

VARIATION IN LEARNER'S ARGUMENTATIVE ESSAYS – A MULTI-DIMENSIONAL COMPARATIVE ANALYSIS

¹Maimoona Abdulaziz, ²Muhammad Asim Mahmood ³Musarrat Azher

¹Department of Applied Linguistics, Government College University Faisalabad,

^{2,3}Allama Iqbal Road, Faisalabad, Pakistan

Corresponding Author: abdulaziz.maimoona@gmail.com

ABSTRACT: *Variation across learner writing has been of key interest to researchers since long and much research has been made on large-sized learner corpora. Corpus-based studies on Pakistani learner writing are but rare. They are either delimited to focusing solely on individual linguistic or grammatical features without taking into account the systemic variations which involve the co-occurrence of sets of features, or are confined to identifying generic elements specified to Pakistani English. The advanced analytical approach towards corpus takes one step further in that it is multidimensional in nature and aims at identifying situational factors on the basis of co-occurring linguistic characteristics in purely quantitative terms. The purpose of this research is to explore linguistic variation between Pakistani Learner writing (PLC), with native (ENL), second language (ESL), and foreign language (EFL) learner writing. It aims at providing a comprehensive description and comparison of sets of characteristic linguistic features of Pakistani learner writing with ENS, ESL, and EFL (Kachruvian division) learner writing, based on two significant dimensions; “Involved versus Informational Production”, and “Overt Expression of Argumentation” [1]. For this purpose, it uses Multi-Dimensional analysis technique to identify and explain linguistic variation on a corpus of learner essays containing around 0.7 million words, extracted from International Corpus Network of Asian Learners of English (ICNALE), and tagged using Biber Tagger. The study reveals that the variation between PLC and the other three varieties is relatively high on these dimensions and goes on to discuss the variation in the light of previous researches on language teaching and learning. It is the first of its kind in exploring distinctive linguistic features associated with Pakistani learner corpus based on quantitative evidence, and comparing them with those of ENS, ESL, and EFL learner corpora.*

INTRODUCTION

Corpus researches have explored variation in learner writing from myriads of perspectives. These researches made worldwide on learner corpora fall into a range of foci; from lexicography (e.g. Siepmann [2]), translation analysis (e.g. Uzar [3]) and text book writing (e.g. Kaszubski [4]), to ELT practices like examining L2 acquisition (e.g. Tono [5]) and error identification (e.g. He [6], Granger [7,8], Van & Schafer [9], Flowerdew, [10]).

These varieties of learner corpora have mostly been compared from discourse, grammatical, genre, and lexical points of view. They cover a wide range of topics including, study of formulaic expressions (e.g. Granger [11]), idioms (e.g. Kaszubski [12]), adjective intensification (e.g. Lorenz [13]), articles and non-count nouns (e.g. Osborne [14]), contrastive rhetoric (e.g. Damascelli [15]), word sequence (e.g. De Cock [16]), tense and aspect (e.g. Eriksson [17]), lexical patterning (e.g. Altenberg & Granger [18]), collocations (e.g. Nesselhauf [19]), connectors usage (e.g. Altenberg & Tapper [20]), causal links (e.g. Lorenz [21]), epistemic modality (e.g. McEnery & Kifle [22]), progressives (e.g. Virtanen [23]), etc.

Corpus-based studies on Pakistani learner writing have so far focused on more or less the same areas, including error identification and analysis of learners at graduation level [24], lexico-grammatical study of noun phrases [25], and comparison of linguistic features of Pakistani English with those of British and American English [26] etc. These researches used corpus software tools like Wordsmith, Antconc, TextStat, etc. for tagging, concordancing and analysis.

Although these researches do not focus on the traditional view of language as a homogenous entity, but rather on identification of features of different registers, genres and

language varieties; most of them however, are either delimited to focusing solely on individual linguistic or grammatical features, or are confined to mere identifying linguistic features specified to Pakistani English. As for the former, Brown & Fraser termed it as “misleading to concentrate on specific, isolated markers without taking into account systematic variations which involve the co-occurrence of sets of markers” [27]. The latter, however, may be taken as the foundation to recent research methods used in corpus linguistics.

The advanced analytical approach towards corpus takes one step further in that it is multidimensional in nature and aims at identifying situational factors on the basis of frequently co-occurring linguistic characteristics in purely quantitative terms. Its ultimate goal is to achieve a comprehensive description of linguistic variation and use in a language [28]. Recently, researches on learner corpus following this multidimensional approach are surfacing, with large innovations to the basic methodology as given in Biber (1988) [1], based on large-scale written and spoken learner corpora like ICLE, BAWE and LINDSEI, SLC, London-Lund Corpus, etc.

Pascual Pérez-Paredes and María Sánchez-Tornel [29] worked on Louvain International Database of Spoken English Interlanguage (LINDSEI) Corpus for a comparison of EFL and NS learners on the basis of interviews. The results reveal stark differences in the use of several linguistic features that had frequently been used by both the population groups. The research holds its significance from language testing and assessment perspective of L2 learner output. Though the study revealed interesting results, it was confined to spoken data only.

Sandra Götz and Marco Schilk [30] compared formulaic sequences in the speech of NS, ESL and EFL speakers of

German language. They calculated the overall type-token ratio of sequences of the type 3-gram and found that there is no significant differences in NS and ESL population group, although NS and EFL showed some marked differences. Further, a comparison of the common core was also made which revealed that the common core was more frequent in ESL and EFL as compared to NS. Lastly, a functional analysis was conducted that showed further differences among the variants, with less variability in EFL vs NS and more in ESL regarding types and tokens. The study is limited in its comparison though. Using multidimensional analysis approach would have given more detailed analysis and description of the underlying grammatical differences of the three varieties.

Bertus van Rooy [31] from North-West University explored 67 linguistic features across a corpus of Black South African English, taking it as a variety of English. The data was extracted from the Tswana Learner English (TLE) Corpus and later compared with Louvain Corpus of Native English Speaking Students (LOCNESS), taking it as representative of Standard English. The aim was to explore whether or not elements of indigenization and systematicity existed in the student writing of Black South African English. A multidimensional analysis was conducted for this purpose on both the varieties of English using Biber's [1] model.

The study continued as van Rooy & Terblanche [32] used the new multidimensional analysis model on both the varieties in order to study the nativeness and non-nativeness differences based on their respective linguistic features. According to this study, the old model had its limitations when it comes to identifying specific dimensions for the two groups. In this regard, the new model identified five new dimensions, viz. Advanced Literacy, Transparency, Informal Style, Contextualization of Information, and Persuasiveness (also identified earlier in Grabe & Biber [33]). An internal evaluation of the features related to both the varieties revealed stylistic variation within academic writing as a register. As much as the new dimensions are characteristic of the two corpora, the overlapping of dimensions in the two methods, old and new, is useful in adding to the validity of the multidimensional approach.

A similar comparative study was conducted on nominalization in L1 and L2 writing and speech [34] where the frequency and use of nominalization was examined across spoken and written registers of the two varieties. It aimed at demonstrating similar frequency of nominalizations in L1 and L2 as well as examining possible functional differences between the types of nominalizations. The results validate study of Biber et al. [35] on nominalization with a more elaborate list of suffixes than Biber's.

Richard Xiao [36] studied World Englishes using a multidimensional approach in order to enhance the MDA framework with semantic components. He introduced a new "enhanced model" in order to explore twelve registers and five varieties of World Englishes. The annotated data was extracted from International Corpus of English (ICE).

