Fach: Englische Philologie

A corpus based comparison of variation in online registers of Pakistani English using MD analysis

Inaugural-Dissertation

zur Erlangung des akademischen Grades „Dr. phil."

an der

Westfälischen Wilhelms-Universität, Münster (Westf.)

vorgelegt von

Muhammad Shakir

aus Pakistan

WiSe 2019/2020

Acknowledgements

Summary

This is a corpus-based study of online registers of Pakistani English, i.e. various types of blogs (individual, news, new media, and technology) and interactive registers (blog comments, Facebook groups, Facebook status updates, and tweets). The study utilises MD analysis – a comparative and quantitative approach to study register variation – to compare the online registers of Pakistani English with similar offline registers (face-to-face conversations, talk shows, interviews, columns, and news reports) and similar registers from U.S. English. Three types of analyses have been performed in this study. The situational analysis shows that Pakistani text producers are generally younger and their communicative purposes include asking for help related to business or study and creative writing. The linguistic analysis shows that generally the discourse produced in Pakistani online registers is formal, literate, and contains lexico-grammatical and semantic features that indicate the presence of abstract and evaluative information. The analysis of code switching instances shows that discourse level items (e.g. tag questions, discourse particles, politeness markers, honorifics, and religious expressions) and single or multi word lexical items related to local culture (e.g. kinship terminology, edibles and wearables etc.) are most commonly used types of resources borrowed form indigenous languages. The study concludes that Pakistani English online registers are limited in terms of communicative functions as compared to their U.S. counterparts. At the same time, Pakistani users show an indigenous colour by using discourse level and lexical resources from local languages in English.

Table of Contents

List of Tables

List of Figures

Abbreviations

Blogs Ind.	Individual blogs
Blogs Tech	Technology blogs
BlogsNM	New media blogs
CA	Cluster analysis
CDA	Canonical discriminant analysis
CMC	Computer-mediated communication
EFA	Exploratory factor analysis
F2F	Face-to-face conversations
FB	Facebook
GloWbE	Global Web-based English Corpus
ICE	International Corpus of English
KW ANOVA	Kruskal Wallis analysis of variance
MCA	Multiple correspondence analysis
MD analysis	Multidimensional analysis
PCA	Principal component analysis
PK	Pakistani English
SFL	Systemic functional linguistics
US	American English

Chapter 1 Introduction

1.1 Background

It was the year 2004 when the first Pakistani English blog directory PakPositive.com was established. The internet was already a decade or so old in the country, with most of the big cities having basic to broadband internet connectivity and many smaller cities also getting connected to the global network. The concept of blogging and free blog making platforms like Blogspot.com and Wordpress.com had been becoming known to the younger generation (in their 20s at that time) since the early 2000s. Now they had a platform where they could post as guest authors and also connect with the budding blogging community in Pakistan by posting a short introduction to themselves and their blog. The number of blogs kept on increasing until approximately 2010, with blogs ranging from topics on self, surroundings, and family to social issues and religion to very specific topics like beauty and cooking. Others started multi-writer blogs where they could post on topics like social issues or on issues about a particular city.

Around the year 2008 the horizon of the Pakistani English blogging community started expanding with the introduction of the first technology blog ProPakistani.com. Such blogs aimed to provide content on a particular topic (technology help and later on technology news in this case) and earn money by showing ads in this process. Around the same time, English newspapers in the country (though some of them were already accessible through their websites) started engaging with their readers by introducing comments sections under news articles and by creating dedicated blog sections where writers could send their English posts. Following the English newspaper *Dawn*, other English newspapers and Urdu news channels started English blog sections on their websites. Many of the bloggers who had initially registered on PakPositive

got engaged in these newly emerged blogging platforms with far wider outreach in terms of audience.

The beginning of the second decade of the 2000s also saw a rise in the popularity of social media platforms like Facebook and Twitter in Pakistan. The social media users not only included 'old' bloggers but also new users who were now teenagers or in their 20s. Unlike the initial blogging community, this set of users was much bigger and these users were not necessarily bloggers. They were much more familiar with mobile internet and smartphones. Consequently, more and more users started creating a digital reflection of their everyday lives owing to the increasing accessibility to the internet. Instead of writing long essay-like blog posts, these users resorted to short texts in the form of 'status updates' on Facebook or 'tweets' on Twitter, or in other words 'microblogging'. By end of the second decade of the 2000s, a sizeable portion of the Pakistani population engages on microblogging platforms in indigenous languages like Punjabi, Sindhi, Pashto, the lingua franca Urdu, and of course in English. In general, their microblogging posts in English cover topics related to business, jobs, study abroad, general and specific help on Facebook, or simply consist of reactions to the political situation in the country on Twitter.

The internet is a new medium of communication, which is different from the traditional mediums of communication using language, i.e. the spoken medium and the written medium. It creates a whole new wealth of situations and forms of communication that are borderless due to the global nature of the internet itself (Deshors and Gilquin, 2018). The English language is one of the most dominant languages of communication on the internet. As it has been mentioned above, Pakistanis actively use the internet as a medium of communication in indigenous languages as well as in English. Traditionally, the English language in Pakistan has been used in

the written medium as the language of administration, judiciary, education, and business communication (Haque, 1983). More recently, it has been being used in the spoken medium in TV and radio as well (cf. section 2.2). The internet is a much more accessible and democratic medium as compared to the traditional mediums of English communication in Pakistan. Until now the use of English on the internet has largely remained unexplored in the context of Pakistan. Hence, the general aim of this study is to look at this phenomenon in a systematic way.

The interest in English on the internet or English computer-mediated communication (CMC) is obviously not new, e.g. scholars like Susan C. Herring have studied this phenomenon since the 1990s (Herring, 1996). The production of English texts by Pakistani and other postcolonial (or non-native) users of English has also caught the interest of scholars of English linguistics, and more specifically those from the paradigm of World Englishes. For example, datasets have been compiled to study variation in the English language produced by users of various countries of the world (Mukherjee and Schilk, 2012; Davies and Fuchs, 2015) that have resulted in studies like Bernaisch, Gries, and Mukherjee (2014) that also include Pakistani English. However, these studies generally do not look at the communicative functions and uses associated with English communication by Pakistani internet users in a nuanced and detailed manner. Lastly, it is important to study digital varieties of English from different geographical contexts – like the present one – to provide insights for theorisation and modelling of World Englishes, as emphasised by Deshors and Gilquin (2018, p. 284).

1.2 Method, Research Questions, and Scope

To describe the communicative functions and uses associated with internet-based communication of Pakistani users, this study utilises the multidimensional (MD) analysis framework developed by Biber and colleagues (e.g. Biber, 1988; Biber and Conrad, 2009). This

framework can be used to operationalise the recommendation of Deshors and Gilquin (2018) to include 'communicative event' in the study of English varieties (on the internet). The concept of register – a variety of language that varies according to the situation of use – is central to this framework. The situational variation manifests in over- or underuse of certain pervasive and frequent lexico-grammatical and semantic features of language that have communicative significance. The linguistic analysis of these features is complemented with situational analysis – e.g. participants, topics, modes of communication etc. – to arrive at functional interpretations of the register varieties under study (cf. section 2.3).

In practical terms, it is proposed here that online (internet-based) and similar offline (non-internet-based) registers (cf. section 4.2) of Pakistani English be studied in comparison to U.S. English (cf. subsection 3.1.2 for a discussion on the selection of U.S. English as the native variety for comparison). The following research questions are formed in this regard:

1. What are the situational differences between online and similar offline registers of both regional varieties?

2. What are the dimensions of variation identified by MD analysis in the registers under study?

3. What are the differences between online and similar offline registers of both regional varieties as per the resulting dimensions?

4. What are the regional differences between Pakistani English registers and their counterparts in U.S. English as per the resulting dimensions?

5. Are there any differences related to the gender of the text producer in online registers of both regional varieties?

6. Is there any chronological variation in technology blogs of both regional varieties as per the dimensions identified?

7. Which text types exist in online registers of both regional varieties as per the resulting dimensions?

8. What is the general communicative purpose that differentiates both regional varieties?

9. Which functions can be associated with the code switching to Urdu and other regional languages in texts from online registers of Pakistani English?

10. How can the linguistic variation be explained in terms of communicative functions, situational differences, and usage?

By answering the above research questions, the study will also contribute to the discussion in the theory and scholarship of World Englishes regarding the nature of variation and change in English in the era of the internet.

In terms of data, the study is limited to those registers that were available for Pakistani users of English on the internet and otherwise (chapter 3 provides details in this regard). The study uses corpus linguistics techniques, more specifically the framework of MD analysis. The linguistic analysis performed in the study (chapters 5 and 6) is quantitative in nature. Lastly, the situational analysis of the data (chapter 4) is observational in nature and does not utilise tools like interviews and questionnaires.

1.3 Overview of the Thesis

The thesis is divided into seven chapters, including the current one. A brief overview of each upcoming chapter is provided below.

Chapter 2 reviews previous relevant research work to provide a context for the research conducted in this thesis. The theory and models related to World Englishes are discussed in the

first section of this chapter. The second section details the relevant historical information about Pakistani English, e.g. educational policy, language politics, and previous research on the variety. The third section reviews relevant research work on the MD analysis framework and relates it to the study of World Englishes. The last section reviews previous related research on internet-based registers with a focus on MD studies.

Chapter 3 is dedicated to the data collection procedures involved in the study. The sources of data collection, the editing and corrections involved in this regard, and the sampling related decisions are discussed in this chapter. Additionally, the chapter also guides towards relevant appendixes which provide information, for example, on the sources or websites used to collect the corpus. Only one issue related to the method of analysis, i.e. the selection of features for MD analysis, is discussed at the end of this chapter. Other details regarding the methodology used in the analysis are discussed in the relevant chapters or in their particular sections.

Chapter 4 is the first of three chapters that analyse the corpus in different ways. This chapter describes the situational features of the registers under study by implementing the first step of a three-step approach of register analysis introduced by Biber and Conrad (2009). A framework for situational analysis is developed by reviewing previous and relevant research. This framework is subsequently applied to describe the situational features of the registers in the corpus. The chapter also provides operational definitions of various registers and important situational differences among registers as well as the differences between similar registers of both regional varieties. The summary section of this chapter provides answers to Research Question (RQ) 1.

Chapter 5 is the main analysis chapter that reports on the linguistic analysis of the data (the second step of register analysis as per Biber and Conrad, 2009). There are three sections in

this chapter that use three quantitative techniques – namely exploratory factor analysis (EFA), cluster analysis (CA), and canonical discriminant analysis (CDA) – to analyse the data using the selected feature set already described at the end of chapter 3. Each section provides relevant background regarding the method of analysis and discusses relevant practical issues before proceeding to the actual analyses and results. The first section is related to MD analysis, which uses EFA for statistical analysis. Subsection 5.1.5 provides details regarding the dimensions identified and, in this process, also answers RQ2. The summary sections of subsections 5.1.6 and 5.1.7 answer RQs 3 and 4. RQs 5 and 6 are addressed in subsections 5.1.8 and 5.1.9. The overall summary of section 5.1 provides a comprehensive overview of the findings of MD analysis and reaffirms the answers to RQs 2-6. The text type analysis in section 5.2 answers RQ7. The application of CDA on the same data in section 5.3 addresses RQ8. The chapter ends with an overall summary of findings of the linguistic analysis, highlighting the differences between online and offline registers as well as the differences between the regional varieties.

Chapter 6 is the third and last chapter that is related to data analysis. It is a supplementary analysis to the main linguistic analysis performed in chapter 5. The chapter looks at the multilingual nature of English communication of Pakistani internet users by examining code switching in predominantly English texts. The chapter provides necessary background by reviewing previous and relevant research work on code switching. After developing a framework of analysis, it provides qualitative results of various functions of code switching instances performed by the users in these data. The findings of this chapter address RQ9.

The last chapter, i.e. chapter 7, discusses the results of previous chapters and concludes the thesis. The first main part of the chapter provides a summary of the results from the previous three chapters. The results are synthesised and interpreted with regards to each register category

in Pakistani English and the variety as a whole. The summary of results partially reiterates the answers to RQs 1-9. The interpretation of these results, which is step three of Biber and Conrad's (2009) register analysis, addresses RQ10. The next main section of the chapter relates the findings of this study with the theory and scholarship of World Englishes. At the end, the limitations of the study are discussed and possible future directions are proposed.

Four appendixes are also part of the thesis. Appendix I provides additional graphs that are related to the linguistic analysis in chapter 5. Appendix II provides details regarding the lexico-grammatical and semantic features used in chapter 5. Appendix III consists of a list of non-standard spellings and their corrections that were applied to the corpus to assist the grammatical tagger. Appendix IV lists the websites and sources that were used to compile the corpus of Pakistani and U.S. English online registers.

Chapter 2 Theoretical and Research Background

The aim of this chapter is to ground the research work being presented here in theory and previous related research studies. In this process the chapter identifies research gaps in previous works to rationalise the present study. Four broader themes are discussed and synthesised together to achieve the above-mentioned goals. The first theme is related to the spread of English and its study under the paradigm of World Englishes. The classification or models of the spread and evolution of World Englishes and their implications for English on the internet are discussed. Pakistani English is the second theme, which is studied under the paradigm of World Englishes. A brief sociolinguistic and socio-political history of Pakistani English and previous research on the variety is reviewed to contextualise the present study. The third major topic is related to the methodological framework used in this study, i.e. MD analysis (Biber, 1988). Issues related to situational variation, text linguistics, the distinction between register, genre, style, and text type, and the relevance of this methodological framework to World Englishes are explained. The last main theme is English communication on the internet. Previous research related to internet registers is reviewed with a special focus on MD analysis and/or World Englishes. Section summaries are provided at the end of each main section to synthesise the main thoughts presented.

2.1 Theoretical Underpinnings: The Paradigm of World Englishes

2.1.1 Introduction

In a few centuries the English language has spread from the tiny British Isles to all over the world. This unprecedented phenomenon has subsequently triggered linguistic and other research scholars to study it under various topics, e.g. historical linguistics, applied linguistics, language contact, macro sociolinguistics among other (Mesthrie and Bhatt, 2008, p. 1).

Kirkpatrick (2014) observes that though the study of varieties of English is old, the field of World Englishes is quite young. He considers Braj Kachru as a key figure in establishing this subdiscipline of sociolinguistics (p. 33). As Kirkpatrick puts it, Quirk (1985) advocated for a single standard English based on native varieties of English, while Kachru (1985) favoured many varieties of English that are equal in status. This debate later on resulted in the establishment of the field of World Englishes.

Kirkpatrick (2007, p. 5) identifies a number of sociolinguistic concepts central to the debates of the World Englishes paradigm. The first of these concepts is the distinction between native and nativised varieties of English. The former are native English varieties like British and American English, while the latter are English as a second language (ESL) varieties like Indian and Pakistani English. Kirkpatrick (2007, p. 7) opines that every variety of English that is used by a distinguishable speech community should be considered a native variety, so there are no purer or better varieties. The second central issue is the distinction between native and non-native speakers of English. As Kirkpatrick observes, it is hard to define the term native speaker in certain contexts, for example in the case of bilingual children. This term in the context of World Englishes, hence, should be avoided. The third issue is the perceived increase in unintelligibility of English with increased differentiation and pluralisation, as argued by purists like Quirk (1985). However, as Kirkpatrick (2007) sees it, speakers of English can choose to be intelligible and communicative when they want, or otherwise can use their localised variety of English to emphasise their identity. Kachru (1988) has also identified several theoretical, sociolinguistic, and pedagogical issues that are now considered central to discussions of World Englishes.

2.1.2 Models of World Englishes

As discussed previously, the English language has spread far and wide in the world. In this process, many localised varieties of English and even new languages, i.e. English-based pidgins and creoles (Kirkpatrick, 2007, p. 13), have come into existence. This has led some authors to coin new terms to describe this phenomenon; for example Mesthrie and Bhatt (2008, p. 3) call it 'English Language Complex'. Apart from the above discussed issues related to World Englishes, a significant portion of theoretical aspects of this paradigm is devoted to constructing theories to better understand variation, change, and spread of the English Language Complex. These theories, or in the words of Kirkpatrick (2007, p. 27), 'models' are classifications which attempt to provide explanations regarding the differences in the uses of the English language in various countries of the world. Other models go one step ahead to identify and generalise the evolutionary processes involved in the development of a variety of English. The former can be called static models of English, while the latter can be considered dynamic models. A brief overview of these models of World Englishes is necessary to provide a theoretical context for the present study. The following subsections provide such an overview by moving from static to dynamic models.

    2.1.2.1 Kachru's three concentric circles model.

Braj Kachru's (1985) three concentric circles model is one of the earliest and most influential classifications of English and its varieties around the world. As per Kirkpatrick (2007, p. 27), it extends an earlier classification mainly used in the language teaching world, i.e. ENL (English as a native language), ESL (English as a second language), and EFL (English as a foreign language). However, the Three Circle model introduces the concept of plurality of the

English language, i.e. Englishes instead of a single English. A graphical representation of this model is presented in figure 2.1.
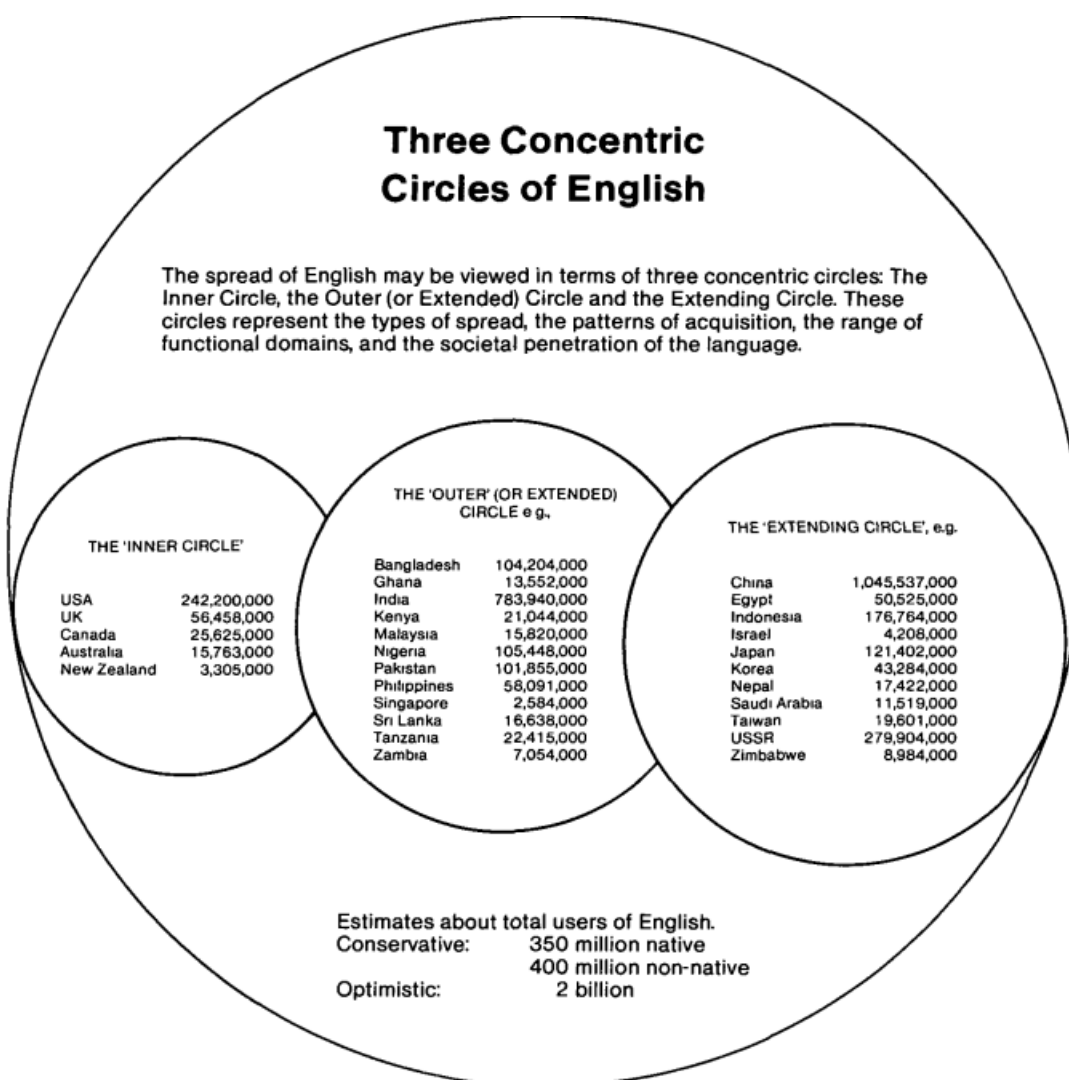
Figure 2.1 Kachru's Three Circle model (Kachru, 1988, p. 5)

One of the main focuses of this model is to highlight the importance of the outer circle or 'institutionalised' varieties of English (Kachru, 1985). He (pp. 12-13) lists the following characteristics of these countries and the varieties of English used by them:

- The countries and regions within the outer circle were colonised by native speakers of English, i.e. Britishers.

- Two main characteristics of these countries can be identified: Firstly, the linguistic repertoire of the users of outer circle consists of English along with one or more other languages or 'codes'; Secondly, English retains an important place in language related policy and decision-making of these multilingual nations.

- These institutionalised varieties have three functional properties:

    o The functions associated with the English language can "be considered traditionally 'un-English' cultural contexts".

    o The English language is used in many domains, but the competence of users can vary. Moreover, these domains can consist of local and international communicative situations.

    o The users of these varieties have started producing literature in English in various forms like essays, short stories, poetry, and novels. In this process, they have created nativised literary traditions.

Despite the paradigm shift introduced by this model, it also faces certain difficulties. As Kachru (1985, p. 14) also observes, countries like Jamaica and South Africa are difficult to classify in the three concentric circles and the functions associated with English in these countries are relatively complex. The other difficulty is the distinction between outer circle and expanding circle. According to the model, whereas English is used in an institutionalised and localised way in the outer circle countries, the main function of English in the expanding circle countries is the language of international communication or lingua franca. Authors like Pennycook (2009) point out that the model is too much bound by the history and geography of nation states. Moreover, Pennycook thinks that in the process of focusing on localised uses of

English, the continuously negotiated status of English in English as a lingua franca (ELF) contexts is ignored.

2.1.2.2 Going beyond geographical and historical boundaries.

Yano (2009) advocates for a transition from "geography-based model of English to person-based model of English speakers" (p. 212) due to increased global mobility, demographic changes in the inner circle due to immigration, and second language speakers developing native speaker-like intuitions etc. A graphical representation of Yano's three-dimensional cylindrical model is presented in figure 2.2.

As the figure explains, Kachru's three concentric circles are used as a starting point, with the expanding circle being the biggest in terms of the number of users. The circles are ordered in terms of proficiency level, which is shown by the central line. Yano's model takes into account English for specific purposes as well as English for general purposes. The ideas of intraregional standard English as well as English as an international language are also incorporated in this model. To be able to communicate at various levels, a user has to master various proficiency levels starting from the expanding circle proficiency to an adult native user's proficiency level and lastly the ability to communicate internationally.

Pennycook (2009) also advocates for a model of English that moves away from geographical boundaries. He draws on Yano's (2001) multi-circle 3D representation of varieties of English to develop a model of his own. He understands 'variety' to be a fluid concept instead of a solid or rigid object that is limited by geographical boundaries. Additionally, he also attempts to incorporate the contextual use as well as users of language in his model.
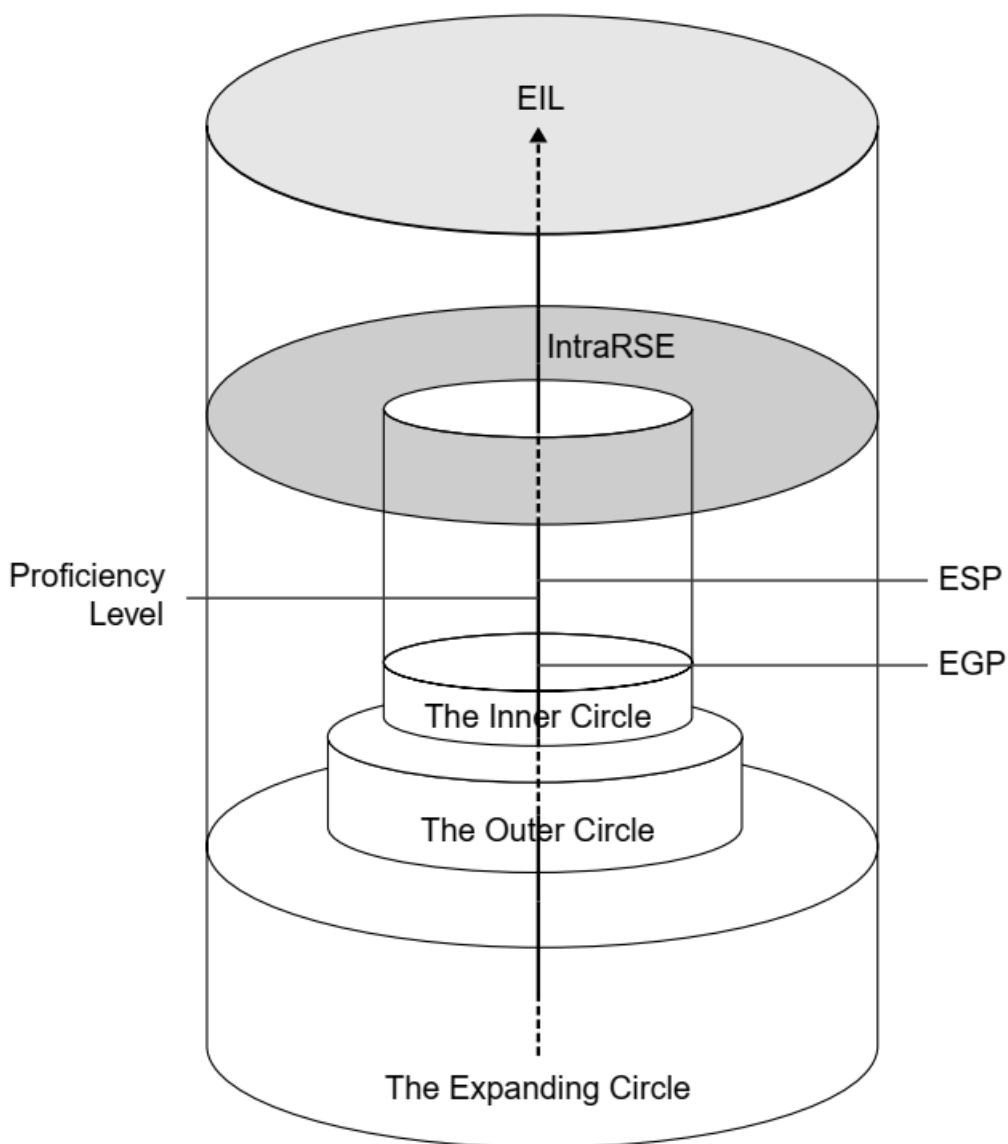
Figure 2.2 Yano's three-dimensional cylindrical model (Yano, 2009, p. 216)

As figure 2.3 explains, the top surface consists of language varieties as an open space without any restrictions of national boundaries. The vertical plane takes into account contextual use of language including non-verbal communication and sign language. Lastly, the third dimension is related to the idiolectal use of the individual user whose linguistic repertoire consists of many codes or languages along with English, with each code connected to certain domains, activities, and ideologies.

Language resources:
The interlingual
plane of communicative
repertoires

Language contexts:
The colinguistic use
of English in
space and time

Speaker location:
The ideolinguistic
uptake of meaning
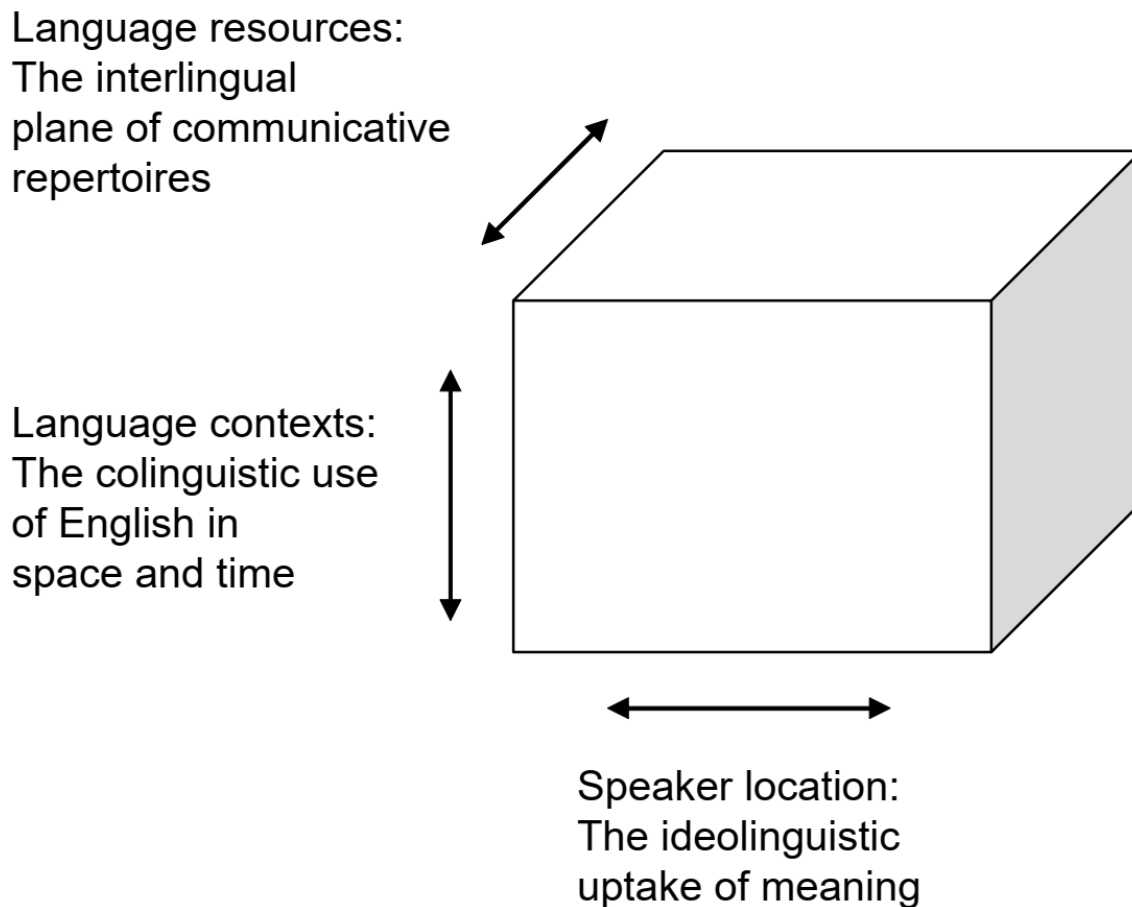
Figure 2.3 Pennycook's 3D transtextual model of English use (Pennycook, 2009, p. 204)

Mahboob (2014a) also proposes a three-dimensional representation of English communication, which consists of users and uses of English(es), and channels of communication. The first dimension focuses on sociolinguistic roles and relationships of the users interacting, including age, gender, social class etc. The second dimension is related to the purpose of language, i.e. if it is related to 'everyday/casual discourses' or related to 'specialised/ technical discourses' (p. 268). The third and last dimension includes the modes of communication, e.g. 'aural, visual, and mixed' modes (p. 269). He identifies eight domains of usage, of which the first four are related to local or localised usage of language: local written casual, local oral casual, local written specialised, and local oral specialised. The next four domains are related to global or international use of language, where localised patterns of

language usage are not utilised: global written casual, global oral casual, global written specialised, global oral specialised. The users of these domains, especially the last four domains related to global communication, construct communities of practice that exceed local geographical boundaries.

The three models discussed in this subsection generally originate from the tradition of studying English as a lingua franca (ELF). Seidelhofer (2009, p. 236) observes that ELF research mainly focuses on Kachru's expanding circle, while World Englishes research has been more concerned with the outer circle. ELF practitioners, instead of focusing on geographical communities, promote the concept of communities of practice "characterized by 'mutual engagement' in shared practices, taking part in some jointly negotiated 'enterprise', and making use of members' 'shared repertoire' (Wenger 1998:72ff.)" (Seidelhofer, 2009, p. 238). To concur with Seidelhofer (p. 239), in the 21st century these communities can be local or global, face-to-face or electronic, within inner, outer, or expanding circles or between them. Consequently, both types of models can complement each other to provide a better understanding of variation in the English language.

2.1.2.3 Globalisation and varieties of English.

Mair's (2013, 2016) World System of Englishes is a static classification of varieties of English around the world. Mair (2016) views the English Language Complex as consisting of a host of varieties that follow a hierarchical system and are ordered in groups and subgroups. In his view, these groups and subgroups – in a direct or indirect way – very efficiently bind together more than one billion users of English around the world. His hierarchical system includes hyper-central, super-central, central, and peripheral varieties of English. American English is considered the hyper central variety due to its global reach and influence through media. Super-

central varieties include standard Englishes like British English as well as non-standard English varieties like Jamaican Creole.

He uses findings from English communication on the internet (forums and blogs) to show that globalisation is weakening nation-based standard varieties. He opines that especially in diaspora groups standard national varieties (e.g. Nigerian Standard English) take a secondary position, while non-standard varieties (e.g. Nigerian Pidgin) gain covert prestige. Moreover, he argues that endonormativity (development of local standards) in outer circle countries is under threat from inside (e.g. due to problematic education system) and outside (due to the influence of super central varieties of English and the hyper-central variety, i.e. American English) (Mair, 2016).

2.1.2.4 Internet communication and post-varieties approach.

van Rooy and Kruger's (2018) model is similar to Pennycook (2009) and other user-based models discussed above. However, they specifically focus on internet communication. They feel that "the empirical challenge to models of Englishes is the extent to which other languages, non-standard Englishes, CMC, hybridity, multiplex identities, and trans-nationalism can be incorporated" (p. 81). To come up with a satisfactory model in this regard, they study online English comments of South African users. Their analysis of the most frequent lexical items reveals that English resources are used most frequently. Further resources include the conventions of English CMC, national and international non-standard forms of English, local innovations (in English), and resources from indigenous languages (p. 103). Their proposed model is a very simple one with three aspects: *resources* that are available to the user based on their knowledge of expressions, words, and constructions instead of a variety as an object,

selection processes that the user applies based on their linguistic and sociolinguistic knowledge, and the textual *diversity* that results from communicative interactions (pp. 103-104).

van Rooy and Kruger in many ways agree with Seargeant and Tagg's (2011) post-varieties approach to internet communication. Seargeant and Tagg analyse MSN chats of internet users who use a mixture of English and Thai in their interactions. In their view, such interactions are neither describable in the categorical terms of, for example, Kachru's three concentric circles, nor can they be classified "as a stable or emergent variety in [their] own right" (p. 510). Hence, they advocate for a post-varieties approach to describe this type of communication. However, they also acknowledge the usefulness of the concept of variety when "the focus is not simply on strategies of communication but also on cultural and political identity" (p. 512) of a community. To sum up, the Kachruvian concept of pluri-centricity of English does not become completely irrelevant in the age of the internet. It, however, needs to be complemented by incorporating the issues of globalised connectivity, multiple influences on varieties of English, and hybridity of linguistic resources.

