# Language and Law

# Linguagem e Direito

VOLUME 11.2

ISSN 2183-3745

# Language and Law
# Linguagem e Direito

ISSN: 2183-3745 (online)
Special Issue, 2024



**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 11(2): https://doi.org/10.21747/21833745/lanlaw/11_2

## Contents / Índice

ARTICLES / ARTIGOS

**Introduction / Introdução**

**Impostors tending to the wild: Purposes of authorship analysis and a specific impostors method in the Karvounakis terrorism case** 1-23
*Isobelle Clarke & Tim Grant*

**Offence-Presumptive Terms: A Troubling Category in Linguistic Offensiveness** 24-54
*Chris Heffer*

**The effect of linguistic detail in police interview transcripts on perceptions of an interviewee** 55-78
*James Tompkinson & Kate Haworth*

**"Could you tell me what that means?" Ordinary and institutional vocabulary in police interviews with rape victims** 79-93
*Nicci MacLeod & Elizabeth Hay*

**The power of uptake: Responses to claims to power on anonymous online fora** 94-114
*Felicity Deamer, Michaela Wolf, Robert Davies & Tim Grant*

**Shifting linguistic identity performance and the acquisition of symbolic capital in an online white nationalist forum** 115-134
*Amy Booth*

**Ideologies of a serial bomber: A longitudinal linguistic appraisal analysis of the writings of the Unabomber** 135-159
*Madison Hunter*

**"This is an extortion note": A rhetorical move-analysis of discourse structure and genre in commercial extortion letters.** 160-178
*Marton Petyko, Lucia Busso, Sarah Atkins, Emily Chiang, Nabanita Basu & Tim Grant*

*Language and Law / Linguagem e Direito*, Vol. *11*(2), 2024
DOI: doi.org/10.21747/21833745/lanlaw/11_2

# Introduction

This collection brings together a diverse yet interconnected set of studies that illuminate the many ways language operates in contexts of crime, conflict, and control. Heffer's exploration of "offence-presumptive terms" urges us to think critically about how certain words can trigger offence by association rather than by inherent meaning, raising questions with implications for workplace disputes, legal judgments, and public discourse. From there, Tompkinson et al. and MacLeod and Hay each turn our attention to the police interview; the former investigating how subtle shifts in linguistic detail can shape perceptions of a suspect and the latter revealing the barriers that arise when victims and interviewers bring different vocabularies and conceptual frameworks to the table. Both studies remind us that in high-stakes encounters, word choice is never neutral, and linguistic mismatches can have profound consequences for the delivery of justice.

The remaining articles extend this theme of language as both a resource and a risk in varied arenas. Clarke and Grant's account of adapting authorship methods in the face of real-world investigative pressures. Deamer et al. and Booth explore the dynamics of linguistic power in online communities; whether in recognising those vulnerable to influence or tracing how individuals accrue symbolic capital over time, even in extremist spaces. Hunter's longitudinal analysis of the Unabomber's writings shows how stance and evaluation shift alongside psychological change, offering a bridge between forensic linguistics and cognitive psychology. Finally, Petykó et al. dissect the rhetorical architecture of commercial extortion letters, revealing how illicit genres are structured for both their intended victims and anticipated law-enforcement audiences.

Read together, this collection provides a rich mosaic of how linguistic evidence, whether in conversation, confession, or coercion, must be understood in its social, psychological, and institutional contexts.

**Guest Editors**
Felicity Deamer
Malcolm Coulthard

# Introdução

Este volume reúne um conjunto diversificado de trabalhos inter-relacionados que ilustram as diversas formas como a linguagem opera em contextos de crime, conflito e controlo. A análise de Heffer sobre os "termos presuntivos de ofensa" convida-nos a refletir criticamente sobre o modo como certas palavras podem desencadear fenómenos de ofensa por associação, mais do que pelo seu significado inerente, suscitando questões que têm implicações em disputas laborais, em decisões judiciais e no discurso público. De seguida, Tompkinson et al. e MacLeod e Hay centram a atenção nas entrevistas policiais; este último trabalho revela os obstáculos que emergem quando vítimas e entrevistadores transportam vocabulários e quadros conceptuais distintos para o mesmo espaço comunicativo, enquanto o primeiro trabalho investiga como variações subtis nos pormenores linguísticos contribuem para moldar perceções sobre um suspeito. Os dois estudos recordam-nos que, em interações de risco elevado, a escolha de palavras nunca é neutra e os desalinhamentos linguísticos podem ter consequências profundas na administração da Justiça.

Os artigos que se seguem ampliam a temática da linguagem como simultaneamente recurso e risco em diferentes domínios. Clarke e Grant descrevem a adaptação de métodos de atribuição de autoria perante as pressões exercidas por casos de investigações reais. Deamer et al. e Booth exploram as dinâmicas de poder linguístico em comunidades online, seja em situações de identificação de pessoas vulneráveis a contextos de influência, seja através do estudo do modo como os indivíduos acumulam capital simbólico ao longo do tempo, mesmo em espaços extremistas. A análise longitudinal de Hunter sobre os textos do Unabomber mostra como o posicionamento e a avaliatibilidade se transformam em paralelo com mudanças psicológicas, estabelecendo uma ponte entre a linguística forense e a psicologia cognitiva. Por fim, Petykó et al. dissecam a arquitetura retórica das cartas de extorsão comerciais, mostrando como são estruturados os géneros ilícitos, quer para as vítimas visadas, quer para os públicos esperados das forças policiais.

No seu conjunto, estes estudos oferecem um vasto conjunto de artigos focados na forma como a prova linguística — seja em situações de conversa, seja em situações de confissão ou coerção — deve ser compreendida nos seus contextos sociais, psicológicos e institucionais.

**Editores Convidados**
Felicity Deamer
Malcolm Coulthard

*Language and Law / Linguagem e Direito*, *Vol. 11*(2), 2024, pp. 1-23
DOI: doi.org/10.21747/21833745/lanlaw/11_2a1

# Impostors tending towards the wild
# Purposes of authorship analysis and a specific impostors method in the Karvounakis terrorism case

Isobelle Clarke [1] and Tim Grant [2]

[1] *Linguistics and English Language, Lancaster University*, i.clarke@lancaster.ac.uk
[2] *Aston Institute for Forensic Linguistics*, t.d.grant@aston.ac.uk

***Abstract***
*Authorship analysis can be used at different stages of a criminal case, including the initial investigation (i.e. to narrow down the search for a suspect), to support a search warrant application, and to provide evidence for a criminal prosecution. Each stage has its own complexities and pressures, requiring some flexibility in the type of method applied. Indeed, it is not unusual in such cases for the conditions to be far from ideal, adding to the complexities of the particular stage of the case. For instance, cases often involve texts that are very short and comparison data that is not within the same genre as the questioned text. In some cases, there are strict security constraints, meaning that some tools and software are prohibited and that the forensic linguist must work alone. HMA v Nikolaos Karvounakis is a domestic terrorism case, which required authorship analyses at all three stages. Each stage had a particular purpose and specific security constraints and complexities, which influenced the kinds of authorship analyses employed. The final stage led to the development of the Specific Impostors method, which is variation of the General Impostors method. This article describes the case history, the methods employed and the outcome of their application to the data and concludes with reflections on the case and the methodology.*
***Keywords:*** *Specific Impostor's Method, n-grams, comparative authorship analysis, domestic terrorism.*

***Resumo***
*A análise de autoria pode ser utilizada em diferentes fases de um processo penal, incluindo a investigação inicial (ou seja, para restringir a procura de um suspeito), para apoiar um pedido de mandado de busca e para fornecer provas para uma ação penal.*

*Cada fase tem as suas próprias complexidades e pressões, exigindo alguma flexibilidade no tipo de método aplicado. De facto, não é raro que, nestes casos, as condições estejam longe de ser ideais, o que aumenta a complexidade da fase específica do processo. Por exemplo, trata-se frequentemente de textos muito curtos e de dados de comparação que não pertencem ao mesmo género do texto questionado. Nalguns casos, existem restrições rigorosas em matéria de segurança, o que significa que algumas ferramentas e programas informáticos são proibidos e que o linguista forense tem de trabalhar sozinho. HMA v Nikolaos Karvounakis é um caso de terrorismo doméstico, que exigiu análises de autoria nas três fases. Cada fase tinha um objetivo particular e restrições e complexidades de segurança específicas, que influenciaram os tipos de análises de autoria utilizados. A fase final levou ao desenvolvimento do método dos Impostores Específicos, que é uma variação do método dos Impostores Gerais. Este artigo descreve a história do caso, os métodos empregues e o resultado da sua aplicação aos dados e conclui com reflexões sobre o caso e a metodologia.*
***Palavras-chave:*** *Método do Impostor Específico, n-gramas, análise comparativa de autoria, terrorismo doméstico.*

## 1. Introduction: Case history

A little before 8pm in the evening of 11th January 2018, a suspicious cardboard box was discovered by a park ranger on a sheltered seat in Princess Street Gardens, a public park in Central Edinburgh in Scotland. A controlled explosion was carried out and initially it was believed that the device was relatively harmless, but later inspection by forensic experts showed that it was a viable bomb and had the capacity to cause serious harm.

There were no clear leads to the identity or the motives of the person who had planted the bomb, until a month later when a journalist received an anonymous email which linked to Maldición Eco-Extremista, a blog from the Mexican eco-anarchist group called *Individualistas Tendiendo a lo Salvaje* (hereafter the ITS). In English their name translates to *Individualists Tending to the Wild*, and their central belief is that the devastation of world ecosystems was linked to scientific and technological progress to an extent that justifies direct and sometimes terrorist actions. On 6th February 2018 the blog published two new posts, both of which contained the same content but one was in English and the other in Spanish. Specifically, the posts contained a clear claim to responsibility for the bomb and a picture of the interior of the bomb (Figure 1).

The blog posts were sent to TG with a generalised request for commentary and a particular focus on whether the author could be identified from the writings. Initial analysis was constrained – as this analysis was part of a live terrorism investigation the text and the focus of the inquiry on the text was classified information – this meant that TG had to work alone and was largely prevented from using online analysis tools. However, through a manual analysis based on back translation of short phrases and idioms using Google translate, TG established that the English language text seemed to be the original and the Spanish language text was based on the English version. Further to this TG concluded that it was unlikely that the English text was written by a native

English speaker, for example, spelling errors that were unlikely to be made by a native speaker with low levels of competence (e.g. "nichilistically"). He examined whether it was likely that the English was influenced by Spanish and concluded that this was highly unlikely, suggesting that the other language influence was from some language other than Spanish. Finally, he concluded that the English version had a very elaborate and unusual style that was distinctive, and thus likely to be amenable to comparative authorship analysis should further texts be produced for comparison. This rather limited analysis was sent to the Scottish police investigators after which nothing much happened for two years.

En extremo escepticismo y experimentación misántropa, más allá de cualquier noción humana, reclamo nichilísticamente los siguientes ataques:

**-El incendio provocado de 2 minibuses que transportan personas mayores.**

¿Por qué? ¿Por qué no le preguntas a los chicos de los libros que lees que te digan por qué? Oh, mierda! Están muertos? ¡Te diré por qué entonces! ¡Porque odio a las personas mayores! Jajajajajajajajaja-jajajajajajajajajajajajajajajaja

**-Un paquete-bomba abandonado completa e indiscriminadamente en una ubicación central de manera selectiva.**


¿Por qué no pienso en la gente "inocente" como uno lo podría pensar? ... Respondo con una pregunta... ¿Creyó la pucha de mi donante de nacimiento cuando fue cogido para ser fertilizado con semen microscópico que crea los vasos que odio? ¿Alguien me pidió que naciera? ¿Alguien sabía en qué me convertiría? ¿Sabes que algunos ven la conciencia como una maldición? Chinguense, patéticos, ¡entonces no saben nada! No busco la justificación de la existencia, tampoco busco a alguien a quien culpar. Busco la violación amoral de la existencia a través de la inyección de vida que pasa desde la Puerta de la Muerte. Odio antihumano es la sangre de mi vida, transformando mi vasija en la Bestia.

In extreme misanthropic skepticism and experimentation, beyond any human notion, I claim nichilistically the following attacks:

**-The arson of 2 mini buses transporting elder people.**

Why? Why don't you ask the guys from the books you read to tell you why? Oh shit! They're dead? I'll tell you why then! Because I hate old people! Hahahahahahahahahah-hahahahahahahhahahahahhahahahahahahaha!!!

**-A package bomb left totally indiscriminately at a central location selectively.**


Why do I not think of the "innocent" people one might think... I answer with a question... Did my birth giver's pussy think when it was fucked to be fertilized with microscopic semen that creates the vessels that I hate? Did anybody ask me to be born? Did anyone know what I would become? Do you know that some see consciousness as a curse? Fuck you, pathetic pricks, you don't know shit then! I do not seek justification for existence, neither do I seek someone to blame. I seek the amoral rape of existence through the injection of life passing from the Death Gate. Anti-human odium is my life's blood, transforming my vessel into the Beast.

**Figure 1. Excerpts of the blog posts from the ITS website and the image of the bomb.**

In the spring of 2020 Police Scotland got back in touch with TG saying that through Europol and the Greek Police a suspect had been identified. This was Nikolaos Karvounakis, a Greek national, living in Edinburgh. Karvounakis was linked to a Greek eco-anarchist group who were in turn linked to ITS. Police Scotland had also identified that Karvounakis was a member of a rock band called *Face of Another*, and that this

band had a website which contained lyrics and other texts supposedly written by Karvounakis. TG was asked to carry out a comparative authorship analysis to give evidence to ask a court to grant a warrant to search Karvounakis' properties, and to seize his computers and other devices. TG was able to carry out an analysis which showed some limited evidence that linked Karvounakis' writings from the *Face of Another* website with the Maldición Eco-Extremista blogs. This along with other evidence from Police Scotland became grounds for a successful application for a search warrant.

Once the warrant had been served this produced a whole lot more text from Karvounakis' computers, including blog posts known to be from the ITS website including the *Paroxysm of Chaos* 'zines' described below ('zine' is an abbreviated term for 'electronic magazine'). Unfortunately, however, the seized materials did not include a copy of, nor was there a link to the original claim of responsibility for planting the bomb from Karvounakis' devices. This meant that the investigation required one final forensic authorship analysis to be carried out. This time the comparison would be between, on the one hand, the set of blog posts seized from Karvounakis' computers, which the police believed they could reasonably show that he had written, and on the other hand, the original 'claim of responsibility' text. The purpose would be to determine whether or not these texts had a common author and therefore provide evidence that could contribute to the criminal case against Karvounakis.

By this time TG had been working with the police team of investigators for more than three years. He was clearly part of the investigation (albeit on the edges of it) and psychologically was not acting as a wholly independent expert. To avoid bias he contacted a second forensic linguist (IC) and provided her with the documents in the case with no story about what had happened, and he did so in a predetermined order following protocols set out in Grant (2022). In what follows we present IC's authorship analysis in which she identifies consistent and distinctive features of Karvounakis' known language. She also shows that these features occur in the 'claim of responsibility' text and this provides some evidence that Karvounakis is a likely author of that claim. Reinforcing the importance of these features in the analysis, IC goes on to show that they hardly occur in the other ITS writers' blogs, and nor do they occur in a comparison with other Greek speakers writing in English or in more general corpora of English.

Through this case history, we can see that the authorship analyses carried out were commissioned with different purposes and constraints and needed to be responsive to these. The next section of this article examines in more detail the variety of purposes across different types of forensic authorship analysis and considers how the design of the analysis can and should (or sometimes should not) bend to these purposes. Following this we discuss the variety of methods applied at each stage and how they serve the specifics of each of these different purposes and of the specifics of the case. Finally, in our conclusion, we reflect on how the required flexibility of method in the light of these specifics can be accommodated even where we also prioritise the push to validation of methods and acting ethically as forensic linguistic experts in investigative and evidential work.

## 2. Varied Purposes of Authorship Analysis

This case involved progressive authorship analyses, first with an investigative purpose, next with the purpose of providing evidence to support an application for a search warrant, and finally with the purpose of providing evidence which would form part of a criminal prosecution. In each of these three stages, the function of the analysis affected its design and outcome.

### 2.1. Stage 1 – Authorship analysis for investigation

Investigative forensic linguistics can involve analysis where the intended purpose is not to directly provide evidence for a court, but rather to help identify a suspect (see Grant & MacLeod, 2020, Chapter 6 for a discussion). Such an analysis is sometimes referred to as simply *investigative analysis* or *intelligence analysis* (as contrasted with evidential analysis). Depending on the Criminal Practice Directions (or equivalent) that apply in different jurisdictions, such analyses may have to be disclosed to defence teams (either as a duty of the prosecution as occurs for UK jurisdictions, or as a response to a subpoena in United States jurisdictions), or a decision might be made by prosecution teams, that an investigative analysis should actually be used as evidence in court. Thus, at the point the expert is approached to work on a case the final value or nature of their contribution will be unknown even when the primary purpose of commissioning the analysis is investigative. This context, alongside the knowledge that disclosure may be required in the future, can constrain the techniques used and the nature of the conclusions that can be expressed.

As described above, the ITS investigation was a live terrorism inquiry and as such came with security constraints. Electronic files of the posts and of the email to the journalist were sent to TG by police courier on encrypted drives, and the investigation was 'indoctrinated' – a term used by UK policing to indicate that materials and any progress in the investigation were classified. TG was required to not seek assistance in the investigation from any colleagues, and to not use online tools to analyse the texts. He also does not read or speak Spanish. These constraints clearly hampered the nature of the analysis that could be undertaken. Colleagues at the Aston Institute of Forensic Linguistics not only include Spanish speakers (including a Mexican Spanish speaker), but also these colleagues carry out research projects directed at the detection of different dialects of Spanish influence on English language texts (e.g. Batel, Abrams, & Pęzik, 2022). TG was able to review the (then unpublished) papers associated with these projects and manually examine the English language texts for artefacts of Spanish influence. Although there were clear indications in the syntax and word choices that the writer was a non-native writer of English, no specifically Spanish interlingual features were found. He was unable to determine what other language(s) might be influencing the written English. These conclusions, fed back to the investigators by way of a brief report, indicated the uncertainty associated with the conclusions. This report was included in the 'unused materials' disclosed by the prosecution and could therefore have been used by a defence team at trial. The risk to the prosecution would have occurred if the defendant had turned out to be a Spanish speaker, in which case the linguistics report could have been of value to the defence.

The limited conclusions of the report were however presumably of little value to the ongoing investigation in identifying a suspect as nothing occurred for a further two years.

### 2.2. Stage 2 – Authorship analysis to a standard of reasonable suspicion

The purpose of evidential analysis is not always to provide evidence beyond a reasonable doubt to a criminal court. Authorship analysis has been used in a variety of civil contexts, in employment and immigration hearings and in other quasi-judicial contexts. One such context is providing evidence to be taken before a judge in an application for a search warrant. In Scotland (as in England and Wales), the legal standard that must be reached to grant a warrant is "Reasonable Suspicion". This more or less equates to the United States standard of "Probable Cause" and is clearly a long way less certain than criminal standards of proof such as "Being Sure" or "Beyond a Reasonable Doubt". This lower burden of proof may raise questions for the analyst in terms of method. There is a difference between, on the one hand, carrying out an analysis with a well-attested method, and the results from the analysis reaching a well-understood lower level of certainty, and on the other hand, carrying out an analysis which is less well-validated, or which has unavoidable weaknesses such that there is less confidence in the outcome. Consideration of the possible problems with the nature of the analysis were an issue with this stage of the case. In comparative authorship analyses there is now increasing evidence that cross-genre authorship analysis is difficult to do well. In recent computational authorship analysis competitions, such as PAN (https://pan.webis.de), the best computational systems that can typically score over 93% accuracy in within-genre verification tasks, drop to barely better than chance in cross-genre attribution (Stamatatos et al., 2022).

In carrying out the analysis to provide evidence to support a warrant application, the same security constraints were applied as in the Stage 1 analysis. This is because, at this stage of the case, the suspect was unaware of the investigation, and because the line of investigation could have led to the elimination of the particular individual from the inquiry. The analysis at this stage was clearly a cross-genre analysis. On the one hand there were the original ideological posts (and the claim of responsibility) from the Maldición Eco-Extremista blog, and on the other hand there were song lyrics from the *Face of Another* website and other brief texts describing the band's music and ideologies. TG used an analysis based on a word n-gram approach. N-gram measures are frequently used in authorship studies and simply take each sequence of units (characters or words) that arise from a text. So, for instance, the phrase 'The cat sat on the mat' has the following 3-grams: "the_cat_sat", "cat_sat_on", "sat_on_the", "on_the_mat". N-grams can also be character sequences (e.g. "t_h_e, h_e_c, e_c_a, etc), but for this analysis, and due to the need for interpretability in court, we only considered word n-grams. Research into n-grams generally suggests that n-grams can be found that are characteristic of specific authors (whether or not these carry interpretable linguistic information). For example, Grieve et al. (2019) point to character n-grams as being the most useful features for discriminating authors in their problem, and yet only provide word n-grams for reasons of interpretability. Importantly, it has been repeatedly shown that collections of n-grams can be used to discriminate between authors (e.g. Grieve, 2007; Grieve et al., 2019; Stamatatos, 2009). Thus, one author may fall into a habitual use of

certain n-grams, whereas a second author never uses those particular strings of words. The purpose of such an authorship analysis is to determine from a text whether strings of n-grams can be derived that demonstrate a consistent within-author pattern of use, which is also distinctive in comparison to other potential authors of the text.

The analysis applied at this stage was computer assisted – TG used a simple R script (R Core Team, 2023) to list the words, bi-grams (two-word strings) and tri-grams (three-word strings) which were present in each set of texts. This was essentially a computer-assisted stylistic analysis which showed some limited evidence that some rare words, rare two-word strings, and rare three-word strings occurred in both the texts known to have been authored by Karvounakis, from the *Face of Another* website and the anonymous texts from the ITS website. TG was also able to show some distinctiveness for these n-grams in that they were either unique or very low frequency in the rest of the texts contained in the blog posts.

In his report, TG made clear that this was a cross-genre analysis and the difficulties with that, and the limitations of the method meant that levels of certainty were low, but also that the finding of individual rare words, bi-grams and tri-grams across the different set of texts did, in his opinion, carry some evidential weight. This evidence, along with other evidence that Police Scotland put before the court[1], gave rise to the judgement that there was sufficient weight of evidence to amount to reasonable suspicion and the warrant was granted.

### 2.3. Stage 3 – Authorship analysis to a criminal standard

The granting of the warrant literally opened the door to a wealth of further evidence. Karvouanakis' premises were searched, and his devices seized and searched. These searches revealed a number of documents that had appeared on the *Maldición Eco-Extremista* blog and provided strong evidence that Karvounakis wrote under the pen name Archegonos. The newly available data and the full analysis is described below. Two principal issues arose as the analysis switched to providing evidence for criminal prosecution.

The first change was, of course, in the standard of proof that the new analysis would have to answer to. Whilst this does not apply to any individual item of evidence or fact in the case – the objective of a prosecution in Scotland is to convince a jury "looking at the evidence as a whole" that they are "satisfied of the guilt of the accused beyond reasonable doubt" (Judicial Institute for Scotland, 2023), it does require a high degree of certainty about the validity of the analysis and its conclusions. How this was achieved is described below.

Further to the increase in the burden of proof, there also arose the issue of potential bias in the analysis and how to mitigate this. As noted above TG had at this stage of the investigation been working on the case for over three years and considered himself more part of the investigation than an independent expert. As in England and Wales, the duties of an expert witness in Scotland are clear, they must understand that their "principal duty is to the court, not to the prosecution or defence. Evidence must come

[1]This comprised evidence from a forensic meteorologist who opined that a cloud formation in a picture from the *Face of Another* website matched a picture of a cloud posted on the Maldición Eco-Extremista blog site.

from the independent thinking of the expert witness. It must not be influenced by those who instructed them. Expert evidence is objective and unbiased" (Crown Office and Procurator Fiscal, 2024). Recognising that his 'independent thinking' might well be compromised and cognisant of the research base on cognitive biases, and specifically confirmation bias (e.g. Edmond et al., 2017; Found, 2015; Stoel, Berger, Kerkhoff, Mattijssen, & Dror, 2015), TG decided to recruit an independent forensic linguist (IC) to produce the evidential analysis which might be used at trial. By this time, Karvounakis had been interviewed and although the police were still concerned with investigative security, permission was obtained to proceed.

TG provided IC with the documents in the case with no story about what had happened, and (following Grant, 2022; Grant & Grieve, 2022 he did so in a predetermined order. First, he provided only the ITS blog posts that Karvounakis was now known to have written. On the basis of these texts, he asked IC to find language features which were common between texts that he had written and that were distinctive from other writers on the ITS blog site. Like the Stage 2 analysis, IC decided to focus on word n-grams in order to systematically account for a larger pool of features. He then informed IC that the writer's English might be influenced by Greek, Karvounakis' native tongue, and encouraged IC to check whether the identified distinctive language features of Karvounakis could be attributed to their language background. Finally, and only when this analysis was complete did TG pass to IC the claim of responsibility blog post. These measures were an attempt to create the conditions for as rigorous analysis as possible and one in which bias was mitigated wherever possible.

Throughout the case history, we thus see the tasks, and the pressures associated with the tasks, moving and changing: in Stage 1, the questions were wide open, but the methods maximally constrained by the security concerns. In Stage 2, those security issues were still a concern, and the authorship analysis was challenged by the small quantity of short texts involved and the cross-genre nature of the comparison. This was possibly mitigated by the lower burden that was at issue for the judge in granting the warrant. For Stage 3, the security issues were considerably mitigated and there was a wealth of within-genre material available to create good comparisons. The challenge was, however, to mitigate bias and provide the strongest possible evidence for criminal prosecution. The method applied to provide this evidence is a variation of the General Impostors method. Thus, before we present our method, we describe the General Impostors approach to provide some context.

## 3. The (General) Impostors Method

The Impostors Method (IM) was originally proposed by Koppel and Winter (2014) to determine whether two texts Q and K are written by the same author, where the Q document is a document of questioned authorship and the K document is a document of known authorship. The IM assesses whether Q is more similar to K than to each one in a set of impostor texts. The comparison is based on a large number of features which are sampled randomly to create 100 separate feature sets. The overall similarity is then expressed as the percentage of those feature sets where Q is more similar to K than to any of the impostors.

Koppel and Winter (2014) suggest that the selection of impostor texts and how many impostors is key in this process. They argue that too few impostors or unconvincing impostors could lead to too many false positives (in Koppel and Winter's terminology[2]), while too many impostors or impostors that are in the same genre as the disputed text could lead to many false negatives. They thus proposed seeking "an optimal combination of impostor quality, impostor quantity, and score threshold" (Koppel & Winter, 2014, p. 182). They suggest generating three groups of impostors for the disputed text:

1. Fixed set – a fixed set of impostor texts that have no relation to the disputed text or the suspected author.
2. On-the-fly – a set of plausible impostors, which are generated by selecting small sets of random medium-frequency words from the known author's text and the disputed text and incorporating them into a search engine query. The top results of each query are aggregated.
3. Shared genre – a set of impostors that are in the same genre as both the disputed and the known author text.

The use of the impostor method is akin to a police lineup where a witness is asked whether a previously seen suspect is present. In visual parades (Darling, Valentine, & Memon, 2008) or indeed in voice lineups (Pautz et al., 2023) the selection of impostors (or 'foils') has been shown to be a crucial aspect of design. In language analysis this often overlooked by advocates of IM. For example, a face lineup where the previously seen suspect was described as having darker skin, and then at the identity parade only a single contender in the lineup was dark-skinned, would clearly create unfair identification. In the linguistic case therefore, the impostors in the IM need to be selected or constructed to be close enough to the Q text and the K text to create a fair competition for closest match. For Koppel and Winter's (2014) three groups of impostors this does not seem to be well accounted for. In the forensic context, it is important that the task is not too easy, as this might be challenged by an opposing side (e.g. in cross-examination). However, the conceptualisation of closeness or distance in style between texts is under-explored, in terms of the full range of linguistic variation and how this is created (see e.g. Grant & MacLeod, 2020). We address this through developing our specific impostors method (below).

With regard to features used in the IM, previous studies have used *inter alia*: the most frequent words (MFW), word and/or character n-grams, function words and part-of-speech tags (e.g. Andrijanić & Bąkowski, 2021; Khonji & Iraqi, 2014). The technique then randomly picks a percentage of the full feature set and then calculates whether Q, based on this set of features, is closer to the suspect's known writings or to the individual impostors using a distance measure. This is repeated for 100 different feature sets.

The IM has been hugely successful with variations of the technique winning first places in PAN 2013 and 2014 (Potha & Stamatatos, 2017). One of the winners was Seidman (2013) who sought to generalise the IM to deal with cases where the task is to assess whether a document is an outlier in a set of known documents (i.e. where there are multiple documents by the candidate author). They tested various methods and

[2]Koppel and Winter (2014) are confusing in their use of the terminology of false negatives and false positives in this context. Too similar impostors could lead to false positives, whilst too dissimilar impostors offers an easy comparison meaning that the known texts of a suspect will be deemed most similar, which could overemphasise the similarity between the Q and K.

found that the best performing approach involved running the IM on all pairs consisting of the questioned document and a single known document and then aggregating the results.

Attempting to address the issue raised above, Potha and Stamatatos (2017) sought to improve Seidman's (2013) General Impostors Method (GIM) by incorporating impostors that were close to the suspect. They acknowledge that in the traditional police lineup, individuals are not selected randomly, but rather they are selected to have characteristics similar to their suspect. They achieve this by submitting queries into Bing search engine using salient words from the set of known documents and download the first results. In addition, they noted how the GIM disregarded cases where at least one impostor text is found more similar to the disputed text in comparison to the candidate author. As a result, they propose ranking the similarities in decreasing order and considering the position of the candidate author's texts. These modifications were found to be competitive and sometimes better than other GIM variations.

As noted, we used a variation of the GIM for Stage 3. We call this the Specific Impostors Method, which we present below.

## 4. Developing a Specific Impostors Method

This section describes the analysis conducted in Stage 3. This analysis was guided by the twin competing hypotheses:

(H1) Karvounakis is the author of the disputed text.

(H2) Someone other than Karvounakis is the author of the disputed text.

Initially, IC was provided with two zines that were posted to the ITS blog: Paroxysm of Chaos I (PoC1) and Paroxysm of Chaos II (PoC2). These zines were comprised of texts written by Karvounakis (under the pen name Archegonos) as well as writings from other individuals. The texts can be characterised as anarchist texts often expressing anti-civilisation attitudes. IC was tasked with identifying a consistent within-author pattern across the documents written by the suspect, and to identify if this within-author pattern is distinctive from the writings of the other individuals. Then, IC was provided with the disputed text for comparison.

### 4.1. Data Description

#### 4.1.1. Comparison texts

#### 4.1.2. Paroxysm of Chaos I

PoC1 is a zine comprised of 17 sections (see Table 1). The zine was compiled under the pen name Archegonos and includes sections written by Archegonos as well as writings from other individuals (some of whom are known to be alive and some of whom are known to be dead). The first chapter is written by 'Archegonos' – our suspect's pseudonym - and is called 'Prelude to the abyss', which serves as the introduction to the publication. In a side note after the introduction it says: "[e]verything included in this zine is something that I see either as theoretical or practical gain and definitely not representative of mine". Files were created based on the presence of an authorial sign-

off. The majority of sections have a sign-off, except for sections 16 and 17. Section 17 is comprised of multiple poems from two different authors and is signed off after each poem within the section (3 poems were from Archegonos and 1 was from Bruno Filippi).

Section 16, however, is unique in its composition of numerous claims of responsibility for illegal acts, yet the sign-offs are often missing or unclear. For instance, in one part of section 16, it describes an incident where ten trucks owned by a prison officer were set on fire in France, but this is not claimed under any specific pseudonym. For some of these subsections, the content makes it seem as if they were written by the main editor of the zine, Archegonos, but this could not be depended upon for the analysis. Because of the missing sign-offs these were grouped into one file for this section and the authors were deemed to be 'unknown' (see Table 1).

One section is signed off as a "Collaboration of anarchists", whilst another section is signed off by 'Fallon and Amelie'. These explicit acknowledgements that some sections are multi-authored, gives at least some weight to the assumption that other signed sections are written by individual authors.

**Table 1. The Paroxysm of Chaos I**

| | Title of section | No. word tokens | Signed Author(s) | Filename |
|---|---|---|---|---|
| 1 | Prelude to the abyss. | 1056 words | Archegonos | Archegonos_1 |
| 2 | Total liberation as an egoist and iconoclastic consideration. | 8364 words | Archegonos | Archegonos_2 |
| 3 | On radical moralism and wildness. | 1309 words | Anonymous | Anonymous_1 |
| 4 | Against the language of militancy. | 400 words | Wolfi Landstreicher | Wolfi-Landstreicher_1 |
| 5 | Concerning "good" and "evil" | 578 words | Émile Armand | Emile-Armand_1 |
| 6 | Sacrilegious laughter. | 962 words | Erinne Vivani | Erinne-Vivani_1 |
| 7 | Into the abyss-chaos. | 240 words | L | L_1 |
| 8 | Veganism from a nihilist and anti-civilization perspective. | 1244 words | Archegonos | Archegonos_3 |
| 9 | A critique, not a program: For a non primitivist anti-civilization critique. | 2359 words | Wolfi Landstreicher | Wolfi-Landstreicher_2 |
| 10 | Mexico : Open letter of Amelie Pelletier and Fallon Poisson. | 482 words | Fallon and Amelie | Fallon-Amelie_1 |

| | | | | |
|---|---|---|---|---|
| 11 | Mexico : Anarchist comrade Carlos Lopez 'Chivo"s reply to International Revolutionary Left "Buenaventura Durruti". | 897 words | Carlos Lopez "Chivo" | Carlos-Lopez_1 |
| 12 | De Profundis Clamavi | 938 words | RadioAzione [Croatia] | RadioAzione_1 |
| 13 | War on the state: The subject of desire. | 1746 words | Saul Newman | Saul-Newman_1 |
| 14 | Tearing down the prison of civilization. | 2773 words | Collaboration of anarchists – Memento Mori (Translated from Greek to English by Archegonos) | C-o-A_1 |
| 15 | Some thoughts on the matter of responsibility claims. | 299 words | Archegonos | Archegonos_4 |
| 16 | Responsibility claims – Moments of war. | 6353 words | Unknown (multiple authors) | Unknown_1 |
| 17 | Individualist poetry. | 801 words | Archegonos | Archegonos_5 |
| | | 199 words | Archegonos | Archegonos_6 |
| | | 255 words | Archegonos | Archegonos_7 |
| | | 580 words | Bruno Filippi | Bruno-Filippi_1 |

### 4.1.3. Paroxysm of Chaos II

PoC2 is the sequel to PoC1 and is made up of 13 sections, which were similarly compiled by Archegonos with sections written by Archegonos or other authors (see Table 2). Like PoC1, texts were created based on an authorial sign-off (as opposed to sections). Most texts had a clear sign-off. Exceptions to this were sections 8 and 10. Similar to PoC1, there was a section compiling numerous claims of responsibility (section 10). These were grouped together in a file with the author(s) specified as unknown (although we suspected parts of this to be written by Archegonos). Section 8 did not have a sign-off. After section 8, there was a note from the publisher (possibly Archegonos), specifying that this text was published "because it makes some points I agree with". We can therefore assume that it was likely not written by Archegonos. Section 12 is the introduction of a group/blog/project called 'Antisocial Evolution'. No other texts or files can be found by this group to connect it to a specific author. Although it seemed that this text in section 12 was likely written by Archegonos, because it was not specifically signed off by this pen name, we did not include it in the known writing samples of Archegonos.

**Table 2. The Paroxysm of Chaos II**

| | Title of section | No. of word tokens | Signed Author(s) | Filename |
|---|---|---|---|---|
| 1 | Prologue – Desecration of universality and the theatrical significance of civilisation's performances | 2720 words | Archegonos | Archegonos_8[3] |
| 2 | Chaotic iconoclasm and incineration of idols (Anticivilisation of paroxysm part 1) | 1746 words | Archegonos | |
| 3 | The coronation of moralism upon the throne of the ghost of Nature (Anticivilisation paroxysm part 2) | 6961 words | Archegonos | |
| 4 | Breaking the windows that sell the products of civilisation (Anticivilisation paroxysm part 3) | 675 words | Archegonos | |
| 5 | My misanthropism (Anticivilisation paroxysm part 4) | 1507 words | Archegonos | |
| 6 | Wildness in the city | 2265 words | Chris Kortright | Chris-Kortright_1 |
| 7 | A Life | 1616 words | Renzo Novatore | Renzo-Novatore_1 |
| 8 | Insurrection or revolution? | 1792 words | Unknown | Unknown_2 |
| 9 | The unique one meets the overhuman II | 1008 words | John F Welsh | John-F-Welsh_1 |
| 10 | Moments of war | 3668 words | Unknown (multiple authors) | Unknown_3 |
| 11 | Poetry of the void | 590 words | Archegonos | Archegonos_9 |
| 12 | Introduction to the project Antisocial Evolution | 2782 words | Antisocial Evolution | Unknown_4 |
| 13 | On being lyrical | 1072 words | E. M. Cioran | E-M-Cioran_1 |

[3] The Archegonos sign-off appeared only at the end of part 4 and not for each section. As a result, these were combined into one file.

Because the known writing samples of Archegonos were in English, the English version of the disputed text was used for comparison. Archegonos' known writing samples amounted to 25,173 words across nine sections of the two Paroxysm of Chaos texts (see Table 2). The only other author to have more than one text was Wolfi Landstreicher, with a total of two texts comprising 2759 words. All other authors only had one text with text lengths ranging from 240 to 6353 words.

### 4.1.4. The Disputed Text – the claim of responsibility

The claim of responsibility was posted to the ITS blog on 6th February 2018 in English and in Spanish. The English version is 1,864-word tokens. Each version of the text in-

cludes the same 5 images, including an image of the bomb. The text claims responsibility not only for the bomb in Princess Street Gardens, but also the arson of two buses. The text includes long stretches of evil laughter, as can be seen in Figure 1 – reflective of the laughter of a fictional villain. Even though the text was posted to the ITS blog, there is also an overt attribution of the attack to ITS, with the author claiming themself to be part of ITS. The Spanish version was a close translation with some evidence of automated translation of English idioms into Spanish, which may have been subjected to correction by a Spanish speaker, but it was not analysed further as part of the case.

#### 4.1.5. Methodology

To investigate H1 and H2 above, IC firstly needed to uncover consistent within-author language features for the suspect's known writing samples, and secondly, needed to assess whether this pattern is distinctive. As described above, IC elected to focus on word n-grams due to validated use in prior research and analyses and because computational n-gram analyses provide a way to take into account all structural features. When considering the criterion of distinctiveness, it is important to take into account the potential authors' linguistic backgrounds as well as the genre and topics of the texts under analysis. A set of n-grams that may be thought to pick out an individual could in fact be common within a certain online community, or common within a certain genre or subject matter. In this case, the suspect is known to be a Greek national, so we needed to check potentially discriminating n-grams against the writings in English of other Greek nationals. Furthermore, as noted previously, based on the compiler Archegonos' introduction, where he says that he selected the PoC1 and PoC2 documents for their "theoretical or practical gain" and/or because they "make[s] some points that I agree with", it was clear that the suspect's writing in the *Paroxysm of Chaos* texts is communicating with, and from, an online community of like-minded writers. It is thus important to establish whether the potentially discriminating n-grams are distinctive in this community, as opposed to the wider community of writers of English.

Consequently, we decided to use the texts from the other authors in the PoC1 and PoC2 as our specific comparative set. This approach had advantages and disadvantages. By using these other authors to create our impostors or foils, we automatically had access to a set of texts from a socially close speech community. We had no information as to whether any of these writers were also speakers of Greek (and so tackled this separately, below). This also limited the amount of text for the impostor group with samples from the other authors ranging from 240 to 6353 words.

The n-grams of size 1 to 5 from each text in PoC1 and PoC2 were extracted. Each list of n-grams of size *n* from each of the texts were collated to form a master list of n-grams of size *n*. For instance, the master bigram list included all 34,434 bigrams from 29 texts (from the PoC1 (20 texts) and PoC2 (9 texts)). Each text was then automatically analysed using R for the frequency of each of these bigrams and the results were recorded in a data matrix, such as Table 3, which presents an extract of this matrix.

**Table 3. Extract from the bigram data matrix recording the occurrence of each bigram in the corpus in each of the texts.**

| | Anonymous_1.txt | Archegonos_1.txt | Archegonos_2.txt | Archegonos_3.txt |
|---|---|---|---|---|
| **i_can** | 2 | 1 | 1 | 1 |
| **can_only** | 4 | 0 | 0 | 0 |
| **only_be** | 2 | 0 | 1 | 0 |

Each text was then compared with every other text in terms of which n-grams were shared (types) as well as the overall number of shared n-grams (tokens). This involved summing the counts of the n-grams for each comparison set. So, taking Table 3 as an example, Anonymous_1 and Archegonos_2 share the bigrams "i_can" and "only_be". There are 3 occurrences of each n-gram across these texts. Thus, the total frequency of shared n-grams from Table 3 between these two texts is 6. Because the lengths of the texts differ substantially (see Tables 1 and 2), the number of shared n-grams were normalised by dividing that number by the sum of the number of word tokens in each text in the pair of texts that were being compared. So, for instance, Anonymous_1 and Archegonos_2 share 249 bigrams in total. Anonymous_1 has 1309-word tokens, whilst Archegonos_2 has a total of 8364-word tokens, meaning that they share 25.74 bigrams per thousand words.

This process was used to establish a consistent within-author pattern for Archegonos. Specifically, where n-grams were used across two or more of Archegonos' texts, even if only in a single instance in each text, we considered this sufficient to indicate a consistent aspect of style across his known texts. For example, in Table 3 "I can" is consistently used across Archegonos 1, 2 and 3, even though in each text it only occurs once. Once IC had established the consistently-used n-grams, the process was repeated to include the n-grams from the disputed text. Each compared pair of texts were then ranked according to their relative frequency for the particular n-gram length. So, the text pair that had the highest relative frequency of shared n-grams (of length *n*) was assigned 1, whilst the text that had the lowest relative frequency of shared n-grams was assigned 465. This was repeated for each length of n-gram. The median of the ranks for all n-gram lengths for each compared pair of texts was retrieved. (The median is the central number of the rankings when all the rankings for a comparison pair are arranged from smallest to largest.) The lowest median indicated the most similar text pair with respect to all n-gram lengths. Median was used for ranking in case the data was skewed by outliers.

In addition to retrieving an overall rate of sharedness between authors, within authors and with the disputed text, we needed to assess distinctiveness of the shared n-grams to answer our hypotheses. For H1, we searched for 1-5-word n-grams which were (1) repeated across Archegonos sections in the PoC texts; (2) exclusive to Archegonos as a writer in the PoC texts, i.e. which occurred only in the Archegonos text and not in the text written by the other writers; and (3) in the disputed text.

With respect to H2, we searched for 1-5-word n-grams which were (1) shared between the disputed and other writers in the PoC 1 and PoC2, but (2) not shared with any of Archegonos' texts. Apart from Wolfi Landstriecher, no other authors had mul-

tiple texts. As a result, we could not check for consistency with respect to our other authors.

Finally, because Karvounakis was a Greek national, the shared and distinctive n-grams between Archegonos and the disputed text were examined for their usage in the Greek section of the International Corpus of Learner English to determine whether any of them were indicative of generic Greek-English idioms[3].

### 4.1.6. Results

This analysis revealed that the texts that are most similar to each other across all n-gram lengths across PoC1 and PoC2 are Archegonos_2 and Archegonos_8, indicating within-author consistency. This is shown in Table 4 along with the top ten most similar ranked file pairings. The next most similar pairs of text are Archegonos_8 and Archegonos_2 linked with the Unknown texts. As discussed above, we had always believed that some of these texts had been authored by Archegonos, but because the sign-offs were missing or unclear we did not want to include them as known to have been written by him. Given that these Unknown author texts when paired with known Archegonos texts share numerous n-grams, we concluded that it is consistent with the proposition that at least parts of these texts were indeed authored by Archegonos. Unknown_1 linked with Unknown_3 are the next in the ranking of most similar text pairs, sharing a high proportion of n-grams. This is perhaps not surprising given that they may well have the same author(s), and also are in a similar register in their announcement of and claims of responsibility to attacks. There remains the possibility that none of Archegonos nor the impostors wrote the Unknown author texts, and this remains a weakness of the IM, which might be mitigated by the non-linguistic evidence in the case.

**Table 4. Top 10 most similar text pairs by ranked n-gram analysis.**

| Paired texts | Unigram rank | Bigram rank | Trigram rank | Fourgram rank | Fivegram rank | **Median rank** |
|---|---|---|---|---|---|---|
| Archegonos_2 + Archegonos_8 | 1 | 1 | 1 | 1 | 1 | **1** |
| Archegonos_8 + Unknown_4 | 2 | 2 | 2 | 2 | 6 | **2** |
| Archegonos_2 + Unknown_1 | 3 | 3 | 4 | 20 | 36 | **4** |
| Archegonos_8 + Unknown_1 | 4 | 4 | 5 | 6 | 12 | **5** |
| Unknown_1 + Unknown_3 | 11 | 9 | 7 | 5 | 7 | **7** |
| Archegonos_2 + Unknown_4 | 5 | 7 | 8 | 13 | 19 | **8** |
| Archegonos_2 + Archegonos_3 | 8 | 11 | 3 | 7 | 35 | **8** |
| Archegonos_2 + Unknown_3 | 7 | 8 | 6 | 19 | 42 | **8** |
| Archegonos_2 + CoA_1 | 9 | 5 | 9 | 11 | 17 | **9** |

When introducing the disputed claim of responsibility into this analysis, the most similar texts to the disputed text were written by Archegonos, specifically Archegonos_8 and Archegonos_2. These pairs of texts rank the 16th and the 30th most similar across all lengths of n-gram with all files included in the analysis (including those attributed

[4]Thanks go to one of our reviewers who writes: "'Nichilistically' is very interesting, because it is obviously not a Spanish spelling. Given the post-hoc knowledge that the suspect was Greek I searched and the word is spelled in Greek with the 'ch' symbol also used for Christ. So a focus on that spelling might have led to a suggestion that the author was Greek".

to sign-offs other than Archegonos). The next most similar text to the disputed text was Unknown_4, which was ranked the 35th most similar pair overall. The highest ranked text by similarity that was not written by Archegonos was by Wolfi Landsteicher_2, which is ranked 45th in the overall table.

Given the strong within-author consistency between Archegonos_2 and Archegnonos_8 and the strong similarity of these texts to the disputed text, it suggests that the disputed claim of responsibility is likely written by Archegonos. However, producing such a metric in a forensic context is problematic, especially in the context of a jury trial as it is not the task of the expert to tell the jury who wrote the text, but rather the linguist should serve as a "linguistic tour guide", showing elements of similarity and difference and allowing the jury to make up their own mind (Grant, 2022; Solan, 1998). As a result, we looked more specifically at the types of shared n-grams to test our hypotheses.

To examine H1, n-grams that were exclusive to Archegonos and the disputed text were explored. This revealed 40 uni-grams, 141 bi-grams, 97 tri-grams, 19 four-grams, and 3 five-grams shared only between Archegonos' texts and the disputed text. Yet, it was important that we only considered those that were used across two or more of his known texts, which in this small number of texts, we considered were part of his consistent style. This left us with 31 uni-grams, 55 bi-grams, 37 tri-grams, and 2 four-grams. Examples included "devours", "nihil and", "the human animal", and "I am a nihilist". An examination of all these n-grams in large general corpora of English, including the British National Corpus 2014 (spoken (Love, Dembry, Hardie, Brezina, & McEnery, 2017) and written (Brezina, Hawtin, & McEnery, 2021) on LancsBox‘ X (Brezina & Platt, 2024)) and the English Web 2021 corpus on SketchEngine Kilgarriff et al. (2014) show them to be rare, as can be seen in Table 5.

**Table 5. The frequencies and relative frequencies per million words (PMW) in the BNC2014 and enTenTen21 of a sample of the n-grams exclusive to Archegonos and the disputed text.**

| | enTenTen21 | BNC2014 |
|---|---|---|
| devours | 14526 (0.24 PMW) | 11 (0.11 PMW) |
| nihil and | 150 (<0.01 PMW) | 1 (0.01 PMW) |
| the human animal | 4352 (0.07 PMW) | 8 (0.08 PMW) |
| I am a nihilist | 43 (<0.01 PMW) | 0 (0 PMW) |

None of the five-grams occurred more than twice in known Archegonos' texts. Rather, they only occurred once in Archegonos' texts and once in the disputed. It is therefore not possible to argue that these are consistent features of Archegonos' known style, but alongside the shared shorter n-grams (bi-grams, tri-grams, etc.) this does seem to provide some additional weight of evidence to link Archegonos with the claim of responsibility. The five-grams included: "unite on the basis of", "for the human being to", and "wouldn't exist without the". An exploration of these five-grams in the same large general corpora of English also show them to be exceedingly rare and therefore provide some measure of distinctiveness (see Table 6).

Because Karvounakis was a Greek national, each exclusive n-gram, which also appeared in the disputed text, was also examined to see if it was characteristic of the Greek section of the International Corpus of Learner English. Our purpose was to determine whether any of these n-grams were in fact indicative of generic Greek-English idioms, rather than an indication of individual style. However, none the exclusive n-grams were found in the ICLE corpus, so we concluded that none were in fact a function of Greek-English language contact.

**Table 6. The frequencies and relative frequencies per million words (PMW) in the BNC2014 and enTenTen21 of the five-grams in the disputed text shared with Archegonos' texts from PoC1 and PoC2.**

| five-grams | Freq. in enTenTen21 | Freq. in BNC2014 |
|---|---|---|
| would n't exist without the | 594 (0.01 PMW) | 0 (0 PMW) |
| unite on the basis of | 170 (0.01 PMW) | 0 (0 PMW) |
| for the human being to | 220 (0.01 PMW) | 0 (0 PMW) |

Additionally, it was important to consider H2 – that someone other than Archegonos wrote the disputed text – by finding n-grams that were exclusive to the disputed text and one of the other authors from PoC1 and PoC2 (but not Archegonos). This revealed in total across the other 19 sections (not written by Archegonos in PoC1 and PoC2) two five-grams, fourteen four-grams, 67 tri-grams, 150 bi-grams, and 80 uni-grams that were shared with the disputed text. This overlap is clearly important, but exploration of the n-grams across the BNC2014 and enTenTen21 deemed them to be more generic than those exclusive to Archegonos' text and the disputed text. For example, Table 7 presents some tri-grams, four-grams and five-grams that were shared by non-Archegonos' texts and with the disputed and their frequencies and relative frequencies in the BNC2014 and the English Web Corpus 2021. These frequencies show them to be more common than those exclusive to Archegonos and the disputed text and therefore carry little evidential weight. In addition, none of the other authors shared as many exclusive n-grams with the disputed text as Archegonos.

**Table 7. The frequencies and relative frequencies per million words (PMW) in the BNC2014 and enTenTen21 of n-grams shared across the disputed text and other authors from PoC1 and PoC2.**

| | Freq. in enTenTen21 | Freq in BNC2014 |
|---|---|---|
| **who want to** (shared with Erinne Vivani and Unknown 4) | 1148768 (18.65 PMW) | 722 (7.06 PMW) |
| **and what is** (shared with Chris Kortright) | 343176 (5.57 PMW) | 460 (4.5 PMW) |
| **do not have** (shared with Chris Kortright) | 8930217 (145 PMW) | 1481 (14.48 PMW) |
| **there are people who** (shared with E. M Cioran) | 66534 (1.08 PMW) | 99 (0.97 PMW) |
| **those who want to** (shared with Erinne Vivani) | 153082 (2.49 PMW) | 127 (1.24 PMW) |
| **I do n't give a** (shared with Wolfi Landstreicher) | 26751 (0.43 PMW) | 157 (1.53 PMW) |
| **not to fall into the** (shared with Unknown 4) | 2386 (0.04 PMW) | 4 (0.04 PMW) |

## 5. Reflections

In general, working on this case required drawing on a diverse set of methods and tools for forensic authorship analysis. The choice of tool was as much determined by the context at the stage of the case as by the structure of the authorship problem. Thus, in the Stage 1 analysis, which attempted to profile the writer, because of the security demands, a hand analysis performed by a single linguist was all that was possible. The outcome of this analysis, however, was not intended to be evidential, but simply to inform the investigation. For Stage 2, security was still a restriction. The analysis was now a cross-genre comparison intended to be taken before a court, but the evidential bar stayed relatively low. Stage 3 was for a criminal standard of proof but was free of the previous security constraints. The comparative analysis was also able to be carried out on within-genre texts drawn from the same speech community. In addition, steps were taken to avoid bias by recruiting a fresh linguist uncontaminated by previous work on the case.

These different contexts and demands influenced the complexity and the degree to which the approach could draw on well-attested methods from the literature. Even in the third stage, the specifics of this analysis required and allowed for variation from the published and tested General Impostors Method. A lot has been written on the necessity for validation of methods in all forms of forensic analysis (Morrison & Enzinger, 2016; Stoel et al., 2015), including forensic authorship analysis (e.g. Grant, 2022; Solan, 2012), and this is clearly crucial. However, as the progress of this case shows, adaptability of method to context is also a key component of at least forensic authorship analysis and this needs more thought and discussion in the literature.

### 5.1. Reflections on impostors methods

The Stage 3 analysis is an authorship verification task, one of the harder tasks of comparative authorship analysis. In contrast with classification tasks, in authorship verifi-

cation tasks, there is potentially an open set of other authors who might have written the disputed text, beyond the single suspect. One of the few published methods for authorship verification is the IM (Koppel & Winter, 2014). This effectively transforms the problem into a quasi-closed set problem by creating sets of random, plausible, and genre-matched impostor texts to the disputed text. Potha and Stamatatos (2017) argue that, like the police line-up, the impostor texts need to be matched to the suspect's known writings, rather than the disputed. We position the methodology we used for Stage 3 as being similar to Potha and Stamatatos' (2017) version of the GIM, in the sense that our impostor texts were socially similar to the known writings of the suspect. We did not follow Potha and Stamatatos' method of searching for internet texts with saliant features. Yet arguably, our impostor texts are even more specific, having been chosen by our suspect to be included in the zines, which he compiled. They are anarchist texts, crossing a similar range of different genres, including poems, claims of responsibility, letters, introductions, and opinion pieces. Additionally, our approach is like Potha and Stamatatos' through our ranking of the pairs of compared texts. This does not entirely counter the argument that the quasi-close set can never be a true closed set. A badly collected or constructed comparison set of impostor documents can statistically emphasise similarity between Q and K documents. Although we tried to avoid this possibility by using the specific impostors provided by the case, this is a risk, or cross-examination issue, that cannot be entirely avoided.

In summary our approach is distinct from other impostor methods in a number of ways: First, we compare Archegonos' known writings to our impostor texts. This allows us to observe our within-author consistency, especially across genres, as well as the influence of genre on our impostor documents. Second, we have not artificially created the set of impostors. The convenience sample of potential impostors was provided naturally by the nature of the documents in the case. Whereas in the GIM, these are created through search engine hits of significant words (either in the disputed text or known writing samples), all of our impostors were found as authors in the PoC zines. This use of specific impostors allows us to assess the distinctiveness of both the disputed text and of Archegonos' known writings from writers drawn from the same very specific community and writing in similar genres. Just as in a line-up, choosing close foils makes the comparison harder and thus the identification of the suspect carries greater evidential weight. The third difference from the GIM is that previous advocates have used statistical distance metrics. As we felt that such approaches would add a layer of complexity in presentation to a jury, we instead used the ranking method described above based on the relative frequencies of shared n-grams. This preserves the usefulness of the IM whilst recognising the importance of the courtroom context for forensic work.

## 6. Outcomes of the case

On the basis of the Stage 3 analysis, TG and IC wrote up an expert witness report in which they were able to assert that:

- the probability of seeing these [language features] in the disputed text is high, if Karvounakis is the author; and that,
- the probability of seeing these [language features] in the disputed text is low if someone other than Karvounakis is the author.

This kind of conclusion did not say that Karvounakis wrote the text, but rather provides an opinion for the jury, which then allows them to draw their own conclusion. In UK jurisdictions, this is now considered best practice for the type of conclusion in an expert witness report. It should be noted that to our knowledge there was no linguist appointed as expert witness for the defence.

Karvounakis was presented with the full range of evidence against him, and this included our linguistic evidence. On the basis of this, he pled guilty to a contravention of Section 57 of the Terrorism Act 2000. Karvounakis maintained that he did not intend the device to detonate and that he had manufactured the device in such a way that it could not have ignited. The judge in the case, Lord Braid (2022), said: "the scientific evidence is that even if what you say is true, there remained a potential for explosion, due to the potential introduction of a hot filament wire in close proximity to a low explosive substance, for example if the fuse was not sufficiently well separated from the filament." Lord Braid goes on to recognise the implications of our analysis saying to Karvounakis that "your present position is not consistent with the claim of responsibility which you posted, or at least caused to be posted, on the internet shortly after the device had been found."

On the 16th February 2022, Karvounakis was sentenced to 8 years and 4 months in prison at the High Court in Edinburgh.

## Acknowledgements

We would like to thank our reviewers for their extremely helpful comments and feedback.

## References

Andrijanić, I., & Bąkowski, J. (2021). Stylometry Then and Now: Authorship Verification of Vivekacūḍāmaṇi and Vedāntic Stotras. *Challenges of Interdisciplinary and Multidisciplinary Approach: New Horizons in Oriental Studies (Prace Orientalistyczne/Oriental Studies)*, *44*, 30–46.