In 2013, however, Richard Xiao along with Yan Cao experimented the model by conducting a contrastive analysis of English abstracts from twelve academic disciplines,

written by NS and NNS writers [37]. The NS corpus contained abstracts from native English writers whereas the NNS corpus contained abstracts written by Chinese writers. 47 out of 163 linguistic features retained after the factor analysis was performed and based on those factors, seven dimensions were identified.

The most distinctive element however lies in its methodology where colligation is integrated at the interpretation stage along with grammatical and semantic features. It therefore moves a step ahead of the general multidimensional analysis in that it elaborates the text by providing linkages across the lexis, grammar, and text.

The methodology was induced with Coh-Metrix when a group of researchers from Georgia State University and Arizona State University analysed a corpus of 1500+ essays [38] in order to examine their functional parameters based on situational parameters. The results were interpreted on four dimensions and associated with essay quality, prompt, and grade level. The study provides insight into the situational parameters affecting writing as well as adds to the reliability of the MDA methodology.

Multi-dimensional analysis of Biber [1] has also been extended to new MD in researches such as that of Egbert [39-41] where new factors are calculated based on a specific register like published academic texts, or web-corpus, etc. The new factors are then interpreted as new dimensions in terms of their functionality.

However, we have confined our methodology to the use of 88 MD as it serves the purpose of this research. This study proposes to find linguistic variation across Pakistani learner writing and learner writing of other varieties of English from a multidimensional perspective, as given by Biber [1]. For this purpose, corpus of Pakistani learner writing has been considered on the one hand. While on the other hand, learner corpus of native and non-native speakers of English has been taken into account, categorized into three different groups based on the Kachruvian division of World Englishes. It goes further in exploring characteristic features of PLC on two dimensions and explaining them with reference to social and pedagogical contexts in the light of previously made researches.

This research focuses on answering one major question: How far does Pakistani learner writing vary from ENS, ESL, and EFL learner writing?

This is aided by the following minor questions:

- How far is PLC similar/different to ENS, ESL, and EFL learner corpora on the cline of Involved versus Informational Production?
- How far is PLC similar/different to ENS, ESL, and EFL learner corpora on the cline of Overt Expression of Persuasion/Argumentation?
- What do these similarities/differences reveal about PLC?

MATERIAL AND METHODS

Corpus Design

The corpus design is based on Kachruvian division of World Englishes into three categories or circles; The Inner Circle, The Outer Circle, and The Expanding Circle, the primary goal of which was to illustrate the unprecedented variability in English [42]. Braj Kachru's Three Circle Model sets out to

demonstrate the types of varieties that have surfaced over the years with the spread of English. Firstly, the Inner Circle consists of regions where English has spread demographically through English immigrants and is now functional as the primary language of those regions; USA, UK, Canada, Australia and New Zealand belong to this category. Secondly, the Outer Circle consists of countries or regions where English was introduced to the indigenous inhabitants (mostly through colonization), now functions as a second language of the local population, and has been institutionalized. The level of competence may vary in the speakers of these regions and may be altered due to impacts of nativization. Pakistani English stands amidst this category among other varieties. Finally, the Expanding Circle inhibits areas or countries where the status of English is that of a foreign language in which the locals of those areas may communicate internationally. In terms of language pedagogy, the terms used for these three categories are English as a Native Language (ENL), English as a Second Language (ESL), and English as a Foreign Language (EFL), respectively [42]. It serves as the basis for this research design to the extent that the categorization of learner corpus, with which Pakistani Learner Corpus (PLC) is to be compared, has been made based on the Kachruvian division.

Population and Sampling

A corpus of around 0.7 million words of learner essays was extracted from the International Corpus Network of Asian Learners of English (ICNALE). Data was bifurcated into four population groups owing to the objectives set by this research and PLC was extracted from the ESL group for the purpose of comparison. The criteria for sample selection involved language proficiency of the students, therefore essays written by students of lower proficiency, as determined by their TOEIC/TOEFL/VTLS scores, were excluded. The sample therefore consisted of the following number of essays/text files for each population group:

Table 1 Number of essays in each population group

Population groups	Sub groups	Sample size	Total sample size
PLC	-	-	178
ENS	-	-	200
ESL	Hong Kong	138	912
	Philippines	374	
	Singapore	400	
EFL	China	236	1242
	Indonesia	172	
	Japan	134	
	Korea	328	
	Thailand	204	
	Taiwan	168	

The essays were collected under controlled writing conditions regarding topic, time, dictionary use, spell and grammar check, and essay length. The essays were argumentative in nature and two essays per learner were collected on the topics of “It is important for college students to have a part-time

job” and “Smoking should be completely banned at all the restaurants in the country”.

Tagging the Corpus

Once sampled, the corpus was sent to Dr Jesse Egbert for running Biber’s tagger software. The files were first converted to ASCII to make them consistent in their character encoding and later tagged for linguistic features as given in Biber [1]. A sample text and its tagged equivalent are presented below:

Some people think that it is very important for college students to have a part-time job.

Some ^dti++++=Some
 people ^nns++++=people
 think ^vb+vprv+++=think
 that ^tht+vcmp+++=that
 it ^pp3+it+++=it
 is ^vzb+bez+vrb++=is
 very ^ql+amp+++=very
 important ^jj++++=important
 for ^in++++=for
 college ^nn++++=college
 students ^nns++++=students
 to ^in++++=to
 have ^vbi++vrb++=have
 a ^at++++=a
 part ^nn++++=part-time
 - ^zz++++=EXTRAWORD
 time ^nn++++=EXTRAWORD
 job ^nn++++=job.

. ^zz++++=EXTRAWORD

Calculating Dimension Scores

Once the data was tagged for linguistic features, the raw counts of the frequencies of each linguistic feature was calculated, and the counts were normalized to per count per 1000 words, a standard set by Biber [1]. The normalized values were then standardized to the mean of 0.0 and a standard deviation of 1.0 so that all features on each dimension may have equivalent weights when their respective dimension scores were calculated. The scores of each group were calculated on five dimensions as set by Biber [1], yet based on the significance of the results, only two dimensions were considered for this study; “Involved versus Informational Production” and “Overt Expression of Persuasion/Argumentation”. The sets of co-occurring features relevant to each dimension are divided into positive and negative features. This division is complementary and does not indicate a particular feature as having a negative not positive effect on the dimension scores, but actually refer to the two extremes of each dimension within which the data is placed. The positive and negative features for the two dimensions are as follows:

Table 2 Positive and negative features of the two dimensions

Dimension 1: Involved versus Informational Production	
Positive Features	Negative Features
Private verbs	Nouns
THAT deletion	Prepositional phrases
Contractions	Word length
Present tense verbs	Attributive adjectives
Second person pronouns	
DO as proverb	

Demonstrative pronouns
General emphatics
First person pronouns
Pronoun IT
BE as main verb
Causative subordination
Discourse particles
Indefinite pronouns
General hedges
Amplifiers
Sentence relatives
WH questions
Possibility modals
Non-phrasal coordination
WH clauses
Final prepositions
Dimension 2: Overt Expression of Argumentation/ Persuasion
Positive Features
Negative Features
Infinitives
No negative features
Predictive modals
Suasive verbs
Conditional subordination
Necessity modals
Split auxiliaries

RESULTS

This section entails statistical details of 88 MD analysis; comparisons of the results obtained from the analysis according to the research questions of this study, their interpretation, and finally conclusions based on these comparisons. It begins with statistical results showing range and standard deviation of the four groups on each dimension, followed by an explanation of significant values obtained from them and their interpretation. It goes on to illustrate comparisons of PLC with ENL, ESL, and EFL, and later with ESL and EFL sub-varieties. These comparisons are backed up with results obtained from ANOVA in order to highlight statistically significant similarities or differences between the groups. Finally, the results are discussed with respect to each comparison and then concluded so that further discussion may be carried out.