2.1.2.5 Explaining the evolution of Englishes.

Apart from static classifications, the other major aim of models of World Englishes is to explain the evolution of varieties of English. As Kirkpatrick (2007, p. 30) notes, there are many similarities in the stages explained by different scholars with overlapping terms/ concepts. Generally, the following stages can be deduced from various scholarly works as mentioned in table 2.1:

- English is introduced to a geographical region by settlers/ colonisers/ globalisation;
- The indigenous population (elites) initially follows foreign (British/ American) standards but local innovations also take place;

- The indigenous population starts recognising local innovations as local standards;

- An expansion in the function of English occurs;

- Further sub-varieties of English are born/ the usage of English either declines or stabilises.

Table 2.1 Developmental stages of varieties of English

Scholar	Phases				
Kachru (1992)	non-recognition	co-existence of local and imported varieties	recognition		
Moag (1992)	transportation	indigenisation	expansion	institutionalisation	(decline)
Schneider (2003, 2007)	foundation	exo-normnative stabilisation	nativisation	endonormative stabilisation	differentiation
Buschfeld & Kautzsch (2017)	**Postcolonial Englishes** foundation	exo-normnative stabilisation	nativisation	endonormative stabilisation	differentiation
	Non-Postcolonial Englishes foundation	exo-normnative stabilisation	nativisation	endonormative stabilisation?	differentiation?
Meer & Deuber (forth-coming)	foundation	exo-normnative stabilisation	nativisation	endonormative stabilisation/ multinormative stabilisation	differentiation

Adapted from Kirkpatrick (2007, p. 33)

As table 2.1 exhibits, Kachru looks at the process of evolution in three simple stages. Moag's five stages of development of varieties of English are based on Hall's (1962) model for the life-cycle of pidgin and creole languages. Moag uses English in Fiji as an example for his model. Schneider (2003, 2007) builds on these and other relevant works to construct a model of English which he calls the Dynamic Model. Schneider (2007, p. 21) thinks that the evolution of varieties follows a uniform process that can be generalised for all postcolonial Englishes. The

construction and reconstruction of identity, which is central to this model, is very closely related to two perspectives of communication: the perspective of the colonisers (settlers) and that of the colonised (indigenous population). He goes one step ahead to identify a set of factors (termed as 'forces' in later derivatives) that come into play at each stage of evolution (cf. table 2.2). He looks at these processes in mostly a linear fashion, i.e. extralinguistic factors like history and politics shape and reshape the identity of the respective groups, which in turn affect language attitudes and language use, and finally that results in local innovations in terms of lexis, grammar, phonology etc.

Table 2.2 Forces affecting the development of varieties of English

Scholar	Forces
Schneider (2007, p. 30)	'(1) extralinguistic factors (history and politics) (2) characteristic identity constructions (3) sociolinguistic determinants of language contact (language contact, language use, and language attitudes) (4) structural effects (at various linguistic levels)'
Schneider (2014, pp. 17-18)	'(1) language policy and English in education (2) attitudes to English (and possible impact on identities) (3) sociolinguistic conditions of using and learning English (4) structural consequences (features)'
Buschfeld & Kautzsch (2017, p. 114)	**Extra-territorial:** 'colonisation, language policies, globalisation, foreign policies, sociodemographic background' **Intra-territorial:** 'attitudes towards colonising power, language policies/ language attitudes, acceptance of globalisation, foreign policies, sociodemographic background'
Meer & Deuber (forth-coming)	**Progressive:** further the evolutionary process of a variety **Conservative:** hinder the evolutionary process of a variety **Translocal:** 'forces that transcend boundaries while relating to local contexts' (p. 17)

The Dynamic Model is certainly a very useful abstraction of the processes involved in the evolution of postcolonial Englishes, which takes into account the diachronic as well as the current status of such varieties of English (Buschfeld and Kautzsch, 2017). Schneider (2014) modifies the original four parameters/ factors affecting the development of English to explain expanding circle scenarios (cf. table 2.2). He, however, also states that the applicability of the

Dynamic Model in the era of globalisation of Englishes is limited and this phenomenon is "distantly related to what the Dynamic Model describes" (p. 27). Schneider further introduces the concept of 'Transnational Attraction', which he describes as "the appropriation of (components of) English(es) for whatever communicative purposes at hand, unbounded by distinctions of norms, nations or varieties" (p. 28). Issues like globalisation and the application of the model on non-postcolonial varieties of English etc. have been very well addressed in a later derivation of the model.

The Extra- and Intra-territorial Forces or EIF Model (Buschfeld and Kautzsch, 2017) is a derivative of Schneider's Dynamic Model. The model aims to explain the evolution of postcolonial as well as non-postcolonial Englishes. The five stages of development (cf. table 2.1) are the same for the former, but the last two stages are yet to be confirmed for the latter type of varieties. Moreover, the model expands the factors/ forces affecting the development of a variety through these stages by introducing the concept of extra-territorial and intra-territorial forces, i.e. the forces that affect a variety from within a country versus those from outside. As table 2.2 reveals, they add some additional factors like foreign policies and globalisation to Schneider's four parameters. The first force (colonisation) is not present for non-postcolonial Englishes. Additionally, they also integrate the static classification of ENL, ESL, and EFL along with the stages of development. Since the three-point scale should be considered a continuum, a variety can develop in both directions from EFL to ENL or from ENL to EFL or somewhere in between.

Buschfeld, Kautzsch, and Schneider (2018) further elaborate the EIF model. They discuss a third dimension which takes into account the idiolectal level and variety-internal variation. The applicability of the model on internet-based Englishes, along with other EFL scenarios, is also discussed. They note that digital Englishes are different from nation-based varieties, because the

speakers have mixed regional backgrounds and proficiency levels. At the same time, an integrated approach for both types of Englishes should be possible. They categorise the factors influencing digital communication in terms of extra- and intra-territorial forces – e.g. in case of online games, globalisation in the form of 'long-distance personal communication', and 'individual enjoyment' respectively (p. 38). However, the term 'territorial', in their opinion, should be avoided in the context of digital Englishes, so the forces should be renamed 'external' and 'internal' (p. 38). They suggest that "digital Englishes can follow similar developmental routes like nation-based varieties" (p. 39), but the extent of applicability of the EIF model in this context is yet to be confirmed.

Another recent derivation of the Dynamic Model comes from Meer and Deuber (forthcoming), who extend the EIF model by introducing a couple of new concepts. Their modified version of the model arises from research work on Anglophone Caribbean countries like Trinidad and Tobago. They are of the view that there are multiple forces and standards at play in this context, which creates an equilibrium of different norms. They draw on Mukherjee's (2007) concept of progressive and conservative forces (cf. table 2.2), which can hinder or further the evolution of a variety. In their view, sometimes it is not possible to distinguish extra- versus intra-territorial forces. They advocate for 'translocal forces' which transcend boundaries. Moreover, they introduce the concept of 'multinormative stabilisation' as an alternative to 'endonormative stabilisation' (cf. table 2.1) at the fourth stage of variety evolution. Though their model is based on the educational context in Trinidad and Tobago, it can be useful in other contexts as well, for example internet-based English communication.

2.1.3 Section summary and research implications

This section started with an introduction of the World Englishes paradigm. Afterwards an overview of Kachru's Three Circle model was presented. It is one of the first approaches which acknowledge the plurality of English, hence Englishes. However, the model is based on the nation state, which is criticised especially by ELF-based approaches. In later sections, usage-based models were reviewed, which focus on users and usage. They advocate focusing on communities of practice, which are based on criteria other than national boundaries. The issue of globalisation, though discussed in previous models as well, is much more central to Mair's World System of Englishes. The influence and power of the hyper-central (American English) and super-central varieties (e.g. British English) is also related to globalised connectivity. In subsection 2.1.2.4, the focus started shifting to internet-based communication and issues of hybridity, multilingualism, and a post-varieties approach. The Dynamic Model and its derivatives look at varieties of English in a more comprehensive way by accounting for the developmental stages of Englishes and the factors influencing the stages in detail. Non-postcolonial Englishes, internet-based Englishes, and accommodating usage-based varieties are major concerns of the EIF Model. Lastly, the concept of an equilibrium of multiple standards has also been discussed.

Central issues for further research and theorising in World Englishes, which can be derived from the above discussion, are identified by Deshors and Gilquin (2018) in a comprehensive manner. Computer-mediated communication (CMC) should be focused on as "[i]t has also led to the creation of many new genres, including emails, tweets or blogging" (p. 283). 'Communicative situations (or communicative events)' (p. 287) should become central to the process of theorisation and the research conducted for this purpose. "[T]he traditional EFL/ESL user has become, in the 21st century, a global communicator" (p. 288) who affects the

development of World Englishes as well. Hence, s/he should also be incorporated in the focus on communicative event/ situation. In terms of methodological implications, they advocate for the use of large corpora and sophisticated statistical techniques to arrive at reliable linguistic evidence for theorisation.

To conclude this section, it has been shown that the models of World Englishes have increasingly recognised the importance of English communication on the internet. Resultantly, scholars like Deshors and Gilquin (2018) have called for further research on this mode of communication. Secondly, it has been shown that the focus on user, usage, and communicative situation has also increased. At the same time, nation-based variety labels, though sometimes severely criticised, are not discarded completely due to their usefulness in the theorisation and description of Englishes. Hence, nation state-based varieties should be studied with a focus on usage and communicative event/ situation.

2.2 Pakistani English

After establishing the necessary theoretical background, this section provides a historical and research background of the variety under study, i.e. Pakistani English. The following subsections provide details in this regard.

2.2.1 Historical domains and usage of English

Pakistani English is an outer circle variety in Kachru's model of English that shares many characteristics of such varieties, as mentioned in subsection 2.1.2.1. After independence in 1947, the country faced a dilemma regarding the selection of an official language. As Haque (1983) explains, complexity of linguistic situation, absence of a neutral local language, and scarcity of resources to train the bureaucracy to use another official language led to the continuous use of

English. After about 35 years of independence, Haque (1983) details various uses of English in the public and private sectors of the country as follows:

> The anchorage of English in Pakistan is that the constitution and the body of the law is codified in English. As a consequence, judgement and precedents, rules and regulations, orders and instructions, standing procedures and other mechanisms of the functioning of the state, and major policy documents of the federal and provincial governments are in English; information, technological, economic, sociological, and statistical, is also largely available in English. English in Pakistan is more the language of Macaulay than of Shakespeare. The large industrial and business sector operates in English. Although introduced in this country through an historical accident, English has become a pattern of life, and its cultural influence continues to be strong. (p. 7)

2.2.2 English in language and education policy

The language and education policy regarding English has changed in the last 70 years according to civil/ military regime changes in the country. Various authors (Abbas, 1993; Mahboob, 2014b; Shamim, 2008) have documented it in detail. A summary is provided below:

- 1947-1971: Mahboob divides this period in two according to regime changes (1947-58, 1958-71). However, both Mahboob and Abbas agree that the newly independent nation could not agree on a national language. Resultantly, the status quo regarding the use of English in the public sector and education largely remained unchanged as compared to the colonial period.

- 1971-1977: This period consists of the first relatively stable civil rule in the country. The eastern part of the country was separated and a new constitution was approved in 1973. Urdu was given the status of official language, which would replace the current official

language, i.e. English, in 15 years. At the same time, the diversity of local languages was also promoted by establishing 'the Institute of Folk Heritage' (Abbas, 1993, p. 150). The tensions between Sindhi speaking natives and Urdu speaking muhajirs (immigrants from India at the time of partition in 1947) reaffirmed the neutral status of English in this period.

- 1977-1988: The first civilian government was soon toppled by the military, which established martial law for the next 11 years. The military dictator promoted Islamisation of the country by introducing Arabic as a subject in middle schools. At the same time a nationalisation campaign was launched by implementing Urdu as a medium of instruction in public-sector schools. Elite class schools were, however, largely spared from the Urdu medium policy (Mahboob, 2014b). At the same time, the country also received aid from foreign agencies like the British Council to start teacher training programs through local universities and public organisations (Abbas, 1993).

- 1988-1999: Shamim (2008, p. 238) calls this the period of 'democratisation of English'. The almost 10 years of democratic regimes in the country did not lead to any big changes in the language policy. In terms of the teaching of English, public-sector schools were initially given an option to introduce English from grade one instead of grade six, and later on it was implemented as a compulsory subject. As Shamim (2008, p. 238) further notes, the government continued to support English medium schools in the private sector and their mushrooming growth throughout the country.

- 1999-2008: Another military dictator seized power in 1999. Unlike the previous military dictator, this regime promoted 'enlightened moderation' and articulated a more 'pro-

English stance' (Shamim, 2008, p. 239). For example, Mathematics and science subjects were proposed to be taught in English from grade six.

In the last decade or so, the spread of English in the education sector has continued. English medium instruction from grade one has been introduced in the public-sector schools of the province of Khyber-PakhtunKhwah (Khan, 2014). 50% public-sector schools in Punjab are also termed as English medium (Rahman, 2016, p. 15). English teacher training programs with the help of foreign agencies like the British Council also continue to date, for example "Over 30,000 teachers to get English training" (2018). The number of students taking British examinations (O/A Levels) has doubled from 2002 to 2013 (Rahman, 2016, p. 17). The language of higher education has been English (Haque, 1983) and will continue to be. About 50,000 Pakistani students go to Western countries for higher education each year (Abidi, 2017), which entails that the demand for English proficiency tests like International English Language Testing System (IELTS) remains high.

The official language status was supposed to be awarded to Urdu as per the 1973 Constitution. However, its implementation as the official language has been being delayed by each government in the last four decades. In 2015, the Supreme Court ordered immediate implementation of Urdu as the official language, which was acknowledged and endorsed by the then prime minister Muhammad Nawaz Sharif (Raj, 2017). On the practical level, however, English remains the de-facto official language of the country with ever-increasing influence in education as well as media, including the internet.

2.2.3 English in media

Three types of media can be listed where English has been and is being used in Pakistan. The first and very traditional medium is the print media that includes English newspapers,

magazines, periodicals, and other publications like books. Urdu has remained the most

dominantly used language in publications, followed by English and other regional and

indigenous languages of Pakistan.[1] As figure 2.4 exhibits, in the last 10 years the number of

publications in English has remained more or less stable around 100.
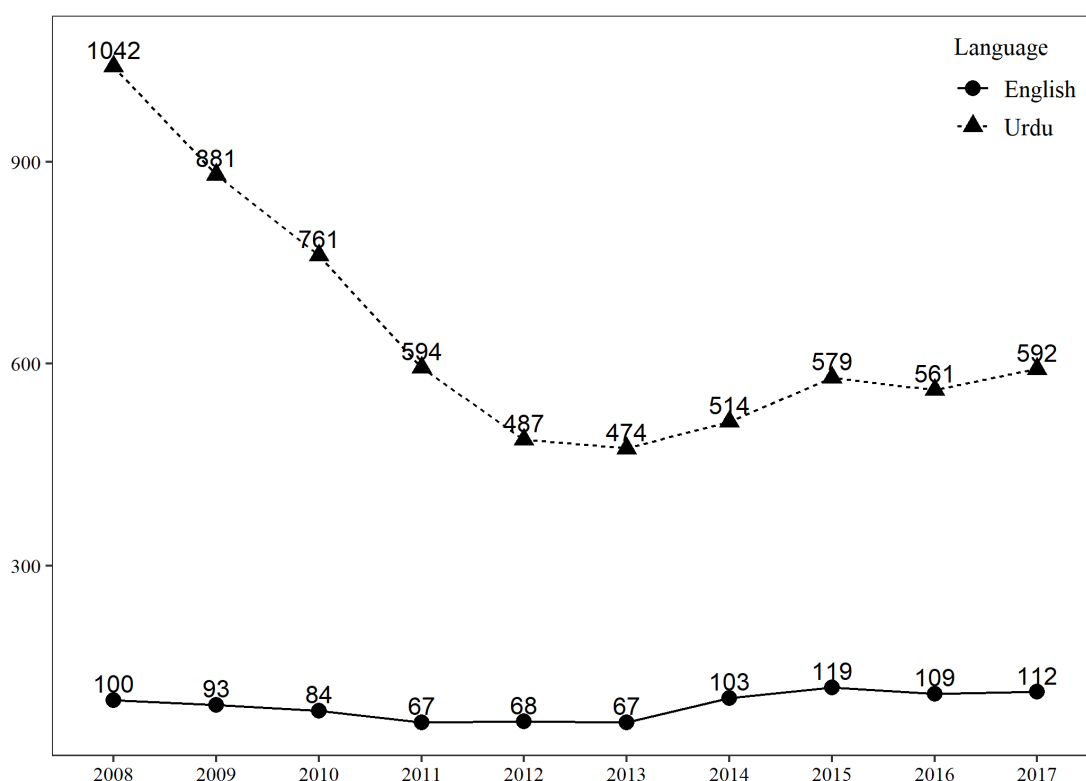
Figure 2.4 English and Urdu newspapers and periodicals published in last 10 years (Pakistan Bureau of Statistics, 2017a)

The second type of medium where English has been being used is radio and TV

programs. Rahman (2016, p. 18) reports the launch of a '24/7 English radio channel (Planet FM

94)' by the state radio, Pakistan Broadcasting Corporation. City FM 89

---

[1] Several factors can be listed for apparent decline in Urdu print media publications, e.g. electronic news channels, availability of Urdu newspapers through official websites, and piracy of Urdu digests and magazines in PDF format on the internet.

(https://www.cityfm89.com/) is another private English radio station, which broadcasts English music and other programs. The state TV channel Pakistan Television (PTV) broadcasted a news bulletin in English in the 1990s. In the 2000s, the federal government issued hundreds of licenses to private cable and satellite TV channels in Urdu and other languages. Two English news channels also started operating around 2008. However, they could not continue their transmission for a very long time due to lack of interest from advertisers (Special Correspondent, 2010; "Express 24/7 bows out amidst revenue drop", 2010). In 2013, the state broadcaster started a dedicated English language news and current affairs channel PTV World (Web Desk, 2013), which has been operating to date.

The internet is the third medium, which has become an avenue of English communication for Pakistani users living in the country and abroad. The internet first arrived in Pakistan in the 1990s (ISPAK, 2014), though it did not spread to all major cities until the mid-2000s. All major public and private sector organisations had their websites set up following the popularity of the internet. Most of these websites have been in English, though some of them now provide an option for other languages as well, mostly Urdu. Many traditional print media, i.e. English newspapers and magazines, also publish their content on their respective websites. Apart from these websites, dedicated English website and blogs (diary like posts on personal as well as commercial topics) have also come into existence. They publish varied content in English, including news stories and articles like print media.

Apart from websites, other communication channels available on the internet – e.g. instant messaging services, social networking websites, and discussion forums – have also increased the possibility of communicating in English at the individual level. Pakistan is one of the fastest growing countries in the world in terms of population. The 2017 census reported a

population of 207 million people with an annual growth rate of 2.4% (Pakistan Bureau of

Statistics, 2017b). A large percentage of this population consists of young people. The people

now in their 30s have grown up being habitual of the internet, instant messaging services like

MSN and Yahoo, and later on SMS texting on cell phones. The people in their 20s are much

more familiar with broadband internet, especially mobile broadband (3G/4G) and social

networking apps available on smartphones. In year 2017-18, there were approximately 56

million 3G and 4G/LTE subscribers in Pakistan (Pakistan Telecommunication Authority, 2018, p.

60). A large proportion of these users, i.e. 36 million, are active monthly users of social

networking websites like Facebook (Farooq, 2019). Many of these users, along with other local

languages like Urdu, also post and communicate in English.

2.2.4 Previous research on Pakistani English

Pakistani English has been studied since the 1980s. The body of research available on the

variety has two identifiable strands, i.e. research studies conducted by local scholars versus those

conducted by international researchers. In the last three decades, not many examples of a

collaboration between these two groups can be found.[2] Generally speaking, a large body of

research on the variety conducted by local Pakistani researchers is not known to the international

---

[2] A number of reasons for this gap can be identified, but a lack of research collaboration is the most obvious one – though there do exist rare exceptions to this, e.g. Manan, David, Dumanig, and Channa (2017). Very few international researchers visit or have visited Pakistan for a research stay or a conference. Moreover, for financial and administrative reasons not many local researchers are or have been able to attend international conferences in the field or go on research stays. A consequence of the lack of collaboration is the quality of studies produced locally. For example, the local research community is either not aware of the latest research trends in the field (e.g. the use of sophisticated statistical techniques in corpus linguistics) or does not have expertise to apply such methods. As a further consequence in this chain reaction, they do not feel confident to send their research papers to internationally renowned academic journals or academic publishers. Also, even if they do, such pieces of academic writing are rejected in most cases for reasons related to quality and non-native language use among others.

research community. Consequently, such research is mostly produced in local (as well as foreign but so-called predatory) academic journals and consumed only within the country. On the other hand, the research from internationally renowned scholars gets published by established academic journals and publishers of the field. Such research, however, is acknowledged and cited by local and international research community alike.

Local research on the variety has been conducted in various subdisciplines of linguistics and on various research themes. Approximately until 2009, the majority of the studies used non-corpus-based methodologies to focus on the following areas: lexico-grammatical description of the variety, including code switching and borrowing;[3] sociolinguistic history of the variety;[4] content and discourse analysis[5]; and attitudes towards (Pakistani) English[6] among others. Since 2009, a growing body of research has also used corpus-based research methods.[7] These studies have focused on describing lexico-grammatical features of the variety using various locally compiled corpora. A sub-group of the corpus-based research has also used MD analysis to study register variation. Lastly, some of the studies and research scholars are also internationally

---

[3] Rahman (1991); Talaat (2002); Mahboob (2004); Anwar (2009); Talaat and Anwar (2010); Sarfraz (2011)

[4] Rahman (2015); Shamsie (2017), Khan (2012)

[5] Mahboob (2009), Mahboob (2015)

[6] Jabeen, Mahmood, and Rasheed (2011); Khushi (2011); Parveen and Mehmood (2013); Khan (2013); Mahboob (2013)

[7] A. Mahmood (2009); R. Mahmood (2009); Jabeen, Rai, and Arif (2011); R. Mahmood, A. Mahmood, and Saeed (2011); Mahmood and Shah (2011); Mahmood and Ali (2011); A. Mahmood, R. Mahmood, and Saeed (2011); Mahmood, Hassan, Mahmood, and Arif (2012); Aziz and Mahmood (2012); Mahmood, Batool, Shah, and Parveen (2013); Tabassum, Shah, and Bilal (2013); Alvi (2013); Shakir (2013); Ali (2013); Ijaz, Mahmood, and Ameer (2014); Jameel, Mahmood, Hussain, and Shakir (2014); Mahmood, Obaid, and Shakir (2014); Ahmad and Ali (2014); Hussain and Mahmood (2014); Gillani and Mahmood (2014); Rasheed and Mahmood (2014); Ahmad and Mahmood (2015a); Rofess and Mahmood (2015); Ahmad and Mahmood (2015b); Alvi, Mehmood, and Rasool (2016); Hussain (2016); Ali and Ahmad (2016); Rashid, Mahmood, and Ahmad (2017); Moghees, Dar, Zaid, and Saeed (2017); Asghar, Mahmood, and Asghar (2018); Azher, Mehmood, and Shah (2018); Iqbal (2018); Sadia and Ghani (2019); Ahmad, Mahmood, Mahmood, and Siddique (2019)

renowned. For example, Rahman (2015) – which was originally published in 1990 – and Talaat (2002) are widely cited research studies on the variety. In terms of research scholars, Rahman, Mahboob, and Talaat are some of the names that appear in literature in relation to Pakistani English.[8]

The studies conducted by international scholars (or in collaboration with them) can also be divided in two groups. The first group generally consists of studies conducted by Baumgardner, a famous international researcher, in the 1980s and 1990s. He studied attitudes towards Pakistani English (Baumgardner,1995) and also published on various lexico-grammatical aspects of the variety (Baumgardner, 1987; 1992; 1996; Baumgardner, Kennedy, and Shamim, 1993). He also edited a book on Pakistani English (Baumgardner, 1993), which included contributions from local researchers as well. The second group of studies started coming into existence after 2010. Around this time, more international researchers began exploring Pakistani English, as the South Asian Varieties of English (SAVE) corpus was created (Mukherjee & Schilk, 2012). Later on, another corpus which includes a Pakistani English component, the GloWbE corpus (Davies and Fuchs, 2015), was also compiled. Many studies have used the Pakistani subcomponents of these corpora to investigate features like verb complementation in Pakistani English.[9]

As mentioned above, local researchers have also been studying the variety using corpus linguistics since 2009 (e.g. A. Mahmood, 2009; R. Mahmood, 2009). More recently, they have also used MD analysis to study register variation in various (mostly written) domains of

---

[8] See also Mahboob (2003, p. 2) and Mahboob (2009, p. 180) for a list of studies on Pakistani English.
[9] Hundt, Hoffmann, and Mukherjee (2012); Schilk and Hammel (2014); Bernaisch et al. (2014); Levshina (2014); Lange (2016); Gries and Bernaisch (2016); Kopaczyk and Tyrkkö (2018)

Pakistani English (cf. footnote 7). However, not many studies have been conducted on English CMC or English on the internet. Moghees, Dar, Zaid, and Saeed (2017) focus on CMC features in Facebook posts. Other studies have explored bilingual aspects like code switching and code mixing (Rafi, 2013; Rafi, 2017; Parveen and Aslam, 2013) on the internet. The use of English on the internet by Pakistani users has largely remained unexplored until now.

2.2.5 Section summary and identification of research gap

This section has provided a socio-historical and sociolinguistic background of Pakistani English as an outer circle variety. It has briefly overviewed the status of English in education and language policy. The section further explored the increase in the usage of English in electronic media (radio and TV) and a new medium, i.e. the internet. Lastly, a summary of previous research studies on Pakistani English was presented.

As it has been noted in previous subsections, a large majority of local and international research studies have focused on written registers – especially newspapers, and newspaper-based corpora – to study lexical and grammatical features in the variety. More recently, spoken registers (e.g. radio and TV programs in English) have also been taken into account in such studies. Radio and TV transmission, however, is mainly dominated by local languages, especially Urdu. Moreover, a very limited set of individuals has access to these programs (cf. chapter 4). The internet, on the other hand, is much more accessible to the general population, especially the younger generation. In addition to local languages, it also provides opportunities to communicate in the English language. This situation implicates two points:

- A potential expansion in the use of English has occurred in Pakistan, especially in terms of interactive uses, e.g. internet-based discussion forums;

- The use of English on the internet has largely remained unexplored in Pakistani context, which presents a research gap.

To sum up, based on sections 2.1 and 2.2, the need to study internet-based English communication has been identified in the light of World Englishes theory as well as previous research on Pakistani English. The next section provides the methodological background on how this is researched in the present study.

2.3 Methodological Framework: Register Variation and MD Analysis

This section provides necessary details regarding the methodological framework that will be used in this study. The concept of register is defined in the next subsection, and later on the MD analysis framework is explained.

2.3.1 Defining register

2.3.1.1 Language and variation.

Ferguson (1994, pp.15-16) links the study of language variation with one of the very basic properties of human language, i.e. conventionalisation. He argues that a community of users implicitly agrees on the meaning of certain expressions that can be used in particular ways under particular conditions and circumstances. This agreement or contract is continuously gained, re-gained, maintained, and changed. The dynamics of maintenance, achievement, and change of this contract are the concern or problem of conventionalisation. This problem, in turn, can be studied in many ways, one of which is to analyse variation and change that happens.

He identifies four types of variation, i.e. dialect, register, genre, and variation at conversational level (Ferguson, 1994, pp. 18-24). The basic assumption behind each type of variational study are provided below:

1. Dialect variation: A group of people that is different from other such groups in terms of location, 'marriage patterns', geographical conditions, religion, or 'other interactional behaviours', with the passage of time develops patterns of language structure and use that are different from other groups.

2. Register variation: A communicative situation that is common in a society with particular functional traits like participants, settings, communicative purposes, topics etc., with the passage of time develops particular patterns of language structure and use – due to particular functional traits – that distinguish its language from the language of other such situations.

3. Genre variation: A type of message that commonly occurs in a society with 'distinct semantic content, participants, occasions of use' etc. with the passage of time develops particular internal structures that differentiate it from other such message types occurring in the same society or community.

4. Conversational analysis: The participants of a conversational interaction build their own patterns of 'turn taking, repairs, code switching', and other such conversational cues that depend on their communicative aims.

The first three types focus on "individual behavior" and "changes in shared patterns of language structure and use", while the fourth type studies a "more local kind of variability" (p. 17). The distinction between dialect and register is quite clear, which, according to Matthiessen (2019, p. 27), Halliday and colleagues term as "the variety according to user" and "the variety according to use" respectively. The contrast between register and genre variation, on the other hand, is not very clear and this issue is again discussed in subsection 2.3.1.4.

2.3.1.2 History of the term 'register'.

The term *register* is borrowed from music to refer to a device that is used to control organ pipes sharing a tonal quality or the range of notes of a voice or instrument (Schubert, 2016, p. 2). The term was first used by Bertram Reid in 1956 in linguistics to refer to situation-dependent linguistic choices made by a speaker or writer (Gray and Egbert, 2019, p. 2). The authors list various theorists and scholars who helped develop models to study register variation in a systematic way; e.g., Hymes presented the SPEAKING model, Halliday introduced field, mode, and tenor, which are three components of a communicative situation, and Brown and Fraser defined various situational characteristics like participants, setting, purpose and so on.

2.3.1.3 Narrow and wide view of register.

Two views regarding the concept of register can be observed in the sociolinguistic literature. Schubert (2016, p. 3) terms them as the narrow and wide views of register. The author notes that the narrow view is normally employed by sociolinguists who limit register to refer to various types of vocabulary, e.g. 'the register of law', 'the register of medicine' etc. The wide view, on the other hand, is not just limited to vocabulary and is one of the basic assumptions of Systemic Functional Linguistics (SFL). He quotes Halliday and Hasan (1976, p. 22) as follows: "The linguistic features which are typically associated with a configuration of situational features – with particular values of the field, mode and tenor – constitute a register". The wide view, hence, also includes situational as well as other linguistic characteristics.

2.3.1.4 Register, genre, style, and text types.

As it has been noted above, there appears to be an overlap between Ferguson's (1994) definitions of register and genre variation. A similar observation has been made by Matthiessen (2019, p. 26), who looks at the history and use of these terms in the research tradition of SFL. As

the author notes, initially the term (literary) *genre* was used for functional variation. The term *register* was popularised in the 1960s and 70s with the work of Halliday and colleagues. In the 1990s the concept of genre was reinvented as a higher contextual level of register. At the same time, a tradition of genre-based research was started in the educational domain. Lastly, the body of research produced by Biber and colleagues (e.g. 1988 and later studies) uses the term *register* solely to refer to functional variation.

Biber and Conrad (2009, p. 16) distinguish between the three relevant concepts of register, genre, and style based on criteria like text length (complete or excerpt), linguistic features, how frequent and where in the text the linguistic features occur, and their interpretation. To study register, text excerpts are used to analyse any lexical or grammatical feature that is frequent or pervasive throughout the text, and the linguistic features are interpreted in terms of communicative functions. The genre perspective focuses on complete texts where specialised features like formatting and 'rhetorical organisation' are examined. The features normally occur only once in the text and usually they are not interpreted functionally. The study of style is similar to the register perspective in terms of textual length, features under study, and their frequency or pervasiveness in the text. However, these features are not interpreted functionally because they are preferred due to their aesthetic value. The view of register adopted by Biber and Conrad is the most relevant for the present study.

Another relevant contrast that is worth mentioning here is that of register and text type. The term *text type* is generally used in corpus linguistics to refer to groups and subgroups of texts in a corpus. For example, Kirk and Nelson (2018) use the term *text types* and *text categories* to refer to textual groupings in the ICE. They also use the term *register/genre* to refer to similar concepts, albeit at a more abstract level.

Biber (1989, p. 5) advocates for a text typology that is based on linguistic criteria instead of "a priori functional grounds". In his typology, the texts are grouped together in such a way that within one group the texts are linguistically similar to each other to the maximum extent, while the texts in different groups are linguistically different from other groups of texts to the maximum extent. Biber refers to the resulting text categories using the term *text types*. They are distinguished from registers, which are text categories based on functional criteria. Each register category will include texts with a range of linguistic features, and can include various text types. For example, Biber (1989) uses 23 registers and identifies 8 text types based on linguistic criteria. Hence, registers and text types dissect a corpus of language texts in two 'complementary ways' (Grieve, Biber, Friginal, and Nekrasova, 2010, p. 315), i.e. based on functional and linguistic criteria respectively.[10]

2.3.2 Study of register in various research designs

2.3.2.1 Three research designs.

Biber (e.g. Biber and Jones, 2009; Biber, 2012) distinguishes three research designs in corpus linguistics based on the observational unit of choice. The observational unit can either be each occurrence, a complete text, or a whole corpus or sub-corpus. Two research goals are also identified: (1) "describing the variants of a linguistic structure" and (2) "describing the differences among texts and text varieties" (Biber, 2012, p. 12). As a result, three corresponding research designs come into existence:

- Type A studies focus on individual occurrences of the phenomenon under study and its linguistic variants;

---

[10] A more practical explanation of texts types is provided in subsection 5.2.1

- Type B studies analyse individual texts and describe differences among text varieties;

- Type C studies examine complete corpora or sub-corpora focusing on either research goal one or two.

Biber (2019, p. 54) calls type A studies 'variationist', type B 'text linguistic', and type C studies 'corpus linguistic' research designs. While the first two are distinct and complementary to each other, the third research design (i.e. corpus linguistic) can either be variationist or text linguistic depending on the research goal.

