, 209 letters signed as ‘Jack the Ripper’ were received but historical evidence suggests that these letters were written by hoaxers. This talk will present the results of an analysis of this ‘Jack the Ripper’ corpus, comprising the totality of the letters associated with the case collected by Evans and Skinner (2001).

The first research question addressed is on the authorship of the two most important of these letters responsible for the creation of the name ‘Jack the Ripper’, which historical evidence suggests were fabricated by journalists. More specifically, the authorship question investigated is whether one single author was responsible for the creation of these earliest letters and whether this author was also responsible for any other of the later letters. Since the Jack the Ripper letters are very short, traditional frequency-based stylometry approaches cannot be used and, instead, a clustering analysis using the Jaccard distance has been adopted, following recent advances in forensic linguistics (Wright 2017; Grant 2013). The results of this analysis support the conclusion that one author was responsible for the two earliest letters and, in all likelihood, for another historically important letter received later.

The second research question revolves around part of the evidence that led to the attribution, a sequence of words seemingly found only in these two Jack the Ripper texts. The question of the nature of this word sequence is investigated using the largest available corpora containing 19<sup>th</sup> century English (COHA, CLMET3, and the Old Bailey Corpus) with a view to understanding the reasons for its distinctiveness. The results of this analysis suggest that this sequence is generated by a series of lexicogrammatical choices that seem to characterise the idiolect of the creator of these earliest ‘Jack the Ripper’ letters. If a normal lexical bundle is a sequence of words that occurs frequently in a register while being widespread enough not to be the result of one idiosyncratic speaker (Biber *et al.*, 1999: 992), we can therefore call this word sequence an *idiolectal lexical bundle*, or a sequence of words that is instead not characteristic of a register but of an individual. The discovery of evidence implying the existence of these bundles raises interesting implications for cognitive and forensic linguistics: their existence is compatible and represents further evidence in favour of usage-based frameworks (e.g. Bybee 2010) while being suggestive of the significant amount of individuality that exists in linguistic production.

Other implications of these results for the Jack the Ripper case and its socio-cultural dimension will also be discussed.

### References

- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *The Longman grammar of spoken and written English*. Harlow: Longman.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge, UK: Cambridge University Press.
- Evans, Stewart P. & Keith Skinner. 2001. *Jack the Ripper: Letters from hell*. Stroud: Sutton.

- Grant, Tim. 2013. TXT 4N6: Method, consistency, and distinctiveness in the analysis of SMS text messages. *Journal of law and policy* 21. 467–494.
- Wright, David. 2017. Using word n-grams to identify authors and idiolects. A corpus approach to a forensic linguistic problem. *International journal of corpus linguistics* 22(2). 212–241.