ALLAHABAD ACADEMIC PRESS
a publishing company

The Research Frontline - Journal

A Peer-Reviewed Quarterly Journal of Interdisciplinary Inquiry and Research

Journal homepage: https://trfjournal.cdfaindia.org/index.php/trfjournal/en/index

Deception as an Institutional Discourse Practice: A Linguistic–Computational Model from Indian Courtrooms

Ayantik Khan Anit Kumar Ghosh

To cite this article: Khan, A., Ghosh, A.K. (2026): Deception as an Institutional Discourse Practice: A Linguistic–Computational Model from Indian Courtrooms, The Research Frontline - Journal (Vol 1, Issue 2, January - March 2026) DOI:https://doi.org/10.66871/trf-j.v1i2.007

To link to this article: https://doi.org/10.66871/trf-j.v1i2.007

© 2026 The Author(s). Published by Allahabad Academic Press,

Published online: 05 May 2026.

Submit your article to this journal

Vol. 1, Issue 2, January - March 2026

Full Terms & Conditions of access and use can be found at https://trfjournal.cdfaindia.org/index.php/trfjournal/en/Open-Access-Policy

The Research Frontline - Journal (Vol. 1, Issue 2, January - March 2026)
https://doi.org/10.66871/trf-j.v1i2.007

ALLAHABAD ACADEMIC PRESS
a publishing company

Deception as an Institutional Discourse Practice: A Linguistic–Computational Model from Indian Courtrooms

Ayantik Khan[1] Anit Kumar Ghosh[2]

[1]Research Assistant (Technical - 2024), DHSS, IIT Delhi, New Delhi [2]Department of Biotechnology, School of Biological Science, MAK Azad University of Technology, West Bengal. India

ABSTRACT

Conventional research on phoney speech in legal circumstances typically views deception as a moral or cognitive defect at the individual level that can be recognised by consistent linguistic indicators. By suggesting that deception is an institutionally created discourse practice that results from the interplay of cognitive load, procedural limitations, and power imbalances in courtroom communication, this research proposes a theoretical reorientation. The study uses a mixed-methods strategy that combines qualitative forensic analysis with natural language processing techniques applied to specific Indian criminal court rulings, drawing on forensic linguistics, discourse analysis, and computational language modelling. Patterns of strategic ambiguity, evasive coherence, emotional modulation, and pragmatic indeterminacy are found in the testimonies of witnesses and accused individuals. To track how institutional forces influence communication behaviour, computational methods such as sentiment trajectory mapping, stance identification, and lexical dispersion metrics are combined with careful language reading. The analysis shows that deceitful discourse in legal contexts functions more as an adaptive, situationally sensible tactic conditioned by juridical norms and interpretive authority than as a sign of personal dishonesty. This study offers a reusable analytical framework for forensic linguistics and legal discourse studies by modelling deception as a situated, procedural, and culturally mediated phenomenon. This framework has implications for judicial interpretation, evidentiary evaluation, and the moral use of computational tools in legal contexts.

ARTICLE HISTORY
Received: 27 January 2026
Accepted: 30 March 2026

KEYWORDS
Institutional Deception, Forensic Linguistics, Legal Discourse Analysis, Strategic Ambiguity, Computational Linguistics, Deception Detection

I. Introduction

One of the most organised institutional settings is the courtroom, where language serves as a strategic tool for building credibility, persuasion, and justification, in addition to being a means of communication. Participants in legal proceedings, including witnesses, defendants, attorneys, and judges, participate in intricate language exchanges where assertions are continuously assessed for veracity and accuracy. As a result, identifying fraudulent communication has become increasingly important in the study of forensic linguistics.

An explicit lie is rarely the only way that deception is represented in legal discourse. Rather, it frequently takes the form of subtle verbal cues, such as discourse markers, hesitation markers, expressions of doubt, negation, and lexical repetition. When speakers try to control their cognitive load, handle psychological pressure, or purposefully alter their comments during cross-examination and

CONTACT: Ayantik Khan ayantik.cstaff.iitd@gmail.com

© 2026 the author(s). Published by Allahabad Academic Press, India. This is an Open access article distributed under the terms of the Creative Commons attribution license (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the accepted Manuscript in a repository by the author(s) or with their consent.

examination, these patterns may show themselves. Therefore, it is possible to think of lying in courtroom speech as both an individual behaviour and an institutionally conditioned communicative practice that is influenced by legal processes, adversarial questioning, and the judiciary's evaluation of credibility.

Recent developments in natural language processing (NLP) and computational linguistics have made it possible for scholars to analyse vast amounts of legal documents and find systematic linguistic patterns linked to dishonesty, uncertainty, and cognitive stress. Scholars can investigate courtroom language in a data-driven way by using automated textual analysis techniques like sentiment analysis, lexical frequency measurement, stance identification, and discourse pattern discovery. When analysing lengthy court records, these methods are especially helpful because hand interpretation could miss tiny but recurrent linguistic cues.

However, there is a direct connection between lying in institutional discourse and cognitive functions like executive control, stress management, and cognitive load management. According to a psychological and behavioural study, people who communicate dishonestly frequently put forth more mental effort, which may affect their language choices and discourse organisation. Therefore, a more thorough knowledge of how misleading communication functions in formal legal systems can be obtained by combining viewpoints from linguistics, computational analysis, and cognitive research.

Relatively little research has examined Indian courtroom conversation, especially using publicly accessible judicial rulings as language corpora, despite the growing interest in computer techniques for fraud detection. Indian court rulings are useful tools for analysing institutional discourse patterns because they provide in-depth narrative reconstructions of witness testimony, cross-examinations, and judicial assessments of trustworthiness. Nevertheless, there are still a few comprehensive computational evaluations of these writings.

By creating a linguistic–computational model to analyse discourse patterns associated with fraud in Indian criminal court rulings, this study fills this gap. The study uses automated natural language processing techniques to find linguistic indications in courtroom narratives related to ambiguity, denial, emotional tone, and cognitive stress using a corpus of chosen criminal cases from Indian courts. The methodology also looks at high-risk discourse segments that exhibit higher linguistic markers related to deceit and investigates how these patterns interact with institutional context and speaker roles.
This study attempts to add to the expanding field of digital forensic linguistics by integrating techniques from computational text analysis, cognitive discourse studies, and forensic linguistics. The results may give a fresh viewpoint on the examination of dishonesty and credibility in institutional discourse by illuminating how linguistic patterns in judicial texts reflect the cognitive and communicative processes of courtroom actors.

II. Literature Review

2.1-Deception in Forensic Linguistics: Forensic linguistics and discourse analysis have traditionally placed a high value on the investigation of language deception. One of the earliest systematic discussions of linguistic deception was presented by Paul Ekman in *"Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage"* (1985), where the author looked at verbal and behavioural signs of dishonesty, such as emotional leaking, reluctance, and speech irregularities. Even while Ekman's research was mainly concerned with psychological behaviour, it set the stage for later language studies of deceptive communication.