Batel, A., Abrams, M., & Pęzik, P. (2022). Native Dialect Influence Detection (NDID) differentiating between Mexican and Peninsular L1 Spanish in L2 English. *Language and Law=Linguagem e Direito*, *9*(1), 120–145. Retrieved 2025-06-27, from https://ojs.letras.up.pt/index.php/LLLD/article/view/12829/11685 doi: 10.21747/21833745/lanlaw/9_1a6

Brezina, V., Hawtin, A., & McEnery, T. (2021). The Written British National Corpus 2014 – design and comparability. *Text & Talk*, *41*(5-6), 595–615. Retrieved 2025-06-27, from https://www.degruyter.com/document/doi/10.1515/text-2020-0052/html doi: 10.1515/text-2020-0052

Brezina, V., & Platt, W. (2024). *#LancsBox X.* Lancaster University. Retrieved from http://lancsbox.lancs.ac.uk

Crown Office and Procurator Fiscal. (2024). *Expert Witness Guide.* Retrieved from https://www.copfs.gov.uk/for-professionals/expert-witness-guide

Darling, S., Valentine, T., & Memon, A. (2008). Selection of lineup foils in operational contexts. *Applied Cognitive Psychology*, *22*(2), 159–169. Retrieved 2025-06-27, from https://onlinelibrary.wiley.com/doi/10.1002/acp.1366 doi: 10.1002/acp.1366

Edmond, G., Towler, A., Growns, B., Ribeiro, G., Found, B., White, D., … Martire, K. (2017). Thinking forensics: Cognitive science for forensic practitioners. *Science & Justice*, *57*(2), 144–154. Retrieved 2025-06-27, from https://linkinghub.elsevier.com/retrieve/pii/S1355030616301289 doi: 10.1016/j.scijus.2016.11.005

Found, B. (2015). Deciphering the human condition: the rise of cognitive forensics. *Australian Journal of Forensic Sciences*, *47*(4), 386–401. Retrieved 2025-06-27, from http://www.tandfonline.com/doi/full/10.1080/00450618.2014.965204 doi: 10.1080/00450618.2014.965204

Grant, T. (2022). *The Idea of Progress in Forensic Authorship Analysis* (1st ed.). Cambridge University Press. Retrieved 2025-06-27, from https://www.cambridge.org/core/product/identifier/9781108974714/type/element doi: 10.1017/9781108974714

Grant, T., & Grieve, J. (2022). The Starbuck Case: Methods for Addressing Confirmation Bias in Forensic Authorship Analysis. In R. Perkins, I. Picornell, & M. Coulthard (Eds.), *Methodologies and Challenges in Forensic Linguistic Casework* (1st ed., pp. 13–28). Wiley. Retrieved 2025-06-27, from https://onlinelibrary.wiley.com/doi/10.1002/9781394266661.ch2 doi: 10.1002/9781394266661.ch2

Grant, T., & MacLeod, N. (2020). *Language and Online Identities: The Undercover Policing of Internet Sexual Crime* (1st ed.). Cambridge University Press. Retrieved 2025-06-27, from https://www.cambridge.org/core/product/identifier/9781108766425/type/book doi: 10.1017/9781108766425

Grieve, J. (2007). Quantitative Authorship Attribution: An Evaluation of Techniques. *Literary and Linguistic Computing*, *22*(3), 251–270. Retrieved 2025-06-27, from https://academic.oup.com/dsh/article-lookup/doi/10.1093/llc/fqm020 doi: 10.1093/llc/fqm020

Grieve, J., Clarke, I., Chiang, E., Gideon, H., Heini, A., Nini, A., & Waibel, E. (2019). Attributing the Bixby Letter using n-gram tracing. *Digital Scholarship in the Humanities*, *34*(3), 493–512. Retrieved 2025-06-27, from https://academic.oup.com/dsh/article/34/3/493/5145759 doi: 10.1093/llc/fqy042

Judicial Institute for Scotland. (2023). *e-Jury Manual.* Retrieved from https://judiciary.scot/docs/librariesprovider3/judiciarydocuments/judicial-institute-publications/export_e-jury_manual_2023-08-15_0911-(1)-(1).pdf?sfvrsn=ff85e46f_1

Khonji, M., & Iraqi, Y. (2014). A slightly-modified gi-based author-verifier with lots of features (asgalf). In *CLEF 2014 Labs and Workshops, Notebook Papers.* CLEF and CEUR-WS.org.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., … Suchomel, V. (2014, July). The Sketch Engine. *Lexicography*, *1*(1), 7–36. Retrieved 2025-06-27, from https://utppublishing.com/doi/10.1007/s40607-014-0009-9 doi: 10.1007/s40607-014-0009-9

Koppel, M., & Winter, Y. (2014). Determining if two documents are written by the same author. *Journal of the Association for Information Science and Technology*, *65*(1), 178–187. Retrieved 2025-06-27, from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.22954 doi: 10.1002/asi.22954

Lord Braid. (2022). *Sentencing Statements.* Retrieved from https://judiciary.scot/home/sentences-judgments/sentences-and-opinions/2022/02/16/hma-v-nikolaos-karvounakis

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014:

Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, 319–344. Retrieved 2025-06-27, from http://www.jbe-platform.com/content/journals/10.1075/ijcl.22.3.02lov doi: 10.1075/ijcl.22.3.02lov

Morrison, G. S., & Enzinger, E. A. (2016, April). New Paradigm for the Evaluation of Forensic Evidence. National Institute for Science and Technology. Retrieved from https://www.nist.gov/system/files/documents/2020/04/03/04_morrisonenzinger_nist_workshop_2016_04_30a_optimized.pdf

Pautz, N., McDougall, K., Mueller-Johnson, K., Nolan, F., Paver, A., & Smith, H. M. J. (2023). Identifying unfamiliar voices: Examining the system variables of sample duration and parade size. *Quarterly Journal of Experimental Psychology*, *76*(12), 2804–2822. Retrieved 2025-06-27, from https://journals.sagepub.com/doi/10.1177/17470218231155738 doi: 10.1177/17470218231155738

Potha, N., & Stamatatos, E. (2017). An Improved Impostors Method for Authorship Verification. In G. J. Jones et al. (Eds.), *Experimental IR Meets Multilinguality, Multimodality, and Interaction* (Vol. 10456, pp. 138–144). Cham: Springer International Publishing. Retrieved 2025-06-27, from https://link.springer.com/10.1007/978-3-319-65813-1_14 (Series Title: Lecture Notes in Computer Science) doi: 10.1007/978-3-319-65813-1_14

R Core Team. (2023). *_r: A Language and Environment for Statistical Computing_*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Seidman, S. (2013). Authorship Verification Using the Impostors Method. In P. Forner, R. Navigli, & D. Tufis (Eds.), *CLEF 2013 Evaluation Labs and Workshop – Working Notes Papers.*

Solan, L. M. (1998). Linguistic experts as semantic tour guides. *International Journal of Speech, Language and the Law*, *5*(2), 87–106. Retrieved 2024-06-14, from https://journal.equinoxpub.com/IJSLL/article/view/6582 doi: 10.1558/sll.1998.5.2.87

Solan, L. M. (2012). Intuition versus Algorithm: The Case of Forensic Authorship Attribution. *Brooklyn Journal of Law and Policy*, *21*(551).

Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*, *60*(3), 538–556. Retrieved 2025-06-27, from https://onlinelibrary.wiley.com/doi/10.1002/asi.21001 doi: 10.1002/asi.21001

Stamatatos, E., Kestemont, M., Potthast, M., Kredens, K., Pęzik, P., Heini, A., … Potthast, M. (2022). Overview of the Authorship Verification Task at PAN 2022. In *CEUR workshop proceedings* (Vol. 3180, pp. 2301–2313). CEUR-WS. org. (Series Title: Lecture Notes in Computer Science) doi: 10.1007/978-3-031-13643-6_24

Stoel, R. D., Berger, C. E., Kerkhoff, W., Mattijssen, E. J. A. T., & Dror, E. I. (2015). Minimizing contextual bias in forensic casework. In M. J. Hickman & K. Strom (Eds.), *Forensic Science and the Administration of Justice: Critical Issues and Directions* (Online-Ausg ed., pp. 67–86). Los Angeles: SAGE.

*Language and Law / Linguagem e Direito*, *Vol. 11*(2), 2024, pp. 24-54
DOI: doi.org/10.21747/21833745/lanlaw/11_2a8

# Offence-Presumptive Terms: A Troubling Category in Linguistic Offensiveness†

Chris Heffer [1]

[1] *Cardiff University*, HefferC2@cardiff.ac.uk

***Abstract***
*This paper argues that there is a set of words, word senses and phrases that are not inherently offensive but that can lead the hearer to presume that they are offensive. The archetypal, though now somewhat problematic, case is the adjective 'niggardly' ('stingy', 'parsimonious'), but I also discuss two cases based on forensic casework: the metaphorical sense of the verb 'kneecap' (bring to its knees); and the now aging catchphrase 'never mind the quality, feel the width' (go for quantity over quality, used ironically). These offence-presumptive terms are a subcategory of Low-Occurrence Ordinary Terms (LOOTs). LOOTs are ordinary enough for the speaker to presume they will be understood and for the hearer to presume that they know or can infer the meaning, while at the same time infrequent enough not to be familiar to many hearers. Crucially, offense-presumptive LOOTs have lexical 'neighbours' in our mental lexicons that are either attested as, or are perceived to be, offensive. I argue that the presumption of offensiveness comes about because hearers apply autonomous intuitive reasoning to the unknown word, word sense or phrase. While in many cases of lexical inferencing, such intuitive reasoning can lead us to make a successful guess about the meaning of a word, in the case of offence-presumptive terms, the hearer is led astray. Moreover, in raising a claim of linguistic offence, claimants prime HR personnel to activate in their own minds an offensive frame for the term. I consider the implications of this novel category for forensic lexical analysis, the judgment of linguistic offence and the ethics of interpersonal communication.*
***Keywords:*** *offensive language, lexical neighbours, forensic lexis, dual-process reasoning, lexical inferencing, phonological similarity,* niggardly*, N-word.*

***Resumo***
*Este artigo defende que existe um conjunto de palavras, significados e orações que não são inerentemente ofensivos, mas que podem levar o ouvinte a presumir que o são.*

† **Content Note** Following standard linguistic practice in the area of offensive language, this article mentions and quotes taboo words including slurs.

*O caso prototípico, embora agora um pouco problemático, é o adjetivo "niggardly" ("mesquinho", "avarento"). Além deste, discuto também outros dois casos baseados em trabalhos forenses: o sentido metafórico do verbo "kneecap" ("ajoelhar-se"); e a expressão agora antiquada "never mind the quality, feel the width" ("prefira a quantidade à qualidade", usada ironicamente). Esses termos presumivelmente ofensivos são uma subcategoria dos Termos Comuns de Baixa Ocorrência (LOOTs). Os LOOTs são comuns o suficiente para que o falante presuma que serão compreendidos e para que o ouvinte presuma que conhece ou pode inferir o seu significado; simultaneamente, também são pouco frequentes para não serem familiares a muitos ouvintes. Na sua essência, os LOOTs, que podem ser considerados ofensivos, têm "vizinhos" lexicais nos nossos léxicos mentais que são comprovadamente ofensivos ou são percebidos como tal. Discuto aqui que a presunção de ofensividade surge porque os ouvintes aplicam um raciocínio intuitivo autónomo à palavra, sentido da palavra ou oração desconhecida. Embora em muitos casos de inferência lexical, esse raciocínio intuitivo possa levar-nos a adivinhar com sucesso o significado de uma palavra, no caso de termos que podem ser considerados ofensivos, o ouvinte é induzido em erro. Além disso, ao apresentar uma alegação de ofensa linguística, os queixosos levam o pessoal de RH a ativar nas suas próprias mentes uma interpretação ofensiva para o termo. Considero ainda as implicações dessa nova categoria para a análise lexical forense, o julgamento da ofensa linguística e a ética da comunicação interpessoal.*
***Palavras-chave:*** *Linguagem ofensiva, vizinhos lexicais, léxico forense, tarefa dupla de raciocínio, inferência lexical, semelhança fonológica, "niggardly", N-word.*

## 1. Introduction

This article develops a novel argument relating to a particularly troublesome category of claimed linguistic offensiveness: the case where a claimant is deeply offended on hearing a word or phrase that is neither objectively offensive from a linguistic standpoint nor intended by the speaker to be offensive. The archetypal (though now highly controversial) case is that of the adjective *niggardly* ('stingy'), which is phonologically similar to the N-word but etymologically and semantically completely unrelated to it. In a famous case in 1999, the director of a municipal agency in Washington, David Howard, told his staff that due to severe financial constraints, he would need to be 'niggardly' with the budget. A couple of Howard's Black staff 'began a whispering campaign that blossomed into a public outcry' (Kennedy, 2002, p. 94) which led to the mayor accepting Howard's resignation on the grounds that he had shown 'poor judgement.' Yet Howard had used the word in its most common context of use (indicating financial constraints) and there was no suggestion that he intended to be offensive. Nor, at the time, was there extensive media discussion of the word of which he could have been aware. While the first

public complaint about the use of *niggardly* appears to have been by a Boston reader in 1995 complaining about The *Economist*'s use of the word in 'productivity growth averaged a niggardly 0.8% a year', the UK-based magazine simply treated the complaint with condescending amusement (Derbyshire, 2002). Reactions to the Howard scandal were as deeply polarised then as controversies about offensive language are today and showed very little understanding about the nature of language and communication. On the 'offence' side, one columnist asked whether Howard could possibly not have noticed that 'he had to pass "nigger" before he could get to the "dly"' (Dickerson, 1999). But this is simply not how we process language: we do not have to 'pass' *cunt* when uttering *country* and a large proportion of informants who are asked what Rice Krispies are made of are unable to respond 'rice' (Wray & Staczek, 2025) despite having to 'pass' the word when naming the brand. Instead, we store words and phrases as single meaning units ('stingy' or 'brand of breakfast cereal') and, even in the case of homonyms, we automatically select the word or sense that is most appropriate in the context. This is why it is not unusual in Wales to see adverts for faggots (Welsh meatballs) such as the following:

> **Faggots** are an everyday favourite here in Pembrokeshire, so we've created what we see as the perfect Pembrokeshire **faggot** – see what you think! (Gwaum Valley Meats at https://gwaunvalleymeats.co.uk/product/gwaun-valley-faggots. Accessed 28/6/25)

Just as the writer of this advert almost certainly was not thinking of the gay slur when writing *faggot*, so Howard was most likely unaware of the phonological proximity to the N-word when uttering the contextually appropriate word *niggardly*, and one wonders how many teachers notice the proximity of *snigger* to *nigger* when telling pupils to 'stop sniggering at the back of the class'. At the same time, the 'defence' reactions also showed ignorance of language. One columnist described Howard as the victim of 'linguistic lynching' who got fired because 'some people in public employ were morons who … didn't know how to use a dictionary' (Snow, 1999). Yet offensiveness does not lie in the dictionary meaning of a word but in how it is used in a specific communicational context (Heffer, in prep). Chris Rock's frequent use of the N-word in his comedy routines presumably does not offend his usual audiences while a White worker might pointedly use *niggardly* to rile a Black co-worker who she knows is offended by the word. Claimants in such cases are often dismissed as hypersensitive or vexatious. Yet, the hurt can be real. Hearing perceived taboo words can give an emotional jolt that has been measured through skin conductance (Dinn & Harris, 2000) and effect on the amygdala (Hansen, McMahon, & De Zubicaray, 2019). So it is perfectly possible for a claimant to feel genuinely harmed by a perceived linguistic offence even if the offence was entirely unintentional and not objectively warranted by the evidence.

Within linguistics, the issue of offence-adjacent 'innocent' words and senses being perceived as offensive tends to be viewed with respect to processes of tabooification in the development of the language as a whole (Allan & Burridge, 2006; Pinker, 2007) and linked to 'verbal hygiene' (Cameron, 1995), or our deliberate attempts to 'clean up' the language. Pinker (2007), for example, explains the tabooification of such words in terms of a linguistic version of Gresham's Law ('bad money drives out good'): 'bad words drive good words out of circulation' because people 'often avoid using innocent terms that they fear might be misheard as profanity' (Pinker 2007, p. 333). Allan and Burridge, in

their introduction to taboo language, write that 'Speakers will not risk appearing to use a taboo term when none was intended; therefore, they are quick to drop the homonyms of taboo terms' and that '[e]ven words and phrases that are similar to non-PC terms are avoided' (Allan & Burridge, 2006, p. 102). They conclude that:

> As is so often the case in issues to do with language, it doesn't matter what the linguistic facts suggest. The reality that *niggardly* has absolutely no etymological connections with *nigger* is of no consequence. What really matters is how speakers perceive their language to be, and if people do start connecting words such as *nitty-gritty* and *niggardly* with the N-word, then this will be the kiss of death for these words. (2006, p. 104)

They go on to point out that *fuk* ('sail') and *feck* ('effect') 'had absolutely nothing to do with *fuck*, but that did not save them' (2006, p. 104-105).

The process of tabooification, though, as can now be evidenced through large language corpora, actually tends to be more gradual, piecemeal and reversible than these writers seem to suggest. Even where the process is near complete for one sense or grammatical form, it may not be for another. For example, *cripple* (n) has taken decades to become more or less universally taboo when referred to a disabled person in the US and the UK, but as this use has dramatically declined, the metaphorical use ('Truss *crippled* the economy') has significantly increased (Google NGram Viewer 'cripple') rather than suffering 'the kiss of death' through its association with the noun. *Fuck* has been taboo in published writing for centuries and in public broadcasting for decades (throughout remaining common in private speech) but it is far less taboo today than it was 50 years ago. The process of tabooification can be restricted to region (e.g. *cock* in the US but not the UK), age (e.g. the 'F-word' has different references when uttered by the old (*fuck*) and young (*faggot*)) and many other factors. Twenty years after Allan and Burridge (2006), *nitty-gritty* and *niggardly* have not suffered the kiss of death: they are contested from time to time but ultimately the 'linguistic facts' do seem to matter and each time a complaint is made the public are reminded of those facts. Similarly, as we have seen with *faggot*, speakers are not always so quick to drop the homonyms of taboo terms but, instead, offensive-innocent pairs often co-exist peacefully.

While linguists might appear resigned to an 'inevitable' if (one suspects) 'regrettable' tabooification process of 'innocent' words, a recent strand of thinking in the philosophy of language actively prescribes the tabooification of slurs and any words associated with them. While eradicationists, who argue that all uses of the N-word are hurtful and wrong and so should be prohibited from being used by Blacks any more than by Whites (Kennedy, 2002, p. 126-128), have been around for a long time, Stojnić and Lepore (2025) take this prohibitionism to a new level. In their 'articulation account', a slur has a 'pejorative potential' that 'robustly projects out of environments that normally render meaning inert – including not only quotation and meaning attributions, but even mere displays of slurs' (Stojnić & Lepore, 2025, p. 123). In other words, one cannot articulate a pejorative word even in academic discussion such as this. Worse still, '[p]ejorative potential can easily "infect" expressions whose articulations only incidentally match, or resemble, those of slurs' (Stojnić & Lepore, 2025, p. 123). Accordingly, the authors self-censor their mentioning of these 'infected' expressions and have to rely on awkward circumlocutions instead:

> Recall, once again, the incidents involving the tokening of the English adverb[1] that happens to both orthographically and phonetically resemble the N-word but is otherwise etymologically and semantically unrelated. (Stojnić & Lepore 2025, p. 123)

Such fetishization of language, though, fails to acknowledge that offence lies in a given act of communication rather than in the word itself. In a linguistic study such as this, the taboo word is an object of study, like a sample under a microscope; just as an excised tumour on a slide might look unpleasant but would not generally be considered harmful to the lab technician, a taboo word extracted from its normal contexts of use and presented as a linguistic sample in an academic discussion should not generally be considered harmful to the academic reader interested in taboo language. I follow here, then, the standard linguistic practice of referring to the linguistic objects of study when needed, though I mostly use the 'N-word' euphemism because of the hypersensitivity relating to that term.

Given the heightened emotions excited by the N-word, and anything apparently related to it, I would not tackle this topic were it not for forensic casework I have been engaged in over the past few years which involved disputed terms such as *kneecap* and *never mind the quality, feel the width*, which are unmarked by known historical controversies that might have affected the perception of offence. In such cases, I observed three phenomena in particular. Firstly, the items were both ordinary rather than specialised terms and yet occurred infrequently in the language as a whole. Secondly, despite probably not knowing the intended meaning of the term or sense, the claimant had such a strong presumption that they knew the term was offensive that they were willing to lodge a formal complaint. Thirdly, despite there being no linguistic or circumstantial evidence that the disputed word or phrase was offensive in the context, HR personnel were happy to advance the cases through disciplinary proceedings. Rather than taking sides in the culture war on linguistic offensiveness, though, the argument I develop here presumes a situation where the claimant was genuinely offended but where the speaker had neither intention of causing offence nor could have known that what they said might cause offence.

The argument proceeds in six steps, which will be teased out in the remainder of the article:

1. Since communication is not a conduit, there is always the risk that a hearer will not understand a word or phrase used by the speaker or will interpret it in a way unintended by that speaker;
2. This risk increases markedly in the case of a class of words I propose to call Low Occurrence Ordinary Terms (LOOTs). The ordinariness of a LOOT means both that the speaker will expect the hearer to understand it and that the hearer will expect to be able to infer the meaning if it is not part of their lexicon. But the low occurrence of a LOOT (for an ordinary term) means that many hearers will not be familiar with it and may well misconstrue it;

[1] Discussions of *niggardly* in the media and non-linguistic academic publications tend to erroneously describe it as an adverb, possibly because of the *-ly* derivational suffix. The adverb used to exist, as in 'I hope you did not feel that I had treated you niggardly' (OED 2003: *niggardly* adv), but this 1959 citation is the last attested for the adverb in the OED, which describes the adverb as 'now rare'.

3. Given the ordinariness of a LOOT, when hearers are not familiar with it, they are likely to use autonomous intuitive processes to infer its meaning, such as background contextual inferencing and lexical association, rather than deliberative reflective processes such as morphological or componential analysis or looking up the word in a dictionary. Such 'Type 1' associative reasoning often succeeds in inferring the correct or approximate meaning of the unknown term, while 'Type 2' reflective inferencing is often erroneous;
4. However, where a LOOT is phonologically similar or identical to a recognised offensive term or sense in our mental lexicons, autonomous processing is likely to incorrectly associate the neutral LOOT with the (often more common) offensive term or sense;
5. Furthermore, when Type 1 associative processing mis-leads the hearer towards the offensive lexical neighbour, the lack of conscious reflection involved is likely to mean that the hearer will simply presume that the neutral term is offensive and that the speaker is being offensive and thus not consciously reflect on whether this is actually the case;
6. Finally, by raising a formal claim that the word or phrase is offensive, HR personnel are metalinguistically primed to frame the offence-presumptive term as offensive.

The following five sections tease out the first five steps in the argument outlined above. I then operationalise the notion of 'offence-presumptive term' so that it can be applied systematically to data. Next, I re-analyse two examples from casework that might be considered offence-presumptive terms: the verb *kneecap* and the phrase *never mind the quality, feel the width*. Then I consider the question of metalinguistic priming in the framing of the term as offensive. Finally, I discuss the implications of this novel lexical category in terms of forensic lexical analysis in general, judging linguistic offence in contexts such as workplace harassment and hate speech in particular, and the ethics of interpersonal communication more broadly.

## 2. Communicational Risk

Were communication a flawless exchange of shared meanings packaged in transparent signs, there would be no problem with miscommunications and misunderstandings. Both the folk linguistic view and early academic approaches to communication shared what might be called a 'Code and Conduit' model of language. The Code consists in lexical signs and grammatical codes that are shared unproblematically among speakers of a language. These shared codes are then meant to be conveyed fluently through a communicational Conduit between interlocutors, as in Saussure's 1983 famous 'talking heads' diagram of the 'speech circuit' or Shannon and Weaver's Bell-(now AT&T)-funded highly influential model of communication (Shannon, 1948) in which communicating was about 'reproducing at one point … a message selected at another point' (Shannon & Weaver, 1964, p. 31).

Over the last fifty years, though, sociolinguistics, linguistic pragmatics and psycholinguistics have troubled this 'conduit metaphor' (Reddy, 1979) of verbal communication. With respect to the Code, while fluent speakers of a language undoubtedly share a common core of linguistic resources consisting in core lexis (e.g. basic verbs, function words) and grammar (e.g. the SVOCA standard word order in English), be-

yond that core there is an enormous amount of individual variation in our knowledge of language. While we talk of 'speech communities' (Gumperz, 1993), 'discourse communities' (Swales, 1990) or 'communities of practice' (Lave & Wenger, 1991), such linguistic communities are sometimes seen by sociolinguists as little more than aggregates of the idiolects of each individual speaker in that community (Hudson, 1996). When we switch focus from the community to the individual, as in forensic authorship attribution, idiolects become paramount (Coulthard, 2004; Grant, 2013). In particular, our individual mental lexicons of words in long-term memory (Jackendoff, 2002) not only differ from others in the words we have stored there, but our individual representations of those words we apparently share with others are unique because of our own particular lifetime experience with those words. Furthermore, the experiences that will have shaped people's understanding of words and grammar in a multicultural and multilingual community are likely to be even more diverse than in more monocultural communities in the past.

Turning to the Conduit aspect, pragmatics has shown that interpretation depends very heavily on context and that speakers can exploit this knowledge by conveying indirect meaning through, for example, implicature and irony (Dynel, 2018; Grice, 1975). But unlike the conduit model of semantic word containers relayed through a communication pipeline, pragmatic meaning is more hazardous. Implicature and irony require 'uptake', or recognition by the hearer of the speaker's communicative intention, and this is by no means guaranteed (L. McDonald, 2020). For example, autistic children and adults are well known to have difficulty distinguishing literal and non-literal language (Happé, 1995). Similarly, psycholinguistics and cognitive neuroscience have come to see language comprehension as context-dependent, predictive and probabilistic (Federmeier, 2007; S. A. McDonald & Shillcock, 2003). The fact that we 'think ahead' when comprehending language has been demonstrated through eye-tracking experiments and through the measurement of Event-Related Potentials (ERP), or brain responses to a specific linguistic stimulus (Federmeier, 2007). But it has also long been known in conversation analysis: micro-analysis of turn-taking (Sacks, Schegloff, & Jefferson, 1974) shows that listeners monitor the trajectory of an ongoing speaker turn for Transition Relevance Places (TRPs) and often try to take the conversational floor once they can predict the end of the speaker's turn (de Ruiter, Mitterer, & Enfield, 2006). However, prediction always comes with risk: the predicted word in lexical comprehension might turn out to be wrong, or listeners might try several times unsuccessfully to take the floor and if they persist without the speaker ceding the floor, it will be interpreted as an unwanted interruption (Kurtić, Brown, & Wells, 2013).

Given the idiolectal nature of language and the predictive nature of language processing, we often find ourselves in a position where we are not familiar with a word or phrase or it seems to be being used in a sense we do not recognise. In other words, the word or word sense is not in our mental lexicons.

## 3. Low Occurrence Ordinary Terms (LOOTs)

The risk of a hearer misconstruing the meaning of a word or phrase significantly increases in the case of a class of words I call Low Occurrence Ordinary Terms, or LOOTs. The idea of a LOOT has emerged from forensic casework, where I have observed that words and phrases in dispute often belong to this class. There are good reasons for this.

There are two significant aspects to LOOTs that increase the risk of communicational trouble: their ordinariness and their comparatively low occurrence.

'Ordinary Terms' are words, word senses or phrases that would be considered to belong to 'ordinary language' rather than specialist, technical, arcane or archaic language. Ordinary terms can be found in everyday contemporary spoken and written genres such as conversation, public talks, fiction and written and broadcast media. Accordingly, rather than being exclusionary, like professional jargon and circumscribed slang, they are inclusionary in terms of potential audiences. The ordinariness of a LOOT, then, has two consequences:

1. The speaker will expect the hearer to be familiar with the term (Howard would have expected his audience to understand the 'ordinary' adjective *niggardly*); and
2. The hearer will expect to be able to infer its meaning if it is not part of their lexicon (his audience would hear 'be niggardly with the budget' as involving ordinary behaviour rather than some obscure technical process)

These consequences do not follow, on the other hand, if a term is recognised as technical or specialist. The speaker, all things being equal, will not expect the hearer to understand the technical or specialist term unless they are part of the technical or specialist community in which it is used. While professionals sometimes use technical terms to perform expertise and erudition in front of lay audiences (Billig, 2013), they are being exclusionary in so doing because they should know that the audience will not be familiar with such terms. Similarly, if the hearer perceives an unfamiliar term to be technical, they are more likely to accept that they just do not understand it rather than presuming that they understand it or can infer its meaning. Some recognized linguistic cues to technical status include word length (Oppenheimer, 2006) (e.g. *antidisestablishmentarianism*), complex morphology (e.g. *photosynthesis*), certain types of derivational morphemes (e.g. *-ology*, *-itis*), nominalizations (Kies, 1985) (e.g. *facilitation*, *enrichment*) and technical discourse contexts such as scientific or academic publications or technical manuals. One reason why legal language is so problematic is that so many legal terms are disguised as ordinary language terms: *trust*, *person*, *sure* etc.

The other aspect of LOOTs, 'Low Occurrence', needs to be qualified in two respects. Firstly, it is low relative to the ordinary status of the LOOT, not in terms of absolute frequencies in the language. The OED's frequency bands (OED, 2025), based on the Google Books database, are helpful in explaining this. The bands run from 1 (very low frequency) to 8 (very high frequency). Band 1 words (e.g. *abaptiston, zarnich*) are effectively obsolete. Band 2 words (occurring fewer than 0.01 times per million words in contemporary discourse) (e.g. *ennead, abactinal* and *absterge*) are described by the OED as 'almost exclusively terms which are not part of normal discourse and would be unknown to most people' (OED, 2025). The lowest frequency band that might contain ordinary terms, then, is Band 3 (0.01 to 0.1 words per million) (e.g. *amortizable, prelapsarian, agglutinative*), which the OED describes as follows: 'These words are not commonly found in general text types like novels and newspapers, but at the same time they are not overly opaque or obscure.' LOOTs mostly belong to this Band 3 range occurring between 0.01 to 0.1 words per million. However, what characterizes a LOOT in comparison with other words in this band is that, though it does not occur frequently,

it occurs *mostly* in everyday genres such as conversation and the media. The combination of presence in everyday genres but low overall occurrence has the following consequences:

1. The speaker, recognizing its presence in everyday genres, is likely to consider the term 'ordinary' and thus understandable to the hearer;
2. The hearer, though, is likely not to be familiar with the term because of its overall low occurrence.

The second qualification is that while the occurrence is low with respect to the language as a whole, as represented by large reference corpora such as the TenTen family (Jakubíček, Kilgarriff, Kovář, Rychlý, & Suchomel, 2013) and large reference dictionaries like the OED, it may occur much more frequently within certain sociolinguistic varieties such as age groups (e.g. *buzz off* [go away] for Boomers, *delulu* [delusional] for Gen Z), geographical regions (e.g. *duck* [term of endearment] in the English midlands), gender (e.g. *mauve*, once favoured by women), register (e.g. *niggardly* might be a more erudite term than *stingy* and so much less common in everyday speech) or topic domain (e.g. *googly* in cricket or *blocking* in theatre). A term might be fairly common within a specific sociolinguistic variety, but, when communicating across varieties, that commonality is lost.

It is important to note with respect to 'Low Occurrence' that frequency is an indispensable but imperfect analytical proxy for likelihood of comprehension. For example, the N-word is far more widely understood than its attested frequencies in corpora would suggest, while *rule of law* is a moderately frequent phrase (4.2 words p/million in enTenTen21, or OED Band 5) yet many believe that it means something like a citizen's need to obey laws –

> The rule of law is not optional in the United States of America. If people want to come here, you WILL obey our laws. (US Secretary of Homeland Security Kristi Noem X 4 Feb 2025)

– rather than the supremacy of rules-based as opposed to arbitrary power, such as respecting judicial decisions and not exceeding the powers of one's branch of government (Bingham, 2011). An analytical alternative to frequency is familiarity, a hearer's subjective perception of recognition of a word, but this is also just a proxy for likelihood of comprehension. Familiarity Rating Scales (Tanaka-Ishii & Terada, 2011) actually measure three separate phenomona: familiarity with the form of the word ('I have frequently seen or heard this word'); *belief* in its meaning ('I know what it means'); and an ability to define it ('I can provide a definition for it'). Yet, as *rule of law* demonstrates, you can be familiar with the form of a term, believe you know its meaning and even believe you can define it while actually demonstrating no understanding of the term at all. Nevertheless, research has shown a clear correlation between frequency and familiarity ratings at least for spoken language (Tanaka-Ishii & Terada, 2011) and frequency is the much easier of the two to measure. For the purposes of establishing low occurrence (of an ordinary term), frequency of use can be operationalised through:

a) frequency counts in very large general corpora of a language like the TenTen Corpus family (Jakubíček et al., 2013);
b) yearly frequencies of words on Google NGram Viewer (Pettit, 2016); and
c) OED frequency bands (OED, 2025).

Corpus-based counts tend to be the most accurate in terms of contemporary usage. The Google counts, based on the enormous Google Books corpus (Juola, 2022), are restricted to usage in written books, but they are useful for tracing how a word falls in and out of favour. Finally, the OED frequency bands, which are based on the Google counts, are useful for giving meaning to the particular frequency of a word or phrase. One significant complicating factor is that we are often interested in the frequency of a particular *sense* of a word and this information can often only be gleaned from close analysis of concordances in a corpus.

## 4. Lexical Inferencing and Dual Processing

What happens, then, when a hearer comes across an unfamiliar LOOT? When we hear a word or phrase that is not in our mental lexicon or represented differently there, three things might happen. We might suffer a 'slip of the ear' (Bond, 1999) and hear a word that is in our lexicon rather than the one that was actually said; for example, hearing *niggard* (stingy person) as the N-word. We might perceive the word correctly, but recognise that we could not understand it without explicit external help, as in the case of technical terms like *nigrosin* (a black pigment). Thirdly, though, we might perceive the word correctly, be unable to retrieve it from our mental lexicon, but then try to infer its meaning. This is the area of lexical inferencing (Haastrup, 1991).

It is helpful at this point to draw on dual process theories of reasoning, according to which cognitive performance is the result of two types of processing, generally referred to as Type 1 and Type 2 (Evans & Stanovich, 2012; Sloman, 1996; Wason & Evans, 1974). Type 1 processes are intuitive and autonomous and so usually faster, while Type 2 processes require 'working memory' (Baddeley, 2007) and so are usually slower. Dual process theories have been particularly applied in such areas as reasoning (Evans, 2012; Evans & Over, 1996) and decision-making (Kahneman, 2013). For example, the Cognitive Reflection Test (CRT) (Frederick, 2005) tests people's ability to suppress Type 1 processing in certain non-intuitive problems. One of the CRT questions is the bat-and-ball problem:

> A bat and ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?

The majority of participants incorrectly answer 10 cents because this is the rapid response suggested by our autonomous intuitive Type 1 processing. In order to get the correct answer, you need to suppress or override the intuitive response and use your working memory to actively reason through the problem reflectively (Type 2): the ball cannot cost 10 cents as then the bat will cost $1.10, making a total of $1.20; so the ball must cost 5 cents and the bat $1.05. The demonstration of such biases deriving from Type 1 processing led to the erroneous belief that Type 2 deliberative processing was necessarily 'better' than Type 1 processing. This idea has been debunked. For example, expert decision-making has been shown to depend greatly on intuitive, rather than just reflective, thinking (Gigerenzer, 2007; Gladwell, 2005). And sometimes explicit reasoning can lead to worse performance (Reber, 1993; T. D. Wilson & Schooler, 1991).

Dual process reasoning has not previously been applied to the area of lexical inferencing, but perhaps only because of a misunderstanding about how communication works. For example, Evans notes that one cannot apply Type 1 processing to *any* au-

tonomous process, such as 'those which enable visual perception and language comprehension' (Evans, 2018, p. 142). Yet, understanding language is not solely an autonomous process and Type 2 lexical reasoning processes are still available to us. For example, with technical terms such as *dysgraphia* an educated hearer might use morphological analysis to understand the term (dys- [faulty, inability] -graph- [writing] -ia [disorder]). But the ordinariness of LOOTs makes this less likely to happen. One of the most common Type 1 processes in such cases, and possibly our main means of acquiring new lexis in naturalistic contexts, is contextual inferencing (Li, 1988), or the reliance on meaning in the surrounding context to predict the meaning of the target word. In language acquisition, this is a very gradual process of intuitively piecing together the contextual evidence for the meaning of a word or phrase (rather like the accumulated contextual evidence leading to an experienced driver's ability to assess the possibility of overtaking under normal driving conditions). But even in individual cases of lexical inferencing, the hearer might successfully arrive at the correct meaning. For example, a hearer not knowing the LOOT loan word *bambino* will probably ascertain the meaning in the following utterance:

> It's the school holidays next week, so I'm taking the **bambinos** to Italy.

Indeed, very few words and phrases in our own language are explicitly learnt through looking in dictionaries, asking for the meaning, or performing structural analyses.

Another fundamental Type 1 lexical inferencing process is to rely on network relations within a 'lexical neighbourhood' in our mental lexicons. Mental lexicons are not organized like dictionaries, but rather as networks of words that are linked phonologically, semantically and grammatically. These networks form 'lexical neighbourhoods' (Vitevitch & Luce, 2016), though the place metaphor 'neighbourhood' is a little misleading since the elements are not necessarily stored in the same place in the brain. In experimental studies of speech processing, phonological neighbourhoods are defined by a one-phoneme metric, according to which the neighbour of a target word is established by the addition, deletion or substitution of a single phoneme e.g. *sit* will have amongst its neighbours *fit*, *sat*, *sin*, *sipped* and *_it* (Greenberg & Jenkins, 1964). There is a well-established consensus that similar-sounding words in a phonological neighbourhood are simultaneously activated and that there is competition for recognition among these words (Marslen-Wilson, 1987; Norris & McQueen, 2008). Psycholinguists have mostly been interested in the dense neighbourhoods of monosyllabic words and in the 'neighbourhood effects' of such density, such as 'slips of the ear', in which the listener misperceives a correctly pronounced word (Bond, 1999). But in our present case we are interested in what happens when there is no representation of the target word, phrase or word sense in the hearer's memory. Mostly we are interested in words of two or more syllables with sparser neighbourhoods, but it is what is in those neighbourhoods that is vital.

When the brain autonomously searches for the meaning of an unknown word or sense, then, it might explore the lexical neighbourhood for phonological, orthographical or sense similarity. This attempt to infer the meaning of a target word by associating it with similar words has been called an 'associating' strategy in the literature on lexical inferencing (Hu & Nassaji, 2014), but there has been no attempt to explain this strategy in terms of reasoning processes. Consider a phrase I have only recently come across: *rustle my jimmies*. This apparently emerged as an Internet meme in around 2010 and

trended until about 2015 (Jimmies, 2024). I had no idea what *jimmies* meant or what rustling them would do. I came across the term out of context and so was not able to rely on contextual inferencing.[2] Yet I still had an intuitive sense that it might mean 'annoy'. Although not explicitly aware of this, my mind probably rapidly, intuitively and autonomously searched the lexical neighbourhood of *rustle*. This produced an immediate success: *ruffle* is a phonological neighbour of *rustle*, with just a one-phoneme substitution of /f/ for /s/:

| | |
|---|---|
| Rustle | /rʌsəl/ |
| Ruffle | /rʌfəl/ |

At this point, another Type 1 linguistic process might have kicked in: collocational prediction. Given the adjective *torrential*, we intuitively predict the next word to be *rain*. Given the word *ruffle*, we are likely to intuitively think *feathers*. This can be proved empirically: by far the most frequent and strongest collocate of *ruffle* is *feathers*. In the 52-billion word enTenTen21 corpus on the Sketch Engine corpus manager (Kilgarriff et al., 2014), *feathers* (18,953 co-occurrences) occurs twice as frequently as *hair* (9,721 co-occurrences) and in terms of collocational strength (the likelihood of the node word predicting the collocate), *feathers* has a LogDice score of 10.58 to *hair*'s 6.13. The reason for *ruffle* so strongly predicting *feathers* is the common expression *ruffle one's feathers*, which means annoy. Thus there is a very strong intuitive lexicogrammatical pattern connecting the two phrases:

| | | | | |
|---|---|---|---|---|
| Verb /rʌ_əl/ | + | my | + | 2-syllable plural noun |
| *Rustle* | + | *my* | + | *jimmies* |
| *Ruffle* | + | *my* | + | *feathers* |

This is probably not coincidental. It is likely that whoever coined the meme *rustle my jimmies* was, consciously or subconsciously, basing it on the pattern of *ruffle my feathers*. So, in this case, Type 1 intuitive reasoning allowed me to arrive at the correct answer (in a barely noticed instant). I still did not know what *jimmies* meant but, as work on formulaic language has long demonstrated (Wray, 2002), we tend to store multiword items as single units of meaning in any case (think Rice Krispies). So, in terms of communication, rather than linguistic curiosity, it was sufficient that I understood that *rustle my jimmies* meant 'annoy'.

But what if the answer does not come intuitively to the hearer through Type 1 processing? At that point they might use Type 2 reasoning by explicitly analysing the meaning of the components of the phrase (this equates with explicitly analysing the morphological components of a single word to establish its meaning). Someone from New England might have the following information in their lexical entries for *jimmy* and *jimmies*:

[2] It was suggested by an MA student on a Research Experience project as an example of a phrase designed to test AI chatbots' handling of potentially offensive words or phrases.

> *Jimmy* (noun): a short crowbar
> *Jimmies* (plural noun): 'tiny rod-shaped bits of usually chocolate-flavored candy often sprinkled on ice cream' (Merriam-Webster 2025)

*Jimmies* are sprinkles, or *hundreds and thousands* in British English. The cotext *rustle my* should preclude the first sense of *jimmy* in the plural (*rustle my crowbars). Sprinkles, on the other hand, are small enough to be rustled ('shake, stir, or otherwise move, causing a rustling [soft, muffled crackling] sound to be produced' OED 'rustle'). However, even knowing this, it is still not clear that the New Englander will arrive at the meaning 'annoy' without making the lexical association with *ruffle one's feathers*. Furthermore, New Englanders might also bring false etymological knowledge to the word *jimmies*: there is a persistent and widespread belief in the region that the word is racist and derives from Jim Crow, even if, as a Snopes fact check concludes (Mikkelson, 2009), 'no substantive evidence demonstrates anything denigrative of African-Americans was tied to the origin of the name.'

So, in the case of *rustle my jimmies*, Type 1 autonomous reasoning, through contextual inferencing, phonological and semantic neighbourhoods and collocational prediction, leads us to the right answer with respect to the meaning of a phrase we have not come across before. On the other hand, Type 2 deliberative reasoning using working memory can lead us astray if we lack relevant entries for key components in the phrase or if we apply false folk etymologies.

## 5. Offence-Adjacent LOOTs

While searching the lexical neighbourhood is often a successful Type 1 lexical inferencing process, where a LOOT is phonologically or graphically similar, or identical, to a recognised offensive term or sense in our mental lexicons, this process, like the intuitive mathematical process in the bat-and-ball problem, is likely to mislead.

Let us return to the archetypal offence-adjacent LOOT *niggardly*, as in 'I was kept on a niggardly allowance' (J. Aiken, *Blackground* 1989, in OED 2003 'niggardly' adj.). *Niggardly* derives from *nig*, a probable early Scandinavian word meaning 'parsimonious, stingy', which is related to *niggle* but completely unrelated to words deriving from post-classical Latin *nigro* ('black') such as *nigritude*, *nigrosine* and the N-word. *Niggardly* is clearly a LOOT. In terms of ordinariness, it is an adjective with a highly productive and common derivational suffix (-ly), it is clearly non-technical and it is found in everyday genres such as novels, letters and speeches. Accordingly, the speaker is more likely to presume that the hearer will understand its meaning and the hearer is more likely to presume that they should be able to infer its meaning. In terms of occurrence, while it was fairly common in the 19th century, with a frequency of 0.7 words per million (Google NGram 'niggardly'), by the end of the 20th century, when the Howard case occurred and others followed, it had fallen in frequency to 0.08 words per million, and thus the typical LOOT frequency of OED Band 3 (0.01-0.1 words p/million). Given that, in 1970, *niggardly* still had a non-LOOT frequency of 0.2 words p/million, we can infer that, by 1999, the date of the Howard case, it was probably more widely known amongst older generations (who were accused of using it) and less by younger generations (who made the complaints). Despite (or perhaps because of) the ongoing dispute over the word, it has remained at around 0.08 words p/million ever since.

*Niggardly*, though, also has both a phonological and orthographic neighbour that is highly offensive: the N-word. By the one-phoneme metric of neighbourliness (Greenberg & Jenkins, 1964), it is *niggard* rather than *niggardly* that is the phonological neighbour of the N-word (/nɪgə/ (British) or /nɪgər/ (US)), through addition of the phoneme /d/ (/nɪgəd/ (British) or /nɪgərd/(US)). However, the derivational suffix -ly is highly productive so that, even if there is, according to the OED, no adjectival form for the N-word (*niggerly* is only given as an alternative spelling of *niggardly*), it is very easy for us to generate this adjective in our minds. Most psycholinguistic work on phonological neighbourhoods has been based on experiments with monosyllabic CVC (Consonant-Vowel-Consonant) words (e.g. cat/bat/rat etc.) so it does not take into account the productivity of morphological derivation (i.e. our capacity to generate derived forms even when these do not currently exist in the language). Orthographically, again if we take *niggard* as the true neighbour of the N-word, we do not quite have a next-door neighbour at one grapheme distance, but we have a neighbour two graphemes down, with a substitution of [e] for [a] and the deletion of [d]: *niggard* -> *nigger_*. Furthermore, this naughty neighbour is considerably more common and familiar to people than the LOOT *niggardly*. The OED gives the frequency of the N-word at 2.5 words per million, putting it in its frequency band 5. However, OED frequencies are based on the Google Books database and thus only written discourse that has been accepted for publication. Given the extremely well-known controversy over the term, the reluctance of publishers to publish the word, and the fact that taboo terms are always far more common in speech than writing (*fuck* was common in everyday speech long before it became acceptable to use it on late-night TV or publish it in some contexts), any published frequency figures almost certainly grossly underestimate people's familiarity with the term.

## 6. Offence Presumption

So we have an unfamiliar LOOT that is a phonological neighbour of an extremely familiar and highly offensive slur. If we are familiar with the LOOT, this is not a problem: as we know the word, there is no need for our language processing to search the lexical neighbourhood, just as the offensive term *cunt* is not usually brought to mind when producing the word *country*. But if we are not familiar with the LOOT, then, as with the bat-and-ball problem, a rapid autonomous Type 1 process (in this case, our searching of the phonological neighbourhood in our mental lexicons) is likely to lead to an erroneous result: that the word *niggardly* is highly offensive. Furthermore, as with the bat-and-ball problem, unless something triggers conscious Type 2 reflection, we will be liable to be confident that our answer is correct. In other words, we shall **presume** that it is correct unless evidence leads us to overturn that presumption (Walton, 2014). In the former case, that evidence might be the framing of the bat-and-ball question as a 'problem', which leads around one-third of undergraduate participants to stop and reflect on their initial intuitive response. In the latter case, it might be the context that enables many hearers to stop and reflect: is the speaker really likely to have used an offensive term when describing the budget in a financial speech? However, at least some hearers will simply presume that the unfamiliar LOOT is related to its more common offensive neighbour.

There is an even more hazardous state of affairs when the LOOT is not phonologically *similar* to a lexical neighbour but is *identical* in form to the offensive term. This

is the case when the same lexeme has both non-offensive and offensive senses but the hearer only has the offensive sense in the lexical entry in their lexicon. Take a diner sitting in a restaurant looking at the menu and expressing his disgust to his fellow diners:

> 'I hate Welsh faggots!'

If the diners are from Wales, they will probably have at least two separate entries for the noun faggot in their lexicons:

> Faggot: gay man (slur)
> Faggots: meatballs made from minced off-cuts and offal

Depending on age and other factors, they might also have some of the other eighteen senses listed by the OED for the noun, such as:

- a bundle of sticks for firewood
- a small bunch of herbs for seasoning a dish
- a derogatory name for a woman (sometimes man) e.g. *old faggot*
- a naughty or mischievous child e.g. *little faggot* (OED, *faggot* n.)

In the context of dining, though, they should automatically and intuitively select the meaning referring to the Welsh dish, which is not always deeply appreciated:

> 'The contorted faces had nothing to do with the indigestibility of the school dinner faggots.' (*Western* Mail 1 March 2004) (cited in OED, *faggot* n. sense I.5)

However, imagine a gay American visitor overhearing the utterance from another table. Like the Merriam-Webster online dictionary in its entry for *faggot* (*Merriam-Webster.com*), they will probably only have the slur in their lexical entry. In the absence of alternative senses for *faggot* in their lexicon, Type 1 processing will rush to the conclusion that the speaker is either racist or homophobic or both. The hearer will simply presume that they have heard a slur rather than a neutral term. That presumption could only be overturned by Type 2 reflection on evidence relating to the context and cotext (perhaps that the company seem polite and well-mannered and that 'faggots' are on the menu) that leads to the realisation that the speaker had a different referent in mind. As with all intuitive reasoning, then, things can go wrong with Type 1 lexical inferencing, and this can lead to serious misconstruals that result in a breakdown in interpersonal relations. It is where Type 1 reasoning goes wrong with respect to lexical inferencing that we end up with offence-presumptive LOOTs.

There is an obvious objection that needs to be raised at this point. Thanks mainly to prominent cases in the US at the turn of the century, *niggardly* has, as we have seen, become controversial and, at least in the US, might be going the way of other neutral terms that have become tabooified. As one columnist put it when the controversies over *niggardly* were at their height: 'Like it or not, the word is now radioactive; having defended it, no one can now use it — especially in racially mixed company — without raising the question of motives, which, however, few will dare voice' (Poniewozick, 1999). This may well be true, but two points need to be kept in mind. Firstly, it is not clear how far high-brow discussions of words like *niggardly* permeate the population at large. Secondly, it is quite possible that it is precisely the offence-presumptive nature of the LOOT that led to the initiation of the tabooification process in the first place. There is no suggestion in the case facts of the Howard case or other early *niggardly* cases that the claimants were objecting to use of an established taboo word. It was quite clearly the

*perceived* link to the N-word that was at issue. Indeed, many prominent Black figures at the time ridiculed the mayor for firing Howard: the distinguished Black scholar of the N-word, Randall Kennedy, described the incident as an 'infamous round of wrongheaded protest' (Kennedy, 2002, p. 96), while the then chair of the NAACP quipped that 'the Mayor has been niggardly in his judgment on this issue' (cited in Kennedy, 2002, p. 96). In other words, in the case of offence-presumptive terms like *niggardly*, it is precisely because they presume offence that they can eventually be perceived by the community at large as offensive.

Another point to bear in mind is the frequency of commonly-believed false etymologies, particularly with regard to putatively racist terms. We saw the case of *jimmies*. Another particularly notable case is that of *nitty-gritty*, as in 'getting down to the *nitty-gritty*, or the important details', which Allan and Burridge twenty years ago thought might have suffered the kiss of death. *Nitty-gritty*, unlike *niggardly*, is not an offence-presumptive LOOT. For a start, the N-word is a fairly distant phonological and orthographical neighbour: both words start with /n/[n] and one of the syllables in both words starts with /g/[g], but *nitty-gritty* has nine phonemes (/nɪtiː-grɪtiː/), of which only three (/nɪ__g/) are shared, and four syllables (nɪ-tiː-grɪ-tiː) of which only one is shared (nɪ). Orthographically, *nitty-gritty* has eleven graphemes of which only four are shared with the N-word ([ni___gr___]). Indeed, it is not even a LOOT. It is an ordinary term: as a rhyming compound, it is extremely unlikely to be perceived as technical and so the listener is likely to presume that they can infer its meaning. But, in terms of occurrence, it does not fall in the typical LOOT band of 0.01 to 0.1 words per million. In the enTenTen21 corpus, *nitty-gritty* occurs at a frequency of 0.28 words p/million, making it a clear Band 4 word, within which 'most words remain recognizable to English speakers and are likely to be used unproblematically in fiction or journalism' (OED Frequency). There is no indication that its use is restricted to a certain demographic. With the exception of being an ordinary term, then, *nitty-gritty* does not appear to share the features of an offence-presumptive term. That is because its claim to offensiveness does not derive from the lexical neighbourhood and a low occurrence but from a false etymology, that it was originally used to refer to the detritus at the bottom of slave ships. While the word is, according to the OED, 'of uncertain origin', its first attestation is dated to 1940 and there is no evidence that it is related to the slave trade. So while offence-presumptive terms like *niggardly* (at least originally) are mistaken as offensive through faulty Type 1 intuitive reasoning, false-etymology terms like *nitty-gritty* are mistaken as offensive through faulty Type 2 deliberative reasoning: someone somewhere must have worked out the false etymology even if the claim might subsequently be taken at face value.

## 7. Analyzing Offence-Presumptive LOOTs

At the core of offence-presumptive LOOTs, then, is the fact that they share, or are perceived to share, a phonological or sense 'neighbourhood' with an offensive term in our mental lexicons. Not being familiar with the target word or sense, the hearer defaults to an offensive phonological or sense neighbour. Accordingly, the lexical item is simply presumed by the hearer to be offensive because they are not aware of a non-offensive alternative. And it is only if contextual or cotextual evidence triggers Type 2 reflection in working memory that the hearer will overcome that presumption. In this section I discuss the three key characteristics of offence-presumptive LOOTs and explore some of

the ways that these can be operationalised in analysis. The first two features (ordinary terms and low occurrence) derive from the fact that they are LOOTs. The third feature (having an offensive neighbour) is fundamental to the offensive-presumptive nature of these terms.

### 7.1. Ordinary Term

The term must be ordinary enough for the hearer, despite being unfamiliar with it, to presume that they can infer its meaning, at least to the extent that it is offensive. Hearers would not generally presume to understand technical terms such as *nigrosin* (a black pigment) or *shagbark* (a tree) and so are unlikely to presume they are offensive. *Nigre* is phonologically and graphically very close to the N-word but it is not offense-presumptive since it is only used in the technical context of soap-making (to refer to the dark-coloured solution of soap and impurities 'which settles out from the pure soap during the final stage of the manufacturing process' OED 'nigre'). Features of form and the discourse context must therefore preclude a technical meaning that hearers/readers should not be expected to know. We can also exclude proper nouns as candidates for offense-presumptiveness: they are represented differently from common nouns in our mental lexicons (Proverbio, Mariani, Zani, & Adorni, 2009); they lack the rich semantic networks of common nouns; and they tend to just store the reference, which makes them more difficult to retrieve (Semenza, 2009). While proper nouns like Dick (Van Dyke), Coon (Carrie, American actress), Cockburn (town in Scotland), Dildo (town in Newfoundland) and Shiitake (mushroom) might be sources of Type 2 explicit humour, proper nouns will not be offence-presumptive unless they are not recognized as proper nouns (e.g Spotted Dick – a traditional British steamed pudding).

### 7.2. Low Occurrence

The term must be infrequent enough to be unfamiliar to many hearers, but frequent enough to be familiar to the speaker and for the speaker to presume familiarity. We vary enormously in our knowledge of words so, at a whole language level, this means that there must be a strong likelihood that the hearer will not be familiar with or understand the term. Commonly understood words and phrases with offensive neighbours, such as *country*, *cocktail*, *bitch* (of dogs), *ride*, *come* and *do*, are not offence-presumptive because the hearer can easily access the non-offensive sense. In these cases, linguistic or contextual cues are required to activate an offensive frame (see section 10). For example, a radio commercial for the UK supermarket chain Somerfield was banned by the industry regulator Ofcom for a husband's utterance to his wife:

> 'I've got nothing against faggots, I just don't fancy them'

Here the issue was not that *faggot* would be offence-presumptive in a US context but that the advertisers were deliberately drawing the attention of the audience, through double entendre, both to the neutral meaning (Welsh meatballs) and to the offensive meaning (gay slur). Furthermore, the double entendre on *fancy* contextually reinforces the double entendre on *faggots*: *fancy* means 'want' but, in informal British English, it also means 'be physically attracted to,' thus making explicit the double entendre of *faggots*.

### 7.3. Offensive Neighbour

The final feature of offence-presumptive terms, and the feature that sets them apart from other LOOTs, is that they must share a phonological/graphological or sense neighbourhood with an offensive term or sense. Aside from *niggardly*, some other LOOTs with offensive phonological neighbours include several N-word neighbours – *niggle*, *snigger*,[3] and *chigger* (a mite common in the US) – *fag end* ('the last and worse part of something'), *gyp* ('mock' as in 'give someone gyp'), *anise* (the plant with aniseeds) and *spick and span* ('neat', 'smart'). LOOT sense neighbours or homonyms include *chink* ('fissure' v 'Chinese'), *fag* (UK, 'cigarette' v 'gay'), *faggot* (UK, 'meatball' v 'gay'), *ejaculate*[4] ('utter suddenly' v 'eject sperm'), *diddle* ('cheat, swindle' v 'copulate'), *mooncalf* ('daydreamer' v 'intellectually disabled'), *fuzzy-wuzzy* ('overly sentimental' v 'any dark-skinned person'), *butters my biscuit* ('pleases me' v 'excites me sexually') and *grinds my gears* ('annoys me' v 'copulates with me'). Sometimes, the non-offensive status of offense-presumptive items cannot be established clearly. This is particularly the case with words in the neighbourhood of the N-word deriving from the *nigro-* root indicating 'black'. For example, the now archaic term *nigritude* could be used in contexts that were not in the least offensive e.g.

> "Our aged friends can well remember when the smoke of London was not equal to one-tenth of last year's nigritude." *Quarterly Review* 168: 372 (1889) (cited in OED 'nigritude')

But there are also contexts where the association with Black people is clear, as in the following lines from a satirical poem by Victorian poet Thomas Hood:

> We've scrubb'd the negroes till we've nearly killed 'em,
> And finding that we cannot wash them white,
> But still their nigritude offends the sight.
> (T. Hood c.1845 in Black Job in Works (1862–3) vol. VI. p.166)

Over time, such contextual associations will tend to lead to the avoidance of the neutral term. The OED, for example, tells us that the N-word used to be used in several non-offensive contexts to refer to, for example: black caterpillars, ladybird larvae and sea cucumbers (II.8.a-c); a steam-driven capstan used on riverboats in the US (II.9.a); a device used to hold and turn logs in a sawmill (II.9.b); a dark brown colour (II.11); and a screen or mask used to deflect or conceal unwanted light in film studios (II.12). I doubt that the word is ever still used in these contexts though.

In the following two sections, I shall consider a couple of examples, based on forensic casework, of what might be seen as offence-presumptive terms: a figurative verb (*kneecap*) and a catchphrase (*never mind the quality, feel the width*). These claims were withdrawn before they got to court and so, in order to protect the reputations of both claimants and defendants, I shall only refer to these cases in the abstract. Although the linguistic reports together amount to almost 200 pages of analysis, including detailed

[3] *Niggle* and *snigger* are at the low end of Band 4 according to the OED, though I suspect they are more common in speech (the frequencies in the British National Corpus Spoken are too low to be significant). This might partly explain why they appear to be less controversial than *niggardly*. *Snigger* contains the entire N-word within it both phonologically and graphically but the fact that it begins with /s/ rather than /n/ might make it less offence-presumptive.

[4] Although *ejaculate* is not offensive in itself, if used in contexts referring to exclamation, a hearer not knowing this sense might find it highly offensive e.g. 'Stop ejaculating while I'm trying to speak!'

analysis of the specific communicational contexts in which the claimed offence took place, since it is not possible to adduce the detailed case-relevant evidence, I will simply ask the reader to accept hypothetical scenarios based on the cases. At the core of these hypothetical scenarios is a situation in which the claimant has genuinely taken serious offence on hearing a word or phrase uttered by the defendant, who in turn is completely oblivious to the possibility that they might have said anything offensive. What is important here is not the rights and wrongs of the specific cases but the potential explanatory value of the proposed category of offence-presumptive LOOTs.

## 8. Presumptive Racism: Figurative *Kneecap* (v.)

In the first case, imagine two experts discussing disciplinary measures with respect to a professional infraction and the defendant, considering that the claimant is being too harsh, saying something like: 'Do you want [the profession's regulatory body] **to kneecap** the poor [professional].' Imagine also that the claimant, unbeknownst to the defendant, is from Ireland and that they feel a strong association between the verb *kneecap* and the Troubles in Northern Ireland in the 1970s and 1980s. The claim, then, is that use of the verb *kneecap* by the defendant is racist.[5] Unable to discuss the details of the case, I would ask the reader to simply *assume* for the sake of argument that there are no circumstantial or contextual linguistic details that point to racism, or rather xenophobia, beyond the denotational and connotational meaning of the verb.

### 8.1. Overview of the Meaning of the *Verb Kneecap*

The verb *kneecap* was coined to describe the shooting of a person in the knee as a form of extrajudicial punishment by paramilitary groups and was, according to the OED, first attested in a newspaper headline in 1975:

> Man 'kneecapped' in Carrickfergus (*Daily Telegraph* 12 August 1975)

During the Troubles, the verb will certainly have developed extremely negative connotations, even if those connotations do not appear to have made it into dictionary entries for the verb. Over time, though, the meaning extended figuratively to hyperbole, as in the following description of attempted tackles in football/soccer:

> He'd be running from the halfway line with defenders trying to **kneecap** him. (*The Daily Telegraph* 22 March 2013)

And then, most commonly, to metaphor:

> This comes as Iran's crude exports have been **kneecapped** by US sanctions, which were re-imposed in November. (*S&P Global Commodity Insights*. 16 May 2019)

We can observe a clear pattern of use through the form's collocations. Figure 1 shows the top 20 collocates, scored by Mutual Information (MI) (the mutual dependency of one word on the other), for the word form 'to kneecap' in enTenTen21:

[5] Strictly speaking, it would be xenophobic rather than racist but the claim was for 'racism'.

| | Word | Cooccurrences | Candidates | MI |
|---|---|---|---|---|
| 1 | Mueller | 8 | 186,343 | 11.76 |
| 2 | competitor | 6 | 358,250 | 10.40 |
| 3 | rival | 8 | 702,303 | 9.85 |
| 4 | Sanders | 5 | 449,772 | 9.81 |
| 5 | opponent | 6 | 880,717 | 9.11 |
| 6 | opponents | 6 | 903,791 | 9.07 |
| 7 | unions | 5 | 809,102 | 8.97 |
| 8 | Republicans | 7 | 1,194,057 | 8.89 |
| 9 | Trump | 14 | 3,401,506 | 8.38 |
| 10 | Obama | 7 | 2,579,886 | 7.78 |

| | Word | Cooccurrences | Candidates | MI |
|---|---|---|---|---|
| 11 | economy | 5 | 4,566,229 | 6.47 |
| 12 | efforts | 5 | 5,259,073 | 6.27 |
| 13 | ability | 6 | 6,929,890 | 6.13 |
| 14 | political | 8 | 9,647,962 | 6.07 |
| 15 | anyone | 6 | 8,580,772 | 5.82 |
| 16 | him | 31 | 44,951,310 | 5.80 |
| 17 | themselves | 5 | 8,602,781 | 5.56 |
| 18 | American | 7 | 15,348,465 | 5.21 |
| 19 | every | 7 | 23,336,672 | 4.60 |
| 20 | me | 17 | 58,332,008 | 4.56 |

**Figure 1. Top 20 Collocates of "To Kneecap" in enTenTen21 by MI Score (Tokens=761; Range=R1-3; Min. Freq.=5)**

We can see that 'to kneecap' collocates strongly with words indicating opponents (*competitor, rival, opponent(s)*) and specified (particularly US) political opponents (*Mueller, Sanders, Trump, Obama*), or opposing parties (*unions, Republicans*), all of whom have been metaphorically rather than physically kneecapped. The *economy, efforts* and *ability* have also been metaphorically debilitated. There are no Troubles-related collocations (though these do exist for the more historical form *kneecapped*). 'To kneecap', therefore, primarily appears to indicate an intention to figuratively debilitate opponents and it seems to have lost its negative connotations associated with the paramilitary groups that first developed the practice.

