



## **Linguistic Variation across Major Disciplinary Groups of Pakistani Academic Writing: Multidimensional Analysis of Doctoral Theses**

**Abdul Aziz \***

[magsiabdulaziz@hotmail.com](mailto:magsiabdulaziz@hotmail.com)

**Habibullah Pathan \*\***

[pathanhabibullah@yahoo.com](mailto:pathanhabibullah@yahoo.com)

**Sajid Ali \*\*\***

[prof.sajidali@gmail.com](mailto:prof.sajidali@gmail.com)

---

**Abstract:** *This study aims at investigating the linguistic variation across two major disciplinary groups (physical sciences and biological & health sciences) of academic written English of Pakistani doctoral theses. Furthermore, the study also compares the language of physical sciences and biological & health sciences of Pakistani doctoral theses with the 23 genres of Biber's (1988) study. The corpus of Pakistani doctoral theses (PakDTh), including the disciplinary groups (Physical sciences and biological & health sciences) as sub-corpora, was built for this study. MD Analysis tagger (v 1.3.) was used to tag and analyze the data. The generated scores were statistically computed using SPSS 20.0 and were qualitatively interpreted through the 5 explored dimensions of Biber (1988) MD analysis. The findings reveal that the language of both the sub-corpora (physical sciences and biological & health sciences) significantly differs at dimensions 1,2 and 3. Whereas, they are relatively similar on dimension 4 and 5. Furthermore, both the disciplines are characterized with significant linguistic variation (along the 5 dimensions) in comparison with 23 genres of Biber's (1988) study. So far as the scope of the present study is concerned, it might be helpful for the researchers working in the field of sociolinguistics, world Englishes and lexicography.*

---

**Keywords:** *Pakistani Academic Written English, Language variation, Multidimensional analysis, Ph.D., Doctoral theses*

---

\* MS Scholar, Mehran University of Engineering and Technology, Jamshoro, Pakistan

\*\* Associate Professor & Director, English Language Development Centre,  
Mehran University of Engineering and Technology, Jamshoro, Pakistan

\*\*\* Assistant Professor, Government Municipal Degree College, Faisalabad, Pakistan

## Introduction

Canagarajah (2002) defines English for Academic Purpose (EAP) as “an international language” used as the source of communication with the researchers of the world. The language used in academic discourse reflects the users’ (readers’ and writers’) social, cultural, linguistic, professional and educational settlements. The use of English language in academic contexts around the world reveals that English has retained the status of a predominant language for academic discourse like other communication processes in which non-native speakers have been found in a challenging situation, particularly in academic writing (Carrió-Pastor, 2014). For which, the practice of English for academic purposes (EAP), with main purpose of training and helping students to learn the language of their target workplaces, plays an important role. EAP is related to academic tasks and activities involving specific communication and academic use of the language. EAP is often considered as a particular element of ESP at tertiary level of education. However, it is also required by both native and non-native students engaged in EAP writing tasks such as assignments, theses, and writing research papers as well as spoken activities (e.g., seminars, conferences and presentations).

Students, admitted to their concerned and particular disciplines for higher-education, are viewed as “novice and potential members of that discourse community” (Hardy, 2015), where they are expected to be surrounded with a range of registers and sub-registers. Most importantly, students’ often fail to differentiate between those registers (Moran, 2013). For this, Biber et al. (2006) put emphasis on students’ understanding towards the academic registers and sub-registers for better teaching and learning effects.

Writing, in any academic context, is a crucial feature of students’ learning and understanding of a discipline which has been influenced by enormous change and development in technology (Hyland, 2013). Students, at their tertiary level of education at universities or fresh graduates at professional workplaces, find themselves surrounded by written texts and face various problems because of their past practices of writing in their previous phase of schooling (Lillis, 2001).

Similarly, Ph.D. candidates, producing a thesis as an academic task and degree requirement, have the same purpose of sharing knowledge through their writing. Theses produced by Ph.D. scholars share their academic knowledge with their disciplinary specificity (Parry, 1998). This specificity, diversity of genres and language use by those involved in academic

practices (e.g., reading and writing research), is explained by Becher and Trowler (2001) and Hyland (2009) as “discourse communities and disciplines”.

New students at post-graduate level in local or international institutions often find it difficult to produce the academic writing texts for research. As described above, they are introduced to new genres and registers based on the communicative purposes of disciplinary specified texts. The international non-native students, enrolling in American and European universities for completing masters or doctoral programs, are required to pass TOEFL or GRE exams with high scores. After arrival, in spite of having high language proficiency level, they take a placement test for academic language skills and according to that evaluation and level of proficiency they are placed in academic writing classes (Cortes, 2007).

Indeed, the teaching of “writing” in EAP at tertiary level is dissimilar to the traditional process and practices because it focuses on improving students’ proficiency in a specific and targeted discourse community (Hardy, 2015). EAP practitioners teach specific types of writing which are demanded in particular situation (academic or professional), which make these contexts or situation based genres crucial and to be recognized by EAP teachers (Hyland, 2013). Undoubtedly, it is important for the teachers of EAP and academic writing to be updated with variety and with the characteristics of the language to be used by candidates in a specific academic context (e.g., Ph.D. theses and dissertation) or specified to the disciplines through empirical research (Hardy, 2015)

Though, this research attempts to analyze the textual variation of English in Pakistani doctoral theses from two perspectives by which the text will be examined comparatively and intra-textually (two major disciplinary groups of Physical sciences and Biological & health sciences) and inter-textually-in comparison with 23 genres from Lancaster-Oslo-Bergen (LOB) and Lundon-Lund Corpus analyzed by Biber (1998). This study will be conducted following Biber’s (1988) framework of Multidimensional analysis (MDA) and a representative corpus of Pakistan doctoral theses, composed of two sub-corpora of 100 theses from each disciplinary group.

The following two research questions will be attempted to be answered:

- 1- How far does the linguistic variation exist across two major disciplinary groups (Physical sciences) of Pakistani doctoral theses along five textual dimensions of Biber's (1988) study?
- 2- How far does the language of Pakistani doctoral theses of (Physical sciences and Biological& health sciences) differ in comparison with the registers of Biber's (1988) study along the five textual dimensions?

### **Literature Review**

In the contemporary language studies, academic language has been focused by many researchers. Many of the studies using computer-based approaches have extensively described individual and specialized linguistic features such as imperatives (Swales, Ahmad, Chang, Chavez, Dressen, & Seymour, 1998), conditional (Carter-Thomas and Rowley-Jolivet, 2008; Ferguson, 2001; Louwerse and Jeuniaux, 2008; Warchal, 2010), personal pronouns (Harwood, 2005; Kuo, 1999; Martínez, 2005; Okamura, 2009; Yeo and Ting, 2014), politeness markers (Alemi and Razzaghi, 2013; Ashtarian and Weisi, 2016; Chejnová, 2014; Hyland and Jiang, 2016; Lestari, Syafrizal and Elfrida, 2014; Purandina, Seken, Budasi and Lin, 2014; Rahmani, Rahmany and Sadeghi, 2014), citation patterns (Budd, 1990; Boshier, 1979; Charles, 2006; Ding, Zhang, Chambers, Song, Wang and Zhai, 2014; Gingras, 2008), vocabulary (Gardner, 2013; Hartshorn and Hart, 2016; Hwang, Lawrence and Snow, 2015) and, collocations (Ackermann and Chen, 2013; Chon and Shin, 2013; Durrant, 2009; Ordem and Bada, 2016). Although the studies mentioned above involved Biber's (1988) Multidimensional analysis, have not yet describe the variation in academic language registers or types considering structural and functional features collectively (see, extensive review of research by Biber, 2004).

Language of various academic registers and sub-registers, such as; textbooks (Biber et al., 2002; Biber et al., 2004; Conrad, 2001), learner's essays, classroom language, lectures, student handbooks, course brochures (Biber et al., 2004), research articles (Biber & Finegan, 1994), student papers and assignments (Nesi, 2008; Hardy, 2015), university prospectus, (Zehra & Shakir, 2015; Nasir & Shakir, 2015), abstracts of research articles (Cao & Xiao, 2013), have widely been researched multi dimensionally in terms of language variation, across-disciplines, register and cultures (Native and non-native).

Surprisingly, theses, one of the most important academic registers, have not been researched by language researchers as much as other academic registers (Groom, 2005; Bunton, 2002; Samraj, 2008). There have been various academic concerns by the research students and their guides regarding thesis writing. According to Mauch and Park (2003), "One of the surprising weaknesses in the thesis or dissertation process is that there is relatively little scholarly literature and a remarkably small number of empirical investigations about it" p.vii. Similarly, Ventola and Mauranen (1996: vii) highlight that very few guidebooks for writing research consider the linguistic analysis of texts that might have to be grasped by academic writers at the beginning. The avoidance in researching academic texts of thesis might be of various reasons. Bunton (2002) provides two explanations for this lack of research on this register; the first is the length of texts and the second is its one-time production with no other chance to rewrite it again improvingly.