Comparison of spread of scores and means of the four population groups along five dimensions

The box plots below present the range of dimension scores of the four population groups on the two dimensions based on 1988 multidimensional analysis. They exhibit the maximum, minimum, and mean scores for each population group, based on the figures given in the table below:

Table 3 Comparison of Mean Dimension Scores

	Mean	Maximum	Minimum	Range	Standard Deviation
Dimension 1 (D1)					
EFL	8.2	49.5	-20.5	70.0	11.0
ESL	1.2	45.5	-22.0	67.5	10.3
ENL	8.5	41.8	-20.3	62.2	12.1
PLC	-1.7	29.8	-20.1	49.9	8.3
Dimension 2 (D2)					
EFL	4.3	33.9	-6.6	40.5	5.6
ESL	4.8	29.7	-6.6	36.3	5.6

ENL	4.9	23.1	-5.5	28.6	5.1
PLC	2.0	18.6	-6.6	25.2	5.5

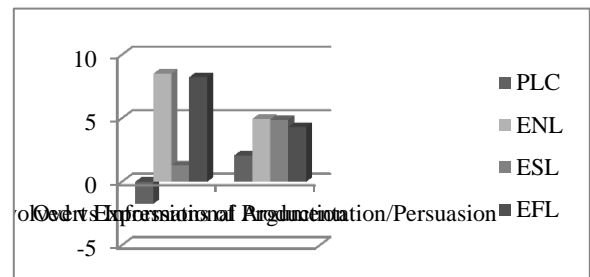
As can be seen, Dimension 1 (D1) holds the largest ranges of the two dimensions, rising above 70 for EFL. The standard deviation is also the highest on D1, with a maximum of 12.1 for ENL. According to Biber [1], the range of scores represents the amount of internal coherence within a group, which in turn indicates the possible range of variation within a group. The figures therefore show that text variation within a group is highest on D1. It is also noteworthy that EFL has the highest ranges on both the dimensions and ESL has the second highest. On D1, EFL has the range of 70.0, while ESL lies close with 67.5. Similarly, on D2, EFL has the highest range of 40.5 and ESL of 36.3. This is indicative of the fact that maximum variation and heterogeneity within any group is found in EFL, followed by ESL. It is therefore necessary to have a closer look at the sub-varieties lying within these two groups for a more detailed comparison.

Comparison of varieties and sub-varieties along the two dimensions based on mean dimension scores

The overall comparison of the mean dimension scores of the four groups obtained from 88 MD analysis has been summarized in Fig 1. Here D1 exhibits bipolar results, where ENL, ESL, and EFL are on the positive side viz. the Involved side, while PLC is on the negative side or Informational side of the plane. ENL and EFL show peak levels of Involved Production, rising above 8, while ESL is at a moderately lower level. The accumulative result, irrespective of the negative and positive side, shows PLC as closest to ESL, and ENL as closest to EFL. D2 illustrates the most variation with respect to PLC as far as the writer's Expression of Persuasion is concerned. Here PLC not only scores the lowest but also stands aloof in the graph, as ENL, ESL, and EFL stand close to each other and score high on Persuasion.

In order to gauge whether or not the variation illustrated in Fig 1 is statistically significant, a two-way ANOVA has been applied to the mean dimension scores of the four groups.

	D1	D2
PLC	-1.74±0.62C	2.08±0.41B
ENL	8.51±0.85A	4.95±0.35A



ESL	1.27±0.34B	4.85±0.19A
EFL	8.25±0.31A	4.30±0.16A
F-value	106.0**	13.3**

** = Highly significant ($P < 0.01$)

Means sharing similar letter in a row are statistically non-significant ($P > 0.05$).

Fig 1 Comparison of the four groups along the two dimensions

The results confirm bipolarity on D1, where the F value reaches 106, $p < 0.01$. This is indicative of the high level of variation between PLC and ENL. On the other hand, ENL and EFL stand closer than any other variety, whereas PLC shows variation from all the other varieties. On D2 too, PLC stands aloof from the rest of the three varieties with a significant F value of 13.3, $p < 0.01$. ENL, ESL, and EFL on this dimension show a great deal of similarity between them as opposed to PLC. We may derive from these results the following conclusions:

1. PLC stands as a distinguished variety on both dimensions.
2. PLC does not show similarity to either ESL or EFL varieties on any dimension.
3. ESL and EFL are similar on D2.

As has been discussed previously, EFL and ESL have the highest and second highest range on the two dimensions, respectively, showing the least amount of homogeneity and internal coherence. It is therefore essential to have a glance into the sub-varieties of both these groups and see whether or not there is any variation between PLC and these sub-varieties. Data from the two groups has therefore been bifurcated and compared with PLC. The comparisons have been made based on the mean dimension scores of the sub-varieties and PLC.

Fig. 2 Comparison of PLC with ESL Sub-Varieties

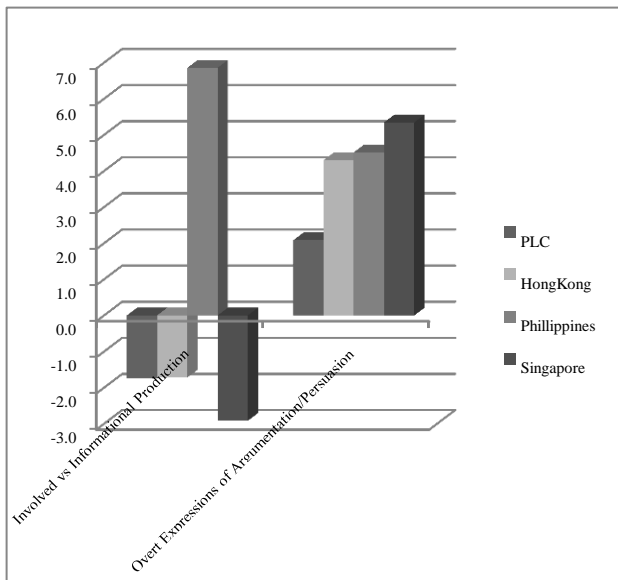


Fig 2 demonstrates the relations among PLC and ESL sub-varieties with respect to the two dimensions. The ESL sub-varieties compared with PLC include Hong Kong, Philippines, and Singapore. The bar graph shows variation not only between PLC and ESL sub-varieties but quite interestingly within ESL sub-varieties, more specifically on D1. Philippines is in fact the only ESL sub-variety that scores high on Involved plane whereas Hong Kong and Singapore score on the negative or Informational side. PLC scores closest to Hong Kong on this dimension. This exhibits

diversity within the ESL sub-varieties. Lesser variation may be seen on D2 between the ESL sub-varieties, specifically between Hong Kong and Philippines, yet PLC stands aloof on this dimension with a comparatively lower score.

In order to gauge this variation, ANOVA has been applied to the mean dimension scores of PLC and the ESL sub-varieties.

