

## Social media and the new frontiers of forensic authorship profiling

Andrea Nini, University of Manchester

Authorship profiling could be defined as the analysis of the language of an anonymous text with the aim of extracting clues that can narrow down the list of suspect authors. Within a forensic linguistic context, this type of analysis is applied in ‘single text problems’, that is, cases of disputed authorship of malicious communication for which a small comparison set of suspect writers is missing (Grant 2008). These cases often involve ransom demands or threatening messages in high-risk situations. Despite the importance of providing intelligence quickly and reliably for cases involving profiling, forensic linguistics has made very little progress in developing objective, reliable and fast methods to tackle these problems. In the majority of the successful cases, forensic linguists have employed *ad hoc* strategies coming from sociolinguistics and dialectology involving extensive manual analyses. Although research in computer science has revealed that it is possible to apply reliable algorithms to profile the anonymous author of a text in terms of their basic demographic details such as gender or age (e.g. Schler et al. 2006, Argamon et al. 2009), the application of these algorithms is not always possible for forensic malicious texts, both because of their register characteristics and their short length. The main problem for forensic authorship profiling is the absence of resources containing sociolinguistic or dialectal information on the use of linguistic structures that can be easily queried automatically. Until a solution to this problem is found, a reliable and general method of profiling the author of an anonymous malicious text will be impossible to develop.

In the present talk, after explaining the state of the art of forensic authorship profiling using real-life cases as examples, it will be proposed that the new frontier of forensic authorship profiling lies in the mining of social media. For the first time in history, linguists can benefit from a massive amount of data often rich in geographical and social details in the form of social media texts published online, such as twitter posts, Facebook posts, forum posts or blogs. By mining the cyberspace for this type of data it is possible to perform on-the-fly and specialised *ad hoc* sociolinguistic studies or dialect surveys that can be key in extracting clues from questioned malicious texts. The problem of the absence of resources can thus be solved using methods to extract sociolinguistic information from large amounts of social media data. The example that will be presented in this talk is the use of geotagged tweets to deduce the geographical origins of the authors of malicious texts using a corpus of about 9 billion words including 890 million tweets collected in 2014. It will be demonstrated that it is possible to extract the dialectal signal of salient lexical items in a text and from that to deduce the most likely geographical origin of an anonymous author.

The talk will outline the progress made so far and suggest new directions. It will be proposed that, besides dialectal information, it is also possible to infer richer socio-demographic details about the anonymous author just using spatial information with the addition of census data. Moreover, it will be pointed out that future research will have to address the problem of the selection of features, which is currently too heavily dependent on the analyst’s expertise.