2.3.2.2 Study of register in variationist linguistics.

Szmrecsanyi (2019, p. 76) defines variationist linguistics as a branch of variation studies which examines alternative ways of articulating the same idea. The job of the researcher is to identify alternative variants and use quantitative methods to arrive at (mathematical) models that explain which factors affect the choice of a given linguistic variant from a set of choices that are equal in terms of meaning and function. This research tradition does not focus on the functional aspect, i.e. the functional link or links between the linguistic features and situational parameters, instead it analyses different ways of "saying the same thing" (Labov, 1972, p. 188).

As it has been noted in the previous subsection, the unit of analysis in this research tradition is the individual occurrence of the phenomenon under study. The job of the researcher, as Szmrecsanyi (2019, p. 78) explains, is to identify the constraints on the linguistic choices of language users. These constraints or restrictions can be language internal (e.g. grammatical category) or language external (e.g. geographical region, race, gender etc.). Register is understood as stylistic variation in Biber and Conrad's (2009) terminology. It is a language external factor which affects the linguistic choices of a language user. Corpus linguistic-based variationist studies consider the text groupings of the corpus under study as register or genre. The

text grouping is used as a predictor in, for example, logistic regression models to explain alternative linguistic choices.

2.3.2.3 Study of register in text linguistics.

Biber (2019, p. 43) finds the roots of the text linguistic approach in the 1970s and 80s when researchers like Van Dijk and Halliday promoted the analysis of texts instead of focusing entirely on the sentence structure.[11] The proponents of this approach favoured the study of the grammar of texts in the same way as sentence-level grammar was being studied. The structural, logical and discourse organisation of texts, cohesion, and coherence were some of the research topics in this tradition. Biber further notes that this research tradition did not focus much on the linguistic description of texts. However, researchers like Hymes – in the framework of ethnography of communication – did describe speech events in terms of "a range of situational characteristics that had functional underpinnings and linguistic correlates" (p. 43). The text linguistic approach, then, aims to describe the linguistic properties of individual texts on the basis of which groups of texts belonging to different registers are compared to study register variation.

The study of registers follows three steps (Biber, 1994; Biber and Conrad, 2009, pp. 6-10). The situational characteristics of register(s) under study are described in terms of participants and their role relationships, setting, medium, topics discussed and so on. 'Pervasive

---

[11] SFL provides a framework to view language as a semiotic system in the tradition of functional linguistics (Matthiessen, 2019, p. 78). The concept of register (and other types of variation like dialect variation) is also theorised in the overall theory of language system. As Matthiessen (2019) and the following discussion from Biber (2019, p. 43) reveals, Biber's approach to register study or 'text linguistics', as he calls it, is an extension of theoretical work performed in SFL. 'Text linguistics' is more concerned with the empirical and practical part of register analysis. There are also other implementations of the SFL view of register study, e.g. Neumann (2014) uses SFL theory to identify lexico-grammatical features to study register variation in English and German translated texts.

and frequent' (Biber, 2019, p. 45) linguistic (both lexical as well as grammatical) features in the texts of target register(s) are analysed. The linguistic analysis is interpreted in terms of communicative functions, which connects the situational and linguistic analyses. So, the "linguistic variation is functional rather than indexical or purely conventional" (p. 45).

2.3.3 MD analysis

MD analysis is a special case of text linguistics. It provides a complete framework of analysis to study register variation. The following subsections elaborate it further.

2.3.3.1 MD analysis versus previous work on register.

Biber (2019, p. 50) identifies three major differences between the MD framework and earlier investigations of register variation.

1. These investigations examine a single parameter which assumes that registers have only one situational difference instead of many.

2. These studies assume that register variation can be studied by analysing dichotomous distinctions like 'high versus low varieties', 'restricted versus elaborated codes', 'formal versus informal registers', 'planned versus unplanned discourse'.

3. Previous researchers recognise that registers are defined by sets of linguistic features that occur together or clusters of features that are associated to each other. However, they do not apply any empirical and/or quantitative methods to identify these sets or clusters of linguistic features that co-occur.

As a result, in MD studies more than one dimensions or parameters are identified empirically, which consist of different co-occurring linguistic patterns that reflect "different functional underpinnings" (p. 50). These dimensions are not dichotomous in nature, but considered as continua of variation in texts of the registers under study. The study of register in

MD analysis is comparative in nature, where many register categories are analysed simultaneously to identify salient patterns of variation in any given category. Lastly, registers can be studied with increasing levels of granularity. Hence, a register like conversations can have sub-registers and each sub-register can further be divided into sub-registers to study minute variational patterns (Biber and Conrad, 2009).

2.3.3.2 Steps involved in a MD analysis.

Biber and Gray (2013, p. 403) say that a MD analysis follows eight methodological steps. A summary of these steps is provided below:

1. Designing a database: Collection of a corpus of texts with enough texts in each register category.

2. Identification of linguistic features: Lexico-grammatical features that have functional significance and are pervasive in the target registers are selected. Egbert and Staples (2019) recommend that previous studies on similar registers should be considered in this regard.

3. Development of computer software for annotation of linguistic features.

4. Tagging: Normally, grammatical taggers are used.

5. Computing Frequencies: Normalised frequencies per thousand words are computed for each linguistic feature. As Biber (2012) notes, alternative ways to say the same thing (e.g. *that* versus *that* deletion in relative clauses) are counted separately for each text (or observational unit). This is unlike the variationist approach where proportions of alternative choices are calculated.

6. Factor Analysis: EFA is a multivariate statistical technique that is used to identify linguistic cooccurrence.

7. Computation of 'dimension' scores to compare and examine prominent linguistic differences and similarities in the registers under study.

8. Functional interpretation of 'factors': Sets of co-occurring linguistic features are interpreted in terms of underlying communicative functions (called 'dimensions' of variation).

Furthermore, there are two types of MD studies. The first type of studies performs all eight methodological steps to identify new dimensions of variation and communicative functions. The second type of studies can use dimensions of variation identified in an existing MD study, e.g. Biber (1988), to describe a new or similar register.

2.3.3.3 MD analysis and other dimensionality reduction techniques.

Dimensionality reduction techniques are statistical methods that are used to reduce a large number of dimensions to relatively fewer ones. They "can be viewed as […] method[s] for latent feature extraction" (Nguyen and Holmes, 2019, p. 1). A number of dimensionality reduction techniques can be used for MD analysis. Most MD studies use EFA to find out underlying dimensions or latent variables.[12] Biber and Egbert (2016) use principal component analysis (PCA), which is mathematically different from EFA, in MD analysis to get similar results. Clarke and Grieve (2017) use another dimensionality reduction technique called multiple correspondence analysis (MCA) to identify underlying functional dimensions in abusive tweets.

---

[12] MD analysis has also been criticised due its use of EFA and other issues. For example, Evert (2018, p. 12) mentions ambiguity in deciding the number of factors to retain and "arbitrary cutoff for feature weights" as potential pitfalls. He also notes that the selection of linguistic features and texts to include affects the results of EFA. However, other dimensionality reduction techniques, as mentioned later on in the same paragraph, and EFA show more or less similar results. Moreover, as mentioned in the previous subsection, feature selection and the interpretation of the resulting dimensions is functionally motivated and an element of subjectivity always remains.

Egbert and Biber use CDA – "a type of regression technique, in which weighted combinations of linguistic variables are used to classify texts into known register categories" (Egbert and Biber, 2016, p. 2) – to arrive at similar communicative functions as those resulting from the use of EFA. Diwersy, Evert, and Neumann (2014) also use discriminant analysis along with PCA to study register variation.

2.3.3.4 Using MD analysis to study World Englishes.

As it has been noted in subsection 2.1.3, the inclusion and analysis of communicative event has been identified as a necessary element for future research in World Englishes. Subsection 2.2.4 shows that the majority of (international) research on Pakistani English uses variationist methodology, as Biber (2012) defines it. MD analysis provides a comprehensive framework that not only takes into account linguistic but also situational characteristics of the registers under study. As Biber (2012) notes, text linguistic and variationist approaches can be used to complement each other. Until now very few studies (e.g. Xiao, 2009; Bohmann, 2017) have used MD analysis to study language variation in varieties of English around the world. These studies have found interesting patterns of variation in Englishes. As the next section shows, MD studies on internet-based registers generally do not include non-native varieties of English. MD analysis, hence, can be used as a method to analyse online registers of Pakistani English, for which a research gap has been identified in subsection 2.2.5.

2.3.4 Section summary

The first subsection started with an explanation of the term *register*. Different types of variations, the narrow and wide view of register, and the distinction between register and other relevant concepts like genre and text type have been discussed. The next subsection overviewed three research designs – namely variationist, text linguistic, and corpus linguistic – and the study

of register in variationist and text linguistics. The third subsection explained MD analysis as a special case of text linguistics, its assumptions (including the comparative study of registers), and how to perform it. Lastly, the suitability of MD analysis to study World Englishes was discussed briefly at the end of the third section.

2.4 Research Background: Previous relevant Studies

The aim of this section is to provide a comprehensive overview of previous relevant research conducted on online and similar offline registers that are under study here – i.e. various types of blogs and interactive registers (Facebook status updates, Facebook groups, tweets, and comments), spoken conversations, opinion columns, and news reports. The focus is to review previous MD studies on these registers. In addition, other relevant types of studies, e.g. genre analysis studies on blogs, are also included.

2.4.1 Blogs

2.4.1.1 The word 'blog'.

According to Grieve et al. (2010), the term *blog* (which is a short form of *weblog*) was popularised in the last couple of years of the previous century. It was also "voted as the new word most likely to succeed by the American Dialect Society" (p. 304). By definition, a blog or weblog is a website which is frequently updated and new entries or posts appear in a reverse chronological order (Herring, Scheidt, Bonus, and Wright, 2004, p. 1). In 2003 around 1.3 million websites called themselves weblogs and around 870,000 of them were being regularly updated (Herring et al., 2004).

2.4.1.2 Blogs versus other (web) registers.

Miller and Shepherd (2004) study blogs in the tradition of genre analysis. They consider the blogs register "as a complex rhetorical hybrid" (p. 16) of a number of other registers that are

much older than blogs. They identify a number of registers as potential ancestors of blogs. Because initial weblogs were used to automatically record the activity of web servers, the register of logbooks kept to record voyage or flight data is one such candidate. The commonplace book is another potential ancestor, which was used as a teaching tool in the Renaissance. It was a notebook for students where they could note down important "passages, epithets, phrases and aphorisms from their reading, organized into headings or places" (p. 13). A number of journalistic registers – pamphlet or broadside from the 17th and 18th centuries, and present-day editorial and opinion columns – are also related to blogs. Unlike the above-mentioned potential ancestors, the aim of these registers is commentary and expression of personal opinion. Lastly, the personal diary or journal can also be considered as an ancestor of blogs. A diary can be divided in two subtypes, intrinsically focused (focusing on self) and extrinsically focused (focusing on surroundings), which has also been inherited by blogs.

Herring et al. (2004) identify the placement of blogs in the ecology of contemporary internet registers. Weblogs fall between standard webpages and asynchronous CMC based on three dimensions, i.e. "frequency of updates, symmetry of communication, [and] multimodality" (p. 10). Standard webpages are not updated very frequently, the relationship between web page author and readers is asymmetrical in nature (i.e. generally one-way exchanges), and they broadcast multimedia. On the other hand, CMC registers (e.g. discussion forums, texting, and chat) are continuously updated, symmetrical in exchanges between participants, and are text-based. Blogs are in between these two ends of the continuum: they are frequently updated, allow limited exchanges in the form of comments but commenters do not have the same rights as a discussion forum poster, and make only limited use of multimedia, i.e. they are mostly text-based.

2.4.1.3 Types of blogs.

Herring et al. (2004, p. 2) – in their review of previous research on blogs – identify three types of blogs: filter blogs, personal journal blogs, and notebook blogs. In the first type, the blogger searches various types of content on the web and then writes a blog post to guide the blog readers as well as comments on external events. So, the function of this type of blogs is 'filtering' content. Personal journal type blogs focus on internal issues (e.g. bloggers' feelings and thoughts), while notebook blogs can either contain external content (like filter blogs) or internal content (like personal journal blogs).

Krishnamurthy's (2002) analysis of community news blogs was conducted immediately after the events of 9/11. Krishnamurthy identifies two dimensions of variation in blogs: individual versus community blogs and personal versus thematic blogs. These two dimensions result in four different types of blogs: personal community blogs, personal individual blogs, thematic community blogs, and thematic individual blogs. Herring et al. (2004, p. 3) note that personal thematic blogs are the same as filter blogs because they select information from the web and provide commentary on it. They further note that certain blog types are less frequent, e.g. community blogs of thematic and personal nature.

Later on, Puschmann (2010) examines another thematic blog type, which he calls corporate blogs. These blogs are published by corporations, in which the company executives post updates regarding their products and services etc.

2.4.1.4 MD studies on blogs.

Unlike the majority of the above-mentioned studies, MD studies mainly focus on linguistic analysis and characteristics of the registers under study. A number of MD studies have included blogs in their datasets or have exclusively focused on blogs. The first attempt to

describe dimensions of variation of web registers is Biber and Kurjian (2007). However, they do not include blogs as a separate category. Grieve et al. (2010) exclusively focus on blogs. Their dataset includes 500 blogs collected from all over the United States. They identify four dimensions of variation in the analysis of this dataset, namely (i) informational versus personal focus; (ii) addressee focus; (iii) thematic variation; and (iv) narrative style. They also identify three text types, which more or less correspond to previously identified blog types: the personal blog type is highly personal, narrative and addressee focused; the commentary blog type focuses on less personal topics and provides a commentary; and a third very rare text type, i.e. expert blogs, uses a "very formal and impersonal style […] to convey information on a particular topic" (Grieve et al., 2010, p. 319).

Grieve et al.'s (2010) study has later on been used to compare two registers and regional varieties of English, i.e. blogs and opinion columns from the USA and Philippine. Hardy and Friginal (2012) conclude that American blogs and opinion columns have more variation as compared to their Filipino counterparts. Filipino blogs are more formal as compared to the informal and personal style of American blogs. The differences between Filipino and American opinion columns are not as salient as compared to those of blogs. The authors conclude that the formal style of Filipino blog writers is probably due to their educated and professional backgrounds as compared to U.S. bloggers.

Two other studies have also conducted MD analyses on online registers. Titak and Roberson's (2013) study includes blogs along with other internet-based registers like comments, Facebook, and Twitter posts etc. The authors observe that blogs are more personal and slightly more interactive as compared to other registers, e.g. Facebook and Twitter posts. Sardinha (2014) uses a similar set of web registers to compare internet and pre-internet registers on Biber's

(1988) five dimensions of variation. In Sardinha's analysis, blogs appear to be similar to press reportage, press reviews, and biographies. The author notes that blogs (and webpages) can be considered digital alternatives of pre-internet written registers.

Biber and colleagues have published a series of research papers (Egbert, Biber, and Davies, 2015; Biber, Egbert, and Davies, 2015; Biber and Egbert, 2016) and a book length study (Biber and Egbert, 2018) on online registers. They use a subset of the GloWbe corpus developed by Davies and Fuchs (2015). Biber and Egbert (2018) apply MD and semantic keyword analyses to describe various online registers, including various types of blogs. A very brief overview of their findings is provided below:

- Personal and travel blogs: These blogs are written by non-professionals and usually read by family, friends, fans etc. They cover topics like parenting and travelling in the case of travel blogs. MD analysis shows that these blogs are oral, involved, and narrative. Biber and Egbert (2018, p. 86) also acknowledge the similarity between personal blogs and personal diaries. However, the online publication of personal blogs makes them different from personal diaries.

- Opinion blogs: The authors note that this category cannot be very well-defined. Opinion blogs can be very similar to opinion editorials due to the use of arguments followed by factual information or narratives on the one hand, and can consist of shorter personal opinions that are supported by little evidence on the other hand (Biber and Egbert, 2018, p. 107). MD analysis shows that they have near zero scores on all dimensions except dimension seven 'literate opinion' which shows that they use stance noun phrases more frequently than other registers (p. 115).

- Religious blogs: These blogs appear to have more informational content as compared to opinion-oriented content. The keyword analysis shows that there are three main semantic domains that are related to religion, i.e. people, concepts, and biblical references (p. 132). The majority of these blogs published from inner circle countries covers topics related to Christianity.

- How-to/Instructional blogs: How-to documents can also appear on websites that are not explicitly identified as blogs. Irrespective of their place of publication, the information provided in these blogs is procedural in nature, guiding the readers regarding the accomplishment of a particular task, for example cleaning a virus from one's computer (p. 136). MD analysis shows that they use features like concrete and process nouns, verbs of activity, auxiliary verbs related to modality, *if* clauses, and second person pronouns (p. 151). Food blogs publishing recipes on cooking various dishes are an example of this type of blogs.

- Information blogs: Information blogs are published by institutions like academic institutes or universities, non-profit organisations, and corporations. They can publish informational documents related to current interest, summarise the findings of recent research studies, and provide information that might be relevant to a product sold by the commercial website not listed on the current page (p. 146). The results of MD analysis exhibit that they are also generally non-distinct in nature, with slightly informational characteristics.

2.4.1.5 Other studies on blogs.

Blogs have also been studied in various other research traditions. Schildhauer (2016) compiles a corpus of personal weblogs starting from the early 2000s and describes the linguistic

and generic development of these blogs. Blogs have also been researched from the point of view of media and journalism related studies, e.g. a book length publication edited by Tremayne (2007). Herring and Paolillo (2006) study blogs for gender variation. Others like Montes-Alcalá (2007) and San (2009) use blogs to study code switching and code mixing. Lastly, blog data has also been used to study variation in Englishes around the world in studies like Kopaczyk and Tyrkkö (2018).

2.4.2 Opinion columns and news reports

As has been noted in previous subsections, blogs have a close relationship with press related registers, i.e. editorials and news reports. Thus, it is worth mentioning the multidimensional characteristics of these two registers based on previous studies.

Biber (1988) includes these two registers labelled as editorials and press reportage respectively. Both of these registers are highly informational on the first dimension of Biber's (1988) study ('involved versus informational production'). News reports are comparatively more informational than editorials. Biber (1988) notes that some news reports are narrative while others are not, which makes them very slightly narrative on dimension two ('narrative versus non-narrative concerns'). Editorials have a negative score on dimension two, while they have a very high score on dimension four ('overt expression of persuasion').

Other studies which have been conducted on online registers are more recent. Titak and Roberson's (2013) study also includes newspaper articles and opinion columns. Their findings show that opinion columns are slightly personal and narrative while newspaper articles are mainly informational. Opinion columns also include reported speech in their data. Biber and Egbert's (2018) corpus also contains news reports. Their findings show that news reports have excessive use of features like communication verbs and various types of nouns. Lastly, according

to Hardy and Friginal (2012), there is not much difference between native and non-native opinion column writers in the case of American and Filipino Englishes.

MD analyses have also been conducted on these two registers in the context of Pakistani English. In case of newspaper opinion columns, Alvi et al. (2016) have compared Pakistani English opinion columns with British English opinion columns using Biber's (1988) five dimensions. Their findings show that Pakistani opinion columns – in comparison to their British counterparts – are more informational on dimension one ('involved versus informational production'), more elaborated on dimension three ('explicit versus situation-dependent reference'), and more abstract on dimension five ('abstract versus non-abstract information'). Lastly, Ahmad and Mahmood (2015a) compare Pakistani and British English news reports. Their comparison on Biber's (1988) five dimensions shows that Pakistani news reports discourse is more informational on dimension one ('involved versus informational production'), slightly more narrative on dimension two ('narrative versus non-narrative concerns'), more elaborated on dimension three ('explicit versus situation-dependent reference'), less overtly argumentative on dimension four ('overt expression of persuasion'), and slightly more abstract on dimension five ('abstract versus non-abstract information') (Ahmad and Mahmood, 2015a, p. 16).

2.4.3 Interactive online registers

2.4.3.1 Computer-mediated communication versus spoken conversations.

Computer-mediated communication (CMC)[13] and spoken conversations have been being compared since the inception of CMC. As Herring (2011, p. 1) notes, early scholars do not

---

[13] The term *computer-mediated communication* or *CMC* has been used in literature to refer to any form of communication that is transmitted electronically using mobile, computer, or any internet- or network-enabled device. Thus blogs, Facebook status updates, discussion forums, tweets, comments, along with text messages etc. all are different forms of CMC. The researcher considers this term to be too

consider CMC to be 'conversation', because it is not produced orally while conversation is "by definition, spoken and heard". However, later on the idea of CMC being 'conversation-like' was accepted by some scholars even for asynchronous CMC modes like blogs, which are less oral as compared to synchronous modes like chat. The similarities between CMC and spoken conversations are "with regard to the employment of 'oral' discourse strategies, the management of interaction, gender and power dynamics, code choice, [and] processes of accommodation" (p. 2). The differences between the two include characteristics like turn taking as even in the most synchronous modes of CMC "turn-taking patterns systematically violate the 'no-gap, no overlap' principle" (p. 2).

Zappavigna (2012, pp. 30-32) comments on the conversational nature of Twitter discourse. She agrees with Herring (2011) that the concept of turn taking cannot directly be applied to such communication channels due to their asymmetric nature and weak social expectation to reply even when a user is directly mentioned. Since the posts remain available long after the original exchange took place, users can continue to take part in an exchange or simply leave it, and the exchanges can also overlap (Zappavigna, 2012, p. 32). She also notes that most definitions of spoken conversations also take into account some form of turn taking which is more relevant from a conversation or discourse analysis point of view instead of a linguistic one. Hence, conversation analysis of a randomised corpus of tweets becomes less relevant, but a linguistic analysis, e.g. of lexico-grammatical structure, still remains possible. To conclude, CMC and spoken conversations have situational similarities as well as differences. A

---

general for the present context. Hence, in the course of this document this term has not been used very often. Instead, specific labels for individual forms of online communication or online registers are used, e.g. different types of blogs or Facebook status updates.

comparative linguistic analysis – i.e. using MD analysis – could provide interesting insights into the nature of these registers.

2.4.3.2 Microblogging: Facebook and Twitter.

Lee (2011, p. 111) defines microblogging as "the writing of short messages on the web designed for self-reporting about what one is doing, thinking, or feeling at any moment". Two of the oldest existing microblogging platforms are Facebook and Twitter, though other platforms like Snapchat and Instagram have also appeared over the years. Microblogging posts on Facebook are called 'status updates'. Lee further notes that status updates combine features from personal blogs, tweets, instant messaging, and text messages. Other features like 'likes' and 'replies' have made status updates more interactive. Facebook posts can be formulaic texts (e.g. 'feeling positive'), re-shares of content from other users, or original posts (Hinrichs, 2016). The content can be text as well as multimedia like videos. As Hinrichs (2016) notes, Facebook posts are semi-public as the user has more options regarding the privacy of the post.

Lee (2011, pp. 115-117) identifies 11 types of Facebook status updates:

1. The first type of status updates responds to the original Facebook prompt 'What are you doing right now?'

2. Participants talk about their daily lives including personal or home and work life.

3. Participants express opinions and judgements about themselves and others.

4. Participants report their positive or negative mood.

5. Like instant messaging, participants may use an 'away message' to let their friends etc. know that they are going somewhere or have been gone.

6. The participants can post open-ended questions to initiate discussions or comments from their visitors.

7. The participants can mention certain audience, hence directing the message to a particular group of Facebook contacts.

8. They can post quotations from famous songs or sayings.

9. They can express their silence and interjection using punctuation and question marks etc.

10. They can create humour using wordplay or by making jokes.

11. The participants can specifically comment on the features of Facebook, expressing their likes and dislikes.

Zappavigna (2012) notes distinct characteristics of Twitter posts, or 'tweets'. Tweets are restricted to a certain number of characters (140 in the past, while 280 characters at the time of writing). Users can mention other accounts or users using @ and hashtags (i.e. words followed by #) to indicate the topic of tweets. This can create "ambient communities of values to coalesce around particular hashtags" (Zappavigna, 2012, p. 191). These communities are not formed as a result of direct interaction but in an indirect way due to the searchable nature of hash-tagged topics.

2.4.3.3 MD studies on Facebook and Twitter posts.

As has been mentioned in the previous subsection on blogs, many MD studies have also included Facebook and Twitter posts and described their characteristics. Titak and Roberson (2013) say that Facebook and Twitter posts have a "nominal and informational style of writing and very limited narrativity" (p. 254). They consider space limitations (limited number of characters in case of Twitter) as one of the possible reasons. Friginal, Waugh, and Titak (2018) use dimensions identified in Titak and Roberson (2013) to study a larger corpus of Facebook and Twitter posts. The corpus is divided in various topic domains like entertainment, politics, and business. Their findings show that Facebook and Twitter posts show opposite trends on most of

the dimensions. Facebook posts have an involved and interactive discourse, and are narrative in comparison to Twitter posts (Friginal et al., 2018, p. 358). They consider post length restrictions and the use of hashtags in tweets as possible reasons of these differences.

Sardinha (2014) compares internet registers to pre-internet registers of Biber (1988). Facebook and Twitter posts cluster with spoken texts in his analysis. Thus, he says that they can be considered spoken language alternatives of internet-based registers (p. 102). The closest pre-internet registers to Facebook and Twitter are unprepared speeches, interviews, and personal letters (p. 102), which shows that they are very different from face-to-face and telephone conversations. Lastly, his findings also show that tweets are distinct from Facebook status updates, which is probably due to production circumstances like limited number of characters allowed in a message, abundant availability on mobile phones and similar devices, and use of Twitter as an instant messaging service (p. 102).

Clarke and Grieve (2017) study racist and sexist tweets using MD analysis. Their analysis identifies three dimensions in these tweets: "interactive, antagonistic, and attitudinal" (p. 8). In their opinion, racist and sexist tweets differ from each other, in that sexist tweets appear to be more interactive and attitudinal. Lastly, they opine that the antagonistic and attitudinal dimensions appear to be more relevant to abusive language because its traits include hostility, expression of opinion, and controversiality.

At least two studies have analysed tweets using MD analysis in the research tradition of World Englishes. Bohmann (2017) studies English tweets collected from 20 different countries in comparison to traditional spoken and written registers from the ICE. His results show that tweets have "a preference for colloquial marking, involved discourse, addressee-orientation and concrete information, and a dispreference for (third person) narrative presentation, factual

assertiveness, stance-marking, and abstract, conceptual language" (Bohmann, 2017, p. 358).

They are similar to spoken registers on some dimensions, while on other dimensions they are

more like written registers.

The second study has been conducted by Coats (2016) who compares a corpus of English

tweets from Finland with a corpus of global English tweets. He identifies two dimensions of

variation, i.e. 'interaction versus specification' and 'narration versus discourse negotiation' (p.

204). In his data, Finish English tweets are more interactive while the tweets from the global part

of the corpus are more informational and narrative.

2.4.3.4 MD studies on discussion forums and comments.

Unlike blogs, Facebook status updates, and tweets, discussion forums[14] are more likely to

have symmetric (two-way) exchanges, as Herring et al. (2004, p. 8) note. Similar assumptions

can be made for blog comments, where readers or blog authors can reply to comments and be

replied to. Many MD studies have looked at the linguistic and situational characteristics of

various forms of discussion forums and comments. This subsection summarises relevant points

from these studies.

Computer chat is similar to discussion forums, in that it allows symmetrical exchanges,

but it is also much more synchronous as compared to discussion forums as well as comments.

Jonsson (2015) has studied ICQ chat using Biber's (1988) dimensions of variation. She calls

computer chat and other similar registers 'conversational writing'. She finds strong similarity

between conversational writing and speech (which she calls 'conversational speech'). She writes

that on dimension one ('involved versus informational production') chat and spoken

---

[14] Facebook groups are a type of discussion forums.

conversations show involved and interactive discourse. On dimension two ('narrative versus non-narrative concerns') computer chat is slightly less narrative as compared to spoken conversations. Both registers are situation-dependent according to dimension three ('explicit versus situation-dependent reference'), while expression of persuasion is very low in both of them as per dimension four ('overt expression of persuasion') (Jonsson, 2015, p. 290).

An early form of discussion forums, i.e. bulletin boards, has been studied by Collot and Belmore (1996) using Biber's (1988) dimensions of variation. The authors explain that the language of bulletin boards is "neither spoken nor written in the conventional sense of the words" (Collot and Belmore, 1996, p. 14) – an observation confirmed by their results. On dimension one ('involved versus informational production'), bulletin boards are similar to interviews, unprepared speeches, and personal letters (p. 22). On dimension two ('narrative versus non-narrative concerns') the register is slightly non-narrative, while on dimension three ('explicit versus situation-dependent reference') the scores are near zero. Collot and Belmore (1996, p. 26) identify four situational features of bulletin boards that are possibly responsible for the linguistic findings: the participants have similar interests and common or shared knowledge; the communicative purpose is to ask for and provide information and to talk about particular issues; there are three types of role relationships among participants, i.e. addressers, addresses, and the audience; and the participants are separated by distance in terms of time and space.

Reddit is a social media platform where users can post and interact with each other in any language, though the majority of the discussions take place in English. It consists of sub-communities called 'subreddits', which are defined by specific topics (Liimatta, 2016). The communication structure of Reddit is similar to bulletin boards and discussion forums. Liimatta (2016) studies English communication on Reddit data using a new MD analysis. He identifies

three dimensions of variation in his corpus of Reddit posts. Dimension two ('personal versus factual focus') consists of features like contractions, predicative adjectives, and *that* deletions which have something to do with "personal, human matters" in the context of the data (Liimatta, 2016, p. 41). Registers on the opposite side of this dimension have more factual focus. According to the author, the third dimension ('informational versus involved style') is similar to Biber's (1988) dimension one. Lastly, the first dimension ('present abstract versus past narrative focused') has different types of modals on the positive side, and past tense on the negative side. Liimatta (2016) concludes that dimensions one and three are universal dimensions identified in many previous MD studies (cf. Biber, 2014), while dimension two ('personal versus factual focus') is a specialised dimension that distinguishes registers on Reddit.

Biber and Egbert's (2018) study also includes various types of discussion forums, which they call interactive discussions. Biber and Egbert (2018, p. 179) identify the following situational features of interactive discussions: the participants are those readers and writers that have an interest in the topic; the production involves limited planning time, no professional editing, and sensitivity to time; and the purpose of the discussions is question answering and discussion on opinions. Interactive discussions have been classified by manual annotators with other oral registers like interviews. These registers have a high score on dimension one, i.e. they have oral and involved discourse. Interactive discussions have the highest score on dimension six ('procedural/explanatory discourse') probably because the main focus of many discussions is to answer questions. The resulting answers contain step-by-step explanations about how to perform something (p. 184). The strongest features associated with interactive discussions include present tense, first person pronouns, subordinating classes (e.g. *if* clauses), and mental verbs that are used by the posters to seek answers for their questions (p. 186).

Lastly, two MD studies have examined various forms of comments:

Titak and Roberson's (2013) MD analysis includes reader comments from various websites like YouTube, Yahoo!, ESPN etc. They find that comments have high scores on dimension one ('personal narrative versus descriptive involved discourse') and two ('involved, interactive discourse'), which indicates the presence of a personal and involved discourse. Their third dimension is related to past versus present orientation, where reader comments incline towards the positive side indicating they are narrative and past oriented.

Ehret and Taboada (2018) analyse reader comments from a corpus of reader comments scraped from Canadian English newspapers in comparison to the ICE (Canada). They perform a new MD analysis using Biber's (1988) 67 lexico-grammatical features. They identify six dimensions of variation, including dimension one ('involved versus informational'), dimension two ('overt expression of opinion'), and dimension three ('formal persuasive presentation'). Their results exhibit that online comments have an informational instead of involved discourse, and are opinionated and non-narrative. The authors conclude that online comments are more like written registers with unique dimensions like overt expression of opinion (dimension two), which consists of linguistic constructions like 'verb *be* + predicative adjectives' to express opinion.

2.4.3.5 Other studies on interactive online registers.

A host of other studies have been produced over the decades in the tradition of CMC studies; e.g. Herring (2011) reviews a number of studies on CMC. Bieswanger (2016) has analysed English discussion forums to find out the effect of synchronicity on linguistic structure of immediately posted versus late posted thread replies. Social media platforms like Twitter have also been extensively studied from a computational linguistic and natural language processing

point of view, e.g. sentiment analysis (Titak and Roberson, 2013, p. 240). Lastly, studies on varieties of English have also utilised these data sources. For example, Dąbrowska (2013) studies language attitudes and sociolinguistic variation in Facebook posts of Indian and Polish English users. Bohmann (2016) studies *because* followed by a complement in Twitter discourse.

2.4.4 Spoken conversations

The studies mentioned in the previous subsections have observed similarities between interactive online registers and spoken conversations. This subsection very briefly overviews the multidimensional characteristics of spoken conversations and other relevant registers.

Biber's (1988) study includes face-to-face conversations, telephone conversations, and interviews. Face-to-face and telephone conversations are characterised by a highly involved discourse on dimension one ('involved versus informational production'). Interviews and spontaneous speeches (public discussions as per Biber and Gray, 2013) have a moderately involved discourse. Face-to-face and telephone conversations incline towards the negative side of dimension two ('narrative versus non-narrative concerns'), i.e. they are non-narrative. As per dimension three ('explicit versus situation-dependent reference'), these registers are situation-dependent as opposed to elaborated discourse. Lastly, dimension five ('abstract versus non-abstract information') characterises these registers as non-abstract.

Biber (2004) identifies three dimensions of variation that are specific to face-to-face conversations. These dimensions are as follows: 'information-focused versus interactive discourse', 'stance versus context-focused discourse', and 'narrative-focused discourse' (p. 22). The resulting dimensions show that while conversations are highly involved, situation-dependent, and non-narrative when analysed in a general corpus of English, the aspects like informational focus and narrativity appear once they are zoomed-in. Biber further identifies six

text types, including two general text types 'unmarked interactive' and 'unmarked context-focused' (p. 23), which make up the largest portion of texts.

Two other studies have included spoken as well as written registers of many varieties of English to conduct MD analyses. Xiao (2009) uses various Asian components of the ICE along with the British English component to identify nine dimensions of variation. His first dimension is similar to Biber's (1988) dimension of variation. The registers with the highest score on the interactive side of dimension one are private and public dialogue related registers. Bohmann's (2017) MD analysis – which is conducted on several varieties of English from the ICE corpora along with Twitter discourse – also shows that private dialogue has a highly involved discourse on dimension one.

