

*International Journal of Language and Legal Studies (IJLLS)**Available Online:*<https://ijlls.online/index.php/journal/index>

ISSN Print: 3105-7950

ISSN Online: 3105-7969

Platform & Workflow by: [Open Journal Systems](#)

From Confessions to Computational Stylometry: A Comprehensive Review of Forensic Linguistics in the 21st Century**Shahab Rehman**

Lecturer in English, Kohat University of Science & Technology KUST

shahab.rehman@kust.edu.pk**Abstract**

The relationship between applied linguistics and law has become a fascinating and diverse area of study that incorporates both theory and practise on the practical application of law and real-life practise in law into the study. This has changed over the years, as, whilst this field has been involved in individual facets of the law language, it has come to be a well-established interdisciplinary discipline termed forensic linguistics. Currently, it incorporates the knowledge of sociolinguistics, discourse analysis, psycholinguistics, and even computational approaches to the study of legal texts, verbal interactions and institutional communication. The review also provides an in-depth review of the current body of knowledge on this subject. It discusses the historical foundations of the field, the theoretical basis of the field, the methodological developments of the field, and the practical issues of the field. We follow the path of the development of forensic linguistic techniques beginning with the early confessions and trademark cases to the contemporary approaches such as corpus approach and computer programmes of authorship determination and legal translation. We also examine the intricacies of court room conversances, the police interrogation, and the barriers to communication experienced by the nonnatives in courtroom situations. In the process, we resolve on the current disputes on the trustworthiness and admissibility of linguistic evidence in the courtroom and emphasise on the importance of interdisciplinary partnership to ensure that legal procedures are more equitable and acceptable. Last, but not least, the report identifies growing trends like quantitative tool usage (e.g. corpus linguistics, stylometry), the development of expert witness procedures, and the increasing effect of globalisation and multilingualism on the language of law. Combining the main results of the powerful research and recent developments, this review will not only trace the existing trends in the research but will identify future possibilities in developing the theoretical progress into the practical revision of the law.

Keywords: *Applied Linguistics, Forensic Linguistics, Legal Discourse, Courtroom Interaction, Corpus Linguistics, Authorship Attribution, Multilingual Legal Communication.*

Introduction

The field of research between applied linguistics and law has experienced a radical change in the last couple of decades. What started as separated, descriptive studies of legal wording has developed to become a multidisciplinary, all-inclusive domain. Initial research mainly focused on the dogmatism of legal texts in terms of style and syntax. As time passed, though, this came to encompass a wider spectrum of matters concerning the evidential aspect of linguistic markers in confessions, to the subtle nature of courtroom discourse and the creation of storeys at court (Coulthard et al., 2016). Modern scholarship extends beyond description

of language of law to the critical analysis of its practical effects: how the choices of language influence the result of judicial reviews, how the participants become empowered or marginalised, and how the language is used to achieve the outcome of legal authority. This change is a manifestation of the increasingly widespread appreciation of the fact that law is at all a language system, and that to achieve a complete comprehension of legal discourse, we need to not only conduct rigorous textual and interactional analysis, but also to appreciate the socio-cultural settings within which law is contained. As a result, the modern academic developments have shifted towards methodologies being a combination of traditional linguistic analysis with empirical and computational methodologies that enable the provision of quantifiable and reproducible evidence in the court of law (Coulthard et al., 2016; Fitria, 2024). This review provides an overview of the research that is being conducted by starting with tracing the historical development of the research, then explaining the major theoretical and methodological advances, and finally defining the major themes, uses, problems, and perspectives of the field of studying language and law.

Scope & Purpose of the study

This paper will critically discuss the historical development of applied linguistics in the legal field as it has come to be known today as forensic linguistics, its historical development, theoretical premises and methodological improvements. It argues to assess fundamental areas of research and effective issues, where the discussion about the reliability and admissibility of linguistic evidence is a major concern. It also determines the new trends and suggests future directions of incorporating linguistic innovation when it comes to legal reform.