Previous studies, focusing theses and dissertations (e.g., Bunton, 2002, Ketabi & Rahavard, 2013; Parry, 1998, Swales, 1990), have been conducted. These studies are noteworthy in the research of academic language from the perspective of genre analysis following Swales' (1990) CARS model, but are having limited significance in the domain of language variation studies. The reason is rhetorical perspectives, the only focus of these studies providing a description of the language use in theses and dissertations.

Ultimately, the review discussed above highlights the importance of the linguistic researches focusing language use in Ph.D. theses. So far, no existing study focusing language variation has embarked to explain the register of Ph.D. theses and also this particular register across disciplines. This research study tries to empirically investigate the language of Pakistani doctoral theses across two major disciplinary groups of Physical sciences and Biological and health sciences, based on Biber's (1988) powerful and statistical model of Multidimensional analysis.

## **Methodology**

### ***Corpus Building***

To the extent of our knowledge, there was not any existing representative corpus of Pakistani doctoral theses for studying the linguistic aspects and variation across major disciplinary groups. Therefore, a corpus of doctoral theses, representing Pakistani academic written English in doctoral theses

(PakDTh), was designed and developed. The standardized criteria (Sinclair, 2004) of corpus building and its features were carefully considered while building the corpus.

The PakDTh corpus consists of 2 sub-corpora representing the academic written English of Pakistani doctoral theses from two major disciplinary groups of Biological and Physical sciences. Both the sub-corpora contain 100 texts each of multiple disciplines under the major groups of the disciplines with 2.7 and 2.9 million words. The Biological Sciences group covers 13 concerned academic disciplines (Applied biological science, biochemistry, biotechnology, clinical medicines, health sciences, human Physiology, microbiology, molecular biology, pathology, pharmacy, plant sciences, zoological science and Biological sciences), whereas the physical sciences group covers only 4 disciplines (chemistry, physics, earth sciences (geology) and mathematics). The inclusion of the disciplines under the group representation has been limited to the list provided in Table 1 due to the availability in Pakistan research repository.

**Table 1.** Description of PakDTh Corpus

| <b>Disciplinary Group</b>                    | <b>Discipline</b>           | <b>Number of Theses</b> | <b>Tokens (words)</b> |
|----------------------------------------------|-----------------------------|-------------------------|-----------------------|
| <b><u>Physical Sciences</u></b>              | Chemistry                   | 46                      | 1,460,924             |
|                                              | Earth Sciences              | 9                       | 217,610               |
|                                              | Mathematics                 | 5                       | 130,340               |
|                                              | Physics                     | 40                      | 940,504               |
|                                              |                             | <b>Σ 100</b>            | <b>Σ 2,749,378</b>    |
| <b><u>Biological and Health Sciences</u></b> | Applied Biological Sciences | 2                       | 48,875                |
|                                              | Biochemistry                | 1                       | 18,048                |
|                                              | Biotechnology               | 7                       | 258,202               |
|                                              | Clinical Medicine           | 8                       | 162,297               |
|                                              | Health Sciences             | 2                       | 114,695               |
|                                              | Human Physiology            | 1                       | 20,678                |
|                                              | Microbiology                | 15                      | 385,539               |
|                                              | Molecular Biology           | 15                      | 371,971               |
|                                              | Pathology                   | 3                       | 86,120                |
|                                              | Pharmacy                    | 17                      | 569,131               |
|                                              | Plant Sciences              | 8                       | 248,625               |
|                                              | Zoological Sciences         | 5                       | 134,197               |
|                                              | Biological Sciences         | 16                      | 504,059               |
|                                              |                             | <b>Σ 100</b>            | <b>2,922,437</b>      |

The textual data for the corpus were collected from research repository of Pakistan. First, a list of the disciplines included in the repository data was generated to make possible representation of the available doctoral theses of disciplines. For the reliable distribution of the subjects, disciplines and sub-disciplines, the topics of the theses and disciplinary information was recorded and carefully considered with reference to the list by Joint academic coding system (JACS 3.0) developed and revised by Higher Education Statistics Agency (HESA) in 2012 and 2013 and the information available at Higher Education Commission Pakistan's Research repository.

At the time of theses selection from Pakistan research repository (PRR), file format was focused, which was necessary for the reliability and accuracy of the data. The file formats (Fully digitized) were considered focusing the time required for converting the data in appropriate file format with accuracy and the availability of resources. All the theses text files (200) of both the sub-corpora selected for the corpus building were submitted from the year 2007 to 2014.

The resulting size of the corpus is 5.67 million words, with biological sciences and physical sciences components containing approximately 2,922,437 and 2,749,378 words, respectively, as shown in Table 1.

### ***Tagging and Analysis***

The research was conducted following Biber's (1988) Multidimensional analysis (MDA). MDA is a corpus based research framework originally pioneered by Biber (1988) to explore and interpret linguistics variation of corpora. MDA uses various statistical and quantitative techniques to find variation across text varieties. Biber (1988) proposes a list of 67 linguistic features categorized into 16 grammatical classes (see, Biber, 1988, p.73). The corpus, targeted to be analyzed or examined, is annotated with lexical and grammatical tagging and using statistical techniques co-occurrences of the linguistic features are investigated for the interpretation of the underlying "dimensions of variation"

The MD Analysis have both linguistic and functional dimensional content. In Biber's (1988) framework, the groups of linguistic features in MDA are identified with the help of factor analysis of the linguistic features (variables) that are correlated within the texts representing a separate dimension as factors. The explanation of the resulted components (factors) "as functional dimensions" is based on the prior theories which state that these sets of the individual linguistic features produce underlying shared

communicative functions. It is assumed that co-occurrence of the linguistic features reflects certain shared linguistic and communicative functions. The patterns of co-occurrences are interpreted to assess the situational, social, and cognitive functions most widely shared by the features.

Each thesis in the corpus is renamed with a coded tag about its attribution to the disciplinary group, as they might be identified easily for further use during the study. In addition, the corpus data was loaded to an automated corpus tagger and analyzer software MAT Tagger V.1.3 (Nini, 2015), which is the replicate of the Biber's (1988) framework of computing and analyzing the data for MDA. The corpus was tagged automatically using Nini's (2015) MAT tagger 1.3. It uses Stanford tagger and MAT tag sets to tag the corpus data with the required linguistic features to produce the normalized frequency information to compute Bibers' (1988) dimension score.

## **Results and Findings**

This section describes the results of Multidimensional Analysis of Pakistani doctoral theses including both the disciplinary groups sub-corpora (Physical sciences and Biological & Health sciences) tagged and analyzed through Nini's (2015) Mat Tagger v.1.3 on the basis of the factors computed with reference to Biber's (1988) five dimensions.

Here, the comparison of mean dimension scores of both sub-corpora (Biological and Physical Sciences) with the dimension scores of seven main genres analyzed by Biber (1988) is presented. The mean dimension scores of the Pakistani doctoral theses were compared to identify the linguistic characterization of Pakistani academic disciplinary groups of biological and physical sciences with reference to five textual dimensions (see section), using a distributional scale. Biber (1988), using complex statistical factor analysis, identified five textual dimensions and underlying sets of co-occurring linguistics features. Factor analysis is used to identify the sets of linguistic features that co-occur with high frequency in a group (on textual Dimensions) with underlying linguistic functions in text. Each textual dimension of Biber's (1988) MDA produces two groups of linguistic features (Factor Loadings): (+ and – features). These linguistic features with positive and negative scores, on textual dimensions represent different communicative functions.

**Table 2.** Mean dimension scores of the MDA of PakDTh Corpus  
(Sub-corpora 1. Biological Science and 2. Physical Sciences)

| Dimensions                                                        | Biological Sciences<br>Pakistani Doctoral<br>Theses |                   | Physical Sciences<br>Pakistani Doctoral<br>Theses |                   |
|-------------------------------------------------------------------|-----------------------------------------------------|-------------------|---------------------------------------------------|-------------------|
|                                                                   | Mean                                                | Std.<br>Deviation | Mean                                              | Std.<br>Deviation |
| <b>1..Involved vs.<br/>Informational<br/>Production</b>           | -23.1778                                            | 1.713882          | -21.0821                                          | 2.853526          |
| <b>2. Narrative Vs.<br/>Non-narrative<br/>Concerns</b>            | -4.1889                                             | 0.843197          | -5.0327                                           | 0.632197          |
| <b>3. Explicit vs.<br/>Situation-<br/>Dependent<br/>Reference</b> | 8.649                                               | 1.319041          | 7.6543                                            | 1.901946          |
| <b>4. Overt<br/>Expression of<br/>Persuasion</b>                  | -6.2067                                             | 0.849676          | -5.9537                                           | 1.205407          |
| <b>5. Abstract vs.<br/>Non-abstract<br/>information</b>           | 3.4734                                              | 1.585638          | 3.2096                                            | 2.017833          |

### ***Findings for RQ1: Comparison across major disciplinary group***

In this section, two major disciplinary groups of Pakistani doctoral theses are compared on the basis of mean dimension scores of both the sub-corpora (Physical sciences; Biological and health sciences). The mean scores of 5 investigated dimensions (as per 88 MD Analysis) of the texts are compared through statistical analysis of variance using SPSS (v.20). The mean scores of the texts along 5 dimensions are given in Table .2.

