



A Corpus-Based Multidimensional Analysis of Linguistic Variation in Pakistani English Novels: A Genre-Functional Approach to ELT

Mahnoor Fatima¹ Shazia Akbar Ghilzai²

¹MPhil Scholar, Department of English, Quaid-i-Azam University, Islamabad, Pakistan,
Email: mahnoorfatima5314@gmail.com

²Assistant Professor, Department of English, Quaid-i-Azam University, Islamabad, Pakistan,
Email: sghilzai@qau.edu.pk

ARTICLE INFO

Article History:

Received: September 11, 2025
Revised: October 05, 2025
Accepted: October 22, 2025
Available Online: November 04, 2025

Keywords:

linguistic variations, Pakistani English novels, MD analysis, corpus analysis, ELT, Lexemes

Corresponding Author:

Shazia Akbar Ghilzai

Email:

sghilzai@qau.edu.pk



ABSTRACT

Corpora has revolutionized language study and transformed the methods used by linguists and academics in English language teaching. This study explores linguistic variation through patterns of co-occurring linguistic features at lexical and grammatical levels in Pakistani English novels. This is a corpus-based study that seeks insight from Biber's model of five-Dimensional Analysis. A dataset of six Pakistani English novels (e.g., Bina Shah's A Season of Martyr, Soniah Kamal's An Isolated Incident, Kamila Shamsie's God in Every Stone, Saba Imtiaz's Karachi You're Killing Me, Mohammed Hanif's A Case of Exploding Mango, and Qurratulain Hyder's River of Fire) has been developed carefully. The dataset has been tagged and analyzed via MAT tagger which replicates Biber's (1988) multi-dimensional analysis framework. This study simultaneously follows the parameters of exploratory quantitative research, focusing on achieving objectivity with the statistical data collected through corpus analysis and by incorporating the researcher's assertions by discussing and interpreting the results. The results indicated that Pakistani English Novels are written in the form of narrative discourse. The language of Pakistani English novels is informational in dimension 1, descriptive in dimension 2, non-argumentative in dimension 3, elaborated in dimension 4, and informal and non-abstract in dimension 5. Therefore, it is closest to the general narrative genre. In the current study, an effort is being made to assist English language teachers and students in Pakistan in comprehending and using corpus-based language teaching pedagogy in classroom settings. This approach can be advantageous for individuals learning the English language and for academic communities seeking to comprehend the diverse applications of vocabulary and their contextual significance.

Introduction

Linguistic diversity, in the world of human communication, manifests as a vibrant array of voices, with each written work serving as a distinct element in the tapestry of expression. Linguistic variety and literary criticism are the two main fields that provide insights into how language behaves or functions within any literary text, especially in the context of linguistic variation. T.S. Eliot's essay "Tradition and Individual Talent," offers perspectives into how a language variance within a literary text can contribute to the linguistic significance and distinctiveness of the text. Linguistic variation in a single text or multiple texts has been a topic of serious interest for linguists and stylistic analysts, and an investigation into linguistic variation can be regarded as of central concern for linguistic studies like stylistics, genre analysis, discourse analysis, and corpus linguistics. Meanwhile, the comprehensive compilation of words and their meanings, known as lexicography, remains an enduring symbol, with its pages carrying the enduring evolution of language throughout history. The field of lexicography has existed for nearly as long as the practice of writing. Since its inception thousands of years ago, writing has primarily fulfilled the practical purpose of facilitating communication among individuals belonging to diverse linguistic communities or using various dialects of a single language. Those demands are subject to ongoing change, just like all living languages. Hence, linguistic variation and its documentation through lexicography have been of great interest to linguists.

Biber's (2006) contribution to analyzing linguistic variations in written or spoken discourse is commendable. He paid particular attention to the lexico-grammatical characteristics and opted for expressing stance. Several studies investigating linguistic variations have been carried out using multidimensional analysis of Biber along with stylistic analysis. This research, by seeking insight from Biber et al., elaborates that there are certain variations in certain kinds of stance expressions. Through the use of MAT tagger and multidimensional analysis, this research paper offered the opportunity to learn more about the types of text and linguistic features that Pakistani English writers use in their novels or literary works. The main question addressed in this research is the co-occurrence of specific linguistic features in six Pakistani English Novels which cause linguistic variance and determine the text type. In numerous literate countries, lexicography has a long-standing legacy, spanning centuries. This heritage involves the creation of word lists and word books using various writing systems, such as hieroglyphs, logograms, or letters. Ever since print culture supplanted scribal culture around five centuries ago and initiated the modern era in European lexicography, the printed book has been the prevailing medium. The dictionary is the most extensively utilized resource for studying languages in educational systems and communities worldwide. Linguistic Variation and Lexicography provide answers to inquiries on the structure, significance, and/or usage of words in one's language or a different language. Electronic corpora and media can help lexicographers, but they do not necessarily make their jobs easier. Man can't manage the amount of textual data computers can store and process. Now, one-volume trade dictionaries use hundreds of millions of rapidly and automatically inputted running words.

Language education and learning are crucial to applied linguistics. The use of dictionaries in language education planning and policy and at all stages of language education justifies the importance of pedagogical lexicography. It focuses on producing and studying dictionaries for first and second/foreign language education and dictionary use, notably by instructors and learners. Mono-, bi-, and multilingual works, children's, school, college, and technical dictionaries are included. Pedagogical lexicography is so extensive that it needs a stronger emphasis. The present study concentrates on second/foreign language teaching and learning. English has emerged as the dominant language for international communication, with people worldwide adjusting and embracing it to fulfill their requirements (Wessendorf, 2015). In recent times, a significant number

of professionals and scholars (Baker, 2012; Aijmer, 2009; Romer, 2010) highly value the contributions that corpus linguistics may make to language instruction. The increasing popularity of using educational corpora and the necessity for study in this field are evident in published research papers, monographs, and books (Campoy, Cubillo, Belles-Fortuno, & Gea-Valor, 2010). The area of English Language Teaching (ELT) has undergone significant advancements, integrating numerous innovative strategies and methodologies to proficiently address the complexities of language instruction. These strategies largely depend on technological advancements and assistance that assists learners. Corpus is a recent creation that serves not only as a tool for various corpus analyses but also holds significance in the field of English Language Teaching (ELT) (Romer, 2010). Applying this teaching method in a non-native environment of English Language Teaching (ELT), such as Pakistan, becomes even more crucial.

The current study is primarily an endeavor to explore linguistic variation and the effectiveness and inevitability of this variation in the selected texts. The secondary concern of this research is to demonstrate the scope of corpus analysis in comprehensively analyzing, interpreting, and understanding any piece of literature or other texts. In the current project, an effort is being made to assist English language teachers and students in Pakistan in comprehending and using corpus-based language teaching pedagogy in classroom settings. The findings of this study shed light on the numerous applications of the available online corpus. The current corpus of five Pakistani English novels provides a wide range of variables. These variables include gender, age, academic position or role, native or non-native speaker, speech event, academic discipline, and interactivity rating. The scope of this study is restricted to the broad application of lexicons, and variable-specificity is not taken into consideration. This is done to provide a comprehensive picture of academic English inside narratives. The aim is to provide a comprehensive depiction of the language as a whole.

Literature Review

The term linguistic variation (or simply variation) refers to regional, social, or contextual differences in the ways that a particular language is used (Trask, 2005). Variation between languages, dialects, and speakers is known as interspeaker variation. Variation within the language of a single speaker is called intraspeaker variation. All aspects of language (including phonemes, morphemes, syntactic structures, and meanings) are subject to variation. Variation in language use among speakers or groups of speakers is a notable criterion or change that may occur in pronunciation (accent), word choice (lexicon), or even preferences for particular grammatical patterns. Variation is a principal concern in sociolinguistics. It has been discovered that variation is typically the vehicle of language change. English has a diverse range of lexical descriptions, which undoubtedly enhances its global stature as a language. The instruction and acquisition of the English language can rely not only on a long-established heritage of diverse grammatical analyses but also on a wide array of dictionaries.

The extensive collection of lexicographical resources is especially valuable for non-native observers who are knowledgeable about the nature, history, development, vocabulary, and grammar of this language. Writing and lexicography are practically ancestral. Since several thousand years ago, it has met the real-life needs of written communication across human populations using different languages or dialects of one language. All live languages vary, and so do those demands. Lexicography has a centuries-old heritage in many literate countries, from word lists and word books in hieroglyphs, logograms, or letters on clay tablets to computers. The printed book has dominated European lexicography since print culture supplanted scribal culture five

centuries ago and started the modern period. The dictionary is the most extensively used language book in schools and society.

Previous researchers have studied Pakistani English to explore Pakistani English as a separate variety. Ali, Samoon, and Ali (2020) examined linguistic variations in Pakistani English in comparison to British English on the morphological and lexical levels, demonstrating that the language is constantly changing in a variety of contexts. This study analyzed different registers such as fiction, newspaper editorials, academic writings, book blurbs, and MPhil and Ph.D. dissertations across the different disciplines. There were two approaches used in computational linguistics in previous studies: a corpus-based approach, which focused on linguistic forms and theory to conduct investigations, and a corpus-driven approach, which relies on corpus data to observe linguistic features, i.e., lexico-grammatical features and linguistic variations. The corpus stylistic approach is also used to investigate corpus data, as corpus-based or corpus-driven research increases the authenticity and specificity of the results. The corpus-based analysis was done through software like Antconc, MAT tagger, Wordsmith tool, POS tagger, and semantic tagger. Pakistani academic writings have remained the focus of attention for researchers in the past few years studied through Biber's multidimensional model. Examples of different works are cited here concerning methodology, the purpose of the study, approach, results, the conclusion of the study, software tools, limitations, and implications of the study.

Rashid, Mahmood, and Ahmad (2017) discussed linguistic choices and syntax variations across research sections in Pakistani academic journals by employing Biber's multidimensional analysis and exploring different sections of research articles along five dimensions. The findings of the study reported that the Pakistani academic journals are informational, non-persuasive, explicit, impersonal, and non-narrative. This research paper reveals that the Abstract in the Pakistani research articles is informational, non-persuasive, and highly elaborated; the Introduction chapter is elaborated and non-narrative; the Literature review is non-narrative; the Methodology is persuasive; the Data analysis chapter is explicit; and the Conclusion chapter is impersonal, least informational, and least non-persuasive among all.

Akoto and Afful (2020) studied linguistic variation through meta-discourse features in the chapters of the Introduction and Literature review of research writings in Pakistan. This study utilizes Hyland's functional meta-discourse model as a research tool and statistically determines the differences in interactive and interactional sub-categories of any section of research writing. The introductions and literature reviews from 10 theses served as the source of data for the corpus. The results of this study showed that these two chapters reported a significant treatment of transitions and Hedges while ranking the meta-discourse items. The researcher quantitatively divided these differences into five indirect subcategories and six interactional categories. This study is theory-oriented, which can help in modifying and verifying the theory in future research. This study was carried out using a mixed approach, integrating statistical and theoretical trends in collecting and analyzing data, which can help us in future studies. The study is significant for its coherent methodology and perfect amalgamation of qualitative and quantitative data collection and analysis techniques.

Azher and Mahmood (2016) have conducted research on Pakistani academic writing using Biber's multidimensional analysis based on the co-occurrence of lexico-grammatical features. A corpus of 8,385,000 million words of Pakistani academic writings was used for examining register variation in the selected M.Phil and Ph.D. theses. The number of selected theses was confined to 235. The researchers made simultaneous use of quantitative and qualitative approaches to explore Pakistani academic writing, which, according to the research findings, was imbued with nouns, present

verbs, nominalization, passives, and conjunctions. These are highly informative, non-narrative, explicit, non-persuasive, and personal discourses. This study highlighted a vacuum to be filled by future studies as it provides information for future researchers, syllabus designers, ESP practitioners, and lexicographers.

