

New Trends in Forensic Linguistics

Tahmineh Tayebi & Malcolm Coulthard

Aston Institute for Forensic Linguistics, Aston University, UK

https://doi.org/10.21747/21833745/lanlaw/9_1ed

Introduction

Editing a special issue on forensic linguistics is not an easy endeavour, not least because of the broad definition of the term ‘forensic’, its disputed origin and the rather obvious fact that, as a field of inquiry, ‘forensic linguistics’ merges the three different academic disciplines of forensic science, law and linguistics. While these challenges could prove daunting, they have actually provided us with the motivation to compile the current Special Issue boosted by a desire to showcase the diverse range of practices and research activities undertaken at the Aston Institute for Forensic Linguistics.

In this editorial introduction to the issue, we have set ourselves the task of providing a brief overview of the field of forensic linguistics, with a special focus on the history of forensic linguistics in the UK. This will be followed by a discussion of current research agendas in the field, and, in particular, the activities that are currently being undertaken at the Aston Institute for Forensic Linguistics. A summary of the articles – all written by current and former members of the Institute – will close this introduction.

The History and Status Quo of Forensic Linguistics

While forensic linguistics broadly covers the intersection of language and the law, its research scope has changed over the last two decades, so it now covers a breadth of topics by exploiting various linguistic tools. While discussing the many recent trends in forensic linguistics is beyond the scope of this introduction, it appears that research conducted under the auspices of forensic linguistics can be broadly treated as belonging to two interrelated strands.

The first strand encompasses hands-on and action-oriented, research-informed ‘practices’ which typically involve, amongst other things, the linguist acting as a practitioner by providing the court of law with expert advice/evidence or by training people involved in the legal process. It is generally agreed that the first example of a forensic linguistic analysis in the UK was Jan Svartvik’s (1968) analysis of the falsified Timothy Evans confession, but Svartvik characterised himself in his report very firmly as a linguist. The early UK cases which brought linguistic analysis to the attention of the public were all concerned with demonstrating police falsification of evidence: the

Appeals of the Birmingham Six, (1991), the Bridgwater Four (1997), and Derek Bentley (1998).

Other examples are the comparative authorship analysis of two high-profile cases involving the disappearance of Danielle Jones in 2001 and Jenny Nicholl in 2005, where linguistic evidence was provided to suggest that a series text messages, sent from the girl's mobile phones after they had disappeared, were unlikely to have written by them. In both of these cases, the linguistic evidence was thought to have been crucial to securing the convictions.

More recently, forensic linguists have been called upon to help police investigators in England and Wales in "taking-over of an individual's online identity for the purposes of intelligence gathering and/or securing an arrest" (MacLeod 2021: 159). In this context, linguists have provided training for undercover police officers, enabling them to identify distinctive features of the victim's linguistic style, so they could impersonate them in online chats with the predator and hopefully set up a meeting (Grant and MacLeod 2018, 2020).

The second strand is about the 'research' that is undertaken not just for the purpose of informing the court of law, but rather with the aim of shedding further light on the various intersections between language and the law. The research strand has traditionally been viewed as comprising three main areas: a) the (written) language of the law, b) the (spoken) language of legal processes and c) authorship analysis and investigative linguistics. However, a fourth area of innovative research is now emerging as a result of the rapidly growing use of modern communication technologies and the Internet (Grant and Tayebi Forthcoming). This strand includes research into different forms of computer-assisted crime such as fraud and deception, cyber harassment/bullying and other forms of criminal behaviour that take place on the dark web, such as child sexual abuse, terrorism, human trafficking and self-harm websites.

What brings research in the latter strand together is the fact that diverse linguistic tools and methods are applied in some form of forensic/legal context, with the aim of either directly or indirectly improving the delivery of justice. Research in this area is motivated by various factors, chief of which are:

- a) previous court cases to learn more about the linguist's role as an expert witness and the ways in which this could be improved (Coulthard 2021; Donlan and Nini 2021; Grant 2017; Grant *et al.* 2022; Picornell 2022),
- b) the use of language at different stages of the judicial process to draw attention to the linguistic and institutional factors and limitations which may hinder the delivery of justice (Coulthard 1996; Fraser 2003; Heffer *et al.* 2013; Haworth 2017, 2018; Heydon 2005; Kredens *et al.* 2021),
- c) questions about the viability and potential of various linguistic tools and methods to address bigger forensically-related issues (Grant 2022; Sousa-Silva 2014; Kredens *et al.* 2019; Wright 2017) and
- d) the investigation of potential online and offline criminal linguistic-type activities, with a view to describing the issue at hand (Culpeper *et al.* 2017; Chiang and Grant 2019), educating and informing current and future preventative measures (Carter 2015; Lorenzo-Dus and Izura 2017; Grant and MacLeod 2020; Lerner 2022).

