Language and Law

Linguagem e Direito

VOLUME 9.1

ISSN 2183-3745

Language and Law
Linguagem e Direito

ISSN: 2183-3745 (online)
Volume 9, Issue 1, 2022

Publisher / Editora
Faculdade de Letras da Universidade do Porto

This publication was supported by project UIDB/00022/2020, FCT-Fundação para a Ciência e Tecnologia, Portugal, through national funds.

Esta publicação é financiada por fundos nacionais através da FCT – Fundação para a Ciência e a Tecnologia, I.P., no âmbito do projeto UIDB/00022/2020.

Language and Law / Linguagem e Direito

Language and Law / Linguagem e Direito is a free, exclusively online peer-reviewed journal published twice a year. It is available on the website of the Faculty of Arts of the University of Porto, at http://ojs.letras.up.pt/index.php/LLLD/ .

All articles should be submitted via the system. See the guidelines for submission at the journal website.

Requests for book reviews should be sent to rssilva@letras.up.pt.

Language and Law / Linguagem e Direito é uma revista gratuita publicada exclusivamente online, sujeita a revisão por pares, publicada semestralmente e disponível no website da Faculdade de Letras da Universidade do Porto, em http://ojs.letras.up.pt/index.php/LLLD/ .

Os materiais para publicação deverão ser enviados através do sistema, e devem seguir as instruções disponíveis na página da revista.

As propostas de recensão de livros devem ser enviadas para rssilva@letras.up.pt.

Indexing and abstracting / Indexação e bases de dados bibliográficas

Language and Law / Linguagem e Direito is covered by the following abstracting and indexing services:

A *Language and Law / Linguagem e Direito* encontra-se indexada e catalogada nas seguintes bases de dados:

CrossRef
EBSCO Legal Source
ERIH PLUS: European Reference Index for the Humanities and the Social Sciences
Google Scholar
Journals for Free
JURN
Portal RCAAP
QUALIS Periódicos (A2)
Sherpa Romeo

PUBLISHED BIANNUALLY ONLINE / PUBLICAÇÃO SEMESTRAL ONLINE

ISSN: 2183-3745
D.O.I.: https://doi.org/10.21747/21833745/lanlaw
D.O.I. Volume 9(1): https://doi.org/10.21747/21833745/lanlaw/9_1

Contents / Índice

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; Larner 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, amd 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 filed 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 Pęzik 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 collostructional 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 infuence 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 l1 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 LawLinguagem 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

The Aston Forensic Linguistic Databank (FoLD)

Marton Petyko, Lucia Busso, Tim Grant & Sarah Atkins

Aston Institute for Forensic Linguistics, Aston University, UK

https://doi.org/10.21747/21833745/lanlaw/9_1a1

Abstract. *The Aston Forensic Linguistic Databank (FoLD) is a permanent, controlled access online repository for forensic linguistic data. We broadly understand forensic linguistics as any academic research with a potential to improve the delivery of justice through the analysis of language. FoLD thus comprises a wide range of datasets with relevance to forensic linguistics and language and law, including commercial extortion letters, investigative interviews in police and other contexts, legal documents, forum posts from far-right online groups, and comment threads from political blogs. This paper outlines how FoLD works and its potential impact on the general discipline of forensic linguistics.*

Keywords: *Databank, Repository, Resource, Data sharing, Data access.*

Resumo. *O Aston Forensic Linguistic Databank (FoLD) é um repositório online de dados de linguística forense, de acesso restrito. Entendemos linguística forense no seu sentido lato, em que toda a investigação académica possui potencial para melhorar a administração de justiça através da análise da linguagem. O FoLD inclui, assim, uma série de "datasets" relevantes para a linguística forense linguagem e direito, incluindo cartas comerciais de extorção, entrevistas policiais e noutros contextos, documentos legais, publicações de grupos de extrema-direita em fóruns online e comentários em blogs políticos. Este artigo apresenta o funcionamento do FoLD e discute o seu potencial impacto na área da linguística forense em geral.*

Palavras-chave: *Banco de dados, Repositório, Recurso, Partilha de dados, Acesso a dados.*

Introduction – What is FoLD?

This paper introduces the Aston Forensic Linguistic Databank (FoLD), and explores the academic rationale and decision-making behind its provision and design. FoLD is a permanent, controlled access online repository for forensic linguistic data (available at fold.aston.ac.uk). FoLD has been developed in the Aston Institute for Forensic Linguistics (AIFL)[1] at Aston University, Birmingham, UK since 2019[2] and it was officially launched at the 15th Conference of the International Association of Forensic Linguists (IAFL) in

September 2021. FoLD is an innovative resource that makes it easier for researchers around the world to access a variety of forensic linguistic datasets. Furthermore, FoLD also accepts data contributions from researchers external from Aston, with the aim of populating a wide and diversified collection of forensic linguistic datasets for the advancement of the discipline.

In this paper we will outline how FoLD works and its potential impact on the general discipline of forensic linguistics. Specifically, the second looks at why openly available forensic linguistic datasets are scarce, why more forensic linguistic data is needed and how FoLD can contribute to addressing this challenge. The third section focuses on existing data sharing practices in corpus linguistics and computational linguistics, two areas that have continuing methodological impact on forensic linguistics. The fourth section has a particular focus on the policies and procedures that inform data submission and data access. Finally, the last section discusses our future plans for FoLD.

Why is FoLD needed? – Access to data in forensic linguistics

In the broadest sense, forensic linguistics can be defined as the application of linguistic knowledge, theory, and methods to legal and criminal contexts (Perkins 2021: 68) with the aim to improve the delivery of justice through the analysis of language (MacLeod and Grant 2017: 173). Although it is evident that forensic linguistics can only fulfil this aim by conducting evidence-based, reliable, and replicable research, access to relevant forensic linguistic data has been notoriously challenging since the conception of the discipline in the 1960s (Larner 2018).

The need for specific datasets for forensic linguistic analysis was recognised in the first edition of the journal then known as *Forensic Linguistics*, which changed its name to *Speech Language and Law*. Coulthard (1994: 29) gives mention to the setting up of a corpus to be "whimsically entitled the Habeas Corpus, which will include suicide notes, threatening letters, transcriptions of threatening and obscene telephone calls, court transactions, witness statements and police interview records".

Unfortunately, this corpus was never built but the plan is striking in the breadth of genre under consideration for such a corpus. Within the call is perhaps the recognition that different types of forensic linguistic analysis have different corpus requirements. On the one hand, analysis of evidential texts, such as 'suicide notes' and 'threatening letters' is considered to be required perhaps to contribute to investigative forensic linguistics and/or providing language as evidence. On the other hand, provision of 'court transactions, witness statements and police interview records' might be used to address the study of linguistics in legal contexts and its institutional practices. Notwithstanding various efforts including databases of threats, such as the now defunct *CTARC* and *FTARC* databases (see e.g., Gales (2011)), and more recent efforts, such as *TextCrimes* (held at AIFL) and *ForensicLing.com* (an open-source collection of resources maintained by Tammy Gales), this early call for a need for such texts has never been fully met.

The urgency for provision of forensic texts has grown. Within investigative forensic linguistics, there is growing recognition (Grant 2022) for the need of validation corpora to meet the needs of, for example, the English and Welsh Forensic Science Regulator (2020), or the United States NAS: National Research Council of the [United States] National Academy of Sciences' Committee on Identifying the Needs of the Forensic Sciences Community (2009) / PCAST: President's Council of Advisors on Science and

Technology (2016) critiques of forensic science. With growing recognition of the challenge of cross genre authorship analysis (Litvinova *et al.* 2018) it may also be that validation is required across different genres of forensic texts as used in evidence. Furthermore, with increased interest in academic forensic linguistic analysis beyond investigative casework, there is commensurate need for spoken and written texts from within the legal system, to enable description and analysis in areas of language and law and critical linguistic approaches to forensic and legal interactions.

The need for more forensic linguistic data is therefore clear. Yet, for a variety of reasons, data sharing is not a regular practice in forensic linguistics, which is a problem that hinders the entire discipline. The scarcity of freely available forensic linguistic data is due to at least three, partly related issues. First, forensic linguistic data often comes from highly institutionalised external parties, such as courts or law enforcement agencies, who must adhere to strict requirements on data protection. Building a working relationship with these external partners and going through the administrative processes necessary to transfer and share data takes a large amount of time and effort, meaning that data collection in forensic linguistics is perhaps more challenging and time-consuming than in many other fields of linguistics. In addition, these external partners can normally only provide the research data with legally binding restrictions on how it can be used, which often prevents researchers from sharing the data with others.

Second, many forensic linguistic datasets, such as dark web fora dedicated to the illegal exchange of child sexual exploitation and abuse material, threatening communications, extortion letters, hate speech corpora, police interviews, and courtroom discourse, contain highly sensitive or disturbing data, which poses an important ethical dilemma for researchers. Whilst data-sharing is seen as good academic practice, publishing the datasets, such as the abovementioned ones, without any restrictions on access runs the risk of making data subjects from these highly sensitive contexts potentially identifiable or may cause serious psychological or other harms to researchers. For these reasons, forensic linguists understandably often abandon the idea of publishing these datasets altogether.

Finally, open access data sharing platforms are simply not tailored to the discipline-specific needs of forensic linguists. Due to the challenging nature of many forensic linguistic datasets, a blanket open-access approach to data sharing in forensic linguistics is unrealistic. Researchers are understandably reluctant to publish their datasets if they are unable to control who gains access to their data and how it is used, but existing standard platforms, such as university repositories or the UK Data Service, do not always have the capacity and discipline-specific expertise to develop the policies and procedures needed for providing controlled access to highly sensitive datasets.

The main purpose of FoLD is to address the challenge posed by the scarcity of available and reusable forensic linguistic datasets by providing a permanent, controlled access platform for sharing all kinds of forensic linguistic data. The uniqueness of FoLD originates from the fact that it is developed and maintained by the Aston Institute for Forensic Linguistics, which has the capacity and discipline-specific expertise to tailor the repository to the needs of the wider forensic linguistic community.

Why not use an existing platform? – Data sharing in corpus linguistics and computational linguistics

Research data, as primary source material that supports academic enquiry and technical analysis, are the foundation of most linguistic projects. The digital age has seen growth not only in the amount of data available to researchers, but also in the infrastructure for sharing data between researchers, with the drive for 'open science' meaning that many studies now make the data that underpins their work available to others.

Since both corpus linguistics and computational linguistics are concerned with the computer-assisted analysis of large and structured collections of electronically stored texts (McEnery and Hardie 2011: 2), the importance of data sharing has long been emphasised in these fields (Ädel 2020: 16). Practices of data sharing developed in corpus linguistics and computational linguistics are potentially relevant to forensic linguistics since these areas have had a continuing methodological impact on forensic linguistics (Coulthard 1994; Cotterill 2010; Wright 2020).

The practice of publishing corpora is considered good academic practice in these fields for at least three reasons. Firstly, the accuracy and reliability of any empirical study can only be scrutinised by other researchers if they can access the data that the findings are based upon (Resnik and Shamoo 2017). This drive for 'open science' and good practice in data sharing is a key reason that many universities, funders and academic journals now expect or at least encourage researchers to make their data freely available (Corti *et al.* 2019). Secondly, building a corpus (or indeed any dataset) takes time and effort, but only published corpora can be reused by others for their own research projects. Finally, some corpus techniques, such as keyword analysis, require large reference corpora, which are often unfeasible to compile for the purposes of a single research project (Pojanapunya and Watson Todd 2018).

Several online platforms have been developed over the years to publish corpora and other datasets (Mieskes 2017). For example, both the *British National Corpus* and the *National Corpus of Contemporary Welsh* have their own web-based interface called *BNCweb* and *CorCenCC online*, respectively, while the web-based corpus analysis system *CQPweb* provides access to more than a hundred corpora. Some other well-known and widely used online platforms for accessing corpora from multiple languages include *Sketch Engine* (Kilgarriff *et al.* 2004), the *Oxford Text Archive*, and *CLARIN*, just to name a few. These platforms often include not only language resources but also software tools for the preparation, collection, and management of corpora.

In the field of computational linguistics, instead, data sharing is often practiced by individual scholars – rather than at the institutional level. Computational linguists who want to make their code or datasets available mostly gravitate toward the use of online public platforms, such as the *Open Science Framework (OSF)* (Foster and Deardorff 2017) or *GitHub* (https://github.com/). However, it is also increasingly common for individual researchers to build their own interfaces and repositories to hold and share specific types of data, such as corpora, datasets, or pipelines. In turn, the presence of this incredible amount of data available online generates unprecedented large-scale linguistic research, and the creation of new platforms and databases (for a discussion see inter alia Bender and Good (2010); Forkel *et al.* (2018)).

Creating new, publicly available resources is an academic endeavour and as such researchers publish peer-reviewed papers to present the new resources to the academic world. For example, the journal *Language Resources and Evaluation* is "the first publication devoted to the acquisition, creation, annotation, and use of language resources, together with methods for evaluation of resources, technologies, and applications." (journal website). In parallel, the biennial conference *LREC* (Language Resources and Evaluation Conference) reunites scholars working in language technologies to share and present advancements in the field.

Academic publications on computational linguistic resources tend to fall into one of two broad categories. Most commonly, resources are built with the specific aim to answer a theoretically motivated research question, or at the very least are grounded in a specific theory/framework. Hence, resulting publications mainly deal with the theoretical implications, analysis tools provided, and case studies (see among many others for example Pustejovsky *et al.* (2017); Petruck (2018); Brunato *et al.* (2020). However, there is also another type of publication for available datasets and resources. This second line of literature describes more in detail the architecture of the resource itself, providing flowcharts and more technical details on the process of data sharing per se. In other words, these publications describe the pipeline, dataset or website and its internal workings and mechanisms (see among others Reichel *et al.* (2016); Shu *et al.* (2020)).

The sheer existence of a wide range of existing data sharing platforms raises the question why forensic linguists would need their own platform in the first place? Whilst developing a new platform arguably requires much more resource than using an existing one, we argue that FoLD is needed because several challenges affecting data sharing in forensic linguistics have not been resolved in corpus linguistics or computational linguistics.

One of the most frequently mentioned challenges is how to deal with sensitive data in linguistic datasets and corpora (Rock 2001; Anthony 2013; Leedham *et al.* 2021). In the social sciences generally, and particularly perhaps in linguistics, commonly used data, such as written texts, recordings of interactions and interviews, nearly always originate from individuals, which becomes particularly problematic when using data from sensitive contexts. A handful of qualitative studies have met these challenges for data-sharing through explicitly consenting data subjects for the use of their data and outlining the level of anonymity that will be achieved, but such methods are not always appropriate or feasible in forensic and legal contexts, particularly when working with secondary data from external organisations, and there are no direct research 'participants' in the traditional sense.

In the context of corpus linguistics, sensitive data almost exclusively refers to personal information of named individuals, such as names, addresses, phone numbers and other contact details (Leedham *et al.* 2021). The standard method for mitigating this problem is anonymisation, i.e., replacing personal information with standard placeholders in a corpus (Rock 2001). Anonymisation is widely used in corpus linguistics, especially in published language resources but it also presents its own challenges. Due to the sheer size of many present-day corpora, manual anonymisation or the manual inspection of the output of automated anonymisation tools has become unfeasible, which

means that it is practically impossible to ensure that all personal information has been removed from published corpora (Baker 2018).

An equally important but less often discussed challenge is how to publish resources that contain distressing or otherwise sensitive language with the potential to cause harm to the researcher or others (de Maiti and Fiser 2021). When facing this challenge, researchers generally follow one of three strategies. The first strategy involves the complete avoidance of publishing corpora containing distressing language. This practice of course minimises the risk of causing harm, but it also prevents other researchers from reusing a potentially valuable resource for their own projects. The second strategy is to remove all potentially distressing language from the corpus and share a heavily redacted version of the dataset with others. This approach again minimises the risk of causing harm at the expense of the integrity and academic usefulness of the corpus. Finally, rather than publishing the corpus as an open-access resource, researchers can provide others with controlled, often heavily restricted access to the dataset they have compiled. This option, however, requires substantial policy and procedure development as the data owner needs to define who can gain access to the corpus and under what circumstances. Policy development is a challenging endeavour in its own right, which is probably one of the reasons why the above-mentioned online corpus platforms do not provide an effective environment for sharing sensitive or potentially distressing language resources with others. This in turn means that researchers often stay on the safe side by not publishing their corpora in any form if they feel there is potential risk involved.

How does FoLD work?

FoLD as an online platform

FoLD (fold.aston.ac.uk) is a simple and easy-to-use online repository that provides access to forensic linguistic datasets (Figure 1). At AIFL, we broadly understand forensic linguistics as any academic research with the potential to improve the delivery of justice through the analysis of language (see Gibbons and Turell (2008); Coulthard *et al.* (2011); Rock (2011)). FoLD thus comprises a wide range of datasets with relevance to the "the application of linguistic knowledge and theory to forensic, legal, or criminal contexts" (Perkins 2021: 68).

Each dataset is represented as a separate tile and has its own dataset page on the website (Figure 2). Dataset pages provide detailed information about every dataset FoLD holds, including a title, a 50-word summary, a 50–500-word detailed description focusing on the content, structure and collection methods of the dataset, up to 5 subject keywords, data collection start and end dates, the language(s) of the dataset, data type(s) (written, spoken-audio, spoken-video, spoken-transcript, and other), access category (open, restricted, controlled, and external; see subsection ***Access categories*** for details), publication license, name and affiliation of the data donor(s), funding information, the AIFL research centre (e.g., Centre for Forensic Text Analysis) that the dataset has most relevance to, and the date of upload. We are also planning to provide every dataset with a Digital Object Identifier (DOI) and a standard citation so that they can be properly referenced in academic publications.

As explained in more detail in subsection ***Access categories***, the access category of a dataset determines whether or not the data files themselves are directly available from the dataset page. Data files for open datasets can be downloaded directly from the

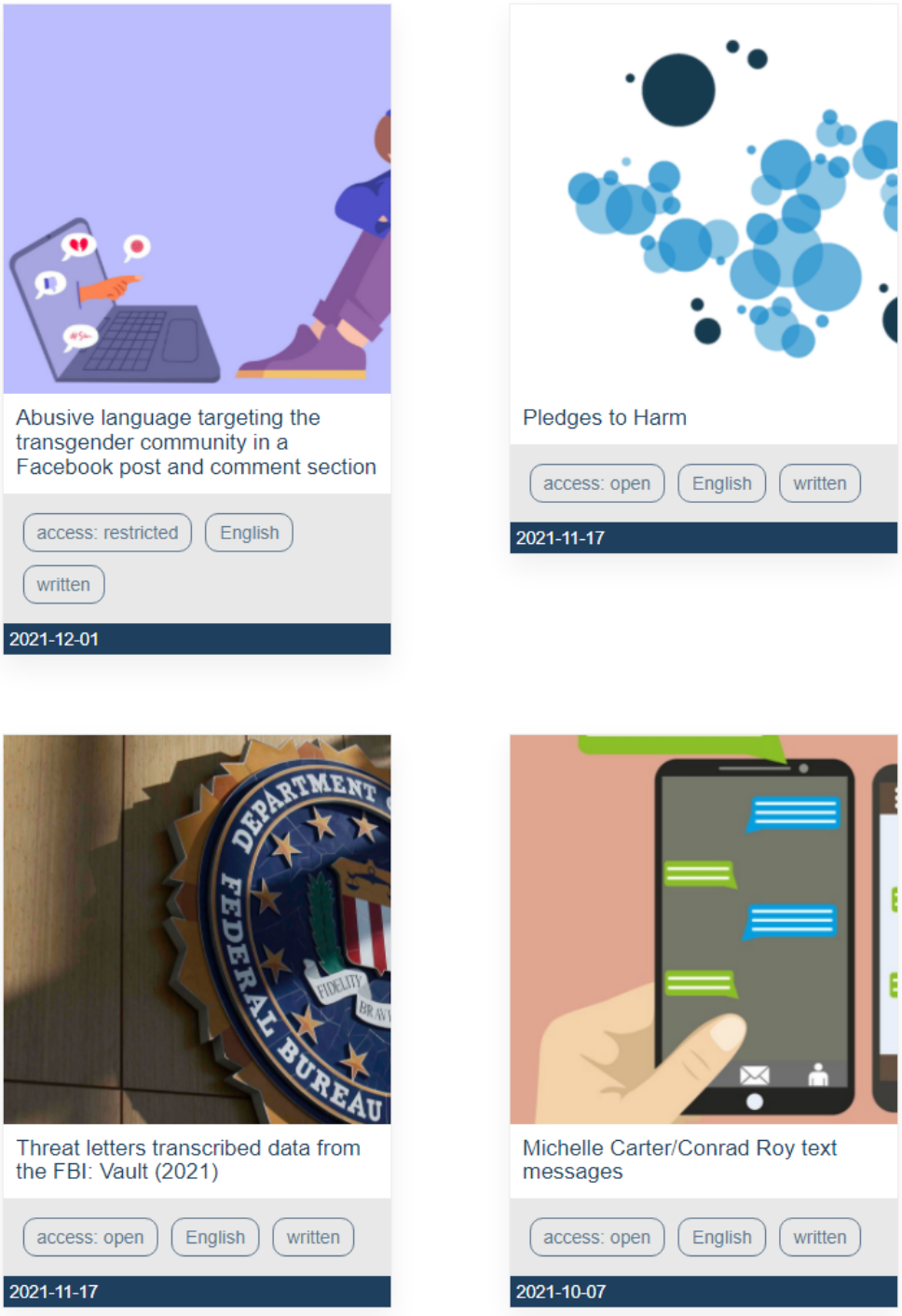

Figure 1. Dataset tiles on FoLD

Operation Heron corpus

Uploaded: 2021-09-29
Access category: Open
Languages: English
Email: t.d.grant@aston.ac.uk
Collected from: 2020
To: 2020

Summary

The dataset is a subset of the abusive letters sent by Margaret Walkers to individuals in the public eye between 2007 and 2009.

Subject keywords: corpus, letter series, abusive language
Data types: Written
Funders: N/A
Associated AIFL centres: Forensic Linguistic Databank (FoLD)
License: Non-Commercial Government Licence for public sector information

Description

The corpus represents an incomplete subset of the entire collection of abusive letters sent by Margaret Walkers between 2007 and 2012. The present datasets spans from January 2007 to April 2009, and consists of 50 letters and 49 envelopes, amounting to 10,650 tokens. The letters are directed to private individuals (50%), healthcare professionals – especially doctors (28%) and to other categories such as Imams, city county officials, hairdressers etc. (22%). The single files are coded with metadata about the file itself: the original number of the document (as given by the police), the type of document (letter or envelope), and the date (in the format MONTH/YEAR). For example, letter n. 1 sent to a medical doctor in January 2007 is coded as "01_letter_doctor_012007". associated publication: Busso, L., Petyko, M., Atkins, S., & Grant, T. (2022). Operation Heron–Latent topic changes in an abusive letter series. Corpora 17(2)

Data Donors

Name: Lucia Busso
Affiliation: Aston Institute for Forensic Linguistics
Link: https://research.aston.ac.uk/en/persons/lucia-busso

Name: Márton Petykó
Affiliation: Aston Institute for Forensic Linguistics
Link: https://research.aston.ac.uk/en/persons/marton-petyko

Name: Sarah Atkins
Affiliation: Aston Institute for Forensic Linguistics
Link: https://research.aston.ac.uk/en/persons/sarah-atkins

Name: Tim Grant
Affiliation: Aston Institute for Forensic Linguistics
Link: https://research.aston.ac.uk/en/persons/tim-grant

Files

Here are the files submitted for this Item.

View/Open
corpus separate env + letters.zip (56.12Kb)

Figure 2. FoLD dataset page

dataset page. In contrast, dataset pages for restricted or controlled datasets only provide information on how users can gain access to the data files by submitting a data access request (see subsection ***Data requests*** for details). In addition, dataset pages for external datasets provide a link to the external website where users can access the data.

Finally, there are ethical considerations for a repository, such as FoLD, that go beyond the potential harm to data subjects – but also the potential for material to cause distress to researchers and users of the site. A great deal of data used in forensic linguistics is likely to contain content that can be disturbing to those that view it. Responsibilities to provide content warnings for such material has been well argued in teaching contexts (e.g., Stringer 2016) and there is a case for providing similar warnings in a repository, so that users have a choice before downloading material.

Access categories

Every dataset in FoLD falls into one of four access categories: open, restricted, controlled, or external. When reviewing newly submitted datasets, the FoLD editorial team makes a decision as to whether the dataset is in scope for FoLD, whether copyright may prevent publications and the most suitable access category for each dataset. Where necessary, for example, if the dataset contains sensitive material, the editorial team can also seek advice from the Aston Institute for Forensic Linguistics Research Ethics Committee. Data donors have the option to select a provisional access category for their own dataset and explain why they think the selected category is the most appropriate, but this category needs to be ultimately approved through the editorial process before the dataset gets published.

The access category of a dataset has significant implications on how the dataset is stored and how researchers can use it.

Open datasets are stored on the FoLD web server, are freely available for download from the FoLD website, and users do not need to be registered on the website or submit a data access request to gain access to these datasets. As a result, researchers can use any software tools of their choice to analyse open datasets. FoLD holds various open datasets reflecting a variety of research interests within and outside the Aston Institute of Forensic Linguistics (AIFL). For example, corpora of legal language (Busso forthcoming; Makouar 2021), datasets of trolling in comment threads (Petykó 2019), police interrogations (Szczekulska and Haworth 2021), or data from old casework such as the *Operation Heron Corpus* (Busso *et al.* forthcoming). FoLD also holds the *Drill Rap Slang Glossary* and the *Youth Slang Glossary* compiled by Professor Tony Thorne (Thorne 2014a,b).

Given the profound importance of data sharing in academia in general and in forensic linguistics in particular, we aim to make as many datasets openly available as possible. However, providing full and unrestricted access to some datasets is not possible because of ethical concerns, copyright or license issues, or constraints established by a Data Sharing Agreement. This is why we have introduced the restricted and controlled access categories.

Similarly to datasets in the open category, restricted datasets are also stored on the FoLD web server. However, they are not immediately available for download, but require some checks to be made on the user's research purpose before being made available for use. Users must register with a FoLD account, submit a data access request

describing their research purpose, and agree and sign up to the terms of use (which may vary according to different conditions under donors Data Sharing Agreements, or under European and UK law Data Sharing Impact Assessments). Once approved, users will be able to download and use the dataset from the FoLD website. In line with these arrangements, users who gain access to restricted datasets are required to adhere to the terms of use when conducting their research project and publishing the results. Restricted datasets currently held on FoLD (as of March 2022) include data from old casework, such as the Amanda Birks' case text messages (Grant 2012), or the 100 Idiolects Dataset, a multichannel corpus of 100+ English speakers used for research on cross-genre idiolect (Heini *et al.* 2021).

Datasets with controlled access contain highly sensitive material that may come from a third party and have even heavier constraints on access and use. Controlled datasets are therefore stored not on the FoLD web server but on an air-gapped, offline computer in our secure data lab at the Aston Institute for Forensic Linguistics. Users who wish to access these datasets must make a detailed application to FoLD and the data owner, as well as potentially gain additional agreement from an external organisation before they can be approved for access. Although information on controlled datasets is detailed in the FoLD repository for users to search, the data itself is not available for download and users may need to visit Aston or agree a secure means of access. Datasets in this category include, for example, scraped data from white supremacist and dark-web child abuse discussion fora (Kredens and Pezik 2021b,a).

Finally, we also use FoLD to signpost the availability of external datasets that are managed by third parties and are not directly available via the FoLD website. These datasets might be open-access or have various access restrictions at the data owner's discretion. External datasets include for example other data repositories, such as the *ForensicLing.com* website managed by Tammy Gales of Hofstra University.

Data submission

As of December 2021, FoLD holds around 20 datasets which have been donated to us by AIFL academic and research staff, PhD and Masters students, and external researchers. FoLD is completely unrestrictive in the sense that anyone who is willing to share their forensic linguistic datasets via FoLD can become a data donor. Given that our aim is to publish as many datasets as possible, we are constantly looking for new relevant datasets, and we actively encourage potential data donors from within and outside AIFL to contact us at aifl_fold@aston.ac.uk to discuss the suitability of their datasets for inclusion into FoLD.

Publishing a dataset via FoLD has several benefits for data donors. Firstly, data donors retain full ownership of their datasets and can withdraw them from the databank at any time if they wish to do so. We never change the content, structure, or access category of any dataset we hold on FoLD without the data donor's explicit permission. At the same time, we provide a permanent, controlled access platform for the datasets and data donors can work with us to set the access category, publication license and any limitations on use. We also help data donors with the everyday practicalities of storing their datasets online. For example, data donors can regularly update their datasets while we ensure that the integrity of the datasets remains intact.

Secondly, FoLD provides online visibility for the datasets donated to us. We work actively to promote FoLD across the global forensic linguistics community to ensure that our data donors' work reaches those interested, thus maximising the impact potential of their datasets on research and teaching.

Finally, many funders require datasets collected for research projects to be deposited in a repository and made available for other researchers. As explained in section ***Why is FoLD needed? – Access to data in forensic linguistics***, historically this has been difficult in forensic linguistics because of the sensitive nature of much of the data it uses. FoLD aims to make data-sharing possible through managing different access categories, meaning that forensic linguists can meet the requirements of their funders' research data policy by using a platform that is tailored to their discipline-specific needs.

As mentioned above, we invite all potential data donors to have an informal discussion with the FoLD team about the suitability of their dataset for inclusion before they submit the dataset to the FoLD website. During this discussion, we establish whether the dataset holds relevance to forensic linguistics and whether there are any obvious ethical or licensing issues that would prevent us from publishing the dataset via FoLD.

Once we have established that the dataset is in principle publishable on FoLD, we ask the data donor to submit their dataset by completing an online form available on the FoLD website (http://fold.aston.ac.uk/static/documents/item_submission_form.pdf). Data donors are required to provide the following information about their datasets: a title, a 50-word summary, a 50–500-word detailed description outlining the content, structure and collection methods of the dataset, up to 5 subject keywords, data collection start and end dates, the language(s) of the dataset, data type(s) (written, spoken-audio, spoken-video, spoken-transcript, and other), access category (open, restricted, controlled, or external), publication license, name, affiliation and contact details of the data donor, funding information, and – if relevant – the associated AIFL research centre (e.g., Centre for Forensic Text Analysis) that the dataset belongs to.

Given that the access category is one of the key considerations that we make when publishing datasets, we also ask data donors to provide a rationale for their requested access category. If the data donor wishes to publish an open dataset with us, we ask them to upload the data files as well. For restricted and controlled datasets, on the other hand, we ask data donors to share their dataset with us for review via other means. Finally, we also ask all data donors to upload an image for their dataset (distributed with Creative Commons licence), which is used on the data tiles and dataset pages to make the datasets visually more recognisable.

Once the dataset has been submitted, we carry out our editorial review before making the data available on FoLD. During this process, we check the metadata and the data files for any inaccuracies and decide on the most suitable access category for the dataset under review. If we have any ethical concerns about the dataset that we are unable to resolve, we seek advice from the AIFL Research Ethics Committee, which works independently from the FoLD team. Whilst we normally follow the advice received from the Ethics Committee, all publication decisions lie with and are the responsibility of the FoLD team.

The ethical aspect of the editorial process is of great importance to a specialist repository such as FoLD. As stated above, sharing and publishing research data through a repository poses ethical and some legal challenges, with the potential for harm to individuals or communities should they become identifiable or if data were to be misused. These challenges prompted many early decisions by the FoLD team about the structure and management of the repository.

The gold standard for sharing data beyond a project is usually to consent research participants for the storage and reuse of their anonymised data (ESRC 2015). Where researchers at Aston University were conducting projects with participants, a description of this ongoing use was indeed provided in information sheets and consent forms, for example:

"During the project your data will be anonymised and will become part of a larger, anonymised dataset. At the end of the project this dataset will be made available in an open access repository (Forensic Linguistic Databank (FoLD)) in the Aston Institute for Forensic Linguistics. Once in the repository, your data will not be able to be re-identified and will be made available for reuse."

However, it is acknowledged that such consent or anonymisation may not exist for pre-existing datasets or even be feasible for many types of research data and in these instances careful consideration and potentially a review by a research ethics committee is needed before a dataset is archived and reused (Summers *et al.* 2019). This issue has perhaps been particularly acute in forensic linguistic contexts, where data may not originally have been intended for research, might be sourced from criminal or sensitive contexts, or be provided by another organisation that requires strict limits on use – long held difficulties for sharing data in the field.

As a means of mitigating these difficulties, an early decision was made to provide datasets through layered access categories (detailed in subsection ***Access-categories***), whereby more complex datasets would not be published openly but could be restricted to authorised researchers, who apply with a particular purpose and can demonstrate ethical approval at their own research institution. This is a process mirrored by the UK Data Service for historic or sensitive datasets such as Dodds *et al.* (2017) (see case study 7.5 in Summers *et al.* (2019)).

Once our editorial review is complete, we inform the data donor about the outcome and we actively liaise with them to reach an agreement on the final version of the metadata, such as the access category and the wording of the title, summary, and description. As a principle, we only publish a dataset if and when the data donor and the FoLD team have managed to agree on the metadata. We also take precautions to ensure that we do not force our opinion about the nature of dataset on the data donor whilst we only accept datasets that we are comfortable to publish. Data donors retain full ownership of their datasets. Finally, data donors can submit new versions of their already accepted datasets. These new versions need to go through the same albeit occasionally simplified editorial process.

Data requests

As explained in subsections ***Data requests*** and ***Access categories***, users who wish to gain access to controlled or restricted datasets need to register on the website and submit an online data access request (http://fold.aston.ac.uk/static/documents/item_request\

_form.pdf). Although the data access request form that users need to complete is the same for both access categories, controlled and restricted datasets differ in that researchers are only allowed to work on controlled datasets on an air-gapped computer in our secure data lab at Aston University while restricted datasets are eventually released to the named requester upon approval.

When completing the data access request form, users need to provide various information about themselves and their research project. This includes the researcher's name, affiliation, academic position or student status, the details of a signatory from their host institution who has the authority to agree to and sign an access agreement on behalf of the requester's host institution, details about the requester's proposed research project with particular focus on the proposed use of the data, plans for publishing and disseminating findings of the project, and ethical approval from the requester's host institution.

Similarly to data submissions, all data access requests are reviewed by the FoLD team. As a team, it was agreed this review should purely assess a legitimate research purpose and not appraise the perceived quality or stance of the research methods. In case of ethical concerns about an application, the editorial team may consult the AIFL Research Ethics Committee and make a decision in line with the conditions or restrictions set by the data donor. When the dataset comes from an external organisation, said organisation might also need to approve access. After all enquiries have been satisfied, we inform the requester about the outcome. If access is granted by all parties, the researcher is then asked to sign an agreement, along with a signatory from either a supervisor (in the case of a student) or the research office at their institution, consenting to the terms of use for the data and any particular limitations in how it can be stored, used and published on. If we are unable to approve the initial request, we provide feedback to the researcher to help them make the necessary adjustments to their proposal and invite them to submit a new request. This practice is to ensure that users, especially students, are not penalised for previously submitted unsuccessful access requests.

Conclusions – What will happen to FoLD now?

Our future plans for FoLD are twofold. Firstly, the website is still under development. Currently, users can manually browse the available datasets, but we are working towards developing a more sophisticated search facility to ensure that users can easily find the datasets they are looking for. We are also in the process to provide all datasets with a Digital Object Identifier (DOI) and a standard citation so that they can be properly referenced in academic publications. In addition, we aim to streamline access to controlled datasets to enable researchers access these highly sensitive but extremely valuable datasets without any unnecessary inconvenience. As mentioned earlier, our utmost priority is to tailor FoLD to the needs of its potential users, including academics and students. In line with this aim, we encourage members of the forensic linguistic community to contact us with their ideas on how FoLD can be improved.

Secondly, whilst FoLD is an innovative data sharing platform that builds on the collective expertise of the AIFL members, it can only become a truly useful resource for forensic linguists and strengthen the discipline if it holds a wide range of relevant forensic linguistic datasets, enabling researchers to build on each other's work. We therefore encourage everyone who wish to share their forensic linguistic datasets with

others to consider publishing them with us. After all, FoLD can only be as good as the datasets it holds.

Notes

[1]The Aston Institute for Forensic Linguistics (AIFL) was founded in 2019. It is a substantial expansion of the former Aston Centre for Forensic Linguistics that was founded in 2008. This expansion was funded via a £6M investment including a £5.4M award from Research England's Expanding Excellence in England (E3) fund.

[2]The FoLD online interface and server is managed by Aston University's student-led software enterprise *Beautiful Canoe* (https://beautifulcanoe.com/)

References

Anthony, L. (2013). A critical look at software tools in corpus linguistics. *Linguistic Research*, 30(2), 141–161.

Baker, P. (2018). Language, sexuality and corpus linguistics. concerns and future directions. *Journal of Language and Sexuality*, 7(2), 263–279.

Bender, E. and Good, J. (2010). A grand challenge for linguistics: Scaling up and integrating models. In *White paper contributed to NSF's SBE 2020 initiative*. 1–5.

Brunato, D., Cimino, A., Dell'Orletta, F., Venturi, G. and Montemagni, S. (2020). Profiling-ud: a tool for linguistic profiling of texts. In *Proceedings of The 12th Language Resources and Evaluation Conference*, 7145–7151: European Language Resources Association.

Busso, L. (forthcoming). Coriells: A specialised bilingual corpus of lay legal communication. In C. Meluzzi and S. Cenceschi, Eds., *Proceedings of the workshop "Linguistica forense: dalla ricerca empirica alla pratica legale"*.

Busso, L., Petykó, M., Atkins, S. and Grant, T. (forthcoming). Operation heron – latent topic changes in an abusive letter series. *Corpora*, 17(2). http://publications.aston.ac.uk/id/eprint/42497/.

Corti, L., Eynden, V., Bishop, L. and Woollard, M. (2019). *Managing and sharing research data: a guide to good practice*. Sage.

Cotterill, J. (2010). How to use corpus linguistics in forensic linguistics. In A. O'Keeffe and M. McCarthy, Eds., *The Routledge Handbook of Corpus Linguistics*. Routledge, 578–590.

Coulthard, M. (1994). On the use of corpora in the analysis of forensic texts. *Forensic Linguistics*, 1(1), 27–43.

Coulthard, M., Grant, T. and Kredens, K. (2011). Forensic linguistics. In R. Wodak, B. Johnstone and P. Kerswill, Eds., *The SAGE handbook of sociolinguistics*. Sage, 529–544.

de Maiti, K. and Fiser, D. (2021). Working with socially unacceptable discourse online: Researchers' perspective on distressing data. In I. Heindrickx, L. Verheijen and L. Wijngaert, Eds., *Proceedings of the 8th Conference on CMC and Social Media Corpora for the Humanities (CMC-Corpora 2021)*, 78–82: Radboud University.

Dodds, C., Keogh, P. and Weatherburn, P. (2017). *Accessing HIV post-exposure prophylaxis: gay and bisexual men in the UK describe their experiences*. Data Collection. Colchester, Essex, United Kingdom: UK Data Service.

ESRC, (2015). Esrc framework for research ethics, economic and social research council. https://esrc.ukri.org/files/funding/guidance-for-applicants/esrc-framework-for-research-ethics-2015/.

Forkel, R., List, J. and Greenhill, S. (2018). Cross-linguistic data formats, advancing data sharing and re-use in comparative linguistics. *Scientific Data*, 5(180205), 1–10.

Foster, E. and Deardorff, A. (2017). Open science framework (osf. *Journal of the Medical Library Association*, 105(2), 203–206.

Gales, T. (2011). Identifying interpersonal stance in threatening discourse: An appraisal analysis. *Discourse Studies*, 13(1), 27–46.

J. Gibbons and M. Turell, Eds. (2008). *Dimensions of Forensic Linguistics*. John Benjamins.

Grant, T. (2012). Txt 4n6: method, consistency, and distinctiveness in the analysis of sms text messages. *Journal of Law and Policy*, 21(2), 467–494.

Grant, T. (2022). *On the idea of progress in forensic authorship analysis*. Cambridge University Press.

Heini, A., Pezik, P. and Kredens, K. (2021). The 100 idiolects project. the aston forensic linguistic databank. http://fold.aston.ac.uk/handle/123456789/17.

Kilgarriff, A., Rychly, P., Smrz, P. and Tugwell, D. (2004). The sketch engine. In G. Williams and S. Vessier, Eds., *Proceedings of the 11th EURALEX International Congress*, 105–116: Université de Bretagne-Sud.

Kredens, K. and Pezik, P. (2021a). Child sexual abuse dark-web discussion forum corpus. the aston forensic linguistic databank. http://fold.aston.ac.uk/handle/123456789/20.

Kredens, K. and Pezik, P. (2021b). White supremacist discussion forum corpus. the aston forensic linguistic databank. http://fold.aston.ac.uk/handle/123456789/19.

Larner, S. (2018). Forensic linguistics. In A. Phakiti, P. Costa, L. Plonsky and S. Starfield, Eds., *The Palgrave Handbook of Applied Linguistics Research Methodology*. Palgrave Macmillan, 703–718.

Leedham, M., Lillis, T. and Twiner, A. (2021). Creating a corpus of sensitive and hard-to-access texts: Methodological challenges and ethical concerns in the building of the xxx corpus. *Applied Corpus Linguistics*, 1(3), 1–10.

Litvinova, T., Seredin, P., Litvinova, O., Dankova, T. and Zagorovskaya, O. (2018). On the stability of some idiolectal features. In *International Conference on Speech and Computer*, 331–336: Springer.

MacLeod, N. and Grant, T. (2017). go on cam but dnt be dirty": linguistic levels of identity assumption in undercover online operations against child sex abusers. *Language and Law/Linguagem e Direito*, 4(2), 157–175.

Makouar, N. (2021). Cyberhate regulation in france: understanding the semantic negotiation and dynamics in parliamentary debates. In *Presentation at 15th Biennial Conference of the International Association of Forensic Linguists*, Birmingham, UK: Aston University.

McEnery, T. and Hardie, A. (2011). *Corpus Linguistics. Method, Theory and Practice*. Cambridge University Press.

Mieskes, M. (2017). A quantitative study of data in the nlp community. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, 23–29.

NAS: National Research Council of the [United States] National Academy of Sciences' Committee on Identifying the Needs of the Forensic Sciences Community, (2009). Strengthening forensic science in the united states: a path forward. https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf.

PCAST: President's Council of Advisors on Science and Technology, (2016). Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

Perkins, C. (2021). The application of forensic linguistics in cybercrime investigations. *Policing: A Journal of Policy and Practice*, 15(1), 68–78.

M. Petruck, Ed. (2018). *MetaNet*, volume 100. John Benjamins.

Petykó, M. (2019). *The discursive construction of trolling on British and Hungarian political blogs*. Lancaster University.

Pojanapunya, P. and Watson Todd, R. (2018). Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*, 14(1), 133–167.

Pustejovsky, J., Ide, N., Verhagen, M. and Suderman, K. (2017). Enhancing access to media collections and archives using computational linguistic tools. In *CDH@ TLT*. 19–28.

Reichel, U., Schiel, F., Kisler, T., Draxler, C. and Pörner, N. (2016). The BAS speech data repository. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 786–791.

Resnik, D. and Shamoo, A. (2017). Reproducibility and research integrity. *Accountability in Research*, 24(2), 116–123.

Rock, F. (2001). Policy and practice in the anonymisation of linguistic data. *International Journal of Corpus Linguistics*, 6(1), 1–26.

Rock, F. (2011). Forensic linguistics. In J. Simpson, Ed., *The Routledge Handbook of Applied Linguistics*. Routledge, 158–172.

Shu, K., Mahudeswaran, D., Wang, S., Lee, D. and Liu, H. (2020). Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big data*, 8(3), 171–188.

Stringer, R. (2016). Reflection from the field. trigger warnings in university teaching. *Women's Studies Journal*, 30(2), 62–66.

Summers, S., Bishop, L., Eynden, V. and Corti, L. (2019). Legal and ethical considerations in sharing data. In L. Corti, V. Eynden, L. Bishop and M. Woollard, Eds., *Managing and Sharing Research Data*. Sage, 2nd ed.

Szczekulska, A. and Haworth, K. (2021). *US police-suspect interrogations*. http://fold.aston.ac.uk/handle/123456789/7: The Aston Forensic Linguistic Databank.

Thorne, T. (2014a). Dictionary of contemporary slang.

Thorne, T. (2014b). The new canting crew. In J. Coleman, Ed., *Global English Slang*. Routledge, 72–81.

Wright, D. (2020). 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*. Routledge, 2nd ed., 611–627.

Ädel, A. (2020). Corpus compilation. In M. Paquot and S. Gries, Eds., *A Practical Handbook of Corpus Linguistics*. Springer, 3–24.

For the Record: Exploring variability in interpretations of police investigative interviews

Felicity Deamer, Emma Richardson†, Nabanita Basu & Kate Haworth

Aston Institute for Forensic Linguistics, Aston University, UK
& †Loughborough University, UK

https://doi.org/10.21747/21833745/lanlaw/9_1a2

Abstract. *Recent research (Haworth 2018) has demonstrated how investigative interview data are (unintentionally) distorted as they pass through the criminal justice system, and the survey-based experiment we present here was designed to test our hypothesis that various aspects of the processing of police-suspect interview data may have an impact on the quality of the official evidential document produced. The quantitative and qualitative findings from this experiment shed light on, and provide a sound evidence base for this claim, rather than leaving it as an untested assumption. The experiment was designed to test each key aspect of the current process of the production of routine written transcripts of investigative interviews (ROTIs), focusing on the conversion from spoken to written format, and the use of different transcription conventions, and it has enabled us to investigate which changes make the most difference in terms of the evidential quality of the end product, in order to effect a change in practice which will reduce or eliminate the effect of those changes. Our findings suggest that when presented with a transcript of a police interview, we are significantly more likely to (1) perceive the interviewee as anxious and unrelaxed, (2) interpret the interviewee's behaviour as being agitated, aggressive, defensive, and nervous, (3) determine that the interviewee is un-calm and uncooperative, and (4) deem the interviewee's version of events to be untrue, than we are if we listen to the original audio recording. Moreover, subjects identified (a) consistency, (b) phrase and lexical choice, (c) emotion (crying/upset), (d) hesitation and/or pauses as significant factors influencing participants' perception and interpretation of the interviewee and their story. This is particularly concerning as the latter two features are not currently routinely included in police transcripts, and Haworth (2018) illustrates multiple ways in which transcripts might differ from the original audio recordings they are intended to replace, with respect to words and phrases, as well as general content. The findings presented in this paper provide a strong motivation for further research into how we capture spoken interaction in legal contexts, and they constitute something of a mandate for reform with respect to the transcription of police interviews in the UK.*

Keywords: *Investigative interviews, Transcription, Entextualisation, Interpretation, Perception.*

Resumo. *Investigação recente (Haworth 2018) mostrou como os dados das entrevistas policiais são (involuntariamente) distorcidos durante o processamento pelo sistema de justiça criminal. O inquérito experimental que apresentamos neste artigo procura testar a nossa hipótese de que vários aspetos do tratamento dos dados das entrevistas policiais com os suspeitos podem influenciar a qualidade da prova oficial. Os resultados quantitativos e qualitativos desta experiência elucidam-nos e fornecem uma fundamentação sólida para esta assunção, até agora não testada. A experiência foi concebida para testar todos os aspetos centrais do atual processo de produção de transcrições do processo de entrevistas de investigação (ROTIs), concentrando-se na conversão da forma oral para a forma escrita e na utilização de diferentes convenções de transcrição, o que nos permitiu investigar quais as alterações que exercem maior impacto em termos de qualidade probatória do produto final, com o intuito de introduzir alterações na prática e assim reduzir ou eliminar o efeito dessas alterações. As nossas conclusões sugerem que, perante transcrições de uma entrevista com a polícia, é significativamente mais provável (1) percecionar o entrevistado como ansioso e inquieto, (2) interpretar o comportamento do entrevistado como agitado, agressivo, defensivo e nervoso, (3) percecionar o entrevistado como pouco calmo e cooperante, e (4) considerar que a versão dos eventos do entrevistado não é verdadeira, contrariamente ao que acontece se ouvirmos a gravação áudio original. Além disso, os participantes identificaram (a) consistência, (b) seleção frásica e lexical, (c) emoção (choro/perturbação), (d) hesitação e/ou pausas como fatores relevantes que influenciam a perceção e interpretação do entrevistado e da sua história, o que é particularmente preocupante uma vez que as duas últimas características não integram atualmente o processo de transcrições policiais. Como mostra Haworth (2018), as transcrições podem diferir, em diversos aspetos, das gravações áudio originais que substituem, não só em palavras e frases, mas também conteúdo geral. As conclusões apresentadas neste artigo constituem um forte incentivo a mais investigação sobre a interação verbal em contextos legais e constituem um apelo à reforma do processo de transcrição das entrevistas policiais no Reino Unido.*

Palavras-chave: *Investigative interviews, Transcription, Entextualisation, Interpretation, Perception.*

Introduction

Police investigative interviews play a critical role in the criminal justice system: they are one of the primary methods of evidence gathering during an investigation and can later serve as crucial evidence during a trial. Standard procedure in England and Wales is that police interviews are audio recorded, then transcribed by clerks employed by the relevant police force. This process is of particular importance given that these are evidential documents, routinely presented in court as part of the prosecution case, yet the original spoken data are (necessarily) substantially altered through the process of being converted into written format. Once a transcript, typically referred to as a ROTI (Record of Taped Interview), has been produced, it is generally heavily relied upon in place of the audio recording, especially in court, making its accuracy all the more important.

Haworth (2018) argues that interview evidence is unintentionally distorted and misinterpreted as it journeys through the criminal justice system; from an initial interaction that takes place in a police interview room to a transcript of an audio recording of that interview being read out in court (typically by the Prosecution) during a trial. Haworth draws attention to the asymmetry between the loose and unregulated practices and procedures associated with the handling of police interview evidence and the strict principles of preservation applied to physical evidence (e.g., DNA, blood spatter, fingerprint evidence), which she argues results from a lack of recognition that changes in the format of linguistic data involve a transformation of the data themselves. Haworth highlights both accidental and intentional discrepancies between the original interview and the version which takes its place as evidence in court. These include poor audio recording and lack of detail and nuance in transcription conventions (e.g. no pauses, intonation, stress emphasis, emotion, overlapping speech etc.), alongside deliberate editing and summarising to condense the official record. Yet the output of each transformational process is treated as an all but identical copy of the previous version (2018: 434).

It is of course, important to recognise that even without any such shortcomings in practice surrounding audio capture and transcription, it isn't possible to create a perfectly accurate written version of spoken language. There will always be something of a translation process involved in converting spoken language into written. Levelt (1983, 1989), among others, illustrated that listeners have a natural tendency to 'repair' any disfluencies they hear in speech, allowing them to make sense of what is being said. Collins *et al.* (2019) looked at what happens if disfluencies are embraced and incorporated into transcripts, focusing on whether filled pauses are perceived in the same way within transcripts as they are in speech. They found that disfluencies in speech (i.e. fillers such as 'um' and 'er') were more likely to be perceived as indicators of uncertainty in the speaker when presented in text (as part of a transcript) than when heard on an audio recording. This suggests that it is virtually impossible to capture and represent speech in written form without some distortion taking place. Moreover, (Fraser 2003) draws our attention to the fact that almost all attempts to convert spoken language into written form are made using an audio recording rather than the original face-to-face interaction (for obvious practical reasons). This means that however good the audio recording is, the talk in question has already been stripped of its meaningful context, and physically present and animated speakers, thus disarming our perceptual and inferential capacities of that critical information. Add to that the fact that there will inevitably be numerous points in any audio recording in which the sound quality is poor enough to leave the listener in doubt with respect to exactly what they are hearing, and it is clear that transcription necessitates a certain amount of informed guess work. The danger is that as transcribers we tend not to be aware of our perceptual inaccuracies, and as with all perception we are not conscious of the role that our predictions and expectations play in our experience of any given perceptual input. (**?**) refers to this as 'the unacknowledged role of the perceiver' (2003: 204), emphasising 'the active role we play in constructing the messages we hear by combining the information in the speech signal with the knowledge in our heads' (2003: 206; see also (Fraser 2014, 2018). (Bucholtz 2009) emphasises similar considerations with respect to the role of the transcriber in the specific context of the processing of spoken data within the legal system. (Coates and Myths 1999) recognise

the distortions that occur when spoken language is captured and converted into text, and argue that transcripts should be treated as nothing more than "an analytic convenience to make [spoken] data accessible to readers".

There is a pressing need for transcription guidelines and training to assist ROTI transcribers in producing ROTIs which encapsulate more of the meaning conveyed by the original spoken interaction, and to enable consistency of interpretation of features such as punctuation and pauses for the reader (i.e., fellow investigating officers, lawyers, courts). We are therefore currently undertaking a project to (1) test our hypothesis that there is a serious unrecognised problem within the criminal justice system with evidential consistency in investigative interview records; (2) collaboratively develop standard guidance for transcription, using transcription conventions which can easily be incorporated into practice; and (3) design new, linguistically-based input to the training for ROTI transcribers. A substantial increase in the accuracy and standardisation of investigative interview evidence (especially in terms of the representation of spoken language features) would enable those who use ROTIs as evidence to be able to interpret punctuation or other visual representation of spoken features consistently when they occur (regardless of who it was produced by) thus removing a major source of potentially subjective and inaccurate interpretation of criminal evidence. However, the above research points to the fact that any such endeavour must be based on evidence, not intuitions or expectations about how spoken and written data are perceived. Our first step, therefore, was to start building that evidence base. The study we present here is the first (as far as we are aware) to directly compare perceptions of and interpretations drawn from a transcript and an audio recording of the same spoken interaction.

The specific aim of the small-scale experiment discussed here was to assess individual perceptions of different versions of the same interview data (audio vs. written transcript). This is the first in a series of studies designed to test our hypothesis that aspects of the processing of police-suspect interview data have a negative impact on the quality of the official evidential document produced. Across this series of experiments, we aim to test each key aspect of the current process of the production of routine written transcripts of investigative interviews (ROTIs), focusing first in this current experiment on the conversion from spoken to written format, before then moving on to testing (in subsequent studies) the use of different transcription conventions. Our long-term intention is to investigate which changes make the most difference in terms of the evidential quality of the end product, in order to effect a change in practice which will reduce or eliminate (the effect of) those changes.

We emphasise that all versions in which interview interaction are recorded will inevitably involve a degree of alteration of the data, and there is no such thing as a 'perfect' transcript Fraser (2003); we also acknowledge that all use of data as evidence involves subjective interpretation on the part of the judge or jury, and that these interpretations will therefore inevitably vary to some extent. What we are aiming towards through this series of experiments (of which this is the first) is assessing which written version is evidentially closest to the original, in terms of introducing the least amount of change in interpretation when compared to the "purest" version available (here, the audio/video).

Method

In order to accurately compare individual perceptions of different versions of the same police interview, we assessed participants' interpretations, impressions, and judgements of both the interviewee themselves and what was said in the interview. These assessments were carried out under experimental conditions in which the mode of presentation was manipulated (transcript or audio recording) in order to assess the effect on the participants' perception of the interviewee, their interpretation of what was said in the interview, and their overall judgement with respect to the truth or falsity of the interviewee's version of events.

Participants

Sixty adult native speakers of English were invited to take part in the study using convenience sampling. All participants were contacts of the research team, and had been identified as (a) not having any prior linguistic training, and (b) not having any knowledge of the research project and its aims and objectives. Following recruitment, participants were emailed a link to the online survey. On opening the survey, participants were all presented with a 3-minute clip[1] from the same police interview (10 minutes long in full), taken from publicly available footage on YouTube of a suspect interview in a UK murder enquiry K.L.E.E. Photography (2015). Thirty participants heard the 3-minute audio recording of the original interview, and the other thirty saw a written transcript of the same clip. The groups were matched for gender and age (approx. even spread from age 18 – 71).

Materials and procedure

The interview clip in the two conditions was exactly the same, only mode of presentation varied between conditions (i.e., written transcript or audio). Participants were encouraged to listen to or to read the interview as many times as they liked, for as long as they liked prior to and while answering the questions that followed. Immediately following having heard or read the interview, participants were presented with a series of questions, some of which elicited quantitative data (i.e., number on a Likert scale), and some of which required an open answer in a text box (lending themselves to qualitative analysis). Participants were reminded that they could continue to listen to or to read the interview while answering the questions.

The transcript was produced with two key considerations in mind, (1) ensuring legibility for a lay audience, and (2) ensuring that it included as much detail as possible, given the first consideration of legibility. With these considerations in mind, we were able to include pauses, stress emphasis, overlapping speech, and emotion (transcribed as 'sniff'); all features which have long been established within linguistics as conveying substantial amounts of meaning, and which are therefore generally included in even relatively simple linguistic transcripts Jefferson (2004). The intention was to include as much detail as we could in the transcript, so that we could compare a 'best possible' transcript (given the first consideration above) with the original audio recording. If we had produced a transcript that might be considered closer to an officially produced police transcript / ROTI, we would not have been able to ask participants the same questions of the data (i.e., concerning linguistic features, since that information would not have been available to those participants in the Transcript condition).[2] There was a legend at the top of the transcript to explain the meaning of each of the transcription conventions.

Transcript: For the record experiment

IR = Interviewer
IE = Interviewee

((Actions)) are indicated by double brackets.

[Speakers talking at the same time] are indicated by square brackets.

Underlining is used to indicate any stress on the word.

Italics indicate 'emotion' (e.g. ((*Sniff*)) would indicate the action of crying).

(0.0) indicates in 10ths of second's pauses or gaps in talking.

```
1     IR:   How do I know that you (0.5 sec) weren't involved?
2                 (2.3 sec)
3     IE:   Again, I shouldn't have any (1.5 sec) DNA reason to be
4           involved. And again (1.7 sec) especially (1.5sec) with my
5           past.
6                 (0.6 sec)
7     IE:   To think that I could (1 sec) allow (0.8 sec) harm to come
8           to somebody else like that ((sniff)) is highly unlikely.
9                 (0.9 sec)
```

Figure 1. Image of transcript

Below are the questions that participants were presented with after having read or heard the interview extract. Questions 1 and 5 were formulated with the intention of exploring how the interviewee's character might be evaluated differently depending on whether the original audio recording was heard, or whether a transcript of the interview was read. In a related way, questions 3 and 9 probe participants' evaluation of the interviewee's story, and whether those evaluations differ depending on the format in which the interview is presented. Questions 2, 4, 6 and 10, allowing free text responses for the purpose of qualitative analysis, were included in order to glean as much information from participants as possible with respect to how the linguistic dimensions of the interview (i.e., what is said and how it is said) influenced or informed their judgements and evaluations of the interviewee and their version of events. Questions 7 and 8 allow participants to describe the interviewee's emotional and behavioural profile, again enabling us to explore potential differences between groups. The options used in these questions were adapted from (Ekman 1992) universal emotion categorisation system to ensure that we offered participants the opportunity to describe the interviewee's emotional state as accurately and as thoroughly as possible. We used multiple choice for these questions in order to ensure that the responses were tractable. Question 11 was included at the end of the survey with a view to establishing whether the format in which the interview is presented to participants impacts on their overall perception of the interviewee's guilt or lack thereof.

1. On a scale of 1 to 5, please specify your level of agreement with the statement below:

   "The interviewee is credible"

"The interviewee is credible"

	1 - I don't agree at all	2	3	4	5 - I agree entirely
"The interviewee is credible"	☐	☐	☐	☐	☐

2. What is it about the language used and/or how it is said that led you to your conclusion about how credible the interviewee is?
3. On a scale of 1 to 5, please indicate how plausible the interviewee's story is.

	1 - Totally implausible	2	3	4	5 - Highly plausible
The interviewee's story is...	☐	☐	☐	☐	☐

4. What is it about the language used and/or how it is said that led you to your conclusion about whether or not the interviewee's story is plausible?
5. On a scale of 1 to 5, please indicate how sincere the interviewee is.

	1 - Totally insincere	2	3	4	5 - Very sincere
The interviewee is...	☐	☐	☐	☐	☐

6. What is it about the language used and/or how it is said that led you to your conclusion about whether or not the interviewee is sincere?
7. On a scale of 1-5, please indicate to what degree the following words could be used to describe the interviewee's emotions at any point during the interview:

	1 - Not at all	2	3	4	5 - Very much
RELAXED	☐	☐	☐	☐	☐
ANXIOUS	☐	☐	☐	☐	☐
FEARFUL	☐	☐	☐	☐	☐
DISGUSTED	☐	☐	☐	☐	☐
SURPRISED	☐	☐	☐	☐	☐
HAPPY	☐	☐	☐	☐	☐
ANGRY	☐	☐	☐	☐	☐
SAD	☐	☐	☐	☐	☐
CONTEMPT	☐	☐	☐	☐	☐

8. On a scale of 1-5, please indicate to what degree the following words could be used to describe the interviewee's behaviour at any point during the interview:

	1 - Not at all	2	3	4	5 - Very much
AGITATED	☐	☐	☐	☐	☐
CALM	☐	☐	☐	☐	☐
PANICKED	☐	☐	☐	☐	☐
FRIENDLY	☐	☐	☐	☐	☐
COOPERATIVE	☐	☐	☐	☐	☐
AGGRESSIVE	☐	☐	☐	☐	☐
DEFENSIVE	☐	☐	☐	☐	☐
ASSERTIVE	☐	☐	☐	☐	☐
NERVOUS	☐	☐	☐	☐	☐

9. In your opinion, is what the interviewee is saying true?

○ Yes
○ No

10. What is it about the language used and/or how it is said that led you to your conclusion about whether or not what the interviewee is saying is true?
11. In your opinion, and based only on what you have heard in the interview, should the interviewee be found Guilty or Not Guilty of being complicit in the murder?

○ Guilty
○ Not Guilty

Analysis

The survey responses were analysed using a mixture of qualitative and quantitative methods.

For the analysis of questions 1, 3, 5, 7, and 8, we analysed the distribution of data for each question in the Audio condition and the Transcript condition. We did this based on measures of central tendency (such as mean, median, and mode) and measures of dispersion (such as standard deviation, inter quartile range, and range.). To test for normality of distribution of responses for each question of each group, Shapiro Wilk test was used (as the number of data points in each group was less than 2000). Given that the datasets deviated from normality assumption, equality of variance for the groups (using Levene's test) was not calculated. Given that normality assumptions were violated and the data is ordinal, a non-parametric test (Mann Whitney U-test) was used to compare the responses for the two groups.

Questions 9 and 11 were dealt with separately because the response to these questions are categorical/nominal and dichotomous. Given that the data (i.e., responses) are nominal and dichotomous, we cannot expect the data to follow a normal distribution. Again, we considered the responses to be unpaired. In order to test whether the distribution of responses is the same for the Audio and Transcript group we used the Chi square test.

Finally, content analysis and inductive thematic analysis were used to derive a coding frame to analyse free text responses to questions 2, 4, 6, and 10 that enabled us to isolate which linguistic features participants identified as contributing to their interpretation of the interview and/or perception of the suspect being interviewed. The coding frame allowed for a nuanced analysis of the specific qualities of the data collected, and was developed iteratively by the research team who met a number of times to scrutinise the text box data and discuss a working list of codes that captured all the linguistic features mentioned in participants' responses. Two researchers separately coded all 60 responses to the free text questions, and then met to discuss and agree on any instances where the two sets of coding didn't match up. The outcome of this coding process is detailed in Table 1.

Code	Audio	Transcript
Clarity	**13**	**23**
In delivery	10	19
In Language	3	4
Content Choice	**37**	**27**
Lexical choice	7	6
Phrase choice	22	18
Repetition	8	3
Emotion	**86**	**43**
Crying/upset	46	27
Genuine+	22	2
Genuine-	8	12
Shock	7	2
Laughter	3	0
Register	**2**	
Formal -	0	0
Formal +	2	0
Sentence Structure	**6**	**6**
Conditionals	4	3
Non-sequiturs	1	0
Unfinished	1	3
Sequencing	**9**	**11**
Interjection	0	1
Question and Answer match	5	4
Quick to answer	4	6
Other		
Hesitation/pausing	14	19
Consistency	3	12
Pace	2	0
Rehearsal	10	8
Sound Quality	1	2
Stress emphasis	0	6

Table 1. Codes and references per condition

Results

Quantitative

The results suggest (see Figure 2 and Figure 3) that there is no statistically significant difference between the responses to question 1 (Z = -1.434, p=.152) and 3 (Z = -.349, p=.727) in the Audio and the Transcript group, indicating that the format in which participants were presented with the interview did not impact on their perception of the interviewee's credibility or the perceived plausibility of the interviewee's version of events[3].

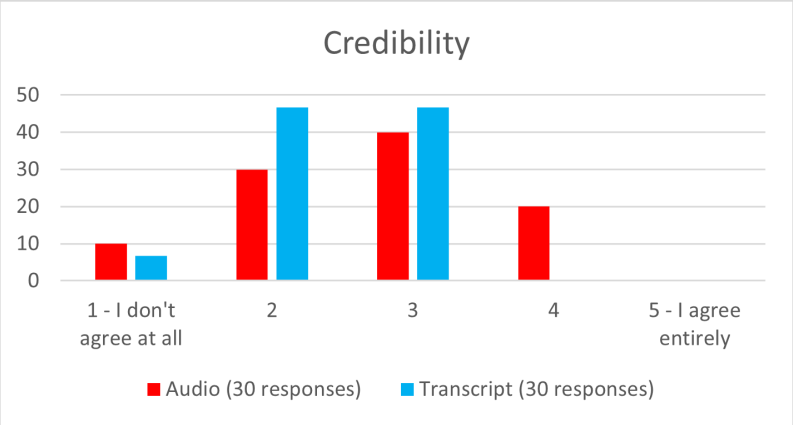

Figure 2. Distribution of responses to question 1 (credibility)

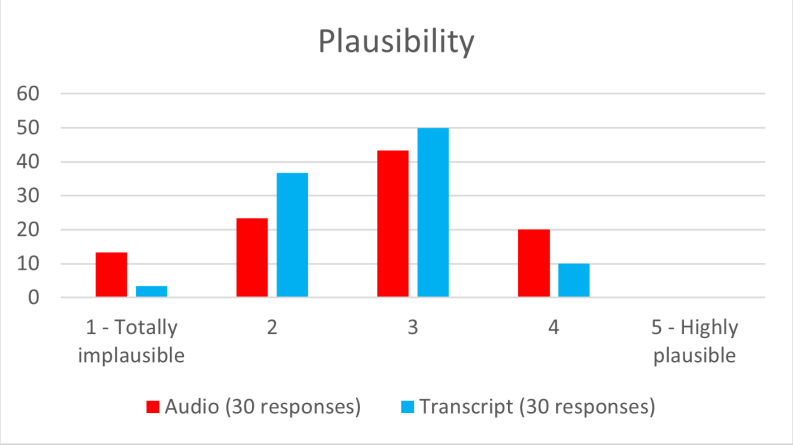

Figure 3. Distribution of responses to question 3 (plausibility)

Results for question 5 (Z=-1.741, p=.082) suggest that with a larger sample size we might find evidence that the format in which participants experience the interview does have an impact on their judgement with respect to how sincere the interviewee is (see figure 4).

The results from question 7 suggest (see Figure 5 to Figure 7) that there is a statistically significant difference (or strong trend towards significance) between the responses in the Audio and the Transcript group, with respect to the attribution of emotional descriptors 'relaxed' (Z= -2.267, p=.023), 'anxious' (Z= -1.984, p=.047), and

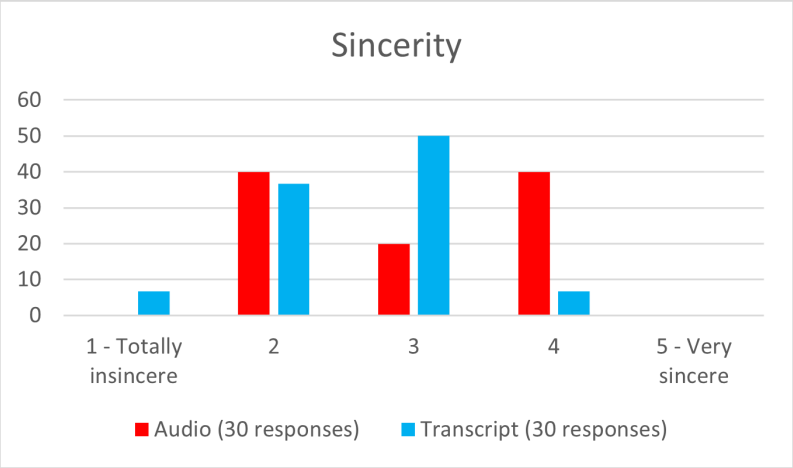

Figure 4. Distribution of responses to question 5 (sincerity)

'fearful' (Z= -1.928, p=.054), suggesting that reading the transcript of the interview makes participants more likely to perceive the interviewee to be unrelaxed, anxious, and fearful than if they were to listen to the original audio recording.

