University of Chitral
JOURNAL OF LINGUISTICS AND LITERATURE

CrossMark

ISSN
p 2617-3611
e 2663-1512

Crossref

Linguistic Variations in the Abstracts of Pakistani Dissertations: A Multidimensional Analysis across Disciplines

Nisa Fatima[1], Ali Raza Siddique[2], *Muhammad Ahmad[3]

[1]BS Student, Department of Applied Linguistics, Government College University, Faisalabad, Pakistan
[2]PhD Candidate, Department of Applied Linguistics, Government College University, Faisalabad, Pakistan
[3]PhD Candidate, Department of Applied Linguistics, Government College University, Faisalabad, Pakistan

A R T I C L E I N F O

A B S T R A C T

This mixed-method, corpus-based study investigates across disciplinary variations in the abstracts of Pakistani dissertations spanning 16 disciplines in the light of Biber's multidimensional analysis approach. While prior research has explored variations in the academic register of Pakistani English, a focused study on the abstract section has largely been ignored. Addressing this oversight, this research is conducted on a specially developed corpus of 72,702 words and analyzed using the MAT tagger. The findings indicate that, although there are evident disciplinary variations, the abstracts predominantly exhibit characteristics of being informational, non-narrative, context-independent, and non-persuasive. Notably, distinct variations emerged across disciplines in D3 (Dimension 3), D4, and D5, with Law as a notable outlier. These results support the idea that Pakistani English is a separate linguistic entity with unique characteristics.

Introduction

Pakistani English (both spoken and written) exhibits variations in different contexts within Pakistan, particularly evident in the realm of academic writing. These variations are prominently observable in abstracts from dissertations, offering an insightful view into the alignment or deviation from British English standards. Exploring this aspect further presents a promising avenue for research. Numerous studies have also drawn attention to the linguistic variations present within the Pakistani English. Past research frequently employed multidimensional analysis techniques to find the distinct linguistic attributes that characterize Pakistani English as a unique variety within the broader spectrum of Englishes worldwide. This study continues in the tradition of such research while also advancing it by investigating the multidisciplinary linguistic variations found in abstracts of Pakistani dissertations using the Multidimensional Analysis Tagger (MAT). The corpus compiled for this purpose spans a wide array of academic disciplines (Table 1), aiming to create a more comprehensive and inclusive dataset. The central argument of this study revolves around the assertion that Pakistani English constitutes an independent form within the family of World Englishes. The primary objective is to study the multidisciplinary linguistic variations that emerge in Pakistani academic writing, with a specific focus on its distinctive register identity.

The term "linguistic features" encompasses various lexico-grammatical elements that shape the language of a text, whether it is spoken or written (Moohebat, Raj, Thorleuchter & Kareem, 2017). Considerable literature has already been devoted to the study of linguistic variation, particularly concerning lexico-grammatical features and their comparative analysis. While some studies have explored differences in rhetorical and generic structures, and textual

patterns. The current study focuses on lexico-grammatical variations, which is consistent with Biber's (1999) claim that this method provides the best way to examine language registers.

Literature Review

Theoretical Background

Past research has utilized various models and frameworks to examine linguistic variations. These include Swales' (1990) CARS model, Halliday's systemic functional linguistics model, Hyland's (2000) rhetorical analysis model, analytical models by Kuhi (2008) and Azar (2012), a fusion of Kanoksilapatham (2007), Nwogu (1991), and Swales' (1990) models, Halliday's thematic organization model (1994) and its revised McCabe (1999) version, and Biber's (1986) MDA model. Among these, the MDA stands out as the most appropriate for studying Pakistani English registers, while models like Swales' CARS, Hyland's, and Halliday's frameworks have limited adaptability here. Therefore, this study employs Biber's MDA.

Definitional Background

Language variations emerge from social, regional, and situational distinctions. Hymes (1984) underscores the significance of the "verbal repertoire" (register variations) in linguistic studies, prioritizing it over the language itself. Register signifies a language variety shaped by its social usage context. Biber and Conrad (2009) argue that terms such as register, genre, and style do not represent distinct variants but offer different angles for studying text varieties. Halliday, on the other hand, presents a broader view of registers without clear demarcations. Biber's (1999) approach diverges by accentuating lexico-grammatical features in texts, avoiding inferences about linguistic features solely from context. Biber's (2000) method examines linguistic commonalities across diverse text types. This study centers on Biber's approach due to its focus on analyzing co-occurring linguistic features in academic prose, especially pertinent within the context of Pakistani English.

Disciplinary Variations in Academic Writing

Numerous studies have delved into understanding the variations in academic writing across disciplines. These investigations range from analyzing structural to linguistic variations in different academic genres. For example, Samraj (2002) assessed the applicability of Swales' CARS model by examining structural variations in research articles from wildlife behavior and conservation biology. The study offered an enhanced CARS model with increased flexibility, though its methodology was critiqued. Another study (Samraj, 2008) employed Swales' CARS model to study structural discrepancies in American master's theses across biology, philosophy and linguistics, and highlighted differences in the representation and discussion of prior work. Similarly, Atai and Habibie (2009) focused on the generic structure of research article introductions from different sub-disciplines of applied linguistics using Swales' CARS model. Their findings underscored the variability of generic structures across sub-disciplines. Another research (Mwinlaaru & Afful, 2012) also analyzed the rhetorical structure of dissertation acknowledgments in the education sector, and observed variation in the form of a consistent three-move structure. Gray (2013) also performed a corpus-based study on linguistic variations in research articles across different disciplines and sub-register and reported variations across disciplines, paradigms, and other parameters. These studies were followed by a number of recent studies. For example, Chen (2017) investigated statements of purpose across five disciplines in China, and revealed the influence of disciplinarity on academic discourse. It was followed by Ebrahimi and Khedri (2019) which studied research article abstracts from chemical engineering and applied linguistics based on Halliday's thematic organization model, and concluded in noting thematic structure differences related to the discipline of the writer. Similarly, Alotaibi (2020, 2021) conducted two studies: the former explored research article abstracts across six disciplines using systematic functional linguistics model, and the latter investigated the rhetorical structure of textbook prefaces across different disciplines using Kuhi's and Azar's models.

Register-Based Studies on Pakistani English

Pakistani English has increasingly captivated the researchers interested in elucidating its unique linguistic traits and registers via MDA. This segment offers a succinct summary of important investigations within this domain.

Mahmood (2009) conducted a comparative study using a corpus of 2.1 million words from 29 Pakistani written English registers and compared against Native American and British English. Despite initially employing multiple taggers, complications led to reliance on the CLAWS tagger. The analysis identified notable features of Pakistani English. Though the study was focused on written registers however the corpus size raised concerns about conclusive outcomes. Azher and Mehmood (2016) compiled an 8.38 million words corpus of Pakistani academic writing from 235 research theses. Employing Biber's tagger and characterized Pakistani academic writing as

informational, non-narrative, non-persuasive, explicit, and impersonal, bearing a distinct linguistic profile. Hussain, Mahmood and Azher (2016) conducted dual studies on the ICE corpus of Pakistani English, both illuminating significant register variations. Their analysis of the 952,563-word corpus spanning 12 registers unveiled four distinct textual dimensions. The results questioned the suitability of Biber's (1988) MDA for non-native English variants.

Abbas (2018) applied MDA to a 2.1 million words corpus of Pakistani Written English, revealing considerable linguistic disparities across registers. A comparison with British Written English affirmed Pakistani English's uniqueness, yet the study focused on written forms and corpus imbalance constrained its scope. Shakir (2020) compared online Pakistani English registers to offline counterparts and American English variants. The study emphasized the distinct nature of online Pakistani English while critiquing its limited communicative capacity. However, it did not delve into structural features. Aziz, Pathan and Ali (2016) analyzed linguistic characteristics in two key disciplines of Pakistani academic writing. Results showed the divergence of this academic language from Biber's genres. While the methodology was robust, the study identified potential enhancements in data range and MDA methodology. Azher, Faiz, Izhar, Nisa and Ali (2019) analyzed 235 M. Phil. and Ph. D. theses from diverse disciplines. Their analysis unveiled substantial disciplinary variations, highlighting the unique character of Pakistani English in academic writing. Nevertheless, the study's focus was limited to academic registers. Rashid and Mahmood (2019) explored linguistic discrepancies across three disciplines in academic writing, illustrating the direct, descriptive, and neutral nature of the Pakistani academic register. While the methodology was strong, data selection and the absence of explicit research questions constrained the study. In conclusion, these studies collectively underscore the distinctiveness of Pakistani English in diverse registers. However, their predominant concentration on written forms underscores the necessity for comprehensive research encompassing both written and spoken domains. Therefore, this study is aimed to fill this gap.

Methods

Data Collection

The initial phase involved the selection of 16 disciplines to encompass a broad spectrum of academic writing. These disciplines comprised 16 subjects related to the four academic disciplines (see Table 1). From each discipline a random selection of ten dissertations was made. These Pakistani dissertations, sourced to study across-disciplinary variations, were accessed from the HEC directory, which is open to the public. The retrieved PDF source files were processed using AntFileSplitter software that segments PDF files. This software was configured to divide and save the data into specific sections, converting them into txt files. However, some dissertations (found in scanned format) were not compatible with AntFileSplitter. For these, Google Lens was utilized to scan and extract the abstract sections, which were subsequently saved as notepad files. Both the AntFileSplitter and Google Lens-derived notepad files were imported into Advanced Renamer, a software that batch renames files based on predefined settings. The naming convention adopted was:

File Name = Incrementing number_Discipline Initial (Field Initial) Researcher
Gender (M or F)_Year of Publication.abstract

For instance, an abstract from agriculture within life sciences, penned by a male researcher in 2014, would be labeled as shown in 1.

1.   1.A(LS)M_2014.abstract

While the files were not subjected to meticulous manual cleansing, formulas and equations that might distort the data were removed. A few files had minor scanner-induced errors. It was observed that not all dissertations had abstract sections, reducing the count from 160 to 147 abstracts, summing up to 72,702 words in total.

Table 1. Description of Data Distribution

Study Field	Discipline	Number of Dissertations	Number of Abstracts	Word Count
Arts & Humanities	English Literature	10	10	3279
	Linguistics	10	10	4111
	History	10	9	5273
	Philosophy	10	8	2768
Life Sciences	Agriculture	10	9	4748

	Biology	10	6	3799
	Food Sciences	10	10	5692
	Psychology	10	10	5354
	Computer Science	10	10	4846
Physical Sciences	Engineering	10	10	4350
	Mathematics	10	8	4201
	Physics	10	10	4958
	Economics	10	10	6139
Social Sciences	Law	10	8	4163
	Politics	10	10	4777
	Sociology	10	9	4244
Total		160	150	72702

Source: Authors

Data Analysis

The MAT tagger by Nini (2019), an adaptation of Biber's (1988) software, was employed for the data analysis. Within the MAT tagger interface, 'zero correction' was activated, and counting was set to 'only VASW tags'. The 'tag and analyse' option was selected, followed by uploading the data for a specific discipline. Default settings were retained for the type/token ratio at 400 tokens. Analysis was conducted on the first five dimensions, omitting the sixth due to its limited and intricate variables. This procedure was consistently applied to all disciplines. Initially, the tagger grammatically labeled the data, encompassing 67 linguistic attributes using the Stanford tagger. Subsequent to tagging, raw frequencies for individual files were ascertained. Given the varying word counts of each file, raw frequencies were normalized using a formula (F-1).

F-1. FNormal frequency of 'x' feature = Raw frequency of 'x' feature/ Word Count
of the file x 1000

Post-normalization, the tagger determined collective mean values, ranges, and standard deviations of the 67 linguistic features. With the derived standard deviation and mean, the zscore was calculated to assess data spread, determine variation-driving features, and compute dimension scores employing F-2.

F-2. $z = x - \mu/\sigma$ (where σ represents standard deviation).

Within each dimension, linguistic features with a zscore of 2 or greater, termed 'interesting variables' were vital for deducing the relative dimension. Dimensions were only considered if they possessed five or more such variables, as those with fewer were challenging to interpret. The dimension scores were then determined using F-3.

Dimension score = sum of +ve features (zscores) - sum of -ve features (zscores)

The positive or negative association of features implied that an increase in one would result in a decrease of the other. In cases where dimensions lacked negative features, the dimension score was calculated simply through F-4.

F-4. Dimension score = sum of zscores of +ve features

Finally, the text type was defined using dimension scores, which was then graphically represented in relation to Biber's established text types.

Results

The data was analyzed based on the first five dimensions, excluding the sixth dimension due to its deemed insignificance. Table 2 shows the descriptive statistics for the chosen dimensions.

Table 2. Description of Linguistic Features across Pakistani Academic Writing

Linguistic Features	Mean Value	Maximum Value	Minimum Value	Range	Standard Deviation

AWL	5.51	6.57	4.18	2.39	0.34587
TTR	185.58	237	90	147	27.32909
AMP	0.1	0.95	0	0.95	0.185117
ANDC	0.47	3.1	0	3.1	0.464069
[BEMA]	1.05	3.59	0	3.59	0.635668
[BYPA]	0.17	1.09	0	1.09	0.216729
CAUS	0.04	1.01	0	1.01	0.121834
CONC	0.04	0.52	0	0.52	0.102422
COND	0.03	0.54	0	0.54	0.085763
CONJ	0.52	1.77	0	1.77	0.417876
[CONT]	0	0.16	0	0.16	0.013242
DEMO	0.91	2.51	0	2.51	0.660605
DEMP	0.13	0.81	0	0.81	0.171391
DPAR	0.01	0.33	0	0.33	0.040114
DWNT	0.11	0.63	0	0.63	0.152418
EMPH	0.21	1.59	0	1.59	0.278199
EX	0.08	0.65	0	0.65	0.145452
FPP1	0.27	4.32	0	4.32	0.641382
GER	0.65	5.14	0	5.14	0.787019
HDG	0	0.19	0	0.19	0.02474
INPR	0	0.21	0	0.21	0.027149
JJ	10.28	17.97	4.63	13.34	2.675685
NEMD	0.05	1.11	0	1.11	0.144923
NN	29.87	42.17	18.4	23.77	4.912995
NOMZ	5.56	11.15	1.85	9.3	1.956568
OSUB	0.2	1.27	0	1.27	0.263323
[PASS]	1.85	3.9	0	3.9	0.790353
[PASTP]	0.09	0.64	0	0.64	0.143226
[PEAS]	0.5	3.5	0	3.5	0.597285
PHC	1.99	5.33	0	5.33	0.967709
PIN	12.73	16.97	7.93	9.04	1.678088
[PIRE]	0.06	0.54	0	0.54	0.110716
PIT	0.65	3.09	0	3.09	0.585178
PLACE	0.18	1.71	0	1.71	0.292151
POMD	0.27	1.79	0	1.79	0.300489
PRED	0.47	1.65	0	1.65	0.426614
[PRESP]	0.11	0.95	0	0.95	0.173554
[PRIV]	1.14	4	0	4	0.723142
PRMD	0.12	2.76	0	2.76	0.348059
[PROD]	0.01	0.68	0	0.68	0.074554
[PUBV]	0.27	1.82	0	1.82	0.302061
RB	1.86	5.07	0	5.07	0.919089
[SERE]	0.1	1.05	0	1.05	0.182993
[SMP]	0.02	0.32	0	0.32	0.058451
[SPAU]	0.35	1.63	0	1.63	0.33849
[SPIN]	0.02	0.46	0	0.46	0.066585
SPP2	0	0.54	0	0.54	0.044691
[STPR]	0.01	0.44	0	0.44	0.058087
[SUAV]	0.41	2.08	0	2.08	0.477066
SYNE	0.04	0.75	0	0.75	0.104753
THAC	0.01	0.26	0	0.26	0.031447
[THATD]	0.09	1.01	0	1.01	0.171506
THVC	0.3	1.41	0	1.41	0.311536
TIME	0.12	0.94	0	0.94	0.206915
TO	1.28	4.05	0	4.05	0.701543
TOBJ	0.07	1.26	0	1.26	0.155657

TPP3	0.46	3.49	0	3.49	0.595152
TSUB	0.17	1.28	0	1.28	0.270032
VBD	2.25	6.57	0	6.57	1.729931
VPRT	4.09	8.89	0.33	8.56	2.026306
[WHCL]	0.01	0.73	0	0.73	0.073722
[WHOBJ]	0.01	0.27	0	0.27	0.045009
[WHQU]	0.01	0.57	0	0.57	0.058882
[WHSUB]	0.17	1.05	0	1.05	0.248845
[WZPAST]	0.34	1.45	0	1.45	0.332438
[WZPRES]	0.45	2.31	0	2.31	0.413062
XX0	0.22	1.63	0	1.63	0.288405

Source: Authors

The Table 2 illustrating descriptive statistics reveals that the majority of mean and maximum values for linguistic features hover around or below 1. Nonetheless, outliers are present, as evidenced by significantly elevated mean and maximum values for certain features, such as attributive adjectives [JJ] (mean=10.28, max=17.97), nouns [NN] (mean=29.87, max=42.17), nominalizations [NOMZ] (mean=5.56, max=11.17), and prepositional phrases [PIN] (mean=12.73, max=16.97). The heightened mean and maximum values of these features suggest a predominantly informational nature of the data, which will be delved into more deeply in subsequent sections. The minimal values for many of the features set at zero, signifying their sparse use within the data. Interestingly, the features with high or non-zero minimum values mirror those with elevated mean and maximum values. This, coupled with the range and standard deviation trends, hints at the data's non-interactive nature. A comprehensive breakdown of these data variations is outlined below.