Ahmed and Mehmood (2016) made a comparison between the features of the Pakistani press and those of Britain. Douglas Biber's tagger is used for the study of the Pakistani new corpus of 1155705 words to analyze linguistic features of the Pakistani language as a non-native register. The present study concluded that the English language of journalism in Pakistan is of an informative and narrative nature. The sections on persuasion and argumentation are not covert, vividly explicit, and have a concrete vocabulary. Pakistani English has its own independent norm. The British press reporter was inclined towards the use of situation-oriented and non-narrative discursive approaches in the news section. This research shed light on the possibility of future research paradigms in the Pakistani press, as they can be compared with print media to evaluate their linguistic features. The diachronic study on press reportage in Pakistani print media can be conducted using the methodology used in the present study and can trace out the variations and varieties in the press of Pakistan.

Qasim and Qasim (2021) probed into the linguistic variations across the Pakistani book blurbs on four textual dimensions in light of Biber's multidimensional analysis. A corpus of the Pakistani book Blurbs was compiled using online and print sources of 1311 words, and the data was further analyzed with factor analysis. The findings explained that the local blurbs showed variations in all dimensions, while the foreign blurbs reflected a wider range of dimensions. The blurbs revealed the subject matters, discipline, and content of the book. This study encompassed the parameters of a mixed approach to analyze book blurbs in the following 4 textual dimensions: abstract informational description versus concrete human focus, interactive stance versus formal reportage of facts, informational density versus elaborated expression, and expression of personal stance and judgment. This study introduced a research gap in the research about the textbooks in light of different perspectives, like probing into discourse style, the vocabulary of the textbooks, etc. The same technology can facilitate the studies of linguistic variations in literature.

Azher, Mahmood, and Shah (2018) used MD analysis for their research on the linguistic variations across research sections of Pakistani Academic writing comprised of a corpus of 235 M.Phil. and Ph.D. dissertations from different universities in Pakistan was compiled to study five textual dimensions that characterize Pakistani academic writing: informational, detached, formal, non-narrative, and evaluating discourse. Rashid and Mahmood (2019) studied linguistic variations across disciplines in Pakistani research articles by using a multidimensional approach. They used a variety of research methods to look at 405 articles in the humanities, 421 articles in the social sciences, and 503 articles in the sciences using Biber's textual dimensions. They found that the register of Pakistani research articles was made up of clear, objective, and expository information. This study showed that the language of the Pakistani research article was impersonal, non-persuasive, explicit, non-narrative, and informational. The MD analysis approach, used in this study can further aid the study of linguistic variations in three disciplines.

Abdul Aziz, Mahmood, and Azher (2016), using MD analysis, studied variations in the learners' argumentative essays. It was much more focused on individual linguistic and grammatical features than on systematic variations. The 0.7 million argumentative essays written by students were looked at from two angles: how they presented information and how they made their arguments clear; and how they described and compared the writing styles and traits of ESL, ENS, and EFL students. It was preferred that this study be strong enough in terms of methodology, techniques,

and data analysis, as it opened room for future studies. It can help teachers and material developers regarding essay writing practices in Pakistan. It can also help in improvising teaching and in the training of learners and teachers.

Akbas (2021) conducted a corpus-based investigation on features of the oral language of fiction. Using four sets of TV dialogues—the Soap Corpora, the TV Corpus, the Movie Corpus, and the Sydney Corpus—he looked at discourse markers, contractions, interjections, and different types of responses. He then compared these to face-to-face conversations in the Santa Barbara Corpus of Spoken American English and narrative fiction and academic writing in the corpus of contemporary American English. According to the analysis, the features of oral language are more frequent in performed fiction than in face-to-face interactions. The findings of the study were shown with quantitative (statistical approach), qualitative (theoretical approach), and graphical representation. This research explored the features of oral language in four corpora by using corpus-based investigation, which can provide several fictional items and has opened up opportunities for future study. This study was limited to exploring only the spoken part of four corpora but this feature can also be studied in written genres.

Ahmed and Ali (2014) studied the impact of Urduized words in Pakistani English fiction in 2014 by analyzing the process of hybridization by employing a corpus-based tool referred to as Antconc. Pakistani English fiction A corpus consisting of one million words had been analyzed through concordance lines, and further, the influence of Urdu on Pakistani English fiction had been explained. The findings of the study addressed the new lexical items in Pakistani English. The mixed-methods approach has been used to present the fact that the English of Pakistan has its features, making it an independent variable. This study used a corpus tool named Antconc to investigate the impact of Urdu words on Pakistani English fiction. By gaining insight from this study, its methodology, and data collection and analysis techniques, future researchers can study the impact of L1 (Urdu) on L2 (English). This study can further guide the process of hybridization.

Ali and Ahmed (2017) endeavor to explore the discursive strategies used in the English Fiction of Pakistan. This study was centered on multidimensional analysis. The researchers used the Corpus Stylistics approach to study the corpora of 91 novels and 277 short stories to explore lexicogrammatical features used in PEF. This research categorized PEF into the following groups: expression of thoughts versus descriptive discourse production, context-oriented discourse, concrete action discourse versus abstract exposition, and narrative versus dialogic discourse. This study analyzed the discourse styles of Pakistani English fiction. The methodology of this study can be further used to study the linguistic variations in PEF.

The advancements in information technology have significantly influenced the development of learners' dictionaries and the future aspirations of dictionary researchers. Ooi (1998: ch. 2) established a valuable distinction between computational linguistics (the construction of lexicons for natural language processing), computational lexicography (the production or utilization of machine-readable dictionaries), and computer corpus linguistics (the principles and practice of compiling a collection of electronic texts that represent actual language). Literary, cultural, and historical studies have caused scholars in this field to become acutely aware of the fact that dictionaries were and have always been cultural products and political tools (for example, Lara 1995; Stark 1999). This awareness has been brought about as a result of developments in these fields. Spence proposed a reformed alphabet that would enable people from lower social classes to achieve sufficient literacy to become politically aware. Lilles (2000) contended that these dictionaries have played a significant role in defining a sense of national identity in addition to linguistic identity.

Krishnamurthy, R. (2008) examined three significant elements of John Sinclair's enduring impact: the collection of linguistic data, the study and compilation of dictionaries, and the concept of utilizing corpora as a driving force in language analysis. The corpus reflected his preoccupation with the characteristics of linguistic evidence. For him, lexicography is the definitive method of describing language at the lexical level. Because the author was so sure that the corpus should have the most impact on the description, he kept trying to use new computer technology to automate the first steps of analysis and put off until later stages the intuitive and interpretive input of linguists. Sinclair's corpus-driven lexicography model has been widely adopted, starting from its initial use at Cobuild, and is now found in most EFL dictionaries, native-speaker dictionaries such as the New Oxford Dictionary of English, as well as many national language dictionaries in emerging or re-emerging speech communities, and bilingual dictionaries like Collins and Oxford-Hachette.

The introduction of corpora has created new opportunities for studying language and has changed the way linguists and academics approach lexicography and English language instruction (ELT). Following the 1980s, English language teachers and academicians became significantly interested in utilizing online language corpora and computer tools. This research by Zahra, and Abbas, (2018) aimed to examine the practical applications of online corpora for English Language Teaching (ELT) and its usefulness in the Pakistani setting, taking into consideration the current trends and requirements of learners. The Michigan Corpus of Academic Spoken English (MICASE) served as the reference corpus for this research. MICASE is a corpus consisting of around 1.8 million words of transcribed speech, equivalent to over two hundred hours of recorded audio.. The study examined the usefulness of lexical elements in terms of their grammatical function and their application in different circumstances. In addition, the reference corpus was used to investigate the many levels of meanings and uses of lexicons by conducting a thorough analysis of right and left collocates. The findings indicated that lexicons can function as nouns, verbs, and adjectives, depending on the specific context of the study. Additionally, the collocates located to the right and left of the lexicons have a noteworthy impact on comprehending the meanings of lexical items in different settings. Therefore, this approach can be advantageous for those learning the English language and for those involved in academic discourse communities who wish to comprehend the diverse applications of vocabulary items and their contextual significance.

This study by Khan, and Ayaz, (2023), investigated how English as a Second Language (ESL) learners make use of digital lexicographical tools in a typical Pakistani setting, as well as their preferences and the difficulties they face when doing so. The significance of digital lexicographical tools has been significantly increased in recent years as the impact of technology on language learning continues to have a significant impact. However, the particular preferences and difficulties encountered by ESL learners when using these tools have not been extensively studied in the Pakistani context. The researchers utilized a mixed-methods approach and distributed a questionnaire to 200 ESL learners from public sector universities to gather information about their preferences, challenges, and frequency of use. In addition, focus group discussions and individual interviews offer a more profound understanding of the participant's viewpoint. The study's findings indicated that ESL learners have a pronounced inclination towards utilizing digital lexicographical tools (DLTs) because of their easy accessibility and convenience. Some of the most widely used distributed ledger technologies (DLTs) include the Google online dictionary, Oxford Dictionary, thesaurus, and Merriam Webster (Mobile Apps). However, the study highlighted the obstacles encountered by the participants, including inadequate internet access, Google advertisements that hinder a smooth and successful consultation process, the use of complex terminology that exceeds their understanding, and a lack of guidance from teachers in

choosing suitable tools. DLTs have provided them with a feeling of self-governance and self-reliance in the process of acquiring knowledge. Additionally, the study recommended seeking guidance from teachers to choose the most appropriate and suitable digital lexicographical tools.

This research, conducted by Lovtsevich, and Sokolov, (2020) examined a shift in the World Englishes paradigm found in four monolingual English-language learner's dictionaries. This study investigated the extent to which modern learner's dictionaries accurately represent the current global dominance of the English language. The dictionary's exclusive focus was on educational learner needs, which appears to disregard the wide range and profound socio-cultural functions of global English in today's world. The authors analyzed the extent to which dictionaries include non-Inner Circle varieties of English. Specifically, they focus on the inclusion of culture-loaded borrowings from Northeast Asian countries such as China, Japan, Korea, and Russia, where English is commonly used for intercultural communication. Terms related to Northeast Asian countries were often chosen arbitrarily and based on Western cultural preferences rather than regional ones. The authors asserted that the portrayal of non-English cultures in learner's dictionaries was driven by specific beliefs and biases, resulting in a limited perspective that fails to address the demands of an interconnected world.

This study by Belhadj and Bani-Khaled, (2022), presented the findings of a questionnaire, an interview, and a classroom observation regarding the attitudes of Algerian EFL secondary school students towards dictionaries and their utilization. A total of 800 students from various academic levels (freshman, sophomore, and junior) took part in the study. The findings indicated that secondary school pupils possessed a moderately favorable disposition towards dictionaries and their applications. Algerian students demonstrate a favorable attitude towards the importance of dictionaries in the process of teaching and learning. Furthermore, the findings revealed that the majority of Algerian students studying English as a foreign language showed a preference for utilizing bilingual dictionaries as opposed to bilingualized and monolingual dictionaries when it came to language and online dictionaries when it came to format. Students demonstrated a level of awareness of the various tactics employed during dictionary look-up procedures. However, they required additional instruction and supervision from primary school teachers. The participants primarily utilized their dictionaries to retrieve word definitions. The study revealed three primary challenges that Algerian learners encounter when using their dictionaries. The challenges were the complexity of comprehending phonetic symbols, determining the correct definition of words, and the insufficiency of illustrated examples provided in the dictionary. The findings also indicated that the participants were not well-trained in the use of dictionaries.