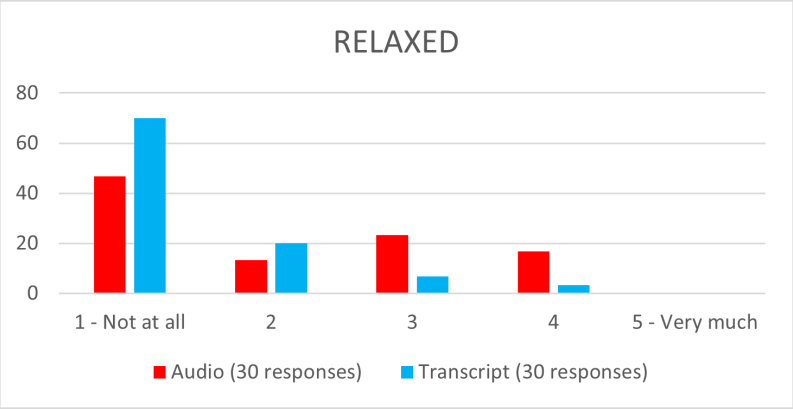

Figure 5. Distribution of responses to question 7 (relaxed)

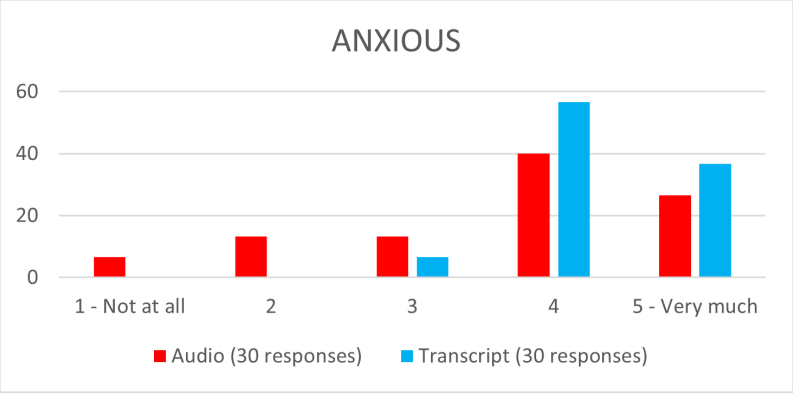

Figure 6. Distribution of responses to question 7 (anxious)

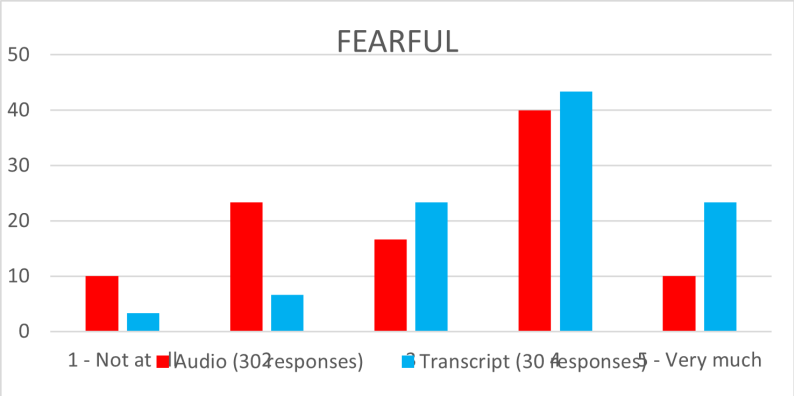

Figure 7. Distribution of responses to question 7 (fearful)

There were no significant differences (see Table 1) between groups with respect to the remaining emotional descriptors 'disgusted' (Z= -.429, p=.668), 'surprised' (Z= -.195, p=.846), 'happy' (Z=-.043, p=.966), 'angry' (Z= -.168, p=.867), 'sad' (Z= -.015, p=.988), 'contempt' (Z= -1.074, p=.283).

Q7_	Relaxed	Anxious	Fearful	Disgusted	Surprised	Happy	Angry	Sad	Contempt
Mann-Whitney U	313.500	326.000	325.500	422.000	437.500	448.500	439.500	449.000	381.000
Wilcoxon W	778.500	791.000	790.500	887.000	902.500	913.500	904.500	914.000	846.000
Z	-2.267	-1.984	-1.928	-.429	-.195	-.043	-.168	-.015	-1.074
Asymp. Sig. (2-tailed)	.668	.846	.966				.867	.988	.283

Table 2. Mann Whitney U Test results for question 7

With regard to Question 8, a further Mann Whitney U test also showed that there was a significant difference in the responses between the Audio and Transcript group with respect to the attribution of behavioural descriptors 'agitated' (Z= -2.95, p=.003), 'calm' (Z= -2.41, p=.016), 'cooperative' (Z= -2.33, p=.020), 'aggressive' (Z= -2.12, p=.034), 'defensive' (Z= -3.32, p=.001), and 'nervous' (Z= -2.39, p=.016) (see figures 8-13), suggesting that reading the transcript of the interview makes participants more likely to perceive the interviewee to be *agitated*, *aggressive*, *defensive*, and *nervous* than if they were to listen to the original audio recording.

There were no significant differences (see Table 2) between groups with respect to the remaining behavioural descriptors 'panicked' (Z= -1.52, p=.128), 'friendly' (Z= -1.74, p=.080), and 'assertive' (Z=-..742, p=.458).

Q8_	Agitated	Calm	Panicked	Friendly	Cooperative	Aggressive	Defensive	Assertive	Nervous
Mann-Whitney U	257.500	298.500	350.000	336.500	302.500	327.500	232.000	402.000	296.000
Wilcoxon W	722.500	763.500	815.000	801.500	767.500	792.500	697.000	867.000	761.000
Z	-2.955	-2.413	-1.522	-1.748	-2.332	-2.121	-3.328	-.742	-2.398
Asymp. Sig. (2-tailed)	.003	.016	.128	.080	.020	.034	.001	.458	.016

Table 3. Mann Whitney U Test results for question 8

A significant difference was recorded (see figure 14 and Table 3) in the distribution of responses to question 9 "In your opinion, is what the interviewee is saying true?"

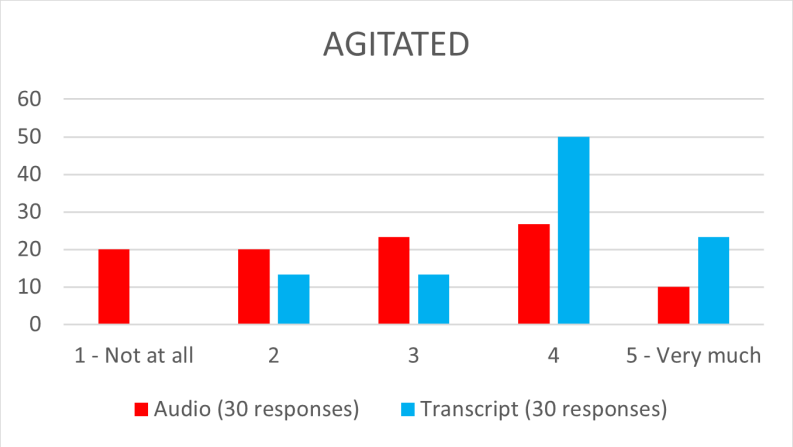

Figure 8. Distribution of responses to question 8 (agitated)

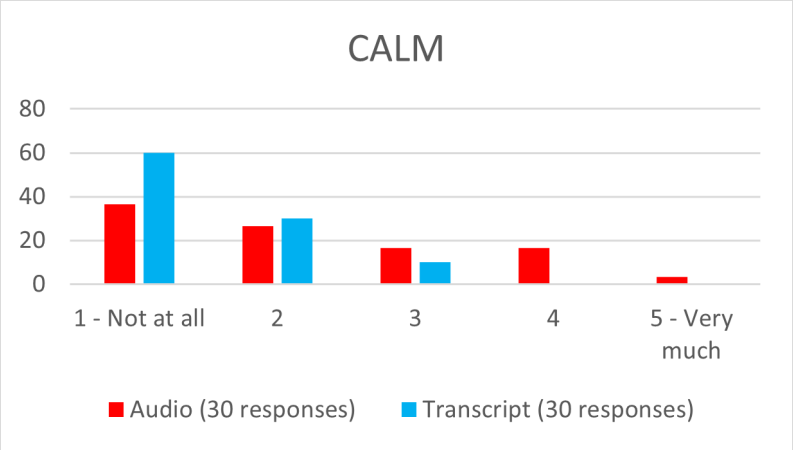

Figure 9. Distribution of responses to question 8 (calm)

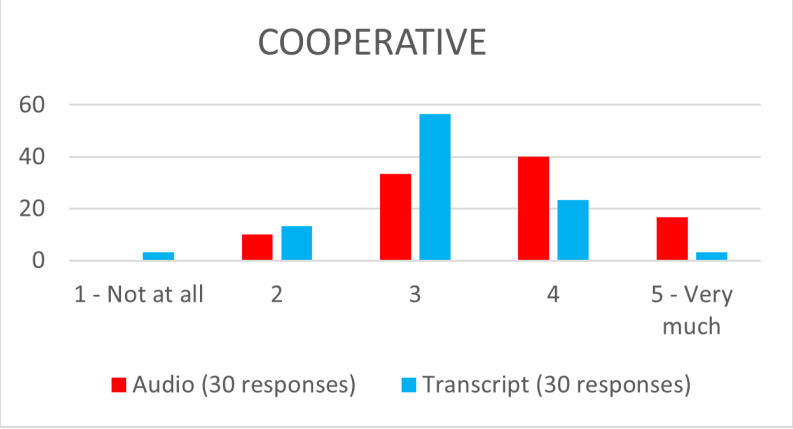

Figure 10. Distribution of responses to question 8 (cooperative)

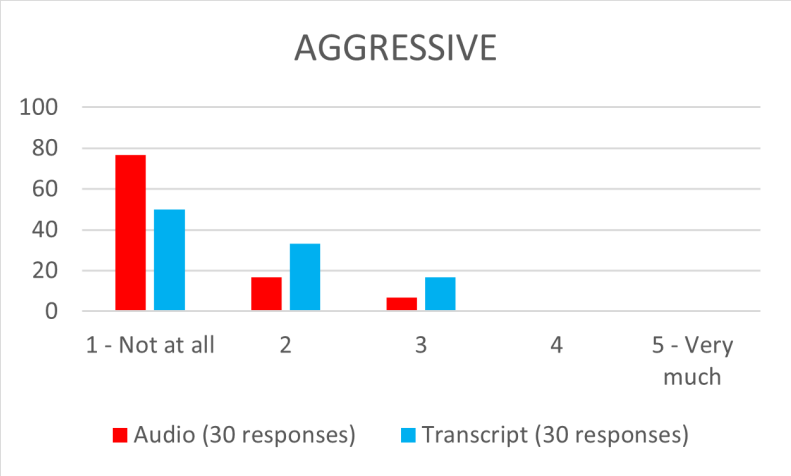

Figure 11. Distribution of responses to question 8 (aggressive)

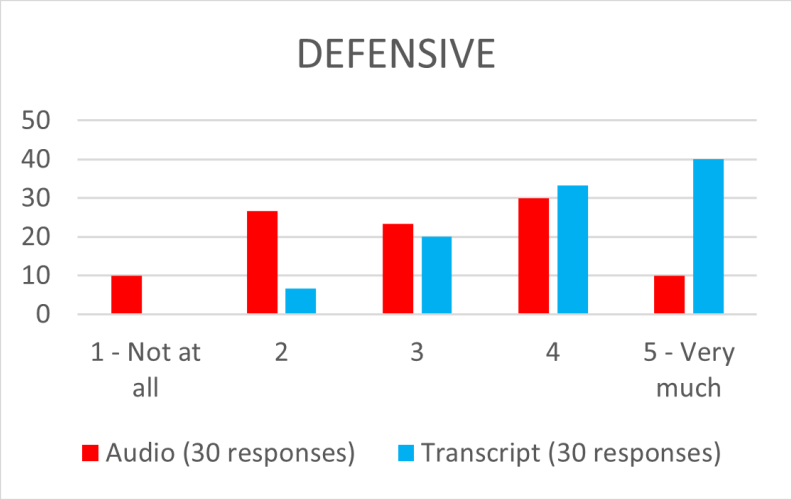

Figure 12. Distribution of responses to question 8 (defensive)

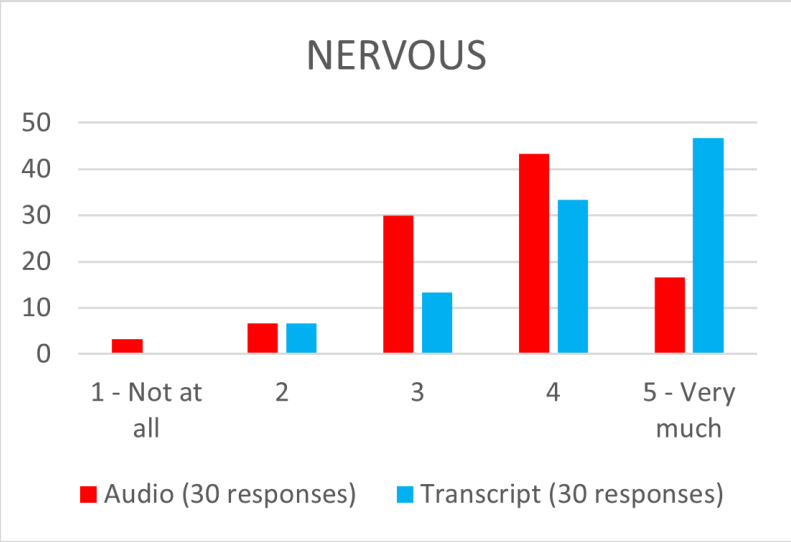

Figure 13. Distribution of responses to question 8 (nervous)

between the Audio and Transcript group ($^2(1) = 4.022$, p=.045), indicating that those who read the transcript of the interview were more likely to judge the interviewee's version of events to be untrue.

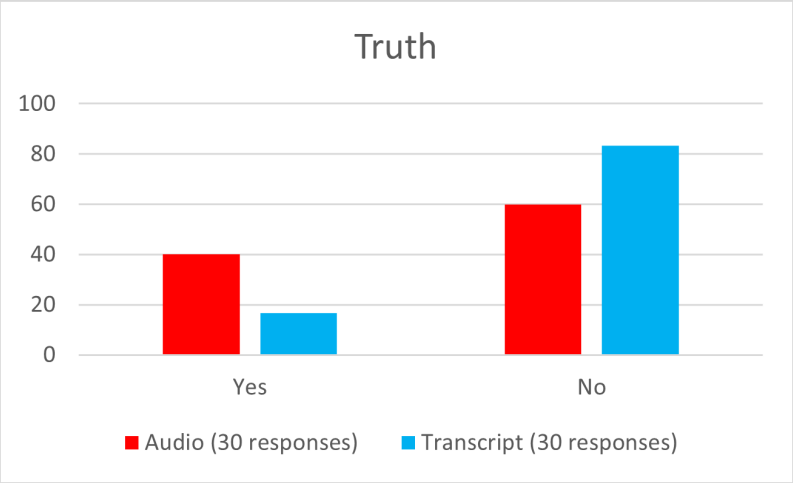

Figure 14. Distribution of responses to question 9 (truth)

No significant difference was recorded in the distribution of responses to question 11 "In your opinion, and based only on what you have heard in the interview, should the interviewee be found Guilty or Not Guilty of being complicit in the murder?" between the Audio and Transcript group ($^2(1) = 0.268$, p=.605).

In summary, participants who read the transcript of the interview were significantly more likely than those who had heard the original audio recording to:

1. Perceive the interviewee as *anxious* and *unrelaxed*
2. Interpret the interviewee's behaviour as being *agitated*, *aggressive*, *defensive*, and *nervous*
3. Determine that the interviewee is *un-calm* and *uncooperative*[4]
4. And ultimately, to deem the interviewee's version of events to be *untrue.*

This study uses a small sample size, and our effect sizes may be a consequence of that. We intend to replicate the findings in a larger study, in which we have (a) a larger sample size in each condition (n=64), and (b) multiple experimental conditions to be compared including *Audio* (the same audio recording used in this current study), *Full Transcript* (the same transcript used in this current study), *ROTI* (a transcript produced in line with standard police practice), *Pauses Removed* (the same as the full transcript condition but with pauses removed), and *Emotion Removed* (the same as full Transcript condition but with markers of emotion removed).

Qualitative

Here, we present the main themes identified from the text responses to the questions: what is it about the language used and/or how it is said that led you to your conclusion about how *credible* the interviewee is (Q2), whether or not the interviewee's story is *plausible (Q4)*, whether or not the interviewee is *sincere (Q6)* and whether or not what the interviewee is saying is *true (Q10).*

Although the quantitative analysis did not find significant difference in responses to the questions of plausibility and credibility, the qualitative analysis enables us to understand further the justifications for scores, based on the language used in the two conditions. We present three main findings, firstly relating to the temporal aspect of the delivery of the suspect's speech, secondly the interpretation of emotion, and finally the impact of an impoverished mode of presentation on participants' judgements of the interviewee and their account.

Temporal aspects of delivery

An analysis of multiple codes ('quick to answer', 'hesitation' and 'rehearsal') outlines how participants reported making use of the speed at which the interviewer's questions were answered by the interviewee to make their judgements. The first of these to be explored here is where the suspect was perceived as being quick to answer the interviewer's questions.

The interviewee was 'Quick to answer'

At the end of the interview extract the interviewer asks the suspect two yes/no interrogative questions, both answered by the suspect with a "no". In the transcript, no pause was indicated prior to the delivery of "no", where pauses are timed and indicated (to the tenth of a second) elsewhere. Participants in each condition reported the speed at which the suspect answers as a justification for their scores.

Extract 1:

> *"The short, sharp "no"s towards the end of the clip seemed insincere." (Audio, Truth No)*

Extract 2:

> *"She was very quick to answer the IR question on line 101." (Transcript, Truth No)*

Extract 3:

> *"She is relatively quick to answer the questions, however doesn't give a lot of detail." (Transcript, Truth Yes)*

Extract 4:

> *"She doesn't hesitate at all - and it backs up the rest of her story that she didn't know what was going on." (Audio, Plausibility 4)*[5]

Most of the references captured by this code relate to this final section of the transcript and whether or not the suspect was telling the truth. If a written record of the investigative interview does not seek to represent the speed at which responses are delivered, then a reader compared with a listener will have unequal opportunity to evaluate the suspect's account in this respect.

'Hesitation' in the suspect's responses

In addition to the speed at which the suspect answered some questions, pauses during the delivery of the suspect's speech (mid-turn pauses) were interpreted by participants as relating to 'thinking' about what to say. Hesitation was used to explain scores mostly on the lower end of the scale for credibility and for why they deemed the suspect not to be telling the truth.

Extract 5:

> *"There were hesitations in her replies as if she was trying to convince herself what she was saying was true as well as trying to convincing the interviewee." (Audio, Credibility 2)*

Extract 6:

> *"Very vague answers, constant hesitation as if they are considering what they say." (Transcript, Credibility 2)*

Extract 7:

> *"[T]here were pauses where it felt like the interviewee was thinking too much about the right thing to say." (Audio, Truth, No)*

However, participants relayed to us that when pauses, interpreted as hesitation, are heard or read in conjunction with 'emotion' in the suspect's speech, this positively impacted their perceptions.

Extract 8:

> *"The nervous, halting, and emotional quality of the delivery make the interviewee sound sincere to me" (Audio Sincerity 4)*

Extract 9:

> *"Not too much hesitation in her voice. Her emotions also come across as genuine, in a way in which I can't articulate." (Audio, Sincerity 4)*

Extract 10:

> *"On the one hand the language, stresses, pauses, could be genuinely indicative of a highly upset individual who is struggling with the circumstances." (Transcript, Credibility 3)*

Whether influencing negatively or positively, these findings argue in favour of including pauses in transcripts, where they feature in the audio, since this demonstrates their importance to the assessment of the interview as evidence.

In addition to pauses, the suspect's use of fillers such as 'uh' and 'um' in their speech also was categorised by one participant as 'hesitation'.

Extract 11:

> *"The constant us[e] of "um" giving herself time to think." (Audio, Credibility 1)*

It is common for these not to be transcribed even in what the police consider to be a "verbatim" transcript. However, the responses from this research indicate that their inclusion could impact on hearers' and readers' interpretation of the suspect's account. The two temporal aspects of delivery discussed so far also feature in the next section where the delivery of the speech is evaluated as if it were, or were not, rehearsed.

The suspect's responses were 'rehearsed'

Participants in both conditions recurrently described the suspect's answers as 'rehearsed', or not. Here we present some of the references which directly attribute that to the temporal aspects of delivery rather than to the content of what is being delivered in the narrative (which we do not focus on in the current paper).

> *"I don't feel the interviewee is sincere as the pauses suggest she has rehearsed what she is going to say, and rather than it being said with true emotion, she has said things for effect." (Transcript, Sincerity 1)*

Extract 12:

> *"In my opinion, what the interviewee is saying sounds rehearsed and unnatural. They don't stutter much, neither do they use "umms" or "urrs", etc. It sounds like something that has been prepared and memorised in order to specifically address the key doubts against them. It seems deliberately vague." (Audio, Truth No)*

Pausing during the speech is described by a participant in extract 13 to be interpreted as planned delivery. In extract 14, we see another participant comment on the lack of disfluency in the response. The responses of participants in relation to plausibility, credibility and sincerity being influenced by a sense of rehearsal does appear to be mitigated again by the presence of what they interpret to be emotion.

Extract 13:

> *"It sounds rehearsed. It sounds like she's reading lines until she starts to cry." (Audio, Sincerity 2)*

Extract 14:

> *"Her emotions appear to be authentic with halting speech, interrupted by crying. She appears not to have pre-rehearsed what she is going to say." (Audio, Sincerity 4)*

In the next section we pay particular attention to the ways in which emotion is perceived and referred to by participants in the two conditions.

Emotion

The code containing the highest number of references in our analysis was 'emotion', with participants making numerous comments about the suspect's emotional state when responding to the questions we asked. This notably included considerations of whether or not the suspect was being genuine, with a number of mentions of the suspect being in shock. Crying seemed particularly marked for participants (mentioned a total of 72 times across both conditions). Interestingly, while participants in both conditions made use of the emotion heard in the suspect's voice, or noted in the transcript, to form judgements about the suspect's account and justify their scores, they did this in slightly different ways. Those in the audio condition interpreted the emotion themselves:

Extract 15:

> *"The interviewee's clear shock and emotional state also makes me believe that she's being honest and is more of a victim in this situation." (Audio, Credibility 4)*

Extract 16:

> *"The interviewee seems most sincere at the beginning of the interview when she is most emotional." (Audio, Sincerity 4)*

Extract 17:

> *"[H]er crying sounded quite real - she seems to be sniffing which is quite hard to make happen." (Audio, Sincerity 2)*

By contrast, those in the Transcript condition described their interpretation as filtered through the transcriber's interpretation and representation of emotion in the transcript.

Extract 18:

> *"The repeated sniffing in the text, and the fluency can be read either way, as genuine emotion, or as an attempt to screen her true feelings." (Transcript, Sincerity 3)*

Extract 19:

> *"I'm told she's emotional and crying at various points." (Transcript, Credibility 3)*

These references to the 'text', or being 'told' about a particular feature, demonstrate that the format distanced participants in the Transcript group from the data in a way which did not seem to occur with the Audio group, by inserting the transcriber in between.

Impoverished mode of presentation

Eight of the 30 participants in the Transcript condition assigned a score of '3', the middle of the scale, for questions relating to interviewee credibility, and plausibility and sincerity of the account. In their text responses they commented that not having access to the audio impacted on their ability to make judgement. They commented how, by reading and not listening, they struggled to be "categorical" in their judgements.

Extract 20:

> *"It's hard to decide, even with the emotion/stress/gaps written down. I feel like I'm still missing information on how she's saying it." (Transcript, Credibility 3)*

More specifically, participants made comment about the difficulty they had in answering the questions based on a reading of rather than a listening to the data. They located the issue as requiring the speaker's 'tone' to make judgements.

Extract 21:

> *"I can't [sic] say without hearing the tone of voice." (Transcript, Sincerity 3)*

Extract 22:

> *"Its [sic] difficult to answer this using the text without hearing tone of voice. "(Transcript, Credibility 3)*

However, references captured by the code 'tone' demonstrate that participants in the Audio group did make judgements based on what they describe as the tone of the suspect's voice.

Extract 23:

> *"I didn't hear anything in the tone of her voice or the words that made me think she was being insincere." (Audio, Sincerity 3)*

Extract 24:

> *"The defensive tone in which she spoke at times in the interview made her testimony appear less sincere." (Audio, Sincerity 3)*

Summary of qualitative analysis

The qualitative analysis makes a strong case for the importance of representing temporal aspects of the speech delivery in the transcript and where possible 'emotion' which (as discussed earlier) we indicated through the transcription of 'sniffs' and where the suspect's voice became altered, or strained, through crying. If pauses, audible sniffs and alterations in delivery were not represented within the transcript (as is typically the case in ROTI transcripts), they would be unavailable to those without access, or the resource to listen to the audio recording. We might hypothesise this to alter the interpretation of the evidence.

Discussion

It is perhaps intuitive that vitally important information is lost in the transformation of spoken interaction into written format. It is somewhat less intuitive that this lost information might have a negative impact on the evidential value of the spoken interaction, but this is what we are seeing evidence of in our data. In this experiment, the aim was to move from the *whether* to the *why*; if we know the mechanisms by which evidential value is negatively impacted, then we can take steps to mitigate that impact.

We might think that in each condition there is a different kind of input (audio vs. written), which (due to their inherent differences) leads to different

inferences/conclusions being drawn. But the idea that there is a stable notion of input is problematic; the inputs themselves will be moulded and shaped by our prior experience (see, for example, increasingly popular Bayesian approaches to cognition Griffiths *et al.* (2008)). Neither the audio recording nor the transcript are stable inputs on which rational inferences operate; top-down information (hypotheses/expectations/prior experience) will determine how we perceive the interview, and will influence the conclusions we draw about the person being interviewed. You could view a transcript as a further distortion of an original interview than the audio recording, having removed all the myriad low level contextual information. However, the reader will instinctively make sense of what they are presented with; they won't simply "suspend judgement" on information that is lacking; they will "fill in". This hypothesis is supported by our findings that participants in the Transcript condition were more, not less, likely to attribute emotional properties to the interviewee, like 'anxious', or 'agitated', even though direct evidence of this is lacking. Put generally, this filling in plays a bigger role in the Transcript condition than in the Audio condition as it is informationally impoverished.

The relevant point here is that the transcript is not just leaving information out, but encouraging a process of filling in**.** The reader doesn't simply fail to see what's there, but might infer things that are not. These "top-down" filling in effects don't only impact rational judgement; they reach down into how the input is experienced at the most basic level[6]. This means that with all the good will (and education/training) in the world, a reader won't be able to avoid the effects of their "priors". Another concern is that these "priors" will vary across participants, meaning that, the more there is this filling in, the less likely responses are to be consistent across participants. In other words, not only would we expect to see less *accuracy* in informationally impoverished conditions (i.e. transcripts), but we will also see more variability, namely, less *consistency* in interpretation.

In adding more detail to our Transcript condition (relative to a standard ROTI), such as emotion, pauses, stress emphasis, it could be argued that we added very explicit 'filling-in' prompts. Just as (Collins *et al.* 2019) found that disfluencies such fillers were more marked in transcripts than in audio recordings, perhaps the detail in our transcription made salient those linguistic features which we typically might not attend to when hearing audio. This raises important questions around whether more transcription detail is in fact better, or whether it makes artificially salient those features which might otherwise be experienced as a 'normal' and unmarked feature of spoken language.

This concern is in fact largely not borne out in our qualitative analysis. It is clear from Table 4 that with respect to pauses and emotion, both features which were detailed in our transcript, there are more mentions in the audio condition than in the Transcript condition, and there are almost no mentions of interjections/overlapping speech in either condition. There were only mentions of stress emphasis in the Transcript condition, albeit only six, so this could indeed be a feature which was made artificially salient when marked in the transcript.

Our qualitative analysis illustrates that the speed of delivery, including pauses and hesitation and representations of emotion were key features that participants identified as contributing to their perception and overall judgement of the interviewee and their

version of events, which suggests that these are aspects of the audio recordings of interviews that we should be looking to retain within standard police transcripts (ROTIs).

The findings discussed here are born out of a small-scale study, and it would be prudent to attempt to replicate these findings on a larger scale, comparing multiple conditions:

Condition 1	Condition 2	Condition 3	Condition 4	Condition 5	Condition 6
Original audio recording	Standard police transcript (ROTI)	Detailed transcript, including pauses, emotion, and overlapping speech	Condition 3 transcript but with pauses removed	Condition 3 transcript but with overlapping speech removed	Condition 3 transcript but with emotion removed

Table 4. Follow-up experimental conditions

However, these initial findings strongly suggest that the onus is on us as linguists to provide the relevant evidence and training to ensure that official police transcripts (ROTIs) contain as much of the vitally important information contained in the original interview as possible, while being mindful of the context and purpose of transcription, as well as the limited resources available to the police Richardson *et al.* (2022), and being conscious of the 'added salience' effects that transcription detail can have.

Acknwoledgements

The research reported in this paper was funded and conducted at Aston Institute for Forensic Linguistics.

Notes

[1]A subsection of the interview was selected to ensure that participants had a manageable amount of information to attend to in close detail when answering the subsequent questions. The specific 3-minute clip was chosen on the basis that it contained the interviewee's main description of events surrounding the killing.

[2]We will run comparisons with an officially produced transcript in later iterations of this project.

[3]We used significance level (i.e,) of 0.05 for all statistical tests. Sample effect is considered to be statistically significant if p value is less than or equal to the chosen value.

[4]As 'calm' and 'agitated' are antonyms, we expected participants who marked high for agitated would mark low for calm. Including these terms allowed us to check for consistency in participants' answers.

[5]The number detailed here refers to where on the sliding scale (1 (not at all)-5 (very much)) participants rated the interviewee on that particular question

[6]See sine-wave speech, Hollow Mask Illusion, McGurk effect etc.

References

Bucholtz, M. (2009). Captured on tape: Professional hearing and competing entextualizations in the criminal justice system. *Text and Talk*, 29(5), 503–523.

Coates, J. and Myths, J. (1999). Lies and audiotapes: Some thoughts on data transcripts. *Discourse Society*, 10(4), 594–597.

Collins, H., Leonard-Clarke, W. and O'Mahoney, H. (2019). Um, er': how meaning varies between speech and its typed transcript. *Qualitative Research*, 19(6), 653–668.

Ekman, P. (1992). An argument for basic emotions. *Cogn. Emot*, 6, 169–200.

Fraser, H. (2003). Issues in transcription: Factors affecting the reliability of transcripts as evidence in legal cases. *Forensic Linguistics*, 10(2), 203–226.

Fraser, H. (2014). Transcription of indistinct forensic recordings: Problems and solutions from the perspective of phonetic science. *Language and Law/Linguagem e Direito*, 1, 5–21.

Fraser, H. (2018). Forensic transcription: How confident false beliefs about language and speech threaten the right to a fair trial in australia. *Australian Journal of Linguistics*, 50, 129–139.

Griffiths, T., Kemp, C. and Tenenbaum, J. (2008). Bayesian models of cognition. In R. Sun, Ed., *The Cambridge handbook of computational psychology*. Cambridge University Press, 59–100.

Haworth, K. (2018). Tapes, transcripts and trials: The routine contamination of police interview evidence. *The International Journal of Evidence Proof*, 22(4), 428–450.

Jefferson, G. (2004). *Glossary of transcript symbols with an introduction*. Amsterdam/Philadelphia: John Benjamins.

K.L.E.E. Photography, (2015). Becky Watts murder - 3rd police interview (long version). https://www.youtube.com/watch?v=hN4fYTXObcg.

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.

Levelt, W. (1989). *Speaking: From intention to articulation*. The MIT Press.

Richardson, E., Haworth, K. and Deamer, F. (2022). For the record: Questioning transcription practices in legal contexts'. *Applied Linguistics*. https://doi.org/10.1093/applin/amac005. Published online 8th Feb 2022.

Killer stance: An investigation of the relationship between attitudinal resources and psychological traits in the writings of four serial murderers

Madison Hunter & Tim Grant

Aston Institute for Forensic Linguistics, Aston University, UK

https://doi.org/10.21747/21833745/lanlaw/9_1a3

Abstract. *An individual's set of psychological traits, or their psychopathology, impacts how they experience the world around them, and language offers resources that allow for that experience to be shared with and communicated to someone else. That language can then be analyzed for patterns and their connections to the psychological traits. In a forensic context, such connections may give valuable insights. There already exist psychological and linguistic approaches to the analysis of forensic texts, but the psychological approach largely lacks grounding in linguistic theory and the linguistic approach does not typically allow consideration of psychological characteristics. What this paper aims to provide is a step toward bridging that gap. In this paper we examine the system of attitude from the Appraisal framework developed by Martin and White (2005) and adapted by Gales (2010) and Hurt (2020) applying this to the writings of four serial murderers with documented mental health diagnoses. Significant patterns in the attitudinal resources were identified quantitatively and examined qualitatively through the lens of the psychological traits that comprised the authors' diagnoses to determine if there was a relationship between them. Despite the obvious limitation presented by the sample size, the results of this study suggest the approach presented in this paper warrants further investigation.*

Keywords: *Appraisal analysis, Stance, Psychopathology, Attitude, Serial murderers.*

Resumo. *O conjunto de características psicológicas de um indivíduo, a sua psicopatologia, tem impacto na forma como vivencia o mundo à sua volta, e a linguagem oferece recursos que permitem que essa experiência seja partilhada e comunicada a outra pessoa. Essa linguagem pode então ser analisada para identificar padrões e as suas ligações com os traços psicológicos. Em contextos forenses, essas ligações podem fornecer conhecimentos preciosos. Já existem abordagens psicológicas e linguísticas para análise de textos forenses, mas a abordagem psicológica carece em grande parte de fundamentação de teorias*

linguísticas e a abordagem linguística não permite normalmente considerar características psicológicas. Este artigo tem como objetivo avançar no sentido de colmatar essa lacuna. Nele analisamos o sistema de atitude, conforme grelha de avaliação desenvolvida por Martin and White (2005) e adaptado por Gales (2010) e por Hurt (2020), aplicando-o aos textos de quatro assassinos em série com diagnósticos de saúde mental documentados. Identificámos quantitativamente padrões significativos nos recursos de atitude, que analisámos qualitativamente à luz dos traços psicológicos identificados nos diagnósticos dos autores para avaliar a existência de uma relação entre eles. Apesar das óbvias limitações impostas pelo tamanho da amostra, os resultados deste estudo sugerem que a abordagem apresentada neste artigo merece uma investigação mais aprofundada.

Palavras-chave: *Análise de avaliação, Posicionamento, Psicopatologia, Atitude, Assassinos em série.*

Introduction

In Wichita, Kansas on the morning of January 15, 1974, Dennis Rader cut the phone line to the Otero family's home before entering (Slevin 2004). He bound and strangled the parents and killed the youngest two of the couple's five children; the three eldest later finding the bodies when they came home from school (Ott 2021). These four murders marked the beginning of a 31-year manhunt for Rader—who became known, infamously, as B.T.K., for Bind, Torture, Kill—during which time he committed six more murders and taunted the police and public with letters and messages claiming responsibility and demanding recognition for his crimes. During his trial, he was assessed by a psychiatrist who determined he met the diagnostic criteria for obsessive-compulsive disorder (OCD) and narcissistic personality disorder (NPD; Ramsland 2016). Rader, of course, is not the only serial murderer to send letters to law enforcement, but cases like his beg the question: was there evidence in his language choices of any of the traits that comprise the two diagnoses he received and what good would it have done if there was?

An individual's collective set of psychological traits, or their *psychopathology*, impacts how they interpret information and interact with the world and the people in it, which, in turn, influences myriad facets of their functioning. For instance—and perhaps what most would think of first—how they react to stimuli or their general behavioral patterns. However, what we are concerned with here is its influence on the perhaps less obvious facet of language. More specifically, how the influence of psychological traits on an individual's experience might be reflected in the linguistic choices made when describing and evaluating aspects of that experience—i.e., when expressing their *stances*. A person's *stance* conveys their "personal feelings, attitudes, value judgments, or assessments" about themselves and the people and things in their world (Biber *et al.* 1999: 966). In this sense an individual's language arguably represents an analyzable manifestation of their world view—a translation of their "internal thoughts and emotions into a form that others can understand" (Tausczik and Pennebaker 2010: 25). We contend, therefore, that through the examination of evaluative language choices we may identify connections between the linguistic resources used to express stances and the psychological traits that influence them.

Of course, the relationship between language and internal psychological processes is a complicated one and there is ongoing debate within the literature as to what linguistic

output actually reveals about cognition. Some have argued that certain features—e.g., cohesion, coherence, or frequency of lexical categories like emotion words—are indicative of aspects of mental health disorders such as symptom/trait severity or functional impairments (e.g., Buck and Penn 2015; Docherty *et al.* 1996; Gawda 2013; Rochester and Martin 1979). Others, like Edwards and Potter (1992), have argued that the relationship is less direct; language enables social interaction and linguistic choices reflect more the social goals of the speaker or writer in a given context than any underlying cognitive processes. The view adopted in this research lies somewhere in the middle; linguistic choices made when recounting experiences are sensitive to, and thus impacted by, numerous factors including situational ones like audience as well the psychological traits that influence those experiences (Bortolan 2019; Fine 2006).

There are many potential applications for stance analysis in forensic settings, such as the assessment of threats (e.g., Gales 2010, 2015) and pledges to harm Hurt (2020) or the analysis of cross-examinations of witnesses in assault trials (Gales and Solan 2017). However, what has yet to be considered is its potential value as a bridge between forensic psychological and forensic linguistic assessments of texts. Forensic psychological assessments often take little account of linguistic evidence as an insight into psychological states and traits and the consideration of psychological traits is generally seen as being beyond the scope of forensic linguistic analyses (Grant 2008). The aim of this research is to demonstrate how stance analysis might be implemented as a way to analyze language evidence for psychological information that is based in linguistic theory. It should be noted—stressed, even—that this research is **not** intended, nor would it be able, to be used by law enforcement to diagnose authors of forensic texts. The approach is designed to produce a linguistically-based account of the psychological *traits* of an author (which may or may not constitute a clinical disorder) that when combined with forensic psychological and linguistic assessments, could be used to help narrow a suspect pool.

The analysis aims to provide detailed account of stance-taking resources. Since stance can be realized at almost every level of language, it can be analyzed in a variety of ways. For example, corpus methods allow for tracking of grammatical stance markers (Biber *et al.* 1999)—such as adverbials or modals—across larger datasets (e.g., Gales 2010), while methods such as Appraisal Analysis (Martin and White 2005) allow for more fine-grained analyses on (typically) smaller datasets (e.g., Hurt and Grant 2019). In fact, it is precisely because it allows for such fine-grained analyses that Appraisal was chosen for this research. It offers a way to track in detail the types of feelings being expressed, at whom/what those feelings are directed, the level of commitment to them, and their intensity (Martin and White 2005). Our broader study engages in the spectrum of areas analyses offered by appraisal, but this paper focuses only on the system of *attitude*, which helps categorize the types of feelings expressed and the people and things at which they are directed.

We thus explore the patterns of resources employed by four infamous serial murderers in writings on the topics of crime, interpersonal relationships/interactions, and childhood memories and how those patterns relate to psychological traits associated with the mental health diagnoses they received.

Data

Selecting Authors

The data collection process began with the identification of authors through two broad selection criteria. The first was that they must be a convicted serial murderer with publicly available writings (e.g., as part of a biography). The second was that they must have a documented mental health diagnosis. This second criterion was included because it provided the most reliable way to remotely gather information on the author's psychological traits as the criteria used to make diagnoses necessarily detail the basic traits that comprise them (Association 2013).

Figure 1 details the four authors selected based on these criteria and the diagnoses they received alongside the source for this information. The official psychiatric reports for all four authors could not be found, but the source of the information for each either cites the original report or was written by the psychiatrist who did the assessment.

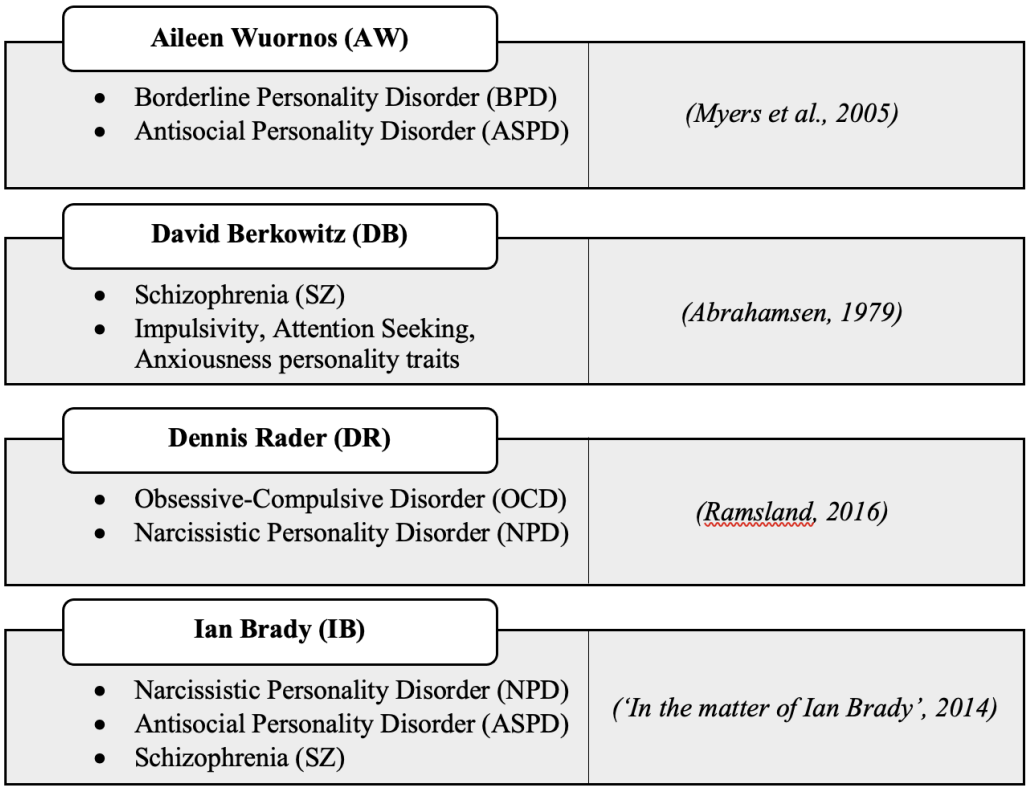

Figure 1. The authors and their diagnoses

Selecting Texts

Once the authors had been identified, the texts were selected from available writings. This meant there were limited choices and decisions about what potentially influential factors could be controlled for had to be made. It proved difficult to completely control for register variables like audience and mode of communication (Biber *et al.* 1999; Halliday and Matthiessen 2014), though efforts were made to limit the magnitude of the differences in these variables as much as possible. What could more easily be controlled for were the length, topic, and *genre* of the texts. *Genre* was the primary

search criterion since it is a set of conventions for the organization and structure of texts that can affect the range of language resources an author sees as available to them (Bhatia 1993; Martin 1997; Swales 1990). For this, we turned to the phenomenological approach to psychopathology, which aims to describe the "conscious, lived experiences" of patients and "what [that] conscious experience—understood in the widest possible sense—reveals to the disciplined observer" (Moran 2019: 205). Thus, texts needed to be *first-person accounts*, which are broadly defined here as a description of an experience or event written by the authors from their own perspective. As a genre, they impose few constraints on organization and content and they allow one to observe "the various ways in which everyday experience can be disrupted" or altered by psychopathology (Bortolan 2019: 1054).

Of course, such a broad definition did not narrow the list of possible texts by much and additional criteria relating to topic and length were developed. It was decided that for topic—the only of the three main register features (Biber *et al.* 1999; Halliday and Matthiessen 2014) that could be controlled for—there needed to be more than one so that any significant findings would not be attributable to subject matter alone. The topics also needed to be general enough to be able to find examples for all four authors but distinct enough to warrant separate classification. The result was three topics, or *sub-genres* (as they are all first-person accounts with variation only in their content), a breakdown of which can be seen in Figure 2.

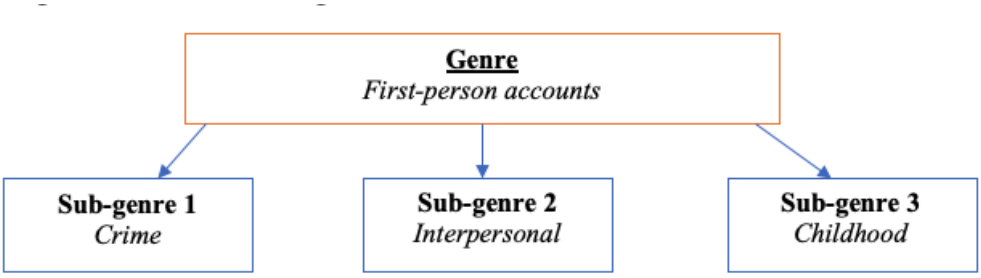

Figure 2. Breakdown of genre

Three texts — one per sub-genre — were collected for each author for a total of 12 texts, each ranging from 500 to 1000 words in length. For AW, two of the texts (*interpersonal* and *childhood*) were letters she had written to a friend who later reproduced them in a book (Wuornos 2011) and the third was a written record of a spoken interaction from a biography (Wuornos and Berry-Dee 2006). Of course, spoken and written modes impact language choices differently (Biber *et al.* 1999), but for AW, it appeared that she wrote and spoke in a very similar, informal way, so it was decided that the value of the text as the choice for the *crime* sub-genre outweighed the risk in using a spoken excerpt. DR's texts were also selected from his biography, in which the author reproduced letters DR had written her with small interjections from her for clarity, though none of the excerpts chosen for DR contained such interjections (Ramsland 2016). For DB, excerpts from letters he had sent the psychiatrist who assessed him during his trial (and who later wrote DB's biography) were pulled from an article on the website for *New York Magazine* ('The Letters of Son of Sam', 2006). Finally, IB's texts were excerpts from a book he wrote

himself (Brady 2001), meaning that unlike the other authors, his putative audience was anonymous to him.

For the *crime* texts, the goal was to find an excerpt where the subjects described one or more of the murders they committed. This, however, was not possible for IB as he did not discuss his own crimes in his book. Instead, for him the text that was chosen included a mention of his interest and engagement in criminal activities but was primarily focused on a generalized description of 'the serial killer'. The *interpersonal* texts all discussed some topic related to social experiences such as a relationship or a specific interaction. Finally, the *childhood* texts all included a recounting of an event or phase of their life during pre-teen or teen years.

Before delving into the analysis section, a few limitations should be addressed. First, while it can be argued that all the texts generally served a similar purpose—i.e., the author's desire to tell their sides of the stories—the inability to control for audience introduces its own set of complications. The relationship between author and audience is generally agreed to be an impactful variable on language choices (e.g., Bell 1984; Halliday and Matthiessen 2014). For DB, DR, and IB, their texts were produced with the knowledge that they—or at least the information contained within them—would be disseminated to the public at large and were all written to or for relative strangers. For AW, this was not necessarily the case; two of the texts were letters written to a close friend and the other was produced as part of an interview in which she provided a confession for her actions, which is typically an inherently adversarial situation (Shuy 1998).

The second issue concerns the subjectivity of the texts as first-person accounts. Phenomenologists argue that subjective accounts allow for a more "comprehensive and faithful account of the phenomena under consideration" (Bortolan 2019: 1054) because the phenomena at issue are subjective themselves. However, the question still remains: how generalizable can any observations be with such person-specific data? If the first-person accounts are considered in isolation, generalization of any findings would, naturally, be impossible. The phenomenological solution to this dilemma is to therefore take accounts from a multitude of sources and identify the features which are *invariant* across them as well as features which are distinctive (Bortolan 2019).

It is acknowledged that with only four subjects and 12 texts, there is not enough data to warrant broad conclusions or generalizations and the different author-audience relationships represent a potential confounding variable. However, in the domain of the study of serial killers, there are relatively few possible subjects to begin with and even fewer with documented diagnoses who have written texts appropriate for analysis. Thus, in this context, and given the novelty of the approach, studying a small number of individuals and texts (even without the ability to control for audience) is a good first step as much can arguably still be gained from the patterns that do emerge, especially considering the detail produced by appraisal analyses.

Appraisal Analysis

Appraisal provides the means to analyze the patterns of linguistic resources authors use to express their *stances* (Martin and White 2005). It is comprised of three systems, each capturing a different type of resource. The system of *attitude* helps identify the core feelings of the stances; the system of *engagement* tracks whether the stance is presented

in a way that opens or closes the conversational floor; and the system of *graduation* acts as an umbrella system, tracking intensity of the feelings and level of commitment to the stance. For the purposes of this paper, we are focusing only on the system of *attitude*.

Attitude is divided into three types—*affect, judgment,* and *appreciation*—each of which encompasses a distinct set of resources for conveying different kinds of *feelings* (Martin and White 2005). *Affect* is considered to be at the core, covering **personal** emotions. It can be further separated into four categories, which are comprised of positive and negative items for each type of emotions. *Un/happiness* encompasses "emotions concerned with 'affairs of the heart' " such as *happiness, sadness, love,* and *hate* (Martin and White 2005: 49). *In/security* covers emotions relating to one's environment including *anxiety, confidence, trust,* and *uneasiness*. *Dis/satisfaction* is concerned with feelings about activities in which one is a participant, or of which one is a spectator such as being *angry, interested, bored,* or *engrossed*. Finally, *dis/inclination* deals with feelings toward the *irrealis*, or things that have not actually happened, such as *desire, fear,* or *keenness*.

The system of *judgment* covers five categories of **institutionalized** (shaped by society) feelings directed at the behaviors and traits of oneself and others (Martin and White 2005). *Normality* is for assessments of normalness and specialness, such as *normal/abnormal* or *stable/erratic*. *Capacity* was originally intended just for assessments of capability, such as *strong/weak* or *successful/unsuccessful*, but it has also proven useful for capturing expressions of violence where one person is being presented as **incapacitating** another person such as with verbs like *kill, harm,* or *attack* (Hurt and Grant 2019). *Tenacity* is for assessments of determination and dependability, such as *patient/hasty*, *loyal/disloyal,* or *brave/cowardly* (Martin and White 2005). *Veracity* is for assessments of honesty, such as *truthful/deceitful* or *discreet/blabbermouth*. Finally, *propriety* is for assessments of morality and ethics, such as *good/evil* or *altruistic/selfish*.

The final system of attitude, *appreciation,* is divided into three main categories: *reaction, composition,* and *social valuation* (Martin and White 2005). *Reaction* captures feelings similar to those seen in *affect,* but the focus is placed on the trigger of the emotion instead of the person feeling it, for instance saying *a boring building* instead of *this building bores me*. *Composition* captures evaluations of a 'thing's' "balance and complexity" (Martin and White 2005: 56), such as things that are *symmetrical* versus *asymmetrical, logical* versus *illogical*, or *simple* versus *complicated.* It has also been found to be useful for capturing references to things that maintain order and things that cause *dis*order, such as weapons (Gales 2010; Hurt 2020). Finally, *social valuation* covers assessments of the value provided by something, such as it being *innovative, authentic, derivative*, or *fake* (Martin and White 2005). Additionally, Martin and White briefly touched upon the link between valuation and judgment, stating that "positive and negative valuations of something imply positive and negative judgments" (2005: 58). Hurt (2020) further explored this and demonstrated how additionally coding items of *valuation* for the *judgment* types they instantiated was a useful expansion of the system, especially in forensic settings.

Instances, or *tokens*, of each of these attitude types can then be coded for additional variables that help generate a representation of an author's perceptions of and interactions with the world and the people and things in it. *Polarity* distinguishes between the positive and negative tokens and helps determine the overall tone of a text

(Martin and White 2005). *Explicitness* differentiates between *inscribed* attitudes that are conveyed directly through the word or phrase used (i.e., based on *denotation*) and *invoked* attitudes, which require shared knowledge or context to interpret (i.e., based on *connotation*). The final two variables help track who does the appraising—which is typically, though not always, the author—and who/what is being appraised. Coding for the *appraiser* makes it possible to not only see how the authors themselves view the world, but also their perception of how others view it and coding for the *appraised* makes it possible to track the distribution of self- and other-directed evaluations.

There are numerous ways to code texts for Appraisal, both by hand and using software. For this study, the software *UAM CorpusTool* (O'Donnell 2019) was used because it includes Appraisal as one of its built-in coding schemes and can provide raw frequencies of all coded features for individual texts and groups of texts. The 12 texts were coded one-by-one with each of the systems of Appraisal as a separate layer. Each word or phrase which conveyed an attitude was coded for the specific *attitude* category it best represented—e.g., *un/happiness* or *normality*—as well as for *polarity, explicitness, appraiser,* and *appraised.* These decisions were then discussed by the authors, especially when potentially contentious, but no formal reliability study was carried out.

The relationship between the stance-taking resources outlined above and psychological traits has yet to be investigated in the literature, at least in the way it is done here, and what research does exist on language and mental health has rarely examined personality disorders or pathological personality traits, with which most of the authors were diagnosed. This in conjunction with the inherent complexities of psychopathology made generating specific hypotheses about the results difficult and a more exploratory analytical approach was taken. That being said, however, the mapping of stance-taking patterns onto psychological traits was not done without some guidance from related (though non-linguistic) research on how the mental health disorders of concern in this study impact world view.

Analysis

Total word counts for the individual authors ranged from 1626 to 2235 words. Given the wide range, token counts were normalized per 1000 words to make the frequencies more comparable. To determine where significant differences in proportions of evaluative resources occurred between authors, Chi-Square tests were used (as the data are inherently categorical in nature; Field 2018) on a pairwise process, for a total of six pairs. Within-author comparisons were also conducted to determine whether any of the significant between-author findings were attributable, at least in part, to disproportionately higher or lower usage of a particular feature in any of the sub-genres. The results from the quantitative tests were then used to identify areas of interest for a qualitative exploration of how the resources were employed by the authors. Table 1 shows the results from the between-author comparisons for all of the *attitude* variables described in the above section. For the sake of brevity, though, only the quantitative results for the variables within the three *attitude* types and *polarity* are considered in this paper.

The cells contain the normalized frequencies of each variable per 1000 words. The highlighted cells indicate that the variable was used by that author at a significantly higher proportion (at the level of 0.05) than the author denoted by the superscript. For

		AW	DB	DR	IB
Affect	un/happiness	1.95	9.23	9.40^{AW}	2.44
	dis/satisfaction	5.85	13.53	8.50	11.23
	in/security	$17.05^{DB,DR}$	9.84	5.37	7.81
	dis/inclination	7.31	14.76	5.82	3.91
Judgment	normality	3.41	11.69^{AW}	13.87^{AW}	11.23
	capacity	39.45	16.61	30.87	38.57
	tenacity	24.35	11.69	14.32	22.95
	propriety	26.30	21.53^{DR}	11.19	40.53^{DR}
	veracity	5.36	9.23	6.26	10.25
Appreciation	reaction	3.41	9.84	3.58	7.32
	composition	11.20	15.99	9.40	21.48
	social-valuation	16.07	22.76	18.34	46.88
Polarity	positive attitude	63.81	55.97	53.69	$125.49^{AW,DB,DR}$
	negative attitude	97.42^{IB}	110.70^{IB}	83.22^{IB}	99.12

Table 1. Between-author comparisons for attitude – frequency per 1000 words

instance, in the row containing *un/happiness*, DR's proportion of use was significantly higher than AW's, thus DR's cell is highlighted and the superscript indicates this relationship to AW. As this table shows, there are several *attitude* variables which were used at significantly different proportions by the four authors. It should be noted that variables are discussed in terms of percentages from this point on, however, because the normalized counts do not always properly represent their proportions. For instance, author A may use overall more attitude than author B, making the frequencies, even when normed, for author A appear slightly higher, but the proportion of total attitude tokens could be exactly the same.

The results pertaining to each author and their individual patterns of use are presented in turn. How these patterns then relate to their psychopathologies is explored in the discussion. Since the literature on analyzing linguistic, and specifically stance, resources in this way is scarce, these interpretations involved extrapolating based mostly on current understandings of how the different mental health disorders and their component traits impact perception of and interaction with the world.

Aileen Wuornos

For overall attitude, AW demonstrates a higher use of *judgment* than any other resource—accounting for 61.15% of total *attitude*—using *affect* and *appreciation* at similar rates—accounting for 19.88% and 18.98% of total *attitude*, respectively. Interestingly, as Figure 3 shows, AW consistently employs positive resources about 40% of the time and negative ones about 60% of the time regardless of *attitude* type.

Beginning with *affect*, AW was found to use *un/happiness* at a significantly lower rate than DR and *in/security* at a significantly higher rate than both DR and DB. Additional chi-square tests comparing proportions when further divided by polarity revealed that AW significantly differed from DB and DR on *-happiness* and from DB on *-security*. Figure 4 displays the distributions of AW's positive and negative affect types as percentages of total attitude tokens.

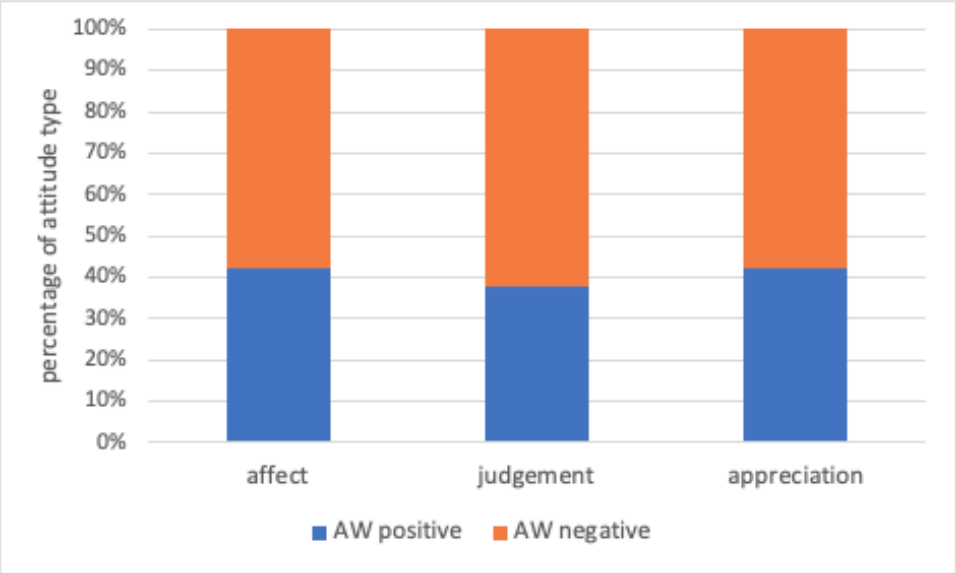

Figure 3. Distribution of positive and negative attitude types for AW

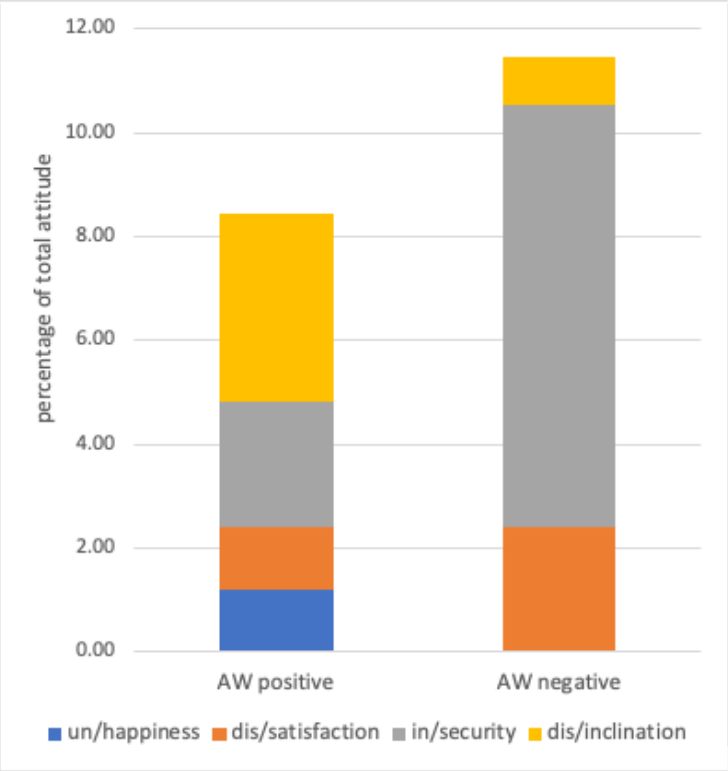

Figure 4. Distributions of positive and negative affect for AW

As Figure 4 shows, AW uses *-security* more often than any other type of affect and does not use any *-happiness.* As mentioned in the above section, *in/security* is concerned with feelings about one's environment as well as to express certainty/uncertainty about knowledge one possesses (Martin and White 2005). Looking more closely at how she uses *-security*, it is revealed that she most often focuses on the aspects of her environment that were most unsettling and that presented a threat of harm—such as when she described feeling *uneasy* and *scared* in the *errie [sic] dark* surrounding a man's house or when she talked about *going through withdrawals* during an interrogation. In the rare instances that she used *+security*, it was to declare a level of certainty about something, such as when she *knew [Mallory was] going to rape* her or when she *wanted to make sure…[she] could use [her] gun.*

In terms of overall *judgment,* AW was found to use significantly less *normality* than DB and DR. However, when divided by polarity, significant differences are found for the variables of *+normality, +tenacity, -normality* and *-capacity* between AW and at least one other author. More specifically, her use of *+normality* was significantly lower than DR's; her use of *+tenacity* was significantly higher than IB's; her use of *-normality* was significantly lower than DB's; and her use of *-capacity* was significantly higher than IB's and DB's. Figure 5 shows a breakdown of the different judgment types in terms of their percentages of total attitude tokens.

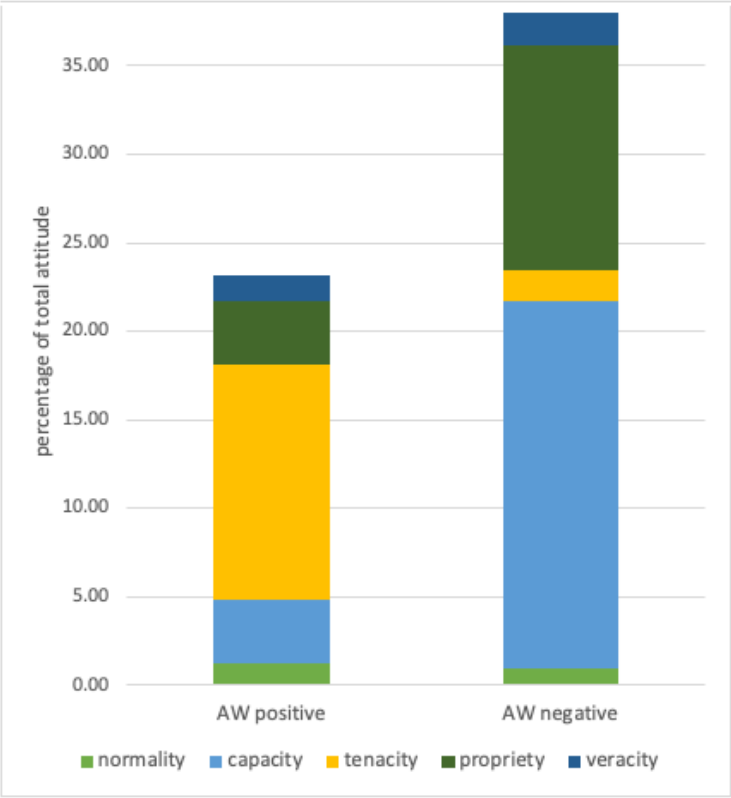

Figure 5. Distribution of positive and negative judgment for AW

Starting with *normality*, positive tokens in this category convey judgments of how special or normal someone is while negative tokens often convey judgments of bizarreness (Martin and White 2005). Despite the low amount of *normality* used by

AW, the function *-normality* served when she did use it was to mark the actions of her victims as out of the ordinary, adding to the support for her assertion that they were intent on causing her harm. For instance, when she asked her first victim when he was allegedly being too rough why he *still [had] to have [his] clothes on* and, referring to the same encounter in a different text, how she had *never had sex like that.*

For *+tenacity,* which accounted for 13.25% of AW's total *attitude* tokens, it was most often used to convey assessments of her own and others' determination to achieve something. When evaluating others, this determination is presented as being directed toward something that would cause her harm, such as how her interrogator *kept stokien [sic]* her despite her apparent distress or how one of her victims was *going to rape* her (a sentiment she repeats multiple times). When evaluating herself, it is more in reference to her resilience in response to the harmful actions of others, such as her *struggling* to break free then *booking it* away from her attacker or her *trying to* recollect and recount past assaults against her to an apparently apathetic interrogator.

Finally, for *-capacity*, which accounted for 20.78% of her total *attitude* tokens, AW often referenced violent acts committed both against and by her—a coding decision based on Hurt and Grant (2019) work demonstrating the usefulness of *-capacity* for tagging violent verbs, where one person is by definition incapacitating another. All three of her texts included at least a small piece devoted to detailing an assault (or intended assault) against her, and in two of those texts, these were immediately followed by her detailing how she retaliated against them. For instance, she describes how one of her victims was *pushing [her] down* and that she was convinced he was *going to rape* her and that another was *going to blow [her] brains out*, which were the reasons she *shot* (and ultimately killed) them. She also talked about a man who was *on top of* her and *slapping* her before she got away and when he pursued, she started *kicking* him until she *felt he was down enough.* She also sometimes used *-capacity* and *+tenacity* together to assert that others intended to—and were determined to—harm her, thus justifying her actions.

It should be noted that AW's use of *-capacity* was one of the three significant between-authors results for which there was a potentially influential significant within-authors finding. Figure 6 shows the distribution of negative *judgment* types across AW's three texts as a percentage of the total *attitude* for each respective text—e.g. *-capacity* makes up about 19% of the total *attitude* seen in her childhood text.

As Figure 6 shows, a significantly higher proportion of *-capacity* is used in AW's childhood text than her interpersonal text, in which *-propriety* instead dominated. Part of what could explain this discrepancy is that the interpersonal text centered around AW's rationalization to her friend about why it was incorrect that she had been classified as a serial killer. This involved detailing what she perceived to be corrupt and/or cruel actions (i.e., judgments of *impropriety*) by various others, such as the interrogators who 'knew' she was distressed but *kept stokien [sic]* her anyway or the *crooked scum* who spread the serial killer story. Ultimately, the *-capacity* in her childhood and crime texts served a similar function to that of the *-propriety* in the interpersonal text; these differences did not change the overall message that she perceived others as being intent on causing her harm, they only indicated a shift in the resources used to communicate it.

For *appreciation*, the third and final main *attitude* type, the only significant finding was that IB employed it—or more specifically *+appreciation*—at a significantly higher

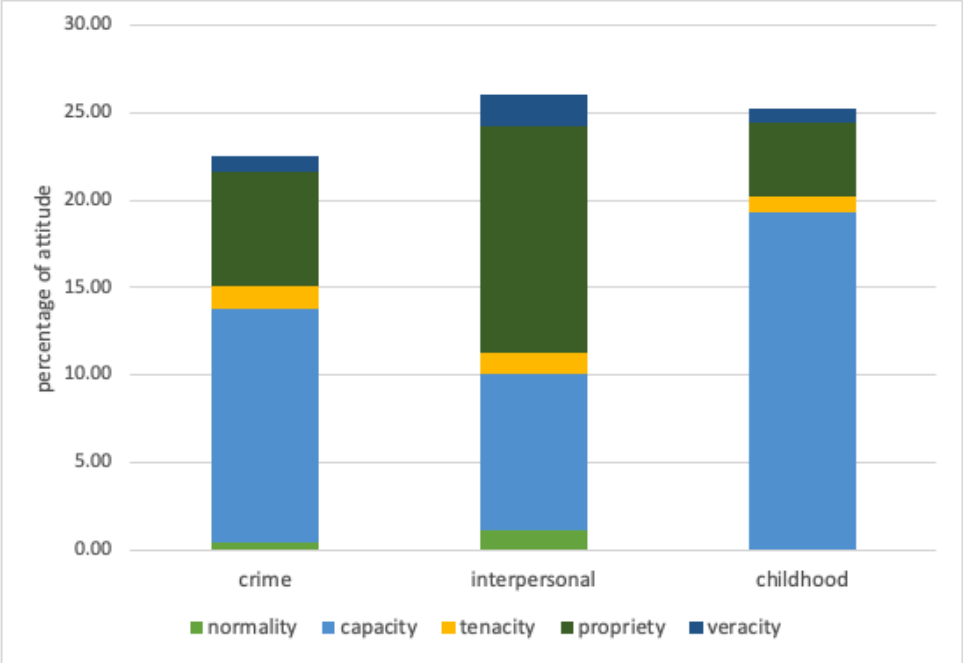

Figure 6. Distribution of negative judgment in AW's three texts

rate than AW. However, when the more specific categories within *appreciation* were analyzed, the significant difference no longer held, and no new significant differences emerged.

David Berkowitz

Similar to AW, DB used *judgment* the most, accounting for 42.43% of total *attitude*, and used *affect* and *appreciation* least, accounting for 28.41% and 29.15% of total *attitude*, respectively. Consistently, about one-third of *attitude* tokens, regardless of type, were positive while the remaining two-thirds were negative, as seen in Figure 7.

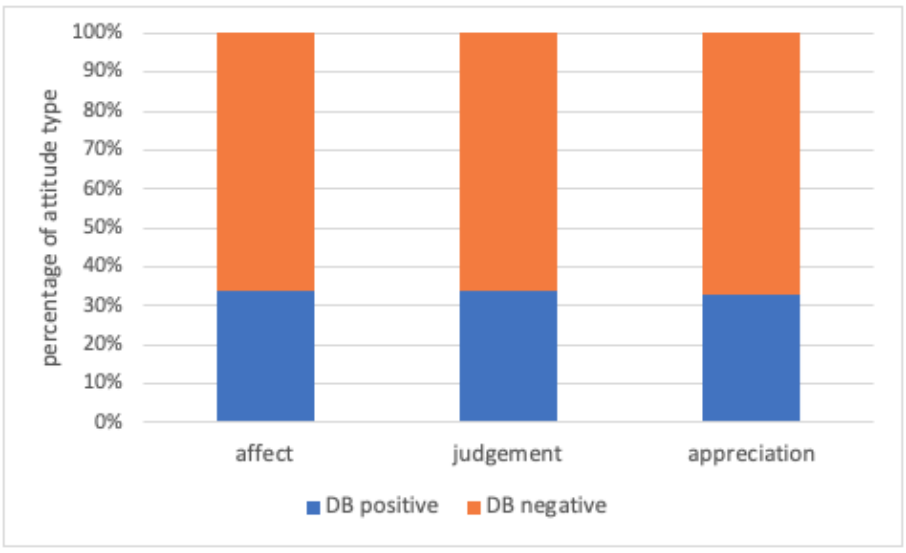

Figure 7. Distribution of positive and negative attitude types for DB

Table 1 shows that, within *affect*, DB uses a significantly lower proportion of *in/security* than AW. When positive and negative *affect* are analyzed separately, it is revealed that this difference holds only for *-security* and that he additionally demonstrates a significantly higher proportion of *-happiness* than AW does. Figure 8 shows the breakdown of DB's use of positive and negative *affect*.