### 8.2. Ordinary Term?

*Kneecap* is clearly an ordinary term. The noun on which it is based is almost universally understood, unlike its medical synonym *patella*. The verb was coined via the very common derivational process of converting an ordinary noun to a verb (e.g. *to fax, to facetime, to studioproduce*). Although the verb initially referred to a very specific form of extrajudicial punishment, it was never used technically and was first attested in the media, probably reporting spoken usage within the community. It later extended in meaning both in the scope of its literal coverage (any deliberate wounding of the leg in any context) but also in its extension to figurative uses (hyperbole and metaphor). The verb appears to be found predominantly in the media, though a lack of sizable spoken corpora mean that it is not clear how extensive its use is in spoken language. In any case, the verb is ordinary enough for the speaker to presume that it does not need explaining and for the hearer to presume that they would be able to know or infer its meaning.

### 8.3. Low Occurrence?

Establishing the frequency of the verb *kneecap* is difficult because neither the OED nor enTenTen20[6] distinguishes between noun and verb uses of *kneecap*. Manual analysis of citations from searches of various corpora reveals, though, that the verb *kneecap* accounts for approximately one-fifth or one-sixth of the total occurrences of *kneecap*. If we divide the frequency of *kneecap* by one-fifth (the estimated proportion of the verb

[6] EnTenTen20 rather than 21 was used for this analysis since that was the latest TenTen corpus available at the time of the analysis.

kneecap to overall *kneecap* in enTenTen20), we arrive at a frequency of 0.068 per million, which places it firmly within the OED's Band 3 and at a similar frequency to *niggardly*. However, we then need to separate the literal and figurative senses. In a random sample of the occurrences of the verb *kneecap* in enTenTen20, 83% were found to be figurative (hyperbole or metaphor) and 75% specifically metaphorical. This might suggest that the figurative sense (the one that is supposedly offence-presumptive) should be more familiar to a claimant because the figurative sense has become the 'dominant' meaning of the verb while the literal sense is now 'subordinate' (Simpson, 1981). However, in the 1993 subcorpus of the English Broadsheet Newspapers 1993-2013 corpus (also on SketchEngine), 75% of the uses refer to kneecapping during the Troubles, thus suggesting that this was the dominant meaning at the time. If a claimant acquired their lexical entry for *kneecap* in Ireland in the 1980s or 1990s, then it is likely that the literal Troubles-related sense is still the one familiar to them, while the low frequency of the verb as a whole means that it is quite possible that they are unfamiliar with the now much more common figurative sense. So, although figurative *kneecap* is now far more frequent overall than literal *kneecap*, an Irishman of a certain age is perhaps going to be more familiar with the literal sense than the figurative one.

### 8.4. Offensive Neighbour?

Clearly we have a *faggot*-type situation here where the hearer may simply not have one of the senses in their mental lexicon and so defaults to the other sense. With *kneecap*, the same lexeme has extended in meaning from a literal sense (shoot in the knees) originally associated with the Troubles in Northern Ireland to a metaphorical sense (bring to its knees) unrelated in any way with the Troubles. The association of *kneecap* with Northern Ireland is clearly still alive within that nation, as is evident from the name of one of their leading (if controversial) bands: Kneecap. Furthermore, the only three references to literal *kneecap* associated with the Troubles in the large Nexis database of current news sources for the period 2020-22 are from Northern Ireland newspapers. Outside Northern Ireland, though, the corpus evidence suggests that the connection between the verb *kneecap* as a whole and the Troubles has been entirely lost.

The question, though, is whether the sense neighbour of the now-dominant figurative use, i.e. literal *kneecap*, can be considered offensive. On the one hand, rather than becoming taboo after the Troubles, the literal sense has become extended to encompass a wider range of crimes than simply paramilitary shooting in the leg. Furthermore, no dictionary marks up *kneecap* as derogatory either in its literal or figurative usage and there is no evidence of the verb *kneecap* being discussed as potentially derogatory in itself. Where terms are known to be offensive, such as the noun use of *cripple*, or believed by some to be offensive, it is generally not difficult to find discussion of this on the Web. On the other hand, for someone who lived through the Troubles, it is not difficult to see how the word might be considered offensive within their particular community. Moreover, that perception of offensiveness might make them more resistant to accepting a non-offensive figurative use of the verb. To the potential objection that the claimant should have realised that the verb must have been being used in a figurative sense, one can respond that almost any verb can be used in a figurative sense but, unless the verb is commonly used figuratively in similar contexts, it can be hazardous to do so. For example, consider if the defendant, instead of using figurative *kneecap*, had

tried to use *shoot* figuratively: 'Do you want them to *shoot* him?' This would have been inexplicable to the hearer because we would not use *shoot* figuratively in this context. So a claimant, lacking a non-offensive figurative sense in their mental lexicon, might apply the same negative connotations of the literal Troubles-adjacent sense of *kneecap* to a context where the more neutral metaphorical sense of *kneecap* is intended. Once again, Type 1 autonomous intuitive reasoning will lead the hearer to default to their perceived offensive literal sense rather than reflect more deliberatively on the more likely pragmatic meaning in the specific context.

In conclusion, the metaphorical sense of the verb *kneecap* is offence-presumptive because it is a Low-Occurrence Ordinary Term that shares a literal sense neighbour which, for a claimant who acquired the word in Ireland in the 1970s or 1980s, has strongly negative offensive connotations associated with paramilitary terrorists in Northern Ireland. Contextually, the defendant probably felt he was using a standardly available verbal resource for conveying his opposition to a proposed harsh form of punishment, while the claimant probably felt that his identity as an Irishman was being invoked offensively through use of an 'offensive' term indexing paramilitary brutality.

## 9. Presumptive Sexual Harassment: *Never mind the quality, feel the width*

The second case involves presumed sexual harassment in a business meeting. Imagine a male defendant, on seeing the agenda move from a section titled 'Quality' to a section titled 'Quantity', quipping 'never mind the quality, feel the width' and a female attendee being deeply offended by the utterance of that phrase.

### 9.1. Overview of the Meaning of the Catchphrase

The catchphrase derives from the popular British 'cultural clash comedy' (Vice 2021: 188) *Never Mind the Quality, Feel the Width* (1967-71) about a Jewish and Catholic tailor who go into business together. According to *Brewer's Dictionary of Phrase and Fable* (Ayto & Crofton, 2011), the title 'may have been a deliberate inversion of a cloth trade saying: 'Never mind the width, feel the quality", though I was unable to find any evidence of this supposedly original saying. On the other hand, contrasting quality with quantity using a phrase beginning 'Never mind the quality…' does precede the sitcom by many decades, as in this advice to nineteenth-century pianists from a music magazine:

> Big tone is the companion to technic. So it is big, **never mind the quality; only give us quantity**. (*The Etude*, June 1888)

Semantically, then, the phrase suggests that quantity is more important than quality. What the sitcom seems to have introduced, though, is an ironic reading so that, in contemporary pragmatic usage, the catchphrase is actually almost always suggesting that we should be prioritizing quality over quantity. This is probably why it seems to be most common in the Hansard archive of UK parliamentary debates (Hansard, 2024). For example:

> Most importantly—this point constantly escapes those who take a **never mind the quality, feel the width** approach to investment—the quality of British investment is vastly improved. (PM John Major, Autumn Statement, 23 January 1990, Commons)

> The way the Minister replied—which is obviously in her brief—was all about the numbers: **never mind the quality, feel the width**. We are talking about quality of life, integration and the chances that a person who gets status would have to thrive in the UK. (Baroness Ludford, Nationality and Borders Bill, 8 February 2022, Lords)

It also explains why it has been used in the titles of academic papers:

> Binns, A. & Potter, R. (1989) Improving the effectiveness of postgraduate supervision: **never mind the quality, feel the width**, *Journal of Geography in Higher Education* 13(2): 211-16.

The phrase is therefore used ironically, though in serious contexts, to critique the privileging of quantity over quality.

### 9.2. Ordinary Term?

The phrase consists in extremely common words and, unlike the technical legal formula *beyond reasonable doubt*, the words in the phrase are being used in their ordinary senses. The hearer needs to understand that the phrase is being used ironically, but this is not uncommon with catchphrases. Although it derives from a title, there is nothing in the phrase itself to indicate that it is the title of a sitcom. Also, the high productivity of the phrase indicates that it has in fact become an established catchphrase rather than just a reference to a sitcom. The examples below depend for their understanding on the reader being able to make the link with the original catchphrase, helped by the phonological (/wɪt/ > /wɪdθ/) and rhyming (/wɪdθ/ > /mɪθ/) near-neighbours:

> **Never mind the quality, feel the wit**. As Sir Noel Coward remarked in rhyme, "The plot of La Gioconda is apt to wander", a somewhat optimistic judgment: in fact, it does not even get out of the front door and on to the pavement. (*The Times* 4 May 1993)

> THE plethora of tributes and retrospectives which have attended their reunion suggests that the Velvet Underground were always as much a media phenomenon as they were a musical one. **Never mind the quality, feel the myth**. (*The Daily Telegraph* 3 Jun 1993)

There is no reason, then, for a hearer to presume that they have not understood the phrase. It uses everyday words, is clearly non-technical, is an obvious catchphrase and is highly productive.

### 9.3. Low Occurrence?

The occurrence patterns make it highly likely both that a hearer would not be familiar with the phrase and that a speaker would expect familiarity. In terms of unfamiliarity, the phrase does not appear in US contexts (there are no occurrences in the Corpus of Contemporary American English (COCA) or in the US domain of enTenTen20) so an American hearer is unlikely to know the phrase. Furthermore, where the phrase does occur in corpora, it ranges in frequency from 0.02 words p/million (enTenTen20 UK Domain), through 0.05 words p/million (British National Corpus; English Broadsheet Newspapers 1993-2021) to a range from 0.05 to 0.1 words p/million in UK Parliamentary debates (Hansard, 2024). In other words, the frequencies are all in the OED Band

3 LOOT frequency range (0.01 to 0.1 words p/million). Turning to expected familiarity, the productivity of the phrase when used in the media and on the web suggests a perception by writers that readers will be able to understand the root catchphrase that is being manipulated. Moreover, there is some evidence of age restriction. While the phrase seems to be most common in the Hansard records of UK parliamentary debates (Hansard, 2024), most of the speakers recorded as using the phrase are now in their 70s or 80s. This means they would have been in their teens or twenties when the sitcom aired in the late 1960s. There is no indication, on the other hand, of gender restriction: many of the users of the phrase in the Lords are baronesses. Accordingly, it is likely that speakers over a certain age would consider the phrase well understood.

### 9.4. Offensive Neighbour?

Extensive lexicographical and corpus searches revealed no offensive sense (even an emerging one) of the phrase. However, there is a recognized sexually allusive phrase in the lexical neighbourhood. Here it is not a phonological neighbour but a pragmatic one involving slightly complex semantic and grammatical relations:

> 'It's not the size that matters, it's what you do with it'

The first half matches semantically with 'never mind the width' and the second half (less successfully) with 'feel the quality'. Then *quality* and *width* need to be inverted:

> It's not the size that matters -> never mind the ~~width~~ -> quality
> It's what you do with it -> feel the ~~quality~~ -> width

Remember that this would all be occurring through Type 1 autonomous intuitive reasoning. It is also possible that an erroneous Type 2 reasoning process might come into play: it has been suggested to me that the mere mention of the sitcom is taboo because the show was supposedly so utterly sexist. However, if the show is now perceived as controversial, it is because of the stereotyping of religious identity (though it was praised by some religious organizations at the time for promoting religious tolerance), rather than excessive sexism. This was no *Benny Hill* (Gibbs, 2023).

In conclusion, although there is no linguistic evidence that the catchphrase *never mind the quality, feel the width* is offensive, its ordinary term status, low occurrence, distribution bias towards the elderly and exclusion of the US, and its sharing not a close but a perceivable neighbour with an attested offensive phrase (at least in the context of a business meeting) can perhaps explain how a claimant might have perceived the phrase to be offensive.

## 10. Priming as Activating an Offensive Frame

At this point we can tackle the last step in the argument. One question that arises in such forensic cases as *kneecap* and n*ever mind the quality, feel the width* (or the early *niggardly* cases in the US or *nitty-gritty* in the UK) is why they are not dismissed as simple misconstruals by managers, HR personnel and the police. One possibility is the legal stipulation that it is the claimant who decides whether conduct is 'unwanted' and the understandable fear, then, of challenging the claimant. Yet, in my experience, even informants completely unconnected with the legal cases tend to see offensiveness in these terms when that is raised as a possibility. Instead, it is possible that those

dealing with such cases are metalinguistically **primed** to activate an offensive framing for the linguistic context. With some genres such as sitcoms, comedy shows and adverts, it is the expectations of the genre that prime audiences to be extremely attentive to ambiguous meanings and double entendres, so that even mildly suggestive material can provoke strong humorous reactions. But explicitly raising awareness of a potential offensive meaning can also prime interlocutors to activate that meaning. When I have informally asked informants whether the phrase *never mind the quality, feel the width* might convey sexual innuendo, if they do not know the phrase they inevitably reply in a somewhat tentative affirmative ('I can see how it could'). They generally cannot put their finger on why that might be the case, just as someone hearing the phrase *rustle my jimmies* might think 'annoy' while not knowing quite why, since this is an instinctive Type 1 process. The more they dwell on it, though, the more likely they are to think that the phrase is offensive. Yet MPs and Lords would not use a phrase in Parliament that they believed might convey sexual innuendo. It seems clear that these informants, like the audience in a pantomime, are being primed to see innuendo.

An example of how one can become hypersensitized to sexual innuendo, and the dangers of this for misconstrual, came in my analysis of the sample concordance lines of *it is not the size that matters…*. With my attention heightened to sexual innuendo, I came across the following example from an interview (all items that could potentially activate a sexual innuendo frame are in italics):

> HME: Why is it important for providers to be part of AAHomecare?
> Steedley: The association is *impotent* if it doesn't have *members*. When *staff* is on Capitol Hill, we have to speak for our *membership*, and our *membership* has to be reflective of the industry. We need *members of all sizes*. ***It's not the size that matters***, but the voice and *passion*. We need *members* to get legislation passed, to get meetings with CMS, to get regulatory *relief*. Otherwise, the program *laid on top of us* will continue to *grow* without any *relief* in sight. (HMEnews.com)

After analyzing many other examples of the phrase that obviously conveyed sexual innuendo, it seemed too much of a coincidence that this passage could be so packed with words and phrases with what might be called Sexual Innuendo Potential without there being any attempt actually to convey sexual innuendo. Although the overall content did not seem to suggest innuendo, I even wondered whether this was from a satirical site. Yet the website HME News (https://www.hmenews.com/) is a very sober site for 'Business News for Home Medical Equipment Providers.' There is nothing in the interview or on the site itself that would suggest that there was any conscious effort to convey sexual innuendo. Indeed, I subsequently realised that, like some examples of *never mind the quality* found on its own, here *it's not the size that matters* almost certainly is not meant to recall the sexually connotative phrase (which requires some variant of 'do with it') at all. Returning to this example now, after a couple of years, it seems extraordinary that I could have seen sexual innuendo in it, but this demonstrates the power of priming.

## 11. Conclusion and Implications for Judging Linguistic Offence

In this article, I have argued that linguistic offence through offence-presumptive Low Occurrence Ordinary Terms (LOOTs) comes about when a hearer is not familiar with a word or phrase and intuitively uses Type 1 lexical inferencing strategies to guess its meaning. In doing so, they draw on an offensive neighbour (lexeme or sense) in their

mental lexicons and transfer the offensiveness of that neighbour to the neutral target word. Accordingly, they are subjectively genuinely offended but objectively the offence is mistaken, just as it would be if the claimant misheard what was actually said. It should be stressed that this is a theoretical argument rather than an empirical claim. While I have drawn on empirical evidence wherever I could find it, there are very clear empirical gaps that I have had to plug in purely theoretical terms. For example, while there is a great deal of psycholinguistic empirical work on lexical neighbourhoods in relation to the question of lexical access to items already in our lexicons, this work has not been extended to cases where the target items are missing from the lexicon and so have to be inferred. Similarly, dual processing theories have not as yet been applied empirically to lexical inferencing, which would seem to be an important avenue to research. The notion of LOOTs is more empirically robust and has been developed over years of casework. However, the precise effects of these LOOTs on offence presumptiveness is currently purely theoretical. It also should be acknowledged that for radical eradicationists like Stojnić and Lepore (2025), it would be sufficient for an offence-presumptive term to be articulated, even in the context of this article, for it to cause offence. However, as with N-word use eradicationism, this appears to have remained a very minority view with respect to the public at large, who tend to be contextualists when it comes to their views of taboo language (Jay & Janschewitz, 2008).

If the argument here is sound, a number of implications follow for forensic lexical analysis in general, for the particular case of judging linguistic offence, or public claims that a word or phrase has caused offence rising to the level of remedy-required harm (Heffer, *in prep*) and for the ethics of interpersonal communication more broadly (Heffer, 2020).

With regard to forensic lexical analysis in general, the lexical category introduced here of the LOOT (Low Occurrence Ordinary Term) appears to be central in cases where the meaning of words and phrases is in dispute. It occurs in trademark cases, where the issue is often whether the average consumer is likely to consider the trademarked term as an ordinary word or phrase in the language or only as a commercial brand, as in *Bambino* (Heffer, 2008). And it occurs, as we have seen, in cases of disputed offensiveness. Empirical work is required, though, to establish precisely what happens in terms of lexical inferencing when a hearer (or reader) encounters an unfamiliar LOOT. It would also be interesting to drill down into specific sociolinguistic variables to test my hypothesis that those who use LOOTs belong to specific discourse communities in which the frequency of the term is much higher than in the general language community (such as elderly UK politicians and policymakers for *never mind the quality…*) but also that hearers who are likely to be offended by these LOOTs also belong to specific communities (such as those who lived through the Troubles for *kneecap* or eradicationists for LOOTs adjacent to the N-word).

With respect to disputes over linguistic offence, claims still need to be judged against the 'linguistic facts' of how a word or phrase is currently being used across society, as evidenced in large corpora of the language. Dictionary entries and particularly etymologies are not sufficient for gaining a clear picture of those facts. We need to search large general reference corpora to gain some idea of how a word or phrase is actually being used. We also need to put our own ideological beliefs about taboo words to one side: a certain word might be considered taboo in a given highly restricted domain,

but it would not be reasonable for someone within that domain to expect others outside it to conform to their beliefs about the word. For example, while London's Metropolitan Police might have bought into the false etymology of nitty-gritty in the 1990s, it was unreasonable to severely and publicly admonish a UK government minister for using the word in a speech he gave to a police conference in 2002 (Hopkins, 2002). A quarter of a century later, the question is still far from settled: while Sky banned its sports journalists from using *nitty-gritty* in 2020, a year later the BBC rejected a complaint about the journalist Laura Kuenssberg using the word (Moore, 2021). *Nitty-gritty*, like *niggardly*, might eventually become taboo across society, at which point this will emerge from the corpus evidence and their etymology will no longer be considered a sufficient defence, but until such point one needs to be conservative in judging the taboo status of a word that is in dispute.

While it is right that legislation on workplace harassment in both the UK and the US makes clear that 'unwanted conduct' 'means "unwanted by the worker" and should be considered from the worker's subjective point of view' (EHRC, 2020), the law also requires, once a case reaches a workplace tribunal, that the claim be deemed 'reasonable' in the circumstances. Yet an enormous amount of time and money can be expended, and lives torn apart, before the reasonableness of the claim is taken into account. It would be much better if, as with patients presenting to A&E, a case could be 'triaged' by HR personnel using a heuristic like the RESPECT framework (Heffer, In prep) to filter out cases where a claim does not correspond to the linguistic facts and has no circumstantial elements to support it. Such triaging would require systematic reflection, via the heuristic, to overcome potential metalinguistic priming. It should be stressed that, in proposing such triaging, I am concerned, as with all researchers in this area, with helping reduce the prevalence of language-based workplace harassment. A study in *The Lancet* found that 1 in 10 employees reported experience with workplace bullying or harassment in the previous year (Bunce et al., 2024), and while language-based harassment only constitutes a part of that total, it still corresponds to too many cases. Seriously tackling harassment in the workplace, though, also means dealing with those cases where current procedures misfire. Like vexatious claims, mistaken claims of linguistic offence do not just seriously harm the unjustly accused, but they also undermine confidence in procedures for identifying and judging the claims and, if common, can lead to a backlash in society.

Finally, with regard to the ethics of interpersonal communication more broadly (Heffer, 2020), since communication can be hazardous and mistakes are going to be made, we need both charity and apology. Applying the principle of charity in interpretation (N. L. Wilson, 1959) means considering the most favourable contextual interpretation of a speaker's utterance before escalating a claim of offence. If the claim of offence is reasonable, it will withstand such steelmanning. Apologizing, and accepting the apology, on the other hand, is the standard way of repairing breakdowns in interpersonal communication. Even when the speaker knows that what they said is objectively not offensive, if it is clear that the hearer is subjectively offended, then an apology is warranted. At the same time, neither the hearer nor third parties should take that apology as an admission of guilt that then warrants escalation of the claim. The more one can avoid interpersonal harm or the escalation of claims of linguistic offence

based on offence-presumptive terms and false etymologies, the more one can focus on those many justifiable cases of grievance that warrant ethical or legal remedy.

## References

Allan, K., & Burridge, K. (2006). *Forbidden Words: Taboo and the Censoring of Language*. Cambridge: Cambridge University Press.

Ayto, J., & Crofton, I. (Eds.). (2011). *Brewer's Dictionary of Modern Phrase & Fable*. Online: Chambers Harrap.

Baddeley, A. (2007). *Working Memory, Thought, and Action*. Oxford: Oxford University Press.

Billig, M. (2013). *Learn to Write Badly: How to Succeed in the Social Sciences*. Cambridge: Cambridge University Press.

Bingham, T. (2011). *The Rule of Law*. London: Penguin.

Bond, Z. S. (1999). *Slips of the Ear: Errors in the Perception of Casual Conversation*. New York: Academic.

Bunce, A., Hashemi, L., Clark, C., Myers, C.-A., Stansfeld, S., & McManus, S. (2024). Prevalence and nature of workplace bullying and harassment and associations with mental health conditions in England: A cross-sectional probability sample survey. *The Lancet*, *24: 1147*.

Cameron, D. (1995). *Verbal Hygiene*. London: Routledge.

Coulthard, M. (2004). Author identification, idiolect, and linguistic uniqueness. *Applied Linguistics*, *25*(4), 431–447.

Derbyshire, J. (2002, September 17). Niggling doubts. *National Review Online*.

de Ruiter, J.-P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, *82*(3), 515–535.

Dickerson, D. (1999, February 5). The last plantation: The 'niggardly' scandal should teach whites to watch their language and blacks to toughen up. *Salon*.

Dinn, W. M., & Harris, C. L. (2000). Neurocognitive function in antisocial personality disorder. *Psychiatry Research*, *97*(2-3), 173–190.

Dynel, M. (2018). *Irony, Deception and Humour: Seeking the Truth about Overt and Covert Untruthfulness*. Berlin: Mouton de Gruyter.

EHRC. (2020). *Sexual Harassment and Harassment at Work* [Equality and Human Rights Commission].

Evans, J. (2012). Dual-process theories of deductive reasoning: Facts and fallacies. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning* (pp. 115–133). Oxford: Oxford University Press.

Evans, J. (2018). Dual Process Theory: Perspectives and Problems. In W. De Neys (Ed.), *Dual Process Theory 2.0* (pp. 137–155). London: Routledge.

Evans, J., & Over, D. E. (1996). *Rationality and Reasoning*. Hove, U.K: Psychology Press.

Evans, J., & Stanovich, K. E. (2012). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.

Gibbs, J. (2023). A modern version of restoration comedy? Double entendre, objectification, fearful men and rakes manqué in the television work of Benny Hill. *Critical Studies in Television*, *19*(4), 468–485.

Gigerenzer, G. (2007). *Gut Feelings: The Intelligence of the Unconscious.* New York: Penguin Books.

Gladwell, M. (2005). *Blink: The Power of Thinking without Thinking.* New York (NY): Little, Brown and Co.

Grant, T. (2013). TXT 4N6: Method, consistency, and distinctiveness in the analysis of SMS text messages. *Journal of Law and Policy*, *58*(1), 467–494.

Greenberg, J. H., & Jenkins, J. J. (1964). Studies in the psychological correlates of the sound system of American English. *Word*, *20*(2), 157–177.

Grice, P. (1975). Logic and conversation. In *Syntax and Semantics 3: Speech Acts* (pp. 41–58). New York: Academic Press.

Gumperz, J. J. (1993). Types of linguistic communities. *Anthropological Linguistics*, *35*(1/4), 130–142.

Haastrup, K. (1991). *Lexical Inferencing Procedures or Talking about Words.* Tübingen: Gunter Narr.

Hansard. (2024). *Hansard Online Archive.* Retrieved from https://hansard.parliament.uk.

Hansen, S. J., McMahon, K. L., & De Zubicaray, G. I. (2019). The neurobiology of taboo language processing: fMRI evidence during spoken word production. *Social Cognitive and Affective Neuroscience*, *14*(3), 271–279.

Happé, F. G. (1995). Understanding minds and metaphors: Insights from the study of figurative language in autism. *Metaphor and Symbolic Activity*, *10*(4), 275–295.

Heffer, C. (2008). *The Perceived Meaning in the UK of the words "Bambino Mio".* Bambino Mio Ltd and Cazitex.

Heffer, C. (2020). *All Bullshit and Lies? Insincerity, Irresponsibility and the Judgement of Untruthfulness.* New York: Oxford University Press.

Heffer, C. (In prep). *Linguistic Offence: The RESPECT Framework for Evaluating Disputed Offensive Language.* Cambridge: Cambridge University Press.

Hopkins, N. (2002, May 15). Why nitty gritty has been ruled a no-no in the police lexicon. *The Guardian.*

Hu, H.-c. M., & Nassaji, H. (2014). Lexical inferencing strategies: The case of successful versus less successful inferencers. *System*, *45*, 27–38.

Hudson, R. (1996). *Sociolinguistics.* Cambridge: Cambridge University Press.

Jackendoff, R. S. (2002). *Foundations of Language: Brain, Meaning, Grammar, and Evolution.* Oxford: Oxford University Press.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen corpus family. In *7th International Corpus Linguistics Conference CL* (pp. 125–127).

Jay, T., & Janschewitz, K. (2008). The pragmatics of swearing. *Journal of Politeness Research. Language, Behaviour, Culture*, *4*, 267-288.

Jimmies. (2024). *That really rustled my jimmies.* Retrieved from https://knowyourmeme.com/memes/that-really-rustled-my-jimmies.

Juola, P. (2022). Google Books Ngrams. In L. A. Schintler & C. L. McNeely (Eds.), *Encyclopedia of Big Data.* Berlin: Springer.

Kahneman, D. (2013). *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux.

Kennedy, R. (2002). *Nigger: The Strange Career of a Troublesome Word.* New York:

Pantheon Books.
Kies, D. (1985). Some stylistic features of business and technical writing: the functions of passive voice, nominalization, and agency. *Journal of Technical Writing and Communication*, *15*(4), 299–308.
Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., … Suchomel, V. (2014). The Sketch Engine. *Lexicography*, *1*(1), 7–36.
Kurtić, E., Brown, G. J., & Wells, B. (2013). Resources for turn competition in overlapping talk. *Speech Communication*, *55*(5), 721–743.
Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
Li, X. (1988). Effects of contextual cues on inferring and remembering meanings of new words. *Applied Linguistics*, *9*(4), 402–413.
Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71–102.
McDonald, L. (2020). Your word against mine: The power of uptake. *Synthese*, *199*(1-2), 3505–3526.
McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, *14*(6), 648–652.
Mikkelson, B. (2009, September 25). Were jimmies named after Jim Crow. *Snopes*.
Moore, M. (2021, January 23). BBC rejects race complaint about use of 'nitty-gritty'. *The Times*.
Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395.
OED. (2025). *Frequency.* Oxford: Oxford University Press. Retrieved from https://www.oed.com/information/understanding-entries/frequency/?tl=true
Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology*, *20*(2), 139–156.
Pettit, M. (2016). Historical time in the age of big data: Cultural psychology, historical change, and the Google Books Ngram Viewer. *History of Psychology*, *19*(2), 141–153.
Pinker, S. (2007). *The Stuff of Thought: Language as a Window into Human Nature*. London: Penguin.
Poniewozick, J. (1999, February 2). The little N-word. *Salon*.
Proverbio, A. M., Mariani, S., Zani, A., & Adorni, R. (2009). How are 'Barack Obama' and 'President elect' differentially stored in the brain? An ERP investigation on the processing of proper and common noun pairs. *PLoS ONE*, *4*(9), e7126.
Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. Oxford: Oxford University Press.
Reddy, M. J. (1979). The conduit metaphor: A case of frame conflict in our language about language. In A. Ortony (Ed.), *Metaphor and Thought* (2nd ed., pp. 164–201). Cambridge: Cambridge University Press.
Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696.
Saussure, F. d. (1983). *Course in General Linguistics*. London: Duckworth.
Semenza, C. (2009). The neuropsychology of proper names. *Mind & Language*, *24*(4),

347–369.
Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.
Shannon, C. E., & Weaver, W. (1964). *The Mathematical Theory of Communication*. Chicago: University of Illinois.
Simpson, G. B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior*, *20*(1), 120–136.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22.
Snow, T. (1999, February 3). Linguistic lynching over 'niggardly'. *Des Moines Register*.
Stojnić, U., & Lepore, E. (2025). *Inflammatory Language: Its Linguistics and Philosophy*. Oxford: Oxford University Press.
Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.
Tanaka-Ishii, K., & Terada, H. (2011). Word familiarity and frequency. *Studia Linguistica*, *65*(1), 96–116.
Vitevitch, M. S., & Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics*, *2*(1), 75–94.
Walton, D. N. (2014). *Burden of Proof, Presumption and Argumentation*. New York (NY): Cambridge University Press.
Wason, P., & Evans, J. (1974). Dual processes in reasoning? *Cognition*, *3*(2), 141–154.
Wilson, N. L. (1959). Substances without substrata. *Review of Metaphysics*, *12*, 521–539.
Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, *60*(2), 181–192.
Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
Wray, A., & Staczek, J. J. (2025). One word or two? Psycholinguistic and sociolinguistic interpretations of meaning in a civil court case. *International Journal of Speech, Language and the Law*, *12*, 1–18.

*Language and Law / Linguagem e Direito*, *Vol. 11*(2), 2024, pp. 55-78
DOI: doi.org/10.21747/21833745/lanlaw/11_2a2

# The effect of linguistic detail in police interview transcripts on perceptions of an interviewee

James Tompkinson [1] and Kate Haworth [2]

[1] *Department of Language and Linguistic Science, University of York*, james.tompkinson@york.ac.uk
[2] *Aston Institute for Forensic Linguistics, Aston University*, k.haworth@aston.ac.uk

***Abstract***
*The question of what constitutes a "linguistically accurate" transcript is highly relevant for transcribers of police interview recordings. Does the inclusion of aspects of speech such as pauses, emphasis and laughter improve the quality of an interview transcript? Or does it make transcripts more difficult to interpret, and affect how interviewees are perceived? In this article, we explore whether the provision of a more linguistically detailed transcript made perceptions of an interviewee more aligned with the corresponding audio recording. Results showed mixed effects, and differences depending on the trait being perceived and the interview content. Our work suggests that the inclusion of additional linguistic detail in transcripts does not systematically make perceptions of interviewees more aligned with audio recordings, although there may be some scope for including the most salient features in transcripts. Our research illustrates the connected nature of speech perception and person perception in the context of evidential records of spoken interaction.*

***Resumo***
*A questão de saber o que constitui uma transcrição "linguisticamente exata" é altamente relevante para os transcritores de gravações de interrogatórios policiais. Será que a inclusão de traços do discurso oral como as pausas, a ênfase e o riso melhora a qualidade da transcrição de um interrogatório? Ou será que torna as transcrições mais difíceis de interpretar e afeta a perceção dos interrogados? Neste artigo, exploramos se o acesso a uma transcrição mais detalhada do ponto de vista linguístico torna as perceções de um entrevistado mais alinhadas com a gravação áudio correspondente.*

*Os resultados revelaram efeitos diversos e diferenças consoante o traço do discurso que estava a ser percecionado e o conteúdo da entrevista. No nosso trabalho verificamos que a inclusão de detalhes linguísticos adicionais nas transcrições não torna sistematicamente as perceções dos entrevistados mais alinhadas com as gravações áudio. Ainda assim, é possível haver alguma margem para a inclusão das caraterísticas mais salientes do discurso oral nas transcrições. A nossa investigação demonstra ainda a interligação da perceção do discurso e da perceção da pessoa no contexto de registos probatórios da interação oral.*

## 1. Introduction

Transcripts are a fundamental and widely used tool within legal systems around the world, and the process of transcription within the legal context has received a great deal of attention from both within and beyond the field of linguistics. A range of transcription contexts have been examined by linguists, from transcripts of covert recordings produced by expert phoneticians (Fraser, 2022; French & Fraser, 2018) to courtroom transcripts (Eades, 1996) and transcripts of police interviews (Harrington, 2024; Haworth, 2018; Haworth, Tompkinson, Richardson, Deamer, & Hamann, 2023; Richardson, Hamann, Tompkinson, Haworth, & Deamer, 2023; Tompkinson, Haworth, Deamer, & Richardson, 2023). The wide-ranging use, variety of contexts, and applications of transcripts leads Fraser (2022) to call for transcription to be treated as a dedicated branch of linguistics. At the most basic level, irrespective of its function or format, a transcript is a written representation of speech. In the absence of an audio recording, transcripts are one method of creating a permanent record of an otherwise momentary event, although a large proportion of transcripts are produced from either an audio or video recording. This makes transcripts different from other records of spoken interaction, such as notes taken by a police officer which may then be converted into a summarised written statement. In this article, we aim to contribute research around the question of whether, and how, police interview transcripts should represent information about 'additional' aspects of speech such as pauses, overlapping talk, laughter and emphasis. We specifically build on the research presented in Deamer, Richardson, Basu, and Haworth (2022) and Tompkinson et al. (2023) to illustrate the effect of excluding or including this type of information on social trait judgements of an interviewee. We focus particularly on the situation in England and Wales, where police interviews with suspects are audio recorded and then a subsequent transcript is produced from the audio recording (Haworth, 2018), rather than other jurisdictions where the term 'transcript' may be used as a catch-all term for any record of spoken interaction. The research presented in this article aims to contribute to the wider debate around the best way to produce transcripts of police interviews.

## 2. Backgroud

### 2.1. The transcription of police interviews

In England and Wales, there is a legal mandate for all police interviews with suspects to be recorded, following the introduction of the Police and Criminal Evidence Act 1984. In addition to the original audio or video recording, it is common practice for Record of Taped Interview (ROTI) or Record of Video Interview (ROVI) transcripts to be produced, with these transcripts used throughout the case for both investigative and evidential purposes (Haworth, 2018). The process is different for interviews with witnesses (see Rock (2020) for a detailed description of how witness interviews are conducted in the UK), and in this article we will therefore concentrate solely on the process for suspect interviews. Our focus is primarily the English and Welsh legal jurisdiction, although the findings could be applicable to other countries and legal systems. We acknowledge, however, that different legal systems might have different processes in place for capturing and recording police interview evidence. For example, Fraser (2022) discusses the police interview transcription process in the Australian legal system, which has some similarities to and some differences from the process in England and Wales.

In England and Wales, ROTI transcripts are predominantly produced by staff employed within individual police forces. Fraser (2022, p. 10) states that police interview transcribers frequently have "contextual understanding of police and legal processes in general, and sometimes of specific cases". However, ROTI transcribers are not routinely provided with linguistic training, nor do they have advanced legal knowledge (Haworth, 2018). ROTI transcripts also frequently contain summarised information (Haworth, n.d.; Richardson et al., 2023), with some police forces displaying preference for verbatim transcripts and others producing transcripts with some speech captured verbatim and some speech summarised (Tompkinson et al., 2022)[1]. This could lead to a question around whether 'transcript' is an appropriate name for ROTI and ROVI documents, if a transcript is to be viewed as an attempt to represent speech in a verbatim manner. A legitimate future question for linguistic research in this area is whether there is a need to separate the terms used for completely verbatim transcripts and those produced with a mix of verbatim and summarised information, given that a summarised record contains the 'voice' of the transcriber. In this article, we continue to use the term 'transcript' as a catch-all description of these documents, which serve as written records of spoken interaction for evidential purposes, while also acknowledging the difference between the two main forms that ROTI and ROVI transcripts take.

Haworth (2018) identifies that transcripts of police interview recordings with suspects are frequently treated as being a direct copy of, or substitute for, the original audio. However, there have been repeated warnings from linguists in recent years that this is a flawed assumption (Haworth, 2018; Fraser, 2022; Tompkinson et al., 2023), and that speech and writing are fundamentally different mediums which cannot be treated interchangeably. Haworth (2018) asserts that the routine practice of taking audio-recorded police interviews, converting them into transcripts, and then having a legal official such

[1]Tompkinson, J., Haworth, K., & Richardson, E. (2022). *For the record: assessing force-level variation in the transcription of police-suspect interviews in England and Wales.* Conference of the International Investigative Interviewing Research Group, Winchester.

as a barrister read the transcript aloud in court, is a problematic process and a form of "routine contamination" of evidence in the legal system.

Haworth (2018) lists six recommendations for the improvement of police interview transcripts, which include the recommendation that the practice of reading transcripts aloud in courtrooms should be abandoned, and a recommendation that people who are required to assess evidence in courtrooms (such as juries or other triers of fact) should be exposed to the original audio recording rather than being left to solely rely on a transcript without the accompanying audio recording. The research presented in this article aims to further investigate another of the proposed recommendations set out by Haworth (2018); that there should be a standard code of practice introduced for the production of police interview transcripts. Haworth (2018, p. 446) states that "this should include a set of standard transcription conventions, to cover features such as overlaps, pauses, and any areas of uncertainty". Haworth (2018) further asserts that more research is required in this area, specifically with respect to understanding which features could be reliably included and represented within a transcript. However, at the time of writing, there is no standardised national guidance for how ROTI transcripts should be produced or formatted in England and Wales (Tompkinson et al., 2023).

The research presented in this article was conducted as part of a wider research project exploring issues around the production of ROTI transcripts. For the Record is a research project led by the Aston Institute for Forensic Linguistics; an overarching summary of the project can be found in Haworth et al. (2023), with the broad aim of the project described as applying "insights from linguistics to improve evidential consistency in police interview transcripts" (Haworth et al., 2023). In addition to research into how interview records are currently produced (Haworth, n.d.; Richardson et al., 2023), the project also questions how well ROTI transcripts serve their current purpose as both investigative and evidential records, and asks what linguistic analysis can offer to improve current transcription processes. It is this last question that the research presented in this article aims to address in relation to the provision of linguistic detail in a transcript.

### 2.2. Producing and standardising transcripts

When attempting to define what a transcript is, Harrington (2024) explains that as a minimum, a transcript should provide details of who is talking and a verbatim record of the speech. Fraser (2022) warns that although a transcript is a written document, separating the process of transcribing from the process of writing is critical. This issue is further explored by Leemann, Perkins, Buker, and Foulkes (2024, p. 130), who highlight the difference between 'real speech' and reported speech in writing, stating that "fluent speech does not resemble the syntactically perfect utterances you might read as reported speech in a novel or as lines in a script for film or theatre." Fraser (2022, p. 2) warns that no transcript can accurately capture every aspect of an audio recording of someone speaking. This links to Bucholtz's 2000 assertion that a transcript can never be neutral, with the process of transcription inevitably requiring the transcriber to make a series of choices about what to include, what to omit, and how to represent aspects of speech in written form. This point is further emphasised by Leemann et al. (2024, p. 129), who describe the process of transcription as a "subjective event". Linked to this, both Fraser (2022) and Richardson et al. (2023) state that there can never be such a thing

as the transcript, only a transcript. This is an important consideration and a necessary reflection on the limitations of the process of transcription, but it nevertheless poses a challenge for the production of consistent evidential transcripts by institutions such as police forces, where there is a clear desirability for a lack of inter-transcriber variation in the production of written records.

Harrington (2024, p. 22) provides a call for more standardisation in transcription practices, arguing that "standardisation is one step towards the ultimate goal of better, more accurate, and more impartial transcripts being presented to juries". However, problems such as what kind of standards transcribers should adhere to, how these standards should be set and achieved, and how a balance could be struck to avoid the risk of standards and guidelines becoming too restrictive and therefore hindering the process of transcript production, remain. This issue is raised by Haworth et al. (2023), who highlight the importance of creating a robust research base upon which any recommendations for standardisation can be made.

### 2.3. What should be included in a transcript?

The form that any transcript takes will be highly dependent on its function and purpose. For example, a transcript of an interaction between a medical practitioner and a patient differs drastically in form and function from a transcript of an indistinct piece of audio which is used as evidence as part of a criminal case. However, there are some issues which will be applicable to transcripts produced in a wide range of settings. In a discussion of how expert phoneticians produce transcripts of indistinct or difficult audio recordings, Leemann et al. (2024) highlight a range of problems that a transcriber might encounter, including how aspects of speech such as pausing, overlap, non-standard grammatical constructions and dialect terms are dealt with. They also crucially highlight the importance of considering the effect that the inclusion, omission, and representation of such features can have on a reader, questioning whether including representations of this linguistic detail could confuse linguistically untrained readers of transcripts (Leemann et al., 2024, p. 130). This issue has also been explored for police interview transcripts, with Richardson et al. (2023) pointing out that in an analysis of a small corpus of ROTI transcripts, there was inconsistency in the way that these kinds of features were represented. Richardson et al. (2023, p. 28) state that "police transcribers are left to use their personal judgement on when to include HOW, in addition to WHAT, was said", and highlight the issues that a reader might face if they are confronted with a transcript which contains unclear or inconsistent representations of aspects of speech. Summarising this issue, Gibbons (2003, p. 30) argues that a transcript must be able to be easily understood by a reader in order to successfully fulfil its primary purpose as a representation of spoken language.

Richardson et al. (2023) highlight the possibility that transcribers could be trained to more consistently represent aspects of speech such as pausing, overlap, and disfluency features in their transcripts, while simultaneously cautioning that this should not happen without explicit consideration of the effect that this could have on readers. This idea is commented on in detail by Fraser (2022), who outlines a range of cautions around the use of Conversation Analysis (CA)-style representations of speech features in legally relevant transcripts. These highlighted issues include difficulties in training transcribers to consistently use CA-style representations, and the possibility that the overt represen-

tation of such features could actually mislead readers and cause them to misinterpret information.

The research presented in this article provides some initial experimental testing to explore the various arguments that have been made in relation to the provision of linguistic detail in transcripts of police interview recordings. In doing so, the work builds directly on the previous research presented in Deamer et al. (2022) and Tompkinson et al. (2023). Using stimuli from a single police interview, these studies showed that there were differences between perceptions of the interviewee depending on whether the interview was presented as a written transcript or as an audio recording.

Additionally, the study in Tompkinson et al. (2023) showed that the overt representation of a specific linguistic feature, silent pauses, resulted in an increased degree of perceptual instability on the part of experimental participants. Perhaps most important was the observation that although accurately marking pauses in transcripts theoretically brought the transcript closer to the original audio, this overt marking created some additional perceptual differences between the two modalities. This creates a specific problem for those tasked with producing police interview transcripts. One the one hand, the inclusion of additional linguistic detail, such as marking pauses, emotion, overlap and emphasis, could be seen as making a transcript more 'accurate', as it reduces the difference between the transcript and the original audio recording through the overt representation of linguistic features. On the other hand, if there is the possibility that the inclusion of such detail will enhance the perceptual distance between audio recordings and corresponding transcripts, then it could be argued that this kind of information should not be included. To more comprehensively address this issue, this article aims to extend the scope of the studies presented in both Deamer et al. (2022) and Tompkinson et al. (2023) and specifically examine how the inclusion and exclusion of additional linguistic features within transcripts can affect how listeners perceive an interviewee.

In their experiment, Deamer et al. (2022) opted to include representations of pausing, emphasis, overlapping speech and emotion markers within the transcript stimulus. The logic for doing so was to compare what the authors describe as a "best possible transcript" (Deamer et al., 2022, p. 29) to the audio recording. However, the study did not then compare these linguistically detailed transcripts with an orthographic transcript which may more closely represent the kinds of police interview transcripts produced by ROTI transcribers. This could also have affected the results within the Deamer et al. (2022) study. For example, the results showed no significant difference between the audio and transcript conditions in perceptions of *sadness*, but it could be assumed that this was because both the audio and transcript contained the relevant emotional markers which would indicate sadness. Indeed, in their call for a follow-up study, Deamer et al. (2022) identified the need to include a plain or orthographic transcript condition to broaden the scope of their findings. This three-way comparison between audio recordings, orthographic transcripts and transcripts containing some additional level of linguistic detail was addressed in Tompkinson et al. (2023), but only in relation to silent pauses. This again limited the scope of the previous research on this topic as other linguistic features were not considered.

Another limiting factor to the research presented in Deamer et al. (2022) and Tompkinson et al. (2023) was that both studies used a single clip from one publicly available

police interview recording. While this has the benefit of providing the high level of control required for exploratory research, it means that the findings are somewhat limited in scope. On this issue, Tompkinson et al. (2023, p. 47) highlight that "future work in this area should also focus on a wider variety of interviewees and interview situations, which would allow an assessment of the relative stability of the findings in this study". The two studies also do not assess whether there is inter-speaker consistency in perceptual judgements of police interview interactions. This is an important consideration which we address in this article, by exploring the relative stability of judgements of an interviewee in two separate audio clips and corresponding transcripts. By doing this, we aim to assess the relative contributions of speech and speaker to judgements of an interviewee.

### 2.3.1. Research Questions

As previously outlined, the research in this article specifically addresses the highlighted limitations of the work presented in Deamer et al. (2022) and Tompkinson et al. (2023). In this study, we aim to provide more detail in relation to the following questions:

1. What are the differences between orthographic transcripts, linguistically detailed transcripts and audio recordings of police interviewees in relation to social trait perception?
2. To what extent do differences in social trait perception between transcripts and audio recordings hold constant for different sections of a single interview recording?

### 2.3.2. Methodology

The stimuli for this study were taken from the West Yorkshire Regional English Database (WYRED), which contains mock police interview recordings as one of the speech elicitation tasks. We opted to use simulated police interview recordings from a publicly accessible research database for this article given the lack of available authentic police interview data for experimental work, to minimise the ethical challenges that are involved with working with real police interview data, and to increase overall researcher control of the data. There are, of course, limitations to working with mock police interview recordings which need to be acknowledged here. The main limitation is that the interactions do not represent the high-stakes and potentially high-stress environment of a real police interview. It is difficult to know the exact degree to which this was a problem for our study, but we considered it a worthwhile trade-off for the benefits that using mock recordings provided.

The full design and composition of the WYRED database is set out by Gold, Ross, and Earnshaw (2018). In the mock police interview task, participants were asked a series of questions about a fictitious crime, with a map used as an aid to elicit certain responses (Gold et al., 2018, p. 2749). The interview we selected consisted of an interaction between a male interviewee and a real life female police officer. We extracted two sections of speech from different parts of the interview. For Experiment 1, we used a sample of speech which was 2 minutes and 26 seconds in length, and for Experiment 2, we extracted a sample which was 1 minute and 59 seconds in length. We considered this to provide enough information upon which participants could evaluate the interviewee, but also be sufficiently short to minimise the risk of participants losing focus during the

experiment. One limitation of this approach is that listeners did not get the full context that would have been provided by a longer extract, but we judged that providing participants with extended audio clips could compromise the experimental design and the level of participant engagement with the study. For each section of audio, we created two versions of the transcript of the interview interaction. The first version was a plain orthographic transcript which contained only the words that were spoken and basic punctuation markers. The second was a more linguistically detailed transcript which specifically marked *pauses, emphasis, overlapping speech, whispered speech* and *laughter*. We aimed to mark these as accurately as possible with respect to the corresponding audio recording, using symbols which were drawn from both Conversation Analysis and previous analysis of representations of these aspects of speech in ROTI transcripts (see Richardson, Haworth, & Deamer, 2022 for a more detailed discussion of these issues). A transcription key was also provided at the top of each transcript so that readers knew what the various symbols represented. We also opted to represent these features in the speech of both the interviewer and the interviewee to avoid creating differences between representations of the two speakers. Again, however, we acknowledge this was another choice that may have affected overall perceptions of the interviewee in our experiment. This issue also relates to a broader point, linking back to Bucholtz's (2000) cautions about transcription choices. It is important to acknowledge that the style, format and choices that we made about representations could differ between transcribers, meaning our version is only one possible representation of the audio recordings. Table 1 shows the representations of the different features and the number of times each feature was marked in the transcripts used for Experiment 1 and Experiment 2.

| Feature | Representation | Number of occurrences in Experiment 1 transcript | Number of occurrences in Experiment 2 transcript |
|---|---|---|---|
| **Pause longer than 0.5 seconds** | **(X.X sec)** | 36 | 23 |
| **Pause shorter than 0.5 seconds** | **(.)** | 6 | 7 |
| **Emphasis** | **Underlined speech** | 18 | 18 |
| **Overlapping speech** | **[Words spoken in overlap]** | 6 | 8 |
| **Whispered speech** | **((whisper))** | 0 | 1 |
| **Laughter** | **((laughs))** | 1 | 1 |

**Table 1. Representations of features in the detailed transcripts, and the number of occurrences of each feature in the transcripts used for this experiment**

To further exemplify the differences between the respective transcripts, examples of the contrasting versions (plain orthographic and linguistically detailed) are displayed in

Figures 1 and 2. Figure 1 shows a section from a plain orthographic transcript, while Figure 2 shows the corresponding detailed transcript.

```
36      IR:   Okay. Could you explain then? We've got CCTV footage of
37            your car with you driving and it looks like a person is in
38            the passenger seat. Would you like to think again if- what
39            that could possibly have been? Or who?

40      IE:   When was this?

41      IR:   On Thursday evening leaving work.

42      IE:   Leaving work. I'm not too sure, I think I might have
43            stopped off or dropped someone off from work.

44      IR:   Okay so maybe a short journey with someone in the car?

45      IE:   Yeah yeah.
```

**Figure 1. Example of a section of a plain orthographic transcript**

```
49      IR:   Okay. Could you explain then. We've got CCTV footage of
50            your car (.) with you driving and it looks like a person
51            in the passenger seat. (1.25 sec). Would you like to think
52            again if- (0.5 sec) what that could possibly have been?
53            (0.8 sec) Or who?

54            (2.9 sec)

55      IE:   When was this?

56      IR:   On Thursday evening leaving work.

57      IE:   ((whisper)) Leaving work. (7.8 sec). I'm not too sure. I
58            think I might have stopped off or dropped someone off from
59            work.

60      IR:   Okay. So maybe a short journey [with someone in the car]?

61      IE:   [Yeah yeah].
```

**Figure 2. Example of a section of a linguistically detailed transcript**

In each experiment, we conducted a perceptual judgement task similar to the designs adopted by Deamer et al. (2022) and Tompkinson et al. (2023). A total of 300 participants were recruited via Prolific (https://www.prolific.com/) to take part in the experiments. 150 participants took part in Experiment 1, and 150 participants took part in Experiment 2. In each experiment, participants were provided with either the audio, the plain orthographic transcript, or the linguistically detailed transcript of the interview clip, and then asked a series of questions about the interviewee. Participants were

UK residents and UK nationals, between the ages of 18 and 70, and equally split between male and female raters. Participants were instructed to either listen to the audio or read the transcript they had been provided with, and provide a judgement of how *sincere, plausible, credible, relaxed, anxious, fearful, disgusted, surprised, happy, angry, sad, contemptuous, agitated, calm, panicked, friendly, cooperative, aggressive, defensive, assertive* and *nervous* the interviewee was. These were the same descriptors used by Deamer et al. (2022) and are based on Ekman's 1992 universal emotion categorisation system. These ratings were provided using a 1-5 Likert-style scale, with 1 representing "not at all…" and 5 representing "very…". Participants were also asked to state whether they thought the interviewee was telling the truth, with *"yes"*, *"no"* and *"don't'* know" possible outcomes. No participant was permitted to assess multiple conditions in the experiments. All data was collected using the JISC Online Surveys experimental platform. The study received Aston University ethical approval, and participants were paid for their participation in the research study at the standard approved Prolific rate.

### 2.3.3. Results

#### *Experiment 1*

Statistical analysis was performed using R software (R Core Team, 2024). Main effect p- values for the numerically judged traits in the experiment were calculated using Kruskal- Wallis significance testing. Post-hoc analysis was conducted using Benjamini-Hochberg adjusted Dunn tests, in order to determine the source and direction of differences within the data.

Statistical analysis of the responses to the question *"do you think the interviewee is telling the truth"* was conducted using a chi-square test. The output of this testing showed that there was no significant main effect of condition (audio, plain orthographic transcript, linguistically detailed transcript) on judgements of whether the interviewee was being truthful ($x^2$ = 7.50, df = 4, p = 0.11). However, post-hoc analysis of the comparisons between each of the conditions did show a significant difference between responses in the audio condition and those in the linguistically detailed transcript condition ($x^2$ = 7.17, df = 2, p = 0.03). The number of responses in each category for the question about whether the interviewee was telling the truth is shown in Figure 3, below. The data illustrates that although the number of respondents who judged the interviewee to be telling the truth was low in all conditions, there was a more notable difference between responses of 'no' and 'don't know'. Twice as many participants in the audio condition (n=21) answered 'don't know' when compared to participants in the linguistically detailed transcript condition (n=10). This suggests that the most negative judgements of whether the interviewee was telling the truth came from participants who read the linguistically detailed version of the transcript.

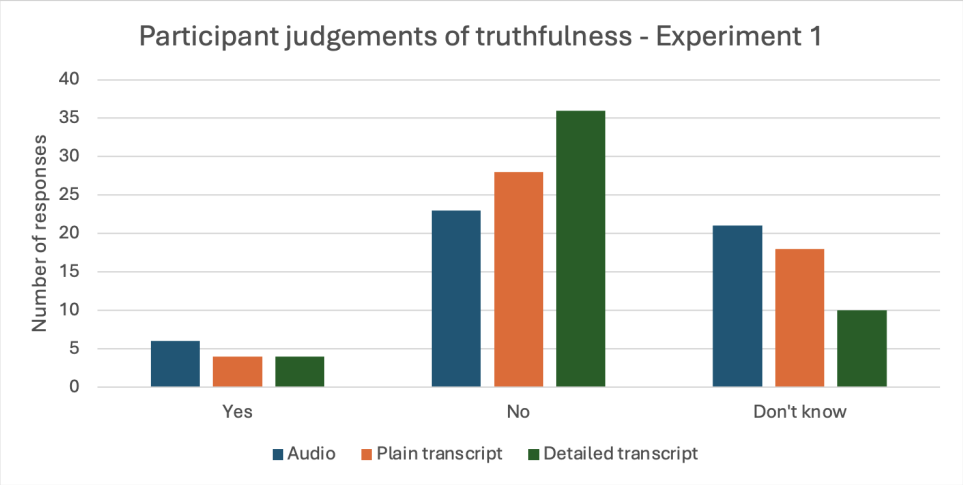



**Figure 3. Participant responses to the question "do you think the interviewee is telling the truth?" in Experiment 1**

Table 2, below, details the results of the main effect significance testing for the numerically rated traits in Experiment 1.

| *Trait* | *KW χ2 =* | *df=* | *p=* |
|---|---|---|---|
| ***Angry*** | 14.4 | 2 | **<0.001*** ** |
| ***Fearful*** | 14.3 | 2 | **<0.001*** ** |
| ***Panicked*** | 13.49 | 2 | **0.001**** |
| ***Plausible*** | 13.45 | 2 | **0.001**** |
| ***Sincere*** | 12.81 | 2 | **0.002**** |
| ***Anxious*** | 12.43 | 2 | **0.002**** |
| ***Nervous*** | 11.79 | 2 | **0.003**** |
| ***Sad*** | 10.32 | 2 | **0.006**** |
| ***Credible*** | 9.43 | 2 | **0.009**** |
| ***Relaxed*** | 8.82 | 2 | **0.01*** |
| ***Calm*** | 7.35 | 2 | **0.03*** |
| ***Aggressive*** | 6.46 | 2 | **0.04*** |
| ***Agitated*** | 6.06 | 2 | **0.048*** |
| ***Surprised*** | 5.92 | 2 | 0.052 |
| ***Cooperative*** | 5.68 | 2 | 0.058 |
| ***Contempt*** | 4.95 | 2 | 0.08 |
| ***Friendly*** | 3.84 | 2 | 0.15 |
| ***Disgusted*** | 3.72 | 2 | 0.16 |
| ***Defensive*** | 2.22 | 2 | 0.33 |
| ***Happy*** | 0.97 | 2 | 0.62 |
| ***Assertive*** | 0.31 | 2 | 0.86 |

**Table 2. Main-effect significance testing for Experiment 1**

The results in Table 2 show that for all traits apart from *surprised, cooperative, contempt, friendly, disgusted, defensive, happy* and *assertive*, there was a significant effect of modality on participants' perceptions of the interviewee. However, this analysis does not show where differences in the data occur. Table 3, below, shows the p-values from the post-hoc pairwise comparison testing and illustrates where significant differences

existed between each of the modality conditions in the experiment which showed a significant effect in Table 2.

| *Trait* | *Audio ~ Orthographic transcript* | *Audio ~ Detailed transcript* | *Orthographic ~ Detailed transcript* |
|---|---|---|---|
| ***Angry*** | **0.006**** | **<0.001***** | 0.15 |
| ***Fearful*** | 0.25 | **<0.001***** | **0.003**** |
| ***Panicked*** | 0.25 | **<0.001***** | **0.004**** |
| ***Plausible*** | 0.14 | **0.01*** | **<0.001***** |
| ***Sincere*** | 0.46 | **0.002**** | **0.003**** |
| ***Anxious*** | 0.47 | **0.003**** | **0.002**** |
| ***Nervous*** | 0.47 | **0.004**** | **0.003**** |
| ***Sad*** | 0.35 | **0.008**** | **0.005**** |
| ***Credible*** | 0.38 | **0.008**** | **0.01*** |
| ***Relaxed*** | 0.43 | **0.01*** | **0.01*** |
| ***Calm*** | 0.15 | **0.01*** | 0.07 |
| ***Aggressive*** | 0.09 | **0.02*** | 0.17 |
| ***Agitated*** | 0.30 | **0.03*** | 0.05 |

**Table 3. Post-hoc testing of significant effects in Experiment 1**

The data in Table 3 shows that there was a significant difference between perceptions of the interviewee in the audio and detailed transcript condition for every trait which had a significant effect in Table 2. This shows that for this interview section, providing more linguistic detail in the transcript did not result in perceptions of the interviewee being more closely aligned with the corresponding audio recording. For perceptions of how *angry* the interviewee was, there was no significant difference between the two versions of the transcript, but a significant difference between both versions of the transcript and the audio recording. This result aligns with the conclusion reached by Deamer et al. (2022) that a difference in modality (written or spoken) can create differences in perceptual judgements of an interviewee. This difference is illustrated in Figure 4, which shows the perceptual judgements of anger in all three conditions and the distribution of ratings provided by participants for each experimental condition. Although most participants rated the interviewee as not being particularly angry in all three conditions, there were far more ratings of 1 (the lowest possible rating) from participants who heard the audio recording compared with those who read either transcript.