Researchers in the field of forensic linguistics have looked at how subtle language patterns, as opposed to overt lies, might be used in deceptive speech. Malcolm Coulthard and Alison Johnson in *"An Introduction to Forensic Linguistics: Language in Evidence"* (2007) emphasised the significance of examining conversation patterns, linguistic structures, and practical decisions in legal contexts. Their

research highlighted how courtroom language is a highly structured type of institutional discourse where linguistic performance is used to negotiate legitimacy.

Similarly, Lawrence M. Solan and Peter M. Tiersma explored linguistic interpretation within legal systems in their book *"Speaking of Crime: The Language of Criminal Justice"* (2005). They contended that institutional norms that affect how judges and legal professionals frame, analyse, and assess testimony impact legal communication. This viewpoint emphasises how speech structure shapes judicial contexts' views of truth and dishonesty.

2.2-Computational Approaches to Deception Detection: As computational linguistics and Natural Language Processing (NLP) have grown in popularity, automated textual analysis has been used to study deception detection. One of the most influential studies in this area was conducted by Mihalcea Rada and Carlo Strapparava in *"The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language"* (2009). To find misleading assertions, the authors created computational models that examined linguistic characteristics like lexical choice, sentiment polarity, and syntactic complexity.

Another significant contribution was made by Jeff Hancock and colleagues in *"On Lying and Being Lied To: A Linguistic Analysis of Deception in Computer-Mediated Communication"* (2007). It showed that false communications frequently have changed emotional tone, decreased self-references, and increased cognitive complexity. These results showed that patterns linked to dishonest speech might be successfully identified using computational linguistic techniques.

James Pennebaker and others conducted an additional study by analysing psychological signals in language using the Linguistic Inquiry and Word Count (LIWC) methodology. In *"The Development and Psychometric Properties of LIWC2007"* (2007), Pennebaker and his associates demonstrated how linguistic markers, such as the use of pronouns, the frequency of negation, and emotive expressions, can reveal underlying psychological states, such as dishonesty and cognitive stress.

Together, these findings show that computational linguistic analysis, especially when applied methodically to huge textual datasets, can detect subtle language patterns linked to dishonesty.

2.3-Courtroom Discourse and Institutional Language: Courtroom communication has been extensively studied as a type of institutional discourse with procedural restrictions and hierarchical interaction. John M. Conley and William M. O'Barr analysed courtroom language in *"Just Words: Law, Language, and Power"* (1998), claiming that power dynamics that influence the presentation and interpretation of testimony are reflected in legal discourse. Their study showed that speakers' perceived credibility may be influenced by their linguistic approaches in legal circumstances.

In a similar vein, Diana Eades investigated linguistic dynamics within legal proceedings in *"Courtroom Talk and Neocolonial Control"* (2008). Eades emphasised how the narrative structure of courtroom language and witness answers can be shaped by institutional questioning techniques. Her research highlighted how crucial it is to examine discourse interactions rather than discrete utterances when researching legal communication.

Computational methods have started investigating judicial writings as language corpora in more recent times. Josh Blackman, Michael Bommarito, and Daniel Martin Katz showed how machine learning methods may be used to analyse legal texts in *"A General Approach for Predicting the Behaviour of the Supreme Court of the United States"* (2017). Their research demonstrated how systematic patterns in legal discourse and reasoning can be found in extensive judicial data.

However, rather than examining the language structure of testimony narratives, the majority of computational studies have concentrated on predicting judicial decisions. This suggests that more research on discourse-level elements in courtroom writings is necessary.

2.4-Cognitive Load and Linguistic Behaviour: According to a psychological study, lying puts a heavy cognitive load on people. Speakers must concurrently manage memory recall, narrative coherence, and behavioural control when they try to invent or manipulate information. Linguistic abnormalities are a common manifestation of this elevated cognitive burden.

Aldert Vrij investigated the connection between deceit and mental effort in *"Detecting Lies and Deceit: Pitfalls and Opportunities"* (2008). According to Vrij, because maintaining false narratives requires mental effort, deceitful communication frequently entails greater hesitancy, linguistic confusion, and simpler sentence patterns.

Similarly, Bella M. DePaulo and associates carried out an extensive meta-analysis in *"Cues to Deception"* (2003). According to their findings, language indications including ambiguity, a lack of information, and discrepancies between repeated tales are often found in false statements.

According to these cognitive viewpoints, linguistic indicators of dishonesty may result from psychological stress and executive control mechanisms in addition to deliberate manipulation. Therefore, the interpretation of conversation patterns related to deceit can be improved by combining linguistic research with cognitive insights.

2.5-Research Gap: While deceit detection, computational language analysis, and courtroom discourse have all been extensively studied separately, comparatively few works have integrated these viewpoints within the framework of **Indian judicial texts**. The majority of current studies on computational deception detection concentrate on social media interactions, online communication, or controlled experimental datasets.

Indian court rulings, which include in-depth accounts of witness testimony, cross-examinations, and judicial assessments of trustworthiness, provide a rich but underutilised language resource. Examining how misleading discourse patterns appear in institutional legal narratives is made possible by these texts.

Nevertheless, there are still a few comprehensive computational assessments of Indian courtroom talk. Studies that combine computational text processing, cognitive discourse indicators, and forensic language analysis to analyse deception-related patterns in court documents are particularly lacking.

In order to close this gap, the current study creates a **linguistic–computational framework for analysing deception in Indian criminal court judgements,** integrating discourse-based interpretation of trial tales with automated natural language processing tools.

III. Methodology

3.1-Research Design: This study analyses deception-related discourse patterns in criminal courtroom narratives using a computational forensic linguistic approach. The study combines techniques from discourse analysis, natural language processing (NLP), and forensic linguistics to look at linguistic markers of cognitive stress, uncertainty, and denial in legal texts.

The methodological framework integrates qualitative discourse interpretation with quantitative computational analysis. The Python programming language was used to create automated scripts that processed and analysed courtroom texts. These scripts locate high-risk conversation segments in the dataset, extract linguistic features, and compute deception-related indicators.

The overall analytical workflow consists of the following stages:

1. Corpus preparation and structuring
2. Extraction of linguistic and psychological markers
3. Deception score computation
4. High-risk segment detection
5. Stance identification
6. Lexical dispersion analysis

The study is able to methodically discover language patterns linked to dishonest communication in courtroom discourse through this multi-stage computer workflow.

3.2-Data Collection: Criminal court judgements from Indian judicial records make up the dataset used in this study. In-depth narrative reconstructions of courtroom events, including witness statements, cross-examinations, and judicial interpretations of testimony, are included in these rulings.

The gathered texts were processed as a corpus for computational analysis after being transformed into a machine-readable format. To enable automated linguistic processing, each text fragment was extracted and arranged into structured datasets in CSV format.

The corpus was divided into more manageable text units representing individual statements or discourse segments during the preparation phase. When feasible, these segments were further annotated with speaker roles, such as:

- Judge
- Witness
- Accused
- Defence counsel
- Prosecutor

The investigation of linguistic behaviour at the speaker and discourse levels was made possible by this segmentation.