Lastly, Hussain's (2016) study on an in-compilation ICE (Pakistan) also includes spoken dialogue related registers of ICE (S1 and S2). Her comparison of Pakistani and British English on Biber's (1988) dimensions shows that the Pakistani private dialogue section (S1A) is less involved on dimension one ('involved versus informational') and slightly more non-narrative on dimension two ('narrative versus non-narrative concerns') as compared to its British counterpart.

2.4.5 Section summary

This subsection surveyed previous research on various registers that are included in this research study. The first subsection reviewed previous research on various types of blogs with a special focus on MD studies. The relationship between blogs and press related registers (opinion columns and news reports) was discussed along with MD research on these ancestral registers. The next subsection started with a comparison of spoken conversations and interactive online registers included in this study. Later on, microblogging related registers (Facebook status updates and tweets) were discussed in the light of previous research. Lastly, multidimensional

characteristics of interactive online registers as well as spoken conversations were identified based on previous research.

This subsection also shown that most MD studies have been carried out on internet-based and non-internet-based registers originating from inner circle countries. Some studies have also been conducted on data from outer circle countries. However, Pakistani English online registers have not been included in these studies (except Bohmann, 2017 who included Pakistani tweets). Lastly, MD studies conducted on Pakistani English registers have also ignored internet-based registers. Thus, this section has confirmed the observation made in subsection 2.2.5, i.e. there is a research gap regarding the study of internet-based registers of Pakistani English.

2.5 Summary and Conclusion

This chapter has contextualised the present research study from four different aspects. The first aspect is related to the theoretical background. Research implications have been identified based on the discussed research work: two of the most important ones are the study of internet-based registers and focus on communicative context. The second aspect concerns the sociolinguistic and research background of the variety under study, i.e. Pakistani English. This aspect has highlighted a research gap regarding the study of internet-based registers in the context of Pakistani English. The third aspect is to connect the already-identified research implications to a methodological framework for the study of internet-based registers. The necessary background of the methodological framework, i.e. MD analysis, and its appropriateness for the present study have been discussed. The fourth and last aspect is to provide the state-of-the-art regarding previous research on the registers under study. Previous studies, especially MD studies, have been reviewed for this purpose. Moreover, the presence of a

research gap regarding the study of internet-based registers of Pakistani English has been

reaffirmed.

Chapter 3 Data Collection

In previous chapters, a detailed discussion has been provided regarding the research gap in the study of Pakistani English in relation to internet-based registers and the need to perform this particular study. Since Pakistani English on the internet has not been explored previously, no corpus of these registers has been compiled. The aim of this chapter is to document the decisions taken while compiling this corpus and designing the study. The following sections provide details regarding the corpus design, identification of online data sources, data downloading and cleaning, text sampling, and finalisation of the sampled texts. The issues related to data tagging and feature selection for MD analysis are also discussed. This chapter does not focus on the practical issues faced, for example, during EFA, which have been detailed in chapter 5. Additionally, the data collection related details of a small case study (cf. subsection 5.1.9) are not included in this chapter.

3.1 Issues before Data Collection

The corpus consists of two types of registers: internet-based registers or online registers and other similar registers or offline registers (cf. section 4.2 onwards). Additionally, similar registers from U.S. English were collected for regional comparison. The following subsections provide more details in this regard.

3.1.1 Selection of register categories

The selection of online registers was primarily motivated by the availability of these registers to Pakistani internet users. In other words, if Pakistani users produced texts in a particular register online, they were included. Hence, four different types of blogs (single- and multi-writer individual blogs, news blogs, technology blogs, and new media blogs), blog comments, Facebook groups and status updates, and tweets were included as online registers.

MD analysis is a comparative approach, i.e. it is necessary to compare the register(s) under study to other (similar) register(s) "to identify the distinctive characteristics of the target register[s]" (Biber and Conrad, 2009, p. 52). To understand the development and characteristics of Pakistani online registers, it was decided to compare them with similar but non-internet/ offline registers. Hence, newspaper columns, news reports, face-to-face conversations, talk shows, and interviews were also selected based on their availability for Pakistani English.

All registers were mainly distinguished based on their situational characteristics. For example, tweets and status updates were distinguished because they were published by the users on their Twitter and Facebook timelines respectively. Individual blogs were distinguished from, for example, news blogs because the former were published on separate websites and the latter appeared on newspaper/ news channel websites. Lastly, technology blogs were given a separate status because of their focus on technology, whereas new media blogs focused on viral content.[15]

3.1.2 Selection of a native variety of English for regional comparison

The idea of register comparison can also be extended to study variation in identical registers of different regional varieties. Such a comparison is necessary to find out how Pakistani online registers differ from their counterparts in a native variety of English. Additionally, it can also shed light on the relationship between online and offline registers between the two regional varieties. For this purpose, U.S. English was selected as the native English variety. British English was the obvious choice, but the unavailability of comparable data especially in the case of offline registers resulted in deciding otherwise. Moreover, this study focuses on, for example, the differences in terms of communicative purposes behind the use of registers in non-native and

---

[15] Operational definitions of all registers included in the corpus are provided in <u>section 4.3</u>.

native English varieties. Since both U.S. and British English are native varieties of English, there should not be much difference in terms of communicative purposes.

3.1.3 Use of existing corpora in data collection

A number of corpora have already been compiled for the register categories included in this corpus. This subsection provides details regarding the inclusion or non-inclusion of these corpora in the process of corpus compilation.

In the case of online registers, only one large corpus consisting of various types of blogs and other web content has been available for some years, i.e. the Global Web-based English Corpus or GloWbE (Davies and Fuchs, 2015). About half of the online data in this corpus consist of various types of blogs. Additionally, the inclusion of Pakistani blogs in the GloWbE made it a natural choice for blog data collection. However, it was not utilised due to a number of reasons. It has been collected programmatically using minimum human intervention, which also resulted in compromises regarding the reliability of data. These compromises include possible unreliability of the geographical origin of blogs (Nelson, 2015) and the inclusion of low-quality texts (Mair, 2015). Additionally, the proliferation of other registers, e.g. the inclusion of web forums in blogs (in the Singapore section of the corpus), was also found to be a problematic area (personal observation while using the corpus for another project). To avoid these problems, blogs were identified using prior knowledge and various blog directories. Additionally, this provided flexibility and control over the data to extract blog comments, which is another register category included in the corpus.

The texts for offline categories were collected from various existing corpora of Pakistani and U.S. English. Three different corpora of Pakistani English were used to collect data for three offline registers. The Pakistani component of the International Corpus of English (Hussain,

2016), which is incomplete at the time of writing, was used to get data for spoken conversations (face-to-face, talk shows, and interviews).[16] Additional texts for talk shows were collected from a colleague at a local university in Pakistan.[17] Ahmad (2016) compiled a corpus of Pakistani English news reports, which was used to extract data for news reports category. The newspaper columns category was compiled using the Pakistani Press Editorials corpus created by Alvi, Mehmood, and Rasool (2016). The data for U.S. offline registers (talk shows, interviews, news reports, and newspaper opinion columns) was mainly collected from the Corpus of Contemporary American English or COCA (Davies, 2008-).[18] Lastly, the Santa Barbara Corpus of Spoken American English (Du Bois, Chafe, Meyer, Thompson, and Martey, 2000) was utilised to extract data for U.S. face-to-face conversations.

3.1.4 Time period

The data collection was completed mainly in the year 2016. The time period of publication/ first availability of texts included in the online part spanned from 2009 to early 2016. The posts of one individual blog in Pakistani data were published between 2004-2007. Moreover, the data for new media blogs and the case study of technology blogs were collected in early 2018. New media blogs also included texts published in the year 2018. The corpora of Pakistani news reports and opinion columns were collected between the years 2014-2015. The data for spoken conversations was also collected after the year 2009. The data extracted from the Corpus of Contemporary American English was published between 2009-2012. The remaining

---

[16] The term *dialogues* could have been a better choice. Nonetheless, the term *conversations* has been used in a broad sense in this study.

[17] The texts were transcribed by M Phil English students at University of Gujrat, Pakistan.

[18] More than half of the texts in newspaper opinion columns were directly downloaded from the newspaper websites. The details are provided in subsection 3.2.5.

texts in newspaper opinion columns were published between 2012-2016. U.S. face-to-face conversations was the only register category with considerably older data, which was collected around 2000.

3.2 Corpus Compilation

This section provides details regarding the identification of sources for online registers, data downloading, text sampling for online and offline registers, text selection and sampling techniques, and the decisions related to text editing.

3.2.1 Identification of sources for online registers

The data collection process started with the identification of sources for Pakistani blogs data. The Pakistani blogs directory Pak Positive (www.pakpositive.com) was used to get access to Pakistani individual (single- and multi-writer) blogs. Additionally, the city oriented multi-writer blog website Metblogs (www.metblogs.com) was utilised. The websites for news blogs (News channels: AAJ, ARY, Dunya, Samaa; Newspapers: Dawn, The Nation, The News, Express Tribune), technology blogs (ProPakistani, TechPK, PakOrbit, TechJuice), and new media blogs (Parhlo, Mangobaaz, Trending PK, Images Blog by Dawn) were identified using Google, Facebook, and prior knowledge. Eleven Facebook groups were also identified through prior knowledge. About 95% of the users selected for status updates were identified from the user lists of Facebook groups, which were already scraped. Other users were selected from the friend list of the researcher. The profiles of Twitter users were identified by examining the livestream of tweets originating from Pakistan. The origin of Facebook and Twitter users was also ascertained to be from Pakistan at the same time. The ratio of male and female users in status updates and tweets (two third males and one third females) was maintained.

Similar to the Pakistani data, the sources identification process for U.S. English online registers started with the search of blog directories. Two blog directories were identified using Google search: Best of the web Blogs (www.blogs.botw.org) and Blog Flux (www.dir.blogflux.com).[19] Individual and multi-writer blogs originating from the U.S. were selected and downloaded. Topical diversity of blogs (beauty blogs, cooking blogs etc.) was maintained as much as possible as it was in the Pakistani individual blogs. Metblogs was also used to identify city-oriented U.S. multi-writer blogs. The websites for news blogs (New York Times, Chicago Tribune, Wall Street Journal, USA Today), technology blogs (AndroidPolice, Gizmodo, TechCrunch, Wired), and new media blogs (Viralnova, Mashable, Distractify, Upworthy) were identified using Google and Facebook. Facebook groups were identified using Facebook search and search terms like U.S. city names (e.g. Los Angeles), popular U.S. sports (e.g. soccer), U.S. presidential candidates for 2016 election (e.g. Bernie Sanders), mobile game (Pokémon go), and political issues (e.g. Black lives matter). The user profiles for status updates and tweets were selected similarly as for the Pakistani data. The same gender ratio was maintained in both status updates and tweets. Users tweeting in another language (e.g. Spanish) in addition to English were not included.[20]

3.2.2 Data downloading

The data downloading process started with Pakistani individual blogs, news blogs, and then technology blogs. The U.S. individual blogs, news blogs, and technology blog AndroidPolice were also downloaded in the same way. The website downloading software Darcy

---

[19] The blog directory Globe of Blogs (www.globeofblogs.com), which was used by Hardy and Friginal (2012) for their data collection, was not accessible at the time of data collection.
[20] Appendix IV contains the list of blog, website, group, and profile links used for data collection.

Ripper (Narrowteq, 2016) was used for this purpose. A small piece of software was written in C# programming language to extract text files from the HTML files downloaded with Darcy Ripper. To reduce the sheer amount of texts in certain Pakistani blogs, several measures were employed. For example, alternative months were deleted in individual blogs with a high rate of publication. In the case of news blog Express Tribune, the extracted texts were limited to one author only to reduce the number of resulting text files from more than 30,000 to around 1,700. Posts published before January, 2009 were also deleted. The author and commenter names for the Pakistani data were extracted separately beforehand. They were examined and foreign names were removed. Finally, blog posts and post comments were extracted in separate text files excluding authors and commenters with foreign names. Reducing the number of posts and reviewing author and commenter names was not performed for most of the U.S. blogs.

The other blog registers (U.S. technology blogs and all new media blogs) and Facebook registers (Facebook groups and status updates) were scraped manually. Equal numbers of posts were copied from each thematic subsection of the remaining U.S. technology blogs and all of the new media blogs. Around 50 posts from each technology and new media blog were copied, from which the final sample was later selected. The comments from technology blogs were also copied in the same way. No comments were extracted from new media blogs. The most recent group posts with a reasonably high number of comments were selected and copied from the Facebook groups. Similarly, individual status updates for Facebook users were copied to one text file per user. In case of Pakistani status updates, the posts with a larger amount of code switching were discarded. The text from shared posts in both Facebook registers was not included. The texts were later cleaned from, for example, commenter names and other robotic texts like button labels.

A C# program was written to download the 50-100 most recent English tweets of the selected Twitter users. Hence, one text file per Twitter user was created like the Facebook status updates.

Median number of words (dotted line) = 517
Mean number of words (solid line) = 614
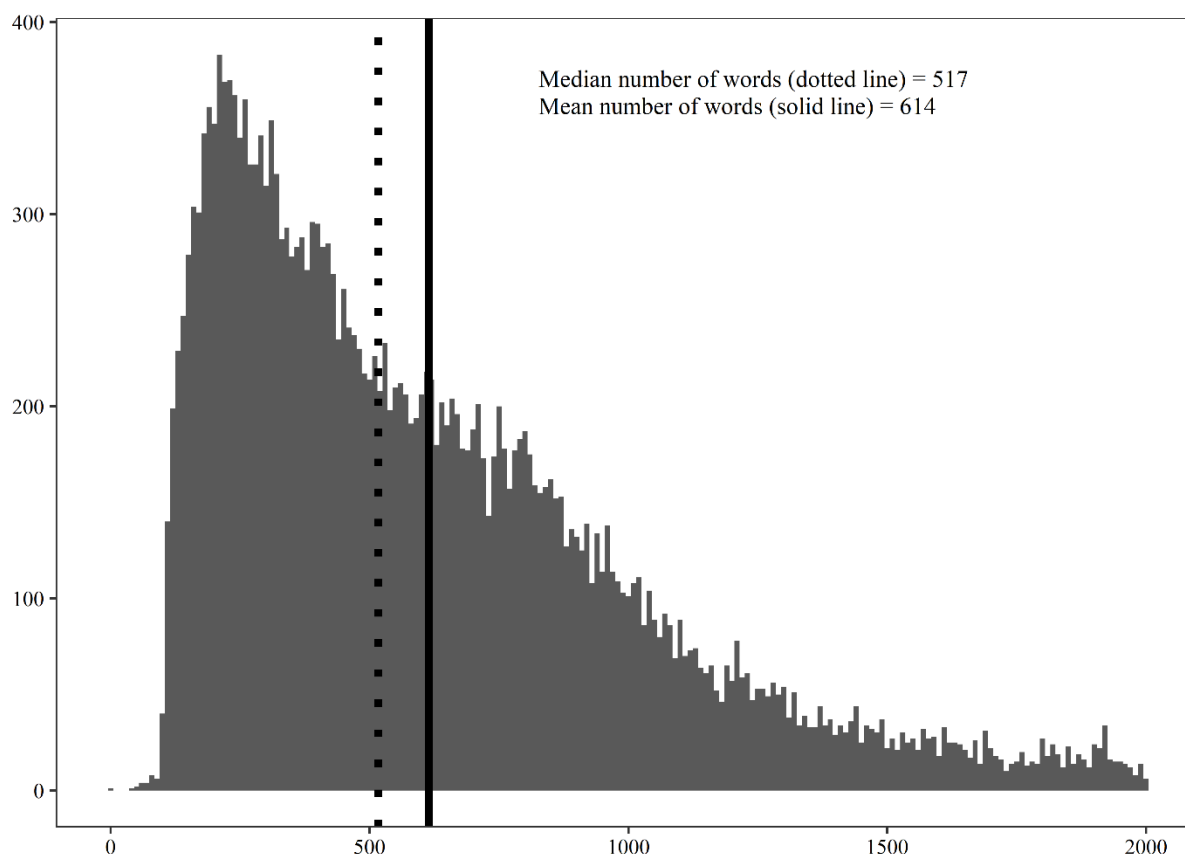
Figure 3.1 Word frequency histogram of online texts

3.2.3 Text and text length

As Biber, Egbert, Gray, Oppliger, and Szmrecsanyi (2016, p. 357) explain, each text is considered a separate observation in text linguistics (and in the tradition of MD studies). Two issues were of central concern during the process of data downloading and corpus compilation, i.e. what should be considered a text and what should be the text length. Initially, a single blog post or a Facebook group discussion/ comment thread was considered as one text. As individual

status updates and tweets contained very short texts, a number of status updates and tweets of one user were combined in individual text files.[21]

As figure 3.1 exhibits, a large number of texts were very short in terms of word frequency/ tokens per text. Individual (single- and multi-writer) blogs generally had short texts that sometimes contained less than 100 words. Pakistani technology blogs had shorter texts as compared to their U.S. counterparts. News blogs, however, generally featured longer texts. Pakistani Facebook group discussions also tended not to be very long. The same was the case with many comment threads on individual blogs of both regional varieties. As a result, it was decided to include those blog posts and comments which contained a minimum of 100 words/ tokens. The same was applied to Facebook groups by copying only those threats which were longer than 100 words. In the case of Facebook status updates and tweets, enough individual posts were collected to meet this requirement.

3.2.4 Sampling techniques

Ideally, a random sampling method was desirable, but it was not possible for many reasons. As it has been explained in previous sections, a large number of texts were downloadable only for individual blogs, news blogs, tech blogs, and comments for the Pakistani data and for some U.S. blogs. Otherwise, texts were manually copied because either automatic and mass downloading was not possible (e.g. Facebook groups, status updates) or it was restricted (e.g. U.S. technology blogs and news blogs like Chicago Tribune). Hence, a mixed sampling approach was adopted to select a reasonably large sample of texts. The first priority

---

[21] Later on, individual blog posts and discussion threads in single text files were deemed too short. As a result, several texts from the same blog were combined to one text file like Facebook status updates and tweets (cf. 3.2.7).

was to select an equal number of texts from each source/ website/ thematic subsection. Where

possible, texts were randomly selected from each blog/ website. As a last resort, opportunistic

sampling techniques were used to collect a minimum number of texts from a source/ website.

3.2.5 Text sampling I

The sampling took place in two stages, and this subsection explains stage I (see

subsection 3.2.7 for stage II). As usual, the process was performed firstly for the Pakistani data,

and was later replicated for the U.S. data. Three groups of online versus offline registers were

identified: individual blogs and news blogs versus columns, technology blogs versus news

reports, blog comments and Facebook groups versus conversations. Facebook status updates and

tweets were not included in these groups. Around 100 texts were collected for the latter two

categories from both regional varieties. The raw texts from each category in the Pakistani online

registers were tagged with Multidimensional Analysis Tagger (Nini, 2014). In-group and

between-group variation was observed to decide the minimum number of texts/ number of words

for each category.

Table 3.1 The summary of sampling stage I

Categories	Pakistani English		U.S. English	
	Words	**Texts**	**Words**	**Texts**
Blogs Ind. (Single & Multi-writer)	394,435	789 Texts (68 Blogs; 10 posts/ SW and 25/ MW blog)	346,105	791 Texts (74 Blogs; 10 posts/ SW blog and $\approx$180 posts from MW blogs)
Blogs News	138,115	200 Texts (8 Newspaper/ channel Blogs; 25 posts/ website)	171,041	308 Texts (4 Newspaper Blogs; $\approx$ 77 posts/ website)
Blogs Tech.	98,215	260 Texts (4 Tech. Blogs; 65 posts/ website)	100,928	140 Texts (4 Tech. Blogs; 38 posts/ website)

Categories	Pakistani English		U.S. English	
Comments	334,447	794 (80 blogs/ websites)[22]	342,517	747 (82 blogs/ websites)
FB Groups	167,358	(11 Groups X 50 Texts ≈ 502 Texts)	166,908	(9 Groups X 50 Texts ≈426 Texts)
Status Updates	71,079	(104 texts/users)	71,554	(108 texts/users)
Tweets	66,657	(115 texts /users @ 50 tweets per user)	70,392	(103 texts /users @ 50 tweets per user)
Columns	543,335	586 Texts (4 Newspapers)	529,650	587 Texts (8 Newspapers)
News Reports	107,697	78 Texts (5 Newspapers)	109,932	93 Texts (10 Newspapers)
Conversations	167,485	85 Texts (F2F, Talk Shows, Interviews)	516,015	111 (7 News channels, 20+ programs)

Table 3.1 explains the number of words and number of texts collected from each register category. A C# program was written to randomly select a given number of blog posts from each blog. For example, to select 10 blog posts from each single-writer blog, the program randomly generated a list of 20 blog posts and the first 10 posts were selected, as long as the text had a minimum of 100 words. The same process was repeated for multi-writer, news, and technology blogs of the Pakistani data. The texts from U.S. single-writer blogs were also selected in the same way. Every text published after 2009 was extracted from U.S. multi-writer blogs (especially from metblogs.com blogs). Only a stratified sampling technique was applied to U.S. news blogs (except Chicago Tribune, where opportunistic sampling technique was used). For example, an equal number of texts was collected from 14 different blogs on USA Today. In the case of comments, more than 50% came from individual blogs in both regional varieties, while

---

[22] Not every blog website contained sufficient number of comments. The number of blogs here just shows total number of websites considered for comment scraping. The actual number was much lower.

others were collected from news and technology blogs. Lastly, as the table exhibits, various measures were adopted for other online registers.

The texts for offline registers (except Pakistani conversations) were also selected using a stratified sampling technique. An equal number of texts was collected from each Pakistani newspaper (Daily Times, Dawn, Frontier Post, The News) and each author. Similarly, news reports were selected by stratifying the texts equally from five newspapers (Daily Times, Dawn Quetta Edition, Frontier Post, The Nation Islamabad Edition, The News Lahore Edition) and eight subcategories (business, crime, politics, cultural etc.). In the case of U.S. columns, 142 texts came from eight different newspapers included in COCA (New York Times, San Francisco Chronicle, USA Today, Atlanta Journal Constitution, Washington Post, Denver Post, Christian Science Monitor, Chicago Sun-Times), while the remaining texts were downloaded from New York Times, USA Today, and Chicago Tribune. Lastly, U.S. talk shows were sampled in the same way from more than 20 TV shows.

Table 3.2 Examples of text editing

Original Version (with Tags)	Post-edited Version
[…] So <}><<->yeh</-><+>yeah</+></}> <}><<->em</-><+>I'm</+></}> doing just fine..<}><<->em</-><+>I'm</+></}> happy with my life..<}><<->em</-><+>I'm</+></}> cool with the way my life is passing on...its all fun here you know...I have so many things to do that I don't find time to think over stupid stuff anymore...<indig>wasay bhi</indig> WHO Cares!! […] *Text# 2.txt (PK, Blogs Ind.)*	[…] So yeah I'm doing just fine..I'm happy with my life..I'm cool with the way my life is passing on...its all fun here you know...I have so many things to do that I don't find time to think over stupid stuff anymore...<*> WHO Cares!! […]
"Many advised me not to take on the issue because my life would be threatened but it reminded me of a Faiz poem, 'aaj bazaar meiN pa bajolaaN chalo,'" said an adamant Taseer in an interview with Arshad Sharif, two days before his death. […] *Text# 513029.txt (PK, Blogs News)*	"Many advised me not to take on the issue because my life would be threatened but it reminded me of a Faiz poem, <quote>" said an adamant Taseer in an interview with Arshad Sharif, two days before his death. […]

Original Version (with Tags)	Post-edited Version
[…] Speaking on the occasion, the President of Samsung Pakistan – Mr. J. H. Lee stated that: "The appointment of Airlink as a distribution partner is a major milestone for further enhancing our distribution capabilities and outreach. […]". Airlink's vast experience in marketing telecom devices will also play a critical role in further elevating the brand-image […] *Text# 514041.txt (PK, Blogs Tech)*	[…] Speaking on the occasion, the President of Samsung Pakistan – Mr. J. H. Lee stated that: <quote>. Airlink's vast experience in marketing telecom devices will also play a critical role in further elevating the brand-image […]
[…] I promised my brother I would post an excerpt from my memoir about him. Here it is. (Excerpt from Chapter 4, The Ghetto Will Follow You, Shades of Tolerance: A Biracial Love Story) In August, just weeks before school would be back in session, and Ronald and I would be back together, Ronald sat in his kitchen eating a hotdog inside a folded slice of bread slathered in mustard. *Text# BI_US_1.txt (US, Blogs Ind.)*	[…] I promised my brother I would post an excerpt from my memoir about him. Here it is. <quote>
[…]<5> Fireworks are literally going off in my neighborhood. <}><-->Lol</-><+>Laugh out loud</+></}>[...] <12> @Xxxxxxxxxxx <}><-->veeeeeery</-><+>very</+></}> cold. […] *Text# 633001.txt (US, Tweets)*	[…]<5> Fireworks are literally going off in my neighborhood. Laugh out loud [...] <12> @Xxxxxxxxxxx very cold. […]
<#> @fasee sab…sir aap halwa khaoo mast ho jao..khali truck ka kya kariyeah gaa…? <#> peshawar say kabul amreeka ke container move karoon ga…aur salary ke 10 % jihad fund main doon ga <#>10%what you are getting now or 10% of what you get on return from kabul….because that would be too much… to wait for…[…] *Text# CM_005.txt (PK, Comments Multi-writer)*	/// /// <#> 10%what you are getting now or 10% of what you get on return from kabul….because that would be too much… to wait for… […]

3.2.6 Editing and review

The sampled texts from online registers of both regional varieties were reviewed and

edited in the next step. The use of raw texts could create serious problems for the grammatical

tagger. The first concern in this regard was non-standard spellings particularly in interactive online registers. Especially in the case of tweets, the use of abbreviations like 'IMO' (in my opinion) could be mis-tagged. The other concern was the quotation of texts from sources other than the blog writer/ user, especially in the case of Pakistani technology blogs and individual blogs of both regional varieties. Press releases from foreign sources were simply pasted in bulk in Pakistani technology blogs. Similarly, long passages from Wikipedia or other sources like books were inserted in U.S. and Pakistani individual blogs' posts. Another relevant issue was robotic and spam texts specially in tweets and comments, and to some extent in Facebook status updates. The last issue was solely related to the Pakistani data, namely code switching and code mixing of local languages especially in interactive online registers. Hence, the sampled texts were reviewed and edited to overcome these problems as much as possible.

The text editor Notepad++ was used to review each and every text file. A spellchecker plug-in was used to identify non-standard spellings as well as non-English words or passages. The spelling correction tags proposed in the ICE written manual (Nelson, 2002) were used to tag non-standard spellings and corrections (e.g. table 3.2, row 1). The process of tagging was semi-automated by writing a simple program in C#. A list of spelling errors and their corrections, which was progressively developed from the data itself and by using online sources, was used as an input to the tagging program.[23] The texts were reviewed category by category, and the corrections list was updated whenever a new spelling error was found. The program was run on

---

[23] Appendix III contains the list of common spelling errors and their corrections used in this process. 95% of the spelling errors came from the data itself. Some corrections related to the omission of apostrophe in contractions were copied from the internet. An additional list of about 4,500 common spelling mistakes was also used, which was acquired from Wikipedia (https://en.wikipedia.org/wiki/Wikipedia:Lists_of_common_misspellings/For_machines).

the next category of texts to tag misspellings automatically. The tagged texts were then reviewed for additional as well as already tagged spelling errors. Newly found errors were again added to the corrections list, while any errors in the automatic tagging were corrected. This process was performed for all online register categories in both regional varieties.

The removal of spam, robotic, and non-blog texts could only be performed manually by scanning the texts by human eye. Tweets and comments containing spam or robotic texts were simply removed. If a lot of tweets from a user consisted of such texts, such a user was replaced with another one. Facebook status updates consisting of news articles and shared posts with quotations were excluded during the data collection process. Statements by foreigners and quotations from non-blog texts were removed and a tag (<quote>) was placed instead (e.g. table 3.2, rows 3 and 4).

Code switching and code mixing was also tackled by skimming and scanning the text files of Pakistani online registers. Quoted texts from Urdu poetry or direct speech in quotation marks was common in blog posts. Such passages were replaced with the quotation tag (e.g. table 3.2, row 2). Tweets, Facebook groups, and comments had the largest number of code switching instances. A phrase or a clause of non-English text in a fairly large text (e.g. a blog post) was tagged (<indig></indig>), as exemplified in table 3.1, row 2. If a very small part of a tweet or a single comment (less than 50%) consisted of non-English text, it was also treated in the same way. Otherwise, the tweet or the comment was replaced by a placeholder (e.g. table 3.1, last row). Arabic phrases like *InshaAllah*, *salam*, noun phrases consisting of proper names, and single word nominal borrowings from local languages were neither tagged nor removed.

The texts were then processed using a C# program to remove tags. The spelling errors and tags were removed and corrected spellings were retained. The tags marking non-English text

along with the text itself were replaced by a generic tag (<*>). It should be noted here that the process of editing and review was far from perfect. Ideally, at least two reviewers would go through each text. However, it was not possible due to the limited time and resources available for this study. Moreover, the texts for offline registers were not reviewed due to the same reason. Especially problematic were the spoken texts from the Pakistani data, which had been transcribed by nonprofessional transcribers (i.e. students). They most certainly would contain transcription errors, but reviewing and re-transcribing them was beyond the scope of this study.

Table 3.3 Data table after sampling II

Category	Pakistani English		U.S. English	
	Texts	Words	Texts	Words
BlogsInd.	68	394,650	74	362,342
BlogsNews	100	72,425	102	58,404
BlogsNM	100	42,341	100	40,233
BlogsTech	107	38,830	100	76,106
Columns	146	136,466	146	136,004
News	78	105,610	93	107,538
Comments	99	334,703	96	343,253
FBGroups	100	163,330	86	163,571
FBStatus	104	67,272	108	68,819
Tweets	115	58,202	103	62,086
Conversations	95	178,557	100	390,566
Total	1112	1,592,386	1108	1,808,922

3.2.7 Text sampling II

Once the data was ready for grammatical tagging and feature extraction, pilot studies were performed to analyse the results of EFA. It was observed that a very large number of texts in certain categories like individual blogs, comments, and Facebook groups could skew the

results. Another concern was regarding a low word frequency in many texts of the same register categories. Additionally, it was also decided to add more register categories in the data set. Firstly, a new subtype of blogs, i.e. new media blogs, was added. Additionally, the U.S. spoken conversations were balanced corresponding to the Pakistani data by adding face-to-face conversations and interviews.

The details of the final data set used in chapter 5 are provided in table 3.3. The number of text files in each register category was aimed to be around 100. As it is observable in the case of individual blogs, each blog was converted into one text file by merging all selected posts of it. The same was applied to comments and Facebook groups. Individual comment threads from each individual blog were merged into one text file. Additionally, news and technology blog comments were merged into 20 text files each. Around 10 discussion threads from each Facebook group were merged into one text file. Hence instead of 50 or so texts, the number was reduced to around 10 texts per group. The number of texts in news and technology blogs was reduced to around 100 texts instead of merging them. For example, every second text in Pakistani news blogs was retained, while every third text in case of U.S. news blogs was retained. Hence, a stratified sampling approach was adopted to retain as many original sources/ websites as possible. Opinion columns were reduced to around 150 texts to make 200+ texts when combined with news reports. Tweets, Facebook status updates, news reports, Pakistani face-to-face conversations, and interviews remained unchanged.

The data collection process for the newly added sub-registers followed the same procedure as described in previous subsections. The texts for new media blogs were manually scraped, sampled, and edited. The first 45 text files from the Santa Barbara Corpus were retained to match the number of texts in the Pakistani face-to-face conversations. Ten broadcast

interviews were also selected from COCA by looking at the metadata of TV talk shows to match their Pakistani counterparts. Lastly, the number of text files in TV talk shows was reduced to 45. Any program having one of the keywords "news, morning news, evening news, news report" in the title was removed to retain only discussion-oriented TV shows. Ten more texts were also added for Pakistani talk shows. However, the number remained limited to 40 due to the unavailability of additional texts.

3.3 Selection of Features

After the preparation of data, a set of linguistic features needs to be selected to study a given register using MD analysis. The purpose of this section is to provide an overview in this regard. As per Biber and Conrad (2009, p. 55), the features for register analysis are pervasive, frequent, and functional in nature. Pervasive means that such a feature is present throughout the text, rather than just in a part of it (e.g. start or end). Secondly, such a feature or group of features is more frequent in a given register as compared to other register(s). Lastly, since registers arise from situational/ functional variation, these features should be attributable to a particular function/ communicative purpose performed by/ in the text. The features can range from morphology, lexis and vocabulary, syntax and grammar, semantics and pragmatics to discourse and beyond. In other words, any linguistic characteristic having the above discussed attributes can be a potential candidate.

Practical difficulties and the limitations of corpus analysis tools, however, make it difficult to extract features at all linguistic levels. For example, prepositional phrases can also function as adverbials in a clause/ sentence. However, it is difficult to recognise the function of a prepositional phrase using a phrase chunker or parser. Hence phrase level adverbials of this kind are excluded from register analysis. Similarly, pragmatic or contextual meanings of a word can

only be ascertained by scanning each and every instance of the given feature. For example, the second person pronoun *you* can be used to address the audience or in place of the indefinite pronoun *one*. If *you* in these two functions is considered as two separate features, their frequencies cannot be extracted from hundreds of texts computationally. As a result, MD and other similar studies rely on various classifications of lexical items (e.g. discourse markers, nouns, adjectives, verbs, prepositions, and semantic groupings of word classes like nouns) and grammatical patterns (e.g. verb complementation patterns). Spelling variations (e.g. standard versus non-standard) and morphology (e.g. prefixes and suffixes) can also be added to this list.

The pioneering MD study (Biber, 1988, pp. 221-245) developed a comprehensive list of features and their functions using previous research on English speech and writing. These 67 lexico-grammatical features have been modified and enhanced in later studies like Biber (2006). Many features like synthetic and analytic negation, and existential *there* have been discarded. Other features like semantic groupings of nouns, verbs, adjectives, and adverbs and their complementation patterns have been added. These additions, which are the result of Biber and colleagues' work on English grammar (Biber, Johansson, Leech, Conrad, and Finegan, 1999), helped capture functions like stance marking in more detail.

At least two other studies have applied a similar methodology in the context of World Englishes. Xiao (2009) used 141 lexico-grammatical features in his analysis. The grammatical and structural features were derived from Biber's (1988) list. However, the semantic categories resulting from the USAS tagging system (Archer, Wilson, and Rayson, 2002) were used to enhance it. The main additions included various semantic classes of verbs, adjectives, adverbs, and expressions related to time, degree, power relationship etc. Many of these semantic categories overlapped with Biber's (2006) semantic classification of nouns and verbs etc.

