

BOOK OF ABSTRACTS

5TH ILLA FOCUS CONFERENCE ON FORENSIC LINGUISTICS:
LANGUAGE AS EVIDENCE

2 SEPTEMBER 2025

VYTAUTAS MAGNUS UNIVERSITY, KAUNAS, LITHUANIA

ORGANISERS: VICTORIA GUILLÉN-NIETO AND DIETER STEIN



International
Language and Law
Association



VYTAUTAS
MAGNUS
UNIVERSITY
Faculty of
Humanities



VYTAUTAS
MAGNUS
UNIVERSITY
Faculty of Law

CONFERENCE ORGANISERS:

VICTORIA GUILLÉN-NIETO

&

DIETER STEIN

Victoria Guillén Nieto graduated from the University of Alicante with a degree in English Philology in 1986. She obtained a Master's degree in Applied Linguistics from the same university in 1987 and a Master's in Forensic Linguistics from Pompeu Fabra University in Barcelona in 2008. Guillén-Nieto earned her doctorate in English Philology at the University of Alicante in 1993. Since 1997 she has been Senior Lecturer in Applied Linguistics at the Department of English Studies. She was Chair of this department from 2003 to 2009. From September 2019 to September 2021, she was President of the International Language and Law Association (ILLA) for Linguistics. She is currently directing the Masters in English and Spanish for Specific Purposes and the Dual Master's in English and Spanish for Specific Purposes and Forensic Linguistics, co-organised by the University of Alicante and the East China University of Political Sciences and Law (ECUPL) in Shanghai. Her latest books include *From Fear to Hate. Legal-Linguistic Perspectives on Migration* (co-edited with Antonio Doval and Dieter Stein) (De Gruyter August 2023); *Hate Speech, Linguistic Perspectives* (De Gruyter 2023); *Language as Evidence: Doing Forensic Linguistics* (co-edited with D. Stein) (Palgrave Macmillan/Springer 2022); *Manual of Romance Forensic Linguistics* (with Dieter Stein) (De Gruyter 2025); and *The Language of Harassment. Pragmatic Perspectives on Language as Evidence* (Lexington Books 2024). Since 2009, she has provided professional linguistic service as an expert witness in Forensic Linguistics in Spain, Switzerland, Sweden, and the USA.

Dieter Stein is emeritus professor of English Language and Linguistics at Heinrich-Heine University Düsseldorf. One main focus of his work is language development, others deal with pragmatics, open access publishing, the linguistics of the Internet, language in the legal domain and forensic linguistics. In addition to teaching at his home universities of Düsseldorf and Cologne, he has taught at several foreign universities including Canada and China as well as UCLA Los Angeles law school and the linguistic institute of the Linguistic Society of America. He has also served as Editor-in-Chief of LSA's online publication portal "eLanguage", as president of the International Language and Law Society, is profesor colaborador honorífico at the University of Alicante and series editor of de Gruyter's "Foundations in Language and Law" series.

TABLE OF CONTENTS

| | |
|---|-----------|
| Cristina Aggazzotti (Johns Hopkins University), Elizabeth Allyn Smith (Universite du Quebec a Montreal) | 4 |
| The relationship between linguistic accommodation and speaker attribution | 4 |
| Ahmad Adha (University of Szeged) | 5 |
| Migrant Workers Do Not Think That is a Lie | 5 |
| Andrea Mojedano Batel (Aston University) | 5 |
| Native Dialect Influence Detection (NDID): Pragmalinguistic differences among Argentinian Spanish, Colombian Spanish, Brazilian Portuguese, and European Portuguese speakers in non-native English suggestions | 5 |
| Ella Fletcher (University of Alicante – IULMA) | 6 |
| The impact of syntactic pre-processing on the accuracy of the automatic authorship attribution of social media texts | 6 |
| Edita Gruodytė (Vytautas Magnus University), Ugnė Urbšytė (Vytautas Magnus University) | 7 |
| The Limits of Political Speech in the Context of ECtHR hate speech Practice: Can Clear Boundaries Be Established? .. | 7 |
| Victoria Guillén-Nieto (University of Alicante) | 8 |
| The pragmatics of emotional manipulation in cases of gender-based violence | 8 |
| Timothy Habick (Reasoning, Inc.) | 10 |
| The Forensic Risks of AI Detection: Bias, Misclassification, and Ethical Issues | 10 |
| Nuria Lorenzo-Dus (Swansea University), Carmen Perez-Sabater (Universidad Politecnica de Valencia), A. Garcia-Montes (Universidad Politecnica de Valencia) | 11 |
| “pues a ver manda foto”: Requesting child sexual abuse material in online grooming interactions | 11 |
| Elena Morandini | 11 |
| Mafia Language as Evidence: Threats and Trust in Mafia Discourse | 11 |
| Jurate Ruzaite (Vytautas Magnus University), Justina Urbonaite (Vytautas Magnus University), Gintare Herasimenkiene (Vytautas Magnus University) | 12 |
| The role of linguistic criteria in hate speech identification: Insights from ECtHR Case-Law | 12 |
| Sara-Sophie Schedel (Phillips-Universität Marburg), Gea de Jong-Lendle (Philipps-Universität Marburg), Mathias Scharinger (Philipps-Universität Marburg) | 13 |
| Would you recognise a fake voice? - A perceptual study on the recognition ability of natural and synthetic Speech | 13 |
| Stefania Selegian (University of Bucharest) | 15 |
| Linguistic markers of promise in corruption offences | 15 |
| Michał Szczyszek (UAM Poznań), Ewa Badyda (Uniwersytet Gdański), Paweł Nowak (UMCS), Jakub Kępiński (UAM), Leszek Szymański (Uniwersytet Zielonogórski) | 16 |
| Plagiarism – a linguistic and legal perspective. Polish court cases with expert linguists’ opinions | 16 |
| PANEL | 18 |
| The Role of Experts and the Court: Theory and Practice | 18 |

CRISTINA AGGAZZOTTI (JOHNS HOPKINS UNIVERSITY), ELIZABETH ALLYN SMITH (UNIVERSITE DU QUEBEC A MONTREAL)

THE RELATIONSHIP BETWEEN LINGUISTIC ACCOMMODATION AND SPEAKER ATTRIBUTION

Purpose statement: We investigate linguistic accommodation in a speech transcript dataset and how it could impact forensic speaker attribution.