#### ***Dimension 1 (D1) - Involved versus Informational Production***

Dimension 1 of Biber's (1988) MD Analysis distinguishes the texts along the polarity of (+) Involved and (-) informational production. The mean scores of this dimension from both the sub-corpora of major disciplinary groups (BHSc and PhSc) of Pakistani doctoral theses corpus (PakDTh) are (-23.1778) and (-21.0821) respectively. Both of the disciplinary groups loaded highly negative on this dimension.

**Table 3.** ANOVA for Dimension-1: (Involved vs. Informational Production)

|                | Sum of Squares | Df  | Mean Square | F      | Sig. |
|----------------|----------------|-----|-------------|--------|------|
| Between Groups | 219.598        | 1   | 219.598     | 39.639 | .000 |
| Within Groups  | 1096.920       | 198 | 5.540       |        |      |
| Total          | 1316.518       | 199 |             |        |      |

$$F(1,198) = 39.639, p = .000 < 0.05$$

There was a statistically significant difference between disciplinary groups on dimension 1 as determined by one-way ANOVA ( $F(1,198) = 39.639, p = .000$ ). The results revealed that the text of Pakistani doctoral theses of physical sciences ( $M = -21.0821, SD = 2.8$  min) was statistically significantly less informational than of biological and health sciences ( $M = -23.1778, SD = 1.7$  min).

### *Dimension 2 (D2) - Narrative versus Non-Narrative Concerns*

D2 of Biber's (1988) MD Analysis differentiates the texts along the polarity of (+) Narrative and (-) Non-narrative concerns. The mean scores of this dimension from both the sub-corpora of major disciplinary groups (BHSc and PhSc) of Pakistani doctoral theses corpus (PakDTh) are (-4.1889) and (-5.0327) respectively. Both of the disciplinary groups loaded highly negative on this dimension.

**Table 4.** ANOVA for Dimension-2: (Narrative vs. Non-narrative Concerns)

|                | Sum of Squares | df  | Mean Square | F      | Sig. |
|----------------|----------------|-----|-------------|--------|------|
| Between Groups | 35.600         | 1   | 35.600      | 64.106 | .000 |
| Within Groups  | 109.955        | 198 | .555        |        |      |
| Total          | 145.555        | 199 |             |        |      |

$$F(1,198) = 64.106, p = .000 < 0.05$$

There was a statistically significant difference between disciplinary groups on D2 as determined by one-way ANOVA ( $F(1,198) = 64.106, p = .000$ ). The results revealed that the text of Pakistani doctoral theses of biological and health sciences ( $M = -4.1889, SD = 0.843197$ ) was statistically significantly with less non-narrative concern than of physical sciences ( $M = -5.0327, SD = 0.632197$ ).

### *Dimension 3 (D3) - Explicit vs. Situation Dependent Reference*

D3 of Biber's (1988) MD Analysis differentiates the texts along the polarity of (+) Explicit and (-) situation dependent references. The mean scores of this dimension from both the sub-corpora of major disciplinary groups (BHSc and PhSc) of Pakistani doctoral theses corpus (PakDTh) are (8.649) and (7.6543) respectively. Both of the disciplinary groups loaded highly positive on this dimension.

**Table 5.** ANOVA for Dimension-3: (Explicit vs. Situation Dependent Reference)

|                | Sum of Squares | df  | Mean Square | F      | Sig. |
|----------------|----------------|-----|-------------|--------|------|
| Between Groups | 49.471         | 1   | 49.471      | 18.469 | .000 |
| Within Groups  | 530.370        | 198 | 2.679       |        |      |
| Total          | 579.841        | 199 |             |        |      |

$$F(1,198) = 49.471, p = .000 < 0.05$$

There was a statistically significant difference between disciplinary groups on D3 as determined by one-way ANOVA ( $F(1,198) = 49.471$ ,  $p = .000$ ). The results revealed that the text of Pakistani doctoral theses of biological and health sciences ( $M = 8.649$ ,  $SD = 1.319041$ ) was statistically significantly highly explicit referential discourse than of physical sciences ( $M = 7.6543$ ,  $SD = 1.901946$ ).

#### *Dimension 4 (D4) - Overt Expression of Persuasion*

D4 of Biber's (1988) MD Analysis differentiates the texts along the polarity of (+) overtly persuasive/argumentative (-) non-persuasive/non-argumentative discourse. The mean scores of this dimension from both the sub-corpora of major disciplinary groups (BHSc and PhSc) of Pakistani doctoral theses corpus (PakDTh) are (-6.2067) and (-5.9537) respectively. Both of the disciplinary groups loaded highly negative on this dimension.

**Table 6.** ANOVA for Dimension-4: (Overt Expression of Persuasion)

|                | Sum of Squares | df  | Mean Square | F     | Sig. |
|----------------|----------------|-----|-------------|-------|------|
| Between Groups | 3.200          | 1   | 3.200       | 2.943 | .088 |
| Within Groups  | 215.321        | 198 | 1.087       |       |      |
| Total          | 218.521        | 199 |             |       |      |

$$F(1,198) = 2.943, p = .088 > .005$$

Statistically, there was insignificant difference between disciplinary groups on D4 as determined by one-way ANOVA ( $F(1,198) = 2.943$ ,  $p = .088$ ). The results revealed that there was no statistical significant difference between the text of Pakistani doctoral theses of biological and health sciences ( $M = -6.2067$ ,  $SD = 0.849676$ ) and physical sciences ( $M = -5.9537$ ,  $SD = 1.205407$ ). Texts of both the disciplinary groups were similarly non-persuasive and argumentative in nature.

#### *Dimension 5 (D5) – Abstract versus Non-Abstract*

D4 of Biber's (1988) MD Analysis differentiates the texts along the polarity of (+) overtly persuasive/argumentative (-) non-persuasive/non-argumentative discourse. The mean scores of this dimension from both the sub-corpora of major disciplinary groups (BHSc and PhSc) of Pakistani doctoral theses corpus (PakDTh) are (3.4734) and (3.2096) respectively. Both of the disciplinary groups loaded moderately positive on this dimension.

**Table 7.** ANOVA for Dimension-5: (Abstract vs. Non-abstract Information)

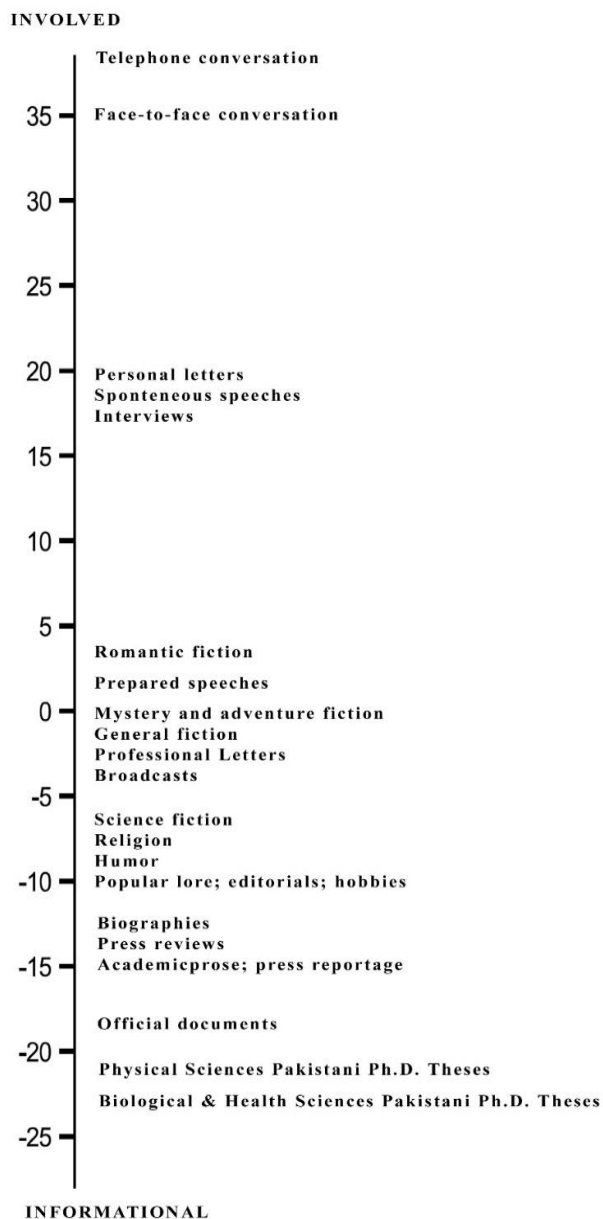
|                | Sum of Squares | df  | Mean Square | F     | Sig. |
|----------------|----------------|-----|-------------|-------|------|
| Between Groups | 3.480          | 1   | 3.480       | 1.057 | .305 |
| Within Groups  | 652.004        | 198 | 3.293       |       |      |
| Total          | 655.483        | 199 |             |       |      |

$$F(1,198) = 1.057, p = .305 > 0.05$$

Statistically, there was insignificant difference between disciplinary groups on D4 as determined by one-way ANOVA ( $F(1,198) = 1.057, p = .305$ ). The results revealed that there was no statistical significant difference between the text of Pakistani doctoral theses of biological and health sciences ( $M = 3.4734, SD = 1.585638$ ) and physical sciences ( $M = 3.2096, SD = 2.017833$ ). Texts of both the disciplinary groups were similarly abstract and impersonal in nature.