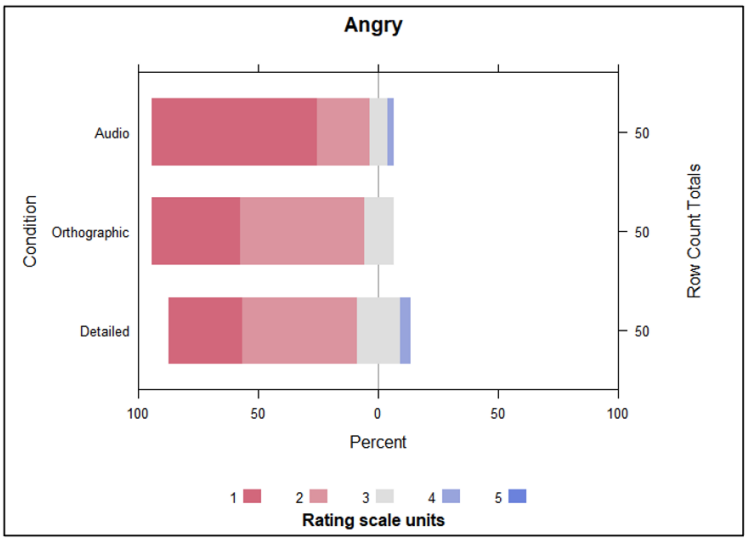



**Figure 4. Listener judgements of how angry the interviewee was in each condition**

Perhaps more notably, for judgements of how *fearful, panicked, plausible, sincere, anxious, nervous, sad, credible* and *relaxed* the interviewee was, the provision of linguistic detail in the transcript made perceptions of the interviewee significantly different to both the audio and plain orthographic transcript, but there was no significant difference between the audio condition and orthographic transcript condition. This means that perceptions of the interviewee were aligned in the plain transcript and audio conditions, but the provision of linguistic detail created more disparate judgements. These effects are illustrated in Figure 5, below.

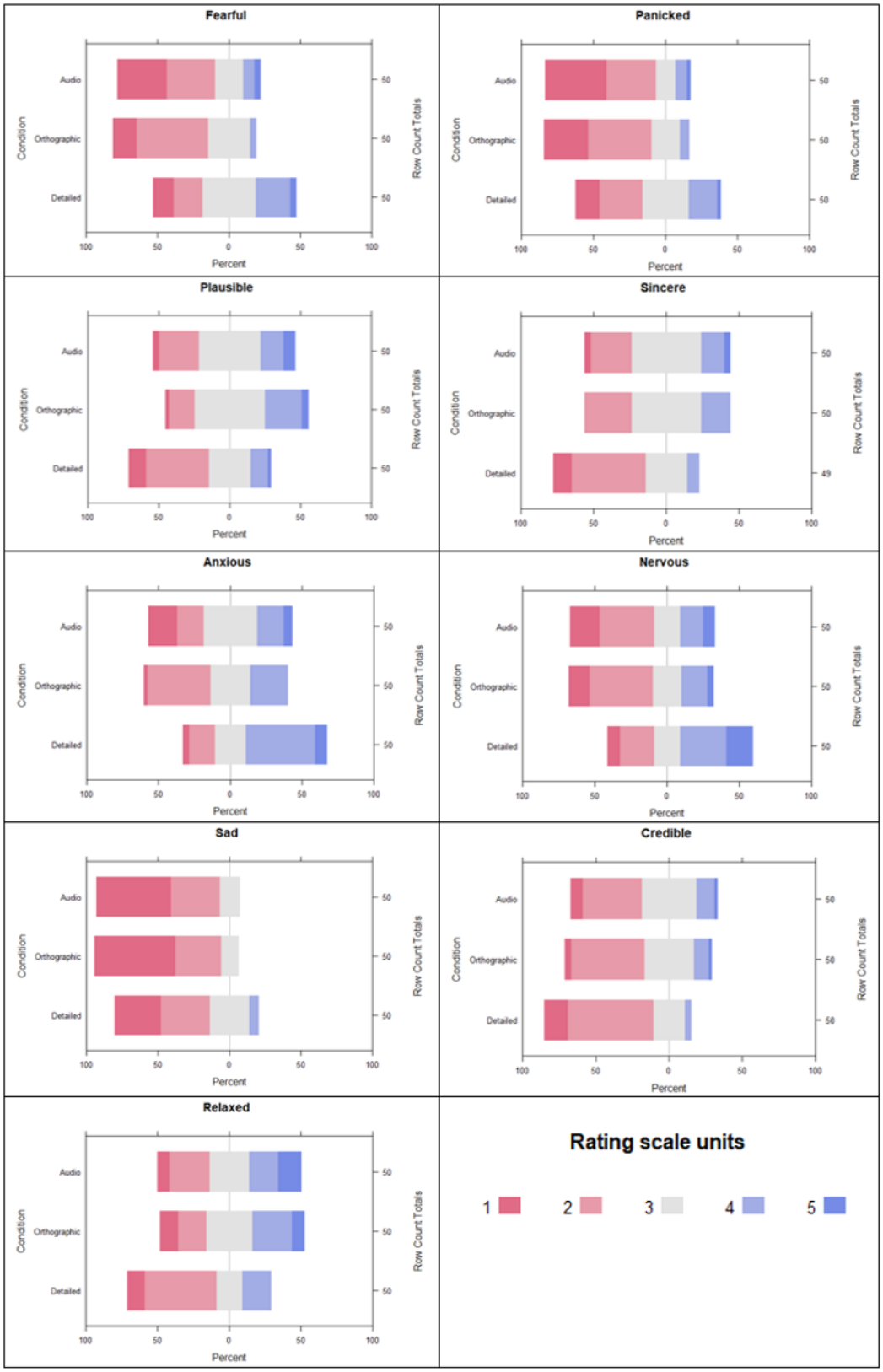



**Figure 5. Listener judgements of how *fearful*, *panicked*, *plausible*, *sincere*, *anxious*, *nervous*, *sad*, *credible* and *relaxed* the interviewee was in each condition**

Finally, for ratings of how *calm, aggressive* and *agitated* the interviewee was, the results showed a significant difference between the detailed transcript and audio conditions, but no other significant differences. Again, this illustrates that the provision of linguistic detail in the transcript created perceptual divergence from the audio rather than the alignment that might have been predicted if it is assumed that the provision of linguistic detail brings a transcript 'closer' to the corresponding audio recording. These effects are shown in Figure 6.

Another consistent effect across the judgement data for Experiment 1 was in the direction of the differences between the three conditions. For every significant trait shown in Figures 4-6, the inclusion of linguistic detail in the transcript created a more negative perception of the interviewee compared to the audio recording. The interviewee was judged to be more *angry, fearful, panicked, anxious, nervous, sad, aggressive* and *agitated* in the detailed transcript condition, and less *calm, relaxed, plausible* and *sincere*. This illustrates that for this interviewee in this section of the interview, the provision of linguistic detail not only created perceptual differences between the transcript and the audio recording, but that these differences were disadvantageous to the interviewee.

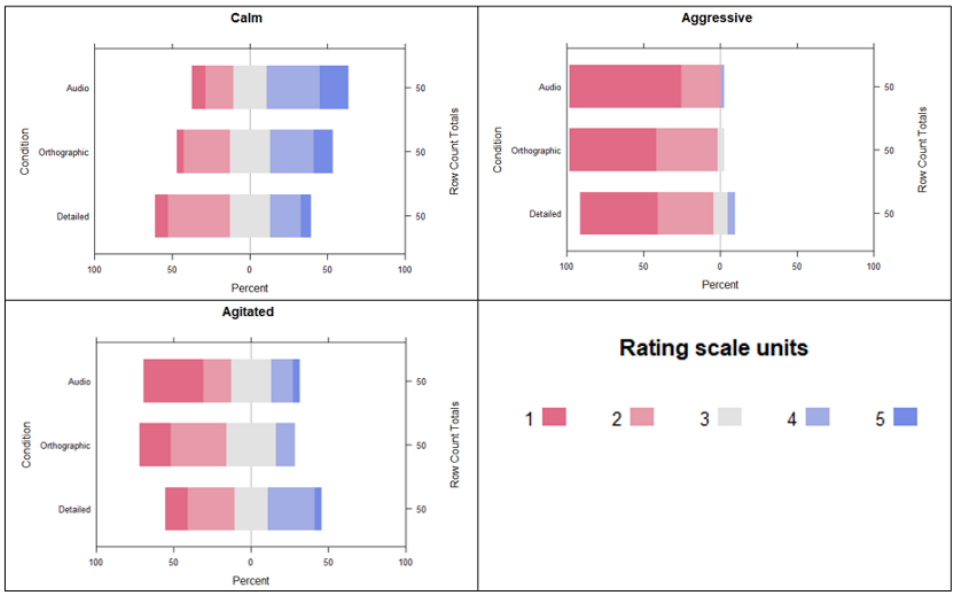



**Figure 6. Listener judgements of how *calm, aggressive* and *agitated* the interviewee was in each condition**

### *Experiment 2*

The research design in Experiment 2 mirrored that of Experiment 1 but used a different section of the same interview. This experiment was designed to test whether the findings from Experiment 1 would be repeated, or whether a different section of the same interview involving the same speakers would produce different results. The same experimental design was used for both experiments, with the same statistical testing procedure used for data analysis.

As in Experiment 1, there was no significant main effect of condition on judgements of interviewee truthfulness for Experiment 2 ($x^2$ = 2.213, df = 4, p = 0.70). The number of responses in each category for assessments of truthfulness is shown in Figure 7, below.

The data in Figure 4 highlights that across all conditions, participants predominantly judged the interview not to be telling the truth, with minimal differences between responses in the individual conditions. Post-hoc analysis showed no significant differences between any of the individual conditions for this question in Experiment 2.

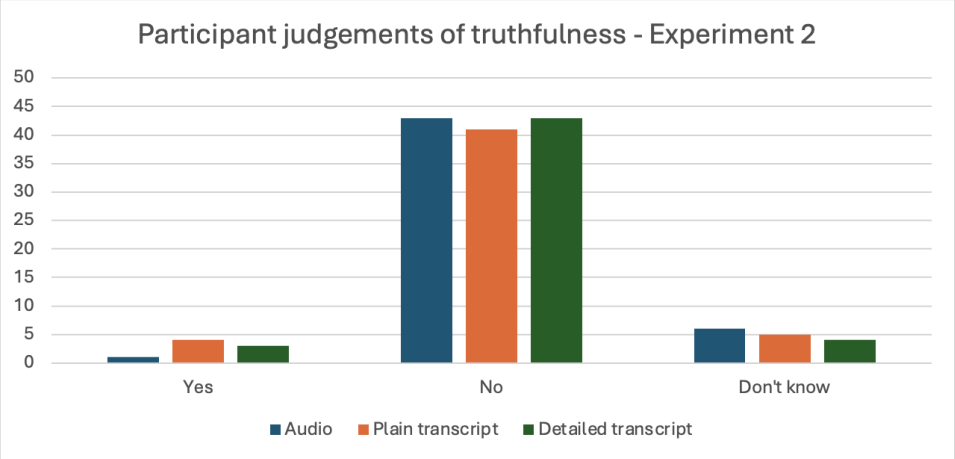



**Figure 7. Participant responses to the question "do you think the interviewee is telling the truth?" in Experiment 2**

Table 4, below, details the results of the main effect significance testing for the numerically rated traits in Experiment 2.

| *Trait* | *KW χ2 =* | *df=* | *p=* |
|---|---|---|---|
| ***Aggressive*** | 24.11 | 2 | **<0.001*** ** |
| ***Calm*** | 15.47 | 2 | **<0.001*** ** |
| ***Agitated*** | 13.09 | 2 | **0.001** ** |
| ***Defensive*** | 11.86 | 2 | **0.003** ** |
| ***Angry*** | 7.75 | 2 | **0.02*** |
| ***Relaxed*** | 7.35 | 2 | **0.03*** |
| ***Disgusted*** | 6.94 | 2 | **0.03*** |
| ***Surprised*** | 4.72 | 2 | 0.09 |
| ***Cooperative*** | 4.14 | 2 | 0.12 |
| ***Anxious*** | 4.06 | 2 | 0.13 |
| ***Fearful*** | 3.73 | 2 | 0.16 |
| ***Happy*** | 3.58 | 2 | 0.17 |
| ***Plausible*** | 3.16 | 2 | 0.21 |
| ***Contempt*** | 3.02 | 2 | 0.22 |
| ***Panicked*** | 2.72 | 2 | 0.26 |
| ***Assertive*** | 2.31 | 2 | 0.31 |
| ***Credible*** | 2.20 | 2 | 0.33 |
| ***Sad*** | 1.68 | 2 | 0.43 |
| ***Sincere*** | 1.60 | 2 | 0.45 |
| ***Nervous*** | 1.24 | 2 | 0.53 |
| ***Friendly*** | 1.07 | 2 | 0.58 |

**Table 4. Main-effect significance testing for Experiment 2**

The results in Table 4 show that there were comparably fewer significant differences between response conditions in Experiment 2 compared Experiment 1. There was a

significant effect of condition on judgements of how *aggressive, calm, agitated, defensive, angry* and *relaxed* the interviewee was. All other traits showed no significant effect of condition on judgements of the interviewee. Table 5, below, shows the p-values from the post-hoc pairwise comparison testing for traits which showed a significant main effect in Table 4.

| *Trait* | *Audio ~ Orthographic transcript* | *Audio ~ Detailed transcript* | *Orthographic ~ Detailed transcript* |
|---|---|---|---|
| ***Aggressive*** | **<0.001*** ** | 0.31 | **<0.001*** ** |
| ***Calm*** | **<0.001*** ** | **0.01*** | 0.06 |
| ***Agitated*** | **<0.001*** ** | 0.10 | **0.02*** |
| ***Defensive*** | **0.002*** | 0.28 | **0.006*** |
| ***Angry*** | **0.02*** | 0.41 | **0.02*** |
| ***Relaxed*** | **0.04*** | **0.01*** | 0.26 |
| ***Disgusted*** | **0.02*** | 0.48 | **0.03*** |

**Table 5. Post-hoc testing of significant effects in Experiment 2**

The most notable feature about the results in Table 5 is that they show different patterns to the results from Experiment 1. In Experiment 2, there were only two traits which displayed a significant difference in participant judgements between the audio and the linguistically detailed transcript conditions. For judgements of how *calm* and *relaxed* the interviewee was, there was a significant difference between the audio condition and both transcript conditions, but no significant difference between judgements in the two transcript conditions. These effects are shown in Figure 8, below. Figure 8 shows that the interviewee was judged to be more *calm* and more *relaxed* in the audio condition compared to the two transcript conditions. This aligns with the direction of difference shown in Experiment 1, where the interviewee was perceived more favourably in the audio condition than either of the transcript conditions.

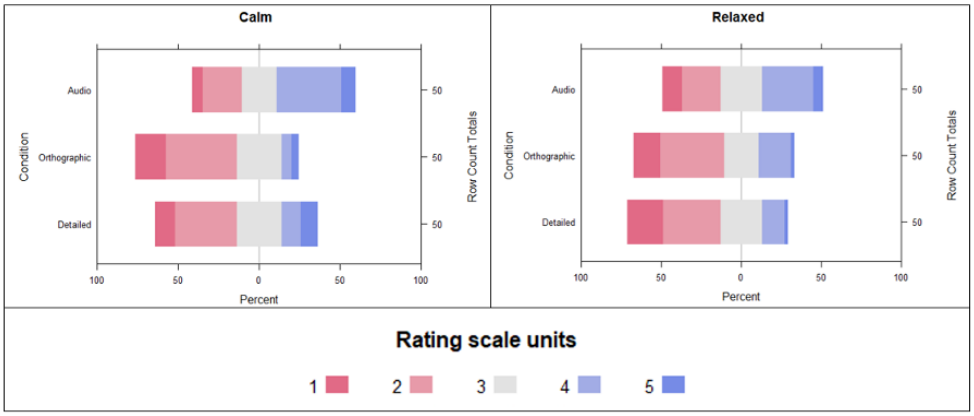



**Figure 8. Listener judgements of how *calm* and *relaxed* the interviewee was in each condition**

For judgements of how *aggressive, agitated, defensive, angry* and *disgusted* the interviewee was, the differences in judgements patterned in the direction that might be expected if the provision of linguistic detail made the audio recording and linguistically detailed transcript more alike. For these traits, there was no significant difference between the audio recording and the linguistically detailed transcript, a significant difference between the audio recording and the plain orthographic transcript, and a signif-

icant difference between the linguistically detailed transcript and the plan orthographic transcript. These effects are illustrated in Figure 9, below.

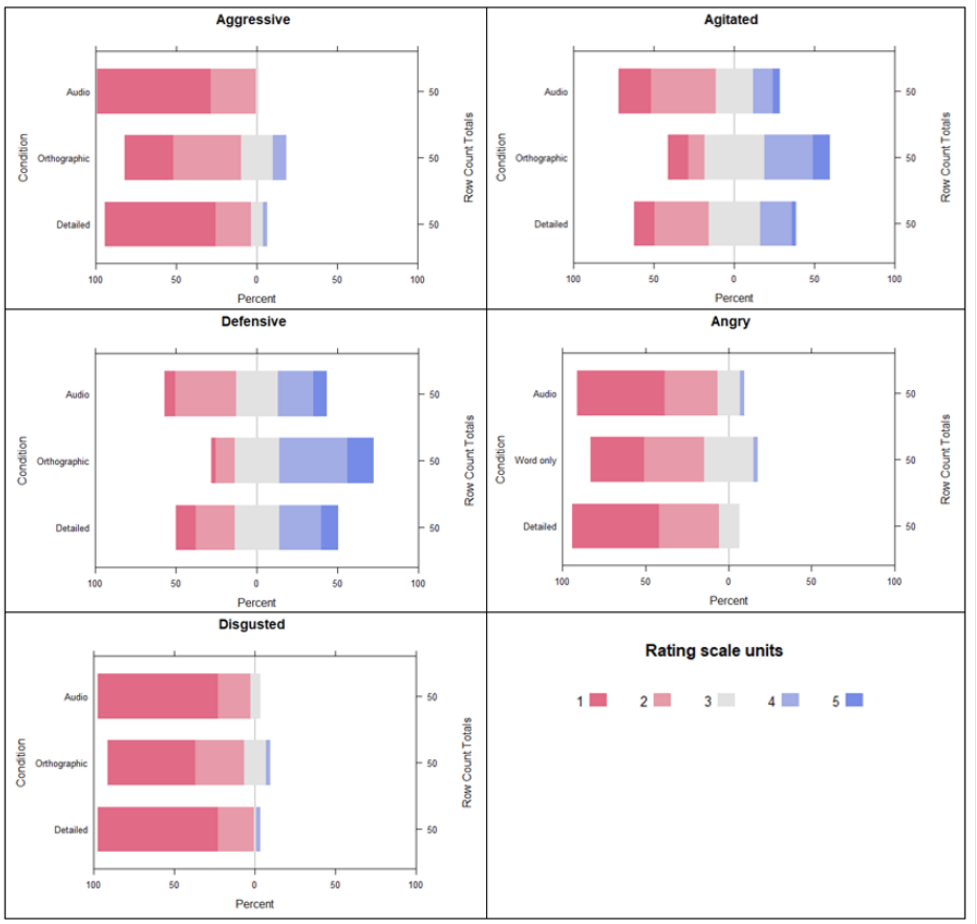



**Figure 9. Listener judgements of how *aggressive, agitated* and *relaxed* the interviewee was in each condition**

Similarly to Experiment 1 and the other significant effects from Experiment 2, the significant differences between the audio recording and the plain orthographic transcript all resulted in a more negative perception of the interviewee in the written modality. The interviewee was perceived to be more *aggressive, agitated, defensive, angry* and *disgusted* in the plain orthographic transcript condition compared to the audio condition. However, unlike in Experiment 1, the provision of the extra linguistic detail in Experiment 2 mitigated the effect and brought perceptions of the interviewee for these traits closer to the corresponding audio recording.

### *Qualitative analysis of judgements in Experiment 1 and Experiment 2*

It could be considered somewhat puzzling that two sections of the same interview, involving the same speakers, could produce such divergence in the patterning of perceptual judgements. There could be many reasons for the differences between Experiments 1 and 2, including random variation and perceptual instability in judgements of interviewees from audio recordings and transcripts, an interaction between the verbal content and the non-verbal information in the two interview clips, or differences in the degree to which participants used the non-verbal aspects of the interviews to inform their de-

cisions about what they were reading or hearing. Tompkinson et al. (2023) illustrated how participants were much more likely to specifically attend to, and base decisions on, pausing behaviour when pauses were marked in a transcript compared to being present in an audio recording. This led Tompkinson et al. (2023, p. 45) to argue that overtly representing linguistic features in transcripts can cause readers to pay more attention to aspects of speech that would have gone relatively unnoticed in the corresponding audio recordings.

Linking this to the results in the current experiments, it is possible that the degree to which participants noticed the features of speech in the audio recordings could explain why there was a difference between the results in the experiments. In both experiments, participants were asked to provide free-text responses to explain which aspects of the interviewee's language use influenced their judgements of sincerity, plausibility, credibility and truthfulness. Analysis of these responses provides an insight into which language features were influencing participants' judgements across the different conditions. Table 6, below, shows the number of times each of the features that were overtly marked in the detailed transcripts (*pausing, emphasis, overlapping speech, laughter* and *whispered speech*), were mentioned in the free-text responses for the different conditions in each experiment.

| | **Experiment 1** | | | | **Experiment 2** | | |
|---|---|---|---|---|---|---|---|
| | **Audio** | **Detailed transcript** | **Plain transcript** | | **Audio** | **Detailed transcript** | **Plain transcript** |
| **Pausing** | 7 | 39 | 1 | | 34 | 37 | 3 |
| **Emphasis** | 3 | 3 | 1 | | 3 | 8 | 1 |
| **Overlap/ interruption** | 2 | 2 | 0 | | 1 | 7 | 10 |
| **Whisper** | 0 | 0 | 0 | | 0 | 1 | 0 |
| **Laughter** | 0 | 0 | 0 | | 0 | 0 | 0 |

**Table 6. Number of mentions of different linguistic features in participants' free-text responses**

The data in Table 6 shows a clear difference between Experiment 1 and Experiment 2 with regards to the number of mentions of pauses as an influencing factor on listeners' judgements of truthfulness, sincerity, credibility and/or plausibility. In Experiment 1, participants in the detailed transcript condition attended to the presence of pauses far more frequently (n=39) than participants who heard the corresponding audio recording (n=7). However, in Experiment 2, a similar number of participants in both the audio (n=34) and detailed transcript (n=37) conditions mentioned pausing as having an influence on their judgements. This difference could potentially explain why the provision of extra linguistic detail in Experiment 1 created more perceptual divergence in listener judgements, whereas in Experiment 2, there was a greater alignment in judgements between the audio and detailed transcript conditions. Analysis of the overall length of the pauses in the speech of the interviewee across the two experiments also showed that in Experiment 1, there were no pauses that were longer than two seconds, whereas in Experiment 2, there were three examples of longer pauses of 2.9 seconds, 5.8 seconds, and 7.9 seconds respectively. This raises the possibility that listeners only attended to the

longer pauses in Experiment 2, and that these were the pauses that were more noticeable and influential on listener judgements. The placement of these pauses within the overall interaction could have also been important in shaping listeners' views. Figure 10 shows the three pauses discussed above, all highlighted in yellow. Two of the pauses come between a question and the answer (lines 57 and 67), and another comes within the middle of an answer given by the interviewee (line 60). Although it is impossible to predict or infer the effect of this on every listener, there was no comparable discourse structure in the extract of audio used for Experiment 1. The specific effect of this kind of linguistic issue could be tested in future research on this topic.

52 IR: Okay. Could you explain then. We've got CCTV footage of
53 your car (.) with you driving and it looks like a person
54 in the passenger seat. (1.25 sec). Would you like to think
55 again if- (0.5 sec) what that could possibly have been?
56 (0.8 sec) Or who?
57 (2.9 sec)
58 IE: When was this?
59 IR: On Thursday evening leaving work.
60 IE: ((*whisper*)) Leaving work. (7.8 sec). I'm not too sure. I
61 think I might have stopped off or dropped someone off from
62 work.
63 IR: Okay. So maybe a short journey [with someone in the car]?
64 IE: [Yeah yeah].
65 (0.8 sec)
66 IR: Can you think who that might have been maybe?
67 (5.7 sec)
68 IE: I think it was the new- the new lass at work.

**Figure 10. Extract from Experiment 2 showing the longer pauses in the transcript**

## 3. Discussion

Our first research question focussed on whether there would be differences between orthographic transcripts, linguistically detailed transcripts, and audio recordings of police interviewees in relation to social trait perception. The results in both experiments showed that there was at least the potential for a change in the format of the police interview evidence to affect perceptions of the interviewee, with differences shown across a range of traits in both interviews. This aligns with the findings of Deamer et al. (2022) and Tompkinson et al. (2023), which both observed differences in listener and reader judgements of police interview evidence depending on whether the perceiver was provided with a transcript of the interview or an audio recording.

However, in this study we have illustrated an additional issue. Our results indicate that there is at least the potential for the overt marking of additional linguistic features using a standardised CA-style system to make listener perceptions more divergent from the corresponding audio recording when compared to a plain orthographic transcript. This was particularly true in Experiment 1, where listeners paid minimal attention to the marked linguistic features in the audio condition, and where judgements were less favourable for the interviewee when participants were presented with a linguistically detailed transcript. This is clearly an undesirable outcome for police interview transcripts, and suggests that marking features in a transcript that might not be clearly perceived in the corresponding audio recording could be a problematic practice, even if those features are marked accurately. The results from Experiment 1 would support the proposition by Fraser (2022) that the provision of CA-style representations in legally relevant transcripts could serve to mislead readers and cause them to misinterpret information within a transcript.

However, the findings from Experiment 2 suggest that when listeners attend to the features that are marked in the transcript in the corresponding audio recording, then perceptions of the interviewee in both conditions can be more closely aligned, and crucially different from judgements made from a plain orthographic transcript. This adds a level of complexity to the discussion of the findings from Experiment 1, suggesting that more research is needed to better understand how listeners respond to different features and their realisations in transcripts. It appears that there may be a 'tipping point', at which the representation of non¬verbal features in transcripts becomes more of a hinderance than a help with respect to social trait perception. If this were the case, the transcripts used in Experiments 1 and 2 would sit on either side of this tipping point. This is particularly interesting for the interview clips used in this experiment as they were both taken from recordings of the same speaker, showing that it is possible for two sections of the same interview with the same speaker to produce different perceptual results. This relates to our second research question and suggests a lack of automaticity in assuming that the same speaker will always be perceived similarly by listeners of audio recorded interviews or readers of transcripts.

## 4. Conclusion

The research in this article has attempted to provide some empirical testing to contribute to the wider debate about whether it is beneficial for transcripts of police interviews to overtly mark aspects of speech such as pausing, overlapping speech, emphasis and other non-verbal cues such as laughter and whispering. By adopting an approach centred around social trait perception, the research in this study has illustrated how the different ways in which police interview evidence can be presented can contribute to judgements of the interviewee.

Although the research is small in scope, using two separate interview extracts from the same speaker, we have aimed to illustrate some of the complexities around the effect that overtly marking linguistic information in transcripts can have on how interviewees are perceived by listeners of audio recordings and readers of transcripts. Our results suggest that caution is required, and more research needed, before recommendations are made that non-verbal cues should be marked in police interview transcripts. The results of this study indicate that the inclusion of additional linguistic detail in transcripts did

not systematically make perceptions of interviewees more aligned with corresponding audio recordings, but also that there could be some features for some interviews which are worthy of representing. However, how this could be done systematically remains a significant challenge, and even if some features are worth representing in transcripts, it is still not clear *how* they should be marked and whether this kind of standardisation could ever be achieved in practice. This issue could be further complicated by the rapid development and potential appetite for the use of automatic transcription systems for the creation of police interview transcripts (Harrington, 2023). This could be a focus for further work in this area, but it would appear to be a significant potential source of variation in transcript production. There is also a wider range of potentially relevant 'additional' features which could be present in police interview recordings which are not considered in the current study as they did not appear in the audio recordings, such as crying or shouting. Again, this issue could be further explored in additional research in this area. Finally, future research could also explore the effect that the provision of video recordings of police interview interactions has on perceptions of interviewees, in comparison to audio recordings and transcripts. More broadly, the research in this article illustrates the connected nature of speech perception and person perception, and highlights some potential complexities regarding perceptions of interviewees in this important and legally relevant context. Given these complexities, we would argue that the initial recommendation by Haworth (2018), that people who are required to assess police interview evidence at any stage of the legal process should be exposed to the original audio recording rather than being left to rely solely on a transcript, is a cautious but sound basis on which to use this important form of evidence.

## References

Bucholtz, M. (2000). The politics of transcription. *Journal of Pragmatics*, *32*(10), 1439–1465. Retrieved 2025-08-13, from https://linkinghub.elsevier.com/retrieve/pii/S0378216699000946 doi: 10.1016/S0378-2166(99)00094-6

Deamer, F., Richardson, E., Basu, N., & Haworth, K. (2022). For the Record: Exploring variability in interpretations of police investigative interviews. *Language and Law=Linguagem e Direito*, *9*(1), 25–46. Retrieved 2025-08-13, from https://ojs.letras.up.pt/index.php/LLLD/article/view/12825/11681 doi: 10.21747/21833745/lanlaw/9_1a2

Eades, D. (1996). Verbatim courtroom transcripts and discourse analysis. In H. Kniffka (Ed.), *Recent developments in forensic linguistic* (pp. 241–254). Frankfurt: Lang.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3-4), 169–200. Retrieved 2025-08-13, from https://www.tandfonline.com/doi/full/10.1080/02699939208411068 doi: 10.1080/02699939208411068

Fraser, H. (2022). A Framework for Deciding How to Create and Evaluate Transcripts for Forensic and Other Purposes. *Frontiers in Communication*, *7*, 898410. Retrieved 2025-08-13, from https://www.frontiersin.org/articles/10.3389/fcomm.2022.898410/full doi: 10.3389/fcomm.2022.898410

French, P., & Fraser, H. (2018). Why "Ad Hoc Experts" should not Provide Transcripts of Indistinct Audio, and a Better Approach. *Criminal Law Journal*, *42*(5), 298–302.

Gibbons, J. (2003). *Forensic Linguistics: an introduction to language in the justice system.* Oxford: Blackwell.

Gold, E., Ross, S., & Earnshaw, K. (2018). The 'West Yorkshire Regional English Database': Investigations into the generalizability of reference populations for forensic speaker comparison casework. In *Interspeech 2018: Speech Research for Emerging Markets in Multilingual Societies* (pp. 2748–2752).

Harrington, L. (2023). Incorporating automatic speech recognition methods into the transcription of police-suspect interviews: factors affecting automatic performance. *Frontiers in Communication*, *8*, 1165233. Retrieved 2025-08-13, from https://www.frontiersin.org/articles/10.3389/fcomm.2023.1165233/full doi: 10.3389/fcomm.2023.1165233

Harrington, L. (2024). *Towards improving transcripts of audio recordings in the criminal justice system* (PhD thesis). University of York.

Haworth, K. (n.d.). Process, procedure and professional practice in the creation of police interview evidence. In T. Grieshofer & K. Haworth (Eds.), *Language and Justice: Communication in Legal Practice.* Cambridge: CUP.

Haworth, K. (2018). Tapes, transcripts and trials: The routine contamination of police interview evidence. *The International Journal of Evidence & Proof*, *22*(4), 428–450. Retrieved 2025-01-28, from https://journals.sagepub.com/doi/10.1177/1365712718798656 doi: 10.1177/1365712718798656

Haworth, K., Tompkinson, J., Richardson, E., Deamer, F., & Hamann, M. (2023). "For the Record": applying linguistics to improve evidential consistency in police investigative interview records. *Frontiers in Communication*, *8*, 1178516. Retrieved 2025-08-13, from https://www.frontiersin.org/articles/10.3389/fcomm.2023.1178516/full doi: 10.3389/fcomm.2023.1178516

Leemann, A., Perkins, R., Buker, G. S., & Foulkes, P. (2024). *An Introduction to Forensic*

*Phonetics and Forensic Linguistics*. London New York: Routledge, Taylor & Francis Group.

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org

Richardson, E., Hamann, M., Tompkinson, J., Haworth, K., & Deamer, F. (2023). Understanding the role of transcription in evidential consistency of police interview records in England and Wales. *Language in Society*, 1–32. Retrieved 2025-01-28, from https://www.cambridge.org/core/product/identifier/S004740452300060X/type/journal_article doi: 10.1017/S004740452300060X

Richardson, E., Haworth, K., & Deamer, F. (2022). For the Record: Questioning Transcription Processes in Legal Contexts. *Applied Linguistics*, *43*(4), 677–697. Retrieved 2024-06-08, from https://academic.oup.com/applij/article/43/4/677/6524571 doi: 10.1093/applin/amac005

Rock, F. (2020). Witnesses and Suspects in Interviews: Collecting Oral Evidence: The Police, the Public and the Written Word. In M. Coulthard & A. May (Eds.), *The Routledge handbook of forensic linguistics* (pp. 112–126). London and New York: Routledge.

Tompkinson, J., Haworth, K., Deamer, F., & Richardson, E. (2023). Perceptual instability in police interview records. *The International Journal of Speech, Language and the Law*, *30*(1), 22–51. Retrieved 2025-08-13, from https://utppublishing.com/doi/10.1558/ijsll.24565 doi: 10.1558/ijsll.24565

*Language and Law / Linguagem e Direito*, *Vol. 11*(2), 2024, pp. 79-93
DOI: doi.org/10.21747/21833745/lanlaw/11_2a3

# "Could you tell me what that means?" Ordinary and institutional vocabulary in police interviews with rape victims†

Nicci MacLeod [1] and Elizabeth Hay[2]

[1] *Aston Institute for Forensic Linguistics*, n.macleod5@aston.ac.uk
[2] *Aston Institute for Forensic Linguistics*, elizabethhay26@gmail.com

***Abstract***
*Perhaps the most significant outcome of the UK government's 2021 'End-to-end' review of how sexual offences are dealt with by the Criminal Justice System is Operation Soteria Bluestone (OSB), a programme of research and change which has resulted in the development of a national operating model rolled out in July 2023 across all 43 police forces in England and Wales. Surprisingly, Achieving Best Evidence (ABE) interviews with victims did not form a major focus of this model, despite them being widely regarded as the most crucial link in the chain. This paper represents part of the final report of a project that aimed to begin to address this gap. The data drawn on are a set of nine ABE interviews collected from the English police force that first piloted OSB (then called Project Bluestone) but that took place during the period immediately preceding the pilot. They are thus a communicative site where potential weaknesses in investigative practice – of the type that OSB set out to tackle immediately – might be expected to be laid bare. In this paper, two issues are identified and discussed: (i) competing vocabularies and (ii) conflicting understandings of demonstrating certain concepts, e.g. lack of consent. I exemplify these conflicts as they play out in the ABE interviews, draw on the philosophical concept of hermeneutical injustice in explaining these, and conclude with some thoughts on how linguistics might usefully contribute to the new model.*
***Keywords:*** *Investigative interviews, Achieving Best Evidence (ABE), rape, Operation Soteria Bluestone, hermeneutical injustice.*

***Resumo***
*O resultado mais significativo da revisão "End-to-end" de 2021 do governo do Reino Unido sobre a forma como os crimes sexuais são tratados pelo Sistema de Justiça Criminal seja a Operação Soteria Bluestone (OSB).*

† **WARNING.** This paper contains extracts from genuine interviews with people reporting rape, some of which are graphic in nature.

*O OSB consiste num programa de investigação e mudança que resultou no desenvolvimento de um modelo operacional nacional implementado em julho de 2023 em todas as 43 forças policiais em Inglaterra e no País de Gales. Surpreendentemente, as entrevistas com as vítimas no âmbito do Achieving Best Evidence (ABE) não constituíram um foco importante deste modelo, apesar de serem amplamente consideradas como o elo mais crucial da rede. Este trabalho representa parte do relatório final de um projeto que tinha como objetivo colmatar esta falha. Os dados utilizados são um conjunto de nove entrevistas do ABE recolhidas da rede de forças policiais inglesas que estiveram à frente do OSB (o antigo "Projeto Bluestone"), mas que tiveram lugar durante o período imediatamente anterior ao projeto piloto. Trata-se, portanto, de um momento de comunicação onde é expectável que sejam reveladas potenciais fragilidades na prática de investigação - do tipo que, desde logo, o OSB se propôs a resolver. Neste artigo, são identificadas e discutidas duas questões: (i) vocabulário concorrente e (ii) entendimentos contraditórios na demonstração de certos conceitos, por exemplo, a falta de consentimento. Exemplificamos estes aspetos tal como se manifestam nas entrevistas do ABE, recorremos ao conceito filosófico de injustiça hermenêutica para os explicar e concluímos com algumas reflexões sobre a forma como a linguística pode contribuir de forma útil para o novo modelo.*
***Palavras-chave:*** *Entrevistas investigativas, Achieving Best Evidence (ABE), violação, Operação Soteria Bluestone, injustiça hermenêutica.*

## 1. Introduction

For many years feminist and legal scholars have decried the so-called 'decriminalisation' of rape (e.g. Baird, 2024; Walker, Foster, Majeed-Ariss, & Horvath, 2020) observing that 'social, economic and political structures support sexual violence' (Scully, 1990, p. 63). Against a longstanding historical backdrop of rape myth acceptance within the justice system as well as society at large (see Anderson & Doherty, 2008) the treatment of sexual offences by the authorities has long been of grave concern. With the 2024 figures showing that a staggering 97.4% of reported rapes in England and Wales do not result in a prosecution, (Home Office, 2024), the treatment of rape by police and legal systems has long been criticized as deeply flawed. This is evidenced by the multiple government reviews conducted over the last fifteen years, for example the joint thematic report by HMCPSI /HMIC (2007) and Stern (2010). In an effort to break the cycle of this so-called 'rhetoric of reform' (Jordan, 2011), one concrete response from the UK government 'end-to-end rape review' (Ministry of Justice, 2021) – which had concluded that 'the current situation is totally unacceptable' (p. 4) – was a large-scale Home Office-funded research programme named Operation Soteria Bluestone (henceforth OSB).

OSB saw a number of police forces across England and Wales give academic researchers unprecedented access to their work, allowing deep-dives into their investigative practices around rape and serious sexual assault (RASSO), resulting in the nationwide rolling out of a new operating model for RASSO investigations (College of Policing, 2023). Setting out a novel 'victim-centred, suspect-focused and context-led' approach

to RASSO investigation, the model is intended to transform police responses to victims/survivors.

Since victims/survivors of sexual offences are classified as intimidated witnesses according to the Youth Justice and Criminal Evidence Act (1999), their police interviews are routinely video recorded in order to enable these same interviews to stand in for the victim during direct examination at a subsequent trial. The interviews are thus conducted in line with the Achieving Best Evidence (ABE) protocol (Ministry of Justice, 2022). This guidance, first rolled out in 2002, is designed to assist interviewers who are tasked with collecting accounts from vulnerable, intimidated, and significant witnesses, and incorporates expert advice, chiefly from the fields of cognitive and social psychology, around enhancing memory retrieval and creating the optimum interpersonal conditions for disclosing high quality evidence. A number of scholars have recognised that the police investigative interview may be the most crucial link in the chain in terms of addressing issues of attrition (the rate at which reported incidents 'drop out' of the system), and for ensuring fair treatment for victims (e.g. Gregory & Lees, 1999; Jordan, 2004). Yet interviews with victims were markedly absent from the discussion when it came to the report of OSB's findings and recommendations (Stanko, 2022). While the Year 1 report highlights the widespread acceptance of rape myth amongst serving officers and foregrounds the importance of police knowledge around sexual offending behaviour, the nature of rape, and its potential impact on different types of victim, it has little to say about how such knowledge might best be integrated into interviewing practice.

## 2. Operation Soteria Bluestone (OSB)

OSB is organised around six pillars, the first three of which relate to everyday policing (suspect-focused investigations; targeting repeat suspect; and a procedural justice approach to victim/survivor engagement), while the second three revolve around drivers for improvement (officer learning, well-being and organisational change; data and performance; examining, understanding, and improving the use of digital material in investigations). As mentioned above, there is no explicit mention of ABE interviews within the resulting recommendations, and they receive only a passing mention as part of the data that were reviewed during year one of the programme (Stanko, 2022). While a new arm of the programme has been commissioned with ABE interviews as its focus[1], this is in its infancy and being conducted from a wholly psychological perspective. It is clear that linguistic input is both absent and necessary.

In terms of where such efforts might be located within the existing OSB structure, there are two obvious themes to which it relates. Pillar One, 'suspect focussed investigations', is intended to represent a shift from focussing on the victim, her actions and her credibility, to those of the suspect. Previous linguistic research into RASSO ABE interviews has uncovered this as a key area of concern. Consider, for example, this extract from MacLeod (2020a). The parties are discussing how the interviewee (IE) came to travel back to her home with two men, one of whom she alleges went on to rape her. IR stands for interviewer:

[1](Milne and Tidmarsh, p.c. 2024)

```
1    IR     and then what happened?

     IE     they said "ah s- can we come back to your house" (.) and I
            said "ok fine it's not very often I get company" (.) didn't
5           have a problem with it.
            (5.0)
     IR     so how had you felt about the night so far with=

     IE     =okay (.) no problems at all.
10          (3.0)
     IR  →  and you'd said yes because you were w- w- enjoying the
            c[ompany,]

     IE       [yeah  ][I felt safe.]
15
     IR                [so then     ]you felt safe •hh what u:m what
            happened after that then how did you get (.) [home?]
```

**Extract 1. Victim-focussed language (MacLeod, 2020a)**

Note how on lines 11-12, despite having been presented with the information that it was the men she was with who had first suggested the idea of them returning home with her, it is not this but the IE's own reasoning for *allowing* them to that the IR has chosen to formulate, and thus foreground in the discourse. This is a prototypical example of focussing on victim behaviour while ignoring that of suspect[s]; any opportunities to explore the latter's potential premeditation of the attack are lost in this instant.

A second area of the existing OSB model into which the evaluation of ABE interviews naturally sits is Pillar 3, 'a procedural justice approach to victim/survivor engagement'. Procedural justice refers to the idea that people's perception of fairness is strongly impacted by the quality of their experiences, and not merely the *end result* of these experiences (Hohl, Johnson, & Molisso, 2022). Thus, a victim/survivor's impression of reporting a RASSO incident will come down to their interactions with the police – the most substantial of which is their ABE interview – rather than whether or not their complaint results in a successful prosecution and conviction. Once again, there is clearly a space here for research around the discourse of the ABE interview.

## 3. Investigative interviews and clashing vocabularies

Interaction in police interview discourse is a well-established focal point of forensic linguistic research, with studies focusing variously on power dynamics (Thornborrow, 2002), question types (Oxburgh, Myklebust, & Grant, 2010) patterns of co-operation (Tkacukova & Oxburgh, 2020), and the co-construction of evidence (Haworth, 2015). Indeed, there are growing efforts to introduce the findings of forensic linguistic research into the training of police interviewers (see, for example, Stokoe, 2014).

For example, Rock (2001) unpicked the processes at work during the collection of witness statements, demonstrating that the final texts are the product of several retellings and negotiations between the witness and the statement-taker. The finished product displays traces of the institutional voice, rather than representing solely the words of the witness, as many might assume. Heydon (2005) seminal work, examining the language of suspect interviews, used tools drawn from interactional sociolinguistics and conversation analysis to further drive home the point that, rather than being un-

altered verbatim accounts from interviewees, the final versions of such interactions are at their heart mediated, co-constructed accounts. Police interviewers, as institutional representatives, bring their familiarity with institutional practices and priorities to the interaction, shaping the narratives in such a way as to give them evidential value.

Interviews with vulnerable individuals, including victims of sexual offences, have proven particularly fertile ground for linguistic enquiry. Antaki, Richardson, Stokoe, and Willott (2015) for example, focus on the potential blame-implicative nature of particular conversational practices, and MacLeod (2020a) elucidates the ways in which particular potentially blame-implicative contributions from interviewers can be mapped directly onto prevalent rape mythology and stereotypes.

Previous work has shown that for police interviewers there is often a tension between using 'the same words as the witness has used' as per the ABE guidelines, (Ministry of Justice, 2022) on the one hand, and using language that is deemed institutionally appropriate and evidentially useful, on the other (MacLeod, 2020b). A long history of research into institutional talk has shown us that it is largely goal oriented and often peppered with technical vocabularies which 'can embody definite claims to specialized knowledge and institutional identities' (Drew & Heritage, 1992, p. 29). This is certainly true of the police, as established by Fox (1993) and later Hall (2008). Yet we need to look beyond technical vocabularies to understand the tensions at play in the RASSO ABE interview. It is a well-established tenet of our discipline that linguistic performance is invariable about *choice*; 'the relevance of lexical choice in institutional contexts is…far more wide-ranging than the use of technical jargon' (Drew & Heritage, 1992, p. 29), 'there are a great many overlapping and competing vocabularies corresponding to different domains, institutions, practices, values, and perspectives' (Fairclough, 1992, p. 77), and we all view language as 'choices among alternatives' (Halliday & Matthiessen, 2004, p. 19). What is laid bare in these ABE interviews is that two speakers' choices are rarely the same, and this can have serious consequences for procedural justice.

A useful lens through which to view these occasions of divergence is the philosophical concept of *epistemic injustice*. First formalised by Fricker (2007), epistemic injustice refers to those wrongs committed against a person 'specifically as a knower' (Fricker, 2007, p. 1). The concept can be broken down into *testimonial injustice*, whereby a diminished level of credibility is afforded to the account of a particular speaker, and *hermeneutical injustice*, which refers to 'a gap in collective interpretative resources put someone at an unfair disadvantage when it comes to making sense of their social experiences' (Fricker, 2007, p. 1). Both of these phenomena are relevant when it comes to victims/survivors' experiences when reporting sexual offences, but we are particularly concerned here with the latter. Hermeneutical injustice is what we observe when 'relations of unequal power can skew shared hermeneutical resources, so that the powerful tend to have appropriate understandings of their experiences ready to draw on...whereas the powerless are more likely to find themselves with...ill-fitting meanings to draw on in the effort to render them intelligible' (Fricker, 2007, p. 148).

In the intensified context of the police interview, itself a microcosm of differential power relations (between police and public, between a long-acknowledged masculine institution (Silvestri, 2017) and a solitary (usually) female victim/survivor, etc.) the phenomenon of hermeneutical injustice becomes all the more magnified. Yet the concept of hermeneutical injustice as it relates to victims/survivors' *communication* of their experi-

ences is largely unexplored. A number of feminist philosophers have discussed its negative impact on victims/survivors' ' *understandings* of what happened to them: Jenkins (2016) notes that 'rape myths...constitute hermeneutical injustice...victims who accept [them] are less likely to understand their own experience of sexual violence accurately' (p. 1), while Jackson (2019) focusses on how historically a lack of interpretative resources for conceptualising date rape prevented victims/survivors' from identifying their experiences as such. MacKenzie (2022) focuses on the impact of hermeneutical injustice in the specific context of rape investigations and trials, noting that here too rape mythologies and systemic sexism and misogyny impact negatively on victims/survivors' ' ability to acknowledge their experiences, with catastrophic impacts on their wellbeing.

This paper makes two novel contributions: first, it shifts the focus from victims/survivors' ' *understandings* of their experiences to their perceived competence at *communicating* them; and second, in focussing on instances of communicative clash in genuine ABE interviews with RASSO victims/survivors the tangible manifestation of epistemic injustice in rape investigations is laid bare.

## 4. Data and Method

The data drawn on here are a set of nine video-recorded ABE interviews conducted with persons reporting themselves as victims of rape between March 2019 and June 2020. Since Project Bluestone was piloted between January and March 2021, we might reasonably expect these to represent a 'before' set. However, the force had already trialled a related initiative in 2009 (REDACTED, 2011), building a specialist RASSO team in order to address their failings as a force when it came to detection and convictions in this area. We might thus expect there to be some residual effects of these earlier efforts; this is a force that had acknowledged they have a problem and has demonstrated motivation to address it.

The data contain a fairly even spread of male and female interviewers and interviewees, and of same-sex and mixed-sex dyads. The interviewees were aged between 22 and 55 at the time of the interview, but had a median age of 27.

The interviews were transcribed to Jeffersonian standards (see Hepburn & Bolden, 2017) by the second author and another research assistant, and then scrutinised for key themes through a discourse analytical lens (see Gee, 2014). While a range of phenomena were identified for further exploration, the focus in this paper is on conflicts of vocabulary and meaning in the talk of interviewees and interviewers. Thus, each instance of participants using differing terms for the same concept were isolated and discussed. Instances of an observable clash in expectations around epistemic bases, i.e. bases for *knowing*, are also drawn out and discussed.

## 5. Lexical choice (hermeneutical resources) in RASSO ABE interviews

There are a number of occasions in the interviews where we can observe the parties drawing on competing vocabularies, or in other words, competing *hermeneutical resources*: the shared meanings people use 'to understand their experience, and communicate this understanding to others' (Romdenh-Romluc, 2017, p. 1). In Extract 2, the IE's use of a common term is problematised by the IR once, and despite the IE's acquiescence to a more formal alternative, is problematised again even more intensely (in

all extracts from the current dataset, demographic information about the interviewee appears after the transcript number in brackets, i.e. F28 = female aged 28).

```
1   IE      no he didn't say anything until he had finished cu- when he
            had cummed in me and stood up (hih)
    IR
        →   okay so when you when you refer to h- hes cum inside you would
5           you would you be able to tell me what what that (0.5) means to
    IE      (.hih) (1.1) h- has he ejaculated

    IR      he's ejaculated [inside my] vagina yeah

10  IE  →                   [okay     ]
            okay so the semen that's come out of his (.) penis did it go
            inside your vagina

            yes
```

**Extract 2. Lexical clash (Transcript 2: F28)**

The IE's use of *cummed in me* (line 2), (a widespread term for ejaculation that would certainly be familiar to the IR and arguably the wider justice system), is picked out and explicitly flagged by the IR as a problem source in need of repair by the IE on lines 4-6: *tell me what that means…has he ejaculated*. The IE confirms this relexicalization with a repetition of it on line 8, adding the clarification that by *in me* she specifically meant *inside my vagina*. One might think the clash should be resolved at this point. However, the IR continues with an even more detailed formulation offered back to the IE to confirm on lines 11-12: *the semen that's come out of his penis did it go inside your vagina*. Unsurprisingly the IE confirms this on line 14. Of course, the IR has an institutional role to play, and this interaction must be converted into something of evidential use. But ABE does set out specific guidance around this phenomenon: 'If the interviewer is seeking elaboration on what the witness mentioned in their free narrative account, the interviewer should as far as possible try to use the *same words* that the witness used' (Ministry of Justice, 2022, p. 92; my emphasis). For this IR, it seems a requirement has been identified to clarify not once, but twice, the meaning of the IE's chosen words: the implication being that they are not appropriate for the institutional context nor for the goal they are intended to achieve. This is not a requirement acknowledged by all IRs, and we can see an opposing pattern occurring elsewhere in the dataset. In Extract 3, the IR wholeheartedly accepts and adopts the IE's chosen vocabulary as they discuss the events in question.

```
1   IR      where did he touch you

    IE  →   he touched me where you would have sex

5   IR      okay

    IE  →   the lady parts hhh

    IR  →   okay so lady parts where you'd have sex
10
    IE      °yep° (0.5) I don't wanna say it

    IR      so you're led down he's kissing you have you still got your
            clothes on at this point
```

```
15
                ((around 5 lines omitted))

20  IR          okay so where did he touch you

    IE  →       down there first

    IR          right okay
25
    IE          um and then that led then into him taking his trousers off

    IR  →       okay so when he does that when you said he's touched you down
                there what do you do
30
                (16 lines omitted)

    IR          °°okay°° so you say you were consenting to being touched and
46      →       you said he's touched you down there what as he touched you
                with

    IE          his hands
50
                (4 lines omitted)

55  IR          °okay so what has he done with his fingers°

    IE          °um well put them inside°

    IR          °okay put them inside where°
60
    IE  →       °my (0.5) vagina°

                (5 lines omitted)
66
    IR          so when you say you consented to being touched and he's put
        →       his fingers inside your vagina how would he know you've
                consented to that
70
    IE          cus cus I didn't say no
```

**Extract 3. Lexical concord (Transcript 8: F27)**

ABE 2022 is clear on how to proceed if an IE displays reticence around talking about particular elements of the offence: 'if the witness has communicated something that the interviewer feels needs to be clarified, but the witness seems reluctant or unable to do so at that point, it may be better that the interviewer return to it later in the interview rather than to press on' (Ministry of Justice, 2022, p. 75). In Extract 3 this is precisely this IR's strategy, and her formulation on line 9 encompasses both of the euphemistic terms used by the IE, *lady parts* and *where you'd have sex*. When the IE introduces a third characterisation, down there (line 22) the IR also adopts this label (line 41). Note that by the time the IR uses the technical anatomical term *inside your vagina* on line 68, this has already been introduced by the IE (albeit with a hesitant pause) on line 61. This extended extract starkly demonstrates that IEs can overcome initial reluctance if treated with patience and understanding.

From the two extracts discussed in this section we can see marked differences in the ways in which IRs deal with the issue of potentially institutionally incompatible language. Outright rejections of particular vocabulary items, as we saw in Extract 3, has the potential to lead to a less co-operative exchange than acceptance and patience, as we see in Extract 4.

## 6. Epistemic bases in ABE RASSO interviews

Previous research (MacLeod, 2010, 2020b) focussing on clashes of understanding in police interviews with rape victims has identified that the two parties often demonstrate contrasting understandings of what constitutes evidence for knowledge, or contrasting *epistemic bases*. In the following extract from MacLeod (2010) for example, the IR is questioning around events that occurred as soon as the IE and the two men who accompanied her home arrived at her address. The IE has already mentioned the topic during the free narrative phase of the interview, and the extract begins with the IR reactivating that topic.

```
1   IR        right (0.5) wh- when you said you'd come in and you'd ran
              straight up the stairs to get into bed or on to the bed
              because you felt ill •hh but you said that Gary: was on the
          →   settee how did you know that he'd sat on the settee?
5             (3)
    IE        cos Steve had come up (.) for pillows for him (.) I knew he
              was on the settee.

    IR    →   •hh so did you actually see him on the [settee]
10
    IE                                               [no    ]

    IR        right=

15  IE        =just Steve come up for pillows
              (1)
    IR        [right]

    IE        [unless]that was before I was sick or after I was sick  (0.5)
20
    IR    →   so at some point (0.5) Steve had came up and asked you for
              pillows and [wh-]

    IE                          [and] some bedding (.) and I said it's all in
25            the cupboard (.) the bedding for the (.) •shih  (1)

    IR    →   and that was how you were aware that Gary was on the se[ttee]

    IE                                                             [aye]
```

**Extract 4. Conflicting epistemic bases, MacLeod (2010)**

What is clear from Extract 4 is that for the interviewee, the fact that *Steve had come up for pillows* (line 7) is evidence enough that 'Gary' was on the settee (line 8). This does not seem to be an acceptable epistemic basis for asserting that knowledge as far as the IR is concerned. She pushes for whether the IE had *actually see[n] him on the settee* (line 10), in response to which the IE repeats that *no…just Steve come up for pillows* (lines 12 and 15). In seemingly only partial acceptance of that basis, the IR produces the formulation so…*Steve had came up and asked you for pillows…and that was how you were aware that Gary was on the settee* (lines 21-22 and 27-28).

This phenomenon is also evident in the current data; as well as having conflicting vocabularies, there is evidence that the IR and IE are often in disagreement about the criteria required in order to *know* something (i.e., what is an acceptable *epistemic basis*). In the following extract, the IE has reported that her attacker engaged in digital penetration, and the extract begins with the IR following this up.

| | | | |
|---|---|---|---|
| 1 | IR | | °okay.° (2.0) (.hhh) >you say he's< (.) and then he started to insert his fingers into your vagina? |
| | | | (1.7) |
| | IE | | ((nods))°yeh° |
| 5 | | | (0.7) |
| | IR | → | how do ya know it was his fingers? |
| | | | (5.6) |
| | IE | | ((shrugs, gestures hands palm up)) um (0.9) (.hh) °just° assuming, |
| 10 | | | (1.6) |
| | IR | | °okay, (0.3) do you know how many fingers?° |
| | | | (2.9) |
| | IR | | ((coughs)) excuse me |
| | | | (1.7) |
| 15 | IE | | no |
| | | | (2.6) |

**Extract 5. Conflicting epistemic bases (Transcript 1, F34)**

The IR flags the IE's claim that her attacker had *insert*[ed] *his fingers into* [her] vagina as warranting an epistemic basis beyond the obvious with her question on line 6, *how do ya know it was his fingers* . For many people, particularly those with a vagina, this question may seem quite nonsensical: akin to asking someone with a mouthful of tea how they know it's tea, or asking someone looking at a cat how they know it's a cat. The IE responds how we might expect, with multiple disfluency features and non-verbal markers indicating she is struggling to produce an appropriate response to the question. Police expectations around epistemic bases, then, are often at odds with victims/survivors' experiences and understandings as manifested in their talk.

In the next extract, the IE is questioned around the epistemic basis for her assertion that *he must have taken my tights off and my dress* (lines 1-2).

| | | | |
|---|---|---|---|
| 1 | IE | | I remember being on the bed and then him I said he must have taken my tights off and my dress just all really really quickly I remember him taking his shirt off really quickly and then he was really cocky about him having a good body and |
| 5 | | | I just thought I don't like that about you I don't like that about somebody at all and then I just remember him obviously bring on top of me it all just happened so quickly |
| | IR | → | why must he have taken your tights and dress off |
| 10 | IE | | I don't remember doing it |
| | IR | → | but do you remember him doing it |
| 15 | IE | | .hhh hhh no |

**Extract 6. Conflicting epistemic bases (Transcript 8, F27)**

When questioned on line 9 about how she knew that it was the suspect who had removed her tights and dress, the IE responds that it was through a process of elimina-

tion, *I don't remember doing it*, on line 11. This seems unsatisfactory for the IR, who in a but-prefaced question pushes for confirmation as to whether or not the IE has a specific memory of *him doing it* on line 13. In the literature the discourse marker *but* has been attributed a contrastive role, functioning to express contradiction with the preceding utterance and to mark the denial of expectation (Blakemore, 2002). Here, it elevates the implication that the pursuit of an answer to this question indicates a rejection of *I don't remember doing it* as adequately addressing the question of *why must he have…*(line 9).

Perhaps exasperated that her answer has not been treated as sufficient (note the rapid in-and-out breaths), the IE concedes that no, she does not remember the attacker removing her tights and dress (line 15). It appears that the clash of expectations around acceptable epistemic bases has the potential to cause conversational trouble and subsequent discomfort for victims/survivors as they attempt to relay their stories.

## 7. Conclusions: conflicting hermeneutical resources and competing epistemic bases = a recipe for epistemic injustice

The vast literature around procedural justice tells us that victims of crime place great importance on their treatment by the police – more so than they do on receiving a desired outcome from their report (e.g. Elliott, Thomas, & Ogloff, 2011). This paper has demonstrated how clashes between police interviewers and victims/survivors of rape, both in terms of vocabulary and in terms of bases for knowledge, have the potential to negatively impact on victims/survivors' experiences of investigative interviews. Together, these implicit disagreements can act as a constraint on the victims/survivors' ability to tell their story unimpeded and on their own terms, thus they constitute a clear form of epistemic injustice.

It is unclear how we might resolve such an injustice, since it is not merely an outcome of conscious linguistic choices on the part of interviewers but is embedded in their world-view: a world-view that is markedly different from that of the victims/survivors with whom they must successfully engage. In line with MacKenzie (2022) we maintain that the solution comes through education. Just as rape mythology can only be dismantled through a sustained campaign of raising awareness of the realities of sexual violence among, for example, jurors (Leverick, 2020), so too might an awareness of the lived experiences of RASSO victims/survivors enhance the ways in which police interviewers elicit their accounts. Operation Soteria Bluestone has already identified the importance of investigators acquiring specialist knowledge 'about sexual offending behaviour, the nature of rape contexts and its impact on different victims from different backgrounds informed by academic research' (Stanko, 2022, p. 24), and what this paper has demonstrated is that such educational efforts should extend to interviewing officers and their approach to gaining quality evidence without sacrificing victim/survivor welfare. We saw in the data that being challenged on the grounds of selected vocabulary or bases for knowledge often led to IEs audibly struggling with the interaction, and efforts to instil a victim-centred approach to RASSO investigations should take these dynamics – dynamics

## Acknowledgements

This work was supported by a BA/Leverhulme Small Grant ref. SRG2220098

## References

Anderson, I., & Doherty, K. (2008). *Accounting for rape: psychology, feminism, and discourse analysis in the study of sexual violence*. London ; New York: Routledge.