Dimension Scores

Linguistic variations across the 16 disciplines are examined through five dimensions delineated by Biber (1985). The average dimension score of a dimension corresponds to a principal text type encapsulated within that dimension. Table 3 presents the average dimension scores for each discipline, organized by dimension.

Table 3. Dimension Scores of 16 Disciplines

Discipline	D1	D2	D3	D4	D5
English Literature	-22.89	-3.42	10.08	-4.74	3.83
Linguistics	-20.08	-3.35	9.13	-4	2.83
History	-22.31	-3.61	10.73	-4.39	4.81
Philosophy	-18.78	-4.92	9.29	-1.04	4.62
Agriculture	-23.55	-4.79	9.4	-4.9	5.55
Biology	-24.45	-4.79	9.4	-5.34	6
Food Sciences	-26.9	-4.79	9.35	-6.77	7.37
Psychology	-23.75	-4.76	10.38	-5.17	2.39
Computer Science	-22.41	-4.02	10.04	-1.82	3.91
Engineering	-22.74	-4.31	9.5	-2.3	5.29
Mathematics	-20.73	-5	11	-1.96	3.8
Physics	-23.12	-4.83	8.66	-4.69	4
Economics	-21.03	-4.78	10.46	-0.88	5.26
Law	-17.86	-2.13	10.18	0.71	5.57
Politics	-19.94	-3.09	8.6	-4.01	4.45
Sociology	-22.67	-3.54	10.34	-5.37	4.47

Source: Authors

*D stands for dimension

The mean dimension scores of the disciplines presented visually in the graphs so that the variations are better discernible.

Interpretation of Dimension 1 Scores: Involved vs. Informational Production

In D1, which differentiates between "involved" and "informational" linguistic styles, the data predominantly tilts towards being informational (Figure 1). This is evidenced by the nature of the scores across the disciplines. The discipline of food sciences, with a score of -26.9, stands at one extreme, while law, at -17.86, is at the other end of the spectrum. This substantial difference of 9.04 underscores the variation in how information is linguistically structured and presented within these disciplines. On the other hand, the minute difference of just 0.1 between computer science (-22.41) and history (-22.31) suggests a close similarity in their linguistic patterns, at least within this dimension. Such variations, from the pronounced 9.04 to the subtle 0.1, reflect the unique linguistic identities and practices inherent to each academic discipline.

Figure 1. Graphical Presentation of Dimension 1 Scores
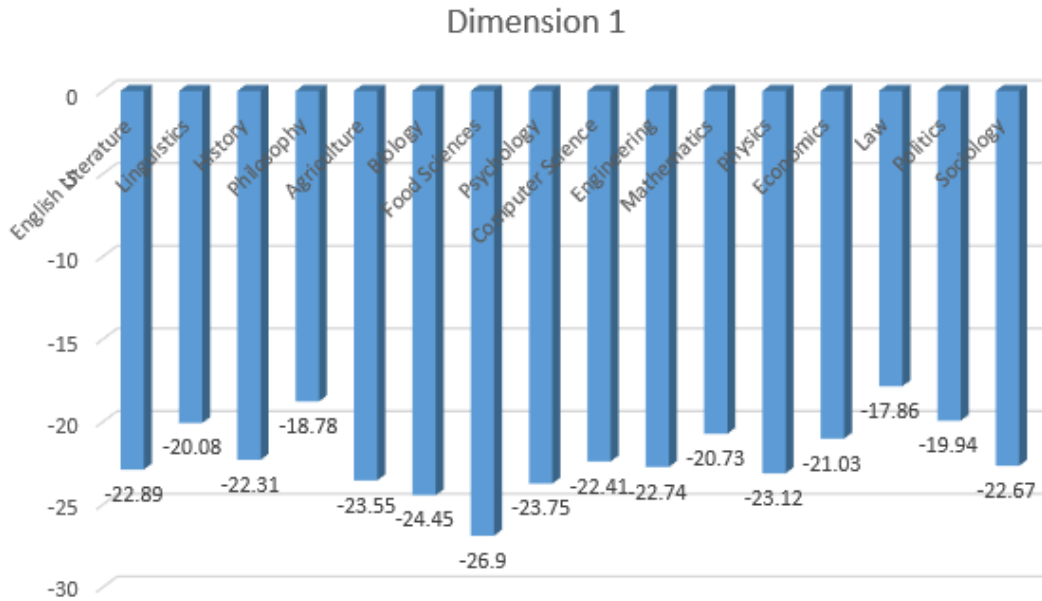
Source: Authors

Interpretation of Dimension 2: Non-Narrative Nature

In D2, which differentiates based on narrative qualities, the data predominantly leans towards a non-narrative style across disciplines (as shown in Figure 2). There appears to be a pattern of consistency in how subjects are approached linguistically within this dimension. While there is a noticeable level of uniformity in scores among several disciplines, there are a few outliers. For instance, the difference between mathematics, with a score of 5, and law, at 2.13, is the most pronounced at 2.87. This suggests that while both disciplines lean towards a non-narrative style, mathematics tends to have a stronger non-narrative inclination compared to law. In contrast, a group of disciplines, namely biology, agriculture, and food sciences, showcases a perfect alignment with an identical score of -4.7. This symmetry indicates that these fields share a very similar linguistic style in terms of narrative qualities within this dimension. Overall, the second dimension reveals a more consistent linguistic pattern across disciplines with only a few anomalies.

Figure 2. Graphical Presentation of Dimension 2 Scores
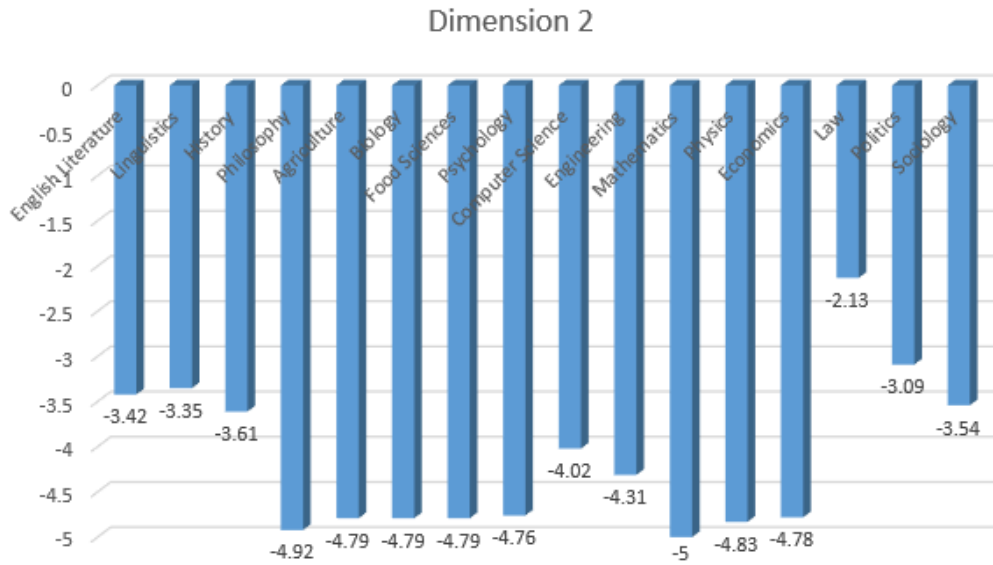
Source: Authors

Interpretation of Dimension 3

In the third dimensional context, the data displays (Figure 3) a characteristic of being independent of the surrounding context. The scores exhibit a higher degree of uniformity compared to the earlier dimensions. Notably, the most significant contrast of 2.4 units is observed between mathematics (11) and politics (8.6). Conversely, the narrowest disparity is witnessed between agriculture and biology, both standing at 9.4, indicating an absolute difference of 0. These disparities collectively indicate a limited extent of fluctuation among the different disciplines within this particular dimension.

Figure 3. Graphical Presentation of Dimension 3 Scores
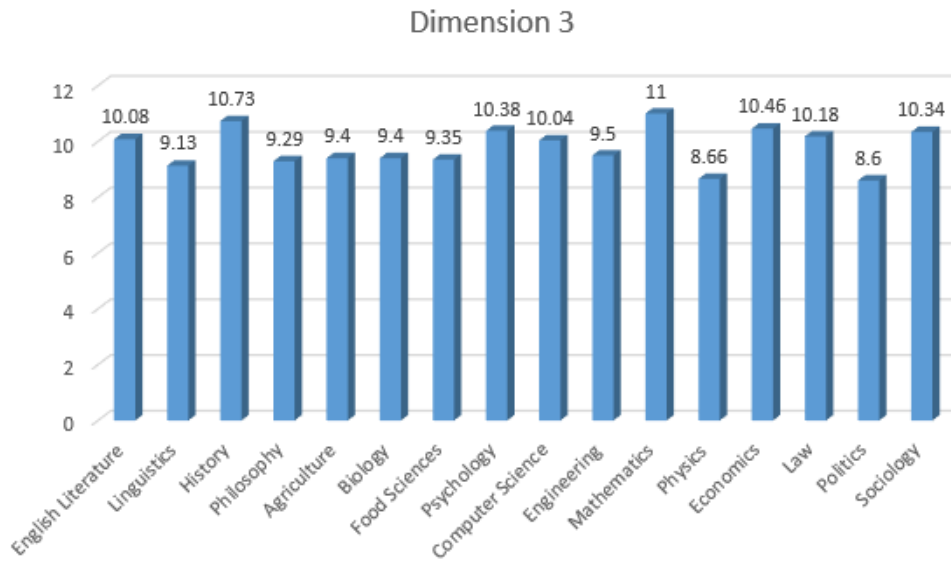
Source: Authors

Interpretation of Dimension 4

Within the context of dimension 4, an intriguing pattern emerges as all disciplines exhibit (see Figure 4) negative values except for a singular positive outlier, although the dataset still lacks a convincing impact. Notably, the

most substantial gap of 7.48 units is evident between law (0.71) and food sciences (-6.77). Conversely, the most marginal disparity of 0.16 is observed between philosophy (-1.04) and economics (-0.88). This distribution showcases a notably broad spectrum of variations across the various disciplines in this dimension.

Figure 4. Graphical Presentation of Dimension 4 Scores
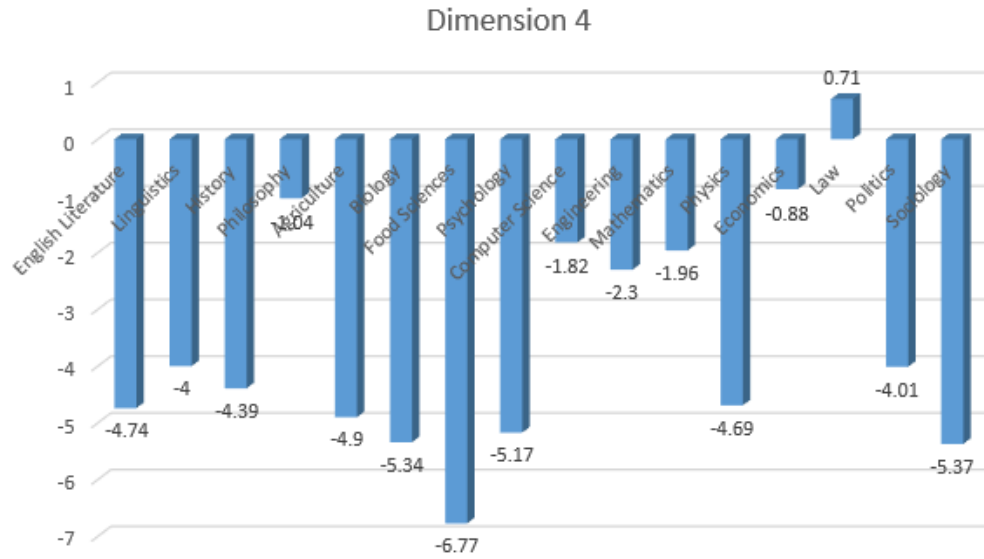
Source: Authors

Interpretation of Dimension 5

Dimension 5 reveals scores (Figure 1) that depict the data as encompassing a spectrum of highly to moderately abstract informational characteristics. Notably, the most substantial contrast of 4.98 units is observed between food sciences (7.37) and psychology (2.39). Conversely, the most minimal disparity is noted between history (4.81) and philosophy (4.62). This observation underscores a marked utilization of linguistic features across the dataset within this dimension, contributing to a significant range of variation in how information is expressed.

Figure 5. Graphical Presentation of Dimension 5 Scores
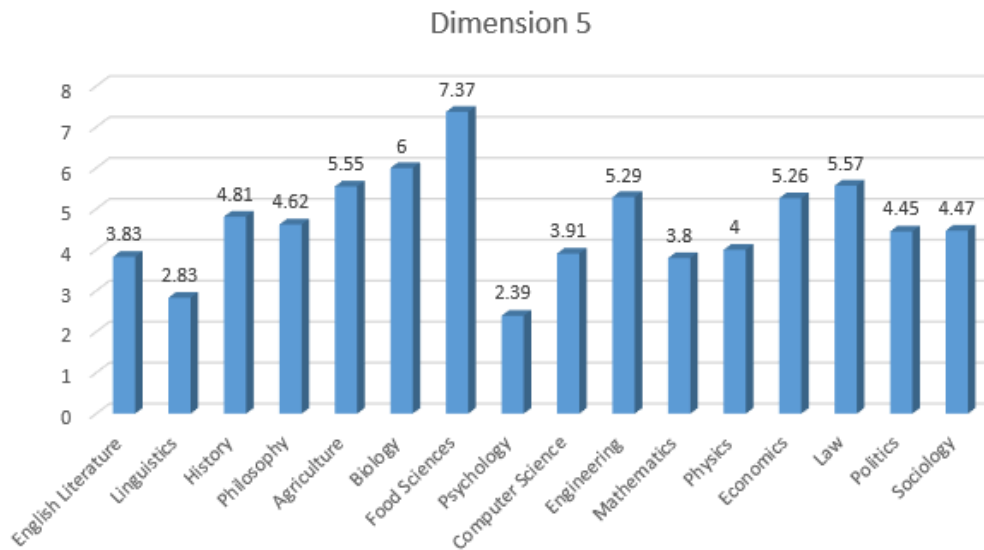
Source: Authors

Discussion

The differences observed across disciplines in five dimensions can be attributed to the inherent nature of their subjects and their discipline-specific writing styles (register). Every discipline caters to a distinct audience. For example, disciplines within the humanities focus on aspects of human life, while science disciplines delve into theories and discoveries, encompassing a broad spectrum of linguistic styles. Each of these fields also presents content using unique moves and sub-moves, resulting in distinct linguistic variations. A detailed examination of how linguistic features contribute to these cross-disciplinary variations across different dimensions will be explored further. Table 4 presents the linguistic features associated with across-disciplinary variations, accompanied by their z-scores from the overall data set.

Table 4. Linguistic Features Responsible for Variation along with Their Zscore

Linguistic Features	Zscore
AWL	2.53
NN	3.32
CONJ	2.51
JJ	2.24
RB	-2.67
[BEMA]	-1.87
NOMZ	2.48
PHC	6.1
[PASTP]	1.95
[SERE]	2.22
[WZPRES]	1.59
[PASS]	1.35
OSUB	0.89
[BYPA]	0.71
TSUB	1.58

Source: Authors

Dimension 1

In this dimension, the scores characterize the data as primarily informational and non-interactive. The word length (AWL), defined as the average length of words in the data, emerges as the principal source of variation among disciplines due to its overuse. Adverbs (RB) rank as the least utilized features in this dimension. Moreover, the use of "be" as a main verb [BEMA] and sentence relatives [SERE] are underrepresented in nearly half of the data. These features typically serve to render a text more interactional.