**Findings for RQ2: Comparison with 23 genres studied by Biber (1988)**

*D1: (Involved vs. Information Production)*



**Figure 1.** Mean dimension scores for Pakistani Ph.D. Theses (Biological and Health sciences; Physical Sciences) in comparison with the 23 genres studied by Biber (1988) on Dimension 1 (Involved vs. Informational)

It is shown in figure 1 that on dimension 1, Pakistani academic text of Ph.D. theses across major disciplinary groups (Physical sciences; Biological and health sciences) obtained the least score, which clearly points that both the groups are informational from communicative and productive perspectives.

The following text samples 1 (Biological and Health Sciences) and 2 (Physical Sciences) from PakDTh show the textual components characterizing the texts with informational (**bold**) and involved (underlined) discourse features.

A **longitudinal study** showed that **high** consumption of **smokeless tobacco** is associated with **metabolic syndromes**, although these **individuals** also had a deprived **socio economic status**, which has a **significant** association with **metabolic disorders**, including **diabetes mellitus** (Norberg et al, 2006, Javed et al, 2007). However, it has also been reported that there is no association between **snus** consumption and the **prevalence of Type 2 Diabetes** (Eliasson et al, 2004). **Gutka** has the **highest** concentration of **nicotine** compared to other forms of **smokeless tobacco**.

#### Text Sample 1. PakDTh-BHSc.OPATH01

**HPLC** is a **preferred technique** for identifying and quantifying **phenolic compounds**. **Phenolic compounds** are non-volatile so their analysis on GC (Gas chromatography) is not generally recommended. But with suitable derivatization, **phenolic compounds** can be analyzed through GC or GCMS. However, **HPLC** is currently most popular and **reliable technique** for **phenolic compounds** (Robards, 2003; Khoddami et al., 2013) with a **wide range** of commercially available columns (Stalikas, 2007). For the analysis of **phenolic compounds** on HPLC, **C18 columns** are mostly used. **Gradient elution** is recommended because of complexity of **phenolic compounds**. Mostly **binary solvents** are used as **mobile phase**, one component is water and other is less **polar organic solvent** such as **acetonitrile** or **methanol**. Reverse **phase LC** is typically used. **Polar compounds (phenolic acids)** elute first then compounds with decreasing polarity like **cinnamic acid derivatives** followed by **flavonoids** (Bernal et al., 2011).

#### Text Sample 2. PakDTh-PHSc.CHEM03

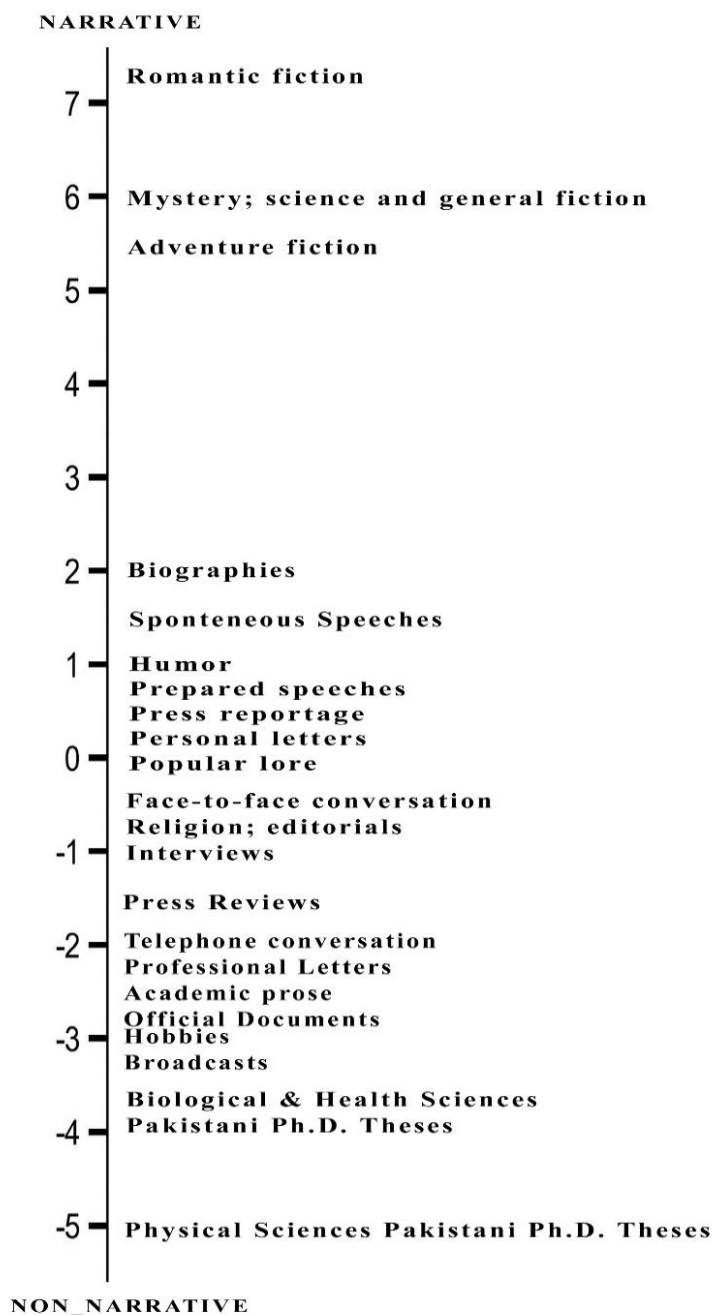
The mean dimension scores on D-1 (Involved vs Informational), achieved by the two sub-corpora of Pakistani doctoral theses (Physical sciences & Biological and health sciences), as sub-registers of academic texts representing the concerned disciplinary major groups, placed closer to the official documents (register). The results produced by MAT tagger revealed that the text analyzed (PakDTh) is characterized with more informational features as compared to all other 23 genres of Biber (1988) examined in LOB and London-Lund corpus. The text of Pakistani Ph.D. in both major disciplinary groups of physical sciences, and biological and health sciences showed significant differences with 11 genres (science fiction, religion, humor, popular lore, editorials, hobbies, biographies, press reviews, academic prose, press reportage and official documents) placed on the negative polarity on dimension 1 (Involved vs. informational production) as per Biber's (1988) findings for LOB and London-Lund corpus. According to Biber's (1988, p.107) "*communicative parameter of production circumstances*", the results of dimension can be considered to assume that the texts produced by Pakistani Ph.D. scholars for their theses are carefully produced and edited with the precision in the selection of lexical items; and textual structures.

Out of 67 linguistic features included in Biber's (1988) MDA, few linguistic features weighed highly negative on dimension 1 which includes nouns, prepositions phrases, TTR (type/token ratio), and attributive adjectives. Whereas, subordination features (Past participial WHIZ deletion & present participle WHIZ deletion), agentless passives and place adverbials were less important than the features discussed above because they were also found with higher weights on other dimensions. High frequencies of these features can be responsible for "*...high informational focus and a careful integration of information in a text. Biber (1988) p.104*".

Similarly, the results and interpretation in the light of the features involved for the resultant dimension polarity indicates that, the text of Pakistani Ph.D. theses of both the disciplinary groups (physical sciences, and biological and health sciences) are integrated with high amount of information. The informational production quality of the text can be assessed and interpreted based on high frequencies of these co-occurring features, which share these functions. In addition to the features with negative loading (weights) discussed above, other linguistic features with positive weights are also recognized as strong features in the texts with involved and interactional production, on dimension 1 (opposite to informational production).

The frequencies of these features (with positive loadings/weights) in informational production texts is less than the features with negative weights. Both of the features with (+ and -) weights have strong corresponding relationship in between and are highly influential on one another. It means that when a text contains negative linguistic features with high frequency rate, conversely the positive linguistic features will occur with low frequency. Biber (1988) has provided a sufficient discussion about the relationship between these features. In summary, Pakistani academic texts (Ph.D. Theses), in both of the disciplinary groups might be informational rather than involved or interactional in communicative and productive concerns due to high frequency of the co-occurring linguistic features sharing functions.

*D2: (Narrative vs. Non-narrative)*



**Figure 2.** Mean dimension scores for Pakistani Ph.D. Theses (Biological and Health sciences; Physical Sciences) in comparison with the 23 genres studied by Biber (1988) on Dimension 2 (Narrative vs. Non-Narrative)

In Biber's (1988) MDA, dimension 2 deals with the difference between narrative vs. non-narrative involvements. PakDTh texts of both the major disciplinary groups (Biological and health sciences; physical sciences) resulted non-narrative in nature, which shows exposition, procedural information and description. On the same time, the physical sciences texts are placed least on dimension 2 as compared to biological and health sciences. Both the textual groups are scientific and experimental descriptions, so they contain experimental information, factual descriptions; and practical or technical information. The text samples 3 and 4 (given below) from both the sub-corpora show how the features, causing narrative concern into the texts, are less frequent.