Methodology

This study adopts a qualitative, literature-based review methodology to critically examine the evolution and current landscape of applied linguistics in legal contexts, evolved into as forensic linguistics. The approach involves:

- **Historical Tracing:** Mapping the chronological development of the field, highlighting key milestones and paradigm shifts.
- **Theoretical Integration:** Synthesizing frameworks from sociolinguistics, discourse analysis, pragmatics, and computational linguistics to contextualize forensic linguistic inquiry.
- **Citation-Based Analysis:** Engaging with seminal and contemporary scholarship (e.g., Coulthard et al., Fitria, Schmitt & Celce-Murcia) to identify influential contributions and evolving research trajectories.
- **Descriptive and Analytical Review:** Providing a structured overview of historical, theoretical, and methodological developments, with attention to interdisciplinary intersections.
- **Thematic Synthesis:** Organizing findings around core research domains such as courtroom discourse, authorship attribution, multilingual legal settings, and legal translation.
- **Application and Challenge Mapping:** Discussing practical implementations of forensic linguistic methods and the challenges surrounding reliability, admissibility, and ethical considerations in legal contexts.

This methodology enables a comprehensive and critical synthesis of the field, bridging foundational scholarship with emerging directions in legal reform and linguistic innovation.

Databases and Sources Consulted

To identify and screen relevant studies, the following databases and sources were consulted: Google Scholar, Scopus, Web of Science, and publisher platforms including Routledge, Wiley, Springer, Taylor & Francis, Cambridge Core, and Multilingual Matters. Additionally, specialized journals and archives such as the International Journal for the Semiotics of Law, Pragmatics and Society, German Law Journal, PRASASTI Journal of Linguistics, LEX RUSSICA, American Anthropologist, Policy Journal of Social Science Review, Michigan Law Review Archive, IJFL Journal, and the Routledge Handbook of Forensic Linguistics were reviewed to ensure comprehensive coverage of domain-specific and interdisciplinary research.

Selection Criteria & Search Strategy

Searches were conducted in Google Scholar, Scopus, and Web of Science, complemented by targeted queries on major academic publishers (Routledge, Wiley, Springer). Search strings combined controlled terms and free text, including: *“forensic linguistics,” “legal discourse,” “authorship attribution,” “corpus linguistics” AND legal, “legal writing” AND readability, “comparative legal linguistics,” “experimental linguistics” AND (law OR treaty interpretation)*. The time window was open (1960s–2025) to capture both foundational and contemporary developments (e.g., Mellinkoff, 1963; Coulthard, Johnson & Wright, 2016; Goźdz-Roszkowski, 2021; Pirker & Skoczeń, 2022; Mattila, 2024; Fitria, 2024).

Inclusion criteria were: (C1) direct relevance to the language–law interface (forensic/applied/legal linguistics); (C2) scholarly credibility (peer-reviewed venues or established academic presses); (C3) contribution to historical foundations and/or recent advances; (C4) methodological breadth (e.g., discourse/pragmatics, corpus/quantitative, experimental, survey); (C5) conceptual scaffolding (handbooks/overviews); (C6) comparative or multilingual perspective; (C7) practical/policy relevance (e.g., courtroom practice, drafting/readability, treaty interpretation); and (C8) access/traceability (DOI or stable link).

Exclusion criteria removed non-scholarly items, works lacking a legal-linguistic focus, purely legal analyses without linguistic method, and redundant items with inferior clarity/scope.

Titles/abstracts were screened for topical fit; full texts were then assessed against C1–C8. Disagreements were resolved by revisiting the criteria and privileging sources that best triangulate theory, method, and application. The final set comprised 15 sources encompassing seminal foundations, methodological exemplars, comparative perspectives, and recent empirical or review work (e.g., Shuy, 2015; Stygall, 2010; Ruiz, 2016; Umiyati, 2020; Khan Maitlo et al., 2025).