Subsequent Graph 1 categorizes abstracts from all six disciplines as "Official Documents", which aligns closely with the text type in D1.

Graph 1. Presentation of Dimension 1
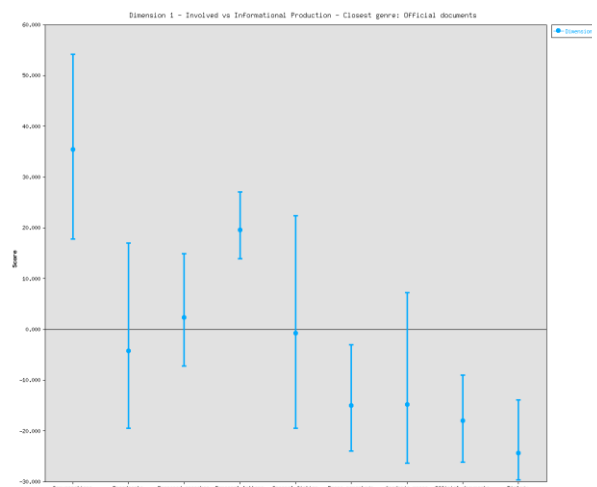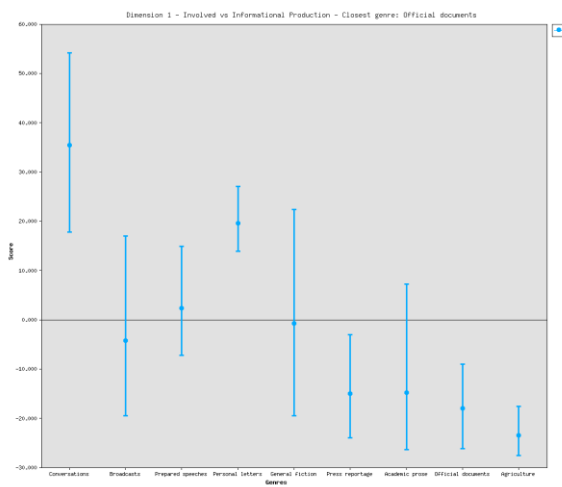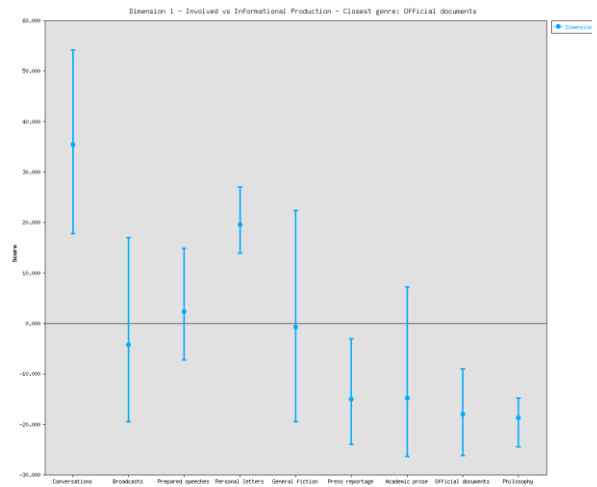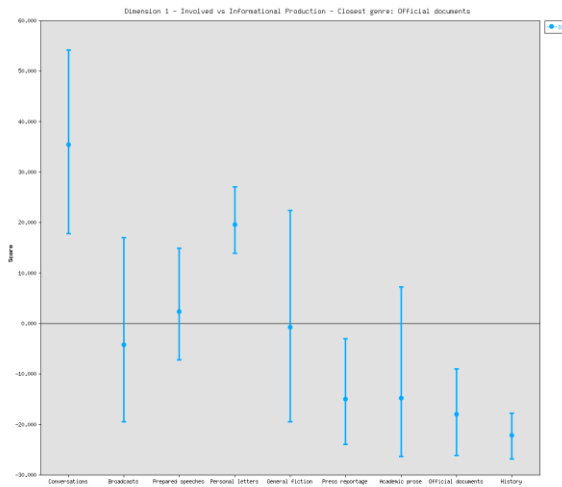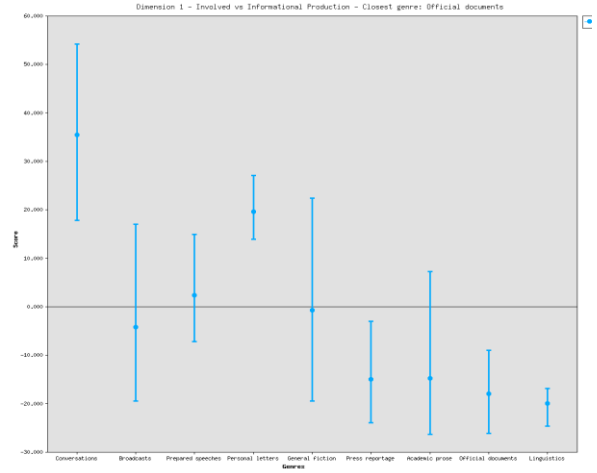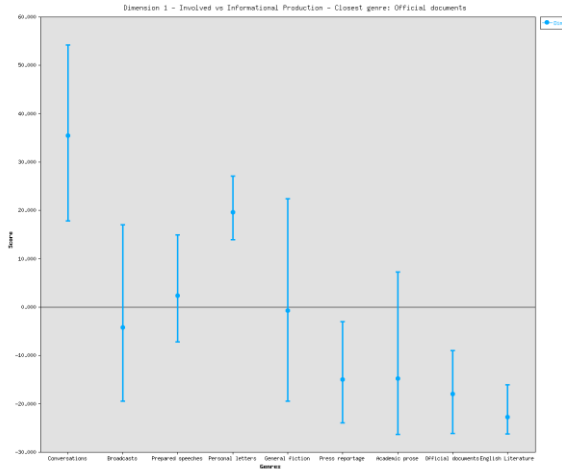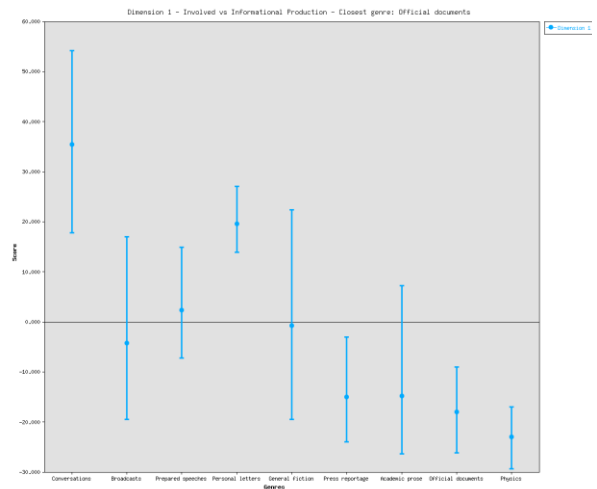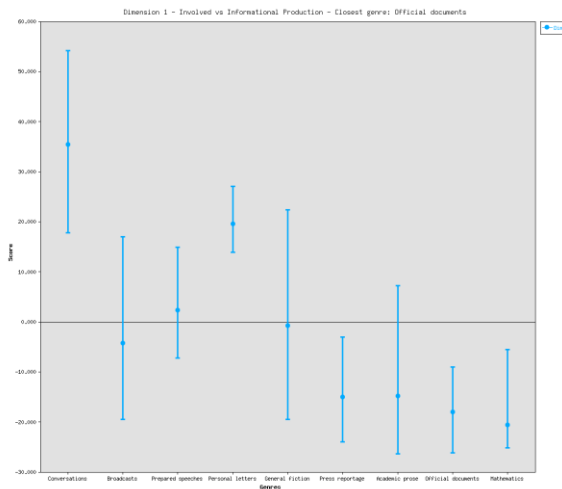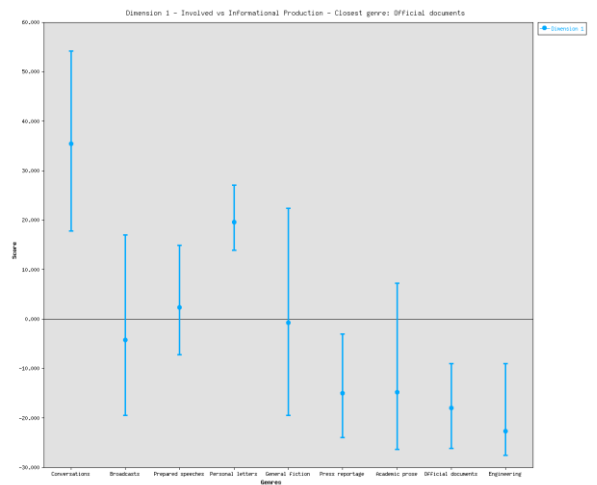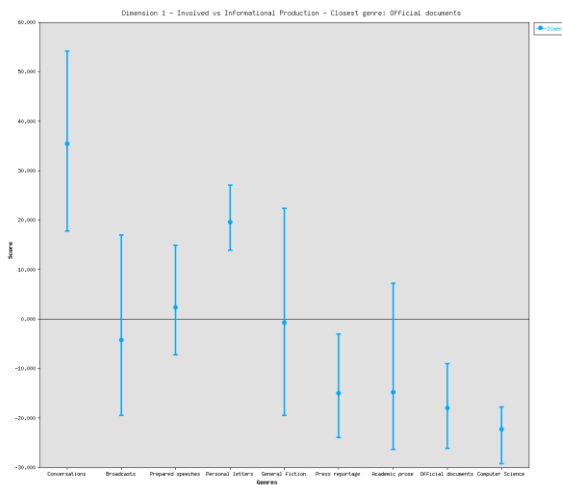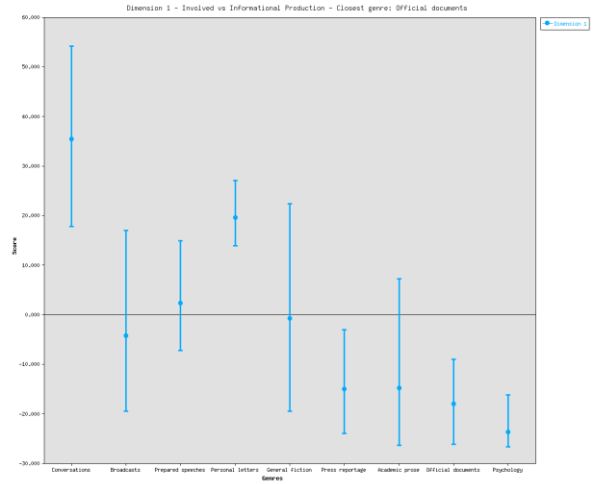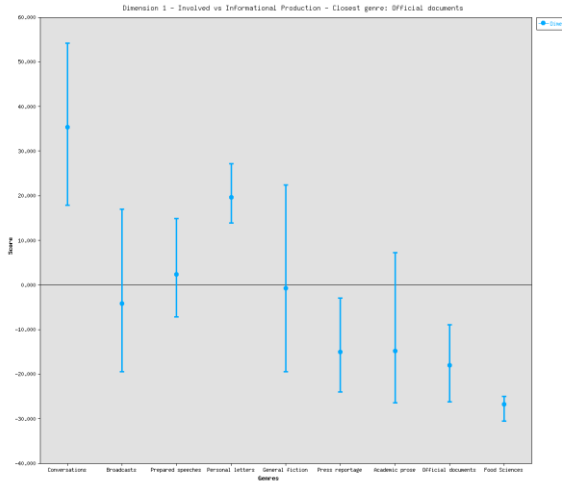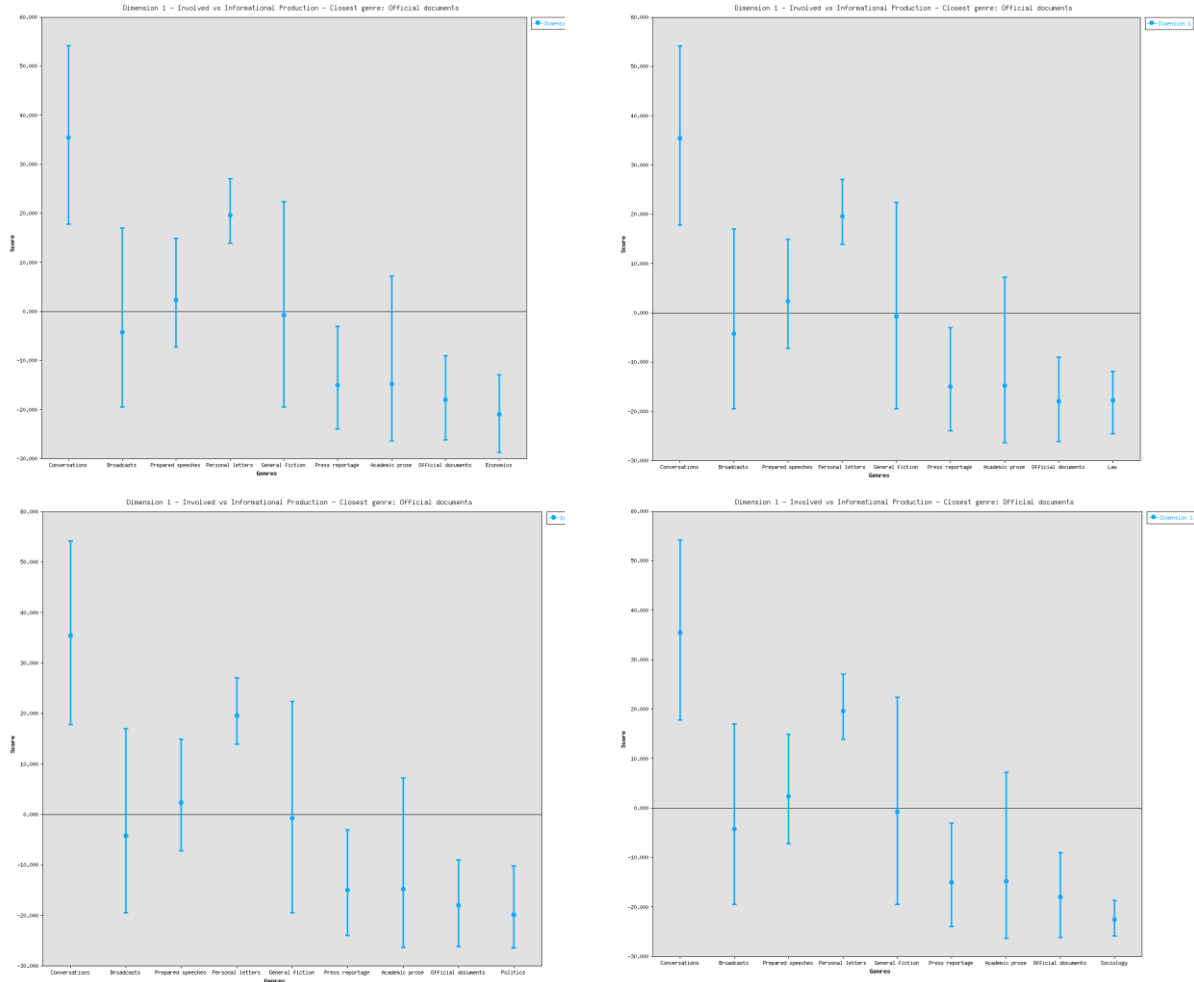
Source: Authors

Nouns serve as the primary conveyors of information in a text. The notation (NN) encompasses all types of nouns, excluding nominalizations and gerunds. On the other hand, attributive adjectives serve to characterize nouns by describing their attributes, thereby enriching the conveyed information. JJ denotes all types of adjectives present in a text, which function to augment the informational content. Nouns are prevalent in the disciplines represented in the current data, with law being the sole exception. In contrast, attributive adjectives appear in more than half of the data. In the ensuing examples from English literature and physics, underlined portions represent NN, while italicized sections indicate JJ. See Excerpts 1 and 2 for example.

Excerpt 1. The data for the study is triangulated, following Wodak et al (2009). It consists of *formal* and *informal* discourses from *various* genres *such* as *official* language bill, an *online* log, sections from interview-based autobiography, an extract from a *political* interview and an opinion section. The study in particular highlights the emergence and interplay between specific discourses that have been mobilized about the *Urdu-speaking* Mohajir immigrants in Pakistan at *different* times and became a part of *discursive* practices.