A higher prevalence of oral mucosal lesions in patients with diabetes **has** been discussed in the literature (Manfredi et al 2004, Seyhanet al 2007, Skamagas et al 2008). However there is **no** consensus as to whether lesions in patients with diabetes are detected more frequently than in patients without diabetes (Sandburg et al 2000, Quirino et al 1995, Guggenheimer et al 2000). A study done by Manfredi et al 2004, Guggenheimer et al 2000, **found** that oral soft tissue diseases **have** been associated more frequently in type 2 Diabetes Mellitus population than in the nondiabetic (control) group. In view of research finding, this study **was** designed to observe the relationship between oral pathological lesions and type 2 diabetes mellitus in Pakistani population. The prevalence of oral mucosal alteration found in this study **was** (60.8%) in type 2 diabetes mellitus and (39.2%) in non-diabetic individuals.

### Text Sample 3. PakDTh-BHSc.OPATH01

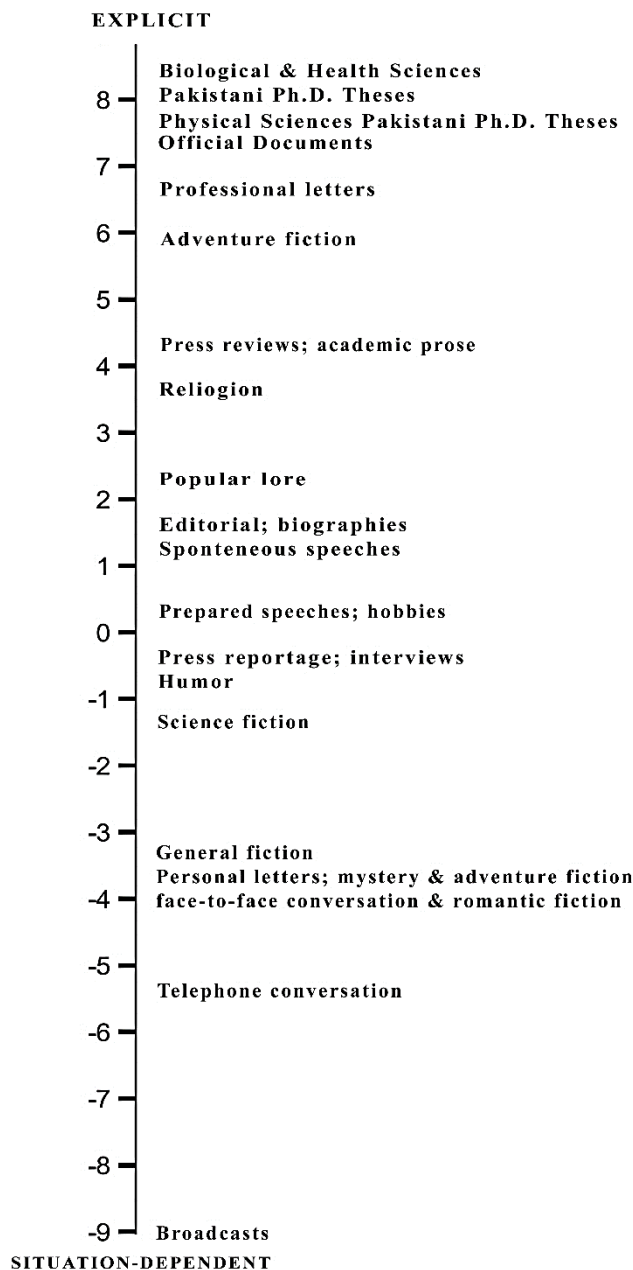
Overall results **showed** that all the plant extracts **showed** antifungal activity but less than standard Terbinafine. All the plant extracts **gave** activity against *A. flavus*, *A. niger* and *F. solani*. But a few extracts **had** not shown activity against *H.spermium*. From the results of disc diffusion assay and MIC, it **was** observed that aqueous mixtures of all the solvents (ethanol, methanol and acetone) exhibited better inhibitory activities against all the four fungal strains in comparison to their pure state. Tiwai et al., (2011) **described** that more flavonoid compounds **were** extracted with 70% ethanol as compared to pure ethanol due to higher polarity of the solvent by the addition of water. Ethanol can easily penetrate through cell membrane to extract phyto-constituents from plant material. **They** also **reported** that all the antimicrobial compounds of plants are saturated or aromatic organic compounds and can be easily extracted through methanol or ethanol. Ganora (2008) also **reported** that biological activities of plant extracts depends on the alcohol and water concentration in the extraction method.

### Text Sample 4. PakDTh-PHSc.CHEM03

The obtained results for mean dimension scores have placed the text of Pakistani Ph.D. theses of both the groups (Physical sciences; Biological and health sciences) closer to broadcasts, one of the 23 genres examined by Biber (1988), on the negative pole along dimension 2 (as shown in Fig.2). The results revealed that the text analyzed (PakDTh) is characterized with more non-narrative concerns as compared to all other 23 genres of Biber (1988) examined in LOB and London-Lund corpus. The text of Pakistani Ph.D. in both major disciplinary groups of physical sciences, and biological and health sciences showed significant differences with the 12 genres (popular lore, face-to-face conversation, religion, editorials, interviews, press reviews, telephone conversations, professional letters, academic prose, official documents, hobbies and broadcasts) placed on the negative polarity on dimension 1 (Involved vs. informational production) as per Biber's (1988) findings for LOB and London-Lund corpus.

The features, on dimension 2-responsible for the non-narrative concerns into the text, are present tense verbs and attributive adjectives. These features are also marked as strong features on dimension 1. The frequency of present tense verbs, as compared to dimension 2, is significantly higher on dimension 1, whereas attributive adjectives have insignificantly higher rate of frequency on dimension 1. Present tense verbs and attributive adjectives dispose of the immediate matters and represent the more frequently elaborated nominal information into a text having concerns other than narrative. In addition to the features with negative loading (present tense verbs and attributive adjectives), other linguistic features with positive weights- past tense verbs, 3<sup>rd</sup> person pronouns, perfect aspect verbs, public verbs, synthetic negation and present participial clauses- can be recognized as the features marking the texts with narrative production, on dimension 2 (opposite to others or non-narrative concerns). The frequencies of these features (with positive loadings) in non-narrative and other production texts are less than the features with negative weights. Concluding, Pakistani academic text (Ph.D. theses), due to less dimension scores along dimension 2 is marked with non-narrative concern.

*D3: (Explicit vs. Situation-dependent reference)*



**Figure 3.** Mean dimension scores for Pakistani Ph.D. Theses (Biological and Health sciences; Physical Sciences) in comparison with the 23 genres studied by Biber (1988) on Dimension 3 (Explicit vs. Situation-dependent)

Dimension 3 distinguishes between two types of discourse: one with “*highly explicit and context-dependent reference*” and the second with “*situation-dependent and non-specific reference*” (Biber, 1988, p.110). The results show that PakDTh texts of both the groups (as shown in figure 3), as compared to Biber’s (1998) analysis of MDA for LOB and London-Lund corpus, achieved significantly high dimension score on positive polarity along dimension 3. This indicates that the text is highly explicit in nature.

Text samples 5 and 6 clearly portrays the features causing explicit references (**bold**) into the texts in contrast to the features relevant to situation-dependent references (underlined).

Oral **pigmentation** due to intrinsic factor: The major intrinsic factor causing oral **pigmentation** is Peutz-Jeghers syndrome. It is an autosomal dominant disorder characterized by intestinal polyposis in **association** with mucocutaneous melanocytic macules. These lesions, often convergent **and** different in size **and** color, appear in almost all cases periorally and on the lips and buccal mucosa. It may involve any oral region, and the degree of **pigmentation** and oral **involvement** vary among affected individuals. Pigmented macules on the cutaneous surfaces covering the **extremities and face** are less frequently observed. The relative risk of developing cancer in this syndrome is increased 15fold compared with that of the general **population**. The cancer primarily involves the gastro intestinal tract (including the pancreas and the luminal organs), the female **and** male reproductive tracts, and the lungs (Giardiello et al 1987)

