



**LEXICAL DENSITY IN PAKISTANI SUPREME COURT
JUDGMENTS: A CORPUS-BASED STUDY USING
ANTCONC AND MAT WITHIN A SYSTEMIC
FUNCTIONAL LINGUISTICS FRAMEWORK**

Danyal Ahmad danyalahmedeeeee@gmail.com BS (Hons) Scholar, Department of Applied Linguistics, GC University Faisalabad, Punjab, Pakistan.

Fajar Rubab fajarrubab378@gmail.com BS (Hons) Scholar, Department of Applied Linguistics, GC University Faisalabad, Punjab, Pakistan.

Ayesha Faiz ayeshah2340@gmail.com BS (Hons) Scholar, Department of Applied Linguistics, GC University Faisalabad, Punjab, Pakistan.

Abstract

In this study, the approach taken is corpus-based method and the theory used is Systemic Functional Linguistics (SFL), which is related to the study of lexical density in Supreme Court judgments of Pakistan. The main purpose is to see how the meaning is created in legal texts by the choice of words, and to quantify the amount of informational compression in the legal language. A set of 25 Supreme Court decisions was collected, with 6,535-word types and 69,626 tokens. The data have been processed in the MAT (Part-of-Speech tagging) and AntConc (frequency and corpus analysis). The result shows that lexical density of Pakistani judicial language is 58.9%, which is relatively high and shows that there is a high level of informational compression in the Pakistani judicial language. The nominalization of legal language is evident from the POS analysis; nouns are followed by verbs, adjectives and adverbs. In the frequency analysis, it is also clear that the Supreme Court judgment uses specific legal terms like court, petitioner, evidence and jurisdiction, indicating the institutional and procedural character of Supreme Court decisions. From the results, it can be concluded that the type of legal discourse is very abstract, formal, and lexically dense, and its meaning is built mainly by means of nominalization and specialized vocabulary. These linguistic characteristics enhance the accuracy and authority of judicial writing but should not diminish the understanding of non-specialist readers. The study is in support of the theoretical claims of Systemic Functional Linguistics, specifically the idea of lexical density in written institutional discourse of M. A. K. Halliday. It also proves the ability of corpus-based tools to analyze legal language. In a broad sense, the researchers are advancing the field of corpus linguistics and legal discourse analysis with empirical data on Pakistani Supreme Court decisions, and by offering a window into the complexity of legal discourse. The abstracts of the legal writings of Pakistani Supreme Court judgments were used in this study to investigate lexical density in legal discourse. In this study, the CSL was used to analyze lexical density of the legal writings of Pakistani Supreme Court judgments.

Keywords: *AntConc, Corpus Linguistics, Judicial Language, Legal Discourse, Lexical Density, MAT, Pakistani Supreme Court Judgments, POS Tagging, Systemic Functional Linguistics.*

Corresponding Author: Fajar Rubab (BS (Hons) Scholar, Department of Applied Linguistics, GC University Faisalabad, Punjab, Pakistan)
Email: fajarrubab378@gmail.com

1. Introduction

It is one of the most specialized and information-rich forms of institutional language: legal discourse. It is complex, involves technical lexical items and is expressionally dense, meaning that it contains expressions that are densely packed and only comprehensible with the knowledge of the context. In the field of law, judicial decisions, especially Supreme Court rulings, are regarded as authoritative works in which the authority of law, the authority of institutions, and the power of language merge. Judicial language, as a discourse, yields an abundant arena for analysis of meaning making through the selection of words. Some of the texts are often nominalized, in noun phrases and with a high level of technical vocabulary and differ significantly from common written or spoken texts. These features make legal sentences ideal for the investigation of corpus-based linguistics. Lexical density is a term used by M. A. K. Halliday in Systemic Functional Linguistics to denote the ratio between the number of lexical (content) words and the overall number of words in a text. Written language, Halliday continues, is more "lexically dense" than spoken language, as it packs meaning into words instead of grammatical forms. This compression is particularly apparent in legal talk, which must be both precise and authoritative. Although judicial language is of immense importance, there is hardly any empirical study that has studied lexical density in Supreme Court judgments in Pakistan using corpus-based approach. The current literature lacks quantitative, computationally assisted analysis of the English in court. It is found that most existing studies focus on general legal English or qualitative discourse analysis, thus, quantitative and computationally supported linguistic investigation is still lacking. In response to this gap, the present study uses the corpus linguistic tools which were used to analyze the lexical density in judicial discourse in Pakistan. This study investigates the overall lexical density and the lexical distribution patterns of a corpus of Supreme Court decisions with the help of the software AntConc and MAT. The study sets out to establish empirical evidence of the role that lexical resources play in legal language complexity and informativeness.