	D1	D2
PLC	-1.74±0.62B	2.08±0.41C
Hong Kong	-1.72±0.72B	4.30±0.44B
Philippines	6.86±0.54A	4.52±0.31B
Singapore	-2.92±0.42B	5.34±0.27A
F-value	85.38**	31.28**

** = Highly significant ($P < 0.01$)

Means sharing similar letter in a column are statistically non-significant ($P > 0.05$).

Statistically, PLC, Hong Kong, and Singapore show similarity on D1, whereas Philippines exhibits extreme diversity and bipolarity, with an F-value of 85.38, $p < 0.01$. It may be gathered from these statistics that the overall score of ESL displayed in Fig 1 as more Involved than Informational is due to the overwhelmingly high score of Philippines, whereas the rest of the two ESL sub-varieties tend to be more informational than involved. On the other hand, PLC and Hong Kong score the closest with -1.74 and -1.72, respectively, on D1. Hong Kong shows similarity with Philippines on D2, scoring closely with 4.30 and 4.52, respectively. While Singapore scores highest on D2 closer to the two ESL sub-varieties, PLC has the lowest score and stands aloof. Therefore in terms of argumentation/persuasion, it may be deduced that PLC varies greatly from all the three ESL sub-varieties. The difference is highly significant, with an F-value of 31.28, $p < 0.01$.

We gather from the above analysis the following information:

1. PLC shows high level of variation from the ESL sub-varieties on D2, whereas on D1, it stands closest to Hong Kong.
2. The highest variation may be found between PLC and Philippines on D1.
3. PLC is comparatively less argumentative than the ESL sub-varieties due to comparatively lower score on D2.
4. The ESL sub-varieties vary largely on D1, where Philippines shows bipolarity in contrast with Hong Kong and Singapore.
5. Hong Kong and Singapore show significant similarity along D1.

A similar comparison has been made between PLC and the EFL sub-varieties on these dimensions, based on their mean dimension scores.

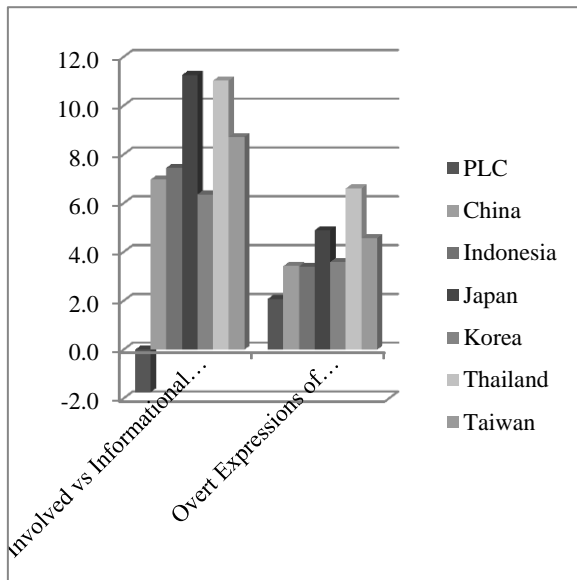


Fig 3 Comparison of PLC with ESL Sub-Varieties

The graph in Fig 3 illustrates a comparison of mean dimension scores of PLC with the EFL sub-varieties namely China, Indonesia, Japan, Korea, Thailand, and Taiwan. Variation may be seen at several levels between PLC and the EFL sub-varieties, as well as within the EFL sub-varieties.

Here D1 demonstrates similar bipolarity of PLC with EFL sub-varieties that has already been seen in the previous comparisons at D1. PLC scores moderately low on the Informational side of the graph on this dimension. Japan scores the highest and stands closest to Thailand in terms of the text being Involved. All the EFL sub-varieties score high on this dimension, ranging between 12 and 6. Variation may also be seen on D2, where Thailand scores the highest and PLC the lowest. China and Indonesia show similar scores and stand closer to PLC than any other sub-variety.

These results are backed up with a two-way ANOVA in order to test the significance of the variation that exists between PLC and EFL sub-varieties as well as within EFL sub-varieties.

Table 4 Comparison of mean dimension scores of PLC and EFL sub-varieties

	D1	D2
PLC	-1.74±0.62D	2.08±0.41E
China	6.98±0.70BC	3.43±0.31D
Indonesia	7.46±0.84BC	3.39±0.38D
Japan	11.26±0.84A	4.89±0.47B
Korea	6.37±0.64C	3.59±0.31CD
Thailand	11.03±0.74A	6.61±0.45A
Taiwan	8.72±0.84B	4.57±0.42BC
F-value	112.10**	13.36**

** = Highly significant ($P < 0.01$)

Means sharing similar letter in a column are statistically non-significant ($P > 0.05$).

The results from ANOVA show the extent to which the differences between PLC and EFL sub-varieties are

significant. The high F-value 112.10, $p < 0.01$ at D1 indicates bipolarity regarding PLC and EFL sub-varieties. This indicates that EFL sub-varieties, unlike ESL sub-varieties, tend to occur unanimously on the Involved side of the plane. Pertaining to this homogeneity within the EFL group, China and Indonesia show negligible differences and so do Japan and Thailand. Taiwan is closer to China and Indonesia than any other sub-variety on D1 and so is Taiwan. China and Indonesia show similarity on D2 with scores of 3.43 and 3.39, respectively, whereas Korea lies closest to them at 3.59. It may also be noticed that Thailand and PLC show the greatest difference on this dimension, which means that Thai learners tend to develop the most argumentative texts as compared to PLC or any other text from EFL sub-varieties. This may be placed in a nut-shell as,

1. PLC stands as a distinguished variety at both dimensions.
2. The EFL sub-varieties vary largely on D1 and somewhat on D2.
3. China shows similarity with both Indonesia and Korea on both dimensions and with Taiwan on D1, whereas Japan shows similarity with Thailand on D1.

Based on the concluding results obtained from the three comparisons, a comprehensive discussion has been conducted in the following section with reference to the linguistic features relevant to the two dimensions.

DISCUSSION

This section construes detailed discussion and arguments based on the results obtained by 88 MD analysis. The analysis has served the primary objective of this research of exploring variation among Pakistani learner writing and ENL, ESL, and EFL learner writing, i.e. of highlighting the similarities and differences not only between PLC and ENL, ESL, and EFL learner writing, but also between PLC and the sub-varieties of ESL and EFL. To this extent, the comparisons have been made on the two dimensions of 88 MDA and some interesting results have surfaced which have been discussed with reference to the linguistic features relevant to each dimension. Further, the discussion has been contextualized with previously made studies on argumentative essays in general and Pakistani learner writing in particular, wherever required, in order to generate more clarity on the matters. On a side note, prominent similarities and differences across ESL and EFL sub-varieties have also been mentioned, though not in detail, as they do not answer the research questions set by this study.