Criteria legend:

- C1 Topical fit (language–law)
- C2 Scholarly credibility
- C3 Temporal value (foundation/recent)
- C4 Methodological breadth
- C5 Conceptual scaffolding
- C6 Comparative/multilingual
- C7 Practice/policy relevance
- C8 Access/traceability

Criterion	Sources that satisfy the criterion
C1 Topical fit	Mellinkoff (1963); Hirsch (2000); Stygall (2010); Coulthard (2011); Shuy (2015); Galyashina (2016); Coulthard et al. (2016); Ruiz (2016); Schmitt & Celce-Murcia (2019); Umiyati (2020); Goźdz-Roszkowski (2021); Pirker & Skoczeń (2022); Mattila (2024); Fitria (2024); Khan Maitlo et al. (2025)
C2 Scholarly credibility	All 15 (peer-reviewed journals/handbooks; Routledge/Wiley/Springer presses; DOIs provided where applicable)
C3 Temporal value	Foundational: Mellinkoff (1963); Stygall (2010). Bridging/mature: Coulthard (2011); Shuy (2015); Galyashina (2016); Coulthard et al. (2016); Ruiz (2016); Schmitt & Celce-Murcia (2019). Recent: Umiyati (2020); Goźdz-Roszkowski (2021); Pirker & Skoczeń (2022); Mattila (2024); Fitria (2024); Khan Maitlo et al. (2025)
C4 Methodological breadth	Discourse/pragmatics: Coulthard (2011); Shuy (2015); Pirker & Skoczeń (2022). Corpus/quantitative: Goźdz-Roszkowski (2021). Readability/legal writing: Stygall (2010). Comparative/translation: Mattila (2024). Survey/attitudes: Khan Maitlo et al. (2025). Field overviews/lit reviews: Umiyati (2020); Fitria (2024); Schmitt & Celce-Murcia (2019); Coulthard et al. (2016)
C5 Conceptual scaffolding	Coulthard et al. (2016); Schmitt & Celce-Murcia (2019); Shuy (2015); Stygall (2010); Umiyati (2020); Fitria (2024)
C6 Comparative/multilingual	Galyashina (2016); Mattila (2024); Pirker & Skoczeń (2022); Ruiz (2016)
C7 Practice/policy relevance	Courtroom/casework/practice: Coulthard (2011); Galyashina (2016); Coulthard et al. (2016). Treaty interpretation/decision-making: Pirker & Skoczeń (2022). Drafting/readability: Stygall (2010). Language planning impacting legal contexts: Ruiz (2016). Professional training/awareness: Khan Maitlo et al. (2025). Comparative implications for access to justice: Mattila (2024)
C8 Access/traceability	All items provide DOIs or stable links in your bibliography (journal/book/publisher records)

Source-Centric Compliance Matrix

Source	C1	C2	C3	C4	C5	C6	C7	C8
Mellinkoff (1963)	✓	✓	✓					✓
Hirsch (2000)	✓	✓	✓	✓			✓	✓
Stygall (2010)	✓	✓	✓	✓	✓		✓	✓
Coulthard (2011)	✓	✓	✓	✓			✓	✓

Shuy (2015)	✓	✓	✓	✓	✓		✓	✓
Galyashina (2016)	✓	✓	✓	✓		✓	✓	✓
Coulthard et al. (2016)	✓	✓	✓	✓	✓		✓	✓
Ruiz (2016)	✓	✓	✓			✓	✓	✓
Schmitt & Celce-Murcia (2019)	✓	✓	✓	✓	✓		✓	✓
Umiyati (2020)	✓	✓	✓	✓	✓			✓
Goźdz-Roszkowski (2021)	✓	✓	✓	✓			✓	✓
Pirker & Skoczeń (2022)	✓	✓	✓	✓		✓	✓	✓
Mattila (2024)	✓	✓	✓	✓		✓	✓	✓
Fitria (2024)	✓	✓	✓	✓	✓	✓	✓	✓
Khan Maitlo et al. (2025)	✓	✓	✓	✓			✓	✓

Number of Studies Scrutinized

Records identified through database and other sources (n = 150)
Records after deduplication and screened by title/abstract (n = 50)
Full-text articles assessed for eligibility (n = 25)
Full-text articles excluded, with reasons (n = 10)
Studies included in qualitative synthesis (n = 15)

Included Studies and relevant Sources

Study	Status	Database/Source
Coulthard, M. (2011). Making a difference: Critical linguistic analysis in a legal context.	Included	Google Scholar, Scopus, Pragmatics and Society (publisher: John Benjamins)
Coulthard, M., Johnson, A., & Wright, D. (2016). An Introduction to Forensic Linguistics: Language in Evidence.	Included	Routledge (Taylor & Francis), Google Scholar
Fitria, T. N. (2024). Forensic Linguistics: Contribution of Linguistics in Legal Context.	Included	Google Scholar, PRASASTI Journal of Linguistics

