

**Liberal Journal of Language & Literature Review**

**Print ISSN: 3006-5887**

**Online ISSN: 3006-5895**

**<https://llrjournal.com/index.php/11>**

**A Corpus-Based Study of Lexico-Grammatical Variation in  
Khyber Pakhtunkhwa Enacted Laws**



<sup>1</sup>Syeda Aqsa Gilani<sup>1</sup>

<sup>\*2</sup>Dr. Syed Shujaat Ali

<sup>3</sup>Maham Khan

<sup>1</sup>Graduate from the Department of English, Kohat University of Science & Technology, Kohat. Email: [aqsagilani24@gmail.com](mailto:aqsagilani24@gmail.com)

<sup>2</sup>Assistant Professor, Department of English, Kohat University of Science & Technology, Kohat. Corresponding Author Email: [s\\_shojaat\\_ali@yahoo.com](mailto:s_shojaat_ali@yahoo.com) & [shujaatali@kust.edu.pk](mailto:shujaatali@kust.edu.pk)

<sup>3</sup>Graduate from the Department of English, Kohat University of Science & Technology, Kohat. Email: [mahamkhankhattak@gmail.com](mailto:mahamkhankhattak@gmail.com)

**Abstract**

In this study, lexico-grammatical variations in a corpus of enacted laws in the province of Khyber Pakhtunkhwa were investigated using the Multidimensional Analysis framework introduced by Biber in his work "Variation across Speech and Writing." The research was conducted over a sample of fourteen categories, comprising seventy enacted laws, including university, finance, ad hoc, benefit and allowance, court, fund, health, land and property, law, loan, statutory bodies, revenue and tax, women, and repeal acts. Using the multidimensional analysis framework, legal texts of these categories of enacted laws were analyzed through six dimensions of linguistic variance. This study highlighted the differences in lexical and grammatical structures embedded within the province's enacted laws. The analysis showed that these fourteen categories significantly differed from one another along all six dimensions of the 1988 model of variation. The study also found that the textual dimensions of these categories are non-narrative, persuasive, situation-independent, abstract, and informative; these variations were not random but rather influenced by a variety of factors, such as context, goal, register, and audience. The findings of this study provide important information about the linguistic variation in Khyber Pakhtunkhwa's enactment discourse and aid in the development of a corpus-focused approach to the analysis of legal language. The findings of this study can aid in the improvement of communication among Khyber Pakhtunkhwa's linguists and legal professionals. It will enhance their ability to communicate and comprehend the usage of language elements in the legal register, or legislated laws. Depending on their requirements and areas of interest, future researchers may decide to focus on one or more dimensions.

**Keywords:** Lexico-grammatical variation, Khyber Pakhtunkhwa's enacted laws, multidimensional analysis.

**Introduction**

Lexico-grammatical variation is a widespread phenomenon in human language. It is a fascinating and complex phenomenon that refers to the systematic and rule-governed differences in word and grammatical usage in various linguistic contexts and registers.

In various communication circumstances, people use language differently, making distinct lexico-grammatical options. The term that is used to define variation in language is called "register". (Moreno, 2006) states that register is the appropriate way of presenting a message depending on the social environment where that message is delivered.

Legal language, as a technical language, is a specific type of language that is used in legal texts including court opinions, statutes, regulations, contracts, and other legal documents. It is distinguished from other written and spoken languages due to its formal and technical vocabulary, complex grammar, specific terminology, and style (Asghar, 2014). (Tiersma, 1999) defines legal language as a particular speaking and writing style that is designed and developed, especially for the field of law. In linguistics, lexicogrammar or lexico-grammar refers to the level of language that combines syntax (grammar) and lexis (vocabulary) into one. At this level, it accounts for the words and grammatical structures being dependent on each other, in which one level interfaces with the other instead of being independent (Sardinha, 2019). Lexico-grammatical variation is the term used to describe differences in how words and grammar are used within or across different groups of speakers or in various situations. This can include variations in word orders, lexicon, verb tenses, and other grammatical elements.

In the context of Pakistan, a multidimensional analysis model was used to investigate a variety of genres. (Shakir, 2014) examined the language of prospectuses from Pakistan, the United Kingdom, and India along Biber's dimension of "involved versus informative production" (Biber, 1988) . (Asim & Alvi, 2016) used multidimensional analysis to examine editorials in Pakistani English newspapers. (Begum, 2016) also conducted a study to investigate university students' academic writing within the Pakistani environment. Several studies have been conducted on various registers for exploration of linguistic specifics and legal register has also been studied generally and comparatively, yet in the context of Khyber Pakhtunkhwa no study has been conducted in this regard.

The present study aims to investigate and explore the multidimensional patterns of lexico-grammatical variation exist among various categories of Khyber Pakhtunkhwa enacted laws.

### **Literature Review**

Legal language is a register of a language that is characterized by various features, such as precision, formality, and complexity. It is used in a variety of legal contexts, including courts, legislatures, and law firms (Shuy, 2003). Legal language is a formalized discourse that requires a special set of skills for its interpretation and production (Gibbons, 2003).

In the early days of law, legal language was often quite simple and straightforward (Smith, 2019). This was because the legislation was less complicated and there was less need for precise language. However, as the law evolved and developed, legal language became more complicated and specialized because of a variety of circumstances, such as an increasing need for precision in the law, the influence of Latin and other languages on the law, and the development of the legal profession.

Latin was the dominant legal language in Europe during the early Middle Ages. This was because Latin was the language of the Church and colleges, as well as the language of Roman law, which was influential in many European legal systems (Shali, 2018). In the 12th century, there was a revival of interest in Roman law, which resulted in an influx of Latin legal words into European languages. The first law schools in Europe were founded in the 13th century, and these schools taught law in Latin. This contributed to the further standardization of legal terminology throughout Europe.

Vernacular languages began to be used more regularly in legal writing in the 14th and 15th centuries. Local languages eventually replaced Latin as the language of law. However, Latin was still used in many legal contexts, such as in international treaties and in Church law (Haugen, 1972). The movement to codify the law was becoming stronger in the 16<sup>th</sup> and 17<sup>th</sup> centuries. National legal codes were created as a result, and these were written in vernacular languages. There was a further standardization of legal language inside countries in the 18th and 19th centuries. This was due to the rise of nation-states and the requirement for a common legal language. There has been an increasing globalization of the legal profession in the 20<sup>th</sup> and 21<sup>st</sup> centuries. As a result, there is a need for a standard legal language that can be used by lawyers and judges from different countries.

Today, Legal terminology is utilized in a wide range of circumstances, such as statutes, regulations, contracts, court opinions, and legal briefs. Non-lawyers or the common people may find legal terminology confusing and difficult to understand due to its complex vocabulary. Several studies have been conducted on legal language. (Gustafsson, 1975) studied binominal and multinomial expressions in legal writing. (Finegan, 1982) focused on forms and functions in testament language.

In another study (Fakuade & Sharndama, 2012) looked at how coherent devices were used in four popular and professional legal texts. The 1976 model by Halliday and Hassan was employed for that purpose. The analysis showed that referring expressions and coordinators are commonly found in both types of texts. In addition to examining linguistic characteristics, (Maazi, 2007) examined forty judgments made in Europe and English/Irish.

Using multidimensional analysis, the study (Asghar & Mahmood, 2018) examined linguistic variance within Pakistani Legal English genres. In Pakistan, according to (Shakir, 2013), linguistic variety was found in print advertising in Pakistani media, demonstrating the use of multidimensional analysis to register investigations that go beyond English legal registers. In addition to the above works, the study of lexico-grammatical variation in legal texts has been widely explored to understand its significance in linguistic analysis and pedagogical implications. For instance, Ali et al. (2019a) investigated conversation analysis, focusing on linguistic and communication principles, which can provide insights into discourse patterns observed in legislative documents. Additionally, exploring nonverbal and paralinguistic features in communication (Ali et al., 2019c) may shed light on the implicit meanings embedded in legal discourse.