3.3-Data Preprocessing and Corpus Structuring: To guarantee consistency and analytical dependability, the textual data was subjected to a number of preprocessing steps before to linguistic analysis.

The preprocessing steps included:

- Text normalisation (conversion to lowercase and removal of irregular spacing)
- Segmentation of courtroom narratives into text units
- Creation of structured datasets using Python scripts
- Cleaning and standardisation of CSV outputs
- Speaker column identification and correction

After processing, the dataset was arranged in a structured way with several variables, such as:

- Speaker role
- Text segment
- Sentiment score
- Negation markers
- Uncertainty markers
- Cognitive load indicators
- Emotional tone

This structured dataset served as the master dataset for further computational analysis.

3.4-Extraction of Linguistic Indicators: Several discourse variables were retrieved from the text corpus using automated NLP scripts in order to find linguistic patterns linked to deceit and cognitive stress.

The following are the main linguistic indicators examined in this study:

1. **Negation markers:** In testimony narratives, negative phrases like "not", "never", and "no" were found to be possible markers of denial tactics.
2.
3. **Uncertainty markers:** Words like "maybe", "perhaps", "I think", and "possibly" that convey hesitancy or uncertainty were examined as markers of cognitive reluctance.
4. **Emotional tone:** To determine whether conversational segments exhibited positive, negative, or neutral emotional orientation, sentiment analysis techniques were employed.
5. **Cognitive load indicators:** Repetition, lexical simplification, and discourse fragmentation are examples of language patterns that were once thought to indicate greater cognitive strain during speech production.

Python-based Natural Language Processing tools were used to automatically extract these linguistic traits.

3.5-Deception Score Calculation: Based on the retrieved linguistic markers, the study developed a composite deception score to measure the probability of deception within conversation segments. Several language factors were combined to determine the deception score, including:

- Negation frequency
- Uncertainty markers
- Emotional tone
- Cognitive load indicators

A numerical score representing the relative occurrence of these indicators was given to each text fragment. Greater levels of communicative stress or strategic linguistic manipulation were thought to be present in segments with higher scores, respectively.

The major dataset for further analysis was the Deception Score dataset, which contained the computed scores.

3.6-High-Risk Segment Detection: Segments with noticeably higher scores were extracted into a different dataset called High-Risk Segments once the deception score was calculated.

These sections show discourse units where several indicators of deceit are present at the same time. Researchers can concentrate on sections of testimony that may require increased cognitive effort, denial behaviour, or narrative inconsistencies by identifying such parts.

To find recurrent linguistic patterns in potentially misleading statements, the high-risk sample was further analysed.

3.7-Stance Identification: A rule-based stance categorisation model was used to analyse the communicative stance speakers took during courtroom conversation.

Every text passage was categorised into one of the following standpoint groups:

- **Assertion –** direct and confident statements without uncertainty markers
- **Denial –** statements containing negation markers
- **Uncertain stance –** statements containing uncertainty expressions

This categorisation made it possible for the study to investigate the linguistic placement of speakers inside trial narratives, especially with regard to accusations and claims of evidence.

3.8-Lexical Dispersion Analysis: The high-risk segments dataset was subjected to a lexical dispersion analysis in order to further explore linguistic patterns within potentially misleading discourse.

The analysis involved:
- Aggregating all high-risk text segments
- Tokenising the text into individual words
- Calculating frequency distributions of lexical items

A frequency distribution model implemented in Python was used to determine which terms appeared most frequently in high-risk speech parts. This technique aids in determining whether specific vocabulary choices are frequently used in statements linked to elevated deception signs.

3.9-Analytical Tools: Python-based scripts created especially for the research workflow were used for all computational analyses in this study. The principal instruments utilised consist of:
- Python programming language
- Pandas library for dataset management
- Natural Language Processing techniques for feature extraction
- Frequency distribution analysis for lexical patterns

Large textual datasets could be processed effectively thanks to these technologies, which also guaranteed the analytical process's repeatability.

3.10-Ethical Considerations: Publicly available judicial documents, which are a component of official legal records, are the only sources for the study. The research process did not involve any private data, experimental procedures, or in-person interviews. As a result, there are no direct ethical issues with participant confidentiality in the study.

IV. Results and Analysis

4.1-Overview of the Analytical Output: Several structured datasets and graphic outputs from the computer analysis show linguistic patterns connected to speech related to deceit in courtroom narratives. The outcomes consist of:
- Speaker-level statistics
- Cognitive load scores
- Correlation analysis between linguistic indicators
- Stance identification results
- High-risk discourse segments
- Lexical dispersion patterns

Together, these findings show how discourse structure in legal texts interacts with linguistic signals like negation, uncertainty, and emotional tone.

4.2-Speaker-Level Linguistic Patterns: The first phase of the analysis looked at language use in courtroom speech across various speaker roles. The speaker types—judges, witnesses, accused people, and legal representatives—were utilised to categorise the dataset.

Linguistic patterns differ considerably based on the speaker's institutional role, according to the speaker-level analysis. For instance, witness statements frequently feature higher amounts of uncertainty expressions and narrative elaboration, whereas judicial discourse is typically more ordered and authoritative. Additionally, witness testimony exhibits more variation in linguistic complexity and emotional tone. The psychological strain of cross-examination and questioning in court may be

reflected in this heterogeneity.

It is common for accused people to use defensive language structures and negation signs more frequently. These tendencies align with denial tactics frequently seen in confrontational judicial settings.

speaker_analysis

Speaker	segments	avg_deception_score	avg_uncertainty	avg_negation	avg_stress
Accused	960	4.271875	0.014583333333333300	1.2729166666666700	2.252083333333330
Judge	40	3.625	0.15	0.85	1.95
Lawyer	92	4.793478260869570	0.021739130434782600	1.5	2.4347826086956500
Unknown	4062	2.9197439684884300	0.022156573116691300	0.8636139832594780	1.5214180206794700
Witness	594	2.983164983164980	0.0	0.7878787878787880	1.7845117845117800

Table 1

4.3-Cognitive Load Indicators in Testimony: Additionally, the study looked at indications of cognitive strain that were obtained from language patterns in discourse segments. The mental work needed to create and sustain a cohesive story during speech production is referred to as cognitive load.

The following were frequently found in segments with high cognitive load scores:

- repeated lexical structures
- fragmented sentence patterns
- increased hesitation markers
- higher frequencies of uncertainty expressions

These language traits imply that while making specific declarations in trial narratives, speakers might feel more cognitive pressure.

Higher cognitive load signs are often seen in speech segments related to defence answers and contested testimony. The difficulty of maintaining narrative consistency during legal questioning may be reflected in this pattern.