As is the case, PLC stands out as a distinguished variety when compared with ENL, ESL, and EFL texts overall. Based on sociolinguistic evidence, as presented in Kachru [42], Pakistan falls in the Outer Circle of World Englishes or the ESL group as in this case. The results based on learner corpora however contradict the sociolinguistic parameters as this. When viewing the overall score of the four groups on the two dimensions (Fig 1), it is evident that PLC shows a great deal of variation from all the other groups on both dimensions. Although there may be a number of interpretations to it, it is highly indicative of the current situation of English language pedagogy and assessment in

Pakistan. Shamim [43] highlighted some major issues in learner writing assessment, one of which was the focus of assessment being on the content knowledge that would display most of the characteristics of a good paragraph, rather than language skills. Another related issue was that both the learners and the teachers were unaware of this issue and deemed the ELT courses fit for meeting the future needs of the students, which is as delusional as it is detrimental. In her study on issues of access and equity in language planning in higher education, Mansoor [44] explores the English proficiency level of Pakistani higher education students and terms it as far below the level required for higher studies and future work. She claims that the reason behind this lag is neither learners' attitudes nor their motivational intensity, but the low quality courses that are outdated and inefficient. According to Mahmood [26], "The current language policy (of Pakistan) is not clear about the choice of the variety of English." No matter if this lag is due to a lack of theoretical knowledge regarding language pedagogy on part of the teacher or due to the maximum marks-attaining goals of the learners, it is highly indicative of the void the learner has to cover in order to fulfill the global linguistic needs required to achieve excellence internationally or what Rehman [45] terms as "international intelligibility".

This intelligibility in its extended form may include English writing skills such as argumentation, exposition, narration, etc. Hence, where PLC scores on the two dimensions emphasize Pakistani learner writing as a distinguished variety, it also suggests that the cause of this variation may be beyond stylistic and lexical choices of the learners as far as language functionality is concerned. For this purpose, a study of linguistic features characteristic of the two dimensions needs to be made.

Dimension 1: Involved versus Informational Production

As we know that the variation among the groups and sub-groups on each dimension are based on some linguistic features relevant to each dimension, the details of which have been given in Table 2. The groups that score high on the first dimension signify that their respective text is highly involved and least informational. It is characterized by variables or linguistic features such as private verbs, THAT deletion, contractions, present tense verbs, first and second person pronouns, the proverb DO, demonstrative pronouns, emphatics, pronoun IT, BE as main verb, causative subordination, indefinite pronouns, hedges, amplifiers, etc. and are less frequent in their use of nouns, prepositional phrases, attributive adjectives, etc. PLC scores as the least involved text, contrary to ENL, ESL, and EFL groups, as is illustrated by the following example texts from each group:

PLC text (W_PAK_PTJ0_146_B1_2), D1 score: -19.85

The most popular example of part time job in our country is the job of call centers. At call centers the people work round the clock in different shifts and without caring for day and night. Mostly students are doing these kind of jobs to increase their pocket money and increase their experience of practical life. Although this is a healthy trend to increase the money and income and a good way to keep the mind

and body busy and working in a positive way even yet it is having some bad effects as well.

ENL text (W_ENS_SMK0_017_XX_1), D1 score: 41.88

We learned in school that every year a smoker may spend thousands of dollars on cigarettes. I have to say that I think this is a waste of money, because cigarettes don't really do anything for you. Alcohol is kind of the same way, that is, it doesn't really do anything for you, but at least when you drink safely and do not drive, you will not affect the health of other people around you.

ESL text (W_PHL_SMK0_028_B1_2), D1 score: 45.52

As a step of prevention, I think that smoking should be completely banned from every restaurant in the country. Why restaurants? It's where the most us commonly go. We Filipino's are food lovers indeed. And most of the people we see in restaurants are children. Why would we want to risk their health with a stupid mistake? We should be implementing this rule. Smoking is very bad to the health. Based on a survey made, 7 out of 10 people die every day due to smoking. Now, can we take that risk? No, we can't. Why do people even smoke? What do they get from it? Can smoking make them richer? No. Can smoking make them smarter? No. Here's one, smoking makes them an idiot. An idiot that takes his life at risk and an idiot who doesn't care about Mother Nature. In order to lessen smokers, Smoking must be banned in every restaurant in the country, not just it gives harm to the smokers' health but also to the people surrounding them. And specially, how will you be able to eat properly and comfortably while smelling such foul and irritating scent? Isn't it a disrespect to those persons eating in the restaurant? So for me, smoking must be banned in restaurants.

EFL text (W_THA_SMK0_393_B1_2), D1 score: 45.81

Smoking is one the main reasons that has killed a lot people already. Everyone knows that smoking does not give anything to our life, but it kills us slowly. Nowadays, smoking is already getting banned in every restaurant. But some restaurant still has a room for smoking. Smoking should be completely banned in restaurant because the smell of cigarette is so bad, and smoking can destroy your own health and everyone around you. Smoking is something that everyone should not try to start it. The first reason why smoking should be completely banned is because of the smell.

The text illustrations of ENL, ESL, and EFL groups are exemplars of positive linguistic features of D1, signifying involvement on part of the writer. There is abundance of first person pronouns, both singular and plural (I, we), as well as second person pronoun (you). These references are typical of highly involved texts like face to face or telephonic conversations, etc. In written genre, they are more used in informal or personal letters, transcribed interviews, etc.

Contractions (e.g. *it's, can't isn't, doesn't*) and the proverb DO (*do anything for you*) are also frequently used. Similarly, the texts are full of present tense verbs (e.g. *say, do, drink, drive, make*), amplifiers (e.g. *completely, indeed, very*), emphatics (e.g. *really*), hedges (e.g. *I have to say I think...*), and possibility modals (*may*). Causative subordination can be seen in the use of conjunctions like 'because' (e.g. *because cigarettes don't really do anything for you, because the smell of cigarette is so bad*). Similarly, these texts are less frequent in the use of negative features of D1. Although nouns are present, they hardly co-occur with either attributive adjectives or prepositional phrases. Type-token ratio is also comparatively lower.

On the other hand, PLC holds negative features of D1 and is hence informational. Frequently used nouns (e.g. *example, people, work, clock, day, night, students*, etc.), attributive adjectives (as in, *most popular example*), prepositional phrases (e.g. *in our country, in different shifts, to increase their pocket money, to keep the mind and body*, etc.) co-occur in this short text making it lesser interactive and more involved. The attributive adjectives here, most of which are classifiers, function as descriptors of the nouns, adding more information to the text. Mostly, they are prominent in genres as academic prose which includes University prospectuses, reading material, journal articles, etc. the objective of which is to provide useful information without much argument. It is therefore unusual for essays, more particularly argumentative essays, to be more informative than interactive. Features like word length and type-token ratio are also prominent in PLC text as compared to the other three texts, as longer words like *popular, different, increase, experience, practical, positive*, etc. have frequently been used. Similarly, in addition to frequent occurrences of negative features, PLC is characterized by relative absence of the positive features of D1. Contractions, THAT deletion, hedges, first and second person pronouns; are but extinct in PLC texts.

The lesser interactive style that is prominent in PLC may refer to one of the two aspects, or both; viz. personal preference of the learner with respect to writing style, and/or under-developed essay writing skills of the writer. There are comparatively lesser chances of the former, as style may vary from learner to learner, in which case, further variation would have emerged, which is quite improbable, given that PLC has the smallest range or the least amount of heterogeneity on this dimension as has already been discussed in reference to Table 3. We may, therefore, have to confine to the latter interpretation derived from the results of D1, i.e. the essay writing skills of the learners may have short comings.

