

NIMRA PERVEZ

Department of Applied Linguistics
Government College University, Faisalabad, Pakistan
ORCID: 0009-0003-6751-5711
nimrapervez471@gmail.com

ALI RAZA SIDDIQUE

Department of Applied Linguistics
Government College University, Faisalabad, Pakistan
ORCID: 0000-0003-1946-2347
aalimalik381@gmail.com

MUHAMMAD AHMAD

Department of Applied Linguistics
Government College University, Faisalabad, Pakistan
ORCID: 0000-0001-6759-9511
ahmad453@yandex.com

ANALYZING LINGUISTIC VARIATIONS IN THE DISCUSSION SECTIONS OF PAKISTANI ENGLISH RESEARCH ARTICLES: A MULTIDIMENSIONAL STUDY

This study investigated linguistic variations to establish the legitimacy of Pakistani English as a distinct variety. The corpus comprised 70 discussion sections from Pakistani English research articles distributed across three academic disciplines (engineering, information technology, and literature). The analysis was conducted using Multidimensional Analysis Tagger v.1.1, which scrutinized 67 grammatical features across five dimensions. The results revealed significant differences in dimensions 1, 3, and 5 which indicated variations in Pakistani English. The results also showed the discussion sections: being highly informative, formal, abstract, and explicit in engineering; informative, and abstract in information technology; and less informative, explicit, and abstract in literature research articles. However, the results displayed fewer variations in dimensions 2, and 4 which suggested them being interesting for further research. Overall, this research contributed to our under-

standing of the linguistic variations in Pakistani English and its distinct characteristics across academic disciplines.

Keywords: academic writing, article discussion sections, linguistic variations, multidimensional analysis, Pakistani English

1. Introduction

Linguistic variations are observable through co-occurring lexico-grammatical patterns that signal the style of the speaker or writer and the nature of language varieties and texts. Previous research has delved into this domain, with some studies focusing on linguistic variations within discussion sections (Jin 2018; Parkinson 2011; Szczygłowska 2022, 2023). A few researchers have combined linguistic variation analysis with move analysis (Al-Shjairi and Al-Manaseer 2022; Amnuai 2019; Holmes 1997; Mu 2021), while others have concentrated on linguistic variations in other sections of research articles (Ahmad et al., 2019; Liu and Xiao 2022; Louvigne et al., 2014; Xie 2020). However, limited research has been conducted in the context of Pakistani English and its linguistic variations (Azher and Mahmood 2016; Azher et al., 2018; Rashid et al., 2017). Additionally, rhetorical structures have been studied alongside linguistic variations in the Pakistani context (Hussain et al., 2019; Yasmin et al., 2019). The current research focuses on linguistic variations as this approach allows for the simultaneous study of multiple linguistic features underscoring (see Khamaiseh 2023) how do the student writers from different context and culture structure information in their writing.

The study of linguistic features follows a specific methodology that involves exploring the methods employed by researchers, which are grounded in contextually dependent frameworks based on logical perspectives. Previous studies have utilized various research methodologies. Corpus-based methodology has been employed to investigate linguistic variations within research articles (Ahmad et al., 2019; Azher and Mahmood 2016; Azher et al., 2018; Hussain et al., 2019; Jin 2018, 2021; Liu and Xiao 2022; Louvigne et al., 2014; Mu 2021; Parkinson 2011; Rashid et al., 2017; Xie 2020), and dissertations (Ahmad et al., 2022, 2023a, 2023b, 2023c). Genre analysis methodology has been used (see Al-Shjairi and Al-Manaseer 2022; Amnuai 2019; Holmes 1997; Szczygłowska 2022, 2023; Yasmin et al., 2019) to examine the rhetorical structures and linguistic differences in the discussion sections. This study employs a corpus-based methodology to achieve both qualitative and quantitative outcomes. The use of corpus-based methodology for this study is based on the notion that corpus-based research has become popular in studying linguistic and stylistic features to offer a detailed understanding of the use of a language in the real world (Alamgir and Shakir 2023).

There have been several studies on discussion sections in research articles that focus on organizational patterns through move-based analysis, micro-analytic studies, or variations in discussion sections and other parts of research articles. However, some gaps remain in the existing research. Firstly, past studies overlook the specific language used in particular genres. Secondly, the studies concentrating solely on individual linguistic features, do not reveal the full extent of language variations within or across genres. This study, therefore, aims to provide a comprehensive description of the linguistic features employed in the discussion sections of engineering and social sciences, along with their interpretation through the corpus-based analysis of linguistic variations. The reason for employing corpus-based analysis for this study is that it helps explore distinctive linguistic features to understand linguistic nuances in academic writing (see Alamgir and Shakir 2023). In addition, the results obtained through corpus-based analyses could be “generalized to the broader context of English language in academia in Pakistan” (Shah et al., 2023: 5).

2. Literature review

2.1. Theoretical framework

For a comprehensive understanding of linguistic variations, research models play a vital role. These models are based on underlying theories that can contribute to the qualitative or quantitative nature of the results. Previous studies (see Ahmad et al., 2019, 2022, 2023a, 2023b, 2023c; Azher and Mahmood 2016; Azher et al., 2018; Fatima et al., 2023; Hussain et al., 2019; Jin 2018, 2021; Khamaiseh 2023; Liu and Xiao 2022; Mu 2021; Parkinson 2011; Rashid et al., 2017; Szczygłowska 2022, 2023; Xie 2020) have employed various research models. Some researchers utilized Biber’s (1988) model to explore differences in linguistic features. Yang and Alison’s model of moves was employed (in Amnuai 2019) for studying the structure of moves. Hopkins and Dudley Evans’ model facilitated genre analysis and the examination of move structures (in Holmes 1997), while Swale’s move analysis model was used in other studies (e.g. Louvigne et al., 2014; Yasmin et al., 2019). These models focus on identifying different moves within texts. However, these models are not sufficient enough to comprehensively address linguistic variations. As the current research necessitates a mixed-method approach, it examines linguistic features using Biber’s (1988) multidimensional model. This model allows analyses systematically, qualitatively, and quantitatively. For systematic analyses, this model employs computer-based software for tagging, and analyzing the data. As far as quantitative and qualitative values are concerned, this model offers ways not only to carry computations of the large amounts of data but also facilitates the

functional interpretations of the results (see Ali 2020; M. Fatima et al., 2023; Siddique et al., 2022). This model is well-suited to this study as it also helps explore multiple linguistic features and facilitates a more comprehensive understanding without introducing ambiguity.

2.2. Definitional background

Linguistic variations are a result of differences in registers underscoring the significance of these specific language registers. These variations can be grouped into dimensions, aiding in the interpretation of the functions shared by patterns of co-occurrence. Quantitative research employs statistical methodologies in order to study the features associated with these dimensions (Biber 1988).

Register variation is observed to emerge in a specific environment in which a particular language is used by the people. This variation occurs as a universal phenomenon resulting in the syntactic variation caused by the physical environment of the use of a language (Ferguson 1983). It is important to mention here that there are two distinct registers of a language, that is, spoken (e.g. conversation, press reportage etc.) and written (e.g. academic prose, textbooks etc.). The users' choice of linguistic features in these registers is context specific. The factor analysis facilitates the successful identification of the said linguistic features, associates them with different dimensions which further help to study variation in the discourse (Biber 2008). Furthermore, the variation reflects a varied use of the language affected by various factors e.g. contextual, regional, social, and so on (Barzan and Heidary 2019). Therefore, these variations are important to understand the dynamic nature of a language.

2.3. Past studies in foreign context

The linguistic variation has been found to remain the focus of paramount importance among the researchers from the field of corpus linguistics. This section presents the review of noteworthy studies conducted in the world particularly in Pakistan.