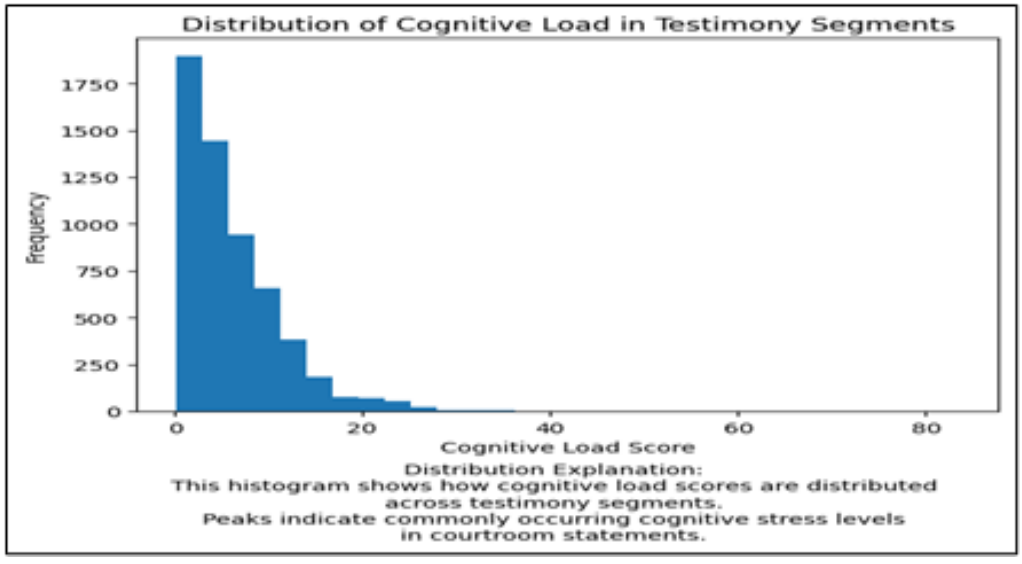

Figure 1

4.4-Correlation Between Linguistic Indicators: A correlation analysis between important variables recovered during the NLP processing stage was carried out to further investigate the relationship between various linguistic aspects.

The variables analysed include:

- sentiment score
- negation markers
- uncertainty markers
- cognitive load score
- deception score

Numerous significant associations are revealed by the correlation data. For example, negation markers frequently coexist with greater cognitive load scores, indicating that language related to denial may demand more mental effort. Similarly, there is a moderate correlation between emotional tone indicators and uncertainty markers.

These results show that deception-related language behaviour arises through interacting linguistic signals rather than depending on a single indicator.

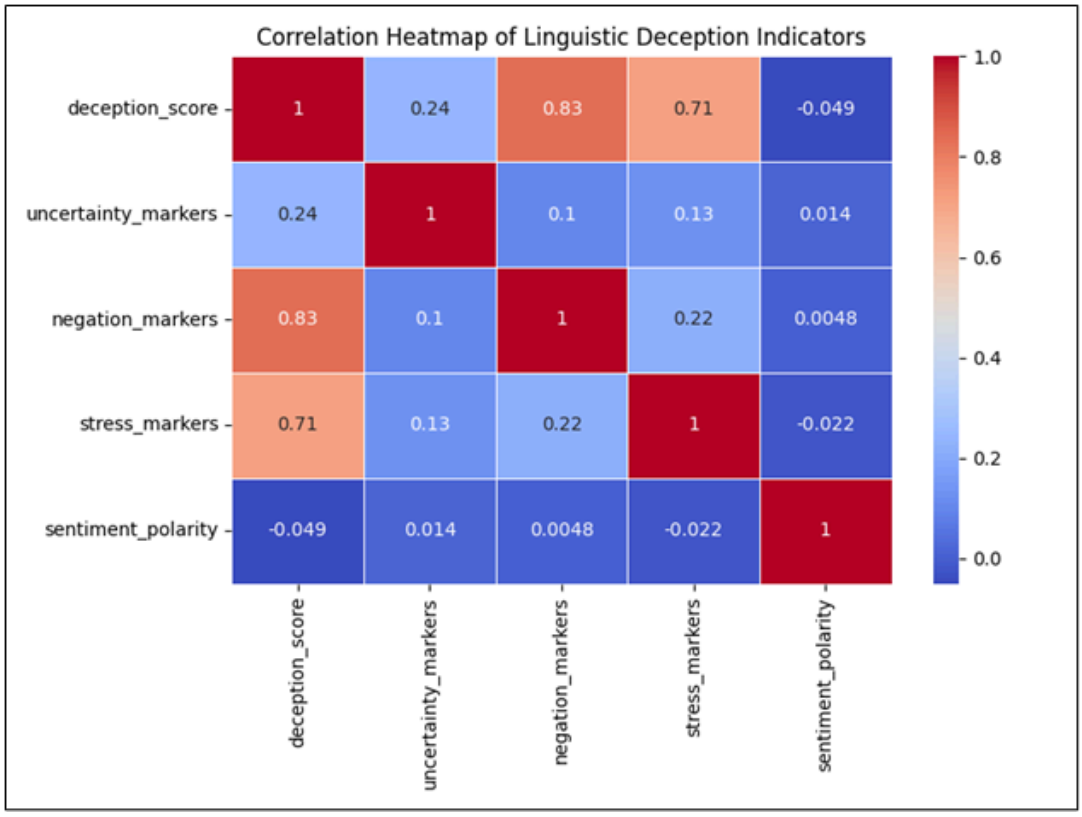

Figure 2: The heatmap visually illustrates the strength of correlations between linguistic features, highlighting clusters of variables that frequently co-occur in courtroom discourse.

4.5-High-Risk Discourse Segments: The identification of high-risk speech segments is one of the study's primary analytical results. Elevated deception ratings from the combined language indicators were used to extract these portions.

High-risk segments typically contain multiple markers simultaneously, such as:

- negation expressions
- uncertainty markers
- emotionally charged language
- elevated cognitive load scores

The existence of several indicators implies that these sections might indicate purposeful linguistic placement within courtroom narratives or areas of communicative strain.