Research problem: Accommodation occurs when conversational participants adapt their communication style (often subconsciously) to that of the other person (Giles et al., 1991). A wealth of literature across fields shows that accommodation (cf. also convergence, alignment) is a conversational default, but, to our knowledge, it has not previously been studied in the context of speaker or authorship attribution. Many attribution systems use style-based features to distinguish individuals from one another, and given that these features may change over the time course of interaction between individuals, this raises the question of how accommodation might impact attribution accuracy.

Research question: How might linguistic accommodation affect speaker attribution in a conversational speech transcript corpus?

Method: We use the openly available toolkit ALIGN (Duran et al., 2019) to measure accommodation in various ways, including POS tag, token, and lemma n-grams on the Fisher Speech Transcript corpus (Cieri et al., 2004). Fisher consists of 10-minute phone calls between strangers compensated for their time and asked to discuss a conversational topic, all factors favoring accommodation. We then explore various aspects of the ALIGN results and discuss them in light of past stylometric attribution results on the same corpus (Aggazzotti et al., 2023).

Impact/state of the art: Since attribution methods often use stylistic features, it is crucial to know whether conversational participants accommodate, as each of their personal styles could begin to resemble the other's to some extent. As there is no previous work specifically evaluating accommodation in the context of speaker or authorship attribution, there is no state of the art in this area, nor even in the general case, as studies using different datasets, metrics, linguistic properties, and conversational units are not comparable.

Keywords: linguistic accommodation; alignment; speaker attribution; speech transcripts; computational linguistics

References

- Aggazzotti, C., Smith E. A., & Andrews, N. (2023). Testing authorship analysis on spoken language transcripts: Establishing a first benchmark. The 4th ILLA Focus Conference on Forensic Linguistics.
- Cieri, C., Graff, D., Kimball, O., Miller, D., & Walker, K. (2004). Fisher English Training Speech Transcripts. Linguistic Data Consortium. <https://catalog.ldc.upenn.edu/LDC2004T19>
- Duran, N., Paxton, A., & Fusaroli, R. (2019). ALIGN: Analyzing Linguistic Interactions with Generalizable techNiques—A Python library. *Psychological Methods*, 24, 419–438. <https://doi.org/10.1037/met0000206>
- Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of Accommodation: Developments in*

AHMAD ADHA (UNIVERSITY OF SZEGED)

MIGRANT WORKERS DO NOT THINK THAT IS A LIE

Recently, an increasing number of foreigners come to Hungary. Linguistic and cultural differences between the migrant workers and the locals might lead to different understanding, for example in the perception of lie. The aim of this research is to find out the perception of lie in different cultural settings. Currently, there is no unanimous definition of lie. Most scholars define lie based on the speaker's false belief, similar with studies with English (Coleman & Kay, 1981) and Hungarian (Németh T. & Adha, 2021) speakers. However, Mandarin (Adha & Li, 2023) and Japanese (Yoshimura, 1995) speakers consider falsehood as the important element to define lie. For the present study, people from Indonesia, Mongolia and Vietnam are expected to have a similar perception of lie with people from China and Japan, and different to the perception of Hungarians. People from each group will take a questionnaire adapted from Coleman & Kay's (1981) study. The questionnaire contains stories based on the presence and/or absence of falsehood, false belief, and deceptive intention. The present research will contribute to our understanding of the universality of the definition of lie, which is fundamental for research in lie detection.

Keywords: lie, deception, Kyrgyz, Mongol, Vietnam

References

- Adha, A., & Li, X. (2023). The analysis of perception of lying by Chinese Mandarin speakers. *East Asian Pragmatics*, 8(1), 109–131.
- Coleman, L., & Paul, K. (1981). Prototype semantics: the English world lie. *Language*, 57(1), 26–44.
- Németh T. E., & Adha, A. (2021). Valóban könnyebb utolérni a hazug embert, mint a sánta kutyát? [Is it really easier to catch a liar than a limping dog?]. *Argumentum*, 17, 488–510.
- Yoshimura, K. (1995). *Ninchi imiron no hoho* [Methods of Cognitive Semantics]. Jibun Shobo.

ANDREA MOJEDANO BATEL (ASTON UNIVERSITY)

NATIVE DIALECT INFLUENCE DETECTION (NDID): PRAGMALINGUISTIC DIFFERENCES AMONG ARGENTINIAN SPANISH, COLOMBIAN SPANISH, BRAZILIAN PORTUGUESE, AND EUROPEAN PORTUGUESE SPEAKERS IN NON-NATIVE ENGLISH SUGGESTIONS

This study tackles cross-linguistic and inter-dialectal differences in non-native English (E2) suggestions produced by native Spanish speakers (S1) from Argentina and Colombia and native Portuguese speakers (P1) from Brazil and Portugal. The theoretical framework and methodology are situated within the field of Native Dialect Influence Detection (NDID), developed in [name and years deleted to maintain the integrity of the review process]. NDID is an emerging subfield of forensic authorship analysis that focuses on determining a speaker's native dialect and differentiating between closely related languages in non-native linguistic data—a task where computational models have demonstrated limited success (e.g., Kochmar, 2011). The

overarching aim of NDID is to estimate group belonging, and it can greatly aid in investigations, benefitting law enforcement agencies or organizations dealing with non-native speakers.