Over the last two decades, a large body of research about the features of different texts and writings has developed. The researchers have narrowed down their inquiry even to the study of a single characteristic, like Hinkel (2004) and Burrough-Boenisch (2003), who probed into the use of tense and aspect in academic writing. Flowerdew (2003) conducted a study about nouns, their types and the variations in the use of nouns, Vassilev (2001) studied variations in the use and significance of modality, and Soler (2002) analyzed the functional significance of adjectives. Moore (2002), Hyland (2000), and Rowley-Jolivet (2002), respectively, explored the linguistic variations in the textbooks and conference papers. The above-mentioned researchers have explored the occurrence of different individual linguistic features like modal verbs, adjectives, nature of language, and discourse style in Pakistani English novels by using corpus tools like Antconc and Biber's tagger. Previous works created the gap in the field of multidimensional study of Pakistani English novels. To date, various methods have been developed and introduced to measure the linguistic variations and styles of writers. In most recent studies, stylistic analysis, along with corpus analysis done with the help of software like Antconc, MAT Tagger, and many

other programs, has been used to study variation in language and style, whether social, contextual, or regional. Antconc is free-source software that lists lexical elements (lexicography) for English language classrooms to teach narrative linguistic aspects and narrative structure.

Research Question

1. What kind of linguistic variation exists in the selected Pakistani English Novels, as revealed through Biber's multi-dimensional analysis approach (1988)?
2. How do the identified linguistic patterns function to construct meaning and reflect genre-specific conventions in Pakistani English novels?
3. What pedagogical implications can be drawn from these corpus-based findings for teaching English language and literary genres in ELT contexts in Pakistan?

Methodology

Corpus linguistics is considered a research tool or methodology. Sandra Kubler and Heiker Zinsmeister write in their book "Corpus Linguistics and Linguistically Annotated Corpora" that the answer to the question of whether corpus linguistics is a theory or a technique is that it can be both. "It depends on how corpus linguistics is applied." There is much software to study corpus—a collection of written text—like Antconc and MAT Tagger. The present study uses Antconc, a freeware analysis toolkit, to find out the frequencies of lexical items and n-grams—the sequence of words in the corpus—and MAT tagger replicated by Andrea Nini to do the register analysis by finding the linguistic variation across Pakistani English novels. This research employs a corpus-based approach, as it employs a predefined analytical framework defined by Biber to investigate specific linguistic features. The annotated corpus provided by the MAT tagger is subsequently utilized to produce a list of lexical terms and their corresponding context using Antconc, for application in the ELT classroom.

Development of Corpus

In the current study, a corpus of six Pakistani novels by various authors, including *A Season of Martyr* by Bina Shah, *An Isolated Incident* by Soniah Kamal, *God in Every Stone* by Kamila Shamsi, *Karachi You're Killing Me* by Saba Imtiaz, *A Case of Exploding Mangoes* by Mohammad Hanif, and *River of Fire* by Qurratulain Hyder, contains 483,432 words. These novels are retrieved and edited according to the purpose of the study. The pdf form of text is converted into notepad files (.txt) by using Antfileconverter. This corpus is investigated by using MAT Tagger, freely available on the internet, to analyze text types. This research is based on the multidimensional model of Biber.

Table 1. Distribution of Corpus

Novel	Writer	Words
A Season of Martyr	Bina Shah	86505
An Isolated Incident	Soniah Kamal	109676
God in Every Stone	Kamilia Shamsi	51636
Karachi You're Killing Me	Saba Imtiaz	66023
A Case of Exploding Mangoes	Mohammad Hanif	22683
River of Fire	Qurratulain Hyder	146909
Total Words		483432

Given table indicates total word token of the corpus determining the corpus size of 483432.

Multidimensional Analysis Tagger and Data Analysis

Multidimensional Analysis Tagger (MAT) is a computer program which is based on multi-dimensional model proposed by Biber (1988), replicated by Andrea Nini. MAT tagger creates different output files which are used for the interpretation of the text, corpus statistics, z-scores, dimensions, and text types in .txt form (notepad files). This program identifies the dimension scores for each dimension (6 dimensions) and returns the tagged version of input data tagged by a tagger based on Stanford tagger using 67 linguistic features of Biber (1988). To calculate dimension scores, occurrence of features and relative frequency per hundred words is developed. Z-score is calculated by using formula;

$$z_x = \frac{x - \mu_B}{\sigma_B}$$

z_x = z-scores of the features in consideration,

x = relative frequency of a feature

μ_B = mean frequency for that feature in Biber's corpus

σ_B = standard deviation of that feature in Biber's (1988) corpus.

Dimension score is calculated by adding and subtracting z-scores of linguistic features following the features' polarities within a dimension.

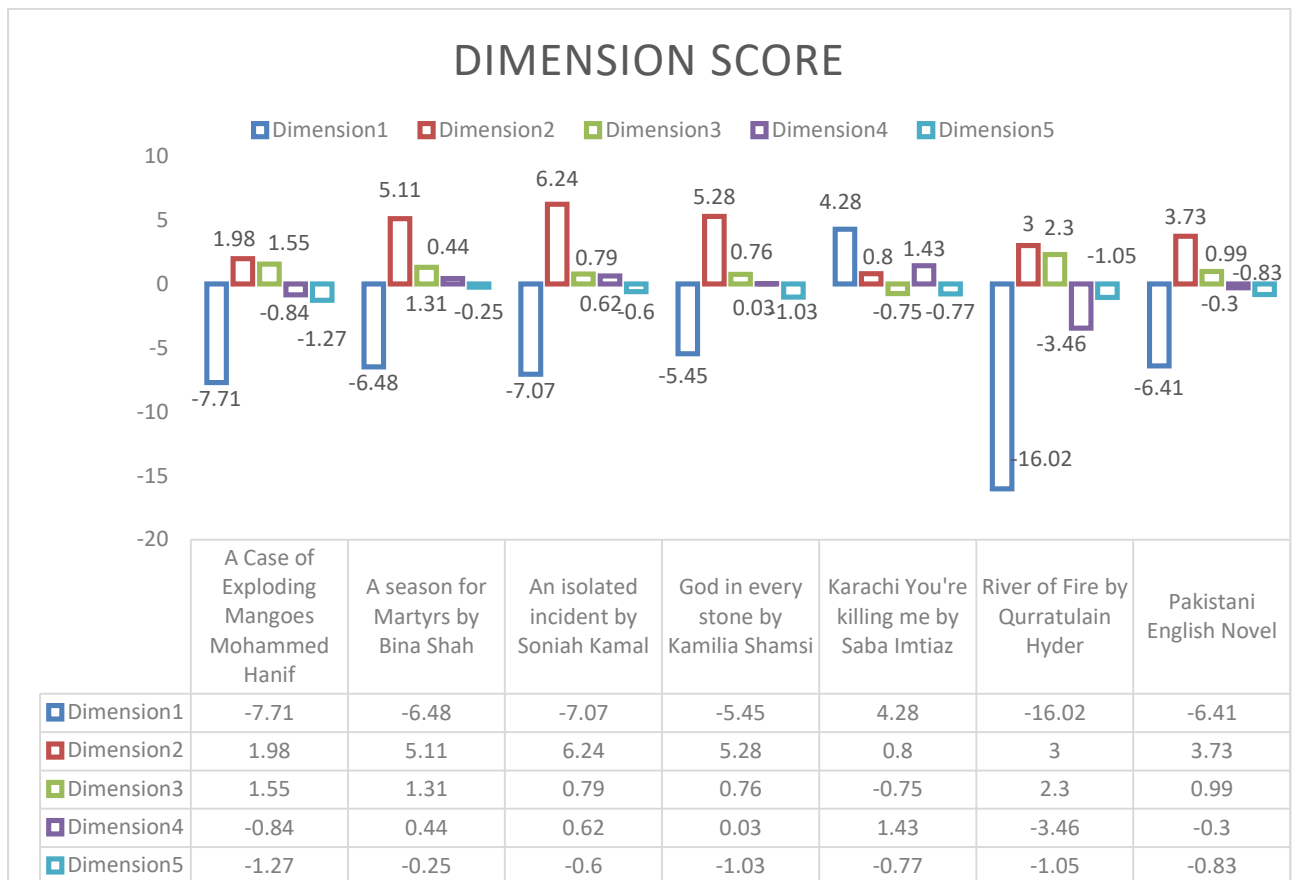
$D = Z$ (Standardized Positive feature) $- Z$ (Standardized Positive feature)

Result and Discussion

This section first presents the results concerning the dimensions that were operationalised in MAT tagger, and then proceeds to discuss their implications to English language teaching (ELT). Based on the tagged corpus, the matter of distribution of the various grammatical and discourse categories identified by MAT (e.g., hedges, conjunctions, and interrogative markers) in the texts are analyzed and how they syntactically and pragmatically pattern in the texts. These patterns consequently afford the empirical grounding of the ELT -oriented discourse, at the end of which the tagged features are converted into the information of pedagogical relevance in designing the materials and dicta, specifically in terms of creating the awareness of the learners of the features of authentic usage, stance-taking, and interactional senses.

Figure 1 demonstrates the comparison of mean dimension scores 6 Pakistani Novels on Biber's five-dimensional analysis. Dimension 1 indicates that the text is more involved than informational. Each novel shows low scores in dimension 1 which demonstrates that the text is complex and contains features like long words, adjectives, and nouns. The lowest score of -16.02 for the novel "River of Fire by Qurratulain Hyder" indicates that the text of this novel is much more complex and denser than all other novels as this novel contains noun phrases. The mean score of Dimension 1 is -6.41, which displays that Pakistani novels are written in a complex language with informational discourse.

Figure 1: Comparison of Five Dimensions of Different Novels of 6 Pakistani Writers.



Low scores with negative values in dimension 2 indicate that the text is non-narrative- structured text written for formal purposes. The mean dimension score of Dimension 2 is 3.73 indicating the text is narrative containing features like third-person pronouns and past tense. The highest score of 6.24 in the novel of Soniah Kamal “An isolated Incident” demonstrates that her writing style is informal and fictional in contrast with other novelists.

All the texts showed explicit trends on Dimension 3. High scores in this dimension indicate that the text is not dependent on the context. All novels showed a positive trend except the novel by Saba Imtiaz “Karachi you're killing me” with a dimension score of -0.75 showing that the text of this novel is to some extent independent of its context. The mean score of Dimension 3 is 0.99 which demonstrates that the text of the novel is a bit more elaborate than situation-dependent.

The highest negative score of -3.46 of the novel “River of Fire” by Qurratulain Hyder " manifests that it makes use of the covert expression of persuasion. The negative scores in two novels (A Case of Exploding Mangoes and River of Fire) indicate that these are non-argumentative. -0.3 is the mean score of dimensions 4, displaying the non-argumentative nature of the text showing prediction, possibility, and necessity by using modal verbs.

Dimension 5 is expressed as Abstract versus Non-Abstract information by Biber. The negative mean score of -0.83 denotes that information given in novels is highly non-abstract. The results of the dimension score of all 6 novels reveal that the novels are fictionally giving us non-abstract information.

From the ongoing discussion, it is clear that Pakistani English novels are highly involved, non-argumentative and non-abstract, having negative values of -6.41, -0.3, and -0.83 in D1, D4, and D5 respectively. Positive scores of 3.73 and 0.99 in D2 and D3 respectively indicate that Pakistani English Novels are narratives and elaborate rather than being dependent on situation references.