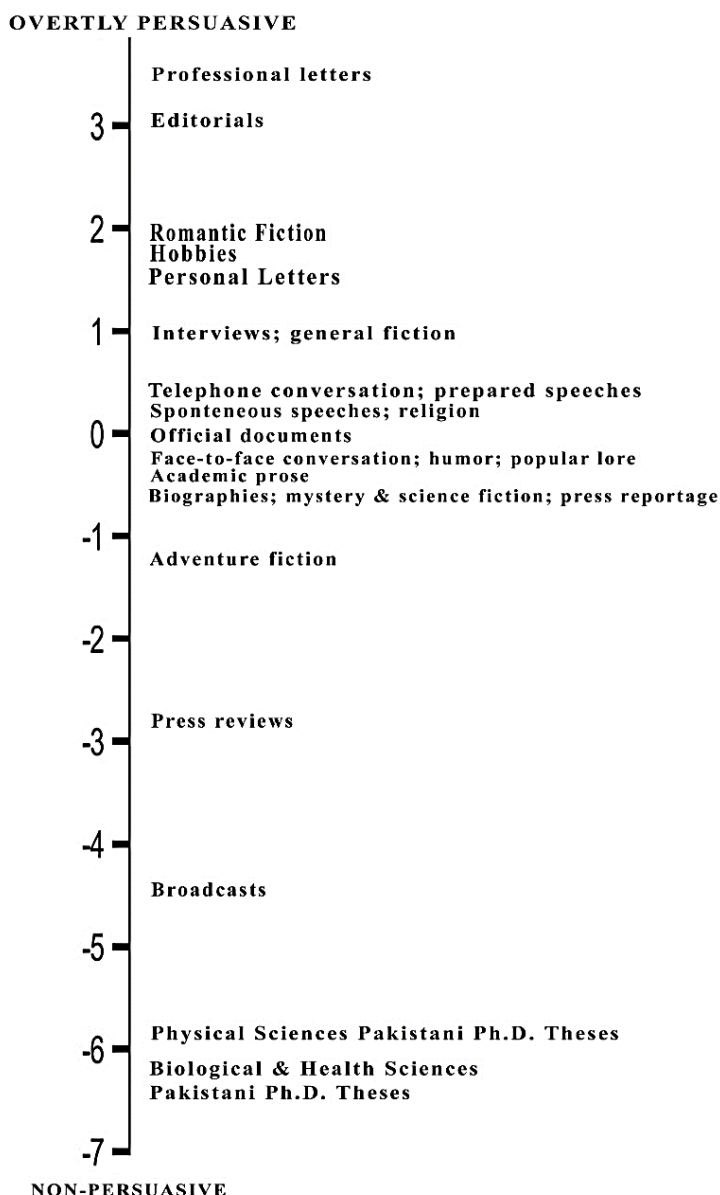
#### Text Sample 5. PakDTh-BHSc.OPATH01

Disk diffusion assay was used for the **determination** of antibacterial **activity** of plant extracts (Bulbul et al., 2011). The required number of Petri plates were washed **and** autoclaved. And they were allowed to cool under laminar air flow. **Solutions** of plant extracts (10 mg/mL) were prepared in 10% dimethyl sulfoxide (DMSO). Nutrient agar **solution** was prepared **and** autoclaved. Aseptically transfer about 20 mL of nutrient agar medium (containing bacterial culture, 50 µL) into each sterile petri plate and allowed to solidify. Dried **and** sterilized discs (10 mm) of wicks sheet were prepared. Then discs containing 100 µL plant extracts **solution** (10mg/mL) were placed on nutrient agar medium. Discs containing standard antibiotic (Rifampicin, 100 µg/mL) and 10% dimethyl sulfoxide were also placed in each petri plate as positive **and** negative control. The plates were then incubated at 37°C for 24h. After **incubation**, the antibacterial **activity** was recorded by measuring the diameter of zone of **inhibition** around the disc using zone reader (mm). The **experiment** was carried out in three replicates and average was calculated.

#### Text Sample 6. PakDTh-PHSc.CHEM03

The results of mean dimension scores have placed the text of Pakistani Ph.D. theses of both the groups (Physical sciences; Biological and health sciences) closer to official documents genre examined by Biber (1988), high on the positive pole along dimension 3 (as shown in Fig.3). The results revealed that the text analyzed (PakDTh) is highly characterized with explicitness as compared to all other 23 genres of Biber (1988) examined in LOB and London-Lund corpus. The text of Pakistani Ph.D. in both major disciplinary groups of physical sciences, and biological and health sciences showed significant differences with the 12 genres (official documents, professional letters, adventure fiction, press-reviews, academic prose, religion, popular lore, editorials, biographies, spontaneous and prepared speeches, and hobbies) placed on the positive polarity on dimension 3 (Explicit vs. situation-dependent discourse) as per Biber's (1988) findings for LOB and London-Lund corpus. The features on dimension 3-marking the explicit and context dependent discourse production, are three; WH relative clauses on object positions, pied piping constructions, and WH relative clauses on subject positions, which appeared highly frequent. Furthermore, two more features, Phrasal co-ordinations and nominalizations, had relatively slighter positive loadings on this dimension. WH relative clauses function as a specifier for the identification of referents overtly into the text, to create more clarity towards the referents for an addressee or receiver (reader) of the text. The relative co-occurrence of WH relative clauses and phrasal co-ordination on dimension 3 shows that the text of Pakistani Ph.D. theses was characterized with situation or context-independence, referential explicitness and integration of informational references.

*D4: (Overtly persuasive vs. Non-persuasive)*



**Figure 4.** Mean dimension scores for Pakistani Ph.D. Theses (Biological and Health sciences; Physical Sciences) in comparison with the 23 genres studied by Biber (1988) on Dimension 4 (Overt Expression of persuasion)

The features marking persuasion, only with positive loadings, make the dimension 4 as an exceptional from other dimensions. The persuasive nature of a text can be either with the writer's or the speakers' persuasion (point of view) explicitly marked or argumentatively structured discourse to influence the receiver (Biber, 1988, p.111). The text of Pakistani doctoral theses has exclusion of both the considerations discussed above. We know that both the text groups of Pakistani Ph.D. Theses are scientific academic discourse and tend to elaborate the contained information in a direct way. The receiver or addressee is exposed to the factual and experimentally observed data and informational contents of the text for its approval or rejection. Pakistani Ph.D. theses text among the groups (Physical sciences; Biological and health sciences) can be considered as a variety of non-persuasive and non-argumentative discourse on the negative polarity of dimension 4 (Overt expression of persuasion). The following text samples exemplify less occurrences of persuasion and argumentation into the PakDTh corpus texts with.

Several terms that describes the characteristic and different origin of halitosis.

Physiologic halitosis: This type is common on awakening morning breath malodour is temporary and is mainly due to normal night time hyposalivation and it is of no significance.(Scully et al, 1994; Sanz et al, 2001; Outhouse et al, 2006; Porter and Scully, 2006; Fukui et al, 2008). Menstruation **may** also lead to halitosis (Kawamoto et al, 2010). Starvation **can** also cause similar malodour. Physiologic malodour **can** easily managed by oral hygiene care like tooth brushing and tongue brushing or scraping and rinsing the mouth with antiseptic mouth wash (Faveri et al, 2006). Tongue scraping alone is found **to** be of no use and doesnot prevent morning breath until it is incorporated with tooth brushing in periodontally healthy individuals (Haas et al, 2007; Allaker et al, 2008).

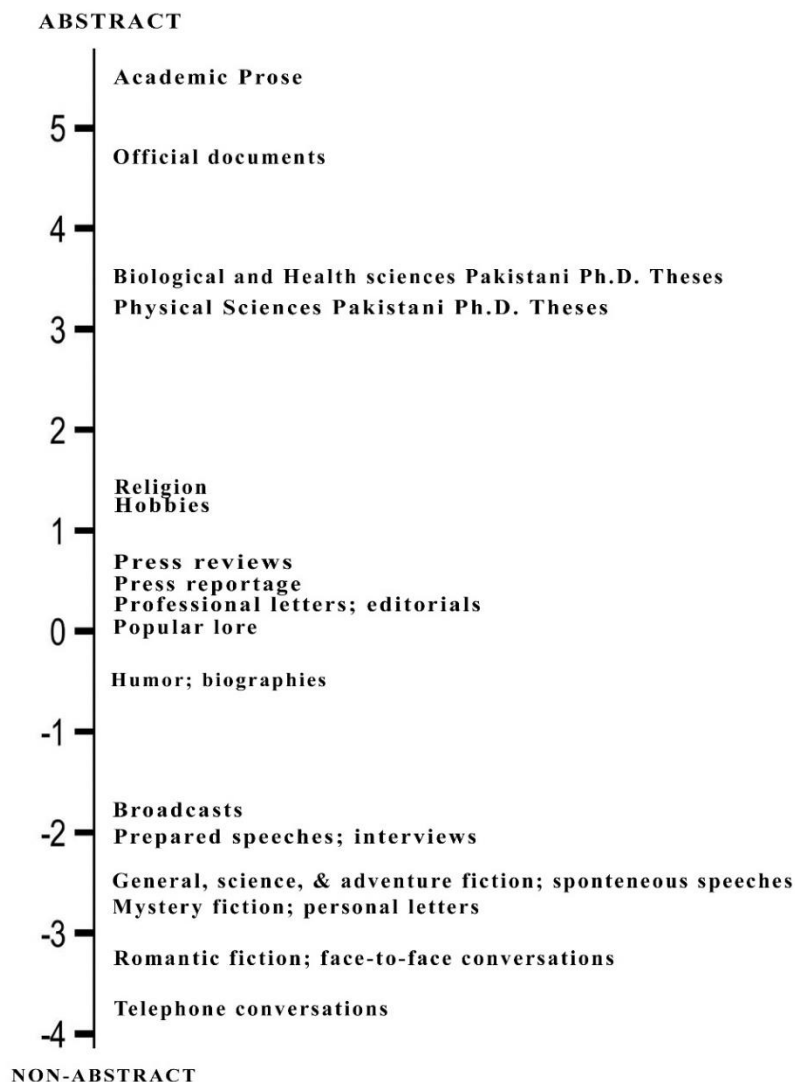
#### **Text Sample 7. PakDTh-BHSc.OPATH01**

For comparison of extraction techniques, aqueous ethanolic extracts of all the plants obtained from four techniques were analyzed through HPLC. Aqueous ethanolic extracts obtained through sonication gave more phenolic contents. This study provides useful information about the phytochemicals, antioxidant and antimicrobial activities of the selected medicinal plants native to Pothohar plateau. The utilization of these indigenous plants as potential source of natural antioxidants and antimicrobial agents is encouraged. However, future in-vivo studies are **recommended** to further evaluate the antimicrobial and antioxidant actions of these plants for therapeutic applications. Moreover, the isolation and characterization of bioactive compounds is also **suggested**.