The role of idiomatic expressions, as discussed by Ali et al. (2019b), highlights the need for precise word choice to ensure clarity in laws and legal processes. Such precision can also align with the findings on semantic density analysis in sacred texts (Ishtiaq et al., 2021a), demonstrating how lexical variations affect comprehension. Furthermore, punctuation and structural accuracy, as observed in writing studies (Ali et al., 2020a), play a vital role in interpreting legal texts. Similarly, variations in transliteration and pronunciation errors (Ishtiaq et al. 2022b) could influence the drafting of multilingual laws, particularly in regions like Khyber

Pakhtunkhwa.

The syntactic structure of languages, including parallels in English, Urdu, and Pashto (Ishtiaq et al. 2022c; Ishtiaq & Gill, 2024), offers a methodological approach to analyze the linguistic complexities in enacted laws. Lastly, discourse analysis of digital and legislative texts (Gill et al. 2025) could provide a feminist and critical perspective on legal representation and inclusivity in legislative discourse.

### **Research Methodology**

#### **Research Design**

The current study is a corpus-based study in which the quantitative method research design is used to examine Khyber Pakhtunkhwa enacted laws by using the (Biber, 1988) multidimensional analysis approach. It quantifies how frequently specific lexical and grammatical patterns and language features occur together. This study's methodology helps to explore lexico-grammatical variances among distinct categories of Khyber Pakhtunkhwa enacted laws.

#### **Data Collection**

The present data is gathered through a primary source in which the researcher self-collected and downloaded the data from the official website of KP code in pdf form. The nature of data collected from the website is an official legal document. Data is ethically valid because it is not a secret document. It is publicly available and is open to all.

#### **Compilation of Corpus**

For the data compilation, the researcher gathered the material from the official website of the KP code (<http://kpcode.kp.gov.pk/>). The research included KP enacted laws, the laws of fourteen different theme-based categories: University, Finance, Ad-hoc, Benefits and allowance, Court, Fund, Health, Land and property, law, loan, statutory bodies, revenue and tax, women, and repeal acts. The rationale for choosing these categories is the nature of the data. The researcher selects those acts that are equivalent in terms of the equal number of pages or quantity of words, in order to establish consistency. The current study's data-collecting tool was a corpus. A sample corpus was prepared for this current research project, consisting of a large number of texts carefully gathered and obtained in PDF format from the KP code website, and then the corpus was created in Word document format. The data was downloaded

from the KP code website and turned into plain text files because MAT only reads text files. A corpus of 70 enacted laws was collected from the official website of the KP code and analyzed using a corpus linguistics tool. The table 1 describes the laws of fourteen different theme-based categories of KP enacted laws.

**Table 1 : Fourteen Categories of KP Enacted Law**

S. No.	Categories	No. of Acts	No. of Pages
1.	University	2	38
2.	Finance	4	38
3.	Ad hoc	8	38
4.	Benefits and allowance,	4	38
5.	Court	3	38
6.	Fund	7	38
7.	Health	3	38
8.	Land and property	6	38
9.	Law	4	38
10.	Loan	4	38
11.	statutory bodies	2	38
12.	revenue and tax	4	38
13.	Women	5	38
14.	repeal act	14	38

#### **Data Analysis**

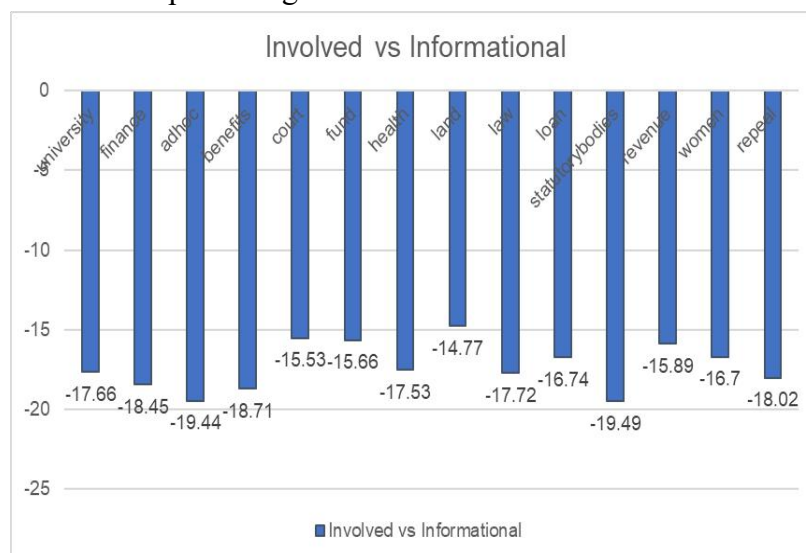
For the analysis, data is tagged and analyzed by using a corpus tool called Multidimensional Analysis Tagger, or MAT (v1.3.3). The Multidimensional Analysis Tagger (MAT) of Biber is a corpus linguistic tool that tags texts automatically and classifies the features of a language. The data were analyzed using Biber's (1988) MD analysis model. It is built on 67 language characteristics and six dimensions i.e., “Involved vs. Informational Production”, “Narrative vs. Non- Narrative Concerns”, “Explicit vs. Situation Dependent Reference”, “Overt Expression of Argumentation/Persuasion”, and “Abstract vs. Non-Abstract Style” (Biber, 1988). Every dimension contains a collection of language features that may be utilized to examine a specific text. There exists a value in every dimension, both negative and positive; this indicates whether the characteristics are available or not.

## Result and Discussion

### Interpretation of Dimension 1

“Informational and involved discourse” make up the first and foremost dimension, which is a continuum with positive and negative values. Biber's 1992 MD model states that on Dimension 1, a text having a high frequency of “*private verbs, that deletion, second-person pronouns, hedges, discourse particles, wh-questions, wh-clauses, etc.,*” (traits with positive scores) indicates that involved discourse is present, which indicates that text is engaging and communicative in nature for instance, as conversation. In contrast, a text having a high frequency of *nouns, prepositions, attributive adjectives, etc* (features with negative scores) indicates that informational discourse is predominant, as in the case of academic prose (Biber, 1988).

The result in Figure 1 reveals variations in university, finance, ad hoc, benefits, court, fund, health, land, law, loan, statutory bodies, revenue, women, and repeal. The result of Statutory bodies shows that it is most informational than other categories. The mean score of statutory bodies is (-19.49) which indicates that it is slightly more informational than the ad hoc category with the mean score (-19.44). On the other hand, the land category, having a mean value of (-14.77), has a lower level of informativeness while producing discourse.