Excerpt 2. This improved behavior of the *plasmonic DSSC* can be attributed to *enhanced interfacial* charge transfer, decrease of charger combination, decrease of series resistance and *plasmonic enhanced* absorption of radiation by the dye. The impedance spectra also revealed *higher photovoltaic* performance of the *plasmonic cell*.

The [WZPRES] tag pertains to present participle forms of verbs (verbs ending in -ing). However, it exclusively addresses participles when they are followed by nouns. This specific attribute from D1 is primarily evident within the field of biology. An illustrative instance from this discipline is provided in Excerpt 3.

> **Excerpt 3.** Numerous roots were emerged from the base of shoots when plantlets were shifted to rooting medium comprised of MS medium <u>containing 1.5mg/l Indole butyric acid</u>. Activities of nitrogen <u>assimilating</u> enzymes, antioxidant enzymes, oxidases and proteases were studied in complete life cycle of Argyrolobium roseum from in vitro developed plant until the regenerated plant produced through indirect organogenesis.

Dimension 2

The scores (Table 4) from this dimension lack significance for data analysis, rendering the dimension itself trivial. Nevertheless, the results obtained depict the data as non-narrative in dimension 2. The [AWL] feature stands out as the sole underused aspect within this dimension. The visual representations provided below predominantly categorize the data as being closest to the text type of broadcasts. Notably, law is situated adjacent to academic prose, while politics corresponds with official documents (Graph 2).

Graph 2. Presentation of Dimension 2

Dimension 2 – Narrative vs Non-Narrative Concerns – Closest genre: Broadcasts
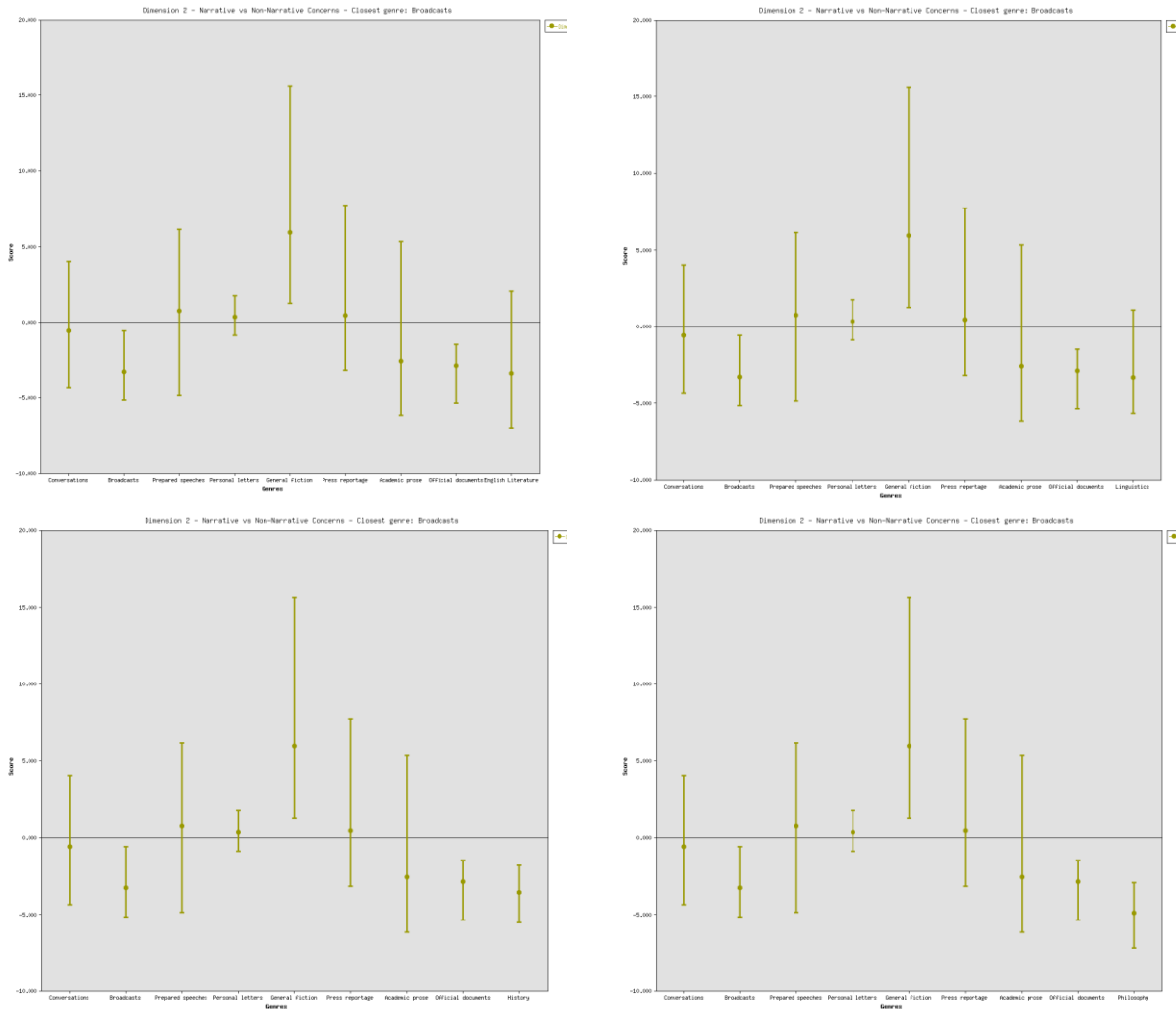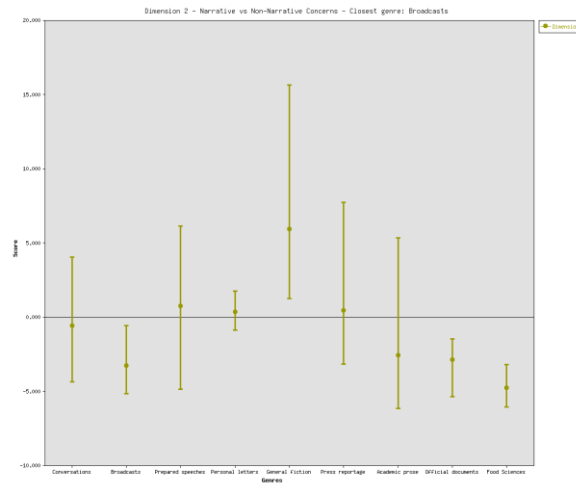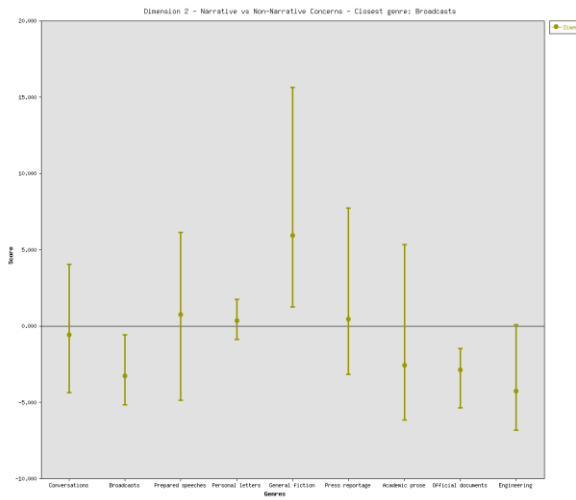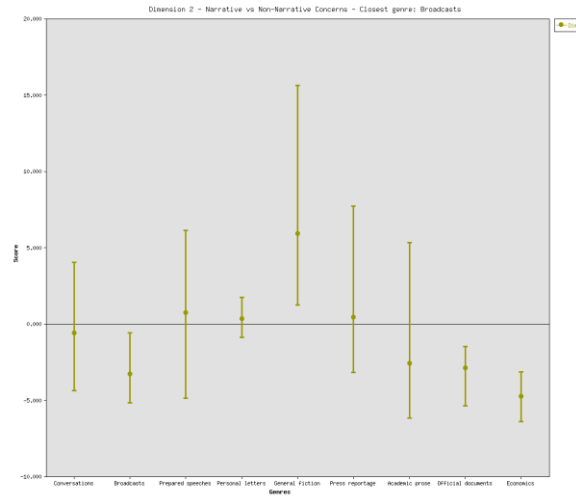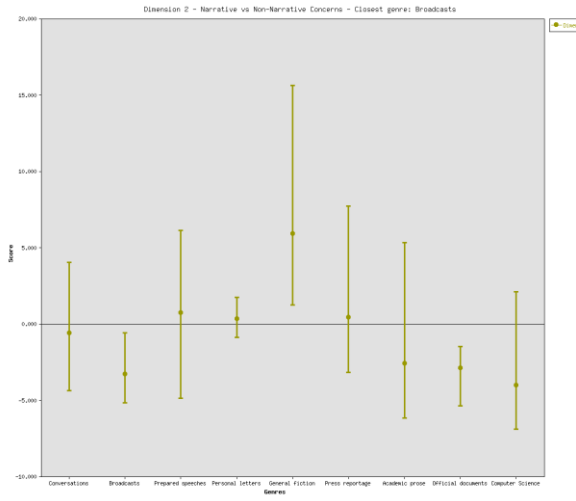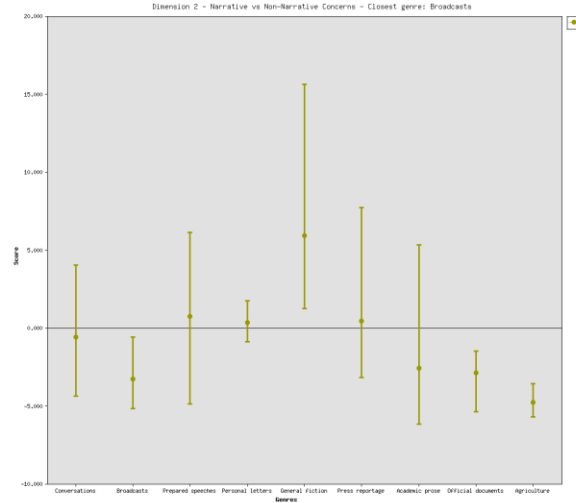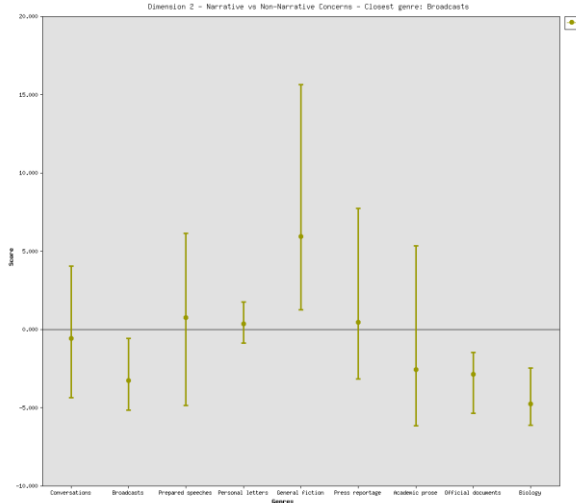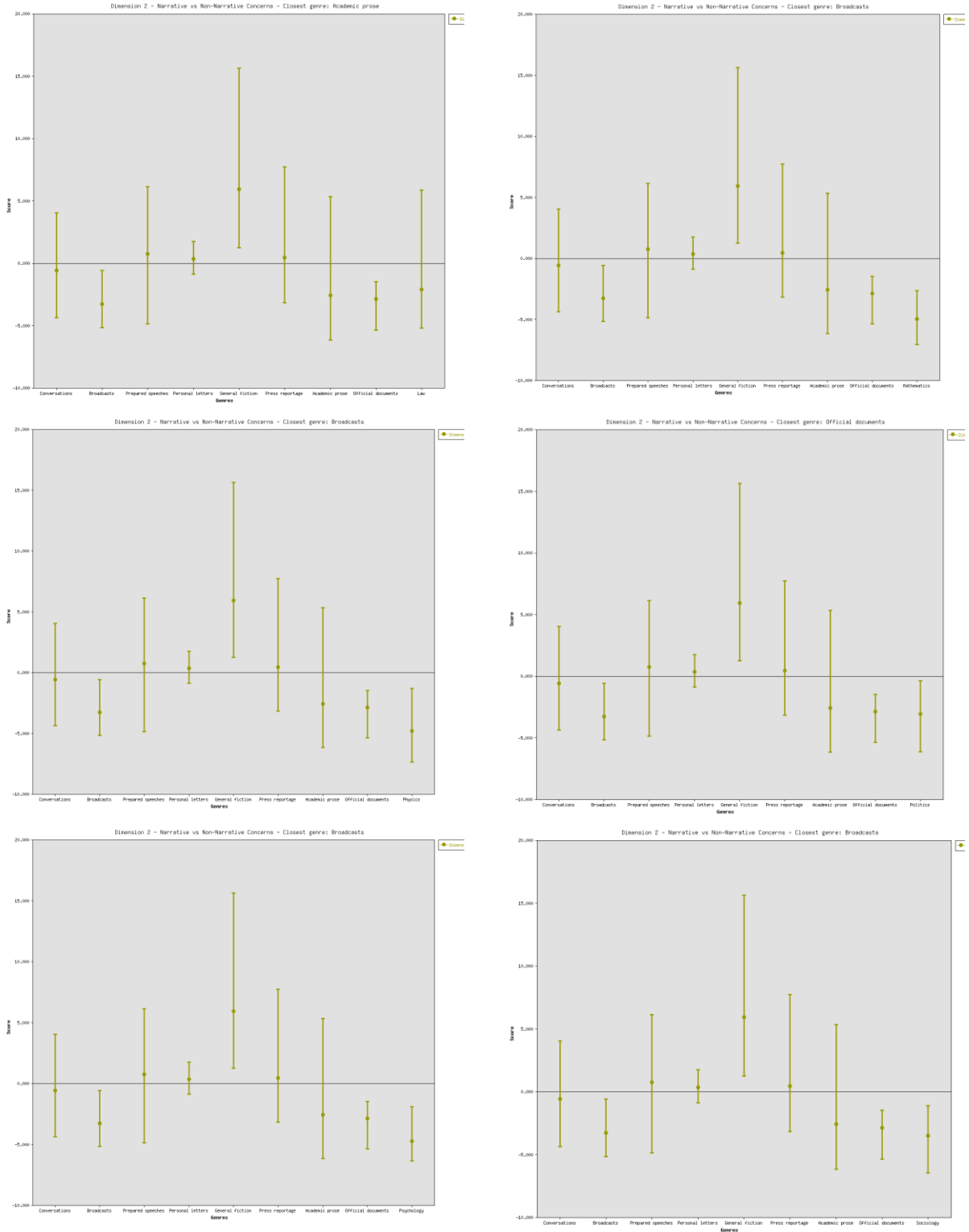
Source: Authors

Attributive adjectives stand out as the sole overused feature within this factor, indicating a strong descriptive tendency in the data. See Excerpt 4 for example.

Excerpt 4. The quantile results show that there is high inequality of income among the rural households. The top quintile households hold about 46 percent of the total income.

Dimension 3

      Dimension 3 revolves around the question of whether the text meaning relies on contextual support. The present dataset demonstrates (Table 4), a context-independent and explicit nature. This assertion is reinforced by the underutilization of adverbs (RB). The provided visual representations position the data in close alignment with official documents, which can be considered the most closely related genre.