First of all we start with the studies conducted in the world. The first noteworthy study in this regard was conducted by Parkinson (2011) in order to study linguistic features in the discussion sections of the research articles and lab reports related to the field of physics. This study employed Biber's (1988) model and studied the features associated with different dimensions with the help of WordSmith 5.0. The results showed the articles and lab reports to be assertive as well as coherent but less articulated. These results suggested improvements in the use of lexico-grammatical features. Jin (2018) is another important study which was conducted on the corpus of research articles retrieved from journals with

high and low impact factors to study linguistic variation in the discussion sections. This study also employed Biber's (1988) model and MAT (Multi-dimensional Analysis Tagger) for the analysis. The results showed frequent use of boosters as well as hedges in the corpus of high impact journal articles and emphasized the importance of the use of lexico-grammatical features. This study used potential statistical methods for analyses. However, its scope was limited to the small dataset. Still there was another study (Jin 2021) by the same researcher which explored linguistic variation in the discussion sections of research articles published in the journals related to the engineering. The researcher used a mixed-method approach involving MAT, Patcount, and SPSS 2.0 software. Three linguistic dimensions were related to the expressions of stance, while another three were related to the narrative, informational, and negation expressions. The results had implications for teaching instructions and academic writing. However, the corpus collection was not been well-defined. Similarly, Mu (2021) explored linguistic variations in three sub-corpora, consisting of international and Chinese local journal articles. The study encompassed 1309 articles in English, 79 in Chinese, and 79 in translation. Using MAT tagger and Biber's (1988) model, the mixed-method approach indicated that both Chinese and English corpora were informational and objective. The results revealed context-dependent and context-independent dimensions, offering pedagogical insights. However, the study's generalizability is limited to two disciplines within social sciences and humanities.

Some studies incorporated move analysis along with linguistic variations in discussion sections. For example, Holmes (1997) emphasized the study of structures and linguistic variations in the discussion sections of 30 research articles (drawn from US journals) from three disciplines i.e. sociology, political sciences, and history. The study employed corpus-based methodology. The results showed similarities in political science and sociology's discussion sections but distinctions in history's discussion sections. While the methodology was appropriate, the study was limited to its applicability as it focused on only three disciplines. Another study (Amnuai 2019) focused on linguistic patterns and moves within the discussion sections of 20 international journal articles in accounting. The study relied on corpus-based methodology employing a mixed-method approach. It found that certain moves were used consistently, while others were optional. This approach could facilitate pedagogical objectives and improve discussion section writing. However, the limited size of the corpus and the focus on accounting make the study less generalizable.

Recently, Al-Shujairi and Al-Manaseer (2022) examined communicative moves and linguistic variations within 50 discussion sections of medical research articles from highly cited medical journals, with each journal contributing to 10 articles. The study used corpus-based methodology, specifically AntConc 3.5.7w. The results indicated the presence of background information and the use

of specific expressions like procedural verbs, first-person pronouns (plurals), and past tense. These findings could enhance the construction of well-organized discussion sections. However, the study's limited scope, confined to the medical science discipline, and the absence of a model are noteworthy limitations. Similarly, Szczygłowska (2022) and Szczygłowska (2023) explored disciplinary variation in the use of lexical verbs, and linking adverbials respectively in the discussion sections of articles written by Anglophone, and Polish researchers from medicine, and psychology disciplines. Results revealed disciplinary differences in the use lexical verbs (Szczygłowska 2022), and linking adverbials (Szczygłowska 2023).

There are also noteworthy studies that observed linguistic variations in sections other than discussion. For example, Louvigne et al. (2014) examined variations in 431 introduction sections of 543 research articles selected from *Optics' International Journal*. Using a corpus-based methodology, the study employed a mixed-method approach with AntConc software. The study revealed differences in moves between abstract and introduction sub-genres and highlighted the importance of deeper analysis of linguistic features for improving introduction section writing. The move analysis was robust, but the use of AntConc software may impact reliability.

Xie (2020) performed a multidimensional analysis using MAT software on 1000 abstract sections of master's theses written by Chinese students over three years to study linguistic variations. This study utilized a mixed-method approach based on corpus-based methodology and Biber's (1988) multidimensional analysis. It identified noticeable differences in dimensions 1, 3, and 5 and fluctuations in the trend of abstraction, with potential implications for English for Academic Purposes teaching. While MAT tagger provided statistical data for solid analysis, the study's generalizability is limited as it focused on abstract sections written by the Chinese students only. Recently, Liu and Xiao (2022) investigated variations in 3000 conclusion sections of research articles from peer-reviewed international journals, with 1500 articles from natural and social sciences disciplines. The study used corpus-based methodology and the MAT tagger based on Biber's (1988) multidimensional analysis model. The results highlighted differences in dimensions 1, 2, and 4 and the presence of possibility and prediction modals in all conclusion sections, particularly in education and sociology disciplines. While the study had a strong methodology, it was limited to its focus on two disciplines only. All of these foreign studies provide valuable insights into linguistic variations in various contexts helping us shape the research landscape in this area.

2.4. Past studies in Pakistani context

Numerous studies have been conducted within the Pakistani context which shed light on linguistic variations and their manifestations. These studies contribute valuable insights to the understanding of this linguistic feature. Here, we present an overview of relevant studies conducted in Pakistan. The first study to mention here is Azher and Mahmood (2016) that delved into lexicogrammatical features across registers by analyzing 235 M. Phil and Ph.D. dissertations using Biber's (1988) Tagger. Employing corpus-based methodology and a mixed-method approach, the study revealed significant differences in dimensions 1 and 4 while noting no significant differences in the other three dimensions. Moreover, there was a significant distinction in the methodology and conclusion sections, depicting Pakistani academic writing as non-narrative and informational. This study holds potential for syllabus designers and researchers. Statistical results and descriptive analysis strengthened the research but its generalizability was limited. Another study (Rashid et al., 2017) explored linguistic variations in academic journal articles published within Pakistan. Utilizing a corpus-based methodology and Biber's (1988) MD analysis model, the study categorized various sections of articles. Findings showed the nature of sections such as abstract, introduction, literature review, methodology, data analysis, and conclusions as informational, non-narrative, least non-narrative, less persuasive, less explicit, and impersonal, respectively. This study lays the foundation for future investigations into linguistic differences. While the statistical analysis was robust, the limitations involved the small data size consideration and confinement to the research article genre. Azher et al. (2018) also examined linguistic variations across registers by systematically collecting 235 theses totaling to 1175 file texts. Employing Biber's (1988) tagger, the study utilized a mixed-method approach to identify significant differences among the dimensions of texts across all sections of the theses. While the research provided strong grounds for further studies and syllabus design, it was constrained by the exclusive focus on one genre (i.e. theses) and therefore its results could not be widely generalized. Similarly, Hussain et al. (2019) observed linguistic variations and moves through a corpus-based methodology by analyzing 100 English research article abstracts from multiple disciplines within hard and soft sciences sourced from ten online journals. Using MAT v.1.1 software, the mixed-method approach revealed the disciplines' shared concern with purpose moves and hard sciences' particular focus on conventional moves while considering other moves as optional. This linguistic analysis carried implications for novice writers seeking acknowledgment. The study's statistical results were suitable, although its limited corpus size and discipline-focused approach hindered broader generalization. Moreover, Yasmin et al. (2019) concentrated on rhetorical structures and linguistic variations within 32 introduction sections of Pakistani

theses in the social sciences and humanities. Employing a corpus-based methodology, the study relied on manual tagging analysis. Analysis, carried through Compleat Lexical Tutor, revealed variations in the frequency of moves. It also showed a strong use of topic generalization in humanities and inter-textual links to previous research in the sciences. Another study by Abid et al. (2022) applied MDA to investigate linguistic variations in the argumentative essays written by Chinese and Pakistani student writers. They obtained a corpus of 400 essays from the ICNALE repository and analyzed through MAT software. The results revealed ample variations in the essay writing styles of Chinese and Pakistani student writers. The results of this study, like those of Szczygłowska (2022, 2023) contribute to an understanding of the cross-cultural variations. Recently, N. Fatima et al. (2023) conducted a mixed-method corpus based study to investigate disciplinary variations in the abstracts of Pakistani dissertations from 16 disciplines. For this purpose, they utilized Biber's (1988) MDA approach. The results indicated clear disciplinary variations, that is, the dissertation abstract sections were predominantly found to be context independent, informational, non-narrative, and non-persuasive. Dimensions 3, 4, and 5 showed notably distinct variations. These findings offered potential benefits for academic writing pedagogy. While the study presents a detailed methodology and descriptive analysis, it is limited to manual tagging and a small corpus size as well as a focus on just two disciplines making broad generalization unreliable.

3. Research methodology

This section outlines the research methodology employed in this study, particularly focusing on how Biber's (1988) multidimensional analysis (MDA) framework guided the data collection and analysis processes.