Antaki, C., Richardson, E., Stokoe, E., & Willott, S. (2015). Police interviews with vulnerable people alleging sexual assault: Probing inconsistency and questioning conduct. *Journal of Sociolinguistics*, *19*(3), 328–350. Retrieved 2025-06-24, from https://onlinelibrary.wiley.com/doi/10.1111/josl.12124 doi: 10.1111/josl.12124

Baird, V. (2024). *The Baird Inquiry*. Retrieved from https://www.greatermanchester-ca.gov.uk/media/9861/the-baird-inquiry.pdf

Blakemore, D. (2002). *Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers* (1st ed.). Cambridge University Press. Retrieved 2025-06-24, from https://www.cambridge.org/core/product/identifier/9780511486456/type/book doi: 10.1017/CBO9780511486456

College of Policing. (2023). *National operating model for the investigation of rape and serious sexual offences*. Retrieved from https://www.college.police.uk/national-operating-model-rasso

Drew, P., & Heritage, J. (Eds.). (1992). *Talk at work: interaction in institutional settings*. Cambridge: Cambridge Univ. Press.

Elliott, I., Thomas, S. D. M., & Ogloff, J. R. P. (2011). Procedural justice in contacts with the police: Testing a relational model of authority in a mixed methods study. *Psychology, Public Policy, and Law*, *17*(4), 592–610. Retrieved 2025-06-24, from https://doi.apa.org/doi/10.1037/a0024212 doi: 10.1037/a0024212

Fairclough, N. (1992). *Discourse and social change*. Polity Press.

Fox, G. (1993). A comparison of 'policespeak' and 'normal speak'. In M. Hoey, J. Sinclair, & G. Fox (Eds.), *Techniques of Description: Spoken and Written Discourse* (1st edition ed., pp. 183–195). Taylor and Francis.

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing* (1st ed.). Oxford University PressOxford. Retrieved 2025-06-24, from https://academic.oup.com/book/32817 doi: 10.1093/acprof:oso/9780198237907.001.0001

Gregory, J., & Lees, S. (1999). *Policing sexual assault*. London: Routlege.

Hall, P. (2008). Policespeak. In J. Gibbons & M. T. Turell (Eds.), *AILA Applied Linguistics Series* (Vol. 5, pp. 67–94). Amsterdam: John Benjamins Publishing Company. Retrieved 2025-06-24, from https://benjamins.com/catalog/aals.5.06hal doi: 10.1075/aals.5.06hal

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An Introduction to Functional Grammar* (3rd ed.). Edward Arnold.

Haworth, K. (2015). The Discursive Construction of Evidence in Police Interviews: Case Study of a Rape Suspect. *Applied Linguistics*, *38*(2), 194–214. Retrieved 2025-06-24, from https://academic.oup.com/applij/article-lookup/doi/10.1093/applin/amv009 doi: 10.1093/applin/amv009

Hepburn, A., & Bolden, G. B. (2017). *Transcribing for Social Research*. Thousand Oaks, CA: SAGE Publications.

Heydon, G. (2005). *The Language of Police Interviewing*. London: Palgrave Macmillan UK. Retrieved 2025-06-24, from http://link.springer.com/10.1057/9780230502932 doi: 10.1057/9780230502932

HMCPSI /HMIC. (2007). *Without Consent: Thematic Report* (Tech. Rep.). Retrieved

from https://assets-hmicfrs.justiceinspectorates.gov.uk/uploads/without-consent-20061231.pdf

Hohl, K., Johnson, K., & Molisso, S. (2022, September). A Procedural Justice Theory Approach to Police Engagement with Victim-Survivors of Rape and Sexual Assault: Initial Findings of the 'Project Bluestone' Pilot Study. *International Criminology*, *2*(3), 253–261. Retrieved 2025-06-24, from https://link.springer.com/10.1007/s43576-022-00056-z doi: 10.1007/s43576-022-00056-z

Home Office. (2024). *Crime outcomes in England and Wales 2023 to 2024* (Tech. Rep.). Retrieved from https://www.gov.uk/government/statistics/crime-outcomes-in-england-and-wales-2023-to-2024/crime-outcomes-in-england-and-wales-2023-to-2024

Jackson, D. L. (2019). Date Rape: The Intractability of Hermeneutical Injustice. In W. Teays (Ed.), *Analyzing Violence Against Women* (Vol. 12, pp. 39–50). Cham: Springer International Publishing. Retrieved 2025-06-24, from https://link.springer.com/10.1007/978-3-030-05989-7_4 (Series Title: Library of Public Policy and Public Administration) doi: 10.1007/978-3-030-05989-7_4

Jenkins, K. (2016). Rape Myths and Domestic Abuse Myths as Hermeneutical Injustices. *Journal of Applied Philosophy*, *34*(2), 191–205. Retrieved 2025-06-24, from https://onlinelibrary.wiley.com/doi/10.1111/japp.12174 doi: 10.1111/japp.12174

Jordan, J. (2004). *The Word of a Woman?* London: Palgrave Macmillan UK. Retrieved 2025-06-24, from http://link.springer.com/10.1057/9780230511057 doi: 10.1057/9780230511057

Jordan, J. (2011). Here we go round the review-go-round: Rape investigation and prosecution—are things getting worse not better? *Journal of Sexual Aggression*, *17*(3), 234–249. Retrieved 2025-06-24, from http://www.tandfonline.com/doi/abs/10.1080/13552600.2011.613278 doi: 10.1080/13552600.2011.613278

Leverick, F. (2020). What do we know about rape myths and juror decision making? *The International Journal of Evidence & Proof*, *24*(3), 255–279. Retrieved 2025-06-24, from https://journals.sagepub.com/doi/10.1177/1365712720923157 doi: 10.1177/1365712720923157

MacKenzie, A. (2022). Why didn't you scream? Epistemic injustices of sexism, misogyny and rape myths. *Journal of Philosophy of Education*, *56*(5), 787–801. Retrieved 2025-06-24, from https://academic.oup.com/jope/article/56/5/787-801/7000125 doi: 10.1111/1467-9752.12685

MacLeod, N. (2010). *Police interviews with women reporting rape: a critical discourse analysis* (Unpublished PhD thesis, Aston University). Retrieved from https://research.aston.ac.uk/en/studentTheses/police-interviews-with-women-reporting-rape-a-critical-discourse-

MacLeod, N. (2020a). The discourse of (re)exploitation: female victims in the legal system. In C. R. Caldas-Coulthard (Ed.), *Innovations and challenges: women, language and sexism.* London New York: Routledge, Taylor & Francis Group. doi: 10.4324/9780429026140

MacLeod, N. (2020b). Tell me in your own words…": Reconciling institutional salience and witness- compatible language in police interviews with women reporting rape. In M. Mason & F. Rock (Eds.), *The Discourse of Police Interviews* (pp. 249–267). University of Chicago Press. Retrieved 2025-06-24, from

https://www.bibliovault.org/BV.landing.epl?ISBN=9780226647821 doi: 10.7208/chicago/9780226647821.001.0001

Ministry of Justice. (2021). *The end-to-end rape review report on findings and actions.* Retrieved from https://assets.publishing.service.gov.uk/media/60ed551c8fa8f50c6ef84fbc/end-to-end-rape-review-report-with-correction-slip.pdf

Ministry of Justice. (2022). *Achieving best evidence in criminal proceedings: Guidance on interviewing victims and witnesses, and guidance on using special measures.* Retrieved from https://www.gov.uk/government/publications/achieving-best-evidence-in-criminal-proceedings

Oxburgh, G. E., Myklebust, T., & Grant, T. (2010). The question of question types in police interviews: A review of the literature from a psychological and linguistic perspective. *International Journal of Speech, Language and the Law*, *17*(1), 45–66. Retrieved 2025-06-24, from https://journal.equinoxpub.com/IJSLL/article/view/10428 doi: 10.1558/ijsll.v17i1.45

REDACTED. (2011). *Operation Bluestone.* London: Home Office.

Rock, F. (2001). The genesis of a witness statement. *The International Journal of Speech, Language and the Law*, *8*(2), 44–72. Retrieved 2025-06-24, from https://utppublishing.com/doi/10.1558/sll.2001.8.2.44 doi: 10.1558/sll.2001.8.2.44

Romdenh-Romluc, K. (2017). Hermeneutical Injustice and the Problem of Authority. *Feminist Philosophy Quarterly*, *3*(3). Retrieved 2025-06-24, from https://ojs.lib.uwo.ca/index.php/fpq/article/view/3086 doi: 10.5206/fpq/2017.3.1

Scully, D. (1990). *Understanding Sexual Violence: A Study of Convicted Rapists.* Harper Collins Academic.

Silvestri, M. (2017). Police Culture and Gender: Revisiting the 'Cult of Masculinity'. *Policing: A Journal of Policy and Practice*, *11*(3), 289–300. Retrieved 2025-06-24, from http://academic.oup.com/policing/article/11/3/289/2965270/Police-Culture-and-Gender-Revisiting-the-Cult-of doi: 10.1093/police/paw052

Stanko, B. (2022). *Operation Soteria Bluestone Year 1 Report.* Retrieved from https://www.gov.uk/government/publications/operation-soteria-year-one-report/operation-soteria-bluestone-yearone-report-accessible-version

Stern, V. (2010). *The Stern Review: an independent review into how rape complaints are handled by public authorities in England and Wales* (Tech. Rep.). Retrieved from https://www.womensaid.org.uk/wp-content/uploads/2016/01/Stern_Review_of_Rape_Reporting_1FINAL.pdf

Stokoe, E. (2014). The Conversation Analytic Role-play Method (CARM): A Method for Training Communication Skills as an Alternative to Simulated Role-play. *Research on Language and Social Interaction*, *47*(3), 255–265. Retrieved 2025-06-24, from https://www.tandfonline.com/doi/full/10.1080/08351813.2014.925663 doi: 10.1080/08351813.2014.925663

Thornborrow, J. (2002). *Power Talk: Language and Interaction in Institutional Discourse.* Routledge. Retrieved 2025-06-24, from https://www.taylorfrancis.com/books/9781317879084 doi: 10.4324/9781315839172

Tkacukova, T., & Oxburgh, G. (2020). Patterns of Cooperation between Police Interviewers with Suspected Sex Offenders. In M. Mason & F. Rock (Eds.), *The*

*Discourse of Police Interviews.* University of Chicago Press. Retrieved 2025-06-24, from https://www.bibliovault.org/BV.landing.epl?ISBN=9780226647821 doi: 10.7208/chicago/9780226647821.001.0001

Walker, T., Foster, A., Majeed-Ariss, R., & Horvath, M. (2020). The justice system is failing victims and survivors of sexual violence. *British Psychological Association*.

*Language and Law / Linguagem e Direito*, *Vol. 11*(2), 2024, pp. 94-114
DOI: doi.org/10.21747/21833745/lanlaw/11_2a4

# The power of uptake: Responses to claims to power on anonymous online fora

Felicity Deamer [1], Michaela Wolf[2], Robert Davies[3], and Tim Grant [4]

[1] *Aston Institute for Forensic Linguistics*, f.deamer@aston.ac.uk
[2] *Aston Institute for Forensic Linguistics*, m.wolf@aston.ac.uk
[3] *Aston Institute for Forensic Linguistics*, r.davies6@aston.ac.uk
[4] *Aston Institute for Forensic Linguistics*, t.d.grant@aston.ac.uk

***Abstract***
*Previous work in this area has identified powerful individuals on anonymous online fora as those who draw on multiple power resources in their interactions, and we were able to show how power operates differently across the different fora (Newsome-Chandler & Grant, 2023). Here we explore the same three datasets from a parenting forum, a white supremacist forum, and a Dark Web CSEA forum, but here we have shifted our focus to track those who are liable to be influenced or persuaded by claims to power. Using a purpose specific speech act coding framework, we have devised a taxonomy of responses to claims to power, and we provide a small-scale analysis of the effects of high- power resource individuals on those they are interacting with. We explore whether such a taxonomy of response types that can be scaled through automation to pick out individuals who may be more vulnerable to persuasion and influence. As well as the taxonomy and coding framework itself, we present findings from a close examination of all the posts and responses, and we explore whether persuadability remains constant or fluctuates according to situation or context, the interactant, or any other observable factor(s).*
***Keywords:*** *Power, Influence, Speech Acts, Online Fora.*

***Resumo***
*Trabalhos anteriores nesta área identificaram como "indivíduos de poder", em fóruns online anónimos, aqueles que recorrem a múltiplos recursos de poder nas suas interações, e conseguiram demonstrar como o poder funciona de forma diferente nos diferentes fóruns (Newsome-Chandler & Grant, 2023).*

*Neste trabalho, exploramos três conjuntos de dados de um fórum de pais, um fórum de supremacia branca e um fórum CSEA da Dark Web; porém, aqui mudámos o nosso foco para seguir aqueles que são suscetíveis de serem influenciados ou persuadidos por reivindicações de poder. Utilizando um quadro de codificação de atos de fala com objetivos específicos, criámos uma taxonomia de respostas a reivindicações de poder e apresentamos uma análise em pequena escala dos efeitos de indivíduos com recursos de poder elevado sobre aqueles com quem estão a interagir. Investigámos se essa taxonomia de tipos de resposta pode ser adaptada através da automatização para selecionar indivíduos que possam ser mais vulneráveis à persuasão e à influência. Para além da taxonomia e do próprio quadro de codificação, apresentamos os resultados de uma análise minuciosa de todas as mensagens e respostas e, além disso, exploramos se a persuasão se mantém constante ou flutua de acordo com a situação ou o contexto, o interagente ou qualquer outro fator observável.*
***Palavras-chave:*** *Poder, Influência, Atos de fala, Fóruns Online.*

## 1. Introduction

Motivated by the need to better understand criminal online spaces, earlier work enabled us to identify powerful individuals on online anonymous fora as those who draw on multiple power resources in their interactions. Earlier findings demonstrate how power is wielded differently across three different fora; a general discussion forum, a white supremacist forum, and a Dark Web CSEA forum (Newsome-Chandler & Grant, 2023). Newsome-Chandler and Grant deliver a nine-resource model (see figure 1 below), and by evaluating how users assert authority—from community to technological expertise—the framework enables nuanced understanding of hierarchies and influence in hidden digital spaces.

**Figure 1. Power resources**

1. Personal experience
2. Broad topic expertise
3. Accredited expertise
4. Community expertise
5. Technological expertise
6. Veteran power
7. Subject of law enforcement
8. Private knowledge
9. Citing a secondary source

By developing a resource-based model of enacted power in interaction, Newsome-Chandler and Grant (2023) were able to observe both the commonalities and differences between the online communities presented in the three datasets. The analysis clearly showed that each forum demonstrates a distinct culture of power. This supports the idea that power is not a uniform or static concept, but one that varies significantly by context. The findings suggest that an individual's ability to enact power depends on their access to relevant power resources in a given context. This access may be influenced by personal history, experience, or by situational factors such as topic or content. Even individuals with a broad set of power resources may find themselves less effective in certain interactions. Thus, a "powerful" individual is best understood as someone who possesses a diverse repertoire of resources and is situated in an interaction that allows them to deploy those effectively.

This previous work focused on how users make claims to power to direct or influence the beliefs or behaviour of others. Here we explore the same three datasets, but we have shifted our focus to the other side of the interaction; onto those who are liable to be influenced or persuaded by such claims to power. The objective is to provide a small-scale analysis of the effects of claims to power on recipients – to create a pragmatics-based descriptive taxonomy of their responses to claims to power. Using a purpose-specific speech act coding framework, we have devised a taxonomy of response types that could potentially be scaled through automation to determine how influence, persuasion, and potentially radicalisation work within distinct anonymous online networks, and to potentially pick out individuals who may be more vulnerable to persuasion and influence.

## 2. Background and theoretical underpinnings

Utterances (and by extension, online posts) can be described as speech acts, meaning that they not only convey information but also perform a particular function or action (Austin, 1962; Põldvere, Felice, & Paradis, 2022). Every utterance can be analysed as simultaneously realising three distinct speech acts; the locutionary act is the production of a particular string of words in a particular grammatical structure encoding a specific meaning, the illocutionary act is the act that is performed in producing the utterance (i.e. asserting, promising, questioning) and the perlocutionary act is the resulting psychological or behavioural effect it has on any interlocutors.

The speaker can control the illocutionary speech act but not its perlocutionary effect. I can 'warn' you, hoping to 'persuade you against', but may in fact have the perlocutionary effect of 'encouraging you to'. A speaker (or poster in our case) can produce an utterance with a particular communicative intention in mind, but her interlocutor has the power to render this illocutionary act a success or failure. The speaker only has control over the illocutionary act; if I say "I assert that Elon Musk…", the listener cannot deny I have asserted, but she can refuse to believe me. Whether or not she believes me or not is the perlocutionary effect, or what McDonald (2020) refers to as the ratification theory of uptake.

Any post on an online forum will contain one or more illocutionary speech acts (e.g. an assertion, a question, a request, etc.) and depending on the size of a thread, there will be any number of responses to that original post. Underlying any given response

is the perlocutionary effect (i.e. whether the responder believes the assertion they're responding to, or whether they accept a request and intend to satisfy or meet it), but it isn't typically possible to determine from a response to a post, the precise perlocutionary effect (i.e. whether an assertion or a promise has been believed).For the purposes of this study, we think about perlocution at the most basic level i.e. whether a response is either accepting or rejecting the speech act(s) performed in the post it is responding to, which McDonald (2020) refers to as 'uptake'.

As mentioned above, there has been a large body of previous work exploring hierarchies of power on these online fora (Newsome-Chandler & Grant, 2023), which has shed light on how individuals express power through recurring claims to specific resources (e.g. broad topic expertise, citing of secondary sources, technological expertise etc.), and how power manifests differently across different online fora.

**Table 1. Examples of claims to power resources (taken from the dataset):**

| | |
|---|---|
| "I came across this piece: [link]" | Citing a secondary source for authority |
| "I honestly don't know. There has been a lot of domestic military activity over the last few years.[...]" | Broad topic expertise |
| "I'd completely ignore the SWs opinions on the matter (unless she has medical degree).[...]" | Community specific initialism |
| "[...]so it does happen but many of the posters are foster carers first and in adoption without fostering first it is very rare, in my experience." | Direct personal experience |
| "I know some people who have lived in various Islamic areas for years.[...]" | Indirect personal experience |
| "There is software, like Eschalot, to create vanity onions." | Direct technological expertise |
| "I work in social care and in this area, each individual assessment is unique and each family different [...]" | Direct accredited expertise |
| "Regular with a name change here[...]" | Community expertise |

1. Their work gives us a better idea of how users attempt to assert and gain power, but in order to understand how those claims to power impact other users on the fora, we need to explore how those interactants respond or act upon those claims to power. A user's power and influence are best measured as the impact they have on others. In other words, an online user is only as powerful as other users enable them to be. With this purpose in mind, this paper sets out to create a taxonomy of responses which captures the following information: Whether or not the response was to a **claim to power** (and if so, which type of claim to power – **Personal Experience, Community Expertise, Citing a Secondary Source etc.**) or to a post that made no claim to power
2. What speech acts were being responded to (i.e. assertion, question, commissive etc.)
3. Whether the response was an acceptance or rejection of the post it was responding to (**Acceptance/Partial Acceptance/Rejection**)

4. Whether that acceptance or rejection was made **explicitly** or **implicitly**

Just as Baker, Vessey, and McEnery (2021) found that the socio-cultural contexts (e.g. function, purpose, conventions, value systems) of different Islamic texts determine the language used and the route to persuasion and radicalisation, earlier preliminary analysis of interactional styles across the three fora illustrates that the different socio-cultural contexts, at least to some extent, determine the manner in which the claims to power resources are drawn upon interactionally. On the white supremacist forum, powerful users are skilful rhetoricians who can persuade others of an ideology.

Whereas on the dark web CSEA forum powerful users are those who effectively preserve the safe space for already radicalised individuals via the use of politeness features, hedging phrases, and expressive language. In line with this, we hypothesise that the different socio-cultural contexts might also determine (1) how persuasive or influential a given claim to power might be (i.e. how readily accepted it is), and secondly, how explicitly users are prepared to offer acceptance or rejection.

## 3. Research questions and data

In order to explore users' responses to claims to power, we set out to manually tag the same subset of 24 threads from each forum (selected on the basis of them starting with an explicit request for advice) which were used in earlier phases of this research (Newsome-Chandler & Grant, 2023) using a taxonomy of responses to explore the following patterns of rejection and acceptance:

1. Do we see any interaction between type of power resource and acceptance and rejection rates (and does this differ across fora?)?
2. Do we see any interaction between number or variety of claims to power and acceptance and rejection rates (and does this differ across fora?)?
3. Are some speech acts more persuasive than others (and does this differ across fora?)?
4. Do explicit and implicit acceptances and rejections pattern in different ways between types of claims to power, and within and between the three fora?

## 4. The coding framework

Previous research has attempted to develop automated approaches to speech act annotation (Core & Allen, 1997). Although computational methods can be used to support manual annotation and is beneficial in terms of time efficiency (Archer, Culpeper, & Davies, 2008, p. 637), it is difficult to fully automate pragmatic annotation. This is due to its complex and context-dependent nature, i.e. implicitness/explicitness, which present difficulties that can be overcome by manual annotation (Yu, Li, Su, & Fuoli, 2023, p. 1; Milà-Garcia, 2018; Kohnen, 2015, p. 64; Archer et al., 2008; McAllister, 2015; McEnery, Xiao, & Tono, 2006). Milà-Garcia (2018) has previously implemented manual annotation to analyse conversational disagreement and agreement in different discursive settings, and has demonstrated its practicality in the field of speech act analysis. Similarly, Rees-Miller (2000) conducted a manual analysis of disagreement in American academic talks and courses. The focus of the study was to investigate a potential relationship between the way disagreement is expressed and the speaker's social power (professor vs. student), the severity of disagreement, and the context disagreement appears in.

Given how there is no 'off the shelf' coding framework which meets the needs of our research, it was necessary to develop our own that was tailored to our specific research data.

### 4.1. The overall process

Step 1: Initial Development of the coding framework
Step 2: Preliminary Coding and Discussion to refine the coding scheme
Step 3: Coding scheme finalised
Step 4: Inter-Rater Reliability measured - coders individually coded an unseen section of data. Inter-rater reliability assessed using percentage agreement and Cohen's kappa.
Step 5: Disagreements discussed and a final coding decision jointly made.
Step 6: Division of Dataset where coders began separately coding individual halves of the remaining corpora using the finalised coding scheme.

### 4.2. Development of the coding framework

As in earlier similar studies, our coding scheme is designed to be as broad as necessary whilst still capturing responses to a diverse range of speech acts as it is important to balance "usefulness" and "ease of consistency" within coding (Archer et al., 2008, p. 631). The first step was to manually identify the types of speech acts that were being responded to (e.g. assertion, assertion to fiction, assertion based on secondary source, request, question, suggestion- expressing a potential future action to another poster to consider, commissive – expressing one's own commitment to an action in the future which can include promises, intentions, and plans). Assertion and suggestion are both typically in declarative form; request and question are both typically in interrogative form. Commissives also tend to be in declarative form, but the difference between suggestions and commissives is that suggestions predicate a future action to the addressee and commissives to the speaker.

Secondly, it was determined whether the speech act in the post being responded to was either accepted or rejected. We also considered whether this was done explicitly or implicitly. An acceptance of a speech act was defined as cases where a responder fulfils an action (e.g. answering a question or fulfilling a request), or where the responder "shares the same point of view [...] as the previous speaker" (Milà-Garcia, 2018, pp. 274–275). On the other hand, rejection was defined as cases where a responder does not fulfil an action (e.g. not doing what is requested or stating that an action won't be fulfilled), or where "the speaker's point of view diverges from that of the previous speaker" (Milà-Garcia, 2018, p. 275). Finally, a partial acceptance was defined as cases where a reply fell between acceptance and rejection and did not fit neatly into either (e.g. replying to an assertion with "I'll think about it"). The questions we asked ourselves when constructing the framework are detailed below.

It is important to note here that this binary classification system in which a response is classed as either accepting or rejecting a claim to power is a very coarse-grained measure of users' evaluations of claims to power. We are inferring evaluations of claims to power based on responses. This is an important starting point, which will enable future research focusing on explicit evaluations of claims to power. Sinclair and Coulthard's (1992) three-part model of spoken discourse offers a useful framework to think about

this with. They observe that spoken discourse can be carved up into three parts or stages;

1. Initiation: this stage involves one speaker initiating the conversation by starting a new topic or responding to a previous topic. This can include asking questions, making statements, or giving directives.
2. Response: In this stage, the other participant in the conversation responds to the initiation made by the first speaker. This can involve answering questions, elaborating on the topic, or showing agreement or disagreement.
3. Feedback: The final stage of the model involves the first speaker providing feedback on the response given by the second speaker. This can involve acknowledging the response, asking for clarification, or further developing the conversation.

Here we are exploring parts 1 and 2 of this discourse model, and have set out to determine (via users' responses) whether the initiation has been accepted, but we haven't had access to, and haven't sought to explore the feedback stage explicitly.

**Figure 2. Steps to constructing the framework**

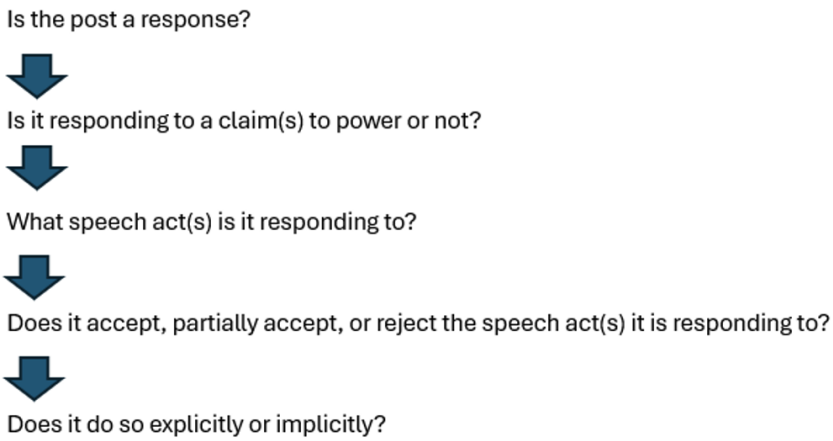
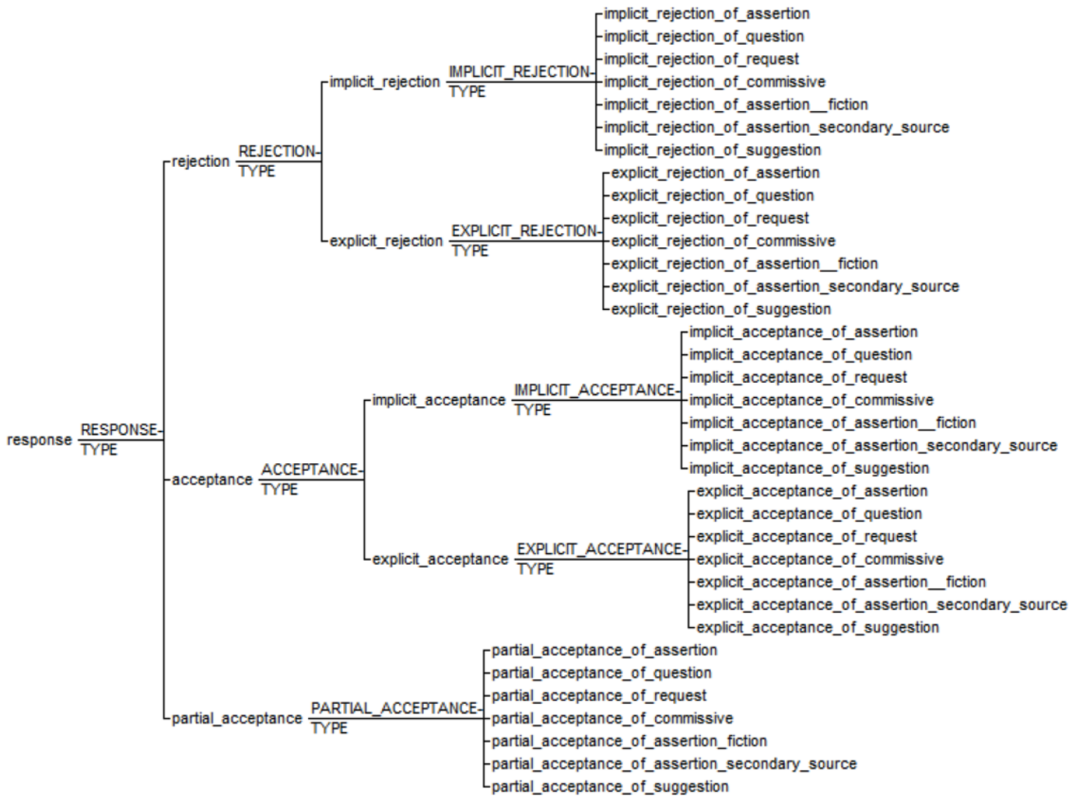


**Figure 3. Coding Scheme Capturing Responses to Speech Acts**




### *Coding examples*

The table below shows demonstrative examples of the coding scheme applied.

**Table 2. Demonstrative examples of Coding**

| | |
|---|---|
| **Implicit Acceptance (Request)** | U1: "Can someone provide me with a link to a useful website, please?"<br>U2: "Hey, please message me privately" |
| **Implicit Rejection (Commissive)** | U1: "I'm going to talk to my social worker about this today."<br>U2: "Maybe think about that again" |
| **Explicit Acceptance (Question)** | U1: "How old is your child? Can they speak yet?"<br>U2: "She's 3 months old, so she definitely can't speak just yet!" |
| **Explicit Rejection (Assertion to fiction)** | U1: "[fictional story]"<br>U2: "This story is horrible" |
| **Partial Acceptance (Suggestion)** | U1: "Maybe you should speak to your husband about this."<br>U2: "I am going to think about it." |

It is important to consider that manual analysis involves an element of subjectivity, and can therefore be prone to errors caused by human factors such as fatigue. To mitigate this, we introduced inter-rater agreement testing (Yu et al., 2023, p. 4).

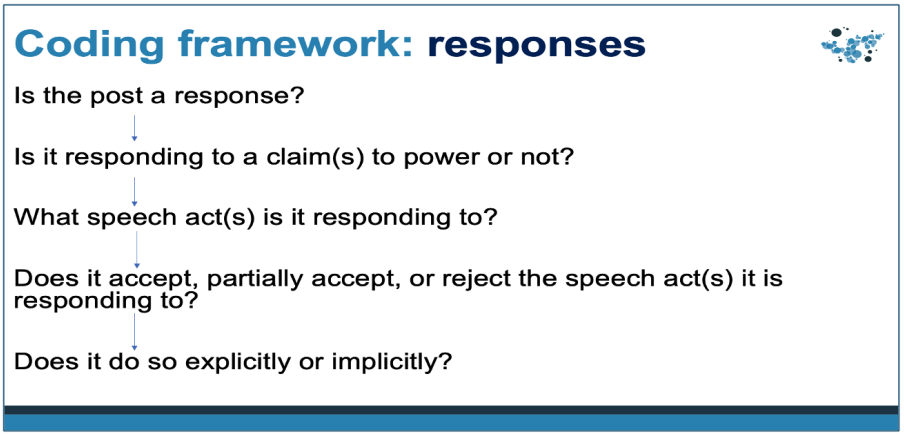
### 4.3. Inter-rater reliability

We used two methods of evaluating interrater reliability: percentage agreement and Cohen's Kappa. Percentage agreement is known to overestimate the reliability between two coders as it takes no account of the chance agreement between the two coders. Cohen's Kappa, however, is calculated by taking the number of possible answers in each coding decision into account,but it makes the assumption that the choices are all independent of one another. In our data this assumption is not true, and this means that Kappa is likely to underestimate reliability. By presenting both figures, we get a range within which 'true' reliability might be expected to sit.

**Table 3. Inter-rater reliability**

| | | Percentage Agreement | Cohen's kappa |
|---|---|---|---|
| Coding of power presence | No Code \| Ambiguous \| No Claim to power \| Claim to Power | 89% | 0.85 |
| | No Code \| n/a \| Personal experience \| Community Initialism \| Vet. \| Accredited expertise \| Broad Topic Expertise \| Subject to Law Enforcement \| Cite \| Personal Knowledge \| Community Expertise | 88% | 0.87 |
| | | | |
| Coding of replies | No Code \| Accept \| Reject \| Partial | 82% | 0.76 |
| | No Code \| Explicit \| Implicit \| NA | 71% | 0.62 |
| | No Code \| Assertion \| Question \| Request \| Commissive \| Thanking \| Assertion-Fiction \| Assertion-sec-source \| Suggestion | 87% | 0.86 |

**Figure 4. Questions guiding coding framework**




***Examples of disagreement***

**Partial and implicit acceptance/rejection (Demonstrative Example)**

| | |
|---|---|
| **U1:** | "All I can speak of is (FORUM), never a member but the "sample" stuff and the leaked stuff is generally plain sick [....] someone. Not to mention I have heard elsewhere what Lux said about the content being trash" |
| **U2:** | "Now I am intrigued. What could possibly be on the site that got you so upset? The posters above seemed to describe it as the same old stuff and nothing too new." |

**Implicit vs. explicit (Demonstrative Example)**

The responder could be asking the question out of interest; "What could possibly be on the site that got you so upset?", although the use of 'possibly' may indicate an element of doubt towards the assertion. We settled on a partial acceptance of the assertion as the meaning of the utterance may lie somewhere in-between.

| | |
|---|---|
| **U1:** | "What made you join this forum?" |
| **U2:** | "Welcome to the forum! I was involved with Czech Nationalism for most of my life but they were made to join the European Union, so I had to focus my efforts elsewhere. I still can't believe our people chose destruction but somehow saw it as an improvement. " |

The response could be understood as an implicit acceptance of the question due to a potential need to infer how the Czech Republic joining the EU could then lead to the poster joining a white supremacist website. However, due to the fact that the poster is still providing a reason for joining (that being because "they were made to join the European Union" and the person "had to focus [their] efforts elsewhere"), the reply was coded as an explicit acceptance of the question.

## 5. Analysis design

After completing the coding for all the threads across the three fora, the raw frequencies for the various categories (e.g. frequency of acceptances across all fora; the frequency of acceptances for each speech act etc.) were extracted from Corpus Tool and input into Excel sheets. Given that the data are frequency counts of categorical variables, chi-square tests were seen as a good choice (Grant et al., 2017) for the more straightforward comparisons (e.g. acceptance rates for posts containing a claim to power compared to those containing no claim to power). In corpus linguistics, chi-square tests are often used to evaluate keyness (Rayson, Archer, Piao, & McEnery, 2004) i.e., to determine whether a feature occurs significantly more often in one corpus over another (Baker, 2006). Similarly, we are exploring frequency of response type (acceptances; partial acceptances; rejections), and use chi-quare tests to determine whether a response type occurs significantly more often in one forum over another, in response to one speech act over another, and/or one claim to power or another. Because chi-square tests become too complex and difficult to interpret with multiple factors and levels (Grant et al., 2017), only two categories can be compared at a time using chi-square. When exploring

the effects of multiple factors (i.e. fora, type of claim to power, response type, degree of explicitness), we use multiway frequency tables (loglinear analysis) to explore a possible association between our factors (i.e. fora, type of claim to power, response type, degree of explicitness). This loglinear analysis makes it possible for us to determine the unique contribution of each factor and its interactions with other factors.

The findings presented and discussed below offer a snapshot of how anonymous users are responding to claims to power delivered via different speech acts, across our three fora. There are numerous ways in which this richly coded dataset could be explored, and the possibilities for future analysis and exploration are numerous, but for the purposes of this paper, we are focusing on whether or not persuasion occurs differently in different online anonymous networks, where the socio-cultural context differ, and where the interactional style and routes to persuasion and/or radicalisation are likely to differ as a result. With this in mind, we will be exploring a series of specific questions, which are detailed below.

1. *Is response type (acceptance, partial acceptance, or rejection) dependent on whether a claim to power (or indeed multiple claims to power) is made in the post that is being responded to?*

Given the pragmatic distinction between a post containing a claim to power and one that doesn't, it is reasonable to expect to find a difference between acceptance and rejection rates of these posts. Moreover, given that previous findings indicated that those making multiple different claims to power are likely to be higher up the hierarchy of power than those making fewer or multiple of the same claims to power, differences between responses to multiple claims to power and one or no claims to power are to be expected.

2. *Do those acceptance and rejection rates pattern differently across the three different fora?*

Given that we know the purpose, conventions, and requirements of the three fora differ significantly, and that some Power Resources (PwRs) are drawn upon differently in our three different fora, we might expect to find that responses to claims to PwRs (or lack thereof) might differ between fora.

3. *Is the type of response (accept, partially accept, or reject) dependent on which speech act is being responded to?*

Given that different speech acts are performing different illocutionary acts (some of which will differ with respect to politeness for example), we might expect to find that the likelihood of a particular speech act being accepted or rejected might differ.

4. *Does the relationship between response type and speech act differ across fora?*

Given that we know the function and interactional style of the three fora differ significantly, and that some PwRs (which are likely to be delivered via particular speech acts) are drawn upon more in one forum than another, we might expect to find that certain speech acts are more likely to be accepted or rejected in one forum over another.

5. *Is the type of response (accept or reject) dependent on which power resource is being responded to?*

In question 1, we explored whether response type is dependent on whether a claim to power, or multiple claims to power are made, but we also want to explore that further by unpacking whether there are differences in acceptance and rejection rates between specific PwRs.

6. *Does the degree of explicitness of response differ across fora?*

We know that the discourse context of each of our online fora differ and the function and purpose of a forum can steer the way in which users interact, as well as the rhetorical and communicative devices they employ to influence and persuade each other. On the basis of this, we hypothesised that we might see more like-minded implicit responses in the fora that facilitate normalisation (i.e. Dark Web and general discussion forum), acceptance, and support, so as to hedge, and indirectly acknowledge and manoeuvre interlocutors. However, it is equally possible that on the white supremacist forum, where users are likely to be vying for a position higher up the hierarchy of skilful rhetoricians (Deamer, Busso, Htait, & Grant, 2023), is also a space in which implicit responses are commonplace, but this time more likely as a face-saving strategy.

7. *Is the degree of explicitness dependent on whether a claim to power (or multiple) has been made, and does any such dependence differ across fora?*

We answer this by comparing the frequency of explicit and implicit responses to claims, no claims, and multiple claims to power, to explore whether responses are dependent on whether or not a claim to power has been made or not.

## 6. Salient findings

### Across fora

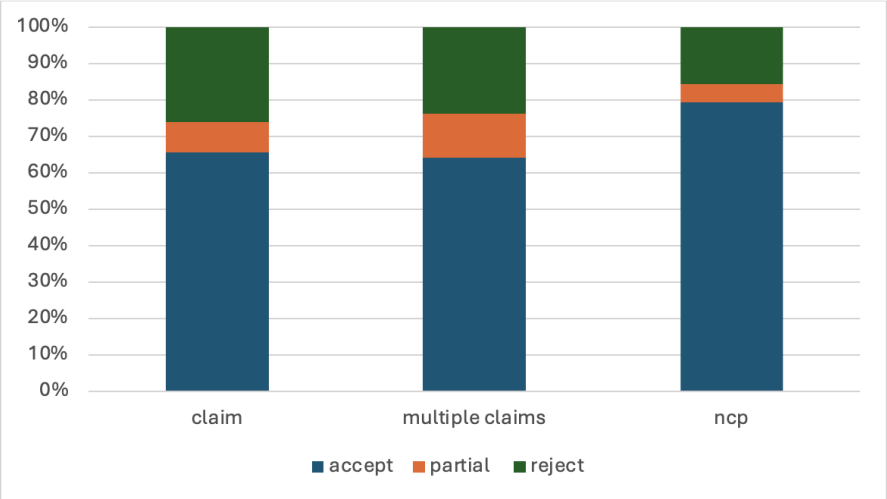
1. **Users are more likely to reject posts containing a claim or multiple claims to power than if no claim to power is made at all, but they do not appear more or less likely to accept or reject a multiple claim to power than a single claim to power**

A Chi-square test of independence revealed that differences in acceptance and partial acceptance rates between posts containing a claim to power and those making no claim to power was not significant (p=.067; p=.093), but there are significantly more rejections of posts containing a claim to power than to those making no claim to power (p=.004).

Moreover, when comparing responses to multiple claims to power and posts making no claim to power, there is a significant difference in frequencies of partial acceptances (p=.002) and rejections (p=.039), but not acceptances (p=.068).

Finally, we can observe that there is no significant variation between responses to posts containing multiple claims to power and those containing just one claim to power (acceptances p=.889; partial acceptances p=.323; rejections p=.690).

**Figure 5. Proportion of responses to claims to power**




2. **Users are more likely to reject an assertion to a secondary source than a standard assertion**

Although proportions of response type do differ, users are not significantly more likely to accept (p=.17) or partially accept (p=.845) an assertion than an assertion to a secondary source, but there are significantly more rejections of assertions to a secondary source than to standard assertions (p=.049).

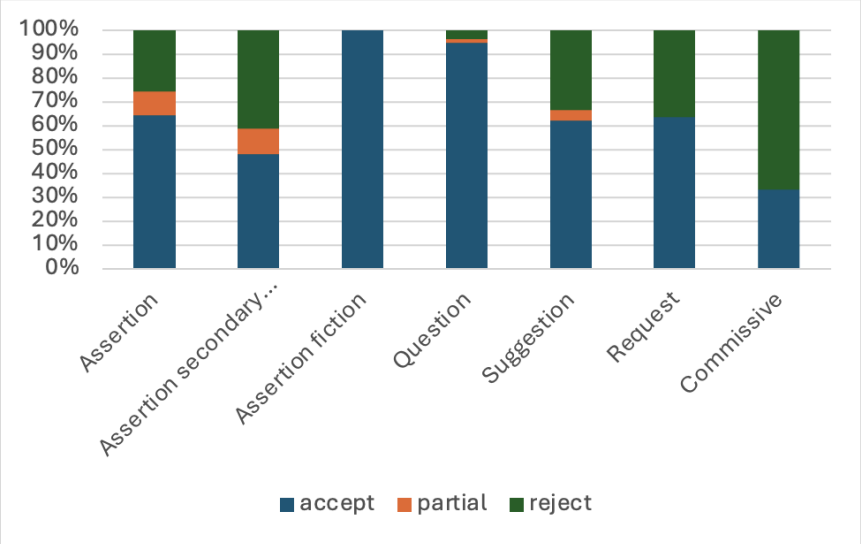
3. **Users are more likely to reject a request than a question**

Similarly, acceptance and partial acceptance rates of questions and requests do not differ significantly (accept p=.295; partial accept p=.686), but rejections did (p<.001).

4. **Users are more likely to accept a question than a standard assertion or an assertion to a secondary source.**

Figure 6, below, illustrates that although proportions of response type do differ, users are not significantly more likely to accept (p=.17) or partially accept (p=.845) an assertion than an assertion to a secondary source, but there are significantly more rejections of assertions to a secondary source than to standard assertions (p=.049). Similarly, acceptance and partial acceptance rates of questions and requests do not differ significantly (accept p=.295; partial accept p=.686), but rejections did (p<.001). Moreover, when comparing responses to assertions and assertions to a secondary source with questions, we see that users are significantly more likely to accept a question than an assertion (accept p<.001) or an assertion to a secondary source (accept p=.001). Rejection rates also differ significantly showing that assertions to secondary source and standard assertions get rejected more than questions (assertion/question p<.001; assert ss/question p<.001). No other comparison between speech acts is significant (likely due to low frequencies).

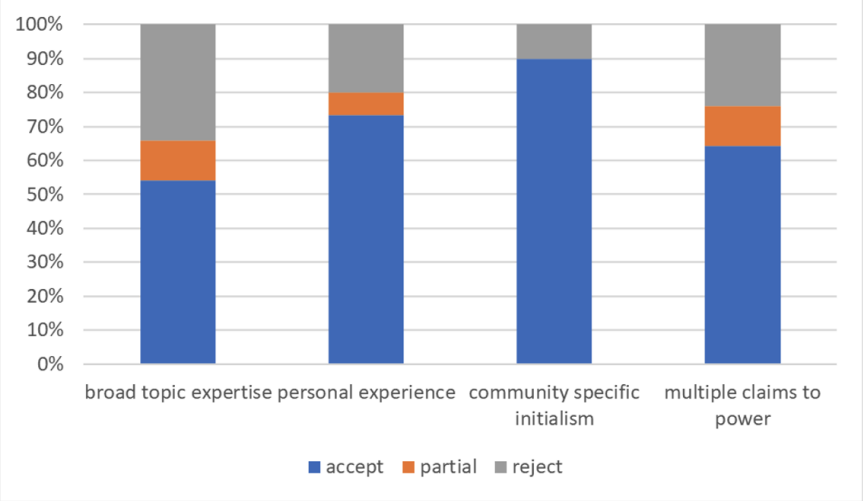
**Figure 6. Proportion of responses to speech acts**




5. **Is the type of response (accept or reject) dependent on which power resource is being responded to?**

Although differences in proportion of response types to different power resources are visible in figure 7, none of these differences were significant (when comparing only the PwRs which had high enough frequency of responses). Again, this is likely due to low frequencies.

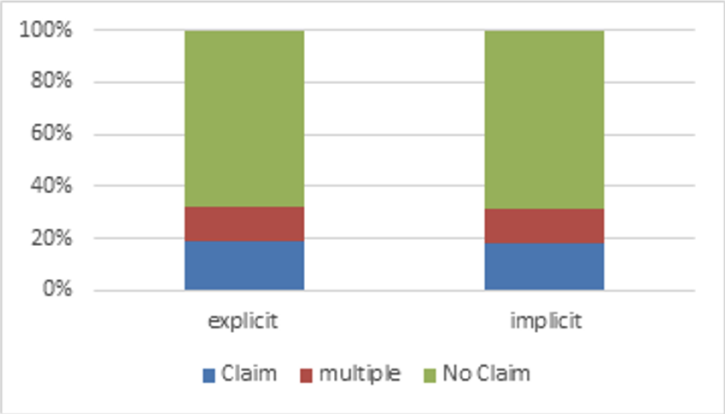
**Figure 7. Proportion of responses to power resources (with high enough frequencies)**




6. **Degree of explicitness is not dependent on whether or not a claim to power has been made or not across fora.**

Again, we ran chi-square tests comparing frequency of explicit and implicit responses to claims, no claims, and multiple claims to power, which show that degree of explicitness is not dependent on whether or not a claim to power has been made or not.

**Figure 8. Proportions of explicit and implicit responses to claims to power**




### Between fora

7. **Response type is dependent on the fora the interaction is taking place on, and on whether one or multiple claims to power are made or not**

We can observe in figures 9-11 that acceptance rates appear to be higher in response to posts containing no claim to power than in response to one or multiple claims to power across all three fora. And users appear more likely to reject multiple claims to power across all three fora, particularly on the dark web. Loglinear modelling allowed us to determine whether response type is in fact dependent not just on whether a claim or multiple claims to power has been made or not (as explored in 5.1 above), but also whether it is dependent on the fora, and whether there is an interaction between those three factors. The multiway frequency analysis reveals that there is a two-way ($p<.001$) but not a three-way ($p=.33$) interaction between factors. Namely, response type is indeed dependent on which fora the interaction is taking place on ($p=.011$), and on whether one or multiple power claims are made or not ($p<.001$), but no three-way interdependence between attributes was found i.e. the association between response type and power claims does not differ between fora.

**Figure 9. Proportion of responses to claims to power White supremacist forum**

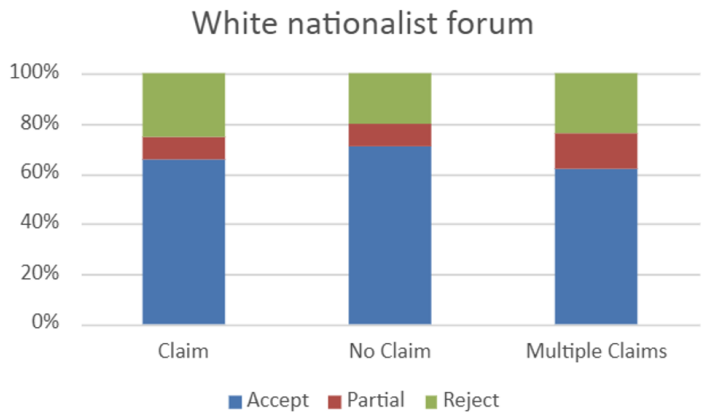


**Figure 10. Proportion of responses to claims to power General discussion forum**

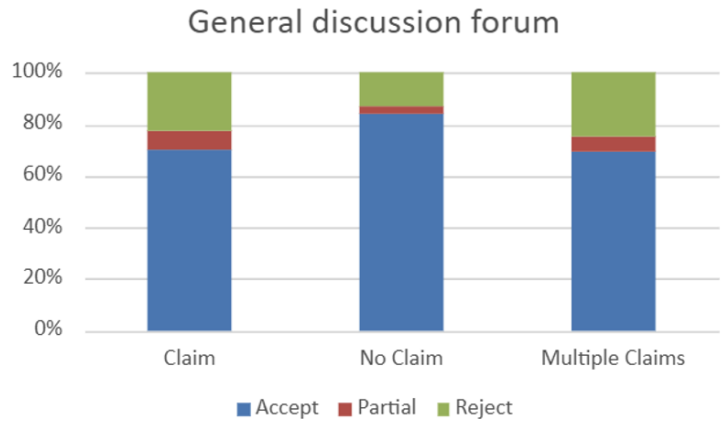
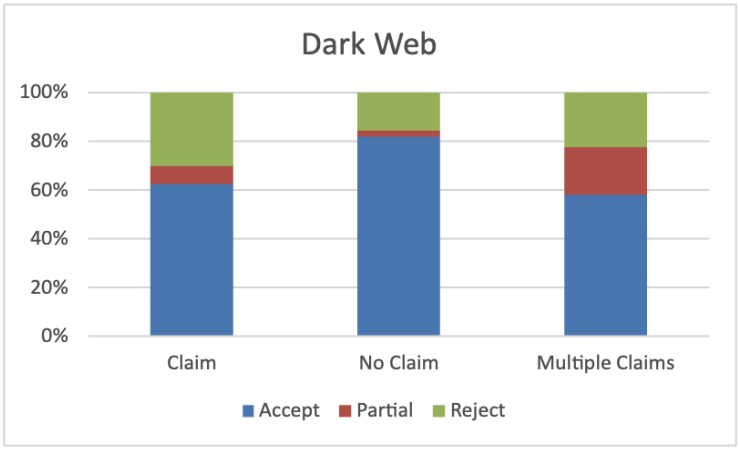



**Figure 11. Proportion of responses to claims to power Dark Web**




8. **The relationship between response type and speech act does not appear to differ significantly between fora**

Loglinear modelling reveals no main effect or higher order interaction between our factors (response type; speech act; fora). As discussed above, when using chi square tests to compare responses to two speech acts at a time, we do see some significant differences, but it is likely that frequencies are too low for differences and interactions to be detected in a loglinear model.

9. **Responses are more often implicit on White supremacist forum than General discussion forum and the Dark web**

We can see from figure 12 that responses are significantly more implicit on the White supremacist forum than General discussion forum and the Dark web, but when we break that down into acceptances and rejections (figure 13), and explore the differences using loglinear analysis, those differences are lost.

**Figure 12. Proportion of explicit and implicit responses between fora**

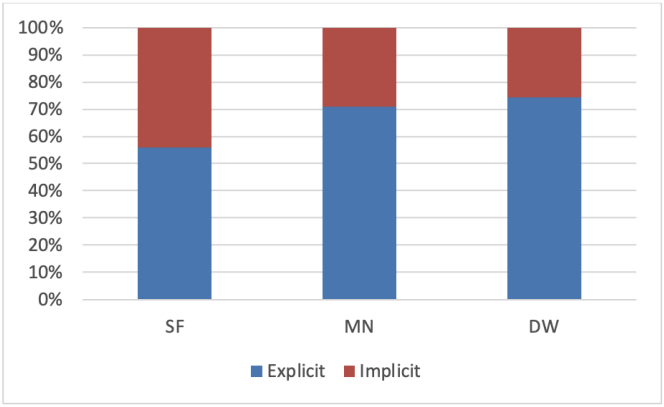
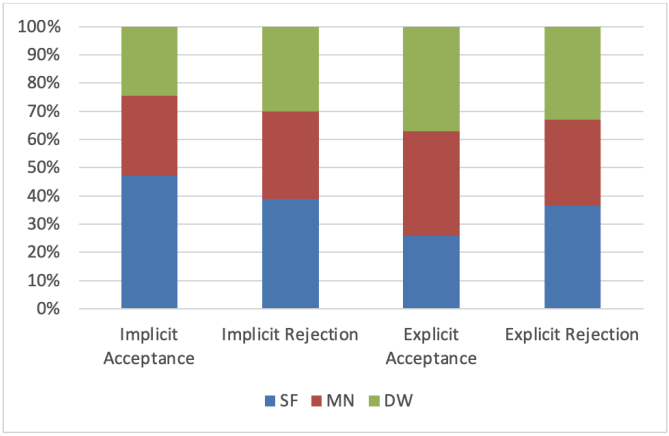



**Figure 13. Proportion of explicit and implicit acceptance and rejection between fora**




## 10. Relationship between degree of explicitness and whether or not a claim (or multiple) to power was made does not differ between fora

Loglinear analysis reveals that there are no interactions between factors i.e. these differences did not vary significantly between fora.

**Figure 14. Proportions of explicit and implicit responses to claims to power White supremacist forum**

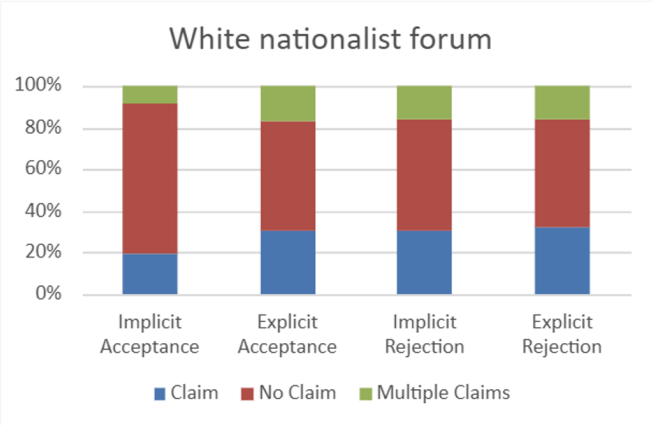


**Figure 15. Proportions of explicit and implicit responses to claims to power General discussion forum**

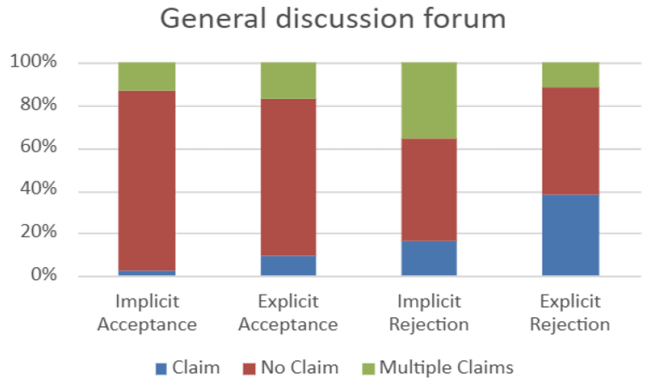
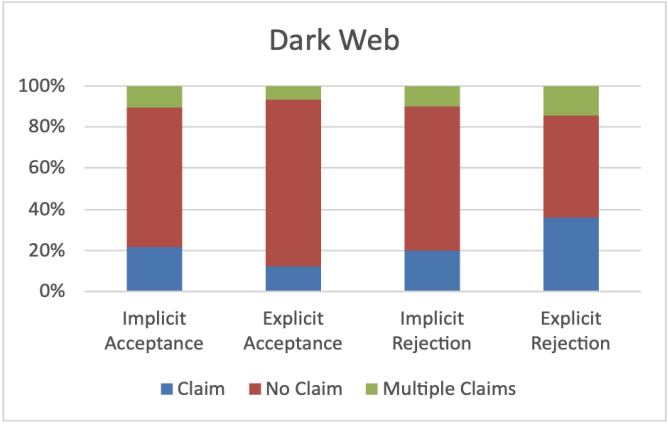



**Figure 16. Proportions of explicit and implicit responses to claims to power dark web**




## 7. Conclusion

Power is hierarchical and inherently bidirectional, where one individual or group seeks to dominate and control another. Having previously identified a number of discrete resources drawn upon by anonymous users to claim power, here we have explored the other half of the bidirectional relationship. Namely, how other users respond to those claims to power, and what that says about (a) the efficacy of specific types of claims to power in different socio-cultural contexts, where routes to persuasion and radicalisation looks somewhat different, and (b) how power dynamics play out interactionally in those different contexts.

Like any hierarchy of power, the power structures on the anonymous online fora examined are unlikely to be set in stone, and some of our findings are perhaps indicative of users jostling for power, saving face, and/or displaying a reluctance to 'lend face' by recognising and acknowledging the power resources of other users. For example, we can see that users are more likely to reject posts containing at least one claim to power than if no claim to power is made at all. We can also see that assertions to a secondary source, in which a user is arguably putting forward a strong display of knowledge and expertise by referencing a robust piece of evidence, are more likely to be rejected by other users than straight forward assertions (not backed up by evidence). Moreover,

agency curtailing speech acts such as requests (a strong act of dominance) are more likely to be rejected than agency enhancing speech acts such as questions (in which the interlocutor is actually invited to make their own claim to power). These findings are perhaps indicative of a broader observation which has been made in other contexts, namely that one is more likely to make a claim to power (i.e. listen to me because.…) or be more forceful (Grant & Woodhams, 2007) when one expects some resistance, or when there is a strong possibility of refusal.

Another notable finding is that both acceptances and rejections are more likely to be implicit on the white supremacist forum than on the general discussion forum or the dark web. This is significant because we know from initial explorations of socio-cultural context and interactional style on the three fora that White supremacist forum is a space in which users who make frequent claims to multiple power resources do so in a manner that could be described as 'hard power', where claims to power are typically formulated via inductive argument without any apparent accommodation to the recipient's frame of mind (e.g. more face-threatening and less hedging). Whereas a powerful user on General discussion forum and the dark web is more likely to consider the psychological state of those they're interacting with, and moderate their knowledge and expertise with emotional awareness. It makes sense then, that recipients of claims to power on White supremacist forum (where users are jostling to move up the hierarchy and create a reputation for themselves as skilful rhetoricians) are less likely to lend face with an explicit acceptance, even if they are in agreement. Moreover, the White supremacist forum primarily functions as a platform for persuasion of an ideology, and much like in any political discourse where success is measured by the growth and popularity of the movement, it makes sense that users are likely to avoid impeding the proliferation of the ideology by explicitly rejecting the claims put forward by other users in support of that ideology.

Further exploration is needed, but response type does appear to be dependent on the fora the interaction is taking place on, as well as on whether at least one claim to power is being responded to. Although we don't see a relationship between degree of explicitness and whether or not a claim (or multiple) to power is made, or a relationship between response type and speech act, this does require further exploration, and expanding our analyses into different corpora may reveal more insights here.

There are many other questions we could ask of this data, and we have explored just a few of them. For example, it would be informative to take a closer look at whether the speech act being responded to dictates the explicitness of the response, and whether that triangular relationship of explicitness, response type, and speech act differs across fora. These analyses would also benefit from the expansion into new and different anonymous fora, beyond the three explored here.

Finally, based on our richly coded dataset from the three fora we have focused on to date, it would be straightforward to identify individuals who have high acceptance rates across the three different fora, and to closely examine all the posts and responses associated with those persuadable and/or easily influenced individuals, and to explore whether throughout their messaging their persuadability remains constant or fluctuates according to situation or context, the interactant or any other observable factor(s).

## References

Archer, D., Culpeper, J., & Davies, M. (2008). Pragmatic annotation. In M. Kytö & A. Lüdeling (Eds.), *Corpus Linguistics: An International Handbook* (pp. 613–642). Mouton de Gruyter.

Austin, J. L. (1962). *How to Do Things with Words*. Oxford: Clarendon Press.

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

Baker, P., Vessey, R., & McEnery, T. (2021). *The Language of Violent Jihad*. Cambridge: Cambridge University Press.