1.1. Background of the Study

Legal discourse is regarded as one of the most specialized and information dense institutional languages. Often technical jargon, nominalization, and structurally complex expressions which make it precise but hard for non-specialist readers to interpret. Supreme Court judgments are very much of importance among legal genres, since they are key cases of legal reasoning and institutional decision-making. Judicial texts offer a valuable source of data for the study of the construction of meaning through lexical selection from a linguistic standpoint. The texts are highly communicative, but they make use of noun phrases and specialized vocabulary that are not common in the day-to-day use of language. This facilitates their use in corpus-based linguistic research to a great extent. Lexical density is one of the concepts used in present research and was first proposed by M. A. K. Halliday of Systemic Functional Linguistics, which deals with the proportion of lexical (content) words in a text. Written language, particularly institutional discourse, is generally more lexical than spoken language, because it compresses information. Although the study of the Supreme Court judgments is very important in Pakistan, very little research has been conducted on the Pakistani Supreme Court judgments corpus-based approach. This study aims to fill this gap in literature by conducting an analysis of lexical density of judicial texts, using the computer tools AntConc and MAT.

1.2. Problem Statement

The legal language is a very complex and technical one, but empirical corpus-based research on Lexical density is lacking in Pakistani Supreme Court judgments. Previous research frequently focuses on descriptive or qualitative research methods, as well as an absence of measurable evidence of the contribution of lexical items to informational density in legal discourse. Moreover, little efforts have been made to use corpus-based analysis tools like AntConc and MAT in the analysis of Pakistani judicial texts along with SFL. The methodological aspect of this is the lack of knowledge about the contribution of lexical characteristics like the noun dominance, nominalization and special vocabulary to the complexity of legal language. Therefore, the linguistic nature of Supreme Court (SC) decisions in Pakistan are under-investigated with respect to their quantitative corpus linguistic analysis.

1.3. Research Gap

Although there is an increasing volume of research on legal discourse and corpus linguistics, there are still some gaps to be filled. There is not much empirical study on the lexical density of Pakistani Supreme Court judgments, first. Firstly, limited empirical studies exist on lexical density in Pakistani Supreme Court judgments. Second, most of the studies conducted is of qualitative nature, but not quantitative, based on a corpus. Thirdly, there is a lack of integration of computational tools, including AntConc and MAT with SFL theory in legal discourse studies. Moreover, there is still no in-depth study which has studied the impact of lexical density in the South Asian judicial context specially

Pakistan. The paucity of studies underscores the need for a systematic study on corpora, grounded in linguistic theory but also using computational analysis. The study looks at the significance of the study. This research is very important because it is related to Corpus Linguistics, Legal discourse analysis, and Systemic functional linguistics. Initially, it is a quantitative study of the lexical density of judgments of the supreme court of Pakistan that provides empirical data on the complexity of a legal text. First, it presents a quantitative study of the lexical density of supreme court judgments of Pakistan, which can provide empirical data on the complexity of a legal text. Second, it extends Halliday's theory of lexical density to a new legal and regional context. Furthermore, the study proves that a combination of AntConc and MAT for linguistic analysis is effective and can be replicated in future studies. The findings also have implications for legal readability, revealing the information complexity of judicial writing, which could impact on the comprehension of non-specialist readers.

1.4. Research Objectives

1. To explore lexical density in the SC judgements using a corpus-based approach in Pakistani context.
2. To explore some Hallidayan aspects of density of lexemes in Legal Texts.
3. To recognize linguistic elements that make a lexical density in judicial texts.
4. To illustrate the use of corpus-based tools in the study of legal discourse.

1.5. Research questions

- i. How does the legal language of Pakistan show Hallidayan characteristics of lexically dense discourse in writing?
- ii. How much lexical density is responsible for the complexity of legal language in Pakistan?
- iii. What are the linguistic features which may enhance the lexical density in Pakistani Supreme Court Judgments?
- iv. What are the possible applications of corpus-based tools in the study of Lexical density in Pakistani legal discourse?

1.6. Rational of the study

The motivation behind this study is the empirical investigation of the linguistic complexity of judicial text in Pakistan using computational approach. Legal language is known to be complex and information rich but there is not much quantitative evidence to back this statement when it comes to the judgments of the Supreme Court of Pakistan. This study combines the two theories of corpus linguistics and the system structural functional linguistics, which will give a systematic and measurable method in analyzing lexical density. The objective identification of the lexical patterns is made possible with the use of AntConc and MAT, while Halliday's theoretical framework is used for interpretive depth. This blend provides empirical rigor and relevance.