Table 2: Dimensions Score and Mean of Dimensions Score

Filename	Dimension1	Dimension2	Dimensi on3	Dimension4	Dimension5	Dimension6	Closest Text Type
A Case of Exploding Mangoes Mohammed Hanif	-7.71	1.98	1.55	-0.84	-1.27	-0.69	General narrative exposition
A season for Martyrs by Bina Shah	-6.48	5.11	1.31	0.44	-0.25	-1.05	General narrative exposition
An isolated incident by Soniah Kamal	-7.07	6.24	0.79	0.62	-0.6	-1.4	General narrative exposition
God in every stone by Kamilia Shamsi	-5.45	5.28	0.76	0.03	-1.03	-1.59	General narrative exposition
Karachi You're killing me by Saba Imtiaz	4.28	0.8	-0.75	1.43	-0.77	-1.48	Involved persuasion
River of Fire by Qurratulain Hyder	-16.02	3	2.3	-3.46	-1.05	-1.64	General narrative exposition
Pakistani English Novel	-6.41	3.73	0.99	-0.3	-0.83	-1.31	General narrative exposition

Linguistic Variations Across Novels

The file named as Z-score shows the statistically under- and over-utilized features of linguistics that explain the systematic difference within the corpus, thus, providing a quantitative evidence of linguistic profiling of the texts. These Z -scores can be used to identify the specific kinds of texts and enable the further interpretation of genre specific stylistic and functional patterns by pointing out which features are much more or less common in certain subsets of the data. The file named Z-score demonstrates the under and overused linguistic features which are responsible for the linguistic variation. These linguistic features determine the text types.

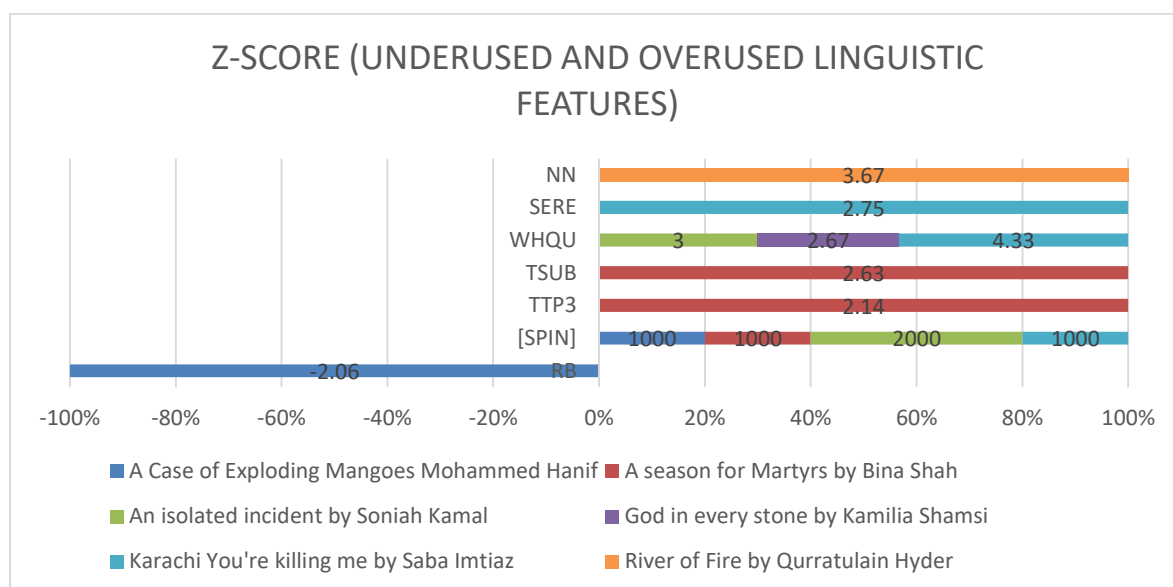
Table 3: Underused and Overused Features in Each Novel (Z-score)

FILENAME	[SPIN]	TPP3	TSUB	WHQU	SERE	NN	RB
A Case of Exploding Mangoes Mohammed Hanif	1000						-2.06
A season for Martyrs by Bina Shah	1000	2.14	2.63				
An isolated incident by Soniah Kamal	2000			3			
God in every stone by Kamilia Shamsi				2.67			
Karachi You're killing me by Saba	1000			4.33	2.75		

Imtiaz							
River of Fire by Qurratulain Hyder						3.67	

67 linguistic features are being analyzed by MAT tagger and their nominalized frequencies are given in the form of Z-scores. Out of those 67 linguistic features, the features having Z-score of more than 2 whether in positive value or negative value are involved in the linguistic variation of the text. Over and Underused linguistic features are given in the above table with their Z-scores. A Case of Exploding Mangos by Muhammad Hanif has 1 underused feature tagged as RB (total adverbs), 2 overused TTR (type-token ratio), and SPIN (Split infinitive). These features make this text a general narrative text. TPP3 (third-person pronoun), TSUB (relative clauses on subj position), WHQU (direct wh question), SERE (sentence relative), and NN (total noun), all these features are responsible for the general narrative exposition text type of 6 Pakistani English Novels being discussed in the present study. Split infinitives showed the highest scores which separate two adverbs.

Figure 2: Z-Score (overused and underused linguistic features)



The given graph shows the trend of overused and underused linguistic features in the text. Examples are taken from the novels indicating the presence of overused and underused features.

A Case of Exploding Mangos by Muhmmad Hanif (SPIN and RB)

I also wish to humbly request that my close arrest is... (to is split infinitive).

.....have the nerve to actually go through it..... (to is split infinitive and actually is an adverb)

Bottle of wine to even think about Jamie. (To is split infinitive and even is an adverb)

From these examples, we can interpret that these two linguistic features co-occurred much in this novel and made this novel a narrative exposition.

A Season for Martyr by Bina Shah (SPIN, TPP3, and TSUB)

Ahmed had been told by Sultan and Allah Bachayo that his gift at chess was his ability to play like a (His is a third-person pronoun co-occurring with noun).

Ahmed heard the words **that chilled** his blood and made his heart almost stop beating in his chest. (That is TSUB co-occurring with VBD)

.... but had **to grudgingly** admit that he admired. (To is SPIN co-occurring with NN)

For calculating dimension scores, the co-occurrence of linguistic features is counted and their relative frequency per hundred words is calculated. Z-scores or standard score of features is calculated by using formula:

$$z_x = \frac{x - \mu_B}{\sigma_B}$$

Given table shows the values or loadings of linguistics features in each novel and these values (mean, standard deviation) are further used to calculate z-score of each feature which help us investigate the dimension score.

Table 4: Mean and Standard Deviation of Normalized Frequencies of Co-occurring Features in Pakistani English Novels

File name	A Case of Exploding Mangoes	A season for Martyrs	An isolated incident	God in every stone	Karachi You're killing me	River of Fire	Mini value	Maxi value	S. D	Mean
Tokens	22923	87390	111771	51441	68534	155321	22923	155321	46672.88	82896.67
AWL	4.39	4.32	4.36	4.32	4.07	4.52	4.07	4.52	0.147241	4.33
TTR	247	232	243	233	244	264	232	264	11.61752	243.83
AMP	0.19	0.1	0.09	0.12	0.13	0.18	0.09	0.19	0.041352	0.14
ANDC	0.38	0.91	0.75	0.84	0.71	0.42	0.38	0.91	0.219583	0.67
[BEMA]	1.37	1.56	1.66	1.56	1.94	1.74	1.37	1.94	0.192916	1.64
[BYPA]	0.03	0.05	0.04	0.03	0.04	0.1	0.03	0.1	0.026394	0.05
CAUS	0.14	0.09	0.11	0.06	0.12	0.08	0.06	0.14	0.028983	0.1
CONC	0.03	0.1	0.05	0.13	0.08	0.05	0.03	0.13	0.037238	0.07
COND	0.31	0.28	0.33	0.31	0.29	0.1	0.1	0.33	0.085088	0.27
CONJ	0.03	0.08	0.1	0.04	0.08	0.07	0.03	0.1	0.026583	0.07
[CONT]	0.8	1.33	1.17	1.57	3.05	0.03	0.03	3.05	1.001234	1.33
DEMO	0.8	0.68	0.59	0.64	0.66	0.56	0.56	0.8	0.083845	0.66
DEMP	0.26	0.35	0.36	0.42	0.52	0.2	0.2	0.52	0.113563	0.35
DPAR	0.02	0.06	0.04	0.06	0.08	0.09	0.02	0.09	0.025626	0.06
DWNT	0.25	0.25	0.21	0.28	0.17	0.14	0.14	0.28	0.053541	0.22
EMPH	0.46	0.55	0.46	0.52	0.9	0.28	0.28	0.9	0.204784	0.53
EX	0.21	0.17	0.17	0.31	0.28	0.15	0.15	0.31	0.065651	0.22
FPP1	3.75	1.22	1.6	1.25	7.02	1.66	1.22	7.02	2.294968	2.75
GER	0.48	0.4	0.49	0.43	0.54	0.43	0.4	0.54	0.051153	0.46
HDG	0.01	0.04	0.02	0	0.06	0.01	0	0.06	0.022509	0.02
INPR	0.06	0.11	0.1	0.08	0.06	0.05	0.05	0.11	0.024221	0.08
JJ	5.69	4.61	4.84	4.44	3.8	5.72	3.8	5.72	0.747101	4.85
NEMD	0.1	0.09	0.1	0.2	0.16	0.13	0.09	0.2	0.042895	0.13
NN	24.2	22.81	23.79	22.49	20.02	31.11	20.02	31.11	3.745488	24.07
NOMZ	1.2	1.15	0.86	1.09	0.74	0.94	0.74	1.2	0.179629	1
OSUB	0.13	0.18	0.15	0.17	0.16	0.09	0.09	0.18	0.03266	0.15
[PASS]	0.59	0.67	0.52	0.42	0.6	0.76	0.42	0.76	0.117587	0.59
[PASTP]	0.07	0.09	0.06	0.06	0.03	0.08	0.03	0.09	0.020736	0.07
[PEAS]	0.85	0.95	1.02	1.03	0.63	1.07	0.63	1.07	0.163921	0.92
PHC	0.48	0.8	0.73	0.47	0.41	0.87	0.41	0.87	0.196435	0.63
PIN	10.03	10.17	9.19	10.83	9.06	9.7	9.06	10.83	0.659545	9.83
[PIRE]	0.03	0.04	0.05	0.08	0.03	0.04	0.03	0.08	0.018708	0.05
PIT	0.97	0.97	1.03	1.52	1	0.55	0.55	1.52	0.308329	1.01
PLACE	0.41	0.58	0.51	0.82	0.48	0.45	0.41	0.82	0.147975	0.54
POMD	0.54	0.56	0.54	0.46	0.68	0.29	0.29	0.68	0.129679	0.51
PRED	0.41	0.57	0.55	0.5	0.64	0.58	0.41	0.64	0.078846	0.54