Where good argumentation usually requires a convincing, affective, and more involved, style of writing in order to present the writer's stance with clarity, a text loaded with heavy nominalizations, such as PLC, may become congested and ambiguous. In fact, compared with the ESL sub-varieties, this phenomenon is quite common, as Singapore and Hong Kong texts illustrated the same tendency on D1 as PLC. It is interesting that previous literature on Pakistani English such as Rehman [45] claims that although standard Pakistani English is identical to standard Indian English, they are by no means similar in lexico-semantic aspects of their written

forms. The results accumulated from quantitative analysis of linguistic features on D1, however, do not support this claim. In fact, the heavily nominalized or information-jam-packed Pakistani learner writing coincides with that of Indian writing which Xiao [36] terms as "nouny", i.e. showing preference to nouns rather than verbs, which makes it less interactive and more elaborate. He further quotes that such writing creates obscurity for the reader rather than impress him/her. For argumentative writing too, the language needs to be interactive, as terseness in expression needs more time for comprehension and the reader may lose interest. This may indicate a need for improvised essay writing pedagogical practices especially for argumentative essay writing; one in which the focus should be on developing and presenting argumentation rather than passing on maximum information about the subject.

Having a closer look at the sub-varieties of ESL and EFL reveals some more interesting results. ESL scores low on this dimension yet lies on the positive side when taken as a whole. However, it may be seen that Philippines is the only one scoring on the positive side of this dimension, while Hong Kong and Singapore show quite the opposite and can be categorized as informational, similar to PLC. The overall score of ESL is positive on this dimension mainly due to the high score of Philippines. This variation within ESL group is large enough to signify a change in writing trends and/or language pedagogical practices in different countries belonging to Outer Circle. The question also arises as to whether or not the Kachruvian division stands the test of time while globalization is persistently making due changes in World Englishes. This quantitative survey of learner essays alone gives sufficient evidence for the existence of this phenomenon.

Dimension 2: Overt Expression of Persuasion/ Argumentation

In terms of Expression of Persuasion/ Argumentation too, there is slight variation within EFL sub-varieties, lesser between ESL sub-varieties, and remarkable differences between PLC and the rest of the varieties. D2 marks elements of persuasion in a text through co-occurrence of linguistic features such as infinitives, modals (predictive, necessity), suasive verbs, split auxiliaries, and conditional subordination. These features work in combination to exhibit the writer's opinion or stance in a convincing and advisable manner. The graph in Fig 1 shows PLC as the least persuasive text as compared to ENL, ESL, and EFL texts. This can be explained better from the following text samples taken from the four groups:

ENL text (W_ENS_PTJ0_032_XX_1), D2 score: 10.15

If we're lucky, our parents are very responsible and have already saved up a lot of money so that we can go to college and focus on school rather than making money ourselves. Actually, I had a part-time job in high school, but in that time, I had a lot more free time and I didn't take school as seriously. Everything is different in college though. Now we're studying to become lawyers or doctors or engineers or psychologists were artists or anything else, and

we are studying so that we can spend the rest of our careers working in these areas. It is important to spend as much time gaining as much knowledge as we can in our own fields so that we can be successful when we have to enter the real world and start working at companies. So no, I don't think college student should have a part-time job, and I think that it will be a while before I start working.

ESL text (W_HKG_PTJ0_075_B2_0), D2 score: 15.36

Having a part-time job would be the one which will teach university students many things that school cannot teach them. Therefore, it would be necessary to university students should have a part-time job at least once in their university life. From part-time job experiences, students may be able to improve effectively their communication skills and interpersonal skills. These skills would not be able to be taught but students should learn by themselves through their experiences. Many companies would value these working experiences since the one who has knowledge on working as a team and able to communicate well in an effective way and who does not have, would have differences. By having a part-time job, students may not only have pocket money but also can have unique experiences. Therefore, university students should have a part-time job.

EFL text (W_CHN_SMK0_118_B1_2), D2 score: 19.46

Surely smoking should be completely banned at all the restaurants in the country. As is known to all, smoking is harmful to our health, so it should be banned at all the public places. Restaurants are one of the public places. Moreover, they are the public places where you have meals. If some people smoke in the restaurants and you are a non-smoker, can you stand your health being ruined while having your dinner? As for people who are sensitive to the smoke of cigarettes, their appetite will be affected by the annoying smoke and they are not likely to enjoy their food. If bosses of restaurants are afraid to lose the guests who are smokers, I suggest they should also consider the guests who are non-smokers.

PLC text (W_PAK_PTJ0_150_B2_0), D2 score: 1.38

People are facing lot of problem because of their earning and it is very difficult for them to manage daily bread. But as we know this is age of science and technology. Everybody wants to get more and more education. But due to insufficient income of their parents and some other worse circumstances they have to do part time jobs to meet their expenses. That's why these many students are doing part time job during their studies. It is very difficult for them to manage two very tiresome task. They are compelled to do part time jobs but it is necessary for them. The adverse circumstances taught them lot of things during their studies. They become very mature at very early age.

The first three texts illustrate a more frequent use of modals; more particularly necessity modals (as in *we should* not have

to provide, *we have* to enter the real world, college student *should* have a part-time job, university students *should* have a part time job, smoking *should* be completely banned, I suggest they *should* also consider the guests who are non-smokers) and predictive modals (like, it *will* be a while before I start working, a part time job *would* be the one, it *would* be necessary, companies *would* value these working experiences, their appetite *will* be affected). The texts illustrate that the necessity modals *should* and *have* occur with either an infinitive verb or a non-perfect tense, while necessity modals denoting logical necessity markedly co-occur with perfect tense, according to Biber et al. [35]. There is no logical explanation for the need of performing a certain task or occurrence of an event. However, it does denote the writer's attempt to persuade the reader of having a part-time job or also considering the guests who are non-smokers etc. as desirable options. Contrary to results achieved from previous literature regarding the use of necessity modals in East Asian students essays [46], it is worth noticing that the use of necessity modals in PLC texts is negligible as compared to the other three varieties that show a more directly persuasive style. The difference in frequency of use of necessity modals between PLC and the other three groups also contradict the results of more recent studies done on student academic writing like that of Russell [47], according to which there is no significant difference in English learner writing and native learner writing with respect to the frequency of use of necessity modals.

Predictive modals, on the other hand, denote the writer's intention, when they co-occur with first person pronoun. However, here they are mostly seen with third person pronouns presenting likelihood of occurrence of some events, like the writer's initiative for working or companies' evaluation of working experiences, etc. Although the texts lack abundance of suasive verbs (e.g. the adverse circumstances *taught* them a lot of things) in particular, the other features on this dimension sufficiently mark explicit attempts of persuasion on part of the writer. It may also be seen that PLC text differs here with respect to the other three texts in that it is quite devoid of either necessity modals or predictive modals. There is however high percentage of infinitives (e.g. in order *to make* ends meet, studying *to become* lawyers, it is important *to spend* much time, not be able *to be* taught, not likely *to enjoy* their food, it is very difficult for them *to manage* daily bread, everybody wants *to get* more and more education, they have *to do* part time jobs *to meet* their expenses, it is very difficult for them *to manage*, they are compelled *to do* part time jobs, etc.) in all the four texts, especially PLC text.