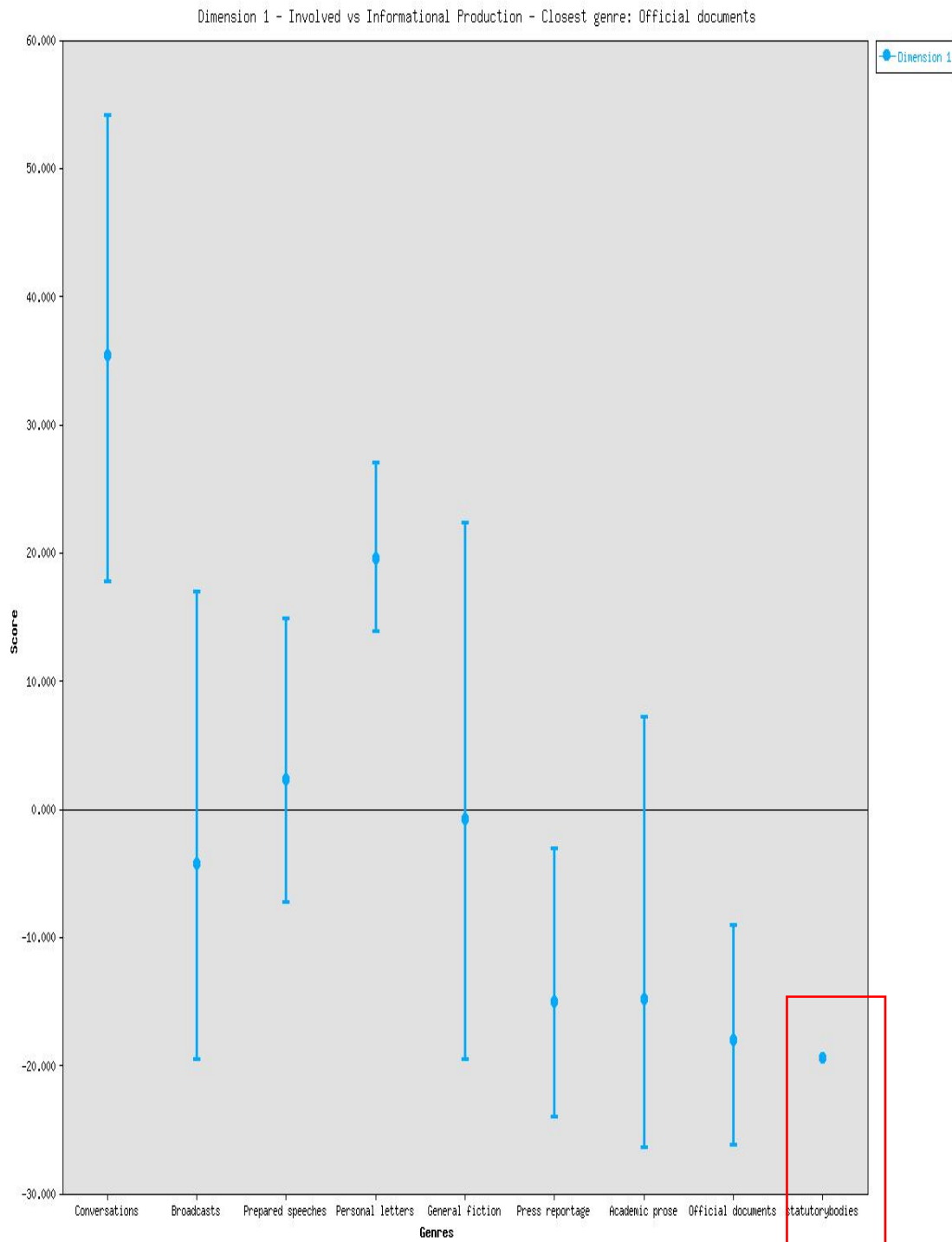


**Figure 1: Involved vs Informational Discourse**

The graphic representation displays the location of the data and offers several statistics-related details. Our corpus data's location has low scores, as indicated by the graphic representation for dimension 1, indicating that text of Statutory bodies

category is informational.

Dimension 1	Mean Score of Statutory bodies Category
	-19.49

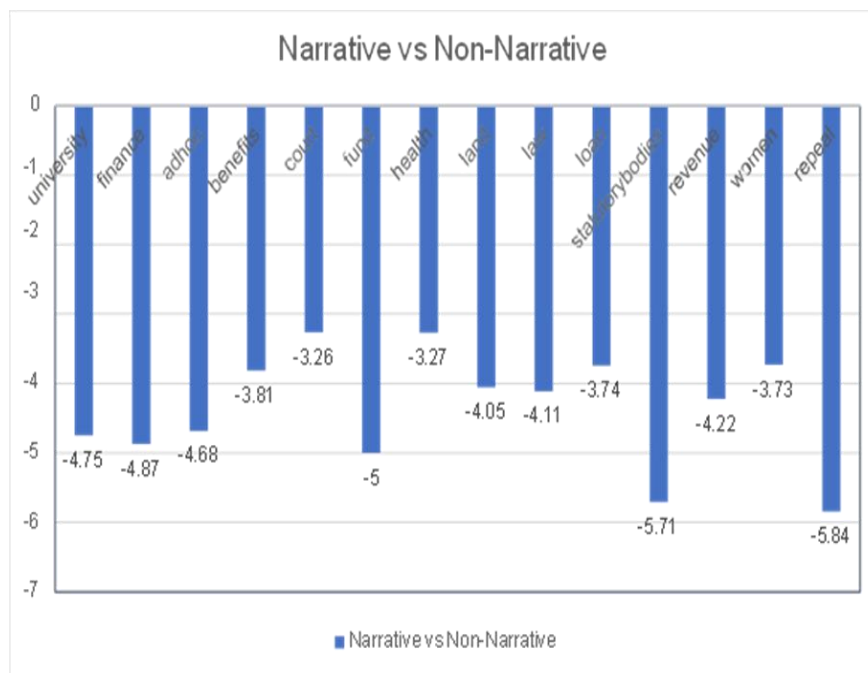


**Figure 2: Graphical Representation of Statutory Bodies Category**

## **Interpretation of Dimension 2**

The second dimension is titled "Narrative vs. Non-narrative Concerns". Positive values signify that particular variety is associated with “narrative discourse”, while negative values show a non-narrative focus for the text. Texts with high positive scores often use “*past tense verbs, third-person pronouns, verb- perfect aspect, and public verbs*. In contrast, texts with negative scores frequently use “*present tense verbs, pronouns it, that deletion, and place adverbial*” (Biber, 1988).

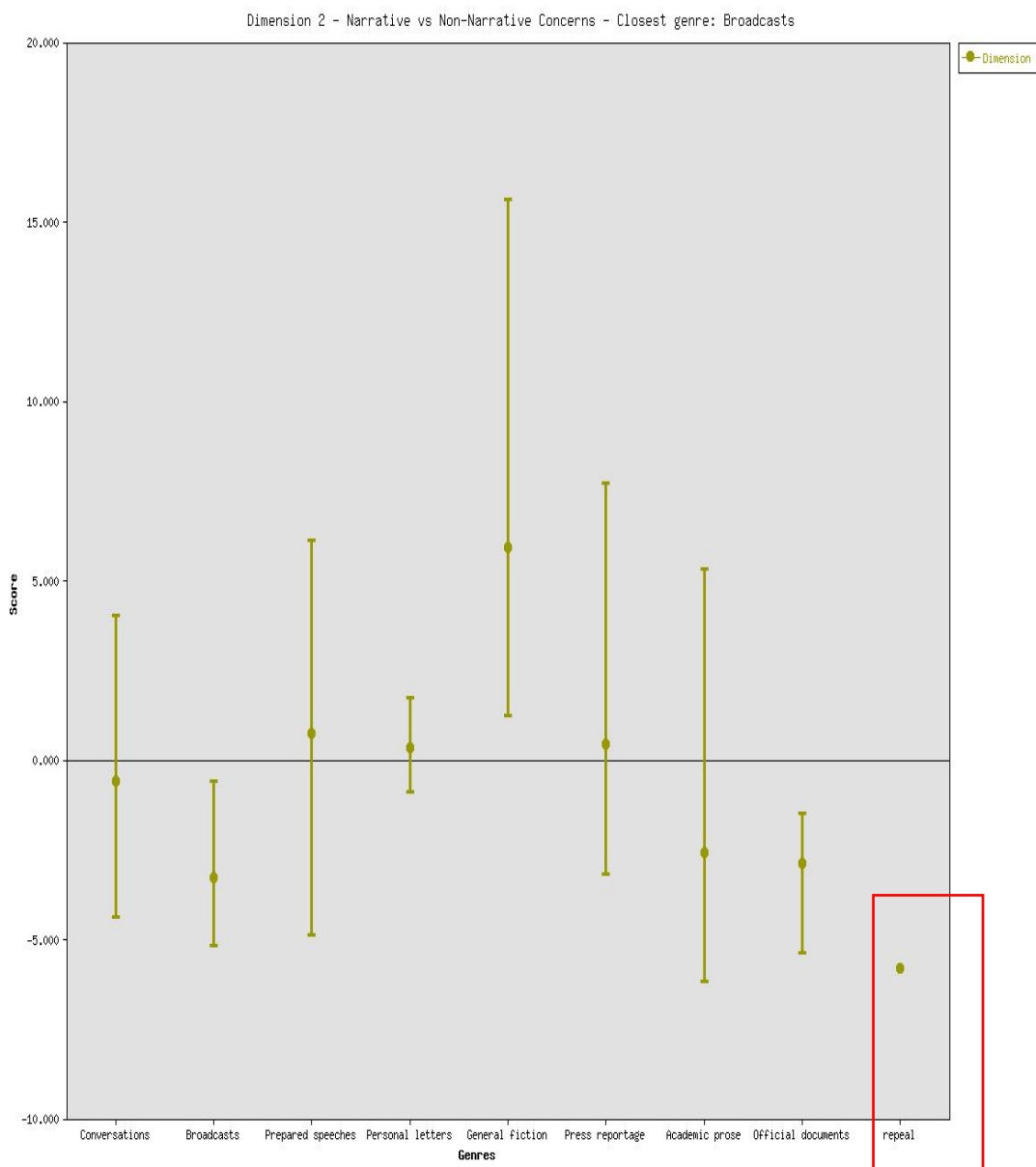
On D2, the result of the Repeal category indicates that it is more non-narrative than other categories. Repeal category, with the mean score (-5.84) produces more non-narrative discourse than statutory bodies category (-5.71). In contrast, court and health categories uses less non-narrative discourse as compared to other categories, with a mean score of (-3.26) and (-3.27). The result of loan and women categories almost show the same trend with mean value (-3.74) and (-3.73).