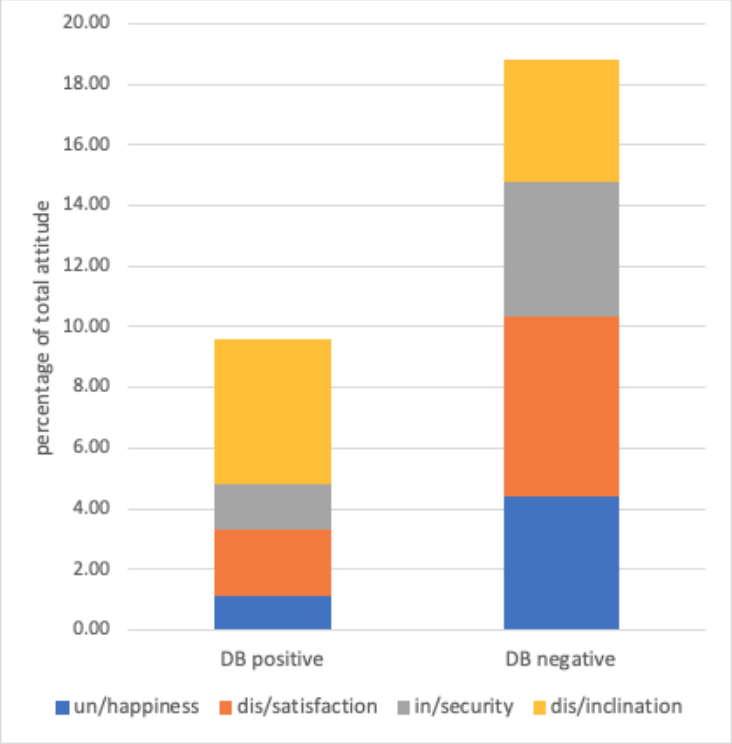

Figure 8. Distribution of and negative affect for DB

-Happiness was most often used in reference to his general mood, such as his feeling *hopeless*, *miserable*, or *unhappy*. However, he also attributed feelings of *unhappiness* to others, such as his biological mother giving birth to him to *spite* his biological father or how he *hurt* his adoptive father with his actions. Often these attributed feelings of *-happiness* occurred in conjunction with his own and alongside instances of *-security,* such as when he talked of his *lonliness [sic]* or the *guilt* he felt about his mother apparently dying in childbirth. These resources combined to provide a sense of an overall negative view of the world and of himself.

In terms of *judgment*, DB used significantly higher proportions of overall *normality* than AW and of overall *propriety* than DR. When analyzing positive versus negative *judgment*, it is found that DB used a significantly higher proportion of *+veracity* than DR, of *-normality* than AW and IB, and of *-propriety* than DR. The distributions of positive and negative judgment can be found in Figure 9. The judgments of *-normality* were typically directed at himself, referring to his *rotten social life* or to his crimes as feeling like *the actions of a stranger*. He used this alongside *-happiness* to convey a generally negative view of himself and his life. He employed *-propriety* tokens often in reference to his own actions as well as in reference to others as justification for his crimes. When self-directed, these tokens captured actions ranging from his childhood crimes of writing *graffiti and curse words* on the walls and starting *garbage and abandoned car fires* to his

later *shootings* as his way to get *revenge*. When outward-directed, he talked of his *inconsiderate neighbors,* people *mentally oppressing* him, and how the people he killed *[weren't] so innocent*.

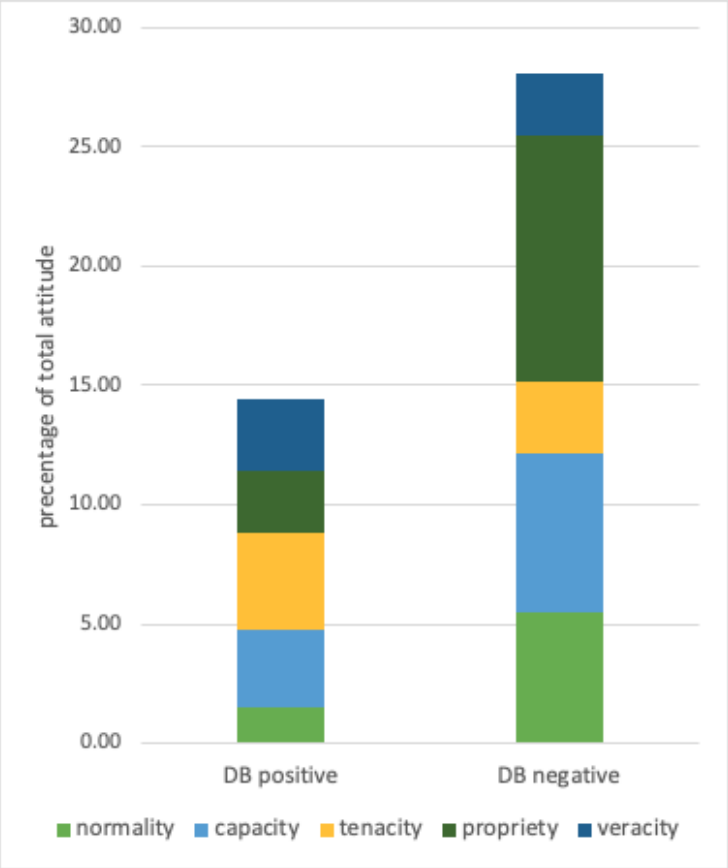

Figure 9. Distribution of positive and negative judgment for DB

There was a potentially influential within-author finding in DB's use of *-propriety* between his *interpersonal* and *childhood* texts. Figure 10 displays the distribution of negative *judgment* across DB's three texts as a percentage of the total *attitude* for each respective text. As the graph shows, the overwhelming majority of his negative *judgment* tokens in the *childhood* text—of which there were far fewer than in the other texts—were *-propriety*. This is because, in this text, he spent a good deal of time describing how he committed a multitude of criminal acts as a minor before he ever committed his first murder. Despite this, the way that he used the *-happiness, -normality*, and *-propriety* tokens was fairly consistent across the texts. *-Happiness* and self-directed judgments of *-propriety* worked hand-in-hand to ostensibly convey a level of remorse for his actions as well as some acceptance of responsibility for harm he caused. Then, the judgments of *-normality* added to this by indicating that he viewed some of his more reprehensible actions as being 'out of the ordinary' for him.

Dennis Rader

Like AW and DB, DR used *judgment* the most, accounting for 55.88% of total *attitude*, and used *affect* and *appreciation* the least, accounting for 21.24% and 22.88% of total *attitude*, respectively. Unlike AW and DB, the proportions of positive and negative tokens were

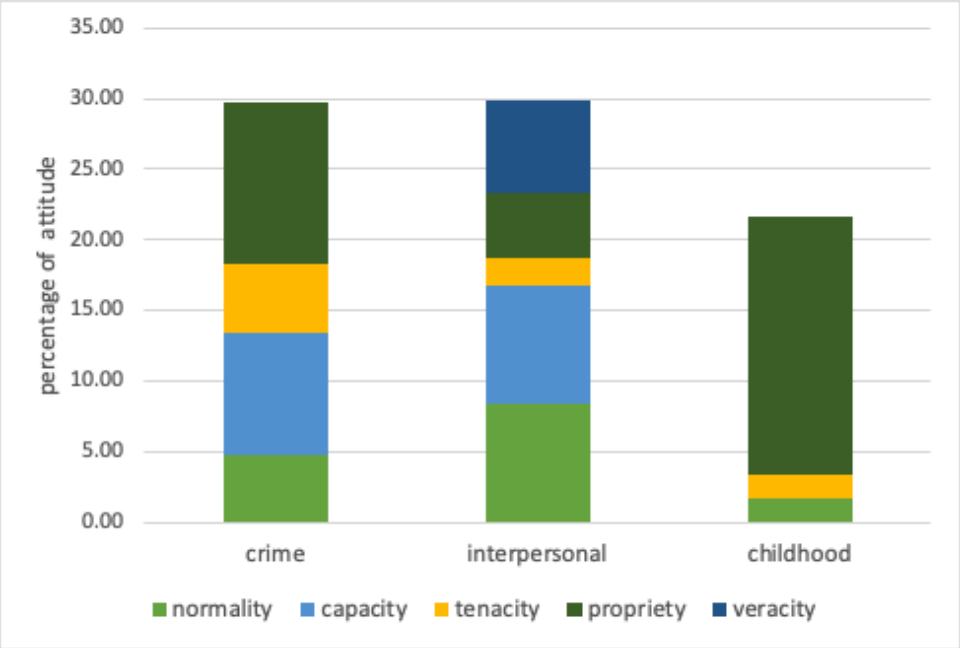

Figure 10. Distribution of negative judgment across DB's three texts

not consistent across the three *attitude* types, as Figure 11 shows. According to this graph, within *affect*, DR used more positive than negative while in both *judgment* and *appreciation*, the opposite was true. In fact, he used the highest proportion of positive *affect* of the four authors, which was significantly higher than IB's. However, this significant difference does not hold when the *affect* categories are analyzed separately.

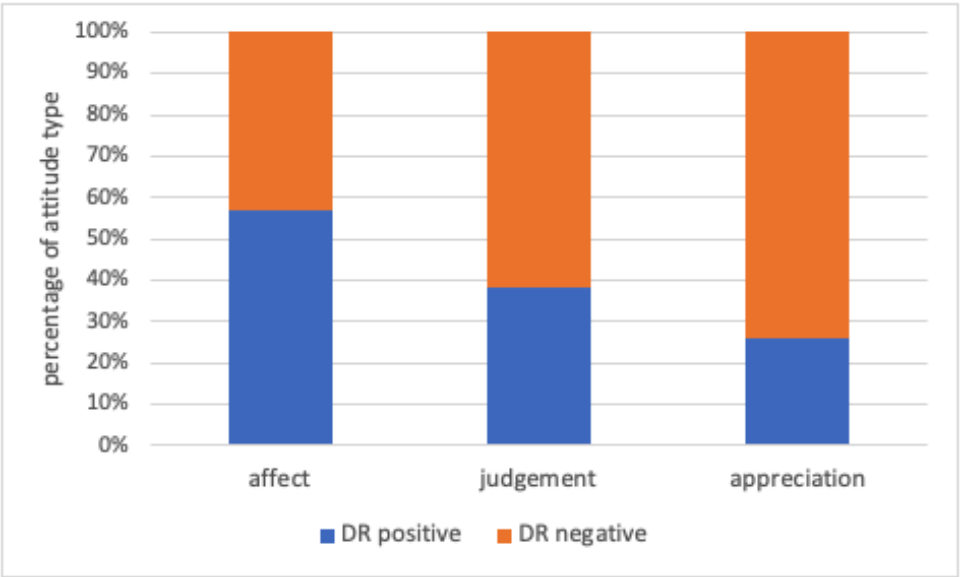

Figure 11. Distribution of positive and negative attitude types for DR

The results in Table 1 show that within overall *affect*, DR used a significantly higher proportion of *un/happiness* than AW and a significantly lower proportion of *in/security* than AW. The significant difference on the variable of *in/security* disappears when broken

down by polarity but does hold for *-happiness.* Figure 12 shows the distribution of the positive and negative *affect* types for DR.

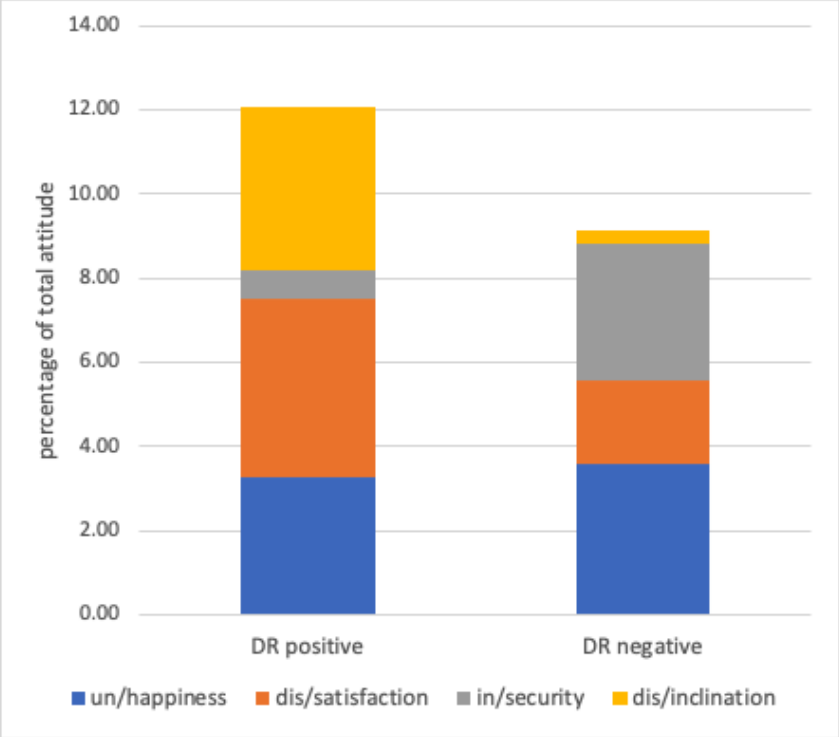

Figure 12. Distribution of positive and negative affect for DR

As Figure 12 shows, DR used similar amounts of positive and negative *happiness*, which he also used in similar ways. With *-happiness,* it was almost exclusively to express negative reactions to interpersonal stresses, such as when he *became physically sick* when his future wife did not come to a gathering or when his *heart was broken* by a childhood girlfriend moving away. His use of *+happiness* was most often feelings directed at people or things in his life rather than moods, such as how he *really loved* a childhood girlfriend or how he found a bookstore he *liked.*

For overall *judgment*, DR was found to use a significantly higher proportion of *normality* than AW and a significantly lower proportion of *propriety* than DB and IB. Within positive *judgment*, it was found that he used a significantly higher proportion of *+normality* than AW and a significantly lower proportion of *+veracity* than DB. Within negative *judgment*, he was found to use a significantly higher proportion of *-capacity* than DB and a significantly lower proportion of *-propriety* than DB and IB. The distribution of the positive and negative *judgment* types can be found in Figure 13.

DR's use of *-capacity* showed a lot of variation. In some cases, it captured instances of his discussions of *bondage*, which was part of his modus operandi. Interestingly, these instances were really the only ones that got close to referencing violent actions against another person. He also used *-capacity* in reference to himself and his capabilities, such as how he was a *slow learner* or his *failed attempts at kidnapping*. His use of *+normality* when used to reference himself or his own behaviors, typically served to demonstrate some aspect of 'normalcy' in his life or at least desire for such things. For instance, he

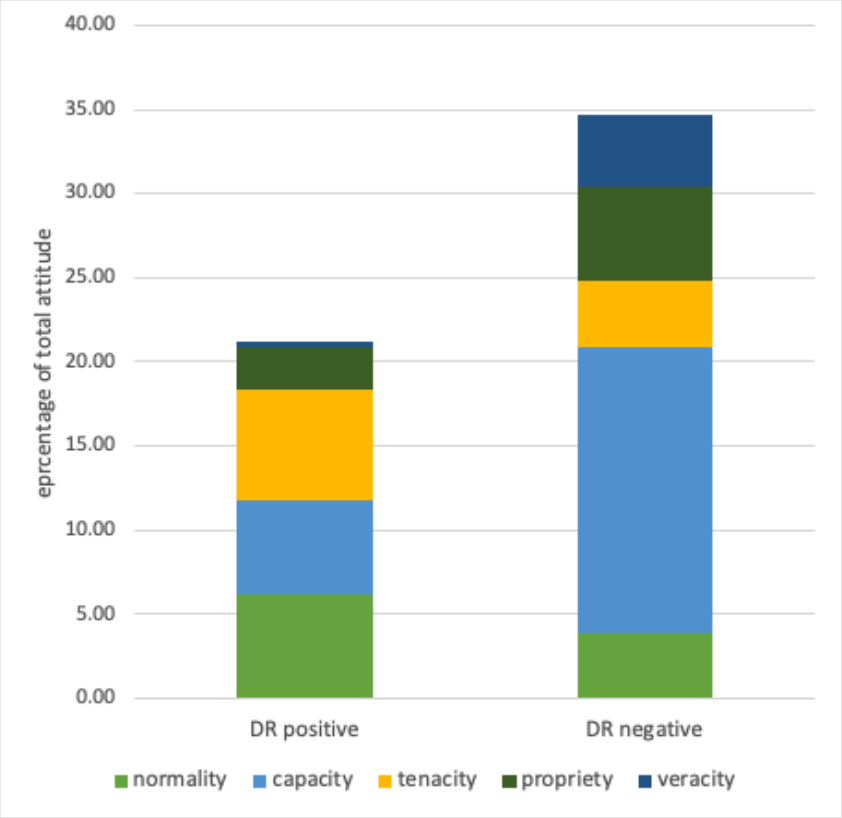

Figure 13. Distribution of positive and negative judgment for DR

talks about trying to *not draw attention* walking away from an attempted kidnapping as well as the various phases of his relationship with his wife from him *propos[ing]* to *looking for rings* to *[finding] a place to live*. He used *-normality* in a similar way, identifying personal experiences or traits that were more 'abnormal', such as him *being awkward* or a *lone wolf* but also pointing out things he found odd in others, such as when talking about a childhood girlfriend saying *but [she] was more of a tomboy*.

When DR used *-propriety*—which only accounted for 5.56% of total *attitude*—it was mostly in reference to himself, though not always explicitly. That is, while he did reference some of his illegal actions (or intentions to commit illegal acts), such as *kidnapping* and *stalking*, he also quite often referred to what he called his *Dark Side*, *Black Hat*, and *dark crack*. In doing so, he deflected some of the guilt for his crimes onto something ostensibly separate from his 'typical', moral self. Taken together, DR appears to be appealing to the audience by suggesting (1) he was not completely in control of his actions, (2) was aware of the extent of his abilities, and (3) had desires for and made an attempt at normal relationships in the past.

Ian Brady

IB was the only of the four authors to demonstrate a higher proportion of positive than negative *attitude* tokens. Interestingly, when separated by *attitude* type, this holds true for *affect* and *appreciation* but not for *judgment*, which was about 53% negative, as seen in Figure 14.

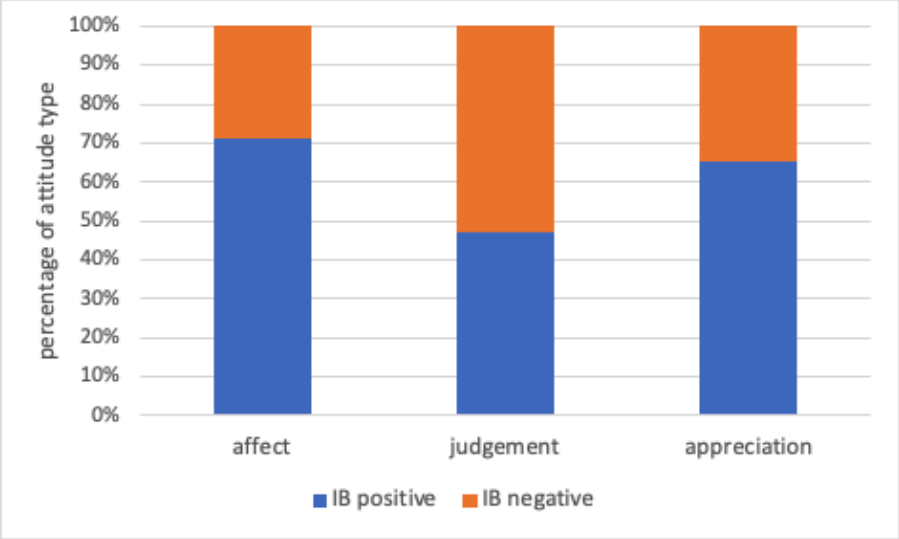

Figure 14. Distribution of positive and negative attitude types for IB

The only significant difference for *affect* between IB and the other authors was that he was found to use significantly less *affect* overall than all three, accounting for only 11.3% of total *attitude* tokens. The distributions of positive and negative *affect* are shown in Figure 15.

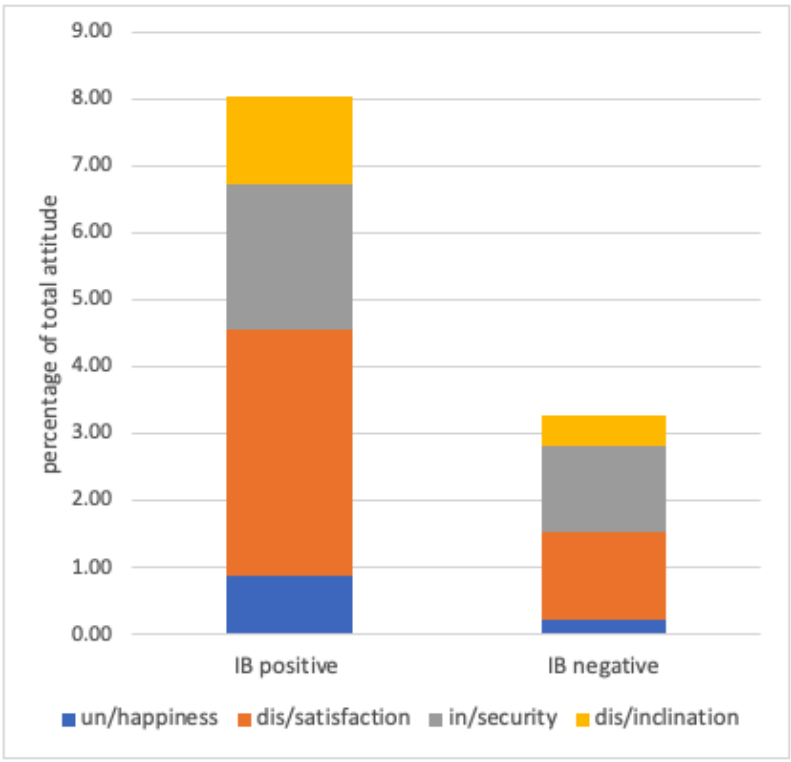

Figure 15. Distribution of positive and negative affect for IB

Perhaps most interesting is that about 60% of these were attributed to someone else—usually some generic third party—or were presented as a 'shared' feeling. For instance,

he talks of 'the serial killer' who *wants more NOW* and of the *common human tendency to luxuriate in fear-induced guilt*. When presented as 'shared' feelings, they are often structured in a way that indicates he views others as being no different from him, such as when he says *we are reluctant to admit that the child is still there deep in each of us* or that we all pretend we *are beyond good and evil and proud of it*. The *affect* tokens representing his own personal feelings were sometimes used to declare a lack of feeling toward something he did, like saying his book is *not an apologia* or that *adverse critism… will not cause [him] to retract one word*. Some of the remaining personal *affect* tokens were positive but directed at negative things, like his *boundless energy for criminal pursuits*, while others were positive feelings about positive things, like him *soaking up the atmosphere* or feeling *truly alive* when he visited childhood 'haunts'.

As Table 1 shows, IB uses a significantly higher proportion of overall *propriety* than DR. When broken down by polarity, it is revealed that this is specifically within *-propriety* and that he also uses a significantly lower proportion of *-capacity* and *+tenacity* than AW and of *-normality* than DB. The distribution of IB's positive and negative *judgment* types can be seen in Figure 16.

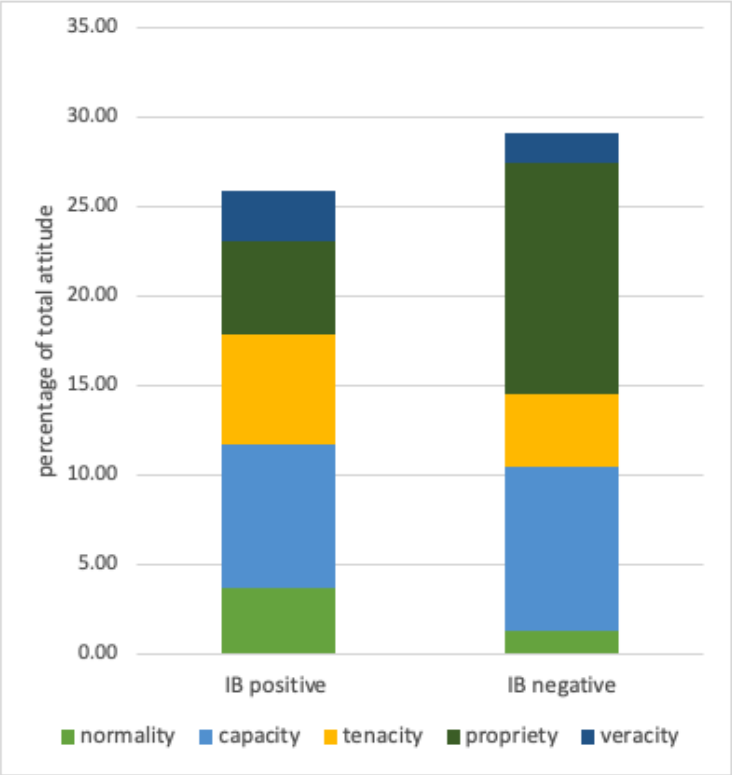

Figure 16. Distribution of positive and negative judgment for IB

IB's apparent preference for *-propriety* presented an interesting pattern. They were rarely self-directed, but when they were, it was not usually to denounce his actions. For example, he explicitly states that his book is *not an apologia* and that *remorse is a purely personal matter* and not something for the public to witness. He also talks of his *boundless energy for criminal pursuits* and his *studious, professional attitude towards crime*, implying a positive view toward illegal acts. Most often, *-propriety* was used in reference to some vague other person/group or society as a whole. For instance, he talks

about the *seed of corruption* being inside everyone, about *human savagery*, and about the *greed* and *hypocrisy* of people in power. This pattern demonstrated not only that he appeared to hold views of ethics and morality that were almost completely opposite of society's but that he also viewed himself as separate and apart from most of society in such matters. In fact, he even said as much, claiming he *[is] no longer of your world,* that he *[is] now simply a curious observer.*

He appeared to combine these morality judgments with judgments of *-capacity* to add to his seemingly self-righteous attitude. He does this by evaluating the capabilities of the masses and rarely turning them toward himself. For example, he talks of the *stultifying lassitude of ordinary modern life* and of a *freedom of thought and expression* he says most free people *will fail to experience.* When these judgments are turned inward, it is usually to talk about his imprisonment with terms that imply he is being incapacitated by someone—for example, when he talks of being in *captivity* or refers to his reader as his captor saying *you contain me till death in a concrete box.*

Discussion

The value of Appraisal in forensic linguistic contexts has been demonstrated in research on, for instance, threatening communications (e.g., Gales 2010; Hurt and Grant 2019) or courtroom discourse (e.g., Gales and Solan 2017). However, it's potential to be used to assess texts for evidence of psychological traits was yet to be explored. This paper demonstrated how Appraisal could be used for such purposes by helping to capture patterns in the linguistic choices of four serial murderers when expressing their stances. These patterns could then be examined in relation to the psychological traits of the author to determine if there is a connection between them. Figure 1 contained the diagnoses received by each of the authors from which the traits that are discussed in the coming paragraphs were derived.

The inherent complexities of psychopathology and the paucity of previous research in this area made it difficult to hypothesize specifically what kinds of patterns would be expected for any individual author. The combination of traits and how they interact necessarily affect the relative import assigned to different aspects of experience (Millon *et al.* 2012), meaning that the impact of any trait could reasonably be communicated using a variety of resource combinations. For instance, consider the trait of *hostility,* which is marked by anger as a frequent or persistent reaction to stimuli (Association 2013). There are numerous ways in which such anger could be expressed, primarily depending on what aspects of their experience the individual focuses on—e.g., use of *-satisfaction* would place the focus on the author's emotional reaction itself (internal), whereas to place the focus more on the trigger of the anger (external), some type of *appreciation* or *judgment* might be used. Due to the lack of any concrete path to guide expectations, a more exploratory approach was taken. That is, the patterns that were identified for each author were examined in relation to their known psychological traits. These are, of course, not necessarily the only possible interpretations of the results, but given the information that is known about the authors, they represent reasonable and defensible evaluations based on observable evidence.

Beginning with AW, the most salient pattern was the high proportions of *-security, +tenacity,* and *-capacity,* which were often used in tandem with each other. *-Security* conveyed the frequent feelings of anxiety AW had about the people and things in her

environment; *-capacity* detailed both the harm others inflicted—or she believed they would inflict—on her and the harm she inflicted on them in response; and *+tenacity* communicated her assessment of the level of determination of others to inflict said harm on her. This pattern suggests a tendency to view the world and others as threatening, which is consistent with what is known about the basic schemas often found in individuals with borderline personality disorder (BPD; e.g., Butler *et al.* 2002). More specifically, though, it is perhaps explained by the BPD traits of *interpersonal hypersensitivity* and *emotional lability* as well as the trait of *hostility,* found in BPD and antisocial personality disorder (ASPD; Association 2013). These traits mark an increased proneness to assuming malicious intent (represented by the combination of *+tenacity* and *-capacity* of AW by others), experiencing heightened negative emotional states and reactions (represented by *-security*), and reacting aggressively (represented by *-capacity* of others by AW).

For DB, the most notable pattern was the high proportions of *-happiness*, *-normality*, and *-propriety.* His use of *-normality* marks his actions as being 'out of the ordinary' for him, as something he perhaps would not have done given due consideration. The combination of the self-directed judgments of *-normality* and *-propriety* alongside *-happiness* portray him as being potentially remorseful for his actions. However, this is then contradicted by his use of other-directed *-propriety* as a way of justifying his crimes. One possible interpretation of this pattern is as a result of the *impulsivity* and *attention seeking* traits noted by Abrahmsen (1979) and the *aloofness* associated with the bygone paranoid subtype of schizophrenia with which he was diagnosed (Association 2000). *Impulsivity* is defined by spontaneous, unplanned actions (Association 2013), which one might later upon reflection view as having been uncharacteristic (*-normality*) for them. The inconsistency of DB's ostensible remorse alongside an assertion that there is justification for his actions may be the result of a combination of *attention seeking* and *aloofness*; the apparent claims of remorse (self-directed *-propriety* and *-happiness*) representing a desire to garner sympathy from the audience and the contradictory 'justifications' suggesting that remorse was a likely carefully crafted illusion.

DR demonstrated most notably high proportions of *-happiness*, *-capacity,* and *+normality.* The high use of *+normality* points to a strong desire to live a 'normal' life, possibly as a way to keep his self-proclaimed 'dark side'—over which he claims to have had no control—at bay. The tokens of *-happiness* then marked instances of distress caused by either a real or perceived threat to some aspect of this coping mechanism. The focus on creating a 'normal' life could also be indicative of preoccupation with how others perceive him and when combined with the self-directed judgments of *-capacity,* suggest deflated self-esteem. The *obsession* feature of obsessive-compulsive disorder (OCD) could help explain the combination of *+normality* and *-happiness. Obsessions* are "recurrent thoughts, urges, or images" that, as a result of being "intrusive and unwanted", cause distress which an individual may try to ignore or cope with using other thoughts or actions (Association 2013: 237). For DR, the obsessive thoughts appeared to revolve around his 'dark side', and creating a 'normal' life for himself acted as the coping mechanism. Aspects of narcissistic personality disorder (NPD) relating to the import of others' perceptions and the effect of self-esteem on emotional regulation (Association 2013) might also help explain the use of all three *attitude* types. That is, his apparent low self-esteem (self-directed *-capacity*) and desire for attention and approval from others

(*+normality*) could have resulted in poor emotional regulation, which could then help explain the distress in response to interpersonal stresses that were often conveyed with *-happiness.*

Finally, IB's high proportion of *-propriety* and overall *positive attitude* presented an interesting pattern. While he did use *-propriety* in reference to his own actions, these were rare and typically accompanied by some positive token, indicating a positive view of his 'improper' actions. This pattern did not hold for other-directed tokens of *-propriety*, which were used to make blanket evaluations of the ethical and moral makeup of vague other persons/groups or society as a whole. The *grandiosity* trait of NPD could well help explain this pattern. *Grandiosity* is marked by condescension toward others stemming from a firm belief of superiority over them (Association 2013). The condescension is evident in the incongruity between the positive view of his personal *impropriety* and the harsh blanket judgments of the morality of others.

While there were some observable connections between patterns of *attitude* and psychological traits for each author, the small size of the dataset limits the generalizability of the findings. However, the fact that there were interpretable connections between the psychological traits and the appraisal coding at all is promising and suggests that using Appraisal to analyze forensic texts for evidence of psychological traits is worthy of further exploration.

Conclusion

This paper has attempted to explore and demonstrate the value of Appraisal as an analytical tool in assessments of forensic texts. It provides a way of tracking the different linguistic resources that are used to express *stances*, resources which correspond to aspects of experience that are impacted by psychopathology. While the sample size was small, there were observable connections between some of the psychological traits of the authors and patterns of *attitude.* These promising results provide support for the merit of the approach presented in this paper. Of course, further research will be needed to better understand the connections between evaluative resources and psychopathology as well as to improve the reliability and efficiency of the analytical approach. Ideally, this future research will be better able to control for the confounding factors which could not be controlled for in this study such as audience and mode of communication. Additionally, being able to compare patterns of violent offenders—like those in this paper—to non-violent counterparts with similar psychopathological profiles would help to determine if any of the patterns are reasonably attributable to a higher propensity for violence.

References

Abrahmsen, D. (1979). Unmasking son of Sam's demons. *New York Times.*

Association, A. P. (2000). *Diagnostic and statistical manual of mental disorders (DSM-IV-TR).* American Psychiatric Publishing, 4th ed. Text Revision.

Association, A. P. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5).* American Psychiatric Publishing, 5th ed.

Bell, A. (1984). Language style as audience design. *Language in Society*, 13(2), 145–204.

Bhatia, V. (1993). *Analysing genre: Language use in professional settings.* Routledge.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999). *Longman grammar of spoken and written English.* Pearson Education Limited, 1st ed.

Bortolan, A. (2019). Phenomenological psychopathology and autobiography. In G. Stanghellini, M. Broome, A. Fernandez, P. Fusar-Poli, A. Raballo and R. Rosfort, Eds., *The Oxford handbook of phenomenological psychopathology*. Oxford University Press, 1053–1064.

Brady, I. (2001). *The gates of Janus: Serial killing and its analysis by the Moors Murderer Ian Brady*. Feral House.

Buck, B. and Penn, D. (2015). Lexical characteristics of emotional narratives in schizophrenia: Relationships with symptoms, functioning, and social cognition. *The Journal of Nervous and Mental Disease*, 203(9), 702–708.

Butler, A., Brown, G., Beck, A. and Grisham, J. (2002). Assessment of dysfunctional beliefs in borderline personality disorder. *Behaviour Research and Therapy*, 40(10), 1231–1240.

Docherty, N., Derosa, M. and Andreasen, N. (1996). Communication disturbances in schizophrenia and mania. *Archives of General Psychiatry*, 53(4), 358. https://doi.org/10.1001/archpsyc.1996.01830040094014.

Edwards, D. and Potter, J. (1992). *Discursive psychology*. Sage Publications.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics*. SAGE, 5th ed.

Fine, J. (2006). *Language in psychiatry: A handbook of clinical practice*. Equinox Publishing Ltd.

Gales, T. (2010). *Ideologies of violence: A corpus and discourse analytic approach to stance in threatening communications (Publication)*. University of California-Davis]. ProQuest Dissertations Publishing.

Gales, T. (2015). Threatening stances: A corpus analysis of realized vs. non-realized threats. *Language and Law / Linguagem e Direito*, 2(2), 1–25.

Gales, T. and Solan, L. (2017). Witness cross-examinations in non-stranger assault crimes: An appraisal analysis. *Language and Law / Linguagem e Direito*, 4(2), 108–139.

Gawda, B. (2013). The emotional lexicon of individuals diagnosed with antisocial personality disorder. *Journal of Psycholinguistic Research*, 42(6), 571–580.

Grant, T. (2008). Approaching questions in forensic authorship analysis. *Dimensions of Forensic Linguistics*, 5, 215–229.

Halliday, M. and Matthiessen, C. (2014). Halliday's introduction to functional grammar.

Hurt, M. (2020). *Pledging to harm: A linguistic analysis of violent intent in threatening language [Unpublished doctoral dissertation]*. Aston University.

Hurt, M. and Grant, T. (2019). Pledging to harm: A linguistic appraisal analysis of judgment comparing realized and non-realized violent fantasies. *Discourse Society*, 30(2), 154–171.

Martin, J. (1997). Analysing genre: Functional parameters. In F. Christie and J. Martin, Eds., *Genre and institutions: Social processes in the workplace and school*. Cassell, 1–39.

Martin, J. and White, P. (2005). *The Language of evaluation: Appraisal in English*. Palgrave MacMillan.

Millon, T., Millon, C., Meagher, S., Grossman, S. and Ramnath, R. (2012). *Personality disorders in modern life*. John Wiley Sons, 2nd ed.

Moran, D. (2019). The phenomenological approach. In G. Stanghellini, M. Broome, A. Fernandez, P. Fusar-Poli, A. Raballo and R. Rosfort, Eds., *The Oxford handbook of phenomenological psychopathology*. Oxford University Press, 1st ed., 205–215.

New York Magazine, (2006). The letters of son of Sam. https://nymag.com/news/crimelaw/sonofsam/20333/.

Ott, T. (2021). Btk killer: A timeline of his murders, reappearance and capture - biography. https://www.biography.com/news/btk-killer-dennis-rader-timeline.

O'Donnell, M. (2019). *UAM Corpus Tool (Version 3.3v)*. http://www.corpustool.com/: WagSoft Systems.

Ramsland, K. (2016). *Confession of a serial killer: the untold story of Dennis Rader, the BTK killer*. University Press of New England.

Rochester, S. and Martin, J. (1979). *Crazy talk: A study of the discourse of schizophrenic speakers*. Plenum Press.

Shuy, R. (1998). *Language of confession, interrogation, and deception*. Sage Publications.

Slevin, P. (2004). Wichita's serial killer reemerges, and so does the fear.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Tausczik, Y. and Pennebaker, J. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54.

Wuornos, A. (2011). *Dear Dawn: Aileen Wuornos in her own words (L)*. Soft Skull Press.

Wuornos, A. and Berry-Dee, C. (2006). *Monster: My true story*. John Blake Publishing.

The importance of being heard: Stories of unrepresented litigants in small claims cases and private family proceedings

Tatiana Grieshofer (previously Tkacukova)

Birmingham Institute of Media and English, Birmingham City University, UK

https://doi.org/10.21747/21833745/lanlaw/9_1a4

Abstract. *The article explores narrativisation practices in small claims cases and private family proceedings, focusing predominantly on cases where at least one of the parties is not represented by a lawyer. By drawing on the data collected during court observations and analysed using the ethnography of communication as the main methodological framework, the study identifies narrative genres across different stages of legal proceedings and illustrates communication barriers experienced by lay court users. The discussion focuses on how formalised narrative genres and the staggered presentation of narratives impact the degree to which court users can use their voice. The article also links the notion of voice projection to procedural justice and suggests that the main narratives should be elicited sooner as part of an open narrative strategy to ensure the court users' voices are heard by the judiciary in the initial stages of the proceedings.*

Keywords: *Legal-lay communication, Narrativisation, Voice projection, Procedural justice, Civil and family proceedings.*

Resumo. *Este artigo aborda as práticas de narrativização em casos de pequenos litígios e processos de família privados, concentrando-se predominantemente em casos nos quais pelo menos uma das partes não é representada por um advogado. O estudo baseia-se nos dados recolhidos durante as observações judiciais e analisados utilizando a etnografia da comunicação como principal enquadramento metodológico para identificar géneros narrativos em diferentes fases do processo judicial e revelar as barreiras impostas à comunicação vivenciadas por leigos no sistema judicial. A discussão centra-se na forma como os géneros narrativos formalizados e a apresentação escalonada das narrativas influenciam a forma como esses leigos podem usar a sua voz em tribunal. O artigo também estabelece a ligação entre o conceito de projeção de voz e a justiça processual, evidenciando que as narrativas principais devem ser invocadas mais cedo como parte de uma estratégia narrativa aberta para garantir que as vozes*

dos atores em tribunal sejam ouvidas pelo sistema judicial nas fases iniciais do processo.

Palavras-chave: *Comunicação leigos-juristas, Narrativização, Projeção de voz, Justiça processual, Processos cíveis e de família.*

Introduction

The exploration of legal-lay discourse has always been at the centre of research into courtroom discourse and, more broadly, spoken interaction in legal contexts (e.g. Heffer 2013). Yet, the most challenging settings for legal-lay communication, i.e. when lay people represent themselves in legal proceedings, remain largely unexplored. In such settings, the differences in institutional powers and recourse to linguistic resources among the legal and lay participants are unequal by default, yet lay people have to perform the role of lawyers. The article focuses on cases where one or neither of the parties is represented by a lawyers and court users have to act in their own behalf. Self-representation is a frequent phenomenon in common law jurisdictions (e.g. Trinder *et al.* 2014; McKeever *et al.* 2018; MacFarlane 2013), with most self-represented litigants often concentrated in small claims cases and private family proceedings as these types of cases are most common and for lay people possibly more manageable than other types of civil proceedings (Trinder *et al.* 2014; Lee and Tkacukova 2017).

Semi-represented and fully unrepresented cases (i.e. cases where one of the parties or neither of the parties is represented) create the conditions in which the discrepancies between legal and lay discourse types are most apparent; establishing effective communication is thus key for ensuring procedural justice and judicial efficiency. A crucial part of legal-lay communication is formed through narrativisation practices embedded in relevant legal proceedings, i.e. the processes of eliciting and presenting narratives. Narrativisation has previously been explored predominantly in criminal contexts, in which lay participants are restricted to the position of active recipients of legal discourse responding to questions related to legal principles (e.g. narrativisation as part of witness examination or trial discourse in Cotterill (2003) or Heffer (2005)). The role of self-represented litigants is, however, much more complex as they have to construct their narratives and ensure that different reiterations of these narratives retain legal coherence (Tkacukova 2016).

The exploration of stories told by self-represented litigants in non-criminal settings has so far focused on the discrepancy between deductive narratives used in US small claims cases and the chronologically organised and overly emotional inductive narratives of self-represented litigants (O'Barr and Conley 1991); the disparity between the legal and lay narrativisation styles leads to delays, misunderstandings and the overall dissatisfaction of self-represented litigants with the court system. Furthermore, the litigants using a powerless speech style (with hedging, hesitations, intensifiers, questioning intonation), as opposed to the powerful speech style used by lawyers, were found to be less successful in their claims (O'Barr 1982). This raises concerns about access to justice for the most vulnerable in the society, especially given the rising numbers of lay court users across different jurisdictions (e.g. Trinder *et al.* 2014; McKeever *et al.* 2018; MacFarlane 2013).

In England and Wales, the number of litigants in person (LIPs), the terms used for self-represented litigants, has risen dramatically since the introduction of cuts to legal

aid by the Legal Aid, Sentencing and Punishment of Offenders Act 2012. In private family proceedings, the number of hearings with at least one unrepresented party has increased from approximately 55% in 2012 to approximately 80% since 2015 (*Family Court Statistics Quarterly: April to June (Ministry of Justice 2021a)*). The judiciary are thus often presiding over cases where neither of the parties is represented, despite the fact that the court processes were designed by lawyers and for lawyers. Some improvements are gradually introduced through the current HMCTS (Her Majesty's Courts and Tribunals Service) reform programme, which aims to enhance the efficiency of courts and increase the accessibility of the legal system. As part of the programme, the user experience is recognised as one of the key aspects of the redesigned system and a considerable part of the digitisation innovations (*HMCTS Reform Update Summer 2019*[1]). Although communication with court users and LIPs is already viewed as an important area which requires further improvements, it is increasingly becoming clear that legal processes and procedures would also benefit from a structural redesign which would put court users at the core of the system (e.g. see suggestions in Hunter *et al.* (2020)). The article argues that narrativisation practices (how and when narratives are elicited and presented) and voice projection opportunities (how and when court users can voice their claims) should be considered as a key part of the potential redesign of the legal system, especially in the types of cases with a markedly high concentration of LIPs, such as private family proceedings.

The study presented here is part of the wider project on communication in legal proceedings with LIPs[2]. The theme of story-telling and voice projection as a key challenge interweaves several aspects of legal-lay discourse, explored through court observations, interviews with lawyers and LIPs and textual data (Grieshofer et al., 2021), but the main focus here lies on court observations as they allow for a more in-depth investigation of voice projection and narrativisation practices (see methodology section). The originality of the article lies in its focus on previously unexplored aspects of legal-lay communication, such as (1) investigating narrativisation practices embedded in civil and private family proceedings; (2) incorporating pre-court stages into the exploration of narrativisation; (3) aligning narrative genres to communicative goals of individual hearings; (4) examining voice projection opportunities for self-represented litigants; and (5) establishing connection between voice projection and procedural justice. Although the main focus is on cases with self-represented litigants, the discussion is equally relevant to fully represented cases as represented clients also need to actively engage with legal proceedings by providing evidence and narrating their stories throughout different stages of the proceedings.

Linguistic barriers to justice for LIPs

Research in socio-legal studies on self-representation has shown that irrespective of their educational or socio-economic background, many LIPs are vulnerable due to the stressful nature of court proceedings and clustering of legal problems with additional financial or health-related difficulties (Trinder *et al.* 2014; Pleasence and Balmer 2019). Furthermore, the lack of accessible advice and information complicates LIPs' understanding of court processes and procedures as many litigants cannot find reliable sources relevant to their case or do not know how to search for information (Lee and Tkacukova 2017). What helps move LIPs' cases forward is an active engagement of a legal professional (usually a judge or, possibly, a legal representative for the other party, a legal representative for

the child or a social worker or another expert involved in the case) as they guide lay court users through the process (Trinder *et al.* 2014: sec 4.4). Apart from the crucial support from a legal professional, the workable hearings tend to have the following characteristics: relate to less complex cases; be at the stage of the proceedings which requires less LIP participation (e.g. directions hearings in the initial stages of the proceedings require less input from court users than a substantive hearing); involve the settlement-oriented and confident LIP who is prepared for the hearing (ibid). Although these are specific factors which can help make hearings more efficient, in broader terms LIPs commonly face intellectual, practical, emotional and attitudinal barriers when accessing the justice system (McKeever *et al.* 2018). Many of these barriers are rooted in the complexity of legal discourse and relate to crucial procedural and communicative stages, such as understanding legal texts, identifying specific legal problems pertinent to relevant legal principles, and, finally, communicating the case following the discursive principles used in the discursive community of legal professionals (Tkacukova 2016).

Even the possibility of full comprehension of legal discourse by the lay court user has been shown to be unrealistic by research in linguistics, applied psychology and law (Assy 2011; Azuelos-Atias 2011; Grieshofer *et al.* 2021; Greene *et al.* 2012; Hiltunen 2012; Masson and Waldron 1994; Mindlin 2005; Pavlenko *et al.* 2019; Ződi 2019). Although the principles of plain language movement are efficient when dealing with some lexical and grammatical complexity (Adler 2012), there are important challenges to simplifying legal texts: firstly, legal texts express complex realities and thus need to rely on some linguistic complexity as explicit expression of legal scope and legislative intention reduces the transparency of the texts (Bhatia and Bhatia 2011); and, secondly, the implicit meaning of procedural and conceptual aspects embedded in the legal system may not always be apparent to lay users even if expressed in simple terms (Assy 2011). For instance, Yeung and Leung (2019) argue that legal homonyms (same words with different legal and ordinary meanings) or even legal terms with phonetic resemblance to common use words can hindering unrepresented litigants' understanding of legal texts.

Beyond the psycholinguistic barriers of processing and comprehending legal texts, many LIPs also struggle with discursive competence (Tkacukova 2016). According to Bhatia (2004: 144), discursive competence in specific institutional settings involves three types of competences: textual, generic and social. As discussed above, the textual competence is hindered by the linguistic complexities of legal discourse as LIPs tend to struggle with the linguistic competence (use of specialised language) and communicative competence (interpretation and production of contextually relevant and legally coherent narrative genres). The degree to which LIPs display textual competence is limited not only due to the restricted comprehensibility of legal texts but also their generally restricted knowledge of law or wider experience with legal texts. The generic competence (the ability to effectively participate in communicative professional practices) and social competence (the ability to communicate effectively by using the linguistic resources appropriate to the institutional role) are closely related to court procedures. It is overcoming challenges with generic and social competence that legal professionals can help LIPs with. As recorded in previous research, LIPs' active participation in the proceedings can be supported through effective elicitation and communication strategies (see Trinder *et al.* (2014) for support offered by legal professionals to LIPs; Tkacukova (2015) for explanations and guidance offered by the

judiciary to LIPs; Tkacukova (2016) for the impact of power relations on LIPs' ability to self-represent). Narrativisation practices and voice projection play an important part in creating the space for lay court users to tell their stories, drawing on varying degrees of discursive competences.

Data and methods

The article draws on 40 court observations of private family hearings and small claims hearings: 10 small claims hearings and 30 private family law hearings, of which 21 were related to child arrangements, five to financial dispute resolution issues, and four to non-molestation orders. In order to examine the narrative development during pre-court preparations and in-court interactions and explore communication goals of different stages of court proceedings, the study presented here also draws on additional data sets collected as part of the wider project on linguistic aspects of access to justice for LIPs: textual data (court forms and guidance documents) and empirical data (questionnaires, interviews and court observations) from all key stages of legal proceedings in civil and private family law contexts in England and Wales. The exploration of diverse datasets from the perspective of communicative challenges, discursive practices and the distribution of linguistic agency among the trial participants has led to the enquiry into the centrality of voice in court processes and its link to procedural justice. The analysis presented here explores the overarching theme of narrativisation and voice projection as interlinked with procedural steps which take place during pre-hearing stages and court hearings. The article thus makes a key contribution to understanding narrativisation embedded in court processes and procedures by introducing theoretical frameworks for analysing narrativisation practices and voice projection, which can be further built on through empirical and experimental investigations in future research.

Given the main data draws on court observations, the methodological approach adopted in the study builds on the ethnography of communication, which enables to accomplish a dual objective: explore the theoretical principles of narrativisation in legal proceedings and at the same time investigate practical aspects of communication and narrativisation practices in context (Hymes 1962; Carbaugh 1989). The data interpretation builds on previous narrativisation frameworks adapted for forensic contexts from everyday naratives (Heffer 2005, 2018; Cotterill 2003; Gibbons 2003); as shown in the following section, these approaches are adapted to reflect the communicative complexities inherent in civil and family prceedings. The notion of voice projection is construed as part of socio-cultural and institutional practice (Heffer 2013, 2018) and draws on Hymes (1996) link between voice and an opportunity to use language, with the caveat that some voices are "acceptable, even valued, in certain roles, but not others" (70) and that the realisation of the voice is "partly at the mercy of others" (xi). The current study expands the understanding of narrativisation practices and voice projection by establishing a link between pre-court and in-court narrative genres and a further link between voice projection and procedural justice.

There are some disadvantages in relying on court observations as the main data source. The observations of court hearings depended on the availability of hearings with LIPs on the days the author attended court and explicit consent from the judiciary and the parties concerned. It was also not possible to gain access to the recordings or even the transcripts of the hearings observed due to the current policy of the Data Access Panel of the HMCTS, which does not allow the release of recordings of hearings

for research purposes. Even the parties can only gain access to verbatim transcripts, rather than the recordings of their hearings. Furthermore, gaining access to transcripts is only possible through purchasing court-approved transcription services. Given that the quality of transcripts is generally insufficient for the linguistic analysis (Walker 1986, 1990; Fraser 2003; Eades 1996), the high costs associated with obtaining the verbatim transcripts create an additional obstacle. The accessibility of the Ministry of Justice (MoJ) and HMCTS data for research purposes, alongside the practices of gathering the data for internal research, have been recently criticised, amongst other reasons, for hindering the collaboration between the academia and justice institutions (Byrom 2019). The understanding of the data requirements for different methodological frameworks (i.e. the importance of accessing audio files for linguistic purposes) should be another essential aspect for the internal data policy of HMCTS and MoJ.

The above-mentioned weaknesses are counterbalanced by the advantages that ethnographic research frameworks employed here offer: the detailed analysis of the participants' linguistic behaviour and mutual interactions and attitudes in the analysed settings; inductive investigation of social and linguistic patterns recurring during diverse interactions; exploration of the data collected in the most naturalistic and realistic environment without artificially pre-defined criteria (Ejimabo 2015; Sangasubana 2011). The observation notes focused on the role of hearing participants and the type of hearing, the main topics discussed, the framing of the topics by the parties, interaction patterns, legal/procedural explanations presented by the judiciary and any arguments presented by the parties. Whenever possible, the notes were verbatim to capture the framing of questions or important arguments (i.e. when non-confidential information was discussed and the speed of speech allowed for a verbatim transcription). The methodological approach adopted here allows to explore current cases common in district courts across the country and provide a representative overview of the most frequent challenges experienced by LIPs and coping strategies employed during legal-lay interactions (rather than searching for singular cases with available transcripts). Exploring narrativisation and voice projection throughout different stages of legal proceedings allows the study to reflect on systemic issues within the legal system and explore the potential role applied linguistics research can have on the justice system.

Narrativisation practices in civil and private family proceedings

At the core of the adversarial legal system is the battle of narratives. As shown in research on criminal law, narrativisation in courtroom settings is characterised by fragmentation, deconstruction and re-interpretation (Cotterill 2003; Harris 2001, 2005; Heffer 2005). In civil and family legal settings, the narratives are equally fragmented, but it is mostly court users, whether represented or not, who are responsible for providing evidence for their stories and constructing micro, or satellite, narratives (see Snedaker, 1991 quoted in Gibbons (2003), p. 155), such as witness statements or responses during witness examination. Interestingly, private family proceedings (and to some extent also civil proceedings) rarely reach the stage of the final hearings as there is strong emphasis on the parties settling their cases (Trinder *et al.* 2014). What is crucial for the narrativisation in civil and family settings is the pre-hearing stages as these are the evidentiary stages during which the main evidence is collected. It is also during these pre-hearing stages that LIPs experience difficulties with constructing their cases due to

practical obstacles and/or lack of comprehension or discursive competence (Grieshofer *et al.* 2021; Tkacukova 2016, 2020; Trinder *et al.* 2014).

Yet, most of the linguistic research has so far focused on narrativisation during court hearings (e.g. Cotterill 2003; Heffer 2005); this is mainly due to the focus on the narratives that emerge during criminal proceedings, which start at the point when most of the evidence is already gathered and the evidence is then put on trial. During criminal cases, stories are thus narrated through the narrative and question/answer discourse types (Harris 2005: 220-221; Cotterill 2003; Coulthard and Johnson 2007: 97). Previous research has mapped out Labov's narrative structure (1972) onto the narrativisation principles within the trials with *abstract* and *orientation* represented in opening statements, *complicating action* in witness examination, *evaluation* in closing arguments, and *resolution* and *code* in verdict and sentencing/release respectively (Cotterill 2003: 24) with *orientation* and *evaluation* being the focal points of the narration (Heffer 2005; Harris 2005). But this structure is not immediately applicable to civil and family proceedings as each hearing in these settings has a specific communicative aim and includes an *abstract* and *orientation* (see Figure 1) whereas *complicating action* or *evaluation* are often introduced throughout the proceedings (e.g. expert reports from social services as part of pre-court investigations).

To reflect on how narratives are constructed in different jurisdictions, it is important to explore discursive practices embedded in the construction of narratives and view narration practices as part of the socio-cultural context (Heffer 2018: 258). Heffer (2013 & 2018: 265) proposes the Narrative Navigation model which illustrates how institutional practices used in forensic contexts relate stories to the relevant audiences within the pre-defined discursive constraints in the institutional context. The model aligns trial genres to embedded narratives and narrative focus, alongside the mode of narration and type of narrator. The trial genres are, however, limited to the oral genres. Within the context of civil and family law hearings, the genres are, however, much more diverse and permeate between written and spoken modes. Given the need to link witness testimony to the written evidence, the principle of orality is weakened in civil and family hearings (cf Hrabovska *et al.* (2021)), which means that it is not sufficient to consider only narrative genres embedded within the hearings and it is necessary to explore pre-court narrative genres.

A related construct, the conceptualisation of legal genres, has so far also been explored predominantly through the lenses of criminal law and the succession of genres within criminal court proceedings or, alternatively, through their link to criminal court proceedings (e.g. Heffer 2005: 67; Gibbons 2003: 132-133). Gibbons (2003), for instance, provides a detailed summary of the dynamic and codified genres involved in trials, including the pre-trial stages; but these pre-trial stages are either characterised by their dynamic nature evolving from legal-lay interaction (e.g. police interviews) or include codified genres which are used for information or as a point of reference (e.g. a will, legislative text). The genres embedded within civil or family proceedings, especially the pre-hearing stages, do not fit within these boundaries (Figure 1).

This study explores narrativisation by, firstly, identifying narrative genres involved in the construction of satellite narratives in pre-court and court stages equally, and, secondly, by exploring the limitations these genres put on the court user's voice projection. Since narratives arise within specific socio-cultural and institutional

constraints, the genre approach to narrativisation is useful for exploring the complexity within related taxonomies of genres or 'genres within genres' (Hyvärinen 2015: 190; Gibbons 2003: 131), especially when narrative genres are aligned with the communicative aims of procedural stages. Drawing on the terminology and concepts established by Heffer (2013, 2018) and Gibbons (2013), Figure 1 shows the diversity of narrative genres used throughout all stages of the proceedings. The focus of Figure 1 is on child arrangements proceedings as these are representative of the most common cases in which lay court users are likely to participate due to the following reasons: the high frequency of child arrangements cases in district courts; the highest concentration of LIPs in these types of cases; and the wide scope for narrativisation due to the personal nature of the cases; court processes representative of other civil or family proceedings (see Tkacukova (2016) for narrativisation in financial remedy proceedings). For illustrative purposes, Figure 1 presents a simplified version of the child arrangements proceedings; many cases require several interim hearings or are disposed of before reaching the final hearing stage (cf flowcharts in the *Guide to Family Court Statistics* (Ministry of Justice 2021b)). Similarly, for brevity, Figure 1 refers to judges presiding over hearings, though it is important to note that child arrangements cases are heard by either magistrates, supported by a legal adviser, or a district judge (if there are any safeguarding concerns).

There are several key characteristics of narrativisation in private family proceedings: (1) the limited number of opportunities for direct narration in the initial stages; (2) limited opportunities for Respondents to provide direct narration without the pre-defined narrative framing from the Applicant; (3) presentation of the initial information through codified and fragmented narrative genres (e.g. court forms); (4) prevalence of procedural genres; (5) presence of genres leading to the adjudicative stage even in pre-court stages (e.g. CAFCASS[3] report); (6) significance of expert-mediated and expert-framed narration; (7) reduced opportunities for an input from the lawyer (even for represented parties). The combination of these complex factors goes beyond creating a narrative disjunction (Coulthard and Johnson, 2007: 111): the process requires lay court users to engage with codified and procedural genres (witness statements, skeleton arguments, court forms) without much information or support. The guidance embedded in court forms, for instance, often lacks clear explanation of court processes or definitions of relevant concepts (Grieshofer *et al.* 2021), which leads to court users searching for more user-friendly, yet potentially biased and inaccurate, advice on social media (Tkacukova 2020).

Furthermore, when constructing satellite narratives through the use of codified written genres, LIPs need to ensure that they meet procedural, discursive and legal criteria in terms of evidence presentation. The genres through which such evidence is elicited are, however, not conducive to the storytelling practices common in everyday situations. The closest genre to storytelling is witness statements filed by the parties, though this mainly applies to applicants; respondents file their statements in response to the applicants' statements, which pre-determines the topics they need to address. Despite the narrative-like qualities of witness statements, they incorporate complex discursive tasks and need to comply with legal rules and directions (Cooper and Mattison 2021) as well as present all the necessary information in an accurate and coherent manner; the quality of witness statements varies even among legal professionals, so

Narrativisation boundaries	Pre-hearing stages		Court hearings	
Court procedures (CPR/FPR), legal framework, discursive practices, justice system narrative (Evidential stage → Submissions)	**Narrative genres**	**Narration**	**Narrative genres**	**Narration**
	Communicative goal: initiating proceedings		**First Hearing Dispute Resolution Appointment (FHDRA)** **Communicative goal:** case management, identifying issues	
	Court forms (procedural, adversarial); Risk identification interview with a Children and Family Court Advisory and Support Service (CAFCASS) worker (procedural, adversarial); CAFCASS safeguarding report (procedural, adjudicative elements)	Codified and Direct (parties or lawyer); Expert-mediated (CAFCASS worker, parties); Expert-framed (CAFCASS)	Out-of-court negotiations; Presenting the case/position (adversarial); Case management and narrowing down issues (procedural, adversarial, adjudicative); Directions, interim court order or consent order (procedural, adjudicative)	Direct (parties or lawyers); Direct (parties or lawyers); Judge-mediated (judge & parties or their lawyers); Directive for narrative scope (judiciary)
	Communicative goal: preparing evidence		**Directions/Dispute Resolution Appointment** **Communicative goal:** case management, narrowing down issues	
	Interviews for the section 7 CAFCASS report (procedural, adversarial); CAFCASS section 7 report (procedural, adjudicative elements).	Expert-framed (CAFCASS worker, parties); Expert-framed (CAFCASS)	Out-of-court negotiations; Presenting the case/position (adversarial); Case management and narrowing down issues (procedural, adversarial, adjudicative); Directions, interim court order or consent order (procedural, adjudicative)	Direct (parties or their lawyers); Direct (parties or their lawyers); Judge-mediated (judge & parties or their lawyers); Directive for narrative scope (judge)
	Communicative goal: preparing further evidence		**Fact finding hearing/Interim hearings (in case of domestic violence allegations)** **Communicative goal:** consider the evidence around domestic abuse allegations	
	Further evidence, statement of facts/issues, witness statements/skeleton arguments (procedural, adversarial); Scotts Schedule, i.e. numbered list of allegations and responses to these (adversarial); Court bundle (procedural, adversarial).	Direct, antagonistic (Applicant, then Respondent); Direct, antagonistic (Applicant, then Respondent); All.	Presenting the case/position (adversarial); Case management and narrowing down issues (procedural, adversarial, adjudicative); Witness examination (adversarial); Decision as to allegations (adjudicative, procedural)	Direct (parties or their lawyers); Judge-mediated (judge & parties or their lawyers); Direct/lawyer-mediated and antagonistic (parties, lawyers, judiciary); Directive for narrative scope (judge)
	Communicative goal: finalising court bundle		**Final hearing** **Communicative goal:** consider all evidence and make the final decision	
	Court bundle (procedural, adversarial)	Direct, expert-framed	Presenting the case/position (adversarial) Case management and narrowing down issues (procedural, adversarial, adjudicative); Witness examination (adversarial) Child Arrangements Order (adjudicative)	Direct (parties or their lawyers) Judge-mediated (judge & parties or their lawyers); Direct/lawyer-mediated and antagonistic (parties, lawyers, judge) Directive for post-proceedings stage (judge)

Figure 1. Narrativisation in child arrangements cases (according to 'Practice Direction 12b – Child Arrangements Programme' (Ministry of Justice 2021c))

it is not realistic to expect LIPs to conform to all the norms. Although there is some guidance available on how to prepare witness statements (e.g. templates given out in court), this support tends to be limited to formal characteristics of witness statements as a genre (the header, paragraphs, statement of truth) and many LIPs struggle with the identification of relevant content (Trinder *et al.* 2014).

In addition to presenting the main opportunity to construct the narrative, witness statements also define the interpretative scope of disputed issues, ensure judicial efficiency and impact the case outcome (Cooper and Mattison 2021) . The centrality of witness statements in civil and family hearings leads to the crucial evidence being provided in a non-interactive manner. Yet, it is the interaction with legal professionals or other experienced experts (e.g. CAFCASS officers) that can help LIPs navigate the proceedings and provide relevant information (Trinder *et al.* 2014). An important narrativisation thread is thus created via expert-framed (e.g. CAFCASS investigations) or judge-mediated genres. Despite the potential guidance these interactions provide, it is important to note that they are imbalanced in terms of power relations and driven by such factors as the specific framing of questions, the choice of topics and control over turn-taking with typically minimal opportunities for court users to introduce new topics (Thornborrow 2002). Another issue with expert-framed genres is that they may introduce potential inaccuracies into satellite narratives through expert reports: in six out of 11 observed hearings, where the CAFCASS report was discussed, the reports contained factual errors or misrepresented some information (the concerns about misrepresentations were expressed by the judiciary, the parties or their lawyers).

Expert-framed narratives precede direct narrative opportunities, i.e. mainly witness statements or the initial presentation of the case at the beginning of each hearing. The fact that oral submissions and opportunities for direct narratives (witness statements) occur in the final stages means that in the initial stages LIPs are repeatedly stopped from telling their story due to the procedural steps that need to be taken before oral submissions. The court observation sample included two hearings (out of four hearings in the initial stages of the proceedings), in which LIPs were told to not tell their story as it did not fit with the aim of the hearing. The adversarial approach and current procedures have previously been criticised as inefficient and insufficiently trauma-aware, especially in cases with an element of domestic violence or serious offence (Hunter *et al.* 2020). The future reforms, such as suggested by Hunter *et al.* (2020), should thus consider giving parties an opportunity to present their direct narratives earlier in the proceedings as this can improve the relevance of the elicited information and ensure the appropriate safeguarding and gatekeeping measures are established earlier in the process (Grieshofer, submitted).

Opportunities for Voice projection

Central to narrativisation is the concept of voice projection, which applies to individuals or groups and communities. It is viewed as a discursive and communicative concept related to a discursive style and at the same time the freedom or right to speak (Heffer 2013: 3; Hymes 1996). It is particularly important to explore the link between the degree of freedom to speak and the impact of the act of speaking. In institutional settings, those with the authority to use their voice can expect that their voice would have an impact on the audience even if the message is not completely comprehended; for instance, jury instructions perform a ritualistic function irrespective of whether they are understood

(Heffer *et al.* 2013). Those with restricted rights to use their voice (e.g. witnesses during cross-examination) are more likely to be subjected to the conditions in which their voices are lost, though this can happen even to speakers in powerful institutional roles (Heffer 2019). It is not only the institutional role, but also the discursive competence with which the voice is projected and the degree to which the appropriate discursive norms, expected in the relevant discursive community, were conformed to that determine the potential impact on the audience (Bernstein 1990). This is especially relevant to LIPs, who find themselves in a precarious institutional position as they act in their own behalf and thus fulfil the role of lawyers, yet often lack the discursive competence due to insufficient knowledge and experience; furthermore, they often do not have the same rights as legal professionals (e.g. LIPs may not be in the position to instruct an expert witness or conduct cross-examination).

Exploring linguistic inequality, Blommaert (2008) highlights that the key to being perceived and understood is linked to the discursive competence of the speaker and the authenticity of their voice. So far, the concept of voice in courtroom discourse has been explored primarily in the context of witness examination by combining stylistic features used to establish factual and character credibility with varied degrees of success of 'responsive understanding' (Heffer, 2013 & 2018). The ambiguity of the LIPs' institutional role, alongside reduced discursive competence (Tkacukova 2016), can potentially impact the degree to which their voices are heard. Given the central role of expert-framed and judge-mediated genres in civil and family proceedings, the pre-defined narrativisation boundaries and delayed presentation of direct narratives, it is important to explore the outcome of LIPs' voice projection and different types of agentive support that can help court users project their voice, irrespective of whether they are represented or not (see Figure 2).

In Figure 2, the inner circle represents the situation in which the voice is heard thanks to the procedural and legal relevance of the message. In practice, this means that points raised by the court user impact the hearing or the course of the proceedings: the points were reflected in the (interim) order or directions or were at least discussed in court. The middle circle refers to the scenario in which the voice was acknowledged, but did not elicit a 'responsive understanding' or could not be taken into consideration due to issues with content relevance for the specific hearing or issues with discursive competence. The outer circle relates to situations in which the voice was used, but the projection failed due to procedural or legal irrelevance. And, finally, the space outside the circles represents circumstances in which the opportunity to use the voice is lost: the court user was prevented from exercising their right to speak or did not wish to say anything.

The degree to which the court user's voice is projected can be supported or challenged through the authoritative voices of other participants in the proceedings, namely the judiciary, legal representatives and experts (e.g. CAFCASS workers), as shown through the triangles in Figure 2. Represented clients are more likely to have their story heard thanks to the discursive competence of their lawyers, though there may be a discrepancy between the narrative presented by the lawyers and the narrative their clients would like them to present, especially if clients have unrealistic expectations or misunderstand law or court processes. Experts (e.g. CAFCASS workers) can also project the court user's voice in their reports, though there is a potential for court users' voices

to be misrepresented (see above on the rate of errors in social services' reports) or even lost due to expert investigation practices (Macdonald 2017). What has become evident in court observations is the active role the judiciary take to ensure LIPs contribute to the proceedings: the judiciary either mediate LIPs' voices (rephrasing what LIPs are saying to clarify) or enable them to project their voices (by asking the questions which are pertinent to the case). As the triangular shapes aim to indicate, even with support, it is still challenging for a party to ensure their voice is heard as this involves efficient engagement with the participants who can frame, enable, mediate or represent their voice. The voice projection in Figure 2 is relevant for all the stages of the proceedings, though the degree to which the voice is projected differs at each stage (e.g. the voice projection opportunities at the FHDRA hearings are much more restricted than at the final hearings).

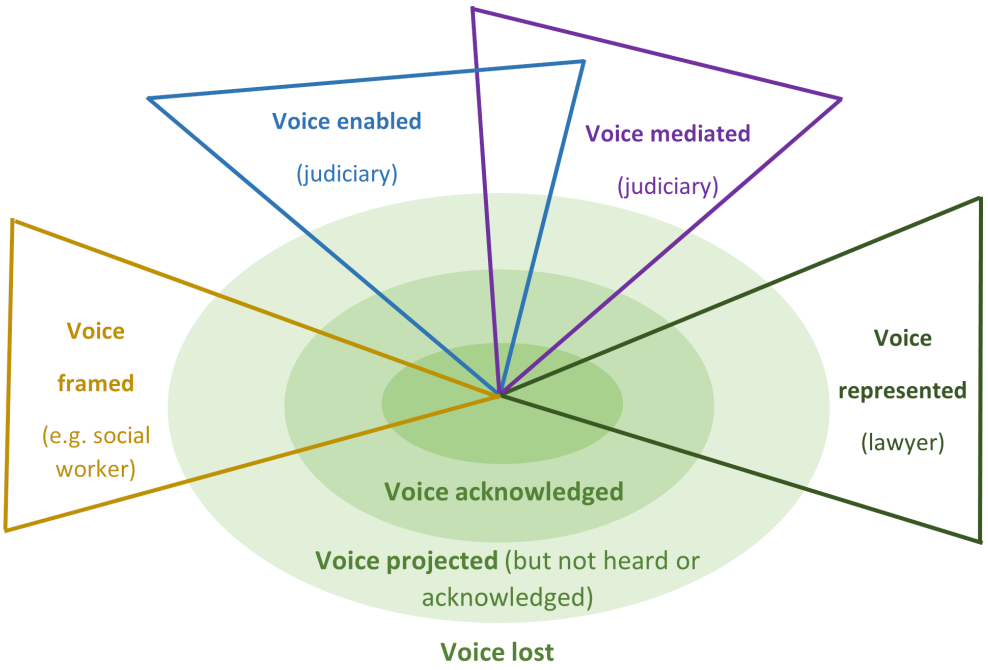

Figure 2. Opportunities and outcome of Voice Projection for represented and unrepresented parties

The court observations showed a multitude of situations in which the voice projection of court users had varied degrees of success. The patterns that emerged and defined the outcome of the voice projection were related not only to procedural and legal relevance but also the 'good character' narrative thread, sometimes even despite the fact that the voice is expert-framed or represented. For instance, in a semi-represented child arrangements case (case 2), the LIP's voice is heard despite the opposing lawyer's arguments against the LIP parent seeing the child (the child is completely non-verbal and possibly autistic, does not respond well to changes and there needs to be a special needs assessment completed before any contact could be resumed). The LIP parent's satellite narrative is presented in a fragmented manner through responses to the lawyer's arguments (the LIP often interrupts the lawyer and asserts their right to speak) and to the judiciary's questions and invitations to speak. Despite the fragmented presentation, the LIP parent manages to make three clearly-defined points by reiterating that they are not

a threat ("I've proved I am not a threat. I want to see my child", "I'm not a risk"); that they want a relationship with the child ("It is frustrating, my family didn't see the daughter for two years", "Even if [the child] is autistic, I'm still a [parent] and want contact"); and by challenging the lawyer's arguments ("It is far-fetched to say [the child] is not comfortable with changes. [The other parent]'s got a new family, partner, the [child] changed friends…"). As a result, the court approved of contact with the LIP parent on the basis that the decision would be reviewed after three initial contact sessions. A number of factors contributed to the LIP's voice being heard: the legal and procedural relevance of arguments made, the assertion of the right to speak, and the pro-contact culture of the family courts as one of the overarching justice system narratives (Hunter *et al.* 2020).