3.1. Corpus collection

The corpus construction phase adhered to a meticulous process, ensuring data quality and alignment with Biber's (1988) MDA principles. This process completed in the following processes.

3.1.1. Targeted source selection

Articles were retrieved from online sources and academic journals (i.e. Pakistan Journal of Engineering and Applied Sciences, NUST Journal of Engineering Sciences, Technical Journal UET Taxila, Journal of Engineering & Applied Sciences) encompassing three disciplines (Engineering, Information Technology, and Literature). This selection reflects Biber's (1988) emphasis on considering register variation (disciplinary writing styles) in the corpus.

3.1.2. Data cleaning for linguistic features

Following initial retrieval and organization, the research articles underwent a thorough cleansing process. This involved removing extraneous elements like headers, footers, author information, and references. This step ensured the corpus primarily focused on the linguistic features of interest, as emphasized by Biber's (1988) MDA.

3.1.3. Segmentation and formatting

Each article was segmented isolating the discussion sections for the analysis. This approach is in line with Biber's (1988) MDA which targets particular sections within a text to examine register variation. Subsequently, the discussion sections were converted into plain text format using AntConverter. The conversion process was facilitated by Nitro Pro 10 software, and the files were renamed using Batch Renamer to ensure compatibility with linguistic analysis tools.

3.1.4. Corpus characteristics

The final corpus composition is summarized in Table 1 showcasing the distribution of data across disciplines and word counts. This transparency aligns with best practices in corpus construction. These comprehensive procedures resulted in a well-organized and refined corpus (see Table 1), ready for the subsequent phases of analysis.

Table 1. Distribution of data across disciplines

Sr. No.	Disciplines	No. of files	Word Count
1	Engineering	21	16,296
2	Information Technology	11	8,392
3	Literature	38	116,174
Total no. of files in the corpus = 70			140,862

3.2. Data analysis

The data analysis for this research underwent a systematic process, involving several key steps. The process began with data tagging and analysis through MAT tagger, employing various measures such as raw frequencies and their transformation into normalized frequencies. This transformation served to address the variation in text file lengths. Z-scores were then calculated, resulting in the computation of dimension scores. The interpretation of these scores was facilitated through illustrative examples derived from AntConc 3.5.9.0.

3.2.1. Multidimensional analysis tagger and feature selection

The data was tagged and analyzed using the MAT tool. Specific configurations were applied, including zero correction and “VASW” features, focusing on the linguistic features relevant to Biber’s (1988) MDA (e.g. noun phrases, verb tenses). Additionally, five dimensions (see Section 3.2.6) were chosen for analysis based on the research objectives and alignment with Biber’s (1988) proposed dimensions (e.g. information density, engagement).

3.2.2. Frequency analysis and normalization

Raw frequencies were initially calculated for all files per 1000 words. However, to account for text length variations and enable fair comparisons, these frequencies were normalized using the Formula 1. This normalization step is crucial in Biber’s (1988) MDA to ensure accurate representation of linguistic features across different sized texts.

Formula 1: Actual frequencies / Number of words in a file x 1000

3.2.3. Z-scores and dimension scores

Z-scores were calculated using the Formula 2 to further enhance the statistical rigor of the analysis. These Z-scores played a vital role in determining the dimension scores, which were computed using Formula 3. The dimension scores, derived by subtracting negative feature scores from positive feature scores, provided a comprehensive measure of linguistic variation within the discussion sections aligning with Biber’s (1988) MDA approach.

Formula 2: $z = x - \mu/\sigma$

Formula 3: D = sum of positive features (z-scores) – sum of negative features (z-scores)

3.2.4. Identifying significant features and interpretation

Z-scores with higher magnitudes were deemed statistically significant, indicating potentially interesting linguistic variations. Based on these scores, both underused and overused features were identified.

3.2.5. Data visualization and interpretation

The findings were presented in tables and graphs, offering a clear visual representation of the linguistic variations captured through the dimension scores. Additionally, to aid interpretation, specific examples were extracted using AntConc, providing concrete illustrations of the observed linguistic patterns.

This aligns with best practices in corpus analysis, where both quantitative and qualitative approaches are used to present the findings.

3.2.6. Description of textual dimensions

Biber's (1988) framework delineates five textual dimensions (described below) essential for analyzing linguistic variation in written discourse, each encapsulating distinct linguistic features reflecting diverse aspects of the language use. This framework facilitates simultaneous analyses to find multiple linguistics features (e.g. grammatical, lexical, and discourse) in the corpora (Biber 1988). Furthermore, this framework provides a useful tool for identifying salient features of different registers (N. Fatima et al., 2023) exploring linguistic features in the academic writing (Abid et al., 2022).

3.2.6.1. Dimension 1: Involvement versus informational production

This dimension measures the writer's level of personal involvement or detachment in the text. Academic writing typically prioritizes informational production, showcasing a detached and objective tone, whereas narrative or persuasive texts often display higher levels of involvement, conveying personal opinions, emotions, or attitudes.

3.2.6.2. Dimension 2: Narrative versus non-narrative concerns

This dimension examines the presence or absence of narrative elements within the text. This dimension helps the researchers delve into how narrative structure and storytelling techniques influence discourse coherence and organization, exploring plot development, character portrayal, and temporal sequencing.

3.2.6.3. Dimension 3: Explicit versus situation-dependent reference

This dimension explores the use of explicit reference markers and their reliance on situational context. It helps the researchers investigate how reference strategies like anaphora and deixis contribute to coherence and clarity across languages and discourse communities.

3.2.6.4. Dimension 4: Overt expression of persuasion

This dimension centers on the overt expression of persuasive intent within the text. It helps the researchers analyzes rhetorical appeals, persuasive language patterns, and argumentative structures across genres such as advertising and political discourse, revealing the persuasive power of language and its influence on audience attitudes.

3.2.6.5. Dimension 5: Abstract versus non-abstract information

This dimension concerns the level of abstraction or concreteness conveyed in the text. Through this dimension the researchers explore lexical choices, syntactic structures, and discourse patterns associated with abstract and concrete language use, elucidating how language mirrors cognitive processes and communicative goals.

4. Results

The results are represented in the Table 2 which includes the Z-scores of the lexico-grammatical features associated with different dimensions. Additionally, it highlights the maximum and minimum z-score values for each feature across the entire corpus. The range of z-scores and the standard deviation are provided, emphasizing their importance in deriving valuable insights from the MAT Tagger analysis.

Table 2. Linguistic distribution across the corpus of academic writing

Linguistic Features	Mean Value	Max Value	Min Value	Range	SD
Word Length	4.99	5.77	4.11	1.66	0.36
Type-Token Ratio	190.83	235	64	171	36.21
Amplifiers	0.18	1.5	0	1.5	0.24
Interdependent Clause Coordination	0.52	1.66	0	1.66	0.37
Be as main verb	1.54	3.49	0	3.49	0.62
By-passives	0.14	0.49	0	0.49	0.12
Causative adverbial subordinators	0.1	0.58	0	0.58	0.13
Concessive adverbial subordinators	0.06	0.29	0	0.29	0.08
Conditional adverbial subordinators	0.11	0.6	0	0.6	0.15
Conjuncts	0.56	2.2	0	2.2	0.45
Contractions	0.08	1.16	0	1.16	0.19
Demonstratives	0.89	2.59	0	2.59	0.48
Demonstrative pronouns	0.3	1.03	0	1.03	0.22
Discourse particles	0.02	0.24	0	0.24	0.04

Table 2. cont.