**Figure 3: Narrative vs Non-Narrative**

The graphic representation displays the location of the data and offers several statistics- related details. Our corpus data's location has low scores, as indicated by the graphic representation for dimension 2, indicating that text of Repeal category is non-narrative.

Dimension 2	Mean Score of Repeal Category
	-5.84



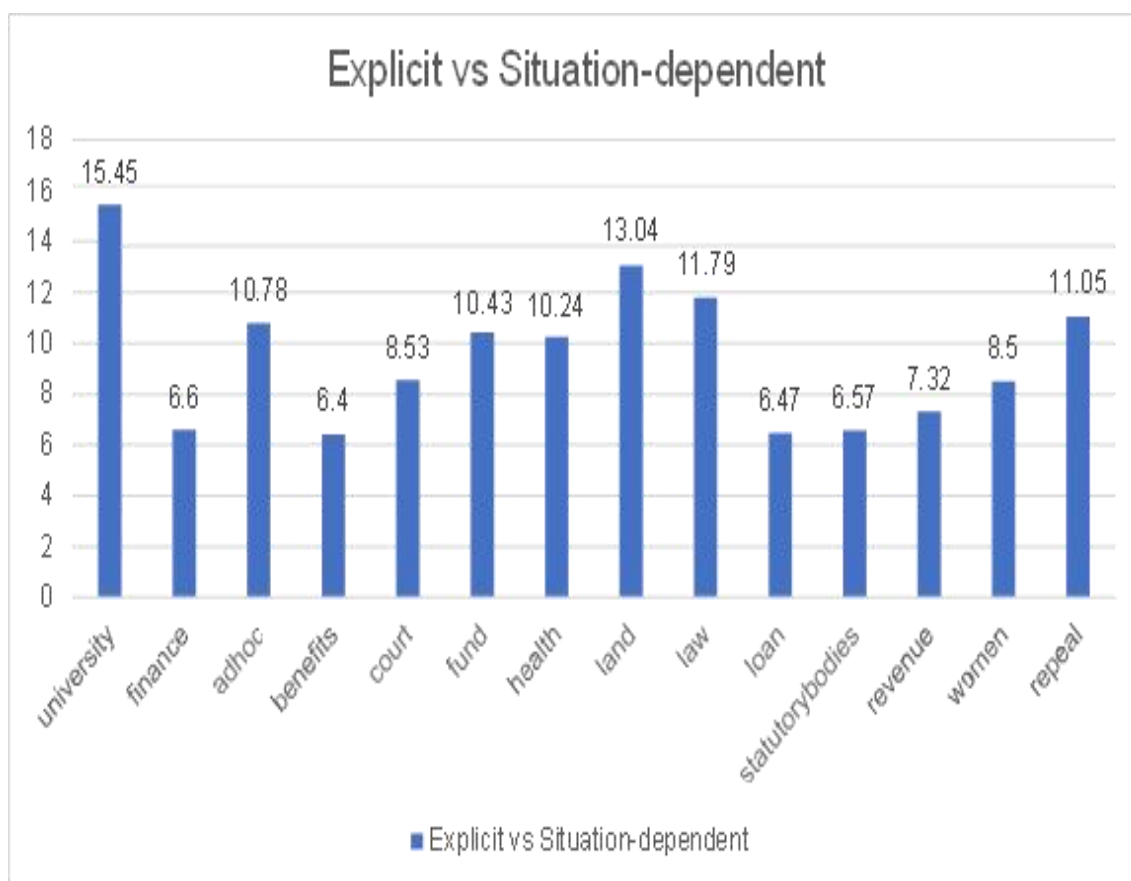
**Figure 4: Graphical Representation of Repeal Category**

### **Interpretation of Dimension 3**

On the third dimension, higher scores suggest explicit discourse, whereas lower values emphasize situation-dependent discourse. “Situation-dependent reference” features include” time adverbials, general adverbs, place adverbials. Some of the explicit reference that contribute to fluent language include relative clauses with wh-

words in object position, relative clauses with wh- words in subject position, nominalizations, phrasal coordination and pied-piping relatives.

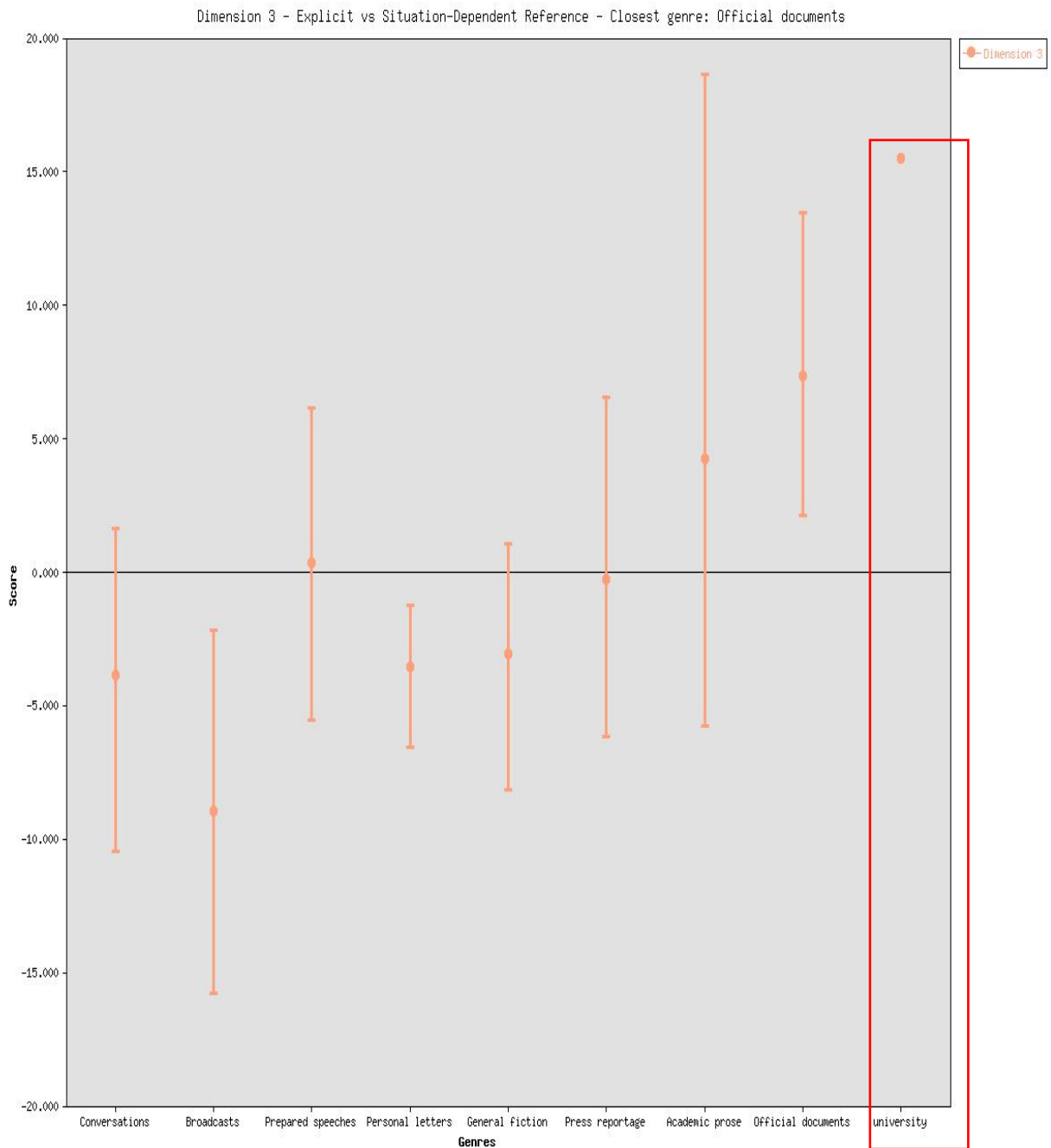
On this dimension, figure shows that university category with the mean value (15.45) is more explicit than the other remaining categories. The results of benefits (6.4) and loan (6.47) categories show almost same trend. Similarly, the results of Court (8.53) and women (8.5) categories also show same result.