### **Text Sample 8. PakDTh-PHSc.CHEM03**

Furthermore, the mean dimension scores of both the major disciplinary groups placed them comparatively closed to the genre of broadcasts according to the analysis of LOB and London-Lund corpus by Biber (1988). The results revealed that the analyzed text (PakDTh) is not characterized with persuasion and argumentation as compared to all other 23 genres of Biber (1988) examined in LOB and London-Lund corpus. The text of Pakistani Ph.D. in both major disciplinary groups of physical sciences, and biological and health sciences showed significant differences with the 3 genres (adventure fiction, press reviews and broadcasts) placed on the negative polarity on dimension 4 (Overt expression of persuasion) as per Biber's (1988) findings for LOB and London-Lund corpus. The linguistic features with positive loadings that co-occur in a text for the production of shared function of persuasion and argumentation (Biber, 1988, p.111) (which are significantly less frequent in PakDThcorpus) on dimension 4 include; prediction modals, possibility modals, necessity models, suasive verbs, conditional subordination and split auxiliaries.

*D5: (Abstract vs. Non-Abstract)*



**Figure 5.** Mean dimension scores for Pakistani Ph.D. Theses (Biological and Health sciences; Physical Sciences) in comparison with the 23 genres studied by Biber (1988) on Dimension 5 (Abstract vs. Non-Abstract)

5

Dimension 5 distinguishes between two varieties of discourse containing information: one with “*abstract or impersonal focus*” and the second with “*non-abstract or non-impersonal focus*” (Biber, 1988, p.113). The results show that the text of Pakistani Ph.D. thesis of both the groups (as shown in figure 5), compared with Biber’s (1998) analysis of MDA, achieved moderate dimension scores on positive polarity along D5. This indicates that the text of the Ph.D. theses is abstract, impersonal and formal in nature. The examined text of PakDTh corpus is scientific academic discourse and is contained with technical style of information illustration which obviously requires impersonal and abstract structure for its communicative functions. The text samples 9 and 10 with (**bold elements**) further illustrate the characteristics of PakDTh corpus texts as abstract in relation to their positive scores on D5.

The virulence of *Candida* species, especially *C. albicans*, **increased** in immunocompromized subjects. Immunosuppression provides an excellent platform for fungal proliferation. Diseases where the immune system **is suppressed** include DM, liver cirrhosis, acquired immune deficiency syndrome and nutritional deficiencies (Pereira et al, 2004). Xersotomia is a common complaint in diabetic subjects predisposes the oral mucosa to opportunistic infections by microorganisms, especially *C. albicans* (Rees, 1994). It **has been** documented that patients with DM harbor high levels of *C. albicans* in their oral mucosa **while** intensive *Candidal* colonization **can be seen** in diabetic patients with elevated glycemic levels (Belazi et al, 2005).

#### **Text Sample 9. PakDTh-BHSc.OPATH01**

Superoxide anions are one of the most damaging free radicals that attack on biomolecules directly or indirectly by forming OH, H<sub>2</sub>O<sub>2</sub>, peroxy nitrite or singlet oxygen during pathological diseases (Pracheta et al., 2011). In the oxidation-reduction reactions of cells, superoxide radicals **are formed** normally, later on their effects **can be magnified** as they produce many types of cell destroying free radicals other oxidizing agents (Subramanian and Vedanarayanan, 2012). SOR scavenging activity **was measured** using NBT (Nitro blue tetrazolium reagent). In this assay, superoxide radicals **are generated** in the reaction mixture. These radicals react with NBT, which **is reduced** to nitrite. Nitrites in the presence of EDTA gives a colour that **can be measured** at 560 nm (Veena and Mishra, 2011).

#### **Text Sample 10. PakDTh-PHSc.CHEM03**

The results of mean dimension scores have placed the text of Pakistani Ph.D. theses of both the groups (Physical sciences; Biological and health sciences) closer to official documents genre examined by Biber (1988), moderately high on the positive pole along dimension 5 (as shown in Fig.5). The text of both the disciplinary groups of physical sciences, and biological and health sciences showed its position along this dimension as moderate compared with the 9 genres of Biber's (1988) analyzed of MD of LOB and London-Lund corpus with positive loadings. Whereas, these texts of the disciplinary groups are placed on D5 with significantly less mean scores than 2 genres of (academic prose and official documents) and higher than 7 genres (popular lore, professional letters, editorials, press reportage, press reviews, hobbies and religion). The linguistic features, which strongly offer impersonal distinction to text with positive weights and high frequency of co-occurrences on dimension 5, are by-passives agentless passives, past participial clauses and past participial WHIZ deletions. Conjuncts and adverbial clauses also support by-passives and agentless passives to produce the relevant textual characteristic (abstractness) on dimension 5.

### **Discussion and conclusion**

This study attempted to analyze linguistic variation of Ph.D. theses registers comparatively with other genres (e.g., telephone conversation, general academic prose, fiction, letters etc.) studied by Biber (1988), and also inter-textual variation across major disciplinary groups. This not only puts insight into the register comparison but also gives chance to introduce a new register of Ph.D. theses in Pakistani English variety and academic discourse.

Some of the existing studies (Nesi, 2015) investigated the cross-disciplinary academic language characteristics and language variation through the MDA approach. As far as my search for the review is concerned, no such evidence has been found for the MDA studies focusing on the language of dissertations or theses in particular. According to the results of this study, Nesi's (2015) cross-disciplinary variation of BAWE (British academic written English) corpus, there is resemblance in the findings in terms of intra-textuality.

The language of PakDTh, among both the sub-corpora including Physical sciences (mean score=-21.0821) and Biological & Health sciences (mean score=-23.1778), was placed at extreme high position with positive polarity on dimension 1 (D1), and closer to official documents (mean score= -18.1). Texts of physical sciences (mean score=-5.0327) and biological & health sciences (mean= -4.1889) showed noticeable inclination towards non-

narrative concerns on dimension 2 (D2) and were placed closer to broadcasts (mean score=-3.3). On dimension 3 (D3), physical sciences (7.6543) and Biological & health sciences (mean=8.649) were positioned at positive polarity (as highly explicit texts) with significantly high scores as compared to other genres of Biber (1988), and were found closed to official documents (mean score=7.3). Physical sciences (mean score=-5.9537) and biological & health sciences (mean score=-6.2067), established highly non-persuasive and non-argumentative characteristics on dimension 4, again they were closed to broadcasts (mean score= -4.4) on this dimension. On dimension 5 (D5), both the groups (physical sciences; biological & health sciences) showed as moderately characterized with abstract features with the mean scores of 3.2096 and 3.4734 respectively, closer to official documents (mean score= 4.7).

The findings conclude that the text of PakDTh including both the sub-corpora in comparison to all the 23 genres studied by Biber (1988), significantly varies.

The findings of Intra-textual variation among the texts of PakDTh corpus indicated that both the major disciplinary groups of physical sciences and Biological sciences hold the resemblance and variation. Both groups are related to scientific academic discourse, but in spite of their mutual communicative and functional aims the texts of both groups vary significantly along dimension 1, 2 and 3 and bear a resemblance to each other on dimension 4 and 5. Unpredictably, the overall language of this register at major disciplinary level goes parallel with the same polarity direction on all the dimensions. The differences are only from perspective of higher and lower density on the above mention 3 (1,2, &3) dimensions.

Taken together, the findings of the study have revealed more strong and empirical evidences about the Pakistani doctoral theses written in English. According to Biber's (1989) text types and textual typology, based on the MD Analysis, the texts of Pakistani Ph.D. theses of both the groups appeared as characterized with highly '*informational exposition*', and resembles to text types of learned and scientific exposition.