In another example (case 16), a pre-final semi-represented hearing, the LIP parent also asserted their right to speak, but did so by interrupting the lawyer and the judiciary to the extent that they had to be continuously reprimanded for constant interruptions. The LIP's fragmented satellite narrative lacked coherence or clear focus due to constantly shifting topics: irrelevant topics (e.g. difficulty with obtaining legal advice due to costs as even lay advisers quoted £60 per hour, difficulties in the past getting CAFCASS support with child care due to the child's special needs, difficulties due to being carer for the partner with special needs) and relevant ones (the other parent not signing up the child for school in the new place of residence; the other parent using cannabis, possibly while caring for the child; notes from the observed contact sessions not being representative of the sessions). The relative informality in which small claims cases and private family proceedings are held means that judges often talk to parties directly to explore core issues in the case. As a result, much of the communication from lay court users is enabled or mediated by the judiciary. To streamline the discussion in the hearing, the legal adviser and the chair of magistrates mediated the LIP's voice projection (by suggesting how the argument on notes form the contact session can be framed) as well as enabled it on multiple occasions (by eliciting responses on why the LIP thought the child is safe with them and should live with them and suggesting what to include in the witness statement), but the LIP was not able to have their voice heard due to misplaced framing of their arguments (lack of school registration and their experience as carer could be reframed into how they can meet the child's education requirements and special needs more efficiently than the other parent). The LIP's voice was acknowledged in relation to one point in the CAFACSS report, which said that their house is "grimy". When the LIP challenged the CAFCASS worker, who was in attendance in court, they retained that the house "was not dirty, but grimy" and after the discussion of the meaning of the word, the court decided to record the difference in opinion, though acknowledged that the condition of the house did not constitute an issue in respect to the child. Despite multiple attempts to be heard, the LIP's voice remained only acknowledged in the expert-framed narrative and did not impact the court of the proceedings as neither of the changes suggested are recorded in the interim order.

Another semi-represented case (case 11) illustrates a situation, in which the voice was allowed to be projected, but could not be acknowledged due to its irrelevance to the substantive matters in the case. The LIP parent decided to withdraw from the case and not pursue contact with the child. The court enabled the LIP to express how they felt, despite the opposing lawyer's objections and despite the case being effectively closed:

"After two years of not seeing [the child], it is not fair for [the child] or me to see [the child]. CAFCASS is absolute shambles and law is blind. Is it not my right to know where [the child] lives, where [the child] goes to school, how often does [the other parent] travel [abroad]? CAFCASS didn't talk to [the other parent], who knows where [the other parent] was." Despite the substantive irrelevance of the emotional narrative, the court recognises that the LIP has the right to project their voice as part of their access to justice journey as this could help to potentially initiate the emotional recovery (cf Bendall (2020)). Deviating from the institutional norms and practices, which do not welcome overly emotional narratives, this example illustrates that there is space for the therapeutic jurisprudence in private family law proceedings (Lens 2016). Though it is important to note that the comparison of hearings led by magistrates to those presided over by district judges shows a pattern of the hearings before magistrates (and a legal adviser) more likely to create interactional space for emotional accounts, possibly because of the role of magistrates as representatives of the public or because there is more variability among the magistrate members.

What the three examples have in common is the LIPs' tendency to express their emotions, with varying degrees of success. Despite of the association of emotional narratives with powerless speech styles (O'Barr 1982), there is a strong argument to be made in favour of supporting LIPs in expressing their narratives without the normative constraints for two reasons: firstly, dismissing emotions as irrelevant may result in important information being left out, and, secondly, using their voice and expressing their story in an authentic way reinforces LIPs' sense of fairness and procedural justice (Toy-Cronin 2019) while also encouraging them to retain active engagement with the process. Providing space for the LIP from the last example to express their concerns in the initial stages of the proceedings could have helped them to feel heard or acknowledged and thus less emotionally withdrawn from the proceedings. All LIPs in the observation sample expressed their emotions as part of the rationale for engaging with or disengaging from the proceedings or as part of the reasoning for the final outcome or even just an excuse for why a direction could not be followed. The embodiment of emotions in the satellite narratives helped LIPs create an authentic voice, which was not observed in the satellite narratives of represented clients. The topic of the authenticity of the voices in semi-represented cases should be further explored from the point of view or judicial perceptions and the effect of voice projection.

Summary of narrativisation practices and voice projection in the context of procedural justice

Focusing on civil and private family proceedings in which at least one of the parties is not represented allows the study to explore narrativisation in the challenging context for legal-lay interaction. Resorting to court observations as the only data collection option for investigating authentic representations of courtroom discourse within the context of England and Wales, the study draws on the ethnographic approach and reflects on narrative practices embedded in court processes and procedures and explores options for LIPs' voice projection.

The findings drawn from the study show that court users, irrespective of whether they are represented or not, have to engage with multiple procedural steps, provide evidence and construct satellite narratives via codified and direct narrative genres as well as engage in expert-framed investigations. Throughout the proceedings, they have

little control over how satellite narratives would be perceived, interpreted or reported as most narrative genres are expert-framed, judge-mediated or codified. It is only in the final stages that court users have an opportunity to tell their direct narratives, but by that stage the narrative scope is already shaped through procedural stages, legal framework and interim orders or directions made on the basis of expert-framed reports and judge-led case management decisions. Court processes (including Civil Procedure Rules or Family Procedure Rules) thus play a defining role in shaping the narratives, which often contradicts with how lay court users would prefer to tell their story (O'Barr and Conley 1991). The awareness of procedural aspects, such as principles of evidence admissibility (Heffer 2018: 257) or the overview of the succession of procedural stages is as important as understanding relevant legal principles. In fact, prior experience with court procedure helps repeat LIPs to represent themselves more efficiently (Trinder *et al.* 2014: 83) and experienced lay advisers have been shown to support lay court users as effectively as lawyers by focusing on the provision of procedural advice (Sandefur 2015) and supporting LIPs with framing their narratives (Tkacukova 2020). Furthermore, the overarching justice system narrative (e.g. the pro-contact culture of child arrangements hearings) also plays an important role in defining narrativisation boundaries and the impact of voice projection, but can be difficult to engage with for anyone who has limited procedural awareness.

For unrepresented litigants, engaging with the narrativisation practices is further complicated due to very little information available on procedural steps and reduced comprehensibility of the guidance documents or explanations of legal principles embedded in the court application process (Grieshofer *et al.* 2021). Unofficial online resources are often too generic and their reliability and accuracy can be difficult to establish for lay people (Tkacukova 2020). Problematic access to support has repercussions for LIPs' development of a clear narrativisation strategy in the pre-court and even court stages, though the judge-mediated genres are useful for eliciting relevant narrativisation threads during court hearings. The current positioning of the main direct narratives towards the final stages complicates the LIPs' narrativisation journey as they have to overcome several procedural stages to gain that opportunity to project their voice. Eliciting direct narratives earlier in the proceedings would allow court users to take advantage of the guidance offered by legal professionals and develop their satellite narratives in response to their arguments and elicitation strategies. Shifting direct narratives towards the initial stages of the proceedings would also minimise the risk of LIPs' narratives being misconstrued through procedural stages and expert-framed narrative genres and possibly encourage court users to keep engaging with the proceedings.

The study argues that it is important to explore language use not only during individual stages, but also investigate the overarching communication processes which create narrativisation practices throughout the entirety of legal proceedings. What is equally important is that language and communication play a crucial role in the execution and perception of justice. Court users are more likely to accept decisions reached by following fair decision-making processes and in which they had an opportunity to participate (Tyler 2000), i.e. they had their voice heard. According to Sela (2018), the perception of procedural justice incorporates four principles: process control (relevant for the stage of presenting the evidence); decision control (with respect

to the choice of the final outcome); interactional justice (encompasses fair treatment with politeness, dignity, respect); and informational justice (incorporates sufficient information about the process and its justification). The four principles rely on efficient language use for reaching a specific communicative aim in the institutional context in which speakers do not have the equal distribution of power or equal access to linguistic resources; the article mainly addresses the principles of process control and interactional justice; other principles are explored in related research (e.g. Trinder *et al.* 2014; Tkacukova 2020; Grieshofer *et al.* 2021).

Both process control and interactional justice are restricted, firstly, due to the complexity of court procedures and lack of procedural information or understanding among lay court users, and, secondly, due to the type of interactions embedded in court processes and procedures (expert-framed narrative genres before direct narrative stage), the delay in eliciting direct narratives and lack of discursive competence among lay court users. This speaks to the core of the issues identified and solutions suggested for private family courts by Hunter *et al.* (2020: 172), which propose to address issues with the design of basic processes by adopting a non-adversarial investigative approach based on open enquiry. The shift away from the adversarial approach to a more investigative and open enquiry based approach presents an opportunity for linguistic research to contribute to changing the communicative strategies embedded within the current narrative practices and evidence elicitation procedures and propose a more user-friendly approach in accordance with the principles of procedural justice.

Notes

[1]<https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/806959/HMCTS_Refo

[2]The study is part of the AHRC funded project *The Language of DIY Justice: Communication Practices & Processes*.

[3]Children and Family Court Advisory and Support Service (CAFCASS)

References

Adler, M. (2012). The plain language movement. In P. Tiersma and L. Solan, Eds., *The Oxford Handbook of Language and Law*. OUP, 67–86.

Assy, R. (2011). Can the law speak directly to its subjects? the limitation of plain language. *Journal of Law and Society*, 38(3), 376–404.

Azuelos-Atias, S. (2011). On the incoherence of legal language to the general public. *International Journal for the Semiotics of Law*, 24(1), 41–59.

Bendall, C. (2020). Should we welcome an end to the 'blame game'? reflecting on experiences of civil partnership dissolution. *Journal of Divorce & Remarriage*, 61(5), 344–365.

Bernstein, B. (1990). *The structuring of pedagogic discourse*. Routledge.

Bhatia, V. (2004). *Worlds of written discourse: A genre-based view*. Bloomsbury.

Bhatia, V. and Bhatia, A. (2011). Legal discourse across cultures and socio-pragmatic contexts. *World Englishes*, 30(4), 481–495.

Blommaert, J. (2008). Bernstein and poetics revisited: Voice, globalization and education. *Discourse & Society*, 19(4), 425–451.

Byrom, N. (2019). Digital justice: HMCTS data strategy and delivering access to justice. https://research.thelegaleducationfoundation.org/wp-content/uploads/2019/09/DigitalJusticeFINAL.pdf.

Carbaugh, D. (1989). The critical voice in ethnography of communication research. *Research on Language & Social Interaction*, 23(1-4), 261–281.

Cooper, P. and Mattison, M. (2021). *Witness statements for the employment tribunal in England and Wales: What are the issues? Project Report*. London, UK: Institute for Crime and Justice Policy Research.

Cotterill, J. (2003). Language and power in court: a linguistic analysis of the o.j. *Simpson trial. Palgrave*.

Coulthard, M. and Johnson, A. (2007). *Introducing Forensic Linguistics: Language in Evidence*. Routledge.

Eades, D. (1996). Verbatim courtroom transcripts and discourse analysis. In H. Kniffka, Ed., *Recent Developments in Forensic Linguistics*. Peter Lang, 241–254.

Ejimabo, N. (2015). The effective research process: Unlocking the advantages of ethnographic strategies in the qualitative research methods. *European Scientific Journal*, 11(23), 356–383.

Fraser, H. (2003). Issues in transcription: factors affecting the reliability of transcripts as evidence in legal cases. *Forensic Linguistics*, 10(2), 203–226.

Gibbons, J. (2003). *Forensic linguistics: An Introduction to Language in the Legal System*. Blackwell.

Greene, E., Fogler, K. and Gibson, S. (2012). Do people comprehend legal language in wills? *Applied Cognitive Psychology*, 26(4), 500–507.

Grieshofer, T., Gee, M. and Morton, R. (2021). The journey to comprehensibility: Court forms as the first barrier to accessing justice. *International Journal for the Semiotics of Law*.

Harris, S. (2001). Fragmented narratives and multiple tellers: Witness and defendant accounts in trials. *Discourse Studies*, 3(1), 53–74.

Harris, S. (2005). Telling stories and giving evidence: The hybridisation of narrative and non-narrative modes of discourse in a sexual assault trial. In J. Thornborrow and J. Coates, Eds., *The sociolinguistics of narrative*. John Benjamins, 215–237.

Heffer, C. (2005). *The language of jury trial: a corpus-aided analysis of legal-lay discourse*. Palgrave Macmillan.

Heffer, C. (2013). *Projecting voice: towards an agentive understanding of a critical capacity. Working Papers in Language and Literature*. Cardiff University.

Heffer, C. (2018). Narrative practices and voice in court. In J. Visconti, Ed., *Handbook of Communication in the Legal Sphere*. De Gruyter Mouton, 256–284.

Heffer, C. (2019). Suppression, silencing and failure to project: Ways of losing voice while using it. In R. Page, B. Busse and N. Nørgaard, Eds., *Rethinking Language, Text and Context: Interdisciplinary Research in Stylistics in Honour of Michael Toolan*. Routledge, 237–253.

C. Heffer, F. Rock and J. Conley, Eds. (2013). *Legal-lay communication: Textual travels in the law*. Oxford University Press.

Hiltunen, R. (2012). The grammar and structure of legal texts. In P. Tiersma and L. Solan, Eds., *The Oxford Handbook of Language and Law*. Oxford University Press, 39–51.

Hrabovska, O., Zakharova, O. and Priazhenkova, N. (2021). The oral factor in adversarial civil proceedings. *Linguistics and Culture Review*, 5(4), 832–844.

Hunter, R., Burton, M. and Trinder, L. (2020). *Assessing Risk of Harm to Children and Parents in Private Law Children Cases: Final Report*. London: Ministry of Justice. https://kar.kent.ac.uk/81894/1/assessing-risk-harm-children-parents-pl-childrens-cases-report_.pdf.

Hymes, D. (1962). The ethnography of speaking. In T. Gladwin and W. Sturtevant, Eds., *Anthropology and human behavior*. Anthropological Society of Washington, 13–53.

Hymes, D. (1996). *Ethnography, linguistics, narrative inequality: Toward an understanding of voice*. Taylor & Francis.

Hyvärinen, M. (2015). Analyzing narrative genres. In A. Fina and A. Georgakopoulou, Eds., *The handbook of narrative analysis*. Wiley Blackwell, 178–193.

Labov, W. (1972). The transformation of experience in narrative syntax. In W. Labov, Ed., *Language in the Inner City: Studies in the Black English Vernacular*. University of Pennsylvania Press, 354–396.

Lee, R. and Tkacukova, T. (2017). *A Study of Litigants in person in Birmingham Civil Justice Centre*. Birmingham: CEPLER Working Paper Series.

Lens, V. (2016). Against the grain: Therapeutic judging in a traditional family court. *Law & Social Inquiry*, 41(3), 701–718.

Macdonald, G. (2017). Hearing children's voices? including children's perspectives on their experiences of domestic violence in welfare reports prepared for the english courts in private family law proceedings. *Child abuse & neglect*, 65, 1–13.

MacFarlane, J. (2013). *The National Self-Represented Litigants Project: Identifying and Meeting the Needs of Self-Represented Litigants: Final Report*. Representing Yourself Canada.

Masson, M. and Waldron, M. (1994). Comprehension of legal contracts by non-experts: Effectiveness of plain language redrafting. *Applied Cognitive Psychology*, 8, 67–85.

McKeever, G., Royal-Dawson, L., Kirk, E. and McCord, J. (2018). *Litigants in Person in Northern Ireland: Barriers to Legal Participation*. Ulster University.

Mindlin, M. (2005). Is plain language better-a comparative readability study of court forms. *Scribes Journal of Legal Writing*, 10, 55–66.

Ministry of Justice, (2021a). Family court statistics quarterly: April to june 2021. https://www.gov.uk/government/statistics/family-court-statistics-quarterly-april-to-june-2021/family-court-statistics-quarterly-april-to-june-2021#children-act---private-law.

Ministry of Justice, (2021b). Guide to family court statistics. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/622664/guide-to-family-court-statistics-june-2017.pdf.

Ministry of Justice, (2021c). Practice direction 12b – child arrangements programme.

O'Barr, W. (1982). *Linguistic Evidence: Language, Power, and Strategy in the Courtroom*. Academic Press.

O'Barr, W. and Conley, J. (1991). Litigant satisfaction versus legal adequacy in small claims court narratives. In D. Papke, Ed., *Narrative and the Legal Discourse: A Reader in Storytelling and the Law*. Deborah Charles, 65–89.

Pavlenko, A., Hepford, E. and Jarvis, S. (2019). An illusion of understanding: how native and non-native speakers of English understand (and misunderstand) their miranda rights. *International Journal of Speech, Language & the Law*, 26(2), 181–207.

Pleasence, P. and Balmer, N. (2019). Justice & the capability to function in society. *Daedalus*, 148(1), 140–149.

Sandefur, R. (2015). Elements of professional expertise: Understanding relational and substantive expertise through lawyers' impact. *American Sociological Review*, 80(5), 909–933.

Sangasubana, N. (2011). How to conduct ethnographic research. *Qualitative Report*, 16(2), 567–573.

Sela, A. (2018). Can computers be fair: How automated and human-powered online dispute resolution affect procedural justice in mediation and arbitration. *Ohio State Journal on Dispute Resolution*, 33, 91–148.

Thornborrow, J. (2002). *Power talk: Language and interaction in institutional discourse*. Longman.

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.

Tkacukova, T. (2016). Communication in family court: Financial remedy proceedings from the perspective of litigants in person. *Journal of Social Welfare and Family Law*, 38(4), 430–449.

Tkacukova, T. (2020). Changing landscape of advice provision: Online forums and social media run by mckenzie friends. *Child and Family Law Quarterly*, 4, 397–420.

Toy-Cronin, B. (2019). Leaving emotion out: Litigants in person and emotion in new zealand civil courts. https://doi.org/10.35295/osls.iisl/0000-0000-0000-1034.

Trinder, L., Hunter, R., Hitchings, E., Miles, J., Moorhead, R., Smith, L., Sefton, M., Hinchly, V., Bader, K. and Pearce, J. (2014). Litigants in person in private family law cases. *Ministry of Justice Analytical Series*. www.justice.gov.uk/publications/research-and-analysis/moj.

Tyler, T. (2000). Social justice: Outcome and procedure. *International Journal of Psychology*, 35(2), 117–125.

Walker, A. (1986). The verbatim record: the myth and the reality. In S. Fisher and A. Todd, Eds., *Discourse and Institutional Discourse*. Ablex, 205–22.

Walker, A. (1990). Language at work in the law: the customs, conventions, and appellate consequences of court reporting. In J. Levi and A. Walker, Eds., *Language in the Judicial Process*. Plenum Press, 203–244.

Yeung, M. and Leung, J. (2019). Litigating without speaking legalese: the case of unrepresented litigants in hong kong. *International Journal of Speech, Language & the Law*, 26(2), 231–256.

Ződi, Z. (2019). The limits of plain legal language: understanding the comprehensible style in law. *International Journal of Law in Context*, 15(3), 246–262.

‘Psy’ expert evidence in the family courts: The potential for corpus-assisted analysis

Lauren Devine, Stephen Parker, Leigh Harrington & Nadia Makouar

Centre for Language and Law, Aston Institute for Forensic Linguistics
Aston University, UK

https://doi.org/10.21747/21833745/lanlaw/9_1a5

Abstract. *This article introduces corpus-assisted linguistic methods as an exploratory means of analysing expert psychologists’ reports used in public family law (child protection) cases. Analysis of this dataset is a new application for corpus linguistics (CL) and the primary purpose of this article is to explore viability and potential for its future research using CL as a core method. For this study we have created and analysed a 25 single-text-type specialised written corpus consisting of 25 expert psychologists’ reports (the Psychology Report Corpus “PRC-25”). The reports are a random sample selected from a population of all psychologists’ reports held in Cafcass files over a 10-year period, representing the first corpus of its kind in a currently under-researched area. Our study uses both an inductive (data-driven) approach to identify significant themes and topics in the reports, and a deductive (legal-intuitive) approach to explore psychologists’ use of legally significant terms, especially risk of and significant harm. We also explore the possibility for using this new methodological protocol to triangulate analysis of a larger and representative corpus of expert psychologists’ reports, and the possibilities for corpus-driven analysis of the genre of written expert evidence text types more generally.*

Keywords: *Expert evidence, Psychologists’ reports, Family justice system, Corpus linguistics, Child protection.*

Resumo. *Este artigo apresenta métodos linguísticos baseados em corpora como forma exploratória de analisar relatórios periciais de psicólogos utilizados em casos de direito da família (proteção de menores). A análise deste dataset constitui uma nova aplicação da linguística de corpus (CL), pelo que o objetivo primordial deste artigo é explorar a viabilidade e o potencial para a sua investigação futura utilizando linguística de corpus como principal método. Para este estudo, criámos e analisámos 25 corpora escritos especializados com um único tipo de texto, constituídos por 25 relatórios periciais de psicólogos (o Psychology Report Corpus “PRC-25”). Os relatórios constituem uma amostra aleatória selecionada a partir de*

uma população de todos os relatórios de psicólogos existentes nos arquivos Cafcass por um período de 10 anos, representando o primeiro corpus do género numa área atualmente sub-investigada. Este estudo utiliza, quer uma abordagem indutiva (baseada em dados) para identificar temas e tópicos relevantes nos relatórios, quer uma abordagem dedutiva (legal-intuitiva) para explorar a utilização de termos juridicamente relevantes por parte dos psicólogos, especialmente riscos e danos relevantes. Também exploramos a possibilidade de utilizar este novo protocolo metodológico para triangular a análise de um corpus maior e mais representativo de relatórios periciais de psicólogos, e as possibilidades de análise baseada em corpora do género de textos prova pericial, em geral.

Palavras-chave: *Prova pericial, Relatórios de psicólogos, Sistema judicial de família, Linguística de corpus, Proteção de menores.*

Introduction

This study was conducted as a pilot, extending a project investigating why Section 31 care order applications in England are consistently high.[1] Interviews conducted with lawyers and the judiciary at this start of the project revealed themes and trends of concern to those involved in decision making in the family courts. In these interviews, psychologists' expert reports were highlighted as an area of concern, particularly by the judiciary who felt more clarity in the use of legal language would assist them in making decisions. This supports findings in previous research, carried out for the Ministry of Justice's Family Justice Council by a team of inter-rating forensic and clinical psychologists using non-corpus methods to conduct largely qualitative research. This research found numerous issues of concern in the reports, including unsuitable and inconsistently used psychological tests, unsupported inferences drawn by the expert psychologists, misreporting of allegations as fact, and unnecessarily pejorative personal comments (Ireland 2012). New standards for expert witnesses in Children's Proceedings in the family courts were drafted as a response to these (and other) concerns to assist with the future focus and relevance of expert psychological evidence in child protection proceedings (Ministry of Justice and Family Justice Council 2013). For the analysis reported here, a pilot corpus, the PRC-25, was built and used to identify and examine the use of selected legal terms in psychologists' reports, as the judicial interviews indicated that the use of legal terms by experts was still an area of concern.

Expert psychologists' reports play a significant role in the court's decision about whether a child remains with their family, or whether parental responsibility is transferred to the State (Children Act 1989, Section 31). Other than Ireland (2012), there is no substantive empirical research into the content, quality, or consistency of these reports as forms of expert evidence, and Ireland's study was not representative. This article explores the potential application of corpus-assisted methods to "problem-solving" practical issues arising in forensic courtroom interactions, such as evaluating the focus and relevance of the expert psychologists' reports which can be applied to larger, representative corpora.

Corpus-assisted methods can be used to provide a quantitative overview of corpora and to contextualise qualitative results. Primarily a computational methodology using statistical procedures to examine large, machine-readable (or electronic) linguistic datasets (corpora), corpus linguistics (CL) is an established methodology in numerous

areas of linguistic interest across written and spoken contexts, including forensic linguistics (Gillings 2022), political discourse (Taylor 2022), news media (Marchi 2022), language evaluation (Scott 2010), and online informal legal advice in child protection cases (Tkacukova 2020). In conceptualising the paper, the authors also drew on the growing body of CL literature which considers and encourages the potential of corpus linguistics in new areas (Brezina and McEnery 2015; Brezina 2018), selected linguistics literature (Lebart *et al.* 1997; Partington 1998; Sinclair 1999; Cutting 2007; Baker *et al.* 2008), and relevant law and related documentation (Children Act 1989; HM Government 2018; Family Procedure Rules, Part 25 2020). As so little previous linguistic research on this data exists, established methodological protocols for their systematic analysis do not exist. Unless corpus assisted methods are used, time and resource limitations mean research leans towards small-scale qualitative studies conducted on an individual-report level, which is the limitation reported by Ireland (2012). As such, CL offers a promising avenue for the analysis of this type of forensic expert evidence and platform from which detailed linguistic analysis can be undertaken. A further limitation of previous research drawing on expert witness reports has been gaining access to and extracting meaningful large-volume data for analysis as the data were created for litigation, not research purposes. Although the detailed work involved in creating a corpus cannot be underestimated, a corpus of such data can be re-used to answer a variety of research questions.

Transparently reported, CL findings are reliable and replicable, offering insights into salient, as well as more subtle linguistic patterns in a corpus. As such, Charteris-Black and Seale (2010) assert that these tools can provide ideal avenues into the exploration of new corpora in principled ways based on statistical saliency rather than exclusively on researcher intuition. However, Baker *et al.* (2008: 274) observe that most corpus analysis does nonetheless require human input at various stages, whether deciding which texts to include when creating a corpus or when interpreting the findings of quantitative tests. Existing CL research points to the analytical synergies achieved by supplementing qualitative approaches with quantitative ones, the most suitable approach depending on the specific research questions (Baker *et al.* 2008; Biber *et al.* 1998).

The Psychology Report Corpus-25 (PRC-25, Table 1 below) is a specialised single text-type written English corpus created by the authors of this paper, comprising 25 expert psychologists' reports used as expert evidence in child protection cases between 2011-2021. The corpus contains 235,111 running words, 11,763 types (the number of distinct words) and 9,436 lemmas (the form of words under which they are registered in a dictionary, for example, "parents", "parenting" and "parental" belong to the same lemma "parent"). The average length of the reports is 11,048 words (longest = 31,799, shortest = 1,311).

Name	**Language**	**Texts**	**Tokens (Total word count)**	**Additional information**
Psychology Report Corpus (PRC-25)	English	25	235,111	Types: 11,763 Lemmas: 9,436

Table 1. The PRC-25

To create the PRC-25, the authors accessed the whole database of expert psychologists' reports compiled across England and Wales between 2011-2021, held by the Children and Family Court Advisory and Support Service (Cafcass).[2] The total population of reports was >10,000, ranging from 4 to 120 pages. Each report was extracted by the authors from the Cafcass files and organised into an encrypted database. The pilot corpus was created from a random sample selected using the Excel RAND() function. The sample is too small be representative of all psychologist reports, but it is sufficiently sizeable that its 235,111 words would be difficult to analyse using manual, qualitative methods alone. Analysis was undertaken using two open access CL tools, #LancsBox v.6 (Brezina *et al.* 2020) and IRaMuTeQ (Loubère and Ratinaud 2014), a data analysis programme using R software and python coding language used for multi-dimensional text and questionnaire analysis. The PRC-25 was also compared with a larger, reference corpus, selected to represent contemporary general English for the same time period. This comparison was carried out to determine how the language used in expert psychologists' reports may be distinct from that of everyday English by statistically comparing the frequency of words in the PRC-25 to their frequency in a general reference corpus Stubbs (2010). Reference corpora are typically very large corpora that represent a chosen language across a variety of contexts within a defined timeframe. For this comparison the BNC Baby was selected, a four-million-word balanced subset of the British National Corpus, 2014 edition designed to be representative of present-day English-use. The BNC Baby 2014 is compiled of written and spoken texts in British English, including newspapers, fiction, academic writing, informal speech, and e-language.

For the analysis, an exploratory two-pronged approach was adopted, drawing on the disciplinary expertise of lawyers and linguists, all with knowledge of both disciplines. The aim of this exploratory analysis was to identify the frequency and general use of pre-selected legal terms used in the reports. These are important terms because the judiciary need experts reports to be "focussed and relevant" to tie their psychological evaluations to the relevant legal thresholds and criteria (Munby 2011). The first prong was a deductive legal-intuitive approach, informed by prior legal knowledge. Concordance and collocation tools (explained in Section 3.1 below) were used in this approach to analyse pre-selected legal terms. The second prong of the analysis was an inductive, data-driven approach. Drawing on no prior contextual knowledge, the main themes and topics in the PRC-25 were explored using keyword analysis. A hierarchical cluster analysis was also undertaken as part of the second approach. Overall, the aim was to consider whether the two-pronged inductive and deductive approaches might provide an effective means of methodological triangulation (Egbert and Baker 2020).

The expert psychologists' reports are intended to be documents to assist the court in their decision making, and as such are not primarily intended to comment on legal terms. However, to be able to contextualise the reports, some reference to the psychologists' findings in relation to the appropriate legal threshold terms might reasonably be expected. In the reports, the psychologists give their opinion in relation to parenting capacity, presenting an entextualised version of the verbal interviews between the expert psychologist and the respondent. This includes the psychologists' interpretation of other reported data, such as accounts from third parties. The role of predictive expert evidence based on the psychologists' estimates of *future* risk plays a significant part in the threshold question for the judiciary. This is a challenging task for both expert

and judge, as protracted social work intervention into private family life creates lengthy and often highly contested versions of complex real-life scenarios, which are not easily translated into items of evidence that will assist judicial decision making. To translate the lived reality of family experience into evidence enabling a court to adjudicate on whether a child should be removed from their family, there are procedural and structural measures in place intended to assist, set out in Practice Direction 12A (Ministry of Justice 2021a) of the Public Law Outline (PLO) and Practice Direction 25A (Ministry of Justice 2021b). It is in the context of this framework that the expert psychologists' reports are prepared and given in evidence.

Some initial observations were made about the reports during the corpus creation process. The authors did not conduct an analysis of register conventions in the PRC-25 but noted that the reports should be presented in the format required by PD25A and adhered to the conventions of the British Psychological Society's Guidance British Psychological Society (2016). The methodological approaches used in the reports were not standardised across the sample, but broadly included (1) collecting antecedents of the parents, including a history of their own childhoods prepared from verbal recollection by the parents; (2) interrogation of the parents in relation to the matters the Local Authority complains of in relation to their parenting ability (but not necessarily presented to the parent as an allegation which requires an answer); (3) assessment of the parents' psychological state; (4) assessment of the parents' willingness and ability to comply with Local Authority requirements; and (5) the concluding opinion of the psychologist about what is likely to happen in the future.

There was a lack of methodological transparency and rigour across the corpus in relation to the selection criteria for the multiple psychological tests used in the reports and no explanation of reliability, or how the conclusions in the reports were informed by the results. There was a lack of standardisation in relation to the final presentation of the individual reports. All the reports typically contained a 'past/present/future' narrative trajectory comprising: 1) the 'histories'; 2) the present situation; and 3) a prediction of what will happen next, but there was no particular order or structure to how this information was presented, and a wide variation in length. For example, one report contained a long continuous section titled "Interview with X parent", whereas another breaks the interview down into multiple themed sections, such as "Early Years", "Attitudes towards the task of parenting", and "Relationship with parents".

Finally, there were insufficiently transparent protocols for reporting the verbal interviews and explaining how inferences are drawn from them. This included gaps in explaining the methods of converting the spoken interviews into written data. It was unclear, therefore, how the experts conducted the interviews themselves, and no information is provided regarding the experts' preparation for the interview, or how they were conducted, recorded, transcribed, or interpreted. Many of these observations are consistent with the findings of Ireland's (2012) study.

Given these observations, it was hoped that corpus-assisted analysis would not only reveal the legal terms in the reports that are germane to the decision-making process. The analysis also considered how they were explicitly used, and whether precise terms were used consistently, as these terms are understood to have a consistent and specific meaning in the legal process.

The legal framework

The reports form a small but important part of a complex process of law and legal process and must be considered in context. The standard of proof in child protection cases is significantly lower than in the criminal courts but nonetheless has significant and long-lasting effects on children and their families (see for example Broadhurst and Mason 2013a). Case statistics are compiled each year, showing a consistently high level (see Figure 1). Experts' reports are important and necessary to assist judicial decision making, but the latest major government review of expert evidence in general found that poor expert evidence is problematic, particularly in relation to the difficulties in communication between expert witnesses and legal professionals (Law Commission 2009, 2011). Although the Law Commission specifically included cases involving child abuse (albeit in the criminal courts), expert psychologists' evidence in relation to child protection cases remained problematic (Ireland 2012). Specific guidance for psychologists giving expert evidence in the family courts exists, but no substantive evaluation of any improvements in psychologists' evidence has taken place since it was published (British Psychological Society 2016).

There are two substantive sub-fields of law involved in child protection cases: Public family law and expert evidence. Public family law operates as a sub-field of family law but is arguably also a sub-set of public law. Expert evidence operates as a sub-field of the broader law of evidence. The positioning of public family law as a sub-field of family law rather than a sub-field of public law, or even a re-named field of its own, raises complex issues of State power and individual vulnerability insufficiently recognised in the family law process (Devine 2017). The litigator in Public family law cases is the State, referred to as the applicant (Children Act 1989), the party defending their position is the party(ies) with parental responsibility for a child (generally the parent(s), almost always the mother), and the victim (alleged victim until a judgment is given) is the child(ren). Regardless of age, the child is simply the subject of the proceedings and cannot halt or dictate their progress or outcome (Children Act, 1989).

Drawing both sub-fields together in this investigation draws attention to the paucity of empirical research in either. This is both a consequence of general difficulties of gaining access to the data needed for analysis and the difficulties in manually analysing large volumes of courtroom data which has not previously been collated for research purposes. Exploring CL methods that will enable swifter consideration of larger volumes of data is hoped to benefit studies of larger specialist corpora of similar data types.

Public family law (child protection) cases

Child protection cases are the mechanism by which children considered to be "at risk of significant harm" can be removed into State care following an application by the Local Authority under Section 31 Children Act 1989. Judgments in child protection cases are focussed on the child's welfare as the paramount consideration (Section 1(1) Children Act 1989). The threshold legal test is set out in Re B (Care Proceedings: Appeal) [2013] and see Cobley and Lowe (2009) for comment on the earlier House of Lords decision).

Experts' reports provide substantive evidence in many cases and are often the sole source of information in relation to what is likely to happen to a child if he or she is not removed. As such, the reports primarily relate to the prediction of future risk to a child if the child is left with their family (or in their current place of residence). Predicting risk is

inherently risky, and normally requires expert forensic input, hence the long tradition of reliance on psychologists: Lady Justice Butler Sloss outlined the need for expert opinion evidence to assist the courts in general when children are involved:

> "Many if not all family cases involving children feature expert opinion evidence ... In cases involving children, expert medical and psychiatric evidence from paediatricians and allied disciplines is often quite indispensable to the Court. As Parker L.C.J. said in Director of Public Prosecutions v A & BC Chewing Gum ltd. [1968]1 Q.B. 159 at 165A, when dealing with children, the court needs 'all the help it can get.' " (Re M & R (Child Abuse: Evidence), 1996)

The use of expert psychologists' evidence in child protection proceedings is thus an important area for research. The prediction is intended to assist judges in deciding whether the risk of making no court order is greater than the risk of making an order. Following Ireland's study (2012), there is no recent substantive empirical research into their content, quality, or consistency, although there is detailed literature in relation to the language of courtroom interactions more generally (for example, Coulthard *et al.* (2016); Aldridge (2010)).

Psychologists' reports for expert evidence are generally commissioned to comment on two main areas: parental (or caregiver) and the child's functioning to establish whether there is a problem (which will have been articulated by the applicant Local Authority acting on behalf of the State), and to give predictive comment on what is likely to happen to the child in the future (risk prediction). Both areas are controversial as they necessarily involve elements of methodological uncertainty (Regehr *et al.* 2010; Camasso and Jagannathan 2000): psychologists often have little to assist them other than verbal recall and accounts of past events. In addition, the evidence-gathering process for the psychologists differs significantly from a clinical evaluation as the primary role of the psychologist as expert is not to offer therapeutic support to the interviewee. In the PRC-25, the respondents reported profound fear and trauma during their interviews in finding themselves facing child protection processes. In addition, most report having suffered historic abuse, and many were facing current domestic abuse situations. This reflects a well-established observation in published literature (Lindley 1994; Broadhurst and Mason 2013b). The full reports also include events involving third parties that could not be independently verified, for example, respondents are routinely asked for retrospective accounts of their own, often traumatic, childhood experiences (Broadhurst and Mason 2013a). In clinical practice, a psychologist is undertaking a therapeutic task which is not compatible with the aims of adversarial litigation where the subjects of their inquiry may be both in urgent need of psychological support (for general explanation about clinical judgment see, for example, Goldberg (1968).

Child protection cases are heard in the Family Courts, where the standard of proof is the 'balance of probabilities' rather than the criminal standard of 'beyond reasonable doubt'. This means that for the Local Authority applicant to obtain a judgment in their favour, their case must be stronger than the respondent's case: more likely than not; more than 50/50. In contrast, for the *respondent* to obtain judgment in their favour their case does not, in theory, need to be stronger, it simply needs to be at least of equal merit to that of the applicant. If both arguments are equally compelling the applicant should fail. In practice, most cases brought by Local Authorities result in an order being made even if it stops short of a care order, which of itself raises interesting research questions

(Redhill and Roe 2021: 31). It could be that Local Authorities are simply particularly good at only litigating where their case is likely to succeed on its merits, but it could also be that the power imbalance and vulnerability of the respondent(s) leave them at significant disadvantage.

Courts follow the procedures in the Public Law Outline (PLO) including a timescale for cases to be heard within 26 weeks (Ministry of Justice, 2021). The PLO was intended to help reduce the number of children removed, but in practice has placed considerable pressure on litigants and the courts to conclude cases in a timescale that makes therapeutic intervention a short-term aim rather than a long-term gain. Lengthy therapeutic intervention is frequently a recommendation of the psychologists' reports, leaving the courts with a conclusion that the intervention cannot be carried out within the time frame and thus the child must be removed. Ironically, it is increasingly the case that cases fail to meet this criterion anyway; in 2019 (Q2) the average case duration was 35 weeks. In 2021 (Q4), the average had risen to 44 weeks (CAFCASS 2022), considerably increasing the load on the courts.

A further factor of concern, also considerably increasing the load on the courts, is an increase in the rate of child protection cases (see Figure 1). Since its national roll-out in 2014, the rate of child protection cases increased year on year until 2016/17. Since then, the rates have not increased but have remained unacceptably high (CAFCASS 2022).

Expert evidence in child protection cases: Psychologists' Reports

The courts increasingly rely on expert evidence to help them understand areas outside their expertise (Law Commission 2011). Child protection cases are exceptionally reliant as "*many if not all family cases involving children feature expert opinion evidence…*" (Re M & R (Child Abuse: Evidence), 1996). Normally witnesses are barred from giving their opinions and are only allowed to testify on factual events which they have witnessed, and to what they know of their own thoughts and actions as they relate to the case (Hollington v Hewthorn 1943). Expert witnesses in child protection cases (and elsewhere) are the exception, as experts are required to give their opinion by using their specialist expertise to help the court to understand facts and evidence, which non-experts in their field would not be able to reliably evaluate otherwise (Family Procedure Rules, Part 25, Section 4.1 2020).

Some types of witness evidence are more 'certain' to a known confidence limit. Physical or documentary evidence can be analysed and explained to the court through the lens of the expert's translation into lay language, or the science can be presented with a quantifiable confidence limit (Law Commission 2011). These types of evidence require explanation to the court, but it is unlikely that other experts in the same field would come to differing conclusions given the objective evidence underpinning their evidence in fields such as engineering or accounting, for example. However, in some areas of expertise, such as psychologists' expert evidence, the data is less certain as it is largely collected from subjective, verbal accounts of past events provided by the respondents.

Existing literature in other areas of forensic linguistics has already established that witness testimony merits linguistic examination (for summary, see Coulthard *et al.* 2016). Extensive literature in child protection research detailing the complex nature of risk prediction also suggests scope for linguistic analysis (Møller Jørgensen *et al.* 2021; Smeeton 2020; Parton 2011; Munro 2010; Gillingham 2006). The courts are tasked with

Figure 1. Child Protection Cases: 2014/15 – 2021/22 (CAFCASS, 2022)

making difficult decisions, which can have a significant impact on people's lives, and a wrong decision can be catastrophic. No feedback is given to the courts on outcomes of their decisions. For all these reasons, great emphasis is given in child protection cases to the importance of experts. In the same judgment referred to above, Lady Justice Butler-Sloss emphasises the importance of expert evidence in child protection cases, stating:

> "… when the judge is of the opinion that the witness' expertise is still required to assist him to answer the ultimate questions (including, where appropriate, credibility) then the judge can safely and gratefully rely on such evidence, while never losing sight of the fact that the final decision is for him." (Re M & R (Child Abuse: Evidence), 1996)

Courts are therefore wanting to be sure that they make the right decisions, looking to the expert to give them as much certainty as possible as to what that decision should be. The general rule for experts is that they can give their opinions but not comment on the ultimate issue (the decision that the court is there to make), on the basis that it is the court, not the expert, which needs to decide the matter at hand. However, and uniquely, in the family courts, the Family Procedure Rules (FPR) do allow experts to comment on the ultimate issue as they are trying to base a decision on a prediction of what might happen in the future. This gives the psychologists' reports significantly more influence over the court's decision than other types of expert evidence.

Methods: a two-pronged approach

For this exploratory study a two-pronged approach to the analysis of the PRC-25 was adopted. Both methods, the deductive legal-intuitive approach, and the inductive, data-driven approach, employ core corpus methods and offer different disciplinary perspectives which lend themselves to a comprehensive overview of a corpus. The tests used were concordance and collocation analysis for the deductive approach, and keyword analysis and hierarchical clustering for the inductive approach. These approaches collectively offer insights into linguistic patterns in the PRC-25 and provide a platform for additional, future qualitative analysis.

Deductive legal-intuitive approach

An advantage of corpus-assisted analysis of large volumes of legal texts and forensic materials such as those considered in the PRC-25 is its ability to respond to specific research questions. For this approach, the authors set out to test the potential of the PRC-25 to respond to pre-set areas of interest, in this case the use of specific legal terms. This type of analysis would respond well to questions asked by subject experts, such as those within the legal profession, the judiciary, psychologists and other experts, policy makers and academics. The selection criteria for these terms specified that they must be in either the primary and secondary legislation (Children Act 1989; HM Government, 2018) and that they must represent terms that experts are routinely instructed to assist the court with. The list for this latter criterion was identified from the written instructions to the psychologists (to which the authors have access as a contextual note attached to the original reports).

The terms, set out in Table 2, are: *best interests;* (HM Government 2018 p. 28 para. 58), *significant harm;* (HM Government 2018, p.28, para.58 (2)A) and Section 31(2)(a) Children Act 1989), *risk of;* (Children Act 1989, c. 41., Section 1(3)(e) England), *reasonable;*

(Children Act 1989, Section 47); *a concern;* (HM Government 2018, p.17 para. 17), and *opinion;* (Family Procedure Rules, Part 25, Section 4.1 2020). Table 2 lists their absolute and relative frequency and dispersion across the 25 texts. The asterisks, denoting wildcard searches on most corpus linguistic software, are included as these searches retrieve words directly matching the characters before the asterisk with any other combination of characters immediately after it. For example, searches for *reasonabl** retrieved instances of *reasonable* and *reasonably* from the PRC-25).

Legal term	Absolute Frequency in PRC-25	Relative Frequency (per 10k)	Reports appeared in PRC-25
Best interest*	20	0.85	11/25
Significant harm	2	0.09	2/25
Risk of*	75	3.19	14/25
Reasonabl*	33	1.4	12/25
A concern*	6	0.26	6/25
Opinion	159	6.763	22/25

Table 2. Pre-selected legal terms

This approach then combined the corpus linguistic tools of concordancing and collocation to analyse the set of pre-selected legal terms.

Concordancing is a tool central to all corpus analysis which enables users to display all examples of a search term or phrase, in its original context in the corpus, with a defined span of words (typically up to 20) to its left and right (Rayson 2015). A such, concordancing enables analysts to view large numbers of examples together in one place, and concordance lines around the search term can be sorted according to different criteria which help highlight patterns in the search term's use in the corpus (Baker 2006; Hunston 2010). Using the Key Word in Context (KWIC) function in the corpus software #Lancsbox, we generated concordance lines in the PRC-25 for each pre-selected legal term, enabling our lawyer-analysts to gauge their general use in the psychologist reports.

The concordance analysis was paired with a separate CL tool, collocation analysis, to identify the collocates of the terms *harm* (n=100, occurring in 13/25 texts) and *risk* (n=254, occurring in 14/25 texts). A collocate is a word that frequently co-occurs with another word in a way that is statistically significant. For example, a strong collocate of *fish* in British English is *chips* (Bailey 2019). *Harm* and *risk* were selected as they appeared the pre-selected legal term list. We undertook collocation analysis to complement concordance analysis. The latter is the most qualitative part of corpus linguistics, relying on the analyst to recognise linguistic patterns and explain their meanings and functions (Bailey 2019: 64). Collocation analysis highlights language usage that can be hard to pinpoint via (legal) intuition alone and is replicable, with the benefit of providing different analysts with the same, reliable results in a way that concordance analysis may not (Xiao 2015).

Inductive data-driven approach

The data-driven approach is more suited to furthering general understanding of this little understood genre by testing the ability of the corpus to respond to purely linguistic inquiries which assume no prior legal knowledge of the PRC-25. This analysis was

conducted using a keyword analysis of the PRC-25, using the BNC Baby 2014 as a reference corpus. Keyword analysis was also combined with hierarchical clustering analysis which visualised the data in a dendrogram. Together they offer an insight into the major themes and topics of the PRC-25.

Keyword analysis is a core corpus linguistic tool that is widely considered to provide a useful starting point for exploring specialised corpora Evison (2010). It provides a data-driven method for approaching corpora in a principled way, based on statistical saliency rather than following researcher intuition (legal or otherwise) alone (Charteris-Black and Seale 2010). Keywords are words that appear significantly more frequently in a specialised corpus, such as the PRC-25, than would be expected when compared with general language use, which is represented by a reference corpus (Baker 2006; Stubbs 2010). As such, keywords point to the "aboutness of a text or a corpus" (Scott and Tribble 2006) helping to identifying a corpora's major themes (Scott 2010), and highlight other linguistic features worthy of further analysis (Bondi 2010). Keyword analysis, therefore, provides an ideal avenue for a data-driven approach to the PRC-25.

Using the corpus software #LancsBox, we have defined a keyword as a word that occurs at least 5 times in the PRC-25 and has a log ratio statistic above 8. Log-ratio, an effect-size statistic that emphasises absolute frequencies, indicates the relationship between word occurrences in a specialised corpus and a reference corpus (Pojanapunya and Todd 2016). A log-ratio was used as an effect-size statistic the PRC-25 is a small pilot corpus of the much larger database of reports. As such, the PRC-25 cannot be treated as representative of all reports. The log-ratio calculation generated an initial list of 223 key words. For the purposes of this first empirical exploration of the PRC-25, which aimed to determine the corpus' main themes, it was decided to focus primarily on lexical keywords. Consequently, one mathematical symbol was removed, as were functional keywords and contractions (e.g. *it's*), and 88 proper nouns to leave a final keyword list of 122 keywords. Concordance lines of each keyword were examined to determine their general usage in the PRC-25 and to group keywords into semantic categories. This process of semantic categorisation, explained in the findings section below, highlights the major themes of the PRC-25 (Scott 2010).

A data visualisation tool, the hierarchical cluster analysis, was used to generate a dendrogram, carried out to consider topics as a discourse category as they appeared across the whole of the PRC-25. These clusters, generated by IRaMuTeQ, highlight the semantic characteristics of prevalent topics in a corpus. This approach is based on Van Dijk (2000)'s consideration for semantic study of text. Topics, defined as semantic macrostructures, regulate the overall coherence of discourse, representing what speakers find most important (Van Dijk 2000: 234). There are many different types of cluster analysis (Everitt *et al.* 2011)) but for CL analysis a hierarchical agglomerative cluster analysis is typically used (Gries 2013: 336ff). For this analysis, individual data points were extracted by the IRaMuTeQ procedure and joined together to create larger clusters. These clusters contained all the data points, displayed in a tree diagram (dendrogram) in Figure 2 (Brezina 2018).

The dendrogram enabled exploration of the PRC-25 by highlighting its internal characteristics, in comparison to the keyword analysis which compared the PRC-25 corpus to a reference corpus. The rationale of this exploratory and free assumption visual method is predicated on revealing themes that tend to be found in similar contexts

(Lebart *et al.* 1997). This analysis complemented the keyword analysis by identifying and classifying additional major thematic groups found across the reports where the clusters are likened to topics in the sense that they highlight the internal semantic organization of the corpus.

Since the role of psychologists in these specific cases is at the intersection of psychology and law, the identification of lexical fields could shed light on the content of these reports and, in particular, topics that psychologists find important. According to Van Dijk (2000: 234), topics are "*[d]efined as semantic macrostructures and [...] represent what speakers find most important, they regulate overall coherence of discourse, how discourse is planned and globally controlled and understood, and what is best remembered by the recipients*". The hierarchical classification extracted "context units", or "text segments", from a lemmatized version of the corpus, each segment containing 40 occurrences/forms. Lemmatization is a process of regrouping words related to a same stem into a single lemma (Brezina 2018: 40). The principle of the procedure implies that the tool cuts out the context units, or text segments, from the lemmatized corpus, removing the inflectional word endings. The PRC-25 was lemmatized only for this clustering procedure step. Each text segment contained 40 occurrences/forms which were clustered according to their vocabularies and distributed according to the reduced form frequencies, bringing together forms found in similar contexts (Lebart *et al.* 1997) before being divided successively into clusters (dichotomously at each step). Thus, the more times that the number of common forms to two given segments is high, the more these two segments are considered close and likely to be grouped together in a same cluster.

Results: synergies and separations

Concordance analysis: pre-selected legal terms

The expression of the pre-selected legal terms in the concordance lines revealed that they were frequently used out of their correct legal context and were either not understood in the reports to have specific legal meaning or were used to imply and comment on a legal meaning, again out of context. The following extracts illustrate sample concordance lines with a brief narrative explanation:

Extract 1:

> *Although I am very clear in my mind that it is not in [the child's]* ***best interests*** *to return to her parents, there is no easy solution to [the child's] alternative placement*

*best interest** (used 20 times in 11 of the reports) was used to support the expert's opinion, although *interests* rather than *best interests* is the appropriate legal test unless there is a conflict of interests between a child(ren) and others (HM Government, 2018). In the full reports, however, the use of *best interests* did not correlate with the required conflict of interest criteria and may have been used out of context.

Extract 2:

> *The Local Authority had concerns that [the child] would be at risk of* ***significant harm*** *due to the controlling nature and domestic violence they believed to be a feature of [the mother's] relationship with [the boyfriend]*

Extract 3:

> *The Local Authority believes that the children have suffered* ***significant harm*** *due to being inadequately parented. It is also the Local Authority's position that [the mother] has failed to meet the*

*significant harm** appeared with contextual accuracy but was in only 2/25 reports, which is surprising as it is the primary threshold legal test. Example 2 aligns with the legal term *risk of significant harm*. Example 3 also uses a legal term *suffered significant harm* but is expressed with a non-legal term *have suffered*. The legal test uses the present tense, *is suffering*, or future tense *is likely to suffer*. Although contextually relevant to the court, the past tense should not influence the court in its decision making unless there is additional direct evidence of present or future significant harm.

Extract 4:

> *The HCR-20 is a standardised tool which has been shown in research to be an effective method of assessing the* ***risk of*** *future violence. However, it does not have a significant number of normed scores available; and therefore, to some extent, the*

Extract 5:

> *[the father] in my opinion presents a low to medium* ***risk of*** *engaging in future acts of violence or inappropriate behaviours which may directly affect any child in his care*

*risk of** was frequently used (75 times in 14/25 reports) and reflects the risk assessment objective of the psychologists in conducting the reports. It is surprising that it does not appear in more reports as it is the primary focus of the psychologists for their reports. It was also of note in the reports that the way in which *risk of* is used does not give the court much assistance in how precisely the risk is quantified or precisely which research is drawn on to support findings.

Extract 6:

> *If an individual were to score highly on either or both measures, compared with the normative sample, it is* ***reasonable*** *to conclude that the subject is dissimulating (faking good or lying).*

*reasonable** was used 33 times in 12/15 reports and was used to add weight to the psychologists' conclusions rather than in its legal context.

Extract 7:

> *This history of sexual victimisation is* ***a concern*** *Although the general research may or may not apply to her, as an individual,*

a concern was used 6 times in 6/25 reports. In all instances it supported a negative conclusion about the parent(s), indicating the psychologist was worried about something. In Example 7, the worry was about the prior victimisation of the mother.

Extract 8:

> *the reader that I am not an expert in 'attachment theory', yet I have been asked to state my professional* ***opinion*** *on this issue. As a general rule, I think it is reasonable to conclude that a child's permanency of care*

opinion was used 159 times in 22/25 reports. Experts are the only witnesses in court allowed to give their opinion, based on facts within their knowledge and expertise. In this example, the expert has been asked to comment outside their expertise and is potentially in breach of their duty to the court, although to their credit they highlight the issue in the text.

Identifying the concordances of the selected legal terms revealed the immediate contextual use of the terms across the PRC-25. The following collocation analysis identified the terms with which *risk* and *harm* co-occur.

Collocation analysis: risk and harm

Table 3 and Table 4 present the collocates of *risk* and *harm* in the PRC-25 respectively. The collocates in the tables have been generated using Mutual Information, an effect size measure commonly used to calculate collocation in corpus research. A possible limitation of MI against other collocation statistics is that it can tend to prioritise low frequency items (Baker 2006). To negate this, a minimum frequency threshold of 10 and a MI score of 5 for *risk* and 3 for *harm* was set. For both terms the calculation took into consideration words 5 to the left and 5 to the right of the search word.

Index	Position	Collocate	Stat (MI)	Freq (coll.)	Freq (corpus)
1	L	pose	9.446482	13	17
2	R	offending	8.779058	13	27
3	L	increased	8.511579	12	30
4	R	factors	8.166082	17	54
5	R	name*	7.87931	16	62
6	R	harm	7.511579	20	100
7	R	harming	7.485583	11	56
8	L	associated	7.460048	11	57
9	R	sexual	6.887546	34	262
10	L	low	6.648395	21	191
11	R	future	6.485583	11	112
12	R	name*	6.203029	13	161
13	R	self	5.919762	14	211
14	M	significant	5.789112	12	198
15	L	level	5.572979	12	230
16	R	assessment	5.355829	25	557
17	L	name*	5.093266	11	294
18	L	any	5.058406	21	575

Table 3. *risk* collocations in PRC-25

There are three quantifiers of risk in Table 3: *low, level*, and *significant*, which may help the court to assess the harm level perceived by the psychologist if these levels are able to be quantified. Quantifying human risk of future behaviour is challenging; in previous linguistic studies concerning risk and quantification, risk was used non-quantitatively (Boholm 2018). The expected collocate for *harm* would be *significant*, based on the legal threshold criterion in the primary legislation *significant harm* (Children Act 1989, Section 47) but this does not appear in the table. This observation merits further

Index	Position	Collocate	Stat (MI)	Freq (coll.)	Freq (corpus)
1	L	deliberate	10.04805	21	46
2	L	self	8.981756	46	211
3	L	physical	8.09183	12	102
4	L	risk	7.512537	20	254
5	L	emotional	7.463799	22	289
6	L	sexual	6.605302	11	262
7	L	from	5.529375	19	954
8	L	of	4.494243	58	5968
9	R	in	3.683248	21	3791
10	R	the	3.523243	43	8673
11	R	is	3.375808	10	2234
12	R	with	3.3038	12	2818
13	L	and	3.277512	33	7892
14	R	a	3.225097	15	3720
15	L	to	3.144219	30	7869

Table 4. *harm* collocations in PRC-25

examination in the context of whether the language used in the experts' reports is sufficiently clear and quantifiable. Legal language is often left deliberately vague and open to interpretation, as terms such as *significant harm* and *risk of* illustrate.

Keyword analysis: semantic categories

The keyword analysis and semantic categorisation of the keywords in the PRC-25 identified eleven main semantic categories (excluding "Other" below), some of which contain sub-groups (Table 5). The keywords in each category are listed with the highest keyness score first and the lowest last. Keywords marked with * appear in more than one semantic category (for instance, *self-worth* is both the focus of a some of the psychological tests that are explained, and it is also discussed in relation to the evaluation of an individual).

These semantic categories provide an overview of the main themes and concerns of the PRC-25. Firstly, the largest category (containing 88 keywords) was Proper nouns. Manually anonymising (replacing names and places with placeholders) the PRC-25 in full before running initial corpus procedures presented an extremely labour-intensive and time-consuming task. The keyword analysis was therefore run using the original reports, permanently redacting all identifying information retrospectively. Unsurprisingly, the names of individuals, organisations, and places were shown to be key in the PRC-25 as they were very unlikely to appear in the reference corpus, the BNC Baby 2014. This does, however, confirm the assumption that individual behaviours, rather than situations giving rise to child protection cases, are a major concern of these reports. This was checked by reading the full reports which confirmed the focus was generally on parent and child behaviours, even where concerns had been noted or raised about the general circumstances of the family, or poor professional responses, which were generally dismissed as irrelevant and not pursued. This is consistent with the role of the parent as respondent in the case. This is reflected in the related semantic category, Generic possessives (see also the Relations subcategory in Other). Examples of the

Semantic Categories	(Subgroups)	Keywords
Proper nouns (names and places)	**-Redacted-**	-Redacted-
Psychological testing	**Test types (technical detail)**	Wechsler, WASI, I.Q., CTQ, Achenbach, PDS, Tully, BDI-II, Bene, DSM-IV, MCMI, IIP, Millon, Centile, PSS, CISS, MCMI-III, Paulhus, Hickman, IQ, Visuo, CSRPI, DAP-IQ, STEEM, Maccoby, IASC, B/G, ASI, BDI, ABAS, Trichotech*, BAI
	Characteristic of test	self-reported, self-report, well-established, non-offending
	Focus of test	self-care, self-deception*, self-awareness*, self-direction, self-sacrificing, self-worth*, self-concept*, non-verbal, non-support, relatedness*, psychopathology*, malingering*
	Measure types	sub-tests, subscale, subtest, subscales, subtests, sub-test
Evaluation of subject(s) (psychological and behavioural description)		avoidant, symptomatology*, parented, life-story*, pre-contemplation*, psychopathology*, absconding, recidivism, self-harm, impulsivity, cooperative, minimisation, self-worth*, self-concept, relatedness*, self-sacrificing, malingering*, self-regulation, nurturance, self-deception*, self-awareness*
Generic possessives		mother's, child's, father's, person's, parent's, individual's, authority's, mum's, client's, grandmother's
Negative verbs		didn't, don't, doesn't, can't, wasn't, won't, hasn't, couldn't, wouldn't
Psychologist's credentials		Psychol, DFES, BPS British Psychological Society (2016), Clin., Solon, Psychopathology*, Symptomatology*, Multiagency
Inter-/intra-report references (i.e. cross-referencing, structure)	**CAFCASS reports themselves**	addendum, A/G, CAFCASS, parent-child*, pre-contemplation, affidavits
	Other relevant forms	C13, UAT, Trichotech*
Programmes		DAP (Integrated Domestic Abuse Programme), DBT (Dialectical Behavioural Therapy), Life-story*
Substances		citalopra, amphetamines
Parenting		parenting
Other	**Relations**	great-grandmother, great-grandfather, half-brother
	Dates	Dob, d.o.b., year's
	Miscellaneous	keyworkers, parent-child, reunification

Table 5. Semantic categories of lexical keywords in the PRC-25

excluded family circumstances included housing, health, poverty, parental vulnerability and/or victimisation and instances of reported professional failures to mitigate family needs.

The second semantic category, psychological testing, points to a second major concern of the reports. This category comprises mainly of the different tests that psychologists have used during their assessments. The sheer number of different tests named points to the non-standardisation of different psychologists' approaches. This reflects the fact that numerous tests are employed by psychologists when conducting clinical evaluations, but the statistical certainty of the outcome of the tests will not necessarily be understood by the courts (Law Commission 2011). As Extract 9 below illustrates, several reports do offer explanatory information about some of the tests used (as indicated by this semantic category's sub-categories, for example Characteristic of test, Focus of test, Measure types).

Extract 9:

> *[name] was psychometrically using the Wechsler Scale of Intelligence, designed for quick and accurately testing of an individual's intellectual functioning from the age of 6 onwards*

Several of the psychological testing keywords, such as *relatedness* and *self-worth,* also appear in the group of keywords concerning the psychologists' actual evaluation of the subject/parent in the Report itself. Therefore, unsurprisingly, descriptions of individuals' behaviour (adjectives: *cooperative, self-sacrificing*, verbs: *absconding*), and psychological states or traits (adjectives: *avoidant*, nouns: *impulsivity*) is another major focus of the PRC-25. It is also relatively unsurprising that the keyword analysis highlighted a semantic category whereby psychologists outline their professional credentials. For instance, the keyword *BPS* British Psychological Society (2016)(n=6) appears when psychologists outline their qualifications:

Extract 10:

> *I am a member of the BPS Division and Neuropsychology and I have completed further training on the use of*

Given that texts in PRC-25 take report form, and the complexity of related documents that psychologists may draw on in their assessments, the semantic category of Inter-/intra-report references, including report section titles and other documents is also unsurprising. Lastly, although the semantic category Parenting contains only one keyword, as would be expected in reports concerning parenting capacity, *parenting* (n=240) is a significant theme. Alphabetically sorted concordance lines immediately to the left and the right of the node *parenting* highlight several sub-themes of parenting referred to in the PRC-25. For instance, three sub-themes include parents' ability (i.e., *parenting ability*, *capacity*, *competence, skills*), measurable standards of parenting (i.e., *parenting assessment*, *performance*, *satisfaction*, *level*, *standard*, *basic*, *poor*, *positive*), and approaches to caring for children (i.e., *parenting behaviours, role*, *styles*). The sub-themes consist of subjective interpretations of behaviours and situations and their importance or otherwise. Future analysis should investigate the process by which the spoken and written materials from which the reports are constructed are entextualised into a re-constructed expert's narrative. From this initial analysis it was observed that the parent-respondents emphasised situations as being the issue of importance, whereas the psychologists emphasised behaviour as the primary focus.

Hierarchical clustering analysis: semantic categories

The results of the hierarchical clustering analysis match with some of the semantic categories in the keyword analysis. This especially applies to the psychological terminology category which matches with the institutional language cluster. Figure 2 shows that the hierarchical analysis revealed four clusters:

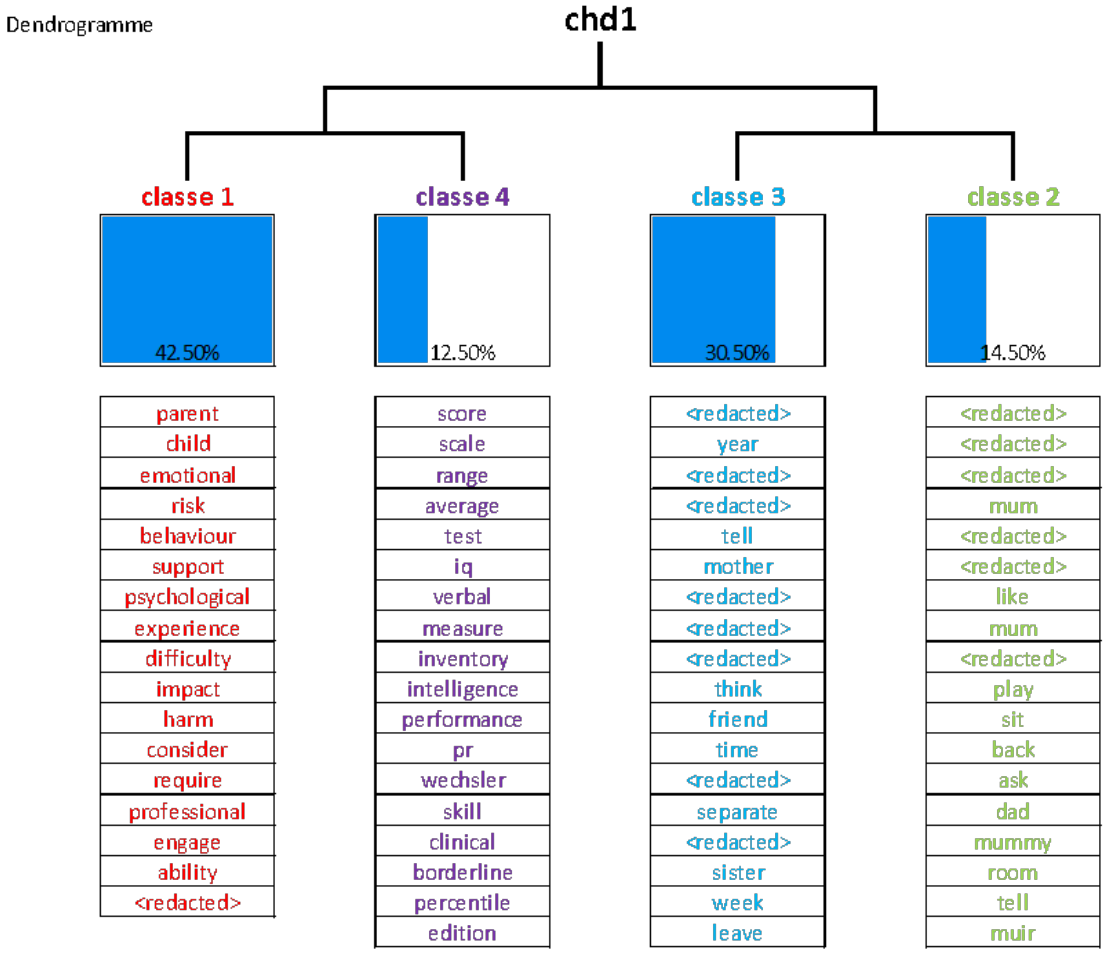

Figure 2. Hierarchical clusters in the PRC-25

Clusters 3 and 2 (blue and green respectively) represent one discourse type in the corpus. This discourse concerns the psychologists' reporting and reformulating the interviews with the parents and children in question and the contents of those interactions. Clusters 2 and 3 contain similar words, which will be explored in more detail in the larger study. As Figure 2 shows, these discourse clusters therefore comprise primarily of proper nouns, such as names and places (redacted here for anonymity), nouns relating to family members and familial relations (mother, dad, sister, friend), and a set of reporting clauses (tell, think, ask). As such, these clusters appear to represent those aspects of the interviews that the psychologist is observing and considers are worthy of reporting. Clusters 1 (red) and 4 (purple) represent a different type of discourse in the corpus: the professional and institutional-legal discourse used by the psychologists. From the small sample in the PRC-25 it was difficult to make general observations about whether they

were linked topically, semantically and/or organizationally in the reports themselves. The clusters did not directly correspond with sections of the reports, as the reports themselves did not follow a set structure and some did not contain clearly defined sections at all, as highlighted in the introduction where general observations were made about the whole reports within the PRC-25.

Cluster 4 (purple) groups the terms relating to quantitative assessment of psychological aptitudes and illustrates the scientific and statistical language (e.g., nouns such as score, measure, test, and percentile). It shows the specificity of this theme that dominates a wide part of the corpus and supports the results observed by the keyword analysis above. Finally, Cluster 1 represents a discourse of assessment developed in a legal and institutional language (e.g., nouns and adjectives such as *risk*, *psychological*, *impact*, *harm,* and verbs such as *consider*, *require*, *engage*).

The analysis included brief focus on the legal and institutional discourse identified by the hierarchical analysis. Table 6 shows the 20 first terms appearing in Cluster 1 (legal and institutional language) ranked on the basis of their Chi^2 score.

