

Nota Introdutória

Rui Sousa-Silva & Malcolm Coulthard

Universidade do Porto, Portugal & Universidade Federal de Santa Catarina, Brasil

A *Language and Law/Linguagem e Direito* tem o prazer de publicar este número especial, dedicado à Linguística Forense Computacional. O número inclui seis artigos originais nesta área redigidos, quer por linguistas forenses, quer por linguistas computacionais/especialistas em ciências dos computadores, oferecendo, assim, um conjunto de artigos muito equilibrado. Acreditamos que este número representará um contributo significativo para o volume de publicações e investigação em linguística forense; por isso, é, para nós, um prazer enorme que, tal como acontece com todos os artigos publicados na *Language and Law/Linguagem e Direito*, os leitores tenham acesso gratuito a este número – não existe acesso pago –, tal como não existem custos para os autores!

O trabalho dos linguistas forenses encontra-se, atualmente, inevitavelmente associado à Linguística Forense Computacional. Quer se trate de analisar a autoria de um documento suspeito, quer se trate de detetar plágio, analisar o significado (questionado) de um texto ou expressão, investigar o discurso da sala de audiências ou da polícia, ou, inclusivamente, desempenhar tarefas de tradução ou interpretação em contextos forenses, a utilização competente de técnicas e ferramentas computacionais é essencial para o desempenho das tarefas de linguista forense. De facto, ao longo das últimas décadas assistimos, não só a uma evolução significativa da natureza das aplicações forenses, como também a um aumento exponencial do volume de texto para análise. Além disso, graças aos mais recentes desenvolvimentos tecnológicos, uma parte significativa da atividade criminal passou a decorrer online, pelo que os métodos tradicionalmente utilizados têm de ser constantemente atualizados, de modo a poderem dar resposta aos novos desafios. A linguística forense computacional oferece as condições ideais para ajudar os linguistas forenses a darem essas respostas.

Este número especial abre com ‘Computational Forensic Authorship Analysis: Promises and Pitfalls’, em que Shlomo E. Argamon faz uma revisão dos diferentes tipos de métodos de análise de autoria computacional e dos seus componentes na perspetiva de um profissional, de modo a assegurar a sua fiabilidade. O autor identifica e discute especificamente algumas das armadilhas com as quais os/as analistas se podem

confrontar na aplicação da metodologia, e fornece algumas orientações destinadas a profissionais.

O número continua com o artigo de Andrea Nini ‘Developing forensic authorship profiling’, que aborda a determinação de perfis de autoria em contextos forenses. Como defende o autor, aos métodos utilizados atualmente falta a transparência de determinadas técnicas computacionais, o que não lhes permite satisfazer os requisitos exigidos pelas aplicações forenses. O artigo descreve uma experiência realizada pelo autor que revela de que modo resultados de estudos efetuados anteriormente relacionados com variação estilística em inglês relativos às categorias de género, idade e classe social também são aplicáveis a textos forenses. O autor termina demonstrando a relevância da investigação em perfis de autoria forense motivada linguisticamente.

O volume prossegue com o artigo de Sheila Queralt ‘The creation of Base Rate Knowledge of linguistic variables and the implementation of likelihood ratios to authorship attribution in forensic text comparison’, no qual explora a questão da fiabilidade em comparação de autoria forense. De modo a garantir uma fiabilidade comparável com outras disciplinas forenses, a autora propõe a implementação de técnicas estatísticas e defende que um método deste tipo auxilia, não só os tribunais, mas também os/as peritos/as linguísticos/as.

Patrick Juola aborda, então, o tema da profissionalização da ciência forense através do desenvolvimento de normas e protocolos. No seu artigo, entitulado ‘The Rowling Protocol, Steven Bannon, and Rogue POTUS Staff: a Study in Computational Authorship Attribution’, o autor aplica um protocolo sistemático de verificação de autoria (utilizado anteriormente na sua análise do caso Rowling) a outro caso mediático: o “Rogue POTUS Staff” (que se auto-descreve como “A equipa de resistência não-oficial no interior da Casa Branca. Puxamos as cortinas para expor as verdadeiras operações no seio desta Administração desastrosa e assustadora.”), oferecendo algumas conclusões sobre a análise.

O artigo seguinte, ‘On the Implications of the General Data Protection Regulation on the Organisation of Evaluation Tasks’, da autoria de Francisco Rangel e Paolo Rosso, centra-se nos desafios que o Regulamento Geral de Proteção de Dados (RGPD) da União Europeia representa para a organização de tarefas de avaliação. Uma vez que estas tarefas, normalmente integradas em congressos de linguística computacional com vista a testar o desempenho de diferentes sistemas informáticos perante o mesmo problema, implicam a recolha e disponibilização de grandes quantidades de dados recolhidos da Internet, em geral, e de redes sociais, em particular, é agora obrigatório que cumpram os exigentes requisitos do RGPD. Os autores baseiam-se na experiência decorrente da organização de competições como o PAN para discutir em particular de que modo a recolha e distribuição dos dados utilizados nessas tarefas cumprem ou, pelo contrário, infringem as regulamentações europeias. Os autores propõem uma metodologia a seguir para organizar tarefas deste tipo, terminando o artigo com a discussão de um caso prático.

O número termina com o artigo ‘Computational Forensic Linguistics: An Overview of Computational Applications in Forensic Contexts’, de Rui Sousa-Silva, que apresenta uma revisão de um número significativo de referências bibliográficas sobre abordagens de linguística computacional que são (ou podem potencialmente ser) utilizadas em

aplicações de linguística forense. Entre estas, incluem-se a análise de autoria, perfis de autoria e estilometria, análise e deteção de plágio e cibercrime, bem como aplicações menos conhecidas, como análise de significados. O artigo termina com uma discussão do potencial e das limitações das abordagens computacionais à análise linguística forense, bem como das implicações futuras para a linguística forense.

Esperamos que este conjunto de artigos proporcione aos leitores uma perspetiva aprofundada sobre a empolgante área da linguística forense computacional e incentive todos aqueles com interesse nesta área a seguir uma destas linhas de investigação. Outros deverão encontrar na investigação apresentada motivos suficientes para utilizar aplicações computacionais em casos práticos de linguística forense. Finalmente, os especialistas em ciências dos computadores (e os linguistas computacionais) irão, esperamos, ficar a conhecer mais pormenoridamente os desafios por detrás da investigação em linguística forense.

Esperamos que a leitura deste número especial lhe proporcione uma experiência gratificante; para nós, editores da revista, foi um prazer editá-lo!

Rui Sousa-Silva
Universidade do Porto
Portugal

Malcolm Coulthard
Universidade Federal de Santa Catarina
Brasil