2. Literature review

2.1. Corpus Linguistics and Legal Discourse Studies

The use of the corpus approach has undergone considerable changes in recent corpus linguistic research, including the analysis of specialized text, especially in the legal field. Corpus-based approaches enable researchers to analyse large amounts of authentic texts and detect regularities in the language of texts, which cannot be easily discerned from the text manually. The study of the corpus has been a major focus in the study of legal discourse, which is institutionalized and formulaic. The findings of earlier studies have shown that legal texts are lexically specific, repetitive, and have domain-specific terms. Research on legal corpus linguistics has revealed that judicial language is found to be very dissimilar from general written language regarding its frequency distribution, its collocational features and syntactic complexity. As the study of legal English demonstrates, for instance, the use of nominal structures and the specific legal vocabulary are largely inherent characteristics of legal texts. The results of these findings can be used as a basis for comprehending legal discourse as a very structured and formulaic register. Moreover, corpus-based legal research has the advantage of focusing on the patterns of the use of language in institutions, which is done by using frequency analysis and concordances. These approaches remove the need to rely on intuition-based analysis and instead are based on empirical evidence from large text corpora.

2.2. Lexical density

The concept of lexical density has been studied extensively in Systemic Functional Linguistics as an important aspect of informational complexity in texts. Lexical density is the rate of lexical (content) words to total words in a text, according to M.A.K. Halliday. Spoken discourse tends to have more meaning per word, or be more lexical, than written discourse.

Research on lexical density in various genres has revealed that academic, scientific and legal texts tend to be more densely lexicon-syntactic than other conversational, or narrative texts. This is because of the high frequency of nouns, nominalization, and technical terms. One of the most lexically dense forms of writing is legal discourse, which conveys complex legal reasoning in compact forms. Lexical density has also been validated as a reliable indicator of complexity in recent studies based on a corpus. It is suggested that the higher lexical density is related to the greater cognitive load in the reader, especially for non-specialists, because dense texts demand more interpretative effort for the reader. But, although a large amount of research has been done into general English and academic discourse, little work has been done on lexical density in South Asian legal discourse, specifically in the context of Supreme Court judgments.

2.3. The complexity of legal language and nominalization.

One of the characteristic phenomena in legal discourse that was identified was nominalization, which is the process of transforming verbs and processes into noun forms. This linguistic characteristic is one of the main reasons for lexical density as it is a way of condensing complex meaning into nominal constructions. In legal texts, nominalizations often serve the function of precision, the formal and institutional authority. For example, “decide” is frequently converted into a noun, e.g., “decision”; “interpret” becomes “interpretation”; “convict” becomes “conviction.” This change makes clauses more informational and less syntactically complex. The use of nominalization has consistently been identified in studies of legal language as one of the main techniques for legal language to become complex. This is especially noticeable in judicial writing, where abstraction and precision are paramount. But there is little quantitative evidence on the contribution of nominalization to overall lexical density as it is found in the linguistic data of legal discourse like Pakistani judicial discourse.

2.4. Legal Studies in South Asian Context (Corpus-Based)

There has been some research work relating to linguistic characteristics of the judicial discourse in South Asian context, but corpus-based studies are still limited. There are indications in the available study literature in regional journals (like JALT and CEEOL journal literature) that the density of the legal lexicon and the frequency of nominalization in South Asian are like those in Western legal systems. This is because most of these studies are descriptive or have small corpora and do not use integrated computer tools like AntConc and MAT. Moreover, no systematic analysis of Supreme Court's judgments in Pakistan has been done through large scale corpus analysis. This gap underscores the need for further empirical research based on both corpus linguistics and Systemic Functional Linguistics for studying the lexical density of legal text in Pakistan. Theoretical framework in this study is systemic functional linguistics or SFL. Theoretical framework in this study is systemic functional linguistics or SFL. The theoretical approach in this research is taken from Systemic Functional Linguistics (SFL) proposed by M. A. K. Halliday. SFL takes the view that language is a semiotic social system, meaning that its meaning depends on context and function. In this perspective, lexical density has been regarded as one of the significant characteristics of written discourse in specific institutional contexts, like law, education and science. Halliday's theory is that written language has a high density of words conveying meaning and that this density comes about due to compression of meaning in the content words and not in the grammatical structure. The legal discourse is in line with this opinion, because it is based mainly on lexical items to encode complex meanings of law. For judging, this is true as nouns and nominal structures occur in large numbers in judicial texts, which suggests that SFL is suitable to analyze the density of judicial text regarding Supreme Court judgments.