This research examines linguistic features at morphological, syntactic, semantic, and pragmatic levels that serve to identify the speech act of suggestions in the four linguistic varieties under scrutiny. Data were collected from the Tripadvisor website and through written discourse-completion tasks. Quantitative methods include creating a list of token n-grams for each variety and ranking n-grams by keyness, through SketchEngine (Kilgarrieff et al., 2014). Qualitative methods involve coding all speech acts, which then underwent linguistic analyses. Importantly, suggestions are relevant in terms of forensic applications because, like threats, they are strategies commonly used to gain others' compliance (Kellerman & Shea, 1996).

Findings show that several linguistic features (e.g., verbs of ability, epistemic constructions, verbal tense and mood, and mitigators) are stable sites in E2 suggestions produced by Spanish and Portuguese native speakers. Crucially, it establishes that Portuguese dialects exhibit internal variation in E2 suggestions and differ from Spanish dialects. These findings underscore NDID's efficacy in distinguishing between closely related languages in non-native data, reinforcing the applicability of NDID in forensic authorship analysis.

Keywords: Native Dialect Influence Detection; authorship analysis; linguistic profiling; native language influence detection; computer-mediated communication

References

Kellermann, K., & Shea, B.C. (1996). Threats, suggestions, hints, and promises: Gaining compliance efficiently and politely. *Communication Quarterly*, 44(2), 145–165.

Kilgarrieff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>

Kochmar, E. (2011). *Identification of a Writer's Native Language by Error Analysis*. Dissertation submitted for the degree of Master of Philosophy in Advanced Computer Science. University of Cambridge, Computer Laboratory, St. John's College. University of Cambridge [Online]. Available at: http://www.cl.cam.ac.uk/~ek358/Native_Language_Detection.pdf (Accessed 31 January 2022).

ELLA FLETCHER (UNIVERSITY OF ALICANTE – IULMA)

THE IMPACT OF SYNTACTIC PRE-PROCESSING ON THE ACCURACY OF THE AUTOMATIC AUTHORSHIP ATTRIBUTION OF SOCIAL MEDIA TEXTS

Authorship attribution has a well-established history as a forensic linguistic discipline with technological advances fomenting a shift towards automatic or computer-assisted approaches. Based on the concept of idiolect, authorship attribution assumes that every individual has a unique and distinguishable way of writing. Despite the strength of syntactic features as representing idiolect, as they are consistent and unconscious (Stamatatos, 2009), very limited research has been conducted into the impact of types of syntactic analysis on the authorship attribution process, especially for computer-mediated communications which typically comprise short texts. Even in approaches which use syntactic information, such as Tyo et al. (2022), we find a recurrence to larger datasets. This project has therefore developed a two-part alternative hypothesis: (1) syntactic information extracted in pre-processing and serving as the input for classification influences

attributive accuracy for short, social media texts, and (2) the type of syntactic information further influences this.

The types of syntactic information tested are plaintext (baseline), part-of-speech (POS) tags, and dependency labels with said information being extracted for each author from a Convokit corpus composed of texts from social media platform Reddit¹ using spaCy-stanza (Qi et al., 2020). This output was subsequently used as the input for classification tool Stylo (Eder et al., 2016) within which the `classify()` function was executed. A 5-fold cross-validation was also implemented to establish a mean accuracy across various textual orientations.

The results obtained have shown a clear preference for syntactic information over the baseline plaintext as the most accurate attributive method, providing evidence to support the first alternative hypothesis. POS tags generally outperformed dependencies, supporting the second alternative hypothesis. These results are particularly relevant when considering future professional applications using short texts, as the established method established was successful even when working with texts smaller than 100 words in length.

Keywords: authorship attribution; idiolect; syntactic analysis; text classification.

References

- Eder, M., Rybicki, J. and Kestemont, M. (2016). Stylometry with R: a package for computational text analysis. *R Journal*, 8(1): 107-21. <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. *CoRR*, abs/2003.07082. <https://arxiv.org/abs/2003.07082>
- Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the Association for Information Science and Technology*, 60(3), 538–556. <https://doi.org/10.1002/asi.21001>
- Tyo, J., Dhingra, B., & Lipton, Z. C. (2022). On the State of the Art in Authorship Attribution and Authorship Verification. *arXiv [Cs.CL]*. Retrieved June 14, 2023 from <http://arxiv.org/abs/2209.06869>

EDITA GRUODYTĖ (VYTAUTAS MAGNUS UNIVERSITY), UGNĖ URBŠYTĖ (VYTAUTAS MAGNUS UNIVERSITY)

THE LIMITS OF POLITICAL SPEECH IN THE CONTEXT OF ECTHR HATE SPEECH PRACTICE: CAN CLEAR BOUNDARIES BE ESTABLISHED?

Freedom of speech, protected under Article 10 of the European Convention on Human Rights, is undoubtedly a cornerstone of any democratic society. While its importance extends across various spheres, freedom of expression holds particular significance in political debate (Bleich and Al-Mateen, 2021). As evidenced by the jurisprudence of the European Court of Human Rights (ECtHR), a broad scope of free speech is essential to enable political parties to present their arguments, critique policies, advocate for ideas, and ultimately persuade voters. Political debate plays a crucial role in fostering the progress of the rule of law (Howard, 2020). However, the growing visibility and influence of politicians also pose significant challenges. Politicians' statements, especially when amplified by social networks and fuelled by advancements in AI and communication technologies, have the potential to spread both accurate and misleading information. This enhanced reach can magnify the harm caused by racist or xenophobic speech, which raises concerns about

¹ <https://convokit.cornell.edu/documentation/subreddit.html> [Last accessed 29-01-2025]

the need for balanced legal regulation. Finding the right equilibrium between protecting freedom of political speech and combating harmful rhetoric is particularly critical in this rapidly evolving technological landscape.