Graph 3. Presentation of Dimension 3

Dimension 3 – Explicit vs Situation-Dependent Reference – Closest genre: Official documents
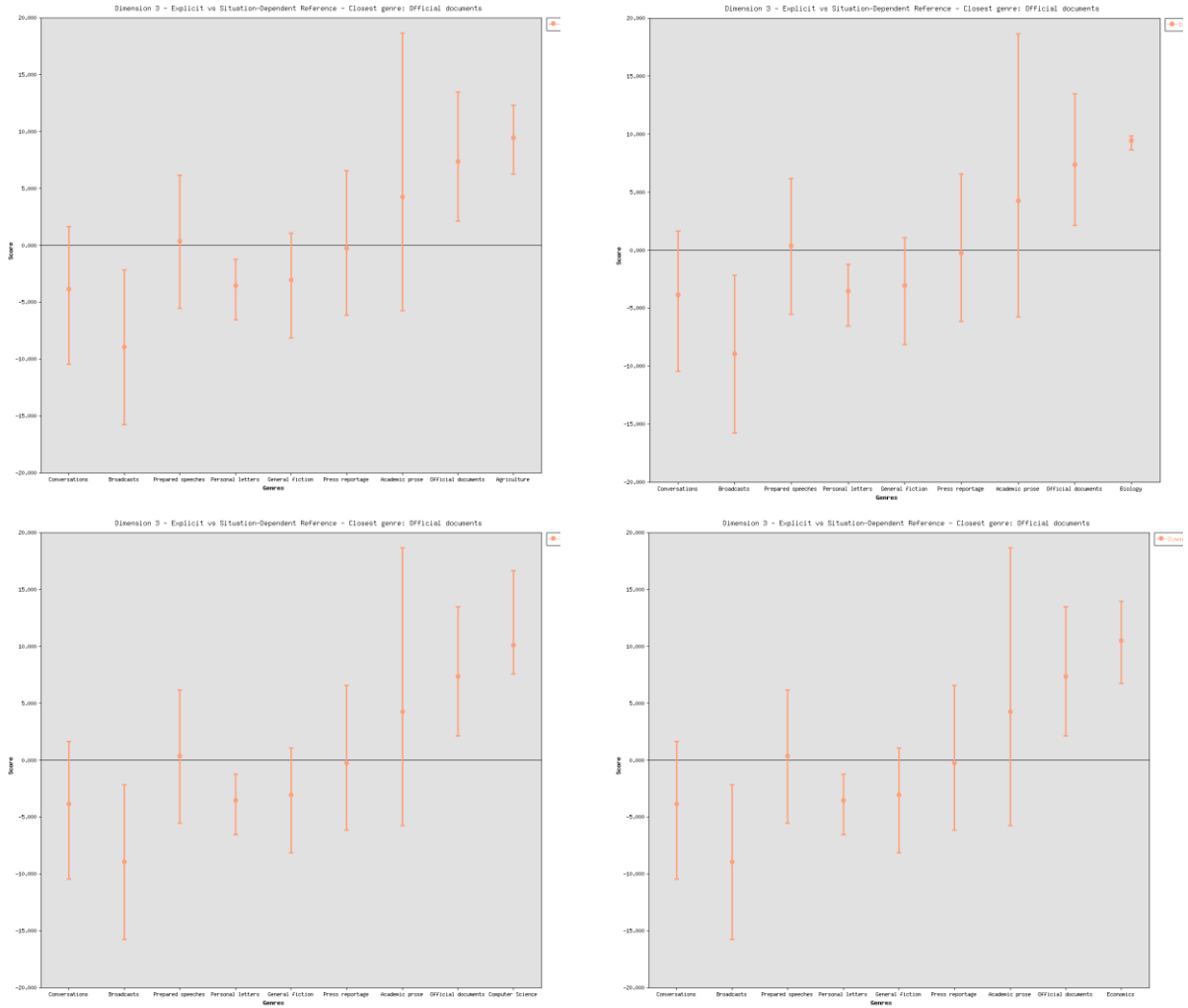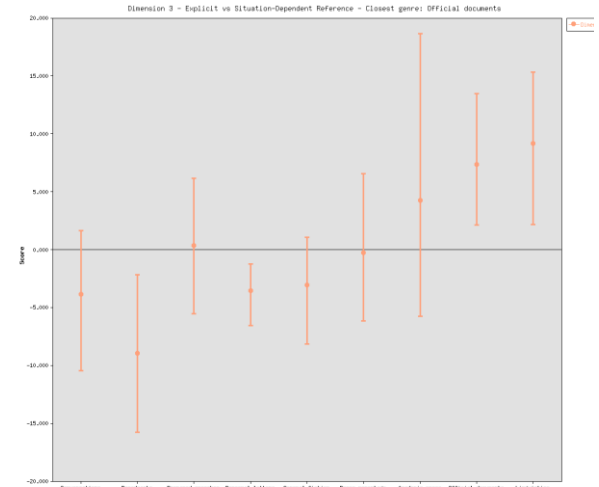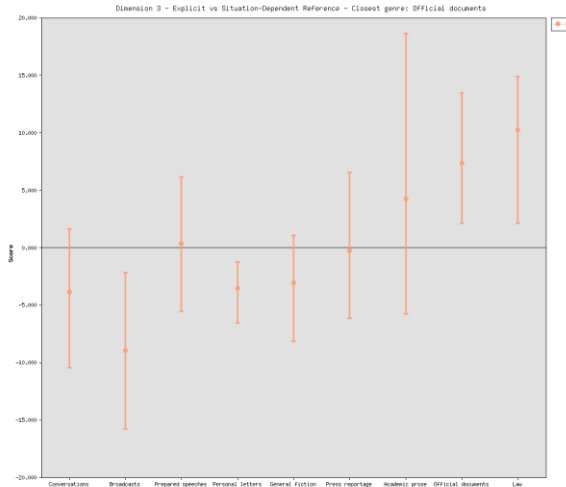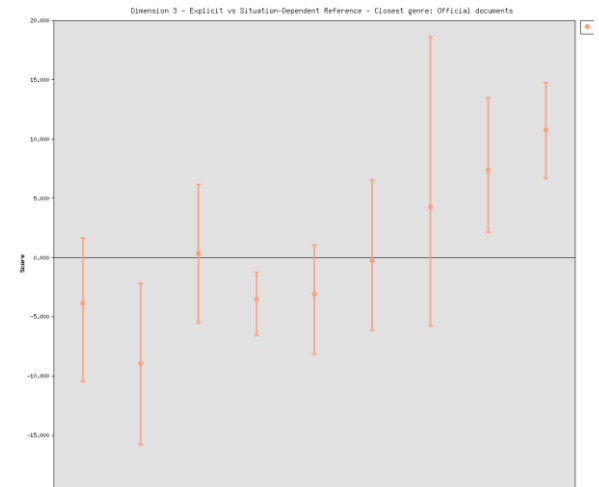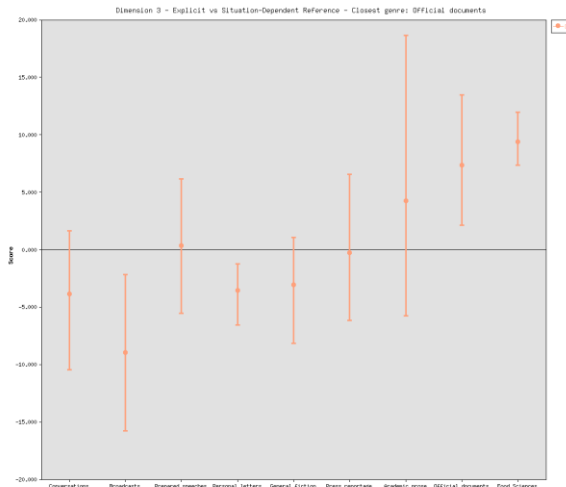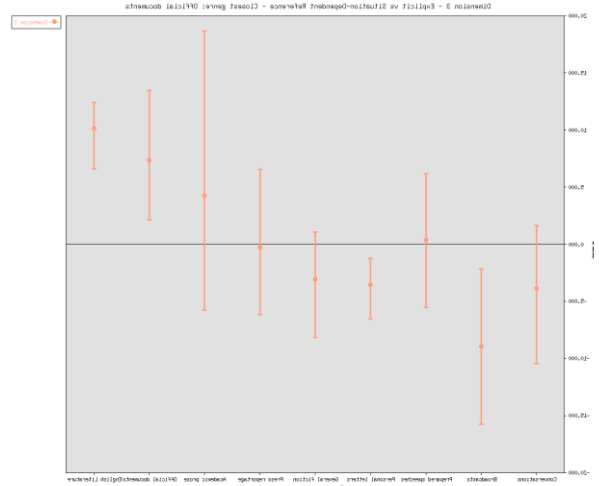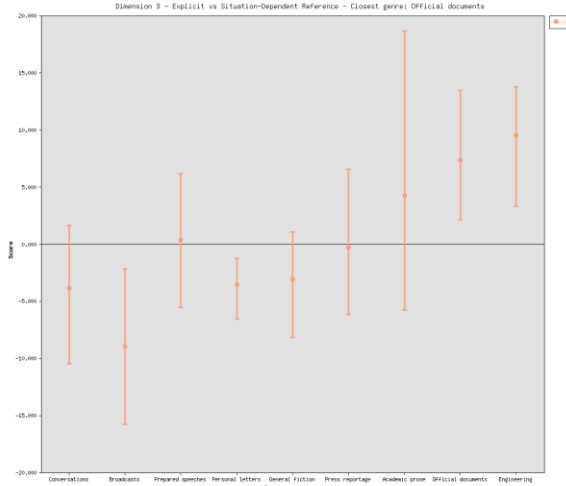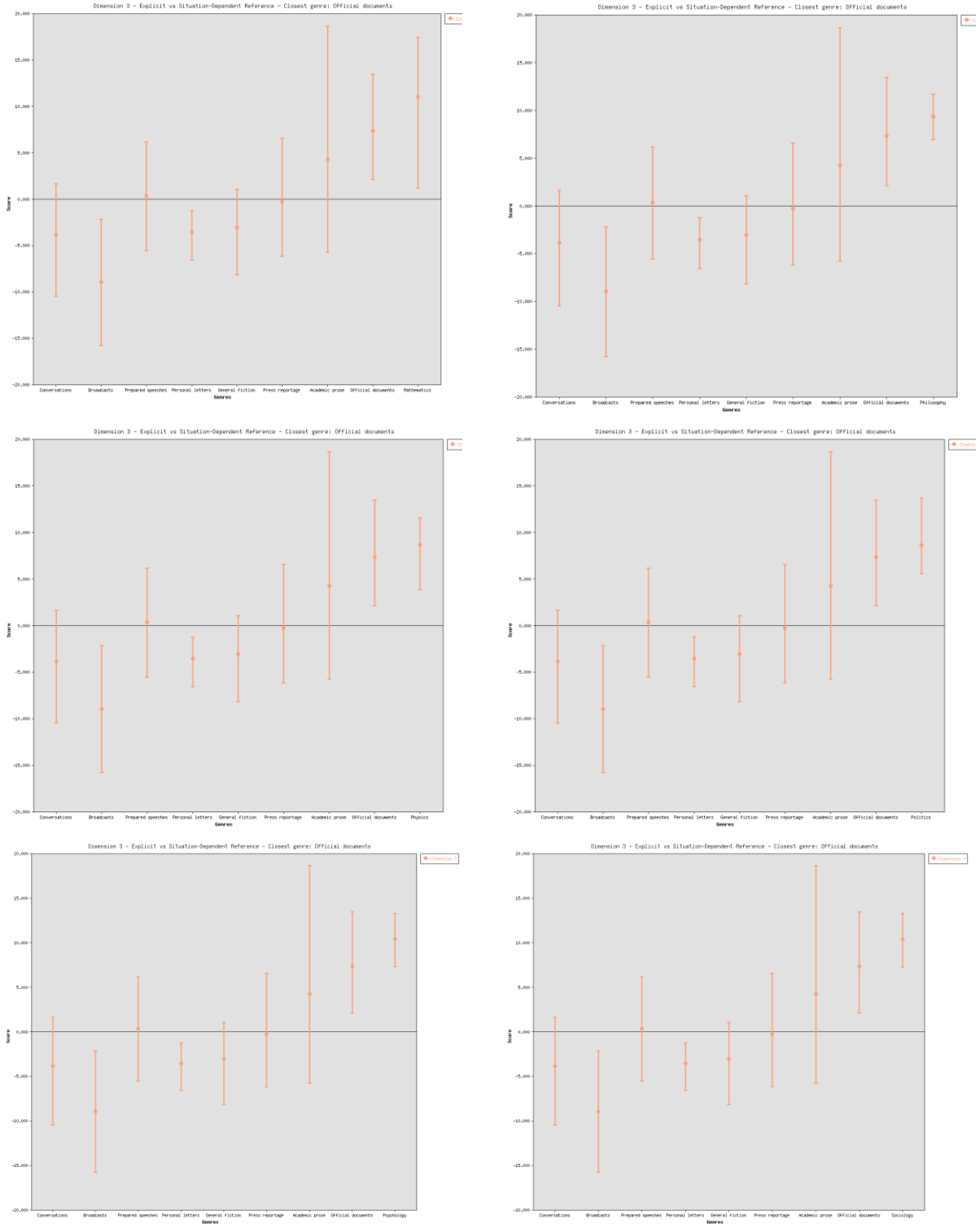
Source: Authors

Nominalizations [NOMZ] and phrasal co-ordinations (PHC) are both present all over the data (Table 4, Graph 3), from this dimension. If one is absent, the other is necessarily present. [NOMZ] includes all of the nouns formed with the help of adjectives by adding suffixes such as -ment, -ity. Whereas (PHC) tags are used for 'and' between similar tags, for phrasing. Both of these features tend to make the information more densely packed. Following

examples are from computer science and philosophy, respectively. The underlined words in Excerpts 5 and 6 are nominalizations whereas italicized are phrasal co-ordinations

> **Excerpt 5.** These factors based on their perceived role in influencing the adoption of Internet banking were further classified and grouped as Motivational, Inhibiting and Situational factors. Motivating factors included commitment, compatibility, availability of complimentary assets, and relative advantage. Inhibiting factors included security risk, technology complexity, access technology infrastructure, market readiness, and society readiness factors. Situational factors included organizational structure, adoption strategy, resources and policy readiness.
>
> **Excerpt 6.** Efficiency and exactness of results worked out are examined from the tables and graphs. The exact values are also simulated to compare and discuss the closeness and accuracy of approximations so obtained.

Dimension 4

This dimension tends to explore the persuasive features of the text and the results (Table 4, Graph 4) show that the current data is highly non-persuasive, due to being academic prose. This dimension is highly insignificant in the results as only 2 variables are overused, split infinitives [SPIN] and suasive verbs [SUAV]. And even out of those two, the scores of split infinitives create ambiguity in the results. Their spread across the data is highly misleading and suasive verbs are prominent in computer science only, as shown in Excerpts 7 and 8.

> **Excerpt 7.** The thesis proposes a technique where a percentage of communication is classified as outlier.
>
> **Excerpt 8.** The proposed algorithm, firstly, reorders and partitions the mesh using an efficient divide and conquer approach and then parallelizes the ALE moving mesh.
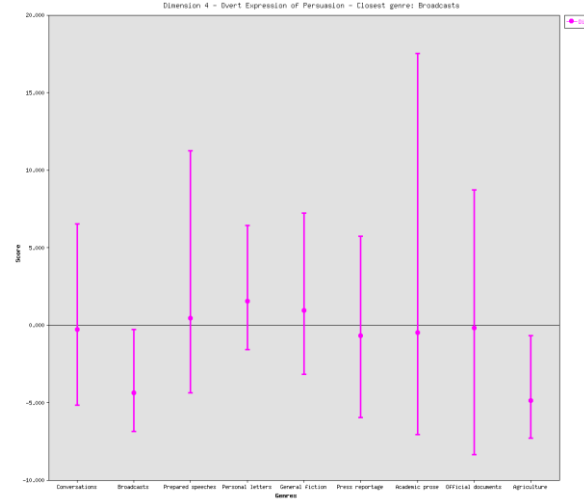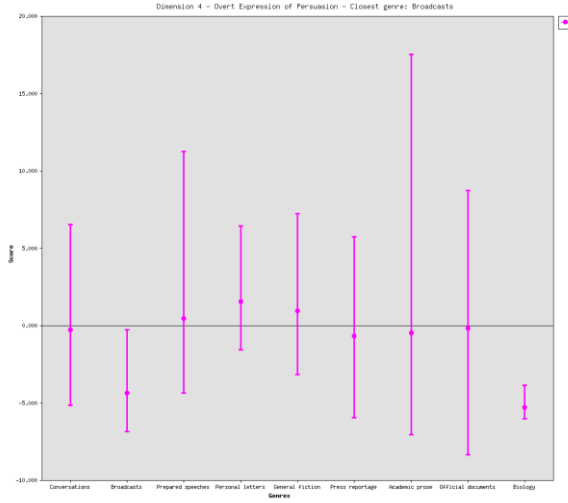
Graph 4 classifies philosophy, engineering, economics, computer science and mathematics into press reportage and law into general fiction, because of the its anomalous score on this dimension. Whereas, rest of the disciplines fall into the category of broadcasts.