3. Methodology

3.1. Research Design

The research design of this study is quantitative corpus-based research design from the view of applied linguistics. It is an attempt to study the lexical density of Pakistani Supreme Court judgments with the help of computational approach. The study is descriptive in nature, since it aims to detect and describe the language patterns without engaging any manipulation of the language. The theoretical approach to the study was Systemic Functional Linguistics (SFL) by M. A. K. Halliday which discussed lexical density. The use of corpus with SFL allows quantifying and functional analysis of legal discourse.

3.2. Data Selection

The data for this study is from 25 Supreme Court judgments from the official records of Pakistan. The judgments were released in 2026 and cover a diverse range of legal types, such as civil, criminal, constitutional, and procedural cases.

Here are the supreme court judgements:

Table: 1

Sr. No	Title	Case number	Hearing date
1	Muhammad Afzal & Usama Mehmood v. Director GEO Information Center Survey of Pakistan, Peshawar and others	Civil Petitions No.5674 and 5708 of 2025	11.02.2026
2	Sunni Masih v. The State	Criminal Appeal No. 47/2023	20.04.2026
3	Mst. Selab Akhtar v. Quwat Khan & Others	C.P.L.A. No. 4792/2025	06.01.2026
4	Taimoor Khizar Hayat alias Khizar and others v. The State through Addl. Prosecutor General of Punjab, Lahore and another	Criminal Petition No. 635 of 2026	12.05.2026
5	Jamshaid Rashid v. The State through Prosecutor General, Punjab and another	Criminal Petition No. 706 of 2026	13.05.2026
6	Muhammad Usman Yousaf v. The State through D.A.G. of Pakistan, Lahore and another	Criminal Petition No. 709 of 2026	12.05.2026
7	Hussain Bux v. The State through P.G. Sindh	Criminal Petition No. 653 of 2026	14.05.2026
8	Khuda Bakhsh v. The State	Jail Petition No. 178 of 2025	13.05.2026

9	Ahmed Saeed @ Bharam @ Nagori v. The State and another	Criminal Appeal No. 649/2022 in Jail Petition No. 385/2019	10.03.2026
10	Director General Federal Investigation Agency (FIA), FIA HQ, Islamabad v. Tahir Amin and others; Hammad Ahmad and another; Shahroom Khan and others; Kainat Mumtaz and another; Rimsha Mumtaz and another; Mehwish Mumtaz and another	C.P.L.As. 1048, 1091, 321, 322, 323 and 324 of 2025 (with C.M.A. 914/2025)	22.01.2026
11	Anwar Ali Shah and others v. Sadiq Ali Shah and others	Civil Appeal No. 1410 of 2019 & CMA No. 7122 of 2019	01.04.2026
12	Fazeelat Nasreen and another v. Maqsood Begum and others	C.R.P. No. 410/2023 in Civil Appeal No. 310 of 2021	10.03.2026
13	Nadeem Akhtar (deceased) through LRs. & Others v. Manzoor Ahmad (deceased) through LRs. & Others	C.P.L.A. No. 1086 of 2019	08.04.2026
14	Muhammad Shafi & others v. The State	Criminal Petition No. 847 of 2023	04.05.2026
15	Qaiser Abbas v. The State, Muhammad Siddique v. The State and another	Jail Petition No. 390/2018 and Criminal Petition No. 593-L/2018	15.04.2026
16	Sher Afzal & another v. The State & another	Criminal Misc. Application No. 1292 of 2025 in Criminal Review No. Nil of 2025 arising out of Criminal Appeals No. 229 & 230 of 2021	26.02.2026
17	Government of Pakistan through its Secretary Pakistan Railway, Islamabad and others v. Shahkar Qudos Hashmi	C.P.L.A. No. 3260 of 2024	28.01.2026
18	Javed Akhtar (deceased) through his LRs. v. Zeenia Satti and others; Abdul Karim Niazi v. Mrs. Zeenia Satti and another	Criminal Review Petition No. 56 of 2022 in Criminal Original Petition No. 13 of 2022; and Civil Appeals No. 428 to 430 of 2021	06.03.2026
19	Shahbaz alias Gillo v. The State etc.	Criminal Petition No. 750-L of 2018	16.04.2026
20	Saddam Hussain v. The State	CrI.P.L.A. No. 1132-L of 2018	27.04.2026
21	Wajahat Bibi v. The State	Criminal Petition No. 2073 of 2025	09.04.2026