Despite extensive discourse on freedom of speech in legal doctrine, there remains a lack of systematic research addressing political speech specifically within the framework of ECtHR jurisprudence. This research responds to concerns raised by academics such as Mchangama and Alkiviadou (2021), who argue that the Court's case law on hate speech restrictions reveals inconsistencies that cast doubt on the extent to which offensive, shocking, or disturbing speech is truly safeguarded. As the line between political and hate speech becomes increasingly relevant, this analysis seeks to identify criteria within ECtHR practice that delineate this boundary.

To achieve this, the research addresses the following key questions:

1. Does political speech receive a distinct scope of legal protection under ECtHR jurisprudence?
2. What criteria does the ECtHR use to distinguish legitimate political speech in relation to hate speech?
3. Under what circumstances can the boundaries of protected political speech be said to have been transgressed?

The research employs a systematic analysis of ECtHR case law alongside a review of relevant academic literature. Findings demonstrate that the ECtHR evaluates the limits of freedom of expression using criteria such as the social and political context of the speech, the speaker's position and status in society, intent, content, form, extent of the speech, and the likelihood of incitement to action against specific groups. The Court's case law consistently underscores that freedom of expression is particularly vital for politicians, warranting restriction only in the presence of compelling justifications.

References

Judgment by the European Court of Human Rights (Grand Chamber), case of Selahattin Demirtaş v. Turkey (No. 2), Application no. 14305/17 of 22 December 2020.

Judgment by the European Court of Human Rights (Grand Chamber), case of Sanchez v. France, Application no. 45581/15, of May 15, 2023.

Bleich, E., & Al-Mateen, S. (2021). Hate speech and the European Court of Human Rights: Ideas and judicial decision-making. *Michigan State International Law Review*, 29(2), 179-212.

Mchangama, J., & Alkiviadou, N. (2021). Hate speech and the European Court of Human Rights: Whatever happened to the right to offend, shock, or disturb? *Human Rights Law Review*, 21(4), 1008–1042.

Howard, E. (2020). Political rhetoric, freedom of expression and hate speech against religious believers. In *European Populism and Human Rights*, 193-214.

VICTORIA GUILLÉN-NIETO (UNIVERSITY OF ALICANTE)

THE PRAGMATICS OF EMOTIONAL MANIPULATION IN CASES OF GENDER-BASED VIOLENCE

In psychology, 'manipulation' is an attempt to subtly or deceitfully influence or control another person's behaviour, thoughts, or emotions. Emotional manipulation is a common and powerful form of this, particularly in instances of gender-based violence. It involves leveraging a person's emotions to sway their behaviour or decisions. For example, the manipulator might appeal to the target victim's feelings to persuade them to undertake actions they would not normally consider. Emotional manipulation can take many forms, including guilt manipulation, victimhood, or emotional blackmail.

Emotional manipulation has not garnered considerable attention as a scientific subject in linguistic research. Some related work includes studies by Whitty and Buchanan (2012), Anesa (2020, 2022), and Carter (2024) on romance scams, which involve the emotional manipulation of the target victim. This paper seeks to offer a theoretical contribution to the study of emotional manipulation within the framework of gender-based violence from a pragmatic perspective. I argue that emotional manipulation is a communicative strategy whereby a speaker (the manipulator) aims to make the addressee (the target victim) perform action *p* by stating a proposition (a distractor proposition) that appeals to their emotions while not wishing to be held accountable for communicating *p*. Hence, in emotional manipulation, the speaker's primary intention is covert and must be conveyed implicitly and indirectly. Moreover, emotional manipulation may function differently depending on the case (Maillat & Oswald, 2009, p. 351). For instance, the communication of a proposition can mislead the addressee into relinquishing a legitimate belief, reinforcing a problematic belief, or preventing them from considering a legitimate belief, all of which can happen without the speaker resorting to lying. The manipulator must ensure that the target victim does not recognise the manipulative act; otherwise, they may be accused of malice, and the desired perlocutionary act could be dismissed. The intended victim will often secure uptake when it is too late to avert psychological harm.

Since emotional manipulation involves implicitness, indirectness, and some degree of deception (cf. Bertuccelli Papi 2003, 2014), it remains inherently ambiguous, equivocal, and elusive, complicating its substantiation in court. This paper, drawing on casework, analyses and compares the strategies employed by manipulators in various instances of gender-based violence adjudicated in Spanish courts. The findings could offer linguistic evidence of emotional manipulation for forensic linguists acting as expert witnesses.

Keywords: deception; emotional manipulation; implicitness; indirectness; gender-based violence.

References

- Anesa, Patrizia (2022). Fighting cybercrime through linguistic analysis. In Victoria Guillén-Nieto & Dieter Stein (eds.), *Language as Evidence. Doing Forensic Linguistics*, pp. 419-438. Palgrave Macmillan.
- Anesa, Patrizia (2020). Lovextortion: Persuasion strategies in romance cybercrime. *Discourse Context & Media*, 35, pp. 1-8.
- Bertuccelli Papi, Marcela (2014). The pragmatics of insinuation. *Intercultural Pragmatics*, 11(1), pp. 1-29.
- Bertuccelli Papi, Marcela (2003). The unsaid in James Joyce's "Araby": A linguistic perspective. *Journal of the Short Story in English*, 40, pp. 1-9.
- Carter, Elisabeth (2024). *The Language of Romance Crimes: Interactions of Love, Money, and Threat*. Cambridge: Cambridge University Press.
- Maillat, Didier & Oswald, Steve (2009). *Defining Manipulative Discourse: The Pragmatics of Cognitive Illusions*. *International Review of Pragmatics*, 1(2), pp. 348-370.

Whitty, Monica & Buchanan, Tom (2012). The online romance scam: A serious cybercrime. *Cyberpsychology, Behavior, and Social Networking*, 15(3), pp. 181-183.

TIMOTHY HABICK (REASONING, INC.)