Term	Eff. s.t	Eff. total	Percentage	Chi2
parent	355	513	69.2	166.63
child	559	929	60.17	145.66
emotional	196	249	78.71	140.54
risk	136	155	87.74	133.9
behaviour	264	385	68.57	115.91
support	169	219	77.17	112.56
psychological	141	173	81.5	111.47
experience	173	233	74.25	100.72
difficulty	223	324	68.83	98.17
impact	84	90	93.33	96.85
harm	122	151	80.79	93.37
consider	151	201	75.12	91.14
require	85	96	88.54	84.85
professional	88	102	86.27	81.59
engage	117	149	78.52	81.49
ability	173	248	69.76	79.26
therapeutic	66	70	94.29	77.86
future	77	90	85.56	69.48
issue	114	151	75.5	69.31
attachment	97	123	78.86	68.17

Table 6. First 20 terms appearing in Cluster 1: legal and institutional language

The table shows also other values such as the number of text segments of this cluster containing the form at least once (Eff. s.t), the total number of text segments in the entire clustered corpus containing the form at least once (Eff. Total) and the percentage of Eff. ST compared to the Eff Total. The Chi^2 value expresses the level of intensity that links the term to the given cluster. It indicates the thematic affiliation of the terms and highlights the proximity and entwinement of the terms relating to psychological vocabulary and those relating to a legal vocabulary. A strong proximity

is observed between the psychologists' observation/conclusion vocabulary (e.g., *risk, ability, emotional, behaviour, experience, difficulty, impact*) and individual denomination (*parent(s), child(ren)*) or verbs (e.g., to *parent*, to *consider*, to *require*). Extracts 11 and 12 are two characteristic examples, meaning they have a high score, that contain the most frequent words associated with Class 1 such as *parent*, *behaviour*, *risk*, which are denoted in italics.

Extract 11:

> in ***light*** of this ***history*** therefore ***child protection professionals*** have ***concerns*** that ***B*** may not have the ***ability*** to ***protect*** *S* from ***future sexual harm*** if ***M*** is ***considered*** to *pose* a ***potential risk*** in this ***regard***

Extract 12:

> given the ***level*** of ***concern*** regarding ST's ***parenting*** and associated ***concerns*** about her ***children's behaviour***, ***alongside ST's difficulties engaging*** with ***professionals***, I do not believe she has the ***ability*** now or in the foreseeable ***future*** to ***meet*** her ***children's care*** needs

In these two examples, the psychologists use logical connectors (e.g., *in light of*, *therefore*, *given the level of concerns*) to justify their arguments regarding causes and consequences or effects in these cases. In these utterances, they seek to justify their observations (*I do not believe*, Extract 12) and translate them into language intended to be accessible to the court.

Discussion

Both the legal-intuitive and the data-driven approach yielded observations made possible within a relatively short analytic time frame by drawing on corpus assisted methods. The analysis illustrates potential to significantly broaden and enrich the scope of legal analysis by supplementing qualitative analysis with quantitative measures, revealing patterns of language use that are hard to identify by (legal) intuition alone Xiao (2015) or that may belie expectations of a corpus. Searching for pre-selected terms, however, proved useful, subject to the noted methodological limitations, occurring at the intersection of quantitative and qualitative analysis.

The authors found the dual approach most useful in relation to the concordance analysis. Such analysis requires a significant input from the analyst, and in the absence of objectively correct practices (which do not exist for this type of data) the organisation and analysis of concordance lines are generally determined by the individual researcher. The legal-intuitive analysis brings value to concordance analysis via contextual legal knowledge, however, basing analysis solely on this knowledge may also equally eclipse other important and/or unexpected linguistic patterns relevant to the analysis of pre-selected legal terms. Using further methods of analysis such as collocation analysis deepens the enquiry. As well as bringing salient but hard to detect linguistic patterns into relief, collocation analysis is also replicable, with the benefit of providing different analysts with the same, reliable results in a way that concordance analysis may not. It presents an essential tool for the analysis of large legal corpora.

Overall, the selected legal terms were expressed less frequently than might reasonably be expected given their prominence in the primary and secondary legislation, and there was a general lack of clarity in the reports to demonstrate experts sufficiently

understand how best to use legal terms to assist the court. There was uncertainty in relation to the way in which they are used and/or omitted in the reports, leaving it unclear whether experts either deliberately use legal terms out of context or are unaware the terms they are using have a specific meaning when used in expert evidence. A data-driven approach would help to identify if, or which, terms are used instead of explicit legal terms by the psychologists and whether the report authors are discussing risk in more implicit ways, or not discussing risk at all.

The single most significant omission in relation to the legal terms is the lack of reference to the terms denoting the legal threshold justifying the legal proceedings, *significant harm,* which only appears in 2 of the 25 reports. This is surprising as it seems a reasonable assumption that salient, threshold legal terms would feature highly in the word frequency list as the purpose of the reports includes providing certainty to the courts (PD12A). The two instances can be contrasted with, for example, the use of *mother*, appearing 580 times in the PRC-25. Although specific analysis of gender bias (or otherwise) is outside the scope of this paper, emphasis on "mother" rather than "father" or "parent" reflects historic research findings arguing gender bias in child protection with over-focus on mothers (Appell 1997; Edleson 1998). Although other phrases can be used to convey the concepts of *significant* and *harm*, in legal materials the precise phrase is imperative and if the reports simply do not engage with this term it is hard to see how they are going to be useful to the courts.

There was also a widespread use of generally unquantified phrases throughout the PRC-25, reflected most starkly in the use of legal terms. Consequently, the terms used in the reports do not necessarily serve to clarify the issues at hand. Our analysis of *harm* and *risk* illustrate this; we found the terms were used without specific reference to how they should be quantified, which reflects their use in general English rather than specialised expert evidence (Boholm 2018). Although there were three statistically meaningful instances where *risk* was quantified (*low, level*, and *significant*), little information was provided as to how that assessment was made. In legislation, some terms are left deliberately vague, enabling judges to interpret them (Li 2017). If expert evidence is unable to establish the thresholds between, for example, *harm, significant harm, no risk, some risk,* or *high risk,* further research on this point is indicated. These are the very issues where the court needs most guidance, hence the experts' ability to comment on the ultimate issue is a unique position. If experts are using legal terms with inherent uncertainty such as *harm, risk, reasonable* or *significant*, particularly where they are paired with terms expressing uncertainty, they are unlikely to provide the certainty the courts may hope for.

Perhaps unsurprisingly, the analysis of the semantic categories shows an over-representation of the terms relating to the vocabulary of psychological evaluation, rather than legal certainty. This was illustrated in the hierarchical clustering analysis which highlighted the coexistence of four predominant themes in the corpus. Two of the clusters the have similarities to those identified by the analysis of semantic categories (vocabulary of assessment, and that relating to proper names and family relations). Another identifies a more complex discourse which mixes both the evaluation discourse and that relating to its translation into a conclusion.

Three observations can be made about the way these psychological features of the Reports were constructed. Firstly, as noted, there was a focus on individual behaviours

and not family circumstance or the causes of those behaviours. As such, these were not problem-solving reports, which would be constructed with the aim of offering clinical input. Secondly, there was a high frequency of references to psychological testing without specific explanation to assist the court in interpreting the meaning of the tests and the results. Thirdly, in the reports, lengthy therapeutic intervention was a recommendation in many of them. Given the timeframe of the Public Law Outline is 26 weeks (albeit frequently exceeded in practice) the authors suggest this may create an impossible dilemma for judges. A question for future research may be whether a case will almost inevitably result in a child's removal in reports where a recommendation is made for therapeutic input beyond the 26-week timescale the courts can allow.

Two structural observations about the methodology of the reports emerged; firstly, there was a lack of standardization across the reports and there was a lack of transparency. In the small sample in the PRC-25 this posed problems for analysis of the reports as individual whole documents. Regardless of how the reports were organized, it was observed that there were frequent mentions of psychological tests, but there was no overt method of extracting how the conclusions in the reports had been drawn from the results of the tests, or discussion of their reliability. Instead, logical connectors were used to link events with narratives and the test results to generate the inferences and draw conclusions. Logical connectors also featured in what was observed to be a (re)construction of the lived experience of the subject(s) of the reports. Used to join two units of language, logical connectors attempt to justify the inferential links made in the reports. The marked use of logical connectors in the reports illustrates precisely why the reports may pose a problem for the court in interpreting their contents in the context of applying threshold tests to the appropriate standard of proof. To clarify this point, psychologists are permitted to comment on the ultimate issue. In the non-representative sample studied here, this did not happen, and the reports did not provide the level of clarity hoped for following the 2013 review Ministry of Justice and Family Justice Council.

This raises two structural issues for further examination: firstly, the question of focus on behaviours rather than situations merits further examination. Secondly, the process of (re)construction of the source material for the reports, including the verbal interviews, requires future analysis to investigate the process of (re)construction of the narratives and their entextualisation. Both these issues concern the level of clarity (or otherwise) in the protocol for reporting interviews, and how inferences can be drawn from them, including the methods of converting the spoken interviews into written data. Of note, from the reports, it is unclear how the experts have conducted the actual interview with the parents. No information is provided regarding the experts' preparation for the interview, how they carried out the interview itself (for example, following a predetermined set of questions), and significantly, how they went about making a record of the interview (for example, synchronously or asynchronously). It is known from linguistic research in forensic and legal contexts (Richardson *et al.* 2022; Andrus 2011; Bucholtz 2009; Fraser 2003) that interviews themselves as pieces of forensic evidence and the processes through which spoken discourse is captured in a written format is a fraught and complicated matter. In this case, the inferences that are made from these reported interviews may be particularly problematic in the conclusions that

psychologists make. This is certainly an aspect of the PRC-25 that requires further attention, combining linguistic and legal expertise.

Overall, the approaches used in this study were found to be complementary, providing overlapping findings which were triangulated to some extent. This demonstrates the synergy between the deductive and inductive elements of the study, confirming the initial hypothesis that this is a promising avenue for exploring expert evidence in a larger corpus. The combination of keyword analysis and hierarchical clustering analysis highlights important observations about salient language in the PRC-25, its main themes, and focusses, and areas for further analysis. Both approaches yielded promising results for the potential of corpus linguistics for developing a methodology to answer specific legal (and other) questions, as well as the potential to interrogate the data to generate new hypotheses. More can be established in terms of this type of report as a genre, including investigation of the register/genre distinctions, and the lexical bundles (phrases) highlighting conventions for the text type. Further detailed linguistic analysis of the reports on a representative sample using CL methods will provide clarity on both the issues raised by this study and will generate further insights into the language used in the reports. Further analysis of a representative sample of the reports will also illuminate the process of entextualisation behind their creation.

Conclusions

This paper demonstrates that the two approaches are complementary, providing overlapping findings, but also highlighting and mitigating mutual methodological and practical shortcomings which can be resolved during and after the creation of a larger and representative corpus. In combining the two approaches to maximise both the legal knowledge, which is vital to the contextualisation of linguistic data, and a principled linguistic approach to the analysis of large datasets, interesting indicative results were generated about the use of legal language in the PRC-25. This demonstrates the synergy and potential for corpus-assisted analysis of written and spoken legal materials.

The PRC-25 offers promising and indicative findings which justify scaling up to a larger study creating a large corpus of psychologists' reports and protocol for development across other areas of expert evidence in a wider range of legal proceedings. The synergy between the approaches is typified by the analytical journey of the terms *risk* and *harm* throughout this paper. Firstly, they were pre-selected legal terms of interest in the corpus. Concordance analysis enabled each use of these terms and what we already knew to be their related phrases (eg *risk of*, *significant harm*), amongst others, to be viewed in their original context. This provided a convenient platform for a legal-intuitive analysis to draw on. Secondly, collocation analysis also brought unexpected associations with these words into relief where they may have been otherwise missed by legal analysis alone. Lastly, keyword and hierarchical analysis designed to enable us to gain insights from the PRC-25 and its main themes and concerns, also identified *risk* and *harm* as salient terms in this corpus, thereby confirming our selection of them for more detailed analysis.

A limitation of previous research with expert evidence transcripts has been the difficulty of accessing the data. To overcome this limitation for future research, the next phase of this research is the preparation of a large and representative corpus for researcher re-use. Although it poses new challenges, the effort required for this

task is justified to enable a deeper and reliable consideration of this type of evidence. The raw reports in the database require considerable work to render them useable for research purposes and anonymisation of individual reports presents a labour-intensive task justifying an automated solution as part of the corpus creation process. A process such as encoding in XML-TEI is envisaged which will allow analysis with corpus tools which recognise this encoding (e.g Lüngen (2017)). Despite these challenges, once created, the corpus can be re-used efficiently to answer a variety of research questions.

Acknowledgements

Enquiries to l.devine@aston.ac.uk. The authors would like to thank Ms Poppy Cooke for her careful work creating the database from which the PRC-25 was extracted. We would also like to thank HHJ Stephen Wildblood QC for his insightful observations directing us towards this field of research. This work was funded by the Nuffield Foundation under grant number JUS/43090 for the project *Theorising the Care Cases Crisis*, and *Mind your Language* funded by Research England under an Expanding Excellence in England (E3) Award. Also, thanks to the reviewers for their helpful comments on an earlier draft.

Notes

[1] *Theorising the Care Cases Crisis* (Grant number JUS/43090) and *Mind your Languag*e, funded by Research England under an Expanding Excellence in England (E3) Award.

[2] The research has been approved by the President of the Family Division of England and Wales. Data has been accessed under a Cafcass Data Processing Agreement has received two ethical approvals from Cafcass and Aston University respectively.

References

Aldridge, (2010). Vulnerable witnesses in the criminal justice system. In M. Coulthard and A. Johnson, Eds., *The Routledge Handbook in Applied Linguistics*. Abingdon. Routledge.

Andrus, J. (2011). Beyond texts in context: Recontextualization and the co-production of texts and contexts in the legal discourse, excited utterance exception to hearsay. *Discourse & Society*, 22(2), 115–136.

Appell, A. (1997). Protecting children or punishing mothers: Gender, race, and class in the child protection system [an essay], 48 s. *C. L. Rev*, 577.

Bailey, A. (2019). *Discourses of dementia in the British press: a corpus linguistic approach*. University of Nottingham.

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.

Baker, P., Gabrielatos, C., Khosravinik, M., Krzyzanowski, M., McEnery, T. and Wodak, R. (2008). A useful methodological synergy? combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the uk press. *Discourse and Society*, 19, 273–306.

Biber, D., Conrad, S. and Reppen, R. (1998). *Corpus Linguistics. Investigating language and use*. Cambridge University Press.

Boholm, M. (2018). Risk and quantification: a linguistic study. *Risk Analysis*, 29(ue 6), 1243–1261.

Bondi, M. (2010). Perspectives on keywords and keyness. In M. Bondi and M. Scott, Eds., *Keyness in Texts*. Amsterdam: John Benjamins, 1–18.

Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press.

Brezina, V. and McEnery, T. (2015). Collocations in context. *International Journal of Corpus Linguistics*, 20(2), 139–173.

Brezina, V., Weill-Tessier, P. and McEnery, T. (2020). Lancsbox 5.x and 6.x. http://corpora.lancs.ac.uk/lancsbox.

British Psychological Society, (2016). Psychologists as expert witnesses in the family courts in england and wales, standards, competencies, and expectations. *Leicester. The British Psychological Society and the Family Justice Council.*

Broadhurst, K. and Mason, C. (2013a). Child removal as the gateway to further adversity: Birth mother accounts of the immediate and enduring collateral consequences of child removal. *Qualitative Social Work*, 19(1), 15–37.

Broadhurst, K. and Mason, C. (2013b). Maternal outcasts: raising the profile of women who are vulnerable to successive, compulsory removals of their children – a plea for preventative action. *Journal of Social Welfare and Family Law*, 35(3), 291–304.

Bucholtz, M. (2009). Captured on tape: Professional hearing and competing entextualizations in the criminal justice system. *Text & Talk*, 29, 503–23.

CAFCASS, (2022). Website: Annual summaries, table of LA rates of care application per 10,000 population. tab: Public care application of excel spreadsheet. https://www.cafcass.gov.uk/download/6250/.

Camasso, M. and Jagannathan, R. (2000). Modeling the reliability and predictive validity of risk assessment in child protective services. *Children and Youth Services Review*, 873–896.

Charteris-Black, J. and Seale, C. (2010). *Gender and the language of illness.* Hampshire: Palgrave Macmillan.

Cobley, C. and Lowe, N. (2009). Interpreting the threshold criteria under section 31(2) of the children act 1989.

Coulthard, M., Johnson, A. and Wright, D. (2016). *An introduction to forensic linguistics: Language in evidence.* London: Routledge, 2nd ed.

J. Cutting, Ed. (2007). *Vague Language Explored.* Basingstoke: Palgrave Macmillan.

Devine, L. (2017). *The Limits of State Power and Private Rights: Exploring the limits of safeguarding referrals and assessments.* London: Routledge.

Edleson, J. (1998). Responsible mothers and invisible men: Child protection in the case of adult domestic violence. *Journal of Interpersonal Violence*, 13(2), 294–298.

Egbert, J. and Baker, P. (2020). *Using Corpus Methods to Triangulate Linguistic Analysis.* New York: Routledge.

Everitt, B., Landau, S., Leese, M. and Stahl, D. (2011). *Cluster Analysis.* Wiley, 5th ed.

Evison, J. (2010). What are the basics of analysing a corpus? In A. O'Keeffe and M. McCarthy, Eds., *The Routledge Handbook of Corpus Linguistics.* Oxon: Routledge, 122–135.

EWCA, (1996). Re M & R (Child Abuse: Evidence) [1996] EWCA Civ 1317; [1996] 2 FLR. 195.

Fraser, H. (2003). Issues in transcription: Factors affecting the reliability of transcripts as evidence in legal cases. *International Journal of Speech, Language and the Law—Forensic Linguistics*, 10, 203–26.

Gillingham, P. (2006). Risk assessment in child protection: Problem rather than solution? *Australian Social Work*, 59(1), 86–98.

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, 2nd ed., 589–601.

Goldberg, L. (1968). Simple models or simple processes? some research on clinical judgments. *American Psychologist*, 23(7), 483–496.

Gries, S. (2013). 50-something years of work on collocations. what is or should be next? *International Journal of Corpus Linguistics*, 18(ue 1), 137–166.

HM Government, (2018). Working together to safeguard children. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/942454/Working_together_to_safeguard_children_inter_agency_guidance.pdf.

Hunston, S. (2010). How can a corpus be used to explore patterns? In A. O'Keeffe and M. McCarthy, Eds., *The Routledge Handbook of Corpus Linguistics*. Oxon: Routledge, 152–166.

Ireland, J. (2012). *Evaluating Expert Witness Psychological Reports: Exploring Quality*. University of Central Lancashire. http://www.fixcas.com/news/2012/IrelandJane.pdf.

KB, (1943). Hollington v hewthorn [1943] kb 587, para 595.

Law Commission, (2009). The admissibility of expert evidence in criminal proceedings in England and Wales, cp190. https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2015/03/cp190_Expert_Evidence_Consultation.pdf.

Law Commission, (2011). Expert evidence in criminal proceedings report, LC325. https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2015/03/lc325_Expert_Evidence_Report.pdf.

Lebart, L., Salem, A. and Berry, L. (1997). *Exploring Textual Data*. Springer Science & Business Media.

Li, S. (2017). A corpus-based study of vague language in legislative texts: Strategic use of vague terms. *English for Specific Purposes*, 45, 98–109.

Lindley, B. (1994). *Families in Court: A qualitative study of families' experiences of the court process in care and supervision proceedings under the Children Act 1989*. London: Family Rights Group.

Loubère, L. and Ratinaud, P. (2014). Documentation iramuteq 0.6 alpha 3 version 0.1., iramuteq. http://iramuteq.org/documentation/fichiers/documentation_19_02_2014.pdf.

Lüngen, H. (2017). Anonymisation of the Dortmund chat corpus 2.1. In E. Stemle and C. Wigham, Eds., *Proceedings of the 5th Conference on CMC and Social Media Corpora for the Humanities (cmccorpora17)*, 21–24, https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/6639/file/Luengen_Beisswenger_Herzberg_Pichler_Anonymisation_2017.pdf: Eurac Research.

Marchi, A. (2022). Corpus linguistics in the study of news media. In A. O'Keeffe and M. McCarthy, Eds., *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 2nd ed., 576–588.

Ministry of Justice, (2020). Family procedure rules. part 25. experts and assessors. https://www.justice.gov.uk/courts/procedure-rules/family/parts/part_25. Last updated 6th.

Ministry of Justice, (2021a). Practice direction 12a - care, supervision and other part 4 proceedings: Guide to case management, last updated.

Ministry of Justice, (2021b). Practice direction 25a – interim injunctions. https://www.justice.gov.uk/courts/procedure-rules/civil/rules/part25/pd_part25a.

Ministry of Justice and Family Justice Council, (2013). *Response: Standards for expert witnesses in Children's Proceedings in the family courts*. London: HM Government.

Munby, S. J. (2011). President of the family division. https://www.judiciary.uk/announcements/tighter-rules-introduced-on-expert-evidence-family-cases/.

Munro, E. (2010). Conflating risks: Implications for accurate risk prediction in child welfare services. *Health, Risk & Society*, 12(2), 119–130,.

Møller Jørgensen, A., Webb, C., Keddell, E. and Ballantyne, N. (2021). Three roads to Rome? comparative policy analysis of predictive tools in child protection services in aotearoa new zealand, england, & denmark, nordic social work research.

Partington, A. (1998). *Patterns and meanings*. Amsterdam: John Benjamins.

Parton, P. (2011). Child protection and safeguarding in England: Changing and competing conceptions of risk and their implications for social work. *The British Journal of Social Work*, 41(ue 5), 854–875.

Pojanapunya, P. and Todd, R. (2016). Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis'. In *Corpus Linguistics and Linguistic Theory [Preprint*.

Rayson, P. (2015). Computational tools and methods for corpus compilation and analysis. In D. Biber and R. Reppen, Eds., *The Cambridge Handbook of Corpus Linguistics*. Cambridge: Cambridge University Press, 32–49.

Redhill, J. and Roe, A. (2021). What do we know about children in the family justice system? nuffield family justice observatory. https://www.nuffieldfjo.org.uk/wp-content/uploads/2021/09/nfjo_infoG_explanatory_note_2021206.pdf.

Regehr, C., Bogo, M., Shlonsky, A. and LeBlanc, V. (2010). Confidence and professional judgment in assessing children's risk of abuse. *Research on Social Work Practice*, 20(6), 621–628.

Richardson, E., Haworth, K. and Deamer, F. (2022). For the record: Questioning transcription processes in legal contexts. *Applied Linguistics*, 1–22.

Scott, M. (2010). Problems in investigating keyness, or clearing the undergrowth and marking out trails. In M. Bondi and M. Scott, Eds., *Keyness in Texts*. Amsterdam: John Benjamins, 43–58.

Scott, M. and Tribble, C. (2006). *Textual Patterns: Key words and corpus analysis in language education*. Amsterdam: John Benjamins.

Sinclair, J. (1999). Concordance tasks. the tuscan word centre. http://www.twc.it/happen.html.

Smeeton, J. (2020). A murky business": A phenomenological ontology of risk in child protection social work. *Qualitative Social Work*, 19(2), 284–300.

Stubbs, M. (2010). Three concepts of keywords. In M. Bondi and M. Scott, Eds., *Keyness in Texts*. Amsterdam: John Benjamins, 19–42.

Taylor, C. (2022). Corpus linguistics in the study of political discourse: recent directions. In A. O'Keeffe and M. McCarthy, Eds., *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 2nd ed., 602–614.

Tkacukova, T. (2020). Changing landscape of advice provision: Online forums and social media run by mckenzie friends. *Child and Family Law Quarterly*, 4, 397–420.

UK Parliament, (1989). Children act 1989 c.41.

UKSC, (2013). Re B (Care Proceedings: Appeal) [2013] UKSC 33.

Van Dijk, T. (2000). On the analysis of parliamentary debates on immigration. In M. Reisigl and R. Wodak, Eds., *The semiotics of racism. Approaches to Critical Discourse Analysis*. 85–104.

Xiao, R. (2015). Collocation. In D. Biber, Ed., *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press, 106–124.

Native Dialect Influence Detection (NDID): Differentiating between Mexican and Peninsular L1 Spanish in L2 English

Andrea Mojedano Batel, Mitchell Abrams & Piotr Pęzik

Aston Institute for Forensic Linguistics, Aston University, UK

https://doi.org/10.21747/21833745/lanlaw/9_1a6

Abstract. *The current investigation addresses a vital lacuna in forensic authorship studies, and more concretely, in Native Language Influence Detection (NLID) research: narrowing down a speaker's native dialect instead of only their native language (L1), which might not be enough when carrying out sociolinguistic profiling tasks. Native Dialect Influence Detection (NDID), the focus of our study, can thus greatly aid at the investigative level. We approach this topic by providing a comprehensive analysis of linguistic features that serve to identify two non-contact dialects of L1 Spanish (i.e., Mexican and Peninsular varieties) when dealing with data written in L2 English, which come from Tripadvisor. Our main aim is to investigate if an author's L2 features can point to their L1 native dialect, rather than only to their native language. Findings point to L1 dialectal transfer of punctuation signs, adjectives of affect, and intensifiers: these linguistic features, even when expressed in an L2, show a culturally bound use. Additionally, we implemented an automatic classifier that achieved an accuracy of 69% in categorizing test data, using only linguistic features that have explanatory power and can aid linguistic theory. This is key for explainability in the forensic context, which Native Language Identification (NLI) studies tend to neglect (Kingston 2019). Results show that L1 Spanish dialects can be differentiated by analyzing L2 English text, pointing to NDID as a fertile approach for narrowing down candidate L1 dialects of a language when analyzing L2 data.*

Keywords: *Native dialect influence detection, Native language influence detection, Authorship analysis, Language variety identification, Spanish.*

Resumo. *A investigação atual aborda uma lacuna essencial nos estudos de autoria forense, mais concretamente na investigação sobre Deteção de Influência da Língua Materna (NLID): afunilar a análise do dialeto materno de um falante em vez de focar apenas a sua língua materna (L1), o que pode não ser suficiente quando se realiza tarefas de caracterização sociolinguística. A Deteção da Influência da Língua Materna (NDID), o enfoque do nosso estudo, poderá assim auxiliar*

significativamente a investigação. Abordamos este tema realizando uma análise abrangente das características linguísticas usadas para identificar dois dialetos sem contacto do espanhol L1 (isto é, variedades mexicanas e peninsulares) em dados escritos em inglês L2 provenientes do Tripadvisor. O nosso principal objetivo é investigar se as características L2 de um autor podem apontar para o seu dialeto nativo L1, e não apenas para a sua língua materna. Os resultados apontam para a transferência dialetal L1 de pontuação, adjetivos afetivos e intensificadores. Estas características linguísticas, mesmo quando expressas numa L2, revelam uma utilização culturalmente marcada. Adicionalmente, implementámos um classificador automático que alcançou uma precisão de 69% na categorização dos dados do teste utilizando apenas características linguísticas que possuem capacidade explicativa e podem contribuir para fundamentar a teoria linguística, fundamental para fornecer fundamentações em contexto forense, o que os estudos de Identificação Língua Materna (NLI) tendem a negligenciar (Kingston 2019). Os resultados mostram que os dialetos espanhóis L1 podem ser diferenciados através da análise de texto L2 em inglês, apontando a NDID como uma abordagem fértil para reduzir os dialetos candidatos L1 de uma língua ao analisar os dados L2.

Palavras-chave: *Deteção de influência de dialeto nativo, Deteção de influência de língua materna, Análise de autoria, Identificação de variedade linguística, Espanhol.*

Native Dialect Influence Detection (NDID): Differentiating between Mexican and Peninsular L1 Spanish in L2 English

This study provides a comprehensive description of linguistic features that serve to identify non-contact dialects of a native language (L1) when dealing with second language (L2) data, carrying out bottom-up quantitative and qualitative analyses of token n-grams and part-of-speech (POS) n-grams, leaving aside other feature analyses for future studies. For this purpose, we have collected a corpus of texts written in L2 English by L1 Mexican Spanish (MxSp) and L1 Peninsular Spanish (PenSp) authors on the Tripadvisor website.[1] The study's main aim is to investigate if an author's L2 features can point to their L1 native dialect, rather than only to their native language. We term this specific subdomain of Native Language Influence Detection (NLID) *Native Dialect Influence Detection* (*NDID*). The current study draws on research from the fields of Second Language Acquisition (SLA), forensic linguistics, contact linguistics, sociolinguistics, and computational linguistics.

Addressing NDID lacunae will aid the forensic linguist in providing "evidence for criminal and civil investigations and courtroom disputes," which is one of the main aims of the field (Coulthard *et al.* 2010: 529). The overarching goal of NDID studies is to estimate group belonging rather than to identify individuals. In the specific case of Spanish, with 483 million native speakers worldwide (Fernández Vítores 2020), knowing that a speaker's L1 is Spanish when dealing with L2 data might not be enough; NDID helps narrow down the possible native and/or dominant regional dialect of a speaker. This type of sociolinguistic profiling, while unlikely to be admissible as evidence in the UK judicial and court systems (Grant 2008; Perkins and Grant 2013), can greatly aid at the investigative level, benefitting law enforcement agencies or organizations, as well as intelligence agencies dealing with non-native speakers (Perkins 2015). Casework in NLID is seldom published due to the very nature of intelligence work (Perkins 2015), but it is

possible to conduct NLID research taking into consideration intelligence applications by collaborating with relevant and interested agencies and departments (e.g., Grant *et al.* (2010)), and the same could be said of NDID as a sub-domain of NLID. Spanish, the L1 explored in this study, is the third most used language on the Web as of March 2020, with almost 364 million Spanish speakers using the Internet, which corresponds to 7.9% of all Internet users (http://www.internetworldstats.com). Hence, there is an abundance of Spanish data on the web; it must come as no surprise that many of these speakers might also express themselves in other languages (e.g., English), and these data should be explored from a forensic linguistic viewpoint.

The present research lies within the framework of forensic authorship analysis, and more concretely, it takes a Native Language Influence Detection (NLID) approach. NLID seeks to reveal an author's native language (L1) from common interlanguage phenomena that may occur in any non-native language they write in, which can be their second language, but also their third, fourth, etc. We will refer to this non-native language output in the present study as L2 in order to simplify matters; however, forensic linguists must consider that a speaker can have more than one native language, as the majority of the planet's population is bilingual (Thomason 2001).

Within the field of computational linguistics, the term *Native Language Identification* (*NLI*) is used, referring to an often wholly automatic computational method for the identification of the language or languages that likely influenced a text or group of texts, based in machine learning and data mining, and introduced by Koppel *et al.* (2005). Since then, a variety of classifier approaches and feature sets have been proposed, the most popular including word n-grams, part-of-speech (POS) n-grams, character n-grams, function words, dependency features, and spelling features. These were all features exploited in the 2013 NLI Shared Task (Tetreault *et al.* 2013). Additionally, production rules were also a successful feature in Wong and Dras (2011).

A frequent problem with NLI studies is that they tend to neglect the features used to classify texts (Kingston 2019). While features for these computational tasks are leveraged mainly to achieve high classification accuracy, they are key for explainability in the forensic context. Some studies, however, do provide a light analysis of linguistically motivated features in NLI tasks (e.g., Koppel *et al.* (2005); Jiang *et al.* (2014); Goldin *et al.* (2018)).

In the following section, we will address relevant studies in the fields of computational linguistics and forensic linguistics in order to provide the reader with information on what has been done already and what still needs to be tackled in the field of NDID.

Previous Accounts of Dialect Identification, with an Emphasis on Spanish

Research dealing with dialect identification lies within the interface between computational linguistics (more specifically, within Natural Language Processing, or NLP) and authorship profiling. The term *Language Variety Identification* (*LVI*) refers to "labelling the text in a native language (e.g., Spanish, Portuguese, English) with its specific variation (e.g., Argentina, Chile, Mexico, Peru, Spain; Brazil, Portugal; UK, US)" (Rangel *et al.* 2016: 1). LVI investigates a dialect of a language written in an L1 and therefore does not directly pertain to NLID studies, yet shares with NDID research the commonality of focusing on the differentiation of dialects of a particular language.

LVI studies point to certain automated features being useful to discern among dialects. Sadat *et al.* (2014) used character n-gram features to discriminate between six different Arabic dialects, obtaining accuracies between 70%-80%. Zampieri and Gebrekidan-Gebre (2012) looked at automatic classification of written Brazilian and European Portuguese, using a character n-gram model and a word n-gram model; their results showed an accuracy of over 90%. Maier and Gómez-Rodríguez (2014) investigated LVI in Spanish tweets from Argentina, Chile, Colombia, Mexico, and Spain. They obtained 60%-70% accuracy by applying language modeling techniques that combined four types of features: character n-gram frequency profiles, character n-gram language models, Lempel-Ziv-Welch compression, and syllable-based language profiles. Rangel *et al.* (2016) used Low Dimensionality Representation (LDR) to differentiate between Argentinian, Chilean, Mexican, Peruvian, and Peninsular Spanish. Instead of selecting the most frequent n-grams, Rangel *et al.* (2016) assigned higher weights to the most discriminative words in each dialect. Their results showed that Peninsular Spanish was the easiest dialect to discriminate and that Latin American varieties were closer to each other than to Peninsular Spanish, and therefore, it was more difficult to differentiate among them. Nevertheless, the highest precisions were attained for Mexico and Peru. While LVI studies help us gauge what kinds of computational features can aid in the task of differentiating among dialects, a frequent problem with these studies, much like with NLI studies, is that they generally neglect the linguistic features used to classify texts and that results from character n-gram analysis cannot usually be explained from a linguistic viewpoint.

The only study, to the best of our knowledge, that has tackled identification of a speaker's native dialect rather than simply their native language by looking at L2 data is Kingston (2019), who analyzed L2 English data from L1 speakers of three French dialects: Metropolitan (i.e., from France), Canadian, and Maghrebi. Because the primary goal of Kingston's study was to "designate potentially distinguishing features of native speakers of different dialects of the same language" (p.26), comparing three dialects where two of them were in contact situations (Canadian French with English and Maghrebi French with Arabic and Tamazight) made it potentially easier to identify dialects vis-a-vis identifying non-contact dialects of a language, which is the aim of our study.

Kingston (2019) secondary aim was to detect the extent to which machine learning classification mirrored the human analyst's findings. She used a decision tree classifier (J48) in the Waikato Environment for Knowledge Analysis (WEKA). Kingston concluded that purely computational methods could achieve very good classification accuracy, but that there was a need for a more inclusive approach taking into account linguistic typology, sociolinguistic data, and semantic and syntactic analysis. The present investigation seeks to address some of these lacunae.

In the present study we chose Peninsular and Mexican Spanish dialects for data analysis because previous investigations have shown that Latin American dialects are closer to each other than to Peninsular (Lipski 2012; Rangel *et al.* 2016). If our results show that we cannot tell the difference between Spaniards' and Mexicans' L2 English output, it might prove more difficult to differentiate English L2 outputs among Latin American dialects. Additionally, Mexican Spanish is the Spanish dialect with the largest number of native speakers in the world (121 million speakers, according to Fernández Vítores (2020)). Being able to examine if there are any differences between Peninsular and Mex-

ican Spanish dialects with regard to speakers' L2 English output will help pave the way for further studies considering other Spanish dialects.

Preference for British or American English as L2 Input and Output

While, to the best of our knowledge, there is no official information as to which dialects of English Mexican and Peninsular Spanish speakers are exposed to, findings from (Caraker 2016) show that several teachers from Central Spain believed Peninsular Spanish students to be more exposed to British English (BrE). Meanwhile, Despagne (2010) reported that in Mexico, television programs are more than 50% in English and the dialect of English that these media mainly use is the American dialect. This points to Mexican Spanish speakers possibly having more influence from American English (AmE) than British English in their L2 English input and output.

Sue Garton (personal communication, July, 2020) believes that most European countries favor standard British English (if there is such a thing, as she notes), while Latin America would be more oriented toward a US variety. However, Garton does not think there are many official policies, and much is likely to depend on the coursebook that is used and the dialect that the teacher speaks. The difference between American and British English when it comes to L2 English instruction is vital, as it could possibly bring different grammatical, lexical, and orthographical choices that we need to consider when carrying out our analysis.

Data and Methodology

This section offers a detailed account of the corpora, the extraction of adequate tokens, and the methodology used in applying both a computational linguistics approach and a forensic linguistics approach to account for dialectal differences in the L2 English output of L1 Peninsular and Mexican Spanish speakers.

The data we present in this study are corpus-based. They come from Tripadvisor, an open internet source, and the genre pertains to computer-mediated communication (CMC). Features that characterize grammar in electronic communication vary systematically across languages, contexts, users, and technological modes (Herring 2012). Herring (2012) notes that non-standard orthography is common in CMC, where users can be lax about orthographic norms. Bieswanger (2008) demonstrated through a systematic comparison of English and German text messages (SMS) that Britons and Germans favor different shortening strategies when texting and that these strategies are used in different proportions. In terms of our Tripadvisor data, by the very nature of the website, the entries are evaluative (which restaurants, hotels, and locations authors liked or disliked), and the entries' main function is communicative, with suggestions for, and requests to, other users. Our findings can thus be transferable to forensic linguistic contexts, because forensic texts can be evaluative and often display requests and/or demands.

We compiled four corpora, listed below:

- A corpus with training L2 English data produced by L1 Mexican Spanish speakers (word count: 37,500, number of entries: 514, number of authors: 38, average entry length: 73 words)
- A corpus with training L2 English data produced by L1 Peninsular Spanish speakers (word count: 40,602, number of entries: 504, number of authors: 37, average entry length: 81 words)

- A corpus with test L2 English data produced by L1 Mexican Spanish speakers (word count: 6,481, number of entries: 53, number of authors: 6, average entry length: 122 words)
- A corpus with test L2 English data produced by L1 Peninsular Spanish speakers (word count: 4,797, number of entries: 50, number of authors: 9, average entry length: 96 words)

We are aware that corpora sizes are small, but it was very difficult to make them larger due to their specialized nature: further studies could build upon this initial research with data from other sources. To ensure that the authors were native speakers of their respective dialects, we first needed to find posts asserting their linguistic identity, such as "I am / I'm Mexican / Spanish / from Mexico / from Spain," "Being from Mexico / Spain," "I live in Mexico / Spain," etcetera. Moreover, Tripadvisor users had to have a geolocator in their account so that we could be more certain they lived in either Mexico or Spain; authors who said that they were from either of these two countries but were geolocated in a different country were discarded because of language contact and language identity issues. Additionally, because we were examining L2 English data, we had to make sure that the participants spoke Spanish as an L1. To do this, we checked their posts to make sure they had written at least one of them in Spanish and that in said post(s), they showed native-like command of Spanish.[2]

For each of the two training corpora, we took the most recent 25 entries per author when there were more than 25 entries for that specific person; if there were less than 25 entries, we took them all; the minimum number of entries per author in the training corpora was one. For each of the two test corpora, we took the five most recent entries per author when there were more than five entries for that specific person; if there were less than five entries, we took them all; the minimum number of entries per author in the training corpora was one.

While we chose texts from all over Mexico, we only chose texts from areas in Spain where Spanish is not in contact with other languages (i.e., Central and Southern Spain) with the purpose of narrowing dialectal variation. It can be argued that Mexican Spanish is in contact with indigenous languages all over the Mexican territory: while this is certainly true, only 6% of the population (approximately six million people) speak an indigenous language, according to the Commission for the Development of Indigenous Peoples (CDI) and the National Institute of Indigenous Languages (INALI). Thus, it is improbable that entries from Mexican Spanish speaking authors come from a Mexican contact dialect.

Before entries in the target and reference corpora were processed for tokenization, tagging, and analysis, the data went through a pre-processing step for cleaning and collecting user post frequency. We kept the raw text for analysis since extra-linguistic conventions of the written online medium (namely, parentheses, quotation marks, ellipses, punctuation, emoticons, and website links) can potentially be influenced by a person's native dialect. In this same vein, we refrained from normalizing the language through lemmatizing, correcting misspellings, and removing capitalization. One exception we made was to replace various instances of accented *í* with *i* for English words, as these were attributed to keyboard mistakes rather than an individual's language, and including these characters in the data would have significantly altered token frequencies. In the final pre-processing step, we collected user post counts to check for feature distri-

bution, so that a feature under analysis was not over-attributed to any one individual in the target corpus.

As a starting point in investigating native dialect influence, we created a list of token n-grams and part-of-speech (POS) n-grams ranked by their keyness. The term *token* in our study refers to words and non-words in our corpora (e.g., word forms, punctuation, digits, abbreviations, etc.) and serves as the basis of our analysis. Even parts-of-speech were leveraged to reveal token realizations and shallow syntactic structures. An English language tokenizer was applied internally through SketchEngine (Kilgarriff *et al.* 2014) before n-gram analysis. An *n-gram* is a contiguous sequence of N words or tokens. For the present study in particular, we collected one to six grams. With these units for analysis, we ranked them by keyness as a way of operationalizing consistency and distinctiveness in the target corpora.

We approached keyness from a corpus linguistic perspective: tokens, POS sequences, and grammatical patterns that are key occur in texts with outstanding frequency as compared with a reference corpus. We believe that keywords are not key by themselves, but because they are frequently used in particular lexical combinations or grammatical patterns, which makes them attractive units of analysis to explore NDID tasks. This study takes two approaches to identify keywords in texts, namely, a keyness metric through SketchEngine software (Kilgarriff *et al.* 2014), and Burrows' Zeta (Burrows 2007) to assess the dispersion of style markers used by authors represented in the corpus. Results were first collected separately and then compared later for analysis, confirming that one approach had not overlooked any important patterns. The keyness metric in SketchEngine software uses the simple maths method (Kilgarriff, 2009) to show how many times a given word is more frequent in a target corpus (corpus 1) than in a reference corpus (corpus 2). This is a simple metric, yet effective enough for our relatively small dataset, comprised of posts of a relatively similar size and of a consistent genre. Alongside keyness scores, we extracted the frequencies to see how much more frequent an n-gram was in one corpus over another. Burrow's Zeta, on the other hand, was calculated as the difference between the proportion of n-word segments in which a given marker such as a word n-gram or POS-gram was found in the Mexican and Peninsular Spanish corpora. This approach was meant to operationalize the more abstract notion of style marker consistency in our study.

The main motivation for extracting token and POS n-grams as linguistic features for analysis, and excluding others (e.g., character n-grams), is that they have explanatory power which can aid both linguistic theory and forensic research. Token n-grams in particular are ideal for this study and reveal noteworthy patterns. Additionally, in computational research, token n-grams have been proven to achieve high accuracy in related NLID and LVI tasks. In such analyses, data scarcity (notably with our smaller target corpus size) becomes a problem, because linguistic items appear in low frequencies, especially when focusing on higher token n-gram sequences (such as bigrams and trigrams). However, even with this limitation in mind, it is possible to identify similarities and differences between observed linguistic patterns in our corpora. We also handled this sparsity issue by leveraging the hidden categories of part-of-speech (POS) tags to reveal more distinct patterns. On the level of morphosyntactic annotation, we analyzed recurrent POS n-gram patterns, which allowed us to capture grammatical patterns that were not directly observable. For our data in particular, we used two methods

for tagging the corpora; one that used the built-in SketchEngine software part-of-speech tagger with the English Penn Treebank tagset, and another that utilizes the POS tagger available in the Spacy (Honnibal and Montani 2015).

Results and Analysis

In the present section, we examine our results, taking a mixed-methods approach to their analysis. First, we discuss token n-grams, divided into intensifiers, adjectives of affect, quantifiers, contracted forms, and punctuation marks. We then discuss POS n-grams, and more concretely, bigrams and trigrams. Finally, we offer a description and analysis of our classification task, providing researchers with a starting point for NDID classification. When reporting descriptive statistics, we include both raw and relative frequencies (unless otherwise stated) for token n-grams and raw frequencies for POS n-grams. Relative frequencies are provided through percentages, as our corpora are small in size.

Token N-grams

Intensifiers

Intensifiers can be classified in intensives and downtoners, and both for AmE and BrE, the most frequent collocation pattern for intensifiers (72%) is with adjectival heads (Bäcklund 1973).[3] In our study, raw and relative frequencies of intensifiers shed light on cross-dialectal semantic and pragma-linguistic differences. In this subsection we discuss the difference in frequency and use of intensives *quite* and *really* and then examine the downtoner *a (little) bit*.

The intensive *quite*. In line with what was expected of Peninsular Spanish speakers, the adverb *quite*, more frequently used in BrE than in AmE (395.14 times per million, according to BNC, vs. 182.71 times per million, according to COCA, respectively), appears in the Peninsular Spanish corpus almost thrice as often than in the Mexican Spanish corpus (N = 58, 0.14% vs. N = 21, 0.05%, respectively).[4] A log-likelihood (LL) test of significance indicated that the difference in use of the intensive *quite* between corpora was significant ($p < 0.05$). Both sets of authors used this adverb most frequently in partially lexically-filled constructions consisting of [be + quite + AdjP/DP], as in (1); yet, the difference in frequency of use can probably be attributed to the assumption that Spaniards are more exposed to a British variety of English, whereas Mexicans generally learn an American English variety.

(1a) *since the broschure [sic]* ***was quite complicated*** *to understand* (PenSp corpus)
(1b) *Marbella* ***is quite a big place*** (PenSp corpus)
(1c) *getting there* ***is quite easy*** (MxSp corpus)

The intensive *really*. In line with what was expected of Mexican Spanish authors, the adverb *really*, more frequently used in AmE than in BrE (902.34 times per million vs. 458.10 times per million, respectively), appears in the Mexican Spanish corpus almost twice as often as in the Peninsular Spanish corpus (151 times vs. 89 times, respectively), mirroring AmE rates of use. A log-likelihood (LL) test of significance indicated that the difference in use of the intensive *really* between corpora was significant ($p < 0.05$). Moreover, while *really* modifies adjectives in a similar way in both MxSp and PenSp, Mexicans also use *really* to modify adverbs, something seldom done by Spaniards in our corpus (N = 9, 0.02% vs. N = 4, >0.01%, respectively), as shown in (2).

(2) *Everything was so amazing and delicious, and served* ***really fast*** (MxSp corpus)

The downtoner *a (little) bit.* Spanish makes ample use of the diminutive morpheme *-ito/-ita*, whose role extends beyond the expression of small size, giving emotional tone to words (Travis 2004). Thus, Travis (2004) argues, not only the lexicon, but also the morphosyntax of a language reflects the cultural values of its speakers. Furthermore, she notes that extensive use of the diminutive is evidence of a cultural value associated with good feelings, a claim that Wierzbicka (1984, 1992) has also made with regard to Russian and Polish. Concerning Spanish dialects, Company Company (2002) has shown that there is wider use of the diminutive in Mexican Spanish compared to Peninsular Spanish, both in terms of overall frequency and range of pragmatic functions. A similar result is provided by Reynoso Noverón (2001), who found that in a corpus of written and oral narratives, Mexicans used the diminutive to encode small size just 28% of the time, whereas Spaniards used it for this same reason 58% of the time: these results show that Peninsular participants in Reynoso Noverón's study tended to use the diminutive morpheme chiefly for encoding small size, whereas Mexican participants generally used it in a wider range of pragmatic functions, and not predominantly for encoding small size.

Standard Spanish makes use of the quantifier *poco* 'a bit' to convey the idea of a small amount. *Poco* can appear with the diminutive morpheme *ito/ita*, i.e., *poquito/poquita*; these lexemes are attested in both Mexican and Peninsular Spanish. Moreover, Mexican Spanish speakers use the quantifiers *tantito/tantita* 'a bit / a little bit' to express the idea of a small amount.

English doesn't encode the diminutive as a morpheme, and Spanish speakers have to make use of other means to express small size in their L2 English output. When examining the construction *a (little) bit* in our corpora —where parentheses indicate optionality— a first observation is that both dialects show, for the most part, similar frequencies (N = 45, 0.11% for Peninsular Spanish and N = 31, 0.08% for Mexican Spanish) and distributional patterns (a log-likelihood (LL) test of significance indicated that the difference in use of the construction *a (little) bit* was not significant). There is one important difference, however: [*be + a (little) bit*] constructions, namely, *a (little) bit* constructions which co-appear with the copula *to be*. This is the most frequent construction in which *a (little) bit* appears in both corpora, but its behavior is quite different depending on the L1 Spanish dialect of the speaker.

The construction [*be + a (little) bit*] shows a raw frequency of six instances in the Mexican corpus and of 24 instances in the Peninsular Spanish corpus. In both, speakers generally use the adverbial form *a (little) bit* to try to downplay a negative trait of something that they're describing, as in (3).

(3a) *The service* ***was a bit slow*** (MxSp corpus)
(3b) *The décor* ***could be a little bit more modern*** (MxSp corpus)
(3c) *january [sic]* ***can be a bit rainy*** (PenSp corpus)
(3d) *The shower* ***is a little bit old*** *but it goes together with the style of the parador [sic].* (PenSp corpus)

When analyzing the differences in use of the [*be + a (little) bit*] construction in both corpora, we found that, similarly to what takes place in Mexican and Peninsular Spanish in general, there is wider use of a diminutive form (in this case, *little*) in the Mexican

Spanish L2 English output compared to the Peninsular Spanish L2 English output (N = 4, 66.6% of all [*be + a (little) bit*] constructions in MxSp vs. N = 3, 12.5% of all [*be + a (little) bit*] constructions in PenSp). We tested for the difference in the frequency of the [*be + a little bit*] construction between corpora with a log-likelihood (LL) test of significance: the LL is 4.65, showing the difference to be significant at the $p < 0.05$ level (or at the 95% level). That is, although Spaniards produce [*be + a (little) bit*] more than Mexicans, they overwhelmingly do so in its [*be + a bit*] form, while Mexicans mainly produce the [*be + a little bit*] construction to convey the same idea of downplaying a negative trait.

Moreover, when analyzing the form *a little bit* by itself (that is, without the copula *to be*), the use of this construction followed patterns found in L1 Spanish: there was a wider use of the construction with the diminutive form *little* in the Mexican Spanish corpus (N = 12) compared to the Peninsular Spanish corpus (N = 8) in terms of range of pragmatic functions: Mexicans, apart from using it to downplay negative emotions —the only function that appeared in Peninsular Spanish use—, also used it once to downplay requests (4a), once to encode positive feelings (4b), and twice to encode quantity, as in (4c).

(4a) *Hi guys* ***I'm looking for a little bit of information*** *regarding a trip me and a few friends are planning* (MxSp corpus)
(4b) *A* ***little bit of everything****: Tulum & Xel-ha* (MxSp corpus)
(4c) *You do need to know* ***a little bit of history*** *of France* (MxSp corpus)

Adjectives of Affect: *Good, Great, Excellent,* and *Amazing*

When examining the lemma *good* (that is, in its different forms *good*, *better*, *best*) we see that it is more frequent in the Mexican Spanish corpus than in the Peninsular Spanish one (N = 392, 1.05% vs. N = 355, 0.87%). We tested for the difference in the frequency of the lemma *good* between corpora with a log-likelihood (LL) test of significance: the LL is 5.95, showing the difference to be significant at the $p < 0.05$ level (or at the 95% level). In example (5) below, we can gauge Mexican authors' use of the adjective *best* in context.

(5a) *This was by far our* ***best meal*** *in Hanoi* (MxSp corpus)
(5b) *This is truly one of the* ***best beaches*** *in Mexico* (MxSp corpus)
(5c) ***Best dinner*** *in town!* (MxSp corpus)

The adjective *best*, like other adjectives we discuss in this section, is a stance marker of affect. Precht (2000, 2003a,b) showed that stance expression is systematically different across cultures and contended that we tend to identify and stereotype people based on their stance use. As Precht (2003a) notes, there has been little research carried out regarding dialectal differences in stance, at least in English. To the best of our knowledge, this seems to be the case for Spanish varieties as well. Her (2003a) study comparing BrE and AmE showed that both varieties had a very similar conversational stance across semantic categories; however, British speakers displayed lower frequencies than Americans for affect markers that expressed emotion and for emphatics, while having higher frequencies for modal verbs. From these results, it appears that Americans directly express emotion more frequently than Britons.

The same could be said for Mexicans vis-a-vis Spaniards in terms of their stance expression: Mexican speakers seem to directly express their emotion, at least in their L2 English output, more often than Peninsular Spanish speakers. Besides producing the lemma *good* in a variety of forms more often than their Spaniard counterparts, Mexicans

also show a significantly higher use ($p < 0.05$) of the following adjectives of affect: *great* (N = 207, 0.55% vs. N = 132, 0.33%), *excellent* (N = 73, 0.19% vs. N = 37, 0.09%), and *amazing* (N = 51, 0.13% vs. N = 36, 0.09%). Figure 1 provides a visual breakdown of the difference in frequencies of adjectives of affect between corpora.

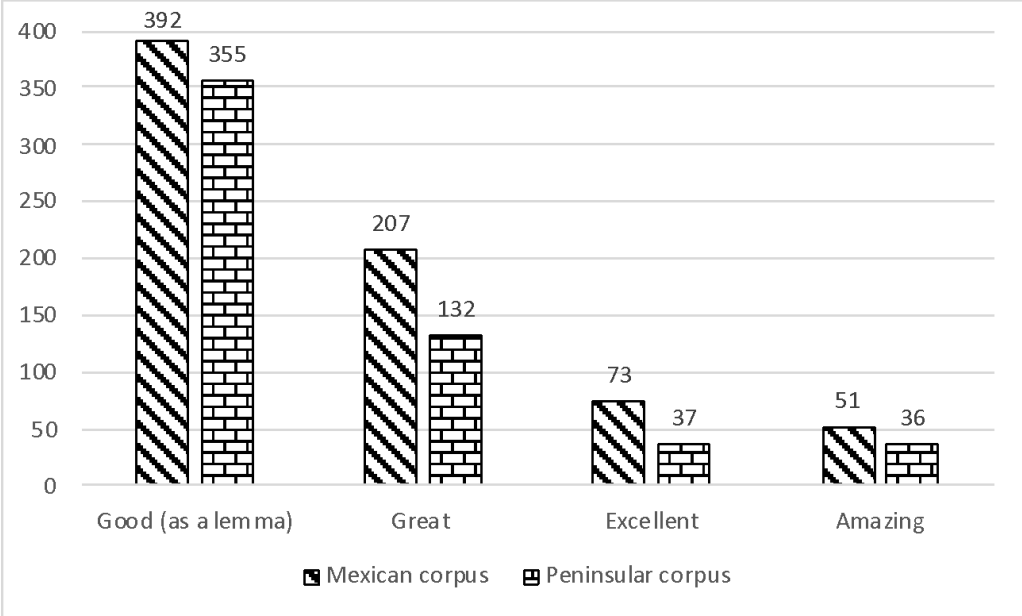

Figure 1. Frequency of Use of Adjectives of Affect by Corpus)

Example (6) below shows different uses of adjectives of affect in both corpora.

(6a) *Belsaita is right and her* ***suggestions are great*** (PenSp corpus)
(6b) *Enjoy! Coba and Tulum* ***are real [sic] great*** (MxSp corpus)
(6c) *A stylish place with* ***excellent Peruvian cuisine****.* (MxSp corpus)
(6d) *The* ***starters are excellent*** *and the pasta with sea food is really impressive.* (PenSp corpus)
(6e) *all of a sudden you are surrounded by an [sic]* ***amazing architecture*** *with a lot of history* (PenSp corpus)
(6f) ***Everything was so amazing*** *and delicious, and served really fast* (MxSp corpus)

One construction with adjectives of affect, however, is only used by Spaniards, with no instances in the Mexican corpus: the token bigram *very good* (7).

(7) ***Very good*** *Chinese food in El Puig at reasonable price.* (PenSp corpus)

The significant difference in frequency of use of adjectives of affect in Spaniards' and Mexicans' L2 English output ($p < 0.05$ in all cases) could be explained by Mexican speakers being more influenced by the way Americans express emotion through affect markers. Yet, as far as we know, there is a lacuna of research concerning Spanish dialectal differences in stance markers, especially with regard to stance markers of affect. Thus, it could also very well be that Mexicans express emotion more frequently than Spaniards: more research is needed in order to draw conclusions.

There is another possible reason for the differences we found in frequency of use of adjectives of affect in both corpora, which does not necessarily invalidate our previous

explanation: Spaniards in the corpus may have a more limited repertoire of use of L2 English adjectives of affect, and consequently make more use of Spanish calques when writing in English. The construction *very good* is a fine example, as it is a word by word translation of Sp. *muy bueno*. Furthermore, *very good* could be understood as a compositional phrase. Compositional phrases, following Snider and Arnon (2012), originate in the grammar while non-compositional phrases originate in the lexicon and are stored together with their idiosyncratic syntactic and semantic features. Since they are derived in a predictable way, compositional phrases do not need to be stored in the lexicon (Snider and Arnon 2012), and thus, should require less processing effort from the speaker who produces them. This could point to Peninsular Spanish speakers in the corpus having a lesser command of English than their Mexican counterparts.

Quantifiers[1]

The augmentative *many* in the construction *many thanks*. The construction *many thanks* is found exclusively in the Peninsular Spanish corpus. Mexican authors either thanked someone on Tripadvisor by writing *thanks* (N = 61, 0.16%) —optionally accompanied by *a lot* or *a ton*—or, to a lesser extent, by writing *thank you* (N = 35, 0.09%) —sometimes co-occurring with *so much* after it. Meanwhile, Spaniards primarily used constructions with *thanks* (N = 77, 0.19%), while *thank you* only occurred on 15 occasions (0.03%) in the Peninsular Spanish corpus. We tested for the difference in frequency of the word *thanks* and the construction *thank you* between corpora with a log-likelihood (LL) test of significance, which indicated that the difference in use of the construction *thank you* to be significant ($p < 0.05$) while the difference in use of the word *thanks* was not significant. Figure 2 provides a visual breakdown of the difference in frequencies of gratitude statements between corpora.

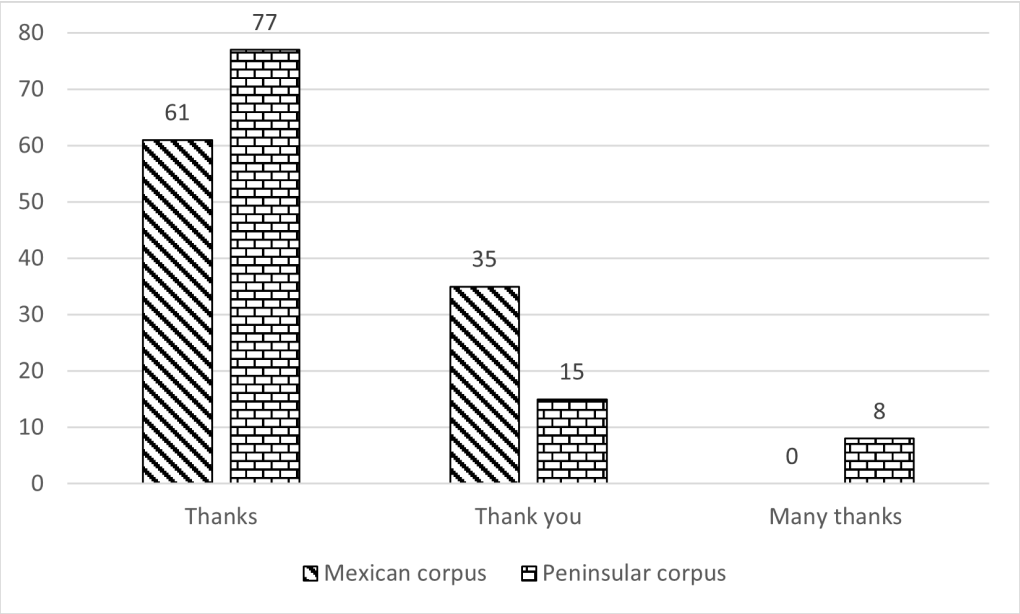

Figure 2. Frequency of Use of Gratitude Statements by Corpora)

[1]The diminutive has already been discussed as a downtoner in the previous subsection.

Of the *thanks* constructions found in the Peninsular Spanish data, six (0.01%) co-appeared with *a lot*, as in *thanks a lot*, and more notably, 16 (0.04%) co-appeared with *many*, as in *many thanks* (8).

> (8) *Could you give any idea??.* ***many thanks*** *again. greetings from Madrid.* (PenSp corpus)

Many thanks, which can be understood as a compositional phrase, only occurs in the Peninsular corpus: the difference in quantities between corpora proved significant ($p < 0.05$). As previously mentioned, compositional phrases do not need to be stored in the lexicon (Snider and Arnon 2012), and thus, should require less processing effort from the speaker who produces them. This, like other examples in the present study (e.g., the construction *very good*) point to Peninsular Spanish speakers in the corpus possessing a lesser command of English than their Mexican counterparts. It could also be argued that Spaniards in the corpus have understood the word *thanks* as Spanish *gracias* (English *thank you*), and because usually this *gracias* appears with the intensifier *muchas* (English *many*) before it, they have made a calque.

This theory is reinforced by a quick comparison of English terms to thank someone. The use of *many thanks* by Spaniards cannot be explained by British English influence alone, as the BNC (BNC Consortium 2007) only shows 2.36 instances per million words (Table 1). Furthermore, in both native dialects of English, *thank you* is the preferred way of thanking someone, whereas in both our L2 English corpora, authors preferred the use of *thanks*.

Term	**BNC**	**COCA**
	Frequency per million words	**Frequency per million words**
Many thanks	2.36	1
Thanks a lot	1.47	4.93
Thanks	62.97	206.27
Thank you	95.55	284.35

Table 1. English Terms Used to Thank Someone: Frequencies

Crystal (2008: 111) defines a contraction as "the process or result of phonologically reducing a linguistic form so that it comes to be attached to an adjacent linguistic form," e.g., *I've* from *I have*, and *haven't* from *have not*, as well as the *wanna* contraction. Both verb contractions and *not*-contractions, following Young (2015), are well represented in standard written texts across various registers and are familiar to non-native speakers (NNSs) of English. Biber (1987), using nine genres from the Lancaster-Oslo-Bergen (LOB) and Brown standardized corpora to compare British and American writing, found systematic differences in the frequencies of contractions: all American genres displayed a markedly higher frequency for this feature than the same British genres, although Biber coded for auxiliary and negative contractions together and unfortunately did not provide quantitative evidence comparing the two contraction strategies. Other more specific analyses of contraction, such as (Algeo 2006), have shown, however, that contraction with the verb *have* is more frequent in BrE. Yaeger-Dror (1997) provided evidence that negative contraction is conditioned by interactional and other register variables, and through her analysis of pragmatic and morphological interpretation of negatives, she

found that negative contraction and auxiliary contraction must be distinguished from each other.

The acquisition of English auxiliary contraction and deletion by second-language learners has been a largely unexplored question (Samar 2003). Research on related topics has shown that copula/auxiliary *be* is one of the first morphemes acquired by both child and adult L2 learners (Krashen 1977; Lightbown 1987), yet, very little is known about how these learners acquire variable contraction of auxiliaries. Odlin (1978) examined the acquisition of English contractions by six Spanish native speakers from Mexico who were studying English as a Second Language (ESL) in Texas. Odlin's results suggested evolutionary stages in the acquisition of contractions characteristic of a certain level of proficiency, where contraction frequency generally correlated with general proficiency (i.e., more proficient students used contracted forms more frequently and with a greater variety of words than less proficient students). The stages of English contraction acquisition would begin with word classes with a single member (such as *it* or *that*), following with other pronouns and *that* and *there*, finally evolving to more complex noun phrases and locatives. The problem with Odlin's stages is that the differences between "word classes with a single member" and "other pronouns" are not clear, and thus, we could not apply his continuum analysis to our results. A further problem is that the acquisition of negative contracted forms (separate from auxiliary contraction) in English as a foreign language is an understudied topic. In the present study we strive to provide a description of linguistic features that serve to identify non-contact dialects of a native language (L1) when dealing with second language (L2) data; taking into consideration human, topic and time constraints on analysis capacity, we compared contraction usage rates between dialects without comparing the two contraction strategies. Future studies in NDID could provide a separate qualitative analysis of negative contraction and auxiliary contraction.

We now turn to our own data, taking previous investigations into account in order to draw certain conclusions. Table 2 allows us to compare contraction usage rates by dialect.

Feature	**Mexican L1 Spanish Corpus**		**Peninsular L1 Spanish Corpus**	
	Number of authors using feature	Segment proportion	Number of authors using feature	Segment proportion
" 's "	23	0.086	non-significant	non-significant
" n't "	24	0.071	21	0.076
" 'm "	15	0.032	20	0.037
" 're "	10	0.013	non-significant	non-significant
" 've "	9	0.012	14	0.017
" 'd "	9	0.01	non-significant	non-significant

Table 2. Usage rates of contracted forms by dialect

Table 2 shows six significant contracted forms in the Mexican Spanish corpus, with three of these contracted forms also being significant in the Peninsular Spanish corpus. We see similar rates of usage and of segment proportions for the significant contracted forms in both corpora. Additionally, while the word *is* was the sixth most frequent unigram found in the Mexican Spanish corpus, this same word was the most frequent unigram in

the Peninsular Spanish corpus, possibly pointing to the comparably lower rates in which contractions occur with the copula *be* in Peninsular authors' L2 English output.

Bigram results in both corpora are also enlightening. Data shows, in general, a more frequent use of contracted forms over their respective full form alternatives in the Mexican Spanish corpus (as examples, *don't* occurs 64 times and *do not*, six; *it's* occurs 62 times and *it is*, 50). In contrast, in the Peninsular Spanish corpus, *it is* occurs 131 times, *It is*, 52 times, and *it's*, 32. Trigrams in our data are no different in this matter. Within the twenty most frequent trigrams in the Peninsular Spanish corpus, we found ten full forms, such as *is not the*, *I do not*, or *I am from*, and zero contracted forms; meanwhile, in the Mexican Spanish corpus, within the twenty most frequent trigrams, we encountered three contracted forms, such as *I don't* and *I'm from*, and only full forms (*if you are* and *would like to*).

These findings demonstrate that Mexican authors make a wider and more varied use of contractions than Peninsular authors. Results also possibly point to Spaniards in the corpus having lower proficiency English levels than Mexicans; having said that, we cannot discount the fact that because Spaniards have probably more input from British English—a dialect with considerably less use of contractions in written registers (Biber 1987) —, their reduced use of contractions could be partly due to British English influence.

Punctuation

Within the punctuation inventory, there are three functional classes (Bredel (2008, 2011), cited in Busch (2021)): syntactic signs, communicative signs, and scanning signs. To guide the grammatical parsing process, syntactic signs are used (i.e., period, comma, colon, and semicolon); to mediate the relationship between writer and reader, communicative signs are used (i.e., exclamation mark, question mark, quotation mark, and parentheses); to indicate that information to decode the message is missing and must be interpreted by the reader, scanning signs are used (i.e., hyphen, apostrophe, and ellipsis dots). As Busch (2021) notes, in some forms of CMC, the use of punctuation signs by collaborative writers in their interactions (such as the ones we find in our TripAdvisor data) acts as contextualization cues, in order to indicate the interpretation of utterances and guide sequential progress.

Previous studies have found that the use of emoji, another form of non-word tokens in CMC, is strongly impacted by cultural background (Gibson *et al.* 2018). Additionally, politeness is culturally bound (Terkourafi 2011). Accordingly, it would be expected for punctuation signs —especially for communicative and scanning signs—to show cross-cultural and cross-dialectal variation.

There are important differences in frequency of use of certain communicative and scanning signs between the Peninsular and Mexican Spanish data in our corpora, shown in Table 3.

Specifically, Mexicans in our data use ellipsis dots (both in its standard three-dot form [...] as well as in repeated forms with four or more dots [....]) more than Spaniards. In (9a), ellipsis dots are used in order to change topic; in (9b), as an ending to the text; in (9c-d), as a way to add suspense and introduce an unexpected outcome, be it positive (9c) or negative (9d); and to delve deeper into a topic (9e). Sometimes, as in (9f), an author can use ellipsis dots in more than one occasion, for different pragma-linguistic

Feature	Mexican Spanish Corpus	Peninsular Spanish Corpus
…	4,04	3,5
…. (4 or more dots)	548	21
!	5,295	4,259
!!!	320	274
?	4,131	2,213
???	182	84
mean	2,419	1,725
SD	2312.60	1871.46

Table 3. Relative frequencies of punctuation signs per million tokens

functions (i.e., as a way to add suspense and introduce an unexpected outcome, and to change topics).

(9a) *Hi, I'm from Guadalajara, and a huge Chivas fan. The games usually start at 07:00 pm. I hope this gives some answers...*
TICKETS: You can buy them during the week (Monday-Friday) at different locations across the city. The day of the game tickets are sold at the stadium until they finish [...] (MxSp corpus)
(9b) *This show is a must when you're in New York. It has it all: great plot, outstanding music, girfted [sic] singers...everything! I definitely recommend Les Mis...* (MxSp corpus)
(9c) *Looking for a great sushi place... This is it!*
We stumbled upon this by chance, but had a mayor surprise (MxSp corpus)
(9d) *Happy birthday... except for the check* (MxSp corpus)
(9e) *Yeah, it's worse due to the dry season, heat and sun...somehow it makes ozone stick around longer* (MxSp corpus)
(9f) *We went there a few days ago because we were visiting the Oceanografic. We just went for a walk to see the buidings and the pools ... Amazing! It looks like you were in a futuristic city ... There are also activities if you go with your kids ...*
You cannot miss this if you visit Valencia! (PenSp corpus)

Moreover, Mexicans in our corpus make use of exclamation marks and question marks, both single and iterated, more often than Spaniards. Example (10) shows some punctuation mark usage within our corpora.

(10a) *The Castle in Chichen.... you can climb it !* (MxSp corpus)
(10b) *The food is very good and the service is very attentive.*
We couldn't try one of their specialties, apparently, they cook Camembert in a very special and delicious way ... next time, hopefully!!
The only thing is that the service is not the fastest in the world... so, take it easy and enjoy the views!!
Congratulations Ca'María team!! (PenSp corpus)
(10c) *I studied at Bristol university for a semester 26 years ago....and I'm coming back!!!* (PenSp corpus)

This analysis of results requires further qualitative examination to clarify the possible reasoning for punctuation mark usage in our data, taking into account the context in which these punctuation marks were used and analyzing them as "contextualization cues" (Busch 2021: 3) that allow language users to "signal contextual presuppositions"

(Gumperz 1982: 131). The literature tackling politeness and requests in the Spanish speaking world points to possible misunderstandings between speakers of different dialects, where (most) issues seem to be found in differing politeness systems, especially between Latin Americans and Spaniards, as Spaniards are perceived as very direct and abrupt by Latin Americans, and Latin Americans are understood by Spaniards to be too formal sometimes (see Márquez-Reiter (2002), for further discussion on the topic).[5] Márquez-Reiter (2002) provides a possible explanation as to why Spaniards are regarded as more direct than Latin Americans, namely, Spaniards' lack of tentativeness when carrying out a request compared to Uruguayan speakers.

The use of punctuation signs in our data signals more tentativeness and more explicit politeness devices by Mexican users, in line with previous research. This seems to be achieved by Mexicans' higher use of ellipsis dots and exclamation and question marks. Our findings align with Bieswanger (2008), who indicated that in CMC, users favor different strategies in different languages and that these strategies are used in different proportions. Our findings also align with Gibson *et al.* (2018), showing that non-word tokens in CMC are impacted by cultural background.