[PRESP]	0.22	0.41	0.35	0.41	0.3	0.18	0.18	0.41	0.096626	0.31
[PRIV]	1.7	1.67	1.47	1.96	1.89	1.16	1.16	1.96	0.292808	1.64
PRMD	0.38	0.88	1.02	0.79	0.63	0.3	0.3	1.02	0.284089	0.67
[PROD]	0.14	0.13	0.13	0.08	0.19	0.09	0.08	0.19	0.039328	0.13
[PUBV]	0.73	0.66	0.76	0.75	1.05	0.76	0.66	1.05	0.135167	0.79
RB	2.94	3.71	3.74	3.58	4.36	3.25	2.94	4.36	0.483267	3.6
[SERE]	0.07	0.07	0.01	0.04	0.12	0.02	0.01	0.12	0.040373	0.06
[SMP]	0.08	0.05	0.06	0.06	0.1	0.08	0.05	0.1	0.018348	0.07
[SPAU]	0.31	0.37	0.32	0.26	0.46	0.32	0.26	0.46	0.068411	0.34
[SPIN]	0.01	0.01	0.02	0	0.01	0	0	0.02	0.007528	0.01
SPP2	1	0.78	1.1	1.14	1.39	0.97	0.78	1.39	0.203437	1.06
[STPR]	0.18	0.13	0.22	0.14	0.3	0.14	0.13	0.3	0.065651	0.18
[SUAV]	0.44	0.33	0.37	0.38	0.48	0.29	0.29	0.48	0.06969	0.38
SYNE	0.15	0.16	0.33	0.21	0.16	0.14	0.14	0.33	0.071949	0.19
THAC	0.03	0.04	0.04	0.03	0.03	0.01	0.01	0.04	0.010954	0.03
[THATD]	0.28	0.31	0.29	0.44	0.41	0.17	0.17	0.44	0.097502	0.32
THVC	0.26	0.24	0.2	0.11	0.13	0.15	0.11	0.26	0.061128	0.18
TIME	0.27	0.43	0.46	0.52	0.55	0.45	0.27	0.55	0.097707	0.45
TO	1.67	1.88	1.68	1.56	2.09	1.01	1.01	2.09	0.364824	1.65
TOBJ	0.21	0.14	0.08	0.05	0.11	0.06	0.05	0.21	0.059805	0.11
TPP3	5.42	7.81	7.5	6.8	3.15	5.24	3.15	7.81	1.743648	5.99
TSUB	0.17	0.25	0.09	0.07	0.12	0.01	0.01	0.25	0.083526	0.12
VBD	5.27	7.79	8.51	7.76	2.57	8.08	2.57	8.51	2.305903	6.66
VPRT	4.72	1.97	2.18	2.11	8.78	3.69	1.97	8.78	2.623657	3.91
[WHCL]	0.11	0.11	0.09	0.13	0.16	0.05	0.05	0.16	0.037103	0.11
[WHOBJ]	0.03	0.03	0.03	0.11	0.03	0.03	0.03	0.11	0.03266	0.04
[WHQU]	0.12	0.11	0.2	0.18	0.28	0.12	0.11	0.28	0.065853	0.17
[WHSUB]	0.16	0.16	0.13	0.39	0.18	0.22	0.13	0.39	0.094587	0.21
[WZPAST]	0.12	0.16	0.17	0.14	0.09	0.14	0.09	0.17	0.028752	0.14
[WZPRES]	0.51	0.31	0.49	0.4	0.41	0.22	0.22	0.51	0.109727	0.39
XX0	1.06	0.87	0.91	0.89	1.2	0.33	0.33	1.2	0.296086	0.88

This table is revealing the normalized frequencies of all the linguistic features used in Pakistani English Novels by different writers. The Z-score of the features is calculated by mean and standard deviation, for corpus or for each novel or text the tagger points out features having a magnitude higher than 2 as Interesting Variables.

Table 5: Interesting variables in Dimension 1

Pakistani English Novel	Feature	Dimension	Feature
-0.52	Amplifiers	D1	AMP
0.46	Independent clause coordination	D1	ANDC
-0.43	Average word length	D1	AWL
-0.23	Demonstrative pronouns	D1	DEMP
-0.27	Discourse particles	D1	DPAR
-0.24	Emphatics	D1	EMPH
-0.28	Hedges	D1	HDG
-0.32	Indefinite pronouns	D1	INPR
-0.65	Attributive adjectives	D1	JJ
1.69	Total other nouns	D1	NN
-0.48	Total prepositional phrases	D1	PIN
-0.19	Possibility modals	D1	POMD
1.9	Type-token ratio	D1	TTR
-1.13	Present tense	D1	VPRT
-1.26	Be as main verb	D1	[BEMA]
-0.5	Pro-verb do	D1	[PROD]

1.13	Sentence relatives	D1	[SERE]
833.33	Split infinitives	D1	[SPIN]
0.48	WH-clauses	D1	[WHCL]
2.47	Direct WH-questions	D1	[WHQU]

These interesting features are mostly responsible for the variation in text across the dimensions. Interesting features within the developed corpus are mentioned below according to the dimensions that are investigated in the study.

Table 6: Interesting variables in Dimension 2

Pakistani English Novel	Feature	Dimension	Feature
1.33	Third person pronoun	D2	TPP3
0.87	Past tense	D2	VBD
1.24	Present participial clauses	D2	[PRESP]

Table 7: Interesting variables in Dimension 3

Pakistani English Novel	Feature	Dimension	Feature
0.29	Concessive adverbial subordinators	D3	CONC
-0.69	Nominalizations	D3	NOMZ
1.06	Phrasal coordination	D3	PHC
0.68	Place adverbials	D3	PLACE
0.27	Predicative adjectives	D3	PRED
-1.68	Total adverbs	D3	RB
-0.21	Time adverbials	D3	TIME
0.98	That relative clauses on subject position	D3	TSUB
-0.23	Pied-piping relative clauses	D3	[PIRE]
-0.57	WH relative clauses on object position	D3	[WHOBJ]

Table 8: Interesting variables in Dimension 4

Pakistani English Novel	Feature	Dimension	Feature
-0.38	Necessity modals	D4	NEMD
0.26	Predictive modals	D4	PRMD
0.28	Infinitives	D4	TO
-0.84	Split auxiliaries	D4	[SPAU]
0.29	Suasive verbs	D4	[SUAV]

Table 9: Interesting variables in Dimension 5

Pakistani English Novel	Feature	Dimension	Feature
-0.33	Conjuncts	D5	CONJ
0.42	Other adverbial subordinators	D5	OSUB
-0.24	By-passives	D5	[BYPA]
-0.56	Agentless passives	D5	[PASS]

1.38	Past participial clauses	D5	[PASTP]
-0.36	Past participial WHIZ deletion relatives	D5	[WZPAST]

Dimension score determines the genre of Pakistani English Novels which is General Narrative Exposition – the text which conveys the information by the use of narration, having low scores on dimension 1 and high score on dimension 2.

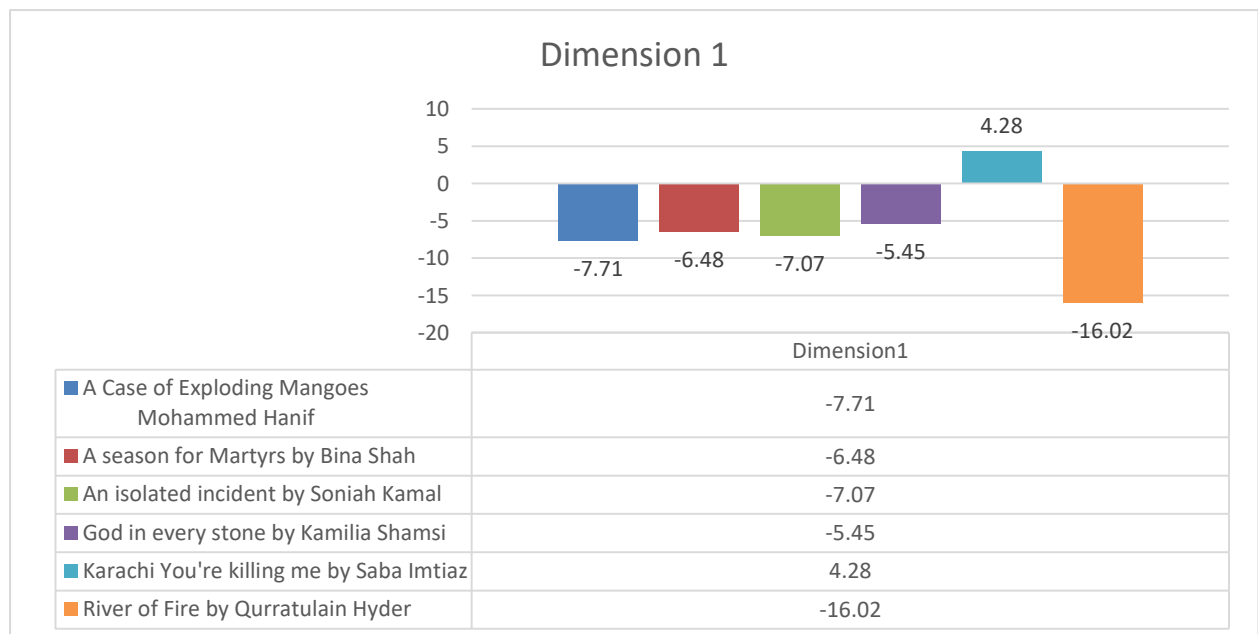
Dimensions

From the given graphs, it is easy to find out major trends of linguistic variations and co-occurring features in Pakistani English Novels by using multidimensional analysis of Biber.

Dimension 1: Involved vs Informational text

This dimension deals with two poles involving an interactive text with positive values and informational text with fully edited and planned text with negative values. The negative side means informational text has nouns, adjectives, and prepositional phrases to show the density of information. The mean score of -6.41 indicates that the information in Pakistani English Novels is dense and has attributes of academic prose, general fiction, and broadcast. 6 Novels by different Pakistani writers showed explicit variation in D1. The comparison reveals that all novels except Saba Imtiaz’s “Karachi You're Killing me” had positive weight in dimension scores, which demonstrates that this novel is more involved than informational. While Dimension score of -16.01 for the novel by Qurratulain Hyder indicates that the information given in that novel is denser than others.

Figure 3: Dimension 1



D1 is informationally dense due to the occurrence of negatively dominant features (AMP, AWL, EMPH, VPRT, PROD), Following tables shows the number of interesting variables in dimension 1.

Table 10: Negative Loadings in Dimension 1

Pakistani English novels (Values)	Feature	Dimension	Feature
-0.52	Amplifiers	D1	AMP
-0.43	Average word length	D1	AWL
-0.23	Demonstrative pronouns	D1	DEMP
-0.27	Discourse particles	D1	DPAR
-0.24	Emphatics	D1	EMPH
-0.28	Hedges	D1	HDG
-0.32	Indefinite pronouns	D1	INPR
-0.65	Attributive adjectives	D1	JJ
-0.48	Total prepositional phrases	D1	PIN
-1.13	Present tense	D1	VPRT
-1.26	Be as main verb	D1	[BEMA]
-0.5	Pro-verb do	D1	[PROD]

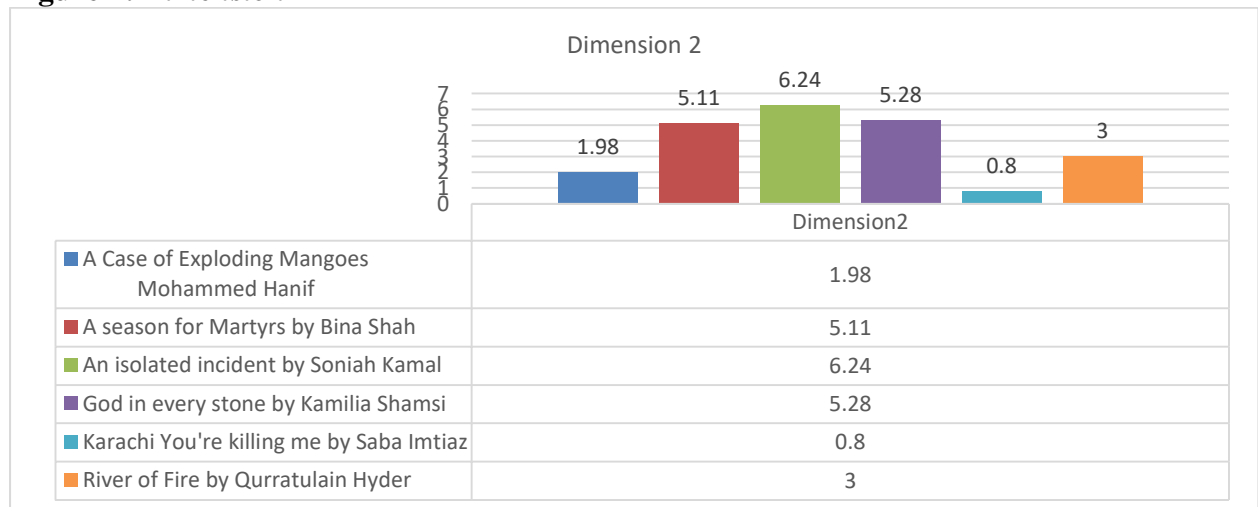
Table 11: Positive Loadings in Dimension 1

Pakistani English novels (Values)	Feature	Dimension	Feature
0.46	Independent clause coordination	D1	ANDC
1.69	Total other nouns	D1	NN
1.9	Type-token ratio	D1	TTR
1.13	Sentence relatives	D1	[SERE]
833.33	Split infinitives	D1	[SPIN]
0.48	WH-clauses	D1	[WHCL]
2.47	Direct WH-questions	D1	[WHQU]

Dimension 2: Narrative vs non-Narrative

This dimension differentiates between narrative and non-narrative text. The result with a negative weight represents that the text is exploratory and descriptive. While the result with positive weight displays that the text is event-oriented containing past tense, third-person pronoun, and perfect aspect verbs. The mean score of 3.73 for D2 manifests that this compiled corpus for this research is narrative and is closely related to general fiction and press reportage. All the novels showed positive dimension scores in D2, depicting that they are more descriptive.