Unlike other studies, the current research has also some limitations. The corpus of Pakistani doctoral theses (PakDTh) used in this study comprises of only two disciplines each represented by 100 number of texts. The addition of more disciplinary groups with more disciplines and sub-disciplines may give more insights into language of this register. Furthermore, the current study, employing old MD Analysis, is also limited

in terms of methodological updates. MAT tagger is used in this study as a tool due to its open and easy access, which only replicates Biber's old MDA rather than the new, updated and expanded MDA approaches (e.g., Biber, 2004; Xiao, 2009).

Concluding, all the above-mentioned limitations, the current study following MDA model for language variation research, may depict new prospects for the future researchers who could research language of undergraduate, masters, M.Phil., Ph.D. or doctoral theses from a multi-dimensional perspective. In addition, the current research may also be useful for the authors and publishers who write guidebooks of writing for researchers.

## References

- Ackermann, K., & Chen, Y. H. (2013). Developing the Academic Collocation List (ACL)—A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12(4), 235-247.
- Alemi, M., & Razzaghi, S. (2013). Politeness markers in English for business purposes textbook. *International Journal of Research Studies in Language Learning*, 2(4), 109-123.
- Ashtarian, S., & Weisi, H. (2016). A Comparative Study of the Use of Affective Markers in Supervision Comments: Case of Native versus Non-native Supervision Comments. *RELC Journal*, 47(2), 161-173.
- Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines*. McGraw-Hill Education (UK).
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers* (Vol. 23). John Benjamins Publishing.
- Biber, D. (1989). A typology of English texts. *Linguistics*, 27(1), pp. 3-44.
- Biber, D., & Finegan, E. (1994). Intra-textual variation within medical research articles. *Corpus-based research into language*, 12, 201-222.
- Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *tesol Quarterly*, 9-48.
- Biber, D., Conrad, S., Reppen, R., Byrd, P., Helt, M., Clark, V., ...& Urzua, A. (2004). *Representing Language Use in the University: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus*. Test of English as a Foreign Language.
- Boshier, R., & Pickard, L. (1979). Citation patterns of articles published in Adult Education 1968-1977. *Adult Education*, 30(1), 34-51.
- Budd, J. M. (1990). Higher education literature: Characteristics of citation patterns. *The Journal of Higher Education*, 61(1), 84-97.
- Bunton, D. (2002). Generic moves in PhD thesis introductions. *Academic discourse*, 57-75.
- Canagarajah, A. S. (2002). *A geopolitics of academic writing*. University of Pittsburgh 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.
- Carrió Pastor, M. L. (2014). Cross-cultural variation in the use of modal verbs in academic English. *Sky, Journal of Linguistics*, 27, 23-45.
- Carter-Thomas, S., & Rowley-Jolivet, E. (2008). If-conditionals in medical discourse: From theory to disciplinary practice. *Journal of English for Academic Purposes*, 7(3), 191-205.
- Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes*, 25(3), 310-331.
- Chejnová, P. (2014). Expressing politeness in the institutional e-mail communications of university students in the Czech Republic. *Journal of Pragmatics*, 60, 175-192.
- Chon, Y. V., & Shin, D. (2013). A Corpus-Driven Analysis of Spoken and Written Academic Collocations. *Multimedia-Assisted Language Learning*, 16(3), 11-38.
- Conrad, S. (2001). Variation among disciplinary texts: A comparison of textbooks and journal articles in biology and history. *Variation in English: Multi-dimensional studies*, 94-107.
- contradictory findings', *Language*, 62(2): 384-414.
- Cortes, V. (2007). Exploring genre and corpora in the English for academic writing class. *The ORTESOL Journal* 25, 8-14.

- Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X., & Zhai, C. (2014). Content-based citation analysis: The next generation of citation analysis. *Journal of the Association for Information Science and Technology*, 65(9), 1820-1833.
- Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, 28(3), 157-169.
- Ferguson, G. (2001). If you pop over there: a corpus-based study of conditionals in medical discourse. *English for Specific Purposes*, 20(1), 61-82.
- Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305-327.
- Gingras, Y., Lariviere, V., Macaluso, B., & Robitaille, J. P. (2008). The effects of aging on researchers' publication and citation patterns. *PloS one*, 3(12), e4048.
- Hardy, J. A. (2015). Multi-Dimensional Analysis of Academic Discourse. In *Corpora and Discourse Studies* (pp. 155-174). Palgrave Macmillan UK.
- Hartshorn, J., & Hart, J. (2016). Comparing the Academic Word List with the Academic Vocabulary List: Analyses of Frequency and Performance of English Language Learners. *JOURNAL OF LANGUAGE TEACHING AND LEARNING*, 6(2), 70-87.
- Harwood, N. (2005). 'We do not seem to have a theory... The theory I present here attempts to fill this gap': Inclusive and exclusive pronouns in academic writing. *Applied linguistics*, 26(3), 343-375.
- Higher education statistics agency.(2012). *JACS codes v3.0*. Retrieved from <https://www.hesa.ac.uk/support/documentation/jacs/jacs3-detailed>
- Hwang, J. K., Lawrence, J. F., Mo, E., & Snow, C. E. (2015). Differential effects of a systematic vocabulary intervention on adolescent language minority students with varying levels of English proficiency. *International Journal of Bilingualism*, 19(3), 314-332.
- Hyland, K., & Jiang, F. K. (2016). "We must conclude that...": A diachronic study of academic engagement. *Journal of English for Academic Purposes*, 24, 29-42.
- Hyland, K. (2006). *English for academic purposes: An advanced resource book*. Routledge.
- Hyland, K. (2013). ESP and Writing. *The Handbook of English for Specific Purposes*, Oxford, Wiley-Blackwell, 95-113.
- Kuo, C. H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for specific purposes*, 18(2), 121-138.
- Lestari, D. P., Syafrizal, S., & Elfrida, E. (2014). An analysis of politeness language patterns in request used in English textbook of second grade of junior high school (Doctoral dissertation, Universitas Bengkulu).
- Lillis, T. M. (2002). *Student writing: Access, regulation, desire*. Routledge.
- Louwerse, M. M., Crossley, S. A., & Jeuniaux, P. (2008). What if? Conditionals in educational registers. *Linguistics and Education*, 19(1), 56-69.
- Martínez, I. A. (2005). Native and non-native writers' use of first person pronouns in the different sections of biology research articles in English. *Journal of Second Language Writing*, 14(3), 174-190.
- Mauch, J., & Park, N. (2003). *Guide to the successful thesis and dissertation: A handbook for students and faculty* (Vol. 62). CRC Press.
- Moran, K. E. (2013). Exploring undergraduate disciplinary writing: Expectations and evidence in psychology and chemistry.
- Nasir, A., & Shakir, A. (2015). Use of Present Tense in Online University Prospectus. *Journal of Literature, Languages and Linguistics*, 9, 64-71.
- Nesi, H. (2008, September). A multidimensional analysis of student writing across levels and disciplines. In *Taking the Measure of Applied Linguistics: Proceedings of the BAAL Annual Conference, University of Swansea* (pp. 11-13).
- Nini, A. 2015. *Multidimensional Analysis Tagger* (Version 1.3). Available at: <https://sites.google.com/site/multidimensionaltagger/>

- Okamura, A. (2009). Use of personal pronouns in two types of monologic academic speech. *The Economic Journal of Takasaki City University of Economics*, 52(1), 17-26.
- Ordem, E., & Bada, E. (2016). Lexical collocations (verb+ noun) across written academic genres in English. *European Journal of Education Studies*.
- Parry, S. (1998). Disciplinary discourse in doctoral theses. *Higher Education*, 36(3), 273-299.
- Purandina, I. P. Y., Seken, I. K., Budasi, I. G., & Lin, D. A. (2014). Politeness in efl classroom interactions and its implications towards efl teaching-learning in SMP Negeri 2 Tabanan in academic year 2013/2014. *Jurnal Pendidikan Bahasa Inggris*, 2(1).
- Rahmani, E., Rahmany, R., & Sadeghi, B. (2014). Politeness strategies and politeness markers in email-request sent by Iranian efl learners to professors. *International Journal of Language Learning and Applied Linguistics World*, 144.
- 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.
- Sinclair, J. (2004). Corpus and text-basic principles. *Developing linguistic corpora: A guide to good practice*, 1-16.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Swales, J. M., Ahmad, U. K., Chang, Y. Y., Chavez, D., Dressen, D. F., & Seymour, R. (1998). Consider this: The role of imperatives in scholarly writing. *Applied linguistics*, 19(1), 97-121.
- Ventola, E., & Mauranen, A. (Eds.). (1996). *Academic writing: Intercultural and textual issues* (Vol. 41). John Benjamins Publishing.
- Warchał, K. (2010). Moulding interpersonal relations through conditional clauses: Consensus-building strategies in written academic discourse. *Journal of English for Academic Purposes*, 9(2), 140-150.
- Yeo, J. Y., & Ting, S. H. (2014). Personal pronouns for student engagement in arts and science lecture introductions. *English for Specific Purposes*, 34, 26-37.
- Zahra, F. T., & Shakir, A. (2015). A Corpus-Based Study of Explicit Versus Situation Dependent Reference in Online University Prospectus. *Research on Humanities and Social Sciences*, 5(7), 128-1