Bohmann (2017) also used Biber's (1988) feature list to develop his own enhanced version of more than 200 lexical, morpho-syntactic, and discourse features. He relied on variationist research on various English dialects (e.g. Grieve, 2016; Kortmann & Szmrecsanyi, 2004) to add more features to the catalogue. In this process, he also changed Biber's feature groupings, e.g. using modal verbs individually instead of three semantic groupings (obligation, necessity, and prediction). However, most of the dialectal features were excluded later, because they were not pervasive and frequent enough. Other features like suffixes, prefixes, non-standard forms of pronouns, double prepositions, and hyphenation etc., were related to nouns (different types of suffixes) or standard (formal) versus non-standard (informal) variants. The features in his final analysis, like those of Xiao, showed a great deal of overlap with the enhanced MD model (Biber, 2006).

At the end, Biber's (2006) enhanced MD model was selected for the following reasons (cf. Appendix II for a list of these features). Firstly, the alternative feature sets included overlapping semantic groupings. Secondly, the grammatical and structural features (e.g. verb complementation, passive voice) remained the same. Thirdly, additional features like the 'discourse bin' category of USAS or the lexical bundles used by Egbert (2014) either did not correlate with other features or were not pervasive enough in the present data. The same would have been the case with Bohmann's (2017) features originating from the dialect and variationist studies of English. His distinction of standard versus non-standard spellings could be useful, but it had to be discarded due to overlapping concerns with Biber's features. Lastly, a grammatical tagger, namely the Biber Tagger, was available to extract feature frequencies of Biber's feature list, which was more convenient and time saving. The data table of feature frequencies resulting from this tagging was used in the EFA in chapter 5.

3.4 Summary and Outlook

This chapter has presented a detailed overview of the data collection and sampling procedures. Issues like the selection of register categories, addition of another regional variety of English for comparison, use of existing corpora for data collection, data downloading, cleaning, and editing, and feature selection have been discussed. The decisions taken in this regard and the motivations behind these decisions have also been documented in detail. The next two chapters, i.e. chapters 4 and 5, apply the MD analysis framework to analyse these data.

Chapter 4 Situational Characteristics of Online and Offline Registers

The study of register in the MD analysis framework consists of three main steps. As per Biber and Conrad (2009, p. 6), the first step is the situational analysis and description of the registers under study. The second step consists of linguistic analysis of the registers. The third step combines both of the previous steps to provide functional interpretations as well as explanations of the variation uncovered after the linguistic analysis. Egbert (2014, p. 39) notes that register analysis is a cyclic process in which the above-mentioned three steps are repeated many times during the process of research.

The aim of this chapter is to provide a situational analysis of the registers included in this study. For this purpose, a framework is defined and described in the following sections. The situational characteristics of the registers are then described using this framework. Additionally, short descriptions of individual registers also follow, which include the operational definitions of the registers as well as the highlights of situational differences between Pakistani registers and their U.S. counterparts. The information provided in this regard is largely based on the researcher's observations as an insider of Pakistani bloggers, online communities, and social media users. Some of this information is also inferred from the corpus itself and the meta information that was available at the time of data collection. Lastly, the comments about the situational characteristics of U.S. English are largely very general and based on the second type of information.

4.1 Defining a Framework for Situational Analysis

Biber (1988, p. 28-29) defines a framework for situational analysis based on previous studies, e.g. Hymes (1974), Halliday (1978), and Brown and Fraser (1979). He distinguishes eight main situational characteristics: "(1) participant roles and characteristics, (2) relations

among the participants, (3) setting, (4) topic, (5) purpose, (6) social evaluation, (7) relations of participants to the text, and (8) channel" (p. 29). These categories and their subcomponents are later on elaborated in the same study as well as in Biber and Conrad (2009, p. 40). While Biber's framework is not limited to a particular type of communication, Herring (2007) specifically focuses on the classification of computer-mediated communication. She presents a faceted classification scheme that consists of technology or medium related factors and situational factors. Her situational characteristics also draw heavily from previous research like Hymes (1974).

The set of situational characteristics devised for this analysis is mainly adapted from Biber and Conrad (2009). Herring's (2007) technology related factors, e.g. message size, have also been helpful in some cases. The registers can be divided in two categories according to their interactivity: various types of blogs, columns, and news reports are least interactive, while comments, Facebook groups, status updates, tweets and different types of spoken conversations are interactive. Two slightly different sets of situational characteristics have been applied to describe these two sets of registers. All of these characteristics are collectively listed below:

- Characteristics of participants/ writers:
    - age (general age group)
    - education level (high school student, university student, university graduate…)
    - ethnicity (mixed, specific ethnicity)
    - gender (male, female, no gender)
    - number of writers (texts are contributed by one, many, or institutional writers)
    - number of participants per 1,000 words (average number of participants in a conversation/ discussion thread)

- o professional background (volunteer or journalist etc.)

- o relationship among participants (friends, strangers, relatives)

- o residence (within country or abroad)

- o social class (lower-middle-class, middle-class, upper-class, other strata…)

- Accessibility: accessibility of the situation to general public

- Audience: size and social class of the blog/ column audience

- Communicative purpose: general purpose of communication, e.g. narrate, report, persuade….

- Editorial oversight: is there a chance that the texts are checked by an editor for publication policy compliance?

- Medium: the means of publication of blogs and similar registers, i.e. online, print, or both

- Message size: is there a limit on number of characters like in the case of Twitter?

- Separate (sub-)domain: do the blogs have a separate identity with a defined theme/ topic? Or the blog posts are assigned to a subsection of the website.

- Setting: public, private, semi-public

- Text length: average length of blog posts/ opinion columns/ news reports

- Topics: general topics discussed in blog posts/ discussions/ interactions

- Turns/ replies per 1,000 words: how many turns/ replies are there in a conversation/ discussion thread?

4.2 Online versus Offline Registers

The distinction between online and offline registers arose from the study of previous research on similar registers as well as due to the inclusion of (sub-)types of registers in this data. For example, Miller and Shepherd (2004, p. 13) suggest that newspaper opinion columns and

editorials are relevant to the family tree of blogs. Pakistani news blogs are published side-by-side with opinion columns on newspaper websites. Pakistani individual blog writers also send their blog posts to be published in these newspaper blog sections. Technology blogs, like news reports, also report about topics related to technology. New media blogs also include news report-like stories. Hence, different types of blogs are considered online registers in the scope of this study. Newspaper opinion columns and news reports are similar but offline registers, because they existed even before the advent of the internet.

Similarly, in the case of interactive registers, Jonsson (2015) finds some similarities between synchronous ICQ chats and spoken conversations. The most interaction-oriented registers included in this study (i.e. Facebook groups and comments) are certainly very different from spoken conversations in many ways including turn-taking (Herring, 2011). The other two interactive online registers, i.e. Facebook status updates and tweets, can also include discussion threads, but they are included here in isolation either as individual status updates or tweets/ tweet replies. However, a certain amount of similarity is expected with spoken conversations due to the interactive/ dialogue-oriented nature of these online registers. Hence, Facebook groups, status updates, tweets, and comments are considered online interactive registers in this study. Spoken conversations, namely face-to-face conversations, TV talk shows, and TV/ radio interviews have been selected as similar but offline registers, because they either include discussions or are dialogic in nature.

4.3 Situational Description of Online and Offline Registers

Tables 4.1 and 4.2 exhibit the situational characteristics of the registers under study as represented in the data. The first table is related to different types of blogs and similar offline registers, i.e. opinion columns and news reports. The second table focuses on interactive online

registers and their spoken counterparts. The information provided in the tables mainly concerns

the Pakistani data unless mentioned otherwise. In addition to these tables, the following

subsections comment on the important situational characteristics of each register category.

Where possible, illustrations are also provided in terms of screenshots.

4.3.1 Individual blogs

Individual blogs exist on subdomains on popular blogging platforms like blogspot.com or

wordpress.com. and are written in English. They might also be created using the popular

blogging software Word Press on privately purchased domain name and hosting. They have two

kinds: single-writer blogs and multi-writer blogs. The only difference between these two types of

blogs is the number of writers: while single-writer blogs have only one dedicated writer, there

might be more than two writers in case of multi-writer blogs. The most noteworthy situational

difference between Pakistani and U.S. individual blogs is their average lifespan.

Table 4.1 Situational characteristics of different blog types and similar registers (columns and news reports)

Situational characteristic	Ind. Blogs	News Blogs	New media Blogs	Tech Blogs	Columns	News reports
Writers						
Number	SW = 1, MW = many	many	many	many	many, institutional	institutional
Age	generally young (<40 years)	generally young (<40 years)	generally young (<40 years)	generally young (<40 years)	younger + above 40 years	--
Gender (percentage)	PK= 33(F), 50(M), 17(N) US= 40(F), 47(M), 13(N)	PK= 33(F), 55(M), 12(N) US= 38(F), 54(M), 8(N)	PK= 14(F), 54(M), 32(N) US= 38(F), 58(M), 4(N)	PK= 15(F), 70(M), 15(N) US= 24(F), 75(M), 1(N)	PK= 16(F), 80(M), 4(N) US= 17(F), 46(M), 37(N)	--
Education level	possibility of high school/ university students	high school/ university students, graduates, highly educated professionals	university graduates more likely	university graduates more likely	highly educated professionals more likely	--
Residence	mostly residing in the country + a few expat bloggers	can include expats	can include expats	more likely residing in the country	can include journalists residing abroad	--
Social class	urban lower middle, middle, and upper class	(upper class +) lower-middle and middle class	urban middle, and upper class (+ other social strata)	(upper class +) middle, and lower middle class	lower middle, middle, and upper class	--
Professional background	mostly volunteers	volunteers + professional (+ journalists)	professional + tech/ social media savvy	professional + tech savvy (/social media savvy + journalists)	professional + experienced writers + journalists (+	--

Situational characteristic	Ind. Blogs	News Blogs	New media Blogs	Tech Blogs	Columns	News reports
					public/ pvt. sector officials)	
Text length	PK= 499 US= 438	PK= 690 US= 555	PK= 423 US= 402	PK= 378 US= 721	PK= 927 US= 902	PK= 1349 US= 1156
Topics	self, surroundings, experiences, social issues, beauty*, religion, creative writing, politics*, law+, cooking*	social issues, religion*, creative writing, politics, word play+, economics+, law+, sports*	viral stories on social media*, social issues*, religion, celebrity gossip*, technology*	technology, science, gadgets, internet, tech companies (same for US English)	politics, social issues, foreign policy, education policy, sports… (same for US English)	criminal, sports, national, international…
Audience	a small community of individuals (generally <40 years) + big cities	wider audience + the general audience of newspapers + (smaller cities)	students and (young) social media users + urban lower middle, middle, and upper-class	students, young technology enthusiasts, entrepreneurs, bloggers, IT professionals + mixed social strata + big/ small cities	traditional audience of English newspapers + general blogs audience	traditional audience of English newspapers + general blogs audience
Communicative purpose	self-expression*, narrate, comment*, opine*, explain/ how-to, creative writing	comment*, opine*, report* creative writing, how-to, (tech/ film) review	informal comment, report, entertain (same for US English)	report, review/ describe, guide, how-to, explain (same for US English)	persuade/ opine, explain, summarise, …. (same for US English)	report, narrate/ recount events (same for US English)

Situational characteristic	Ind. Blogs	News Blogs	New media Blogs	Tech Blogs	Columns	News reports
Editorial oversight	No	yes	yes	yes	yes	yes
Separate (sub-) domain	Yes	PK= no US= yes in some cases	yes	yes	no	no
Medium	Online	online	online	online	online + print	online + print

Note: SW = Single-writer, MW = Multi-writer, F = female, M = male, N = no gender, ( ) = less certainty, * = also in U.S. English, + = only in U.S. English

Table 4.2 Situational characteristics of interactive online registers and similar registers (face-to-face conversations, interviews, and talk shows)

Situational characteristic	Comments	FB groups	FB status	Tweets	Face-to-face	Interviews	Talk shows
Participants							
Age	generally young (<40 years)	generally young (<40 years)	generally young (<40 years)	generally young (<40 years)	generally young (<40 years)	generally young (<40 years)	younger + above 40 years
Social class	urban lower middle, middle, and upper class (Ind.) + lower strata (N + Tech)	urban lower middle, middle, and upper class (+ small cities)	urban lower middle, middle, and upper class	urban lower middle, middle, and upper class	lower middle and middle class (+ other strata)	lower middle, middle, and upper class (+ other strata)	lower middle, middle, and upper class (+ other strata)
Background	students + young entrepreneurs + tech enthusiasts + bloggers + (political activists + other)	uni. students + PhD scholars + young entrepreneurs + tech enthusiasts, bloggers + pet owners + English teachers (+ social activists + other)	entrepreneurs + tech enthusiasts + bloggers (+ English teachers + famous personalities)	students + political/social activists (journalists + bloggers + expats + other)	(post-) graduate students	(part-time) radio hosts + celebrities + artists + famous personalities	journalists/ anchors + celebrities + retired/ in-service public & private professionals + politicians + analysts + highly educated individuals (+ students)

Situational characteristic	Comments	FB groups	FB status	Tweets	Face-to-face	Interviews	Talk shows
Residence	mostly residing in the country + (some expats)	can include some expats	mostly residing in the country + (a few expats)	mostly residing in the country + (a few expats)	majority residing in the country	majority residing in the country	majority residing in the country + (a few expats)
Ethnicity	mixed	mixed	mixed	mixed	generally Punjabi	mixed	mixed
Relationship	generally strangers	strangers (+ friends)	strangers (+ friends)	strangers (+ friends)	generally friends, relatives[+]	acquaintances/ colleagues/ strangers	acquaintances/ colleagues/ strangers
Number per 1000 words	PK= 13 (Ind.), 16 (N), 21 (Tech) US= 15 (Ind.), 17 (N), 18 (Tech)	PK= 11.35 US= 28.21	--	--	PK= 1.4 US= 0.58	PK= 1.31 US= 1.93	PK= 1.16 US= 1.89
Turns/ replies per 1000 words	PK= 18 (Ind.), 18 (N), 28 (Tech) US= 17 (Ind.), 20 (N), 23 (Tech)	PK= 18.73 US= 48.71	--	--	PK= 47.48 US= 42.87	PK= 27.1 US= 22.24	PK= 30.58 US= 22.89
Message size	generally no limit	around 8000 characters	around 5000 characters	140 characters	context dependent	context dependent	context dependent
Topics	personal, social issues, education,	foreign study, admissions, job ads,	job ads, business, advice,	self, politics, social issues, religion,	religion, health, university	professional life, personality….	politics*, foreign policy, international

Situational characteristic	Comments	FB groups	FB status	Tweets	Face-to-face	Interviews	Talk shows
	beauty, (religion, politics, law, cooking, tech)*	general help, pets, technology, start-ups, (study, Pokémon, politics, community, sports, food)*	religion, social issues, (politics, self)*	education, business, (job ads, student life)	life, education, other social issues		relations, social issues, current affairs*, showbiz*, (health, beauty*)
Communicative purpose	interact*, respond/ react*, debate/ persuade*, ask (for help/ tips)	ask (for help), discuss, inform, advertise (jobs/ events)	advertise, advise, interact*, express*	advertise, react, inform, express*, interact*	interact*, discuss, (debate/ persuade), explain*, narrate*	answer/ respond	debate/ persuade, answer, discuss, explain*, inform* (/report*)…
Setting	public	PK= generally closed US= generally public	generally public	public	generally private	public	public
Accessibility	anyone with an internet connection and username	anyone with an internet connection and username/ membership	anyone with an internet connection and username	anyone with an internet connection and username	limited opportunities	limited to certain social strata	limited to certain social strata

*Note: Ind.= Individual blog comments, N= News blog comments, Tech= Technology blog comments, ( ) = less certainty, * = also in U.S. English, + = only in U.S. English*

Table 4.3 Average life of individual blogs in years

Region	First Post Published	Last Post Published	Average Life
PK	2009	2014	5
US	2008	2015	7

Adopted from Shakir and Deuber (2019)

As table 4.3 shows, Pakistani individual blogs have a shorter lifespan as compared to their U.S. counterparts. Most Pakistani individual blogs were started at the end of the 2000s. Social media platforms like Facebook, Twitter, and Tumblr also started getting popularity around 2010-2012. Apparently, individual blogs lost their popularity with the increasing use of microblogging and other social media websites.

Multi-writer blogs are a special case of individual blogs, where several writers come together to run a single blog website. The most prominent subtype of multi-writer blogs included in this data comes from Metblogs.com, a project to promote city-oriented blog journalism. The writers are groups of individuals or other bloggers living in the same city and covering local issues in their blog posts. Pakistani metro blogs include three city blogs from Islamabad, Karachi, and Lahore – the first being the federal capital and the latter two being the most populated cities in the country. The U.S. data, in contrast, includes more than 10 metro blogs from various U.S. cities. Pakpositive.com is also a noteworthy case in Pakistani multi-writer blogs, where Pakistani bloggers have been submitting their blog posts since around 2004. Lastly, as figure 4.1 reveals, Metroblogs.com had more than 50 city-oriented blog websites at its peak. These blogs were mostly abandoned around the year 2010 in the U.S. data, while their Pakistani counterparts lived on until around 2012.

Figure 4.1 A screenshot of Metblogs.com main page

4.3.2 News blogs

News blogs are blog posts that are published on English newspaper websites under the heading of 'Blogs' or a subdomain, e.g. 'blogs.xyz.com'. In the case of the Pakistani data, all of these blog posts are published under *Blogs* sections on newspaper websites. Pakistani news channels that do not have their own corresponding newspapers also provided English blog sections on their websites. However, these blog sections are not continuously updated any more like the newspaper blog sections. Pakistani individual blog writers also send their blog posts to be published in the blog sections of newspaper websites. Additionally, sometimes newspapers also ask their audience to share their thoughts regarding a particular issue. The social media team of Express Tribune Blogs does that, as it can be seen in the screenshot provided in figure 4.2. In case of the U.S. data, the blog posts exist in the form of thematic blogs in three out of four newspaper websites, e.g. economy blog, law blog, wordplay blog. In the case of USA Today, it appears that the blogs section is not maintained any more, as it is not accessible from the main website.

Figure 4.2 A screenshot of Express Tribune Blogs social media post

4.3.3 New media blogs

New media blogs focus on interesting news stories and other content in English that their readers might find engaging and humorous. The blog posts are written in a very informal way. They frequently employ visual media – like memes, videos, animated gifs, screenshots, and quotations of social media user posts from Twitter and Facebook – to complement the textual content. In many cases, the blog posts practically consist of several one-line comments on a series of such non-textual elements. In the case of Pakistani blogs, code switching is also employed in memes and the blog post text. These websites also maintain a very strong social

media presence by sharing interesting posts to keep their audience engaged on the social media page. Different new media blogs may have varying degrees of focus on the types of content described above. Hence, they might also publish on topics like technology, which is generally the domain of technology blogs. Out of eight blog websites included in this category, at least two from each regional variety can be considered as focusing mainly on this type of content. Most of the Pakistani new media blogs emerged after 2014. Some of their U.S. counterparts, e.g. mashable.com, are much older in comparison. Blog posts also include advertisement sections.

Relax, the rumor that PTA is blocking all phones is false.
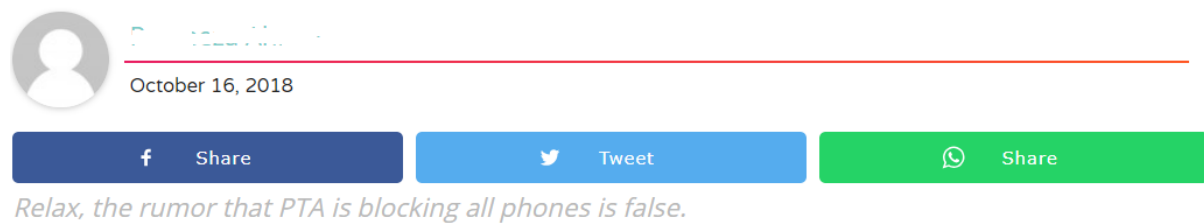
There have been WhatsApp forward messages going around saying that if you do not complete your cell phone's IMEI registration process, it will mean that PTA will block your phone.

Source: Broadway Video

This was further aggravated by PTA sending official messages out saying that non-compliant will stop working after 20th October, 2018.

Figure 4.3 A screenshot of Mangobaaz.com post

Figure 4.3 provides a screenshot of a blog post from the Pakistani new media blog

mangobaaz.com, which copies a number of layout elements from U.S. website mashable.com.

One of several animated gifs included in the original post is also observable in the screenshot.

4.3.4 Technology blogs

Technology blogs publish content related to technology, gadgets, computers, mobile

phones, deals to buy electronic gadgets etc. Most of the Pakistani blogs and all of their U.S.

counterparts also maintain social media pages, where they share the latest blog posts. In the case

of the U.S. data, additional content like video podcasts and reviews are also shared on social

media in addition to the original websites. Pakistani technology blogs also publish, for example,

on offers by local telecom companies. The interest to know about such offers and topics like new

smartphones, operating systems etc. might attract an audience from the lower strata of the society

as well. Their posts also include advertisement sections in between, above, or below the content.

4.3.5 Opinion columns and news reports

As the names suggest, opinion columns and news reports are published in the respective

sections of English newspapers from both regional varieties on the internet as well as in the print

editions. Most of their situational characteristics have been described in table 4.1. In the case of

U.S. opinion columns, the sub-register of letter to the editor also exists in about 5% of cases. The

Pakistani data consists almost exclusively of newspaper columns.

4.3.6 Comments

Comments are responses and reactions in the English language posted under the blog

posts of individual blogs (single- as well as multi-writer), news blogs, and technology blogs.

There may be differences in the social backgrounds of commenters from individual blogs as

compared to news and technology blogs, because of the latter two having a wider reach among

the audience. Conversely, the commenters of individual blogs are more likely to know each other as compared to the commenters from other types of blogs. Technology blogs in the Pakistani data are different from their U.S. counterparts, in that the commenters may also ask the blog writers for help regarding their technology related problems. Resultantly, Pakistani technology blogs have more users and more replies as compared to their U.S. counterparts.

4.3.7 Facebook groups

Facebook groups are topic specific discussion groups in English on the social media platform. Apart from the difference in terms of topics and communicative purposes as described in table 4.2, Pakistani Facebook groups generally are created at the national level. The members belong to urban centres all around the country or reside abroad. Some groups, like Study Life in Germany for Pakistani Students, have more than 100,000 members. U.S. groups, on the other hand, are mostly local, for example focusing on a specific community. Some groups related to topics like politics might also be temporary in nature. As a result, the number of members in these groups is generally less than 10,000. Additionally, two situational differences are observable between both regional varieties: firstly, discussion threads in the Pakistani data involve fewer contributors and replies as compared to their U.S. counterparts. Secondly, Pakistani participants are generally younger as compared to their U.S. counterparts, who are more likely to include older members as well.

4.3.8 Facebook status updates

Facebook status updates are the posts by the respective users on their Facebook walls in English. Apart from the situational characteristics described in table 4.2, Pakistani status updates users are more likely to be younger as compared to their U.S. counterparts. Status updates do not include comments and replies.

4.3.9 Tweets

Tweets are Twitter posts or post replies in English by the respective users. The maximum length of Twitter posts was 140 characters at the time of data collection in 2016, which was later increased to 280 characters. The most noteworthy difference between both regional varieties is again the diversity of age groups in the U.S. data, where Pakistani users are mostly youngsters.

4.3.10 Spoken conversations

Spoken conversations have three sub-registers: face-to-face conversations, interviews, and talk shows. The Pakistani face-to-face conversations are discussions between university postgraduate students. The participants of the U.S. counterparts, on the other hand, have more diverse backgrounds and intimate relationships among themselves. The other notable difference is the choice of topics, which are more abstract in the case of the Pakistani conversations. The U.S. data has fewer speakers per thousand words and more turns per speaker as compared to the Pakistani data. Resultantly, overlapped, phrasal, or word level utterances occur less frequently in the Pakistani data, and the turn taking process happens quite smoothly. The communicative purposes also seem different between both regional varieties, where Pakistani participants are generally interested in discussing or opining, while their U.S. counterparts also engage in explanation and narration etc.

The Pakistani interviews are taken from an English FM radio channel, City FM 89. The interviewer is an RJ/DJ or radio presenter, while the interviewee is a famous personality or a celebrity. Conversely, the U.S. data for interviews consists of TV shows where the hosts sometimes conduct exclusive interviews with politicians, celebrities or other famous personalities.

The talk shows are English broadcast discussions aired on public and private TV channels in Pakistan. Their U.S. counterparts come from various TV shows included in the Corpus of Contemporary American English (Davies, 2008-). The main situational differences between both regional varieties in this subcategory are related to the format of programs and the kinds of topics covered during the discussions. The Pakistani talk shows rarely include, for example, video packages or news reports, which would later be used for discussion or comment by the participants. The U.S. TV shows, on the other hand, sometimes do include such parts. Moreover, some Pakistani TV shows follow a debate-like format, where a topic is given to the participants and they are supposed to argue for or against it. Additionally, a reasonable number of programs included in the data focus on topics that are related to international relations and/or foreign policy etc. Though the U.S. TV shows do include topics like politics, the range of topics is certainly more diverse as compared to the Pakistani data.

4.4 Summary and Conclusion

A detailed situational analysis has been carried out in this chapter by establishing a framework for situation analysis, defining each register category, and describing their situational characteristics. The registers have been divided in two groups: non-interactive and interactive registers. Another distinction has been made in terms of online versus offline registers. Non-interactive registers consist of individual blogs, news blogs, new media blogs, and technology blogs versus their similar offline counterparts, i.e. opinion columns and news reports. The situational analysis has distinguished salient differences between different blog types and their offline counterparts. Similarly, interactive registers – comments, Facebook groups, status updates, and tweets – and their offline counterparts – face-to-face conversations, interviews, and talk shows – can also be distinguished in terms of situational characteristics.

The most prominent situational differences between online versus offline registers of Pakistani English appear to be as follows:

- Accessibility: easy access to internet-based registers versus limited opportunities to produce texts in opinion columns or take part in, for example, face-to-face conversations and talk shows;

- General age group: young internet users as compared to potentially older writers (of opinion columns) and participants (of talk shows);

- Social background: generally less-known internet users versus well-known journalists, highly educated professionals, celebrities, and other famous personalities.

Additionally, the most noteworthy situational differences between online registers of Pakistani and U.S. English appear to be in terms of topics, communicative purposes, and the characteristics of the participants. Moreover, the U.S. data also seems to be more diverse with regard to these characteristics.

This chapter has provided a situational basis for the linguistic analysis that is carried out in chapter 5. It will also be helpful in the functional interpretation of the linguistic results and in explaining the differences between register categories and regional varieties.

Chapter 5 Data Analysis

The aim of the present chapter is to try and understand the linguistic and consequently the functional characteristics of Pakistani English online registers in comparison to their U.S. counterparts, as well as similar offline registers. This is achieved using a comprehensive analysis of the data using three quantitative techniques, namely MD analysis, or in other words EFA, CA, and CDA. Correspondingly, there are three parts, and each one provides an overview of the method of analysis used, a detailed analysis of the data, and finally a short summary concluding the section. Where applicable, nonparametric ANOVA (Kruskal Wallis ANOVA) and post-hoc Dunn's test with Bonferroni correction are applied to test the significance of differences in various subcategories of the data.[24]

5.1 Multidimensional Analysis

5.1.1 Two types of MD studies and motivation for a new MD analysis

There are eight methodological steps in an MD study (Biber and Gray, 2013). The studies that involve a 'new' MD analysis perform all 8 methodological steps, which start from data collection and end at interpretation of dimensions resulting after an EFA. Other studies, however, use dimensions already identified in a previous MD study (p. 403). Biber's (1988) study was performed on a general corpus of spoken and written English. It has been used by several previous MD studies to compare new registers on these general dimensions of variation. These dimensions are also used here as a point of departure to explore linguistic and functional variation in the data. Like in most MD studies, the first dimension, i.e. involved versus informational production, is the strongest of all. Most spoken-like or oral registers scored high on

---

[24] R package dunn.test (Dinno, 2017) was used for this purpose.

the positive side, while written and informational registers had high scores on the negative side (Biber, 1988). Figure 5.1 shows online and offline registers of the present data on this dimension.
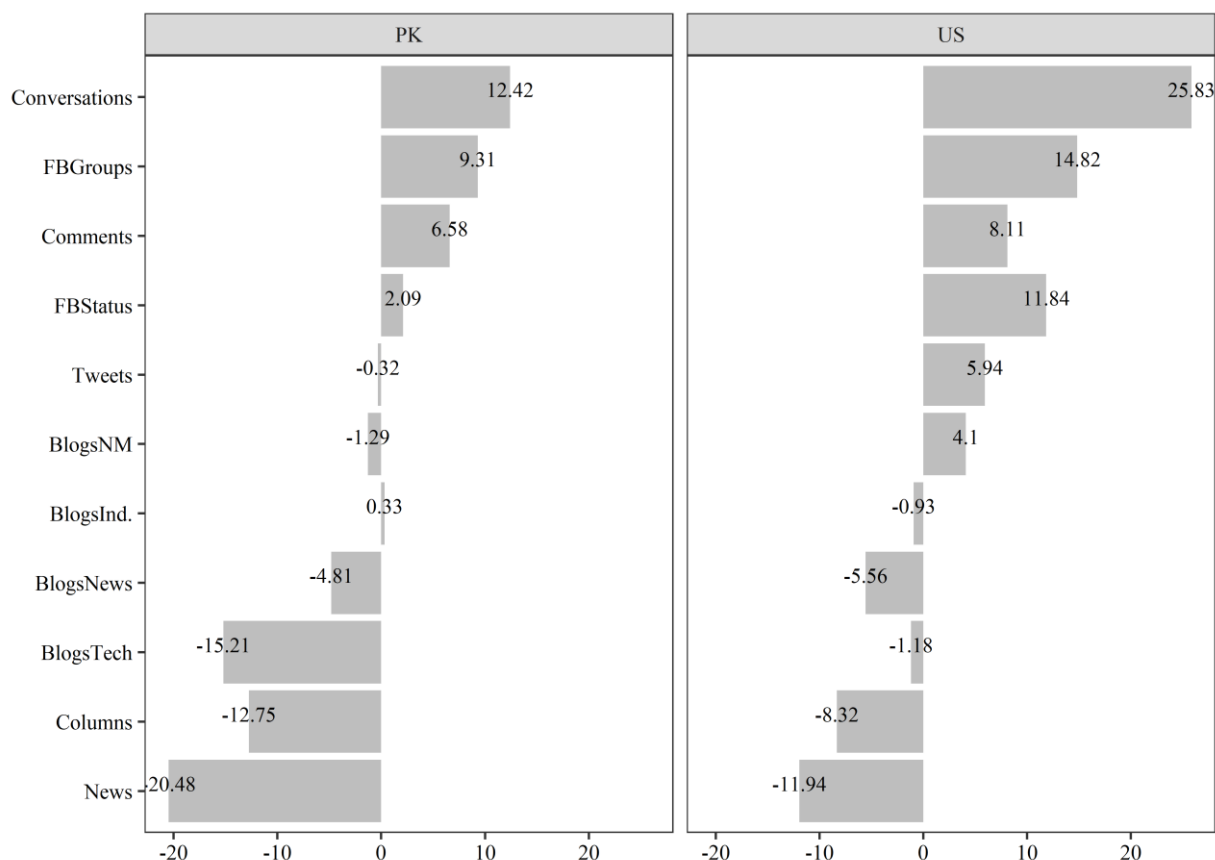
Figure 5.1 Online and offline registers on Biber's (1988) dimension 1 Involved versus Informational Production

The figure places offline registers on opposite ends, while online registers are in between the two extremes, i.e. involved versus informational orientation. The spoken-like online registers like Facebook groups and comments have high scores on the positive side, while blogs have high scores on the negative side. A general regional trend is also visible, where Pakistani registers are more informational as compared to their U.S. counterparts. While dimension 1 provides a clear overview of regional and functional variation in the data, other dimensions might not be so relevant (cf. Appendix I figures 1-4). For example, dimension 2 is not much relevant because most of the registers fall on the negative, i.e. non-narrative, side.

There are a number of reasons which provide motivation for a new MD analysis. The point of relevance is of course the very first one. Apart from the first dimension, other dimensions in Biber's (1988) model might not be much relevant to the current set of registers. A new MD analysis, based on a new EFA, can generate dimensions that are specific to these registers. The second reason is the set of variables used in the above mentioned study. Biber used 67 lexico-grammatical features in his pioneering study. However, later on he expanded this set to include more than 130 features (Biber, 2006). These features add semantic classes of nouns, verbs, adjectives, and adverbs along with various types of dependent clauses. The extent of detail provided by this expanded set of features enables a more fine-grained analysis of the data and identification of communicative purposes. The new dimensions identified are also helpful in the second part of the quantitative analysis, where a CA is performed based on dimension scores.

5.1.2 New MD analysis: data screening and selection of linguistic features

The process of an EFA starts with the selection of variables and screening the resulting data table. Egbert and Staples (2019) provide a detailed guide to perform an EFA for MD studies. Since this study limits itself to the set of features tagged by the Biber tagger used in previous studies (Biber, 2006; Biber, 1988), the selection of features was an easy process. As per the recommendations of Egbert and Staples, around 148 lexico-grammatical features were reduced to around 120 features by excluding overlapping categories (e.g. all adjectives, all verbs, all nouns etc.). The next step was to check for collinearity and factorability of the data, which was performed using SMC and Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy respectively.[25] The data was also checked for linguistic features with very low mean scores, very

---

[25] All statistical and mathematical operations were performed in programming language R (R Core Team, 2013).

high standard deviations and zero occurrence in more than 50% of the texts.[26] Additionally,

feature communalities were also taken into consideration during the iterative process of EFA and

features with very low values (less than 0.05) were removed.[27] Lastly, certain linguistic features

were combined into one category to compensate for their low occurrence and/or low

communality values, e.g. *to* clauses controlled by stance nouns and stance adjectives were

combined into one category. At the end, a set of 93 linguistic features was deemed fit for EFA.

5.1.3 New MD analysis: performing EFA

The EFA was performed using psych package (Revelle, 2017) in R. The first step in an

EFA is to select a factor extraction method. Since the data was not normally distributed, principal

axis factoring was used as previous studies have indicated this method to be robust against

normality violations (Bohman, 2017; Costello and Osborne, 2005; Egbert and Staples, 2019).