22	Aftab Zafar v. The State; Hamza Jahangir v. The State through A.G. Islamabad & another; Muhammad Yasir v. The State through A.G. Islamabad & another; Mukhtar Ahmad Khan Rathore v. Asad Mumtaz and another	Jail Petition No. 136 of 2025 and Criminal Petitions No. 575, 679 & 756 of 2025	02.03.2026
23	Baz Khan; Mst. Nusrat Bibi; Mehboob Alam; Abdul Rasheed; Syed Ubaid ur Rehman; Muhammad Ramzan; Muhammad Sohail; Asif Khan; Arshad Subhan Khalil; Lal Akbar v. The State through Prosecutor General Balochistan, etc.	Criminal Appeals No. 43, 359-L, 360-L of 2020, 232 & 233 of 2022; Criminal Petitions No. 140-K of 2018, 1137 of 2021, 593-L of 2017, 939-L & 940-L of 2023; Jail Petitions No. 711 of 2019 & 564 of 2023; and Criminal Misc. Application No. 1334 of 2024	17.02.2026
24	Dr. Hassan Fatima v. Pakistan Red Crescent Society and others	C.P.L.A. No. 769-K of 2025	23.12.2025
25	Amir Waseem v. Provincial Police Officer/Inspector General of Police, Khyber Pakhtunkhwa, Peshawar and others	Civil Petition No. 3691 of 2024	05.05.2026

The selection criteria were based on:

1. Public access to the case records.
2. The importance of Supreme Court discourse in its representativeness.
3. A range of different legal cases are included.
4. A variety of legal case types are featured.
5. Accessibility in Digital format.

All judgments were downloaded from the official website of the Supreme Court of Pakistan and then converted into plain text format for corpus analysis.

3.3. Preprocessing and compilation of corpus

A specially made corpus was built up after data collection. The preprocess stage consisted of:

1. Switching PDF files to text files
2. Removing non-linguistic elements (page numbers, headers, footers)
3. Eliminating formatting inconsistencies
4. Standardizing text encoding

The cleaned texts were joined together into one corpus of 69,629 tokens and 6,536 types. The main dataset for the lexical and statistical analysis was this one.

3.4. Analytical Tools

Two main computational tools were used in this study:

3.4.1. AntConc

The statistics and frequency analysis of the corpus and the exploration of the lexical field were done in AntConc.

1. Total token and type counts
2. Word frequency lists
3. Identification of recurring legal vocabulary
4. Corpus-level lexical patterns

AntConc was especially useful in determining the most common legal terms and for lexical analysis.

3.4.2. MAT (Multidimensional Analysis Tool)

MAT was used as a Part-of-Speech (POS) tagging tool. It was used to decide the grammatical category of lexical items, such as:

1. Nouns (NN, NNP, NNS, NNPS)
2. Verbs (VB, VBD, VBG, VBN, VBP, VBZ)
3. Use of adjectives (JJ, JJR, JJS)
4. The use of adverbs (RB, RBR, RBS)

The classification allowed the calculation of the lexical density and the analysis of the linguistic structure of legal writing.

4. Analysis and discussion of data

The present study uses corpus-based approach to study the lexical density in Pakistani Supreme Court judgements. The corpus comprised a set of 25 Supreme Court decisions that were gathered from official judicial documents. The corpus is preprocessed and cleaned, resulting in 69,626 tokens and 6,535 types. The Multidimensional Analysis Tool (MAT) for Part-of-Speech (POS) tagging was used for analyzing the corpus, and AntConc was used for corpus statistics and frequency analysis. This integrated approach allowed for the structured analysis of lexical structures and informativeness in legal texts.

4.1. Corpus Characteristics

This corpus includes a variety of constitutional, civil, and criminal decisions, and it spans a wide range of legal issues that are significant in modern Pakistan. The corpus is large enough (69,626 tokens) to reveal common patterns and structures of the language. There are 6535 types of words, which reveal a significant amount of lexical variation, that is, variation in the words used to address various legal issues in Supreme Court decisions.

Yet, even in this lexical variety, there are institutional and procedural patterns in judicial writing. Legal communication has its own special nature and the specific use of legal terms leads to certain lexical features that make legal text stand out from other genres of written communication.

4.2. Part-of-Speech Analysis

The present corpus was tagged for POS by using MAT to find the distribution of the lexical categories. The analysis was conducted for four important lexical classes: nouns, verbs, adjectives and adverbs. The tagging results showed that the largest lexical category in the corpus is the nouns followed by verbs, adjectives and adverbs. The high nominalization level is striking because it suggests that nominal expressions play a prominent part in legal discourse and represent legal entities, institutional actors, legal procedures and abstract concepts of law. This is a feature of formal written discourse, especially legal, in which the meanings of longer sentences are usually included in noun phrases. Verbs were identified as the second most common lexical category, used mainly to provide information regarding judicial acts, judicial decisions, and the processes. Other, less common, words, such as adjectives and adverbs, are used to qualify legal concepts and to modify judicial reasoning, in ways that can lead to precision and specificity. The general distribution of POS shows that the Supreme Court decisions are conceptually and institutionally rather than narratively oriented. This inclination has a strong impact on the informativeness of legal discourse.

