Linguistic Analysis of Pakistani Book Blurbs on New Textual Dimensions

Shahla Qasim[1]
Aleem Shakir[2]
Ahmad Bin Qasim[3]

Abstract

Register analysis is an emerging area in the field of applied linguistics, which is increasingly becoming a focus of attention of researchers and linguists the world over. Book blurb register has also attained singularity, specifically in context of rapidly expanding book marketing industry. Blurbs are short descriptions about the content and quality of the book along with endorsements by critics and prominent literary figures. Previous studies on book blurbs established their findings on individual linguistic features and the analysis based upon individual linguistic features is invariably found deficient in producing scientific and accurate results with respect to linguistic characterization of registers. Present study set out to analyze linguistic characterization of Pakistani book blurb register by applying a multidimensional (MD) approach which emphasizes the distributional patterns of co-occurring linguistic features and their shared communicative functions in register analysis. A standardized principled corpora of 1311 Pakistani book blurbs was built by using online and print sources. Sample data was compared on four linguistic dimensions, generated as a result of applying new factor solution matrix on the whole book blurb data. The four new linguistic dimensions identified were: dimension1, Abstract informational description versus concrete human focus; dimension 2, Interactive stance versus formal reportage of facts; dimension 3, Informational density versus elaborated expression; dimension 4, Expression of personal stance and judgment. The findings extend considerable contribution to the domain of register variation in general and Pakistani book blurbs register in particular. The new dimensions identified in present study may be exploited by upcoming research works to establish linguistic characterization of new Pakistani registers.

Keywords: *book blurb register, MD approach, new factor solution matrix*

Introduction

The status of book blurbs as a distinct register is well established. Blurbs illustrate the content of given books for their readers benefits and highlight some of the positive features for their authors' and publishers' benefits. Their appealing presentation of selected data is meant to elicit an appreciative response from potential book buyers (Gesuato, 2007, p. 84).

Book blurbs are extensively studied in recent times but a substantial volume of research is focused on exploring the generic and rhetorical structure of book blurbs in terms of move and steps (e.g., Bhatia, 2004; Gea-Valor, 2005; Onder, 2013). Analysis of linguistic features (an essential requirement of register analysis) is not executed at length. A couple of research works devoted to linguistic analysis (e.g., Basturkmen, 1999; Cacchiani, 2007; Marciulioniene, 2009), have established their findings on the study of individual linguistic features and as demonstrated by Biber (1988), analysis based on individual linguistic features does not accurately investigate the complex nature of textual relations across different registers. Therefore, the findings of these studies need to be tested quantitatively to exclude the element of idiosyncratic assumptions. Given the methodological inadequacies of previous linguistic studies, the present study on book blurbs establishes its findings on a multidimensional analytical approach, which explores the specific distributional patterns of co-occurring linguistic features in text and is, so far, considered to be the best empirical approach to explore linguistic variation of different registers. A multidimensional analysis of book blurb register is carried out to investigate the following research question:

Q. What are the distributional patterns of co-occuring linguistic features across Pakistani book blurb register on new linguistic dimensions?

Concept of Register Variation

Variation is inherent and all-pervasive in human languages. This variation may be of three types, “variation associated with constraints in the linguistic environment, variation associated with the social or demographic characteristics of speakers, and variation associated with situations of use” (Biber, 1995, p. 316). Register variation studies may take into account register features and register markers. Register features are core lexical and grammatical characteristics found to some extent in almost all the texts and registers in varying frequency, whereas register markers are specific and constrained to a target register (Biber, 1995, p. 29). The analysis of register features instead of register markers can reliably distinguish among two or more registers. Biber and Conrad (2009) stated the inevitability of the analysis of register features in these words, “Linguistic component of a register analysis requires identification of the pervasive linguistic features in the variety: linguistic characteristics that might occur in any variety but are much more common in the target register” (p. 6).

Register features are dominant indicators of register variation. They have a tendency of occurring frequently in the form of groups and considerable differences are found in their relative distributional patterns across registers. The divergence of distributional pattern of linguistic features depends upon the discourse function of features with respect to

diverse situational properties of each register. An empirical approach which employs quantitative techniques is crucially important to identify register features.

Register features of Book blurbs have not been explored so far, from quantitative perspective. Existing studies mainly depend upon qualitative interpretations of individual linguistic features informed by hypothetical assumptions.

Linguistic Studies on Book Blurbs

Following is given a brief account of research studies which approached language of book blurb text through analyzing individual linguistic features.

A significant study conducted by Cacchiani (2007) compared book reviews and book blurbs in terms of key words usage. The frequency of key words (superlatives and intensifiers) used in the selected corpus of book reviews (BR) and book blurbs (BB) was obtained by using Word Smith Tool. The lexical devices were found to be used in blurbs more dominantly than reviews to obtain semantic and pragmatic intensification. Cacchiani's study (2007) further concluded that book blurb text was a persuasive text with dominant use of adjectives (p. 11). Gea-Valor and Ros (2009) examined the use of evaluative language in blurbs which qualify the book in hand. Evaluative language is mostly realized through positive adjectives such as sensitive, powerful and astonishing. The study claimed that the use of syntactic structure (one of the most+ adjective) was reminiscent of advertising discourse (p. 212). The study did not use any statistical or empirical method to authenticate its findings. Likewise, Marciulioniene's study (2009) also affirmed that the language of blurb of English fiction was highly adjectival (p. 64). The study provided a detailed account of structural and linguistic changes occurred in blurb writing over a course of time, but the findings were not supported by any statistical analysis, rather depended only on rate of occurrence of linguistic features. Therefore, the results generalized only on the basis of frequency are not reliable and valid.