The selection of factor rotation method was the next step. Promax, which is an oblique rotation

method, is recommended by previous MD studies. This rotation method allows the resulting

factors to be correlated to a certain extent. Since everything in language is connected and

correlated with each other (Biber, 1988), this method was selected. The next step was to identify

the number of factors to extract. A number of methods are available for this purpose, e.g. parallel

analysis, scree test, and retaining factors with eigenvalues above the threshold value of 1

(Costello and Osborne, 2005). An examination of scree plot, factors above eigenvalue 1, and the

interpretability of resulting factors have been recommended by previous studies like Biber and

---

[26] Since average length of texts in some categories was less than 400 words, the absence of certain linguistic features was expected. However, features occurring in a very small percentage of texts were marked for removal.

[27] "A communality is the variance accounted for by the factors. The greater the communalities, the greater the variance explained by the factors." (Egbert and Staples, 2019, p. 129)

Egbert (2016). The same criteria were applied here, i.e. a scree plot was generated and then various factor solutions ranging from three to nine factors were examined for interpretability and variance explained. Finally, a factor solution with five factors was extracted. The final factor solution includes 63 linguistic features, the variance explained is 23%, and the KMO value is 0.83 (Meritorious).
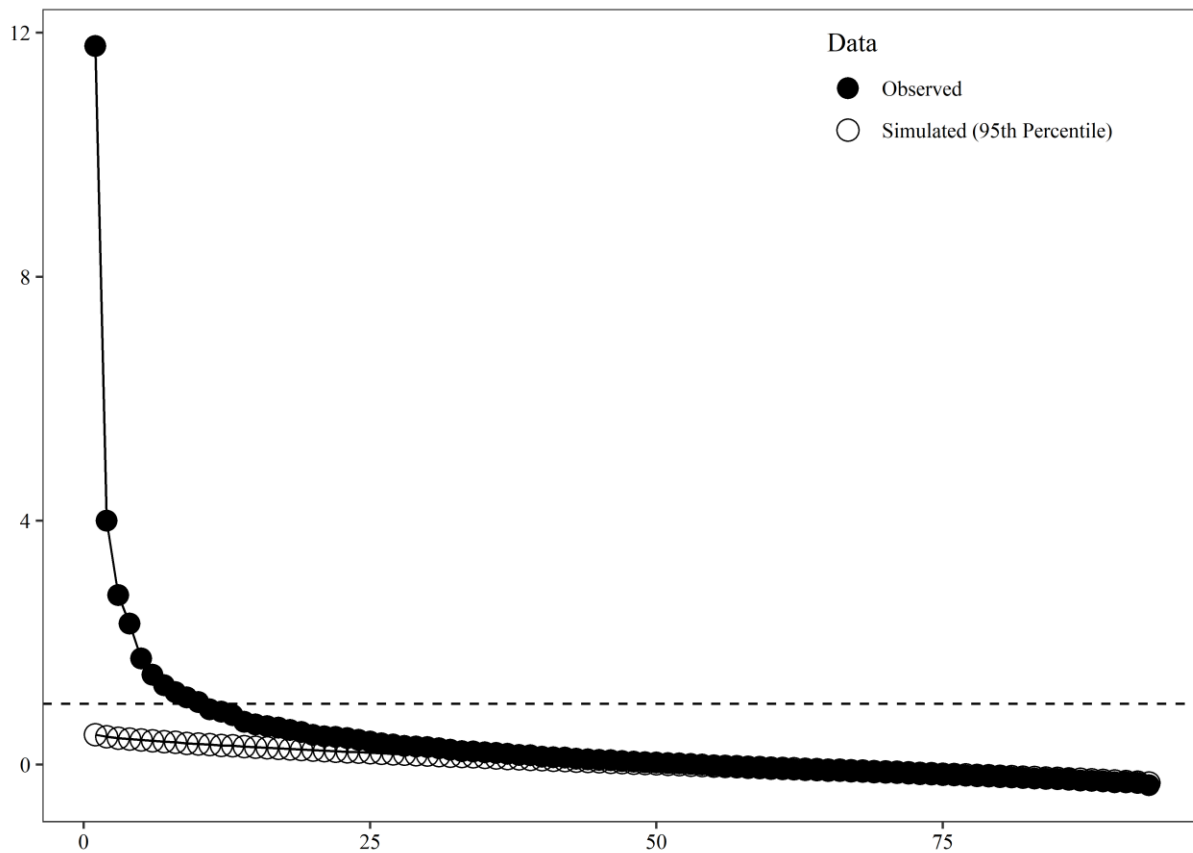
Figure 5.2 Scree plot used to identify number of factors

Figure 5.2 presents a scree plot with eigenvalues on the y-axis and the number of factors on the x-axis. Around eight factors lie above the minimum threshold limit of 1. However, after the fifth factor the variance explained by the factors becomes very small and consequently it becomes harder to interpret these factors. Hence, five factors were selected as the optimal solution. Table 5.1 provides an overview of the correlation among the resulting factors. The

highest correlation is between factor one and the other four factors. A high correlation among factors is not surprising considering the interconnected nature of language itself. Additionally, this phenomenon has been observed in previous MD studies as well (e.g. Biber and Egbert, 2016; Bohman, 2017).

Table 5.1 Correlation among resulting factors

	Factor 1	Factor 2	Factor 3	Factor 4
Factor 2	0.30			
Factor 3	-0.26	-0.05		
Factor 4	-0.18	0.11	0.14	
Factor 5	0.26	0.21	-0.07	-0.08

5.1.4 Calculation of dimension scores

After the completion of the factor analysis, the next step was to calculate factor/ dimension scores for each text and subsequently for each category for further quantitative analysis. Biber (1988) and several other MD studies have used a dimension score calculation method called "sum scores standardised variables" by Distefano, Zhu and Mîndrilă (2009, p. 8). According to the authors, the calculation procedure is as follows: the variable frequencies are converted to z scores with a mean of 0 and standard deviation of 1; the individual variable scores are then summed up. The variables with factor loadings below a threshold can also be discarded in this procedure. Most MD studies have used a similar method with a couple of exceptions. The sum of variables on the negative side is subtracted from the sum of variables on the positive side. Secondly, when the variable is loaded on more than one factor, it is only calculated in the factor score of the factor where it has the highest loading. Distefano et al. (2009) also list a number of other methods, including weighted sum scores, regression-based scores, and Bartlett's method.

These other methods also consider the factor loading assigned to each variable while calculating the dimension score. After observing the graphs produced by Biber's (1988) calculation method and regression-based dimension scores, it was decided to continue using Biber's method for several reasons. The first reason was that the method chosen did not affect the overall distribution of text groupings in a box plot or bar plot. The only difference between these two methods was that Biber's dimension scores were much more inflated as compared to, for example, regression-based dimension scores. The second reason was the ability to calculate dimension scores for the texts that were not included in the original EFA. Hence, the dimension scores were calculated using Biber's (1988, p. 93) method.

Table 5.2 Factor solution with feature loadings and examples

Category	Feature	Example
	Dimension 1 Oral…	
Adjectives	attitudinal adjectives in other contexts (0.58)	afraid, amazed, aware
	evaluative adjectives (0.34)	good, bad
	predicative adjectives (0.30)	This book is *interesting*.
	epistemic adjectives in other contexts	amazing, appropriate, conceivable
Adverbs	emphatics (0.41)	a lot, for sure, really
	factive adverbs (0.31)	undoubtedly, obviously
Clauses (finite)	*that* clauses controlled by factive verbs (0.33)	identify, prove + *that* clauses
	WH clauses controlled by factive verbs (0.31)	identify, prove + *WH* clauses
	WH clauses (0.30)	
Clauses (non-finite)	*to* clauses controlled by verbs of desire (0.38)	want, desire + *to* clauses
Conjunctions	subordinating Conjunctions – conditional (0.46)	if, unless
Reduced forms	*that* deletion (0.52)	He thinks ϕ the glass is empty.
	contractions (0.49)	isn't, amn't, aren't, he's
Pronouns	2nd person pronouns (0.71)	you, your
	1st person pronouns (0.63)	I, my, me
	nominal pronouns (0.73)	everything, someone

Category	Feature	Example
	demonstrative pronouns (0.36)*	this, that, these, those
	pronoun *it* (0.33)	it
Verbs (auxiliary)	pro-verb *do* (0.52)	He *does* not …
	modals of possibility (0.44)	can, could, may, might
	verb have (0.42)	has, have
	modals of prediction (0.39)	will, would, shall, be going to
	modals of necessity (0.37)	must, should, have to
	verb *be* (0.34)	is, was, were, are
Verbs (main)	mental verbs (0.70)	feel, like, hear, remember, believe
	activity verbs (0.35)	send, go, give
	attitudinal verbs in other contexts (0.32)	agree, anticipate without *that* clause
Verbs (tense/ aspect)	present tense (0.94)	
Other	discourse particles (0.43)	well, now, anyway
	WH questions (0.30)	*Why* would you…
	…versus Literate	
Adjectives	attributive adjectives (-0.39)*	an *interesting* book
Nouns	nominalisations (-0.37)*	ending in -tion, -ment, -ness, -ity
	proper nouns (-0.49)	
Passive voice	*by* passives (-0.30)	The window *is broken by the worker*.
	agentless passives (-0.34)	The snake *is killed*.
Verbs (tense/ aspect)	perfective aspect verbs (-0.35)	has/have/had + V-en
	past tense verbs (-0.37)*	V-ed
Other	post nominal passive modifiers (-0.34)	the book *written by John* was…
	word length (-0.38)*	generally lengthy words
	prepositions (-0.73)	of, off, in, about
	Dimension 2 Conversational Style	
Adverbs	general adverbs (0.48)	all words with adverb tag
	adverbs of likelihood (0.38)	apparently, evidently, kind of
	hedges (0.32)	at about, something like, almost
Conjunctions	coordinating conjunctions – clause level (0.54)	and, or, but It is funny, *and* it is ironic.
	adverbial conjuncts (0.53)	there, so, anyway, though, however
Pronouns	demonstrative pronouns (0.36)	this, that, these, those
	pronoun *it* (0.31)*	it
Other	stranded prepositions (0.35)	I find it so difficult to get *at*

Category	Feature	Example
Nouns	common nouns (-0.58)	
	pre-modifying nouns (-0.61)	noun noun combinations
Other	word length (-0.43)	generally lengthy words
Dimension 3 Abstract Evaluative Information…		
Adjectives	attributive adjectives (0.46)	an *interesting* book
	topical adjectives (0.41)	commercial, environmental, political, social
Nouns	nominalisations (0.47)	ending in *-tion, -ment, -ness, -ity*
	cognitive nouns (0.44)	concept, fact, idea, knowledge
	abstract nouns (0.43)	education, effect, function, method
	process nouns (0.37)	process, meeting
	stance nouns in other contexts (0.33)	evidence, importance, problem (without *that* clause)
	stance nouns (0.31)*	evidence, importance, problem + *that* clause
Other	prepositional phrases with stance nouns	evidence, importance, problem + *prepositional phrase*
	word length (0.31)*	generally lengthy words
…versus Non-Abstract Information		
Adverbs	adverbs of place (-0.33)	above, beside, outdoors
Reduced forms	contractions (-0.33)*	isn't, amn't, aren't, he's
Nouns	concrete nouns (-0.31)	phone, drug, picture, truck
	proper nouns (-0.35)*	
Verbs	activity verbs (-0.33)*	buy, make, get, go, give
Dimension 4 Reporting Style		
Clauses (finite)	*that* clauses controlled by communication verbs (0.60)	say, tell, call + *that* clause
	that clauses controlled by verbs (0.37)	all verbs + *that* clauses
Verbs (main)	communication verbs (0.74)	say, tell, call, ask, write…
	communication verbs in other contexts (0.54)	say, tell, call, ask, write… without *that* clauses
Verbs (tense/ aspect)	past tense (0.32)*	V-ed
Adjectives	attributive adjectives (-0.30)*	an *interesting* book
Dimension 5 Narrative Focus		
Clauses (finite)	*that* clauses controlled by factive verbs (0.30)*	identify, prove + *that* clauses
Reduced forms	*that* deletion (0.31)	He thinks ϕ the glass is empty
Pronouns	3rd person pronouns (0.42)	he, she, they, him, her, hers, them, their

Category	Feature	Example
Nouns	human nouns (0.36)	family, guy, individual, kid, man, manager, member, parent, teacher, child
Verbs (main)	mental verbs (0.30)*	feel, like, hear, remember, believe
Verbs (tense/ aspect)	past tense (0.60)	V-ed
Verbs (auxiliary)	verb *be* (-0.30)*	is, are, was, were
	modals of prediction (-0.36)*	will, would, shall, be going to

Note: Feature loadings => 0.30 included; features marked with * were not included in the calculation of dimension scores.

5.1.5 From factors to dimensions: interpreting the factors

Table 5.2 presents the factor solution with features in each factor and their loadings. Additionally, some examples are also provided in front of each feature.[28] The factors that result after an EFA are interpreted functionally in MD analysis (Biber and Conrad, 2009), hence the interpreted factors are referred to as linguistic dimensions. The following subsections analyse the co-occurring linguistic features in the light of example texts, dimension scores of the register categories included, and previous MD studies to find appropriate functional labels for each dimension.

5.1.5.1 Dimension 1 Oral versus Literate Discourse.

Dimension 1 consists of 40 lexico-grammatical and semantic features, out of which 30 occur on the positive side and 10 have negative loadings. The distribution of features on the positive and negative sides of this dimension is similar to the first dimension of variation observed by several previous MD studies, e.g. Biber (1988), Grieve et al. (2010), and other studies listed in Biber (2014, pp. 17-20). As Biber (2014) has observed, this dimension is nearly

---

[28] See Appendix II for lexico-grammatical features adopted from Biber (2006). Other features have been adopted from Biber (1988).

universal across many languages and discourse domains. This opposition is realised through the use of two distinctive ways of discourse construction, i.e. clausal versus phrasal features (p. 16). The features on the positive side include different types of lexical and auxiliary verbs, adverbs and stance adverbs, discourse particles, hedges, stance adjectives, pronouns, and various types of complement clauses. On the other hand, the features on the negative side include prepositional phrases, attributive adjectives, passive voice, nouns, nominalisations, and other noun modifiers.
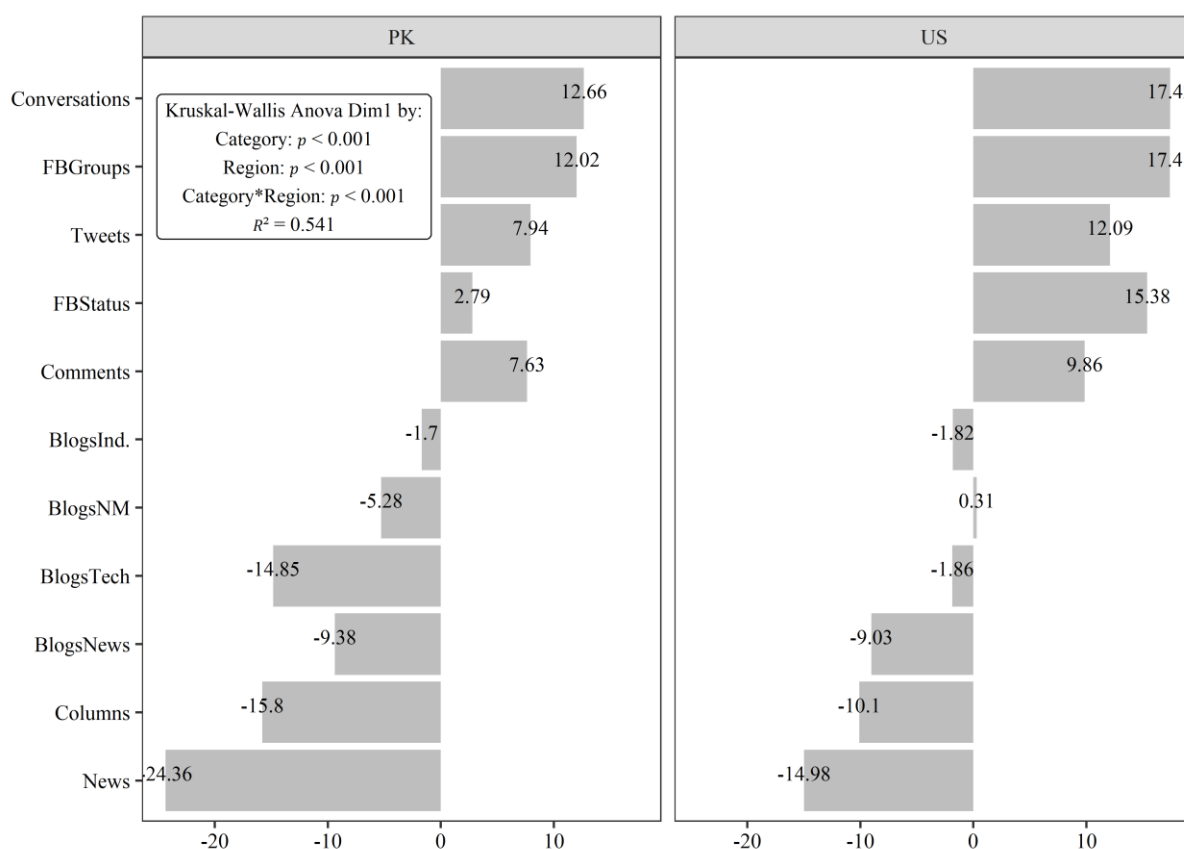
Figure 5.3 Online and offline registers on dimension 1 Oral versus Literate Discourse

Figure 5.3 shows how the above discussed linguistic features divide the registers under study into involved/ interactive/ oral versus informational/ formal/ literate discourse. Like in numerous previous MD studies, spoken conversations are highly interactive, personal, and involved along with other online interactive registers like Facebook groups, status updates, and comments. Individual and new media blogs are in between oral and literate discourse, while

technology and news blogs are on the negative side, along with columns and news reports

representing a highly literate/ formal discourse. The regional trends are also clear in the figure,

where Pakistani registers are generally less oral as compared to U.S. registers. Both categorical

variables (Category, Region, and their interaction) are significant predictors of variation in

dimension scores, as shown by the KW ANOVA results in the information box. Additionally, the

variance explained by the categorical variables and their interaction is above 50%, which shows

that this is a strong dimension of variation.

Table 5.3 Sample texts with high dimension scores on dimension 1 Oral versus Literate

Discourse

Text Samples
[…] awww... **you're right** about that **how much everybody miss you** here in Pakistan, **whenever we gathered** at anybodies place **you** guys that **we** all the time **talk** about!! **we** all **really really** missing **you** guy!! **Find** a **nice** job and **visit** Pakistan as soon as possible! Inshallah Allah **will give us** chance **we sure will visit** Canada!! :)<#> Hmm.. dear every coin **has** only two coins and **yours' may** not be exception, but depiction of **your** thoughts in these words **is very** heart touching. Today first time **i am looking** at **your** blog(s), but **honestly** speaking, **you have good** blogs.. **Keep it** up and good luck in life :) […] *File#CI-012, Register: Comments, Variety: Pakistani English, Dimension Score: 35.28*
I find myself wondering what are my dreams and goals in life. Many times **I feel discouraged** and **want to** give up. Getting through each day **is hard**. **Feeling** like **I'm** not **good** enough to achieve the things **I want to** achieve. **Wishing I could** just **be better**. A better mom and wife. Better friend and a better Christian.**It's honestly hard** to put this out there but **it's what is me** and **what I've** dealt with **my** whole life. **I've** never been **very confident** or easily been **able** to say to **myself You are GOOD** ENOUGH! […] *File#632032, Register: FB status, Variety: U.S. English, Dimension Score: 60.4*
Ooredoo, the **Qatar** based telecom giant and cent percent stake holder **of** wi-tribe **Pakistan**, **has announced** to **have completed** the sale **of** wi-tribe **to NB Offshore Investment** Ltd **for** around USD 9 million or Rs. 940 million. **Without** revealing much details, **Ooredoo said** that transaction **was completed in** cash payment that is **in** line **with** the value **of** the asset. Our sources suggest that deal **between Ooredo** and **HB Group** — the backers of NB Offshore Investment, **was locked** almost six months back and today's announcement **is made** only **after** complete takeover by **HB Group with** new management **in** place. […] *File#514060, Register: Blogs Tech, Variety: Pakistani English, Dimension Score: -40.29*

Text Samples
The man **<u>identified as</u>** the producer **<u>of</u>** an Internet film that **<u>sparked</u>** violent protests **<u>across</u>** the Muslim world **<u>has been arrested</u> <u>in</u>** Los Angeles for violating probation **<u>for</u>** a 2010 bank fraud conviction. Update at 7:17 p.m. ET: The **<u>Los Angeles Times</u>** summarizes the terms of **<u>Nakoula</u>**'s probation: He **<u>was ordered</u>** not to own or use devices **<u>with</u>** access **<u>to</u>** the Web **<u>without</u>** approval **<u>from</u>** his probation officer -â?? and any approved computers **<u>were</u> <u>to be used</u> <u>for</u>** work only. "Defendant shall not access a computer **<u>for</u>** any other purpose," according **<u>to</u>** the terms **<u>of</u>** his probation. […] *File#613202, Register: Blogs News, Variety: U.S. English, Dimension Score: -41.9*

Table 5.3 provides sample texts from both regional varieties with high dimension scores on both sides of the dimension.[29] Some of the features are highlighted to show the interplay of linguistic features and their functional interpretation. The text from Pakistani English on the positive side represents two comments from an individual blog. The comments are highly interactive and focused on the addresser and addressee. The text from U.S. English is taken from Facebook status updates. It is also highly personalised, where the writer is expressing their feelings and attitudes. The texts on the negative side are taken from technology and news blogs. The highlighted features, which include passive voice, prepositional phrases, past tense, and proper nouns, show that these texts are focused on conveying information and represent literate/ formal discourse. Based on the discussion in the above three paragraphs, this dimension can be labelled as 'Oral versus Literate Discourse'.

5.1.5.2 Dimension 2 Conversational Style.

Dimension 2 consists of 11 linguistic features, out of which eight occur on the positive side and the rest are on the negative side. Two features with the highest loadings on the positive

---

[29] The text samples have been taken from edited texts. More details regarding this can be found in chapter 3.

side are related to the grammatical category of connectors, i.e. coordinating conjunctions (*and, or, but*) and linking adverbials (*so, however, in addition to* etc.). Coordinating conjunctions connect clauses in a loose way, as Biber (1988, p. 106) observes in the case of non-phrasal *and*. Linking adverbials, on the other hand, primarily "state the speaker/writer's perception of the relationship between two units of discourse" (Biber et al. 1999, p. 875). They have several semantic categories, of which result/ inference type – especially *so* – occurs abundantly in conversations. Adverbs of probability and hedges show the level of uncertainty, and limited lexical choices available to the language users specially in conversational settings (Chafe and Danielewicz, 1986 as cited in Biber, 1988, p. 240). The use of general adverbs indicates the presence of clausal/ verbal as opposed to phrasal/ nominal features.
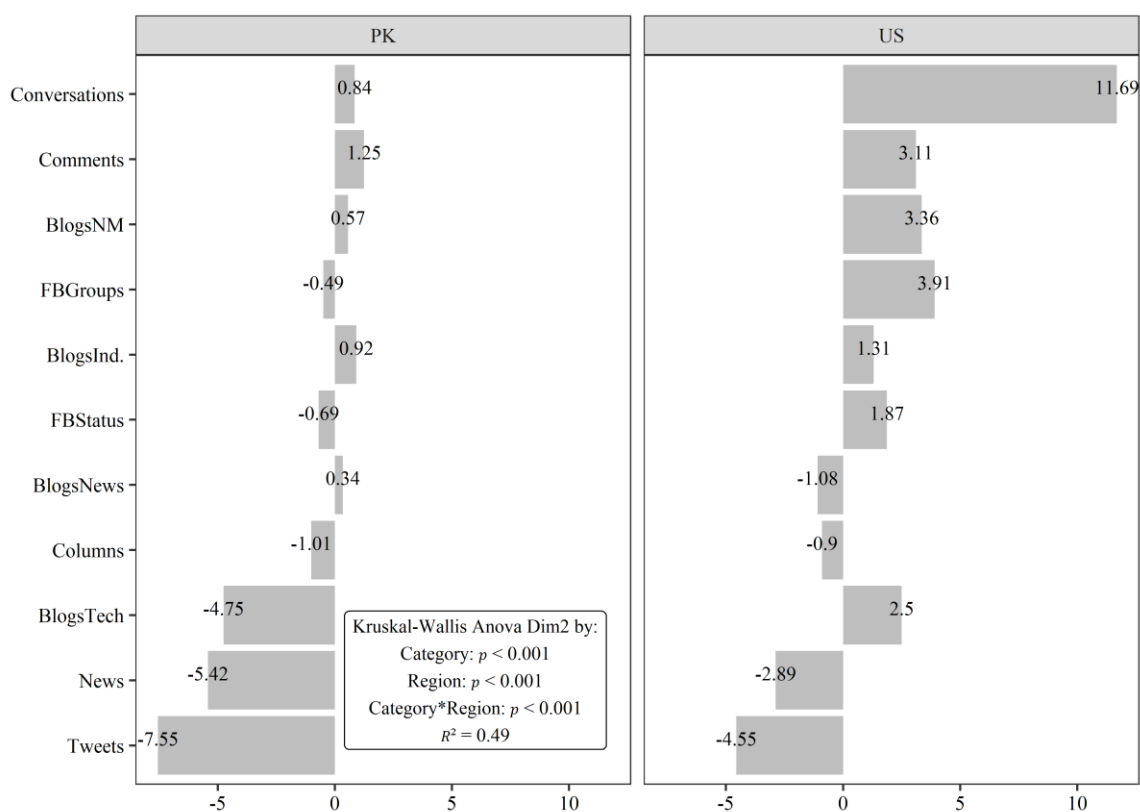
Figure 5.4 Online and offline registers on dimension 2 Conversational Style

Demonstrative pronouns and pronoun *it* indicate the presence of unspecified reference (Biber, 1988, p. 106), and hence refer to context specific references. Final (stranded) prepositions occur mostly in conversations and fiction (Biber et al., 1999, p. 106). Biber (1988, p. 113) observes that stranded prepositions along with demonstrative pronouns indicate the presence of informal and unplanned type of discourse. The features on the negative side indicate the presence of a nominal type of discourse, and are not discussed in detail here.

Figure 5.4 shows the distribution of registers on dimension 2. All of the textual categories are significant predictors of variation in the dimension scores, and consequently the variance explained is also high. The regional trends show that U.S. registers are skewed towards the positive side of the dimension, while Pakistani registers generally do not have high dimension scores. Most importantly, the texts with the highest dimension scores belong to face-to-face conversations in the U.S. data. Correspondingly, the example text in table 5.4 also reveals a highly fragmented discourse, with the use of context focused references, limited lexical choices, and hedges etc. The second sample text from new media blogs, and a high mean dimension score of new media and technology blogs show that some of these colloquial features are also adopted by U.S. online text producers. The sample texts with high dimension scores on the negative side include long lists of @ mentions in Twitter discourse, which are tagged as nouns or noun noun expressions by the tagger. In conclusion, the texts with the highest dimension scores are either informal and face-to-face conversations, or they are those texts that incorporate some of these conversational features to create an informal style. Hence, the label 'Conversational Style' appears to be appropriate for this dimension.

Table 5.4 Sample texts with high dimension scores on dimension 2 Conversational Style

Text Samples

[...] The way we do ***this***,
Yeah.
... is ***first of all***,
we make a,
a guesstimate,
.. of how many calories,
.. ***typically***,
 that you would need.
D- we- we didn't do ***this*** last time did we.
Hm-m.
Okay.
 Alright.
Yeah,
I think .. we were gonna wait,
until we,
.. cause you were gonna change some things **maybe**,
after you saw Doris.
So,
 first of all we make a good guesstimate.
Of how many calories you need.
And that's ***always*** a tough one.
Um,
Mhm.

.. most women,
 unless you work out a lot.
.. You **probably** aren't burning ***much more*** than **like** eighteen hundred,
 to **maybe** .. two thousand calories.
Mhm.
 So ,
what we can do ,
um,
.. I would say,
I'm gonna go ahead and make **it** out about eighteen hundred.
.. Just because,
uh,
you're at a good weight,
you're about five five,
.. um,
... ... think that would be a prudent place to,
.. to start out from.
.. I know I won't uh,
.. cause you to gain weight,
and,
.. you **probably** won't lose any either. […]

File#831041, Register: Conversations, Variety: U.S. English, Dimension Score: 41.14
There's an overwhelming amount of evidence that ***strongly*** suggests Keanu Reeves is immortal.
First, there's the fact that the man ***clearly*** doesn't age. ***Like***, at all.
Now you might notice ***in addition*** to photographs of the man, there are some classic paintings in that little timeline, and that's because, well, there are some too-similar-to-be-a-coincidence classical portraits that are ***undeniably*** Keanu.
Now all of us might not be as lucky as Mr. Reeves to have found the fountain of youth and life for all eternity, ***however***, thanks to the Google Arts and Culture App, people can find their own classical painting doppelg?ngers. […]
File#615061, Register: Blogs New media, Variety: U.S. English, Dimension Score: 18.6
[…] <#> When one door closes, another opens.. Don't stress it. #WednesdayWisdom
<#> @Xxxxxxxxxxxxxx Walaikum assalam Sir Ameen Congrats to joining bol news We are with you and support you Stay blessed
<#> @XxxxxXxxxxxXXX @Xxxxxx_xx @XxxxxxXxxxxx Excellent Well Said Sir Absolutely right
<#> Congrats to you @xxxxxxxxxx Sir May ALLAH pak grant you more success Stay blessed @XxxxxXxXx
<#> @xxxxxxxxxxxxxx We welcome to you @XxxxxxxxXXX   Sir in #Karachi

Text Samples
<#> @XxxXxxXxxxx @xxxxxxxxxxxxxxx @xxxxx786_xxx @XxxxXxxxxxXXX @xXxXxxxxXxxx @XexxxxxXxxxx72 @X_X_Xxxxxx @xxxxxxxxxxxx Congrats to all […]

File# 533076, Register: Tweets, Variety: Pakistani English, Dimension Score: -20.99

[…] <#> @Xxxxxxx0828 @xxxxxxxxxxxxxxx @XxxXxxXxxx @XXX913 @xxxxxxxXX @xxxxx_xxx @XxxXxxxxxx @Xxxxxxx628 @xxxxxxxxxxxxxxx Hey, Denise.
<#>   @xxxxxxxxxx77   @XxxxxxxX   @XzXxxxx67   @xxxxxxxxxx   @Xxxx117 @Xxxxxxxx4XXX @xxxxxxxxxx @XxxxxXxxXxx @xxxxxxxxxx Always, Grace
<#> @xxxxx_xxxxxxx @Xxxxxxxxxx @XxxxxXxxXxx @XxxXxxxxXxxx @xxxxxxxxxx @xxx7576 @XXxxxxxx @xxxxxxxxxxxx @xxxxxxxxxx Absolutely, my friend.
<#> @XxxxxxxXexxx That's an actual photo. Did change the cap-tion.
<#> @xxxxxxxxxxxx Fuckhead, to use the scientific term
<#> @xxxxxxxxxx "Who is that woman? I've never seen her before in my life, […]

File# 633004, Register: Tweets, Variety: U.S. English, Dimension Score: -28.5

5.1.5.3 Dimension 3 Abstract Evaluative versus Non-Abstract Information.

Dimension 3 consists of 15 linguistic features, with 10 features on the positive side and the remaining five on the negative side. The features occurring on the positive side include various types of nouns and noun modifiers, including nominalisations, abstract, process, cognitive, and stance nouns, attributive and topical adjectives, and prepositional phrases controlled by stance nouns. The various types of nouns are related to processes and concepts, which deal with abstract information. Other features, i.e. stance nouns and adjectives indicate the presence of stance/ personal opinion, evaluation, and description on the writer's part. The features on the negative side include nominal features like proper and concrete nouns, and verbal features like activity verbs, adverbs of place, and contractions. Activity verbs take a subject in the semantic role of agent (Biber et al., 1999, p. 361), and can involve physical actions like *give, send, get*. Adverbs of place indicate the place of action, as do proper nouns, which can also include names of places. Concrete nouns are also related to physical as opposed to abstract information. Lastly, contractions are a feature of an informal and interactive language use.