4.3. Lexical Density Analysis

The main purpose of this study was to calculate the lexical density of Supreme Court verdicts of Pakistan. Lexical items were identified and lexical density was calculated as a ratio of lexical items to the total number of tokens in the corpus. In conformity with Halliday's framework, nouns, verbs, adjectives and adverbs, identified in the process of POS tagging, were used as lexical items.

The lexical density was calculated with the formula:

$$\text{Lexical Density} = \left(\frac{\text{Lexical Words (Nouns + Verbs + Adjectives + Adverbs)}}{\text{Total Tokens}} \right) \times 100$$

There were 41,002 lexical items in total in the corpus. The lexical density was calculated after dividing the number of tokens (69,626) with the number of the language items used (52,458), giving a lexical density of 58.9%.

This is a good result since nearly sixty percent of the corpus is made up of content bearing lexical items. Such a percentage indicates that there is a high level of information density within the text, as well as that the language of judges is also highly dependent on lexical resources in terms of meaning construction. The results indicate that Supreme Court decisions do not use a large number of grammatical devices to elaborate, but rather a dense lexical structure to transmit complex legal reasoning.

4.4. Lexical patterns in legal discourse

Through the application of AntConc, Frequency analysis was used to identify the lexical patterns that are present in the text, which are common features of legal writing. Constitutional terminology; evidential procedures; judicial decision making; law; legal structures; and legal practice were the most abundant terms in the corpus. From the

analysis of these lexical patterns, legal texts rely on a specific group of specific terms to present a coherent and precise picture of legal interpretation. Lexical density also is directly affected by the repetition of legal terms since many legal terms abound with institutional meanings and complex concepts. Texts of judges are not usually long explanations of legal concepts but instead use technical language to convey much information in a small linguistic area. Another phenomenon in the corpus that is of note is the tendency to nominalize verbs. In law, abstract nouns are often used instead of action verbs. This linguistic approach allows legal discourse to represent processes, procedures and legal relations as conceptual entities. Thus, more information is conveyed in the judicial language in a more compact and formal way.

4.5. Analytical findings are discussed as the work proceeds

The analysis shows that Pakistani Supreme Court rulings have a high lexical concentration and informational density. Judges' speech is mainly nominal in structure, given the fact that both the number of nouns and the wealth of legal jargon suggest that the structure of judicial speech is largely nominal. With such a linguistic structure, legal texts can be more efficient in communicating difficult legal information, while also being precise and authoritative.

In addition, a lexical density score of 58.9% indicates that there is a significant amount of semantic content in legal writing. The results corroborate the view of judicial language as a highly specialized register, which aims at conveying legal reasoning in a compact lexical structure. The large number of content words is due to the purpose of legal communication, which demands accuracy and interpretive clarity. The overall result of the corpus-based analysis points to the fact that Pakistani Supreme Court judgments are written linguistic discourses that are extremely lexicalized. The language they use has a high density of information, a high degree of nominalization, and a great deal of legal jargon, which makes legal language unique.

5. Results, Findings, and Discussion

5.1. Corpus Statistics

The current work investigated a corpus of 25 judgements extracted from the Supreme Court of Pakistan and organized it into a structured dataset for analysis by corpus linguistic methods. The total size of the corpus was 69,626 tokens and 6,535-word types, which led us to believe that there was a fair amount of variation in the number of words used in the judicial texts. This is a variation of the Supreme Court's discourse characteristics, which are formal, institutional, and legally specialized. MAT (a tool for Part-of-Speech tags) and AntConc (a tool for frequency analysis and lexical extraction) were used to process the corpus.

Table: 2

Measure	Value
Number of Judgments	25
Total Tokens	69,626
Total Types	6,535
Total Lexical Items	41,002
Lexical Density	58.9%

5.2. POS Tag Distribution

The POS analysis revealed that there is a strong dominance of the lexical category, notably nouns, followed by verbs, adjectives and adverbs.

The distribution of detail is as follows:

Table: 3

Lexical category	Code	Tokens	File hits	File types	Frequency list
Noun	NN	69629	12771	6536	12771
	NNP	-	5352	-	5352
	NNS	-	3189	-	3189
	NNPS	-	72	-	72
Verb	VB	69629	1822	6536	1822
	VBD	-	2133	-	2133
	VBG	-	1205	-	1205
	VCN	-	3711	-	3711
	VBP	-	583	-	583
	VBZ	-	1959	-	1959
Adverb	RB	69629	2850	6536	2850
	RBR	-	38	-	38
	RBS	-	8	-	8
Adjective	JJ	69629	5215	6536	5215
	JJR	-	73	-	73
	JJS	-	21	-	21

The dominance of nouns clearly shows that the judgments of the Supreme Court in Pakistan are heavily based on nominal structure instead of verbal structures or narratives. This shows that legal talk is institutionalized and that meaning is condensed in abstract legal entities and concepts.

