Introduction

Rui Sousa-Silva & Malcolm Coulthard

University of Porto, Portugal & Federal University of Santa Catarina, Brazil

Language and Law/Linguagem e Direito is delighted to publish this Special Issue devoted to Computational Forensic Linguistics. It contains a balanced collection of six articles by both forensic linguists and computer scientists/computational linguists which are exclusively dedicated to the area. We think the issue will come to be seen as a significant addition to the body of research into forensic linguistics and so we are especially pleased that, as with all articles published in *Language and Law/Linguagem e Direito*, readers will have free access – there is no paywall and no cost for authors!

The work of forensic linguists is nowadays inescapably linked to Computational Forensic Linguistics. Whether the forensic linguistic task consists of analysing the authorship of a questioned document, detecting plagiarism, analysing the (disputed) meaning of a text or text excerpt, investigating courtroom or police discourse, or even translating and interpreting in forensic contexts, a competent application of computational tools and techniques is crucial. Indeed, over the last decades not only has the nature of forensic applications evolved dramatically, but so has the volume of text needing to be analysed increased exponentially. Additionally, thanks to more recent technological developments, a significant proportion of criminal activity has started taking place online, so methods used in the past have had to be constantly updated to handle the new challenges. Computational forensic linguistics is ideally placed to assist forensic linguists address these challenges.

This Special Issue opens with 'Computational Forensic Authorship Analysis: Promises and Pitfalls'. In this article, Shlomo E. Argamon surveys from a practitioner's perspective the different types of computational authorship analysis methods and their components, with a view to ensuring reliability. The author identifies and discusses specifically some of the pitfalls potentially faced by an analyst when applying the methodology, and eventually offers guidance to practitioners.

The issue continues with Andrea Nini's article 'Developing forensic authorship profiling', which approaches authorship profiling in a forensic context. As the author argues, current methods lack the transparency offered by certain computational

techniques, and so fail to meet the standards required for forensic applications. The article reports an experiment he conducted to show how previously established findings related to stylistic variation in English for gender, age and social class also apply to forensic texts. The author concludes by demonstrating the relevance of linguistically-motivated research into forensic authorship profiling.

The volume continues with Sheila Queralt's article 'The creation of Base Rate Knowledge of linguistic variables and the implementation of likelihood ratios to authorship attribution in forensic text comparison', in which she explores the issue of reliability in forensic authorship comparison. In order to guarantee reliability that is comparable with other forensic disciplines, the author proposes the implementation of statistical techniques and argues that such a method assists, not only the courts, but also the linguistic experts.

Patrick Juola then approaches the topic of professionalisation of forensic science through the development of standards and protocols. In his article, entitled 'The Rowling Protocol, Steven Bannon, and Rogue POTUS Staff: a Study in Computational Authorship Attribution', the author applies a systematic protocol for authorship verification (previously used in his analysis of the Rowling case) to another high-profile case: the "Rogue POTUS Staff" (self-described as "The unofficial resistance team inside the White House. We pull back the curtain to expose the real workings inside this disastrous, frightening Administration.").

The next article, 'On the Implications of the General Data Protection Regulation on the Organisation of Evaluation Tasks', by Francisco Rangel and Paolo Rosso, focuses on the challenges that the General Data Protection Regulation (GDPR) of the European Union presents to the organisation of evaluation tasks. As these tasks, which are frequently hosted as part of computational linguistics conferences to test the performance of different computer systems, involve collecting and making available large volumes of data collected from the Internet, in general, and from social media platforms, in particular, they must now meet the stringent requirements of the GDPR. The authors build upon experience gained from the organisation of tasks such as PAN to discuss especially how the collection and distribution of the data used in those tasks comply or fail to comply with European regulations. They propose a methodology to follow when organising such tasks and conclude with a discussion of a practical case.

The volume ends with the article 'Computational Forensic Linguistics: An Overview of Computational Applications in Forensic Contexts', by Rui Sousa-Silva which reviews a significant body of the available literature on computational linguistics approaches that are (or can potentially be) used in forensic linguistics applications. Such applications include authorship analysis, authorship profiling and stylometry, plagiarism detection and analysis and cybercrime, as well as less high-profile applications such as meaning analysis. The article concludes with a discussion of both the potential and the limitations of computational approaches to forensic linguistic analysis, and the future implications for forensic linguistics.

We hope that this collection of articles gives readers an insight into the exciting field of computational forensic linguistics and encourages all those who share an interest in the area to follow one of these lines of research. Others should find in the research presented reasons for employing computational applications in forensic linguistics

casework. Finally, computer scientists (and computational linguists) will hopefully gain a deeper understanding of the challenges driving forensic linguistics research.

We hope you find reading this special issue a rewarding experience – it's been a pleasure editing it!

Rui Sousa-Silva
University of Porto
Portugal

Malcolm Coulthard
Federal University of Santa Catarina
Brazil