The research landscape of forensic linguistics is changing as researchers are increasingly making use of insights from neighbouring fields and different areas of linguistics to

conduct research which challenges preconceived notions about the use of language and its perception in the above contexts. Increasing numbers of studies are benefitting from inter/multidisciplinary approaches (McAuliffe *et al.* (2022) and Devine *et al.*, this issue, and Hunter and Grant, this issue); computational techniques (Sousa-Silva (2018) and Busso, this issue); and experimental designs (Deamer *et al.*, this issue). Furthermore, methodological frameworks and analytical tools have evolved in line with the requirements of these new topics (see Picornell *et al.* 2022. One particular trend is the use of established techniques of corpus linguistics (CL), which has come a long way since it was first used by Coulthard (1994) in the authorship analysis of Derek Bentley's disputed statement (see Wright 2021 and Gillings 2022 for a review of CL in forensic linguistics). Today, with the advancements made in corpus techniques and improved data collection and storage, corpus linguistics is being increasingly applied to the analysis of statutory interpretation (Solan and Gales 2017), Native Language Influence Detection (NLID) (Kredens *et al.* 2019; Perkins and Grant 2018, and Mojedano Batel, this issue) and authorship analysis (Nini 2018; Grant 2013). Furthermore, corpus methods are increasingly being used to complement qualitative (critical) discourse and pragmatic analysis in the study of legal language (Mattioli and McAuliffe 2021, and Harrington *et al.*, this issue), courtroom language (Tkacukova 2015) and various forms of online aggressive and criminal behaviours (Krendel *et al.* 2021; Parvaresh and Tayebi 2018).

In addition, the use of larger data collections that can be subjected to a new level of scrutiny and inquiry – something which may not have been possible previously due to the lack of established databanks – is finding its way into the analysis of forensic linguistics. While accessing relevant forensic linguistic data has always been notoriously difficult for researchers, with the emergence of new databanks (see Petyko *et al.*, this issue), more research can now be conducted using previously used and validated corpora to answer some of the more current challenging issues relating to forensic linguistics.

It should also be noted that, while there is a general consensus that both strands of forensic linguistics exist in the UK, it is not easy to identify when forensic linguistic practices actually began or who was the first person to self-identify as a forensic linguist. Regardless of who initiated what, it is undeniable that nowadays forensic linguistics has turned into a vibrant, sought-after and popular field of inquiry. One place that brings all the aforementioned activities under one roof is the Aston Institute for Forensic Linguistics, which was founded in 2019. The Institute evolved from the former Aston Centre for Forensic Linguistics, which was founded in 2008.

The Aston Institute for Forensic Linguistics (AIFL)

As a vibrant research institute with the mission of *improving the delivery of justice through the analysis of language*, AIFL is home to five research centres, each engaged in a wide range of research. The Centre for Forensic Text Analysis has as its main objective to uncover individual variations in language use with a view to enhancing forensic author identification. To achieve this aim, special attention is paid to the notions of idiolect and genre and their relevance to authorship analysis. The Centre for Language and Law conducts research on the intersection of language and the law, in particular the problems that are experienced in the digital age with a focus on inequalities and injustice. The Centre for Spoken Interaction in Legal Contexts focuses primarily on police interviews and courtroom interaction. The Centre for Forensic Data Science

(CFDS) is involved in a wide range of projects aimed at improving the way in which forensic evidence in general is evaluated. In the linguistic areas the main concern is improving the computerised analysis of voice samples. Finally, the Forensic Linguistic Databank (FoLD), provides researchers with varying degrees of access to such data as 'malicious communication data', 'investigative interview data' and 'forensic evidence validation data', which involve both spoken and written texts.

The articles in this Special Issue are written by current and former members of AIFL, and showcase some of the work that has been undertaken at the Institute. They indicate the depth and breadth of research topics and the range of methodological tools being used, in particular the use of big data and corpora as well as computational linguistic analysis.