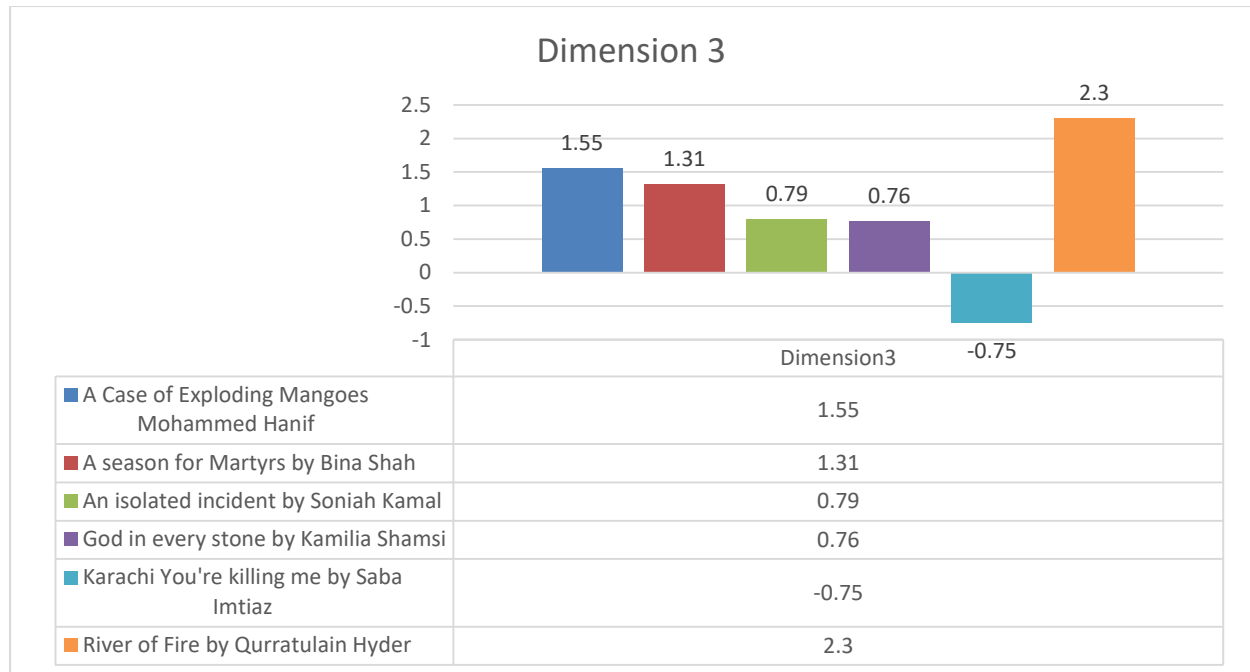
Figure 4: Dimension 2



Dimension 3: Context- Independent discourse vs Context-dependent Discourse

The features with positive values in this dimension include relative clauses to demonstrate explicitly informational text. The temporal and place adverbials along with adverbs represent context-dependent discourse. The present result of the mean score of Pakistani English Novel depicts that the corpus is more highly elaborated and informational than situation-dependent and is closely related to prepared speeches. Comparative analysis of D3 indicates that the only novel which had a negative value is “Karachi You're Killing Me” by Saba Imtiaz demonstrating that the novel is situation-dependent with a dimension score of -0.75.

Figure 5: Dimension 3



High scores in this dimension indicate that the text is not dependent on the context. Dimension 3 is explicit because of the use of positive features like TSUB, PHC, PLACE. Both features are equally present in the text but the average of the values indicates that positive features are dominant.

Table 12: Negative Loadings in Dimension 3

Pakistani English Novel (Value)	Feature	Dimension	Feature
-0.69	Nominalizations	D3	NOMZ
-1.68	Total adverbs	D3	RB
-0.21	Time adverbials	D3	TIME
-0.23	Pied-piping relative clauses	D3	[PIRE]
-0.57	WH relative clauses on object position	D3	[WHOBJ]

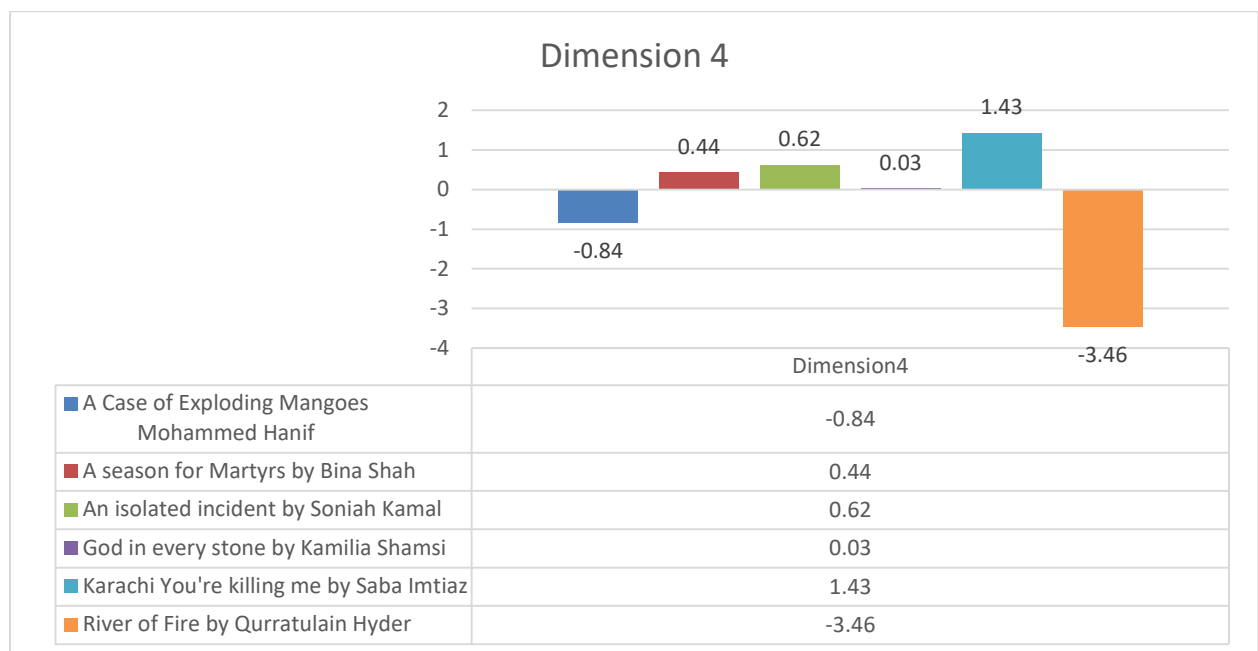
Table 13: Positive Loadings in Dimension 3

Pakistani English Novel (Value)	Feature	Dimension	Feature
0.29	Concessive adverbial subordinators	D3	CONC
1.06	Phrasal coordination	D3	PHC
0.68	Place adverbials	D3	PLACE
0.27	Predicative adjectives	D3	PRED
0.98	That relative clauses on subject position	D3	TSUB

Dimension 4: Overt Expression of Persuasion

Biber remarks: “this dimension marks the degree to which persuasion is marked overtly, whether the overt making of the speaker’s point of view or an assessment of the advisability or likelihood of an event presented to persuade an event” (1988, P.111). The closest genre of this dimension is the conversation as it is having a mean score of -0.3 which indicates that the text is non-argumentative.

Figure 6: Dimension 4



-0.3 is the mean score of dimension 4, displaying the non-argumentative nature of the text showing prediction, possibility, and necessity by using modal verbs because of the use of negative features (NEMD, SPAU)

Table 14: Negative Loadings of Dimension 4

Pakistani English Novel (Values)	Feature	Dimension	Feature
-0.38	Necessity modals	D4	NEMD
-0.84	Split auxiliaries	D4	[SPAU]

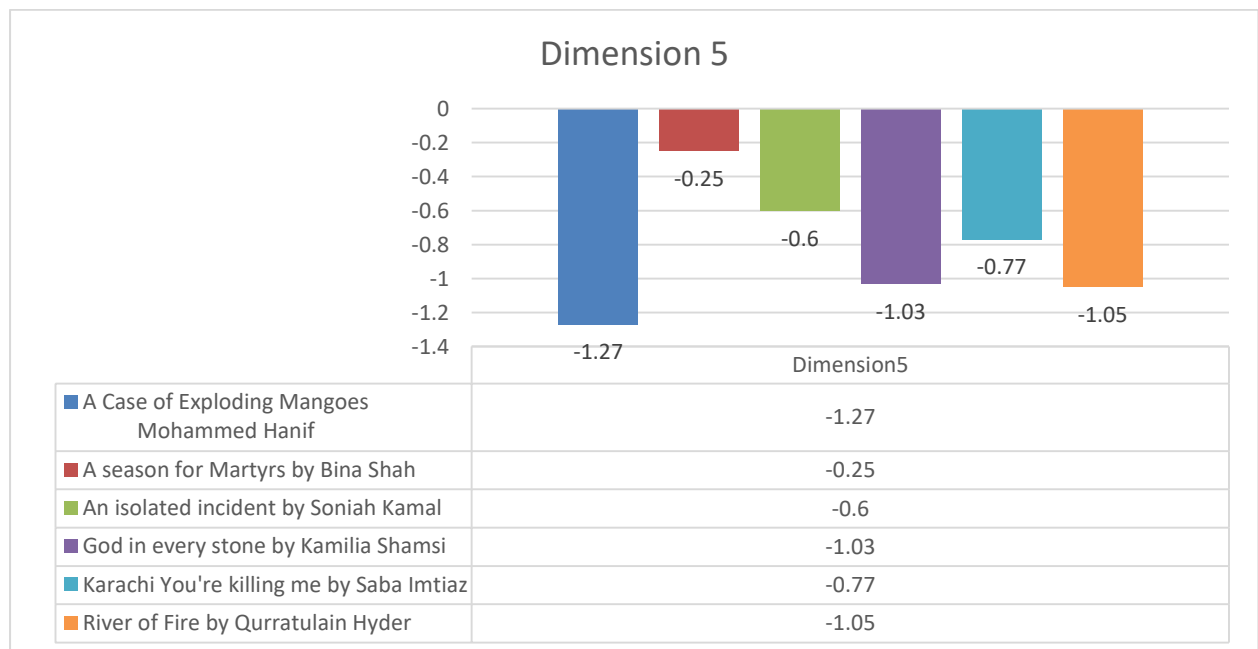
Table 15: Positive Loadings of Dimension 4

Pakistani English Novel (Values)	Feature	Dimension	Feature
0.26	Predictive modals	D4	PRMD
0.28	Infinitives	D4	TO
0.29	Suasive verbs	D4	[SUAV]

Dimension 5: Abstract vs. non-Abstract Information

This Dimension displays the informational discourse which is formal and abstract. The linguistic features like adverbial, past participial WHIS deletions, and predicate adjectives make the logical relationship between the information. The negative mean score manifests that the corpus is non-abstract and is related to the broadcast genre. The novel by Muhammad Hanif showed the highest negative value in D5 indicating that his style of writing is more informal than other writers.