Core, M., & Allen, J. (1997). Coding dialogs with the DAMSL annotation scheme. In (pp. 28– 35). Cambridge, MA.

Deamer, F., Busso, L., Htait, A., & Grant, T. (2023, March). (Unpublished government report). *Hierarchies of Power*, *5*.

Grant, T., Clark, U., Reershemius, G., Pollard, D., Hayes, S., & Plappert, G. (2017). *Quantitative Research Methods for Linguists: A Questions and Answers Approach for Students* (1st ed.). Routledge. Retrieved 2025-08-11, from https://www.taylorfrancis.com/books/9781351722988 doi: 10.4324/9781315181707

Grant, T., & Woodhams, J. (2007). Rape as Social Activity: an Application of Investigative Linguistics. In J. Cotterill (Ed.), *The Language of Sexual Crime* (pp. 1–15). London: Palgrave Macmillan UK. Retrieved 2025-08-11, from http://link.springer.com/10.1057/9780230592780_1 doi: 10.1057/9780230592780_1

Kohnen, T. (2015). Speech acts: a diachronic perspective. In K. Aijmer & C. Rühlemann (Eds.), *Corpus Pragmatics: A handbook* (1st ed., pp. 52–83). Cambridge University Press. Retrieved 2025-08-11, from https://www.cambridge.org/core/product/identifier/9781139057493%23c01504-2-0/type/book_part doi: 10.1017/CBO9781139057493.004

McAllister, P. G. (2015). Speech acts: a synchronic perspective. In K. Aijmer & C. Rühlemann (Eds.), *Corpus Pragmatics: A handbook* (1st ed., pp. 29–51). Cambridge University Press. Retrieved 2025-08-11, from https://www.cambridge.org/core/product/identifier/9781139057493%23c01504-1-0/type/book_part doi: 10.1017/CBO9781139057493.003

McDonald, L. (2020). Your word against mine: the power of uptake. *Synthese*, *199*(1-2), 3505–3526. Retrieved 2025-08-11, from https://link.springer.com/10.1007/s11229-020-02944-1 doi: 10.1007/s11229-020-02944-1

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: an advanced resource book*. New York: Routledge.

Milà-Garcia, A. (2018, September). Pragmatic Annotation for a Multi-Layered Analysis of Speech Acts: A Methodological Proposal. *Corpus Pragmatics*, *2*(3), 265–287. Retrieved 2025-08-11, from http://link.springer.com/10.1007/s41701-018-0037-z doi: 10.1007/s41701-018-0037-z

Newsome-Chandler, H., & Grant, T. (2023). Developing a Resource Model of Power and Authority in Anonymous Online Criminal Interactions. *Language and Law / Linguagem e Direito*, *10*(1), 110–130. Retrieved 2025-08-05, from https://ojs.letras.up.pt/index.php/LLLD/article/view/12841/12478 doi: 10.21747/21833745/lanlaw/10_1/a4

Põldvere, N., Felice, R. D., & Paradis, C. (2022). Advice in Conversation: Corpus Pragmatics Meets Mixed Methods. In P. Culpeper & M. Haugh (Eds.), *Elements in*

*pragmatics* (1st ed., pp. 1–80). Cambridge University Press. Retrieved 2025-08-11, from https://www.cambridge.org/core/product/identifier/9781009053617/type/element doi: 10.1017/9781009053617

Rayson, P., Archer, D., Piao, S. L., & McEnery, T. (2004). The UCREL semantic analysis system..

Rees-Miller, J. (2000). Power, severity, and context in disagreement. *Journal of Pragmatics*, *32*(8), 1087–1111. Retrieved 2025-08-11, from https://linkinghub.elsevier.com/retrieve/pii/S0378216699000880 doi: 10.1016/S0378-2166(99)00088-0

Sinclair, J. M., & Coulthard, M. (1992). Towards an analysis of discourse. In M. Coulthard (Ed.), *Advance in Spoken Discourse Analysis.* London: Routledge.

Yu, D., Li, L., Su, H., & Fuoli, M. (2023). Assessing the potential of AI-assisted pragmatic annotation: The case of apologies. Retrieved 2025-08-11, from https://arxiv.org/abs/2305.08339 (Publisher: arXiv Version Number: 5) doi: 10.48550/ARXIV.2305.08339

*Language and Law / Linguagem e Direito*, *Vol. 11*(2), 2024, pp. 115-134
DOI: doi.org/10.21747/21833745/lanlaw/11_2a5

# Shifting linguistic identity performance and the acquisition of symbolic capital in an online white nationalist forum

Amy Booth [1]

[1] *Aston Institute for Forensic Linguistics*, a.booth@aston.ac.uk

***Abstract***
*This article contributes to a small existing literature which seeks to identify linguistic power resources in harmful online communities, using a dataset taken from a long-running, high profile white nationalist forum to longitudinally explore users' development of power resources over time. By improving understanding of the nature of power in such communities, this work can aid law enforcement in the investigative prioritisation of high- status community members. To do so, I combine Newsome-Chandler and Grant's (2023) linguistic resource model of power with Bourdieu's (2004) concept of symbolic capital, tying the former into a broader, more generalisable understanding of available power resources, while anchoring the latter in a more concrete model of how power may operate discursively in local contexts. My findings show that, surprisingly, users draw on fewer traditional discursive power resources over time, producing less explicitly assertive and extreme discourse. They also begin to treat the forum as a social, rather than solely ideological, space, and increasingly focus criticism on their white (nationalist) in-group, rather than racialised out-groups. I also show that users rely on non-linguistic forms of symbolic capital as secondary, non-linguistic power resources.*
***Keywords:*** *Language and power, linguistic capital, white nationalism, far-right, computer-mediated communication.*

***Resumo***
*Este artigo contribui para a escassa literatura existente no âmbito da identificação de recursos linguísticos associados a uma linguagem de poder presente em comunidades online potencialmente perigosas. Para tal, recorri a um conjunto de dados retirados de um reconhecido fórum nacionalista branco para explorar longitudinalmente o desenvolvimento dos recursos de poder dos utilizadores ao longo do tempo.*

*Ao melhorar a compreensão da natureza do poder nessas comunidades, este trabalho pode ajudar as autoridades policiais a dar prioridade à investigação de membros da comunidade com estatuto elevado. Neste sentido, relaciono o modelo de recursos linguísticos de poder de Newsome-Chandler e Grant (2023) com o conceito de capital simbólico de Bourdieu (2004), interligando o primeiro a uma compreensão mais ampla e generalizável dos recursos de poder disponíveis e assentando o segundo num modelo mais concreto de como o poder pode funcionar discursivamente em contextos localizados. É possível concluir que, surpreendentemente, os utilizadores recorrem a menos recursos tradicionais de poder discursivo ao longo do tempo, produzindo, de forma clara, um discurso menos assertivo e extremista. Começam também a tratar o fórum como um espaço social, em vez de apenas ideológico, e centram cada vez mais as críticas no seu grupo interno branco (nacionalista), em vez de em grupos externos racializados. Demonstro, ainda, que os utilizadores recorrem a formas não linguísticas de capital simbólico como recursos de poder secundários e não linguísticos.*
***Palavras-chave:*** *Linguagem e poder, capital linguístico, nacionalismo branco, extrema-direita, comunicação mediada por computador.*

## 1. Introduction

Recent years have brought increased attention to the far-right movement, largely due to the electoral successes of far-right politicians and parties in the western world and beyond. This has been accompanied by a global rise in deadly far-right extremism, (Europol, 2023; Institute for Economics and Peace, 2023; Pitcavage, 2023), and is compounded by the increasing normalisation of intolerant ideologies in mainstream political discourse (Mondon & Winter, 2020). Law enforcement authorities are increasingly concerned with the presence of far-right actors in online spaces (Europol, 2023). While the far-right has maintained an online presence since the earliest days of the internet (Levin, 2002), the ubiquity of the internet today has allowed the far-right to normalise and spread its message of hate to new audiences. Indeed, in the U.K. today, the majority of convicted extremists were, at least partially, radicalised online (Kenyon, Binder, & Baker-Beall, 2021).

The highly populous, but often anonymous, nature of far-right online spaces creates problems for the law enforcement agencies who monitor them, creating an investigative need to prioritise users who may pose the greatest threat. Understanding how to recognise powerful or high-status users in harmful communities, including extremist communities, allows law enforcement to better prioritise their investigations, and to more convincingly perform as powerful users in undercover operations such as account takeovers (Grant & MacLeod, 2020).

To this end, a limited amount of previous research has explored the discursive construction of powerful and authoritative identities in harmful online communities (Newsome-Chandler & Grant, 2023; see also Chiang, 2024). Being able to perform such

an identity relies on the acquisition of locally normative linguistic and pragmatic *resources* (Grant & MacLeod, 2020), which can only be done over time and through participation in the community. Despite this, the acquisition of such resources has rarely been explored longitudinally, and particularly not with reference to extremist or other harmful online communities.

This article addresses this research gap, drawing both on the resource-constraint theory of linguistic identity performance (Grant & MacLeod, 2020; Newsome-Chandler & Grant, 2023) and on the sociological lens of *capital* (Bourdieu, 2004), a symbolic resource which is acquired over time and through participation, and which renders the holder more powerful in a given context. Combining these two similar, but distinct, notions offers an interdisciplinary perspective which can anchor linguistic identity performances more strongly in existing sociological theories of interpersonal power within groups. I demonstrate this by applying, for the first time, the lens of capital to the primarily linguistic construction of powerful identity performances in the context of a far-right online forum, using a longitudinal inductive discourse analytic approach to explore how users begin to perform power over time.

## 2. Shifting identity performance in online communities

Beyond far-right or other harmful contexts, a small literature exists which explores how users' language use changes over time in online communities, moving from the powerless performance of a new community member to the powerful performance of an established member. This literature can offer insights into the nature of the linguistic identity resources which users may acquire over time through their participation in an online community. However, these studies do not usually longitudinally investigate the same individuals over time, so do not directly demonstrate users' acquisition of resources. Instead, as a proxy, researchers have typically explored the discourse of up to two groups of users at the same point in time. In particular, the language of newcomers or 'newbies', and of users' first ever posts to the community, have received significant attention in the literature. Less commonly, scholars have directly compared newcomers' language with that of the rest of the community. While not explicitly focusing on longer-term users, these studies typically highlight the relative power and expertise of such members, and as a result they are often referred to as experts or 'veterans' (Newsome-Chandler & Grant, 2023), a term I will adopt here.

### 2.1. Newcomer language

Newcomers are commonly stigmatised due to their lack of knowledge of community norms (Golder & Donath, 2004; Graham, 2016), and may threaten the face of more experienced members (Brown & Levinson, 1987). Newcomer discourse is therefore often characterised by the use of negative politeness strategies, mitigating this face threat by emphasising the newcomer's powerless and even subservient identity. This includes the use of low epistemic commitment (Benwell & Stokoe, 2006; Golder & Donath, 2004) and self-deprecation (Benwell & Stokoe, 2006), including explicit reference to their newcomer identity and inexperience (Chiang, 2024). This performance of inexperience and powerlessness is further reflected through the frequent asking of questions, described by Golder and Donath (2004, p. 13) as the "single most prevalent behaviour" exhibited

by newcomers (see also Ang & Zaphiris, 2010; Chiang, 2024; Newon, 2016; Nguyen & Rosé, 2011).

On the other hand, newcomers may also perform what Chiang (2024) calls the 'dual identity' of the *competent newbie*, making claims to legitimacy within the community by highlighting their prior experience or highlighting common ground between themselves and the community (Benwell & Stokoe, 2006; Graham, 2016). This is a somewhat more powerful identity, creating greater alignment with the more experienced community, while still acknowledging the user's newcomer status. This is often achieved by providing biographical narratives in an initial introductory post. These describe the user's reasons for joining the community, emphasising their newness, but also what they share with the community (Chiang, 2024; Graham, 2016; Nguyen & Rosé, 2011). This more personal style contributes to a higher rate of first person singular pronouns than among veteran members (Danescu-Niculescu-Mizil, West, Jurafsky, Leskovec, & Potts, 2013; Nguyen & Rosé, 2011). This feature highlights the sometimes dual nature of newcomer identity performance; although used in the demonstration of the connection between the user and the community, the singular pronoun form nonetheless serves to individualise the user and set them aside from the collective.

### 2.2. Veteran language

Unsurprisingly, veteran language is often characterised in ways that directly oppose that of newcomers. While newcomers typically avoid imposing themselves and their views on others and most commonly ask questions, veterans perform a much more powerful, knowledgeable identity: depending on the nature of the community, they may give advice, often drawing on their own personal experience as the source of their expertise (Angouri & Sanderson, 2016; Newsome-Chandler & Grant, 2023; Rudolf von Rohr, Thurnherr, & Locher, 2019); provide technical information (Singh, 2012); or offer instructions (Ang & Zaphiris, 2010; Newon, 2011). In addition to providing their own, veterans may also have the authority to assess other users' contributions to the discussion and community (Newon, 2011; Rudolf von Rohr et al., 2019). The few existing longitudinal studies of the development of expertise highlight that, as users spend more time in the community, the frequency of question-asking decreases (Ang & Zaphiris, 2010), while those questions that are asked become more detailed and informed (Singh, 2012). Veteran users may also highlight their knowledge of community, or topic-specific, vocabulary (Danescu-Niculescu-Mizil et al., 2013; Newon, 2011; Newsome-Chandler & Grant, 2023; Nguyen & Rosé, 2011).

Alongside performances of knowledge and expertise, veteran users are also typically more community-oriented. They show concern with building rapport between members (Angouri & Sanderson, 2016), sometimes through a lighter or more joking tone (Ang & Zaphiris, 2010). In contrast to newcomers' use of first person singular pronouns, first person *plural* pronouns are more common among veterans; taking a rare longitudinal approach, Danescu-Niculescu-Mizil et al. (2013) show that use of the singular may decrease as plural usage increases. First person plural usage among veterans has several functions, including allowing the user to act as the voice of the community (Rudolf von Rohr et al., 2019); signifying increased identification with the group (Danescu-Niculescu-Mizil et al., 2013); and displaying increased concern with rapport and emotional connections within the community (Nguyen & Rosé, 2011). On a less

inclusive note, veteran users' focus on community may involve gatekeeping and maintenance of community boundaries, and sometimes even active hostility towards perceived 'outsiders' (Graham, 2016; Honeycutt, 2005). In this way, veteran users position themselves as having the authority to determine the boundaries of the community and acceptable behaviour within it.

## 3. The acquisition of linguistic symbolic capital

Newsome-Chandler and Grant (2023) review the ways in which the relationship between language and power has been explored, highlighting the unique situation of anonymous or pseudonymous contexts such as those commonly found in harmful online environments. In these contexts, traditional sources of power – such as institutional power, or macro-level social categories including age, gender, and class – are often invisible. Newsome-Chandler and Grant (2023) therefore propose a *resource model* of power for the exploration of such online settings. Their work draws heavily on Grant and MacLeod's (2020) resource-constraint theory of linguistic identity performance, which posits: a) that identities are not essential or inherent, but are constructed discursively through interaction with others (see also Bucholtz & Hall, 2005); b) that the range of identity performances available to individuals is determined and constrained by the *resources* that the individual has access to; and c) that resources are acquired from a number of sources, including, significantly, the individual's previous socio-discursive experiences. Applying this to the performance of power, then, Newsome-Chandler and Grant (2023) note that power in anonymous online spaces is primarily or solely interactional in nature, created and performed through the locally pertinent linguistic and pragmatic power resources that each individual has acquired and is able to draw upon for their identity performance. As this implies, access to resources – and therefore to performing power – varies across individuals. Resources of power also vary across contexts, with different individuals able to produce powerful performances in different environments. Over time, and through repeated engagement and interaction, members of an online community are able to acquire and subsequently draw on new interactional resources in order to perform an identity that is situationally appropriate and socially valued in the context of that particular community. This results in that individual being perceived by others in the community as powerful, someone "to be listened to" (Newsome-Chandler & Grant, 2023, p. 113).

This formulation of power draws on, and shows strong parallels with, Bourdieu's (1991b, 2004) notion of *symbolic capital*; for Bourdieu (2004, p. 16), capital and power "amoun[t] to the same thing.". The idea of symbolic capital is to recognise that power and authority are drawn not only from financial wealth (economic capital), but also from the possession of prestigious cultural and social resources (symbolic capital). Bourdieu (2004) describes a number of subtypes of symbolic capital, including *social capital*, which is drawn from one's position in a social network, and *cultural capital*, which is the individual's familiarity and comfort with prestigious forms of cultural expression. Cultural capital can be *institutionalised*, formally recognised or reified by some institution; for example, an academic degree functions as a certification that the holder has cultural capital, allowing the degree-holder to 'demonstrate' their capital without the need to perform it interactionally.

However, cultural capital is not necessarily institutionally ratified. Symbolic capital of all kinds is a social, relational phenomenon. Those who *have* symbolic capital in a given context are able to recognise and implicitly ratify others' capital in that context, thus perpetuating the capital that they themselves have (Bourdieu, 1991a). Non-institutionally ratified cultural capital may be *embodied*, reflected in the "long-lasting dispositions of the mind and body" (Bourdieu, 2004, p. 17) that allow individuals to enact their cultural capital unconsciously, as second nature. Embodied cultural capital is not easily acquired; it takes time and effort to consistently make the 'right' choices in the 'right' contexts until it is done unthinkingly, allowing the individual to be admitted to the community and recognised as legitimate (Bourdieu, 1991a).

Bourdieu (1991b) recognises *linguistic capital* as a subtype of cultural capital. Linguistic capital can be understood as an embodied form of capital, concerning the repertoire of linguistic resources individuals have access to (Grant & MacLeod, 2020) which allow them to perform appropriately and in a socially valued way across interactions in a given cultural or subcultural context. As Thompson (1991, p. 13) summarises, the possession of linguistic capital gives individuals "a feel for the game" of successful performance in an interactional context. The notion of linguistic capital derives from Bourdieu's criticism of the abstract, grammatically-focused Chomskyan view of linguistic competence, which disregards actual linguistic performance. Bourdieu argues that speakers also have a "practical competence [...] by virtue of which they are able to produce utterances that are appropriate in the circumstances" (Thompson, 1991, p. 7), and which therefore are "likely to be listened to" (Bourdieu, 1991a, p. 55), mirroring Newsome-Chandler and Grant's (2023) definition of (linguistic) power. Individuals with linguistic capital have a high level of practical competence, possessing the ability to draw on more resources from their larger repertoire, performing comfortably in a wider range of interactions. In this way, the socially dominant – those with greater (linguistic) capital – have a higher degree of interactional agency, and are able to recognise, and even set the standard for, what Bourdieu (1991b) calls *legitimate language* – that is, situationally appropriate and socially valued discourse.

The notion of linguistic capital has primarily been employed at the level of whole societies, characterising the relative capital associated with different languages or dialects and their relationships with socioeconomic class. Despite Bourdieu's focus on *practical* competence, the wider literature on linguistic capital often seems to prioritise *grammatical* competence, typically focusing on the capital associated with proficiency in English in international and transnational contexts (Morrison & Lui, 2000; Roth, 2019; Silver, 2005). Observations about distinctions in practical competence exhibited by those with equal grammatical competence are rare, although Harrison (2009) notes one participant's observation that her fluent Philippine English was considered inferior to the Australian English used by others working in her institution.

In this article, I take a different approach to linguistic capital. This approach moves away from exploring distinct (varieties of) languages with reference to the macro-social category of class; as noted above, the anonymous and disembodied nature of online discourse means that such social categories are often hidden. Instead, I explore how different forms of expression in the same language can index symbolic capital in a more local context. In doing so, I bring the concept of linguistic capital in line with more recent trends in sociolinguistics, which increasingly move away from exploring macro-

social categories in favour of a focus on discursive performance and indexing of more local identities (Eckert, 2012).

In this section, I have emphasised the connections between Newsome-Chandler and Grant's (2023) resource model of powerful identity performance and Bourdieu's (1991b) linguistic form of symbolic capital. Both address the role of discourse in power and authority, each speaking to their respective disciplines of linguistics and sociology, and thus bringing unique perspectives on the same issue. I argue that applying both lenses, as I do in this article, can therefore bring new strengths to each. Linking the resource model of power into a more generalised theory of power prevents linguistic analysis from becoming siloed, more fully recognising the relational nature of power by acknowledging that language is one of many resources which may be reciprocally recognised to construct relations of power (albeit one that is of central importance in anonymous online contexts). Meanwhile, using a resource model of power to explore linguistic capital demonstrates how capital can operate in more local contexts, which so far remains unexplored, and gives practical competence – the production of situationally appropriate and valued discourse, as distinct from technical grammatical competence – the central role that Bourdieu (1991b) intended it to have.

## 4. Data and methods

The data for this study was taken from a large and long-running self-identified 'white nationalist' forum. *White nationalist* is one of many terms used to refer to the broad movement which is also commonly described as *far-right*, *white supremacist*, and *right-wing extremist*, whose members espouse racist and other illiberal views. Preferred terminology is not settled and has shifted over time, although many scholars currently favour *far-right* (Mudde, 2019), which I have used so far in this article. I have chosen, however, to use *white nationalist* to describe this particular forum community. I believe this term balances accuracy, by capturing the research subjects' own self-identification as well as the racially exclusionary nature of their beliefs, with the complex ethics of representing human data subjects whose views I find abhorrent (e.g. Pasieka, 2019). Such ethical concerns have also influenced my decision not to name the site from which my data is drawn. This offers a degree of protection to the site's users in the form of greater anonymity. However, it also recognises the dangerous nature of the community; by not publicly associating my name with that of the forum, I attempt to shield myself from the dangers facing researchers who become known to extremist communities (Conway, 2021).

The full corpus from which the data for this study is taken contains every post made to the forum from over 120,000 user accounts between 2001 and 2018. Here, however, I focus on a much smaller sample. This article is based on work from a much broader project reported in Booth (2023), where smaller samples were drawn from the full dataset. Full details of this process of data filtering can be found in Booth (2023), particularly pages 71-74, and the process is summarised below.

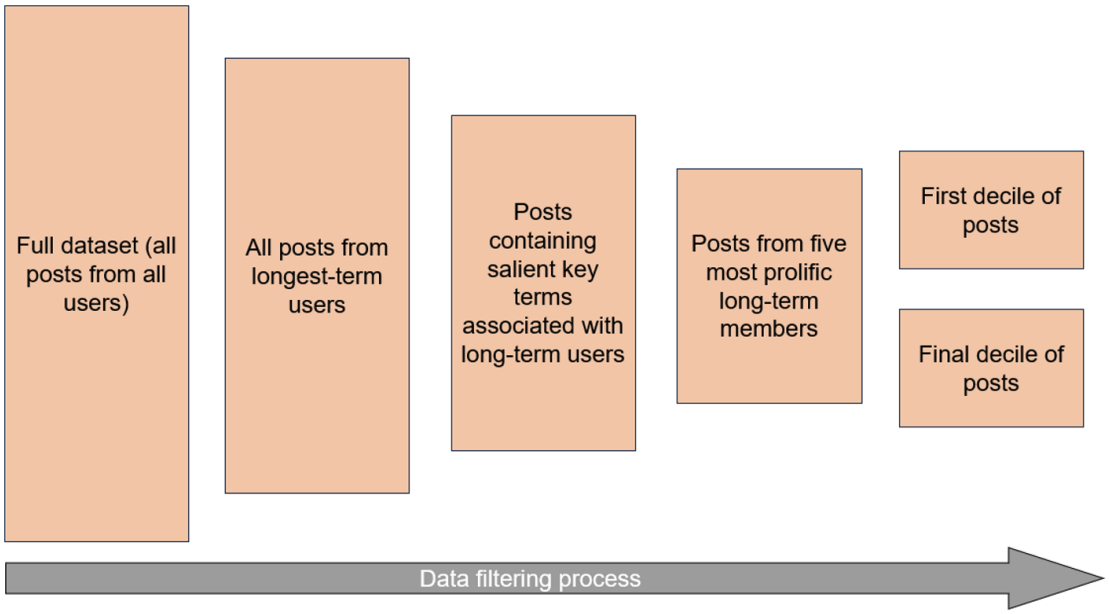



**Figure 1. Summary of data filtering process for the present study. Further details can be found in Booth (2023, pp. 71-74).**

The dataset was split into four groups of users based on the duration of their engagement with the forum. Using Sketch Engine, keyness analysis was used to identify terms and constructions that were particularly salient to the longest-term users by comparison to all other users. These terms were explored in the first and final chronological deciles of posts made by a sample of five long term users (that is, in the earliest tenth of these users' posts, and in the latest).

The posts in the present dataset therefore have the following characteristics:

- Written by one of the five most prolific members within the longest-term user category (that is, users who posted over a period of seven or more years);
- Within the first or final decile (tenth) of posts that these users made across the duration of their engagement with the forum;
- Contain one or more of the terms identified as salient for the purposes of an earlier study reported in Booth (2023).

Using this dataset, I was able to explore the difference between users' postings at the beginning and end of their engagement with the forum. As we have seen above, users acquire identity resources through sustained engagement with a particular social milieu which allow them to perform in more situationally appropriate, and therefore powerful, ways. Exploring the first and final deciles of users' posts therefore allows an insight into the resources of (linguistic) symbolic capital that they have accrued over time. To do so, I coded each post manually, focusing on the content and function of the discourse. This analysis was primarily inductive, allowing themes to emerge from the data. While Newsome-Chandler and Grant (2023) provide a valuable starting point for the identification of linguistic power resources, I do not use their typology of resources as a framework for deductive analysis here. Newsome-Chandler and Grant (2023, p. 129) themselves note that their typology is not intended to be seen as an "exhaustive list" of power resources even for the three fora they investigate, and is based on an exploratory study of small sets of data. My aim in this paper is to go beyond the resources

identified by Newsome-Chandler and Grant (2023), exploring how these may fit into a wider framework for the performance of power that incorporates both discursive and other resources. Nevertheless, some of the coding themes identified do, unsurprisingly, overlap with those reported elsewhere in the literature, including in Newsome-Chandler and Grant (2023).

Following annotation and compilation of results, I used the most frequently occurring themes to characterise each user's discourse at both the deciles. I then compared the two deciles for each user, identifying themes which were shared or unique to each decile. Finally, I compared themes from the first and final deciles across all five users in order to understand how users change over time. These convergences reveal what is considered to be situationally appropriate and socially valued discourse, and therefore allow us to understand the nature of power and linguistic capital in this forum.

## 5. Analysis

As expected, all five users display evidence of change in the linguistic and pragmatic resources which they exploit to engage with the community between the first and final decile of their posting careers. While each user naturally maintains a degree of individuality in their identity performance, a number of similar themes are identified across multiple users. The following analysis is organised around the major changes observed over time across users. In the subsequent Discussion, I consider what the findings reported here reveal about the nature of linguistic capital and authority in this white nationalist forum community.

### 5.1. Veteran users are less assertive in expressing their views

Contrary to expectations based on the literature (e.g. Ang & Zaphiris, 2010; Danescu-Niculescu-Mizil et al., 2013; Newon, 2011), veteran users produce less overt performances of expertise, authority, and commitment to their point of view. For most users, claims of knowledge or strong assertions are not entirely absent in the final decile, but are much less frequent, suggesting a lower level of concern among users with positioning themselves as knowledgeable and asserting their own authority by the later stages of their forum engagement.

The data suggests that new users draw on a much wider range of resources than veterans to appear knowledgeable. Several users offer evidence, including personal experience, for the claims they make. While this 'evidence' is sometimes of a dubious nature, its inclusion nonetheless shows users' concern with a perceived need for validation.

(1) *A study of history shows that England was as responsible for WW 1 as much as any country* (User 1, first decile)
(2) *The majority of real [nationality]*[1]*, and I know plenty of them, look nothing like Europeans but more like Arabs* (User 3, first decile)
(3) *I carry different guns depending where I'm going and so far have never had a problem with the cops* (User 4, first decile)

[1]Booth (2023) makes it clear that this user claims this nationality, so I also anonymise it here for ethical reasons.

Other resources used by new users in the construction of an authoritative, knowledgeable identity include the explicit assumption that their own view is surely correct (examples 4-5) and citing an example of having been right previously (example 6). User 1 positions themselves as an authority on the views of all white nationalists (example 7) or even all white people (example 8), and therefore able to speak on behalf of these groups.

(4) *We are on the side of the facts and conversley the facts are on our Side* (User 1, first decile)
(5) *when you know you're correct in what you're saying, (us)* (User 3, first decile)
(6) *I quietly predicted that the media would start the "child abuse" comments when P[russian] Blue*[2] *was on last week. I was one week too early.* (User 3, first decile)
(7) *We are White Nationalists and state senators do not represent us.* (User 1, first decile)
(8) *White people know from our everyday dealings with negros, arabs and hispanics that they do not measure up in everyway to White people* (User 1, first decile)

Giving advice to other users is also unexpectedly more common in the first than the final posting decile. The advice given is often tangential or even unrelated to white nationalist matters, but nonetheless serves to position users as knowledgeable.

(9) *You mean AOL as a service will not let you in to [forum]? Have you done a google search or so to let you to [forum]?* (User 2, first decile)
(10) *If you ever go to NH [New Hampshire], go to North Salem route 93 and you will find a place called America's Stonehenge* (User 4, first decile)
(11) *But if it's stabbing and slashing you want. I recommend the Rex Applegate/WE Fairbairn (German Boker-sounsen) double edge, SOG Daggert1*[3] (User 5, first decile)

This distinction between newcomer and veteran language use is typically reflected in a lower frequency of these authoritative themes in the final decile of posts. However, User 4 performs an explicitly uncertain or knowledge-seeking identity in their final decile of posts. This behaviour, which threatens the user's own face and position as an authority, is unexpected from a veteran user.

(12) *On the news they said it was nearly a 1,000 yard shot with a AK47, I'm thinking about this, can that really happen?* (User 4, tenth decile)
(13) *Maybe someone out there smarter than me can put more light on this.* (User 4, tenth decile)

The findings shared in this section are surprising in light of the literature, which suggests that knowledgeable and authoritative performances are typically the preserve of veteran users.

Given the selected users' very longstanding engagement with the forum (at least seven years), it seems possible that, by the final posting decile, these users feel that they no longer need to 'prove themselves'. When viewed on the forum website, each post is accompanied by metadata about its author's account, including, significantly, the date they joined the forum. Veteran users may therefore rely on their long tenure in the community as a kind of *institutionalised cultural capital*, endowed by the site's

[2]Prussian Blue is a former neo-Nazi musical duo whose members were young twin sisters.
[3]A variety of combat knife.

infrastructure rather than through their own linguistic performance. The importance of this long engagement is explicitly demonstrated by some users who draw on what Newsome-Chandler and Grant (2023, p. 116) call *veteran power*, deriving from "long-term membership of a forum and/or of the wider community or subject area." This is exclusive to the final posting decile for User 1, but appears in both deciles for User 4. In the first decile, User 4 employs this resource alongside others to perform a typically knowledgeable identity; its presence in the final decile, alongside a much less authoritative linguistic performance, suggests the power of time as a more institutionalised form of capital that users in their final posting decile have access to.

(14) *As you can see I have been posting on [forum] for 13 years, have joined several Nationalist groups and given away all kinds of money* (User 1, tenth decile)
(15) *Am I ready, you bet. I've been waiting for years.* (User 4, first decile)
(16) *I know I grew up a long time ago, maybe it was another world. [...] Obozo*[4] *wants change, but the country has changed way too much already.* (User 4, tenth decile)

### 5.2. Veteran users are less extreme in their expression

The previous finding concerned the assertion of an authoritative identity, which was unexpectedly more common among newcomers. Relatedly, the data also shows that users typically express more extreme views in the first decile of their posts (I define *more extreme*, following McCauley and Moskalenko (2017), as showing greater inclination to action, especially violent action, in the name of the white nationalist cause).

This typically takes the form of rallying other white nationalists, or white people more broadly, to various kinds of action, or of indicating that violence may be the only solution to the social problems that users perceive. User 1 (examples 17-19) is unique in sustaining this behaviour throughout both of the posting deciles explored here. However, as we will see below, in the final decile this is increasingly tied into a sense of despair at other white people and white nationalists and what this user perceives as these groups' inaction in the face of an existential threat. In the final decile, then, User 1's focus appears to be appeals for unity among the white (nationalist) community, with calls for violence secondary to this.

(17) *A White Homeland will not be achived unless we stop writing "our Congressman" and start taking matters into our own hands* (User 1, first decile)
(18) *The question is are we finally ready to unify and take action* (User 1, tenth decile)
(19) *The White race must march forward all at one time and totally united* (User 1, tenth decile)
(20) *there will only be one way to solve our problems as a race, [...] and that way is a very uncomfortable way.* (User 3, first decile)
(21) *I'm afraid that only blood will make these invaders go home and stay there.* (User 4, first decile)

User 4 is perhaps especially emphatic in their expression of extreme ideas; notice that example 21 is the most explicit advocation of violent action. This user also openly mocks less extreme actions such as voting, and strengthens their calls to action by appealing both to history and to readers' emotions.

[4]Former United States President, Barack Obama.

(22) *Off course if we become that lazy or chicken we could always vote for a 3rd. party that doesn't take much courage.* (User 4, first decile)
(23) *if the founders were here they would have already marched on Wasington with a gun in one hand and a rope in the other* (User 4, first decile)
(24) *our childern and grand childern will ask why we did nothing when we could have, and they will hate you.* (User 4, first decile)

As for the previous finding, this pattern suggests that veteran users may rely on their status as long-standing forum members, and no longer feel the need to overtly highlight their commitment to the white nationalist cause by performing a more extreme identity. This also aligns with findings reported below, suggesting that users display an increasingly casual relationship with the forum and its community over time, and/or have become increasingly jaded or disappointed with the movement and so no longer see the value in attempting to rally its members.

### 5.3. Veteran users are more sociable

For veteran users, the forum seems to take on new meaning as a more social, rather than solely ideological, space. This is often reflected in the tone of users' posts, which is typically more ironic or joking in nature in the final decile. While much of this content contains ideological elements, for example taking the form of racist jokes or mocking perceived political correctness, the tone is nonetheless more ironic than is typical among newcomers, suggesting that veteran forum members are increasingly interested in the 'fun', social side of interaction offered by the community.

(25) *What do you call a black man in a three piece suit?* (User 2, tenth decile)
(26) *Are you sure we can still use the word boy anymore? I mean no one wants to be a racist, well almost no one.* (User 4, tenth decile)
(27) *And we all know that the very worst thing that could happen to a White man is being called a racist or bigot.* (User 5, tenth decile)

Typically, this simply contrasts with an absence of such a tone in the early posting decile. In the case of User 5, however, this user often takes a highly solemn and serious tone in their earlier posts; in turn, this is absent in their final decile of posts.

(28) *The media will pound the masses with the evils of Secession. For this strikes at the very heart of the empire's existence. [...] The legs of the beast are shaking from the top heavy strain of bureaucrats and usurpers.* (User 5, first decile)
(29) *True, the cannons are silent, but the shadow of oppression still lingers in every nook and cranny of our Southern homeland.* (User 5, first decile)

Some users engage in 'off-topic' talk, focusing on popular culture or lifestyle. This discourse is often entirely non-ideological, or with only tangential connections to white nationalist ideologies. Although found in both the early and the late decile for some users, across the dataset this is more common in the final decile, and is unique to this decile for User 3. Such off-topic talk may serve to build rapport and stronger bonds between community members (Angouri & Sanderson, 2016).

(30) *I don't care for Michael Moore or his liberal politics,I'll wait and see this when it comes out on dvd.* (User 2, tenth decile)
(31) *I played full court basketball with a group of friends in a gym once or twice a week until I was thirty* (User 3, tenth decile)

(32) *Look at the label on these supplements, they say with a proper diet and exercise you will lose fat and weight.* (User 4, tenth decile)

User 3 presents a unique case among this sample of users. We have already seen that, as a newcomer, this user seeks to position themself as knowledgeable and informed (examples 2, 5, 6). As a veteran, this user's discourse is heavily characterised by hostility and ad hominem attacks against other members, often through the medium of debate. This debating context naturally entails self-positioning as knowledgeable and assertive. I argue, however, that contrary to User 3's behaviour as a newcomer, which has a broader function of representing themself as knowledgeable to readers, as a veteran, User 3 is much more strongly motivated by interaction and interpersonal relationships. Recall that User 3 also engages in off-topic talk in the later decile, further supporting this interpretation of a more interactional focus.

(33) *Who do you think you're kidding here* (User 3, tenth decile)
(34) *your position doesn't hold water, as you're showing bias and alterior motvies.* (User 3, tenth decile)
(35) *You're a pathetic, cowardly little man who is all talk.* (User 3, tenth decile)

Combined with the previously-reported finding of less violent discourse among veteran users, this trend suggests that, while users may be brought to the forum by a desire to discuss their white nationalist ideologies, longer-standing users may additionally seek a more social environment. Users' increasingly joking or ironic discourse may suggest a greater sense of comfort or ease within the community, as well as mirroring the decrease in violently extreme discourse. Together, these findings suggest that social capital plays a greater role in veteran users' engagement with the forum, with users' relationship to and position within the forum's social network becoming a more significant element of their identity performances.

### 5.4. Veteran users criticise their in-group

Unsurprisingly for a white nationalist forum, both newcomers and veterans express hateful views towards groups racialised as 'non-white'. However, there is a tendency for users to increasingly shift the focus of their criticism towards their own in-group(s) over time. This is particularly marked for User 1, whose hateful discourse towards racialised groups is almost entirely replaced by criticism of their fellow white people or even white nationalists.

(36) *there might be 500,000 muslims in this country right now that are willing to go on some kind of crime spree or commit terrorist acts.* (User 1, first decile)
(37) *They will start to harm White people from the minute they get there.* (User 1, first decile)
(38) *I am beginning to think there is some kind of fundamental flaw in White People that keeps them from reaching their full potential* (User 1, tenth decile)

Veteran criticism of the white in-group often centres around a perception that white people are apathetic to the alleged threat posed by racialised out-groups, and are therefore to blame for their own perceived oppression. In this way, white people with more liberal political or social views begin to be treated as a "functional out-group" (Berger, 2018, p. 63) in relation to white nationalist extremist in-group identity.

(39) *more than can be said for lazy and stupid Whites in America who are doing nothing towards protecting and defending their race.* (User 1, tenth decile)
(40) *I can't understand liberals and their attitudes supporting a subversive lifestyle.* (User 2, tenth decile)
(41) *Maybe that is the trouble with white folk, they don't speak up when we see something that is wrong.* (User 4, tenth decile)

Some veteran users' criticism extends to their fellow white nationalists, including fellow members of the forum, with frequent reference made to division within the white nationalist movement. We have already seen that User 1 increasingly focuses on unity within the movement as a prerequisite for action (examples 18-19), lamenting their perception that a divided movement will be ineffective in its goals. This user also blames this disunity for the perceived unpopularity of white nationalism among white people.

(42) *If WN [white nationalism] was strong and unified other Whites would begin to follow but they aren't because we don't have our own house in order* (User 1, tenth decile)

The focus on division within the white nationalist movement is also used to gatekeep white nationalist identity, as well as membership of this particular forum. By defining 'proper' conduct or beliefs for members of these groups, users implicitly position themselves as a 'good' member of the forum and of the wider white nationalist community, as well as having the power to exclude others who do not meet the standard they set. Complementing this behaviour, these veteran users also demonstrate their knowledge of forum rules and norms (examples 45 and 48). This serves as a further technique to gatekeep, while also performing a powerful identity with the authority to enforce rules.

(43) *WN in the U.S. can help the overall picture simply by being the best we can be which few of us are* (User 1, tenth decile)
(44) *If you approve of race mixing then you are on the wrong forum.* (User 1, tenth decile)
(45) *I appreciate your post however it is generally known that we speak and write in English on this forum.* (User 1, tenth decile)
(46) *Beware the ones among us that claim we're "white supremacists", they're not White Nationalists and are here to destroy us.* (User 3, tenth decile)
(47) *You obviously are posting on the wrong website and would be better served posting on a site which is more in line to your views.* (User 3, tenth decile)
(48) *Get over to OV [Opposing Views][5] where you belong, you antagonistic, phony WN.* (User 3, tenth decile)

In the final posting decile, some users even begin to criticise white nationalist organisation leaders or other high-profile figures in the movement. In the case of User 1, this is a direct reversal of the behaviour shown in the early decile, where they praise such organisations. The later criticisms, then, further demonstrate this user's apparent increasing disillusionment with the perceived state of the white nationalist movement. For other users, mention of these high-profile figures is simply absent in the first decile. The appearance of this behaviour among veteran users may suggest that these users feel

[5]The forum rules state that visitors who oppose white nationalist ideology must only post in the Opposing Views subforum.

more secure in their position within the forum, and therefore empowered to challenge the status quo by criticising powerful or popular figures within the white nationalist movement.

(49) *Consider joining the National Alliance.* (User 1, first decile)
(50) *I have a hard time understanding how Will Williams*[6] *can do so many things that are in bad taste and judgement* (User 1, tenth decile)
(51) *I wish something can transpire and move the NationalAlliance into a new direction. Gliebe*[7] *has got to go.* (User 2, tenth decile)
(52) *How dare we not like what Hitler did to our people and loved ones and be WN's.* (User 3, tenth decile)

## 6. Discussion and conclusions

The analysis showed that users' discourse becomes less assertive and violent in the tenth decile of posts. Fewer traditional discursive displays of authority and expertise are used to overtly emphasise the depth of users' commitment to the movement and its ideology, with users apparently taking their status in the forum for granted without feeling pressure to perform it linguistically. Over time, ideological discourse became increasingly accompanied by 'off-topic' talk and an ironic tone, suggesting that long-term users may not be solely concerned with or motivated by ideology, but also by the social relations they have established within the community. Where ideological discourse did appear in the final decile, its focus had frequently shifted to criticism of groups and individuals closer to the deictic centre of white nationalist identity, including other white people and white nationalists, often serving to gatekeep or negotiate the boundaries of these identities.

Comparing the findings reported here to the existing literature on newcomer and veteran language shows that members of this white nationalist community draw on some unexpected resources for their identity performances, in particular with regard to the performance of authority and commitment. This is particularly striking given the extremist nature of the forum, where we might expect to witness a process of radicalisation (Kenyon et al., 2021) resulting in the expression of increasingly extreme views over time. Instead, veteran performances of authority appear to have a pacifying effect on users' discourse, while newcomers express themselves with much more zeal. These new members are decidedly closer to Chiang's (2024) 'competent newbie' than to the deferential newcomer described by most scholars, and give traditional discursive performances of power drawing on typical power resources such as assertiveness and displays of knowledge.

As newcomers, however, users have little symbolic capital. By dint of their newness and inexperience, they do not have "a feel for the game" (Thompson, 1991, p. 13) of the norms of interaction in this particular social context, and therefore lack the resources to behave in a fully situationally appropriate and socially valued way. As a result, newcomers risk threatening their own face and that of others, and, to avoid this, are more likely to feel the need to impress others and 'prove themselves' as legitimate members of the community. In this forum, they do so by drawing on the resources which they

[6]Leader of the (now largely defunct) National Alliance neo-Nazi organisation.
[7]Erich Gliebe, former leader of the National Alliance.

*do* have: those which, as the literature tells us, are typical of powerful discourse across other contexts.

Veteran users, on the other hand, have greater familiarity and comfort with the community and its norms, and therefore have greater symbolic capital within it, distinct from the self- conscious use of traditional power resources by newcomers. Alongside the resources of linguistic capital which they have acquired over time through their participation in the community (Grant & MacLeod, 2020), veteran users are also able to claim other forms of symbolic capital. The date when a user joined the forum is presented beside every post they make, alongside their username and avatar. For each of the five users explored here, this date is at least seven years prior to their final decile posts, placing them in the longest-standing group of users. The join date therefore serves as a form of *institutional cultural capital*: a resource, derived from the infrastructure of the site, upon which veteran users can rely to demonstrate their status without needing to perform it linguistically. In addition, the more social nature of veteran users' discourse points to their *social capital*, suggesting that these users are more tightly enmeshed in a "durable network" (Bourdieu, 2004, p. 21) of their fellow forum members than they were as newcomers. While the existence of this network is reflected in users' humorous, off-topic, or ad hominem talk, this discourse alludes to a form of capital that goes beyond the purely linguistic, alluding to a social network and the user's position within it. Veterans, then, are able to rely on other forms of symbolic capital beyond the purely linguistic; they do not need to use typical power resources to perform their legitimacy within the community and have their capital and authority recognised. These users can therefore 'get away with' drawing on a wider range of resources without fear of threatening their own, or others', face (unless this is their purpose; consider User 3's overt hostility and insults). Individuals with symbolic capital have the power to set the standard of legitimate language (Bourdieu, 1991b); therefore, through association with capital, veterans' less traditionally powerful linguistic performances in fact become powerful in this forum context. In this way, forum veterans "valoriz[e] their own capital" (Bourdieu, 1991a, p. 7) and ensure their continued high status. These resources are unavailable to newcomers until they undergo what Bourdieu (1991a, p. 8) calls a "metamorphosis," a recognition and acceptance of the local "rules of the game" as perpetuated by veteran users.

In their review, Newsome-Chandler and Grant (2023) highlight the lack of availability of traditional power resources such as macro-social categories in anonymous online contexts, and argue for the primacy of discourse as an identity- and power-constructing resource in such spaces. Indeed, I have shown that, through time and participation, users are able to acquire new linguistic resources which allow them to perform power in a situationally appropriate way for the context of this white nationalist forum. However, I have also shown that, alongside the powerful *linguistic* resources with which Newsome-Chandler and Grant (2023) are concerned, veteran identity performances are also able to draw on non-linguistic elements – namely, institutional cultural capital and social capital – to construct powerful identities, and to use language to make these pertinent in a primarily linguistic environment.

This demonstrates the value of linking Newsome-Chandler and Grant's (2023) linguistic resource model of power (broadly construed as a model of power expressed through discursive resources, and not limited to the particular set of resources they

identify in their exploratory study) into the broader view of power offered by Bourdieu's (2004) notion of symbolic capital. Symbolic capital functions as a semiotic system: certain forms of cultural expression signify capital, allowing individuals to claim authority and to be recognised as powerful by others. Language, as a semiotic system *par excellence*, is naturally central to the expression of symbolic capital. While Bourdieu (1991b) introduced the idea of linguistic capital as a *practical competence* in producing situationally appropriate and socially valued utterances, this has remained empirically underexplored in subsequent literature. In this article, I have demonstrated both how power is constructed and deployed using linguistic resources, and how language may be used to highlight forum users' other, non-linguistic resources, such as institutionalised cultural capital and social capital. In this way, I argue, linguistic capital, expressed through the kinds of linguistic power resources introduced by Newsome-Chandler and Grant (2023), 'slots into' studies of symbolic capital in a hierarchical relationship, representing just one facet of the many to be explored within this general semiotic system of symbolic capital. While of particular importance in anonymous and pseudonymous online contexts such as the one I explore here, the ubiquity of language ensures that a consideration of linguistic capital stands to enrich any study of symbolic capital.

Beyond these theoretical implications, the findings reported here also have practical value, showing how power is performed in a large, long-running white nationalist forum and therefore adding to the body of knowledge used to prioritise powerful individuals in harmful communities. My analytical focus has primarily been linguistic, showing that socially-oriented discourse and criticism of white and white nationalist in-groups are associated with powerful performances, and that assertive performances of knowledge and violent ideation are, unexpectedly, not. However, I have also shown that other resources, such as the user's join date and their position in the forum's social network, may also be valuable starting points for prioritisation. The partially unexpected findings of this study, which is almost unique in both a) its longitudinal approach and b) its focus on a white nationalist community, show the value of these more authentic approaches for the identification of powerful and high-priority individuals in harmful online communities; the findings reported here could not have been predicted based on the existing literature.

## References

Ang, C. S., & Zaphiris, P. (2010). Social roles of players in MMORPG guilds. *Information, Communication & Society*, *13*(4), 592–614. Retrieved 2025-08-05, from http://www.tandfonline.com/doi/abs/10.1080/13691180903266952 doi: 10.1080/13691180903266952

Angouri, J., & Sanderson, T. (2016). 'You'll find lots of help here' unpacking the function of an online Rheumatoid Arthritis (RA) forum. *Language & Communication*, *46*, 1–13. Retrieved 2025-08-05, from https://linkinghub.elsevier.com/retrieve/pii/S0271530915000749 doi: 10.1016/j.langcom.2015.10.001

Benwell, B., & Stokoe, E. (2006). *Discourse and identity*. Edinburgh University Press.

Berger, J. M. (2018). *Extremism*. MIT Press.

Booth, A. (2023). *Collective identity and careers in a white nationalist forum* (PhD thesis). Aston University.

Bourdieu, P. (1991a). The peculiar history of scientific reason. *Sociological Forum*, *6*(1), 3– 26.

Bourdieu, P. (1991b). The production and reproduction of legitimate language (G. Raymond & M. Adamson, Trans.). In J. B. Thompson (Ed.), *Language and symbolic power* (pp. 43–65). Polity. (Reprinted from Ce que parler veut dire: l'économie des échanges linguistiques, pp. 23-58, by P. Bourdieu, 1982, Librairie Anthème Fayard.)

Bourdieu, P. (2004). The forms of capital. In S. J. Ball (Ed.), *The RoutledgeFalmer reader in sociology of education* (pp. 15–29). RoutledgeFalmer. (Reprinted from Handbook of theory and research for the sociology of education, pp. 241-258, by J. C. Richardson, Ed., 1986, Greenwood Publishing Group)

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

Bucholtz, M., & Hall, K. (2005). Identity and interaction: a sociocultural linguistic approach. *Discourse Studies*, *7*(4-5), 585–614. Retrieved 2025-08-05, from https://journals.sagepub.com/doi/10.1177/1461445605054407 doi: 10.1177/1461445605054407

Chiang, E. (2024). "I read the rules and know what is expected of me": The performance of competence and expertise in 'newbie' offenders' membership requests to dark web child abuse communities. *Discourse, Context & Media*, *57*, 100744. Retrieved 2025-08-05, from https://linkinghub.elsevier.com/retrieve/pii/S2211695823000776 doi: 10.1016/j.dcm.2023.100744

Conway, M. (2021, February). Online Extremism and Terrorism Research Ethics: Researcher Safety, Informed Consent, and the Need for Tailored Guidelines. *Terrorism and Political Violence*, *33*(2), 367–380. Retrieved 2025-08-05, from https://www.tandfonline.com/doi/full/10.1080/09546553.2021.1880235 doi: 10.1080/09546553.2021.1880235

Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., & Potts, C. (2013). No country for old members: User lifecycle and linguistic change in online communities. In *22nd international conference on world wide web.* Rio de Janeiro, Brazil.

Eckert, P. (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*, *41*(1), 87–100. Retrieved 2025-08-05, from https://www.annualreviews.org/doi/10.1146/annurev

-anthro-092611-145828 doi: 10.1146/annurev-anthro-092611-145828

Europol. (2023). *European Union terrorism situation and trend report 2023.* (Tech. Rep.). Retrieved from https://www.europol.europa.eu/publication-events/main-reports/european-union-terrorismsituation-and-trend-report-2023-te-sat

Golder, S. A., & Donath, J. (2004). Social roles in electronic communities. In *Internet research 5.0.* Brighton, UK.

Graham, S. L. (2016). Relationality, friendship, and identity in digital communication. In A. Georgakopoulou & T. Spilioti (Eds.), *The Routledge handbook of language and digital communication* (pp. 305–320). Routledge.

Grant, T., & MacLeod, N. (2020). *Language and online identities: The undercover policing of internet sexual crime.* Cambridge University Press.

Harrison, G. (2009). Language politics, linguistic capital and bilingual practitioners in social work. *British Journal of Social Work*, *39*(6), 1082–1100. Retrieved 2025-08-05, from https://academic.oup.com/bjsw/article-lookup/doi/10.1093/bjsw/bcm153 doi: 10.1093/bjsw/bcm153

Honeycutt, C. (2005). Hazing as a process of boundary maintenance in an online community. *Journal of Computer-Mediated Communication*, *10*(2), Article JCMC1021.

Institute for Economics and Peace. (2023). *Global terrorism index 2023: Measuring the impact of terrorism* (Tech. Rep.). Retrieved from https://www.visionofhumanity.org/wp-content/uploads/2023/03/GTI2023-web-170423.pdf

Kenyon, J., Binder, J., & Baker-Beall, C. (2021). *Exploring the role of the internet in radicalisation and offending of convicted extremists.* Retrieved from https://www.gov.uk/government/publications/exploring-the-role-of-the-internet-in-radicalisation-and-offending-of-convicted-extremists

Levin, B. (2002). Cyberhate: A legal and historical analysis of extremists' use of computer networks in America. *American Behavioural Scientist*, *45*(6), 958–988.

McCauley, C., & Moskalenko, S. (2017). Understanding political radicalization: The two-pyramids model. *American Psychologist*, *72*(3), 205–216. Retrieved 2025-08-05, from https://doi.apa.org/doi/10.1037/amp0000062 doi: 10.1037/amp0000062

Mondon, A., & Winter, A. (2020). *Reactionary democracy: How racism and the populist far right became mainstream.* Verso.

Morrison, K., & Lui, I. (2000). Ideology, linguistic capital and the medium of instruction in Hong Kong. *Journal of Multilingual and Multicultural Development*, *21*(6), 471–486. Retrieved 2025-08-05, from http://www.tandfonline.com/doi/abs/10.1080/01434630008666418 doi: 10.1080/01434630008666418

Mudde, C. (2019). *The far right today.* Polity.

Newon, L. (2011). Multimodal creativity and identities of expertise in the digital ecology of a World of Warcraft guild. In C. Thurlow & K. Mroczek (Eds.), *Digital discourse: Language in the new media* (pp. 131–153). Oxford University Press.

Newon, L. (2016). Online multiplayer games. In A. Georgakopoulou & T. Spilioti (Eds.), *The Routledge handbook of language and digital communication* (pp. 289–304). Routledge.

Newsome-Chandler, H., & Grant, T. (2023). Developing a resource model of power and authority in anonymous online criminal interactions. *Language and Law / Linguagem e Direito*, *10*(1), 110–130. Retrieved 2025-08-05, from https://ojs.letras.up.pt/index.php/LLLD/article/view/12841/12478 doi: 10.21747/21833745/lanlaw/

10_1/a4

Nguyen, D., & Rosé, C. P. (2011). Language use as a reflection of socialization in online communities. In *11th workshop on languages in social media.* Portland, OR, USA.

Pasieka, A. (2019). Anthropology of the far right: What if we like the 'unlikeable' others? *Anthropology Today*, *35*(1), 3–6.

Pitcavage, M. (2023). *Murder and extremism in the United States in 2022.* Retrieved from https://www.adl.org/resources/report/murder-and-extremism-united-states-2022

Roth, S. (2019). Linguistic capital and inequality in aid relations. *Sociological Research Online*, *24*(1), 38–54. Retrieved 2025-08-05, from https://journals.sagepub.com/doi/10.1177/1360780418803958 doi: 10.1177/1360780418803958

Rudolf von Rohr, M.-T., Thurnherr, F., & Locher, M. A. (2019). Linguistic expert creation in online health practices. In P. Bou-Franch & P. Garcés-Conejos Blitvich (Eds.), *Analyzing digital discourse: New insights and future directions* (pp. 219–250). Palgrave Macmillan.

Silver, R. E. (2005). The discourse of linguistic capital: Language and economic policy planning in Singapore. In *Language Policy* (Vol. 4, pp. 47–66).

Singh, V. (2012). Newcomer integration and learning in technical support communities for open source software. In *17th ACM International Conference on Supporting Group Work.* Sanibel Island, FL, USA.

Thompson, J. B. (1991). Editor's introduction. In J. B. Thompson (Ed.), *Language and symbolic power* (pp. 1–31). Polity.

*Language and Law / Linguagem e Direito*, *Vol. 11*(2), 2024, pp. 135-159
DOI: doi.org/10.21747/21833745/lanlaw/11_2a6

# Ideologies of a serial bomber: A longitudinal linguistic appraisal analysis of the writings of the Unabomber

Madison Hunter [1]

[1] *Aston University, Aston Institute for Forensic Linguistics, Birmingham, UK*, m.hunter5@aston.ac.uk

***Abstract***
*Forensic linguistic and psychological assessments of language evidence each offer useful information about authors, but conclusions reached by one are often considered beyond the scope of the other. This paper explores an interdisciplinary approach intended to help bridge the gap between the two which uses the Appraisal framework as the analytical method but examines the findings through the lens of cognitive psychological theory. The connection between the two disciplines is made through the shared concept of 'schemas' – beliefs that impact the interpretation of and response to incoming information. More specifically, this study uses the linguistic operationalization of schemas – i.e., something observable through language choices – alongside the cognitive psychological conceptualization – as underlying different psychological traits and symptoms and producing evaluative/perceptual biases. It is argued that the impact of schemas on one's experience will be reflected in the language one uses to describe and evaluate aspects of that experience. The analysis is conducted on the writings of Ted Kaczynski – the Unabomber – which span almost 30 years and are separated into three time periods: (1) before likely onset of psychiatric symptoms; (2) after onset but before voluntary self-isolation; and (3) just before his first bombing to a few years before his arrest. This dataset affords a unique opportunity to examine not only the potential relationship between language and psychopathology, but also how it may change over time. Results suggest a connection between stance-taking patterns and schemas underlying different psychological traits/symptoms, with some changes over time being observed. While further research is needed, the findings provide evidence in support of an interdisciplinary approach like the one explored here.*
***Keywords:*** *Appraisal analysis; forensic text analysis; stance; psychopathology.*

***Resumo***
*A análise linguística forense e psicológica das provas linguísticas oferecem informações úteis sobre os autores.*

*Contudo, as conclusões alcançadas por cada uma delas são frequentemente consideradas fora do âmbito uma da outra. Este artigo explora uma abordagem interdisciplinar destinada a ajudar a colmatar o fosso entre estes tipos de análise, utilizando o quadro de avaliação como método analítico e examinando; porém, optou-se por analisar as conclusões através da lente da teoria psicológica cognitiva. A ligação entre as duas disciplinas é feita através do conceito partilhado de "esquemas" - crenças que têm impacto na interpretação e na resposta à informação recebida. Mais especificamente, este estudo utiliza a operacionalização linguística dos esquemas - ou seja, algo observável através das escolhas linguísticas - juntamente com a concetualização psicológica cognitiva - como estando subjacente a diferentes traços e sintomas psicológicos e produzindo enviesamentos avaliativos/percetivos. Defende-se que o impacto dos esquemas na experiência de uma pessoa se reflecte na linguagem que esta utiliza para descrever e avaliar aspetos dessa experiência. A análise é efetuada com base em textos escritos de Ted Kaczynski - o Unabomber - produzidos ao longo de 30 anos e que estão divididos em três períodos de tempo: (1) antes do aparente surgimento de sintomas psiquiátricos; (2) após este surgimento, mas antes do auto-isolamento voluntário; e (3) imediatamente antes do seu primeiro atentado bombista e alguns anos antes da sua detenção. Este conjunto de dados oferece uma oportunidade única para examinar não só a potencial relação entre a linguagem e os sintomas psicopatológicos, mas também como esta se pode alterar ao longo do tempo. Os resultados sugerem uma ligação entre os padrões de posicionamento e os esquemas subjacentes a diferentes caraterísticas/sintomas psicológicos, tendo sido observadas algumas alterações ao longo do tempo. Embora seja necessária mais investigação, os resultados fornecem provas que apoiam uma abordagem interdisciplinar como a que foi aqui explorada.*

***Palavras-chave:*** *Avaliação, análise de autoria forense, posicionamento, psicopatologia.*

## 1. Introduction[1]

In May of 1978, a package was left in a parking lot on the University of Illinois at Chicago campus with the return address of a Northwestern University professor. Upon it being returned to him, the professor claimed he had not sent the package and reported it to security. When it was later opened, it exploded, injuring the security guard's hand (Fitzgerald, 2004). Over the next 17 years, 15 more bombs would be planted by the *Unabomber*—a moniker created from the associations of early victims, i.e., 'universities and airlines'—causing a total of three deaths and 23 injuries and sowing widespread panic across the United States (Federal Bureau of Investigation, n.d.). For much of that 17-year period, there was a paucity of forensic evidence, which resulted in few leads

[1]This paper has been adapted from a chapter of my doctoral thesis (Hunter, 2022).

and little substantive progress in the investigation. Beginning in 1993, however, the bomber began sending letters to various recipients expounding his ideologies and in 1995, he mailed a 35,000-word manifesto to the *Washington Post* and *New York Times*, demanding it be published (Fitzgerald, 2004). After publication, a man named David Kaczynski reported suspicions to the FBI that his brother might be the author, and he and his mother later provided investigators with letters the brother had sent them throughout the years to compare linguistically to the Unabomber writings. There were a striking number of similarities between the two sets of documents, which proved to be a crucial piece of evidence in securing the search warrant that, when executed, ended the 17-year manhunt with the arrest of the Unabomber, Ted Kaczynski.