Linguistic Features	Mean Value	Max Value	Min Value	Range	SD
Downtoners	0.21	0.85	0	0.85	0.18
Emphatics	0.36	1.26	0	1.26	0.28
Existential there	0.16	0.84	0	0.84	0.19
First person pronouns	0.49	3.11	0	3.11	0.64
Gerunds	0.56	1.98	0	1.98	0.49
Hedges	0	0.05	0	0.05	0.01
Indefinite pronouns	0.02	0.16	0	0.16	0.04
Attributive adjectives	7.93	13.29	4.35	8.94	1.82
Necessity modals	0.08	1.18	0	1.18	0.17
Total other nouns	27.64	41.76	17.28	24.48	4.19
Nominalizations	3.8	7.15	1.11	6.04	1.35
Other adverbial subordinators	0.24	1.2	0	1.2	0.26
Agentless passives	1.51	5.2	0.34	4.86	1
Past participial clauses	0.08	0.75	0	0.75	0.15
Perfect aspect	0.56	2.51	0	2.51	0.51
Phrasal coordination	1.53	3.68	0.47	3.21	0.68
Total prepositional phrases	12.85	22.29	9.04	13.25	2.63
Pied-piping relative clauses	0.07	0.39	0	0.39	0.08
Pronoun it	0.83	2.26	0	2.26	0.54
Place adverbials	0.2	1.08	0	1.08	0.21
Possibility modals	0.44	2.22	0	2.22	0.4
Predicative adjectives	0.71	2.17	0	2.17	0.45
Present participial clauses	0.12	0.49	0	0.49	0.13
Private verbs	1.52	3.76	0.2	3.56	0.72
Predictive modals	0.18	0.9	0	0.9	0.22
Pro-verb do	0.05	0.29	0	0.29	0.07
Public verbs	0.55	1.93	0	1.93	0.45
Total adverbs	2.25	4.79	0.48	4.31	0.9

Table 2. cont.

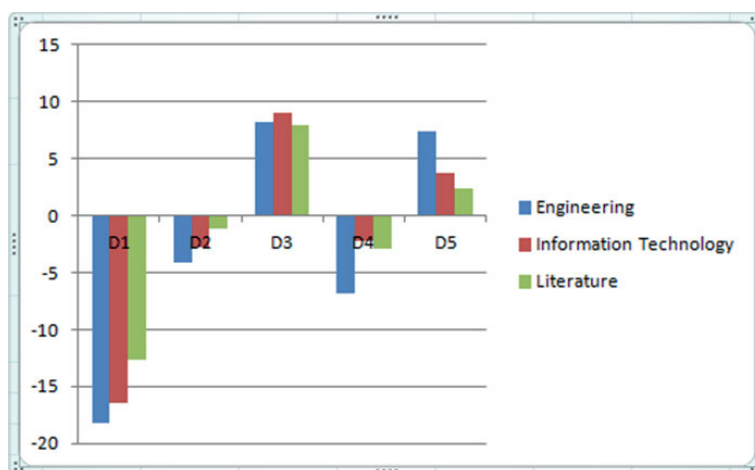
Linguistic Features	Mean Value	Max Value	Min Value	Range	SD
Sentence relatives	0.1	0.58	0	0.58	0.13
Seem appear	0.06	0.39	0	0.39	0.09
Split auxiliaries	0.27	0.81	0	0.81	0.18
Split infinitives	0.01	0.14	0	0.14	0.03
Second person pronouns	0.13	1.08	0	1.08	0.26
Stranded preposition	0.03	0.37	0	0.37	0.06
Suasive verbs	0.26	1.08	0	1.08	0.24
Synthetic negation	0.12	0.66	0	0.66	0.14
That adjective complements	0.03	0.33	0	0.33	0.06
Subordinator that deletion	0.11	0.77	0	0.77	0.14
That verb complements	0.42	1.58	0	1.58	0.34
Time adverbials	0.16	0.81	0	0.81	0.18
Infinitives	1.14	2.76	0	2.76	0.65
That relative clauses on object position	0.16	1.5	0	1.5	0.21
Third person pronouns	2.33	10.87	0	10.87	2.56
That relative clauses on subject position	0.16	0.75	0	0.75	0.16
Past tense	1.92	6.42	0	6.42	1.53
Present tense	5.63	10.82	1.19	9.63	2.11
WH-clauses	0.03	0.24	0	0.24	0.05
WH relative clauses on object position	0.04	0.33	0	0.33	0.07
Direct WH-questions	0.02	0.19	0	0.19	0.04
WH relative clauses on subject position	0.2	0.77	0	0.77	0.19
Past participial WHIZ deletion relatives	0.25	1.5	0	1.5	0.24
Present participial WHIZ deletion relatives	0.25	1.32	0	1.32	0.24
Analytic negation	0.49	2.11	0	2.11	0.42

4.1. Dimension scores

The dimension scores for the three disciplines have been presented in Table 3 and Graph 1.

Table 3. Dimension scores across disciplines

Disciplines	D1	D2	D3	D4	D5
Engineering	-18.22	-4.1	8.16	-6.86	7.44
Information Technology	-16.39	-2.74	8.95	-2.26	3.74
Literature	-12.71	-1.19	7.87	-2.84	2.38



Graph 1. Comparison of average dimension scores in discussion sections among three Disciplines in Pakistani English research articles

4.2. Interpretation of dimensions

The analysis of dimension scores in this study revealed varying degrees of significant differences among the three disciplines: engineering, information technology, and literature. In dimension 1, a substantial difference of 5.51 was observed between engineering and literature, while the difference between information technology and literature was 3.68. Notably, the difference between engineering and information technology was statistically insignificant (1.83). In dimension 2, distinctions were also noticeable: a difference of 2.91 existed between engineering and literature, while the gap between engineering and information technology was 1.36, and that between information technology and

literature was 1.55. These variances underscore the unique lexico-grammatical patterns within each discipline. Shifting to dimension 3, slight differences were observed. A minimal 0.79 difference was noted between information technology and engineering, while a slightly larger difference of 1.08 emerged between information technology and literature. The smallest difference, 0.29, was identified between engineering and literature in this dimension. Dimension 4 exhibited noteworthy variations. A substantial difference of 4.6 was found between engineering and information technology, while literature showed a considerable difference of 4.02 from the engineering discipline. However, only a slight difference was detected between information technology and the literature discipline. In dimension 5, significant differences were evident. Engineering displayed notable distinctions from both information technology (3.7) and literature (5.06). In contrast, the difference between information technology and literature was statistically insignificant (1.36). Overall, the dimension scores' differences ranged from 0.29 in dimension 3 to 5.51 in dimension 1, highlighting the unique linguistic variations across these disciplines.

5. Discussion

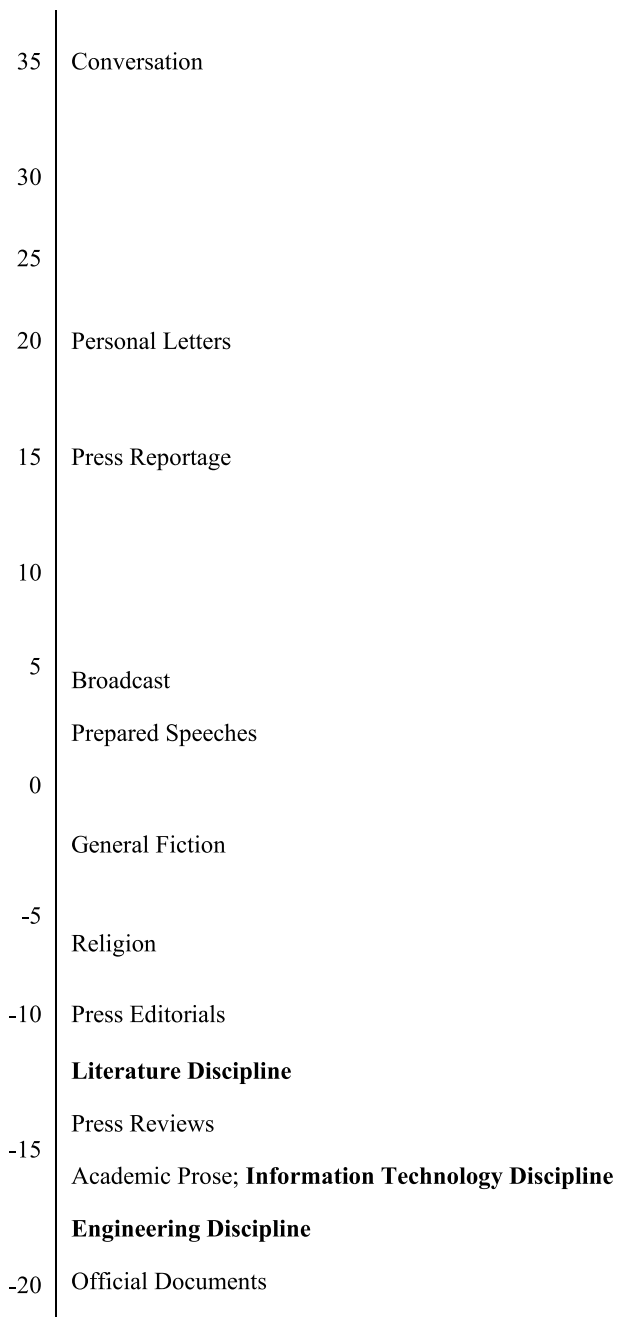
5.1. Linguistic variations across five dimensions

The study primarily focuses on linguistic variations within the discussion sections to ascertain whether the text in each discipline tends to be informational or interactional, narrative or non-narrative, explicit or situation-based, persuasive, and abstract or non-abstract. These variations across the disciplines are visually presented, offering graphical insights into the outcomes across the five dimensions.