Part-of-Speech (POS) N-grams

We must consider syntactic complexity as an important measure of L2 proficiency (Larsen-Freeman 1978; Lu 2011) that can assist us in gauging differences between authors, and possibly, between different L1 dialects when producing L2 output. Ortega (2015) defines syntactic complexity (also called *syntactic maturity* or *linguistic complexity*) as "the range and the sophistication of grammatical resources exhibited in language production" (p. 82). Biber and colleagues (Biber *et al.* 2011, 2014), when examining formal academic registers, indicated that spoken registers showed grammatical complexity in their clausal elaboration (i.e., subordination with high-frequency mental verbs such as *think*, *know*, and *say*) while written registers displayed grammatical complexity through phrasal compression, and specifically through complex noun phrases and complex phrases. Tyler and Evans (2001) showed that prepositions are significant markers of linguistic complexity, as more advanced L2 learners start being able to use lower-frequency prepositions and prepositional phrases and use these in more idiomatic extended senses. Kyle and Crossley (2018: 334) suggest using fine-grained indices of syntactic complexity, such as phrasal complexity indices, as they "may provide a clearer understanding of the relationship between syntactic complexity and L2 writing" than coarse-grained indices that measure complexity at the clause or sentence level.

Taking these studies into consideration, we now proceed to analyze syntactic complexity through POS bigrams and trigrams. POS n-grams that were most distinctive, that is, that were ranked as distinctive by both the keyness metric and by the Burrows' Zeta metric (versus only appearing as distinctive in one of the two metric lists) are the ones that we discuss below.

POS Bigrams

POS bigrams provide information about differences in how Spaniards and Mexicans express themselves. The construction [VERB ADP], realized as phrasal verbs, is distinctive of the Mexican Spanish corpus. The presence of phrasal verbs points to Mexican authors in our corpora possessing higher lexical sophistication in English than Spaniard authors. The most frequent examples in our data include the verbs *to go to, to look for, to be from,*

to stay in, to go to, to get to, to feel like, to be in, to ask for, and *to live in*. Phrasal verbs are said to be inherently difficult for learners of English as a second language (ESL) to master, with research indicating that ESL students often avoid them and frequently make mistakes in their attempt to use them (Bronshteyn and Gustafson 2015).

Another construction distinctive of the Mexican Spanish corpus is the bigram [NOUN NOUN], where most frequent examples point to Mexican authors' ability to use complex nouns such as *day trip, front desk, observation deck, bus station, science museum, train tickets, train schedule*, and *wine list.* As Biber and colleagues (Biber *et al.* 2011, 2014) have argued, the main source of linguistic complexification in written academic language is phrasal compression through complex noun phrases and complex phrases. It seems that we can transpose this feature of linguistic complexity not only to academic writing, but also to more informal registers, such as the CMC genre of Tripadvisor entries.

POS Trigrams

A distinctive POS trigram in our Peninsular Spanish data is the construction [ADJ NOUN PUNCT], with one of its realizations being *Many thanks!* (N = 5). As has been discussed previously, this phrase is possibly a calque from Sp. *Muchas gracias*, and as a compositional phrase, it does not need to be stored in the lexicon (Snider and Arnon 2012), requiring less processing effort from the speaker who produces it. Possibly, Peninsular Spanish authors in our corpus have a less idiomatic command of English than Mexican Spanish authors, relying more on simpler structures and trying to compositionally define many constructions. This can also be perceived by Spaniard authors' distinctive use of the [AUX ADV ADJ] construction, with which, instead of describing a noun with an adverb and an adjective before it (e.g., *a very*$_{[ADV]}$ *good*$_{[ADJ]}$ *restaurant*$_{[NOUN]}$), they described a noun in a compositional manner (the restaurant$_{[NOUN]}$ is$_{[AUX]}$ very$_{[ADV]}$ good $_{[ADJ]}$), as in (11).

(11a) *The staff* ***is very good****, helpful and friendly.* (PenSp)
(11b) *Price* ***was very good*** *too* (PenSp)
(11c) *The food is great, the service* ***is very attentive*** *and the atmosphere is calm and very pleasant* (PenSp)

The [PRON AUX VERB] construction was the fourth most frequent trigram in both the Peninsular Spanish (N = 415) and Mexican Spanish (N = 360) corpora. The most frequent examples of this construction in the Mexican Spanish corpus, however, show a wider variety of verbs, verb tenses and pronominal persons than its Peninsular Spanish counterpart, as can be seen in Table 4 below. More specifically, Mexican Spanish —in opposition to Peninsular Spanish—includes the *be going to* future tense (*you are going to*), use of *should*, use of contracted forms (*'m*), and use of the first-person plural pronoun *we*.

Classification task

To supplement our corpus analysis, we investigated the discriminatory power of grammatical features in an automatic classification task. Specifically, we tested whether a particular combination of features could predict if a Tripadvisor post was produced by an L1 speaker of Peninsular Spanish or Mexican Spanish. Our primary goal was to understand how linguistic features can aid the forensic linguist in discerning against dialects, not to tweak the classification model to achieve the highest accuracy possible. Because linguistic features are key for explainability in the forensic context, we avoided using

Peninsular Spanish Corpus	**Raw Frequency**	**Mexican Spanish Corpus**	**Raw Frequency**
You will find	12	You should try	9
You can find	12	I would like	8
You are looking	10	You would expect	7
I would recommend	10	I'm planning	7
I would like	9	You can go	6
I would say	8	You can take	6
You can see	7	I was thinking	5
You can take	6	I would recommend	5
You can go	6	We would like	5
You can have	5	You are going	5

Table 4. The [PRON AUX VERB] construction by dialect

opaque features such as character n-grams or word embeddings (Mikolov *et al.* 2013) for our classification task.

We trained and developed on the target corpora and saved the classifier weights for our held-out test set. We implemented a logistic regression model in Python with the Scikit-learn library (Pedregosa *et al.* 2011), using one-hot encoding as a vectorization method for each feature. One-hot encoding accounts for the presence of a feature in a text, notwithstanding its frequency. This approach was a good fit for our experiment, as texts were not very long; longer texts might require frequency encoding.

Since we had a binary classification task, we considered a 50% accuracy result to be our chance baseline. We also reported on precision, recall[6], and F-1 (the harmonic mean of the precision and recall score) to offer a holistic evaluation. Since the F-1 score is the weighted average of the precision and recall score, it punishes extreme scores and offers a better understanding of the performance of the models.[7]

Our first baseline models leveraged token unigrams and token bigrams as features. Since these are typical features in text classification, they offer a starting point for seeing how well we can classify the texts without sophisticated feature engineering. This model considered all tokens in the text as features—either a single token context or a two-token context. For instance, for the text "*you should try*", unigrams represent the individual tokens (*you, should, try*) and bigrams the two-token sequences (*you should, should try*). Consequently, this allowed the classifier to consider tokens beyond the scope of our analysis (e.g., content words) and encoded other linguistic patterns that we did not examine in detail (e.g., short token sequences or shorter syntactic patterns).

As displayed in Table 5, both unigram and bigram classifier variants yielded an accuracy of 65%, an improvement on the chance baseline (50%), demonstrating that there are linguistic features in the text that can discriminate between the two L1 dialects in L2 English. The unigram and bigram classifiers show identical precision, recall, F-1, and accuracy scores because both models are similarly influenced by the presence of the same tokens—either the presence of a single token (unigram) or a two token sequence (bigram)—, but how these tokens are represented does not seem to make a difference for predicting a label in our data. Despite the predictive power of these baseline models,

content words can be misleading features. For example, many words in our corpus give away the speaker's L1 dialect by mentioning places such as *Mexico*, *Madrid*, *Mexico City*, *Spain*, *Monterrey*, *Sevilla*, *Cancun*, and *Guadalajara*. The classifier weighs these words heavily, yet such features are topic dependent.

Next, we turned to feature engineering with the linguistic categories from our qualitative analysis, namely, punctuation, adjectives of affect, intensifiers, and contracted forms. We also considered the performance of combining these features to improve classification accuracy. We did not take POS n-grams into account for feature engineering, because these features seem to point to authors' non-native proficiency rather than to idiosyncratic linguistic features carried over from their L1 dialect. Taking POS n-grams into consideration would have been misleading, as it might well be possible for Peninsular Spanish speakers to have higher non-native proficiency than Mexican Spanish speakers in other data sets.

Our best performing classifier used punctuation features and scored just above chance with an accuracy of 55%. The remaining features, however, did not perform well overall, with accuracy scores ranging from 36%-50%. Interestingly, the combination of features did not perform well either.

Notably, many of these features aside from punctuation are not prevalent in every text. This feature sparsity issue leaves many texts in the test set without any of the features from the training step and leads the classifier to make poor decisions. Combining features also increases the number of features for the model to consider, but in our case, this introduced more noise into the system and accuracy did not improve.

Punctuation was a distinctive feature that was present in our texts, but it did not occur enough to be a reliable classification feature. This was the case for other features (namely, adjectives of affect, intensifiers, contracted forms, and a combination of these): from our corpus analysis we note frequency differences that are distinctive between the two classes but not common enough to make a difference in automatic classification, as can be observed in Table 5.

Feature	Precision	Recall	F-1	Accuracy
token unigrams	**0.65**	**0.65**	**0.65**	**0.65**
token bigrams	**0.65**	**0.65**	**0.65**	**0.65**
punctuation signs	**0.6**	**0.55**	**0.51**	**0.55**
adjectives of affect	0.36	0.43	0.36	0.43
intensifiers	0.36	0.36	0.36	0.36
contracted forms	0.24	0.49	0.32	0.49
combination	0.42	0.45	0.38	0.45

Table 5. Classification results for the baseline n-gram model and for models with linguistic features from qualitative analysis

While result scores were low overall, the primary focus of this section of our study is to gain insight into how the linguistic features we qualitatively analyzed from our corpus perform in a classification task. Therefore, we restricted our classification models to using these features and not other traditional features that tend to work well in NLID related tasks but are difficult to explain from a linguistic viewpoint.

To improve our classification task but maintain transparency, we considered another feature that is commonly used in authorship attribution and forensic linguistic contexts: function words. The motivation for applying function words was that they are ubiquitous enough to fix our feature sparsity issue and, unlike content words, they are not topic dependent. Therefore, we offer another round of experiments that considered function words as a feature and combined them with our main linguistic features. We provide classification results in Table 6.

Feature	Precision	Recall	F-1	Accuracy
function words	0.58	0.58	0.58	0.58
function words + punctuation	**0.69**	**0.69**	**0.69**	**0.69**
function words + contracted forms	0.58	0.58	0.58	0.58
function words + adjectives of affect	0.59	0.59	0.59	0.59
function words + intensifiers	0.53	0.53	0.53	0.53
combination	0.55	0.55	0.54	0.54

Table 6. Classification results with function words as features and function words in combination with linguistic features from qualitative analysis

The classifier that used function words alone performed better than chance and better than the previous baseline classifier. The rest of the classifier variants that combined one of the linguistic features with the function word features improved accuracy, all above chance. Punctuation was a powerful feature on its own, and with function words, it yielded the highest accuracy of all (69%). This improvement surpasses the baseline token unigram and baseline classifiers. Interestingly, combining all of these features performs worse than the other variants, and combining adjectives of affect and contracted forms with function words has little to no impact on the models. These features do not seem to correlate well together, even though they may yield different scores in isolation. Additionally, we note that all results have the same precision and recall score. This is more expected for balanced datasets. As a result, this also leads to a similar F-1 score and accuracy score.

Our best performing classifier can predict between a text produced from an L1 speaker of Peninsular Spanish and Mexican Spanish when writing in L2 English with a unique combination of features that are transparent, topic independent, generalizable, and prevalent. While using an engineering approach, and specifically a logistic regression model, is not always applicable to the forensic context because of text length and explainability issues, we show that linguistically informed features can improve an automatic approach for this task. Therefore, such a model potentially serves as an analysis tool that balances an automatic approach and its linguistic explainability, which can be helpful when tackling automatic language and dialect identification tasks. These results, however, only pertain to the scope of this specific genre and text size and need further research with data from L2 English texts written by authors with other L1 Spanish dialects. Indeed, these models would struggle with shorter texts where these features are possibly sparser.

While the accuracy seems relatively low and there is still room for improvement, these experiments are a starting point for NDID classification. There was not one feature that carried the prediction but, from among the features we analyzed, punctuation

seemed to be the most effective, suggesting that other NDID classification tasks within online genres should consider punctuation types and punctuation repetition.

Conclusion and future perspectives

We presented a first effort to identify two Spanish L1 dialects, Mexican and Peninsular, when analyzing texts written in L2 English. The mixed-approach methodology we used to tackle NDID provides a comprehensive linguistic description of features that serve to identify Mexican and Peninsular dialects of Spanish. As Kingston (2019) remarks, a frequent problem with NLI studies is that they tend to neglect the features used to classify texts; our study, on the contrary, has taken linguistic features into consideration as they are key for explainability in the forensic context.

In terms of linguistic features in our data, a careful analysis from results revealed that frequency alone is not necessarily explanatory; as an example, while *a (little) bit* is more frequent in the Peninsular Spanish corpus, it shows a wider array of pragma-linguistic functions in the Mexican Spanish corpus, mirroring what happens in speakers' L1 dialects. Thus, qualitative analyses are needed to help dissect between dialects more accurately. Another revealing conclusion from our study is that stance markers, which vary culturally, can aid the forensic linguist in discovering the L1 dialect of a language from L2 output. Additionally, use of punctuation signs in the data follows patterns found in previous studies (Bieswanger 2008; Gibson *et al.* 2018), where CMC users from different cultural backgrounds show different communicative strategies and where use of non-word tokens is culturally bound. Importantly, results showed that we can tell the difference between Spaniards' and Mexicans' L2 English output, so that it might be possible to differentiate English L2 outputs among Latin American Spanish dialects. Having said this, inter-medial and inter-modal studies are needed to further advance knowledge in NDID.

We also implemented a classifier to detect the L1 Spanish dialect of an anonymous author writing in L2 English. In the experiments we carried out on the test data, our model with function words and punctuation as features achieved accuracy of 69% in categorizing unseen Tripadvisor entries. The applicability of this method to different data requires further investigation, as we need to see if it is transferable to different genres. Furthermore, the question of how short a text can be and still allow accurate categorization still needs to be answered.

This investigation contributes to the field of authorship attribution studies in general and to NLID and NDID studies in particular by being the first of its kind to address NDID in non-contact dialects of any language. Future studies should apply the same methodology to other dialects of Spanish to see if these linguistic features apply to all dialects of Spanish in general. Finally, another issue that commands additional research is whether the same linguistic features can detect native dialect influence in L2 English (or any other L2) texts with an L1 that is not Spanish.

Notes

[1]According to its Wikipedia entry, Tripadvisor is "an American online travel company that operates a website and mobile app with user-generated content, a comparison shopping website, and offers online hotel reservations as well as bookings for transportation, lodging, travel experiences, and restaurants" (https://en.wikipedia.org/wiki/TripAdvisor).

[2]We are aware that gauging native proficiency can be complicated. In this specific case, we were looking for accuracy (that is, the ability to produce grammatically correct sentences) and grammatical range. The author who carried out the analysis is a native Spanish speaker and has also carried out L2 proficiency level assessments.

[3]Intensifiers are adverbs that magnify meaning, scaling a quality up (Ito and Tagliamonte 2003).

[4]All data for AmE frequencies in the study were obtained from the *Corpus of Contemporary American English* (COCA, Davies (2008)), while data for all BrE frequencies were obtained from the *British National Corpus* (BNC Consortium 2007).

[5]As Márquez-Reiter (2002) points out, these are quite generalized cultural perceptions: Latin America is a vast, culturally diverse geographical area.

[6]The recall score refers to the number of predicted labels divided by the number of labels in the dataset. The precision score refers to the number of predicted labels divided by the number of those predicted labels that actually belong to the label.

[7]Since we have more of an even class distribution—an even number of testing input from Mexican Spanish and Peninsular Spanish— we anticipate the accuracy score being sufficient for this study.

References

Algeo, J. (2006). *British or American English? A handbook of word and grammar patterns*. https://doi.org/10.1017/CBO9780511607240: Cambridge University Press.

Biber, D. (1987). A textual comparison of british and american writing. *American Speech*, 62(2), 99–119.

Biber, D., Gray, B. and Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *Tesol Quarterly*, 45(1), 5–35.

Biber, D., Gray, B. and Staples, S. (2014). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, 37, 639–668.

Bieswanger, M. (2008). 2 abbrevi8 or not 2 abbrevi8: A contrastive analysis of different space- and time-saving strategies in english and german text messages. *Texas Linguistics Forum*, 50. http://studentorgs.utexas.edu/salsa/proceedings/2006/Bieswanger.pdf.

BNC Consortium, (2007). *The British National Corpus*. Oxford Text Archive, xml ed. http://hdl.handle.net/20.500.12024/2554.

Bredel, U. (2008). *Die Interpunktion des Deutschen. Ein kompositionelles System zur Online-Steuerung des Lesens*. Tübingen: Niemeyer.

Bredel, U. (2011). *Interpunktion*. Heidelberg: Winter.

Bronshteyn, K. and Gustafson, T. (2015). The acquisition of phrasal verbs in l2 english: A literature review. *Linguistic Portfolios*, 4(1), 91–99.

Burrows, J. (2007). All the way through testing for authorship in different frequency strata. *Literary and Linguistic Computing*, 22, 27–47.

Busch, F. (2021). The interactional principle in digital punctuation. *Discourse, Context & Media*, 40, 100481.

Bäcklund, U. (1973). *The collocation of adverbs of degree in English*. Almqvist & Wiksell.

Caraker, R. (2016). Spain and the context of english language education. *Research Bulletin*, 92, 23–35.

Company Company, C. (2002). Gramaticalización y dialectología comparada: Una isoglosa sintáctico-semántica del español. *Dicenda: Cuadernos de Filología Hispánica*, 20, 39–71.

Coulthard, M., Grant, T. and Kredens, K. (2010). Forensic linguistics. In R. Wodak, B. Johnstone and P. Kerswill, Eds., *The SAGE Handbook of Sociolinguistics*. Sage Publications, 529–544.

Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics*. Blackwell Publishing, 6th ed.

Davies, M. (2008). The corpus of contemporary american english (coca): 385 million words, 1990-present. https://www.english-corpora.org/coca/.

Despagne, C. (2010). The difficulties of learning english: Perceptions and attitudes in mexico. *Canadian and International Education / Education Canadienne et International*, 39(2), 55–74.

Fernández Vítores, D. (2020). El español: una lengua viva. informe. https://www.cervantes.es/imagenes/File/espanol_lengua_viva_2019.pdf.

Gibson, W., Huang, P. and Yu, Q. (2018). Emoji and communicative action: the semiotics, sequence and gestural actions of 'face covering hand. *Discourse, Context & Media*, 26, 91–99.

Goldin, G., Rabinovich, E. and Wintner, S. (2018). Native language identification with user generated content. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, 3591–3601, https://aclanthology.org/D18-1000.pdf.

Grant, T. (2008). Approaching questions in forensic authorship analysis. In J. Gibbons and M. Turell, Eds., *Dimensions of Forensic Linguistics*. John Benjamins Publishing Company, 215–229.

Grant, T., Kredens, K. and Perkins, R. (2010). Identifying an author's native language phase 2 + finding and training the bilingual language expert.

Gumperz, J. (1982). *Discourse strategies*. Cambridge University Press.

Herring, S. (2012). Grammar and electronic communication. In C. Chapelle, Ed., *Encyclopedia of applied linguistics*. Wiley-Blackwell.

Honnibal, M. and Montani, I. (2015). spacy: Industrial-strength natural language processing (nlp) with python and cython). https://spacy.io.

Ito, R. and Tagliamonte, S. (2003). Well weird, right dodgy, very strange, really cool: Layering and recycling in english intensifiers. *Language in society*, 32(2), 257–279.

Jiang, X., Guo, Y., Geertzen, J., Alexopoulou, D., Sun, L. and Korhonen, A. (2014). Native language identification using large, longitudinal data. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC '14)*, 3309–3312: European Language Resources Association (ELRA).

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. and Suchomel, V. (2014). The sketch engine: ten years on. *Lexicography*, 1(1), 7–36.

Kingston, J. (2019). *Other Language Influence of Metropolitan, Canadian, and Maghrebi French in Colloquial Written English [Unpublished Master's Thesis*. Birmingham, UK: Aston University.

Koppel, M., Schler, J. and Zigdon, K. (2005). Automatically determining an anonymous author's native language. In P. Kantor, G. Muresan, F. Roberts, D. Zeng, F.-Y. Wang, H. Chen and R. Merkle, Eds., *Intelligence and Security Informatics: IEEE International Conference on Intelligence and Security Informatics*, 209–217: Springer.

Krashen, S. (1977). The monitor model for adult second language performance. In M. Burt, H. Dulay and M. Finocchlaro, Eds., *Viewpoints on English as a second language*. Regents Publishing Company, 152–161.

Kyle, K. and Crossley, S. (2018). Measuring syntactic complexity in l2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, 102(2), 333–349.

Larsen-Freeman, D. (1978). An esl index of development. *TESOL quarterly*, 12, 439–448.

Lightbown, P. (1987). Classroom language as input to second language acquisition. In C. Pfaff, Ed., *First and second language acquisition processes*. Newbury, 169–187.

Lipski, J. (2012). Geographical and social varieties of spanish: An overview. In J. Hualde, A. Olarrea and E. O'Rourke, Eds., *The handbook of Hispanic linguistics*. https://doi.org/10.1002/9781118228098: Wiley-Blackwell, 1–26.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level esl writers' language development. *TESOL quarterly*, 45(1), 36–62.

Maier, W. and Gómez-Rodríguez, C. (2014). Language variety identification in spanish tweets. In *Proceedings of the EMNLP'2014 Workshop on Language Technology for Closely Related Languages and Language Variants*, 25–35.

Mikolov, T., Yih, W.-t. and Zweig, G. (2013). Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 746–751.

Márquez-Reiter, R. (2002). A contrastive study of conventional indirectness in spanish: Evidence from peninsular and uruguayan spanish. *Pragmatics*, 12(2), 135–151.

Odlin, T. (1978). Variable rules in the acquisition of english contractions. *TESOL Quarterly*, 12(4), 451–458.

Ortega, L. (2015). Syntactic complexity in l2 writing: Progress and expansion. *Journal of Second Language Writing*, 29, 82–94.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830.

Perkins, R. (2015). Native language identification (nlid) for forensic authorship analysis of weblogs. In M. Dawson and O. Marwan, Eds., *New threats and countermeasures in digital crime and cyber terrorism*. https://doi.org/10.4018/978-1-4666-8345-7.ch012: IGI Global, 213–234.

Perkins, R. and Grant, T. (2013). Forensic linguistics. In J. Siegel, P. Saukko and M. Houck, Eds., *Encyclopedia of Forensic Sciences*. Elsevier, 174–177.

Precht, K. (2000). *Patterns of stance in English*. Northern Arizona University.

Precht, K. (2003a). Great versus lovely: Stance differences in american and british english. In P. Leystina and C. Meyer, Eds., *Corpus Analysis: Language structure and language use*. Brill Rodopi, 133–151.

Precht, K. (2003b). Stance moods in spoken english: Evidentiality and affect in british and american conversation. *Text & Talk*, 23(2), 239–257.

Rangel, F., Franco-Salvador, M. and Rosso, P. (2016). A low dimensionality representation for language variety identification. In *International Conference on Intelligent Text Processing and Computational Linguistics*, 156–169, Cham: Springer.

Reynoso Noverón, J. (2001). La pragmática como evidencia en el contacto español-lenguas indígenas. el diminutivo en el español actual. In C. Matute and A. Palacios, Eds., *El indigenismo americano (II). Actas de las Segundas Jornadas Internacionales sobre Indigenismo Americano*. Universitat de València, Facultat de Filología, 213–222.

Sadat, F., Kazemi, F. and Farzindar, A. (2014). Automatic identification of arabic language varieties and dialects in social media. In *Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP)*, 22–27: Association for Computational Linguistics and Dublin City University.

Samar, R. (2003). Aux-contraction in second language speech: A variationist analysis. *Cahiers Linguistiques d'Ottawa*, 31.

Snider, N. and Arnon, I. (2012). A unified lexicon and grammar? compositional and non-compositional phrases in the lexicon. In D. Divjak and S. Gries, Eds., *Frequency effects in language representation*. https://doi.org/10.1515/9783110274073.127: De Gruyter-Mouton, 127–163.

Terkourafi, M. (2011). From politeness1 to politeness2: tracking norms of im/politeness across time and space. *Journal of Politeness Research*, 7(2), 159–185.

Tetreault, J., Blanchard, D. and Cahill, A. (2013). A report on the first native language identification shared task. In *Proceedings of the eighth workshop on innovative use of NLP for building educational applications*, 48–57, The Association for Computational Linguistics. https://aclanthology.org/W13-1700.pdf.

Thomason, S. (2001). *Language Contact: An Introduction*. Georgetown University Press.

Travis, C. (2004). The ethnopragmatics of the diminutive in conversational colombian spanish. *Intercultural Pragmatics*, 1(2), 249–274.

Tyler, A. and Evans, V. (2001). Reconsidering prepositional polysemy networks: The case of over. *Language*, 77(4), 724–765.

Wierzbicka, A. (1984). Diminutives and depreciatives: Semantic representation for derivational categories. *Quaderni di semantica*, 5(1), 123–130.

Wierzbicka, A. (1992). *Semantics, culture, and cognition: Universal human concepts in culture-specific configurations*. Oxford University Press.

Wong, S. and Dras, M. (2011). Exploiting parse structures for native language identification. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 1600–1610.

Yaeger-Dror, M. (1997). Contraction of negatives as evidence of variation in register specific interactive rules. *Language Variation and Change*, 9, 1–36.

Young, J. (2015). *Exploring Japanese learners' perception, production, and beliefs concerning spoken English contractions*. University of Illinois at Urbana-Champaign.] ProQuest Dissertations Publishing.

Zampieri, M. and Gebrekidan-Gebre, B. (2012). Automatic identification of language varieties: The case of portuguese. In *KONVENS2012-The 11th Conference on Natural Language Processing*, 233–237: Österreichischen Gesellschaft für Artificial Intelligende (ÖGAI).

An investigation of the lexico-grammatical profile of English legal- lay language

Lucia Busso

Aston Institute for Forensic Linguistics, Aston University, UK

https://doi.org/10.21747/21833745/lanlaw/9_1a7

Abstract. *The article presents a study on the lexico-grammar of the genre of English legal-lay language (Tiersma 1999), using the English subcorpus of the CorIELLS corpus (Busso forthcoming). The study explores four grammatical constructions (in Goldberg 2006's Construction Grammar sense): nominalisations heading prepositional phrase attachments, modal verb constructions, participial reduced relative constructions, and passive constructions. Specifically, we use collostructional analysis (Stefanowitsch 2013), followed by a vocabulary analysis using English core vocabulary as a reference (Brezina and Gablasova 2015), and a comparative frequency analysis with corpora of legal language and general-domain written prose. Results of this first part of the study foreground how legal-lay language is quantitatively different from both neighbouring genres, suggesting that it might be considered a "blended" genre. We further explore the data in terms of accessibility for speakers, using readability metrics and a survey on English participants. Both methods show that legal-lay language is at an intermediate level of complexity between legal jargon and general-domain prose; however, we further note that readability metrics generally underestimate speakers' ability to comprehend legal-lay language.*

Keywords: *Construction Grammar, Legal lay language, Italian, English, Quantitative corpus linguistics.*

Resumo. *O artigo apresenta um estudo sobre a léxico-gramática do género linguagem jurídica para leigos (Tiersma 1999), utilizando o subcorpus inglês do corpus CorIELLS (Busso forthcoming). O estudo explora quatro construções gramaticais (no sentido da Gramática da Construção de Goldberg 2006): nominalizações que regem sintagmas preposicionais, construções com verbos modais, construções relativas restritivas participiais e construções passivas. Especificamente, recorremos à análise colostrucional (Stefanowitsch 2013), seguida de uma análise de vocabulário utilizando o vocabulário principal do inglês como referência (Brezina and Gablasova 2015), e uma análise de frequência comparativa com corpora de linguagem jurídica e textos escritos de linguagem*

geral. Os resultados desta primeira parte do estudo destacam como a linguagem jurídica para leigos é substancialmente diferente dos dois géneros que lhe estão próximos, o que sugere tratar-se de um género "híbrido". Exploramos, ainda, os dados em termos de acessibilidade para os falantes, recorrendo a métricas de compreensibilidade e a um inquérito realizado junto de participantes ingleses. Os dois métodos mostram que a linguagem jurídica destinada a leigos se encontra num nível intermédio de complexidade entre o jargão jurídico e a linguagem geral; contudo, notamos ainda que as métricas de legibilidade subestimam, habitualmente, a capacidade de os falantes compreenderem a linguagem jurídica para leigos.

Palavras-chave: *Gramática construtiva, Linguagem jurídica para leigos, Italiano, Inglês, Linguística de corpus quantitativa.*

Introduction

Comprehension and readability of legal documents – especially if aimed at non-specialists – has been at the centre of the debate in both applied linguistics and legal studies (Tiersma 1999; Frade 2007; Haapio 2011). Particularly, many scholars have advocated a clearer, plainer language in the drafting of legal texts for the lay public (Charrow and Charrow 1979; Schiess 2007).

There is often a difficult trade-off to manage when deciding the level of linguistic complexity to embed in a LLL text, which has to obtain both legal precision and linguistic clarity. In fact, writing a legal document that is at the same time clear and understandable and respects the intricacies of the law is not an easy task (Ződi 2019). However, while a certain level of complexity due to the topic is generally considered vital to reduce vagueness as much as possible (Gotti 2014), a lack of comprehensibility leads to linguistic and legal problems alike (Haapio 2011; Conklin *et al.* 2019).

The need for more comprehensible language in legal settings has been present among English scholars for a long time and resulted in the *Plain Language* movement (Bhatia 1983; Adler 2012) , the most prominent example of interdisciplinary effort to simplify access to complex texts. As Adler (2012: 3) specifies,

> " '[p]lain language' means language and design that presents information to its intended readers in a way that allows them, with *as little effort as the complexity of the subject permits*, to understand the writer's meaning and to use the document."

In the UK, the *Plain English Campaign* (founded in 1979) has been "campaigning against gobbledygook, jargon and misleading public information." (Plain English Campaign website).

In general, scholars advocating for a simplification of legalese argue that syntactic, semantic, and pragmatic complexity hinder comprehension for the lay reader. This issue is crucially relevant in contemporary societies, where different (often binding) legal documents regulate many parts of every-day life. The online world in fact constantly exposes us to legal texts – which require from the user a basic understanding of legal concepts for a variety of purposes (e.g., the terms and conditions of websites, legal notices of online banking services, etc.). The overwhelming importance of legal-lay language is perfectly exemplified by the Cambridge Analytica scandal, in which

legally binding terms and conditions of an app used by Facebook stated that they were harvesting data from users who authorised it and their friends. The app then would transfer the data to the political consultancy Cambridge Analytica which could assemble psychological profiles of voters based on their online presence (Romm 2018).

Terms and conditions are the most used example of legal-lay language; however, it is here argued that the term expands wider than that, including all types of texts with legal content but aimed at a non-specialist audience (Tiersma 1999; Williams 2010; Busso 2022, forthcoming). Example 1 and 2 represent two concordances of the word "contract" extracted respectively from the corpora of legal-lay and specialist legal language used in this article.

1. A contract for the provision of an account with the functions described in these Terms and Conditions is concluded when we confirm that we have set up an account for you either via e-mail or through a message delivered through the App.
2. The Court shall have jurisdiction to give judgment pursuant to any arbitration clause contained in a contract concluded by or on behalf of the Community, whether that contract be governed by public or private law.

Specifically, it is here argued that specialised legal jargon and legal-lay language (henceforth: LLL) can be considered – at least in some respects – different. In fact, the inaccessible nature of legal texts has been widely studied (Chovanec 2013). Complex and highly specialised syntax and lexicon are the most noticeable features of this genre, playing an almost 'ritualistic' role in identifying it (Coulthard and Johnson 2007: 37). But while specialist legal language remains principally used by professionals with years of legal education, LLL has instead the specific aim to be read and – more critically – understood by lay readers.

As mentioned, legal language has been extensively researched by linguists. Most recently, many scholars have started to use computational models to analyse the genre (Hamann *et al.* 2016; Fanego and Rodríguez-Puente 2019; Van Boom *et al.* 2016; Frankenreiter and Livermore 2020). However, the linguistic analysis of LLL as a separate textual type is still an under researched area (Lintao and Madrunio 2015; Conklin *et al.* 2019). The present contribution aims at filling this gap by providing an exploratory analysis of a corpus of LLL in English, following a similar procedure to the study outlined in Busso (2022) for Italian. Specifically, combining evidence from quantitative text-based and experimental methods, this article addresses the following research questions, which focus on different level of linguistic analysis:

I. How specialised is the LLL lexicon? (lexico-grammatical level)
II. Does LLL exhibit linguistic features that are measurably different from specialist legal jargon and general-domain written language alike? (syntactic-semantic level)
III. How comprehensible is LLL with respect to legal and general-domain written language? (semantic-pragmatic level)

Construction Grammar as a reference framework for corpus-based analysis

Construction Grammar, a family of linguistic theories advocating a Usage-Based model of language, understands language as composed of complex units called *constructions*

(Goldberg 2006, 2019). Constructions are conceptually cognate to the Saussurean notion of *sign* as a "two-sided psychological entity" (Saussure 1916: 63) that combines a particular form, i.e., the 'signifier' (or 'signifiant'), with a particular meaning, i.e., the 'signified' (or 'signifié'). Crucially, Construction Grammar extends the idea of arbitrary form-meaning pairings to all levels of grammatical description – from lexical items, to abstract phrasal patterns.[1] Since constructions at the lexical level are not ontologically different from abstract grammatical constructions, Construction Grammar does not see syntax and lexicon as qualitatively different as in rule-based models of language (Pollard and Sag 1994). All constructions – from lexical items to fully schematic syntactic structures – are included in the *constructicon* of a language, i.e., the full inventory of constructions. In other words, constructions differ among themselves only in terms of length, complexity, or level of schematicity.

Since Construction Grammar is part of the constellation of Usage-based models – i.e., models that argue that knowledge of usage is inseparable from grammar (Bybee 2015) – observational data such as corpus data play a crucial role in many studies that adopt such a framework (Gries 2013; Hilpert 2013). Furthermore, it has been argued that Construction Grammar can prove to be useful for the analysis of genre (Groom 2019). A constructionist approach hence offers tools for an approach to text analysis that allows for a cohesive and unified account of features at different levels of linguistic complexity.

The present study uses corpus data aligning itself to general constructionist tenets. That is, the study foregrounds usage of form-meaning patterns at different levels of abstraction, and analyses their structure, function, and frequency. While the study of morpho-syntactic patterns in legal and bureaucratic language is by no means uncommon in the literature (Goźdź Roszkowski and Pontrandolfo 2015; Goźdź-Roszkowski and Pontrandolfo 2017), there is no study in the literature – to the best of the author's knowledge – that explicitly uses Construction Grammar as a means to explore the linguistic structure of legal or legal-lay language. Moreover, analyses of phraseology in legal contexts are mainly qualitative in nature, while the present work employs quantitative methods.

The next sections will provide an in-depth description of the data and of the various methods of analysis used: collostructional analysis, lexical analysis, and contrastive frequency analyses comparing LLL to legal and written prose corpora. The last section draws general conclusions from the analyses performed.

Data: the *CorIELLS* corpus

As data, the study employs *CorIELLS* (CORpus of Italian and English Legal-lay textS), a specialised bilingual corpus of LLL in Italian and English (Busso 2022)[2]. In line with our working definition of LLL (see Introduction section), different textual types were included in the corpus. Particularly, the types of document selected follow two general criteria for inclusion. They are all: (a) freely available online, to approximate the types of LLL people are exposed to on the Internet, and (b) varied, to obtain a sample as representative as possible for the genre in question (Biber 1993; Almut 2010). The final selection includes four major categories of document:

- TERMS AND CONDITIONS AND/OR TERMS OF USE OF WEBSITES. 45 in total or for each country websites were manually selected from the Alexa list of the

500 most visited websites in Italy and the UK in 2019; only web services with legal notices in both languages were included.

- EUROPEAN LEGISLATION SUMMARIES. These texts are "short, easy-to-understand explanations of the main legal acts passed by the EU and intended for a general, non-specialist audience" *(EUR-lex* website).[3] A selection of texts was collected from the official website *EUR-lex* in both their Italian and English versions. 247 summaries per language (all summaries from 2019 and 2020).
- BANK CONTRACTS. Freely accessible legal documents for standard current accounts were selected from 15 banks in Italy and the UK.
- UTILITIES. Standard contract terms for 5 energy suppliers, 5 Wi-Fi suppliers, and 5 pay-by-the-month phones in Italy and the UK were selected.

The documents were semi-automatically retrieved, cleaned, and downloaded using the web scraping *Bootcat* toolkit (Baroni and Bernardini 2004). Size of the corpus amounts to 1.85M words. Composition of the general corpus can be seen in Table 1.

Document type	**Number of texts**	**English subcorpus (800K words)**	**Italian subcorpus (1M words)**
Bank contracts	15 per language	19%	27%
Utilities contracts	15 per language	25%	23%
Terms and conditions	45 per language	34%	27%
EurLex summaries	247 per language	22%	23%

Table 1. Composition of CorIELLS and of its English and Italian subcorpora

In this article, only the English subcorpus will be analysed. For a similar study using the Italian subcorpus of CorIELLS, see Busso (2022).

Construction selection

Lexical bundles and grammatical patterns are a common object of study in the analysis of legal and bureaucratic language (Goźdź Roszkowski and Pontrandolfo 2015; Goźdź-Roszkowski and Pontrandolfo 2017; Yunus and Ab Rashid 2016). However, as mentioned in the Introduction, this paper takes the analysis of phraseology a step further, considering lexico-grammatical patterns as constructions, i.e., linguistic units.

Four constructions were selected for two theoretically motivated reasons. Firstly, constructions were selected at different levels of abstraction to obtain a balanced representation of the lexico-grammatical nature of the corpus. Secondly, the selection was carried out capitalizing on previous research on legal and bureaucratic texts; only constructions unanimously recognized by the literature as highly characteristic of legal language and LLL were selected (Williams 2004; Chovanec 2013; Haigh 2013).

i. Lexical/phrase level: Nominalizations heading prepositional chains (henceforth: NOM_PP). Nominalizations are lexical constructions broadly defined lexically as the "process via which a prototypical verbal clause (…) is converted into a noun phrase" (Givón 1993: 287) . They have been long recognised as being "overwhelmingly used in legislative provisions" (Bhatia 1993: 148). This type of construction is especially used instead of verb phrases (VP), which are usually

scarce in English legal texts (Williams 2013). That is, events are preferentially encoded through deverbal nominalizations, typically embedded in long PP-attachment chains, as in example 1.[4]

1.
2.
3. Mandatory collective management of rights for retransmissions of radio and television programmes by means other than cable.

ii. Phrase Level: Modal verbs (henceforth: MOD). Phraseological patterns Vmod + V composed by a modal verb and any finite or non-finite form of a verb, as shown in example 4 below. MOD are generally understood as 'grammaticized constructions' (Langacker 2013: 14; see among others Cappelle and Depraetere 2016; Torres–Martínez 2019). The literature has long recognized modality as one of the distinguishing features of legal and bureaucratic language (Tiersma 1999; Aher 2013).

1.
2.
3.
4. We must be satisfied of your identity and can refuse instructions if we doubt your identity.

iii. Phrase/Clause Level: Reduced participial relative clause (henceforth: PART). These constructions contain a present (or past) participle that 'replaces' a relative pronoun and main verb (Quirk *et al.* 1985). Present participial constructions are typical of the morpho-syntax of legal English (Janigová 2008).

1.
2.
3.
4.
5. The 'application publisher' means the entity licensing the application to you as identified in the Store.

iv. Discourse level: Passive constructions (henceforth: PASS) (Jaeggli 1986)Passive constructions are a distinctive feature of legal and legal-lay texts (Bulatović 2013), often used to omit the agent of the sentence, as in example 4 below. It has been claimed in the literature that an excessive use of passives leads to highly cognitively demanding texts (Yokoyama *et al.* 2006).

1.
2.
3.
4.
5.
6. Payments (…) will be sent on the next working day.

The chosen constructions are used for all following analyses. To retrieve all instances of them in the corpus, general CQL queries were carried out on *SketchEngine* (Kilgarriff *et al.* 2004).[5]

Collostructional analysis

Collostructions vs collocations

Collostructional analysis (Stefanowitsch, 2013) is a family of quantitative methods that measure the statistical preference or dispreference (in terms of association strength) that words exhibit to constructions. It is an extension of traditional collocational analysis using Construction Grammar tenets; the term *collostruction* itself is in fact a blend of the two words 'collocation' and 'construction'. It significantly differs from traditional collocation methods since it does not measure the association of words to other words, but of words to syntactic patterns.

Since meaning of abstract constructions is understood to emerge from the meaning of its fillers, collostructional analysis contributes to the identification of the meaning range of constructions. In other words, using collostructional analysis helps to discover how a construction is used. Words that are found to be significantly attracted to the analysed constructions are called *collexemes*.

Collostructional analysis is composed of three types of methods: *simple, distinctive*, and *covarying* collexeme analysis. In this paper, simple and covarying collexeme analysis will be used. Simple collexeme analysis (Stefanowitsch and Gries 2003)(henceforth: SC) is the clearest reinterpretation of collocational analysis in a grammatical perspective. It measures the statistical co-occurrence relation of a lemma to a slot in a construction (typically an argument structure construction). Co-varying collexeme analysis (Stefanowitsch and Gries 2005)(henceforth: CC) is used instead to quantify the association of lemmas in one slot to lemmas in another slot of a single construction.

These two methodologies are employed on the four constructions selected (see ***Construction selection*** subsection). More specifically, simple collexeme analysis was carried out for NOM_PP, PART, and PASS. The investigated slots are respectively the deverbal noun, the present participle, and the main verb. Covarying collexeme analysis is instead performed for MOD, retrieving association strength for modal and main verb. Both analyses are conducted using the R package *collostructions* (Flach 2018).

Simple collexeme (SC) and Covarying collexeme (CC) analysis

To perform SC, CQL queries of the general constructions were performed on the web corpus tool *SketchEngine* and a frequency list of all the lemmas in the fillers under consideration was extracted. Data were then manually checked and cleaned from noise. The analysis was carried out on all lemmas occurring with a frequency equal to or higher than 5 for NOM_PP and PART, and on the first 100 occurrences for PASS. The final dataset consists of 50 occurrences for NOM_PP[6], 94 for PART, and 100 for PASS. SC requires a comparison between the frequency of the lemma in the construction and the frequency of the same lemma in the corpus, hence general frequencies for the selected lexical item were also retrieved with simple searches on *SketchEngine.* CC was conducted on the remaining construction MOD, to explore the attraction of modal + main verb in the construction. For the CC analysis, a frequency list of all the pairings of the two words in the two slots with their frequency of occurrence was retrieved. The list was manually cleaned and resulted in 1915 individual pairings of modal+ verb, and 494 significantly associated covarying collexemes. Appendix 1 reports the significant results for the two analyses.

Comparative analysis

Accessibility of CorIELLS: a comparison with the New General Service List

Having found the most significantly attracted lexical items to the 4 constructions, we explore the degree of lexical specialization in the collexemes (to answer RQ 1). To do this, we check each collexeme against the English core vocabulary in the *New General Service List* (Brezina and Gablasova 2015) (henceforth: NGSL). The NGSL is a list of ~2500 words obtained by comparing overlaps across four corpora (*Lancaster-Oslo-Bergen Corpus, British National Corpus, Corpus of British English*, and *EnTenTen12*). It aims to represent the core vocabulary of contemporary English, covering more than 80% of the text in the source corpora.

For this reason, we approximate absence from the NGSL as an indication of lexical specialization. Although the literature has acknowledged that the distinction between general and specialist lexicon is not straightforward (Bonin *et al.* 2010), this working distinction between highly accessible and less common lexicon is sufficient for the purpose at hand. Table 2 outlines the composition of the dataset and the results of the analysis in percentages.

Constructions	Collexemes	%presence	%absence
MOD	94	75.5%	24.5%
NOM_PP	50	66%	34%
PART	94	79.8%	20.2%
PASS	100	77%	23%

Table 2. Size of the dataset and percentages of presence/absence from NGSL

Results show that for 3 out of 4 constructions, between 20 and 25% of collexemes are not present in the NGSL, with PART being the most accessible (20.2% of specialized collexemes). NOM_PP instead shows a significantly higher percentage (34%) of specialized lexical items.[7]

The picture painted by these preliminary results is of a 'blended' genre: constructions are highly associated with accessible lexical items and highly specialized collexemes alike. This finding supports our hypothesis that decades of research on the accessibility of legal language has made LLL an autonomous and independent textual type, with idiosyncratic elements and lexico-grammatical features.

The nature of CorIELLS: a comparison with legal jargon and written prose

So far, the linguistic features of LLL have been discussed as they are found in CorIELLS. However, it is essential to also contrast LLL to other textual types to foreground how this genre is (or isn't) different from its 'parent' genre, specialized legal language.

Hence, we carry out a comparative analysis which contrasts LLL with two other genres: specialist legal jargon and general domain written prose. To do so, we use two specialized subcorpora: for legal language, an ad-hoc subcorpus of the *EurLEX* (Baisa et al., 2016) (Baisa *et al.* 2016) corpus including legislative documents in English ranging from the 90s to 2015 (henceforth: EUR)[8]; for general written language, the *BNC* imaginative subcorpus ((BNC Consortium 2007); henceforth: BNC_imag). Narrative was chosen as a proxy for non-specialist written prose since fiction is inherently aimed at

large and varied audiences, and hence the use of highly specialised registers is rare. At the same time, fiction is a written genre – akin in this sense to legal language and LLL alike. All the corpora were accessed via the *SketchEngine* web interface.

Data for the comparative analysis are the above mentioned statistically associated collexemes (see Simple collexeme (SC) and Covarying collexeme (CC) analysis and Appendix 1). Frequencies of the same lexico-syntactical patterns were retrieved from both BNC_imag and EUR using CQL queries. The boxplots in Figure 1 visually represent (log transformed) frequency distributions of collexemes in constructions across the three corpora.[9] As can be seen, NOM_PP and PASS (as abstract grammatical patterns) are used very similarly in LLL (in green) and specialized legal jargon (in blue). PART and MOD instead display idiosyncratic patterns of behaviour in each corpus.

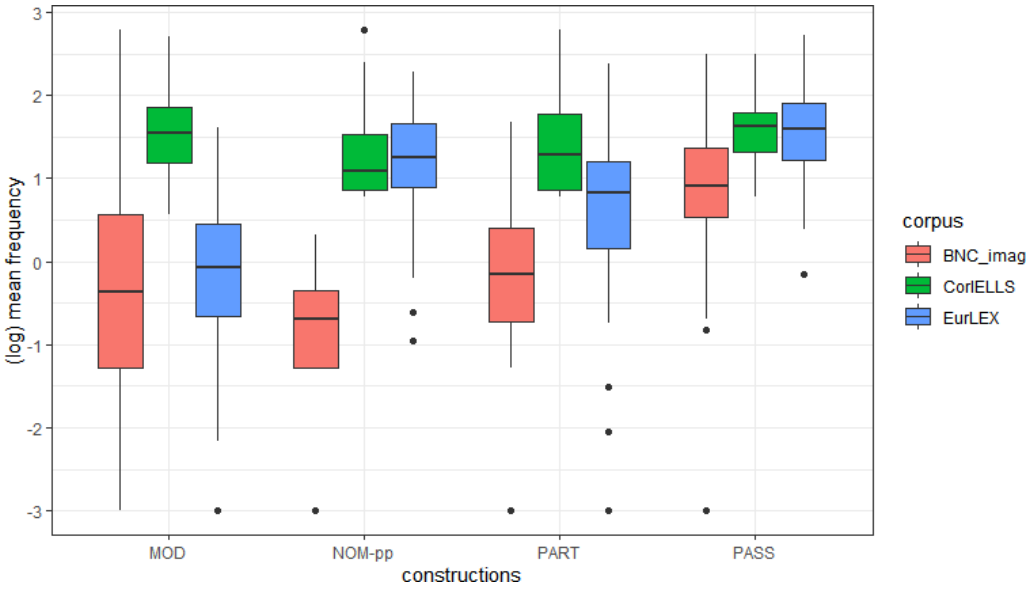

Figure 1. Frequency distributions for each construction

To test for the statistical significance of those trends, linear mixed effect modelling was used (Kuznetsova *et al.* 2017). Data were log-transformed to fit into a log-normal distribution. The model's predictors include corpora and construction type, in interaction types (i.e., in R syntax *corpus * construction*). The random intercept structure of the model includes the variable of collexeme – i.e., the different lexical items tested (in R syntax *(1| collexeme)*). Model selection was performed via Likelihood Ratio Test (Singmann *et al.* 2020). Contrasts for the variable of construction type were sum coded, i.e. "each coefficient compares the corresponding level of the factor to the average of the other levels" (Fox and Weisberg 2011: 130). In this way, the reference level for the variable (the intercept) is the overall average value for the predictor. Since we do not have a theoretically driven motivation to compare all constructions to one specific construction, this choice is the most methodologically sound. For the variable of corpus, instead, EUR was chosen as the reference level, as we are interested in analysing LLL as compared to specialized legal jargon. Therefore, all levels of all variables are statistically contrasted to average frequency mean of the four constructions in the EUR corpus.
Findings (see Table 3[10]) confirm that CorIELLS displays general overall frequency

patterns which are significantly different from legal language and written language alike, and three constructions show with idiosyncratic behaviour with respect to legal jargon.

corpus	Predictors	Estimates	CI	p
EUR	(Intercept)	0.81***	0.72 – 0.91	**<0.001**
	MOD	-0.97***	-1.13 – -0.81	**<0.001**
	NOM-pp	0.41***	0.21 – 0.61	**<0.001**
	PART	-0.15*	-0.30 – 0.00	0.05
	PASS	0.71***	0.56 – 0.86	**<0.001**
BNC imag	MOD	0.76***	0.55 – 0.96	**<0.001**
	NOM-pp	-1.24***	-1.50 – -0.98	**<0.001**
	PART	0.06	-0.14 – 0.26	0.6
	PASS	0.43***	0.22 – 0.63	**<0.001**
	MOD	1.08***	0.88 – 1.28	**<0.001**
Coriells	NOM-pp	-0.57***	-0.83 – -0.31	**<0.001**
	PART	0.12	-0.08 – 0.33	0.2
	PASS	-0.64***	-0.84 – -0.43	**<0.001**
Comparison EUR-BNC_imag and EUR-Coriells	BNC_imag	-1.09***	-1.21 – -0.96	**<0.001**
	CorIELLS	0.62***	0.49 – 0.74	**<0.001**
Marginal R2 / Conditional R2: 0.489 / 0.561				

Table 3. Results of the statistical model

Particularly, in LLL modal verbs constructions (MOD) are used significantly more than in legal language (as shown by the absence of a negative sign in the estimates column), while nominalisations (NOM_PP) and passive constructions (PASS) are used significantly less. Since both NOM_PP and PASS are highly characteristic of specialised legal language, the result confirms that there are structural differences between the grammar of legal language and the grammar of language with legal content directed at a wider audience.

These findings align with our hypotheses: LLL exhibits lexico-grammatical features which are not totally ascribable to specialist legal jargon. Subcategorization preferences for the sample of constructions considered here point to a 'blended' genre, a result which is comparable with findings on Italian using the same procedure outlined in Busso (2022).

Readability of CorIELLS: is LLL more readable?

The analysis carried out on lexico-grammatical properties of LLL has provided preliminary evidence for our hypothesis of LLL as an independent and 'blended' genre between specialist legal jargon and general written prose.

We further tested this hypothesis by conducting an exploratory analysis of the *readability* of LLL with respect to the other 2 genres (specialist legal jargon and general domain written prose). Readability is here defined – following the literature – as "how easily written materials can be read and understood" (Richards and Schmidt 2013). Therefore, our definition of readability relates to text comprehension rather than processing (e.g., Kate *et al.* 2010).

To investigate text comprehension we employ readability metrics, which are widely used in the scientific literature (and beyond) to assess the reading ease/difficulty of a

document. Readability measures are a useful tool, although their theoretical foundations are considered to be weak (Davison and Kantor 1982). Generally, these metrics rely on superficial text-based features such as number of words per sentence, or number of characters or syllables per word – as a proxy of respectively syntactic and lexical complexity. While both important components of readability, sentence and word length are by no means exhaustive measures of readability, which comprises several other features such as cohesion, lexical sophistication, and discourse structures (Snow 2002; Crossley *et al.* 2008).

However, a number of studies report strong correlations with text comprehension criteria (Chall and Dale 1995), and have been adopted vastly in academia and beyond. Such formulas are manifold, with well over 200 different readability scores developed since the 1920s (DuBay 2004).Particularly, different fields in linguistics have variously applied a multitude of readability formulas: L2 learning (Crossley *et al.* 2011; Xia *et al.* 2019), NLP (François and Miltsakaki 2012; Crossley *et al.* 2019; Smeuninx *et al.* 2020), psycholinguistics (Dębowski *et al.* 2015; Howcroft and Demberg 2017), language teaching (Carrell 1987; Zalmout *et al.* 2016), etc. Given that readability scores have been proven useful in research despite being far from perfect measures (Conklin *et al.* 2019), we here employ classic readability scores that will be compared to native speakers' judgments to compare text-based measures of comprehension with data collected from actual speakers.

Three readability indexes were chosen: the *Flesch-Kincaid formula* (henceforth: FK, (Flesch 1979), the *Automated Reading Index* (henceforth: ARI, (Senter and Smith 1967)), and the *Coleman-Liau Index* (henceforth: Col, (Coleman and Liau 1975)). The reason for using these particular scores is their cross-comparability, as they all employ a numerical scale based on the American school system: the higher the value, the more years of education are allegedly required to understand a given text (Figure 2).

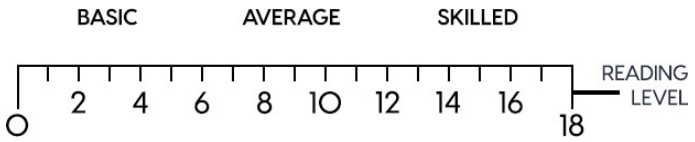

Figure 2. Grade levels used for the 3 readability measures (adapted from readable.com)

For this part of the analysis, a random sample of 20 concordances instantiating each construction was selected (total: 80 concordance lines per corpus, 240 in total). The concordances were chosen using the GDEx function on SketchEngine[11], and manually refined to have concordances of comparable length (between 100 and 200 characters, mean length 156.9 characters). Moreover, the concordances often instantiate more than one construction at a time (see Table 4). This is inevitable when working with chunks of text and not with single sentences or clauses. However, data selection was careful to include concordances with an overwhelming majority of occurrences of one construction as instances of that construction.

Table 4 reports examples of concordance lines extracted from each corpus.

The three abovementioned readability measures (FK, ARI, and COL) were calculated for the whole dataset (Rinker 2020).

	BNC_imag	**CorIELLS**	**EUR**
MOD	I thought Mr. Braden should be reminded that there were ladies present, but instead I said, "I don't know if the ladies enjoy this kind of talk very much."	If there is any inconsistency between this Part A and any other Part of these Terms and Conditions the provisions of that Part shall prevail.	For those purposes, the certification body may accompany the paying agency when it carries out secondary level on-the-spot checks.
NOM_pp	He'd appeared confident of meeting his commitments with the tourists at lunchtime at The Randolph, and then again during the afternoon.	Directive on the strengthening of certain aspects of the presumption of innocence and of the right to be present at the trial in criminal proceedings.	Agreement on the Accession of the Republic of Austria to the Convention implementing the Schengen Agreement of 14 June 1985.
PART	It was now a warm, clear night with just a soft breeze rustling the ropes and canvas of the small boats berthed in the marina far below.	We will organise a day for installation which is convenient for both of us and we will send you a letter confirming the date of your engineer appointment.	Thus, a horizontal law implementing a European directive would take precedence over conflicting provisions contained in national legislation.
PASS	Doreen was the type of girl who always sounded as though her nasal passages were obstructed or her throat sore.	Details of your normally available download speed and minimum download speed will have been provided to you at point of sale.	Any moneys recovered from loan losses for which payment has been made under guarantees called shall be credited to the Trust Account.

Table 4. Concordances examples for all four constructions

The three sets of readability scores were averaged to obtain a "meta-measure". Figure 3 plots the distribution of the averaged readability scores per corpus. As can be easily seen, the intermediate 'mixed' character of LLL seems to hold also in terms of readability, although the median value for both CorIELLS and EUR is very high (respectively, 13.9 and 14.9) with respect to BNC_imag (10.5).

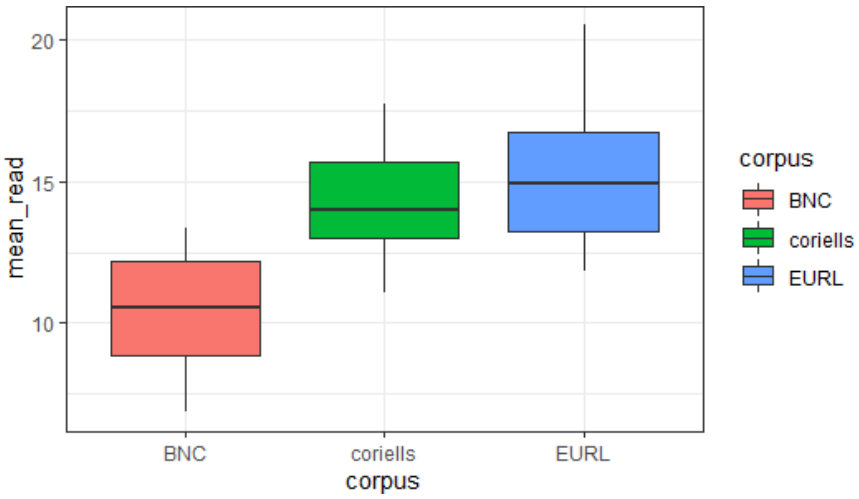

Figure 3. Boxplot of (averaged) readability per each corpus

Statistical significance was again assessed with linear mixed effect modelling, using LRT for model selection. The final model includes the variable of corpus (sum coded) as a predictor and the variable of concordance as the intercept random factor.

Table 5 outlines estimates for fixed effects and Figure 4 plots such estimates.

Predictors	Estimates	CI	standardized CI	p
(Intercept)	13.35***	12.81 – 13.90	-0.16 – 0.21	**<0.001**
BNC_imag	-2.79***	-3.55 – -2.04	-1.22 – -0.70	**<0.001**
CorIELLS	0.88*	0.08 – 1.68	0.03 – 0.58	**<0.05**
EUR	1.91***	1.16 – 2.67	0.40 – 0.92	**<0.001**

Table 5. Results for the model

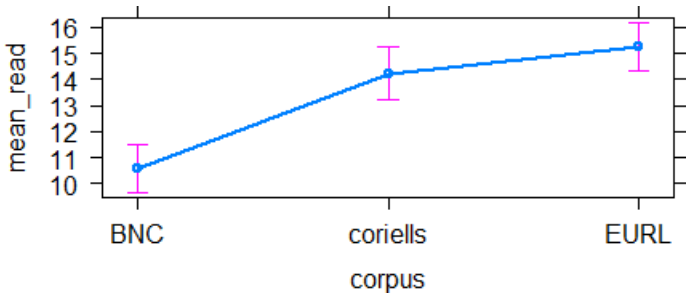

Figure 4. Estimates plot from the model

The statistical model confirms the trend found in the raw data: LLL shows a readability 0.88 grades higher than the overall mean, while legal jargon requires almost 2 grades more to be understood, while general written prose almost 3 grades less.

Difficulty of CorIELLS: can speakers understand LLL?

Readability metrics are a useful proxy for text comprehension and – to some extent – lexical and syntactical complexity. However, as we have seen in the previous paragraph, they have several limitations. Therefore, we compare the text-based analysis of readability with data collected from native speakers of English, which were presented a survey using a selection of the same concordance lines. Specifically, a random sample of concordance lines was extracted from the dataset used for the readability analysis. A total of 80 stimuli (20 BNC_imag + 20 EUR + 40 CorIELLS, 10 per subcorpus) was selected.[12] The stimuli for the survey are sentences between 100 and 200 characters long (normalised per length, mean= 156.8, st. dev= 21.6) that include one (or more) of the grammatical constructions analysed. Similarly to Table 4 above, examples from all 3 corpora and for all 4 constructions are reported in Table 6.

The survey was presented to 50 native British English speakers using the *SurveyMonkey Audience* platform[13]. Due to non-completion of task, data from 7 subjects had to be excluded from all following analyses, leaving a total of 43 participants (24 F, 19 M, median age group:31-45).

Before the survey, informed consent and a brief sociolinguistic questionnaire asking for information on gender, age, and education level were presented.[14] Stimuli were preceded by the following instructions:

> "How difficult to understand is the following sentence(s)? Use the slider to indicate how complex and difficult to understand you find the following texts, and list from 1 to 4 how well you think you understood its meaning; keep in mind that you will be show excerpts of longer texts."

Participants were then presented with the stimuli in random order, and ratings were formulated against a graded scale from 1 to 100.

Mixed models were chosen once again here as a statistical technique to control for the random effect of participants and stimuli selection. During model selection via LRT, the effect of the different constructions was found to be non-significant ($p = .48$), hence the final model only includes the predictor of corpus ($p < .0001$), with CorIELLS set as the intercept level. Ratings are log-transformed to fit a normal distribution. Random structure includes intercepts for both participant and stimuli (in R syntax, *(1|participant) + (1|stimuli)*).

Not surprisingly, results are in line with all previous analyses: general-domain prose appears to be significantly less difficult than LLL (-0.4), and legal jargon significantly more difficult (0.15) (see Table 7). This is somewhat an expected result, but still important in itself: native speakers' intuition and text comprehension confirms the corpus-based analyses described in the previous paragraphs.

Bridging the gap between readability and speakers' judgments

To compare results from the two analyses, readability scores and difficulty ratings were normalised on a common scale from 1 to 10. Figure 5 plots the (aggregated) normalised results.

The raw data from both experiments (text-based readability and human judgments) show very similar trends. However, to see if the descriptive trend can be generalised, a two- way ANOVA was carried out, with experimental condition (i.e., survey or

	BNC_imag	**CorIELLS**	**EUR**
MOD	She couldn't take their mother's place, of course, but for Liz's sake she must try to do everything she possibly could for the little girls.	The Content you submit must not include third- party intellectual property such as copyrighted material unless you have permission from that party or are otherwise legally entitled to do so.	The authorities of the Côte d'Ivoire shall communicate, before the entry into force of the Agreement, all information concerning the bank account to be used for the payment of the fees.
NOM_ pp	Even so, it was plain from the mixture of resentment and hostility on his face that her words had wounded him.	Provisions on the application and development of the Schengen acquis, relating to the abolition of checks at internal borders and movement of persons.	The Commission has examined France's application for the approval of amendments to the specification of the protected designation of origin 'Olives noires de Nyons'.
PART	The three boys sat under heavy guard in a glow-globe-lit room hung with a tapestry depicting the march across the wastes three centuries earlier.	Our 5G services may be affected by the number of people using the 5G service, maintenance and upgrades, faults from other networks, the weather, other environmental factors or degradation.	The data are based upon the "special trade" system, according to which, external trade comprises goods crossing the customs border of the country.
PASS	In the first crime, he had been robbed of something on which he had set his heart, in the second he was robbed of his life.	Any claim dispute or matter arising under or in connection with this User Agreement shall be governed and construed in all respects by the laws of England and Wales.	Bee-keeping products can only be sold as organic products if the general conditions on feeding, care and housing have been observed for at least one year.

Table 6. Example stimuli for all constructions in the 3 corpora

Predictors	Estimates	CI	standardized CI	p
CorIELLS	1.72	1.67 – 1.77	1.67 – 1.77	**<0.001**
BNC_imag	-0.39	-0.48 – -0.31	-0.48 – -0.31	**<0.001**
EUR	0.15	0.07 – 0.24	0.07 – 0.24	**<0.001**
Marginal R2 / Conditional R2 0.300 / 0.512				

Table 7. Results of the model

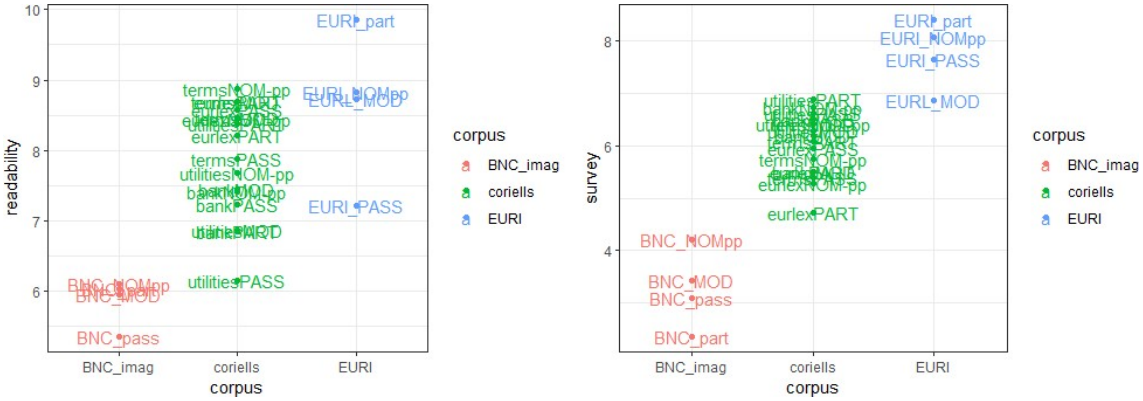

Figure 5. Normalised aggregated readability scores and difficulty ratings from the survey

readability) in interaction terms with the variable of corpus. ANOVA was chosen as a statistical method to estimate how the mean of readability and human judgments' scores is affected by the levels of the two independent categorical variables "experimental condition" and "construction".

Main effects report a significant difference between survey and readability data (F value =74.5, p-value= <.0001) and across corpora (F value=53, p-value= <.0001). A marginally significant effect is also found in the interaction between the two variables (F value =2.7, p-value= .07). Hence, readability measures seem to underestimate the accessibility of texts, as the higher estimates indicate (see Figure 5).

A *post-hoc* Tukey HSD test reveals that pairwise comparisons of corpora across conditions reach statistical significance for BNC_imag and CorIELLS (Table 7). In other words, readability scores are significantly higher for both written prose and LLL (Figure 6, Table 8), but no difference is found in the assessment of legal jargon. Here, we hypothesize that the higher accuracy of reading metrics in evaluating legal language with respect to LLL and general domain prose could lie in the 'tuning' of the metrics themselves. In fact, readability scores have been traditionally employed to analyse the accessibility of highly specialist genres (Formisano 2015).

Conclusions

The present paper has presented some preliminary quantitative analyses on English legal-lay language (LLL) using an ad-hoc compiled specialised corpus, CorIELLS. Several types of analysis were carried out on a sample of 4 lexico-grammatical constructions (Goldberg 2019): nominalisations heading prepositional phrase attachments, modal verb constructions, participial reduced relative constructions, and passive constructions.

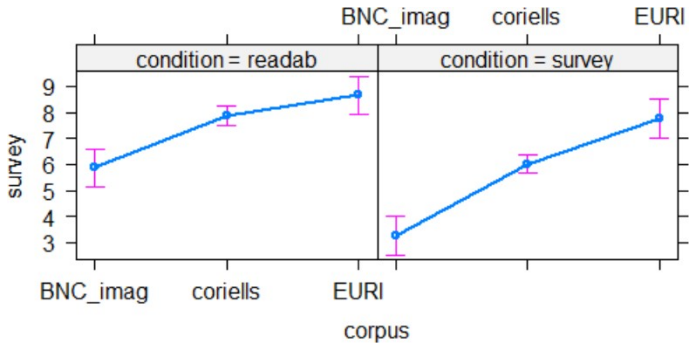

Figure 6. Effects of the ANOVA

Corpus (comparison survey*readability)	**Difference**	**lower**	**upper**	**p-value**
BNC_imag	-0.6	-0.85	-0.35	**<.0001**
CorIELLS	-0.27	-0.39	-0.14	**<.0001**
EUR	-0.11	-0.36	0.14	0.8
Marginal R2 / Conditional R2 0.300 / 0.512				

Table 8. Relevant pairwise comparisons of the post-hoc Tukey HSD test

A first exploratory part of the study set out to examine specifically the lexico-grammatical features of legal-lay English. The subcategorization preferences of the selected constructions were investigated using simple and covarying collexeme analysis (Stefanowitsch 2013). Collexemes that were found to be significantly associated with each construction were then checked against the NGSL (Brezina and Gablasova 2015) to determine the degree of specialisation of LLL. Findings indicate that NOM_pp is the construction with the most specialist lexicon out of the four constructions (34% of terms do not present in the core vocabulary).

The second part of the study aims to compare LLL to specialist legal jargon and general-domain written prose. Specifically, the frequency of statistically associated collexemes found in CorIELLS was compared with the frequency of the same structures in two other specialised corpora: the BNC imaginative subcorpus (BNC, 2007), and the English version of *EurLEX* (Baisa *et al.* 2016). Results support our hypothesis that LLL displays linguistic features quantitatively different from the other two genres. Findings point to LLL being a 'blended' genre, similarly to what was found for Italian (Busso 2022) A similar result is obtained by analysing readability scores of a sample of concordances. Interestingly, the pattern holds true also in native speakers' judgments of the same set of concordances. Even though the pattern is the same, the statistical comparison of survey responses and readability scores indicates that speakers consider legal-lay language more accessible than text-based metrics seem to suggest.

To conclude, the present study has presented the first quantitative in-depth exploration of legal-lay language, taking both a corpus-driven and an experimental perspective. Specifically, the investigation of lexico-grammatical characteristics of LLL suggests that it possesses idiosyncratic charactersistics that differentiate it from specialist legal language and general-domain written language: idiosyncratic lexical choices, and intermediate readability and comprehensibility. A comparison of text-based

readability and survey data also suggests that readability metrics might underestimate the readers' ability to understand texts. However, further research in this direction is needed to confirm this preliminary finding.