Yang (2013) investigated the use of key words in blurbs of academic text book blurbs across eight disciplines of hard and soft sciences. Yang's study relied on the analysis of individual linguistic features and provided implied hypothetical justifications of their prevailing or less prevailing existence in book blurb text without adopting any quantitative strategy to approach accurate and reliable findings. Onder (2013) examined the promotional elements in blurbs of Amazon UK and Okouko TR. The study revealed that the frequency of nouns and evaluative adjectives was closer to each other in both corpora. Frequencies of adverbs and verbs were found different. Verbs were the lowest frequency in Amazon and adverbs were the

lowest frequency in Okouko. Onder (2013), though, considered the issue of varying number of blurbs (Amazon, 40 and Okouko, 55) and applied a two tailed Z Test to find significant and non-significant differences, yet the issue of varying length of blurbs in both corpora (Amazon, 16,392 & Okouko, 10,145) was not taken into account, consequently the study faced validity threat.

Basturkmen (1999) examined key words in text book blurbs and obtained their frequencies in the sub corpus (Move 1 and 2). Basturkmen's (2009) study, though classified individual linguistic features under three categories to reach better decisions but even then the problem is not resolved and the required explicitness and accuracy could not be attained (possible only through using MD analysis approach). Kathpalia (1997) examined the difference in the use of lexico-grammatical features in the corpora of local blurbs (published in Singapur) and international blurbs. The study pointed out a clear difference with respect to the use of evaluative language. In international blurbs, evaluative language permeated throughout the blurbs like spreading waves. While local blurbs tended to confine the evaluative language to a focused move indicating the value of book. The contrastive analysis conducted by Kathpalia (1997) lacked numerical figures and empirical approach, essentially required for a quantitative analysis.

In Pakistani context, only one study conducted by Qasim and Shakir (2016) analyzed linguistic variation of 200 online blurbs of Pakistani fiction and non-fiction books on five linguistic dimensions based on Biber (1988) multidimensional analysis. A comparative analysis of the linguistic characterization of book blurbs and other promotional varieties (direct mail letters, non-profit grant proposals, print advertisements) led to the conclusion that contrary to the claims of previous studies, book blurbs did not show close resemblance with advertising text on dimensions 1, 2 and 3, but with respect to dimensions 4 and 5, they displayed mutual similarity. The results of this study are limited in the sense that were entirely based upon five dimensions of Biber (1988) MD analysis and the new factor solution matrix was not performed to generate new dimensions specific to a new group of texts in a different discourse domain.

New Multidimensional Studies

A review of previous studies has elucidated the fact that scant attention has been directed so far to explore the distributional patterns of co-occurring linguistic features of book blurb register. The current study set out to conduct a quantitative analysis of the linguistic characterization of a specific register of Pakistani book blurb by applying a new multidimensional approach. New multidimensional analysis is based upon new linguistic dimensions generated as a result of dealing with a new group of blurb texts

in a different discourse domain. It involves the application of a new factor solution matrix. A different corpus would certainly generate new sets of co-occurring linguistic features defining unique dimensions indicative of specific shared communicative functions. A few new dimensions were identified through a new factor analysis employed by Biber (2006), D3: Procedural versus Content Focused Discourse and D4: Academic Stance. Biber (2008) discovered a new dimension labeled as, Stance Focused versus Content Focused Discourse. The study conducted by Shakir (2013) explored three new dimensions in Pakistani corpus of advertising text. These are, D1: Oral versus Literate, D2: Expression of Organizational Policy versus Other Concerns, & D3: Impersonal versus Audience Centered Style. Hardy and Romer (2013) explored four new dimensions in the Michigan Corpus of upper level students. These are, D 1: Involved Academic Narrative versus Descriptive Academic Discourse, D2: Expression of Opinions and Mental Processes, D3: Situation-Dependent, Non-Procedural Evaluation versus Procedural Discourse and D4: Production of Possibility Statement and Argumentation.

Gray (2011) explored linguistic variation of research articles across six disciplines and identified four new dimensions: D1: Academic Involvement and Elaboration versus Informational Density, D2: Contextualized Narration versus Procedural Discourse, D3: Human versus Non-Human Focus, D4: Academese. Egbert (2015) discovered five new dimensions based on the multidimensional analysis of academic writing. The five dimensions were labelled as, D1: Affective Synthesis versus Specialized Information Density, D2: Definition and Evaluation of New Concepts, D3: Author-Centered Stance, D 4: Colloquial Narrative and Dimension 5: Abstract Observation and Description. Getkham (2013) examined the distributional patterns of linguistic features across different sections of research articles of applied Linguistics by using New MD analytical approach and identified six dimensions which are labelled as follows, D1: Established Knowledge/Expression of Ownership, D2: Expression of Purposes, D3: Evaluative Stance, D4: Expression of Generality, D5: Framing Claims, and Dimension 6: Conceptual complexity.

MD Approach and its Suitability for Present Study

The current study on linguistic variation of book blurbs subsumes to multidimensional approach (Biber, 1988) as its methodological framework. The logic underlying this decision is that no other approach, so far, is available which possesses such an ample capacity to explore quantitatively the linguistic similarities and differences across various registers. The central importance is given to the notion of linguistic co-occurrence patterns and the empirical methods of their identification.The idea of co-occurrence of linguistic features is not a newly perceived idea. It has already

been acknowledged and employed by a number of earlier researchers and linguists (Brown and Fraser, 1979; Chafe, 1982; Chafe and Danielewicz, 1987; Ervin -Tripp, 1972; Hymes, 1974). Halliday (1988) emphasized the essential presence of co-occurring linguistic features in a register and defined register as a "cluster of associated features having a greater- than-random . . . tendency to co-occur (p. 162).

Multidimensional approach developed by Biber (1988) is highly appropriate to distinguish among registers because "in this approach co-occurrence patterns are identified quantitatively, based on the actual distribution of linguistic features in a large corpus of text" (Biber *et al.*, 1998, p. 145). Therefore, the studies based on the examination of individual features and the conclusions derived from them may be considered just subjective perceptions and hypothetical speculations.