Galyashina, E. I., & Галяшина, Е. И. (2016). Forensic Linguistics In Legal Proceedings.	Included	Google Scholar, LEX RUSSICA (Russian Law Journal)
Goźdź-Roszkowski, S. (2021). Corpus Linguistics in Legal Discourse.	Included	Springer, Scopus, International Journal for the Semiotics of Law
Hirsch, S. F. (2000). Just Words: Law, Language, and Power.	Included	Google Scholar, American Anthropologist (Wiley)
Khan Maitlo, S., Ahmed Abbasi, I., Ali Jatoi, Z., & Ahmad, A. (2025). Quantifying Awareness And Attitudes Towards Forensic Linguistics.	Included	Google Scholar, Policy Journal of Social Science Review
Mattila, H. E. S. (2024). Comparative Legal Linguistics: Language of Law, Latin and Modern Lingua Francas.	Included	Routledge (Taylor & Francis), Google Scholar
Mellinkoff, D. (1963). Mellinkoff: The Language of the Law.	Included	Google Scholar, Michigan Law Review Archive
Pirker, B., & Skoczeń, I. (2022). Pragmatic Inferences and Moral Factors in Treaty Interpretation.	Included	Cambridge Core, German Law Journal
Ruiz, R. (2016). Orientations in language planning.	Included	Multilingual Matters (Publisher), Google Scholar
Schmitt, N., & Celce-Murcia, M. (2019). An overview of applied linguistics.	Included	Routledge (Taylor & Francis), Google Scholar
Shuy, R. W. (2015). Discourse Analysis in the Legal Context.	Included	Wiley Online Library
Stygall, G. (2010). Legal writing: Complexity: Complex documents/average and not-so-average readers.	Included	Routledge Handbook of Forensic Linguistics
Umiyati, M. (2020). A Literature Review of Forensic Linguistics.	Included	Google Scholar, IJFL Journal

Historical Development and Emerging Directions in Forensic Linguistics

Language and law The archaeology of law Scholarly analysis of legal text as an aesthetic and structural artefact, however, began with Mellinkoff (1963) in his seminal study *The Language of the Law*, which emphasised the archaism and formalism of legal language and its impossibility to ordinary comprehension (Hirsch, 2000). In the 1970s the discipline expanded exponentially as sociolinguists, anthropologists, and sociologists came to realise that legal authority is not exercised in the form of written texts, but through oral activities and regulation by the everyday interaction of people with each other, including interviews with police, courtroom dialogue and informal negotiations (Coulthard et al., 2016; Hirsch, 2000). Preliminary forensic analysis of controversial confessions, trademark cases, and inconsistencies in police interrogations indicated that linguistic anomalies can be trusted as the reliable tool in detecting coercion or deception and set the foundation of forensic linguistics as the practise-oriented field (Coulthard et al., 2016; Stygall, 2010). The discipline underwent institutionalisation of expert testimony, the establishment of professional associations, and the implementation of strict mixed methods in the following decades and

turned individual theoretical investigations into a respectable source of legal judgments (Fitria, 2024; Pirker & Skoczen, 2022).

Scholarly literature today is based on a wide range of theoretical approaches that do not view legal language as a conduit but rather as an effective tool in terms of creating and maintaining social orders. The Foucaultian discourse analysis has helped to understand institutional power in the legal context, and the sociolinguistic frameworks have demonstrated how discourse production and reception are influenced by gender, race, ethnicity, and class (Hirsch, 2000; Fitria, 2024). Theories of pragmatics and the speech act have helped to understand how performative utterances such as police cautions, oaths and verdicts have a legal power that does not always correspond with the understanding of the lay people (Fitria, 2024; Galyashina and Galiashina, 2016). In terms of methodology, the research has evolved through the research of small-scale qualitative research to the large-scale mixed-method designs that combine corpus linguistics, stylometry, conversation analysis, ethnographic observation, and forensic phonetics (Gozdz-Roszkowski, 2021; Schmitt and Celce-Murcia, 2019; Umiyati, 2020). The current existence of computational advances facilitates pattern detection in large scale by statistical measures and machine learning, and multimodal analysis includes gaze, gesture, and prosody to give more detailed accounts of courtroom interaction (Pirker and Skoczen, 2022). These advances have largely empowered empirical rigour and evidential credibility, and these issues have long been issues of judicial concern in terms of reliability and replicability.