**Figure 5: Explicit vs Situation-Dependent**

The graphic representation displays the location of the data and offers several statistics- related details. Our corpus data's location has high scores, as indicated by the graphic representation for dimension 3, indicating that text of University category is explicitly elaborated.

Dimension 3	Mean Score of University Category
	15.45



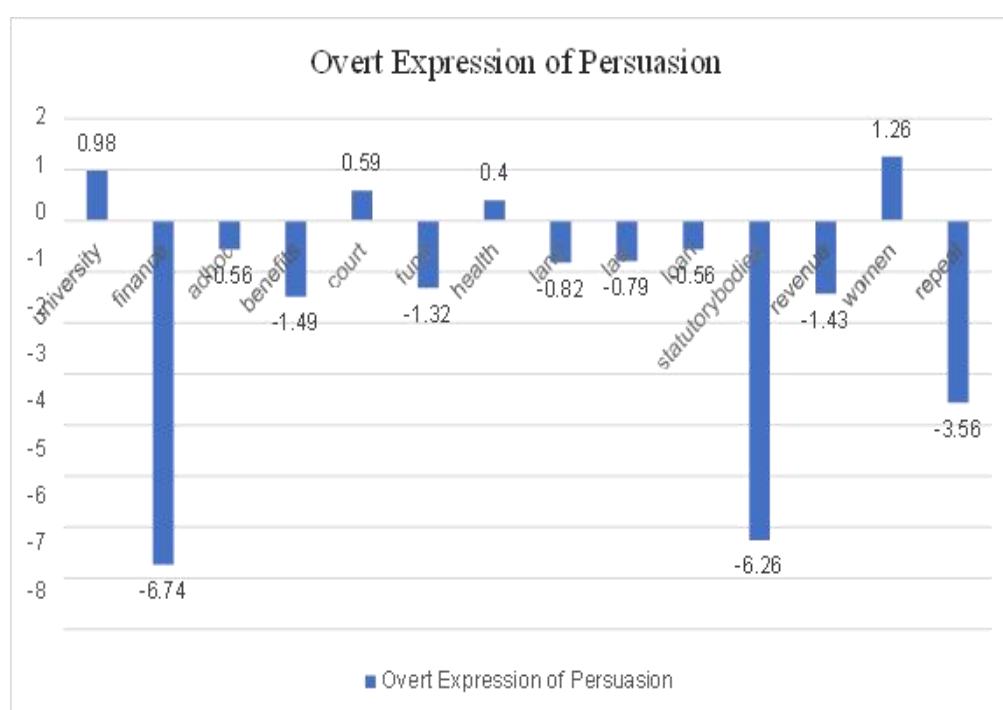
**Figure 6: Graphical Representation of University Category**

#### **Interpretation of Dimension 4**

The term "overt expression of argumentation/persuasion" refers to dimension four.

The research conducted by Biber in 1992 found that only positive weight features are present in dimension four. Argumentative discourse on this dimension's positive polarity is produced by linguistic elements such as “infinitives, split auxiliaries, prediction modals, conditional subordinations, necessity modals and suasive verbs”.

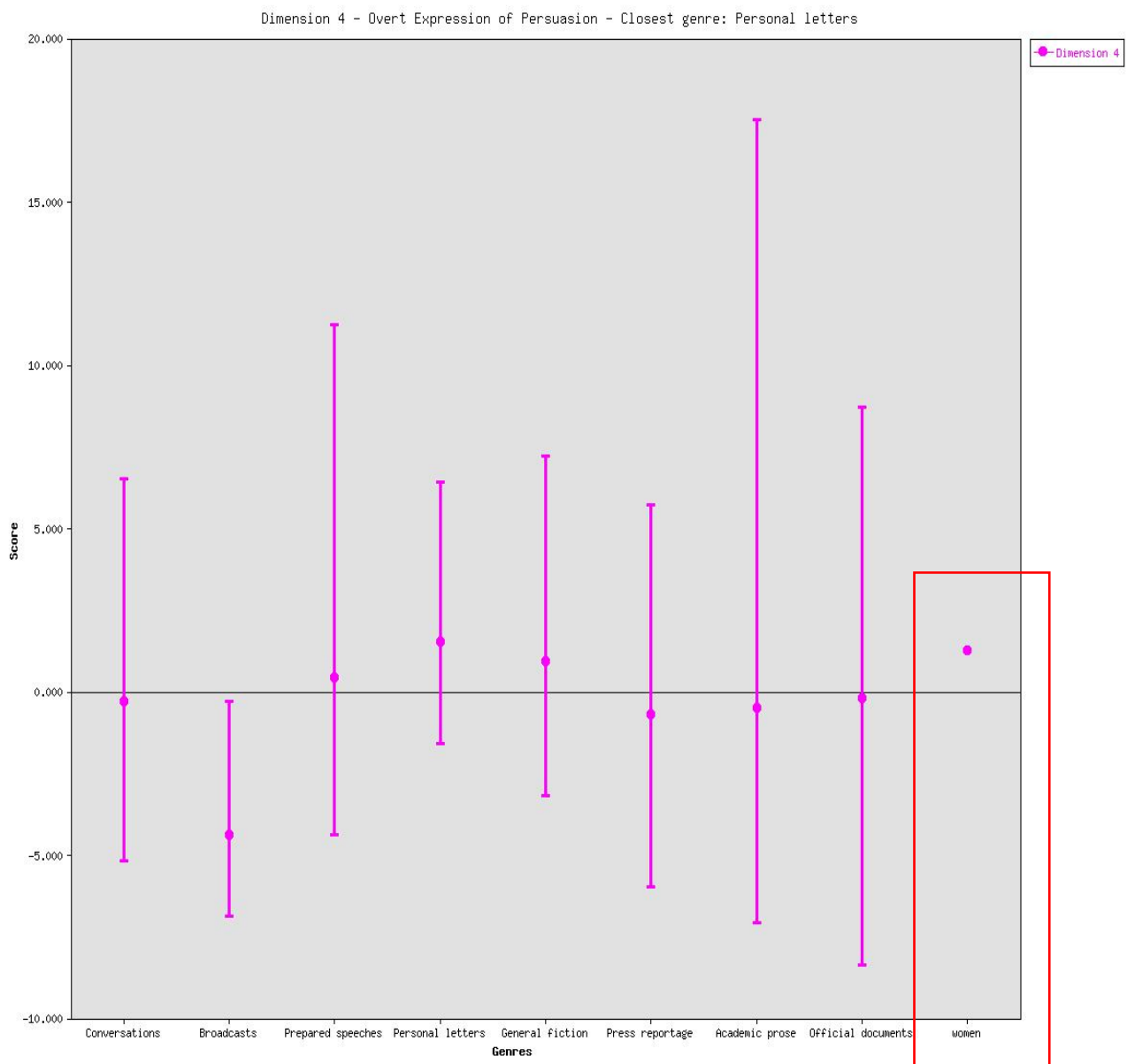
However, the figure shows both positive and negative values. The women category (1.26) indicates that it is the most persuasive and argumentative than other categories. Likewise, the finance category (-6.74) is more non-argumentative than the category of statutory bodies (-6.26) and it is most non-argumentative than the remaining other categories. Ad hoc and loan categories indicated the same trend with a mean score (-0.56). The graph reveals that all of the categories on this dimension, with the exception of university, court, health, and women, have negative values, indicating that they lack argumentative elements.