Figure 5.5 provides an overview of the registers under study on dimension 3. Generally written and non-interactive online registers, with the exception of comments, have a high score on the positive side, while interactive online registers and spoken conversations have lower scores. The regional trend is very clear, where the Pakistani data is skewed towards the positive side and the U.S. data has high dimension scores on the negative side. All categorical predictors are significant, and the R^2 value is reasonably high.
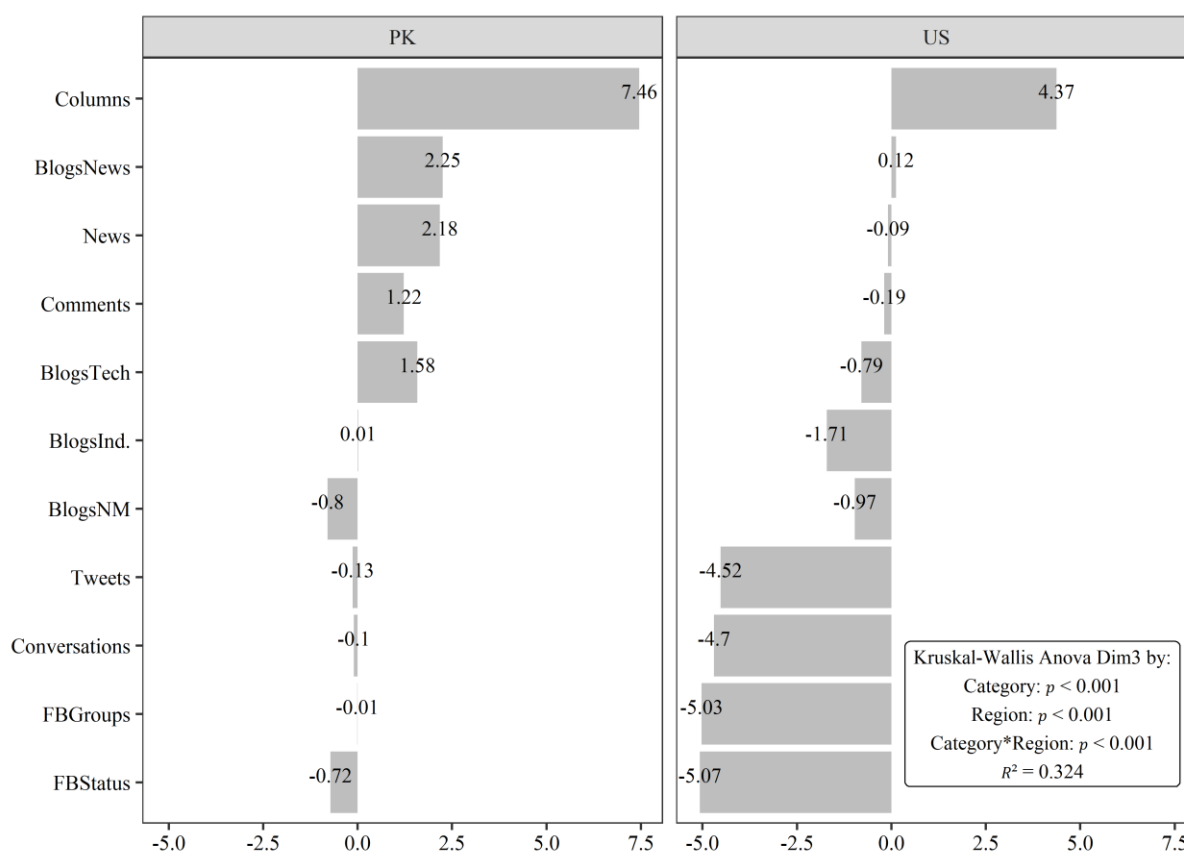
Figure 5.5 Online and offline registers on dimension 3 Abstract Evaluative versus Non-Abstract Information

Table 5.5 presents two sample texts with high dimension scores on the positive side from Pakistani English, while two texts with high dimension scores on the negative side are taken from U.S. English. The first text from Pakistani English is an excerpt from opinion columns. The text discusses a rather abstract topic, i.e. national security, along with the author's opinion on

this. The second text is a job ad taken from Pakistani Facebook groups, which lists a number of desired qualities. Both of these texts present information that is abstract as well as evaluative in some aspects. The texts from U.S. English, on the other hand, deal with rather non-abstract topics and information. The first text is taken from news blogs and presents information about a car. The second text, taken from Facebook groups, presents an exchange of information regarding a computer game (Pokémon go). Considering the features loaded on this dimension, the distribution of registers, and the sample texts, this dimension can be labelled as 'Abstract Evaluative versus Non-Abstract Information'.

Table 5.5 Sample texts with high dimension scores on dimension 3 Abstract Evaluative versus Non-Abstract Information

Text Samples
FINALLY, the prime minister has appointed a **full-time national security** adviser. The move was long overdue given Pakistan's **dire internal security** and **fast**-**changing geostrategic** environment at a time when there is an **urgent need for** a more **coherent strategy** to deal with **multiple** challenges. Although the government had established a **national security** secretariat, it has never been fully functional. It was indeed hard for Sartaj Aziz to juggle between the two most **critical** offices of **national security** and **foreign affairs** adviser. **National security** is too **serious** a matter to be dealt with on a **part-time** basis[…] *File#71292, Register: Columns, Variety: Pakistani English, Dimension Score: 22.52*

[…] We need a **Fresh** .NET developer. The **selected** candidate will participate in the **complete software development** life cycle from **requirements definition** and analysis to **deployment**.
Will learn C#, Jquery, Angular JS, XML, SQL Server, <link>
Will learn **Web** technologies, such as HTML, XHTML, XML, **Advanced** CSS, JavaScript, jQuery, AJAX
Will learn **knowledge of** Webservice
Some experience in developing, releasing, and supporting **Web applications**.
Ability to write and understand **complex** code.
Analysis, design and **development** of **software** products.
Constantly updating **technical knowledge** and **skills** by reading manuals and accessing new **applications**.
Exceptional familiarity with object oriented **development principles**
Social Media **Marketing Experience** would be plus. […]
File#FBG_PK_4_3, Register: FB Groups, Variety: Pakistani English, Dimension Score: 15.62
SEASIDE, Calif. --- How small can a **car** get? This one looks like a washing **machine** on **wheels**.
And it does**n't** have much more power.

Text Samples
Steven Mandell of **Glendora**, **Calif**., says his 1973 **Acoma Mini Comtesse** dishes out 1.8 horsepower from its 47 cubic-centimeter, one-cylinder **engine**. It**'s** enough to **get** it up to about 20 kilometers per hour. There**'s** actually a speedometer to prove it, he points out. Not enough? There**'s** also some **foot** pedals to allow the driver -- yes, there**'s** thankfully only room for one -- to help it along. […] *File#613126, Register: Blogs News, Variety: U.S. English, Dimension Score: -15.9*
[…] I**'ve** never **caught** an ivysaur and this gem is in my **box**.... It says I **got** it 6 days ago and I have more bulbs candy than I did. It treats it as though I just **caught** it. It**'s** date is days ago... So it**'s** clearly a glitch or a **gift** from **Niantics** because I**'m** cute. Anyone else have anything they did**n't** **catch**? <img> <#> You do**n't** think an egg hatched while you were**n't** looking? Because that**'s** how my **Lapras** was born. <#> I**'m** pretty sure they do**n't** hatch stage 2. […] *File#FBG_US_7_1, Register: FB Groups, Variety: U.S. English, Dimension Score: -12.22*

5.1.5.4 Dimension 4 Reporting Style.

Dimension 4 has only six lexico-grammatical features, which include communication verbs, e.g. *say*, *talk*, *announce*, *that* complement clauses controlled by communication and general verbs, and past tense, while the only negative feature is attributive adjectives. The concentration of such features indicates the presence of reporting clauses. As figure 5.6 reveals, in both varieties the register category with the highest scores is news reports. Spoken conversations, news blogs, and columns in U.S. English also have relatively high scores on this dimension. The information box shows that this is not a very strong dimension, with a small R^2 value and region as a nonsignificant predictor. However, the regional trends regarding various registers are different, as it can be seen in the reverse bars in the case of columns and news blogs. A look at the sample texts in table 5.6 also confirms that the texts report what is said by someone else. Hence, this dimension can be labelled as 'Reporting Style'.
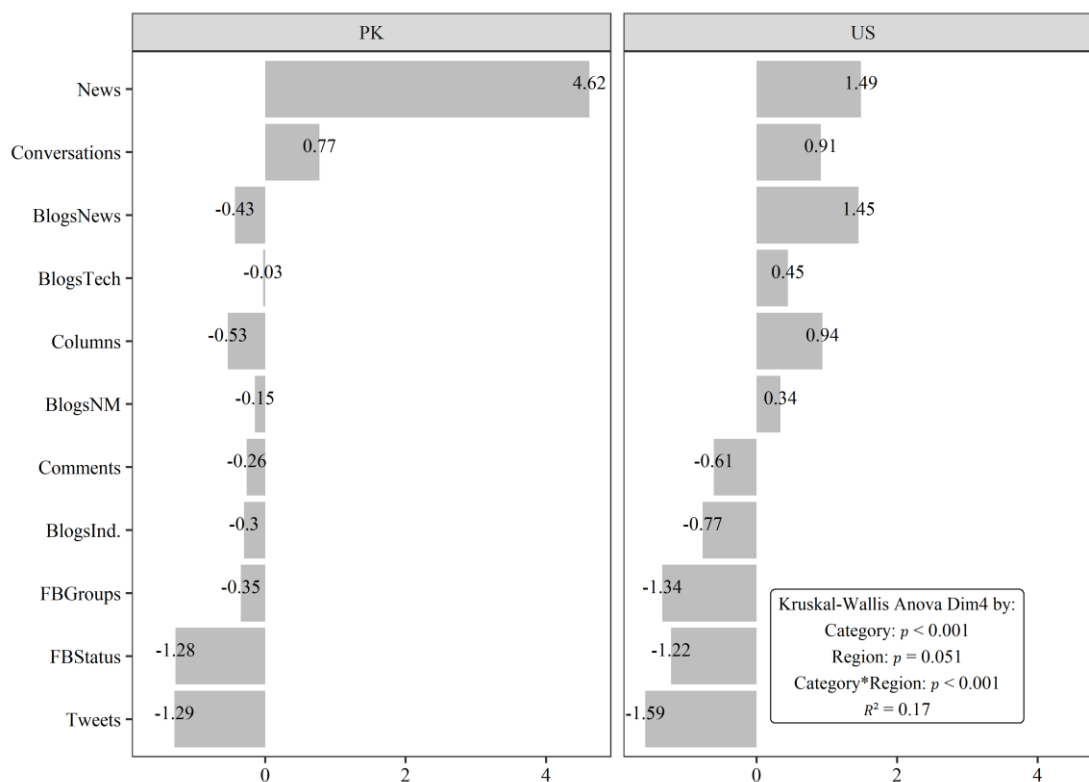
Figure 5.6 Online and offline registers on dimension 4 Reporting Style

Table 5.6 Sample texts with high dimension scores on dimension 4 Reporting Style

Text Samples
Punjab Chief Minister Muhammad Shahbaz Sharif has **said that** Punjab government is implementing a comprehensive programme for speedy economic development of the province and Development Strategy 2018 is being formulated with consultation of experts for increasing growth rate in Punjab. He **said that** there is a need for institutional reforms for achieving the targets of economic development and the government is taking effective measures for capacity-building of government institutions. He **said that** rapid progress is being made towards setting up of e-government with the help of information technology. He **said that** lakhs of new job opportunities would be created during next four and a half years. He was **talking** to a delegation of World Bank, here on Friday. *File#720048, Register: News, Variety: Pakistani English, Dimension Score: 18.25* Prime Minister Muhammad Nawaz Sharif has **said that** international business community's confidence to invest in Pakistan has been restored due to the incentives-based transparent policies of the government. The Prime Minister **said** this while meeting Mr Jean Yves Charlier, CEO VimpelCom who called on him at PM House today. The Prime Minister **said that** the government's clear policies have especially revolutionized the […]

Text Samples
The Prime Minister **<u>said that</u>** Pakistan offers an ideal environment for telecom sector as the country has one of the largest cellular phone subscribers. […]   *File#514034, Register: Blogs Tech, Variety: Pakistani English, Dimension Score: 15.17*

5.1.5.5 Dimension 5 Narrative Focus
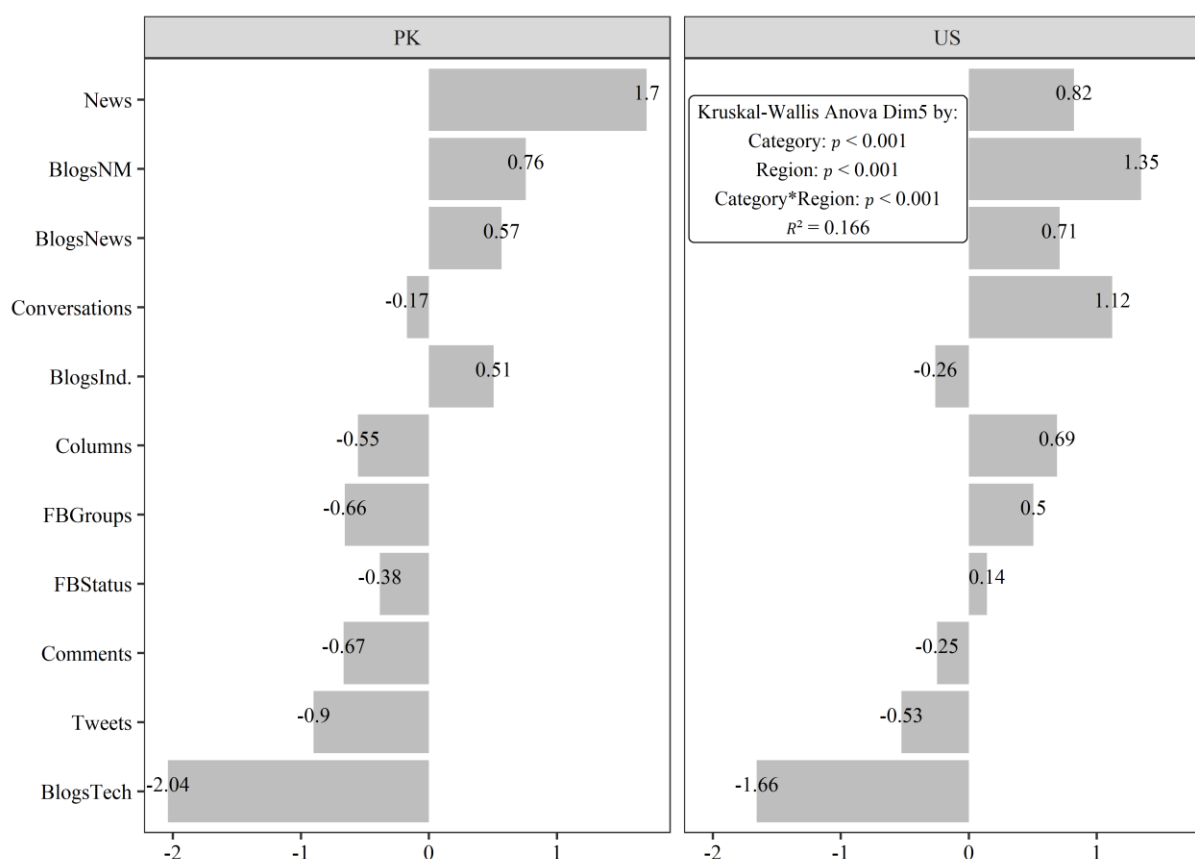
Figure 5.7 Online and offline registers on dimension 5 Narrative Focus

Dimension 5 consists of eight linguistic features, of which only two are on the negative side of the dimension. The top three features on the positive side are past tense verbs, 3rd person pronouns, and human nouns (e.g. *daughter, mother, guy, neighbour, observer*). As Biber (2014, p. 25) has observed, these features together with some other linguistic features create the second nearly universal dimension of variation, which represents the functional aspect of describing past events in a language. Other linguistic features like mental verbs, *that* deletion, and *that* clauses

controlled by factive/ mental verbs can be attributed to the presence of mental activities like

thinking, wondering, accepting etc. The features on the negative side, on the other hand, include

predictive modal verbs and the verb *be*, which are attributable to non-narrative discourse.

Figure 5.7 shows the distribution of registers on dimension 5. The highest scoring register

is news reports in Pakistani English and new media blogs in U.S. English. Both of these registers

include news stories, which generally narrate past events. The element of narration is also

present in Facebook groups and spoken conversations of U.S. origin, while the reverse trend

prevails in the Pakistani data. The variance explained by the categorical predictors on this

dimension is lower than 17%, which shows that this is not a strong dimension. However, all

textual categories are significant predictors of variation on this dimension.

Table 5.7 Sample texts with high dimension scores on dimension 5 Narrative Focus

Text Samples
Alexis Preston **was** just 23 years old when **she accepted** the fact that **her** eight-month-old **daughter**, Lexi, **would** have to grow up without **her**. Diagnosed with stage four breast cancer, Preston **knew she was** going to die. That's why when **she met** Heather Salazar in 2002, **she did** everything **she could** to convince the mom-of-three and her **husband** to adopt Lexi. With no one else to turn to, the young Ohio **mother pinned her** hopes on the Salazars to give her **baby** a loving home. It only **took** three days for the family to agree. "It **was meant** to be," Salazar **said**. "**She was meant** to be ours." […] *File#615011, Register: Blogs New media, Variety: U.S. English, Dimension Score: 15.14*
He lay on the rough floor, watching the red rivulets on **his** arm, in a spidery formation, pool down to **his** hand and then drop by drop plop to the floor. Red – his blood, **he** never **knew** this red. It **was** fascinating and yet, at the same time **his** heart seemed to be crushed within **his** chest as if a steel band **had been tightened** around it, the thumping betraying the terror **he felt**. **He wondered** when **they would** be back. […] **He braced** for more and predictably as **they started** with the beatings, **he did** what **had** always **been ingrained** in **his** mind from childhood - to start reciting the 'Ayatul Kursi' in the face of the unknown fear. **He had** never **known** such pain. […] *File#513048, Register: Blogs News, Variety: Pakistani English, Dimension Score: 9.71*

Table 5.7 contains two text samples from new media and news blogs which have high

dimension scores. The text from new media blogs of U.S. origin narrates the story of a cancer

patient. Although the text from news blogs of Pakistani origin does not have as high a dimension score, it includes some sections that narrate the events. The short story incorporates 3rd person pronouns and past tense in the process of storytelling. To conclude the discussion, dimension 5 can be labelled as 'Narrative Focus'.

5.1.6 Comparing online and offline registers: blogs versus columns and news

This section provides a detailed comparison of blogs, opinion columns and news reports. Firstly, comparisons are performed between regional varieties for similar register categories. The second type of comparisons are performed between the respective online and offline register categories of each regional variety. The purpose is to show regional variation, as well as register variation within each regional variety. Due to non-normality and unequal group sizes, nonparametric ANOVA, i.e. Kruskal Wallis ANOVA, and post-hoc Dunn's test with Bonferroni correction are used for group comparisons. The box plot in each subsection shows the detailed distribution of dimension scores on each register category, along with the regional comparison of register categories. The table provides register comparisons within each regional variety. The results of the post-hoc tests are provided in terms of z statistic, and p values within brackets. Significant group differences have been highlighted.

5.1.6.1 Blogs versus columns and news on dimension 1.

Figure 5.8 shows that significant regional differences only exist for technology blogs, columns, and news reports, where the Pakistani registers have a highly literate discourse as compared to their U.S. counterparts. Apart from that, Pakistani individual blogs, new media and technology blogs have more in-group variation, as shown by the spread of whiskers. The same is the case with news blogs in U.S. English.
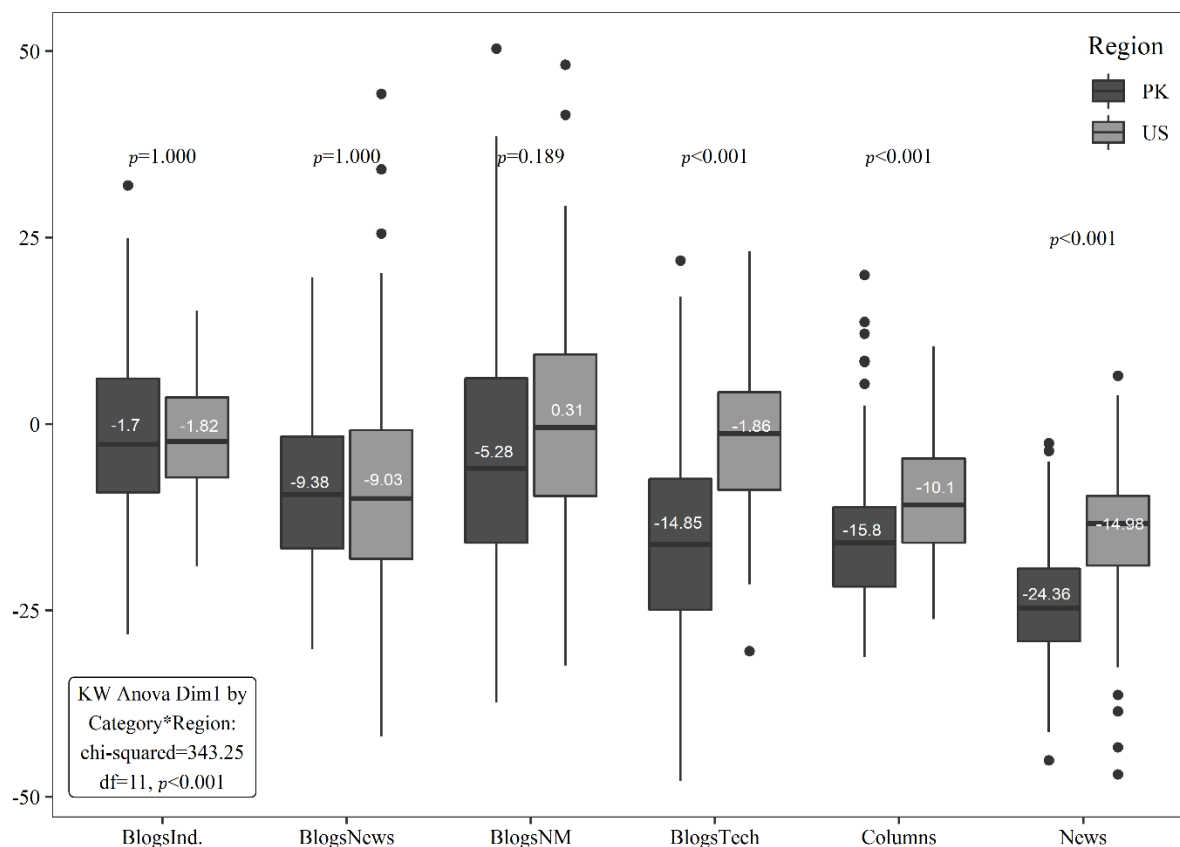
Figure 5.8 Group comparisons: Blogs versus columns and news on dimension 1 Oral versus Literate Discourse

Table 5.8 shows group comparisons among different blog types, columns, and news reports. All blog types in Pakistani English have significant differences in comparison to Pakistani columns and news reports. The same is the case with U.S. English with the exception of news blogs, which are not much different from U.S. columns and news reports. Looking at blog comparisons, technology blogs is the most unique blog type in Pakistani English, which is not the case in U.S. English.

Table 5.8 Blogs versus columns and news: Region-wise group comparisons on dimension 1 Oral versus Literate Discourse

	BlogsInd.	BlogsNews	BlogsNM	BlogsTech
PK: Kruskal-Wallis chi-squared = 158.1302, df = 5, *p* < 0.001				
BlogsNews	3.284 (0.008)			
BlogsNM	2.170 (0.225)	-1.239 (1.000)		
BlogsTech	6.362 (<0.001)	3.382 (0.005)	4.642 (<0.001)	
Columns	7.215 (<0.001)	4.183 (<0.001)	5.533 (<0.001)	0.571 (1.000)
News	10.612 (<0.001)	8.237 (<0.001)	9.397 (<0.001)	5.199 (<0.001)
US: Kruskal-Wallis chi-squared = 128.8953, df = 5, *p* < 0.001				
BlogsNews	4.889 (<0.001)			
BlogsNM	0.050 (1.000)	-5.250 (<0.001)		
BlogsTech	0.274 (1.000)	-5.006 (<0.001)	0.244 (1.000)	
Columns	5.715 (<0.001)	0.535 (1.000)	6.224 (<0.001)	5.958 (<0.001)
News	7.649 (<0.001)	3.104 (0.014)	8.218 (<0.001)	7.979 (<0.001)

5.1.6.2 Blogs versus columns and news on dimension 2.

As per figure 5.9, technology blogs and news reports are the only registers that are significantly different from their counterparts in U.S. English. Moreover, most of the Pakistani registers have a non-conversational, i.e. nominal, type of discourse like on dimension 1. Lastly, most of the U.S. registers have an opposite trend along with a relatively more variety, as shown by the in-group variation in terms of the spread of whiskers, especially in new media blogs, technology blogs, and news reports.
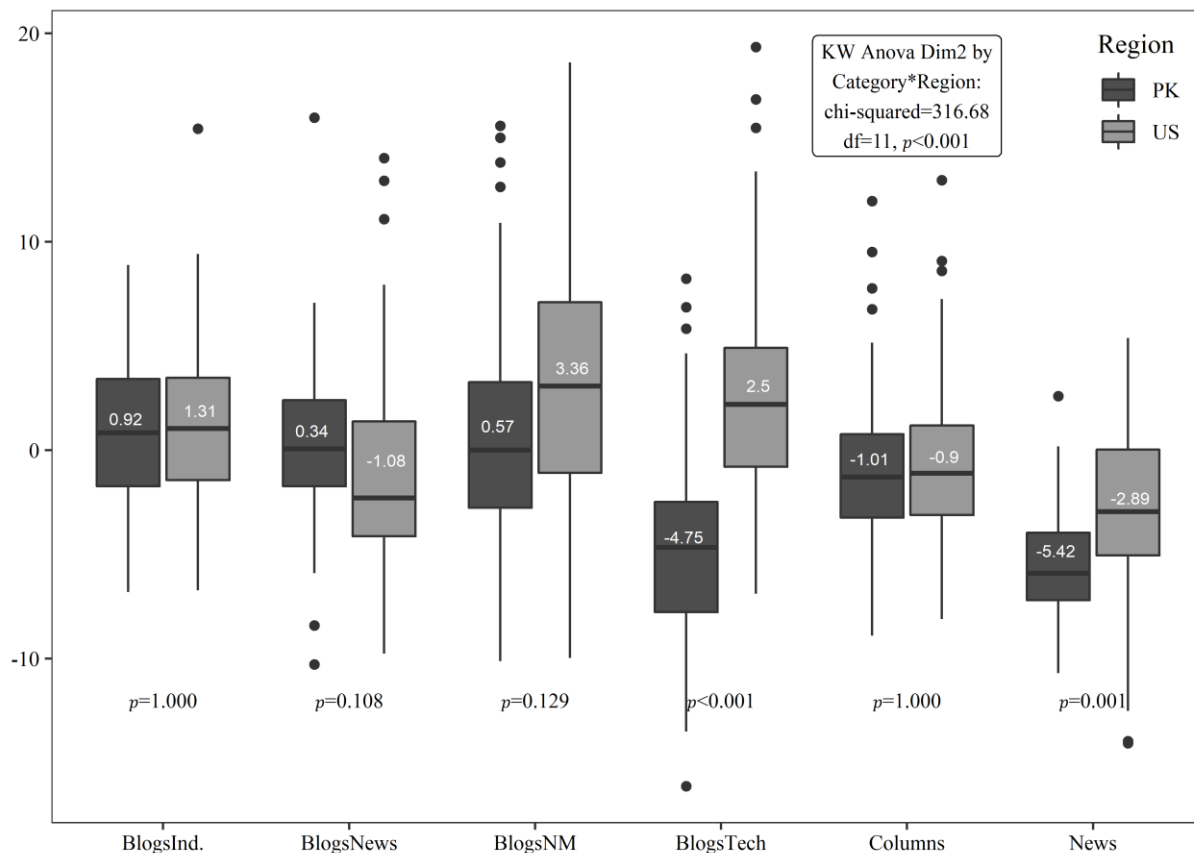
Figure 5.9 Group comparisons: Blogs versus columns and news on dimension 2
Conversational Style

The group comparisons, as shown by table 5.9, on dimension 2 again present Pakistani

technology blogs as the most unique register among different blog types, which is not the case

with U.S. technology blogs. Most blog types in Pakistani English are significantly different from

news reports. In U.S. English, on the other hand, apart from the similarities between news blogs,

columns, and news reports, all other blog types have significant differences in relation to offline

registers.

Table 5.9 Blogs versus columns and news: Region-wise group comparisons on dimension 2 Conversational Style

	BlogsInd.	BlogsNews	BlogsNM	BlogsTech
PK: Kruskal-Wallis chi-squared = 181.7397, df = 5, *p* < 0.001				
BlogsNews	0.838 (1.000)			
BlogsNM	1.244 (1.000)	0.451 (1.000)		
BlogsTech	8.143 (<0.001)	8.133 (<0.001)	7.674 (<0.001)	
Columns	3.108 (0.014)	2.501 (0.093)	2.009 (0.334)	-6.338 (<0.001)
News	9.285 (<0.001)	9.325 (<0.001)	8.903 (<0.001)	1.864 (0.467)
US: Kruskal-Wallis chi-squared = 111.9592, df = 5, *p* < 0.001				
BlogsNews	3.966 (0.001)			
BlogsNM	-1.362 (1.000)	-5.787 (<0.001)		
BlogsTech	-1.181 (1.000)	-5.590 (<0.001)	0.196 (1.000)	
Columns	3.575 (0.003)	-0.740 (1.000)	5.539 (<0.001)	5.325 (<0.001)
News	6.152 (<0.001)	2.459 (0.104)	8.101 (<0.001)	7.909 (<0.001)

5.1.6.3 Blogs versus columns and news on dimension 3.

Both regional varieties do not differ significantly from each other, as shown by figure 5.10, apart from newspaper opinion columns. However, a general trend of higher dimension scores in Pakistani English, and lower dimension scores in U.S. English is observable.
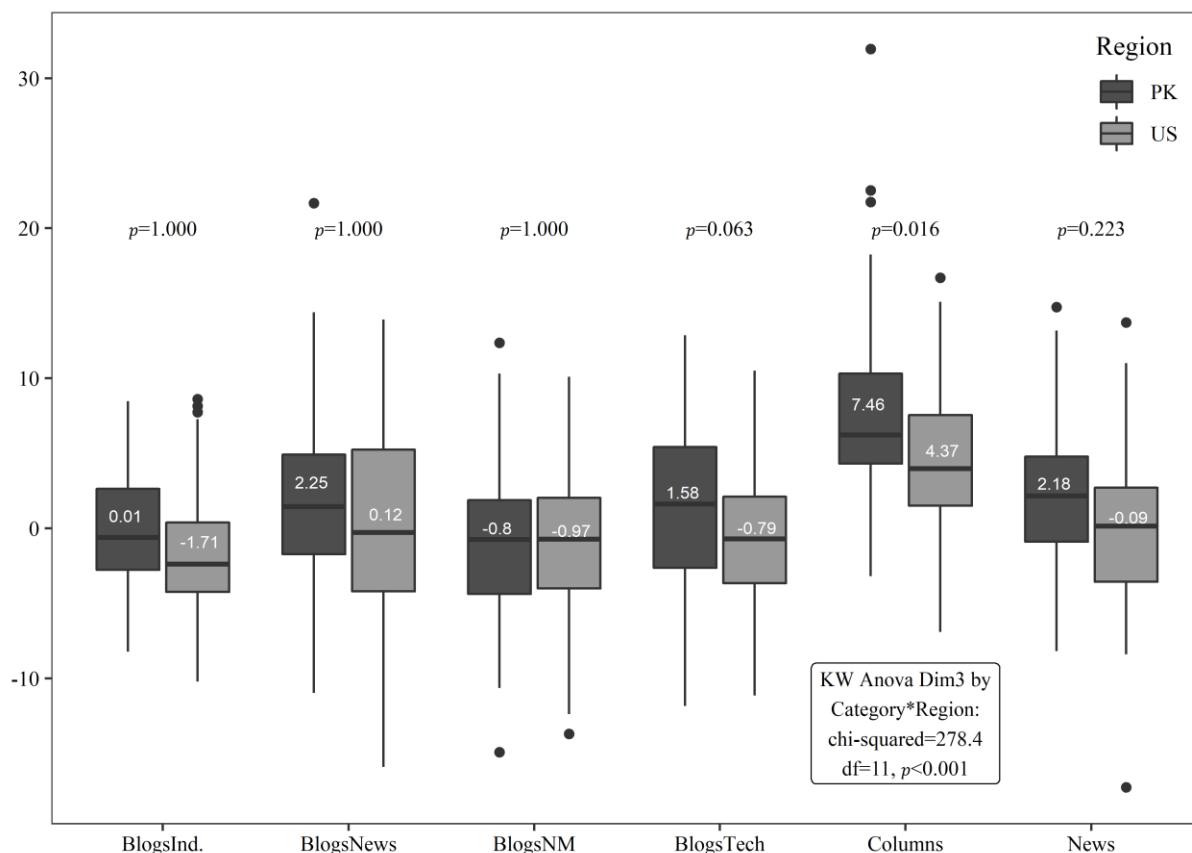
Figure 5.10 Group comparisons: Blogs versus columns and news on dimension 3 Abstract Evaluative versus Non-Abstract Information

The comparison between register categories within each regional variety, as depicted in table 5.10, reveals nonsignificant differences between most online registers. It shows that blogs generally deal with less abstract information. On the other hand, columns differ significantly from blogs in both regional varieties, which indicates that columns contain a high amount of abstract information, along with evaluation and personal opinion.