New MD analysis (used in current study) is based upon new linguistic dimensions generated as a result of dealing with a different discourse domain. It involves a new statistical analysis which is incorporated to encompass the linguistic variations found in a specific corpus yet unexplored. A different corpus would certainly generate new sets of co-occurring linguistic features defining unique dimensions indicative of specific shared communicative functions. A few new dimensions were identified through a new factor analysis employed by Biber (2006, 2008), Getkham (2010), Shakir (2013), Hardy and Romer (2013), Gray (2011), Egbert (2015).

A review of previous studies conducted in the fashion of Old and New multidimensional analytical approach has clearly elucidated the fact that not a single study has been carried out so far to explore the co-occurring distributional patterns of linguistic features of book blurb register which possesses a remarkable importance in today's ever expanding world of book marketing. The current study set out to conduct a quantitative analysis, following New MD approach in order to unfold the linguistic variation of the specific register of Pakistani book blurb text.

Collection of Data and Corpus Compilation

Present study covered a comprehensive data of Pakistani book blurbs, provided by local as well as foreign publishing houses, available in print and online form. The corpus comprised a total of 1311 sample book blurb texts containing 0.4 million words. Based on the insights drawn from previous concerned literature, the book blurb data was categorized with respect to three independent situational variables; 1) publisher location - to find out linguistic variation across Pakistani blurbs provided by foreign publishers and local publishers, 2) publication medium - to explore linguistic variation between traditional print book blurbs and online blurbs, 3) literary

source - to explore linguistic variation between blurbs written for fiction books and blurbs written for non-fiction books. However, the results of present study deal with the book blurb corpus on the whole.

Data was collected in two phases. In the first phase, online book blurbs for books written by Pakistani authors were collected from the websites of different local and foreign publishing houses. A java program was also written to extract blurb description and other important required information about the title of book, name of author, publication year and the name of publishing house. Print book blurbs for Pakistani books were collected from the libraries of different public institutions in the form of camera images. For conversion of images to text files, the third party GUI flavors of Google Tesseract namely Sunny Page OCR and Tesseract4java were tried, but they did not work effectively to extract text from images. Instead, Text Fairy (a flavor of Google Tesseract tools for Android) worked comparatively well.

All collected print and online book blurbs were edited and scrutinized. Out of more than 2000 book blurbs, 1311 were finally selected for analysis. To assign codes to the categories, a Java script was written. Methods implemented within Java String Class were used for text manipulation of blurb data. The Java script contained the details of the codes to be assigned to the main categories, sub-categories and sub-sub categories. The Java script parsed and analyzed the blurb data and extracted the initial category code, the author name, the book title, the publisher detail and year of publication. A final category code was created from this information. Lastly, the program renamed the blurb files to respective category codes.

Data Analysis

Multidimensional analysis of the blurb corpus was based on new factor solution matrix. Methodological process involved tagging of linguistic features, acquiring raw frequencies of linguistic features, turning raw frequencies into normalized frequencies, standardization of normalized frequencies, secree plot, factor analysis, computation of dimension scores and finally, identification of new linguistic dimensions and their labelling corresponding functional interpretation. Data was sent to Jessy Egbert, Brigham Young University for complete tagging and analysis. Sequential progression of analytical steps were as follows.

Step 1 Tagging of Linguistic Features

First important step was the tagging of data for all linguistic features. A grammatical tagger (developed by Biber, 1993) was used to identify the grammatical category of each linguistic feature in all book blurb texts included in study.

Step 2 Computation of Normalized and Standardized Frequency Scores

After the tagging process of the linguistic features in sample texts was done, the raw frequencies of all linguistic feature were computed. The obtained raw frequency scores were normalized to a text length of 1,000 words to avoid skewing effect in case of varying length of texts included in analysis. Normalization of raw frequencies was obtained through this formula: Raw frequency of linguistic feature divided by total number of words in text, multiplied by the sum 1000. Biber (1988) recommended the process of normalization that it is very important for having accurate results included in a study unaffected by various length of texts.

Standardization of normalized frequencies was another important step to be carried out. The normalized frequencies were standardized to the mean of 0.0 and a standard deviation of 1.0 (as practised by studies based on 1988 MD analysis). Formula devised for this purpose was to subtract the individual normalized scores from the mean normalized scores and the resulting scores to be divided by the standard deviation. The logicjustifying this process of standardization was to ensure that all linguistic features on dimension had equal weightage in the calculation of dimension scores.

Step 3 Scree Plot

In order to determine the optimal number of factors to be extracted through New factor solution matrix, a Scree plot of Eigen values was examined. It indicated percentage of maximum shared variance extracted by each factor.

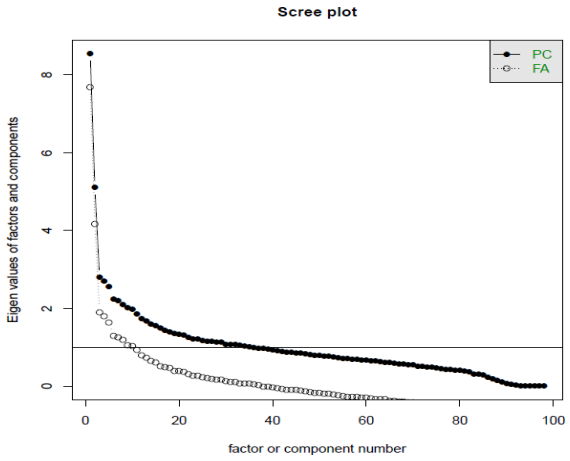

Figure 1: Scree Plot

Biber (1988) demonstrated that "Scree plot is a plotof Eigen values which are direct indices of the amount of variance accounted for each factor" (p. 82). According to the scree plot, factor 1 accounted for maximum amount of variance i.e., 7.8%, factor 2, for 4.1%, factor 3 for an additional

1.9%, etc. The scree plot shows a sharp break between factors 5 and 6, indicating a point at which additional factors would have a minimum loading and consequently would not contribute significantly to overall analysis. So the first 5 factors, which extracted the relatively larger groupings of co-occurring linguistic features, were considered at initial stage. Later on fifth factor was not included in interpretation because the feature loading of fifth factor did not commensurate with the determined cut-off point. The rest of the factors had little significance because they accounted for little amount of shared variance.

