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 (Grant, 2013; Wright, 2017). 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 found that led to the attribution, a sequence of words seemingly found only in these two Jack the Ripper texts. This question of the nature of this word sequence is investigated using modern and 19th century corpora 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 the earliest 'Jack the Ripper' letters. This result raises interesting implications for cognitive and forensic linguistics, including the possibility that *idiolectal lexical bundles* might exist, consistently with some of the predictions of usage-based frameworks (e.g. Bybee, 2010). Other implications of these results for the Jack the Ripper case, its socio-cultural dimension, and modern linguistics will also be discussed.

References

Bybee, J. (2010) *Language, Usage and Cognition*. Cambridge, UK: Cambridge University Press.

Evans, S. P. and Skinner, K. (2001) *Jack the Ripper: Letters from Hell*. Stroud: Sutton.

Grant, T. (2013) 'TXT 4N6: Method, consistency, and distinctiveness in the analysis of SMS text messages', *Journal of Law and Policy*, 21, pp. 467–494.

Wright, D. (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), pp. 212–241. doi: 10.1075/ijcl.22.2.03wri.