Graph 4. Presentation of Dimension 4
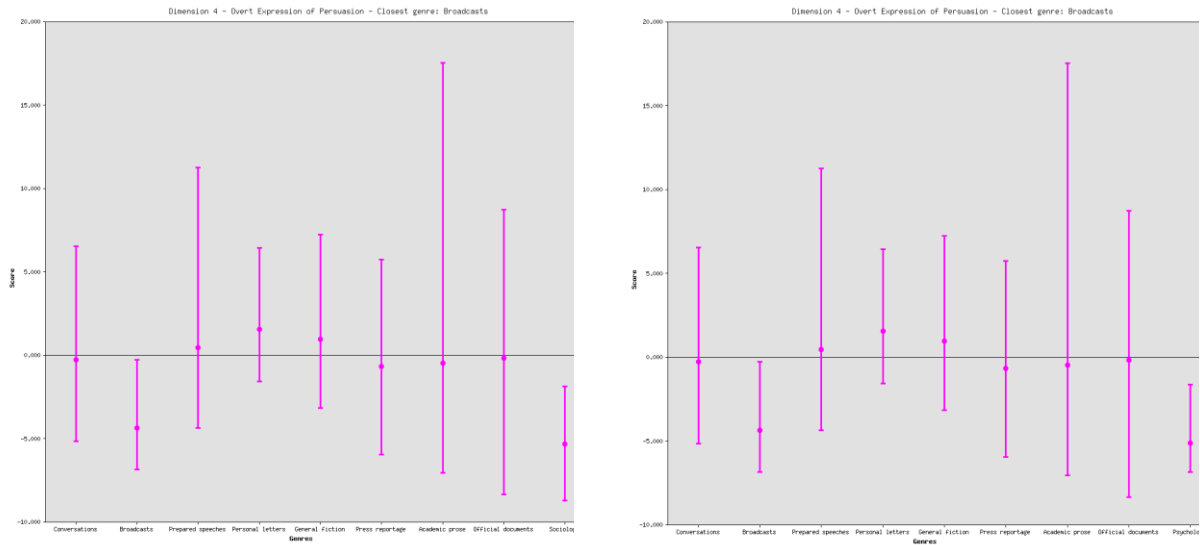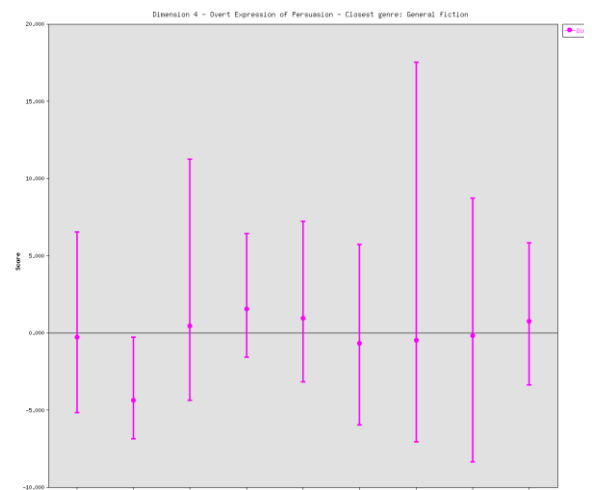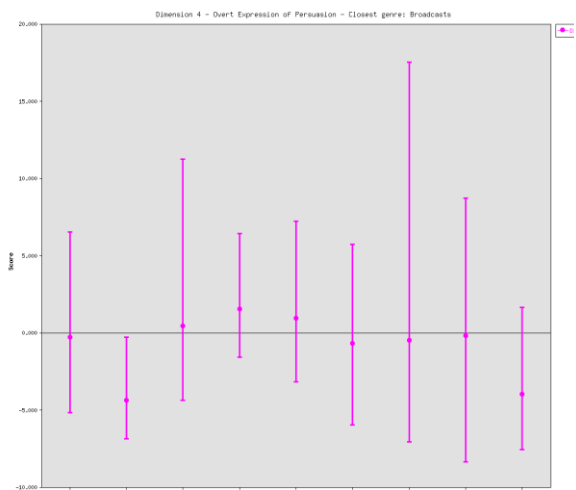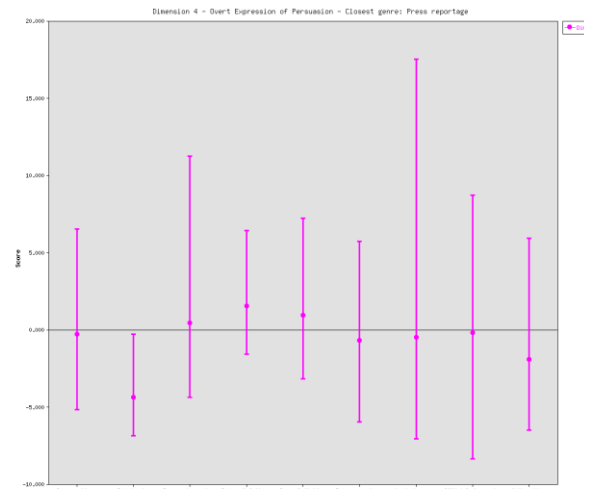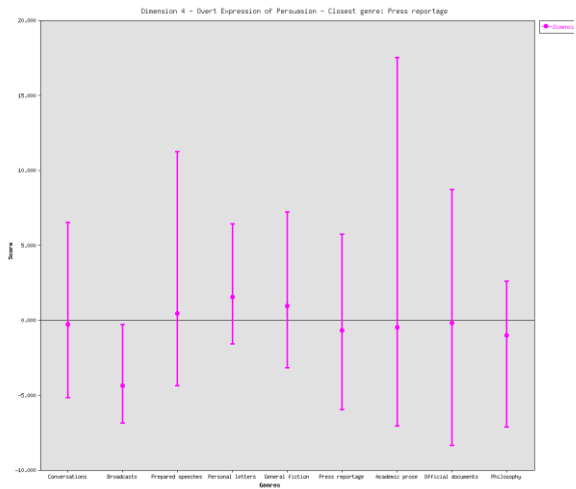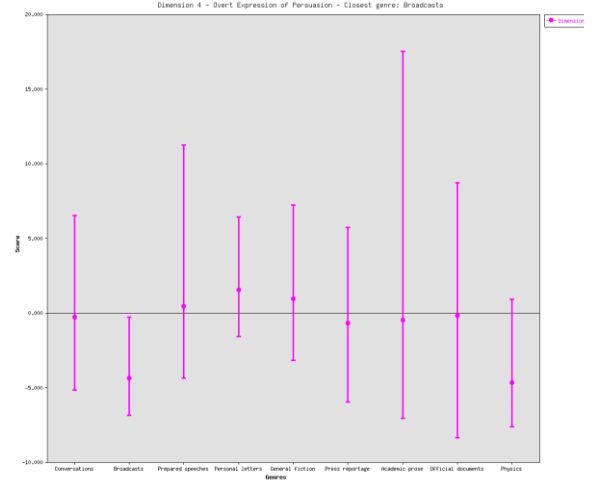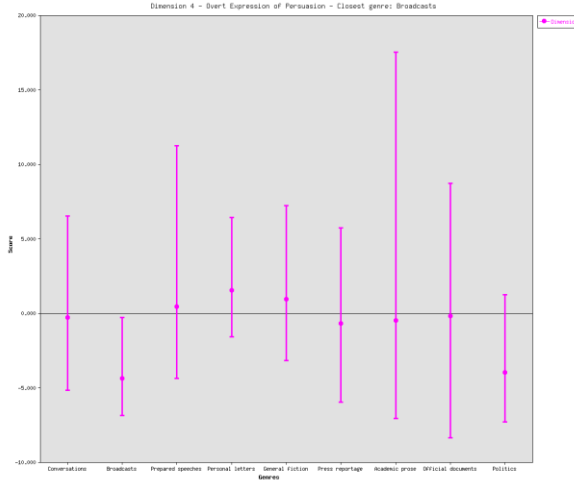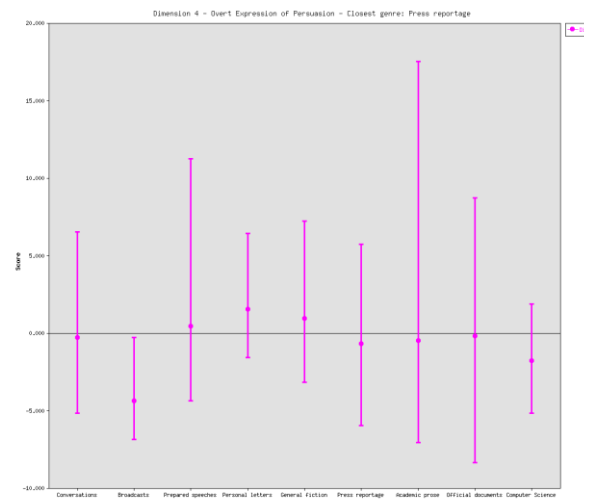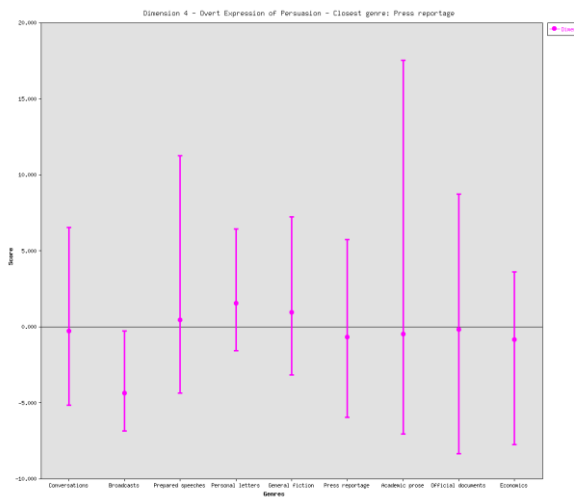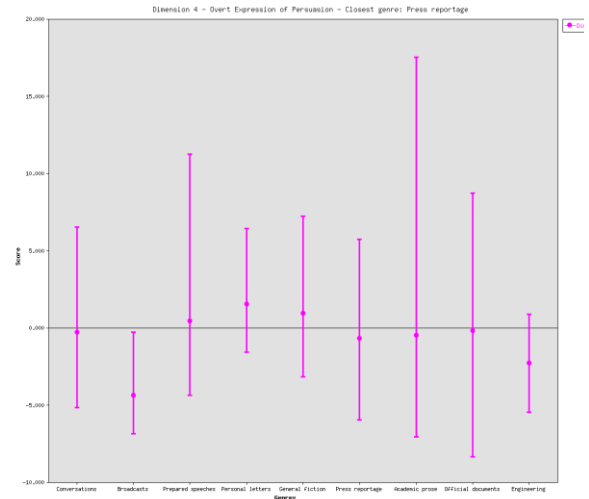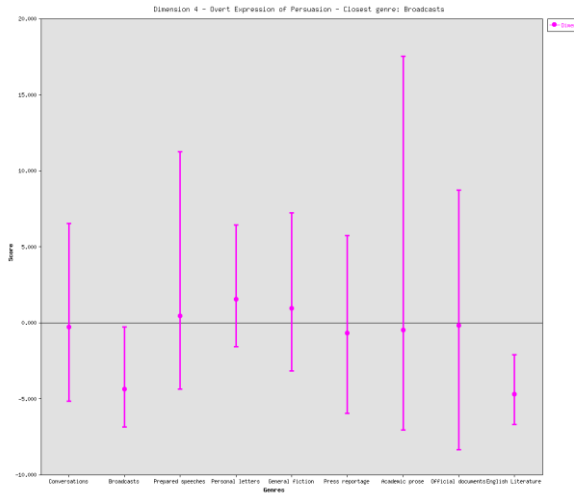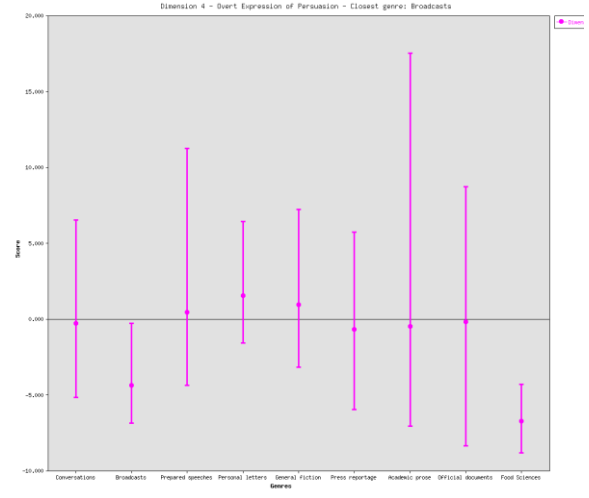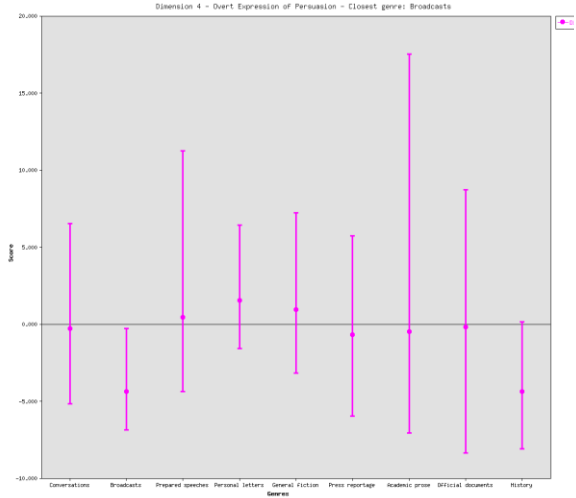
Source: Authors

Dimension 5

This dimension classifies (Graph 5) the present data as abstract, having conceptual features. There are no underused features in this dimension. The given graphs classify agriculture, sociology, economics, engineering, food sciences and law as academic prose, and psychology as press reportage, while the rest fall under official documents as closest text type.