Among the major areas of research include criminal investigations, authorship attribution, courtroom discourse, plain-language reform and multilingual justice. Analysis of confessions and police interviews is still used to reveal coercion and vulnerability (Coulthard et al., 2016), and the authorship attribution issue with the help of the idiolectal features of the signature helps to solve ransom note and online threatening problems (Fitria, 2024; Gozdz-Roszkowski, 2021a). Courtroom discourse studies show that there are consistent power imbalances and disjunctive narratives at the disadvantage of witnesses (Schmitt and Celce-Murcia, 2019), and that the literature undergoing cross-cultural and multilingual research provides evidence of dangerous misinterpretation when dealing with non-native speakers and minority-language groups (Fitria, 2024; Shuy, 2015; Mattila, 2024). In spite of these achievements, there are still problems: not all evidence is admissible in a court of law when it comes to the use of evidence using the Daubert-style standard, methodological standardisation has not been fully achieved, and not all original recordings can be accessed to analyse them (Coulthard, 2011; Pirker and Skoczen, 2022). The hostile courtroom atmosphere tends to diminish linguists to clarification duties, not conclusive ones, whereas political forces commonly damage language rights frames (Khan Maitlo et al., 2025).

In the future, the domain is poised at the edge of revolutionary development characterised by increased levels of computationalization, larger domain specific corpora, and multimodal digital analysis. Machine learning has already been suggested to be more precise in authorship verification, detecting deception, and recognising discourse patterns, and there are efforts to improve legibility and persuasiveness, such as the Plain Language Movement and new form of expert testimony (e.g., concurrent "hot tub" evidence) (Schmitt and Celce-Murcia, 2019). Further studies on the language as a communicative resource and as a tool of power in a globalised and technologically disrupted world will be explored through comparative legal linguistics and studies of new online genres, such as online hate speech, encrypted communication, AI-generated texts, and others (Gozdz-Roszkowski, 2021; Mattila, 2024). Cross-disciplinary cooperation, uniformity, and judicial training are still required to

implement these innovations into fair judicial practise, where linguistic mastery will still be useful to the administration of justice and not only detail the intricacies (Coulthard et al., 2016; Khan Maitlo et al., 2025).

Conclusion and Outlook

The intersection of applied linguistics and law has grown into a rich interdisciplinary discipline that has gone much further than what traditional descriptions of legal language entail. Since its inception in specialising in individual phenomena include controversial confessions and trademark wars, forensic linguistics has expanded to implementing advanced methods including large-scale corpora, computational stylometry, machine learning, and multimodal discourse analysis to reveal unanticipated power dynamics within police interviews, trial interactions, and multilingual legal proceedings. This development has been to make linguistic evidence not a sidelining interest, but a tool that cannot be ignored in the pursuit of protecting fairness, exposing coercion and maintaining equal access to justice. Through exposing the ways in which language creates authority, credibility and culpability, the field has strongly proved to be applicable in the real world, and has shaped the outcome of cases, influenced policy changes, and improved learners to understand legal processes in various jurisdictions.

Nevertheless there are still significant obstacles. The admissibility and perceived accuracy of linguistic expert testimony remains a matter of wide variation, undermined by the lack of methodological standardisation, judicial educative preparation in linguistic principles and continued misgivings about transparency in the data and analysis methods. The gap between academia and practise in the court requires long-term interdisciplinary cooperation, better guidelines on using expert evidence, and the willingness of the institutions to include linguistic expertise in the process of legal education and adjudication. In the future, the swift development of computational forensics, the growth of multilinguality and multimodality corpus, and the increasing interest in plain-language initiatives is likely to bring unprecedented accuracy and inclusiveness. With the law systems struggling with globalisation, electronic evidence, and growing linguistic pluralism, applied linguistics will be even more central in facilitating clarity, responsibility, and fairness that language is a tool to help, rather than hinder, fair legal justice in the ever-evolving complex global environment.