5.1.1. Variations in discussion sections in relation to dimension 1

The Figure 2 provides an overview of dimension 1 within the three disciplines: engineering, information technology, and literature. The closest text type is interpreted by comparing these results with other text types from Biber's (1988) Multidimensional framework.

In terms of dimension 1, characterized by involved versus informational discourse, the discussion sections of research articles predominantly exhibit highly informational discourse. Among the three disciplines studied, the engineering discipline (-18.22) contains the highest degree of informational content, surpassing that of information technology (-16.39) and literature (-12.71) research article discussion sections. This difference arises from the overuse of features such as nouns (NN) and agentless passives [PASS] and the



Graph 2. Comparison of discussion sections of Pakistani English research articles with Biber's (1988) other genres on dimension

underuse of total adverbs (RB). Notably, the most significant variance in dimension scores is found between the discussion sections of engineering and literature research articles. The literature discipline stands out from the other two disciplines due to its overuse of the linguistic feature “sentence relatives” [SERE]. Employing Euclidean distance, the discussion sections of engineering articles are closely related to official documents, whereas those of information technology closely resemble the genre of press reportage, and the discussion sections of literature research articles bear a strong resemblance to academic prose. In sum, the text within the discussion sections closely aligns with the characteristics of scientific exposition. The co-occurring features in the three disciplines are given in Table 4.

Table 4. Standardized values of linguistics features

Co-occurring Patterns	Statistical Values of Interesting Variables
Nouns (NN)	2.7
Passives [PASS]	0.83
Sentence relatives [SERE]	2.28
Adverbs (RB)	-2.45

Examples 1

After the nuclear devastation, which **also** brings about the tragic **end** of her first **love**, **Hiroko** makes the **decision** to pursue **Konrad’s past** and travels to **India** alone, an almost unimaginable **thought** at the **time**. **Shamsie** makes it clear to the **reader**, almost immediately, that **Hiroko** is a woman who defies **norms** and resists **stereotypes**, and this **aspect** of her personality becomes deeply pronounced in her associations with the **Burtons**, a sophisticated and highly educated **English family** living in **India** during the **time** of the “**Empire**”.

Example 1 has been extracted from the discussion section of a literature research article. In this text, bold words represent nouns, italicized words indicate sentence relatives, and adverbs are highlighted in both bold and italics. The recurring nouns, such as “love” “decision” “Hiroko” “Konrad’s past” “India” “Shamsie” “English family” and “empire” frequently co-occur with other patterns in the text. This co-occurrence reflects the highly informational nature of the data. These nouns are content words used to convey meaning within the text, as demonstrated in the above example. The increased use of sentence relatives, such as “which” which is a feature of involved discourse, in the literature discipline’s discussion sections indicates that despite their informational focus, these sections incorporate more interactive elements than engineering and

information technology. This is done to captivate the readers and engage them in the text. The example also illustrates the underuse of the linguistic pattern “adverb” with the word “also”. This study is consistent with the findings of Biber (1988) and Rashid et al. (2017), indicating that the discussion sections in Pakistani research articles tend to be predominantly informational in nature.

The Figure 2 illustrates that conversation has very high mean scores, while official documents, academic prose, and press reportage have the lowest mean scores. Personal letters and prepared speeches have moderately high mean scores, whereas broadcast and general fiction have moderately low scores. The present study exhibits relatively low scores on dimension 1, aligning it with the genre of official documents, which is characteristic of academic writing. The dimension score for Biber’s (1988) academic prose (-14.09) differs from the current research (-14.08) by only 0.01, signifying that both exhibit similar patterns of variation across the linguistic features required for conveying information.

5.1.2. Variations in discussion sections based on dimension 2

The MAT analysis (see Figure 3) indicates that there are slight but statistically significant differences between the discussion sections of the engineering, information technology, and literature disciplines concerning dimension 2, which relates to the distinction between narrative and non-narrative discourse.

Based on dimension 2, which examines narrative versus non-narrative discourse, the discussion sections in research articles primarily contain non-narrative discourse. Notably, the engineering and literature disciplines lack interesting variables, while the information technology discipline exhibits an overuse of the negative score for the linguistic feature “AWL” emphasizing the non-narrative nature of the text. When considering Euclidean distance, the discussion sections in engineering are more closely related to broadcasts, those in information technology are closer to academic prose, and those in literature research articles bear a resemblance to conversation. Overall, the text in discussion sections aligns closely with scientific exposition. The co-occurring patterns with their respective scores are given in Table 5.

The Figure 3 illustrates that general fiction has the highest mean score, while broadcasts, academic prose, and official documents have low scores. Personal speeches, personal letters, and press reportage exhibit moderately high scores. On the other hand, conversation has moderately low scores. The ongoing study closely aligns with the genre of academic prose, as it exhibits low scores on this dimension. The small difference (0.42) between Biber’s (1988) academic prose (-2.6) and the dimension scores of Pakistani English discussion sections (-2.18) indicates a significant resemblance in the use of linguistic features that serve non-

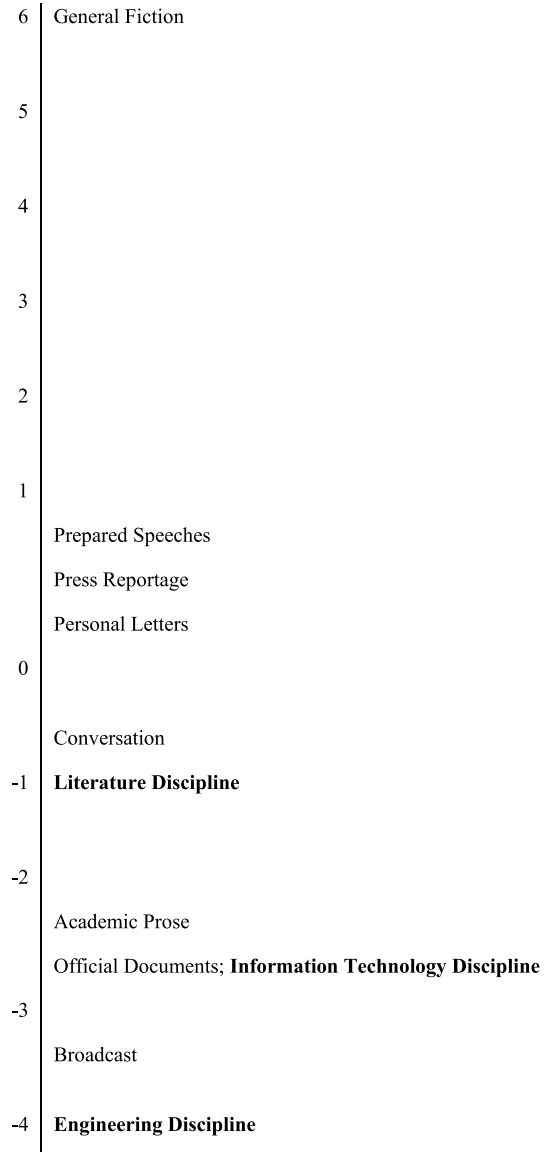


Figure 3. Comparison of discussion sections of Pakistani English research articles with Biber's other genres on Dimension 2

Table 5. Co-occurring linguistic pattern

Co-occurring Pattern	Scores
Word Length (AWL)	1.23

narrative functions in the text. These findings concur with those of Biber (1988) and the study by Mahmood et al. (2017), as the discussion sections are primarily non-narrative in nature, with literature displaying the least non-narrative discourse.

5.1.3. Variations in discussion sections according to the dimension 3

The results regarding this dimension reveal slight differences among the disciplines. All three disciplines exhibit explicit writing in their discussion sections. The Figure 4 represent the dimension scores of discussion sections for each discipline.