5.3. Lexical Density Results

The standard formula in corpus linguistics was used to determine the lexical density (lexical items/total tokens):

$$\text{Lexical Density} = \left(\frac{\text{Lexical Words (Nouns + Verbs + Adjectives + Adverbs)}}{\text{Total Tokens}} \right) \times 100$$

Substituting values:

Total lexical words = 41,002

Total tokens = 69,626

Lexical Density=4100269626×100

Final Score: Lexical Density = 58.9%

This means that over half of the corpus is made up of content words and that the informational compression of judicial writing is high.

5.4. Frequency Analysis

The word frequency analysis from AntConc showed high occurrence of legal and institutional terms. The most common words are:

Table: 4

Sr. No	Legal term	Frequency
1	Court	543
2	Petitioner	274
3	Evidence	207
4	Order	176
5	Section	164
6	Accused	135
7	Jurisdiction	124
8	Judgment	122
9	Record	117
10	Rule	114
11	Trial	99
12	Appeal	92
13	Sentence	70
14	Bail	67
15	Relief	43
16	Tribunal	37
17	Inquiry	37
18	Conviction	32
19	Constitution	32
20	Maintainability	14

Judicial discourse is very formulaic and domain specific, as indicated by the repetition of institutional terms. These words have rich legal connotations, and they are key components in the development of judicial reasoning.

5.5. Interpretation of Findings

The findings show that Pakistani Supreme Court decisions have a high level of lexical density (58.9%), indicating a high degree of compression of information. M. A. K Halliday's framework suggests that written institutional discourse tends to have a higher lexical density since meaning is contained in lexical items rather than being distributed

across grammatical structures. The high frequency of nouns (21,384 occurrences) indicates a high rate of nominalization, which is characteristic of legal text. The encoding of legal reasoning is mainly accomplished by employing abstractions like jurisdiction, conviction, and constitution; these replace dynamic verbal processes. This change is a major element of text compressiveness and informativeness. The proportion of verbs is relatively low, suggesting that judicial texts are more conceptual than are narrative. In common with other kinds of discourse, legal discourse builds meaning through abstract legal categories and institutional references rather than by describing actions. Less common but still used are the adjectives and adverbs that give precision and qualifications of legal meaning. In practice, however, their function is subordinate to that of nouns and verbs, and this reinforces the idea that legal discourse is noun-based.

6. Discussion of Lexical Density in Legal Discourse

The Supreme Court judgments in Pakistan have a mean Lexical density value of 58.9%, which falls in the high density written institutional texts category. This can be compared with other corpus-based studies in the field of written language in the legal domain, which have revealed that legal writing is one of the most information-intensive, being a domain of writing. The present study, however, extends the knowledge by empirically supporting it by using the Pakistani judicial corpus as a field not utilized when it comes to corpus-based studies. The results reveal that the general pattern of legal language in Pakistan is in line with the global trends of legal language, which uses specialized words and nominalization. The lexical density is also high, and it also has an impact on readability. There is a high load of information for non-specialist readers in the dense lexical structure, where meaning is tightly packed into technical vocabulary and not broken down into meaning by way of grammatical explanations. This helps to strengthen the notion of precision and institutional control in the law, more than on the accessible.

6.1. Role of Corpus Tools in Analysis

The use of AntConc and MAT was indispensable for this study. Systematic POS tagging was made possible with MAT, and frequency extraction and corpus statistics was made possible by AntConc. These tools gave structural and quantitative accuracy in the analysis. The use of methodological triangulation improves the validity of the findings, and shows that corpus-based approaches are successful tools for legal discourse analysis. It also looks at how computation can be used to reveal patterns in the language which may not be evident from a handwritten analysis.

6.2. Key findings summary

The most important findings of the study are summarized below:

1. The lexical density of Pakistani Supreme Court Judgments is 58.9%.
2. High nominalization as a group of nouns as the most dominant lexical category.
3. Legalese is very specific and repetitious.

4. Verbs, adjectives and adverbs in meaning construction are subordinate.
5. Judges write in an institutional, formal, and dense manner.

Overall, the results reaffirm that Supreme Court decisions in Pakistan are dense, complex texts in terms of language. Based on this, it can be inferred that Supreme Court's judgments are generally complex and dense language texts. Lexical density, noun dominance, and legal jargon illustrate that judicial writing is not meant to be easy to read, comprehend, or understand. The results are indeed very positive, as they provide support to the Theory of SFL and the idea that legal language is a very compact mode of written institutional communication.