References

- Coulthard, M. (2011). Making a difference: Critical linguistic analysis in a legal context. *Pragmatics and Society*, 2(2), 171–186. <https://doi.org/10.1075/PS.2.2.03COU>
- Coulthard, M., Johnson, A., & Wright, D. (2016). *An Introduction to Forensic Linguistics: Language in Evidence* (2nd Ed). Routledge. <https://www.routledge.com/An-Introduction-to-Forensic-Linguistics-Language-in-Evidence/Coulthard-Johnson-Wright/p/book/9781138641716>
- Fitria, T. N. (2024). Forensic Linguistics: Contribution of Linguistics in Legal Context. *PRASASTI: Journal of Linguistics*, 9(1), 118–134. <https://doi.org/10.20961/PRASASTI.V9I1.71527>
- Galyashina, E. I., & Галяшина, Е. И. (2016). Forensic Linguistics In Legal Proceedings. *LEX RUSSICA (РУССКИЙ ЗАКОН)*, 12(1), 136–145. <https://doi.org/10.17803/1729-5920.2016.118.9.136-145>
- Goźdz-Roszkowski, S. (2021). Corpus Linguistics in Legal Discourse. *International Journal for the Semiotics of Law*, 34(5), 1515–1540. <https://doi.org/10.1007/S11196-021-09860-8/METRICS>
- Hirsch, S. F. (2000). Just Words: Law, Language, and Power. *American Anthropologist*, 102(1), 169–170. <https://doi.org/10.1525/AA.2000.102.1.169>

- Khan Maitlo, S., Ahmed Abbasi, I., Ali Jatoy, Z., & Ahmad, A. (2025). *Policy Journal of Social Science Review Quantifying Awareness And Attitudes Towards Forensic Linguistics: A Survey Of Young Legal Professionals*. 3(2). <https://journalofsocialsciencereview.com/index.php/>
- Mattila, H. E. S. (2024). Comparative Legal Linguistics : Language of Law, Latin and Modern Lingua Francas. In *Comparative Legal Linguistics*. Routledge.
<https://doi.org/10.4324/9781003578017>
- Mellinkoff, D. (1963). Mellinkoff: The Language of the Law. In *Michigan Law Review* (1st ed., Vol. 63). Little Brown.
<https://repository.law.umich.edu/mlr><https://repository.law.umich.edu/mlr/vol63/iss1/11>
- Pirker, B., & Skoczeń, I. (2022). Pragmatic Inferences and Moral Factors in Treaty Interpretation—Applying Experimental Linguistics to International Law. *German Law Journal*, 23(3), 314–332. <https://doi.org/10.1017/GLJ.2022.22>
- Ruiz, R. (2016). Orientations in language planning. *Honoring Richard Ruiz and His Work on Language Planning and Bilingual Education*, 13–32.
<https://doi.org/10.21832/9781783096701-004/HTML>
- Schmitt, N., & Celce-Murcia, M. (2019). An overview of applied linguistics. In *An Introduction to Applied Linguistics* (pp. 1–16). Routledge. <https://doi.org/10.4324/9780429424465-1>
- Shuy, R. W. (2015). Discourse Analysis in the Legal Context. In *The Handbook of Discourse Analysis* (pp. 822–840). John Wiley & Sons, Ltd.
<https://doi.org/10.1002/9781118584194.CH38>
- Stygall, G. (2010). Legal writing: Complexity: Complex documents/average and not-so-average readers. In *The Routledge Handbook of Forensic Linguistics* (pp. 51–64). Taylor and Francis.
<https://doi.org/10.4324/9780203855607-14/LEGAL-WRITING-COMPLEXITY>
COMPLEX-DOCUMENTS-AVERAGE-AVERAGE-READERS-GAIL-STYGALL
- Umiyati, M. (2020). A Literature Review of Forensic Linguistics. *IJFL (International Journal of Forensic Linguisti)*, 1(1), 23. <https://doi.org/10.22225/.2.1.1603.1-6>