high_risk_sample

Case	Speaker	Text Segment	deception_score	uncertainty_markers	negation_markers	stress_markers
S4	Lawyer	Because:- (1) A girl or a woman in the tradition bound non-permissive Society of India would be extremely reluctant even to admit that any incident which is likely to reflect on her chastity had ever occurred; (2) She would be conscious of the danger of being ostracised by the Society or being looked down by the society including by her own family members, relatives, friends, and neighbours; (3) She would have to brave the whole world; (4) She would face the risk of losing the love and respect of her own husband and near relatives, and of her matrimonial home and happiness being shattered; (5) If she is unmarried, she would apprehend that it would be, difficult to secure an alliance with a suitable match from a respectable or an acceptable family; (6) It would almost inevitably and almost invariably result in mental torture and suffering to herself; (7) The tear of being taunted by others will always haunt her; (8) She would feel extremely embarrassed in relating the incident to others being over powered by feeling of shame on account of the upbringing in a tradition bound society where by and large sex is taboo; (9) The natural inclination would be to avoid giving publicity to the incident lest the family name and family honour is brought into controversy; (10) The parents of an unmarried girl as also the husband and members of the husband's family of a married woman, would also more often than not, want to avoid publicity on account of the fear of social stigma on the family name and family honour; (11) The fear of the victim herself being considered to be promiscuous or in some way responsible for the incidentBharwada Bhoginbhai Hirjibhai vs State Of Gujarat on 24 May, 1983 Indian Kanoon - http://indiankanoon.org/doc/207774/ 4regardless of her innocence; (12) The reluctance to face interrogation by the investigating agency, to face the court, to face the cross examination by Counsel for the culprit, and the risk of being disbelieved, acts as a deterrent.	39.0	1	8	25
S4	Lawyer	Because:- (1) A girl or a woman in the tradition bound non-permissive Society of India would be extremely reluctant even to admit that any incident which is likely to reflect on her chastity had ever occurred; (2) She would be conscious of the danger of being ostracised by the Society or being looked down by the society including by her own family members, relatives, friends, and neighbours; (3) She would have to brave the whole world; (4) She would face the risk of losing the love and respect of her own husband and near relatives, and of her matrimonial home and happiness being shattered; (5) If she is unmarried, she would apprehend that it would be, difficult to secure an alliance with a suitable match from a respectable or an acceptable family; (6) It would almost inevitably and almost invariably result in mental torture and suffering to herself; (7) The tear of being taunted by others will always haunt her; (8) She would feel extremely embarrassed in relating the incident to others being over powered by feeling of shame on account of the upbringing in a tradition bound society where by and large sex is taboo; (9) The natural inclination would be to avoid giving publicity to the incident lest the family name and family honour is brought into controversy; (10) The parents of an unmarried girl as also the husband and members of the husband's family of a married woman, would also more often than not, want to avoid publicity on account of the fear of social stigma on the family name and family honour; (11) The fear of the victim herself being considered to be promiscuous or in some way responsible for the incidentBharwada Bhoginbhai Hirjibhai vs State Of Gujarat on 24 May, 1983 Indian Kanoon - http://indiankanoon.org/doc/207774/ 4regardless of her innocence; (12) The reluctance to face interrogation by the investigating agency, to face the court, to face the cross examination by Counsel for the culprit, and the risk of being disbelieved, acts as a deterrent.	39.0	1	8	25
D3	Unknown	Section 299 Section 300 A person commits Subject to certain culpable homicide if the exceptions culpable homicide act by which the death is is murder if the act by which caused is done -- the death is caused is done -- INTENTION (a) With the (1) With the intention of intention of causing death; or causing death; or (b) With the (2) With the intention of intention of causing such bodily causing such injury as the offender bodily injury as knows to be likely to is likely to cause cause the death of the death; or person to whom the harm is caused; or (3) With the intention of causing bodily injuryVinod Kumar @ Raju vs State Nct Of Delhi on 2 July, 2025 Indian Kanoon - http://indiankanoon.org/doc/182937635/ 9 to any person and the bodily injury intended to be inflicted is sufficient in the ordinary course of nature to cause death; or KNOWLEDGE (c) With the (4) With the knowledge knowledge that that the act is the act so imminently is likely to cause dangerous that it must death in all probability cause death or such bodily injury as is likely to cause death, and without any excuse for incurring the risk of causing death or such injury as is mentioned above.	30.5	4	7	12
D3	Unknown	Section 299 Section 300 A person commits Subject to certain culpable homicide if the exceptions culpable homicide act by which the death is is murder if the act by which caused is done -- the death is caused is done -- INTENTION (a) With the (1) With the intention of intention of causing death; or causing death; or (b) With the (2) With the intention of intention of causing such bodily causing such injury as the offender bodily injury as knows to be likely to is likely to cause cause the death of the death; or person to whom the harm is caused; or (3) With the intention of causing bodily injuryVinod Kumar @ Raju vs State Nct Of Delhi on 2 July, 2025 Indian Kanoon - http://indiankanoon.org/doc/182937635/ 9 to any person and the bodily injury intended to be inflicted is sufficient in the ordinary course of nature to cause death; or KNOWLEDGE (c) With the (4) With the knowledge knowledge that that the act is the act so imminently is likely to cause dangerous that it must death in all probability cause death or such bodily injury as is likely to cause death, and without any excuse for incurring the risk of causing death or such injury as is mentioned above.	30.5	4	7	12
D2	Unknown	The other aspects like not calling the police official from the concerned police station, no photography of the spot of recovery or not calling the crime team or not making any attempt to pick up chance prints from the handle of the weapon also create furtherState vs Naresh Orthjers on 11 March, 2025 Indian Kanoon - http://indiankanoon.org/doc/88884465/ 20doubt about the alleged recovery.	20.5	0	9	7
S7	Unknown	Phulpatta Devi and RenuRamesh Harijan vs State Of U.P on 21 May, 2012 Indian Kanoon - http://indiankanoon.org/doc/104710706/ 6did not sleep in the house of Ramesh, however, as he was living permanently in the city and did not say that he was present on that day in the village, his evidence cannot be taken into consideration so far as this issue is concerned.	20.5	0	11	4
S7	Unknown	Phulpatta Devi and RenuRamesh Harijan vs State Of U.P on 21 May, 2012 Indian Kanoon - http://indiankanoon.org/doc/104710706/ 6did not sleep in the house of Ramesh, however, as he was living permanently in the city and did not say that he was present on that day in the village, his evidence cannot be taken into consideration so far as this issue is concerned.	20.5	0	11	4
D2	Unknown	The other aspects like not calling the police official from the concerned police station, no photography of the spot of recovery or not calling the crime team or not making any attempt to pick up chance prints from the handle of the weapon also create furtherState vs Naresh Orthjers on 11 March, 2025 Indian Kanoon - http://indiankanoon.org/doc/88884465/ 20doubt about the alleged recovery.	20.5	0	9	7

Table 2

Examples of language patterns seen in the high-risk segments include attempts to reframe or misinterpret preceding remarks, hedging expressions, and repeated denials. In the institutional context of the courtroom, speakers frequently use these linguistic techniques to control the credibility of their testimony.

During cross-examination, when participants must answer direct and frequently difficult questions from opposing counsel, certain patterns commonly arise. Speakers may use negation markers, uncertainty expressions, and stress-related linguistic signs in these situations, which raises the computer analysis's deception ratings.

Because these segments include language cue clusters linked to cognitive burden and defensive communication techniques, the computational model classified them as high-risk. For instance, frequent negation (e.g., "I did not", "I never", etc.) and uncertainty markers (e.g., "perhaps", "I think", "it seems") may indicate attempts to minimise accountability or steer clear of conclusive statements.

These results imply that language traits associated with deceit in courtroom discourse are not dispersed at random but rather typically surface in situations when there is interactional pressure, especially during adversarial questioning. The discovery of these patterns shows how effective computational linguistic methods are in identifying speech structures linked to possible dishonesty in court testimony.