During the nearly two-decade investigation, numerous versions of a *behavioral* profile had been generated and just before Kaczynski's arrest, the FBI also enlisted the services of a forensic linguist to generate a *linguistic* profile (Shuy, 2014). A *behavioral* profile – sometimes referred to as a psychological or criminal profile – typically compares aspects of a crime to past crimes of a similar nature to determine potential behavioral or psychological traits of the author (Douglas, Ressler, Burgess, & Hartman, 2004). *Linguistic* profiles, conversely, involve analyzing and describing how features of a suspect's language use compare to features identified in previous sociolinguistic research to be "characteristic of specific societal groups" (Shuy, 2014, pp. 76-77).

The behavioral profiles generated about the Unabomber varied, sometimes drastically, but ultimately theorized that the perpetrator was middle-aged and from the Chicago area (which were accurate), was high school educated, potentially with some university experience (which was inaccurate) and was a "loner with low self-esteem" who had a "strong sense of superiority" (Shuy, 2014, p. 81). The linguistic profile had similarly hypothesized the Unabomber was older and from Chicago but predicted a much higher education level and speculated about aspects of religious background and other geographical background information (all of which turned out to be accurate).

In this case, none of the profiles that were generated ultimately contributed to the identification of Kaczynski – his brother's tip and the subsequent comparative analysis of the Unabomber manifesto/letters and Kaczynski's known writings provided the necessary evidence in the end (Fitzgerald, 2004). Despite this, the case does demonstrate how combining two distinct approaches can form a more complete picture of the perpetrator as each one may yield information that the other does not. More to the point, it arguably represents an excellent case in which to explore the potential benefits of an interdisciplinary approach which merges aspects of both methodologies and bridges the gap between them.

One such possible approach is proposed in this paper; the analytical method is based in linguistic theory and the identified language patterns are examined through the lens of a psychological theory to interpret how they may relate to the author's underlying psychopathology (i.e., their collective set of psychological traits). The reason Kaczynski's case represents a particularly useful testing ground is that there is information available on at least parts of the profiles generated in the investigation (Shuy, 2014), a report from a psychiatric competency assessment performed during his trial (Johnson, 1998), and a timeline of major events in his life (including changes in his mental health). This means:

1. the findings of the analysis conducted here can be examined in reference to the past profiles to see what value may have been added;
2. the relationship between the language patterns and specific psychological traits can be explored; and
3. the language patterns can be examined at different time points to determine whether changes in the patterns may correspond to changes in Kaczynski's mental health or major life events (e.g., his first bombing).

The next section contains an overview of the theoretical framework (both linguistic and psychological) used in this study and explains how the relationship between language and psychopathology is conceptualized and operationalized. After that, a brief introduction to Kaczynski's background and the findings of the psychiatric competency assessment (Johnson, 1998) is provided before introducing the data and methods. In the final sections of this paper, the key findings are outlined and their implications are discussed (primarily in terms of the potential utility of the proposed approach in forensic contexts).

## 2. Theoretical Framework

The base assertion underlying this research is that an individual's psychopathology impacts their lived experience – i.e., how they interpret information and interact with the world (Beck, 2015; Bortolan, 2019) – and language is a means through which that experience can be shared with others. More specifically, it is argued that the language used to describe and evaluate (i.e., express feelings/attitudes – or *stances* – about; Biber, Johansson, Leech, Conrad, & Finegan, 1999) one's experiences represents one possible way of examining, albeit indirectly, the underlying psychological processes that impact them. This notion is not completely unfounded, though the nature of the relationship between language and psychopathology is inherently complex and has been a topic of debate in the literature. At one extreme is the argument that certain linguistic features—such as lexical categories or syntactic complexity—are indicative of aspects of mental health like symptom/trait severity, functional impairments, or personality traits (e.g., Buck & Penn, 2015; Gawda, 2013; Pennbaker & King, 1999). At the other extreme is the argument for a less direct relationship; language facilitates social interaction and linguistic choices therefore more likely reflect the social goals of the language user in a given context than any underlying cognitive processes (Edwards & Potter, 1992, 1993).

The viewpoint taken here combines aspects of the aforementioned perspectives – that is, language used when recounting one's experiences will be sensitive to, and therefore impacted by, several factors which influence those experiences, including the social/contextual (e.g., audience or genre) and the psychological (e.g., psychological traits). To explore this empirically, though, requires an interdisciplinary theoretical framework as neither psychology nor linguistics offers one that comprises all the necessary components. For this research, that framework was created using features of phenomenological psychopathology, cognitive psychology, and systemic functional linguistics (SFL).

Phenomenological psychopathology emphasizes the value of analyzing first-person accounts of lived experiences, arguing that they allow one to observe "the various ways in which everyday experience can be disrupted" or altered by psychopathology (Bortolan, 2019, p. 1054). This helped to create the data selection criteria to ensure the

language that was analyzed was appropriate for the intended task. Cognitive psychology then provides a way to specifically characterize the impact of psychopathology by describing the underlying *schemas* (beliefs and assumptions) responsible for disrupting and altering experience (e.g., Beck & Haigh, 2014). Different schemas are thought to underlie different mental health disorders, traits, and symptoms (e.g., Beck, 2015), but they are themselves considered abstract and therefore not directly observable. They instead can be observed through, for instance, self-report measures or assessment scales and sometimes through language output – though the focus here tends to be on thematic content rather than patterns of linguistic resources and choices (Beck, Freeman, & Davis, 2015).

This represents the first piece of the puzzle, broadly connecting psychological characteristics to language output. However, the psychological approach to the analysis of the language is insufficient for our purposes here in that it does not offer the level of detail about the language patterns that we want. For that, we turn to the linguistic perspective of schemas – which serves as the bridge between the psychological and linguistic theories used here – and SFL, which supplies the analytical method. In linguistics, schemas have broadly the same definition – i.e., they are beliefs/assumptions that influence how one interprets and responds to incoming information (e.g., Shuy, 2015) – and are considered observable through analyzing patterns of linguistic choices. This represents the other piece of the puzzle but is in its own way lacking in that linguistic schemas are**not** tied to any psychological characteristics. Therefore, to fully explore the relationship between language and psychopathology, we need a combination of the two perspectives; namely, the *conceptualization* from psychology (which links the schemas with psychological traits) and the *operationalization* from linguistics (which links schemas with language output).

What SFL offers is a way of approaching the analysis that focuses on the linguistic resources that may be used to describe and evaluate one's experiences, which as argued above, would represent one possible observable manifestation of schemas. To be more specific, this research makes use of the Appraisal framework (J. R. Martin & White, 2005), which is comprised of three interacting systems encapsulating the various resources that are chosen from when expressing stances. The first system – *attitude* – captures the resources for expressing the feelings and judgments at the core of the stances. The second system – *engagement* – captures resources used to position one's views with respect to possible alternatives and convey commitment to one's own positions. The third system – *graduation* – acts as an umbrella over the other two, capturing the resources for scaling instances of each (i.e., increasing/decreasing the intensity of the feeling or the level of commitment to the position).

In using the Appraisal framework, it is possible to gain an immensely detailed account of the patterns of linguistic resources that an individual uses to discuss and evaluate themselves, others, and their world/experiences. These patterns can then be examined through the lens of the cognitive psychological literature to determine whether there are any clear connections to the schemas associated with different psychological traits and symptoms. A more comprehensive overview of the framework is provided below in section 4 after the discussion of Kaczynski's psychological background and the cognitive psychological literature relating to it.

## 3. Psychological Background

Kaczynski was born in Chicago in 1942. He started university at the age of 16, obtaining his graduate and doctoral degrees in mathematics by the age of 25, and subsequently getting hired by the University of California-Berkeley before abruptly resigning two years later in 1969 ("Unabomber (Ted Kaczynski)", 2018). In 1971, he moved to an isolated cabin in the Montana woods, from which he constructed the bombs he used in his 16 attacks spanning from 1978 until his arrest in 1995. After his arrest, Kaczynski's defense team hired a psychiatrist to conduct an evaluation of his competency to stand trial. The psychiatrist concluded that Kaczynski was indeed competent, but noted in the report that he exhibited signs of the (now bygone) paranoid type of schizophrenia and paranoid personality disorder (PPD; Johnson, 1998). Johnson opined that the PPD was likely present *before* the onset of the schizophrenia symptoms, which she estimated to have been around 1966, when Kaczynski was 24 years old and completing his graduate studies.

Johnson noted that while the severity of Kaczynski's symptoms fluctuated over time, he did not appear to ever experience full remission and that exacerbations of symptom severity had been preceded by "depressed mood, insomnia, increased distractibility, and intensification of sexual identity problems" (Johnson, 1998, p. 45). She also reported various symptoms with which Kaczynski presented, including impaired social and occupational functioning—which was evidenced by his lack of close friendships and romantic relationships as well as abrupt resignation from his faculty position in 1969 and later decision to isolate himself in his cabin. The predominant symptom, though, was the presence of delusions centered around two main themes: (1) believing he was being controlled by technology and (2) that experiencing psychological verbal abuse by his parents led to his interpersonal dysfunction.

For context, a brief introduction to the relevant literature relating to the diagnoses Kaczynski received is warranted. The fifth edition of the *Diagnostic and Statistical Manual* (DSM-5; American Psychiatric Association, [APA], 2013) is used here. It provides an idea of what symptoms/traits may be present with the two diagnoses Kaczynski received and contains the most up-to-date information about them as of this writing. (An earlier edition was used when Kaczynski was evaluated, and this was consulted to ensure that the criteria were not significantly different from that in the *DSM-5*.) Based on this and Johnson's (1998) report, the relevant schemas from cognitive psychological research can be explored in more depth. The two diagnoses are discussed in the following subsections, starting with the diagnostic criteria before outlining core schemas and processing biases associated with the diagnoses and their component symptoms/traits.

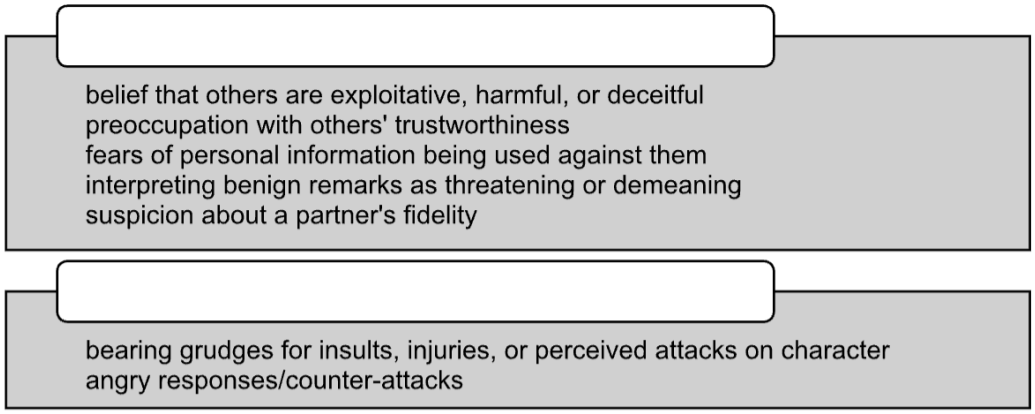
### 3.1. Paranoid Personality Disorder (PPD)

Personality traits are "enduring patterns of perceiving, relating to, and thinking about the environment and oneself that are exhibited in a wide range of social and personal contexts" (APA, 2013, p. 647). Each trait is accompanied by schemas that impact interpretation of and response to various situations (Beck, 2015). 'Pathological' traits (which comprise personality disorders, or PDs), are *maladaptive* and *inflexible* – in contrast to 'normal' traits which are adaptive and flexible – in that their role in the interpretation and response will remain the same even in situations for which they are inappropriate

(Beck, 2015, p. 24). For example, heightened 'competitiveness' might be adaptive for an athlete trying to win a game, but might be considered maladaptive if it causes someone to alienate their friends by trying to best them in every activity.

Paranoid personality disorder (PPD) is marked by by a "pervasive distrust and suspiciousness of others such that their motives are interpreted as malevolent" (APA, 2013, p. 649). In the *DSM-5*, the seven categorical (or behavioral) diagnostic criteria can all be subsumed under the pathological personality traits of *suspiciousness* and *hostility* (something suggested in past research; e.g., Beck, 2015; Millon, Millon, Meagher, Grossman, & Ramnath, 2012). For the purposes of this study, it is more useful to focus on these pathological personality traits as they relate more clearly and directly to the processes that might be impacted by the underlying schemas than the behavioral criteria do. How the behavioral criteria relate to the two pathological traits is shown in Figure 1.

**Figure 1. Relation between pathological traits and categorical criteria for PPD**
**Adapted from APA (2013, p. 649)**




On two scales or belief questionnaires used to assess the presence of schemas associated with personality traits and disorders, subscales were proposed and validated for PPD which center around the heightened suspiciousness and mistrust of others and their intentions (Arntz, Dreessen, Schouten, & Weertman, 2004; Beck et al., 2001; Bhar, Beck, & Butler, 2012; Fournier, DeRubeis, & Beck, 2012). For instance, holding the belief that others aim to take advantage of them; that they must always be on guard; or that when others are friendly, they are trying to manipulate them (Bhar et al., 2012; Fournier, 2015). Individuals with PPD may view themselves as righteous and clever, but at the same time vulnerable to mistreatment (Beck, 2015). Others, conversely, are viewed as mischievous, deceptive, manipulative, and exploitative (Beck, 2015; Renton & Mankiewicz, 2015), resulting in hypervigilance, often seeing threats in even the most benign actions (Renton & Mankiewicz, 2015, p. 248). As a result, PPD is often accompanied by anger or constant anxiety (Beck, 2015). These beliefs are also often held with a high degree of conviction, though not necessarily to the extent of a full-blown *delusion* (Renton & Mankiewicz, 2015) – a fixed belief considered incontrovertibly true regardless of contrary evidence (APA,2013).

Studies examining the impact of PPD specifically on cognitive processes are limited, though the effects of paranoia and persecutory delusions (like Kaczynski's) – i.e., beliefs that others are intent on causing one harm (APA,2013) – have been investigated. It has

been shown that similar reasoning biases are present in PPD as in persecutory delusions (Thompson-Pope & Turkat, 1988), which are discussed in the next subsection.

### 3.2. Schizophrenia

Schizophrenia is believed to exist on a spectrum with other psychotic disorders. For diagnosis in the *DSM-5* (APA,2013), abnormalities in at least two of five domains associated with psychotic disorders must be present – one must be delusions, hallucinations, or disorganized thought/speech – alongside impairment in social, occupational, and/or personal functioning (p. 99). The diagnosis of the paranoid type of schizophrenia from the earlier editions of the *DSM* – which was removed from the *DSM-5* – had similar criteria, but it specifically required the presence of one or more delusions or hallucinations (APA, 2000). As stated above, Kaczynski showed evidence of at least two delusions as well as impairment in social and occupational functioning (Johnson, 1998), though here we are focused on the delusions and their associated schemas.

Delusions take time to fully solidify and even after they have, they can still be elaborated upon and refined and an individual's investment in them can vary over time (Fineberg et al., 2015; Freedman, 2010). The impact on the interpretation of incoming information from the schemas underlying delusions helps to reinforce the beliefs and ultimately solidify and maintain the delusions (Beck & Rector, 2005; V. Bell, Halligan, & Ellis, 2006; Garety, Kuipers, Fowler, Freeman, & Bebbington, 2001). The content of these schemas is then argued to be evident in the themes of an individual's delusions (Beck & Rector, 2005). *Persecutory* delusions – like those noted for Kaczynski (Johnson, 1998) – are most common (APA,2013) and involve the individual believing that others intend to harm or harass them. The paranoid ideation central to this type of delusion is also present in PPD – though usually at a slightly lower intensity (Renton & Mankiewicz, 2015).

Paranoid ideation and persecutory delusions have been found to be associated with externalizing and reasoning biases (e.g.,Kinderman & Bentall, 1997; Langdon, Corner, McLaren, Ward, & Coltheart, 2006; Langdon, Ward, & Coltheart, 2010; J. A. Martin & Penn, 2002; Murphy, Bentall, Freeman, O'Rourke, & Hutton, 2018). The former result in an increased tendency to blame external circumstances or other people for negative events (e.g., Bentall, Corcoran, Howard, Blackwood, & Kinderman, 2001; Langdon et al., 2006, 2010; J. A. Martin & Penn, 2002), which is thought to contribute to the creation and maintenance of the paranoid world view (Bentall et al., 2001). The latter include a greater tendency to 'jump to conclusions' after minimal information (e.g., Garety et al., 2005; Langdon et al., 2010) and deficits in *theory of mind* (ToM), which is the ability to infer others' intentions (e.g., Langdon et al., 2006, 2010). Bentall et al. (2001) posited that the ToM deficit might result in a greater chance of attributing negative characteristics to others for negative events over circumstances because of a diminished capacity for considering alternative explanations.

The aim of the next few sections is to present the data and findings of the current study, and then discuss how those findings might align with the information presented in this section. No specific hypotheses are posited as this study is exploratory and schemas could theoretically manifest through myriad combinations of resources – e.g., **others are out to get me** could present through describing others' actions (*they chased me*) or one's own emotional states (*I felt uneasy walking home*). Thus, the patterns of

linguistic resources identified in the analysis are presented first and then explored to see how they relate to the possible schemas and associated information processing biases discussed above.

## 4. Appraisal Analysis

As mentioned above, the Appraisal framework consists of three interacting systems, each encompassing the linguistic resources for a different aspect of a stance (J. R. Martin & White, 2005). The first system, *attitude*, encompasses the resources for expressing personal emotions and societally-shaped assessments of people and 'things'. Attitudinal resources point to the kinds of feelings the author has about themselves, other people, and aspects of their experience, which constitute a core aspect of cognitive schemas (e.g., Beck, 2015; Beck & Haigh, 2014). Thus, not only do they help identify the schematic content, but they also help distinguish between schemas which may take similar linguistic form but differ in the core sentiment and direction of the belief (Beck et al., 2015). The second system, *engagement*, comprises the resources for encoding commitment to or certainty about a proposition and positioning of the author with respect to other persons and viewpoints, allowing observation of how authors frame their evaluations and engage with their audiences (J. R. Martin & White, 2005). *Engagement* resources helps identify, for instance, the views the authors hold most strongly, the views they expect others hold, and the knowledge they expect others to share with them, which may help identify which schemas are most active. Finally, the system of *graduation* covers the resources for adjusting the degree of commitment/certainty of positions expressed via engagement, the intensity of *attitudes*. For the purposes of this paper, the focus is placed on *attitude* and *engagement* with reference to the scaling within each system when relevant in the discussion of each (as opposed to a separate section on *graduation*).

Within attitude, there are three main types: (1) *affect* for personal emotions, (2) *judgment* for institutionalized assessments of behaviours and traits of oneself or others, and (3) *appreciation* for institutionalized assessments of objects and phenomena (J. R. Martin & White, 2005). *Affect* has four subcategories: *un/happiness* for emotions relating to "affairs of the heart (J. R. Martin & White, 2005, p. 49) like *love/hate* or *happy/sad*; *dis/satisfaction* for emotions about activities we participate in or watch like impressed/angry; in/security for emotions about our environment or confidence in our knowledge like *comfortable/uneasy* or *convinced/unsure*; and *dis/inclination* for emotions about irrealis things (i.e., things that have not happened yet). Judgment has five subcategories: *normality* for assessments of how normal/abnormal someone is; *capacity* for assessments of physical/mental capabilities (and assessments of removing or providing capacity like *kill/assist*); *tenacity* for assessments of determination, dependability, and willpower; veracity for assessments of honesty; and *propriety* for assessment of ethics/-morals. Finally *appreciation* has three main subcategories: *reaction* for assessments of the feelings evoked by things (similar to *affect*, but with the focus placed on the object like a *captivating story*); *composition* for assessments of how well put together something is; and *valuation* for assessments of the social value of the stance object. In line with Hurt (2020), *valuation* has been expanded here to include the categories of *judgment* as the same values can be assigned to things as well as people; for instance, one may evaluate a person as *truthful* (+veracity), but they may also place the focus on a proposition instead, saying X is *untrue* in which case it is *-veracity* via valuation. All to-

kens of *attitude* are also coded for four other features: *polarity* (positive versus negative), *explicitness* (explicit/denotation versus implicit/connotation), *appraiser* (from where the feeling originated), and *appraised* (at whom/what the feeling is directed).

Within *engagement*, there are two main types of utterances: *heteroglossic*, in which alternatives are referenced by either opening the dialogic space to those alternatives (*expanding*) or closing it off from them (*contracting*); and *monoglossic* in which no alternatives are referenced (J. R. Martin & White, 2005). In *expanding* utterances, an author may either *entertain* their proposition as one of many (e.g., *I might/may/could probably do X*) or *attribute* the proposition to someone else (e.g., *A claims/says/believes X*). In *contracting* utterances, an author may proclaim a proposition is the most warrantable by *pronouncing it* (e.g., *The truth is X*), concurring with it (e.g., *We all know X*), or *justifying* it (e.g., *X is true because*); they may also achieve a similar effect by *disclaiming* alternatives through either *denying* the proposition (e.g., *X is not true*) or presenting a *counter* to it (e.g., *But X is the case*).

## 5. Data and Methods

For this study, Kaczynski's evaluative language patterns are analyzed *over time*. The writings were pulled from a copy of larger collection of writings compiled by his defense team during his trial to aid the psychiatrist who evaluated his competency. The whole collection spans almost four decades and consists of around 3,600 pages of documents with those authored by Kaczynski himself – mostly journal entries and letters written to family, friends, and various strangers – accounting for an overwhelming majority. The collection used here comprises transcribed copies of the original documents – all of which can be be accessed in person at the University of Michigan library and some of which (though none of the ones used in this research) have recently become available online[2]. Full ethical approval was obtained to use these documents for this research from the author's university.

### 5.1. Dataset

The dataset used in this study reflects three main stages of Kaczynski's mental health history based on a timeline proposed in Johnson's (1998) report and a general understanding of the progression of both schizophrenia (e.g., APA, 2013; Freedman, 2010) and PPD. Broad periods were chosen over more specific ones because he never really sought treatment and thus had few or no prior records to use, making a more detailed breakdown impractical (Johnson, 1998). The first time period consists of the years before the likely onset of the symptoms of schizophrenia (i.e., pre-1966) but after evidence of PPD likely appeared. Thus, during this first stage, there may be evidence of paranoid ideation, but at a lower intensity than might be expected after the onset of delusional thinking (e.g., Beck & Rector, 2005; Renton & Mankiewicz, 2015). The second time period consists of the years just after the schizophrenia symptoms began, but before Kaczynski completely isolated himself (i.e., between 1966 and 1969), which marked an exacerbation of symptoms and the social and occupational impairment (Johnson, 1998). During this stage, it is likely that the delusions would have been in the process of formation and their content would have been evident, though with potentially lower apparent

[2]https://harbor.klnpa.org/california/islandora/object/cali%3A885

investment than one might expect later on in their development (Federal Bureau of Investigation, n.d.; Fineberg et al., 2015). The third time period consists of the few years just before his first bombing until his arrest (i.e., 1975 to 1996), during which time the delusions would have likely become fully-formed (Johnson, 1998).

The primary selection criterion for texts in this dataset was that they needed to be *first-person accounts* (i.e., any texts discussing personal beliefs or experiences and not those that detailed logistical information or the like as those are unlikely to carry much evaluative language). This provides the best opportunity to observe how Kaczynski described and evaluated his experiences and the people/things involved in them – that is, the events most likely to have been impacted by information processing biases associated with cognitive schemas (e.g., Beck & Haigh, 2014). Beyond that, an attempt was made to control for three main register features – i.e., topic, audience, and mode of communication – as each of these can impact language choices (A. Bell, 1984; Biber et al., 1999; Halliday & Matthiessen, 2014). Ultimately, topic was too difficult to control for as there were considerable changes in Kaczynski's circumstances between time periods, but audience and mode were both possible to control. Only letters to family were available across all three time periods, so these were focused on during selection. Finally, word count for each time period was also controlled, to an extent, to try to ensure the time periods were as balanced as possible. The aim was to collect two to three texts per time period, totalling between 1500 and 2000 words. A breakdown of the dataset – including the specific family member(s) acting as the audience, the date of the letter, and the word count – can be found in Table 1.

**Table 1. Breakdown of texts in dataset**

| | Date | Audience | Word Count |
|---|---|---|---|
| **Time period 1 (TP1)** | January 1961 | Parents | 546 |
| | April 1962 | Parents | 479 |
| | March 1963 | Parents | 727 |
| | | | **1752** |
| **Time period 2 (TP2)** | February 1967 | Parents | 1102 |
| | August 1968 | Parents | 964 |
| | | | **2066** |
| **Time period 3 (TP3)** | March 1975 | Parents | 850 |
| | January 1982 | Brother | 536 |
| | November 1990 | Mother | 769 |
| | | | **2155** |
| | | **Total (TP1-TP3)** | **5973** |

## 5.2. Analysis

Each of the texts was coded with the features of all three systems of the Appraisal framework in the annotation software *UAM CorpusTool* (O'Donnell, 2019). Once this was done, the raw frequencies of all features were exported so they could be analyzed quantitatively. Based in part on the work of Hurt (2020), chi-square tests were used to determine the *key variables* – those which were used significantly more in one text or time period over another (Baker, 2006) with the conventional significance threshold of $p<0.05$. This helped narrow the focus for the subsequent qualitative analysis, which was necessary for determining how the various stance-taking resources were actually used

in context and in turn, gain a better understanding of the patterns of these resources in the different texts and time periods. These patterns are then discussed in reference to their possible relationship to different schemas associated with Kaczynski's psychological traits/symptoms.

The primary focus of the quantitative analysis is on between-time period comparisons (i.e., comparing each time period to the other two), but within-time period analyses (i.e., comparing each text to all other texts) were also conducted to aid in the qualitative analysis. The within-time period analyses helped identify and account for fluctuations in evaluative resource patterns in the shorter intervals between individual texts that may reflect subtle changes in symptom severity that might not have been known given Kaczynski did not receive treatment before his arrest (Johnson, 1998). Since chi-square tests become too complex and difficult to interpret with multiple factors and levels (Grant et al., 2017), only two time periods or texts were compared at a time. For the between-time period analyses, this resulted in three comparisons (TP1 x TP2; TP1 x TP3; TP2 x TP3). There were 28 within-time period comparisons, so these are not discussed individually, but the distributions across the eight individual texts are examined alongside the broader time period distributions when relevant. The findings of the analysis alongside a discussion of how they may relate to different underlying schemas are presented in the next section.

As the three time periods were slightly different in size, frequencies for *attitude* and *engagement* were normalized to make them more comparable; these normalized frequencies were used to conduct the analyses. For *attitude*, tokens tend to be single words, so the frequencies were normalized per 1000 words as this represents the highest (neat) common denominator. In line with Hurt (2020), *engagement* frequencies were normalized per 100 **instances** rather than words as they tend to be clause/sentence length.

## 6. Results

For the sake of brevity, the results are presented first followed by a subsection relating to observations made about the patterns and of what they might be evidence. The discussion of how these results address the question of Appraisal's utility as an investigative tool is the main focus of section 7.

Before outlining the key findings, a few prefatory comments are warranted. The analysis did not reveal many significant quantitative differences (i.e., *key variables*), which suggests changes in Kaczynski's mental state between and within time periods did not have a significant impact on the distribution of evaluative resources he used over time. However, there were trends in the distributions that, while nonsignificant, are worth discussing alongside a few notable patterns that emerged in the qualitative analysis of *how* resources were employed. Thus, while the statistical results are reported and discussed in this section, the primary focus is on these trends and patterns.

### 6.1. Attitude

As outlined above, *attitude* resources help capture the different types of feelings and judgments authors express in their stances (J. R. Martin & White, 2005). Table 2 shows the features for which there was a significant difference (at least $p<0.05$) between at

least two of the time periods. The greyed cells indicate which time period had the higher proportion than the time period(s) indicated by the superscript.

**Table 2. Significant between-time period differences for attitude features**

| (Sub-)System | Feature | Time period 1 | Time period 2 | Time period 3 |
|---|---|---|---|---|
| **Attitude** | *Affect* | 25.11 | 21.30 | 38.52[1,2] |
| *Judgment* | *Capacity* | 53.65[3] | 38.72 | 21.81 |
| | *Veracity* | 0.57 | 0.48 | 6.03[1,2] |
| *Appreciation* | *Reaction* | 2.85 | 8.23 | 7.89[1] |
| | *Valuation-veracity* | 10.27 | 4.36 | 15.78[1,2] |
| **Explicitness** | *Invoked* | 30.82[3] | 19.85 | 12.99 |

Frequency per 1000 words

As the table shows, there are only six features within attitude for which there was a significant difference found between time periods. Interestingly, there is a significantly higher proportion of invoked meanings in TP1 over TP3, which suggests less reliance in later years on implication and more reliance on direct and explicit evaluations. As explicitness is coded on every attitude token, the use of implicit versus explicit meanings is discussed throughout this section when it becomes relevant.

With *affect*, roughly 25% of tokens in each time period were *attributed* to someone else. Within these examples, there appears to be a distinction between tokens denoting an 'internal state' versus those denoting an external and observable "surge of emotion" (J. R. Martin & White, 2005, p. 47). Consider the examples in Table 3.

**Table 3. Attributed affect**

| | Example | Text |
|---|---|---|
| | **'Internal' Tokens** | |
| 1 | If they are **interested** [attributed +inclination] they would send an application right away | 1961 |
| 2 | I think any eminent professor of math or science **would agree** [attributed +satisfaction] with me on this point here. | 1963 |
| 3 | The student "activists" are rebelling because they **want** [attributed +inclination] to follow the flaming torch of left-liberalism | 1967 |
| 4 | You could readily be excused for submitting such a story diffidently and **with grave doubts** [attributed -security]... but that **wasn't the frame of mind** [negated attributed -security] in which you submitted it | 1982 |
| | **'External' Tokens** | |
| 5 | ...many people are **clamouring** [attributed -satisfaction] that boxing should be outlawed | 1962 |
| 6 | The student "activists" are **rebelling** [attributed -satisfaction] | 1967 |

The 'internal' tokens represent Kaczynski's inferences about the emotional states of others—something that cannot truly be known to him as an outsider. The 'external' tokens, on the other hand, represent his own personal observations and descriptions of others' behaviors (lines 5 and 6). Internal tokens outnumbered external ones overall, but in the earlier texts, mitigating language was used more often around the internal tokens than in the later texts. For instance, in lines 1 and 2, TK introduces the attributed tokens with *if* and *I think* as if acknowledging them as mere possibilities. Conversely, in lines 3 and 4, the tokens are presented as the only warrantable possibilities through

declaring the reason for the rebellion in 3 and rejecting the possibility of an insecure or uncertain *frame of mind* in 4. Therefore, not only was he inferring the internal states of others, but over time, he appeared to become more confident in those inferences.

Across the three time periods, tokens of attributed *affect* were also consistently paired with other value-laden items that implied Kaczynski's own stances about the person(s) to whom the emotions were attributed. Examples are shown in Table 4.

**Table 4. Attributed affect with judgment and valuation**

| | Example | Text |
|---|---|---|
| 7 | The student "activists" [-valuation: veracity]... **want** [attributed +inclination] to stand up for [+propriety] the Right vs. the Wrong, because they **want** [attributed +inclination] reform [+composition] the world... | 1967 |
| 8 | ...but mainly because they **want** [attributed +inclination] to be able to have liquor, women, and marijuana in their dormitory rooms [-propriety] | |
| 9 | ...because they **want** [attributed +inclination] to...**do away with** [attributed -satisfaction] old-fashioned tradition-bound things like hard studying [+valuation: tenacity] | |
| 10 | I don't understand women – they seem so inconsistent [-tenacity]. I **don't think they quite know** [attributed negated +security] what they are doing themselves | 1975 |
| 11 | Also at about that age you **insisted** [attributed +security] that your injured hand was stronger than the other... In adulthood you admitted that this was self-deception [-valuation: veracity] | 1982 |

In lines 7-9 (all from the same text), Kaczynski talks about the wants of *student "activists"*. The scare quotes around *activist* convey disagreement with the accuracy of that descriptor for the individuals. The positive judgments that follow in line 7 are canceled out by the series of negative ones immediately after, like the assessment of impropriety in line 8 or of laziness implied in line 9 through saying the *activists* want to *do away with* things that require effort like *hard studying*. A similar pattern is observed with tokens of *in/security*, mostly within the *1975* and *1982* texts. For example, in line 11, when TK states that his brother's insistence about his injured hand is one admitted instance of *self-deception* [-valuation: veracity], or in line 10 where he calls women *inconsistent*, adding *I don't think they quite know what they are doing themselves* after one gave him her number but then later ignored his calls.

*Capacity* represented the highest proportion of *judgment* in all three time periods, encoding assessments of physical and mental abilities. These were either evaluations made directly about a single person/group, like lines 14 and 15 in Table 5 or through the use of comparative adjectives, like in lines 12, 13, and 16 (which imply a positive evaluation of one person and a negative evaluation of the other).

**Table 5. Self- and other-directed capacity**

| | Example | Text |
|---|---|---|
| 12 | ...he seems to **have considerable difficulty** [-capacity] with the last assignment. Also he **asked me a question** [-capacity] before the hour exam which showed I **understood** [+capacity] partial differentiation **better than he** [-capacity] | 1961 |
| 13 | So why did **he do so much better** [+capacity] on the hour exam **than I** [-capacity] did? It proved [+valuation: veracity] that this system of examinations is unfair [-valuation: propriety] | |
| 14 | I am inclined to **doubt the competence** [-capacity] of H.S. teachers to teach these subjects properly | 1963 |
| 15 | Obviously, you are **incapable** [-capacity] of the slightest self-control | 1975 |
| 16 | I **can** [+capacity] live on straight venison, being **smarter** [+capacity] **than** [-capacity] the local game warden | 1975 |

In instances of self-directed -*capacity*, there was usually additional context that implied the failure was not in Kaczynski's control. For instance, in line 12, he negatively evaluates a classmate's capabilities, but in line 13, Kaczynski admits that classmate tested better than him, which he claims is because the system is *unfair*, not a failure on his part.

Resources of *capacity* also worked in conjunction with resources of *tenacity* to convey an overall negative view of others, but usually indirectly through use of positive items against a broader negative backdrop. For example, when he expressed his belief that a Communist society was made up of people *who are completely dedicated* [+tenacity] *to the whole*. 'Dedication' is still a positive token but within the broader context, the dedication is described as being at the cost of individuality, saying they *have no desires* [negated +inclination] *of their own*. Kaczynski achieved a similar result in other ways, such as by placing it after countering language – e.g., *instead of* [counter] *facing the problem* [+tenacity] – or within a hypothetical, like saying his brother *could readily be excused for submitting such a story diffidently* [-tenacity]. In some cases, he also combined *tenacity* with other types of negative judgment, intensifying the negative evaluation, such as when he said to his parents their *insane* [-normality], *mindless* [-capacity] *persistence* [+tenacity] in sending magazines he did not want was *irritating* [-satisfaction].

The resources of *veracity* within both *judgment* and *evaluation* were used at a much higher rate in the later texts than any of the earlier ones. Those relating to Kaczynski himself helped depict him as the protagonist of the narrative who was just trying to be honest (even when what followed was generally negative evaluations of others). This included him saying he was just *pointing out* [+valuation: veracity] things that bothered him or that he was *warning* [+veracity] his brother, David, that his goals would not be easy to achieve. Even some of the instances of the self-directed -veracity achieved the same effect, such as saying how he was tired of having to *conceal* [-veracity] his opinions (i.e., he had been dishonest, but for the sake of David's feelings). Outward-directed evaluations of *veracity* via both *judgment* and valuation, on the other hand, were primarily used to indicate the perceived truth value of a proposition or to assess another person as deceitful. Within *judgment*, these assessments were mostly directed at Kaczynski's brother, calling him a *habitual rationalizer* and talking about his *habitual self-deception* (tokens that were repeated throughout the texts in which they occurred).

Within *valuation*, an interesting pattern arose in how truth values were indicated for propositions originating from Kaczynski versus those originating from others. Those relating to Kaczynski's beliefs were most commonly positive tokens and implied he was being direct and honest (e.g., *pointing out* various things or preceding a statement with *frankly*). For those originating from others, tokens marked them as deceitful to some extent, such as talking about David's *rationalizations* or saying that David had made *concessions* and *admitted* to various things, implying that statements containing different (and presumably false) assertions had been made previously. Overall, the use of *veracity* further suggests a stance of mistrust and suspiciousness of others and what they say.

## 7. Observations and Interpretations

There are two main observations of note here. The first is that many of the patterns above offer insight into how Kaczynski assigns praise and blame for events. Overall, more negative events are mentioned than positive ones and the blame for such negative events is often placed on external sources. Through the use of attributed *affect* and forms of *judgment* and *valuation*, portrayed others as responsible for the negative events that happened to them or as having caused some negative event that affected others, including those that affected him. For instance, he attributed his poor exam performance to an *unfair* [-valuation: propriety] evaluation system. Interestingly, in one instance of a positive event, Kaczynski portrayed himself positively for initiating a conversation with a woman who gave him her number *without hesitation* [negated -security]. However, as shown in line 10, when the woman did not return his phone calls, he called women *inconsistent*, claiming they do not *quite know* [negated +security] *what they are doing themselves* instead of acknowledging any alternative possibility that would have reflected negatively on him. If Kaczynski did portray himself negatively, he often used resources that would lessen the blame. For instance, in the text mentioned above where TK describes his brother's *self-deception* and *rationalizing*, Kaczynski first explains that he had been dishonest to spare David's feelings. However, he then continues to juxtapose David's 'dishonesty' (*self-deception* and *rationalizing*) with his own honesty (e.g., for *pointing out* the *self-deception* and *rationalizing* ), which effectively absolves him of wrongdoing for expressing such negative views of his brother in the letter.

The second observation relates to Kaczynski's shift in his views of society and his role within it. To be sure, Kaczynski's views of others and societal rules/norms were not positive during any of the three time periods, but he expresses different amounts of tolerance for certain rules between the earlier and later texts. In the first two time periods, there is disapproval of certain practices and institutions (e.g., blaming the *system of examinations* for his bad grade [1961] or expressing negative views of student activists [1967]), but he was still engaging with them. In the third time period, there is a noticeable shift to him explicitly negatively evaluating societal practices and institutions and also expressing a refusal, or at least a strong reluctance, to adhere to or engage with them. For example, he described the idea of finding a partner as *succumbing* [-capacity/-tenacity] to *hedonism* [-tenacity] and a *defeat* [-valuation: capacity] for him (*1975*) or when he claimed that he no longer wished to *conceal* [-veracity] *his opinions* or *put up with* his brother's *irritating* [-reaction] *traits* to spare his feelings.

This avoidance of assigning responsibility for negative events to himself and instead assigning it to others and situational factors aligns with the externalizing biases associated with both PPD (mostly the paranoid ideation aspect; Langdon et al., 2010; Murphy et al., 2018) and persecutory delusions (e.g., Beck & Rector, 2005; Kinderman & Bentall, 1997; J. A. Martin & Penn, 2002). The pattern regarding society also aligns with the nature of schemas associated with PPD and persecutory delusions (e.g., Beck, 2015; Beck & Rector, 2005; Kinderman & Bentall, 1997). There is also evidence of PPD core beliefs around oneself being righteous and clever and others being exploitative and deceptive (Beck, 2015; Renton & Mankiewicz, 2015). For instance, he views himself as being capable (e.g., line 16) – or at least not being at fault for failures (e.g., his poor exam performance) – and as having been altruistic in the past by not sharing his negative views of his brother. However, he describes others as having ulterior motives for positive behaviors (e.g., student activists in lines 7-9) or being consistently deceptive, like his brother.

### 7.1. Engagement

The system of *engagement* comprises the resources for communicating commitment to or certainty about a proposition and for the author to align or disalign (i.e., agree or disagree) with their own propositions or with other persons or viewpoints (J. R. Martin & White, 2005). There were no *key variables* identified between time periods within this system, but there are a few interesting qualitative patterns that are worth discussing.

The first thing to note is that Kaczynski's letters tended to have an overall argumentative tone; he would essentially play out an entire argument on a topic within the letter, though not necessarily with the recipient. Rather, alternatives would be acknowledged in some way (usually using *expansion* and *proclaim* resources), then his own interpretations and evaluations of situations and people would be presented or reinforced through *counters* and *denials* – which comprised at least 40% of the *contractions* in each text. Interestingly, it was with the *disclaim* resources that Kaczynski's own views were most often implied; that is, instead of *proclaiming* his views as most warrantable, he implied this through either *denying* the validity of alternatives or by presenting a 'more suitable' position than the acknowledged alternative. In doing so, his views are ultimately portrayed as less flawed than others or as being a more accurate depiction of events. To illustrate, consider the following examples in Table 6.

**Table 6. Disclaim tokens**

| | Engagement token | Text |
|---|---|---|
| 17 | *You will say why should anyone have the right to box* [acknowledge]*...their fight does not interfere with anyone else's rights* [deny]*...thus* [concur] *boxing can't justly be outlawed* [deny] | 1962 |
| 18 | *Their usual argument is that college draft deferment is unfair to the "culturally deprived"* [acknowledge]*...but* [counter] *if you consistently follow that line of reasoning* [entertain] *you have to maintain* [entertain] *that no one should be punished for any crime* [deny] | 1967 |
| 19 | *I recall suggesting to you* [entertain] *that you were only telling yourself this because it was an attractive idea* [counter]*. But still* [counter] *you insisted that your left hand was stronger* [distance] | 1982 |

In all three examples, there is some kind of acknowledgement of a position before a *counter* or a *denial* of that position. In lines 17-18, the first position originates from some vague other person or group, and the *counter* or *denial* serve to point out flaws Kaczynski's sees in those positions. The example in line 19 shows how these resources bolstered his own positions by demonstrating how David had continued with his 'self-deception' despite Kaczynski's efforts to prevent it. This pattern of usage not only depicts Kaczynski as more knowledgeable than others, but also paints them in a negative light—as people who fight to take 'rights' away and as impractical people with flawed arguments.

Instances of *entertain* were realized through a variety of formulations, though mostly with questions, hypotheticals/conditionals, hedged evaluations, and subjective statements. These functioned to convey a lower level of commitment to the position and/or reduce the intensity of the evaluation contained within them. Consider the examples in Table 7.

**Table 7. Expansion tokens**

| | Engagement token | Text |
|---|---|---|
| 20 | *I am convinced that exams cheat me of my due* | 1961 |
| 21 | *I am inclined to doubt the competence of H.S. teachers to teach these subjects properly* [entertain]*...they give a probably sketchy and possibly inaccurate coverage of advanced things* [entertain] | 1963 |
| 22 | *I am inclined to think* [entertain]*...that the current turmoil is not primarily caused by any desire for personal power as such* [deny]*. I suspect that the top leaders in China are sincerely devoted to the cause* [entertain] | 1967 |
| 23 | *She had seemed so friendly when I talked to her on the street* [entertain]*. So anyway I wrote her off* [pronounce] *and merely sent her a somewhat sarcastic note* [counter] *which probably led her to conclude that I was an escapee from a mental institution* [entertain] | 1975 |

Line 20 in Table 6 appears to have a high level of commitment but being *convinced* of something does not necessarily make it so and this highly subjective framing still leaves open other possibilities. In lines 21-23, the mitigating language of the *entertainments* weakens the negative evaluations presented within and alongside them. For instance, in line 21, the *am inclined to think* weakens the intensity of the assessment of the competence of the teachers. In line 23, the *probably* at the beginning of the last clause mitigates the intensity of the evaluation Kaczynski was assuming the woman had made of him, leaving open that other views might have been possible. Interestingly, as mentioned above, mitigating language became less common in instances of attributed *affect*, going from *entertaining* the possibility of people having these feelings like *if they are interested* [entertain] or *even if he should decide* [entertain] to higher commitment statements that imply Kaczynski has knowledge of others' feelings like *that wasn't the frame of mind in which you submitted* [deny].

Primarily in the texts where Kaczynski talked about his opinions of others and their actions, *attributions* were common. Within *attributions*, there are two possible options: *acknowledgements* which simply report the views of others such as *X* said or *Y believes* ; and *distancing* statements which imply the author's view toward the attributed proposition such as *X claimed* or *X's B.S. Acknowledge* tokens usually outnumbered *distance* tokens slightly. Kaczynski used both to introduce others' viewpoints; the *acknowledge-*

*ments* were often followed by evaluations of the person he cited, while the distancing attributions contained an evaluation of the person within them, which was then typically followed by yet another evaluation of them. Consider the examples in Table 8.

**Table 8. Attribute tokens**

| | Engagement token | Text |
|---|---|---|
| 24 | *Skip that B.S. about…* [distance]. *if they are interested* [entertain] *they would send the application right away* [entertain]. *I will not send those letters* [deny] *because they would be embarrassing* [justify] | 1961 |
| 25 | *Some of the slogans Diem and our people there have been using* [acknowledge] *are just as hypocritical as any of the twisted nonsense the Communists put out* [counter] | 1962 |
| 26 | *They don't want to have to study too hard* [acknowledge]. *Of course, there are a few who have deep and sincere convictions* [concur], *but most of them are just a bunch of jerks* [counter] | 1967 |
| 27 | *You have lately given some faint signs of admitting your moral fallibility* [distance] *though not nearly to the extent you should* [counter] | 1975 |
| 28 | *"No", you said, "I think I could become a published author. It wouldn't be hard"* [acknowledge]. *By this time, I trust you know better* [entertain] | 1982 |

As the examples in Table 8 show, *attributed* utterances were often accompanied by (or contained) evaluations of the persons to whom they were attributed and more often than not, these evaluations were negative. In line 24, the *distance* token contains a negative evaluation of the content his parents had produced, which is reinforced through the justification shortly thereafter. Similarly in line 27, saying that his parents showed *faint signs* of *admitting* suggests that the *moral fallibility* was something Kaczynski was already aware of, but they had not necessarily acknowledged as yet. In the lines 25, 26, and 28, the *acknowledgements* present the information in a more unbiased manner but are immediately followed by negative evaluations contained in *counter* propositions. This further reinforces the overarching stance that Kaczynski is more knowledgeable and that others hold impractical, unrealistic, and/or flawed beliefs.

## 8. Observations and Interpretations

The main observation of note from *engagement* patterns pertains to the argumentative tone that Kaczynski took in his letter, something which became more intense over time - e.g., consider the difference between how the views are expressed in lines 1 and 2 versus lines 3 and 4 in the *attitude* section. This suggested a belief that Kaczynski was more knowledgeable than others and that his opinions were more well-reasoned and practical and less flawed than theirs (which was also apparent through the range of negative evaluations he made of those others, as shown in 6.1). These patterns align with the externalizing biases and schemas associated with PPD and delusions (the predominant symptom of schizophrenia that Kaczynski experienced; Johnson, 1998). In particular, Kaczynski's portrayal of himself as always right (and also **in** the right) aligns with the core beliefs in PPD that one is righteous and clever (Renton & Mankiewicz, 2015). The increased commitment to his views over time and portrayal of his views as most warrantable aligns with delusions (Fineberg et al., 2015; Hinzen, Rosselló, & McKenna, 2016) and the way they develop over time (Fineberg et al., 2015; Freedman, 2010). More specifically, the amount of hedging observed in the earlier texts might reflect the start

of changes to Kaczynski's perception and interpretation of stimuli associated with the 'pre-delusional' stage, when there is a vague sense that something is different without certainty about what it is (Henriksen & Parnas, 2019). The last time period occurred well after the onset of the delusions, though, at a point when they would have more likely solidified, which would have been accompanied by more certainty and stability in his beliefs (Fineberg et al., 2015).

## 9. Appraisal as an investigative tool

As stated in the introduction, one of the goals of this study was to explore how the approach presented here could contribute to an investigation. The comparison that follows is, of course, being done in hindsight and is based on the analysis of a small corpus. It is therefore intended primarily to offer support for why this study provides a good initial 'proof of concept' of this approach as a potential investigative tool. What is explored here is where the findings might have been useful in supplementing the psychological and linguistic profiles generated in the original investigation and how extrapolations could be made from the findings based on the literature.

There were several psychological profiles throughout the nearly two-decade investigation, and the linguistic profile was only requested shortly before Kaczynski was found (Shuy, 2014). The linguistic profile contended the Unabomber was well-educated, around fifty-years-old, had likely grown up Catholic, might have at one time been in academia (likely in the hard sciences), and had lived in Northern California though likely grew up in or around Chicago—all of which turned out to be accurate. An early version of the psychological profile speculated that the Unabomber was young, uneducated, with low self-worth, and a desire to harm animals. In later versions, the age was increased slightly, they vacillated between hypothesizing that he had low and high self-esteem, and speculated that he was likely intelligent and might have had some education, though not a lot. One thing that remained consistent was the belief that the increasing lethality of his bombs indicated escalating anger and frustration directed at his victims and that only his arrest or death would stop the bombings (Fitzgerald, 2004).

With this information in mind, what might the approach proposed in this research have been able to offer? It is worth noting that the texts analyzed in this study did not include the letters or manifesto used in the original profile development, but they contained similar ideologies to the manifesto and Unabomber letters (this in fact being the reason that Kaczynski's brother suspected him; Shuy, 2014, p. 82). Given that his evaluative patterns were demonstrated above to be fairly consistent over time, it is possible that the most salient patterns, at the very least, would have been present in the Unabom texts, as well.

The analysis identified patterns which aligned with aspects of paranoid ideation and delusional beliefs, such as Kaczynski exhibiting a strong mistrust and suspiciousness of others and institutions, consistently externalizing blame for negative events, and adopting an overall argumentative and slightly condescending tone (e.g., Beck & Rector, 2005; Langdon et al., 2010; Renton & Mankiewicz, 2015). Using the literature on such traits, it might then be possible to extrapolate other information, as well. For instance, the psychological profile stated that Kaczynski was likely a loner with a heightened sense of superiority (Shuy, 2014), both of which are evident in the evaluative patterns. How-

ever, a desire for social isolation could be indicative of, for example, paranoid, schizotypal (Renton & Mankiewicz, 2015), or antisocial personality traits (Mitchell, Tafrate, & Freeman, 2015), and a heightened sense of superiority might be evidence of narcissistic (Behary & Davis, 2015), antisocial (Mitchell et al., 2015), or paranoid personality traits (Renton & Mankiewicz, 2015). Thus, it is important to consider all of the patterns that occur together to narrow down the most likely active schemas. In Kaczynski's case, the desire for social isolation appeared connected to the mistrust and suspiciousness of others and their motives, as evident in his externalization of blame and consistent negative evaluations of others shown in 6.1. This is more in line with paranoid PD (Beck, 2015) and delusional thinking (Beck & Rector, 2005) than with the strong discomfort with or disdain for social relationships seen in schizotypal PD (Renton & Mankiewicz, 2015) or the avoidance of intimacy seen in antisocial PD (Mitchell et al., 2015). Kaczynski's overall tone in his letters could be argued to suggest a heightened sense of superiority. He often portrayed his views as more warrantable than others' views and he depicted himself as righteous or not at fault while others were assigned blame and character flaws. However, the juxtaposition of himself as righteous and others as flawed is more in line with core beliefs of PPD (Renton & Mankiewicz, 2015) than with, for instance, grandiosity beliefs associated with narcissistic PD (Behary & Davis, 2015). The level of commitment evident in his statements in the later texts is also in line with delusional thinking patterns in that delusions necessarily manifest in contractive utterances that leave no real space for alternatives (Hinzen et al., 2016). If such space was left, it would suggest the beliefs were not held to be incontrovertibly true by the individual, which is a necessary aspect for their diagnosis (APA, 2013).

## 10. Limitations and Conclusions

The above is intended to demonstrate what one might be able to extrapolate about an author based on the patterns identified in their texts. This is, however, only based on the analysis of a corpus of writings from one author and therefore is not intended to be generalized to other individuals, as yet. To demonstrate the utility of the approach for a wider range of data, a larger analysis would be required which compares multiple individuals with a variety of diagnoses. It would also need to be demonstrated that the approach can be used on a variety of linguistic data types as it is not possible to control for variables in non-experimental situations to the extent they were controlled for here. Interrater reliability studies would be a potentially beneficial next step to address this latter point and increase the replicability of the coding approach.

Finally, since Appraisal is a manual method, it necessarily limits the size of the dataset that can be analyzed. Further, a method like Appraisal is not as suitable for capturing non-functional stance features. Thus, conducting the analysis using other methods, such as a form-to-function approach (e.g., Biber et al., 1999), in addition to Appraisal (similar to Gales, 2010) would be a potentially fruitful expansion of this research. Expanding to other stance features would also potentially help find more relationships between stance and certain symptoms/traits which may have schemas that are much more varied in their content across individuals, like delusions (Beck & Rector, 2005).

Despite the limitations, this study provides some evidence in support of a link between stance and psychopathology and demonstrates the potential utility of an assessment approach that combines linguistic and psychological theories, showing that this

approach warrants further research. With a better understanding of the relationship between psychopathology and stance would come expansions of and improvements to the approach to increase its consistency and reliability as an investigative tool.

## References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (DSM-IV-TR ed.). American Psychiatric Publishing.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th edition ed.). American Psychiatric Publishing.

Arntz, A., Dreessen, L., Schouten, E., & Weertman, A. (2004). Beliefs in personality disorders: a test with the Personality Disorder Belief Questionnaire. *Behaviour Research and Therapy*, *42*(10), 1215–1225. Retrieved 2025-07-02, from https://linkinghub.elsevier.com/retrieve/pii/S000579670300233X doi: 10.1016/j.brat.2003.08.004

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

Beck, A. T. (2015). Theory of personality disorders. In A. T. Beck, D. D. Davis, & A. Freeman (Eds.), *Cognitive therapy of personality disorders* (3rd edition ed., pp. 19–62). Guilford Press.

Beck, A. T., Butler, A. C., Brown, G. K., Dahlsgaard, K. K., Newman, C. F., & Beck, J. S. (2001). Dysfunctional beliefs discriminate personality disorders. *Behaviour Research and Therapy*, *39*(10), 1213–1225. Retrieved 2025-07-02, from https://linkinghub.elsevier.com/retrieve/pii/S0005796700000991 doi: 10.1016/S0005-7967(00)00099-1

Beck, A. T., Freeman, A., & Davis, D. D. (2015). General principles and specialized techniques in cognitive therapy of personality disorders. In A. T. Beck, D. D. Davis, & A. Freeman (Eds.), *Cognitive therapy of personality disorders* (pp. 97–124). Guilford Press.

Beck, A. T., & Haigh, E. A. (2014). Advances in Cognitive Theory and Therapy: The Generic Cognitive Model. *Annual Review of Clinical Psychology*, *10*(1), 1–24. Retrieved 2025-07-02, from https://www.annualreviews.org/doi/10.1146/annurev-clinpsy-032813-153734 doi: 10.1146/annurev-clinpsy-032813-153734

Beck, A. T., & Rector, N. A. (2005). Cognitive Approaches to Schizophrenia: Theory and Therapy. *Annual Review of Clinical Psychology*, *1*(1), 577–606. Retrieved 2025-07-02, from https://www.annualreviews.org/doi/10.1146/annurev.clinpsy.1.102803.144205 doi: 10.1146/annurev.clinpsy.1.102803.144205

Behary, W. T., & Davis, D. D. (2015). Narcissistic personality disorder. In A. T. Beck, D. D. Davis, & A. Freeman (Eds.), *Cognitive therapy of personality disorders* (3rd edition ed., pp. 299–324).

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

Bell, V., Halligan, P. W., & Ellis, H. D. (2006). Explaining delusions: a cognitive perspective. *Trends in Cognitive Sciences*, *10*(5), 219–226. Retrieved 2025-07-02, from https://linkinghub.elsevier.com/retrieve/pii/S1364661306000763 doi: 10.1016/j.tics.2006.03.004

Bentall, R. P., Corcoran, R., Howard, R., Blackwood, N., & Kinderman, P. (2001). Persecutory delusions: A review and theoretical integration. *Clinical Psychology Review*,

*21*(8), 1143–1192. Retrieved 2025-07-02, from https://linkinghub.elsevier.com/retrieve/pii/S0272735801001064 doi: 10.1016/S0272-7358(01)00106-4

Bhar, S. S., Beck, A. T., & Butler, A. C. (2012). Beliefs and personality disorders: an overview of the personality beliefs questionnaire. *Journal of Clinical Psychology*, *68*(1), 88–100. Retrieved 2025-07-02, from https://onlinelibrary.wiley.com/doi/10.1002/jclp.20856 doi: 10.1002/jclp.20856

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

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

Buck, B., & Penn, D. L. (2015). Lexical Characteristics of Emotional Narratives in Schizophrenia: Relationships With Symptoms, Functioning, and Social Cognition. *Journal of Nervous & Mental Disease*, *203*(9), 702–708. Retrieved 2025-07-02, from https://journals.lww.com/00005053-201509000-00007 doi: 10.1097/NMD.0000000000000354

Douglas, J. E., Ressler, R. K., Burgess, A. W., & Hartman, C. R. (2004). Criminal profiling from crime scene analysis. In H. Campbell & D. DeNevi (Eds.), *Profilers: Leading investigators take you inside the criminal mind* (pp. 13–33). Prometheus Books.

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

Edwards, D., & Potter, J. (1993). Language and causation: A discursive action model of description and attribution. *Psychological Review*, *100*(1), 23–41. Retrieved 2025-07-02, from https://doi.apa.org/doi/10.1037/0033-295X.100.1.23 doi: 10.1037/0033-295X.100.1.23

Federal Bureau of Investigation. (n.d.). *The Unabomber. Retrieved December.* Retrieved from https://www.fbi.gov/history/famous-cases/unabomber

Fineberg, S. K., Deutsch-Link, S., Ichinose, M., McGuinness, T., Bessette, A. J., Chung, C. K., & Corlett, P. R. (2015, January). Word use in first-person accounts of schizophrenia. *British Journal of Psychiatry*, *206*(1), 32–38. Retrieved 2025-07-02, from https://www.cambridge.org/core/product/identifier/S0007125000237367/type/journal_article doi: 10.1192/bjp.bp.113.140046

Fitzgerald, J. R. (2004). Using a forensic linguistic approach to track the Unabomber. In J. H. Campbell & D. DeNevi (Eds.), *Profilers: Leading investigators take you inside the criminal mind* (pp. 193–221). Prometheus Books.

Fournier, J. C. (2015). Assessment of personality pathology. In A. T. Beck, D. D. Davis, & A. Freedman (Eds.), *Cognitive therapy of personality disorders* (pp. 63–85). Guilford Press.

Fournier, J. C., DeRubeis, R. J., & Beck, A. T. (2012, April). Dysfunctional cognitions in personality pathology: the structure and validity of the Personality Belief Questionnaire. *Psychological Medicine*, *42*(4), 795–805. Retrieved 2025-07-02, from https://www.cambridge.org/core/product/identifier/S0033291711001711/type/journal_article doi: 10.1017/S0033291711001711

Freedman, R. (2010). *The madness within us*. Oxford University Press.

Gales, T. (2010). *Ideologies of violence: A corpus and discourse analytic approach to stance in threatening communications* (Doctoral dissertation). University of California-

Davis, ProQuest Dissertations Publishing.

Garety, P. A., Freeman, D., Jolley, S., Dunn, G., Bebbington, P. E., Fowler, D. G., … Dudley, R. (2005). Reasoning, Emotions, and Delusional Conviction in Psychosis. *Journal of Abnormal Psychology*, *114*(3), 373–384. Retrieved 2025-07-02, from https://doi.apa.org/doi/10.1037/0021-843X.114.3.373 doi: 10.1037/0021-843X.114.3.373

Garety, P. A., Kuipers, E., Fowler, D., Freeman, D., & Bebbington, P. E. (2001). A cognitive model of the positive symptoms of psychosis. *Psychological Medicine*, *31*(2), 189–195. Retrieved 2025-07-02, from https://www.cambridge.org/core/product/identifier/S0033291701003312/type/journal_article doi: 10.1017/S0033291701003312

Gawda, B. (2013). The Emotional Lexicon of Individuals Diagnosed with Antisocial Personality Disorder. *Journal of Psycholinguistic Research*, *42*(6), 571–580. Retrieved 2025-07-02, from http://link.springer.com/10.1007/s10936-012-9237-z doi: 10.1007/s10936-012-9237-z

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th edition ed.). Routledge.