Figure 7: Dimension 5



The negative mean score of -0.83 denotes that information given in novels is highly non-abstract due to these features (CONJ, PASS).

Table 16: Negative Loadings of Dimension 5

Pakistani English Novel (Values)	Feature	Dimension	Feature
-0.33	Conjuncts	D5	CONJ
-0.24	By-passives	D5	[BYPA]
-0.56	Agentless passives	D5	[PASS]
-0.36	Past participial WHIZ deletion relatives	D5	[WZPAS T]

Table 17: Positive Loadings of Dimension 5

Pakistani English Novel (Values)	Feature	Dimension	Feature
0.42	Other adverbial subordinators	D5	OSUB
1.38	Past participial clauses	D5	[PASTP]

Corpus and ELT

Corpus-based language instruction is a growing trend in English language teaching (ELT).

The corpus can serve as a valuable instrument for studying several aspects of language, such as grammatical structures, frequency patterns, contextual language use, and terminologies. Lexicography traditionally focuses on the semantics and usage of words. Corpus linguistics has revolutionized lexicon study by investigating the frequency of distinct terms. (Frequency), Collocation refers to the systematic relationships between words, specifically how some words invariably occur together. Do phraseological structures and words exhibit systematic relationships with specific registers, dialects, and communities?

Lexicography is the scholarly examination of lexemes, which are the basic units of meaning in language, along with their corresponding grammatical categories. In the beginning, lexeme and grammar were examined separately, but in lexicography, they are researched as a unified category, intricately intertwined. An example of such a methodology is the examination of synonyms and antonyms. Students may find it challenging to grasp that "significance" and "importance" have the same meaning. However, by independently searching for these terms in a corpus and studying them within a contextual framework, their knowledge will be improved, ultimately resulting in enhanced learning. Learners frequently acquire synonyms, yet in their enthusiasm to utilize novel and superior vocabulary, they may employ ill-suited synonyms in academic or non-academic writing. Therefore, a corpus enables learners to efficiently utilize lexicons. Alternatively, instructors can prompt students to correct their errors by doing a corpus search to identify the appropriate contextual usage of both words.

Many linguistic features have been identified from the analysis that was shown earlier. These linguistic structures and features can be taught to learners of English as a foreign language (ELT) for them to comprehend the narrative structure and lexemes that are utilized in literary work of this kind. Here is a list of linguistic variables that English Language Teaching (ELT) students need to be taught based on the interesting features found in multidimensional analysis that explain how text changes across dimensions:

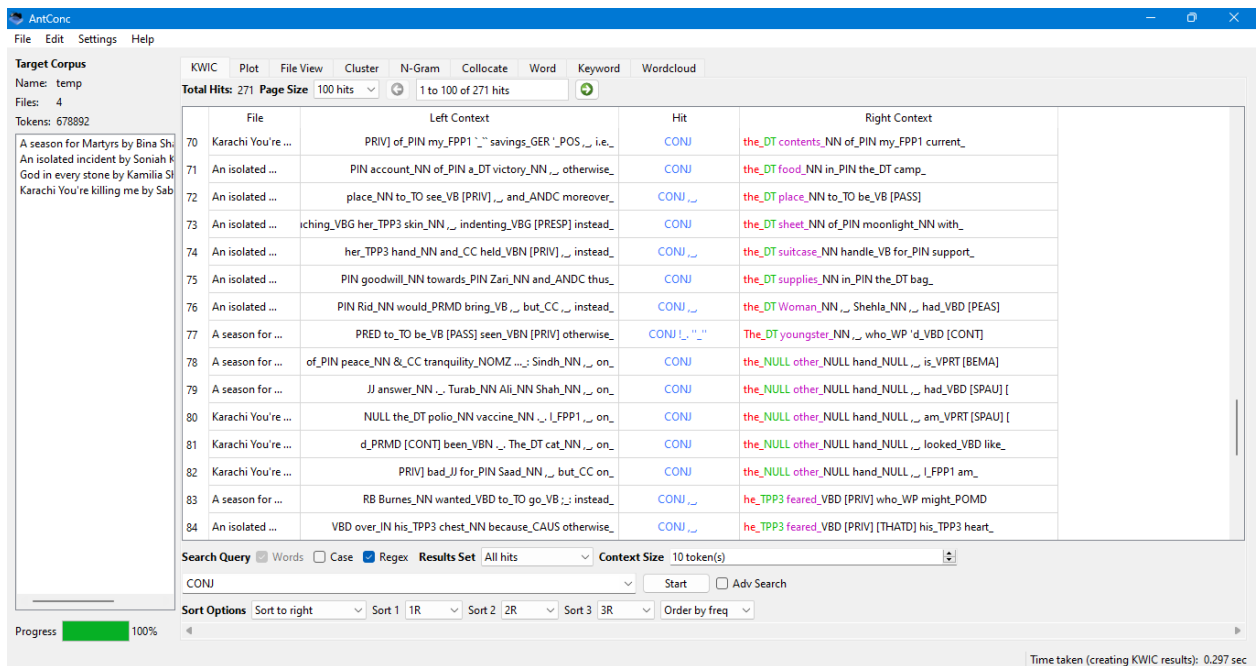
- Amplifiers
- Independent clause coordination
- Average word length
- Demonstrative pronouns
- Discourse particles
- Emphatics
- Hedges
- Indefinite pronouns
- Attributive adjectives
- Total other nouns
- Total prepositional phrases
- Possibility modals
- Type-token ratio

- Present tense
- Be as main verb
- Pro-verb do
- Sentence relatives
- Split infinitives
- WH-clauses
- Direct WH-questions
- Third person pronoun
- Past tense
- Present participial clauses
- Concessive adverbial subordinators
- Nominalizations
- Phrasal coordination
- Place adverbials
- Predicative adjectives
- Total adverbs
- Time adverbials
- That relative clauses on subject position
- Pied-piping relative clauses
- WH relative clauses on object position
- Necessity modals
- Predictive modals
- Infinitives
- Split auxiliaries
- Suasive verbs
- Conjuncts
- Other adverbial subordinators
- By-passives
- Agentless passives
- Past participial clauses
- Past participial WHIZ deletion relatives

Contextualization refers to the review of the context in which a word is used in order to comprehend its various dimensions of meaning (Bhatia, 2004). It enhances the comprehension, utilization, and comprehension of lexical elements. Merely stating the definition or surface-level meaning of a word is inadequate for learners, as it fails to provide them with the necessary understanding to effectively incorporate it into their everyday language. Nevertheless, if learners are instructed to utilize the selected corpus of Pakistani English novels to search for a specific word and subsequently analyze its frequent co-occurrences and the corresponding context in which it is used, this approach may result in more advanced learning.

Here are some examples of lexical items that have been extracted using Antconc in order to examine the contextual meanings of those lexical features.

Figure 7: CONJ (Conjunctions used in the text)



In the present research, the usage of AntConc was used to produce keyword-in-context (KWIC) concordance lines based on the tagged corpus that gives empirical evidence to dictionary compilation. In the figure 7, the concordance view is shown, where the rows indicate an instance of a token marked by the MAT tagger, where the central column displays the target item (e.g., the part-of-speech tag CONJ) and the left and right columns on the left and right respectively display the co-text of the target item. This structure reveals the syntactic environments, collocation patterns and discourse functions of each tagged item, and thus it makes it possible to extract the most important linguistic characteristics in a systematic way, including common word combinations, typical positions within the clause, and common semantic functions of the prosodies. The patterns found in corpora are then used to shape the patterns of dictionary headwords, sub-senses and notes of usage whereby the entries are based on actual use of language as opposed to inner sense.

Figure 8: WHQJ (Direct WH-questions)

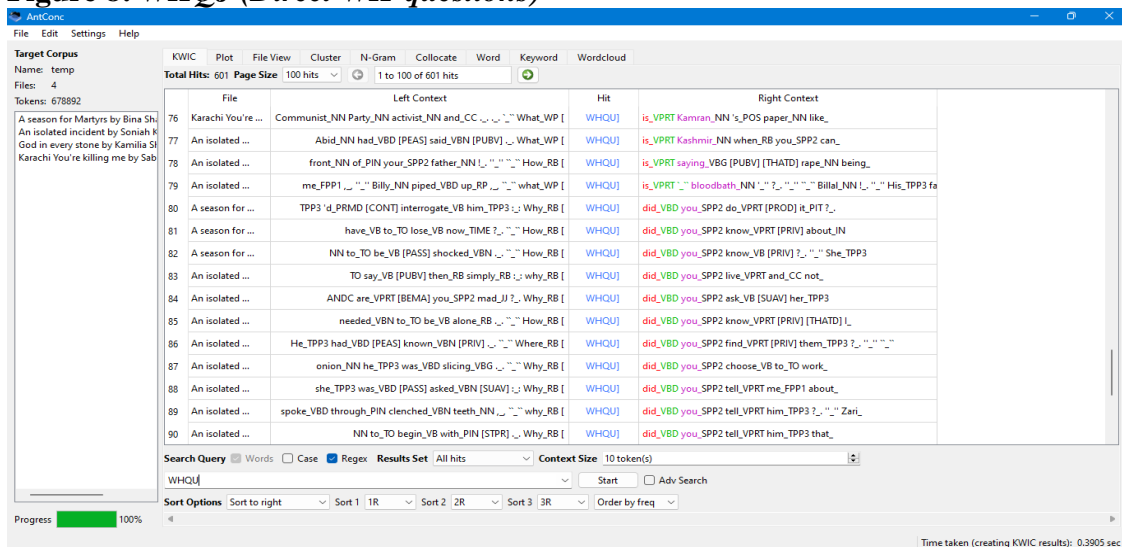


Figure 8 illustrates how the structures of WH questions are actually manifested in the corpus e.g. which auxiliaries co-occur with WHQU, what subject position, and what common verb form. These trends will assist in identifying structural properties of interrogatives, including the combinations of tense/aspect, frequent lexical collocations, and structure of the clause, and encode them as grammar information and patterns of usage in your dictionary entries of WH-words and similar constructions.