THE FORENSIC RISKS OF AI DETECTION: BIAS, MISCLASSIFICATION, AND ETHICAL ISSUES

AI-detection tools claim to distinguish human-authored text from AI-generated content, yet they function as black-box classifiers reliant on statistical tendencies rather than transparent forensic methodologies. This paper examines how such tools misclassify highly proficient second-language writing as AI-generated due to commonalities in lexical richness, syntactic complexity, and low perplexity—hallmarks of both AI-generated text and meticulously crafted human prose.

The case of *Patel v. Tulane University* illustrates the serious consequences of such misclassification. Tulane's medical school falsely accused a student of AI plagiarism after detection software flagged her original essay as AI-composed. Independent forensic linguistic analysis revealed that the detector's assessment was flawed, exposing how AI tools sometimes penalize grammatically sophisticated second-language writing.

Forensic linguistics requires transparent, replicable methodologies for text classification—criteria that AI detection tools fail to meet. Their opaque algorithms, unregulated decision-making processes, and susceptibility to false positives challenge the admissibility of their results as forensic evidence. Nonetheless, universities increasingly rely on these tools to police academic integrity—often without understanding the detectors' biases, error rates, and limitations. The misuse of AI-detection tools can jeopardize academic standing, employment, immigration status, and professional reputations.

This paper highlights the ethical and forensic risks of uncritical reliance on AI detection tools. The disproportionate impact on vulnerable groups, including second-language writers and nontraditional students, underscores the need for responsible oversight. Forensic linguists have long focused on authorship identification, but the evolving landscape now demands expert oversight in AI methodologies, evidentiary reliability, and forensic best practices.

By critically evaluating AI-detection tools, we can expose their limitations and help prevent flawed automation from dictating authorship disputes. This case study serves as a cautionary example of the dangers of overreliance on AI detection, advocating for linguists' essential role in shaping ethical AI policies and ensuring that AI-authorship allegations are evaluated through principled forensic methodologies.

Keywords: AI detection bias; False positives in authorship attribution; Forensic linguistic methodology; Ethical risks in AI governance; Admissibility of AI-generated evidence

References

Liang, W., Yuksekogonul, M., Mao, Y., Wu, E., & Zou, J. (2023). *GPT detectors are biased against non-native English writers*. *Patterns*, 4(7), 100779.

[Author] (2016). *Rapport Final: Établissement des Procédures d'Examens des Baccalauréats Général et Professionnel (Djibouti National Exam Review for the World Bank)*.

OpenAI. (2023). *How AI text generation works: Language models and probability distributions*.

Phrasly AI Detector. (2024). *AI content detection and evaluation methodology*.

Hung, C.-Y., Hu, Z., Hu, Y., & Lee, R. K.-W. (2023). "Who Wrote it and Why? Prompting Large-Language Models for Authorship Verification." arXiv preprint arXiv:2310.08123.

NURIA LORENZO-DUS (SWANSEA UNIVERSITY), CARMEN PEREZ-SABATER (UNIVERSIDAD POLITECNICA DE VALENCIA), A. GARCIA-MONTES (UNIVERSIDAD POLITECNICA DE VALENCIA)

"PUES A VER MANDA FOTO": REQUESTING CHILD SEXUAL ABUSE MATERIAL IN ONLINE GROOMING INTERACTIONS

It takes an average of 45 minutes to engage a child in high-risk grooming interactions within social gaming environments (WeProtect Global Alliance, 2023). This typically includes groomers' requesting child sexual abuse material (CSAM) – the focus of this presentation. We discuss the findings of a study of the discourse of, and around, groomers' first CSAM requests. This is important given the link between obtention of coerced, 'self-generated' CSAM and both sextortion and barriers to child disclosure. Our data consists of 70 online grooming chatlogs in Spanish (103,396 words). We ask two main research questions:

- 1) How and when do groomers first request CSAM?
- 2) How and when do children respond to groomers' requests?

Having identified all the CSAM requests in the data, we conduct a discourse analysis of the first CSAM request sequence per chatlog. This includes drawing upon speech act theory (Blum-Kulka et al., 1989; Author, 2023) to identify the level of in/directness in the requests' head act and facework in any accompanying support moves. The analysis also entails examining children's facework in the request sequences.

Our results show a mean of 5.8 requests before the first CSAM is sent. Groomers favour directly worded requests (58%), supported with a strategic facework. Children's facework around these requests generally consists of positive politeness (78%). Our findings may be used to inform the development of targeted interventions against online grooming.

Keywords: online child sexual grooming; CSAM; discourse analysis; requests; facework

References

Blum-Kulka, S. et al (1989). *Cross-cultural pragmatics: Requests and apologies*. Ablex.

Lorenzo-Dus, N. (2023). *Digital grooming: Discourses of manipulation in cybercrime*. Oxford University Press.

WeProtect Global Alliance (2023). *Assessing the scale and scope of child sexual exploitation and abuse online, to transform the response*. WeProtect Global Alliance. <https://www.weprotect.org/global-threat-assessment-23/#full-report>

ELENA MORANDINI

MAFIA LANGUAGE AS EVIDENCE: THREATS AND TRUST IN MAFIA DISCOURSE

Mafia discourse serves as a strategic tool for coercion, deception, and control. Threats and trust-building mechanisms operate simultaneously to enforce hierarchy, regulate transactions, and ensure loyalty. However, these interactions are often implicit, masked by indirect speech acts and coded phrases. Forensic linguistics has not systematically examined how Mafia discourse encodes threats and trust as a structured communicative system despite their role in organized crime. This paper analyzes how threats and trust-building strategies co-exist in Mafia discourse, investigating whether trust expressions always indicate genuine loyalty or serve as a strategic tool, particularly in informant-led interactions. Existing forensic linguistic studies on threats and coercion have primarily focused on legal discourse (Shuy, 1993), police interrogations (Tiersma & Solan, 2012), and explicit threats (Muschalik, 2018). Meanwhile, computational linguistic methods have been applied to detect deception and persuasion in criminal language (Adams & Jarvis, 2006) but computational identification of implicit threats within Mafia speech remains unexplored.