Step 4 New Factor Solution Matrix

Factor analysis is a statistical tool of MD approach which is basically used to reduce large number of original variables to a reduced number of derived variables. In current study, factor analysis was used to reduce large number of frequencies of linguistic features to small sets of factors. Each factor represented a set of linguistic features which co-occurred with a high frequency in text. The co-occurring linguistic features may be of two types: positive features and negative features. Both types of linguistic features are in complementary relationship with each other. It means that if positive features co-occur frequently in a target text, there will be relatively low occurrence of negative features and vice versa.

The new factor solution was based upon 87 linguistic features which met minimum requirements for frequency and variance across texts. These 87 features were selected from a list of 147 linguistic features. Many of the 147 features were overlapping categories, so only those features were included which did not overlap with any others. The selected features not only included the 88 MD analysis features but also the features added to the program later.

The factor analysis performed in this study used the statistical program R (R Development). A principal factor analysis was run because of the exploratory nature of study. This was done by using the R function 'fa' (factor analysis) within the "psych" library, with "fm" (factor method) set to "pa" (principal axis).

Table 1: Factor Solution for the New MD Analysis of Book Blurbs

Linguistic Features	**D1**	**D2**	**D3**	**D4**	**D5**
Contractions	-0.15	**0.25**	-0.05	-0.07	-0.13
vb_present	0.04	**0.78**	-0.05	0.01	0.01
pro_2	-0.17	**0.3**	-0.02	-0.04	-0.1
pro_1	-0.23	**0.3**	-0.01	-0.02	-0.05

pro_it	0.27	0.11	**0.3**	0	-0.03
mod_poss	0	**0.52**	-0.01	-0.09	0.05
coord_conj_cls	-0.02	-0.06	0.07	0.03	**-0.28**
nn_all	-0.14	-0.01	**-0.94**	-0.08	-0.13
Prep	0.1	**-0.48**	0.1	-0.05	0.13
jj_attr	**0.68**	-0.08	0.07	-0.07	-0.13
vb_past	-0.43	**-0.46**	0.31	0.1	0.16
pro_3	**-0.71**	-0.09	-0.01	-0.02	-0.01
wh_rel_subj	0	0.06	**0.33**	-0.06	0.12
coord_conj_phrs	**0.32**	0.08	0.11	-0.07	-0.1
nn_nom	0.34	-0.16	**-0.44**	0.09	0.03
Infinitive	0.05	-0.01	0.18	**0.55**	-0.02
mod_pred	0.01	**0.51**	0	-0.06	0.01
mod_necess	-0.02	**0.34**	0.01	-0.09	0.01
split_aux	-0.07	0.16	-0.05	-0.09	**0.26**
th_vb	-0.07	**0.31**	-0.07	0.21	-0.02
mod_all	-0.01	**0.81**	-0.03	-0.15	0.05
conj_all	0.11	0.03	0.21	0.02	**-0.28**
passive_all	-0.06	0.04	0.16	-0.07	**0.86**
wh_rel_all	0.01	0.03	**0.44**	-0.09	0.11
jj_all	0.66	0.01	0.18	-0.07	-0.22
pro_all	**-0.78**	0.1	-0.03	-0.04	-0.06
vb_all	-0.27	0.31	0.31	**0.32**	0.19
jj_att_other	0.11	0.4	**0.12**	-0.07	-0.05
all_def_art	0.26	**-0.28**	0.4	0.02	-0.02
nn_common	0.17	0.22	**-0.38**	-0.06	0.02
nn_proper	**-0.49**	-0.18	-0.43	-0.12	-0.18
nn_premod	0.07	0.1	**-0.65**	0.02	0.1
th_vb_stance_all	-0.11	0.29	-0.13	**0.58**	-0.06
th_stange_all	-0.08	0.29	-0.1	**0.56**	-0.04
to_vb_stance_all	0.05	-0.15	0.05	**0.87**	-0.06

to_stance_all	0.08	-0.19	0.08	**0.87**	-0.07
nn_human	**-0.33**	0.07	0.01	-0.09	-0.04
nn_process	**0.38**	0.04	-0.17	0.1	0.12
nn_cog	**0.25**	0.14	0.03	0.04	0.06
nn_abstact	**0.42**	0.11	0.07	0.07	0.03
jj_topic	**0.44**	-0.1	0.05	0.06	-0.03
vb_mental	-0.06	**0.38**	0.05	0.26	-0.03
word_length	0.53	-0.04	**-0.52**	0.08	0.01

The co-occurrence patterns of linguistic features on first four dimensions were taken into consideration. Fifth dimension, due to inadequate loading of less than five linguistic features did not meet the criteria, hence not interpretable.

Step 5 Computation of Dimension Score

The next step was the computation of dimension score for each text. Dimension scores were computed by subtracting sum of standardized scores of negative features from the sum of standardized score of positive features. The standardized scores of positive features were taken as dimension scores where negative scores did not occur at all.

The criteria for the inclusion of a linguistic feature in the computation process was settled with the help of previous studies. The features having larger loadings (indicating maximum occurrence of the linguistic feature) was retained, whereas the features having lower weight/ loading, were not included in computation process. The minimum loading of any linguistic feature in 88 MD analysis was settled at +/- 35, whereas in the new MD analysis the ratio of minimum loading was reduced to +/- 30 (following Shakir, 2013). It means that any linguistic feature either positive or negative having lesser than 30 weight on any dimension was not included. The minimum loading was referred to as cut-off point or salient loadings. In case a feature had salient loadings on more than one dimension/ factor, it was included in the factor score of that factor on which it had the highest loading (Biber, 1988, p. 93).