In dimension 3, which concerns explicit versus situation-dependent discourse, the discussion sections of research articles exhibit explicit discourse. The information technology discipline (8.95) displays the highest degree of explicit content in its discussion sections, in contrast to engineering (8.16) and literature (7.87). This is primarily due to the overuse of linguistic feature phrasal coordination (PHC) and the underuse of total adverbs (RB). Additionally, an overuse of nominalization features is observed in the discussion sections of information technology. Based on Euclidean distance, all three disciplines closely align with the genre of official documents. Overall, the text in the discussion sections leans towards scientific exposition. The co-occurring features are given in Table 6 and highlighted in Example 2.

Examples 2

Messages and reminders can be sent to patients regarding their pills using mobile phones. Similarly, infant and maternal **mortality** can be reduced by delivering advice through SMS pregnant mothers *and* health **information** to **community** health workers. The **availability** these ICT **facilities** may be used to get medical advice from health care providers at remote **locations** through the support centers working round the clock.

5.1.4. Variations in discussion sections according to the dimension 4

In dimension 4, there is a significant difference observed between the engineering discipline and both information technology and literature. The Figure 5 represents the closest text type of each discipline:

The engineering discipline (8.89) exhibits a distinct pattern from information technology (-4.48) and literature (-3.45) in the discussion sections. Engineering research articles show a marked preference for features like common nouns (NN) and lexical words (LW) while displaying less usage of personal nouns (PN) and proper nouns (NP). On the other hand, information technology and literature diverge from this pattern with an overuse of proper nouns (NP) and personal nouns (PN) in their discussion sections. These differences imply that the

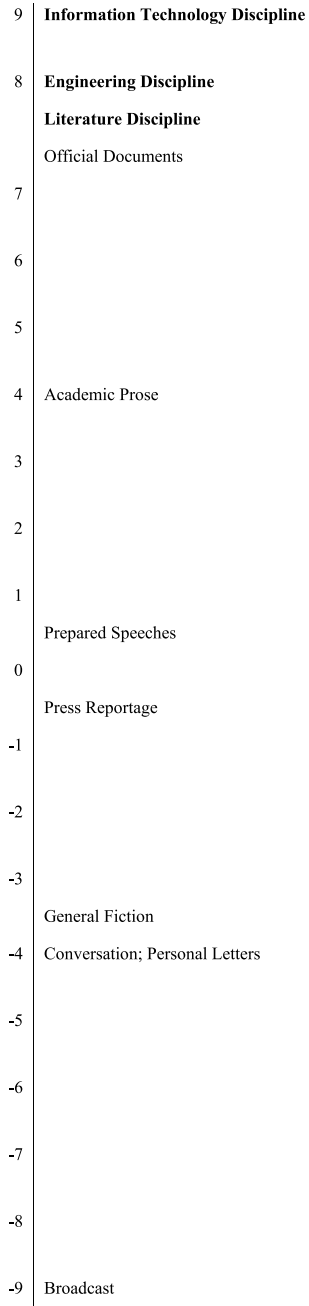


Figure 4. Comparison of discussion sections of Pakistani English research articles with Biber's other genres on dimension 3

Table 6. Co-occurring linguistic pattern

Co-occurring Patterns	Statistical Values of Interesting Variables
Phrasal Coordination (PHC)	4.4
Adverbs (RB)	-2.45

engineering discussions are significantly more context-independent and involved in nature compared to information technology and literature.

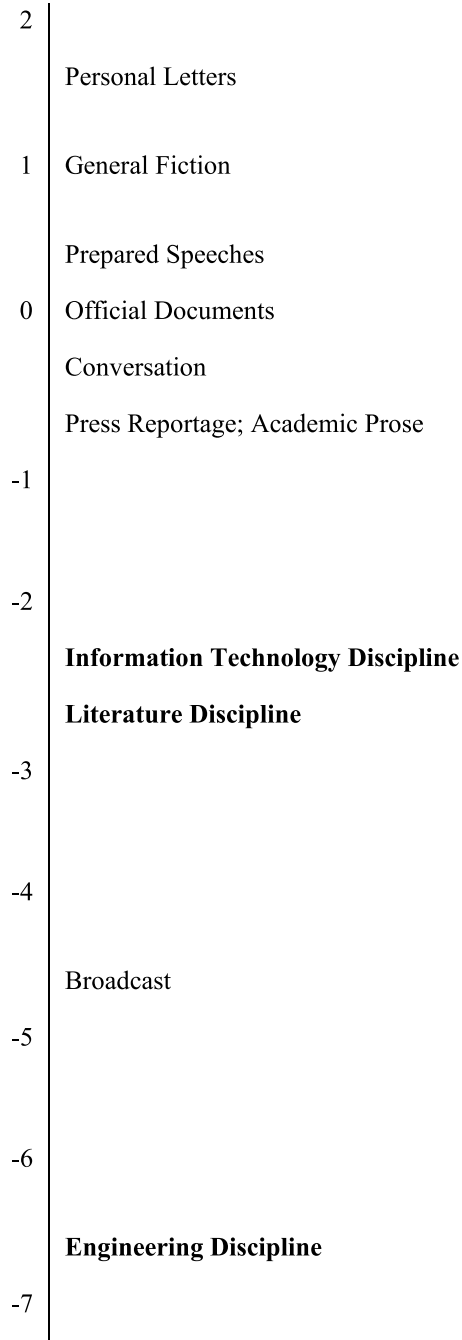
In the Figure 5, we can observe that official documents and academic prose have high mean scores, while broadcasts, conversation, personal letters, and general fiction have low scores. Personal speeches and press reportage have moderately high scores, whereas the ongoing study falls into the high score range, closely related to the genre of academic prose. The difference of 13.37 between the dimension scores of academic prose (-8.74) and the current study (4.63) suggests that Pakistani English discussion sections, especially in the engineering discipline, tend to exhibit context-independent and involved language features. These results support the findings of previous studies such as Biber's (1988) work, indicating that Pakistani research articles utilize language features in a context-independent manner, particularly in engineering discussions.

5.1.5. Variations in discussion sections according to the dimension 5

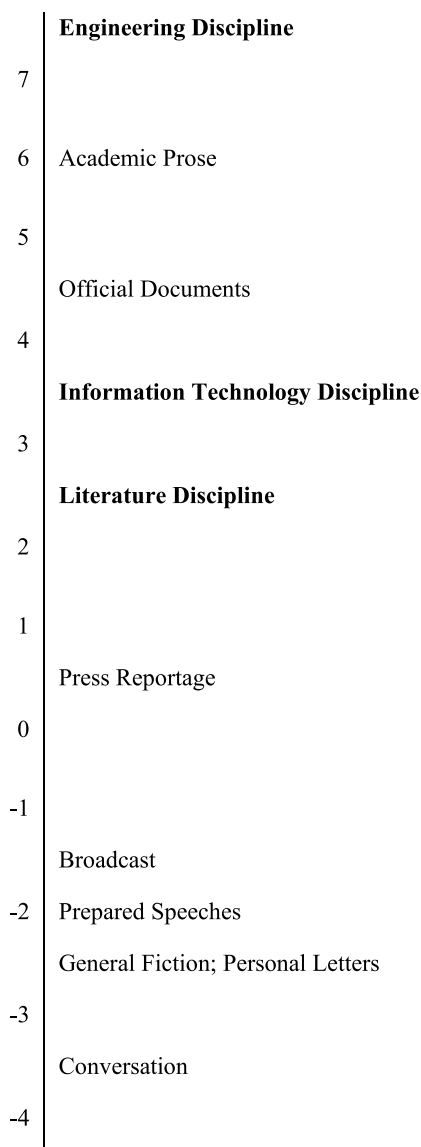
In dimension 5, there are significant differences in the discussion sections among the three disciplines i.e. engineering, information technology, and literature. The Figure 6 illustrates the characteristics of text used in the discussion sections of these disciplines.

The engineering discipline (8.89) exhibits a marked difference from information technology (-3.07) and literature (-5.06) in their discussion sections. Engineering discussions contain a higher usage of nominalizations and nouns (NN) while using fewer adverbs and pronouns (PP). In contrast, information technology and literature show a tendency to use more adverbs and pronouns (PP) and fewer nominalizations and nouns (NN) in their discussion sections. These differences suggest that the engineering discussions are significantly more abstract and less persuasive in comparison to information technology and literature.