PLC's lowest score on this dimension may pertain to the fact that it has the least use of two important features marking overt expression of persuasion/ argumentation; split auxiliaries (present in other texts as, *have already saved*, I *didn't take* school as seriously, I *don't think*, students *may not only have* pocket money, I suggest they *should also consider*, etc.) and conditional subordination. Whereas ENL, ESL, and EFL learners build up their arguments by extending different propositions using complex structures such as conditional coordinate clauses, PLC learners show preference to explanatory style of writing without developing much

argument on the topic no matter if the topic at hand is argumentative in nature. The following examples may be considered for a more comprehensive explanation of this phenomenon:

- 1) *If we're lucky, our parents are very responsible and have already saved up a lot of money so that we can go to college and focus on school rather than making money ourselves.*
- 2) *If some people smoke in a restaurant and you are a non-smoker, can you stand your health being ruined while having your dinner?*
- 3) *If bosses of restaurants are afraid to lose the guests who are smokers, I suggest they should also consider the guests who are non-smokers.*
- 4) *People are facing lot of problem because of their earning and it is very difficult for them to manage daily bread.*

The proposition in the first extract taken from ENL text, lies in the main clause preceded by the conditional clause *If we're lucky* which presents a semantically 'open condition' [35], i.e. not specifying whether or not the condition is fulfilled or not, but expressing at the same time the possibility of its fulfillment. Similarly, the condition in EFL text extract 2) *If some people smoke in a restaurant and you are a non-smoker* also generates a hypothetical situation in the mind of the reader which is again semantically 'open' and may lead the reader to believe in the proposition carried by the main clause, which, in this case, is a question rather than a statement, the objective of which is neither to inquire nor provoke thoughts on the matter but to illustrate harmful effects of smoking in a convincing manner. It is more of a rhetorical device used for persuading the reader through thoughtful imagination. The semantic category is similar in the conditional clause in extract 3), though the main clause in this case contains communicative verb *suggest* which highlights explicitly the writer's recommendation based on the argument. The crux of this may therefore be interpreted as, the conditional clauses used in argumentation development are confined to 'open' category only serving a variety of purposes, like expressing possibility, creating hypothetical dystopia, and giving suggestions; all directed towards persuading the reader towards a particular point of view.

Extract 4) taken from PLC text, on the other hand, uses conjunctions for explanatory purpose which may be seen in the fuller PLC text, where the nature of writing is more expository than argumentative. Lack of modals, split auxiliaries, and conditional subordination signify the missing element of argumentation in PLC texts. Although attempts for persuasion are made based on reasoning yet there is an ambiguity as to the stance taken by the writer in favor of or against the topic at hand. Some PLC texts have even demonstrated a change in stance in the middle of their argument. Imtiaz & Mahmood [48] highlighted similar instances based on genre analysis of Pakistani learners' argumentative writing and identified them as a new move, Contradiction, which according to them "showed the inconsistency with or relevance to the Proposition" (p.100). A lack of stance or contradictions within an argument may

indicate learners' inability to comprehend the basic objective of argumentative writing. Therefore, in spite of the fact that PLC lies on the positive side of D2 due to the frequency of its co-occurring linguistic features, it barely holds the "rational and affective appeal" [49] required in order to achieve the purpose of argumentative writing, i.e. to persuade the reader towards a certain point of view.

Previous studies on L2 essay writing have somewhat similar results. Hinkel [50] states that L2 writer's essays are "primarily knowledge telling, full of recounts of personal experience and exemplification without argumentation" and rely on less academically valued linguistic features such as simple conjunctions etc. A proposed solution to enhance argumentative skills in learners would be to implement a genre-based approach towards writing pedagogy in which explicit instructions regarding writing conventions such as stance taking, appropriate organizational patterns, frequently used sentence structures and even statistically co-occurring linguistic features be given. For instance, the use of split auxiliaries may be explained explicitly as an acceptable form since it creates the least disruption in the flow of language. This may be backed up with references from grammar books as Murphy's [51] which clearly endorses non-native writer's use of placing adverbs after the auxiliary, and so on.

As far as ESL and EFL sub-varieties are concerned, PLC scores lowest when compared with both, in conformity to the results of the overall comparisons of the four groups, as has already been discussed. Some of the sub-varieties show similarity in the amount of argumentation in the text; China is similar to Indonesia and Korea, and Hong Kong to Philippines.

CONCLUSION

The research aimed to explore variation across PLC and three different varieties of World Englishes; ENL, ESL, and EFL, further divided into their respective sub-varieties. The comparisons were made using multi-dimensional analysis technique in the light of two significant dimensions, viz. Involved versus Informational Production, and Overt Expression of Persuasion/Argumentation. The variations on these dimensions are complementary in determining some linguistic characteristics of argumentative essay writing of Pakistani learners. The first dimension revealed PLC text as informational in contrast to the other three texts, more specifically to ENL text, which showed high levels of involvement. However, its comparisons with the sub-varieties gave a slightly different picture in terms of ESL texts, as two of the ESL sub-varieties, Hong Kong and Singapore, are found to be more informational than involved. Features of informational texts therefore are highly attributive of ESL texts, although exceptions are there. In terms of expression of persuasion and argumentation, Pakistani learner essays present the least amount of arguments as compared to the other varieties and sub-varieties. Several reasons have emerged from a glance at the previous studies made on Pakistani learner writing, the crux of which is that it has more to do with the argumentative skills development of the learners than with their stylistic preferences, which has been interpreted as lack of proper essay writing pedagogy in Pakistani educational system. The research not only offers

insight to syllabus designers, material developers, and teachers regarding essay writing practices of Pakistani learners but also offers suggestions for improvised teaching in this regard.