4.6-Stance Identification Results: Discourse fragments were divided into three communicative categories using the stance identification model:

- Assertion
- Denial
- Uncertainty

According to the findings, assertive statements dominate judicial narratives, especially in the sections that describe factual findings and judicial reasoning. Nonetheless, denial and ambiguity postures are more common in accused comments and witness evidence.

While uncertainty postures sometimes include hedging statements like perhaps, maybe, or I think, denial stances are closely linked to negation markers.

These findings imply that posture classification can offer valuable insights into the linguistic positioning of speakers during courtroom discussions.

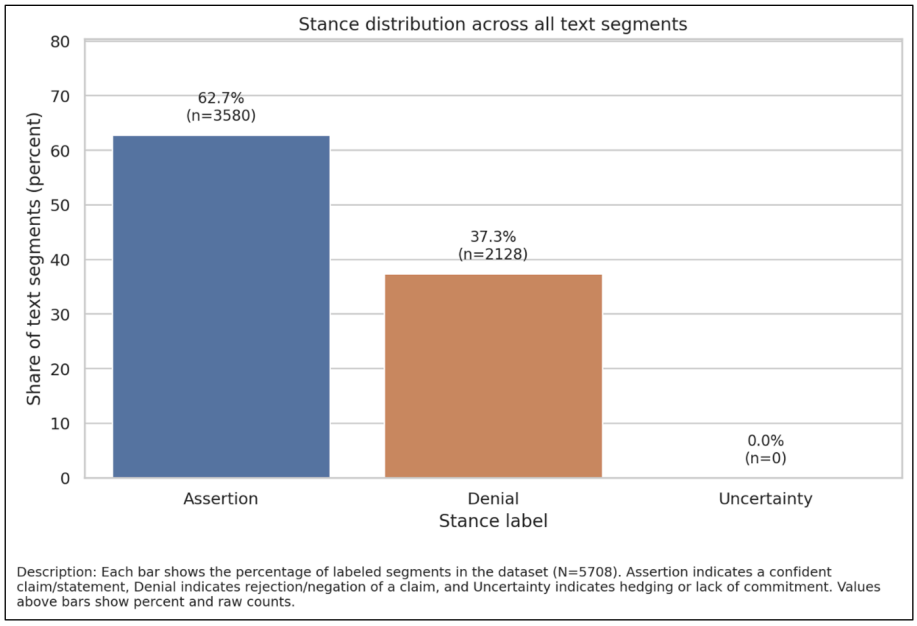

Figure 3

4.7-Lexical Dispersion in High-Risk Segments: A lexical dispersion analysis was performed on the high-risk dataset to better explore language behaviour in potentially misleading conversations.

The study determined the most common lexical items in high-risk conversation segments. Themes of rejection, doubt, and evidentiary explanation are frequently associated with these phrases.

Typical lexical patterns consist of:

- references to memory or recollection
- repeated use of negation terms
- references to events, time, and actions

When reacting to disputed accusations, speakers may employ particular language tactics, as indicated by the recurrence of such lexical elements.

The concentration of negation markers and cognitive verbs indicates that speakers frequently use linguistic strategies linked to managing uncertainty and justifying their statements. This lexical grouping corresponds with deception research findings showing that deceptive speakers often depend on repetitive discourse markers and self-referential structures to uphold narrative consistency under cognitive strain. Therefore, the lexical frequency pattern supports the broader computational findings of this study, indicating that high-risk segments exhibit distinctive lexical signatures that may serve as indicators of deception within forensic discourse.

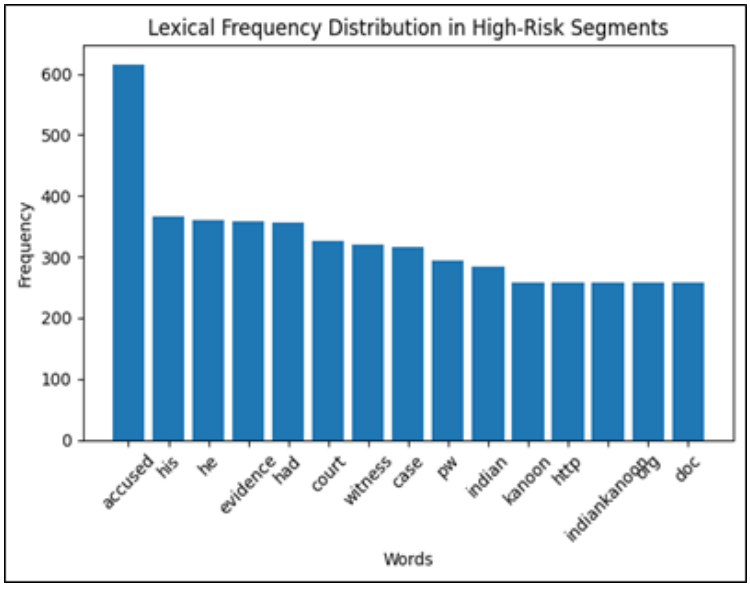

Figure 4 illustrates the lexical frequency distribution of the most recurrent words identified in high-risk deception segments extracted from the courtroom transcript corpus. The visualisation highlights the dominant lexical items that appear most frequently in segments associated with elevated deception scores. Words such as “not”, “did”, “know”, and “said” occur with high frequency, reflecting patterns of negation, defensive framing, and narrative reconstruction commonly observed in deceptive discourse.

The concentration of negation markers and cognitive verbs indicates that speakers frequently use linguistic strategies linked to managing uncertainty and justifying their statements. This lexical grouping corresponds with deception research findings showing that deceptive speakers often depend on repetitive discourse markers and self-referential structures to uphold narrative consistency under cognitive strain. Therefore, the lexical frequency pattern supports the broader computational findings of this study, indicating that high-risk segments exhibit distinctive lexical signatures that may serve as indicators of deception within forensic discourse.

4.8-Interpretation of Findings: When considered collectively, the findings show that several interacting linguistic cues, rather than a single diagnostic feature, are responsible for the emergence of deception-related speech patterns.

In courtroom narratives, portions of elevated communication stress can be identified through the use of negation markers, uncertainty expressions, emotional tone, and cognitive load indicators.

A systematic, large-scale analysis of courtroom discourse is made possible by the combination of computer text analysis and forensic linguistic interpretation. The study identifies recurrent linguistic patterns that could go overlooked in conventional qualitative analysis by analysing big textual datasets.

By showing how natural language processing techniques can be applied to institutional legal texts to investigate complex communicative phenomena like deception, credibility construction, and discourse positioning, these findings add to the expanding field of computational forensic linguistics.

V. Discussion

The results of this work show how computational linguistic analysis can identify patterns of strategic communication in judicial testimony, adding to the expanding corpus of research on deception -

detection and institutional discourse. The current study demonstrates how language cues linked to cognitive pressure and communicative adaptability manifest in legal procedures by fusing corpus-based techniques with discourse analysis. The findings imply that dishonesty in courtroom language is both an institutionally shaped communicative practice and an individual psychological reality, impacted by adversarial questioning, legal strategy, and the formal limitations of judicial proceedings.