Graph 5. Presentation of Dimension 5
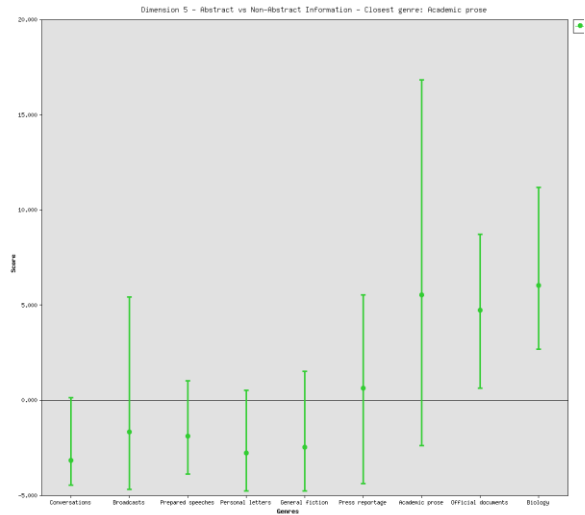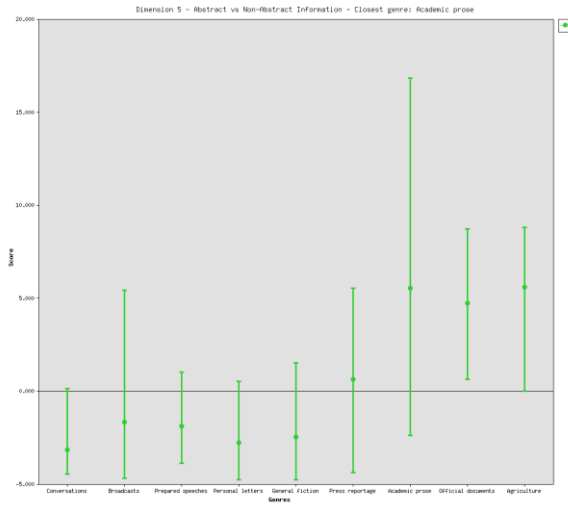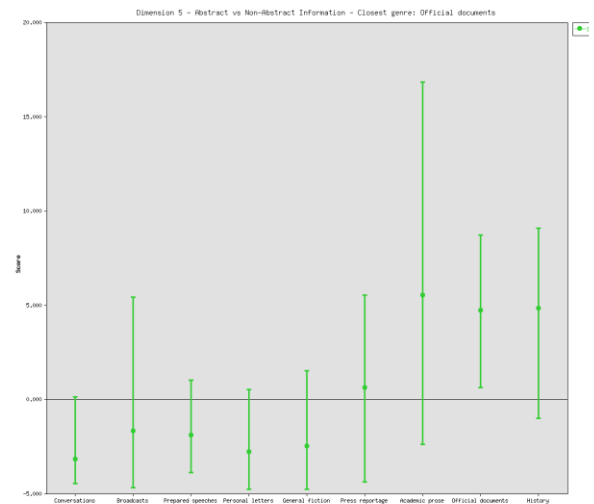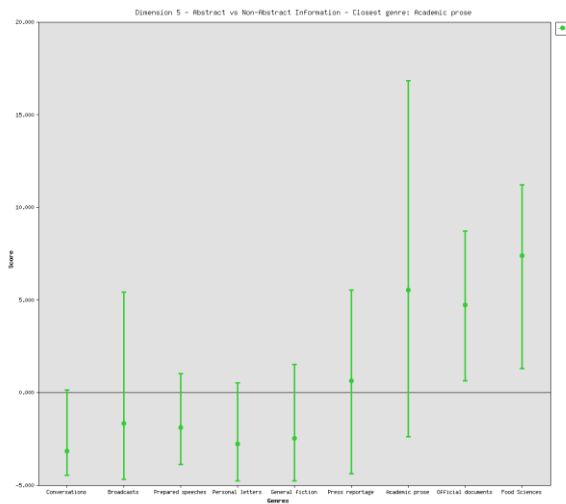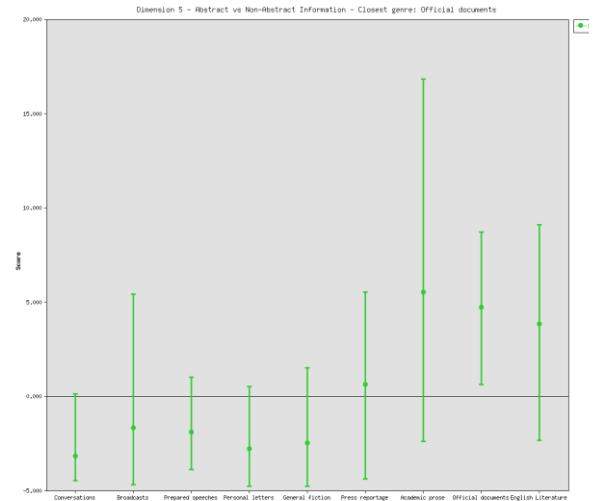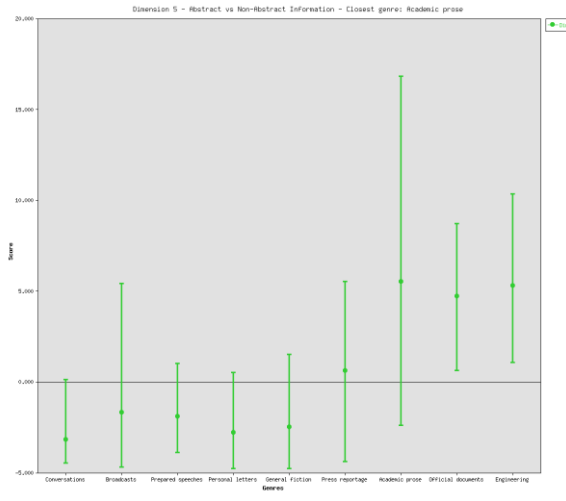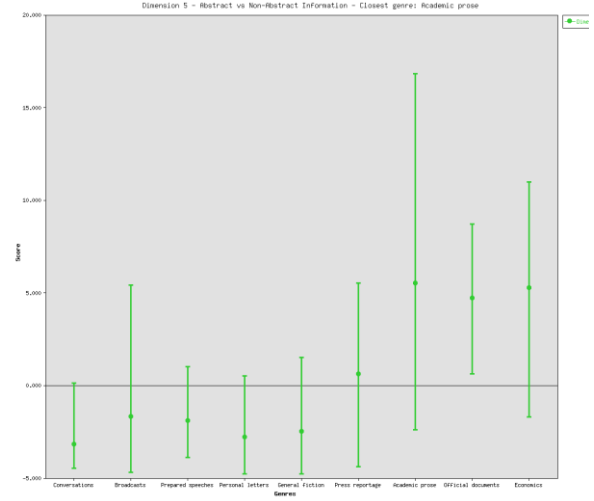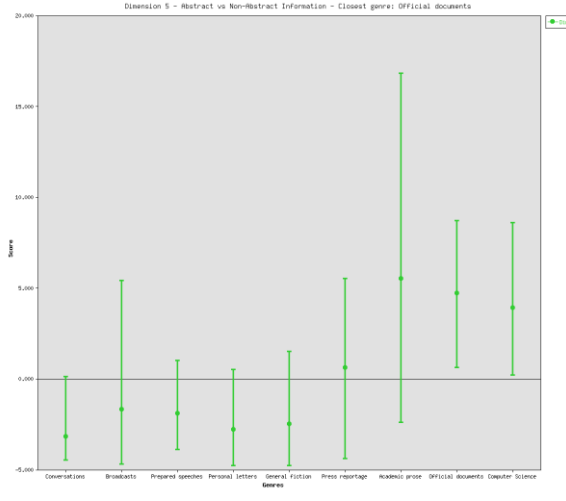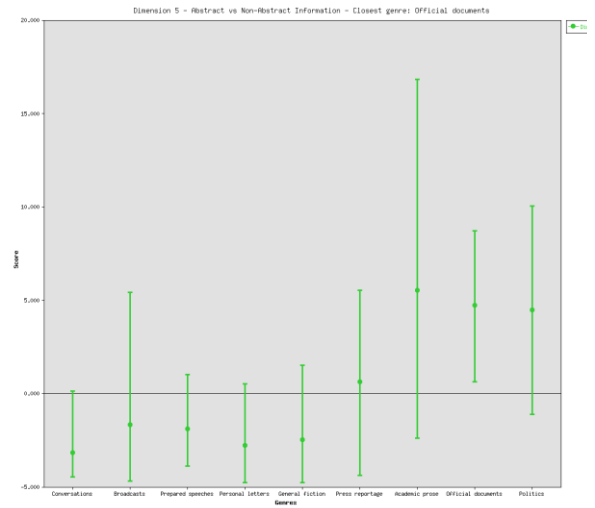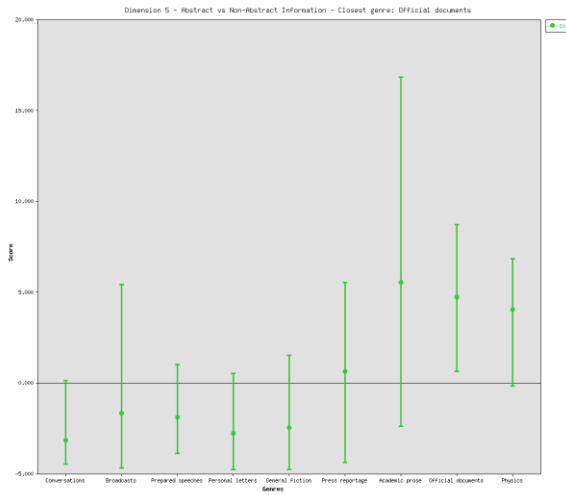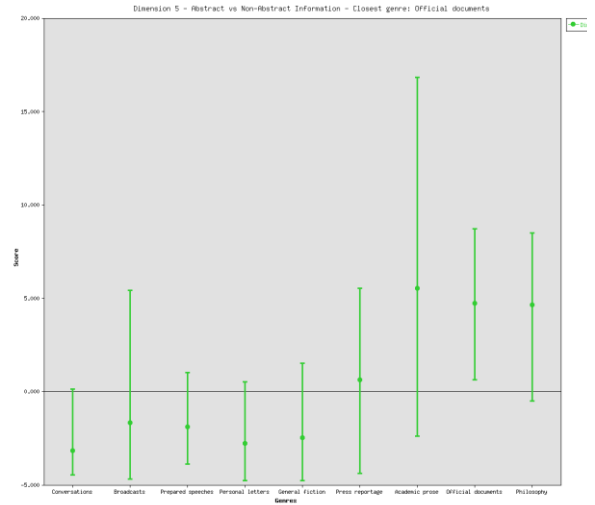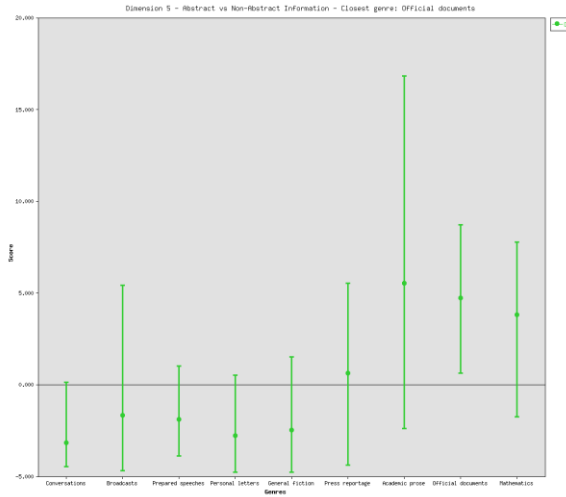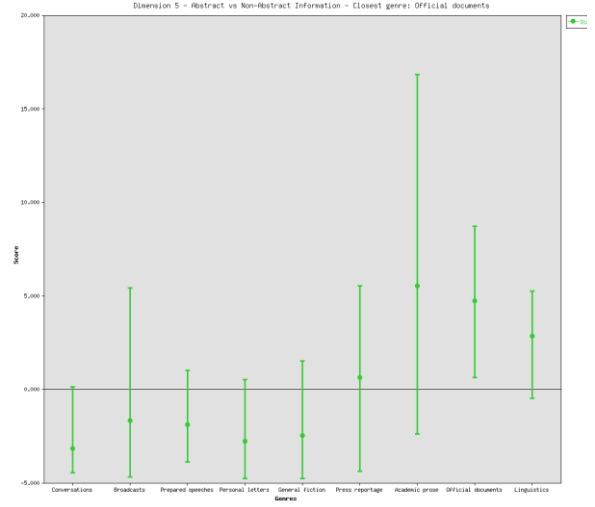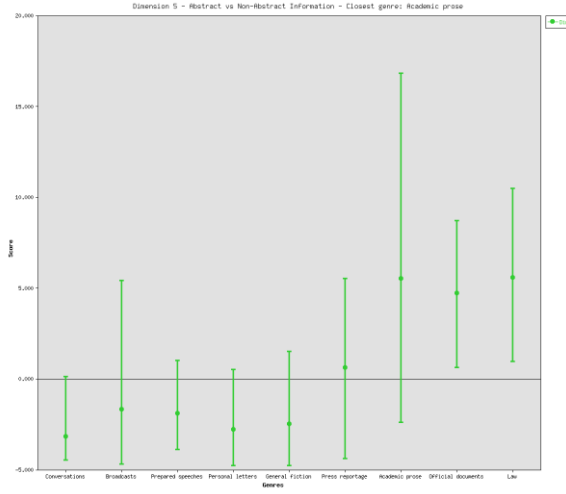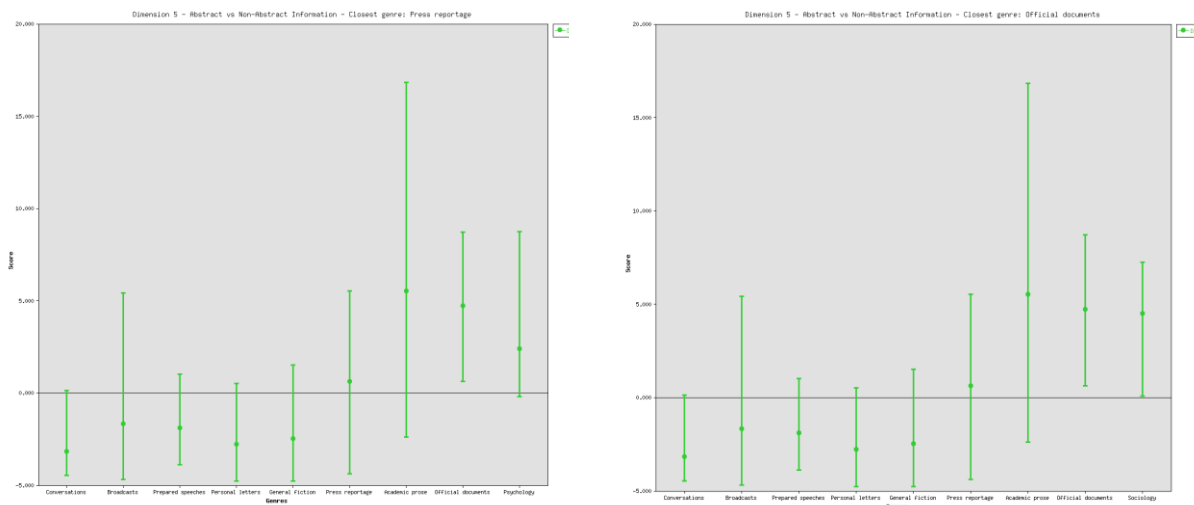
Source: Authors

Conjuncts [CONJ] are the most overused feature in this dimension, occurring in almost every discipline. Its function is to add information. Excerpt 9 shows conjuncts from mathematics.

Excerpt 9. We therefore propose that the performance of a substitution box is not just depending on the nature of the bijective Boolean function, however, it is affected by the degree 8 irreducible polynomial as well, which generates the maximal ideal of the principal ideal domain.

Past participles [PASTPs] are the second most occurring feature, in almost half of the data. They are tagged if followed by a preposition or adverb. See Excerpt 10 as an example from linguistics.

**Excerpt 10.** Then, the sequence/order of article acquisition is determined with regard to the SOC (Supplied in Obligatory Contexts), TLU (Target Like Use), and UOC (Used in Obligatory Contexts) measures.

By passives [BYPA] are marked when followed by prepositions. An example from physics is shown in Excerpt 11.

Excerpt 11. High performance TiO2 photo anodes undoped and doped with silver nano particles of size about 15 nm were fabricated by chemical route and were employed in dye-sensitized solar cells (DSSCs).

The presence of other adverbial subordinators (OSUB) in data is confirmed by an example from philosophy (Excerpt 12).

Excerpt 12. Central contention of this chapter is that while Islam does not minimize the importance of "Know thyself", it does lay heavy emphasis on "Choose thyself".

Agentless passives [PASS] include any form of 'be' followed by a verb or nominal form. It adds to the abstract features of the data and occurs only in engineering (Except 13).

Excerpt 13. Investigation into elliptical frequency diverse arrays (EFDA) reveals that, much better range selectivity and reduced side lobe levels can be achieved.

That relative clause on subject position [TSUB], is discussed by Biber and tagged in the data but is not classified in any of the dimensions. So, it has been discussed separately, because it is one of the major variation responsible features. It is tagged when 'that' is followed by a verb. Excerpts 14 and 15 present examples from sociology.

Excerpt 14. Furthermore, rural development activities with socio-political as well as technical dimensions demand a better coordination among the social, political and technical aspects that can be well realized by the establishing local level developmental organizations.

Excerpt 15. The bourgeois public sphere existed in a 'nation-state', having an industrialist economic orientation and social system that reflected values of that age.

Comparison of Present Data with That of Biber

The results detailed above position the current dataset as highly informative, markedly descriptive, straightforward, objective, abstract, and explicit across the five dimensions of Biber's model. This suggests that the data predominantly falls under the category of learned exposition and, to a certain extent, scientific exposition. These characteristics align with those of Biber's academic prose, which is primarily defined by scientific and learned exposition. A comparison of the dimension scores between Biber's academic prose and the current data is provided in Table 5.

Table 5. Comparison of Dimension scores across Pakistani Academic Writing and Biber's Academic Prose

Dimensions	Biber's Academic Prose	Pakistani Academic Writing	Difference
D1	-14.09	-22.14	8.05
D2	-2.6	-4.12	1.52
D3	4.2	9.78	-5.58
D4	-0.5	-3.6	3.1
D5	5.5	4.59	0.91

Source: Authors

Dimension 1

In this dimension, conversations score (Table 5, Graph 6) the highest, suggesting a predominant use of involved text features. Meanwhile, broadcasts, personal letters, prepared speeches, and general fiction follow with high to moderate scores. On the contrary, text types like official documents and academic prose, which garner the lowest scores, are characterized by their prevalent use of informative text features. The study indicates that the data, with its low score, aligns closely with official documents. The marked difference of 8.05 between Biber's dimension score and this data underscores the unique characteristics of Pakistani English, highlighting its information-dense nature.