REFERENCES

- Biber, D., "Variation across speech and writing." Cambridge University Press (1988)
- Siepmann D., "Discourse Markers across Languages. A contrastive study of second-level discourse markers in native and non-native text with implications for general and pedagogic lexicography." London: Routledge (2005)
- Uzar R., "A Corpus Methodology for Analysing Translation," In *Cadernos de Tradução* 2002 Universidade Federal de Santa Catarina. 237-265 (2002)
- Kaszubski P., "Enhancing a writing textbook: a national perspective," *Learner English on Computer*. London & New York: Addison Wesley Longman, 172-185 (1998)
- Tono, Y., "Multiple comparisons of IL, L1 and TL corpora: the case of L2 acquisition of verb subcategorization patterns by Japanese learners of English". *Corpora and language learners*, 17: 45-66 (2004)
- He, A., "A Corpus-based Analysis of Middle School Students' English Spelling Errors," *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching. Language Learning and Language Teaching*, Amsterdam & Philadelphia: Benjamins (2002)
- Granger S., "Use of Tenses by Advanced EFL Learners: Evidence from an Error-tagged Computer Corpus," *Out of Corpora - Studies in Honour of Stig Johansson*. Amsterdam & Atlanta: Rodopi, 191-202 (1999)
- Granger S., "Error-tagged learner corpora and CALL: a promising synergy," *CALICO (special issue on Error Analysis and Error Correction in Computer-Assisted Language Learning)* 20(3): 465-480 (2003)
- Van Rooy B. and Schäfer L., "Automatic POS tagging of a learner corpus: the influence of learner error on tagger accuracy," In Archer et al. (eds.), 835-844 (2003)
- Flowerdew J., "A taxonomy of learner errors in signalling nouns," Paper presented at ICAME 26 (*International Computer Archive of Modern and Medieval English*) - AAACL 6 (*American Association of Applied Corpus Linguistics*), University of Michigan (2005)
- Granger S., "Prefabricated patterns in advanced EFL writing: collocations and formulae," *Phraseology: theory, analysis and applications*. Oxford: Oxford University Press, 145-160 (1998)
- Kaszubski P. "Tracing idiomaticity in learner language – the case of BE," *Proceedings of the Corpus Linguistics 2001 Conference*, Lancaster University (UK), 312-322 (2001)
- Lorenz G., "Adjective Intensification - Learners versus Native Speakers. A Corpus Study of Argumentative Writing," *Language and Computers: Studies in Practical Linguistics* 27, Amsterdam & Atlanta: Rodopi (1999)
- Osborne J., "Articles and non-count nouns in learner English: Perception and production," *Practical Applications in Language and Computers (PALC 2003)*. Frankfurt: Peter Lang. 359-369 (2004)
- Damascelli A., "Argumentative Essays Written by English and Italian University Students. Aspects of Contrastive Rhetoric: the Use of Logical Connectors," Unpublished MA dissertation, University of Torino, Italy (1999)
- De Cock S., "Preferred sequences of words in NS and NNS speech," *Belgian Journal of English Language and Literatures (BELL), New Series* 2: 225-246 (2004)
- Eriksson, A., "Tense and Aspect in Learner Writing. Advanced Swedish Learners' use of Tense and Aspect in English Argumentative Text", Doctoral thesis, University of Gothenburg (2004)
- Altenberg B. and Granger S., "The grammatical and lexical patterning of make in native and non-native student writing," *Applied Linguistics* 22(2): 173-194 (2001)
- Nesselhauf N., "The Use of Collocations by Advanced Learners of English and Some Implications for Teaching," *Applied Linguistics* 24(2): 223-242 (2003)
- Altenberg B. and Tapper M., "The use of adverbial connectors in advanced Swedish learners' written English," *Learner English on Computer*. London & New York: Addison Wesley Longman, 80-93 (1998)
- Lorenz G., "... and state your reasons clearly: causal links in native and non-native discourse," *Anglistentag 1998 Erfurt. Proceedings*. Trier: Wissenschaftlicher Verlag, 61-78 (1999)
- McEnery T. and Kifle N.A., "Epistemic modality in argumentative essays of second-language writers," *Academic Discourse*. London: Longman, 182-215 (2002)
- Virtanen, T., "The progressive in NNS and NS student compositions: evidence from the International Corpus of Learner English," *Corpus-based Studies in English. Papers from the seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17)*, Stockholm, Amsterdam: Rodopi, 299-309 (1997)
- Ijaz, M.T., Mahmood, M.A., and Ameer, A., "A Corpus Based Study of the Errors Committed by Pakistani Learners of English at Graduation Level," *Journal of Education and Practice*, 159-162 (2014)
- Mahmood, R., "A Lexico-grammatical Study of Noun Phrase in Pakistani English," PhD thesis, Department of English, Bahauddin Zakariya University Multan (2009)
- Mahmood, M.A., "A Corpus Based Analysis of Pakistani English" PhD thesis, Department of English, Bahauddin Zakariya University Multan (2009)
- Brown, P. and Fraser, C., "Speech as a marker of situation," *Social markers in speech*. Cambridge: Cambridge University Press, 33-62 (1979)
- Biber, D. and Conrad, S., "Register, Genre, and Style," New York: Cambridge University Press (2009)
- Perez-Paredes, P., and Sanchez-Tornel, M., "A multidimensional analysis of learner language during story reconstruction in interviews." *Learner Corpora in*

- 424 ISSN 1013-5316;CODEN: SINTE 8 Sci.Int.(Lahore),28(4),413-415 ,2016
- Language Testing and Assessment*, London: John Benjamin Publishing Company, 141-162 (2015)
30. Götz, S. & Schilk, M., "Formulaic sequences in spoken ENL, ESL and EFL. Focus on British English, Indian English and learner English of advanced German learners," *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*, Amsterdam: John Benjamins, 79-100 (2011)
 31. Van Rooy, B., "A multidimensional analysis of student writing in Black South African English," *English World-Wide*, **29 (3)**: 268-305 (2008)
 32. Van Rooy, B. and Terblanche, L., "A Multidimensional Analysis of a Learner Corpus," *Corpus Linguistics: Refinements and Reassessments*, Brill/Rodopi, 239-254 (2009)
 33. Grabe, W. and Biber, D., "Freshman student writing and the contrastive rhetoric hypothesis," *Paper presented at SLRF7*, University of Southern California (1987)
 34. Terblanche, L., "A comparative study of nominalisation in L1 and L2 writing and speech," *Southern African Linguistics and Applied Language Studies* **27(1)**: 39-52 (2009) DOI: 10.2989/SALALS.2009.27.1.4.752
 35. Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E., "Longman Grammar of Spoken and Written English," Essex: Longman Publishers, (2007)
 36. Xiao, R., "Multidimensional analysis and the study of world Englishes," *World Englishes*, **28 (4)**: 421-450 (2009)
 37. Cao, Y., & Xiao, R. "A multi-dimensional contrastive study of English abstracts by native and nonnative writers," *Corpora*, **8(2)**: 209-234 (2013) DOI: [10.3366/cor.2013.0041](https://doi.org/10.3366/cor.2013.0041)
 38. Crossley, S. A., Allen, L., & McNamara, D. S., "A multidimensional analysis of essay writing: What linguistic features tell us about situational parameters and the effects of language functions on judgments of quality," *Multi-Dimensional Analysis, 25 years on: A Tribute to Douglas Biber*. Philadelphia, PA: John Benjamins, 197-238 (2014)
 39. Egbert, J. "Style in nineteenth century fiction: A Multi-Dimensional analysis," *The Scientific Study of Literature*, **2(2)**: 167-198 (2012)
 40. Egbert, J. "Student perceptions of stylistic variation in introductory university textbooks," *Linguistics and Education*, **25**: 64-77 (2014).
 41. Egbert, J., "Publication type and discipline variation in published academic writing - Investigating statistical interaction in corpus data." *International Journal of Corpus Linguistics*, 1-29 (2015)
 42. Kachru, B. B., "Institutionalized second language varieties," *The English language today*, 211-226 (1985)
 43. Shamim, F., "English as the language for development in Pakistan: Issues, challenges and possible solutions," *Dreams and Realities: Developing Countries and the English Language* (2011)
 44. Mansoor, S., "Language Planning in Higher Education - Issues of Access and Equity," *The Lahore Journal of Economics*, **8(2)**: 17-42 (2003)
 45. Rahman, T. "Pakistani English," *Islamabad: NIPS*, Revised edition of 1990 edition, 1-100 (2014)
 46. Hinkel, E., "The effects of essay topics on modal verb uses in L1 and L2 academic writing," *Journal of Pragmatics*, **41**, 667-683 (2009)
 47. Russell, M. K. "A Comparison of Linguistic Features in the Academic Writing of Advanced English Language Learner and English First Language University Students," Masters Thesis, Portland: Portland State University (2014)
 48. Imtiaz, Z. and Mahmood, A., "Genre Analysis of Argumentative Essays of Pakistani EFL Learners." *Journal of Education and Practice*, 95-100 (2014)
 49. Connor, U. and Upton, T. A., "Discourse in the Professions - Perspectives from corpus linguistics," *Studies in Corpus Linguistics*, John Benjamins Publishing Company, 16-334 (2004)
 50. Hinkel, E., "Second language writers' text: Linguistic and rhetorical features," *Mahwah, NJ: Lawrence Erlbaum Associates, Inc.* (2002)
 51. Murphy, R., "Essential Grammar in Use: A Self-Study Reference and Practice Book for Elementary Students of English : with Answers," Cambridge, U.K.: Cambridge University Press (1997)