The Figure 6 illustrates that personal letters, personal speeches, and general fiction have high mean scores, whereas academic prose, official documents, and conversation have low scores. Broadcasts and press reportage have moderately high scores. The current study falls within the high score range, indicating its similarity to personal letters, personal speeches, and general fiction. The difference of 0.01 between the dimension scores of academic prose (-14.09) and the present study (-14.08) suggests that the linguistic features in Pakistani English discussion sections align closely with academic prose. These results



Graph 5. Comparison of discussion sections of Pakistani English research articles with Biber's other genres on dimension 4



Graph 6. Comparison of discussion sections of Pakistani English research articles with Biber's other genres on dimension 5

indicate that Pakistani research articles in engineering exhibit high levels of abstraction and low levels of persuasion in their discussion sections, especially when compared to information technology and literature.

In summary, the results of this study have shown that Pakistani English research articles' discussion sections exhibit specific linguistic features in each of

the five dimensions. While these features vary between the three disciplines examined in the study, it is clear that the engineering discipline stands out as having the most distinct linguistic characteristics in comparison to information technology and literature. These findings provide valuable insights into the unique linguistic patterns present in Pakistani research articles and contribute to our understanding of academic writing in Pakistani English.

In dimension 5, which pertains to abstract versus non-abstract informational discourse, the discussion sections of research articles contain highly informational discourse. Notably, the engineering discipline (7.44) is characterized by the highest amount of abstract informational content. This is attributed to the overuse of linguistic features such as other adverbial subordinators (OSUB), Conjuncts (CONJ), agentless passives [PASS], and Past Participial Clauses [PASTP]. On the other hand, there is an underuse of features like total adverbs (RB) in engineering discussion sections.

The information technology discipline exhibits overuse of the linguistic feature Conjuncts (CONJ), contributing to the abstract informational nature of its discussion sections. Meanwhile, the analysis of the literature discipline's discussion sections suggests that they lack the use of linguistic features that play a significant role in creating variations.

By employing Euclidean distance, the discussion sections of the engineering discipline are found to be closely related to official documents, while those of Information Technology closely resemble the genre of press reportage. In contrast, the discussion sections of literature research articles bear a closer resemblance to academic prose. Overall, the text in discussion sections leans toward academic prose, reflecting their informative and abstract nature.

The results (Table 8) highlight the distinctive linguistic patterns present in the discussion sections of research articles across the three disciplines. While there are variations, the engineering discipline particularly stands out as having the most abstract and informational discourse, followed by information technology and literature. These insights provide valuable information about the linguistic characteristics of Pakistani research articles, especially in the context of academic writing in Pakistani English.

Table 7. Co-occurring patterns of interesting variables

Co-occurring Patterns	Statistical Values of Interesting Variables
Other adverbial Subordinators (OSUB)	1.3
Conjuncts (CONJ)	2.72
Agentless Passives (PASS)	0.83
Past Participial Clauses (PASTP)	1.76
Adverbs (RB)	-2.45

Example 3

The effect of different operating parameters on the performance and efficiency of RO membrane has been studied by changing one parameter at a time while the other parameter is kept constant. Experiments were carried out to investigate the effect of operating pressure on the performance of RO process.... *However* it has been noted from literature that a limiting value of feed pressure is reached beyond which permeate flux cannot be increased... On the other hand as the feed water pressure increases from 5 bar to 25 bar, salt rejection increases from 90.6% to 94.7% (calculated from equation 1).

Example 3 is extracted from the discussion section of an engineering discipline research article. It illustrates the technical and formal use of linguistic patterns, which align with abstract informational discourse. The bold word represents an adverbial subordinator, the italicized word represents a conjunct, and the underlined words represent agentless passives. Additionally, the bold and underlined word represents a past participial clause. In this example, the co-occurrence of the adverbial subordinator (OSUB) “while” and conjunct (CONJ) “however” with other words marks logical relations among the clauses. The use of agentless passives [PASS] with words like “is” “were” and “been” and past participial clauses [PASTP] using the term “calculated” emphasizes the information with less focus on the agent and more on the verb’s patient, indicating a reduced emphasis on the agent’s role.

In dimension 5, high scores are observed in academic prose and official documents. Press reportage exhibits moderately high scores. In contrast, general fiction, personal letters, personal speeches, broadcasts, and conversations have moderately low scores. The ongoing study shows high scores, aligning closely with the genre of official documents. A difference of 1.34 exists between the dimension scores, revealing that Biber’s (1988) academic prose (5.5) and the discussion sections written by Pakistani article writers (4.16) exhibit variations. The results indicate a higher presence of abstract informational discourse. Comparing these results to dimension scores obtained in Rashid et al.’s study (2017), the discussion section of the engineering discipline shows much higher dimension scores. In contrast, Information technology and literature discipline exhibit slightly lower dimension scores, indicating they are less informational compared to the engineering discipline’s discussion sections and the results of the mentioned study for this section. The dimension scores of Biber’s (1988) academic prose and the current study are provided in the Table 8, along with the differences between their dimension scores, which can aid in understanding these variations.

Based on these dimension scores and their differences, the existing variations between academic prose and the text of the discussion sections, which are a sub-genre of academic writing, can be determined.

Table 8. The dimension scores of Biber's (1988) academic prose and the current study

Dimension scores	Dimension scores of Biber's Academic prose	Dimension scores of three disciplines in the present study across the corpus	Difference between the dimension scores
D1	-14.09	-14.8	0.71
D2	-2.6	-2.18	0.42
D3	4.2	8.05	3.85
D4	-0.5	-3.99	3.49
D5	5.5	4.16	1.34

Collectively, the data closely aligns with learned exposition. Dimensions 1, 2, and 5 contain features that strongly resemble this exposition. Pakistani English, as a non-native variety, bears a greater resemblance to British English in these dimensions. In contrast, dimensions 3 and 4 exhibit a higher presence of linguistic features that differentiate the discussion sections from Biber's (1988) academic prose. This divergence in linguistic elements used by Pakistani research article writers in composing the discussion sections can be attributed to various external factors that influence their lexical choices.

6. Conclusion

The examination of linguistic variations in the discussion sections of research articles using the MAT Tagger has revealed significant differences in the use of linguistic features within dimensions 1, 3, and 5. However, dimensions 2 and 4 did not exhibit pronounced variations, lacking the presence of specific linguistic features responsible for these differences. Within dimension 1, all disciplines showed a high frequency of nouns and agentless passives, indicating the strongly informational nature of the text. In dimension 2, information technology displayed non-narrative characteristics, with no notable variations found in the discussion sections of engineering and literature research articles. Dimension 3 exhibited a shared pattern of discourse across the disciplines, with slight variations in textual formality, reflecting the use of phrasal coordination and nominalizations. For dimension 4, the presence of split infinitives did not contribute to the variation, likely due to their irregular distribution in just 18% of the text files. In dimension 5, the results pointed to the abstract informational nature of the text, primarily due to the overuse of conjuncts, passives, and past participial clauses. The discussion sections were also found to be rich in nouns, agentless passives, nominalizations, phrasal

coordination, conjuncts, passives, and past participial clauses, supporting the conclusion that these sections are highly informational and involve explicit and abstract discourse. These findings align with the objectives of the study and provide validation for the nature of the text within these various disciplines. While Kachru's model situates Pakistan within the outer circle of English language usage, making Pakistani English speakers norm followers, the research results indicate that Pakistani English is a distinct non-native variety that closely resembles native English. Therefore, Pakistani English should be considered closer to the borderline of the inner circle due to its characteristics that align with native English, warranting recognition as a distinct variety (see N. Fatima et al. 2023; Siddique et al., 2022).

Nonetheless, this study has certain limitations, particularly its focus on discussion sections within only three disciplines (engineering, information technology, and literature), limits its generalizability. The relatively small corpus size, consisting of 70 text files, and the utilization of MAT Tagger, while reliable, is not as dependable as Biber's multidimensional tagger. This research can serve as a valuable resource for future scholars investigating linguistic variations within research article discussion sections and can offer guidance to novice writers regarding lexical choices within specific registers. Furthermore, the study's results may be compared with discussion sections from other disciplines or academic writing genres. The corpus created for this research has the potential for further studies and the development of specialized dictionaries for specific genres using tools like AntConc or WordSmith. This multi-dimensional analysis underscores the need for additional research by Pakistani scholars to examine the linguistic variety within Pakistani research articles, and comparative studies will help identify distinctive features of Pakistani English as a non-native variety.

