Small words in an email can reveal a person's identity

October 30 2017

Credit: CC0 Public Domain

It's possible to identify the author of an email by analysing as little as two words, research by Nottingham Trent University suggests.

Dr David Wright, an expert in forensic linguistics, examined thousands of emails to show it's possible to identify someone by analysing small sequences of words and prove them as the author.

The research aims to address the challenges experts face when analysing language evidence in court proceedings or in reports.

Computer scientists use methods such as algorithms and statistical analysis to measure the similarity between texts. However it can be difficult for experts to explain why these techniques could and should distinguish between people's unique writing styles.

As part of the study, Dr Wright analysed thousands of emails from 12 employees at a former energy company and correctly identified authors 95% of the time, when the email samples were longer than around 1,000 words.

He did this by comparing how often employees used particular sequences of words in their emails.

These word sequences were between two and six words long and were as basic as "Please review and let's discuss" and "A clean and redlined version."

The research is based on thousands of emails from American energy company Enron.

More than 1.7 million emails from the company were released into the public domain and have since been used for research purposes.

By analysing these emails, Dr Wright also found that the way people join small words together is unique to them and is influenced by the different speech and writing they are exposed to in their lifetime.

Dr Wright focused on a case study of one employee in the study, who was a lawyer at the company.

He compared their emails against samples from 175 other employees and discovered that their most distinctive phrases were the five sequences of words "A clean and redlined version" and "Please review and let's discuss."

While other lawyers at the company used phrases beginning with "Please review," they didn't use it in exactly the same way as the lawyer, suggesting that these particular clusters of words were unique to themselves.

Dr Wright, of the university's School of Arts and Humanities, said: "The repetitiveness of these phrases shows that the individual has developed their own tried and tested phrases, which they know will work to get a job done while working in their role of a lawyer.

"This shows that when faced with written evidence in cases, of which authorship is disputed, clues to the writer's identity can reside in small, common, everyday phrases. This may lead to improving the reliability of evidence given to the courts, and ultimately the delivery of justice."

Provided by Nottingham Trent University