Figure 9: HDG (Hedges)

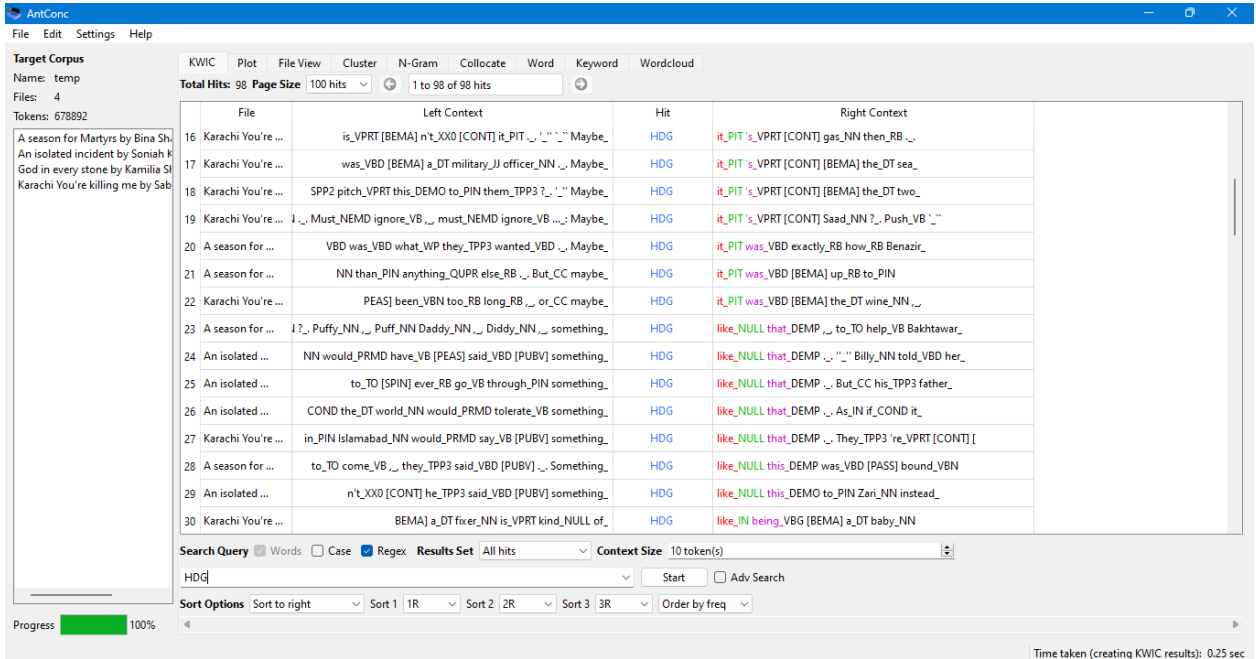
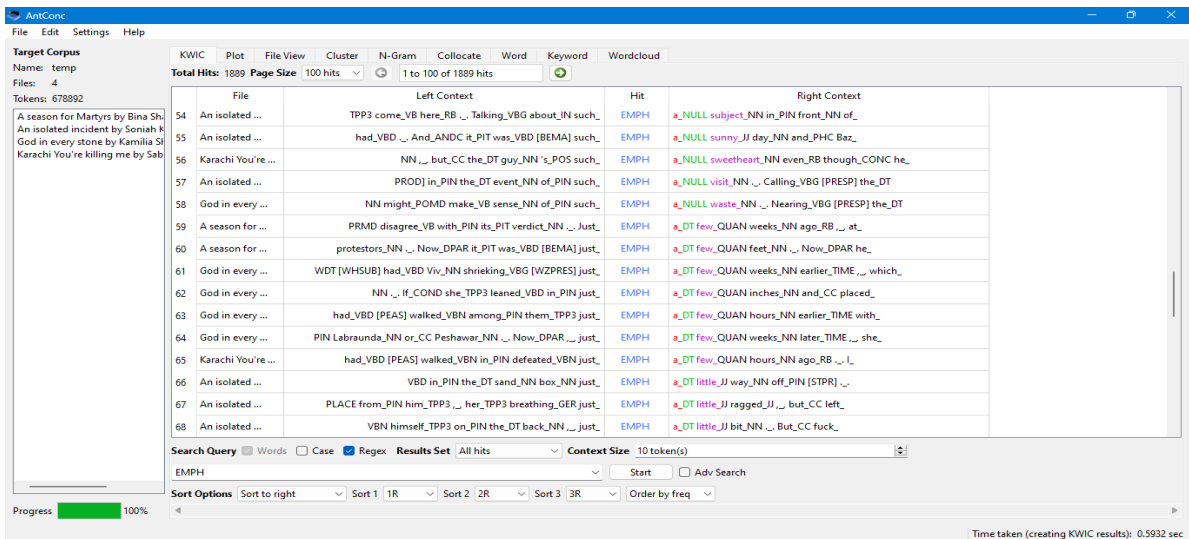


Figure 9 shows the key-word-in-context (KWIC) result of the tag HDG (hedges) in the corpus of the MAT-tagging. Under this interpretation, every line of the concordance has an instance of HDG in its central column of hits that is surrounded by its immediate left and right co-text, which is made up of other morpho-syntactically annotated tokens.

Figure 10: EMPH (Emphatics)



Through the examination of these regular structures and collocational patterns of emphatics shown in figure 10, the study becomes able to draw generalisation of grammatical nature in the functioning of linguistic feature and its distribution, which in turn can be used to make the creation of the dictionary entries and description of the use based on corpus evidence and not introspective intuition. Learners can extract the lexemes and their context from the offered examples. For instance, under the CONJ tag, learners can extract all the conjunctions used in the corpus, such as "and," "instead," and "thus." Corpora can be used to teach grammatical phrases such as noun phrases and direct Wh questions, as well as other grammatical structures. English language Teachers can employ either data-driven or pedagogical-based approaches to teach English proficiency to learners using corpora.

Conclusion

The present study is focused on the linguistic variation across 6 Pakistani English Novels by different writers by applying a five-dimensional analysis of Biber, using MAT tagger. From the ongoing discussion, the Pakistani English Novel is written in the form of narrative discourse. The mean score of -1.31 indicates that the text of Pakistani novels is informational, descriptive, non-argumentative, elaborated, informal, and non-abstract. It is closest to the general narrative genre. The objectives of this research are achieved by analyzing the z-score of linguistic features which are responsible for the linguistic variation in the text: i.e., NN (noun), WHQU (direct wh questions), RB (adverbs) TPP3 (third-person pronoun), TSUB (relative clauses on subj position), AND SERE (sentence relative).

Due to its status as a developing country, Pakistan is not included in the list of technologically sophisticated nations. As a result, traditional pedagogical methods are typically employed for English Language Teaching (ELT) at all educational levels. By adopting the strategies employed in developed nations, Pakistan can introduce modern technology and corpus-based pedagogy in English Language Teaching (ELT). This would enhance the quality of ELT in Pakistan. Corpus-based English Language Teaching (ELT) can serve as a valuable tool for acquiring knowledge in different aspects of language. The abundance of data can offer students and scholars numerous alternatives and perspectives to examine each item within its context, so fostering the contextual understanding of language, a facet often overlooked in traditional grammar. Corpus-based strategies are valuable in English Language Teaching (ELT) for enhancing vocabulary, comprehending meaning in context, using idiomatic expressions, mastering phrasal verbs, and improving grammar.

The current study discussed a comparative analysis of linguistic variations across Pakistani English Novels by using a compiled corpus of 6 Pakistani novels by different writers. The scope of this study is confined to a limited number of data; the same methodology can be applied to the investigation of a wider variety of literary genres. The present can help future researchers to analyze linguistic variation and to study the discourse style of Pakistani English Writers by investigating linguistic features. It can help analyze the nature of the language used by writers in their texts. The result of MD analysis can also be compared with stylistic analysis. This study can help find out different lexical grammatical features used by writers which are the basic cause of variation in their language.

References

1. Abdulaziz, M., Mahmood, M. A., & Azher, M. (2016). Variation in learner's argumentative essays—a multidimensional comparative analysis. *Science International*, 28(4), 413-424.

2. Ahmad, S., & Mahmood, M. A. (2015). Comparing Explicit Features of Pakistani Press Reportage with British Press Reportage: A Multi-Dimensional Analysis. *NUML Journal of Critical Inquiry*, 13(2).
3. Akbas, E. (2012). Exploring Metadiscourse in Master's Dissertation Abstracts: Cultural and Linguistic Variations Across Postgraduate Writers. *International Journal of Applied Linguistics & English Literature*, 1(1), pp. 12-26.
4. Akoto, O. Y., & Afful, J. B. A. (2020). Variations In Metadiscourse Use in English Language Introduction and Literature Review Thesis Chapters. *Language Literacy: Journal of Linguistics, Literature, and Language Teaching*, 4(2), pp. 390-408.
5. Ali, S., & Ahmad, S. (2017). Discourse styles of Pakistani Fiction in English: A Multidimensional Analysis. *Balochistan Journal of Linguistics ISSN-No. 2312-5454 (2017 Issue 5)*, 9.
6. Ali, S., & Ahmad, S. (2016). Discourse style variation among the leading novelists of Pakistani fiction in English: A multidimensional analysis. *ELF Annual Research Journal*, 18(1), 227-246.
7. Ali, S., & Shehzad, W. (2019). Linguistic variation among South Asian englishes: A corpus-based multidimensional analysis. *Journal of Nusantara Studies (JONUS)*, 4(1), 69-92.
8. Azher, M., & Mehmood, M. A. (2016). Exploring Variation across Pakistani Academic Writing: A Multidimensional Analysis. *NUML Journal of Critical Inquiry*, 14(2).
9. Azher, M., Mehmood, M. A., & Shah, S. I. (2018). Linguistic Variation Across Research Sections of Pakistan Academic Writing: A Multidimensional Analysis. *International Journal of English Linguistics*, 8(1), pp. 30-43.
10. Belhadj, Z., & Bani-Khaled, T. (2022). The Place of Pedagogical Lexicography among Algerian EFL Secondary Schools' Students. *Jerash Journal for Research & Studies*, 23.
11. Biber, D., Johnson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English* (Vol. 2). London: Longman.
12. Biber, D. (2006). Stance in spoken and written university registers. *Journal of Academic Purposes*, 5(2), pp. 97-116.
13. Coleman, J. (2006). Lexicography. *The Handbook of English Linguistics*, 581-600.
14. Halliday, M. A. K. (2003) Introduction: On the 'architecture' of human language. In J. J. Webster (ed.) *On Linguistics and Language*. Volume 3 in *Collected Works of M. A. K. Halliday*, 1-29. London and New York: Continuum.
15. Khan, M. Y., & Ayaz, A. (2023). Lexicographical Landscape: An Investigation into Digital Tools, Preferences, Challenges and Usage Patterns among Pakistani Graduate ESL Learners. *Annals of Social Sciences and Perspective*, 4(2), 541-553.
16. Krishnamurthy, R. (2008). Corpus-driven lexicography. *International Journal of Lexicography*, 21(3), 231-242.
17. Lovtsevich, G. N., & Sokolov, A. A. (2020). World Englishes and learner lexicography: View from the expanding circle. *Russian Journal of Linguistics*, 24(3), 703-721.
18. Qasim, S., & Qasim, A. B. (2021). Linguistic Variation across Pakistani Book Blurbs on Four Textual Dimensions: A New Multidimensional Analysis. *CORPORUM: Journal of Corpus Linguistics*, 3(2), 22-50.
19. Rashid, A., & Mahmood, M. A. (2019). Linguistic variations across disciplines: A multidimensional analysis of Pakistani research articles. *Global Social Sciences Review*, 4(1), 34-48.
20. Römer, U. (2010). Using general and specialized corpora in English language teaching: Past, present and future. *Corpus-based approaches to English language teaching*, 18-35

21. Wessendorf, S. (2015). 'All the people speak bad English': Coping with language differences in a super-diverse context.
22. Zahra, T., & Abbas, A. (2018). Pedagogical Implications of Corpus-Based Approaches to ELT in Pakistan. *Journal of Education and Educational Development*, 5(2), 259-275.