7. Conclusion

The present study aimed to investigate the lexical density of Pakistani Supreme Court judgments with the help of Systemic Functional Linguistics (SFL) approach and corpus-based method. This study employed MAT for POS tagging, AntConc for frequency and corpus analysis and produced a comprehensive linguistic description of the construction of meaning in judicial discourse with the analysis of a corpus of 25 judgments.

8. Key Findings of the Study

This research has come up with a main result that shows that the Supreme Court of Pakistan judgments are highly lexical dense having 58.9% informational compression. This indicates that almost half of the corpus is made up of lexical (content) words, which is indicative of dense legal discourse that is rich in information. The result of the POS analysis revealed that all the nouns were more dominant than the verbs, adjectives and adverbs. In the judicial language, it is remarkable to observe the use of nouns to the extent of using abstract legal terms such as constitution, conviction and jurisdiction. This nominality is used to a great extent in the complexity and compactness of legal texts. The frequency analysis also showed the high degree of institutionalization and specialization of the language found in the law, such as the occurrences of the words court, petitioner, evidence and judgment. The Supreme Court decisions have a procedural and authoritative nature, as indicated in these lexical items.

9. Theoretical Implications

The results are very supportive of the theoretical statements made by M. A. K. Halliday on lexical density in written institutional discourse. The structure of Pakistani Supreme Court judgments is a clear illustration that the lexically dense nature of written language is more dominant than spoken language as Halliday argues. The study also establishes that the use of legal discourse follows lexical compression with complex legal reasoning being represented in dense noun phrases and not in elaborated grammar. The study also demonstrates the applicability of SFL to the study of institutional texts.

9.1. Linguistic and Practical Implications

On the linguistic level, the study reveals a rate of occurrence of nominal, abstract and highly lexicalized legal discourse. That is complexity of structure that aids in the accuracy and strength of judicial language but also impedes the comprehension of it by non-specialist readers. The results show that language use in the legal sphere is challenging for the public due to its high lexical density and the usage of technical terms. This reflects the significance of lawyers focusing on the legal readability, and potential legal simplifications, in legal communication.

9.2. Methodological Contribution

The study also contributes to the methodology used as it demonstrates that the use of the combination of Corpus tools (AntConc and MAT) in linguistic analysis is effective. For systematic POS tagging, MAT has been used and for reliable frequency extraction and corpus analysis, AntConc has been used. These tools were incorporated into a complete and automated system for the lexical density analysis of legal texts. These tools were combined in order to achieve a complete and repeatable system for the lexical density analysis of legal texts.

10. Limitations of the Study

In the future, the database might be expanded to include more laws and analyzed using fewer cases from lower courts or international legal systems to obtain more general insights into the variation of legal discourse.

11. Recommendations for Future Research

1. This comprises a comparative study of the Pakistani and international legal discourse.
2. Measure of Lexical density and Readability assessment.
3. Legal speech and discourse, e.g., courtroom speech.
4. The employment of state-of-the-art computational approaches, such as tagging complex systems by employing machine learning techniques.
5. English Japanese legal translation of documents.

Finally, this study confirms that the Pakistani Supreme Court judgments are indeed very lexical dense, nominal and institutionally specialized texts. Results suggest that the lexical items used in the production of meaning, authority and precision in legal discourse are strong characteristics. This research combines corpus linguistics and Systemic Functional Linguistics as one of the contributions to the studies of legal language from the empirical and theoretical aspects. It also highlights the importance of the corpus tools in showing subconscious patterns in legal language and offers a good foundation for future research on legal language.

References

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

- Biel, Ł. (2010). Legal terminology in English: A corpus-based study of judicial language. *Journal of Specialized Translation*, 14, 1–20.
- Gales, T. (2009). Review of patterns of linguistic variation in American legal English. *Corpus Linguistics and Linguistic Theory*.
- Gianluca, P. (2019). Corpus methods in legal translation studies. In *Corpus-based approaches to translation and interpreting* (pp. 23–45). Routledge. <https://doi.org/10.4324/9781351031226-2>
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Edward Arnold.
- Halliday, M. A. K., & Matthiessen, C. (2014). *Halliday's introduction to functional grammar* (4th ed.). Routledge.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
- Stubbs, M. (1996). *Text and corpus analysis*. Blackwell.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins.
- Zhang, X. (2022). Corpus-based analysis of legal discourse and readability. *Humanities and Social Sciences Communications*, 9, 1–12. <https://www.nature.com/articles/s41599-022-01485-x>