Graph 6. Comparison of Scores across Pakistani Academic Writing and Biber's Academic Prose
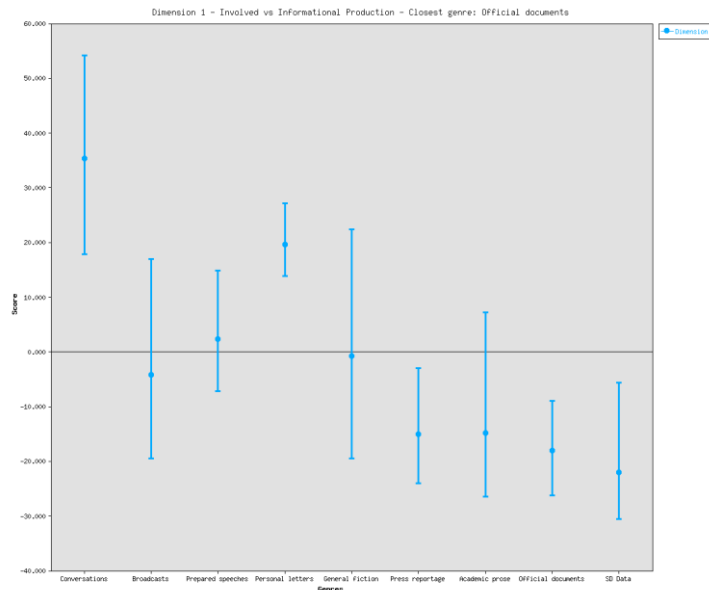
Source: Authors

Dimension 2

In this dimension, general fiction takes the lead with the highest score, while broadcasts and official documents trail with the lowest scores. Other genres find their positions within this range. A high score is indicative of a prevalence of narrative features, while a low score suggests the opposite, leaning towards non-narrative elements. The data at hand, with its moderately low scores, closely aligns with the broadcast genre, as the study suggests. The score difference of 1.52 in this dimension is notable, sitting between significant and negligible (see Table 5 and Graph 7).

Graph 7. Comparison of Scores across Pakistani Academic Writing and Biber's Academic Prose

Source: Authors

Dimension 3

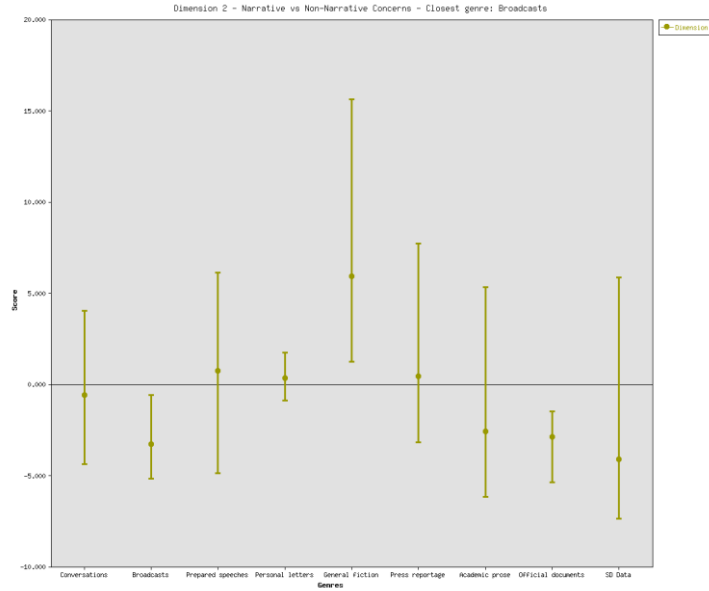In dimension 3, academic prose secures the highest score, while broadcasts settle for the lowest. Other genres lie within the range between these two. A high score on this dimension signifies text that is explicit and independent of context, and the reverse is true for a low score. Yet, the current data, boasting a notably high score on this factor, aligns more closely with official documents. A significant discrepancy of 5.58 in the dimension scores underscores the uniqueness of Pakistani English as a distinct variety (see Table 5 and Graph 8).

Graph 8. Comparison of Scores across Pakistani Academic Writing and Biber's Academic Prose

Dimension 4
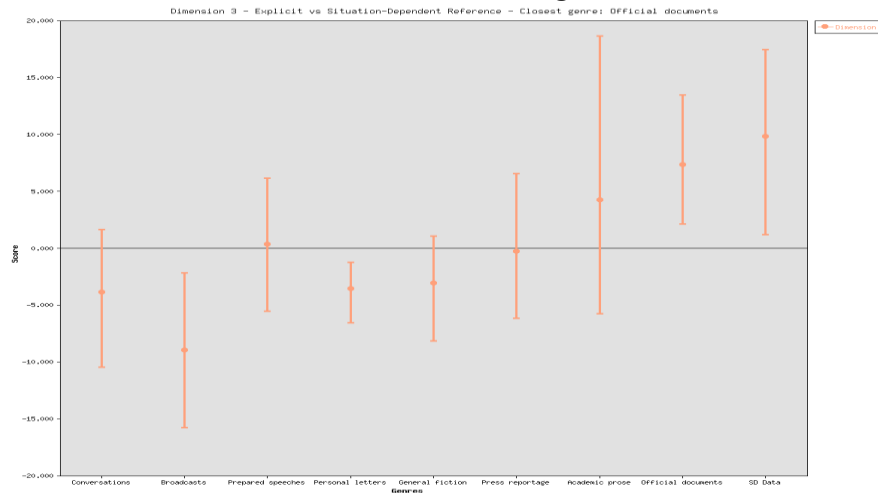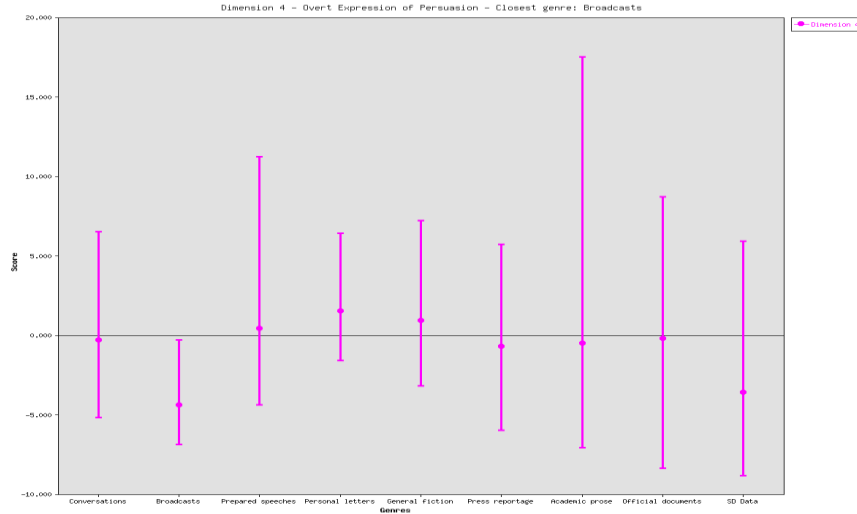
On this dimension, academic prose registers the highest score (Table 5, Graph 9), while broadcasts notch the lowest, with other genres falling in between. A higher score denotes greater subjectivity in the text and its features, and the opposite holds true for a lower score. The current data, possessing a moderate score, aligns most closely with broadcasts. A marked score difference of 3.1 further bolsters the assertion that Pakistani English diverges significantly from British English.

Graph 9. Comparison of Scores across Pakistani Academic Writing and Biber's Academic Prose

Source: Authors

Dimension 5

In this dimension, academic prose scores the highest, whereas conversations register the lowest (Table 5, Graph 10). A high score indicates a greater degree of abstraction. The current dataset, boasting a moderately high score, closely aligns with official documents. A minor discrepancy of 0.91 in the scores further substantiates the notion that, despite its distinctiveness, Pakistani English has roots in British English.

Graph 10. Comparison of Scores across Pakistani Academic Writing and Biber's Academic Prose

Source: Authors

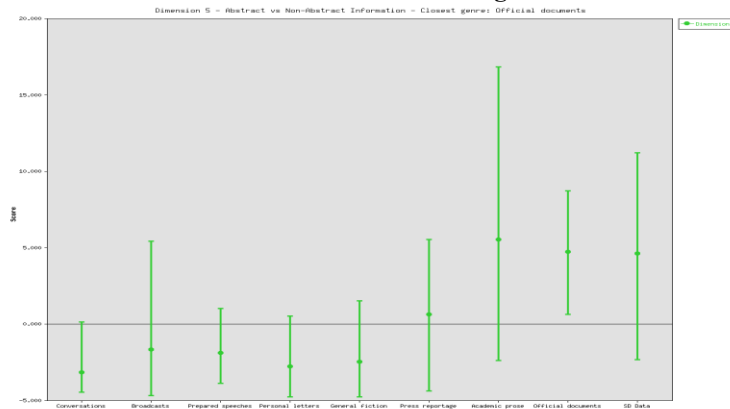**Conclusion**

Based on the findings of the present study, distinct variations are evident across disciplines from four fields. While the majority of the disciplines exhibit a similar pattern abstract from law notably diverged, specifically on D4.

78

Generally, disciplines tend to be informational on D1, non-narrative on D2, context-independent on D3, non-persuasive on D4, and abstract on D5. However, across-disciplinary variations began to emerge subtly on D3 and became more pronounced by D5. These variations highlight the diverse lexico-grammatical features employed in different academic fields. Comparing the results with Biber's data reveals significant disparities between the mean dimension scores of both datasets. The vast array of lexical and grammatical features in academic writing, along with its deviation from Biber's academic prose, underscores Pakistani English as a distinct variety. This variant exhibits unique features and a broad spectrum, aligning with the observations made by Mehmood (2009).

This research strongly advocates for an in-depth examination of other written and spoken registers of Pakistani English using MDA. Such analyses can help identify their defining characteristics as separate registers. The study also holds significant implications for the instruction of English for Academic Purposes in the Pakistani context. It serves as a foundation for subsequent research in this domain and offers valuable insights for curriculum developers. The findings can assist in identifying and integrating the appropriate lexical variety for each discipline, facilitating the creation of a more effective and tailored syllabus.

References

Abbas, H. (2018). *Linguistic variation across Pakistani English written registers: A multidimensional analysis*. (Master Thesis). Department of Applied Linguistics, Government College University, Faisalabad, Pakistan.

Abbas, H., Mahmood, M. A., & Asghar, S. A. (2020). Register variation in Pakistani written English: A multidimensional analysis. *Pakistan Journal of Society, Education and Language, 6*(1), 148-164.

Afful, J. B. A., & Mwinlaaru, I. N. I. (2012). *Sub-disciplinary variation and rhetoric in dissertation acknowledgements written by education students: The case of the University of Cape Coast, Ghana*. Retrieved on August 28, 2023 from https://ir.ucc.edu.gh/xmlui/handle/123456789/6559.

Aleem, S. (2013). *Linguistic variation across print advertisements in Pakistani media: A multidimensional analysis* (Doctoral Dissertation). International Islamic University, Islamabad, Islamabad, Pakistan.

Alotaibi, H. S. (2020). The thematic structure in research article abstracts: Variations across disciplines. *Cogent Arts & Humanities*, *7*(1), 1-12.

Alotaibi, H. S. (2021). The rhetorical structure of preface sections in textbooks: Variations across disciplines. *International Journal of Arabic-English Studies*, *21*(2), 165-182.

Atai, M. R., & Habibi, P. (2009). Exploring sub-disciplinary variations and generic structure of applied linguistics research article introductions using CARS Model. *Journal of English Language Pedagogy and Practice*, *2*(5), 26-51.

Azher, M., & Mehmood, M. A. (2016). Exploring variation across Pakistani academic writing: A multidimensional analysis. *NUML Journal of Critical Inquiry*, *14*(2), 86-113.

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), 30-43.

Azher, M., Faiz, R., Izhar, A., Ali, S., & Awan, R. U. N. (2019). Revealing disciplinary variation in Pakistani academic writing: A multidimensional analysis. *International Journal of English Linguistics, 9*(2), 258-272.

Azher, M., Ali, S., & Mahmood, M. A. (2021). Linguistic variation across disciplines in Pakistani academic writing: A multidimensional analysis. *Jahan-e-Tahqeeq, 4*(1), 43-67.

Aziz, A., Pathan, H., & Ali, S. (2017). Linguistic variation across major disciplinary groups of Pakistani academic writing: Multidimensional analysis of doctoral theses. *ARIEL-An International Research Journal of English Language and Literature, 27*, 27-60.

Biber, D. (1985). Investigating macroscopic textual variation through multifeature/multidimensional analyses. *Linguistics, 23*, 337-360.

Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language, 62*(2), 384-414.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison.* UK: Cambridge University Press.

Biber, D. (1999). *Longman grammar of spoken and written English*. UK: Longman.

Biber, D., Conrad, S., & Reppen, R. (2000). *Corpus linguistics: Investigating language structure and use*. UK: Cambridge University Press.

Biber, D., & Conrad, S. (2009). *Register, genre and style*. Cambridge: Cambridge University Press.

Cao, Y., & Xiao, R. (2013). A multi-dimensional contrastive study of English abstracts by native and non-native writers. *Corpora*, *8*(2), 209-234.

Chen, S. (2017). Disciplinary variations in academic promotional writing: The case of statements of purpose. *Functional Linguistics, 4*(1), 1-14.

Crystal, D. (1995). *A dictionary of linguistics and phonetics.* USA: Blackwell Publishing.

Ebrahimi, S. F., & Khedri, M. (2011). Thematicity in research article abstracts: A cross-disciplinary study. *Educational Quest-An International Journal of Education and Applied Social Sciences, 2*(3), 287-292.

Gholipour, B., & Saeedi, Z. (2019). Cross disciplinary rhetorical-linguistic variations in physical education research article abstracts in English as a lingua franca for academia context. *Iranian Journal of English for Academic Purposes*, *8*(4), 64-85.

Gray, B. (2013). More than discipline: Uncovering multi-dimensional patterns of variation in academic research articles. *Corpora*, *8*(2), 153-181.

Halliday, M. A. K. (1989). *Spoken and written language*. Oxford: Oxford University Press.

Halliday, M. A. K., & Hasan, R. (1990). *Language, context, and text: Aspects of language in a social-semiotic perspective.* UK: Oxford University Press.

Halliday, M. A. K. (1994). *An introduction to functional grammar.* London: Edward Arnold.

Heidari Kaidan, Z., Jalilifar, A., & Don, A. (2021). On the significance of disciplinary variation in research articles: Perspectives from nominalization. *Cogent Education, 8*(1), 1-18.

Hopkins, A., & Dudley-Evans, T. (1988). A genre-based investigation of the discussion sections in articles and dissertations. *English for Specific Purposes, 7*(2), 113-121.

Hussain, Z., Mahmood, M. A., & Azher, M. (2016). Register variation in Pakistani English: A multidimensional approach. *Science International, 28*(4), 391-402.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. Harlow: Pearson Education.

Hymes, D. (1979). *Soziolinguistik: zur Ethnographie der kommunikation*. Frankfurt am Main: Suhrkamp.

Hymes, D. (1984). Sociolinguistics: Stability and consolidation. *International Journal of the Sociology of Language, 45*, 39-45.

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.

Khany, R., & Tazik, K. (2010). A comparative study of introduction and discussion sections of sub-disciplines of applied linguistics research articles. *Journal of Research in Applied Linguistics, 1*(2), 97-122.

Mahmood, M. A. (2009). *A corpus-based analysis of Pakistani English*. Bahauddin Zakariya University, Multan, Pakistan.

McCabe, A. M. (1999). *Theme and thematic patterns in Spanish and English history texts*. Retrieved on August 28, 2023 from: http://www.wagsoft.com/systemics/archive/McCabe.phd.

Moohebat, M., Raj, R. G., Thorleuchter, D., & Kareem, S. B. A. (2017). Linguistic feature classifying and tracing. *Malaysian Journal of Computer Science*, *30*(2), 77-90.

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.

Ryvitytė, B. (2003). Research article introductions: Variations across disciplines and cultures. *Kalbotyra*, *53*, 93-100.

Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for Specific Purposes, 21*(1), 1-17.

Samraj, B. (2008). A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes, 7*(1), 55-67.

Shakir, M. (2020). *A corpus-based comparison of variation in online registers of Pakistani English using MD analysis* (Doctoral Dissertation). Universität Münster, Germany.

Suntara, W., & Usaha, S. (2013). Research article abstracts in two related disciplines: Rhetorical variation between linguistics and applied linguistics. *English Language Teaching*, *6*(2), 84-99.

Swales, J. M. (1990). *Genre analysis*. Cambridge: Cambridge University Press.

Teich, E. (2003). *Cross-linguistic variation in system and text.* Berlin: Mouton de Gruyter.

Trask, R.L. (1999). *Key concepts in language and linguistics*. London and New York: Routledge.