The Articles

The issue begins with an article by Petyko, Busso, Grant and Atkins which introduces the Aston Forensic Linguistic Databank (FoLD). The article starts by explaining the reasons why a forensic linguistic databank is needed and what has hindered its production to date. The article introduces FoLD and explains the workings of this permanent, online repository for forensic linguistic data, as well as the levels of access that are available, ranging from open access to severely controlled. The article also discusses plans for the improvement and augmentation of the Database and explicitly encourages other research centres to deposit their data in FoLD, not only to enrich the database itself, but through so doing to help other researchers, particularly those beginning their careers. The article ends by assessing the likely impact of the Database on the discipline of forensic linguistics

The next article, by Deamer, Richardson, Basu and Haworth, is a survey-based experimental study which tests the hypothesis that the method of transcribing police-suspect interviews can affect the quality of officially produced evidential documents. Based on results from one of the first experimental studies in the field, the authors demonstrate that when a reader is presented with a transcript of a police interview, interpretations can differ because of the reader's perceptions of the addressee and of the context and these perceptions can in turn be crucially affected by the nature of the transcript. The findings indicate that perceived emotion and even overt markers of hesitation and pauses are important factors in how the reader interprets the interview. This is particularly significant as these two features are not currently routinely included in police transcripts, which could lead to transcripts mis-representing meanings of the original audio recordings. This article is one of the first to reveal how experimental studies can be used to answer important forensic linguistic questions.

The third article is by Hunter and Grant. It addresses the relatively unexplored topic of how an individual's psychological traits could have an impact on how they experience the world around them. The article aims to develop a multidisciplinary approach by bridging the gap between psychological approaches and linguistic theories. To achieve this, the authors adopt what is known in linguistics as the Appraisal Framework to analyse texts written by a number of serial murderers that have been diagnosed with mental health problems. The analysis reveals that appraisal theory could be used to capture patterns in the linguistic choices made by serial murderers. The identified patterns are then further examined in relation to the psychological traits of each author

and the connection between them are also scrutinised.

The next article is by Grieshofer (previously Tkacukova), who uses the ethnography of communication as the main methodological framework to explore narrativisation practices in small claims cases and private family proceedings. By drawing on real data collected during court observations, Grieshofer explains the narrative genres that exist across the different stages involved in legal proceedings and how lay court users experience communication barriers. The study also discusses the importance of an open narrative strategy to ensure that the court user's voice is heard during the initial stages of proceedings. She also touches on the link between the notion of voice projection and procedural justice, and suggests that "main narratives should be elicited sooner as part of an open narrative strategy".

Continuing on from the theme of family proceedings, Devine, Parker, Harrington and Makouar's article introduces corpus-assisted linguistic methods as an exploratory means of analysing reports written by expert psychologists that are used in public family law (child protection) cases. The main purpose of this article is to explore the viability of corpus linguistics as a core method of analysis. The study has an interdisciplinary design in that it "uses an inductive (data-driven) approach to identify significant themes in the reports, and a deductive (legal-intuitive) approach to explore psychologists' use of legally significant terms". The study confirms the need for corpus-informed investigations in the analysis of written expert evidence.

The next three articles, while addressing different topics, all have corpus and computational analysis as part of their methodology. Mojedano Batel, Abrams and Pezik focus on the issue of Native Language Influence Detection (NLID). The main objective of the study is to take NLID to another level by investigating whether an author's L2 (English) features can reveal not simply their first language (L1), in this case Spanish, but their native dialect, in this study either Mexican or Peninsular Spanish. The analysis reveals that distinctive L1 features, such as use of punctuation marks, and choice of adjectives of affect and intensifiers, do indeed transfer to L2, thus the identification of culturally determined usages allows dialect identification. This study devised a computational analysis to automatically classify test data samples and it achieved an accuracy of 69%. These results demonstrate that NLID is possible at dialect level.

Busso's article also uses corpus analysis, this time to study the lexicogrammar of English legal-lay language. More specifically, the study uses collocation analysis – which is "an extension of traditional collocational analysis using Construction Grammar tenets" – and English core vocabulary as a reference. The results of the quantitative analysis show that legal-lay language differs from both of its neighbouring genres, suggesting that it might be best labelled as a 'blended' genre. The qualitative analysis, which uses readability metrics and surveys English participants, indicates that legal-lay language lies between legal jargon and general-domain prose, and, in terms of its complexity, is at an intermediate level. This study ends with the important finding that "the readability scores indicate that speakers consider legal-lay language more accessible than text-based metrics seem to suggest".