Notes

[1] Different constructions differ for *schematicity*, and are distributed on a gradient cline ranging from lexical items to abstract argument structure patterns:

a. Lexical level: Word e.g. avocado, anaconda, and

b. Complex word e.g. daredevil, shoo-in

c. Complex word (partially filled) e.g. [N-s] (for plurals)

d. Phrase level: Idiom (filled) e.g. give the devil his dues, going great guns

e. Clause level: Idiom (partially filled) e.g. jog <someone's> memory

f. Covariational Conditional [The Xer the Yer] e.g. the more you think about it, the less you understand

g. Discourse level: Ditransitive (double object) [Subj V Obj1 Obj2] e.g. he gave her a fish taco, Passive [Subj aux VPPP (PPby)] e.g. the armadillo was hit by a car

[2] The corpus is freely available online on the Forensic Linguistic Databank (Petyko *et al.* 2022) https://fold.aston.ac.uk/

[3] These documents are originally drafted in English and later adapted by specialised translators and legal experts in each language of the European Union, as prescribed in EU style guides (Inter institutional Style Guide, 2015:54-62)

[4] All following examples are taken from *CorIELLS*.

[5] The CQL searches for the 4 constructions are as follows:

MODAL: [tag="MD"] [] {0,1} [tag="V.*" & tag!="VVN"]

NOMINALIZATIONS PP CHAIN: [tag= "N.*"] [tag= "IN/that|IN" & word!="and"] []?[tag="N.*"] [tag!= "SENT|SYM"] {0,2} [tag= "IN/that|IN"& word!= "and"] [ ]? [tag="N.*"] [tag!= "SENT|SYM"] {0,2} [tag="IN/that|IN" & word!="and"]

PARTICIPIAL: [tag= "NN.*"] [tag="VHG|VBG|VVG"] [tag= "DT"]

PASSIVE: [tag= "MD"]? [tag= "VB.*|VH.*"] [word= "been|being"]? [tag="VVN.*"]

[6] The general CQL simply retrieves nouns. Deverbal nouns were manually selected from the general frequency list.

[7]Interpretation of these findings was done bearing in mind that due to sparsity of linguistic data, it is inevitable for core vocabulary to cover a high percentage of the lexicon (Zipf 1949).

[8]Unfortunately, there is no easy way of knowing which of the documents in the EurLEX corpus were initially drafted in English and which one were translated from another official language.

[9]Boxplots represent data range in quartiles. The black line that divides the box into two parts is the median value (middle quartile), which marks the "mid-point" of the data. Half the frequency values are greater than or equal to this value and half are less. The first and fourth quartile are represented as the "whiskers" of the plot, while the second and third quartiles by the box.

[10] Adjusted R2 values are automatically retrieved with the package *sjplot* (Lüdecke 2021). R2 values describe the amount of variance in the data that is explained by the model. In this case, more than 56% of the variance is explained by the predictors in this model.

[11] GDEX stands for Good Dictionary Examples, a function the user can select in KWIC searches in Sketch Engine. GDEX automatically identifies sentences that are illustrative and representative of the query.

[12]The survey presented a subset of all concordance lines to avoid fatigue in participants and promote completion of the task.

[13]Available online at http://www.surveymonkey.com

[14]Sociolinguistic variables will be explored in further research, but for the purposes of this study, we will only consider corpus and construction as independent variables of interest.

References

Adler, M. (2012). *The Plain Language Movement*. Oxford: Oxford University Press.

Aher, M. (2013). *Modals in Legal Language*. University of Osnabrück.

Almut, K. (2010). Building small, specialised corpora. In O. A. and M. M, Eds., *The Routledge handbook of corpus linguistics*. London: Routledge, 66–79.

Baisa, V., Michelfeit, J., Medveď, M. and Jakubíček, M. (2016). European Union language resources in sketch engine. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2799–2803.

Baroni, M. and Bernardini, S. (2004). BootCaT: Bootstrapping corpora and terms from the web. In M. Lino, Ed., *Proceedings of the Fourth Language Resources and Evaluation Conference, LREC2004 Lisbon*, 1313–1316.

Bhatia, V. (1983). Simplification v. easification— the case of legal texts. *Applied Linguistics*, 4(1), 42–54.

Bhatia, V. (1993). *Analyzing Genre: Language Use in Professional Settings*. London: Longman.

Biber, D. (1993). Representativeness in corpus design. *Literary and linguistic computing*, 8(4), 243–257.

BNC Consortium, (2007). *The British National Corpus*. XML Edition, Oxford Text Archive.

Bonin, F., Dell'Orletta, F., Venturi, G. and Montemagni, S. (2010). Singling out legal knowledge from world knowledge. an nlp–based approach". In *Proceedings of the 4th Workshop on Legal Ontologies and Artificial Intelligence Techniques*, 39–50.

Brezina, V. and Gablasova, D. (2015). Is there a core general vocabulary? introducing the new general service list. *Applied Linguistics*, 36(1), 1–22.

Bulatović, V. (2013). Legal language: The passive voice myth. *ESP Today*, 1(1), 93–112.

Busso, L. (2022). Lexicon and grammar in legal-lay language: a quantitative corpus study on Italian. *Studi Italiani di Linguistica Teorica e Applicata (SILTA)*, LI(1), 5–33.

Busso, L. (forthcoming). CorIELLS: a specialised bilingual corpus of English and Italian legal-lay language. In *Presentation at the workshop Forensic linguistics between scientific research and legal practice, LIV International Conference of the Italian Linguistics Society*.

Bybee, J. (2015). *Language change*. Cambridge: Cambridge University Press.

Cappelle, B. and Depraetere, I. (2016). Modal meaning in construction grammar. *Constructions and Frames*, 8(1), 1–6.

Carrell, P. (1987). Readability in esl. *Reading in a Foreign Language*, 4, 21–40.

Chall, J. and Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Cambridge, MA: Brookline Books.

Charrow, R. and Charrow, V. (1979). Making legal language understandable: A psycholinguistic study of jury instructions. *Columbia law review*, 79, 1306–1347.

Chovanec, J. (2013). Grammar in the law. In C. Chappelle, Ed., *The Encyclopedia of Applied Linguistics*. Oxford: Blackwell, 1– 8.

Coleman, M. and Liau, T. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, 60, 283–284.

Conklin, K., Hyde, R. and Parente, F. (2019). Assessing plain and intelligible language in the Consumer Rights Act: a role for reading scores? *Legal Studies*, 39(3), 378–397.

Coulthard, M. and Johnson, A. (2007). *An introduction to forensic linguistics: Language in evidence*. London: Routledge.

Crossley, S., Allen, D. and McNamara, D. (2011). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a foreign language*, 23(1), 84–101.

Crossley, S., Greenfield, J. and McNamara, D. (2008). Assessing text readability using cognitively based indices. *Tesol Quarterly*, 42(3), 475–493.

Crossley, S., Skalicky, S. and Dascalu, M. (2019). Moving beyond classic readability formulas: New methods and new models. *Journal of Research in Reading*, 42(3-4), 541–561.

Davison, A. and Kantor, R. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly*, 17(2), 187–209.

DuBay, W. (2004). The principles of readability, impact information. http://www.nald.ca/library/research/readab/readab.pdf.

Dębowski, , Broda, B., Nitoń, B. and Charzyńska, E. (2015). Jasnopis–a program to compute readability of texts in Polish based on psycholinguistic research. In B. Sharp, W. Lubaszewski and R. Delmonte, Eds., *Natural Language Processing and Cognitive Science*. Venice: Libreria Editrice Cafoscarina, 51–61 ,.

T. Fanego and P. Rodríguez-Puente, Eds. (2019). *Corpus-based research on variation in English legal discourse, 91*. Amsterdam: John Benjamins.

Flach, S. (2018). Collostructions: An R implementation for the family of collostructional methods. R package version 0.1.2.

Flesch, R. (1979). *How to write in plain English: A book for lawyers and consumers*. New York: Harper.

Formisano, V. (2015). XVIII. new methodological approaches to readability corpus linguistics and dictionaries of science and technology. In G. Dotoli, C. Saggiomo, R. Spiezia and C. Boccuzzi, Eds., *La lisibilité du dictionnaire*. Paris: Hermann, 273–289.

Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression*. New York: Sage Publications.

Frade, C. (2007). Power dynamics and legal English. *World Englishes*, 26(1), 48–61.

Frankenreiter, J. and Livermore, M. (2020). Computational methods in legal analysis. *Annual Review of Law and Social Science*, 16, 39–57.

François, T. and Miltsakaki, E. (2012). Do NLP and machine learning improve traditional readability formulas? In *Proceedings of the First Workshop on Predicting and Improving Text Readability for target reader populations*, 49–57.

Givón, T. (1993). *English grammar: A function-based introduction*, volume 2. Amsterdam: John Benjamins.

Goldberg, A. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Goldberg, A. (2019). *Explain Me This*. Princeton: Princeton University Press.

Gotti, M. (2014). Linguistic insights into legislative drafting. *The Theory and Practice of Legislation*, 2(2), 123–143.

Goźdź Roszkowski, S. and Pontrandolfo, G. (2015). Legal phraseology today: corpus-based applications across legal languages and genres. *Fachsprache: Internationale Zeitschrift für Fachsprachenforschung- didaktik und Terminologie*, 37(3), 130–139.

S. Goźdź-Roszkowski and G. Pontrandolfo, Eds. (2017). *Phraseology in Legal and Institutional Settings: A Corpus-based Interdisciplinary Perspective*. London: Routledge.

Gries, S. (2013). Data in construction grammar. In T. Hoffmann and G. Trousdale, Eds., *The Oxford handbook of Construction Grammar*. Oxford: Oxford University Press, 93–108.

Groom, N. (2019). Construction grammar and the corpus-based analysis of discourses: The case of the WAY IN WHICH construction. *International Journal of Corpus Linguistics*, 24(3), 291–323.

Haapio, H. (2011). Contract clarity through visualization: Preliminary observations and experiments. In *Proceedings of the 15th International Conference on Visualization*: IEEE Computer society.

Haigh, R. (2013). *Legal English.* London: Routledge.

Hamann, H., Vogel, F. and Gauer, I. (2016). Computer assisted legal linguistics. In F. Bex and S. Villata, Eds., *Legal Knowledge and Information Systems: JURIX 2016: The Twenty-Ninth Annual Conference*, 195–198, Amsterdam: IOS Press.

Hilpert, M. (2013). *Constructional change in English: Developments in allomorphy, word formation, and syntax.* Cambridge: Cambridge University Press.

Howcroft, D. and Demberg, V. (2017). Psycholinguistic models of sentence processing improve sentence readability ranking. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, volume 1, 958–968: Long Papers.

Jaeggli, O. (1986). *Passive.* Linguistic inquiry.

Janigová, S. (2008). *Syntax of-ing Forms in Legal English*, volume 439. Bristol: Peter Lang.

Kate, R., Luo, X., Patwardhan, S., Franz, M., Florian, R. and Mooney, R. (2010). Learning to predict readability using diverse linguistic features. In *Proceedings of the 23rd International Conference on Computational Linguistics*, 546–554: Association for Computational Linguistics.

Kilgarriff, A., Rychlý, P., Smrz, P. and Tugwell, D. (2004). *Itri-04-08 The Sketch Engine.* Information Technology.

Kuznetsova, A., Brockhoff, P. and Christensen, R. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.

Langacker, R. (2013). Modals: Striving for control. In J. Marín-Arrese, M. Carretero, J. Hita and J. Auwera, Eds., *English modality: Core, periphery and evidentiality.* Amsterdam: De Gruyter Mouton, 3– 55.

Lintao, R. and Madrunio, M. (2015). Analyzing the lexical structures of a Philippine consumer-finance contract. *Journal of Teaching English for Specific and Academic Purposes*, 2(3), 359–370.

Lüdecke, D. (2021). sjPlot: Data visualization for statistics in social science. R package version 2.8.7. https://CRAN.R-project.org/package=sjPlot.

Petyko, M., Atkins, S., Busso, L. and Grant, T. (2022). The forensic linguistic databank. *Journal of Language and Law.*

Pollard, C. and Sag, I. (1994). *Head-driven phrase structure grammar.* Chicago: University of Chicago Press.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language.* London: Longman.

Richards, J. and Schmidt, R. (2013). *Longman dictionary of language teaching and applied linguistics.* London: Routledge.

Rinker, T. (2020). qdap: Quantitative discourse analysis package 2.4.2. https://github.com/trinker/qdap.

Romm, T. (2018). Facebook didn't read the terms and conditions for the app behind Cambridge Analytica. *The Washington Post*. https://www.washingtonpost.com/news/the-switch/wp/2018/04/26/facebook-didnt-read-the-terms-and-conditions-for-the-app-behind-cambridge-analytica/ accessed on 31/3/2022.

Saussure, F. (1916). *Course in general linguistics*. London: Duckworth.

Schiess, W. (2007). The art of consumer drafting. *Scribes Journal of Legal Writing*, 11, 1–24.

Senter, R. and Smith, E. (1967). Automated readability index. Amrl-Tr. aerospace medical research laboratories (u.s). *Wright-Patterson Air Force Base*, 1–14.

Singmann, H., Bolker, B., J., A., F., and Ben-Shachar, M. (2020). afex: Analysis of factorial experiments. R package version 0.28-0. https://CRAN.R-project.org/package=afex.

Smeuninx, N., Clerck, B. and Aerts, W. (2020). Measuring the readability of sustainability reports: A corpus-based analysis through standard formulae and NLP. *International Journal of Business Communication*, 57(1), 52–85.

C. Snow, Ed. (2002). *Reading for understanding: Toward an R & D program in reading comprehension*. Santa Monica, CA: Rand.

Stefanowitsch, A. (2013). Collostructional analysis. In T. Hoffmann and G. Trousdale, Eds., *The Oxford handbook of construction grammar*. Oxford: Oxford University Press.

Stefanowitsch, A. and Gries, S. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–43.

Stefanowitsch, A. and Gries, S. (2005). Co-varying collexemes. *Corpus Linguistics and Linguistic Theory*, 1(1), 1–43.

Tiersma, P. (1999). *Legal language*. Chicago: University of Chicago Press.

Torres–Martínez, S. (2019). Taming english modals. *English Today*, 35(2), 50–57.

Van Boom, W., Desmet, P. and Dam, M. (2016). If it's easy to read, it's easy to claim'—the effect of the readability of insurance contracts on consumer expectations and conflict behaviour. *Journal of Consumer Policy*, 39(2), 187–197.

Williams, C. (2004). Legal English and plain language: an introduction. *ESP Across Cultures*, 1, 111– 124.

Williams, C. (2010). Functional or dysfunctional? the language of business contracts in english. *Rassegna Italiana di Linguistica Applicata*, 217–227.

Williams, C. (2013). Changes in the verb phrase in legislative language in english. In B. Aarts, J. Close, G. Leech and S. Wallis, Eds., *The Verb Phrase in English: Investigating Recent Language Change with Corpora (Studies in English Language)*. Cambridge: Cambridge University Press, 353–371 ,.

Xia, M., Kochmar, E. and Briscoe, T. (2019). Text readability assessment for second language learners. arXiv preprint arXiv:1906.07580.

Yokoyama, S., Okamoto, H., Miyamoto, T., Yoshimoto, K., Kim, J., Iwata, K. and Kawashima, R. (2006). Cortical activation in the processing of passive sentences in L1 and L2: An fMRI study. *Neuroimage*, 30(2), 570–579.

Yunus, K. and Ab Rashid, R. (2016). Colligations of prepositions: Essential properties of legal phraseology. *International Journal of Applied Linguistics and English Literature*, 5(6), 199– 208.

Zalmout, N., Saddiki, H. and Habash, N. (2016). Analysis of foreign language teaching methods: An automatic readability approach. In *Proceedings of the 3rd workshop on natural language processing techniques for educational applications*, volume NLPTEA2016, 122–130.

Zipf, G. K. (1949). *Human behavior and the principle of least effort.* Boston: Addison-Wesley.

Ződi, Z. (2019). The limits of plain legal language: understanding the comprehensible style in law. *International Journal of Law in Context*, 15(3).

Appendix 1: significantly associated collexemes from SC and CC analysis

Simple Collexeme Analysis

CONSTRUCTION	COLLEXEME	CORPUS. FREQ.	OBS.	EXP.	COLL.STR. (LOGL)	SIGNIF
NOM-pp	use	3518	507	48.1	1553.591	*****
	charge	610	204	8.3	988.6324	*****
	purchase	231	67		301.9074	*****
	transfer	412	59	5.6	178.0778	*****
	refund	178	38	2.4	145.6235	*****
	display	171	32	2.3	113.7878	*****
	change	603	48	8.2	92.50596	*****
	accordance	532	34	7.3	52.88384	*****
	conclusion	112	16	1.5	48.1655	*****
	protection	334	23	4.6	38.60167	*****
	report	155	16	2.1	38.25826	*****
	processing	164	14	2.2	28.66795	*****
	access	671	29	9.2	27.75447	*****
	payment	2093	60	28.6	26.67419	*****
	application	484	22	6.6	22.63027	*****
	consideration	31	6	0.4	21.74256	*****
	relief	36	6	0.5	19.90199	*****
	impact	78	8	1.1	19.02998	****
	transmission	79	8	1.1	18.84251	****
	provision	496	20	6.8	17.21279	****
	booking	144	10	2	16.91684	****
	assistance	97	8	1.3	15.89314	****
	notice	919	28	12.6	14.30254	***
	obligation	557	20	7.6	14.15943	***
	indemnification	39	5	0.5	13.99496	***
	information	2859	64	39.1	13.58679	***
	participation	75	6	1	11.59979	***
	connection	532	18	7.3	11.40833	***
	assessment	68	5	0.9	8.93997	**
	accommodation	82	5	1.1	7.38728	**
	loss	628	17	8.6	6.52525	*
	procedure	300	10	4.1	6.15402	*
	notification	186	7	2.5	5.37681	*
	supplier	340	10	4.6	4.71098	*
	agreement	2352	45	32.1	4.65463	*

CONSTRUCTION	COLLEXEME	CORPUS. FREQ.	OBS.	EXP.	COLL.STR. (LOGL)	SIGNIF
	addition	167	6	2.3	4.24951	*
	restriction	173	6	2.4	3.98334	*
PART	make	2264	519	30.9	2088.002	*****
	require	704	277	9.6	1452.694	*****
	arise	370	211	5.1	1314.178	*****
	relate	514	231	7	1288.317	*****
	provide	2482	369	33.9	1149.601	*****
	govern	260	164	3.6	1070.565	*****
	regard	194	128	2.7	853.4201	*****
	pay	1615	260	22.1	849.2509	*****
	give	1144	211	15.6	746.8929	*****
	use	3518	326	48.1	722.0198	*****
	apply	1523	224	20.8	690.6339	*****
	take	1146	188	15.7	620.1748	*****
	send	738	157	10.1	601.9633	*****
	set	1025	169	14	558.9743	*****
	receive	1018	155	13.9	487.7473	*****
	process	324	90	4.4	396.9423	*****
	hold	384	84	5.2	326.5629	*****
	carry	353	75	4.8	286.8725	*****
	result	151	56	2.1	284.5307	*****
	enter	377	66	5.2	225.8185	*****
	remove	284	53	3.9	188.245	*****
	determine	173	42	2.4	172.5715	*****
	read	222	45	3	167.5523	*****
	request	312	51	4.3	167.3401	*****
	exclude	200	43	2.7	165.4476	*****
	grant	271	48	3.7	165.311	*****
	ask	556	63	7.6	161.7833	*****
	meet	219	42	3	151.5138	*****
	label	72	28	1	145.4475	*****
	confirm	196	38	2.7	137.9314	*****
	follow	626	60	8.6	135.4979	*****
	affect	321	44	4.4	128.9828	*****
	establish	349	45	4.8	126.5791	*****
	post	299	41	4.1	120.21	*****
	offer	319	42	4.4	119.8267	*****
	cover	307	41	4.2	118.1031	*****
	amend	313	41	4.3	116.5653	*****
	show	187	31	2.6	102.5493	*****
	depend	193	31	2.6	100.5764	*****
	share	261	33	3.6	91.55111	*****
	message	44	17	0.6	88.01939	*****
	contain	280	33	3.8	87.12635	*****
	display	283	32	3.9	81.98178	*****

CONSTRUCTION	COLLEXEME	CORPUS. FREQ.	OBS.	EXP.	COLL.STR. (LOGL)	SIGNIF
	allow	490	40	6.7	78.82668	*****
	handle	59	15	0.8	63.11487	*****
	originate	23	10	0.3	54.73058	*****
	operate	293	25	4	51.1886	*****
	ship	23	9	0.3	46.87404	*****
	remain	256	22	3.5	45.32067	*****
	include	2379	77	32.5	44.83004	*****
	act	582	32	8	42.07543	*****
	involve	140	16	1.9	41.29394	*****
	tamper	40	9	0.5	35.47752	*****
	implement	158	15	2.2	33.5704	*****
	block	116	13	1.6	33.07002	*****
	work	388	23	5.3	32.96131	*****
	fall	34	8	0.5	32.3056	*****
	belong	49	8	0.7	26.20415	*****
	go	244	16	3.3	25.54625	*****
	exploit	39	7	0.5	24.27617	*****
	comprise	20	5	0.3	20.84979	*****
	exceed	138	10	1.9	17.63473	****
	accompany	52	6	0.7	15.58929	****
	travel	52	6	0.7	15.58929	****
	seek	106	8	1.4	14.66188	***
	maintain	169	10	2.3	14.29531	***
	appear	67	6	0.9	12.79405	***
	build	48	5	0.7	12.03575	***
	host	54	5	0.7	10.96101	***
	copy	55	5	0.8	10.79606	**
	live	118	7	1.6	10.034	**
	report	120	7	1.6	9.84651	**
	begin	65	5	0.9	9.32652	**
	visit	137	7	1.9	8.4067	**
	indicate	74	5	1	8.22817	**
	enable	188	8	2.6	7.47512	**
	open	193	8	2.6	7.18361	**
	address	90	5	1.2	6.64826	**
	deal	95	5	1.3	6.22998	*
	describe	213	8	2.9	6.12416	*
	order	114	5	1.6	4.88416	*
	start	158	6	2.2	4.67928	*
	exercise	122	5	1.7	4.41124	*
PASS	entitle	243	241	3.3	2051.113	*****
	find	384	264	5.2	1798.968	*****
	deem	166	113	2.3	764.802	*****
	lose	155	103	2.1	688.8913	*****
	limit	420	117	5.7	517.032	*****

CONSTRUCTION	COLLEXEME	CORPUS. FREQ.	OBS.	EXP.	COLL.STR. (LOGL)	SIGNIF
	return	257	99	3.5	512.5376	*****
	base	265	97	3.6	490.106	*****
	bind	96	69	1.3	479.4927	*****
	consider	195	78	2.7	410.9374	*****
	terminate	365	88	5	360.5733	*****
	prohibit	126	60	1.7	342.8743	*****
	authorise	299	78	4.1	333.0243	*****
	register	278	75	3.8	325.7948	*****
	calculate	92	51	1.3	312.7697	*****
	accept	464	83	6.3	287.7518	*****
	add	211	61	2.9	274.3903	*****
	conduct	83	45	1.1	273.1023	*****
	resolve	175	54	2.4	250.9118	*****
	design	94	44	1.3	249.3809	*****
	bring	121	47	1.7	244.0819	*****
	place	173	52	2.4	238.4778	*****
	treat	92	41	1.3	227.1079	*****
	agree	1364	113	18.6	226.1232	*****
	oblige	49	33	0.7	221.9553	*****
	commit	82	37	1.1	206.1294	*****
	protect	234	52	3.2	203.775	*****
	issue	171	46	2.3	199.4166	*****
	delay	53	31	0.7	194.9057	*****
	close	250	50	3.4	184.7813	*****
	activate	60	31	0.8	183.9247	*****
	notify	392	59	5.4	183.8804	*****
	deliver	178	44	2.4	182.5272	*****
	convert	94	35	1.3	178.1086	*****
	obtain	198	44	2.7	172.3971	*****
	collect	331	51	4.5	161.295	*****
	store	151	38	2.1	159.109	*****
	record	82	31	1.1	158.891	*****
	intend	153	38	2.1	158.0131	*****
	cancel	618	64	8.4	153.6224	*****
	list	212	41	2.9	148.6182	*****
	restrict	188	39	2.6	147.0925	*****
	connect	172	37	2.4	142.3869	*****
	supply	366	49	5	141.4103	*****
	cause	233	41	3.2	140.6321	*****
	refuse	181	37	2.5	138.3981	*****
	submit	312	45	4.3	136.4086	*****
	install	140	33	1.9	133.4536	*****
	view	117	31	1.6	133.3064	*****
	identify	165	33	2.3	121.912	*****
	update	209	36	2.9	121.9004	*****

CONSTRUCTION	COLLEXEME	CORPUS. FREQ.	OBS.	EXP.	COLL.STR. (LOGL)	SIGNIF
	inform	186	34	2.5	119.2593	*****
	suspend	254	37	3.5	112.8603	*****
	choose	319	38	4.4	101.1104	*****
	fail	213	32	2.9	99.55127	*****
	tell	562	45	7.7	87.17637	*****
	exclude	200	27	2.7	78.31423	*****
	govern	260	29	3.6	73.5456	*****
	establish	349	31	4.8	65.70088	*****
	use	1396	61	19.1	59.40919	*****
	end	902	42	12.3	44.70849	*****
	act	582	30	8	36.46254	*****
	amend	313	18	4.3	24.92328	*****
	enter	377	17	5.2	17.28779	****
	do	2403	59	32.8	17.16697	****
	cover	307	15	4.2	17.01284	****
	process	324	15	4.4	15.82365	****
	post	299	14	4.1	14.99884	***
	confirm	196	11	2.7	14.80436	***
	allow	490	17	6.7	11.29948	***
	send	738	22	10.1	10.69717	**
	include	2379	51	32.5	9.12327	**
	refund	378	13	5.2	8.49806	**
	share	261	10	3.6	7.91758	**
	take	1146	27	15.7	6.85899	**
	require	704	18	9.6	5.90119	*
	display	102	5	1.4	5.6929	*
	request	311	9	4.3	4.08111	*
	display	171	6	2.3	4.07035	*
	determine	173	6	2.4	3.98334	*
	purchase	219	7	3	3.95745	*

Covarying Collexeme Analysis

SLOT1	SLOT2	FS1	FS2	OBS	EXP	COLL.STR. (LOGL)	SIGNIF
can	find	2902	224	193	34.1	566.8055	*****
would	like	366	58	58	1.1	467.9355	*****
must	ensure	1684	117	69	10.3	187.4634	*****
will	refund	6900	109	101	39.5	155.9121	*****
shall	be	1134	4484	430	267.1	124.8021	*****
can	guarantee	2902	38	36	5.8	120.8059	*****
will	try	6900	78	71	28.3	103.7473	*****
must	comply	1684	67	38	5.9	98.70649	*****
may	require	5113	132	89	35.4	95.07449	*****
can	ask	2902	204	89	31.1	94.7817	*****
may	include	5113	179	110	48.1	94.76683	*****
should	contact	521	224	41	6.1	94.48616	*****
will	tell	6900	226	152	81.9	90.48314	*****
must	pay	1684	292	81	25.8	89.18392	*****
could	damage	215	22	13	0.2	87.79367	*****
shall	deem	1134	79	33	4.7	85.21312	*****
may	assign	5113	53	46	14.2	84.23325	*****
must	follow	1684	38	26	3.4	81.27988	*****
will	continue	6900	142	104	51.5	80.99488	*****
will	notify	6900	171	119	62	79.03737	*****
shall	govern	1134	82	32	4.9	77.67637	*****
may	charge	5113	254	130	68.2	68.44959	*****
can	change	2902	227	83	34.6	62.76964	*****
may	arise	5113	38	33	10.2	60.44768	*****
may	suspend	5113	77	53	20.7	59.10666	*****
would	compromise	366	12	9	0.2	57.96643	*****
will	apply	6900	416	225	150.8	56.03956	*****
will	send	6900	210	129	76.1	55.37221	*****
would	have	366	732	48	14.1	55.01709	*****
may	offer	5113	66	46	17.7	52.71069	*****
may	change	5113	227	111	61	50.52549	*****
will	give	6900	316	176	114.5	50.21346	*****
can	use	2902	592	156	90.2	50.12886	*****
will	be	6900	4484	1824	1625	49.46702	*****
will	let	6900	75	57	27.2	49.45472	*****
must	inform	1684	79	29	7	46.38278	*****
shall	limit	1134	72	23	4.3	45.89666	*****
may	vary	5113	66	44	17.7	45.62345	*****
should	read	521	40	13	1.1	44.89637	*****
will	treat	6900	50	41	18.1	44.31236	*****
can	cancel	2902	162	59	24.7	44.04661	*****
shall	remain	1134	83	24	4.9	43.15618	*****

SLOT1	SLOT2	FS1	FS2	OBS	EXP	COLL.STR. (LOGL)	SIGNIF
should	check	521	52	14	1.4	42.60423	*****
would	cause	366	39	11	0.7	41.91893	*****
may	result	5113	60	40	16.1	41.46047	*****
can	prove	2902	18	15	2.7	41.26745	*****
shall	survive	1134	29	14	1.7	40.79821	*****
can	choose	2902	69	33	10.5	40.74453	*****
could	have	215	732	31	8.3	39.83457	*****
would	prefer	366	5	5	0.1	39.58395	*****
can	do	2902	192	64	29.3	39.19543	*****
will	start	6900	44	36	15.9	38.64947	*****
can	contact	2902	224	71	34.1	38.48193	*****
may	refuse	5113	71	44	19.1	38.43534	*****
can	purchase	2902	29	19	4.4	37.5146	*****
should	know	521	15	8	0.4	37.35178	*****
might	happen	199	22	7	0.2	36.88316	*****
might	need	199	261	17	2.7	35.54358	*****
shall	prevail	1134	10	8	0.6	35.42177	*****
must	keep	1684	81	26	7.2	34.86547	*****
must	sign	1684	16	11	1.4	34.46942	*****
may	differ	5113	13	13	3.5	34.20785	*****
can	access	2902	66	30	10.1	33.98812	*****
can	learn	2902	9	9	1.4	33.88424	*****
shall	preclude	1134	6	6	0.4	33.8794	*****
can	get	2902	53	26	8.1	33.42554	*****
may	request	5113	66	40	17.7	33.07173	*****
can	withdraw	2902	29	18	4.4	32.93825	*****
may	terminate	5113	113	59	30.3	32.70531	*****
shall	conduct	1134	28	12	1.7	31.52166	*****
would	expect	366	12	6	0.2	31.11062	*****
can	transfer	2902	132	46	20.1	31.10582	*****
should	note	521	6	5	0.2	30.68053	*****
can	make	2902	367	96	55.9	29.64284	*****
will	explain	6900	19	18	6.9	29.63363	*****
can	close	2902	75	31	11.4	29.63221	*****
will	need	6900	261	137	94.6	28.96456	*****
must	repay	1684	16	10	1.4	28.49675	*****
shall	cooperate	1134	13	8	0.8	28.47256	*****
may	restrict	5113	36	25	9.7	28.3646	*****
must	destroy	1684	8	7	0.7	28.1379	*****
should	exercise	521	25	8	0.7	27.28285	*****
must	meet	1684	24	12	2.1	27.22406	*****
can	recall	2902	7	7	1.1	26.35031	*****
may	monitor	5113	25	19	6.7	26.19971	*****
can	obtain	2902	31	17	4.7	25.96613	*****

SLOT1	SLOT2	FS1	FS2	OBS	EXP	COLL.STR. (LOGL)	SIGNIF
should	direct	521	8	5	0.2	25.61219	*****
may	enable	5113	30	21	8.1	24.2416	*****
may	amend	5113	24	18	6.4	24.12743	*****
might	have	199	732	24	7.7	24.00311	*****
must	return	1684	51	17	4.5	23.93731	*****
can	refer	2902	46	21	7	23.9348	*****
will	remain	6900	83	52	30.1	23.88316	*****
would	complicate	366	3	3	0.1	23.73417	*****
should	review	521	43	9	1.2	22.66149	*****
can	promise	2902	6	6	0.9	22.58423	*****
would	prevent	366	33	7	0.6	22.3406	*****
will	confirm	6900	31	24	11.2	21.94073	*****
may	have	5113	732	253	196.6	21.89427	*****
can	see	2902	38	18	5.8	21.82263	*****
will	receive	6900	132	74	47.8	21.53734	*****
need	help	3	35	2	0	21.4965	*****
must	provide	1684	369	60	32.6	21.13179	*****
will	assume	6900	28	22	10.1	20.99686	*****
must	file	1684	18	9	1.6	20.40606	*****
will	cost	6900	10	10	3.6	20.30966	*****
will	expire	6900	10	10	3.6	20.30966	*****
may	revise	5113	11	10	3	20.23183	*****
may	suffer	5113	20	15	5.4	20.10122	*****
may	appear	5113	22	16	5.9	20.0701	*****
ought	have	3	732	3	0.1	19.56294	*****
could	lead	215	14	4	0.2	19.41475	****
would	jeopardise	366	4	3	0.1	19.27399	****
must	adhere	1684	6	5	0.5	19.0452	****
may	expose	5113	13	11	3.5	19.02817	****
shall	apply	1134	416	48	24.8	18.9655	****
must	register	1684	33	12	2.9	18.89035	****
shall	have	1134	732	73	43.6	18.57709	****
may	delegate	5113	7	7	1.9	18.41359	****
may	encounter	5113	7	7	1.9	18.41359	****
shall	entitle	1134	71	15	4.2	18.39109	****
will	endeavor	6900	9	9	3.3	18.27786	****
might	exacerbate	199	2	2	0	18.26393	****
must	satisfy	1684	9	6	0.8	18.22048	****
must	present	1684	16	8	1.4	18.13518	****
may	impose	5113	21	15	5.6	18.09136	****
shall	serve	1134	12	6	0.7	17.97708	****
may	need	5113	261	101	70.1	17.53838	****
can	email	2902	9	7	1.4	17.47563	****
may	contain	5113	37	22	9.9	17.30909	****

SLOT1	SLOT2	FS1	FS2	OBS	EXP	COLL.STR. (LOGL)	SIGNIF
can	recover	2902	14	9	2.1	17.28422	****
can	borrow	2902	7	6	1.1	17.17271	****
can	foresee	2902	7	6	1.1	17.17271	****
would	consider	366	48	7	0.9	17.13897	****
may	modify	5113	28	18	7.5	17.11652	****
might	suffer	199	20	4	0.2	16.88158	****
might	arise	199	38	5	0.4	16.81933	****
should	report	521	25	6	0.7	16.7389	****
could	claim	215	8	3	0.1	16.47053	****
will	handle	6900	8	8	2.9	16.24625	****
may	upgrade	5113	16	12	4.3	16.077	****
would	break	366	13	4	0.2	15.95374	****
must	establish	1684	14	7	1.2	15.86518	****
must	specify	1684	14	7	1.2	15.86518	****
could	cause	215	39	5	0.4	15.8366	****
would	risk	366	2	2	0	15.81739	****
shall	deal	1134	14	6	0.8	15.73564	****
would	create	366	38	6	0.7	15.59174	****
can	control	2902	18	10	2.7	15.55951	****
shall	determine	1134	26	8	1.5	15.28585	****
will	depend	6900	65	39	23.6	15.1364	***
may	increase	5113	39	22	10.5	15.09441	***
will	post	6900	35	24	12.7	15.07687	***
can	afford	2902	4	4	0.6	15.05381	***
must	submit	1684	57	15	5	14.90264	***
must	maintain	1684	11	6	1	14.88876	***
could	disable	215	10	3	0.1	14.88306	***
must	design	1684	5	4	0.4	14.5924	***
shall	exclude	1134	21	7	1.3	14.50793	***
will	bind	6900	11	10	4	14.50672	***
should	consult	521	2	2	0.1	14.40167	***
should	fly	521	2	2	0.1	14.40167	***
should	pack	521	2	2	0.1	14.40167	***
should	speak	521	2	2	0.1	14.40167	***
can	produce	2902	8	6	1.2	14.2471	***
will	process	6900	64	38	23.2	14.134	***
may	share	5113	58	29	15.6	14.03645	***
will	calculate	6900	21	16	7.6	13.94432	***
can	visit	2902	6	5	0.9	13.74221	***
can	end	2902	138	38	21	13.72649	***
may	decide	5113	38	21	10.2	13.62585	***
could	affect	215	106	7	1.2	13.6079	***
might	break	199	13	3	0.1	13.57267	***
should	tell	521	226	17	6.2	13.52632	***

SLOT1	SLOT2	FS1	FS2	OBS	EXP	COLL.STR. (LOGL)	SIGNIF
will	govern	6900	82	46	29.7	13.39147	***
must	review	1684	43	12	3.8	13.07601	***
can	view	2902	17	9	2.6	13.01498	***
may	assert	5113	8	7	2.1	13.00998	***
will	deduct	6900	27	19	9.8	12.97525	***
will	attempt	6900	18	14	6.5	12.96417	***
will	begin	6900	18	14	6.5	12.96417	***
will	cease	6900	18	14	6.5	12.96417	***
will	communicate	6900	18	14	6.5	12.96417	***
must	adopt	1684	13	6	1.1	12.47142	***
might	involve	199	16	3	0.2	12.23762	***
should	ensure	521	117	11	3.2	12.22537	***
could	submit	215	57	5	0.6	12.22368	***
can	enforce	2902	21	10	3.2	12.21516	***
may	access	5113	66	31	17.7	12.1992	***
will	investigate	6900	15	12	5.4	12.05844	***
shall	obligate	1134	8	4	0.5	11.97888	***
would	constitute	366	36	5	0.7	11.76027	***
could	result	215	60	5	0.7	11.75423	***
can	elect	2902	12	7	1.8	11.69988	***
might	interest	199	5	2	0.1	11.59623	***
will	credit	6900	17	13	6.2	11.45125	***
could	last	215	5	2	0.1	11.29049	***
can	chat	2902	3	3	0.5	11.28948	***
can	complain	2902	3	3	0.5	11.28948	***
shall	erase	1134	2	2	0.1	11.28648	***
would	pay	366	292	15	5.6	11.2774	***
may	update	5113	33	18	8.9	11.26035	***
can	inspect	2902	7	5	1.1	11.1034	***
will	bill	6900	19	14	6.9	11.03176	***
will	work	6900	19	14	6.9	11.03176	***
will	deem	6900	79	43	28.6	10.84535	***
shall	dispose	1134	9	4	0.5	10.82635	**
shall	indemnify	1134	9	4	0.5	10.82635	**
may	reduce	5113	31	17	8.3	10.79346	**
may	owe	5113	7	6	1.9	10.66541	**
may	undermine	5113	7	6	1.9	10.66541	**
can	accept	2902	42	15	6.4	10.64658	**
must	report	1684	25	8	2.2	10.63286	**
must	connect	1684	15	6	1.3	10.59521	**
may	consolidate	5113	4	4	1.1	10.52034	**
may	exempt	5113	4	4	1.1	10.52034	**
may	import	5113	4	4	1.1	10.52034	**
may	participate	5113	4	4	1.1	10.52034	**

SLOT1	SLOT2	FS1	FS2	OBS	EXP	COLL.STR. (LOGL)	SIGNIF
must	tell	1684	226	35	20	10.4605	**
shall	construe	1134	5	3	0.3	10.44711	**
can	book	2902	10	6	1.5	10.44454	**
must	respect	1684	7	4	0.6	10.40532	**
can	manage	2902	13	7	2	10.38579	**
can	switch	2902	5	4	0.8	10.38011	**
may	choose	5113	69	31	18.5	10.37983	**
shall	bear	1134	15	5	0.9	10.35645	**
must	obey	1684	4	3	0.4	10.24331	**
must	proceed	1684	4	3	0.4	10.24331	**
will	acknowledge	6900	5	5	1.8	10.15252	**
will	compensate	6900	5	5	1.8	10.15252	**
will	re-credit	6900	5	5	1.8	10.15252	**
will	redirect	6900	5	5	1.8	10.15252	**
may	delay	5113	9	7	2.4	10.12834	**
may	wish	5113	9	7	2.4	10.12834	**
can	call	2902	20	9	3	9.98746	**
must	give	1684	316	45	27.9	9.9867	**
may	add	5113	53	25	14.2	9.97709	**
must	notify	1684	171	28	15.1	9.94272	**
would	mean	366	27	4	0.5	9.88797	**
must	operate	1684	21	7	1.9	9.83176	**
must	activate	1684	2	2	0.2	9.70365	**
must	seat	1684	2	2	0.2	9.70365	**
must	subscribe	1684	2	2	0.2	9.70365	**
must	tamper	1684	2	2	0.2	9.70365	**
must	travel	1684	2	2	0.2	9.70365	**
must	trust	1684	2	2	0.2	9.70365	**
could	interfere	215	7	2	0.1	9.68984	**
will	correct	6900	11	9	4	9.64517	**
will	respond	6900	11	9	4	9.64517	**
must	exceed	1684	27	8	2.4	9.52999	**
can	request	2902	66	20	10.1	9.52951	**
can	read	2902	40	14	6.1	9.5007	**
might	want	199	8	2	0.1	9.39141	**
will	pass	6900	22	15	8	9.24075	**
can	download	2902	8	5	1.2	9.22441	**
may	submit	5113	57	26	15.3	9.20654	**
would	encourage	366	5	2	0.1	9.20312	**
might	lack	199	1	1	0	9.12697	**
could	misuse	215	8	2	0.1	9.09076	**
could	relate	215	8	2	0.1	9.09076	**
can	discuss	2902	11	6	1.7	9.07676	**
could	harvest	215	1	1	0	8.97192	**

SLOT1	SLOT2	FS1	FS2	OBS	EXP	COLL.STR. (LOGL)	SIGNIF
could	mislead	215	1	1	0	8.97192	**
could	overburden	215	1	1	0	8.97192	**
might	delay	199	9	2	0.1	8.87487	**
may	use	5113	592	191	159	8.76927	**
may	ask	5113	204	74	54.8	8.75666	**
will	convert	6900	47	27	17	8.72581	**
shall	inform	1134	79	12	4.7	8.67143	**
must	agree	1684	55	12	4.9	8.50043	**
must	disconnect	1684	13	5	1.1	8.42179	**
may	appeal	5113	6	5	1.6	8.36932	**
may	subcontract	5113	6	5	1.6	8.36932	**
may	harm	5113	12	8	3.2	8.26756	**
must	log	1684	5	3	0.4	8.19678	**
should	make	521	367	20	10	8.12266	**
will	count	6900	4	4	1.4	8.12164	**
will	honour	6900	4	4	1.4	8.12164	**
will	migrate	6900	4	4	1.4	8.12164	**
will	scan	6900	4	4	1.4	8.12164	**
will	undertake	6900	4	4	1.4	8.12164	**
would	violate	366	18	3	0.3	8.09176	**
can	avoid	2902	6	4	0.9	8.0763	**
can	phone	2902	6	4	0.9	8.0763	**
can	sit	2902	6	4	0.9	8.0763	**
may	reject	5113	10	7	2.7	8.07055	**
may	incur	5113	29	15	7.8	8.04451	**
may	record	5113	29	15	7.8	8.04451	**
will	administer	6900	10	8	3.6	8.03678	**
may	allow	5113	65	28	17.5	7.93793	**
would	hate	366	1	1	0	7.90601	**
would	outweigh	366	1	1	0	7.90601	**
would	shield	366	1	1	0	7.90601	**
would	struggle	366	1	1	0	7.90601	**
may	accompany	5113	3	3	0.8	7.88982	**
may	edit	5113	3	3	0.8	7.88982	**
can	help	2902	35	12	5.3	7.76844	**
would	affect	366	106	7	2	7.66312	**
shall	constitute	1134	36	7	2.1	7.60735	**
shall	affect	1134	106	14	6.3	7.59033	**
shall	consist	1134	3	2	0.2	7.59	**
shall	deprive	1134	3	2	0.2	7.59	**
shall	procure	1134	3	2	0.2	7.59	**
shall	relieve	1134	3	2	0.2	7.59	**
can	award	2902	2	2	0.3	7.52574	**
can	escape	2902	2	2	0.3	7.52574	**

SLOT1	SLOT2	FS1	FS2	OBS	EXP	COLL.STR. (LOGL)	SIGNIF
can	flex	2902	2	2	0.3	7.52574	**
can	foretell	2902	2	2	0.3	7.52574	**
can	litigate	2902	2	2	0.3	7.52574	**
can	resell	2902	2	2	0.3	7.52574	**
can	sort	2902	2	2	0.3	7.52574	**
can	telephone	2902	2	2	0.3	7.52574	**
can	trace	2902	2	2	0.3	7.52574	**
can	unlock	2902	2	2	0.3	7.52574	**
might	affect	199	106	5	1.1	7.50978	**
will	renew	6900	12	9	4.3	7.47947	**
will	charge	6900	254	113	92	7.38018	**
could	expect	215	12	2	0.1	7.3647	**
could	harm	215	12	2	0.1	7.3647	**
could	subject	215	12	2	0.1	7.3647	**
might	expose	199	13	2	0.1	7.33043	**
must	call	1684	20	6	1.8	7.27446	**
should	believe	521	1	1	0	7.19896	**
should	disagree	521	1	1	0	7.19896	**
should	integrate	521	1	1	0	7.19896	**
should	intensify	521	1	1	0	7.19896	**
should	preserve	521	1	1	0	7.19896	**
should	reuse	521	1	1	0	7.19896	**
should	ring	521	1	1	0	7.19896	**
should	talk	521	1	1	0	7.19896	**
should	trigger	521	1	1	0	7.19896	**
can	earn	2902	20	8	3	7.15655	**
must	reimburse	1684	10	4	0.9	7.0603	**
would	run	366	8	2	0.2	7.05173	**
must	set	1684	75	14	6.6	7.04056	**
could	impact	215	13	2	0.1	7.03829	**
might	cause	199	39	3	0.4	7.00401	**
will	commence	6900	16	11	5.8	6.96221	**
will	reach	6900	16	11	5.8	6.96221	**
shall	waive	1134	8	3	0.5	6.95977	**
must	protect	1684	6	3	0.5	6.79405	**
should	refer	521	46	5	1.3	6.66058	**
may	violate	5113	18	10	4.8	6.57504	*
can	show	2902	42	13	6.4	6.54622	*
may	redeem	5113	11	7	3	6.49206	*
may	sell	5113	11	7	3	6.49206	*
can	buy	2902	7	4	1.1	6.48411	*
can	revoke	2902	7	4	1.1	6.48411	*
could	interpret	215	15	2	0.2	6.46549	*
must	cover	1684	11	4	1	6.28479	*

SLOT1	SLOT2	FS1	FS2	OBS	EXP	COLL.STR. (LOGL)	SIGNIF
can	refuse	2902	71	19	10.8	6.22618	*
could	disrupt	215	2	1	0	6.22194	*
shall	sever	1134	9	3	0.5	6.21006	*
may	adjust	5113	9	6	2.4	6.19952	*
may	develop	5113	5	4	1.3	6.14134	*
may	opt	5113	5	4	1.3	6.14134	*
may	lose	5113	21	11	5.6	6.12203	*
can	go	2902	14	6	2.1	6.1042	*
will	abide	6900	3	3	1.1	6.09096	*
will	alert	6900	3	3	1.1	6.09096	*
will	defend	6900	3	3	1.1	6.09096	*
will	drop	6900	3	3	1.1	6.09096	*
will	guide	6900	3	3	1.1	6.09096	*
must	achieve	1684	3	2	0.3	6.06956	*
must	adapt	1684	3	2	0.3	6.06956	*
must	attend	1684	3	2	0.3	6.06956	*
must	declare	1684	3	2	0.3	6.06956	*
must	fulfil	1684	3	2	0.3	6.06956	*
must	suggest	1684	3	2	0.3	6.06956	*
will	come	6900	11	8	4	6.05325	*
may	invite	5113	7	5	1.9	6.02524	*
may	search	5113	7	5	1.9	6.02524	*
will	show	6900	42	23	15.2	5.96328	*
may	bring	5113	43	19	11.5	5.95787	*
may	enter	5113	32	15	8.6	5.84956	*
must	reflect	1684	7	3	0.6	5.73603	*
will	deliver	6900	21	13	7.6	5.68753	*
will	aim	6900	6	5	2.2	5.64536	*
will	state	6900	6	5	2.2	5.64536	*
shall	condemn	1134	1	1	0.1	5.64241	*
shall	confer	1134	1	1	0.1	5.64241	*
shall	excuse	1134	1	1	0.1	5.64241	*
shall	fall	1134	1	1	0.1	5.64241	*
shall	inure	1134	1	1	0.1	5.64241	*
shall	measure	1134	1	1	0.1	5.64241	*
shall	prejudice	1134	1	1	0.1	5.64241	*
shall	recredit	1134	1	1	0.1	5.64241	*
shall	twitch	1134	1	1	0.1	5.64241	*
will	collect	6900	47	25	17	5.60288	*
may	cancel	5113	162	57	43.5	5.45926	*
would	subject	366	12	2	0.2	5.39012	*
might	qualify	199	3	1	0	5.3497	*
will	write	6900	32	18	11.6	5.29098	*
can	claim	2902	8	4	1.2	5.28485	*

SLOT1	SLOT2	FS1	FS2	OBS	EXP	COLL.STR. (LOGL)	SIGNIF
can	instruct	2902	8	4	1.2	5.28485	*
can	turn	2902	8	4	1.2	5.28485	*
could	include	215	179	6	2	5.26471	*
may	advertise	5113	2	2	0.5	5.2596	*
may	concern	5113	2	2	0.5	5.2596	*
may	contract	5113	2	2	0.5	5.2596	*
may	divert	5113	2	2	0.5	5.2596	*
may	exonerate	5113	2	2	0.5	5.2596	*
may	experience	5113	2	2	0.5	5.2596	*
may	filter	5113	2	2	0.5	5.2596	*
may	interrupt	5113	2	2	0.5	5.2596	*
may	persist	5113	2	2	0.5	5.2596	*
may	pool	5113	2	2	0.5	5.2596	*
may	recoup	5113	2	2	0.5	5.2596	*
may	stipulate	5113	2	2	0.5	5.2596	*
should	discontinue	521	9	2	0.2	5.25398	*
should	wish	521	9	2	0.2	5.25398	*
can	combine	2902	5	3	0.8	5.22023	*
would	accrue	366	2	1	0	5.17214	*
would	license	366	2	1	0	5.17214	*
shall	exceed	1134	27	5	1.6	5.04565	*
will	initiate	6900	8	6	2.9	4.98522	*
will	take	6900	312	132	113.1	4.95154	*
could	use	215	592	13	6.7	4.92638	*
shall	execute	1134	5	2	0.3	4.9242	*
shall	strike	1134	5	2	0.3	4.9242	*
must	misuse	1684	8	3	0.7	4.89667	*
may	transfer	5113	132	47	35.4	4.88902	*
must	attack	1684	1	1	0.1	4.85128	*
must	bid	1684	1	1	0.1	4.85128	*
must	compile	1684	1	1	0.1	4.85128	*
must	conform	1684	1	1	0.1	4.85128	*
must	equip	1684	1	1	0.1	4.85128	*
must	focus	1684	1	1	0.1	4.85128	*
must	possess	1684	1	1	0.1	4.85128	*
must	pre-approve	1684	1	1	0.1	4.85128	*
must	recruit	1684	1	1	0.1	4.85128	*
must	reside	1684	1	1	0.1	4.85128	*
must	reverse-engineer	1684	1	1	0.1	4.85128	*
must	stamp	1684	1	1	0.1	4.85128	*
must	study	1684	1	1	0.1	4.85128	*
must	supervise	1684	1	1	0.1	4.85128	*
must	top	1684	1	1	0.1	4.85128	*
could	block	215	23	2	0.3	4.8357	*

SLOT1	SLOT2	FS1	FS2	OBS	EXP	COLL.STR. (LOGL)	SIGNIF
will	provide	6900	369	154	133.7	4.8251	*
may	engage	5113	10	6	2.7	4.82148	*
would	rely	366	14	2	0.3	4.7976	*
can	open	2902	20	7	3	4.74398	*
might	put	199	26	2	0.3	4.66168	*
will	earn	6900	20	12	7.2	4.64575	*
must	conflict	1684	4	2	0.4	4.52848	*
might	mean	199	27	2	0.3	4.52554	*
should	answer	521	2	1	0.1	4.48176	*
should	detect	521	2	1	0.1	4.48176	*
should	evaluate	521	2	1	0.1	4.48176	*
should	feature	521	2	1	0.1	4.48176	*
should	lower	521	2	1	0.1	4.48176	*
will	affect	6900	106	49	38.4	4.45892	*
may	exchange	5113	8	5	2.1	4.441	*
may	introduce	5113	8	5	2.1	4.441	*
may	want	5113	8	5	2.1	4.441	*
would	result	366	60	4	1.2	4.41832	*
must	publish	1684	27	6	2.4	4.39818	*
could	mean	215	27	2	0.3	4.25721	*
can	exclude	2902	21	7	3.2	4.2385	*
may	impact	5113	13	7	3.5	4.2176	*
may	replace	5113	13	7	3.5	4.2176	*
may	entitle	5113	71	27	19.1	4.21575	*
can	grant	2902	17	6	2.6	4.14203	*
shall	debit	1134	6	2	0.4	4.13877	*
shall	indicate	1134	6	2	0.4	4.13877	*
may	commit	5113	6	4	1.6	4.13225	*
may	link	5113	6	4	1.6	4.13225	*
may	reproduce	5113	6	4	1.6	4.13225	*
will	disassociate	6900	2	2	0.7	4.06045	*
will	eliminate	6900	2	2	0.7	4.06045	*
will	lift	6900	2	2	0.7	4.06045	*
will	misappropriate	6900	2	2	0.7	4.06045	*
will	oversee	6900	2	2	0.7	4.06045	*
will	spread	6900	2	2	0.7	4.06045	*
will	strive	6900	2	2	0.7	4.06045	*
will	uphold	6900	2	2	0.7	4.06045	*
could	construe	215	5	1	0.1	4.05834	*
could	contribute	215	5	1	0.1	4.05834	*
can	demonstrate	2902	3	2	0.5	4.03718	*
can	enjoy	2902	3	2	0.5	4.03718	*
can	influence	2902	3	2	0.5	4.03718	*
can	spend	2902	3	2	0.5	4.03718	*

SLOT1	SLOT2	FS1	FS2	OBS	EXP	COLL.STR. (LOGL)	SIGNIF
will	remit	6900	5	4	1.8	4.0173	*
may	hand	5113	4	3	1.1	4.01627	*
may	launch	5113	4	3	1.1	4.01627	*
may	regard	5113	4	3	1.1	4.01627	*
can	join	2902	6	3	0.9	3.96305	*
can	offset	2902	6	3	0.9	3.96305	*
can	order	2902	6	3	0.9	3.96305	*
may	close	5113	75	28	20.1	3.932	*
will	pay	6900	292	122	105.8	3.86619	*
should	settle	521	13	2	0.4	3.84752	*

Commercialising disadvantage: the neoliberal discourses of commercial bail bond websites

Leigh Harrington, Stephen Parker, Lauren Devine & Nadia Makouar

University of Manchester and Aston Institute for Forensic Linguistics
Aston University, UK

https://doi.org/10.21747/21833745/lanlaw/9_1a8

Abstract. *The commercial bail bond industry is one of the most profitable aspects of America's highly marketized criminal justice system that is increasingly shaped by neoliberal structures and ideologies. Drawing on a specialised corpus of "Home" and "About Us" pages from bail bond websites, this paper is the first empirical linguistic examination of commercial bail bonds discourse grounded in its legal context. Using corpus-assisted critical discourse analysis, we examine how bail bond companies 1) discursively present and promote their services, 2) represent the legal system and its processes, and 3) construe arrest and detention to prospective service users. The findings show that bail bond companies position their services as an unobjectionably common (Brookes and Harvey 2017a) part of legal and financial self-management by normalising, legitimising, and idealising their use whilst seeking to minimise the power-imbalance between themselves and their often financially and socially disempowered 'clients'. By grounding our linguistic analysis in a legal context, we demonstrate that these discourses simultaneously serve whilst oppress those they purport to help, offering an example of a local form of structural violence that subtly perpetuates neoliberal agendas and a two-tier justice system.*

Keywords: *Bail bonds, Legal studies, Critical discourse analysis, Corpus linguistics, Neoliberalism, Service users.*

Resumo. *O setor das fianças é um dos aspetos mais lucrativos do sistema de justiça criminal altamente mercantilizado dos Estados Unidos, que é cada vez mais moldado por estruturas e ideologias neoliberais. Este artigo baseia-se num corpus especializado de páginas "Início" e "Sobre Nós" de websites de fianças comerciais para proporcionar uma primeira análise do discurso de base linguística empírica aplicada a fianças comerciais, assente no seu contexto legal. Recorrendo a uma análise crítica do discurso baseada em corpus, analisamos de que forma as empresas de fianças 1) apresentam e promovem discursivamente os seus serviços, 2) representam o sistema jurídico e os seus processos e 3) formulam a prisão e a detenção*

para potenciais utilizadores do serviço. Os resultados mostram que as empresas de fianças posicionam os seus serviços como uma parte da autogestão jurídica e financeira inquestionavelmente comum (Brookes and Harvey 2017a), normalizando, legitimando e idealizando a sua utilização, ao mesmo tempo que procuram minimizar o desequilíbrio de poder entre eles próprios e os seus "clientes", muitas vezes desprovidos de poder financeiro e social. Sustentando a nossa análise linguística num contexto jurídico, demonstramos que estes discursos servem, ao mesmo tempo que oprimem, aqueles que supostamente têm como objetivo ajudar, proporcionando um exemplo de uma forma local de violência estrutural que perpetua subtilmente os planos neoliberais e um sistema de justiça de duplo nível.

Palavras-chave: *Fianças, Estudos jurídicos, Análise crítica do discurso, Linguística de corpus, Neoliberalismo, Utilizadores de serviço.*

Introduction

The commercial bail bond (henceforth BB) industry in the United States is one of the most discussed and contested aspects of the American criminal justice system. It generates considerable public debate, providing the onus for numerous broad-reaching movements to reform pretrial and bail systems across America (Stevenson 2018). The industry also attracts attention from academic research in a variety of disciplines, including law, criminology, politics, economics, sociology, and psychology. Multidisciplinary research on bail has variously examined issues of debt, profitability, and social inequality (Page 2017; Justice Policy Institute 2012; Western and Pettit 2010; Harris *et al.* 2010; Pettit and Western 2004), prison funding, incarceration, rehabilitation, and recidivism (Harper *et al.* 2021; Ortiz and Jackey 2019; Alper *et al.* 2018; Shannon *et al.* 2017), pretrial decision-making and the calculation of risk (Stevenson 2018), disenfranchisement (Meredith and Morse 2017; Justice Policy Institute 2012), and the power and regulation of BB agents and bounty hunters (Johnson and Stevens 2013; Baker *et al.* 2008b). However, one significant area of the commercial bail industry that has until now been overlooked is the public-facing discourse of BB companies themselves, specifically the language used by these companies in addressing prospective users of their services.

In analysing the language with which BB companies directly interact with potential users of their services, this paper is the first empirical linguistic investigation of commercial BB discourse grounded in its legal context. We examine how BB companies discursively present and promote their services, how they represent the legal system and its processes, and how they construe arrest and detention and, in doing so, how they perpetuate inequalities. Taking commercial BB discourse as a microcosm of the neoliberal ideologies and agendas that pervade the American and other criminal justice systems globally, we discuss the impact of neoliberal frameworks on the multi-jurisdiction question of the tension between neoliberalism and fairness in relation to the administration of justice.

The following section begins by describing the increasingly neoliberal landscape in which contemporary justice systems operate and its legal context. We then go on to explain the US commercial BB industry which is the focus of this paper.

Background

Neoliberalism in criminal justice systems

The administration and delivery of justice in the United States and in jurisdictions such as Canada, the UK, Australia, New Zealand, and many European countries is increasingly shaped by and bound up in structures and ideologies that bear the characteristics of neoliberalism. This political, economic, and ideological reconfiguration, which Walmsley (2019: 249) refers to as the "neoliberal project", has taken place over the past 40 years in the US. Here, we understand neoliberalism to be a "variegated set of ideas, policies, and practices" (Walmsley 2019: 251) which signify the ideological shift in purpose of the state away from one that is responsible for the full employment and protection of its citizens "against the exigencies of the market" and towards ensuring "individual responsibility and protection of the market itself" (Wrenn 2015: 453). As such, its three core organising tenets are privatisation, deregulation, and the retrenchment of the welfare state (Wrenn 2015).

The impact of neoliberal structures on the already highly inequitable and marketized American criminal justice system is clear to see. The US incarcerates more people than any other country in the world; Harper *et al.* (2021), referencing Shannon *et al.* (2017), estimate that 3% of the total US adult population will be incarcerated over the course of their life, meaning the criminal justice system represents an incredibly prolific and lucrative venture for those who run it. Walmsley (2019: 256), for instance, points to the interrelation of carceral expansionism and neoliberal economic restructuring, where coercive shifts in criminal justice policies have opened "new channels of profitability for financial institutions" and enabled the emergence of consumer credit industries.

The consequences of these shifts are, for instance, directly reflected in the use of a bail system, and a proliferation since the early 1980s of the use of legal financial obligations (LFOs), a series of fines and fees applied to offenders themselves to cover the financial burden of America's growing criminal justice system. The shift of costs onto defendants and offenders is central to poverty-trap theories of criminal justice, where they act as a form of regressive taxation (Stevenson 2018). These theories assert that such systems compound and perpetuate the existing debt burden of low-income families who are already disproportionately more likely to be involved in the criminal justice system (Harper *et al.* 2021: 250). Ultimately, they create a two-tier justice system (Page 2017) which can affect pre-trial decision-making (Stevenson 2018; Justice Policy Institute 2012) and political enfranchisement, as well as contribute to an increasing distrust of the criminal justice system (Harper *et al.* 2021; Western and Pettit 2010; Pettit and Western 2004).

The oppressive design of the monetised components of American criminal justice are reinforced by neoliberally motivated moves at the level of policy and legislation. For instance, in the bail industry, the American Bail Coalition has repeatedly fought reform in pretrial policy to restrict or de-fund alternatives to bail, such as pretrial services, whilst promoting (and even drafting) industry-friendly legislation (Justice Policy Institute, 2012, p. 3). Likewise, the bail industry is also backed by huge insurance companies and trade associations "with the money and political power needed to maintain their place in the criminal justice system" (Justice Policy Institute, 2012, p. 3). Finally, the general underfunding of rehabilitation programmes means America has high recidivism rates; 7/10 inmates will re-enter the criminal justice system at some point in their

lifetime (Alper *et al.* 2018). This points to a "revolving door" of reoffending which is desirable to those who profit from it (Ortiz and Jackey 2019: 484).

It is unsurprising that numerous critics view the monetised design of the US criminal justice system as a form structural violence, designed to maintain, perpetuate, and profit from legal, political, and economic inequality (Ortiz and Jackey 2019; Hallett 2012). Influenced and shaped by neoliberal agendas and ideologies, this system is deeply entwined in American society, culture, and politics. One linchpin in this landscape is the bail bond industry, which is the focus of this paper.

The bail bond industry

The US is one of only two nations to have a commercial BB industry, the other being the Philippines, though BB practices are banned in the District of Columbia and the states of Illinois, Kentucky, Massachusetts, Oregon, and Wisconsin. Bail has been used for those awaiting trial in the American criminal justice system informally for a few hundred years (Baker *et al.* 2008b) with the first BB company emerging around 1898 (Justice Policy Institute, 2012, p. 5). Since then, the commercial bail industry has grown into one of the largest and most profitable parts of the criminal justice system – there are around 14,000 bail agents in operation across America who extract $2 billion a year from America's most disadvantaged families (Page 2017), though publicly-available statistics on BB practices are typically hard to come by (Justice Policy Institute, 2012, p. 33).

In bail systems, after arrest, courts can release defendants, detain them, or require them to pay an amount of money to reduce the risk that they will fail to appear in court at a later date. Often defendants cannot afford bail, so private companies (bail bondsmen) charge a non-refundable premium of the bail to the defendant and pay the full amount on the defendant's behalf to secure their pretrial release. Typically, fees of between 7 and 20% of the surety value are charged (Shouse California Law Group 2022; Bail Agent Network 2022; Breston 2019). Figure 1 and Table 1 illustrate the variations in the premiums that bail agents charge across the 50 US states (AboutBail 2022).

Private Bail Fees	No. of States
No Private Bail	6
No Max set (inc. Texas)	14
10%	9
15%	3
20%	1
10% Max	4
12% Max	1
15% Max	3
10% mandated	2
At least 6.5%	1
Other	7

Table 1. Variation in Bail Agents' Premiums by US State*

* 50 States plus Washington DC.

Co-signers, which are usually defendants' family members or friends, act as de-facto underwriters, assuming responsibility for ensuring the defendant makes it to court, placing

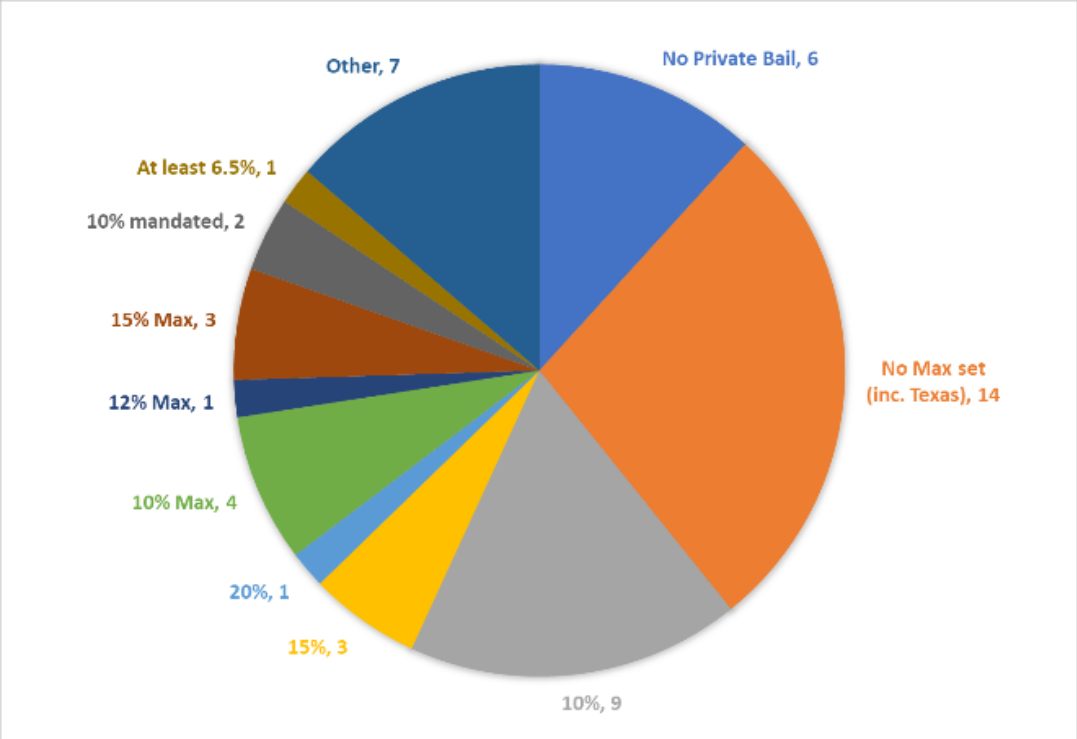

Figure 1. Variation in Bail Agents' Premiums by US State*

themselves at considerable jeopardy if that is not the case. Likewise, in many States, including Texas, an attorney can act as a surety at their own discretion (Texas Occupations Code, Sec. 1704.163). If defendants fail to appear, the court can recover the full amount from the co-signers (Page 2017). If a defendant absconds, BB recovery agents, or "bounty hunters" as they are colloquially known, are often employed by BB companies to track down and bring the defendant into custody. If they cannot locate the defendant, the responsibility lies with the co-signer. Bounty hunters are less regulated than the police and other law enforcement personnel; because of their status as private actors they have "broad powers of apprehension, surpassing those powers that are granted to governmental officials when conducting arrests, searches, and seizures" (Johnson and Stevens 2013: 190).

The commercial bail industry is symptomatic of a two-tiered justice system (Page 2017) where those who can afford to buy their freedom (however temporary) are rewarded and those who cannot are penalised; it is, "one for the haves and one for the have-nots" (Justice Policy Institute, 2012, p. 12). As such, it is highly divisive on many levels, presenting a complicated and contested legal, social, and economic lens through which issues surrounding neoliberalism can be focalised. Some (Baker *et al.* 2008b: 125) view BBs as an integral part of the American criminal justice system, helping to relieve the costs associated with pre-trial incarceration and serving the "utilitarian interests and needs of both the defendants and the courts" by offering legal advice, reminding the former of court dates, and recommending attorneys (though the latter is barred in Texas and would cause a BB company's licence to be revoked (Texas Occupations Code, Sec 1704.252 (Texas State Capitol 2021: 11)). Others (Harper *et al.* 2021; Lartey 2020; Stevenson 2018; Justice Policy Institute 2012) propose the elimination of cash bail and LFOs and

advocate for improved access to legal services and financial guidance for defendants both when in the criminal justice system and when re-entering the community.

The very pervasiveness of the for-profit bail industry in contemporary American society means it has established itself as an embedded, "everyday part of the criminal justice system"; the Justice Policy Institute (2012, p. 6) observes that commercial bail proponents use "the language of tradition to portray the practice as integral to the operation of the American criminal justice system", and hence the absence of BBs is unimaginable. It is precisely the language of commercial bail bond companies that this study is concerned with.

Bail bond website discourse

One significant area that has been overlooked by an extensive body of multidisciplinary research on the commercial BB industry is the public-facing discourse of BB companies themselves. As such, this paper is the first to empirically analyse the language BB companies use to address prospective users of their services via their websites, placing that analysis in its socio-legal context.

The public-facing language of BB of websites is a powerful resource at the disposal of BB companies in the context of the power imbalances inherent within the American criminal justice system. We draw on Foucault's (1969: 54) explanation of discourse "as practices that systematically form the objects of which they speak". Therefore, we conceptualise language use as a form of social practice which represents, reflects and makes sense of reality whilst also informing, construing, and shaping the reality that is lived out day to day in society (Baker and McEnery 2005: 198). Consequently, the language of BB websites is inextricably situated in and constitutive of its discursive context.