**Figure 7: Overt Expression of Persuasion**

The graphic representation displays the location of the data and offers several statistics- related details. Our corpus data's location has high scores, as indicated by the graphic representation for dimension 4, indicating that text of Women category is persuasive in nature.

Dimension 4	Mean Score of Women Category
	1.26



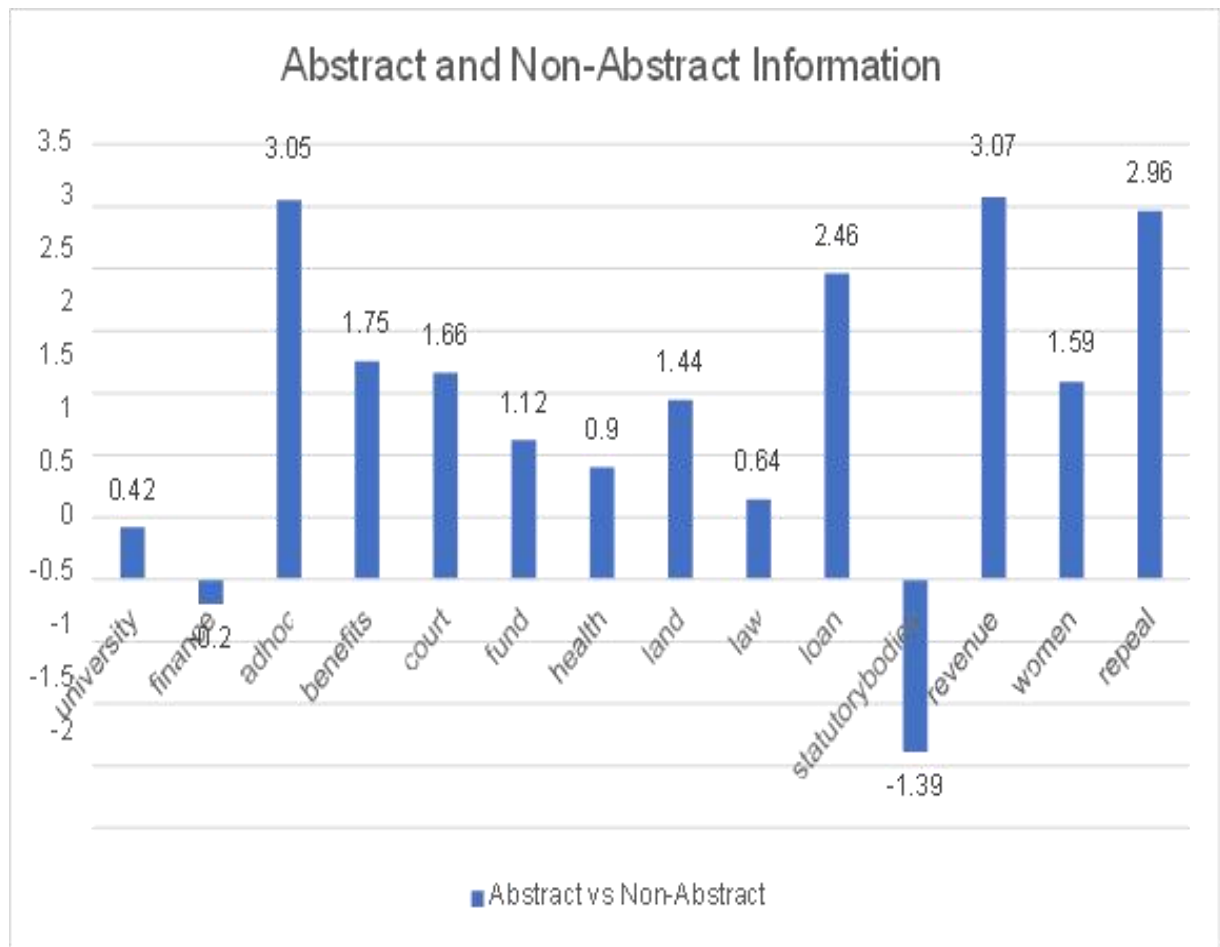
**Figure 8: Graphical Representation of Women Category**

### **Interpretation of Dimension 5**

The distinction between “Impersonal (Abstract) and Non-impersonal (non-abstract style)” is labelled as Dimension 5. The features of abstract information are “conjuncts, agentless passives, past participial clauses, by passives, past participial WHIZ deletion relatives, other adverbial subordinators”.

In D5, the result indicates that the revenue category (3.07) generates the most abstract discourse among all categories. The table shows that the revenue category is slightly higher abstract than ad hoc category (3.05). Statutory bodies produce the most non-