5.1-Deception Research and Linguistic Indicators: Deceptive communication frequently incorporates recognisable linguistic cues, including enhanced negation, hedging, and indirect wording, according to long-standing research in deception studies. Research in this area, such as Jeffrey Hancock and colleagues' *Deception Detection from Linguistic Cues (2005)*, has demonstrated that distancing techniques and linguistic structure changes intended to conceal accountability or lessen commitment to factual claims are frequently found in deceptive statements.

These findings are corroborated by the current analysis. Particularly in situations when speakers were answering direct questions, segments with higher deception scores regularly revealed clusters of negation markers, uncertainty expressions, and stress-related signs. These patterns imply that attempts to control credibility through linguistic modification rather than just outright lying may be a part of misleading discourse. In other words, rather than using straightforward factual contradiction, deceit in courtroom language frequently takes the form of discursive techniques of mitigation and reframing.

5.2-Cognitive Load and Communicative Pressure: Cognitive Load Theory, which contends that dishonest communication demands more mental effort than truthful communication, is another significant theoretical viewpoint pertinent to the findings. According to academics like Aldert Vrij in Detecting Lies and Deceit (2008), lying raises cognitive demand since people have to concurrently invent information, keep an eye on the coherence of their own narratives, and foresee any obstacles from others.

This framework is compatible with the linguistic indications found in the computational study. Increased stress markers and complex negation patterns were frequently associated with elevated deception scores, which may indicate the cognitive strain encountered during high-pressure interactions. The cognitive load on speakers is expected to increase in courtroom settings where utterances are scrutinised and subject to cross-examination. Linguistic indicators linked to cognitive overload may therefore become more noticeable, especially when speakers try to balance conflicting narratives or answer unexpected queries.

These results support the notion that identifying misleading statements should not be the exclusive focus of deception detection. Instead, deceit might be best viewed as a cognitively demanding communicative process where speakers modify their language use in response to institutional demands.

5.3-Forensic Linguistics and Institutional Communication: Courtroom speech has long been acknowledged in the field of forensic linguistics as a unique communicative setting with uneven power dynamics and formal rules. Legal language frequently reflects institutional interests rather than just conversational norms, according to academics like Malcolm Coulthard and Alison Johnson in *An Introduction to Forensic Linguistics* (2007).

The study's findings support this viewpoint. Variations in the distribution of indicators connected to deception were revealed by linguistic variance among speakers, including judges, attorneys, witnesses, and accused people. In argumentative sequences, legal professionals—especially attorneys—frequently showed higher deceit ratings. This might just be a reflection of the strategic framing inherent in adversarial legal argumentation rather than a sign of deliberate dishonesty.

On the other hand, testimony ascribed to accused people and witnesses often included linguistic elements linked to narrative reconstruction and ambiguity, indicating a different communication orientation. These variations emphasise how crucial it is to understand deception signs in the institutional framework of legal discourse, where communicative pressure, legal procedure, and rhetorical strategy interact in intricate ways.

5.4-Courtroom Discourse and Interactional Dynamics: Legal procedures are marked by structured interactional patterns, especially during examination and cross-examination, according to research in courtroom discourse analysis. According to Janet Cotterill's 2002 book Language in the Legal Process, cross-examination frequently puts speakers under aggressive questioning and forces them to react fast and accurately.

The current study demonstrates that linguistic patterns linked to greater deception scores are often produced in such interactional circumstances. High-risk speech segments frequently surfaced during conversations where speakers had to defend or clarify earlier claims. In these circumstances, speakers may use techniques like hedging, negation, and repetition as ways to respond to outside pressure while preserving narrative coherence.

These findings imply that intentional attempts to deceive, as well as the interactional constraints inherent in courtroom questioning, may give rise to deception-related linguistic traits. As a result, it is important to carefully consider the larger communicative context in which the assertions are made when interpreting computational detection techniques.

5.5-Implications for Computational Forensic Analysis: There is a lot of promise for forensic linguistic research when computer methods are combined with discourse analysis. Computational models can help researchers discover conversation segments that require more in-depth qualitative analysis by methodically finding linguistic patterns linked to cognitive stress and communicative adjustment.

But the results also highlight how crucial contextual interpretation is. Linguistic signs of dishonesty should be interpreted as signals of communicative conflict or strategic discourse management rather than as conclusive proof of dishonesty. Computational techniques can offer important insights into how institutional communication functions in judicial settings when paired with qualitative analysis and legal knowledge.

Overall, the study's findings show that the cognitive and institutional dynamics of legal discourse are intimately linked to language characteristics associated with lying. Researchers can gain a better understanding of how language serves as a strategic and cognitive tool in legal procedures by computationally analysing these patterns.

VI. Limitations and Future Research

Although the results of this study offer insightful information about linguistic patterns associated with lying in courtroom discourse, a number of caveats need to be noted. In order to properly understand the findings and direct future computational forensic linguistics research, it is critical to acknowledge these limitations.

6.1-Dataset Scope and Representativeness: English-language court records from Indian court rulings make up the majority of the corpus used in this study. These texts are written reconstructions of spoken testimony rather than verbatim transcripts of live courtroom interaction, despite the fact that they offer rich instances of courtroom discourse. The linguistic trends found in the research may be influenced by the fact that judicial rulings frequently summarise or paraphrase witness testimonies and legal arguments.

Additionally, the dataset is restricted to a certain geographic jurisdiction and legal framework. Because of this, the linguistic traits found in this study might represent the norms of Indian legal discourse and might not be entirely applicable to other legal systems or linguistic groups.

In order to overcome this constraint, future studies could use audio-based courtroom recordings, multilingual legal corpora, or direct courtroom transcripts, which might offer a more thorough depiction of spontaneous speech patterns.

6.2-Interpretation of Deception Indicators: Interpreting verbal cues linked to dishonesty is another restriction. The computational model finds patterns that are frequently connected to cognitive pressure or defensive communication tactics, such as negation marks, uncertainty expressions, and stress indicators. These indicators, however, should not be taken as conclusive proof of dishonest intent.

Speakers often use sophisticated linguistic techniques in institutional settings like courtrooms because of interactional pressure, rhetorical goals, or legal restrictions. Therefore, higher deception scores may not always indicate intentional deceit but rather communication stress or argumentative framing.

Therefore, in order to effectively differentiate between strategic legal rhetoric and truly false communication, future research should integrate computational analysis with qualitative discourse analysis and legal interpretation.

6.3-Model Simplification and Feature Selection: This study's computer model is based on a set of pre-established language indicators, such as stress-related lexical patterns, negation markers, and uncertainty expressions. These characteristics are only a small portion of the language elements that could affect courtroom discourse, even though they offer a helpful beginning point for identifying possible deception signs.

More advanced models could incorporate additional features such as:

- syntactic complexity and sentence structure
- semantic role analysis
- discourse coherence patterns
- prosodic and acoustic features from spoken testimony

Future research may be able to create more advanced deception detection frameworks that can capture more complex linguistic and cognitive patterns by incorporating such features.