Results and Discussion

This section presents the results of multidimensional analysis of book blurbs based on new factor solution. New factor solution matrix is performed on 87 features, which meet minimum requirements for frequency and variance across texts. Based on the co-occurrence patterns, merely 44 of those features were retained (i.e., they had a factor loading that reached the minimum cutoff of .30). The section also describes the

distributional pattern of the sets of co-occurring linguistic features grouped on new factors identified as new dimensions. It offers functional interpretation of these groups of co-occurring features that what specific shared functions they have to perform collectively, which cause their existence on one dimension. The dimensions are labelled in accordance with the shared function of the group of features placed on it. Text evidences are presented with positive features bolded and negative features underlined. Four dimensions of New Factor Analysis are labelled as:

D 1- Abstract Informational Description versus Concrete Human Focus

D2- Interactive Stance versus Formal Reportage of Facts

D3- Informational Density versus Elaborated Expression

D4- Expression of Personal Stance and Judgment

Explanation of Factors

Factor 1:
Abstract Informational Description versus Concrete Human Focus

Co-occurring distributional pattern of linguistic features on dimension 1, Abstract Informational Description versus Concrete Human Focus, is given below.

Table 2: List of Linguistic Features on First Dimension of New Factor Analysis of Book Blurbs

Positive Linguistic Features	Scores	Negative Linguistic Features	Scores
Attributive Adjectives	0.68	Human Nouns	-0.33
All Adjectives	0.66	Proper Nouns	-0.49
Word_length	0.53	3rd Person pronoun	-0.71
Topical Adjectives	0.44	All Personal Pronouns	-0.78
Abstract Nouns	0.42		
Process Nouns	0.38		
coord_conj_phrs	0.32		
Cognitive Nouns	0.25		

Factor 1 shows the co-occurrence of 8 positive and 4 negative features. Positive features comprise attributive adjective, all adjective, topical adjectives, word length, abstract nouns, process nouns and cognitive nouns. The negative features on factor 1 include human nouns, proper nouns, third person nouns, and all personal pronouns. The highest load on positive polarity is achieved by attributive adjectives (0.68) and all adjective (0.66). Attributive adjective are usually associated with nouns and their major function is to modify nouns. They are used to qualify both animate

and inanimate nouns to denote informational density. Attributive adjectives accompanied with abstract noun (0.42) and word length (0.53), formulate an exclusive combination which shows integration of abstract informational description. Another important linguistic feature co-occurring with this strong positive group is topical adjective (0.44), which performs the same function of attributive adjectives. "Attributive adjectives are a more integrated form of nominal elaboration than predicative adjectives or relative clauses, since they pack information into relatively few words and structures" (Biber, 1988, p. 105).

Following excerpt from a local print non-fiction academic blurb text indicateS the existence of positive and negative features. Positive linguistic features are focused to perform the function of imparting abstract information about the contents of a book on medicine and how it would prove beneficial to meet the requirements of postgraduate examination. Additional information regarding author's qualification and expertise in the specific field of Cardiology has also been described.

Example 1

LPNASSMed.Ahmad.2013.791.

Cardiology **Ward Rounds** is an **open source** of Cardiology for **young minds** to explore, learn **and** acquire **knowledge** in order to meet the **needs** of **postgraduate examination**. This **comprehensive anthology** of **cardiovascular medicines** simplifies **diagnosis and** management in the form of **bullet** or numbered **lists**, **tables and acronyms** to make it **easy**-to-remember. Dr. Shafique Ahmed is a **graduate** of Quaid-e-Azam Medical College**,** Bahawalpur. He is the associate **professor and** the **head of department** Cardiology at the Bahawal Victoria **Hospital** Bahawalpur. With an FCPS **degree** in Cardiology awarded by the **College of Physicians and Surgeons,** Pakistan and **advanced training** in Angiography/Angioplasty, Echocardiography and **cardiac pacemakers**, he is also **examiner** of FCPS**-II** Cardiology **and** a **contributor** to **renowned journals** of Cardiology.

The highest negative weight on factor 1 is attained by all personal pronouns (0.78) and third person pronoun (0.71). Use of all personal pronouns reflect a substantial focus on the evaluators' remarks regarding book and author. Use of third person pronouns refer to the individuals not in immediate contact. Third person pronouns are generally used with past tense and perfect aspect verbs to reflect a narrative discourse. But here on the negative polarity of factor 1, the use of third person pronoun combined with proper nouns (with a weight – 0.49) and human noun (with a weight -0.33) indicate a concern for concrete descriptions of human beings or characters outside of immediate interaction.

Following excerpt taken from an online fiction blurb by foreign publisher provides evidence of the incorporation of negative linguistic features to outline briefly the major moves of leading human characters.

Example 2

FOFNOV.Hanif.2008.181

Intrigue and subterfuge combine with **bad luck and good** in this **darkly comic debut** about **love,** betrayal, **tyranny**, **family**, **and** a conspiracy trying its damnedest to happen. Ali Shigri, Pakistan **Air Force pilot** and **Silent Drill Commander** of the Fury Squadron, is on a **mission** to avenge his father's **suspicious death**, which the **government** calls a **suicide**. Ali's target is none other than General Zia-ul-Haq, **dictator** of Pakistan. Enlisting a **rag-tag group of conspirators**, including his cologne-bathed roommate, a hash-smoking **American lieutenant**, and a **mango-besotted crow**, Ali sets his **elaborate plan** in motion. There's only one problem: the line of would-be Zia **assassins** is longer than he could have possibly known.

In view of shared communicative function performed by the complementary grouping on this factor, the dimension is labeled as "Abstract Informational Description versus Concrete Human Focus."