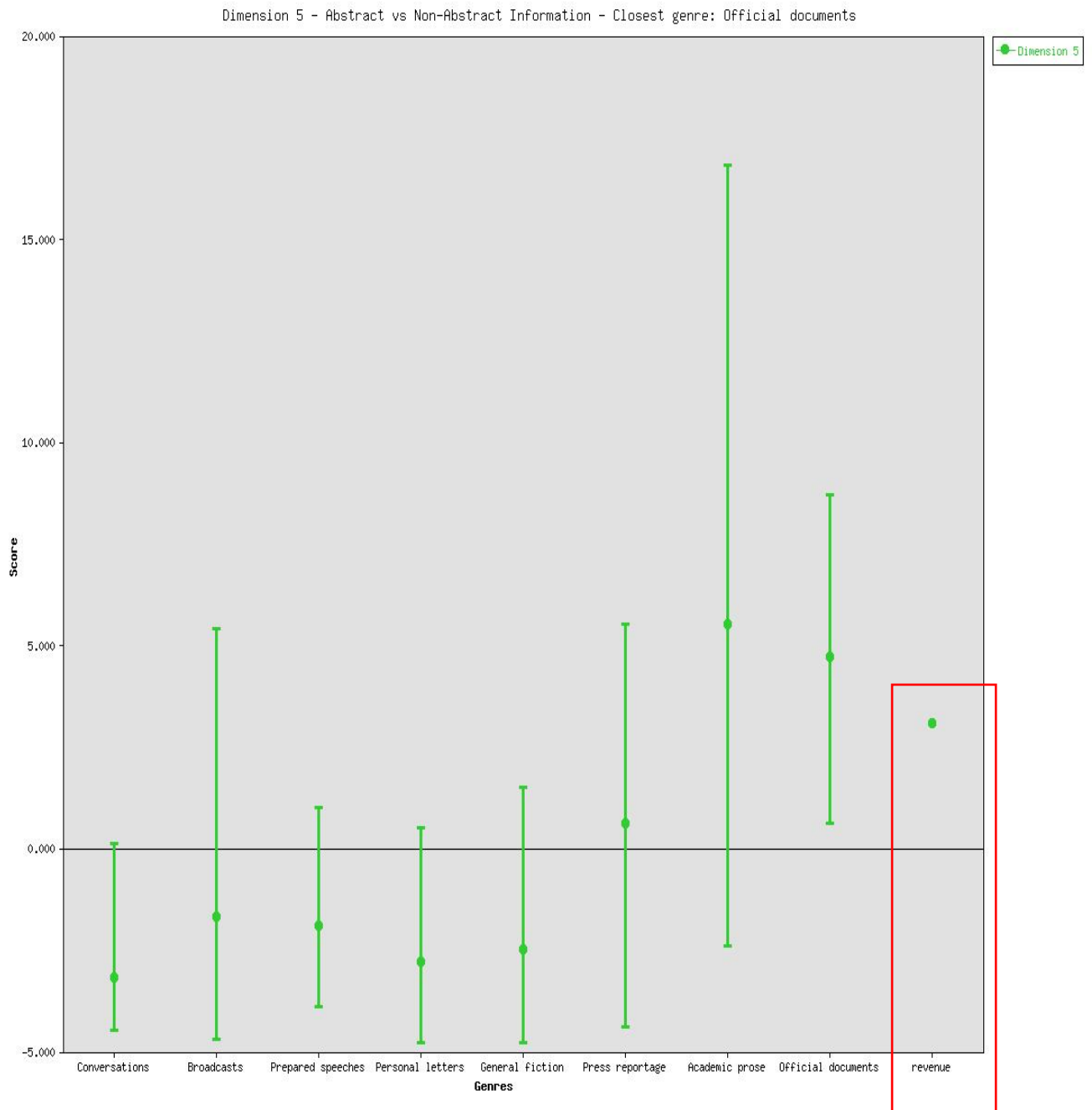
abstract discourse having a mean score of (-1.39) among all categories. The tables show that only statutory bodies and finance categories show negative scores that indicate non-abstract discourse. The figure shows that all of the categories, on this dimension except for statutory bodies and finance, have low values demonstrating the lack of abstract qualities.



**Figure 9: Abstract vs Non- Abstract**

The graphic representation displays the location of the data and offers several statistics- related details. Our corpus data's location has high scores, as indicated by the graphic representation for dimension 5, indicating that text of Revenue category is impersonal in nature.

Dimension 5	Mean Score of Revenue Category
	3.07



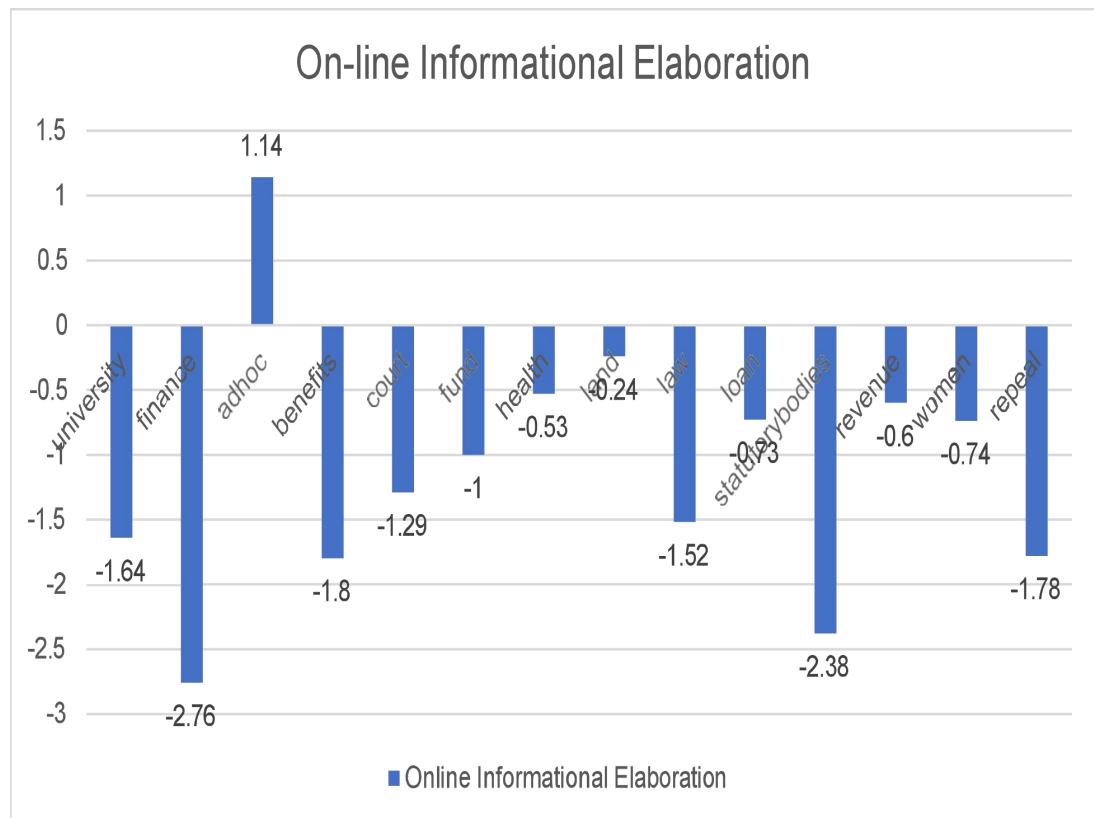
**Figure 10 Graphical Representation of Revenue Category**

### **Interpretation of Dimension 6**

Dimension 6 is labelled as “Online informational elaboration”. High values on this dimension suggest that time limitations were used in the production of the information conveyed. "On-line informational elaboration" exhibits traits such as “verb complements in "that" clauses, demonstratives, relative clauses in object position, and adjective complements in "that" clauses” (Biber, 1988).