6.4-Future Directions for Computational Forensic Linguistics: Despite these drawbacks, the current work shows how computational methods can be used to analyse institutional discourse in legal contexts. This work could be expanded in a number of ways by future research.

First, larger and more varied corpora might be created to study linguistic patterns associated with deceit in various legal systems and cultural situations. Second, speech segments might be automatically classified using machine learning approaches based on linguistic and contextual feature combinations. Finally, interdisciplinary collaboration between linguists, legal scholars, and computational researchers could contribute to the development of more reliable analytical frameworks for courtroom discourse analysis.

Future research may improve our knowledge of how language reflects institutional dynamics, strategic communication, and cognitive processes in the judicial system by combining computational techniques with conventional discourse analysis.

VII. Conclusion

Through the use of computational linguistic techniques on judicial texts, this study investigated deception as a discursive phenomenon within the institutional context of courtroom communication. The study showed how deception-related patterns can be found through systematic computer analysis by analysing a corpus of criminal case judgements and identifying language elements linked to cognitive strain and communication pressure.

The results revealed that discourse segments with higher deception scores frequently contained clusters of negation markers, uncertainty expressions, and stress-related linguistic indicators. These characteristics were more noticeable in exchanges that resembled cross-examination frameworks and were linked to aggressive questioning and argumentative speech. Such findings support existing research suggesting that deception is often reflected not through isolated linguistic markers but through combinations of linguistic signals that emerge under cognitive and interactional pressure.

The research also revealed significant variations among courtroom participants. While other speakers had somewhat lower levels of deceit-related indicators, speech attributed to attorneys and accused individuals frequently indicated higher deception ratings. These variations suggest that deception-related linguistic patterns may be shaped by the institutional roles and communicative strategies of courtroom participants, rather than being solely the product of individual psychological factors.

From a methodological perspective, the study demonstrates the value of combining computational text analysis with forensic linguistic frameworks. Large amounts of legal language were extracted and analysed thanks to the automated pipeline created in this study, which made it possible for researchers to find speech patterns that might not be readily apparent through manual analysis alone. These methods provide new opportunities for analysing institutional discourse in legal contexts and add to the expanding field of computational forensic linguistics.

Overall, the findings indicate that deception in courtroom discourse should be understood as a complex interaction between cognitive processes, linguistic strategies, and institutional constraints. In order to facilitate more in-depth assessments of communication within legal systems, computational approaches offer a helpful tool for identifying discourse segments that call for additional qualitative examination.

Larger multilingual corpora, the incorporation of audio features from oral testimony, and the application of sophisticated machine learning algorithms to enhance the detection of deception-related verbal patterns are some ways that future research may build on this work. Computational forensic linguistics may continue to advance our knowledge of how language functions within the frameworks of court communication through such interdisciplinary initiatives.

References

1. Angeliki Lazaridou, Dinu, G., & Baroni, M. (2015). Hubness and Pollution: Delving into Cross-Space Mapping for Zero-Shot Learning. https://doi.org/10.3115/v1/p15-1027
2. Chen, H., Miranda, R., Zeng, D. D., Demchak, C., & Therani Madhusudan. (2003). Intelligence and Security Informatics. Springer.
3. Cotterill, J. (2003). Language and Power in Court : a Linguistic Analysis of the O.J. Simpson Trial. Palgrave Macmillan.
4. Coulthard, M. (2007). An Introduction to Forensic Linguistics. Routledge. https://doi.org/10.4324/9780203969717
5. DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. Psychological Bulletin, 129(1), 74–118. https://doi.org/10.1037/0033-2909.129.1.74

6. Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., Monaco, A. P., & Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. Nature, 418(6900), 869–872. https://doi.org/10.1038/nature01025
7. Fisher, S. E., & Scharff, C. (2009). FOXP2 as a molecular window into speech and language. Trends in Genetics, 25(4), 166–177. https://doi.org/10.1016/j.tig.2009.03.002
8. Fitzpatrick, E., Bachenko, J., & Fornaciari, T. (2015). Automatic Detection of Verbal Deception. In Synthesis Lectures on Human Language Technologies. Springer International Publishing. https://doi.org/10.1007/978-3-031-02158-9
9. Graham, S. A., & Fisher, S. E. (2013). Decoding the genetics of speech and language. Current Opinion in Neurobiology, 23(1), 43–51. https://doi.org/10.1016/j.conb.2012.11.006
10. Granhag, P. A., & Strömwall, L. A. (Eds.). (2004). The Detection of Deception in Forensic Contexts. Cambridge University Press. https://doi.org/10.1017/cbo9780511490071
11. Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2007). On Lying and Being Lied To: A Linguistic Analysis of Deception in Computer-Mediated Communication. Discourse Processes, 45(1), 1–23. https://doi.org/10.1080/01638530701739181
12. Hancock, J. T., Woodworth, M. T., & Porter, S. (2011). Hungry like the wolf: A word-pattern analysis of the language of psychopaths. Legal and Criminological Psychology, 18(1), 102–114. https://doi.org/10.1111/j.2044-8333.2011.02025.x
13. Juan Oyarzo Alvarado. (2023). Raymond W. Gibbs Jr.: Metaphor Wars: Conceptual Metaphors in Human Life. Onomázein Revista de Lingüística Filología Y Traducción, 61, 249–254. https://doi.org/10.7764/onomazein.61.13
14. Levine, T. R. (2014). Truth-Default Theory (TDT). Journal of Language and Social Psychology, 33(4), 378–392. https://doi.org/10.1177/0261927x14535916
15. Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying Words: Predicting Deception from Linguistic Styles. Personality and Social Psychology Bulletin, 29(5), 665–675. https://doi.org/10.1177/0146167203029005010
16. Ott, M., Choi, Y., Cardie, C., & Hancock, J. (2011). Finding Deceptive Opinion Spam by Any Stretch of the Imagination (pp. 309–319). Association for Computational Linguistics. https://aclanthology.org/P11-1032.pdf
17. Plomin, R., & Kovas, Y. (2005). Generalist Genes and Learning Disabilities. Psychological Bulletin, 131(4), 592–617. https://doi.org/10.1037/0033-2909.131.4.592
18. Tiersma, P. (2012). A History Of The Languages Of Law. 12–26. https://doi.org/10.1093/oxfordhb/9780199572120.013.0002
19. Vernes, S. C., Newbury, D. F., Abrahams, B. S., Winchester, L., Nicod, J., Groszer, M., Alarcón, M., Oliver, P. L., Davies, K. E., Geschwind, D. H., Monaco, A. P., & Fisher, S. E. (2008). A Functional Genetic Link between Distinct Developmental Language Disorders. New England Journal of Medicine, 359(22), 2337–2345. https://doi.org/10.1056/nejmoa0802828
20. Vrij, A. (2011). Detecting Lies and Deceit : Pitfalls and Opportunities. John Wiley & Sons.
21. Wagner, A., & Cheng, L. (2011). Exploring courtroom discourse : the language of power and control. Ashgate.