Factor 2:
Interactive Stance versus Formal Reportage of Facts

Table 3: List of Linguistic Features on Second Dimension of Book Blurbs

Positive Linguistic Features	Scores	Negative Linguistic Features	Scores
All modals	0.81	Past Tense Verbs	-0.46
Present Tense Verb	0.78	Prepositions	-0.48
Possessive Modals	0.52		
Prediction Modals	0.51		
Attitudinal Adjectives in other Contexts	0.4		
Mental Verbs	0.38		
Necessity Modals	0.34		
"That" complement clause controlled by verbs	0.31		
2nd Person Pronoun	0.30		
1st Person Pronoun	0.30		
Contractions	0.25		

Factor 2 comprises 11 positive and 2 negative features. The number of substantial loading of positive features (8) on dimension 2 is equal to the number of positive features assembled on dimension 1. These features,

hierarchically organized in terms of weight, are all modals (0.81), present tense verbs (0.78), possibility modals (0.52), and prediction modals (0.52), attitudinal adjectives in other contexts (0.4), mental verbs (0.38), and necessity modals (0.34) "that" complement clause controlled by verb (0.31). Linguistic features having weight equal to cut-off point are 1st person pronoun (0.30), second person pronoun (0.30), and a little lesser weight than cut-off point is of contractions (0.25). The features on negative polarity are prepositional phrases and past tense verbs. Prepositions have higher negative weight (-0.48) than past tense (-0.46).

Positive linguistic features on this dimension reflect an involved and interactional expression. High concentration of all modals (0.81) is especially marked with an interpersonal focus. Possibility modals (can, could, may, might) are incorporated to pronounce permission and slight probability that certain event will occur. Prediction modals (will, shall, would) are used to prophesy in advance that certain event will or will not occur in future. They demonstrate possible future consequences. Necessity modals (ought, must, should) indicate obligation or necessity of a specific action. Cluster of modals generally function to present multiple possible actions and arguments for or against an option and finally derive toward the most suitable one to convince the readers.

Along with modals, dimension 2 is characterized by heavy reliance on present tense forms (0.78). Present tense verb count in Biber's tag count includes count for imperatives, uninflected present tense and third person singular verb. Present tense generally deals with the action and events of immediate situation. Oches (1979) associated present tense with unplanned speech style. Weber (1985) explained that cognitive verbs which describe the mental process of the speaker usually occur in present tense. Mental verb (0.38) are frequently used to present stance expression, the inner feelings and attitude of the writer. First person pronoun and second person pronoun, though occur with lesser weight, tend to co-occur with interactive features to indicate a less formal, involved style used mostly to share personal experience. First person pronouns are specifically used to 'bring the author, the reader, and often mankind in general into discourse, creating a sense of interaction and relating points and ideas to the readers as members of mankind' (Gray, 2011, p. 160). Overall, the positive feature on dimension 2 reflect an affective, involved style.

Following text evidence which serves as Example 3 is extracted from a foreign online fiction (Children book) blurb. Positive features are in boldface and negative features are underlined. Positive features are utilized to emphasize interpersonal stance which indicates an informal and involved style.

Example 3

FOFCB.Khan. 2005.382

"Ami **loves** her chicken better than **me**. She **calls** her Bibi**, I call** her silly." Rani's mother **loves** Bibi the chicken more than her. At least **that's** what Rani **thinks**. That silly chicken **gets** all the attention, and Rani just **can't** stand it. Even worse, Bibi **seems** to know **she's** the favorite! But when Bibi **disappears** one afternoon, Rani **realizes** how sad her mother **is**. **Will** Rani's jealousy disappear, too? Set in **rural Pakistan** and illustrated with **lively**, **expressive illustrations**, this **original take on sibling rivalry is hilarious** and **poignant** at the same time.

The negative feature, past tense form and prepositional phrases reflect an expression which simply depicts reportage of facts and specifics. Past tense forms are primary surface markers of narrative "removed situation." Complementary groups of linguistic features situated on this dimension may lead to the interpretation of this dimension as Interactive Stance versus Formal Reportage of Facts.

Following blurb text (Example 4) excerpt is focused to report details of the notable literary services and distinguished career achievements of a former foreign minister of Pakistan and the author of a book on history. Negative features are used more frequently than positive features to describe past events associated with the life of the author.

Example 4

LONAHHis.Sattar.2013.373.

With a thirty-nine-year **career span** in the Ministry of **Foreign Affairs**, Abdul Sattar was twice Pakistan's **Foreign Minister**, from July to October 1993, and from 1999 to 2002. He was **Foreign Secretary** from 1986 to 1988 and twice Pakistan's Ambassador to India. He also served as Ambassador to the USSR and **Permanent Representative** to the IAEA in Vienna. As a **Distinguished Fellow** at the United States Institute of Peace, he wrote a research paper 'Reducing Nuclear Dangers in South Asia' which was published in the Nonproliferation Review in 1994, and later in Dawn. His other research paper, 'Shimla Pact: Negotiating under Duress', was published in journals in Islamabad and New Delhi in 1995. He also contributed the section on foreign policy in the book Pakistan in Perspective, 1947-1997, published by Oxford University Press on the fiftieth anniversary of Pakistan.

Factor 3:

Informational Density versus Elaborated Expression

The following table provides details regarding Factor 3:

Table 4: List of Linguistic Features on Third Dimension of Book blurbs

Positive Linguistic Features	Scores	NegativeLinguistic Features	Scores
All Nouns	0.94	Pronoun It	-0.3
Premodifying Nouns	0.65	Wh pronoun relative clause subject position	-0.33
Normalizations	0.44	all_def_art	-0.4
Common Nouns	0.38	All wh relative Clauses	-0.44

Positive features on factor 3 show a high preponderance of nouns. All nouns (.94), pre-modifying nouns (.65), normalizations (.44) and common nouns (.38) co-occur frequently on this dimension reflecting informational density. Maximum weight of all nouns (0.94, the highestever on any dimension) lead to the assessment of the blurb text in nominal terms. It is possible to elaborate noun phrases through both "pre-modifiers" (which come before the head noun) and "post-modifiers" (which come after the head noun). There are two important kinds of pre-modifiers in English; Attributive adjectives (for example, specific event, major project) and Nouns as Pre-modifiers (the traffic jam, the committee recommendation). Nominalizations (words ending in tion, ment, ness and ity) are used to condense high amount of information in a fewer words. Use of nominalizations reflect a compact and lucid style. They generally occur with passive constructs and prepositions to impart dense informational material.