This research investigates how informants may leverage trust to prompt self-incriminating responses. The study uses Critical Discourse Analysis, Speech Act Theory, and corpus-based forensic linguistics to analyze a corpus of Mafia wiretaps obtained from the RICO Act via PACER. Methods include manual annotation with CATMA, and computational analysis using AntConc, Voyant Tools, and spaCy. This approach examines lexical markers, collocations, and syntactic structures related to threats and trust, distinguishing between coercive and performative forms of trust within these interactions. Preliminary results suggest that Mafia discourse encodes threats within indirect speech acts, embedding coercion in assurances of loyalty. However, trust may not always reflect internal cohesion, as informants could manipulate it to steer conversations. These findings enhance forensic linguistics by improving wiretap analysis and linguistic profiling in organized crime investigations.

Keywords: Mafia discourse; computational forensic linguistics; criminal organizations; threats; trust

References

- Adams, S. H., & Jarvis, J. P. (2006). Indicators of veracity and deception: An analysis of written statements made to police. *Int. J. Speech, Lang. Law*, 13, 1–22. doi:10.1558/sll.2006.13.1.1
- Muschalik, J. (2018). *Threatening in English: A Mixed Method Approach*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Shuy, R. W. (1993). *Language crimes: The use and abuse of language evidence in the courtroom*. Cambridge, Massachusetts, USA: Blackwell Publishers Ltd.
- Tiersma, P. M., & Solan, L. M. (2012). *The Language of Crime*. Unpublished, Brooklyn Law School. doi:10.1093/oxfordhb/9780199572120.013.0025

JURATE RUZAITE (VYTAUTAS MAGNUS UNIVERSITY), JUSTINA URBONAITE (VYTAUTAS MAGNUS UNIVERSITY), GINTARE HERASIMENKIENE (VYTAUTAS MAGNUS UNIVERSITY)

THE ROLE OF LINGUISTIC CRITERIA IN HATE SPEECH IDENTIFICATION: INSIGHTS FROM ECTHR CASE-LAW

The identification of hate speech presents a challenge in both legal and social contexts (e.g., Fino, 2020; Gorenc, 2022), with linguistic criteria playing a key role in distinguishing harmful expressions from legitimate free speech, though significant uncertainty remains regarding how they can be formalized. This study

examines how the European Court of Human Rights (ECtHR) uses linguistic features in its hate speech rulings. Despite existing research on language use in ECtHR judgments (e.g. Peruzzo, 2019; Nikitina, 2023), the role of linguistic criteria in hate speech identification remains underexplored.

For this analysis, 51 cases in English from the HUDOC database were categorized by thematic focus: ethnic hatred (11 cases), religious hatred (6 cases), totalitarianism (4 cases), violence (20 cases), genocide denial (1 case), LGBTQ rights (8 cases), and one general case. This classification reflects the varying justifications used by the Court to identify hate speech based on its target. The study investigates cases with different outcomes: those in which textual items were considered outside the protection of Article 10 due to potentially hateful content, those that did not meet the threshold, and cases lacking a unanimous decision.

The research addresses the following questions: (1) How is linguistic expertise utilised in the decision-making of ECtHR cases?; (2) What language does the Court use to comment on the linguistic properties of hate speech, and how does linguistic argumentation vary across different types of cases?; (3) What levels of analysis are applied (micro, meso, macro)?; and (4) How do linguistic and contextual factors interact? To answer these questions, the court judgments are examined using a primarily qualitative approach based on the theoretical framework of discourse analysis and pragmatics.

In the analysis, the primary focus is on the Court's reliance on linguistic categories – such as epithets, metaphors, and covert or overt language – and how linguistic analysis informs judicial decision-making. The results reveal the nuanced approach the ECtHR takes, highlighting the interplay between language use and contextual factors, including the speaker's intent, the audience, and the societal impact of the expression. The argumentation used by the Court is of general nature and tends to avoid technical terminology, but it clearly addresses all the three levels of analysis, micro, meso, and macro, and pays special attention to discriminatory social practices (meso-level).

Keywords: hate speech, the European Court of Human Rights (ECtHR), case-law, linguistic criteria

References

Fino, A. (2020). Defining hate speech: A seemingly elusive task. *Journal of International Criminal Justice*, 18(1), 31–57. <https://doi.org/10.1093/jicj/mqaa023>

Gorenc, N. (2022). Hate speech or free speech: an ethical dilemma? *International Review of Sociology*, 32(3), 413–425. <https://doi.org/10.1080/03906701.2022.2133406>

HUDOC, <http://hudoc.echr.coe.int/eng>

Nikitina, V. (2023). *Human rights discourse: Linguistics, genre, and translation at the European Court of Human Rights*. Routledge. <https://doi.org/10.4324/9781003434788>

Peruzzo, K. (2019). National law in supranational case-law: A linguistic analysis of European Court of Human Rights judgments in English. *Journal of Language and Law*, 8(2), 1–24. <http://hdl.handle.net/10077/28071>

SARA-SOPHIE SCHEDEL (PHILLIPS-UNIVERSITÄT MARBURG), GEA DE JONG-LENDLE (PHILIPPS-UNIVERSITÄT MARBURG), MATHIAS SCHARINGER (PHILIPPS-UNIVERSITÄT MARBURG)

WOULD YOU RECOGNISE A FAKE VOICE? - A PERCEPTUAL STUDY ON THE RECOGNITION ABILITY OF NATURAL AND SYNTHETIC SPEECH

The rapid advancements in AI-driven speech synthesis present an increasing challenge to human detection. Particularly in criminal contexts, such as the “grandparent scam”, highly realistic synthetic voices are used to gain trust and deceive victims. The German Federal Network Agency registered over 11.000 complaints related to this scam in only the first half of 2024. The increasing accessibility of affordable speech-software increases the risk of voicebased deception, especially affecting vulnerable populations like elderly people or children (e.g. sexual online grooming).