References:

- Abid, A., H. Manzoor, and A.R. Siddique 2022. Cross-cultural examination of argumentative English essays: A multidimensional analysis of Pakistani and Chinese learners. *Linguistic Forum* 4(4): 40-48. <http://doi.org/10.53057/linfo/2022.4.4.6>.
- Ahmad, M., M.A. Mahmood, and A.R. Siddique 2019. Organisational skills in academic writing: A study on coherence and cohesion in Pakistani research abstracts. *Languages* 4(4): 1-26. <https://doi.org/10.3390/languages4040092>.
- Ahmad, M., M.A. Mahmood, and A.R. Siddique 2022. Exploring disciplinary variation in Pakistani academic writing: A corpus-based research on doctoral dissertations. *Pakistan Languages and Humanities Review* 6(4): 51-60. [https://doi.org/10.47205/plhr.2022\(6-IV\)06](https://doi.org/10.47205/plhr.2022(6-IV)06).

- Ahmad, M., M.A. Mahmood, and A.R. Siddique 2023a. Features characterizing academic writing: A corpus-based research on dissertations from hard and soft science disciplines. *City University Research Journal of Literature and Linguistics* 6(1): 61-78.
- Ahmad, M., M.A. Mahmood, and A.R. Siddique 2023b. Determining the L2 academic writing development stage: A corpus-based research on doctoral dissertations. *International Review of Applied Linguistics in Language Teaching*. <https://doi.org/10.1515/iral-2023-0028>.
- Ahmad, M., M.A. Mahmood, and A.R. Siddique 2023c. Variation in academic writing: A corpus-based research on syntactic features across four disciplinary divisions. *Novitas-ROYAL (Research on Youth and Language)* 17(2): 50-65. <https://doi.org/10.5281/zenodo.10015816>.
- Alamghir, A., and A. Shakir 2023. Effect of research design on distribution of prototypical participant-oriented lexical bundles: A corpus-based study of quantitative research articles. *Linguistic Forum* 5(4): 57-73. <http://doi.org/10.53057/linfo/2023.5.4.3>.
- Ali, S. 2020. *Multidimensional corpus based analysis of newspaper reportage: A comparative study of Pakistani, other South Asian, and British newspapers* (Doctoral dissertation). Department of English, Air University, Islamabad, Pakistan.
- Al-Shujairi, Y.B.J., and F.A.J. Al-Manaseer 2022. Backgrounding the discussion section of medical research articles. *Open Journal of Modern Linguistics* 12(1): 71-88. <https://doi.org/10.4236/ojml.2022.121008>.
- Amnuai, W. 2019. The textual organization of the discussion sections of accounting research articles. *Kasetsart Journal of Social Sciences* 40(2): 389-394. <https://doi.org/10.1016/j.kjss.2017.10.007>.
- Azher, M., and M.A. Mahmood 2016. Exploring variation across Pakistani academic writing: A multidimensional analysis. *NUML Journal of Critical Inquiry* 14(2): 86-113.
- Azher, M., M.A. Mahmood, and S.I. Shah 2018. Linguistic variation across research sections of Pakistan academic writing: A multidimensional analysis. *International Journal of English Linguistics* 8(1): 30-43. <http://doi.org/10.5539/ijel.v8n1p30>.
- Barzan, P., and B. Heidary 2019. Language variation. <https://doi.org/10.13140/RG.2.2.36761.36969>.
- Biber, D. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D. 2008. Corpus-based analyses of discourse: Dimensions of variation in conversation. In V. Bhatia, J. Flowerdew and R.H. Jones (eds.), *Advances in Discourse Studies*, 110-124. London and New York: Routledge.
- Fatima, M., A.R. Siddique, M. Ahmad, and M.A. Mahmood 2023. Exploring linguistic variation in Pakistani English newspaper editorials through multidimensional analysis. *Newspaper Research Journal* 44(4): 425-451. <https://doi.org/10.1177/07395329231194005>.

- Fatima, N., A.R. Siddique, and M. Ahmad 2023. Linguistic variations in the abstracts of Pakistani dissertations: A multidimensional analysis across disciplines. *University of Chitral Journal of Linguistics and Literature* 7(1): 50-80. <https://doi.org/10.33195/btm19817>.
- Ferguson, C.A. 1983. Sports announcer talk: Syntactic aspects of register variation. *Language in Society* 12(2): 153-172. <https://doi.org/10.1017/S0047404500009787>.
- Holmes, R. 1997. Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes* 16(4): 321-337. [https://doi.org/10.1016/S0889-4906\(96\)00038-5](https://doi.org/10.1016/S0889-4906(96)00038-5).
- Hussain, Z., S. Qasim, and M.A. Mahmood 2019. Linguistic properties of Pakistani research article (RA) abstracts from multiple disciplines. *AL-Qalam* 24(2): 165-183.
- Jin, B. 2018. A multidimensional analysis of research article discussion sections in the field of chemical engineering. *IEEE Transactions on Professional Communication* 61(3): 242-256. <https://doi.org/10.1109/TPC.2018.2817002>.
- Jin, B. 2021. A multi-dimensional analysis of research article discussion sections in an engineering discipline: Corpus explorations and scientists' perceptions. *Sage Open* 11(4): 1-17. <https://doi.org/10.1177/21582440211050401>.
- Khamaiseh, M. 2023. Toward a model for analyzing the rhetorical move structure of the master thesis introductions in applied linguistics. *Linguistic Forum* 5(4): 1-29. <http://doi.org/10.53057/linfo/2023.5.4.1>.
- Liu, J., and L. Xiao 2022. A multi-dimensional analysis of conclusions in research articles: variation across disciplines. *English for Specific Purposes* 67: 46-61. <https://doi.org/10.1016/j.esp.2022.04.003>.
- Louvigné, S., J. Shi, and S. Sharmin 2014. A corpus-based analysis of the scientific RA genre and RA introduction. In *Proceedings of the 2014 International Conference on Advanced Mechatronic Systems*, 123-127. IEEE. <https://doi.org/10.1109/ICAMechS.2014.6911636>.
- Mu, C. 2021. A multidimensional contrastive analysis of linguistic features between international and local biology journal English research articles. *Scientometrics* 126(9): 7901-7916. <https://doi.org/10.1007/s11192-021-04102-x>.
- Parkinson, J. 2011. The Discussion section as argument: The language used to prove knowledge claims. *English for Specific Purposes* 30(3): 164-175. <https://doi.org/10.1016/j.esp.2011.03.001>.
- Rashid, A., M.A. Mahmood, and S. Ahmad 2017. Linguistic variation across research sections: A multidimensional analysis of Pakistani academic journal articles. *Global Language Review* 2(1): 15-37. [https://doi.org/10.31703/glr.2017\(II-I\).02](https://doi.org/10.31703/glr.2017(II-I).02).
- Shah, S.A.A., Kifayat ullah, and N. Ahmed 2023. Vagueness in communication: A corpus driven study of academic discourse. *Linguistic Forum* 5(3): 1-11. <http://doi.org/10.53057/linfo/2023.5.3.1>.
- Siddique, A.R., M.A. Mahmood, and M. Ahmad 2022. Diachronic variation in the language of Pakistani English newspapers: A multidimensional analysis. *CORPORUM: Journal of Corpus Linguistics* 5(2): 109-127.

- Szczygłowska, T. 2022. Lexical verbs of epistemic modality in academic written English: Disciplinary variation. *Linguistica Silesiana* 43: 91-111. <https://doi.org/10.24425/linsi.2022.141219>.
- Szczygłowska, T. 2023. Adverbial cohesion in English-medium academic prose: Disciplinary and linguacultural considerations. *Linguistica Silesiana* 44(2): 81-108. <https://doi.org/10.24425/linsi.2023.146648>.
- Xie, S. 2020. Multidimensional analysis of master thesis abstracts: A diachronic perspective. *Scientometrics* 123(2): 861-881. <https://doi.org/10.1007/s11192-020-03408-6>.
- Yasmin, T., M.A. Mahmood, and I.H. Butt 2019. Rhetorical analysis of introduction sections from humanities and sciences research theses authored by Pakistani scholars. *NUML Journal of Critical Inquiry* 17(2): 79-93.