As a means of representing and shaping reality, discourses are underpinned by ideologies. The former provides a means by which the latter can be expressed, actualised, and reproduced in communicative practices (Flowerdew and Richardson 2018; Wodak 2001); discourses serve as "narrative and framing" vehicles through which ideologies are "presupposed, assumed, normalized, and reinforced through repeated variations of a theme" (Chun 2018: 7). As such, dominant or powerful groups in society, especially those in control of public discourse, such as commercial bail proponents, can use discourses in the reproduction of dominant knowledge and ideologies in that society to meet their own demands and to instantiate and sustain social inequalities (Dijk 2005). Likewise, discourses can also be sites of resistance against these dominant systems of belief (Burr 2015). As such, certain aspects of an event, concept, or object can be selectively emphasised, while others are downplayed or concealed.

When viewed as social practice that reflects and informs reality, BB website discourse offers a unique insight into the complex nexus of neoliberal ideologies and agendas in the American justice system and how they may be sustained or challenged through language (Fairclough 2015). The following analysis of BB website discourse provides insights into how the neoliberal agenda interacts with the criminal justice system by examining: firstly, how BB companies discursively present and promote their services to prospective service users; secondly, how BB companies represent the criminal justice system and its processes to prospective service users; and thirdly, how BB companies construe arrest and detention to prospective service users. The latter is important given the extremely high incarceration and recidivism rates in the US.

The Federal and State legal framework for Texas

This first examination of BB website discourse is based on data from the state of Texas, selected using a random State generator. Here we outline the legal context of bail in Texas, focussing on the regulation of BB companies and their advertising. The legal context is important to ground the neoliberal framework in which we situate our linguistic analysis. As this is the first analysis of its kind, particular attention has been paid to the complex, relevant legislation which inevitably controls the parameters of the ways in which bail bond companies advertise their services.

The overarching Federal basis across the US for allowing defendants bail, be it publicly or privately organised, derives from the 8th Amendment of the US Bill of Rights (Full Constitution of The United States) which states: "Excessive bail shall not be required, nor excessive fines imposed, nor cruel and unusual punishments inflicted". As Texas has recently changes its bail laws, including a recent State-specific problem for disadvantaged defendants, it has a particularly complex legal framework in relation to licencing and advertising. This merits an explanation of the background against which the BB companies advertise their services.

Senate Bill 21 was passed in the final days of the second special session of 2021 and aims to further restrict release without cash bonds by restricting release on personal bonds. The Bill, which was signed into law as the Damon Allen Act:

> "…[will] block those accused of violent crimes from being released from jail without paying bail. Opponents say it discriminates against poor people."

The overall regulation of Texan BB sureties derives from the Occupations Code, Title 10, Chapter 1704. Although this is a State regulation, it is devolved to the Counties, complicating how the BB industry is managed at State level. Devolved regulation controls licencing, location, and the fees that can be, or must be, charged. Advertising methods are controlled by overlapping Federal and State laws. Federal regulation is found in the US Federal Trade Commission Act (2006), enforced by the Federal Trade Commission (FTC). "[U]nfair or deceptive act[s] or practice" is outlined in section 45(a)(1) of the Federal Trade Commission Act and FTC can request sanction from the Attorney General (and State Attorneys). The AG's duty is "to cause appropriate criminal proceedings to be brought" (US Gov 2006: § 56(5)(b)). In practice, the Commission often uses "§ 57b Civil actions for violations of rules and cease and desist orders respecting unfair or deceptive acts or practices" (US Gov 2006).

Sanctions include cease-and-desist orders if the violation is "*one which a reasonable man would have known under the circumstances was dishonest or fraudulent, the court may grant relief*", including but not limited to "*rescission or reformation of contracts, the refund of money or return of property, the payment of damages, and public notification respecting the rule violation or the unfair or deceptive act or practice*" (US Gov 2006: § 57b(b)). This may dissuade commercial companies from making unreasonable claims or statements (in their advertising). However, as BB companies' clients are accused of a criminal offence(s), normal company/client power dynamics are altered; users are less likely to challenge unreasonable claims or statements in advertising should they identify them.

This can be contrasted with the usual commercial/client relationship where clients are typically able to take their business elsewhere or seek legal redress. Defendants or

their co-signers are stigmatised by criminal accusation and face the threat of incarceration if the bond is breached. A co-signer faces financial repercussions. Access to justice is limited for those who cannot afford to pay for high quality legal representation to challenge unfair practices; state-funded legal assistance is very limited - "*more than 50 percent of those seeking [state funded] help are turned away*" (US, DoJ).

All BB companies must be licenced, and county Bail Bond Boards can revoke licences for breaches of the code (Sec. 1704.151). Sec. 1704.109 regulates "*SOLICITATION AND ADVERTISEMENT*" to protect: "(1) the public from: (A) harassment; (B) fraud; (C) misrepresentation; or (D) threats to public safety; or (2) the safety of law enforcement officers". This code also limits unsolicited contact with those arrested by means other than for "*Class C misdemeanor[s]*". Sec. 1704.252 (10) allows a Board to revoke a licence if they "*solicit bonding business in a building in which prisoners are processed or confined*", thereby limiting advertising near to detained persons. Likewise, Sec. 1704.304 controls who can recommend bail bond companies and where recommendations can be made. BB companies, for example, cannot recommend attorneys, and police officers, sheriffs, deputies, constables, jailers, or employees of a law enforcement agency, judges, employees of a court or public officials cannot recommend BB companies. Soliciting BB sureties is not permitted in police stations, jails, prisons, detention facilities, or "*other places of detainment for persons in the custody of law enforcement*" and *"[a] person may not place a device in a place of detention, confinement, or imprisonment that dispenses a bail bond in exchange for a fee*".

Advertising online bypasses these in-person restrictions. Thus, although the general accessibility of websites arguably improves access to justice, the BB websites themselves provide a particularly crucial resource for BB companies to present and promote their services. Hence, we focus our analysis on website advertising.

Data and Method

This section outlines the data and methods used for its analysis.

The BBW Corpus

We created a specialised corpus from twelve BB company websites (the BBW corpus), randomly selected from a Google search for "bail bonds companies" in Texas. Texas was also selected randomly using a State generator. Websites are a widespread and accessible medium for advertising financial services and therefore provide an important public-facing platform for BB companies (Cook 2001). Brookes and Harvey (2017a) note that websites can be particularly attractive to prospective users of alternative financial services, such as payday lenders, because the relative anonymity of the internet "insulates" against the stigma and stress that using these services may carry. Additionally, given strict regulation of soliciting contact with defendants in person, online advertising is a particularly important resource for BB bonds companies in Texas. Likewise, BB company websites provide a primary source of information for defendants and their families.

Specifically, the BBW corpus comprises the two website pages that all twelve websites had in common, "Home" and "About Us" (or just "About"), and the content of which were most relevant to our research questions. An initial survey found that these pages were the best for capturing, for instance, data where BB companies dedicated language to presenting and promoting their services. "Home" pages are the main landing point for

visitors and prospective customers; they are crucial for communicating key messages and creating traction. "About Us" pages typically provide a broad background on the company, including its values and objectives, as well as the bail services it provides. In total, the BBW corpus comprises 23 web pages from 12 websites, totalling 11,271 words.

Whilst the BBW corpus is a specialised corpus compiled through the principled selection of specific text type, it is not a statistically representative sample of all BB website discourse and we do not treat it as such in our analysis (Koester 2010). Creating a corpus that would be wholly representative of BB website discourse is unrealistic given the volume of BB companies in operation in the US, and the variation of their regulation across different states. Therefore, we have intentionally created a small BB website corpus which we expect to yield important initial insights into how BB companies present their services, legal processes, and arrest to the public in Texas at least. Additionally, as the first empirical investigation of this discourse type, a small corpus enables greater familiarity with each text and lends itself to the more qualitative aspects of critical discourse analysis which are explained below.

Corpus-assisted critical discourse analysis

This study examined our specialised BBW corpus through corpus-assisted critical discourse analysis, combining the computational methods of corpus linguistic tools with the analytical perspectives of critical discourse analysis (CDA). The use of corpus-assisted CDA is now well established and has, for instance, been successfully applied in research on representations of different groups of people (Baker and McEnery 2005), diseases and illnesses (Siiner 2019; Hunt and Harvey 2015), societal issues, like marriage (Kania 2020), air pollution (Wang 2018), violent jihadist discourse (Brookes and McEnery 2020) and language ideologies (Wright and Brookes 2019). Our data is grounded in this genre of societal issues, particularly those which can be affected by issues of power imbalance, and as such is ideally suited for corpus-assisted CDA.

Corpus linguistics is a computational methodology that uses software to examine large amounts of electronically stored linguistic data and identify salient, as well as non-obvious, linguistic patterns. As such, many corpus linguistic methods are primarily quantitative, providing a "map" of a corpus and a valuable entry point into data by identifying linguistic patterns or features of interest (Baker *et al.* 2008a: 293). Whilst this "map" can provide insight into salient patterns of language use, it cannot help to explain the meanings and functions of them, nor can it place them in their wider context beyond the corpus. Likewise, purely quantitative techniques cannot readily analyse how such patterns might be realised in more subtle, nuanced, complex, and even coded ways (Baker and Levon 2015: 234). To explain and meaningfully interpret their findings, quantitative corpus linguistic methods are often supplemented and expanded by more qualitative techniques that do account for such nuance and complexity, as well as the social, political, historical, and cultural contexts of the data (Baker *et al.* 2008a: 293).

Critical discourse analysis refers to a collection of approaches to examining discourse in ways that *are* intimately concerned with connecting texts and their linguistic features with their social, political, historical, and cultural contexts, as well as who produced them, for whose benefit, and with what purpose (Baker 2010). As such, CDA approaches, which traditionally prioritise qualitative methods, aim to identify and reveal precisely those implicit, subtle and "hidden features of language use" that may not be readily available to many quantitative corpus methods, as well as those that are more

transparent (Flowerdew and Richardson 2018: 21). In doing so, CDA investigates "critically social inequality as it is expressed, constituted, legitimized [sic], and so on, by language use" (Weiss and Wodak 2003: 15); it sets out to discredit claims to authority made through language use (Flowerdew and Richardson 2018: 21). The "critical" in CDA refers to raising awareness of the status quo (Flowerdew and Richardson 2018), and understanding and explaining how discourses relate to issues of power, dominance, ideology, and control. This includes how discourse relates to aspects of the social world, including practices, norms, structures or systems, and how these are sustained or challenged through language (Fairclough 2015). This is because critical discourse analysts conceptualise discourse as social practice that both reflects and construes reality.

There is a "methodological synergy" (Baker *et al.* 2008a: 274) and complementarity in the methods normally used by corpus linguistics and CDA research. Corpus methods help to "tie down" (Baker and Levon 2015: 233) qualitative interpretations typical of CDA research whilst the results of statistical corpus methods also need to be interpreted up-close (Cicourel 1969). In seeking to make the implicit explicit, corpus-assisted CDA is well positioned to aid our investigation of "neoliberal common-sense beliefs" in the context of the American justice system, by asking "who benefits and who is left behind in our societies" (Chun 2018: 14).

Keywords

To ensure our route into the BBW corpus was based on statistical saliency rather than researcher intuition, we began with a keyword analysis (Charteris-Black and Seale 2010). Keywords are words that appear significantly more frequently in a specialised corpus (such as the BBW) than would be expected when compared to their frequency in a reference corpus, pointing to the "aboutness" of a text (Scott and Tribble 2006). Reference corpora are usually very large and represent a chosen language across a variety of contexts within a certain time period; they act as a standard against which a specialised corpus is statistically compared to identify the keywords. This study's reference corpus is the American English 2006 corpus (AmE06) (Baker *et al.* 2006), a 1-million-word corpus of published general written American English, which is compiled of 500 texts, including newspapers, fiction, academic writing, and general prose, mainly published in 2006. Whilst other, larger, and more contemporary reference corpora are available, such as the 15-million-word Open American National Corpus (2022), corpus linguistic literature stresses that the best reference corpus is determined by its suitability to the study corpus in terms of size, rather than its superior size (Gabrielatos 2018; Berber-Sardinah 2000). Hence, as our study of BB discourse is intentionally small-scale, we have chosen the AmE06 as an appropriately sized reference corpus with which to create keywords (Baker 2013).

Given the relatively small size and non-representative nature of the BBW corpus, we used the effect size statistic, log-ratio, to calculate keywords in this study using the corpus software #LancsBox 6.0. Log-ratio indicates the relationship between word occurrences in the specialised corpus and the reference corpus, and emphasises absolute frequencies (Pojanapunya and Todd 2016). The log-ratio calculation generated 44 keywords which are listed in Table 2 below. Each keyword had a minimum frequency of 5 in the BBW corpus and a log-ratio keyness score of 5+. This list excludes proper nouns (e.g. Dallas), as well as numerical values, such as telephone numbers, and contractions (e.g. won't).

Analysis

We examined concordance lines of each of the 44 keywords to determine their overall usage in the BBW corpus and sorted them into semantic categories. When grouped into semantic categories, keywords point to the major themes of a corpus and act as a signposting tool from which the researcher can begin to identify discourses (Scott 2010). Where an appropriate semantic group already existed, a term was added to it. Where there was not, a new group was created to accommodate it. As Table 2 shows, this bottom-up process identified five main themes in the BBW corpus and numerous subgroups. The keywords are listed with decreasing keyness scores.

Semantic Categories	**(Subgroups)**	**Keywords**
BB Company Referents		bail, bondsman, bondsmen, bonds, bond, bonding, agents
BB Services	Accessibility/ Availability	24-hour, affordable, convenience, mobile, nationwide
	Credibility	licensed, knowledgeable, reliable, professional
	Privacy	confidential, confidentiality, discrete,
	Speed	fastest, fast, quick
	Money	payment, flexible, financing
	Object/ product of services	release, assist, customer
Legal	Concepts	notary, defendant, warrants, defendants, processed, paperwork
	Offences	DUI, DWI, misdemeanor, tickets
Arrest/ Detention		jails, jail, arrests, arrested
Family		loved

Table 2. Semantic categories of top 44 lexical keywords in BBW corpus

The five main semantic categories provide an insight into BBW corpus' main themes. Unsurprisingly, the first category concerns referents to BB companies. *bail*, *bondsman*, and *bondsmen* make up the top three keywords. The prominence of these terms indicates how specialised the BBW corpus is in comparison to general written American English. The description of BB services category is the largest, with six subcategories which mainly contain adjectives with typically positive associations. Hence, this initial survey of the data suggests BB services are routinely framed in an appealing and favourable light, something unsurprising given the primary function of "Home" and "About Us" pages we noted in our first area of interest (the way in which BB companies discursively present and promote their services to prospective service users). The construal of the semantic category Arrest/Detention directly maps on to our third area of interest (the way BB companies construe arrest and detention to prospective service users given the extremely high incarceration and recidivism rates in the US). The semantic category, Legal, aligns with our second area of interest (the way BB companies

represent the criminal justice system and "normalise" their services). The fifth keyword category, Family, has only one term, the adjective *loved*, as in the noun phrase *loved ones*.

The remainder of the analysis interrogates these keywords and their semantic categories in their discursive context to address all three of our areas of interest. As will be shown, the three are closely interrelated.

Representations of bail bond services

This section presents an analysis of the keywords in the semantic category "BB Services" to address how BB companies discursively present and promote their services to prospective users. The dominant discourse in the data frames securing a bail bond and a defendant's release as highly routinized and streamlined. For instance, the adjectives *fast* (n=48) and *quick* (n=24) in the keyword subcategory "Speed" demonstrate the swiftness of this process:

1. We provide **fast** bail bonds in the following counties (Text#1)
2. Our licensed bail bond agents specialize in **quick release** from all jails (Text#5)

Similarly, BB services are highly convenient; they are both readily available (in terms of time and location) (Examples 3-4) and offer accessible (Example 5) (see Accessibility/Availability, Figure 2) and *flexible* ways of paying (Example 6):

3. Call our Dallas Bail Bond agents **24 hours a day, 7 days a week** (Text#8)
4. We provide a variety of easy payment options as well as **mobile** bond delivery (Text#5)
5. We guarantee you the best **affordable** bail bonds. (Text#7)
6. make sure to ask how you may benefit from our **flexible** programs (Text#10)

The speed, convenience, and simplicity of bail services is typified by Figure 2, which conceptualises the BB process in four steps: 1) defendants or their families contact the company at the point of arrest, 2) the relevant paperwork is completed, 3) payment of the non-refundable premium is made, and 4) the defendant is granted bail. In most cases, these simplified representations presuppose the success of BB applications.

The second step in particular streamlines the BB process by eliding defendants' and co-signers' important legal and financial commitments and obligations into one simple umbrella term, *Paperwork* (n=5) (see Table 2, Legal). Two thirds of the websites in the BBW corpus had secondary pages that provided more information on these issues (e.g. "How bail works" and "Post-arrest guide"). However, that this information is largely absent on the Home and About Us pages, which communicate key messages and have the most traction, may indicate that BB companies may not want this information to be available initially in case it deters prospective service users. Equally, BB companies may simply not consider this information to be immediately relevant to prospective service users or they may assume that some (repeated) service users are already familiar with this process. In addition, as Examples 7-9 show, BB companies position themselves as actively pruning (*minimize*), downplaying (*basically*, *just*), and accelerating (*speeding up*) the paperwork stage. The following section will illustrate how legal processes are constructed as inaccessible and impenetrable in the BBW corpus.

7. We'll **minimize** paperwork and never throw away time (Text#2)
8. **Basically,** the individual with the warrant **just** needs to come to our office [...] and **complete the bonding paperwork** (Text#12)

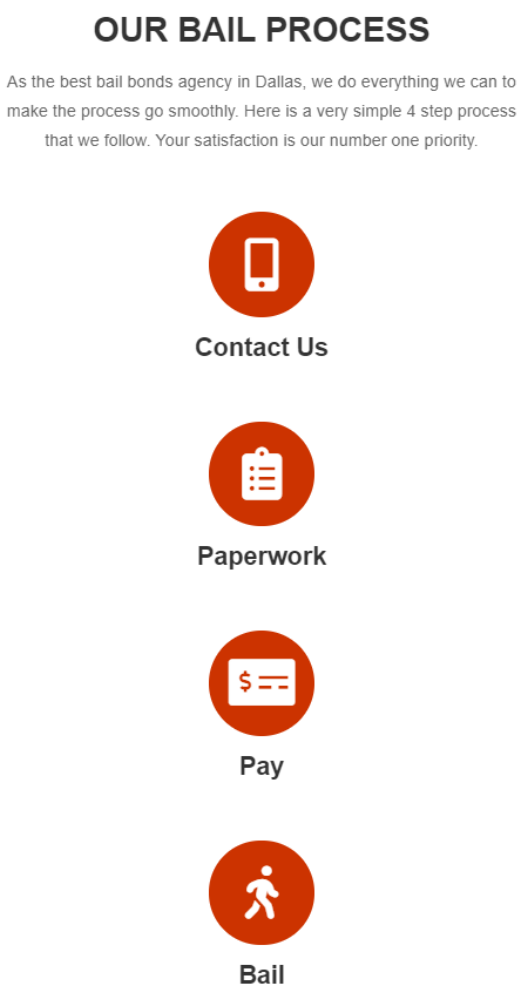

Figure 2. Simplified four-step bail process (Text#6)

9. paperwork takes between one and four hours, so our bail bond service focuses on **speeding up that process**. (Text#5)

Through a discourse of speed, convenience, and simplicity, securing bail is highly routinized to the point of near automation, which is comparable to Brookes and Harvey's (2017a) findings on payday loan lenders. This discourse has the effect of normalising the use of BB services, a process which Gavriely-Nuri (2018: 7) explains positions social objects as "forming part of normal life and normal society". This normalisation is reinforced across the corpus by numerous quantifications of 1) time in the industry and 2) previous customers, which Figure 3 and Figure 4 illustrate.

	Node	
YOU CAN TRUST! For more than 10	years	(since 1995), A-Affordable Bail Bonds has been
released from jail quickly. With over 24	years	of experience, we've earned an excellent reputation
Keller and Tarrant County. Over our 26	years	in business, we've earned an excellent reputation
Licensed Bail Bondsmen After more than 26	years	doing bail bonds in the Dallas/Fort Worth
Qualified Bail Bondsmen With more than 26	years	in the Dallas/Fort Worth bail bond industry,
for your privacy. With more than 26	years	in the Keller bail bond industry, we
been in business since 1990. Our 30	years	of experience make us more than qualified
the bail bonding profession. During his 33	years	of experience in this proud profession, he
us out from others. With over 35	years	of combined experience under our belt, we
Houston Metropolitan Area for more than 60	years,	maintaining a reputation for high-quality service and
Chair of the COP for nine consecutive	years	until 2000. In February 2004, he was
been in the Bail Bonds industry for	years	in Texas. Our professional Dallas County Bail
Hwy. 67 and Polk St. We have	years	experience in the Texas bail bonds industry.
family together again. We have dedicated many	years	to providing excellent service to the community.

Figure 3. Concordance lines of *years*

	Node	
put forth unparalleled service to work for	thousands	of satisfied customers every week. About Us
people who truly love you. We've helped	thousands	across Houston with the bail bond process.
Our Reputation In Houston Precedes Us With	thousands	of happy customers and the title of

Figure 4. Concordance lines of *thousands*

Such quantifications evidence each individual company's Credibility (see Table 2) and legitimacy. Concurrently, they also are testament to the common use of BB services in general. In this way, prospective users are reassured that their individual legal and/or financial predicaments are entirely unexceptional to BB companies who have encountered similar situations thousands of times before. Consequently, using BB services is constructed as a normal and accepted form of every day legal and financial self-management.

In analysing concordance lines of the keywords *bail* and *release*, we also identified a pattern in the data where BB services are positioned as protecting and upholding defendants' constitutional rights. For instance, Examples 10-13 illustrate how BB companies refer to the defendants' entitlements to liberty when describing their underpinning values.

10. We **believe** in the **right to bail and our mission** is to help you get out of jail [...] (Text#1)
11. **Our mission statement**. [...] **we believe** that all people are innocent till proven guilty and **we support your constitutional right to be released on bail** before you go to court (Text#10)

12. Our professional services will **help you** get back your freedom and give you **the second chance you no doubt deserve**. (Text#2)
13. We at [company name] **believe** everyone deserves **second chances** (Text#10)

The *right* to bail, a *second chance* (Examples 10, 12, 13), and preserving personal freedom are central to these companies' beliefs (Examples 10, 11, 13); they are commonly held truths of which there is *no doubt* (Example 12). In Examples 10 and 11, the BB companies' raisons d'être (their *mission*) is to uphold the constitutional right to bail; overall, BB services provide a necessary and morally just function, to *help* (Examples 10, 12) and *support* (Example 11) this basic entitlement. As BB services are positioned as stemming from highly virtuous and altruistic principles and motivations which empower and benefit its users, the BBW data legitimises not only each company's provision but also the very basis of the BB industry itself.

This analysis has so far shown how BB services are normalised through discourses of speed, convenience, and simplicity, and are legitimised or justified by a discourse of protecting defendants' constitutional rights. We will now briefly demonstrate how the experience of (the keyword) *release* that follows bail is represented in the BBW corpus as restorative and transformative.

Firstly, the representation of release typically implies that life will immediately return to normal for defendants. The nouns *normalcy* [(sic) normality] and *routine*, and the verbs *regain*, *restore*, and *continue* in Examples 14-16 below illustrate how securing bail enables a return to the status quo and a state of seemingly unbounded freedom:

14. To **regain your freedom** and **restore normalcy**, get help fast from [...] (Text#5)
15. we're dedicated to helping you get your loved ones out of jail and **back to their daily routine** as quickly as possible (Text#11)
16. Once done, you can **continue to enjoy your freedom** and peace of mind without worrying about huge debt incurred via bail bonds (Text#2)

This return to normality is also focalised through the notion that BB services reunite families:

17. Our well-experienced professional bonding agents won't stop until you're **back** where you **should be** –at home with your loved ones! (Text#2)
18. We want your family together **again**. (Text#10)

These constructions contain adverbs (*again*) and modal verb phrases (*should be*) that imply that the natural and default mode is for families to be together, and that BB services ensure that this is upheld. Moreover, in the BBW corpus, securing a bail bond is represented as providing defendants with new opportunities:

19. whenever you or your loved ones get arrested, to secure their freedom, and **move on to better things** [...] (Text#2)

and a new lease of life:

20. Call now at [number] for **a better life break** (Text#2)

The discourse of the restorative or transformational properties of BB services is highly idealised; the experience of bail is coloured in "positive tones" (Gavriely-Nuri 2018: 7). It purports that the purchase of freedom is permanent, absolving defendants of all their

problems. Restorative or transformative conceptions of release, premised on the notion that BB companies *save the day* (Text#4), obscure the reality of the bail experience, namely that this freedom is conditional, and, in some cases, may only temporarily defer incarceration. Additionally, this conception of BB services eschews that the time spent on bail might be rife with the negative human consequences of awaiting trial such as anxiety and stress, which may be compounded by further financial hardship. The appealing transformative property of BBs also mirrors that of British payday lenders. Brookes and Harvey (2017b: 177) assert that transformative propositions on payday lending websites "de-stigmatise" payday lenders by legitimising their incredibly high interest loans on the basis that their loans enable borrowers to "improve their personal, social and financial standing".

Like the discourse of protecting and upholding defendants' constitutional rights, this restorative or transformational discourse is another means through which BB services are legitimised as morally virtuous organisations. Such attempts to legitimise BB services and their motivations mirrors justification, another strategy for normalisation which Gavriely-Nuri (2018: 7-8) asserts "aims to depict the social object as just, rational, worthy of support".

Representations of the legal system

Despite the legal framework in which commercial BB companies operate, it is of note that whilst the BBW corpus comprises BB website "Home" and/or "About Us" pages for all the selected BB companies' websites, not all websites contain other pages that focus specifically on the legal and financial aspects of using BB services, including the obligations of defendants and co-signers.

The top keyword in the legal semantic category, *notary*, was used exclusively in the noun phrase *notary public.* Concordancing found that *warrant* (noun) was typically something the BB companies could *verify*, *help with*, *resolve* or simply *lift*. *Paperwork* is discussed in the previous section regarding the normalisation of BB services.

Considering that naming conventions raised by other disciplines are bound up in issues of power (Donzelot 1979), *defendant* (n=22) (and the plural *defendants* (n=5)) provides an important insight into how BB websites position the (prospective) users of their services. Users of BB services are also referred to by the keyword *customer* (n=14) in the corpus (see Object/Product of Services semantic category), as well as *client* (n=16). *Defendant* almost always occurs with the determiner *the* (n=15), having no pronominal possessors. *Customer* and *client*, however, are often pre-modified by the first-person possessive pronoun *our* (n=5 and n=7 respectively), which directly affiliates these individuals with BB companies. Interestingly, one instance in the data attempts to explicitly distinguish between *defendant* and *client* by removing agency from the latter, stating that:

21. To us you're not a **defendant**. You're another **client** who has found themselves in some unfavourable circumstances (Text#2)

We examined concordance lines of *defendant*, *customer*, and *client* in this study's reference corpus (AmE06) to determine their general usage in contemporary written American English. Unsurprisingly, *defendant* primarily occurs with legal terminology (*trial*, *law*), while *customer* reflects a commercial discourse (*service, base, consumer, benefits, investment, satisfaction*). The general written American English use of *client* revealed a

mix of commercial usage (*retail*, *trader*, *interview*) and therapeutic usages (*treatments*, *sensations*, *counsellor*). *Client,* in particular, implies an equal power balance between the BB company and the defendant. The BBW corpus refers to users and prospective users of BB services using one term that is generally rooted in legal discourse and a second that is rooted in commercial discourse in comparable frequencies, whilst the third pervades commercial and therapeutic discourse types. This elision between legal and commercial realms is indicative of the wider agenda and positioning of BB companies.

As there were only 6 keywords in the Legal semantic category (excluding the subcategory Offences), we widened the investigation by using other terms related to the keywords, such as *law*, and employing collocation analysis to consider other salient lexical patterns surrounding these keywords. Collocation assesses the strength of the relationship between two words in a corpus and the significance of that relationship (Rayson 2015). For instance, the title of the keyword category, "legal", collocated with the noun *system* to give the noun phrase *legal system* (n=5) (where collocates were calculated using Mutual Information, with statistic value above 3, and a minimum frequency of 5). Concordance analysis revealed that the *legal system* is typically presented as a complicated entity (Example 22) that is difficult to *navigate* (Example 23):

22. we will be there to offer all our services and lead clients through **the complicated legal system** which we'll make simple and easy to follow. (Text#2)
23. Our intimate knowledge of the **legal system** and procedures will help you **navigate** this difficult situation with ease (Text#11)

The structure of these examples follows the format that (1) BB companies or their assets, *we*/*our knowledge*, (2) perform processes, *lead clients*/*help you*, that (3) simplify the legal system or make it more accessible, e.g. *easy to follow*/*with ease*. Consequently, the legal system is positioned as inaccessible or impenetrable without BB companies' assistance and specialist knowledge. Perhaps the best example of this discourse is Example 22 when placed in its broader discursive context, as below:

24. we will be there to offer all our services and lead clients through the **complicated legal system** which we'll make simple and easy to follow. **Bureaucracy** is a pain and we know exactly how to deal with it (Text#2)

The *complicated legal system* is directly associated with the noun *bureaucracy*, which is used here to denote ineffective or excessively complicated official procedures. Hence, the legal system is represented as an inaccessible and inconvenient process to defendants (*a pain*) to which this BB company provides a neat resolution – they know how to *deal with* it.

Expanding the search criteria further, we found that *system* (n=14) also co-occurs with the following terms which are interchanged with *legal*: *bail systems* (n=1), *court systems* (n=1), *judicial system* (n=2), *criminal justice system* (n=1), and, simply, *the system* (n=2). As Examples 25-28 show, these terms are also framed as inaccessible:

25. The Houston **court and bail systems** may appear very **confusing and complicated** to most people (Text#2)
26. We can [...] help **familiarize** you with **the judicial system** (Text#11)
27. Our **dedicated** bail bondsmen know how **the system works** (Text#6)
28. We founded [company name] with the goal of educating our clients about **surviving** in **the judicial system** (Text#2)

Once again, to lay people, the legal system is presented as a complicated and confusing entity (Example 25) that requires familiarisation (Example 26) or specialised staff to understand (Example 27). In Example 28, through the intransitive verb *surviving*, the *judicial system* is associated with words typically semantically related to events of natural disaster, conflict, and destruction. For instance, in the AmE06 reference corpus, concordance lines of *survive* show that it is associated with terms such as *storm*, *drought*, *battle*, *disease*, and *combat*. Hence, the construction in Example 28 implies that without the BB company's guidance defendants will perish in the judicial system. Additionally, as in Examples 10-13, where BB companies are positioned as having highly virtuous and altruistic principles and motivations, in Example 28, the purpose of this company is to educate their clients.

Representations of arrest and detention

The presentation of Arrest and Detention, another key semantic category, in the BBW corpus was a complex and contested topic. The analysis identified one discourse whereby arrest is positioned as a non-routine, exceptional and potentially embarrassing experience. Firstly, through concordance analysis of *arrest* (n=37) we found a recurring pattern that acknowledges that, for both defendants and their co-signers, arrest: 1) induces feelings of fear:

29. Arrests are **scary and stressful**, but you're not alone (Text#11)
30. Getting arrested, booked and jailed can be a **terrifying, confusing and dangerous** experience (Text#6)

and 2) causes stress:

31. Choosing [a bail bond agency] while you're **emotionally stressed** by the arrest of a loved one can feel overwhelming (Text#11)
32. We understand that dealing with an incarcerated loved one can be a traumatic, **stressful**, and frustrating experience (Text#1).

Secondly, arrest is framed as invoking feelings of embarrassment and hence it should be handled sympathetically (i.e., a *sensitive issue*):

33. When it comes to **sensitive issues** such as getting arrested, you need a reliable company (Text#7)
34. Being arrested by the police is troubling and, often times, **embarrassing** situation (Text#1)

The potential embarrassment caused by arrest was also identifiable in the Privacy subcategory of keywords that describe Bail Services. The adjectives in this subcategory typically emphasise these companies' discretion and confidentiality. For instance:

35. We offer **confidential**, fast, and quality, service to all our clients. (Text#12)
36. Call [company name] for **confidential service** that is lightning fast, dependable and **discreet**. (Text#6)

Confidentiality is expected of companies in this industry. However, when paired with a pattern of arrest causing fear, stress, and embarrassment, BB companies' conscious selection of adjectives like *confidential* and *discreet* may also imply that using their services is something that their customers may wish to conceal.

The discourse of arrest as a non-routine, exceptional, and embarrassing event for many prospective bail bond users was supported by analysis of the keyword *jail* (n=102). In the BBW corpus, being detained in jail is highly undesirable; it is *not a pleasant place to be* (Text#10) and a *hell hole* (Text#4). Examples 37-38 are explicitly directed at defendants' co-signers. Example 37 exemplifies imperatives in the data which implore friends and family to act to help their *loved one* (*Do not leave*). In turn, incarceration is positioned as something that defendants must be saved from.

37. **Do not leave** your loved one in jail, call us and we will get them out QUICK (Text#10)
38. By using a bail bond agency, **you can avoid leaving your loved one** in jail where they are vulnerable to violent incidents, pressure from police officers, and even contagious diseases like COVID-19. (Text#11)

Significantly, we also identified a counter discourse to the discourse of arrest as a non-routine, exceptional, and embarrassing experience, whereby arrest and detention are subtly naturalised. Gavriely-Nuri (2018: 7) asserts that naturalisation, which aims to represent a social object as "a force or event independent of human agency, or as an inevitable outcome of the laws of nature" is another strategy for achieving the normalisation of it.

As will be illustrated, this counter discourse is primarily realised through lexical items relating to time that modify *arrest* and its related concepts. Examples 39-41 demonstrate those constructions in the BBW corpus which presuppose or anticipate arrest:

39. Nobody plans **when** to go to jail. But **whenever it starts** to get messy, we'll always be there to help you (Text#2)
40. No one can predict life's challenges, including **the time at which** you or a loved one might be arrested (Text#5)
41. We know you don't make appointments to go to jail, so we're open **anytime** you're inconvenienced. (Text#2)

In Example 39, the adverb *when* presupposes that arrest will take place by framing it as a probability. Meaning "at what time", it positions the timing of arrest as unforeseen and problematic, rather than the fact of it. Without *when*, arrest in this clause is a simply an unplanned but hypothetical situation; [negative] + "to go". The conjunction *whenever* (n=5) also presupposes arrest. Meaning "at whatever time", *whenever* also modalises arrest (euphemised here as *get messy*) as a probability rather than indicating a possible event. For instance, compare *whenever* with alternatives, such as the conjunction *if* or the verb *should*. In Example 40, the time phrase *the time at which* performs a similar function. Again, the time of arrest is problematized, rather than whether it will occur at all. In this way, arrest is presupposed once more. Example 41 takes the same format, stating that detention is not prearranged, and hence BB services are available when such circumstances arise. Additionally, it is notable that, contrary to arrest being a scary and stressful experience, in Example 41 it is trivialised as a mere inconvenience.

The presupposing or anticipation of arrest is also discernible in Example 42 below where the threat of arrest and detention is positioned as an impending or looming part of everyday life:

42. Arrest and detention **can happen any time** of the day or night. (Text#6)

Finally, this theme is extended in Examples 43-45 which illustrate a pattern in the data whereby arrest and detention are positioned as recurring rather than discrete events. This is realised through both repeated use of BB services, as signalled by more time phrases (*Next time* and *once again*), and quantification (*multiple*, *more than one*):

43. **Next time** you or someone you know is in need of fast, professional bail bonding services, call Ace Bail Bonds (Text#3)
44. we hope you will leave us feeling happy. So if the need arises you will call us to help **once again** (Text#10)
45. our [...] Bail Bonds agents **regularly post multiple bonds** for [...] jails on a single individual who has been arrested for **more than one offense** (Text#12)

The adverb *regularly* in Example 45, which frames the use of bail bonds service as common, is indicative of the cumulative effect of these subtle patterns through which arrest and detention are naturalised.

Discussion and conclusions

This article set out to explore how commercial BB companies 1) discursively present and promote their services, 2) represent the legal system and its processes, and 3) construe arrest and detention to prospective service users on their websites. Unsurprisingly, as an accessible form of advertising, these websites overall represent BB services and their use by the public in an attractive and appealing light. However, this analysis has also unearthed that this favourable representation is underpinned by discourses that perpetuate neoliberal agendas and ideologies and has highlighted the power imbalances inherent within the relationship between the BB companies and their service users. As the findings presented above are particularly rich and their contextualisation complex, we focus this discussion on three problematic ways in which these inequitable ideas pervade everyday public discourse. It is important to remember that the users of BB services have not been convicted of a crime and have the right to be presumed innocent during their interactions with BB services and indeed more broadly in relation to their interactions with the criminal justice system.

Firstly, our analysis suggests that BB companies routinely pass off the commercialisation of freedom from incarceration, from which they routinely profit, palatably, perhaps insidiously, as moral resolve. In presenting themselves to prospective service users, BB companies make efforts to legitimise themselves as morally virtuous organisations that are just and worthy of support (Gavriely-Nuri 2018), positioning themselves as performing important and needed social and legal functions and serving the "utilitarian interests and needs of both the defendants and the courts" (Baker *et al.* 2008b: 125). This includes upholding defendants' constitutional right to bail and guiding laypeople through the process of obtaining bail, educating them about a legal system that is framed as otherwise wholly inaccessible, and helping offenders where arrest and detention is a frightening and stressful experience. However, it must be remembered that this is a for-profit commercialised industry that both relies and thrives on the continued legal, financial, and social hardship of others. The most vulnerable in society will be disproportionately affected by both an increased likelihood of contact with the criminal justice system and an increased likelihood of lacking the financial, educational, and social means to be able to mitigate this contact on an equal footing with a non-vulnerable citizen. Of course, it could be argued that, by definition, any accused person in the criminal

justice system becomes vulnerable no matter the level of their financial, educational, and social resources. However, the question of equal access to justice is a known issue within the criminal justice system in many jurisdictions so it is reasonable to conclude that the impact is compounded for defendants already occupying vulnerable or disadvantaged groups in society (see for example Rhode (2004); Cappelletti *et al.* (1982)). The Organisation for Economic Co-operation and Development (OECD) (2016: 12) note that "Unequal access to justice is expensive. Evidence is mounting that unresolved legal problems are costly both to the individuals directly affected and to society as a whole", concluding that "unequal access to justice also diminishes public confidence in the justice system which can further erode the social fabric".

Secondly, BB websites may have the effect of normalising BB services, not only as an "everyday part of the criminal justice system" (The Justice Policy Institute, 2012, p. 6) but also as an unexceptional mode of legal and financial self-management. Akin to payday loan lenders as alternatives to mainstream financial services (Brookes and Harvey 2017a,b), the use of bail is construed as highly routinized, owing to these companies' speed, convenience, and simplicity. These processes are so streamlined that the "Home" and "About Us" pages are able to downplay defendants' and co-signers' important legal and financial commitments and obligations to BB companies, whilst promoting their services in a more philosophical, moral, context. That bail services are positioned as "unobjectionably common" (Brookes and Harvey 2017a: 242) and something on which the public can so readily rely, directly reflects, and compounds the dependence that many low-income families trapped in cycles of indebtedness and poverty can and often do have on these companies (Harper *et al.* 2021; Stevenson 2018; Western and Pettit 2010; Pettit and Western 2004).

Thirdly, just as the use of BB services are positioned and presented on the websites as unexceptional, the arrest and detention of prospective users of their services is also naturalised (Gavriely-Nuri 2018). Discourses on BB websites that presuppose arrest and downplay it as an unexceptional or inconvenient experience are symptomatic of the exceptionally high rates of incarceration and recidivism in the US; when arrest is framed in public discourse as a commonly recurring life event, it echoes the US's generally low success with rehabilitating offenders (Alper *et al.* 2018) and precisely the desirable "revolving door" of reoffending that keeps BB companies in business (Ortiz and Jackey 2019: 484; Hallett 2012). This point highlights the inherent tensions between the State's (presumed) objective of reducing recidivism and the (presumed) objective of the BB companies to remain in business, for which they require a steady stream of users of their services.

Critics have already described actions by the bail industry, including lobbying legislation and defunding alternatives to bail, as forms of structural violence, designed to compound the legal, political, and economic inequality from which the powerful profit (Ortiz and Jackey 2019; Hallett 2012). This paper has highlighted that powerful groups may also subtly enact structural violence through "everyday" public-facing discourse, such as the language of BB websites (Dijk 2005). Public discourses that simultaneously serve whilst oppressing those they purport to help noxiously perpetuate neoliberal agendas and ideologies, and ultimately, maintain a (lucrative) two-tier justice system (Page 2017).

Given the state-by-state variability of regulations of the bail industry in the US, we have necessarily conducted this first examination of BB website discourse on one randomly selected state (Texas). Consequently, we have not treated the BBW corpus, which is relatively small, as representative of all BB discourse. In future corpus-assisted BB research we would be interested to examine how the financial and legal obligations of defendants and co-signers are represented in a larger corpus of BB website pages besides the ones analysed here, as well as in terms and conditions documents. However, we believe that our findings are indicative of the broader commercial BB industry in the U.S. The significance of our study lies in its highlighting of issues inherent in the commercialisation of services offered to marginalised and vulnerable groups as a result of coercive State processes. The issues raised in our study apply to other areas of coercive State processes across multiple jurisdictions where neoliberalist policies and their implementation are rising. They will continue to do so, particularly during periods of austerity or rising inflation when governments look to reduce State costs of implementing the complex legal frameworks involved in both policing and protecting society. We see the application of this analysis across key areas of the criminal justice and social justice systems. Bail bond services, adult incarceration services, juvenile incarceration, probation, and child protection are key areas of privatisation, but we anticipate there are many other examples for which this type of analysis is applicable.

References

AboutBail, (2022). How much does bail cost? aboutbail. https://www.aboutbail.com/pages/bail-cost.

Alper, M., Durose, M. and Markman, J. (2018). *2018 update on prisoner recidivism: A 9-year follow-up period (2005-2014) (Bureau of Justice Statistics Special Report, NCJ250975)*. U.S. Department of Justice.

ANC, (2022). Open american national corpus. https://anc.org/.

Bail Agent Network, (2022). *How Do Bail Bonds Work In Texas? Bail Agent Network*. https://www.bailagentnetwork.com/texas/.

Baker, P. (2010). *Sociolinguistics and Corpus Linguistics*. Edinburgh University Press.

Baker, P. (2013). Visiting with the brown family. esrc centre for corpus approaches to social science (cass). http://cass.lancs.ac.uk/visiting-with-the-brown-family/.

Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T. and Wodak, R. (2008a). A useful methodological synergy? combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the uk press. *Discourse & Society*, 19(3), 273–306.

Baker, P. and Levon, E. (2015). Picking the right cherries? a comparison of corpus-based and qualitative analyses of news articles about masculinity. *Discourse & Communication*, 9(2), 221–236.

Baker, P. and McEnery, T. (2005). A corpus-based approach to discourses of refugees and asylum seekers in un and newspaper texts. *Journal of Language and Politics*, 4(2), 197–226.

Baker, P., Potts, A. and Cristofaro, M. (2006). The american english 2006 corpus. https://www.lancaster.ac.uk/linguistics/about/people/paul-baker.

Baker, S., Vaughn, M. and Topalli, V. (2008b). Review of the powers of bail bond agents and bounty hunters: Exploring legalities and illegalities of quasi-criminal justice officials. *Aggression and Violent Behaviour*, 13, 124–130.

Berber-Sardinah, T. (2000). Comparing corpora with wordsmith tools: How large must the reference corpus be? In *Proceedings of the Workshop on Comparing Corpora*, volume 9, 7–13.

Breston, D. (2019). What you need to know about the texas bail bond process. In *How Do Bail Bonds Work in Texas?* https://www.houston-criminalattorney.com/bail-bonds-in-texas/.

Brookes, G. and Harvey, K. (2017a). The discourse of alternative credit: A multimodal critical examination of the cash converters mobile app. In A. Mooney and E. Sifaki, Eds., *The Language of Money and Debt*. Palgrave Macmillan, 233–256.

Brookes, G. and Harvey, K. (2017b). Just plain wronga? a multimodal critical analysis of online payday loan discourse. *Critical Discourse Studies*, 14(2), 167–187.

Brookes, G. and McEnery, T. (2020). Correlation, collocation and cohesion: A corpus-based critical analysis of violent jihadist discourse. *Discourse & Society*, 31(4), 351–373.

Burr, V. (2015). *Social Constructionism*. Routledge.

Cappelletti, M., Garth, B. and Trocker, N. (1982). Access to justice, variations and continuity of a world-wide movement. *Rabels Zeitschrift Für Ausländisches Und Internationales Privatrecht / The Rabel Journal of Comparative and International Private Law*, 46(4), 664–707.

Charteris-Black, J. and Seale, C. (2010). *Gender and the language of illness*. Palgrave Macmillan.

Chun, C. W. (2018). Neoliberalism, globalization and critical discourse studies. In J. Flowerdew and J. Richardson, Eds., *The Routledge Handbook of Critical Discourse Studies*. Routledge, 549–565.

Cicourel, A. (1969). *Method and Measurement in Sociology*. The Free Press.

Cook, G. (2001). *The Discourse of Advertising*. Routledge.

Dijk, T. (2005). *Discourse and Racism in Spain and Latin America*. John Benjamins.

Donzelot, J. (1979). *The Policing of Families*. Pantheon Books.

Fairclough, N. (2015). *Language and Power*. Routledge.

Federal Trade Commission, (2006). Federal trade commission act incorporating u.s. In *SAFE WEB Act amendments of 2006 (Unofficial version)*. https://www.ftc.gov/sites/default/files/documents/statutes/federal-trade-commission-act/ftc_act_incorporatingus_safe_web_act.pdf.

Flowerdew, J. and Richardson, J. (2018). Introduction. In *The Routledge Handbook of Critical Discourse Studies*. Routledge.

Foucault, M. (1969). *Archaeology of Knowledge*. Routledge.

Full Constitution of The United States, (-).

Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. *Corpus Approaches To Discourse: A critical review*, 225–258.

Gavriely-Nuri, D. (2018). Cultural approach to cda (ccda): From theory to practice. In J. Flowerdew and J. Richardson, Eds., *The Routledge Handbook of Critical Discourse Studies*. Routledge, 171–187.

Hallett, M. (2012). Reentry to what? theorizing prisoner reentry in the jobless future. *Critical Criminology*, 20(3), 213–228.

Harper, A., Ginapp, C., Bardelli, T., Grimshaw, A., Justen, M., Mohamedali, A., Thomas, I. and Puglisi, L. (2021). Debt, incarceration, and re-entry: A scoping review. *American Journal of Criminal Justice*, 46, 250–278.

Harris, A., Evans, H. and Beckett, K. (2010). Drawing blood from stones: Legal debt and social inequality in the contemporary united states. *American Journal of Sociology*, 115(6), 1753–1799.

Hunt, D. and Harvey, K. (2015). Health commmunication and corpus linguistics: Using corpus tools to analyse eating disorder discourse online. In P. Baker and T. McEnery, Eds., *Corpora and Discourse Studies*. Palgrave Macmillan, 134–154.

Johnson, B. and Stevens, R. (2013). The regulation and control of bail recovery agents: An exploratory study. *Criminal Justice Review*, 38(2), 190–206.

Justice Policy Institute, (2012). For better or for profit: How the bail bonding industry stands in the way of fair and effective pretrial justice. https://justicepolicy.org/research/for-better-or-for-profit-how-the-bail-bonding-industry-stands-in-the-way-of-fair-and-effective-pretr

Kania, U. (2020). Marriage for all ('ehe fuer alle')?! a corpus-assisted discourse analysis of the marriage equality debate in germany. *Critical Discourse Studies*, 17(2), 138–155.

Koester, A. (2010). Building small specialised corpora. In A. O'Keeffe and M. McCarthy, Eds., *The Routledge Hanbook of Corpus Linguistics*. Routledge, 66–70.

Lartey, J. (2020). New york rolled back bail reform. what will the rest of the country do? the marshall project. https://www.themarshallproject.org/2020/04/23/in-new-york-s-bail-reform-backlash-a-cautionary-tale-for-other-states.

Meredith, M. and Morse, M. (2017). Discretionary disenfranchisement: The case of legal financial obligations. *The Journal of Legal Studies*, 46(2), 309–338.

OECD, (2016). Understanding effective access to justice [workshop background paper]. https://www.oecd.org/gov/Understanding-effective-access-justice-workshop-paper-final.pdf.

Ortiz, J. and Jackey, H. (2019). The system is not broken, it is intentional: The prisoner reentry industry as deliberate structural violence. *The Prison Journal*, 99(4), 484–503.

Page, J. (2017). Desperation and service in the bail industry. *Contexts*, 16(2), 30–37.

Pettit, B. and Western, B. (2004). Mass imprisonment and the life course: Race and class inequality in u.s. incarceration. *American Sociological Review*, 69, 151–169.

Pojanapunya, P. and Todd, R. (2016). Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. In *Corpus Linguistics and Linguistic Theory*.

Rayson, P. (2015). Computational tools an methods for corpus compilation and analysis. In D. Biber and R. Reppen, Eds., *The Cambridge Handbook of Corpus Linguistics*. Cambridge University Press, 32–49.

Rhode, D. (2004). *Access to Justice*. Oxford University Press.

Scott, M. (2010). Problems in investigating keyness, or clearing the undergrowth and marking out trails. In M. Bondi and M. Scott, Eds., *Keyness in Texts*. John Benjamins, 43–58.

Scott, M. and Tribble, C. (2006). *Textual Patterns: Key words and corpus analysis in language education*. John Benjamins.

Shannon, S., Uggen, C., Schnittker, J., Thompson, M., Wakefield, S. and Massoglia, M. (2017). The growth, scope, and spatial distribution of people with felony records in the united states, 1948–2010. *Demography*, 54(5), 1795–1818.

Shouse California Law Group, (2022). *Bail Bonds in Texas*. Shouse Law Group. https://www.shouselaw.com/tx/defense/laws/bail/.

Siiner, M. (2019). Let me grow old and senile in peace: Norwegian newspaper accounts of voice and agency with dementia.

Stevenson, M. (2018). Distortion of justice: How the inability to pay bail affects case outcomes. *The Journal of Law, Economics, and Organization*, 34(4), 511–542.

Texas State Capitol, (2021). Occupations code, title 10. occupations related to law enforcement and security, chapter 1704. regulation of bail bond sureties. https://statutes.capitol.texas.gov/Docs/OC/htm/OC.1704.htm.

U.S. Government, (2006). Incorporating u.s. safe web act amendments of 2006.

Walmsley, D. (2019). Neoliberalism, mass incarceration, and the us debt–criminal justice complex. *Critical Social Policy*, 39(2), 248–267.

Wang, G. (2018). A corpus-assisted critical discourse analysis of news reporting on china's air pollution in the official chinese english-language press. *Discourse & Communication*, 12(6), 645–662.

Weiss, G. and Wodak, R. (2003). Introduction: Theory, interdisciplinarity and critical discourse analysis. In G. Weiss and R. Wodak, Eds., *Critical Discourse Analysis: Theory and Interdisciplinarity*. Palgrave Macmillan, 1–34.

Western, B. and Pettit, B. (2010). Incarceration and social inequality. *Daedalus*, 139(3), 8–19.

Wodak, R. (2001). What cda is about: A summary of its history, important concepts, and its developments. In R. Wodak and M. Meyer, Eds., *Methods of Critical Discourse Analysis*. Sage, 1–13.

Wrenn, M. (2015). Immanent critique, enabling myths, and the neoliberal narrative. *Review of Radical Political Economics*, 48(3), 452–466.

Wright, D. and Brookes, G. (2019). This is england, speak english!": A corpus-assisted critical study of language ideologies in the right-leaning british press. *Critical Discourse Studies*, 16(1), 56–83.

Appendix

List of websites included in the BBW corpus (all accessed October 2020).

Text Number	Website
Text#1	http://a-affordablebail.com
Text#2	https://abetterbailbond.net/
Text#3	https://www.acebailbonddfw.com/
Text#4	https://atlasbail.com/
Text#5	https://www.awayoutbonds.com/bail-bonds-keller-tx.html
Text#6	https://www.a1bonding.net/
Text#7	http://badboybailbondsservice.com
Text#8	https://dallascountybailbondstx.com/
Text#9	http://justbailbond.com/
Text#10	https://www.sunrisebailbond.com/
Text#11	https://deltabail.com/
Text#12	https://southernbailbonds.com/

Language as Evidence. Doing Forensic Linguistics

Book review by Karoline Marko

University of Graz, AT

https://doi.org/10.21747/21833745/lanlaw/9_1r1

Language as Evidence. Doing Forensic Linguistics
Victoria Guillén-Nieto & Dieter Stein (Eds.) (2022)
Cham, Switzerland: Palgrave MacMillan

The presence of linguists as expert witnesses in court has grown over the past years, which has sparked a discussion about the separation of forensic linguistics from legal linguistics that can be traced back at least to Kniffka (1990), and later to Fobbe (2011). In this regard, the publication of this edited volume reflects the tenor of the time that is also visible in the recent name change of the *International Association of Forensic Linguists* to *International Association for Forensic and Legal Linguistics*. As argued in the introduction to this edited volume, the difference between legal linguistics and forensic linguistics manifests itself in different training requirements: "the legal linguist is a theoretician, but the forensic linguist has the practical task of having to act, to appear at court, know the rules of conducting and executing this part of the law, different in different legal cultures [...] and also acquire a rhetoric for presenting evidence at court" (p. 16) – the present book thus aims at providing the tools and knowledge necessary for the work of forensic linguists.

In this respect, the book fills a specific gap in the existing literature in that it combines a discussion of a variety of topics related to linguistic evidence, such as authorship studies and deception detection, as well as genres such as suicide notes, romance scams, and online terrorist threats. Previous books have either addressed a larger scope of forensic linguistics including legal linguistics (e.g., Coulthard *et al.* 2017; Solan and Tiersma 2005), have focused on one particular type of evidence such as authorship attribution (e.g., Dern 2009; Love 2002), or have discussed legal linguistics or language in the legal

process separately (e.g., Eades 2010; Kaplan 2020). The present book thus situates forensic linguistics not only in the context of linguistics but also in the wider context of and in relation to other forensic sciences. The aims of this book are to provide a comprehensive overview of the available expertise in forensic linguistics, while simultaneously serving as an introduction to commonly used methods and scientific procedures in the field. The present volume contains contributions by several recognized experts in linguistics who cover a variety of languages, legal systems, and countries – a fact that deserves to be highlighted given the otherwise still largely Anglo-centric perspective prevalent within the discipline. Chapters 1 to 5 reflect on the role of the linguist inside the courtroom and cover central issues involved in providing evidence in different legal systems. The remaining chapters discuss well-established areas of forensic linguistics, such as authorship analysis and speaker identification, as well as developing and emerging areas, such as online crimes. Each of these chapters offer a discussion of the theoretical backgrounds of the respective topics, as well as definitions, descriptions of state-of-the-art methodologies, case studies which exclusively discuss authentic data, as well as suggestions for further research. They also contain step-by-step guides to the application of the introduced methodologies.

In Chapter 2, Janet Ainsworth describes ethical issues from the perspective of a lawyer-linguist, i.e., she does not only highlight ethical issues faced by linguist expert witnesses but also relates them to ethical issues that can arise due to differences between linguists' and lawyers' ethical obligations. Andrew Hammel (Chapter 3) and Mercedes Fernández-López (Chapter 4) describe the role of the linguist expert witness in the common law and civil law traditions, respectively. While the former offers a historical overview of the common law tradition, the role of expert witnesses, and the emergence and development of both the Frye and Daubert standards, the latter hints at "substantial differences" in continental legal systems (p. 86) and draws attention to various appointment procedures in countries like Germany, France, Italy, Poland, and Spain. Subsequently, Magdalena Szczyrbak (Chapter 5) analyzes expert testimony in the courtroom with discourse analytical tools and thereby manages to provide insight into the tension between legal systems and sciences like linguistics with regard to subjectivity, speaker commitment, and stance.

In Chapter 6, Martina Nicklaus and Dieter Stein proceed to introduce lie detection tools and methods. Starting from a theoretical outline of definitions of lying and deception, the authors present a case study in which they draw attention to the need for more pragmatically oriented approaches to the detection of deception. The following three chapters are dedicated to authorship analysis. In Chapter 7, Eilika Fobbe provides an in-depth overview of largely qualitative authorship identification methods, of important concepts like style and idiolect, as well as feature selection. The case study of anonymous extortion letters does not only provide a guide to the application of qualitative authorship analysis but also explains the use of probability scales to express the results of an analysis. To complement this qualitative approach, Hans van Halteren (Chapter 8) proceeds to outline tools for automatic authorship analysis. He provides a historical overview of the development in this field before moving on to a discussion of linguistic features and their selection and use. A part of the presented case study sets out to test and evaluate an innovative deep learning approach for authorship analysis. In Chapter 9, Gea de Jong-Lendle provides a comprehensive overview of speaker identification by

describing different methodologies, tools, and approaches. A description of the historical development and salient cases in which phonetic evidence played a crucial role is complemented by a detailed description of key features of investigation and a presentation of the most important methods used in the field. Victoria Guillén-Nieto (Chapter 10) concludes the discussion of authorship analysis by outlining the field of plagiarism detection. Following a theoretical overview of central concepts in this field, such as plagiarism, copyright infringement, the moral and legal rights of authors, the chapter uses a case study to introduce the use of different tools for plagiarism detection in a step-by-step manner.

Subsequently, Monika Zaśko-Zielińska (Chapter 11) describes the linguistic analysis of suicide notes by first providing a discussion of suicide letters as a genre, as well as their super-, micro-, and move structures. The methodology is then applied to and illustrated with a case study of a genuine suicide letter and one whose authenticity is questioned. With the aim to "draw attention to the need to refine the concept of cybercrime and highlight its multifaceted character" (p. 433), Patrizia Anesa (Chapter 12) illustrates how linguistics can contribute to the analysis of cybercriminal activities exemplified by romance scams. The case study uses Applied Societal Discourse Analysis in order to reveal linguistic strategies employed by scammers. The final chapter by Julien Longhi (Chapter 13) addresses online terrorist threats and outlines a methodology that calls for a combination of "precise linguistic criteria and computer tools" (p. 442). The chapter concludes with a case study that is based on data provided by the French Gendarmerie and applies textometric and innovative deep learning methods.

Overall, the book is an important and timely contribution to the field of forensic linguistics that can serve as an introduction to the field of language as evidence. In particular, the book can be a useful resource for advanced students of linguistics, (early career) researchers, and anyone interested in deepening their knowledge in forensic linguistics. Students and researchers who have not had an opportunity to attend specialized classes in forensic linguistics might profit from this volume as they are not only introduced to a variety of methodologies but are also guided through individual analyses with the help of case studies. However, in order to make the most of the information provided in the book, some basic knowledge and understanding of theoretical concepts and methodologies are useful.

References

Coulthard, M., Johnson, A. and Wright, D. (2017). *An Introduction to Forensic Linguistics: Language in Evidence*. London and New York: Routledge.

Dern, C. (2009). *Autorenerkennung*. Stuttgart: Boorberg.

Eades, D. (2010). *Sociolinguistics and the Legal Process*. Bristol: Multilingual Matters.

Fobbe, E. (2011). *Forensische Linguistik. Eine Einführung*. Tübingen: Narr.

Kaplan, J. (2020). *Linguistics and Law*. London: Routledge.

H. Kniffka, Ed. (1990). *Texte zu Theorie und Praxis Forensischer Linguistik*. Tübingen: Niemeyer.

Love, H. (2002). *Attributing Authorship. An Introduction*. Cambridge: Cambridge University Press.

Solan, L. and Tiersma, P. (2005). *Speaking of Crime*. Chicago, IL: University of Chicago Press.

The Routledge Handbook of Forensic Linguistics

Book review by Georgina Heydon

RMIT University, Australia

https://doi.org/10.21747/21833745/lanlaw/9_1r2

The Routledge Handbook of Forensic Linguistics
Malcolm Coulthard, Alison May & Rui Sousa-Silva (Eds.) (2020)
London & New York: Routledge

The second edition of *The Routledge Handbook of Forensic Linguistics* (2020), edited by Malcolm Coulthard, Alison May and Rui Sousa-Silva, is a substantial contribution to our field, both in scope and significance and will likely prompt many readers to reflect upon the history of forensic linguistics and the place of this edition in that history. Readers will likely recall the popularity of the first edition in 2010, which provided broad coverage of the field with plentiful reference to general scholarship in linguistics, applied linguistics and the legal processes giving it a wide appeal. However, no sooner do we start to consider the history or development of this field than we run into a range of definitional problems: what exactly do we mean by 'forensic linguistics'? What is the distinction between forensic linguistics and language and the law? These questions have been addressed implicitly in the *Handbook* through its structure and scope; however, the contents of the *Handbook* also makes it clear that such vagueness has been an asset to our community because it has allowed us to define ourselves dynamically, adjusting the boundaries of what is considered 'language and the law' or 'forensic linguistics' when emerging technologies or a changing legal landscape introduce new possibilities for linguistic analysis. As a second edition, with 10 years since the first edition, we can expect a publication such as this to provide insights into the changes that have occurred in the field over that time. New readers might be less concerned about such evolutions, and for them, the contents will be of intrinsic value regardless of how they compare to the earlier edition. But for fans of the original Handbook who are seeking out this second edition, it is pertinent to ask: what is new here? Have the changes in the field, indicated by papers presented at conferences and colloquia for instance, been captured by the second edition?

The introduction to the second edition provides detail of the additional content which indicates that the changes reflect priorities of the editors: firstly, that seminal texts included in the original edition are retained and new texts do not replicate such chapters; secondly, that the scope of the book is widened to reflect a more geographically diverse field and thirdly, that advances in technology affecting forensic linguistic analysis are adequately covered in the new material. In addition, a number of chapters from the original text have been updated rather than being replaced with a wholly new chapter, on the basis that the original work has not diminished in value, but rather, there were further insights that could be added given developments in the field.

Returning to the question of scope, this volume divides the 43 chapters into three broad sections. Section I is titled *The language of the law and the legal process*. It includes four subsections: Legal language and legal meaning; Witnesses and suspects in interviews and investigations; Language in the courtroom; and Lay participants in the judicial process. Section II presents *The linguist as expert in the legal process*, thus the two 'sides' of forensic linguistics, as they are most commonly divided (research into legal language and forensic linguistic casework) are clearly represented in these two main sections. However, a third section is added to this volume which provides an opportunity to move beyond this binary segmentation of our field and is aptly titled *New directions*.

In the first subsection of Section I, the 'language of the law' is initially presented in its most referential form, with chapters addressing meaning in legal language in section 1.1. Although the first chapter by May, Holt, Al Saeed and Sani takes on the fundamentally interactional concept of questions, and gives them a pragmatic treatment, the next three chapters present analyses of meaning in written legal language, including complex documents (Stygall), US supreme court opinions (Finegan and Lee) and translation in the EU Court of Justice (McAuliffe). The final chapter in this section, 'The category of *ordinary meaning* and its role in legal interpretation', which is a new addition from Hutton, will be welcome to any forensic expert witness who has experienced the frustration of having one's complex semantic and syntactic analysis of a disputed text dismissed by a judge who instead reaches for the Shorter Oxford and proclaims that meaning can be found in a dictionary and is therefore not a field requiring expert testimony. The author's conclusion that a greater acceptance of linguistic expertise as part of the process of legal interpretation of language will resonate with many readers.

Moving through Section I, the book then turns to the *legal processes* described in the section's title, presenting first the analysis of language involved in the police and investigation process. This subsection opens with a chapter on the Miranda warning from Ainsworth, updating her seminal work on the topic which has been a mainstay for those of us working on police interview language (Ainsworth 1993). After chapters from Rock, Gaines and Lowry-Kinberg and Haworth on aspects of police interviewing, MacLeod's chapter introduces us to undercover work in online identity synthesis which calls to mind the work of Roger Shuy, examining sting operations in the US (see for instance Shuy 2017), although MacLeod brings us into the internet age with her analysis. The subsequent sections on legal processes then move naturally to the courtroom in two parts. Firstly section 1.3 presents a series of chapters on the interactions between legal professionals in the courtroom and includes some very familiar names alongside more recent contributors to this field: Drew and Ferraz de Almaida, Heffer, Chaemsaithong, Greenlee, and Matoesian and Gilbert. Section 1.4 provides a separate space for consid-

eration of interactions in courts and tribunals that involve lay participants, and has an emphasis on vulnerable or disadvantaged speakers navigating legal processes in complex circumstances. I must admit a degree of personal bias in highlighting the important new contribution from my fellow Australian, Peter Gray, whose final chapter in this section will introduce linguists around the world to the extraordinary cultural and linguistic phenomenon of Aboriginal land rights claims in remote northern Australia — the oldest continuous living culture and languages in the world meets the impenetrable jargon of a foreign, settler code of law. Gray extends the work on Aboriginal languages and the law by Eades (2004), Cooke (1996), Gibbons (2014) and others to become our on-the-ground correspondent from that extraordinary front line of linguistic exploration and cultural negotiation.

Section II, dealing with forensic linguistic analysis, expert opinions and case work, does not divide quite so easily into subsections as Section I. Although there are the expected contributions relating to the key categories of voice analysis, by Jessen, Watt and Brown, and Fraser (on transcription), authorship analysis, by Coulthard, McMenamin, Grant, Sousa-Silva and Woolls, and trademark disputes, by Butters and Dumas, the latter two chapters plus an opening chapter on expert linguistic witnesses and adversarial courts by Solan are included with the speech analysis chapters in section 2.1 *Expert and process*. As a subdivision of the expert witness field, this feels somewhat contrived, but there is no doubt as to the value of each of these chapters regardless of their position in the book's structure. Moreover, the expected 'expert witness' topics described above are augmented by chapters on interpreting and multilingualism by English, Hale, and Kredens, Monteoliva and Morris in section 2.2, *Multilingualism in legal contexts*. These analyses, of language proficiency, court interpreting and law enforcement interpreting respectively, represent a growing area of interest in academia more broadly, and forensic linguistics particularly, as language rights and migration combine to pressure institutions away from entrenched monolingualism. One of the most prominent examples of this change is in South Africa, and the omission of a South African perspective on this topic is surprising here, especially given the increased participation of African delegates in forensic linguistics conferences in recent years. This point will be addressed more fully below. Regardless, in Section II, readers will find a wealth of experience from a range of forensic contexts, and it is a reminder that a volume such as this provides a valuable resource for newcomers of first-hand knowledge about preparing and delivering courtroom testimony.

The final of the three main sections presents an array of new applications of forensic linguistic analysis and is, for this reviewer, perhaps the most exciting aspect of this volume. Established topics covered in the previous two sections are of course a valuable resource for everyone in the field, but for those of us who have been researching in forensic linguistics for several decades, Section III reminds us that there are still new avenues to explore and will doubtless inspire upcoming doctoral researchers. Importantly, the chapters in this section demonstrate the new collaborations that are being established between linguists and other academic disciplines that engage with language in a justice setting. While forensic linguists have a long history of engagement with judges and lawyers — though still not as much influence on legal training as many would like — and some of us have worked directly with police, there are few linguists engaging with literature on prisons, crisis negotiations or online abuse, all of which are topics of

intense interest in criminology and justice studies. In this section, we are presented with new work in each of these areas including Archer and Todd on police crisis negotiation, Gales on discourse of corrections and parole and Hardaker on internet trolling. Additionally, the chapter on investigative linguistics by Grieve and Woodfield provides an overview of the range of ways in which linguists can contribute to police investigations beyond the expert testimony described in Section II. The two chapters on corpus analysis for forensic and legal linguistics by Wright and Gries are another indication of the new directions in which linguistics is moving — big data and online language sources. These are exciting and much-needed developments and well-represented in these six chapters, with a seventh summary chapter from the editors providing their vision of a future state — 'a utopian future' — where forensic linguistic research findings contribute to the establishment of universal language rights in the administration of justice.

When delving into the fascinating micro-level analysis of language represented by many of the chapters in this volume, it can become a little hard to see the wood for the trees: has linguistics contributed to improvements in the administration of justice as a whole? The summary from the editors, Coulthard, May and Sousa-Silva, provides clear instances of progress, and this is important and helpful for readers. As they point out, we have indeed seen improvements in the way that courts understand linguistic evidence, more accessible and better trained interpreters, increased use of digital recording of interviews by police and recognition of the disenfranchisement of a great many people before the law simply because of its use of archaic and complex language. But of course, we still have a long way to go in many parts of the world, and in many parts of those justice systems where some progress has already been made. A third volume, already anticipated by the authors, needs to extend the consideration of forensic linguistic research further into the global South, to make greater efforts to engage with intersections of culture and justice such as in countries where customary law, religious law and non-adversarial systems create entirely different interactional contexts from evidence gathering to sentencing. Although a few chapters and authors in this volume bring diverse perspectives, the collection is still heavily dominated by research into the English and American legal systems.

Further diversity of the subject matter and broadening the scope of this Handbook will bring its own challenges though. The final remarks of the editors point to this challenge of diversity for a volume such as this. As a collection of individual research papers and case reports from 53 authors, the reader is presented with an encyclopaedic coverage of the field. Each chapter presents a unique research or analytical endeavour, with its own conceptual framing, methodology and analytical tools. Ideological positions are not prominently discussed and there are varying levels of critical self-reflection across the contributions. It is always challenging for such a collection to address the question of cohesion in these aspects of each contribution, but at a time when social science and humanities scholars are moving away from positivist, a-critical presentations of research findings, perhaps a future third edition of this book will take a stronger editorial stance on critical evaluation of the role of the linguist in institutions of justice that can be oppressive and structurally inequitable. Movements to decolonise justice or 'defund the police' have focused heavily on the exclusionary or prejudiced language of state institutions, and the academic engagement with these movements deserves greater attention from our field. This new edition of the Handbook has brought us new methods, new ap-

plications and new ways of thinking about established research in forensic linguistics. It is exciting to imagine where this will take us next.

References

Ainsworth, J. E. (1993). In a different register: The pragmatics of powerlessness in police interrogation. *The Yale Law Journal*, 103(2), 259–322.

Cooke, M. (1996). A different story: narrative versus' question and answer'in aboriginal evidence. *International Journal of Speech Language and the Law*, 3(2), 273–288.

Eades, D. (2004). Understanding aboriginal english in the legal system: A critical sociolinguistics approach. *Applied linguistics*, 25(4), 491–512.

Gibbons, J. P. (2014). *Language and the Law*. London & New York: Routledge.

Shuy, R. W. (2017). *Deceptive ambiguity by police and prosecutors*. Oxford: Oxford University Press.