This study examines listeners' ability to distinguish natural from synthetic voices. The experiment was designed to imitate the characteristics of a grandparent scam. Two groups of listeners were tested: familiar vs. unfamiliar (control group). Stimuli were generated using ElevenLabs and Speechify: two text-to-speech systems that are affordable and do not require high-level programming skills. Four stimuli conditions were compared: studio-quality vs telephone vs noise vs telephone&noise.

Analysis of the results shows that, as expected, listeners familiar to the voice consistently outperformed the non-familiar control group with overall recognition rates of 67% vs. 62% (significant). The fake voices produced by ElevenLabs were significantly more challenging for the listeners compared to those produced by Speechify – recognition rates were 51% and 73% respectively. Comparing stimuli quality: significant differences were found for the natural voices (unaltered 76%, background noise 49%, telephone filter 62%, and noise&telephone 70%). For synthetic voices rates rates were: 63%, 61%, 61%, and 61% respectively (nonsignificant).

Another notable finding was that listeners aged 70+ performed significantly worse than younger listeners, achieving an overall recognition rate of 56% vs. 71%.

This study shows that current developments in the software industry have resulted in highquality spoof-voices, produced by affordable speech programs, that are hard to distinguish from real voices. Detection rates may be even lower, when the victim is elderly and under stress.

Keywords: grandparent scam; voice spoofing; AI voices; speech synthesis; Text-to-Speech (TTS).

References

Bundeskriminalamt. 2020. Einzeltrick: Betrug am Telefon mit neuer Masche.

https://www.bka.de/SharedDocs/Kurzmeldungen/DE/Warnhinweise/200317_CoronaEinzeltrick.html. (27 October, 2024.)

Bundesnetzagentur. 2024. Beschwerden zu Rufnummernmissbrauch auch 2023 auf hohem Niveau – Schwerpunkt bei Einzeltrick-Bekämpfung.

https://www.bundesnetzagentur.de/SharedDocs/Pressemitteilungen/DE/2024/20240115_Rufnrmissb.html. (26 October, 2024.)

ZEIT ONLINE GmbH. 2024. Bundesnetzagentur verzeichnet hohe Beschwerdezahl wegen Telefonbetrugs. <https://www.zeit.de/digital/2024-07/telefon-betrug-sms-chatnachrichtbundesnetzagentur-kriminalitaet>. (26 October, 2024.)

ElevenLabs. 2025. Create the most realistic speech with our AI audio platform. <https://elevenlabs.io>. (8 January, 2025.)

Speechify. 2024. <https://myvoice.speechify.com/>. (8 January, 2025.)

The promise is one of the ways in which corruption crimes are carried out. At the moment of uttering the promise, the speaker commits him/herself to performing an action that is not materialized at the moment of speech, but which belongs to the realm of future possible/probable actions. For a commissive act to succeed, certain conditions must involve both the interlocutors and the propositional content. Evidence from criminal trials indicates that promises can be expressed directly, through performative verbs, or indirectly, through diverse linguistic strategies.

This study explores the linguistic markers of promise in corruption offences, focusing on how promises are crafted and interpreted within legal frameworks, particularly in bribery and influence-peddling cases under Romanian law. The research problem addresses the need to better understand the linguistic strategies by which promises are conveyed, particularly in intercepted conversations used as legal evidence.

Existing research on promises has primarily focused on their linguistic and pragmatic dimensions in general contexts. However, there is a gap in understanding how these acts are realized in corruption offences, where promises often take indirect forms and are embedded in complex communicative settings. Therefore this study tries to answer the following questions:

1. How are promises linguistically expressed in corruption offences?
2. What strategies are used to convey promises directly or indirectly?
3. How does Romanian legislation shape the interpretation of these promises?

The methods include a pragmatic analysis of intercepted conversations from bribery cases in Romania, identifying linguistic markers of promise such as performative verbs and indirect strategies. The findings reveal that promises in corruption offences are rarely explicit (e.g. "I promise") and more often conveyed through indirect means, such as conditional statements. Context plays a critical role in interpreting these promises, as does the legal framework in evaluating their acceptance or rejection. This research enhances understanding of corruption-related speech acts, providing valuable insights for linguists and legal practitioners.

Keywords: corruption offences; linguistic markers; promises; speech acts

References

- Bărbuță, Ion (2017), *Verbele performative în limba română*, Buletin de Lingvistică 18, 22–34.
- GALR = Valeria Guțu Romalo (ed.) (2008), *Gramatica limbii române*, București, Editura Academiei.
- Ionescu-Ruxăndoiu, Liliana (2003), *Limbar și comunicare. Elemente de pragmatică lingvistică*, București, All.
- Pérez-Hernández, Lorena (2001), *Illocution and Cognition: A Constructional Approach*, La Rioja, Universidad de la Rioja, Servicio de Publicaciones.
- Solan, Lawrence/Tiersma, Peter Meijes (edd.) (2012), *The Oxford Handbook of Language and Law*, Oxford, Oxford University Press.