Henriksen, M. G., & Parnas, J. (2019). Delusional mood. In G. Stanghellini, M. R. Broome, A. V. Fernandez, P. Fusar-Poli, A. Raballo, & R. Rosfort (Eds.), *The Oxford handbook of phenomenological psychopathology* (pp. 743–752). Oxford University Press.

Hinzen, W., Rosselló, J., & McKenna, P. (2016). Can delusions be understood linguistically? *Cognitive Neuropsychiatry*, *21*(4), 281–299. Retrieved 2025-07-02, from https://www.tandfonline.com/doi/full/10.1080/13546805.2016.1190703 doi: 10.1080/13546805.2016.1190703

History.com Editors. (2018). *Unabomber (Ted Kaczynski).* Retrieved from https://www.history.com/topics/crime/unabomber-ted-kaczynski

Hunter, M. (2022). *Violent ideologies: An investigation of the relationship between linguistic evaluative patterns and psychopathology in three types of violent offender* (Doctoral thesis). Aston University.

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

Johnson, S. C. (1998). *Psychiatric competency report of Dr. Sally C. Johnson* (Psychiatric report No. Case No. S-96-259 GEB). Retrieved from http://www.karenfranklin.com/files/Kazynski-Johnson_Report-09.11.98.pdf

Kinderman, P., & Bentall, R. P. (1997). Causal attributions in paranoia and depression: Internal, personal, and situational attributions for negative events. *Journal of Abnormal Psychology*, *106*(2), 341–345. Retrieved 2025-07-02, from https://doi.apa.org/doi/10.1037/0021-843X.106.2.341 doi: 10.1037/0021-843X.106.2.341

Langdon, R., Corner, T., McLaren, J., Ward, P. B., & Coltheart, M. (2006). Externalizing and personalizing biases in persecutory delusions: The relationship with poor insight and theory-of-mind. *Behaviour Research and Therapy*, *44*(5), 699–713. Retrieved 2025-07-02, from https://linkinghub.elsevier.com/retrieve/pii/S0005796705001178 doi: 10.1016/j.brat.2005.03.012

Langdon, R., Ward, P. B., & Coltheart, M. (2010). Reasoning Anomalies Associated With Delusions in Schizophrenia. *Schizophrenia Bulletin*, *36*(2), 321–330. Retrieved 2025-07-02, from https://academic.oup.com/schizophreniabulletin/article-lookup/doi/10.1093/schbul/sbn069 doi: 10.1093/schbul/sbn069

Martin, J. A., & Penn, D. L. (2002). Attributional Style in Schizophrenia: An Investigation in Outpatients With and Without Persecutory Delusions. *Schizophrenia Bulletin*, *28*(1), 131–141. Retrieved 2025-07-02, from https://academic.oup.com/schizophreniabulletin/article-lookup/doi/10.1093/oxfordjournals.schbul.a006916 doi: 10.1093/oxfordjournals.schbul.a006916

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave MacMillan.

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

Mitchell, D., Tafrate, R. C., & Freeman, A. (2015). Antisocial personality disorder. In A. T. Beck, D. D. Davis, & A. Freeman (Eds.), *Cognitive therapy of personality disorders* (pp. 346–365). Guilford Press.

Murphy, P., Bentall, R. P., Freeman, D., O'Rourke, S., & Hutton, P. (2018). The paranoia as defence model of persecutory delusions: a systematic review and meta-analysis. *The Lancet Psychiatry*, *5*(11), 913–929. Retrieved 2025-07-02, from https://linkinghub.elsevier.com/retrieve/pii/S2215036618303390 doi: 10.1016/S2215-0366(18)30339-0

O'Donnell, M. (2019). *UAM Corpus Tool*. WagSoft Systems.

Pennbaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*(6), 1296–1312.

Renton, J. C., & Mankiewicz, P. D. (2015). Paranoid, schizotypal, and schizoid personality disorders. In A. T. Beck, D. D. Davis, & A. Freeman (Eds.), *Cognitive therapy of personality disorders* (3rd edition ed., pp. 244–275). Guilford Press.

Shuy, R. W. (2014). *The Language of Murder Cases: Intentionality, Predisposition, and Voluntariness*. Oxford University Press. Retrieved 2025-07-02, from http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199354832.001.0001/acprof-9780199354832 doi: 10.1093/acprof:oso/9780199354832.001.0001

Thompson-Pope, S. K., & Turkat, I. D. (1988). Reactions to ambiguous stimuli among paranoid personalities. *Journal of Psychopathology and Behavioral Assessment*, *10*, 21–32.

*Language and Law / Linguagem e Direito*, *Vol. 11*(2), 2024, pp. 160-178
DOI: doi.org/10.21747/21833745/lanlaw/11_2a7

# "This is an extortion note": a corpus-driven genre analysis of commercial extortion letters†

Marton Petyko[1], Lucia Busso [2], Sarah Atkins [3], Emily Chiang [4], Nabanita Basu[5], and Tim Grant [6]

[1]*Aston Institute for Forensic Linguistics, Aston University*
[2]*Aston Institute for Forensic Linguistics, Aston University*, l.busso@aston.ac.uk
[3]*Aston Institute for Forensic Linguistics, Aston University*, s.atkins@aston.ac.uk
[4]*Aston Institute for Forensic Linguistics, Aston University*, e.chiang2@aston.ac.uk
[5]*Northumbria University*, nabanita.basu@northumbria.ac.uk
[6]*Aston Institute for Forensic Linguistics, Aston University*, t.d.grant@aston.ac.uk

***Abstract***
*This study presents a corpus-assisted genre analysis of 39 British commercial extortion letters, the largest dataset of its kind. We use a combined methodology that triangulates evidence from move analysis, corpus linguistics, and clustering algorithms. We innovate on traditional move analysis by using clauses as the unit of analysis and interrater reliability testing to identify 11 rhetorical moves. Corpus linguistic methods, including n-gram and collocation analysis, reveal that while certain moves (Threat and Demand) are highly prevalent and "core" for the genre of extortion letter, there is no consistent structural pattern across texts. Instead, rhetorical moves occur flexibly in texts, with no recurrent patterns, reflecting their status as an "illicit genre" produced outside conventional discourse communities. We further use clustering algorithms to explore whether we can reliably differentiate letters based on moves frequency. This multi-method approach demonstrates that rhetorical move prevalence, rather than sequence, can serve as a robust basis for genre classification in extortion communications. Furthermore, the study offers methodological innovations for forensic linguistics and genre.*
***Keywords:*** *extortion letters; malicious communications; move analysis; corpus linguistics; hierarchical clustering.*

†The present paper is the last scholarly work of Dr. Marton Petyko, who passed away in November 2023 and is greatly missed by his colleagues and friends at the Aston Institute for Forensic Linguistics. Paragraph 4, 5, and 5.1 are still largely his own writing, and he carried out the corpus linguistic analysis here presented. Lucia Busso is responsible for the move maps and paragraph 2 and 5.2, and the writing of paragraph 5.3. The analysis in 5.3 is done by Nabanita Basu. The Move Analysis, the coding, and the writing of the rest of the paper is the result of a collaboration among the first 5 authors.

***Resumo***
*Este estudo apresenta uma análise de género textual baseada num corpus de 39 cartas de extorsão comercial britânicas, o maior conjunto de dados deste género. Utilizamos uma metodologia combinada que reúne e relaciona dados de análise de movimentos, linguística de corpus e algoritmos de agrupamento ("clustering"). Destacamo-nos da análise tradicional de movimentos, uma vez que utilizamos orações como unidade de análise e testes de fiabilidade interavaliadores para identificar 11 movimentos retóricos. Os métodos linguísticos do corpus, incluindo a análise de n-gramas e de colocações, revelam que, embora certos movimentos (Ameaça e Exigência) sejam altamente prevalecentes e "nucleares" para o género da carta de extorsão, não existe um padrão estrutural consistente nos textos. Em vez disso, os movimentos retóricos ocorrem de forma flexível nos textos, sem padrões recorrentes, refletindo o seu estatuto de "género ilícito" produzido fora das comunidades discursivas convencionais. Além disso, utilizamos algoritmos de agrupamento para explorar a possibilidade de diferenciar as cartas de forma fiável com base na frequência dos movimentos. Esta abordagem multimétodo demonstra que a prevalência de movimentos retóricos, e não a sequência, pode servir como uma base sólida para a classificação de géneros em comunicações de extorsão. Além disso, o estudo oferece inovações metodológicas para a linguística forense e para o estudo dos géneros textuais.*
***Palavras-chave:*** *Cartas de extorsão; Comunicações maliciosas; Análise de movimentos; Linguística de corpus; Agrupamento hierárquico.*

## 1. Introduction

Malicious and threatening communications, of which written extortion letters are a particular type, have been a key area of study for forensic linguistics. Scholars have analysed different types of malicious communication in different ways, using SFL and Appraisal (Hurt, 2020; Reczek, 2023), Biber's Multidimensional Analysis (Nini, 2017, 2019), Speech Act theory (Spitzberg & Gawron, 2016) and Move Analysis (Abaalkhail, 2015; Chiang & Grant, 2019). Threatening language characteristics have also been analysed using a plethora of different data types, from Late Modern English threatening letters (Neumaier, 2025) to spoken threats (Tompkinson, 2023), to a large corpus of different types of threats (Gales, 2015).

However, these analyses have not focused on the genre of extortion letters specifically, and the question of whether this text type constitutes a genre has remained relatively underexplored. Some studies have identified that the form or 'text type' of extortion communications is only minimally shaped by norms, with writers borrowing from other genres, such as business letters, but with the texts open to a large degree of individual variation (e.g., Fobbe, 2011, 2020). However, such studies are often necessarily limited to small numbers of texts, mainly because of the difficulty of accessing texts be-

yond single case studies, and there is no study – to the best of the authors' knowledge – that examines more than a single digit number of texts.

The current study is therefore the first of its kind, as we analyse the largest existing collection of 39 written commercial extortion communications (letters and emails), as all other works (to the best of the authors' knowledge) analyse single digit amounts of documents. The documents all come from a UK context, and span from 2008 to 2019. With this in mind, we address the following research questions:

1. What are the recurring communicative functional units in commercial extortion communications?
2. Can we describe the discourse structure of this genre according to these units?
3. How homogeneous/varied is the collection of texts?

In the remainder of the paper we first discuss the two key theoretical concepts for our analysis: genre (section 2) and rhetorical moves (section 3). We then introduce the data (section 4) and present the methodology and results from our Move Analysis in section 5, looking at move prevalence (5.1), patterns (5.2), and clustering (5.3).

## 2. Genre analysis and forensic texts

The concept of genre is widely debated in the academic literature, with three major research traditions shaping genre studies (Xia, 2020). The English for Specific Purposes (ESP) approach, notably Swales (1990), views genre as a class of communicative events with shared purposes, shaped by discourse communities. Systemic Functional Linguistics (SFL, Martin, Christie, Rothery, & Reid, 1987), sees genre as a staged, goal-oriented social process. Finally, Rhetorical Genre Studies (RGS), following Miller (1984), frames genre as a response to recurring rhetorical situations, focusing on the typification of rhetorical actions.

Despite their differences – ESP's focus on community, SFL's on social processes, and RGS on rhetorical recurrence – there is broad consensus that genre is inherently social and contextual. Understanding genre, across these frameworks, involves analysing how texts reflect the practices and goals of the communities that produce and use them (Bhatia, 2016; Corbett, 2006). Biber and Conrad (2009) emphasise this social dimension, defining genre as a conventionalised construct, characterised by consistent linguistic features and a unified communicative purpose.

This study adopts Swales (1990) ESP-based definition, particularly useful for analysing recurring rhetorical moves in extortion communications. Swales defines genres as classes "of communicative events, the members of which share some set of communicative purposes", recognised by expert members of the parent discourse community (Swales 1990, p. 58). Central to his model are the notions of 'discourse community' ('socio-rhetorical networks that form to work towards sets of common goals', Swales, 1990, p. 9) and 'communicative purpose', which drives the language activities of the discourse community.

However, the notion of 'discourse' communities outlined above, potentially becomes problematic with forensic linguistic genres. Malicious communications, such as extortion letters, are unlikely to stem from a 'socio-rhetorical network' of authors (Swales, 1990, p. 9) nor from repeated exposure to examples within a discourse community. Bojsen-Møller, Auken, Devitt, and Christensen (2020) address this by proposing a fifth

category - 'illicit genres' - to Miller's 2017 four genre typology. Illicit genres are not learned institutionally, lack formal instruction, and are socially marginal (Bojsen-Møller et al., 2020, pp. 34–35). Consequently, these genres are recognised as such not by a clear community of authors and users, but rather by "uptake communities", that is audiences, such as law enforcement or victims - those who label and respond to these texts, whether or not they are the intended audience. 'Uptakes' can be expected or unexpected, welcome or unwelcome (Freadman, 2012), which, in the case of illicit genres, Bojsen-Møller et al. 2020 suggest, results in a 'dual addressivity', simultaneously targeting victims as recipients while anticipating interception by authorities. This notion of 'uptake community' underpins our analysis of genre, where we focus on functions and actions accomplished within the texts.

## 3. Move analysis

Move analysis is a functional framework for examining the conventional discourse structures within genres. 'Moves' are rhetorical or discoursal units that serve coherent communicative functions, often realized through combinations of lower-level strategies (Biber, Connor, & Upton, 2007; Swales, 2004). Originally developed for academic writing, move analysis has since been applied to a range of genres, e.g; marketing texts (e.g., Campbell & Naidoo, 2017), online reviews (Skalicky, 2013) and crowdfunding texts (Liu & Deng, 2016). Despite this utility, theoretical and methodological challenges persist, particularly in adequately defining 'communicative function' for identifying moves, with similar terms such as rhetorical function, communicative purpose and communicative intention, are often used interchangeably (Askehave & Swales, 2001; Bhatia, 1996). While individual moves are often richly described, methods for identifying them remain opaque, limiting replicability. Moreno and Swales (2018) propose a bottom-up approach, identifying steps and sub-moves first and only later categorizing these into higher-level moves. However, their functional approach lacks consistent, formal criteria.

Another challenge lies in using moves to define genre membership. Early studies often relied on the presence of 'obligatory' and 'optional' moves, particularly in academic genres (Biber et al., 2007; Swales, 2004). However, in more fluid genres, researchers have tended towards more flexible terms, for example Samraj and Gawron (2015, p. 95), found that no move occurs in 100% of suicide notes. Instead, they propose identifying 'core' moves and 'minor' moves based on frequency. Building on Samraj and Gawron's concept, Chiang (2018) extends this to online child abuse conversations, using move frequency to assess "coreness" and their relevance to the text type. These studies suggest that 'occluded' genres, lacking fixed moves or structure, require a more flexible model of move analysis .

Our study refines move analysis for forensic linguistic application in two ways. Firstly, by using clauses as the unit of analysis and introducing inter-rater agreement as a novel method to identify and validate rhetorical moves. Secondly, we follow Chiang (2018) in using a three-tiered classification system for moves – 'core' (nearly 100% of texts), 'typical' (at least 50% of texts) and 'atypical' (under 50% of texts). This approach enhances the reliability of analysis for 'occluded' forensic genres.

## 4. Data and methods

The 39 written extortion communications were provided by a UK law enforcement partner. All had been previously classified and investigated as 'extortion communications' by the police, indicating an uptake community that had recognised and labelled these texts as belonging to the same genre. The texts all relate to cases that the law enforcement partner classed as 'historic' (at least 12-months old) and either resolved ('historic–complete'), or unresolved but no longer under investigation ('historic–incomplete').

The study examines the rhetorical structure of the letters, without inferring authorial or case-specific details, due to limited metadata (e.g. on number of perpetrators or extortion outcomes). The only reliable metadata was the police subdivision into individual communications and series (i.e. multiple letters by same author).

Ethical considerations were central, particularly in protecting the confidentiality of victims and organizations mentioned for the 'historic–incomplete' cases that had not entered the public domain. Ethical approval was obtained from Aston University as well as compliance with the law enforcement partners' protocols. Anonymisation was key to preventing traceability, involving secure storage and systematic redaction of identifiers (e.g. replacing names with <MaleFirstName1>, <Town>, etc). This process, developed in consultation with the police, followed best practices in anonymising sensitive linguistic data. The final dataset comprises 39 texts totalling 9,295 words, with texts ranging from 31 to 951 words (average length: 270 words, standard deviation: 222.2). Figure 1 plots text length range in the corpus.

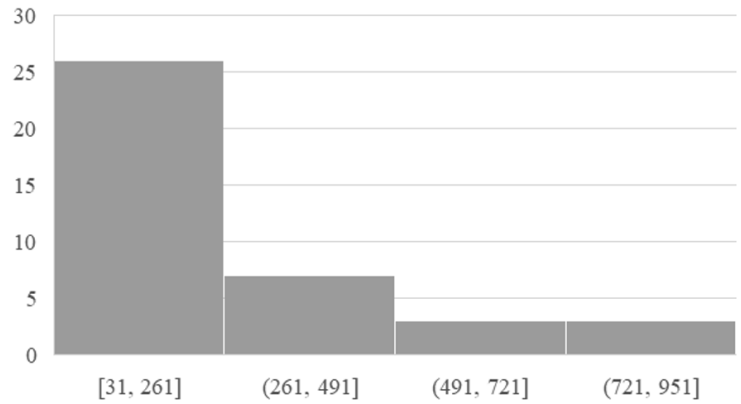



**Figure 1. Text length in the corpus. As it can be seen, texts are generally short, between 31 and 261 words long**

The dataset was analysed combining qualitative, corpus-based, and computational methods. Specifically, the letters were coded for rhetorical moves using a new and more reliable system that uses clauses as the basic unit of analysis and that involve iterative coding and reliability testing. We first selected five texts randomly and broke them down to clause units. Then four members of the research group independently analysed these extortion communications and created preliminary lists of recurring moves. We compared these preliminary lists and developed a refined set of 11 key moves. Once the move set had been finalised, we wrote a codebook with definitions and examples for every move identified (Section 5). Four members of the research team then coded

moves in the five selected texts and we then calculated the level of interrater reliability for this coding process (section 5.1).

Section 5.2 introduces a corpus-based analysis of move n-grams and collocations to explore how moves pattern in the letters. Section 5.3 uses clustering algorithms to group letters in the corpus in statistically stable and coherent groups based on move prevalence.

## 5. The rhetorical moves in commercial extortion communications

Using an iterative annotation method, four coders from the research group independently coded five randomly selected texts, broken down into clauses ( 12% of the corpus, 20% of the total clauses). Inconsistencies were discussed and a refined, final set of 11 key moves was agreed. This move codebook includes *Opening, Sign-Off, Demand, Instructions, Justification, Threat, Demonstrating Credibility, Consequences, Additional Persuasion, Statement of Purpose, and Pre-Announcement.*

**Openings** and **Sign-Offs**, similar to other types of written communications, mark the beginning and the end of the analysed text. Some display more conventional formality, such as *Dear <supermarket> directors* [S02L01C01][1] or *Good luck as I await your reply* [S22L01C49], while others are informal, such as *Hello*, [S17L01C01] and *Bye luvvie* [S18L01C27].

**Demands** are central to extortion letters, urging the addressee to make a payment, *You must pay £50,000 in cash on Sunday 14th May* [S08L01C13] and *I demand a sum of 250.000 £ in my bitcoin account by the end of December* [S02L01C02] or respond to the author (*Get back to me now if you are ready to pay some fees to spare your life* [S22L01C31] and *Do not ignore this threat* [S13L01C08]). Authors often instruct the addressee not to contact the authorities (*Do not go to the police* [S23L01C12]).

**Instructions** frequently accompany demands, guiding the addressee on how to deliver payment or perform additional actions. These instructions can be highly detailed and sometimes make up a large part of the text, as the step-by-step instruction clauses in this text illustrate:

> *At 1400hrs you will drive to the <Supermarket1> car park and stop at the unloading only at the front of the store. Wait for 5 mins then head for <StreetName1>. You will then drive up to <HouseNumber1> <StreetName1> until you see an off licence on your left. Take an immediate left turn at this off licence and park up against the shop wall. You will see a brick construction halfway across the wall 4 bricks high and rectangle in shape. It surrounds a steel grate. There are bricks inside this. Lift up the corner brick and place package there. Get in your car and drive off. [S15L01C50-60]*

**Justifications** provide reasons for the demand, acting as a persuasive technique on the authors' part to push the addressees towards compliance. For example, extortionists may claim that their victim owes them money (*I feel a little disappointed that you do not seem to be willing to pay me back my £10000 you owe me* [S10L01C03]), previously caused harm (*you kindly made me unemployed* [S17L01C06]) or committed a

[1]The shorthands between square brackets refer to the location of the examples in our corpus. For example, S02L01C01 means that the example comes from the first clause of the first letter of the second letter series. (S stands for series, L for letter and C for clause.)

crime (*Several weeks ago you purchased items of jewellery that were stolen during a robbery at a couple's home. Five days later the lady suffered a stroke and to all intents and purposes their lives have been devastated. I must state that I don't hold you responsible for these events only to the facts of you buying stolen property* [S15L01C1-3]), thus framing demands (and their subsequent threats) as legitimate and morally justifiable.

**Threats** are the other fundamental move of extortion communications, used to coerce compliance. Threats were classed as a statement that expresses the extortionist's willingness or intent to cause physical harm (*Get back to me now if you are ready to pay some fees to spare your life. If you are not ready for my help, then I will carry on with my job straight up* [S19L01C32-34]; *You will not be harmed, but your wife, Sons, business partners and employees will be* [S08L01C7-8]) or damage the business interests of the targeted company (*On the 5th December 2013 your <supermarket> products contaminated with salmonella will be simultaneously placed in 25 major retail outlets around the country* [S13L01C01]).

**Demonstrating Credibility** was performed to support threats, with extortionists claiming access to the necessary tools (*We have devices ready to cease your production for one day – one week – one month – three months and more* [S11L01C12]), knowledge (*As I am writing to you now my men are monitoring you and they are telling me everything about you* [S22L01C27-28]), expertise (*we would detect [the tracking device] before we got close* [S09L01C48]), and determination (*Secondly I am trained to follow through all plans I make and will not be deterred from anything or anyone in the in the pursuit of those plans* [S15L01C15-16]). These portray the authors as experienced criminals, often with claims to membership of organised crime groups.

**Consequences** contrast the costs of non-compliance (*However if you put us to this trouble we double what we want to 200,000 pounds* [S26L01C21-22]) with the benefits of compliance, such as promises that no further demand will be made (*When you deliver the money I will consider the matter closed. There will be no further contact between us* [S15L01C31-32]). This move also covers a 'cost/benefit analysis' of compliance, where authors tend to minimise the costs (*<company name> as a company is financially stable and should be able to pay* [S12L02C7-8]) and maximise the benefits of compliance (*you can quickly protect the public* [S26L01C17]).

**Additional Persuasion**, encompasses a variety of persuasive techniques, grouped into a single move for practical operation of the analysis. This included non-negotiability (*This is a serious requirement and is not negotiable* [S09L01C7-8]), appeal to reason (but stupid you are not [S05L01C24]), or to the futility of resistance (*You cannot win* [S10L02C69]), appeals to moral duty (*so the faith or your employees and customers is in your hands* [S02L01C05]), and follow-up promises (*We will be in touch* [S07L01C19]).

**Statement of Purpose** declare the communicative purpose of the texts (*Subject: extortion message* [S20L01C06]), suggesting some awareness of 'genre' by extortionists themselves.

**Pre-Announcements** introduce the move that will follow, such as the heading *Instructions* [S08L01C19] before the instructions.

### 5.1. Prevalence of moves

Once the codebook had been finalised, the sample of texts was then re-coded, achieving high interrater reliability using percentage agreement (88% of codes with 75% agreement), following Rau and Shih's 2021, who argue this is more appropriate for this type of unit-boundary coding[2].

A trained, external coder was then introduced, who first coded the same five texts.[3] The high level of agreement was maintained, with 84% of clauses achieving agreement above 75% and 68.7% in 100% agreement between all coders. The coder then independently annotated the rest of the corpus, with ongoing discussions with the research team to ensure consistency.

Table 1 shows the range (percentage of texts where a move appears at least once) and frequency (percentage of clauses for each move across all texts) of the 11 rhetorical moves across the 39 texts, divided into 1,062 clauses.

**Table 1. The range and frequency of moves in the corpus**

| **Move** | **Range** (n = 39) | **Frequency** (n = 1,062) |
|---|---|---|
| Threat | 92.3% | 15.7% |
| Demand | 92.3% | 9.3% |
| Consequence | 87.2% | 15.4% |
| Opening | 87.2% | 5.6% |
| Instruction | 84.6% | 14.8% |
| Demonstrating credibility | 82.1% | 19.4% |
| Additional persuasion | 64.1% | 7.5% |
| Sign-off | 51.3% | 2.5% |
| Pre-announcement | 46.2% | 2.8% |
| Justification | 35.9% | 6% |
| Statement of purpose | 20.5% | 0.8% |

Results show threats and demands are the two core moves, found in 92.3% of texts. Demands appear in every stand-alone written text and multi-text series, while Threats also occur in every multi-text series and in almost all stand-alone communications, except one in which the author implicitly implies intimidation by referencing damage to an organised crime member's car.

Beyond these two core moves, other typical moves (Consequences, Instructions, Demonstrating the Credibility, Openings, Additional Persuasion, and Sign-Offs) appear in more than three-quarters of the texts. The prevalence of discussing Consequences and Demonstrating Credibility, along with Additional Persuasion, suggests that while threats are a fundamental persuasive tool, extortion communications employ a broader range of persuasive techniques.

[2]Percentage agreement is calculated in a pairwise manner, i.e. between two annotators at a time. That is, the coding of Annotator 1 is compared to the coding of Annotator 2, then to the coding of Annotator 3, and so on. If the codings of the two annotators coincide, a value of 1 is assigned to the cell. If the codings diverge, it is assigned 0. This process is repeated for every annotator, for a total of 6 comparisons (1 vs 2, 1 vs 3, 1 vs 4, 2 vs 3, 3 vs 4). The total is then divided by the total number of comparisons and assigned a percentage score. A score of 100% is given to instances in which all 4 annotators agree, 75% is given to cases in which 3 out of 4 annotators agree, and so on.

[3]We follow the rule of thumb followed by the literature that values greater than 0.70 are typically acceptable for consistency estimates of interrater reliability (Stemler, 2004).

Less surprising is the widespread presence of Openings and Sign-Offs, reflecting norms in written communications more generally, while the prevalence of Additional Persuasion aligns with the generally persuasive nature of extortion communication.

Finally, three of the moves identified, Pre-Announcement, Justification, and Statement of Purpose, can be considered atypical, present in less than 50% of the texts. Justification therefore emerges as the least common persuasive technique.

The mosaic plot in Figure 2 visualises the frequency of each move within individual texts, showing that frequency of moves, especially the core and typical moves, varies greatly across individual texts. For example, Demonstrating Credibility (in orange) is rather dominant in texts 13, 25 and 39, but it becomes much less prevalent in texts 9, 32 and 36.

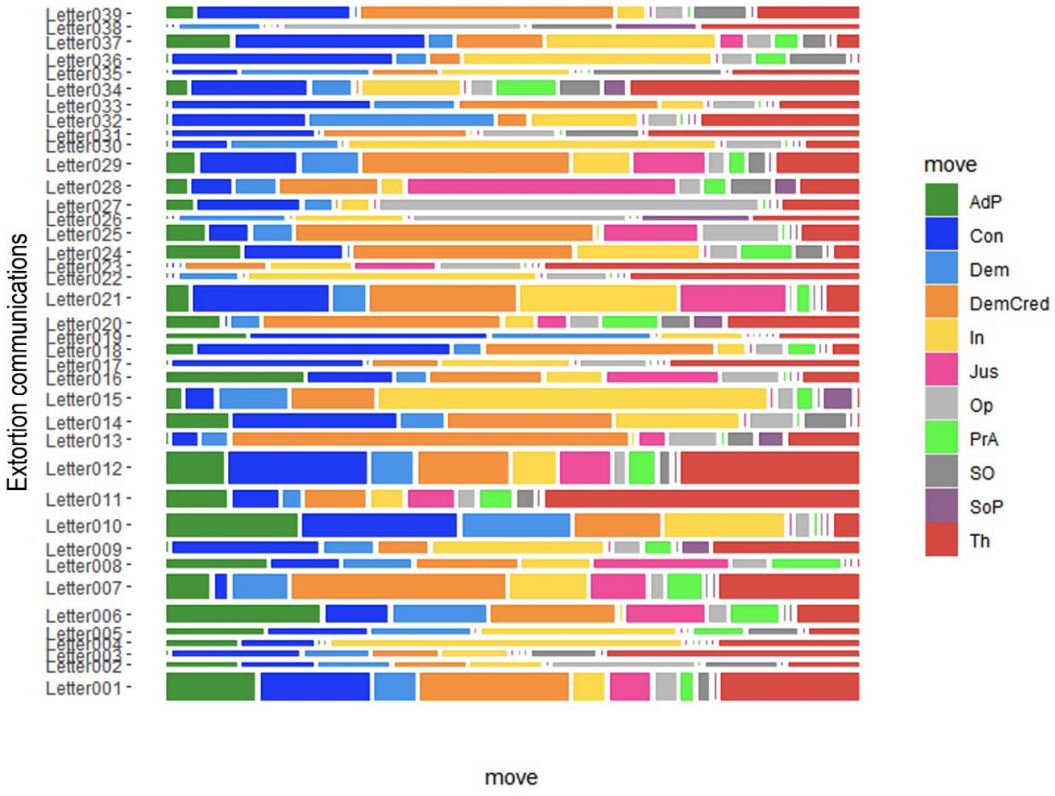



**Figure 2. Prevalence of moves within individual texts. Thickness of bars refer to length of the text (longer texts have thicker bars). The moves are shown in the legend in alphabetical order: AdP - Additional Persuasion, Con - Consequences, Dem - Demand, DemCred - Demonstrating Credibility, In - Instructions, Jus - Justification, Op - Openings, PrA - Pre-announcement, SO - Sign offs, SoP - Statement of Purpose, Th - Threat.**

## 5.2. Move patterns

Having coded moves across the corpus, we use corpus linguistics to explore patterns in the sequencing of moves. The importance of using corpus linguistic methods to enhance and complement discourse studies has been highlighted in the literature (Biber et al., 2007; Upton & Cohen, 2009), as it enables

> "detailed but generalizable analyses of discourse structure across a representative sample of texts from a genre. (….) It is then possible to identify the sequences of move types that are typical for a genre, and against that background it is also possible to identify particular texts that use more innovative sequences of move types." (Upton & Cohen, 2009, p. 588).

Recurrent structures of moves have been found in many professional contexts, such as promotion letters and job application letters (Bhatia, 1993), and business letters of negotiation (Dos Santos, 2002). However, when the method is applied to forensic texts, linear structures in the moves are not always found. For example, Samraj and Gawron's 2015 study of suicide notes found texts only required one move from a set of 'core' moves to be included in the 'genre' and that they displayed a great variation in structure.

Having access to a relatively large corpus of extortion letters from a variety of different authors gives us the perfect testbed to explore whether any recurrent patterns of moves can be found in this illicit genre. To do so, we employ corpus linguistics methods using LancsBox (Brezina, Weill-Tessier, & McEnery, 2020). For these analyses we consider moves as a whole, and not clauses, as units. That is, a move that spans multiple clauses is considered to be one unit (leaving us with 612 moves across 39 texts). So, example 1.(a)-(c) below (taken from text 008) consists of 3 consecutive clauses which were all coded as one move (Threat).

1. (a) *The consequences will be Very serious.*
   (b) *You will not be harmed,*
   (c) *but your wife, Sons, business partners and employees will be.*

We analyse patterns of moves using ngram and collocation analysis. We use ngram analysis to explore sequences of two or three adjacent moves, while collocation analysis is instead focused on the co-occurrence of two non-adjacent moves within a 2L 2R window span (i.e., 2 moves to the left 2 moves to the right). The ngram analysis of move sequences allowed us to find fixed patterns, whereas we used the collocation approach to explore looser associations between two moves, where two moves can stand directly next to each other or can have any third move between them. The minimum raw frequency in both ngrams and collocation analysis was 1. After applying the above criteria, we found 392 move sequences and 142 collocations.

**Table 2. The ten most frequent move ngrams in the corpus**

| **Move sequence** | **Frequency** | **Range** (n = 39) |
|---|---|---|
| Consequence -Threat | 22 | 43.6% |
| Demonstrating Credibility -Threat | 22 | 30.8% |
| Consequence - Additional persuasion | 20 | 33.3% |
| Threat - Consequence | 20 | 35.9% |
| Demand - Instruction | 19 | 38.5% |
| Demand - Consequence | 19 | 33.3% |
| Demonstrating - Credibility Consequence | 18 | 33.3% |
| Consequence - Demonstrating credibility | 17 | 33.3% |
| Instruction - Demand | 16 | 33.3% |
| Instruction - Consequence | 15 | 30.8% |

Table 2 shows the ten most frequent move sequences in the corpus, all of which are bigrams (i.e. sequences of two moves). As expected, the most frequent sequences are made up of the most frequent moves. For example, the joint most frequent sequence, *Consequence-Threat* and *Demonstrating Credibility-Threat*, feature the three most frequent moves: *Demonstrating Credibility* (19.4%), *Threat* (15.7%), and *Consequence* (15.4%). This suggests that there are no preferred sequential patterns of moves in our corpus, and that on the contrary moves occur quite freely in texts.

Moreover, none of the sequences can be considered typical, as even the most widespread one, *Consequence-Threat*, only appears in 43.6% of the texts, while the rest occur in around only one third of the extortion notes. The ngram analysis also suggests that the moves do not follow any specific order in the analysed texts. For example, *Consequence* is followed by *Threat* 22 times while *Threat* is followed by *Consequence* 20 times. Similarly, *Instruction* comes after *Demand* 19 times and *Demand* after *Instruction* 16 times.

These results overall indicate that while the analysed extortion communications do have core and typical moves, these moves do not form recurrent structural patterns. In other words, while genres with a strong community of practice generally show a fixed order of recurrent move sequences (e.g. academic articles, Swales 1990), the generic structure of this text type appears to be driven by function rather than by a fixed order of recurrent move sequences. In this sense, the common features of the genres are more functional than formal. This raises interesting questions concerning the notion of genre based on observation of the same functions across texts, rather than the ordering or arrangement of these functions. This finding is in line with other works on genres that do not have a clear community of practice, such as the already mentioned work by Samraj and Gawron (2015) on suicide notes and this clearly deserves further investigations.

**Table 3. The ten most frequent move collocations in the corpus**

| Move collocation | Frequency |
|---|---|
| Consequence […] Threat | 66 |
| Consequence […] Demonstrating credibility | 55 |
| Demand […] Consequence | 53 |
| Demonstrating credibility […] Threat | 49 |
| Demand […] Instruction | 46 |
| Consequence […] Additional persuasion | 45 |
| Threat […] Demand | 43 |
| Instruction […] Consequence | 38 |
| Threat […] Instruction | 38 |
| Instruction […] Demonstrating Credibility | 36 |

Table 3 reports the ten most frequent move collocations in the analysed texts, which reinforce the key findings of the ngram analysis. While the most frequent moves form the top collocations, the collocations themselves are not particularly frequent, albeit more common than the sequences, suggesting that the moves do not form typical patterns in the texts.

Figures 3 and 4 show move maps of the corpus. Move maps are colour-coded visual representations of the move structure and patterns in each text of the datasets (Chiang & Grant, 2017, 2019). In Figure 3, each move is represented by a different colour,

and horizontal bars represent individual texts. It is immediately clear by looking at the colour distribution that very little stable patterning is present, apart from Openings and Sign-Off. On the contrary, there is great variety in the possible order that the moves can take across different texts. Figure 4 is a different take on a move map, showing the positions each individual move can take in the texts. It importantly reveals that one of the main reasons why we were unable to find any typical move patterns is that most moves do not have any typical position in the texts in the first place.

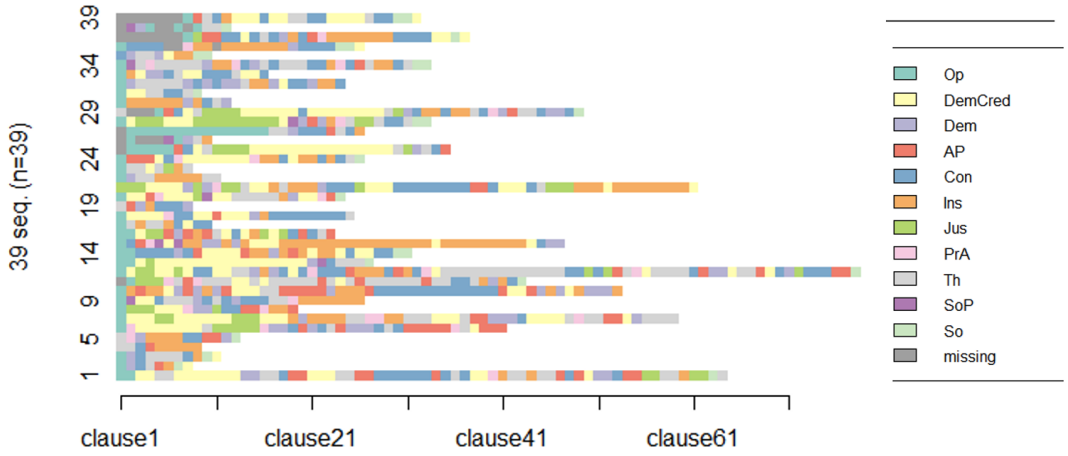



**Figure 3. Move maps of commercial extortion communications. The moves in the legend are shown in alphabetical order: AdP - Additional Persuasion, Con - Consequences, Dem - Demand, DemCred - Demonstrating Credibility, In - Instructions, Jus - Justification, Op - Openings, PrA - Pre-announcement, SO - Sign offs, SoP - Statement of Purpose, Th - Threat.**

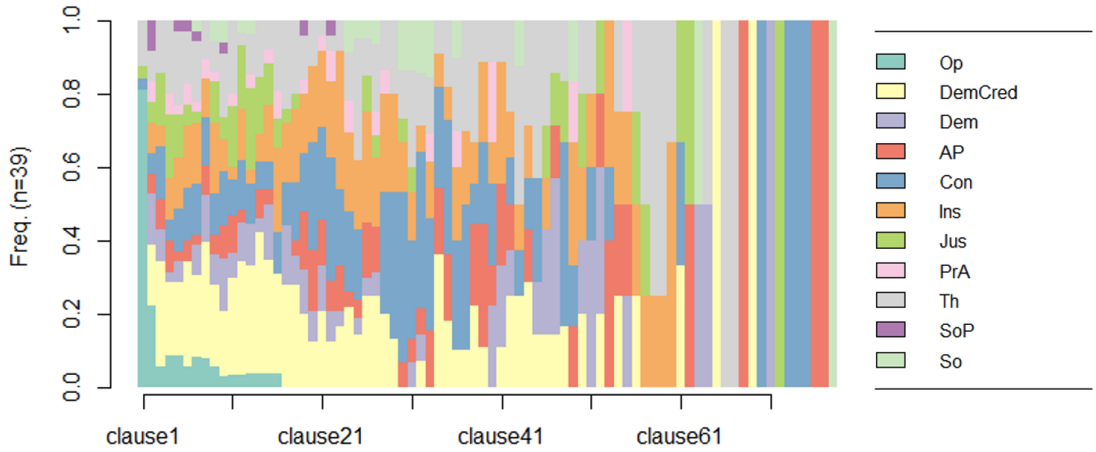



**Figure 4. Positions of individual moves across communications. The moves in the legend are shown in alphabetical order: AdP - Additional Persuasion, Con - Consequences, Dem - Demand, DemCred - Demonstrating Credibility, In - Instructions, Jus - Justification, Op - Openings, PrA - Pre-announcement, SO - Sign offs, SoP - Statement of Purpose, Th - Threat.**

Apart from Openings and Sign-Offs, which of course appear at the start and end of the communications, all the other moves can take any position. Threats for example

(in gray at the top of the graph) can appear in virtually any position, from clause 1 to clause 70.

Overall, then, although the texts in this collection of extortion notes hold a similar communicative purpose and two core moves decisively emerge from the data, there is no consistent sequence or pattern to those moves. The relationship between the texts is perhaps of a somewhat 'looser kind' (Swales, 1990, p. 49) but they share a family resemblance through the core moves contained (Swales, 1990, p. 52), perhaps partly accounted for by the lack of a coherent discourse community within which these texts are written.

### 5.3. Clusters of moves

Since no dominant move pattern was found in the data, we turned to explore whether groups of texts can be identified within the corpus based on move frequency (and not positioning). The use of computational techniques that use rhetorical moves as input data is an innovative method first introduced by Chiang and Grant (2019). Here, we use clustering algorithms based on move frequency to explore whether texts can be grouped into statistically relevant and meaningful clusters based on move presence, absence, or frequency.[4] In other words, we want to see if moves can be reliably used not only as a descriptive tool, but also as a reliable variable for text differentiation. If that is the case, our more principled approach to move analysis can be potentially used as an investigative tool (similarly to what Chiang and Grant, 2019 propose).

As a preparatory step, we adopt term frequency-inverse document frequency (TF-IDF), which determines how relevant a word is to a particular document in view of the corpus under consideration. This is done by multiplying the frequency of the word in a document by the inverse document frequency of the word across the entire set of documents. In our case, instead of word frequency we use move frequency. Each text is hence defined by a set of 11 features, where each feature is the product of move frequency within each text, and the inverse frequency of texts in which each move appears. To prevent small feature values from being overshadowed by large ones, all features were scaled to mean value of zero and standard deviation of one, as a means of normalisation. We also estimated products of features by taking any 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 features at a time (Naseer, Noori, Qureshi, & Hong, 2016). Empirical selection was performed with a view to identifying explainable, logical patterns in the dataset. This procedure revealed that a combination of 6 and 7 features was optimal, hence we augmented our dataset with a combination of these feature products taking any 6 and 7 features at a time. Hence, the dataset consists of 39 texts each defined by a total of 803 features FORMULA.

To ensure reliability, three clustering algorithms were applied: *hierarchical clustering* (Murtagh & Legendre, 2014), *k-means clustering* (Hartigan & Wong, 1979) and *partition around medoids clustering* (*PAM*, Kaufman & Rousseeuw, 1990). The number of clusters tested ranged from 2 to 6, reflecting the corpus's limited size. Performance was evaluated using both interpretability and statistical metrics – connectedness, compactness, separation, and stability – using the R packages 'clValid' and 'fpc'(Brock, Pihur, Datta, & Datta, 2008; Hennig, 2020). These metrics assess (respectively) how well texts

[4] R code is available as supplemental material.

are grouped, how distinct the clusters are, and how consistent results remain under resampling. Clustering was based on Euclidean distance and Ward's linkage, a method favoured in linguistic studies for producing balanced clusters.

Agglomerative hierarchical clustering with two clusters emerged as the most effective method, producing robust and interpretable groupings. Notably, both hierarchical and k-means clustering yielded identical results: one cluster contained two specific texts (S14L20 and S21L28), while the other included the remaining 37 (see Figure 4). This convergence, despite differing methodologies, suggests a strong underlying structure in the data (Handl, Knowles, & Kell, 2005).

The resulting clusters also demonstrated high stability scores (0.96 and 0.84), reinforcing the reliability of the findings and the presence of meaningful patterns in the corpus. Results show that cluster 1 consists of texts displaying all 11 moves, while texts in cluster 2 have at least one of the moves missing. This means that the cluster algorithm was able to group the texts based on the presence or absence of atypical moves. When we explored which of the atypical moves provide greater separation between the two clusters, we find that combinations including 'Justification' and 'Statement of purpose' make the greatest contribution. That is to say that the top 10 feature combinations which include those two moves increased the separation between the clusters.

This result suggests that the moves identified in the corpus have at least some predictive power. The difference between core and atypical moves seems to be crucial in separating groups of letters, which – in datasets with more reliable metadata – can be useful for investigative contexts.

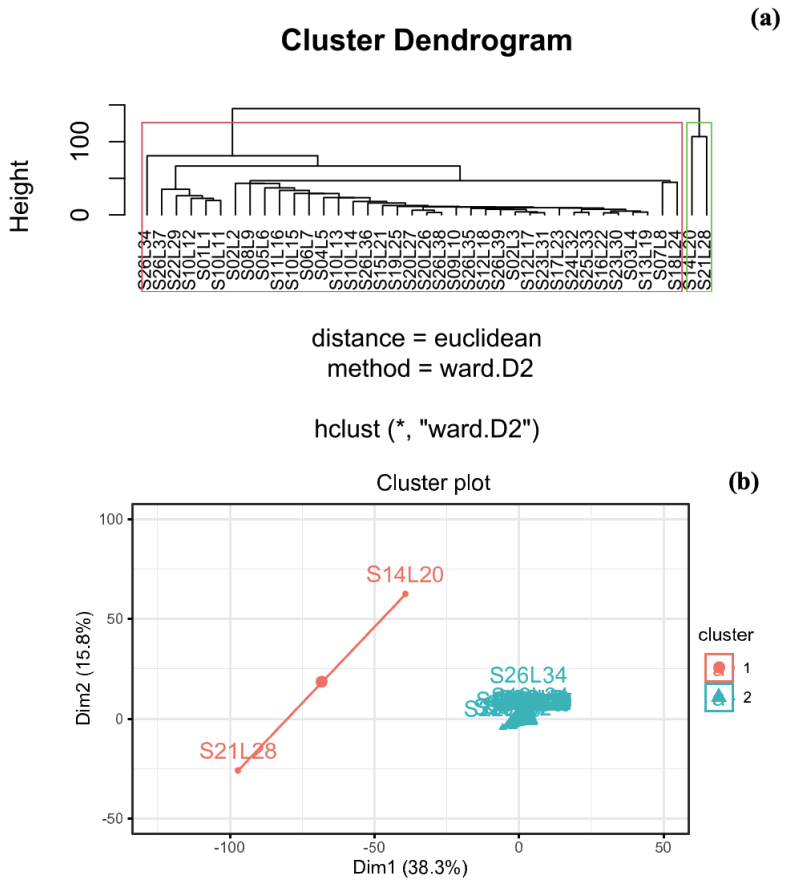



**Figure 5. (a) Cluster dendogram for hierarchical agglomerative clustering. The labels represent identifier for each text in the corpus. The red and green rectangles represent the two clusters identified by the algorithm. (b) Scatter plot of the clusters identified by K-means (K=2) algorithm. Cluster 1 is represented in red and cluster 2 is represented in blue. The tightly overlapping labels for cluster 2 indicate that these data points all occupy the same area in the two-dimensional space.**

## 6. Conclusions

This study provided an innovative approach to applying Move Analysis to the illicit genre (Bojsen-Møller et al., 2020) of commercial extortion letters, providing a noteworthy profile of the functional features in this genre. Specifically, we analyse the biggest corpus so far of commercial extortion letters, EXCrOW, with an improved version of Move Analysis. We tackle the lack of methodological clarity of the method by introducing clauses as the minimum unit of analysis and performing iterative coding and interrater reliability testing.

We found 11 rhetorical moves in our data: Opening, Sign-Off, Demand, Instructions, Justification, Threat, Demonstrating Credibility, Consequences, Additional Persuasion, Statement of Purpose, and Pre-Announcement. The two most prevalent moves identified are Threats and Demands, which appear in 92.3% of the texts. Beyond these, however, we have also found six typical moves which appear in more than half of the texts: Consequences, Openings, Instructions, Demonstrating Credibility, Additional Persuasion, and Sign-Off. Our analysis therefore highlights that while demands and threats are, perhaps inevitably, the most fundamental persuasive tools of the genre, authors

also resort to a plethora of other communicative techniques, such as discussing Consequences of compliance and non-compliance, Demonstrating Credibility and Additional Persuasion. This suggests that in assessing the level of threat posed by an extortion letter, a more nuanced approach should be taken, considering not only explicit or implicit threats alone.

With regards to the discourse structure of commercial extortion communications, corpus linguistic analyses and move maps show that there are no discernible structural patterns in the data. Corpus-based analyses of moves – as discussed in Upton and Cohen (2009) – are essential for a more generalisable but at the same time fine-grained study of discourse structures and function. Extortion communications, like other illicit genres that lack a community of practice, seem to be better described in terms of prevalence rather than recurrent patterns of moves. In other words, while traditional genres like business letters or academic articles generally display a fixed structure, with identifiable structural patterns of moves, extortion communications exhibit a flexible internal structure. This is confirmed by our computational analysis, that using clustering algorithms demonstrates that meaningful distinctive groups of texts can be identified based on the presence or absence of rhetorical moves. We interpret these result as indicating that extortion as a genre is mainly driven by function rather than by strict internal formal constraints. This functional nature of extortion letters is due – we claim – to the absence of a coherent discourse community.

Our findings have important implications for the study of forensic genres at large. In fact, corpus-based analyses of recurrent patterns of discourse units (Biber et al., 2007; Upton and Cohen, 2009) work remarkably well for quantifying discourse information for most genres, but forensic genres – such as suicide notes or extortion letters – escape the notion of genre as made of recurrent and predictable patterns of discourse units. Rather, the rhetorical moves in these text types combine freely with no discernible pattern. Therefore, we argue that using clustering based on move prevalence – rather than methods rigidly based on sequencing – is an alternative quantitative analysis tool for illicit genres.

In terms of methodological implications, this paper intends to inform the methods of future studies that employ rhetorical moves. Firstly, using a consistent and easy-to-define unit of analysis (such as clauses) provides for a higher and better reliability to move analysis, making the method more principled and replicable. Secondly, the use of multiple coders and interrater reliability testing ensures that the coding of rhetorical moves remains principled and potentially replicable. Finally, we also combine qualitative move analysis with quantitative methods such as move maps, ngrams and collocation analysis of moves, and cluster analysis. This demonstrates that a principled move analysis can be successfully combined with quantitative tools to explore move patterns and positions within individual texts as well as distinctive groups of texts in a corpus.

## References

Abaalkhail, A. (2015). *Rhetorical moves in an occluded genre: A qualitative analysis of suicide notes* (PhD dissertation). Carleton University.

Askehave, I., & Swales, J. M. (2001). Genre identification and communicative purpose: a problem and a possible solution. *Applied Linguistics*, *22*(2), 195–212.

Retrieved 2025-08-12, from https://academic.oup.com/applij/article-lookup/doi/10.1093/applin/22.2.195 doi: 10.1093/applin/22.2.195

Bhatia, V. K. (1993). *Analysing Genre*. Longman.

Bhatia, V. K. (1996). Methodological Issues in Genre Analysis. *Hermes, Journal of Linguistics*, *16*, 39–59. Retrieved 2025-08-12, from https://tidsskrift.dk/her/article/view/25383 doi: 10.7146/hjlcb.v9i16.25383

Bhatia, V. K. (2016). *Critical genre analysis: Investigating interdiscursive performance in professional practice*. Routledge.

Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the Move: Using corpus analysis to describe discourse structure* (Vol. 28). Amsterdam: John Benjamins Publishing Company. Retrieved 2025-08-12, from http://www.jbe-platform.com/content/books/9789027291912 doi: 10.1075/scl.28

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.

Bojsen-Møller, M., Auken, S., Devitt, A. J., & Christensen, T. K. (2020). Illicit Genres: The Case of Threatening Communications. *Sakprosa*, *12*(1), 1–53. Retrieved 2025-08-12, from https://journals.uio.no/sakprosa/article/view/7416 doi: 10.5617/sakprosa.7416

Brezina, V., Weill-Tessier, P., & McEnery, T. (2020). *#LancsBox.*

Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). **clValid** : An *R* Package for Cluster Validation. *Journal of Statistical Software*, *25*(4). Retrieved 2025-08-12, from http://www.jstatsoft.org/v25/i04/ doi: 10.18637/jss.v025.i04

Campbell, K. S., & Naidoo, J. S. (2017). Rhetorical Move Structure in High-Tech Marketing White Papers. *Journal of Business and Technical Communication*, *31*(1), 94–118. Retrieved 2025-08-12, from https://journals.sagepub.com/doi/10.1177/1050651916667532 doi: 10.1177/1050651916667532

Chiang, E. (2018). *Rhetorical moves and identity performance in online child sexual abuse interactions* (Unpublished doctoral dissertation). Aston University.

Chiang, E., & Grant, T. (2017). Online grooming: Moves and strategies. *Language and Law / Linguagem e Direito*, *4*(1), 103–141.

Chiang, E., & Grant, T. (2019). Deceptive Identity Performance: Offender Moves and Multiple Identities in Online Child Abuse Conversations. *Applied Linguistics*, *40*(4), 675–698. Retrieved 2025-08-12, from https://academic.oup.com/applij/article/40/4/675/4952155 doi: 10.1093/applin/amy007

Corbett, J. (2006). Genre and genre analysis. In K. Brown (Ed.), *Concise encyclopedia of pragmatics* (pp. 286–292). Elsevier.

Dos Santos, V. (2002). Genre analysis of business letters of negotiation. *English for Specific Purposes*, *21*(2), 167–199. Retrieved 2025-08-12, from https://linkinghub.elsevier.com/retrieve/pii/S0889490600000284 doi: 10.1016/S0889-4906(00)00028-4

Fobbe, E. (2011). *Forensische Linguistik. Eine Einführung*. Narr.

Fobbe, E. (2020). Text-Linguistic Analysis in Forensic Authorship Attribution. *International Journal of Language & Law*, *9*, 93–114. Retrieved 2025-08-12, from https://www.languageandlaw.eu/jll/article/view/78 (Publisher: International Language and Law Association (ILLA)) doi: 10.14762/JLL.2020.093

Freadman, A. (2012). The Traps and Trappings of Genre Theory. *Applied Linguistics*, *33*(5), 544–563. Retrieved 2025-08-12, from https://academic.oup.com/applij/

article-lookup/doi/10.1093/applin/ams050 doi: 10.1093/applin/ams050

Gales, T. (2015). The stance of stalking: a corpus-based analysis of grammatical markers of stance in threatening communications. *Corpora*, *10*(2), 171–200. Retrieved 2025-08-14, from https://www.euppublishing.com/doi/10.3366/cor.2015.0073 doi: 10.3366/cor.2015.0073

Handl, J., Knowles, J., & Kell, D. B. (2005). Computational cluster validation in post-genomic data analysis. *Bioinformatics*, *21*(15), 3201–3212. Retrieved 2025-08-12, from https://academic.oup.com/bioinformatics/article/21/15/3201/195678 doi: 10.1093/bioinformatics/bti517

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, *28*(1), 100. Retrieved 2025-08-12, from https://www.jstor.org/stable/10.2307/2346830?origin=crossref doi: 10.2307/2346830

Hennig, C. (2020). *Minkowski distances and standardisation for clustering and classification of high dimensional data.* arXiv. Retrieved 2025-08-14, from https://arxiv.org/abs/1911.13272 (Version Number: 2) doi: 10.48550/ARXIV.1911.13272

Hurt, M. (2020). *Pledging to Harm: A linguistic analysis of violent intent in threatening language* (PhD dissertation). Aston University.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis* (1st ed.). Wiley. Retrieved 2025-08-12, from https://onlinelibrary.wiley.com/doi/book/10.1002/9780470316801 doi: 10.1002/9780470316801

Liu, J., & Deng, L. (2016). A genre analysis of web-based crowdfunding discourse. *Asian ESP Journal*, *12*(2), 171–202.

Martin, J. R., Christie, F., Rothery, J., & Reid, I. (1987). Social process in education: A reply to Sawyer and Watson and others. In *The place of genre in learning: Current debates* (pp. 58–82). Deakin University.

Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech*, *70*(2), 151–167. Retrieved 2025-08-12, from http://www.tandfonline.com/doi/abs/10.1080/00335638409383686 doi: 10.1080/00335638409383686

Miller, C. R. (2017). "Where Do Genres Come From?". In C. R. Miller & A. R. Kelly (Eds.), *Emerging Genres in New Media Environments* (pp. 1–34). Cham: Springer International Publishing. Retrieved 2025-08-12, from http://link.springer.com/10.1007/978-3-319-40295-6_1 doi: 10.1007/978-3-319-40295-6_1

Moreno, A. I., & Swales, J. M. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes*, *50*, 40–63. Retrieved 2025-08-12, from https://linkinghub.elsevier.com/retrieve/pii/S0889490617303332 doi: 10.1016/j.esp.2017.11.006

Murtagh, F., & Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, *31*(3), 274–295. Retrieved 2025-08-12, from http://link.springer.com/10.1007/s00357-014-9161-z doi: 10.1007/s00357-014-9161-z

Naseer, N., Noori, F. M., Qureshi, N. K., & Hong, K.-S. (2016). Determining Optimal Feature-Combination for LDA Classification of Functional Near-Infrared Spectroscopy Signals in Brain-Computer Interface Application. *Frontiers in Human Neuroscience*, *10*. Retrieved 2025-08-12, from http://journal.frontiersin.org/Article/10.3389/fnhum.2016.00237/abstract doi: 10.3389/fnhum.2016.00237

Neumaier, T. (2025). "I have come to the conclusion that you must die": Threats in Late

Modern English threatening letters. *Journal of Historical Pragmatics*, *26*(2), 262–287. Retrieved 2025-08-12, from http://www.jbe-platform.com/content/journals/10.1075/jhp.23016.neu doi: 10.1075/jhp.23016.neu

Nini, A. (2017). Register variation in malicious forensic texts. *The International Journal of Speech, Language and the Law*, *24*(1), 99–126. Retrieved 2025-08-14, from https://utppublishing.com/doi/10.1558/ijsll.30173 doi: 10.1558/ijsll.30173

Nini, A. (2019). The Multi-Dimensional Analysis Tagger. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-Dimensional Analysis: Research Methods and Current Issues* (pp. 67–94). Bloomsbury Publishing PLC.

Rau, G., & Shih, Y.-S. (2021). Evaluation of Cohen's kappa and other measures of interrater agreement for genre analysis and other nominal data. *Journal of English for Academic Purposes*, *53*, 101026. Retrieved 2025-08-12, from https://linkinghub.elsevier.com/retrieve/pii/S1475158521000709 doi: 10.1016/j.jeap.2021.101026

Reczek, A. P. (2023). *Forensic Applications of Appraisal Theory and Genre to Threatening and Malicious Language* (PhD dissertation). Aston University.

Samraj, B., & Gawron, J. M. (2015). The suicide note as a genre: Implications for genre theory. *Journal of English for Academic Purposes*, *19*, 88–101. Retrieved 2025-08-12, from https://linkinghub.elsevier.com/retrieve/pii/S1475158515000454 doi: 10.1016/j.jeap.2015.04.006

Skalicky, S. (2013). Was this analysis helpful? A genre analysis of the Amazon.com discourse community and its "most helpful" product reviews. *Discourse, Context & Media*, *2*(2), 84–93. Retrieved 2025-08-12, from https://linkinghub.elsevier.com/retrieve/pii/S2211695813000160 doi: 10.1016/j.dcm.2013.04.001

Spitzberg, B., & Gawron, J. (2016). Toward Online Linguistic Surveillance of Threatening Messages. *Journal of Digital Forensics, Security and Law*. Retrieved 2025-08-12, from http://commons.erau.edu/jdfsl/vol11/iss3/7/ doi: 10.15394/jdfsl.2016.1418

Stemler, S. E. (2004). A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Interrater Reliability. *Practical Assessment, Research & Evaluation*, *9*(4). Retrieved 2025-08-12, from https://openpublishing.library.umass.edu/pare/article/id/1540/ (Publisher: University of Massachusetts Amherst) doi: 10.7275/96JP-XZ07

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

Swales, J. (2004). *Research genres: Explorations and applications*. Cambridge University Press.

Tompkinson, J. (2023). *Spoken threats from production to perception*. Cambridge University Press.

Upton, T. A., & Cohen, M. A. (2009). An approach to corpus-based discourse analysis: The move analysis as example. *Discourse Studies*, *11*(5), 585–605. Retrieved 2025-08-12, from https://journals.sagepub.com/doi/10.1177/1461445609341006 doi: 10.1177/1461445609341006

Xia, S. A. (2020). Genre Analysis in the Digital Era: Developments and Challenges. *ESP Today*, *8*(1), 141–159. Retrieved 2025-08-12, from http://doi.fil.bg.ac.rs/volume.php?pt=journals&issue=esptoday-2020-8-1&i=7 doi: 10.18485/esptoday.2020.8.1.7