Finally, Harrington, Parker, Devine and Makouar offer the first empirical linguistic examination of commercial bail bond discourse. Drawing on a specialised corpus of 'Home' and 'About Us' pages from bail bond websites, this article uses corpus-

assisted critical discourse analysis to examine how bail bond companies use a variety of discursive practices to sell their products and services, thereby presenting the services they offer as ‘normal’. The study backs up its linguistic analysis by grounding it in a legal context. The authors argue that the way in which the bail bond discourses are constructed tends to both “serve whilst oppress those they purport to help”, and “subtly perpetuate neoliberal agendas and a two-tier justice system”.

Conclusion

All in all, forensic linguistics as a field of inquiry appears to be in its prime, as indicated by the ever-increasing number of projects, the growing number of students taking courses in forensic linguistics, of linguistics graduates with relevant backgrounds seeking employment in forensic-related contexts and the increasing call for linguists as experts in courtrooms. Written by current and former members of one of the world’s leading research institutes for forensic linguistics, the articles in this Special Issue are a testament to the vitality and strength of forensic linguistics in improving, amongst other things, how we conceptualise and understand the relationship between language and the law, all with an ultimate aim of helping ‘to improve the delivery of justice’ and indeed much more.

References

- Carter, E. (2015). The anatomy of scam communications: An empirical analysis. *Crime, Media, Culture*, 11(2), 89–103.
- Chiang, E. and Grant, T. (2019). Deceptive identity performance: Offender moves and multiple identities in online child abuse conversations. *Applied Linguistics*, 40(4), 675–698.
- Coulthard, M. (1994). On the use of corpora in the analysis of forensic texts. *International Journal of Speech, Language and the Law*, 1(1), 27–43.
- Coulthard, M. (1996). The official version: Audience manipulation in police records of interviews with suspects. In C. Caldas-Coulthard and M. Coulthard, Eds., *Texts and Practices: Readings in Critical Discourse Analysis*. London & New York: Routledge, 166–178.
- Coulthard, M. (2021). Experts and opinions: In my opinion. In M. Coulthard, A. May and R. Sousa-Silva, Eds., *The Routledge Handbook of Forensic Linguistics*. London & New York: Routledge, 2 ed., 523–538.
- Culpeper, J., Iganski, P. and Sweiry, A. (2017). Linguistic impoliteness and religiously aggravated hate crime in England and Wales. *Journal of Language Aggression and Conflict*, 5(1), 1–29.
- Donlan, L. and Nini, A. (2021). A forensic authorship analysis of the Ayia Napa rape statement. In I. Picornell, R. Perkins and M. Coulthard, Eds., *Methodologies and Challenges in Forensic Linguistics Casework*. NJ & West Sussex: Wiley Blackwell.
- Fraser, H. (2003). Issues in transcription: factors affecting the reliability of transcripts as evidence in legal cases. *Forensic Linguistics*, 10(2), 203–226.
- Gillings, M. (2022). How to use corpus linguistics in forensic linguistics. In A. O’Keeffe and M. McCarthy, Eds., *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 589–601.
- Grant, T. (2013). Txt 4n6: method, consistency, and distinctiveness in the analysis of sms text messages. *JL & Pol’y*, 21, 467.