MICHAŁ SZCZYSZEK (UAM POZNAŃ), EWA BADYDA (UNIwersYTET GDAŃSKI), PAWEŁ NOWAK (UMCS), JAKUB KĘPIŃSKI (UAM), LESZEK SZYMAŃSKI (UNIwersYTET ZIELONOGÓRSKI)

PLAGIARISM – A LINGUISTIC AND LEGAL PERSPECTIVE. POLISH COURT CASES WITH EXPERT LINGUISTS' OPINIONS

Research problem: due to increasing public awareness of the existence of the legal protection of intellectual property, more and more plagiarism cases are being brought before Polish courts. The paper examines the phenomenon of the so-called linguistic plagiarism from three intersecting perspectives: the common view of its essence, the regulations of this issue in the Polish legal system, and the authors' own experience resulting from their function as court experts appointed to present linguistic opinions on plagiarism cases. The common and linguistic views of plagiarism – as reflecting its potential understanding by the plaintiff – are reconstructed from lexicographic and corpus data, while the legal understanding stems from legal definitions of the copyright and related rights (as well as the jurisprudence) in Poland. We assume a **hypothesis** about a far-reaching divergence of these (i.e. common, linguistic and legal) perspectives, and we **pose the question** of the fields and breadth of these divergences and perspectives. We attempt to answer this question with a legal and linguistic analysis of authentic court (both civil and criminal) cases. Discussing them, we indicate the text types that are subject to judicial evaluation. Thus, we use, on the one hand, **linguistic methodology** from different areas of linguistics – from (ortho)graphical analyses, through morpho-syntactic analyses, to semantic and pragmatic analyses (in line with the procedures for court experts' opinions), but, on the other hand, we invoke **jurisprudent methodology** from the field of the copyright law. Our **aim** is to characterize the perspective from which a linguistics court expert's opinion is most often compiled in these cases, and to recall a typical range of court's questions to the expert, which allows us to reconstruct an outline of the issues that are of particular interest to the court in these cases. Finally, as a **summary**, we present analyses of court verdicts and their justifications in the cited cases, showing a correlation between legal (judicial) and linguistic processing. Another **outcome** of juxtaposing the perspectives of: plaintiffs' claims, the overtone of experts' opinions, and the justifications of verdicts in plagiarism cases that enable us to assess the convergence or divergence of common, linguistic and legal views of plagiarism is an establishment of a field of interest for both lawyers and linguists concerned with plagiarism, and thus to institute a *de facto* new area of research. **The results of our findings** will constitute a certain novelty in the Polish state of research on the phenomenon of plagiarism and court procedures involving court experts in plagiarism cases, which, in our opinion, will contribute to the constitution of forensic linguistics in Poland as a sub-discipline of applied linguistics.

Keywords: plagiarism in Polish copyright law, linguistic analysis of plagiarism, the role of the linguist as a court experts, forensic linguistics in Poland

References

- Daśko N. (2020). *Karnoprawne i kryminologiczne aspekty przestępstwa plagiatu*, „Forum Prawnicze”, Nr 3.
- Feluś, A. (2011). *Z zagadnień językoznawstwa kryminalistycznego*. Poznań: Wyższa Szkoła Pedagogiki i Administracji im. Mieszka I.
- Hołyst, B. (2000), *Możliwości i granice kryminalistycznej ekspertyzy lingwistycznej*. W: B. Hołyst *Kryminalistyka*. Warszawa: Wydawnictwo LEXISNEXIS, s. 120-135.

Kępiński J. (2017). *Ochrona prawnoautorska dla wzorów przemysłowych. Glosa krytyczna do wyroku Sądu Najwyższego z 6 marca 2014 roku, V CSK 202/13*, „Acta Iuris Stetinensis (Roczniki Prawnicze)”, vol. 18, nr 2.

Lingwistyka kryminalistyczna. Teoria i praktyka. (2019). red. Monika Zaśko-Zielińska i Krzysztof Kredens, ss. 384-421, Wrocław: Uniwersytet Wrocławski.

Pankalla, O. (2019). *Rola biegłych językoznawców w polskim procesie karnym* (niepublikowana praca magisterska). Poznań.

Sołtysiak G. (2009). *Plagiat. Zarys problemu*. Warszawa.

Szczyszek M. (2021). *Plagiat jako przedmiot analizy biegłego językoznawcy*, „Kwartalnik Stowarzyszenia Sędziów Polskich "IUSTITIA"”, nr 4 (46), s. 212-216.

PANEL

THE ROLE OF EXPERTS AND THE COURT: THEORY AND PRACTICE

Coordinators:

Eilika Fobbe (Federal Criminal Police Office (BKA)) & Gea de Jong-Lendle (Philipps-University Marburg)

Participants: Timothy Habick (Reasoning, Inc.), Jūratė Ruzaitė (Vytautas Magnus University), and Victoria Guillén Nieto (University of Alicante)

Theoretically, the expert and judicial representatives' roles seem straightforward. The primary function of an expert witness (sometimes known as a forensic expert) is to assist the courts in adequately interpreting case-relevant information by giving an expert opinion in their area of expertise. This opinion should be offered independently and impartially. An expert witness can be anyone with knowledge or experience of a particular field or discipline beyond that to be expected of a layman.

The role of the court is to decide controversies. In case the court decides that additional information is required to carry out that task, an expert witness may be instructed to conduct an analysis and write a report. In the context of forensic linguistics and phonetics, it is language—both written and spoken—that forms the basis for the analysis.

So far so good?

There are a number of practical and methodological problems that can arise. Of these, we will focus on the following three areas: 1) the instructions formulated by the court, 2) the expectations of the client and 3) the interpretation of the reported results.

First, the instructions accompanying the expert report assignment may be unclear, which can result in misunderstandings about the assignment's objectives. How should the linguist navigate this uncertainty without exceeding the boundaries of their expertise or role? Additionally, these assignments may reflect the potentially misguided expectations of clients, which may concern specific issues at hand and/or the general understanding of scientific principles. How can the linguist effectively address these challenges?

Finally, the linguistic interpretation of the reported results must align with their forensic value and legal evaluation. This alignment can present pitfalls for both linguists and legal professionals. What contributions can linguistics offer in tackling these concerns?

While many of these issues are well-known, they continue to pose significant challenges, and their relevance to the work of linguists in forensic settings remains critical.

This panel aims to discuss some of the problems relevant to expert witness work, identify possible causes, and propose practical solutions. Examples will be used to illustrate the various problems and to place them in a broader conceptual framework.

We welcome participation from both experienced expert linguists and those who are new to the field.

Keywords: expert role; forensic linguistics; forensic phonetics; interpretation of evidence; court instructions