The result shows that only ad hoc category with mean value (1.14) shows positive

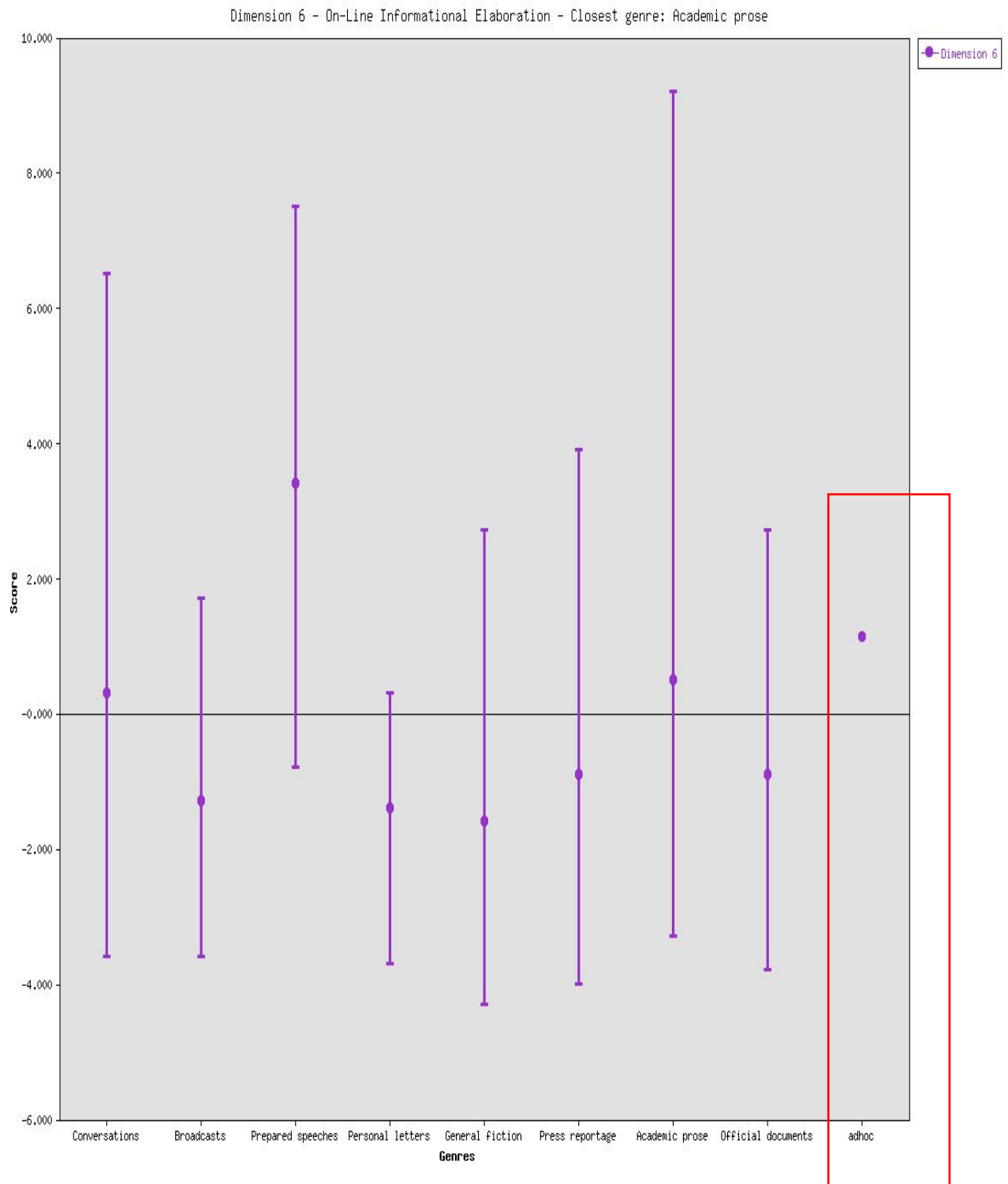
scores among all the other categories which means that information is generated under specific time limitation. In contract, the finance category (-2.76) shows the most negative score than the statutory bodies (-2.38). The graph reveals that all of categories except ad hoc on this dimension has negative values, indicating the lack of online informational elaboration qualities in them.



**Figure 11: On-line Informational Elaboration**

The graphic representation displays the location of the data and offers several statistics- related details. Our corpus data's location has high scores, as indicated by the graphic representation for dimension 6, indicating that text of Ad hoc category is generated under specific time limitation.

Dimension 6	Mean Score of Ad hoc Category
	1.14



**Figure 12 Graphical Representational of Ad hoc Category**

### **Conclusion**

Dimension 1 revealed that all the legal categories are informational with negative

values in higher frequency. The statutory bodies category was found to be more informational than other categories. Dimension 2 showed that every category has a low score, indicating that non-narrative content is their main focus. The repeal category indicates that it was more non-narrative than other categories. Dimension 3 revealed that all the categories are highly explicit and elaborated. The university category was highly elaborated and explicit among all other categories. Dimension 4 showed both positive and negative values. The women category was the most persuasive and argumentative than other categories. In contrast, the Finance category was more non-argumentative in nature. Dimension 5 showed both positive and negative values. The revenue category produces the most abstract discourse whereas the statutory bodies category showed more negative scores. Dimension 6 showed that only ad hoc category showed positive scores among all the other categories which means that the information is generated under specific time limits. In contrast, the finance category showed the most negative score.

### **References**

- Ali, S. S., Amin, M. T., & Ishtiaq, M. M. (2020a). Punctuation errors in writing: A comparative study of students' performance from different Pakistani universities. *sjesr*, 3(1), 165-177.
- Ali, S. S., Farukh, A., & Ishtiaq, M. (2018). Arabic vs English characterization: Comparative analysis of Hind's art of characterization with that of Geoffrey Chaucer. *AL-Qalam*, 23(1), 377-397.
- Ali, S. S., Farukh, A., & Ishtiaq, M. (2019a). Conversation analysis of Muhammad (PBUH) in the light of the model principles laid down by the disciplines of linguistics and communication. *Al Qalam*, 24(1), 300-320.
- Ali, S. S., Farukh, A., & Ishtiaq, M. (2019b). Value of animal idioms in bringing about native-like proficiency among EFL learners. *Global Regional Review*, 4(4), 360-368.
- Ali, S. S., Ishtiaq, M., & Khan, M. (2019c). Conversation analysis of Muhammad (PBUH) for exploring his effective use of nonverbal communication including paralinguistic features. *Rahat-ul-Quloob*, 3(2 (2)), 75-86.
- Ali, S. S., Ishtiaq, M., & Shahid, M. (2020b). Perceptions of Pakistani learners of English about standard British and American English: An exploratory

# **Liberal Journal of Language & Literature Review**

**Print ISSN: 3006-5887**

**Online ISSN: 3006-5895**

- approach. *sjesr*, 3(2), 99-106.
- Ali, S. S., Shahid, M., Ishtiaq, M., Hussain, M., & Khan, M. A. (2021). A study of undergraduate students' perceptions regarding code-switching as a teaching strategy. *Elementary Education Online*, 19(3), 3651-3651.
- Asghar, S. A. (2014). Linguistic Variation in Pakistani Legal English A Multidimensional Analysis. Government College University, Faisalabad.
- Asghar, S. A., & Mahmood, M.A. (2018). A Multidimensional Analysis of Pakistani Legal English. *International Journal of English Linguistics*; Vol. 8, No. 5.
- Asim & Alvi. (2016). A Multidimensional Analysis of Pakistani Press Editorials. *Dialogue*, 270- 284.
- Begum, M. (2016). Linguistic variation across academic writing of Pakistani university Students: A Multidimensional Analysis. Government College University, Faisalabad: (Unpublished doctoral dissertation).
- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Fakuade, G., & Sharndama, E. C. (2012). A comparative analysis of variations in cohesive devices in professional and popularized legal texts. *British Journal of Arts and Social Sciences*, 300-318.
- Finegan, E. (1982). Form and function in testament language. *Linguistics in the professions*.
- Gibbons, J. (2003). *Forensic linguistics: An introduction to language, crime, and the law*. Wiley.
- Gustafsson, M. (1975). Some syntactic properties of English law language.
- Haugen, E. (1972). *The Ecology of Language*. Stanford University Press.
- Maazi, D. (2007). The construction of argumentation in judicial texts: Combining a genre and a corpus perspective. *Argumentation*.
- Moreno, R. G. (2006). A new approach to register variation: the missing link.
- Sardinha, T. B. (2019, June 21). *Lexicogrammar*. Retrieved from online library. <https://onlinelibrary.wiley.com/doi/epdf/10.1002/9781405198431.wbeal0698.pub2>
- Shakir, A. (2014). Study of Information Generating Linguistic Features in Online. *Research on Humanities and Social Sciences*, 122-127.

# **Liberal Journal of Language & Literature Review**

**Print ISSN: 3006-5887**

**Online ISSN: 3006-5895**

- Shakir. (2013). A Linguistic variation across print advertisements in Pakistani media: A multidimensional analysis. (Unpublished doctoral dissertation). International Islamic University, Islamabad.
- Shali, S. K. (2018). Legal Language and Communication: Discourse and Social Implementation. International Journal of Law Management & Humanities.
- Shuy, R. W. (2003). Language in Legal Discourse. Continuum.
- Smith, J. C. (2019). The history of legal language. Cambridge University Press.
- Tiersma, P. M. (1999). Legal Language. Chicago: University of Chicago Press.