Following blurb extract has been taken from a foreign print non-fiction academic book of social science. The blurb describes contextual information regarding recent international fluctuations which have transformed the overall world scenario. The information is densely packed by using positive linguistic features such as pre-modifiers and normalizations. Positive features are bolded.

Example 5

FPNASSPS.Rizvi.1993.282.

Momentous changes have taken place in the **international system** in **recent years**. The end of the **Cold War**, the **disintegration** of the erstwhile **Soviet Union**, and the **collapse** of the **socialist bloc** have radically altered the **familiar contours** of the **post-War bipolar international order**. These and **related developments** will have a **profound impact** on **global security alignments, regional conflicts** and on the **patterns** of **international economic linkages** in the **years** to come.

The **central purpose** of this **timely book** is to examine the **consequences** of these **developments** for **South Asia**. One of its **distinguishing features** is that the **issues** and **problems** are examined from the **viewpoint** of the **region** itself as a **corrective** to the **dominant discourse**

which tends to reflect only the **perceptions** and **preoccupations** of **Western scholars**.

The negative feature, bearing the highest weight on dimension 3 is indicated as all 'wh' relative clauses (-0.44). 'Wh' pronoun as a relative clause on subject position (-o.33) has weight slightly larger than cut-off point. Wh relative clauses are generally used for structural elaboration and explicit reference in planned discourse. They are more typical in written discourse. Other linguistic feature on negative polarity of this dimension with lesser weight include Pronoun 'it', which is the most generalized pronoun used to stand for both animate beings and abstract concepts. The following example text is taken from a foreign online fiction (Novel) blurb, showing negative features underlined.

Example 6

FOFNOV.Ali.2006.187.

The fourth novel in **Tariq Ali's** 'Islam Quintet' charts the life and loves of the medieval cartographer **Muhammed al-Idrisi**. Torn between his close **friendship** with the **Sultan** and his **friends** who are leaving the **island** or plotting a **resistance** to **Norman rule**, **Idrisi** finds **temporary solace** in the harem; but his **conscience** is troubled . . . A Sultan in Palermo is a **mythic novel** in which pride, greed, and lust intermingle with **resistance** and **greatness**. It echoes a **past** that can still be heard today.

Given the co-occurrence pattern of positive and negative features, the interpretive label "Informational Density versus Elaborated Expression" is suggested for dimension 3.

Factor 4

Expression of Personal Stance and Judgment

Table 5: List of Linguistic Features on Fourth Dimension of Book Blurbs

Positive Linguistic Features	**Scores**
to_vb_stance_all	0.87
to_stance_all	0.87
th_vb_stance_all	0.58
th_stange_all	0.56
Infinitive	0.55
vb_all	0.32

The interpretation of Factor 4 is relatively direct and straightforward, because it comprises only positive features. Negative features do not appear on this dimension. Linguistic feature with heaviest loading is indicated as 'To complement clause controlled by stance verbs (0.87). Other equally strong representative features include 'sum stance 'to'

complement clause (0.87). Features with comparatively lesser weights are 'sum stance 'to' complement clause controlled by verbs (0.58), 'sum stance 'that' complement clauses (0.56), infinitives (0.55) and all verbs (0.32, not including auxiliary verbs). All these features are directed towards expressing personal stance and judgment about a specific individual or entity. Heavy representative co-occurrence of stance features in book blurb text characterizes it as a discourse full of personal feelings and expressions.

The following blurb excerpt contains a number of instances indicating stance expressions:

Example 7

FOFNOV.Abdullah.2009.317.

"Saffron Dreams" is a tale of love, tragedy, and redemption from the award-winning author of "Beyond the Cayenne Wall." From the darkest hour of American history emerges a mesmerizing tale of tender love, a life interrupted, and faith recovered. ArissaIllahi, a Muslim artist and writer, discovers in a single moment that no matter how carefully you map your life, it is life itself that chooses your destiny. After her husband's death in the collapse of the World Trade Center, the discovery of his manuscript marks Arissa's reconnection to life. Her unborn son and the unfinished novel fuse in her mind into one life-defining project that becomes, at once, the struggle for her emotional survival and the redemption of her race. "Saffron Dreams" is a novel about our ever evolving identities and the events and places that shape them. It reminds us that in the midst of tragedy, our dreams can become a lasting legacy.

In view of shared functions performed by the distributional pattern of co-occurring features on dimension 4, the label "Expression of personal stance and judgment" is used.

Factor 5 does not fulfil the criteria required for interpretation of any factor (five independent features). Factor 5 appears with only four features, (all passives, split auxiliary, coordinating conjunction clauses and all conjunctions). Although the loading of all passives (0.86) is much higher than the cut-off point (+/-30) but loading of rest of the three features is lower. Therefore, factor 5 is found to be relatively uninterpretable.

Conclusion

The study identified linguistic characterization of Pakistani book blurb register on textual dimensions. A multidimensional analytical approach was used to mark distributional patterns of co-occurring linguistic features employed to perform shared communicative functions. Four unique interpretable dimensions were generated resulting from applying a new factor soluion matrix to a comprehensive blurb corpus. The study

concluded that Pakistani book blurb register possesses a distinct and unique linguistic structure diverse from registers of other domains.

Factor 1 showed the co-occurrence of 8 positive and 4 negative features. Positive features comprised attributive adjective, all adjective, topical adjectives, word length, abstract nouns, process nouns and cognitive nouns whereas the negative features on factor 1 included human nouns, proper nouns, third person nouns, and all personal pronouns. In view of shared communicative function performed by the complementary grouping on this factor, the dimension was labeled as "Abstract Informational Description versus Concrete Human Focus." Factor 2 comprised 11 positive and 2 negative features. Positive linguistic features on this dimension reflected an involved and interactional expression. The negative feature, past tense form and prepositional phrases reflected an expression used to depict reportage of facts and specifics. Complementary groups of linguistic features situated on this dimension led to the interpretation of this dimension as "Interactive Stance versus Formal Reportage of Facts." Heavy loading of "nouns" as co-occurring positive features on factor 3 reflected an overall trend of informational density, whereas the high frequency of "all wh relative clauses" as negative features on factor 3 echoed an elaborated style of expression. Given the co-occurrence pattern of positive and negative features, the interpretive label "Informational Density versus Elaborated expression" was suggested for dimension 3. The interpretation of Factor 4 was relatively direct and straightforward, because it comprised only positive features. Negative features did not appear on this dimension. Heavy representative co-occurrence of stance features in book blurb text characterized it as a discourse full of personal feelings and expressions.