- Grant, T. (2017). Duppying yoots in a dog eat dog world, kmt: Determining the senses of slang terms for the courts. *Semiotica*, 2017(216), 479–495.
- Grant, T. (2022). *Te Idea of Progress in Forensic authorship analysis*. Cambridge: Cambridge University Press.
- Grant, T., Grieve, J., Picornell, I., Perkins, R. and Coulthard, M. (2022). The starbuck case: Methods for addressing confirmation bias in forensic authorship analysis. In I. Picornell, R. Perkins and M. Coulthard, Eds., *Methodologies and Challenges in Forensic Linguistic Casework*. NJ & West Sussex: Wiley Blackwell, 13–28.
- Grant, T. and MacLeod, N. (2018). Resources and constraints in linguistic identity performance—a theory of authorship. *Language and LawLinguagem e Direito*, 5(1), 80–96.
- Grant, T. and MacLeod, N. (2020). *Language and Online Identities: The Undercover Policing of Internet Sexual Crime*. Cambridge: Cambridge University Press.
- Grant, T. and Tayebi, T. (Forthcoming). Forensic linguistics. In L. Wei, Z. Hua and J. Simpson, Eds., *The Routledge Handbook of Applied Linguistics*. London & New York: Routledge.
- Haworth, K. (2017). The discursive construction of evidence in police interviews: Case study of a rape suspect. *Applied Linguistics*, 38(2), 194–214.
- Haworth, K. (2018). Tapes, transcripts and trials: The routine contamination of police interview evidence. *The International Journal of Evidence & Proof*, 22(4), 428–450.
- C. Heffer, F. Rock and J. Conley, Eds. (2013). *Legal-lay communication: Textual travels in the law*. Oxford: Oxford University Press.
- Heydon, G. (2005). *The language of police interviewing. Hampshire*. London: Palgrave Macmillan.
- Kredens, K., Monteoliva-García, E. and Morris, R. (2021). Interpreting outside the courtroom: ‘a shattered mirror?’ interpreting in law enforcement contexts outside the courtroom. In M. Coulthard, A. May and R. Sousa-Silva, Eds., *The Routledge Handbook of Forensic Linguistics*. London & New York: Routledge, 2 ed., 502–520.
- Kredens, K., Perkins, R. and Grant, T. (2019). Developing a framework for the explanation of interlingual features for native and other language influence detection. *Language and LawLinguagem e Direito*, 6(2), 10–23.
- Krendel, A., McGlashan, M. and Koller, V. (2021). The representation of gendered social actors across five manosphere communities on reddit. *Corpora*, 17(2).
- Larner, S. (2022). Facilitating children’s informal disclosures of sexual abuse: The role of online counsellors at a national children’s helpline. *Journal of child sexual abuse*, 31(3), 276–296.
- Lorenzo-Dus, N. and Izura, C. (2017). “cause ur special”: Understanding trust and complimenting behaviour in online grooming discourse. *Journal of Pragmatics*, 112, 68–82.
- MacLeod, N. (2021). Assuming identities online: Authorship synthesis in undercover investigations. In M. Coulthard, A. May and R. Sousa-Silva, Eds., *The Routledge Handbook of Forensic Linguistics*. London & New York: Routledge, 159–173.
- Mattioli, V. and McAuliffe, K. (2021). A corpus-based study on opinions of advocates general of the court of justice of the european union: changes in language and style. *International Journal of Legal Discourse*, 6(1), 87–111.
- McAuliffe, K., Mattioli, V. and Muntean, L. (2022). Through the lens of language: Using interviews and corpus linguistics analysis to examine opinions of advocates general.

- In M. R. Madsen, F. Nicola and A. Vauchez, Eds., *Researching the European Court of Justice: Methodological Shifts and Law's Embeddedness*. Cambridge: Cambridge University Press, 158–186.
- Nini, A. (2018). An authorship analysis of the jack the ripper letters. *Digital Scholarship in the Humanities*, 33(3), 621–636.
- Parvaresh, V. and Tayebi, T. (2018). Impoliteness, aggression and the moral order. *Journal of Pragmatics*, 132(91–107).
- Perkins, R. and Grant, T. (2018). Native language influence detection for forensic authorship analysis. identifying 11 persian bloggers. *International Journal of Speech Language and the Law*, 25(1), 1–20.
- Picornell, I. (2022). Casework in forensic linguistics: Looking outward. In I. Picornell, R. Perkins and M. Coulthard, Eds., *Methodologies and Challenges in Forensic Linguistic Casework*. NJ & West Sussex: Wiley Blackwell, 162.
- I. Picornell, R. Perkins and M. Coulthard, Eds. (2022). *Methodologies and Challenges in Forensic Linguistic Casework*. NJ & West Sussex: Wiley Blackwell.
- Solan, L. M. and Gales, T. (2017). Corpus linguistics as a tool in legal interpretation. *BYU Law Review*, 1311.
- Sousa-Silva, R. (2014). Detecting translingual plagiarism and the backlash against translation plagiarists. *Language and Law/Linguagem e Direito*, 1(1), 70–94.
- Sousa-Silva, R. (2018). Computational forensic linguistics: an overview of computational applications in forensic contexts. *Language and Law/Linguagem e Direito*, 5(2), 118–143.
- Svartvik, J. (1968). *The Evans statements: a case for forensic linguistics*. Göteborg: University of Göteborg.
- Tkacukova, T. (2015). A corpus-assisted study of the discourse marker 'well' as an indicator of institutional roles: Professional and lay use in court cases with litigants in person. *Corpora*, 10(2), 145–170.
- 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), 212–241.
- Wright, D. (2021). Corpus approaches to forensic linguistics: Applying corpus data and techniques in forensic contexts. In M. Coulthard, A. May and R. Sousa-Silva, Eds., *The Routledge Handbook of Forensic Linguistics*. London & New York: Routledge, 611–627.

Tahmineh Tayebi & Malcolm Coulthard
Aston Institute for Forensic Linguistics, Aston University
UK