Factor 5 was not interpretable because it did not fulfil the criteria required for interpretation of any factor (five independent features). Factor 5 appeared with only four features, (all passives, split auxiliary, coordinating conjunction clauses and all conjunctions), therefore excluded.

The findings of present study may prove to be helpful for the researchers working in the field of register analysis in general and book blurbs in particular. The findings may also provide substantial insight to the researchers who want to apply multidimensional analytical approach to characterize other registers in linguistic terms. The new dimensions generated in the present study may be exploited to establish linguistic characterization of other similar registers.

The book blurb corpus built by this study can be maneuvered to examine diachronic evolution of blurbs in different decades in order to cater for the changing demands of readership and the ever expanding book

industry. The results of this study may be utilized and compared with other multidimensional studies in native and non-native contexts.

References

Basturkmen, H. (1999). A content analysis of ELT textbook blurbs: Reflections of theory-in-use. *RELC Journal*, *30*(18), 17-38.

Basturkmen, H. (2009). Back cover blurbs: Puff pieces and windows on cultural values. In K. Hyland & G. Diani (Eds.), *Academic evaluation: Review genres in university settings* (pp. 68-83). Basingstoke: Palgrave.

Bhatia, V. K. (2004). *World of written discourse*. London. Continuum.

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

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

Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes, 5*, 97-116.

Biber, D. (2008). Corpus-based analysis of discourse. Dimension of variation in conversation. In V. Bhatia, J. Flowerdew & R. Jones (Eds.), *Advances in discourse studies* (pp. 100-114). London: Routledge.

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

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use.* Cambridge: Cambridge University Press.

Brown, P., & Fraser, C., (1979). Speech as a marker of situation. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech* (pp. 33-62). Cambridge: Cambridge University Press.

Cacchiani, S. (2007). From narratives to intensification and hyperbole: Promotional uses of book blurbs. *Proceedings of the Corpus Linguistics Conference* URL: http://ucrel.lancs

Chafe, W. L. (1982). Integration and involvement in speaking, writing and oral literature. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 35-54). Norwood, NJ: Abex.

Chafe, W. L., & Danielewicz, J. (1987). Properties of spoken and written language. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending oral, and written language* (pp. 83-88). New York, NY: Academic Press.

Egbert, J. (2015). Publication type and discipline variation in published academic writing: Investigating statistical interaction in corpus data. *International Journal of Corpus Linguistics*, *20*, 1-29.

Ervin-Tripp, S. (1972). On sociolinguistic rules: Alteration and co-occurrence. In J. J. Gumperz & D. Hymes (Eds.), *Directions in Sociolinguistics* (pp. 213-250). New York, NY: Holt.

Gea-Valor, M. L. (2005). Advertising books: A linguistic analysis of blurbs. *Ibérica, 10*, 41-62.

Gea Valor, M. L., & InigoRos, M. (2009). On the dynamic nature of genre: A diachronic study of blurbs. In K. Hyland & G. Diani (Eds.), *Academic evaluation: Review genres in university settings* (pp. 199-216). Basingstoke: Palgrave, MacMillan.

Gesuato, S. (2007). Evaluation in back-cover blurbs. Special issue of *Textus* XX: 83-102.

Getkham, K. (2010). A corpus-based study of Applied Linguistics research articles: A multidimensional analysis. In I. Moskowich-Spiegel Fandino, B. Garcia, I. Martin & P. Sandino (Eds.), *Proceedings from language windowing through corpora*: Visualizacion del lenguaje a traves de corpus. Part I. CILC 2nd International conference on corpus linguistics. University of a Coruna, Spain.

Gray, B., (2015). *Linguistic variation in research articles: When discipline tells only part of the story*. Amsterdam: John Benjamins.

Halliday, M. A. K. (1988). On the language of physical science. In M., Ghadessy, (Ed.), *Registers of written English: Situational factors and linguistic features* (pp. 162-178). London: Pinter.

Hardy, J., & Romer, U. (2013). Revealing disciplinary variation in student writing: A multidimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, *8*(2), 183-207.

Hymes, D., (1974). *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvanian Press.

Kathpalia, S. S. (1997). Cross-cultural variation in professional genres: A comparative study of book blurbs. *World Englishes, 16*, 417-26.

Marciulioniene, V. (2012). Publisher's blurb on English books of fiction*:* A diachronic genre analysis. Retrieved on 06 June, 2012 from http://archive.minfolit.lt/arch/9501/9982.pdf

Oches, E. (1979). Planned and unplanned discourse. In T. Givon (Ed.), *Discourse and syntax* (pp. 51-80). New York: Academic Press.

Onder, N. (2013). Generic structure and promotional elements in best- selling online book blurbs: A cross-cultural study. *Ibérica*, *25*, 171- 194.

Qasim, S., & Shakir, A. (2016). Linguistic variation of Pakistani fiction and non-fiction book blurbs: A multidimensional analysis. *ELF Annual Research Journal, 18*, 185-206.

Shakir, A. (2013). *Linguistic variation across print advertisement in Pakistani media: A multidimensional analysis* (Unpublished doctoral dissertation). International Islamic University, Islamabad.

Weber, E. G. (1985). From feelings to knowledge: Verbs of cognition. NWAVE XIV, Georgetown. In D., Biber, *Variation across speech and writing*. (p. 224). Cambridge: Cambridge University Press.

Yang, W. (2013). Keyness in academic text book blurbs: Lexical variations across disciplines. *Taiwan International ESP Journal*.