Table 5.10 Blogs versus columns and news: Region-wise group comparisons on dimension 3 Abstract Evaluative versus Non-Abstract Information

	BlogsInd.	BlogsNews	BlogsNM	BlogsTech
PK: Kruskal-Wallis chi-squared = 144.1687, df = 5, *p* < 0.001				
BlogsNews	-2.423 (0.115)			
BlogsNM	0.721 (1.000)	3.495 (0.004)		
BlogsTech	-1.998 (0.343)	0.511 (1.000)	-3.043 (0.018)	
Columns	-8.588 (<0.001)	-6.779 (<0.001)	-10.59 (<0.001)	-7.473 (<0.001)
News	-2.446 (0.109)	-0.169 (1.000)	-3.440 (0.004)	-0.648 (1.000)
US: Kruskal-Wallis chi-squared = 106.6321, df = 5, *p* < 0.001				
BlogsNews	-2.847 (0.033)			
BlogsNM	-1.407 (1.000)	1.555 (0.899)		
BlogsTech	-1.532 (0.941)	1.419 (1.000)	-0.136 (1.000)	
Columns	-8.384 (<0.001)	-5.903 (<0.001)	-7.555 (<0.001)	-7.407 (<0.001)
News	-2.278 (0.17)	0.557 (1.000)	-0.965 (1.000)	-0.832 (1.000)

5.1.6.4 Blogs versus columns and news on dimension 4.

Figure 5.11 provides information regarding variation between Pakistani and U.S. news blogs, columns, and news reports. News blogs, columns, and news reports have significant differences in both regional varieties. Additionally, a reverse trend between news blogs, new media blogs, technology blogs, and columns versus news reports is observable, where Pakistani blog types and columns have less reporting style as compared to their U.S. counterparts. However, the reverse is true for news reports in both regional varieties.
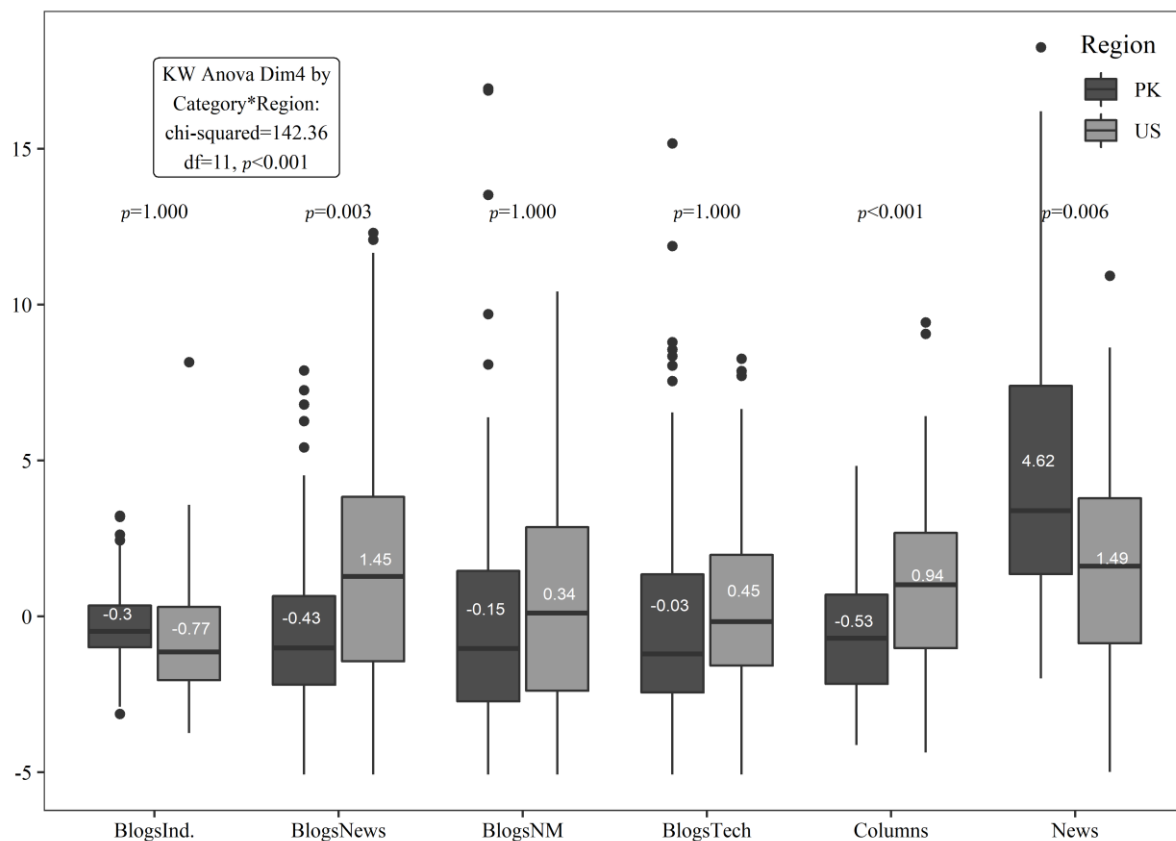
Figure 5.11 Group comparisons: Blogs versus columns and news on dimension 4 Reporting Style

As table 5.11 reveals, there are no significant differences between Pakistani blog types, and only news reports are significantly different from blog registers. Hence, Pakistani blogs apparently do not follow the reporting style of news reports. U.S. blogs, on the other hand, are slightly different, as it can be seen from the nonsignificant differences between news blogs and technology blogs versus news reports. This indicates the presence of news reports-like style in both blog types.

Table 5.11 Blogs versus columns and news: Region-wise group comparisons on dimension 4 Reporting Style

	BlogsInd.	BlogsNews	BlogsNM	BlogsTech
PK: Kruskal-Wallis chi-squared = 92.1822, df = 5, *p* < 0.001				
BlogsNews	1.096 (1.000)			
BlogsNM	1.478 (1.000)	0.424 (1.000)		
BlogsTech	1.069 (1.000)	-0.047 (1.000)	-0.478 (1.000)	
Columns	1.039 (1.000)	-0.152 (1.000)	-0.615 (1.000)	-0.104 (1.000)
News	-6.008 (<0.001)	-7.739 (<0.001)	-8.136 (<0.001)	-7.808 (<0.001)
US: Kruskal-Wallis chi-squared = 33.7587, df = 5, *p* < 0.001				
BlogsNews	-4.561 (<0.001)			
BlogsNM	-2.436 (0.111)	2.294 (0.163)		
BlogsTech	-2.730 (0.047)	1.974 (0.363)	-0.319 (1.000)	
Columns	-4.399 (<0.001)	0.532 (1.000)	-1.959 (0.376)	-1.611 (0.804)
News	-4.938 (<0.001)	-0.507 (1.000)	-2.746 (0.045)	-2.433 (0.112)

5.1.6.5 Blogs versus columns and news on dimension 5.

As depicted in figure 5.12, only Pakistani columns are significantly different from U.S. columns. Otherwise, there are no significant differences in similar registers of both regional varieties. However, U.S. registers, with the exception of individual blogs and news reports, are generally more narrative as compared to their Pakistani counterparts.
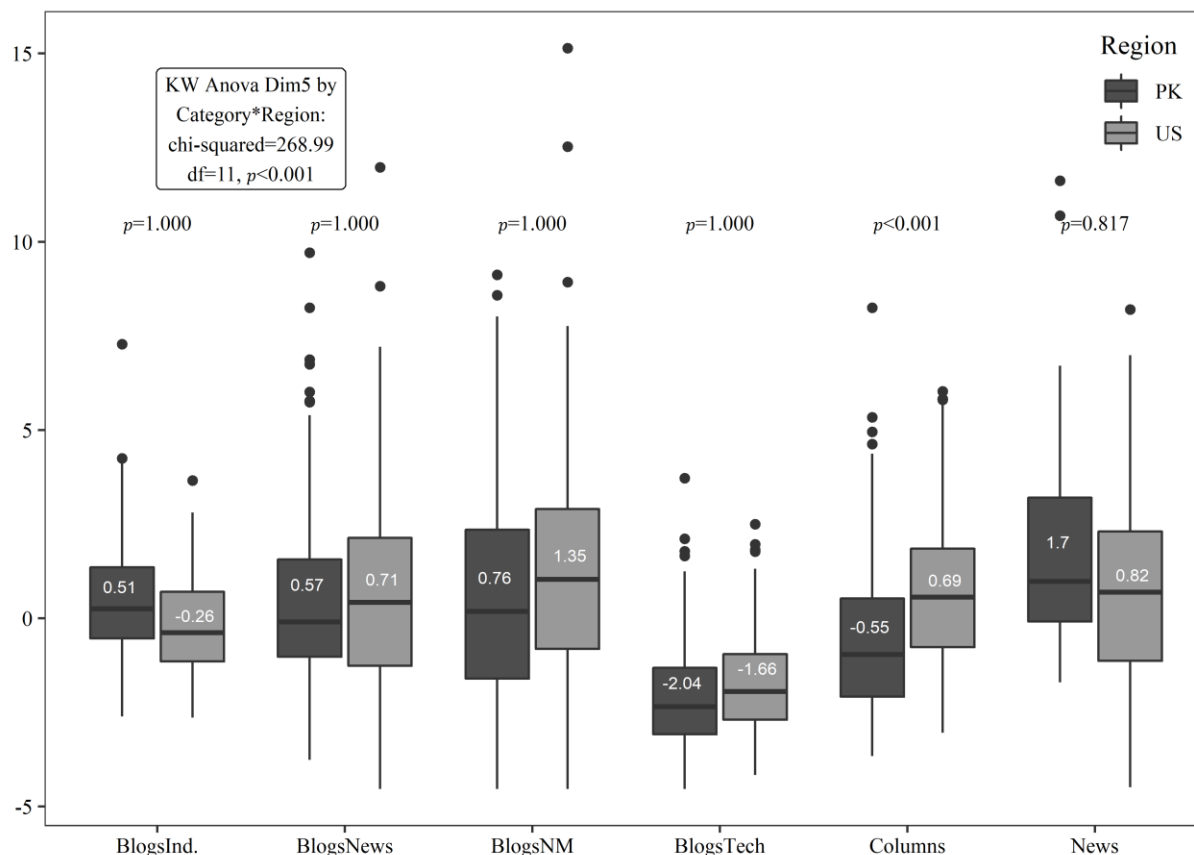
Figure 5.12 Group comparisons: Blogs versus columns and news on dimension 5 Narrative Focus

Table 5.12 shows register variation within each regional variety. Technology blogs in both regional varieties are significantly non-narrative in comparison to all other online and offline registers. Other blog types in Pakistani English are generally significantly different from news reports and columns. U.S. blog types, however, are not significantly different from news reports and columns generally.

Table 5.12 Blogs versus columns and news: Region-wise group comparisons on dimension 4 Reporting Style

	BlogsInd.	BlogsNews	BlogsNM	BlogsTech
PK: Kruskal-Wallis chi-squared = 152.5569, df = 5, *p* < 0.001				
BlogsNews	0.654 (1.000)			
BlogsNM	0.657 (1.000)	0.003 (1.000)		
BlogsTech	8.012 (<0.001)	8.194 (<0.001)	8.190 (<0.001)	
Columns	3.919 (0.001)	3.640 (0.002)	3.636 (0.002)	-5.243 (<0.001)
News	-2.174 (0.223)	-3.068 (0.016)	-3.071 (0.016)	-10.77 (<0.001)
US: Kruskal-Wallis chi-squared = 110.1566, df = 5, *p* < 0.001				
BlogsNews	-2.040 (0.31)			
BlogsNM	-3.435 (0.004)	-1.529 (0.946)		
BlogsTech	4.748 (<0.001)	7.388 (<0.001)	8.873 (<0.001)	
Columns	-2.795 (0.039)	-0.677 (1.000)	0.985 (1.000)	-8.682 (<0.001)
News	-2.573 (0.076)	-0.623 (1.000)	0.874 (1.000)	-7.836 (<0.001)

5.1.6.6 Blogs versus columns and news: section summary.

To conclude this subsection, the group comparisons have revealed that Pakistani and U.S. blogs are mostly not significantly different from each other. Only Pakistani technology blogs are significantly more literate and abstract evaluative on dimensions 1 and 3, respectively, while Pakistani news blogs are significantly more non-narrative on dimension 4. Additionally, a general trend of literate, non-conversational, abstract evaluative, non-narrative, and non-reporting style is observable in other Pakistani blogs.
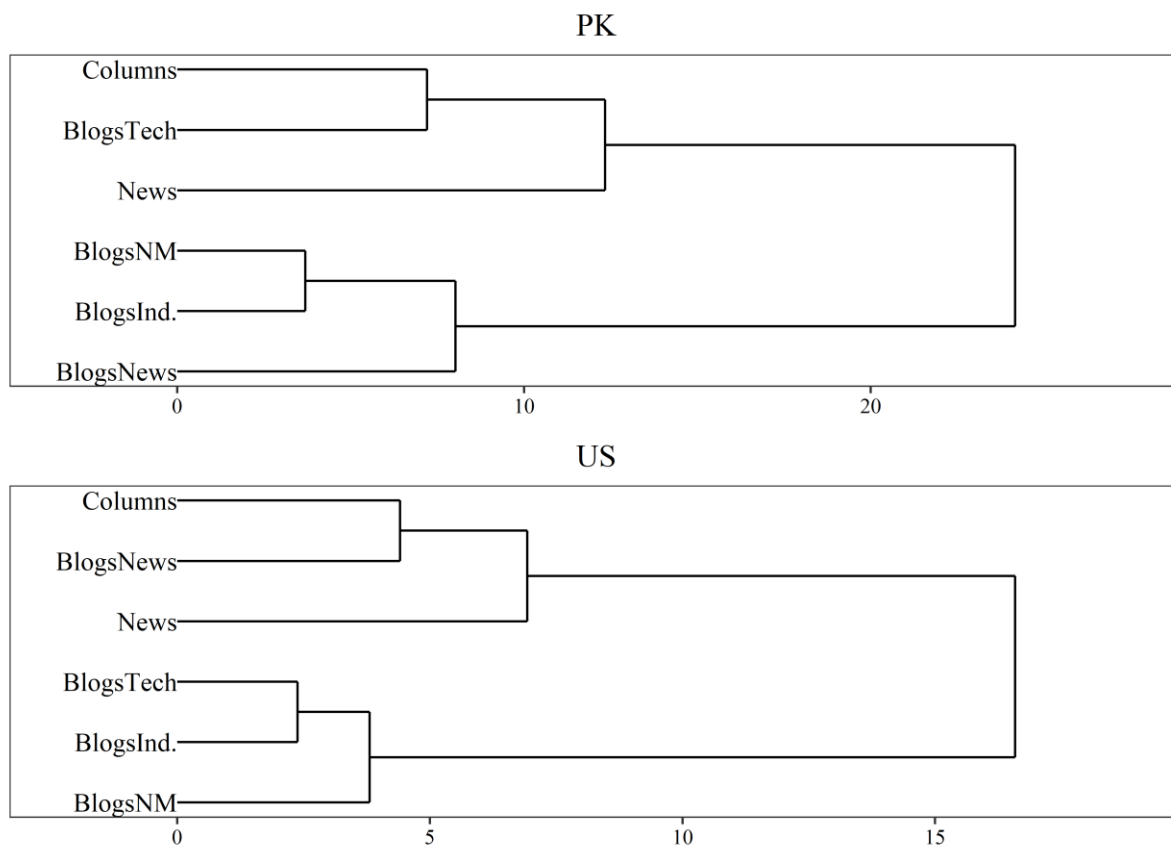
Figure 5.13 Summary of comparisons between blogs versus columns and news reports in both regional varieties using hierarchical cluster analyses of mean dimension scores[30]

The results can also be considered in a different perspective, when regional trends are observed in the comparisons of blogs versus columns and news reports. Figure 5.13 reveals slightly different clusters of similarity in both regional varieties. The first cluster in each variety includes columns and news reports. Columns are generally literate and abstract evaluative, while news reports are generally narrative and reporting oriented. The second cluster entirely consists of blogs, which are generally more oral, conversational, and non-abstract, as has been seen previously in this subsection. Technology blogs in Pakistani English, which are located in the

---

[30] Hierarchical cluster analysis has been performed using hclust() function in R.

first cluster, do not follow the blogging style observable in the other three blog types in Pakistani English, while the reverse is true for U.S. technology blogs. Similarly, news blogs in U.S. English, which are located in the first cluster, are more like columns and news reports instead of the other three types of blogs in U.S. English, and vice versa for Pakistani news blogs. Individual and new media blogs, however, show more or less similar trends in both regional varieties.

5.1.7 Comparing online and offline registers: online interactive registers versus spoken conversations

This section provides group comparisons between online interactive registers and spoken conversations in both regional varieties. The spoken conversations have been subdivided in interviews, face-to-face, and talk shows. The comparison follows the same pattern as in the previous section.

5.1.7.1 Interactive registers on dimension 1.

Figure 5.14 shows the detailed distribution of interactive registers on dimension 1. Online registers, i.e. Facebook groups and Facebook status updates are significantly different between both regional varieties. As observed in previous sections, Pakistani interactive registers are less oral as compared to U.S. registers, with the exception of interviews.
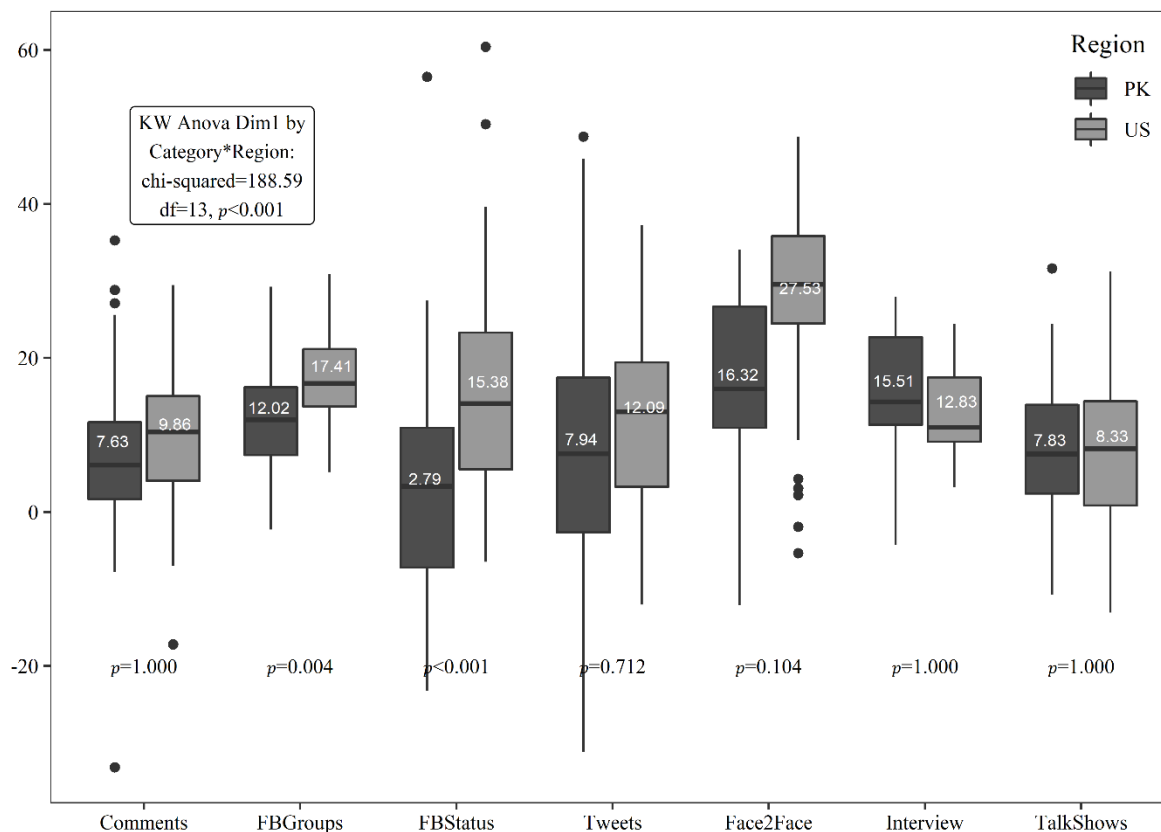
Figure 5.14 Group comparisons: online interactive registers versus spoken conversations on dimension 1 Oral versus Literate Discourse

Comparisons between online and offline registers in table 5.13 show that face-to-face conversations are generally significantly different from online registers in both regional varieties. Pakistani Facebook groups differ from any type of online registers. U.S. talk shows also have significant differences from Facebook groups and status updates. Generally, there is more similarity in Pakistani online and offline registers, as compared to their U.S. counterparts.

Table 5.13 Online interactive registers versus spoken conversations: Region-wise group comparisons on dimension 1 Oral versus Literate Discourse

	Comments	FBGroups	FBStatus	Tweets
PK: Kruskal-Wallis chi-squared = 62.048, df = 6, *p* < 0.001				
FBGroups	-3.371 (0.008)			
FBStatus	2.275 (0.241)	5.693 (<0.001)		
Tweets	-0.616 (1.000)	2.878 (0.042)	-2.984 (0.03)	
Face2Face	-4.663 (<0.001)	2.008 (0.469)	6.489 (<0.001)	4.288 (<0.001)
Interviews	-2.338 (0.204)	-0.898 (1.000)	-3.308 (0.01)	2.097 (0.378)
TalkShows	-0.218 (1.000)	2.336 (0.205)	-1.936 (0.555)	-0.237 (1.000)
US: Kruskal-Wallis chi-squared = 89.6621, df = 6, *p* < 0.001				
FBGroups	-5.294 (<0.001)			
FBStatus	-3.222 (0.013)	2.312 (0.218)		
Tweets	-1.724 (0.889)	3.707 (0.002)	1.505 (1.000)	
Face2Face	-7.746 (<0.001)	3.334 (0.009)	5.340 (<0.001)	6.463 (<0.001)
Interviews	-0.731 (1.000)	1.626 (1.000)	0.633 (1.000)	-0.005 (1.000)
TalkShows	0.821 (1.000)	5.079 (<0.001)	3.383 (0.008)	-2.199 (0.293)

5.1.7.2 Interactive registers on dimension 2.

Pakistani registers have lower scores on conversational style, as shown by figure 5.15. Unlike dimension 1, most registers are significantly different in both regional varieties, with the exception of interviews and tweets.
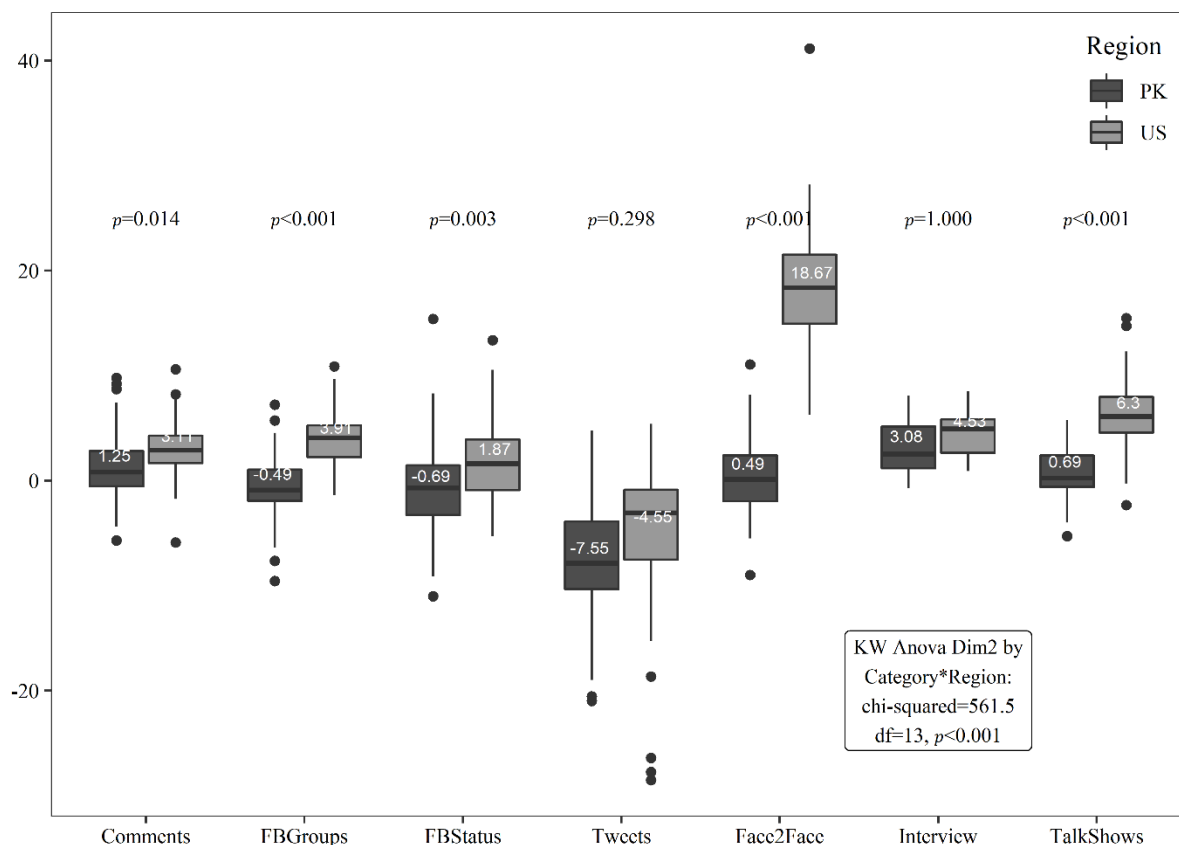
Figure 5.15 Group comparisons: Online interactive registers versus spoken conversations on dimension 2 Conversational Style

Table 5.14 reveals that tweets are most significantly non-conversational in both regional varieties. Additionally, U.S. face-to-face conversations significantly differ from all online interactive registers. U.S. talk shows also differ from online interactive registers, with the exception of Facebook groups. In summary, it can be said that Pakistani interactive registers are linguistically and functionally less diverse on dimension 2 as compared to their U.S. counterparts.

Table 5.14 Online interactive registers versus spoken conversations: Region-wise group comparisons on dimension 2 Conversational Style

	Comments	FBGroups	FBStatus	Tweets
PK: Kruskal-Wallis chi-squared = 194.1342, df = 6, *p* < 0.001				
FBGroups	3.384 (0.007)			
FBStatus	3.604 (0.003)	0.187 (1.000)		
Tweets	12.18 (<0.001)	8.707 (<0.001)	8.604 (<0.001)	
Face2Face	1.632 (1.000)	1.038 (1.000)	1.192 (1.000)	7.831 (<0.001)
Interviews	-1.324 (1.000)	-2.771 (0.059)	-2.856 (0.045)	6.399 (<0.001)
TalkShows	0.622 (1.000)	-1.942 (0.548)	-2.093 (0.381)	8.464 (<0.001)
US: Kruskal-Wallis chi-squared = 292.862, df = 6, *p* < 0.001				
FBGroups	-1.412 (1.000)			
FBStatus	2.283 (0.236)	3.667 (0.003)		
Tweets	9.035 (<0.001)	10.210 (<0.001)	6.981 (<0.001)	
Face2Face	-7.990 (<0.001)	6.705 (<0.001)	9.940 (<0.001)	15.250 (<0.001)
Interviews	-1.066 (1.000)	-0.433 (1.000)	-2.040 (0.434)	4.939 (<0.001)
TalkShows	-3.599 (0.003)	-2.394 (0.175)	-5.469 (<0.001)	10.811 (<0.001)

5.1.7.3 Interactive registers on dimension 3.

Dimension 3 most distinctively divides both regional varieties, as can be observed from the significant differences on most register categories in figure 5.16.
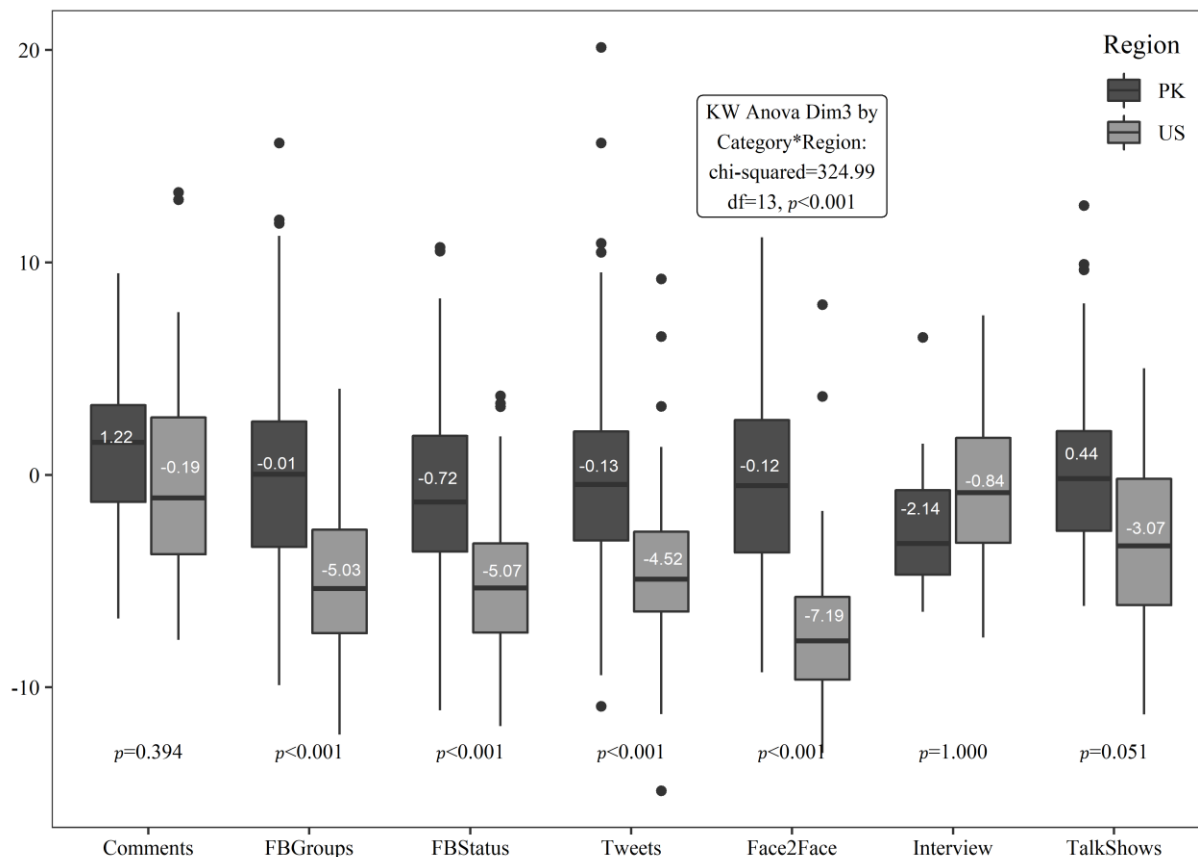
Figure 5.16 Group comparisons: Online interactive registers versus spoken conversations on dimension 3 Abstract Evaluative versus Non-Abstract Information

Table 5.15 highlights that most Pakistani interactive registers are similar in terms of abstract and evaluative informational discourse, except comments which are significantly higher than at least three other registers. U.S. comments are also abstract evaluative, and significantly different from most other registers. U.S. Facebook status updates, however, are significantly non-abstract information oriented. Lastly, U.S. face-to-face conversations are unlike any online register, as shown by significant differences.

Table 5.15 Online interactive registers versus spoken conversations: Region-wise group comparisons on dimension 3 Abstract Evaluative versus Non-Abstract Information

	Comments	FBGroups	FBStatus	Tweets
PK: Kruskal-Wallis chi-squared = 19.6125, df = 6, *p* < 0.001				
FBGroups	2.669 (0.08)			
FBStatus	3.708 (0.002)	1.016 (1.000)		
Tweets	3.069 (0.023)	0.309 (1.000)	-0.739 (1.000)	
Face2Face	2.320 (0.213)	-0.216 (1.000)	0.580 (1.000)	0.020 (1.000)
Interviews	2.846 (0.046)	1.706 (0.923)	1.280 (1.000)	-1.59 (1.000)
TalkShows	1.651 (1.000)	-0.370 (1.000)	-1.136 (1.000)	0.607 (1.000)
US: Kruskal-Wallis chi-squared = 122.936, df = 6, *p* < 0.001				
FBGroups	7.279 (<0.001)			
FBStatus	7.871 (<0.001)	0.161 (1.000)		
Tweets	6.675 (<0.001)	-0.916 (1.000)	-1.140 (1.000)	
Face2Face	9.464 (<0.001)	-3.419 (0.007)	-3.414 (0.007)	-4.269 (<0.001)
Interviews	0.485 (1.000)	-2.753 (0.062)	-2.853 (0.046)	2.373 (0.185)
TalkShows	3.257 (0.012)	-2.677 (0.078)	-2.906 (0.038)	2.007 (0.47)

5.1.7.4 Interactive register on dimension 4.

As per figure 5.17, only Facebook groups are significantly different between both regional varieties, and just barely so. The general trend, however, shows that Pakistani registers have higher scores on reporting style, with the exception of interviews and face-to-face conversations.
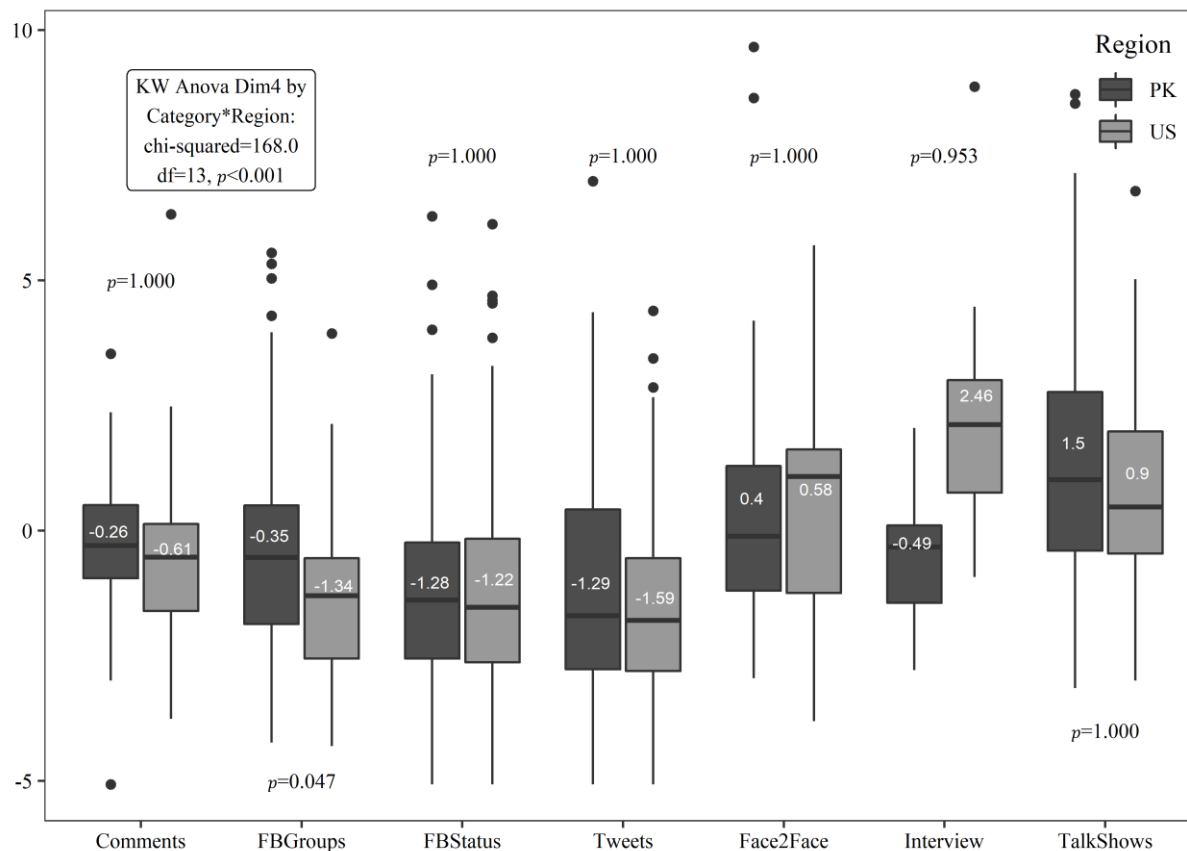
Figure 5.17 Group comparisons: Online interactive registers versus spoken conversations on dimension 4 Reporting Style

Table 5.16 shows that only Pakistani talk shows are significantly different from online registers. The same can be observed from figure 5.17, where Pakistani talk shows have the highest score on reporting style. The differences between U.S. online and offline registers, on the contrary, are highly significant, with the exception of comments versus face-to-face conversations. The elements of reporting style are generally absent from U.S. online interactive registers, while the reverse is true for offline registers.

Table 5.16 Online interactive registers versus spoken conversations: Region-wise group comparisons on dimension 4 Reporting Style

	Comments	FBGroups	FBStatus	Tweets
PK: Kruskal-Wallis chi-squared = 69.6044, df = 6, *p* < 0.001				
FBGroups	1.340 (1.000)			
FBStatus	4.675 (<0.001)	3.330 (0.009)		
Tweets	4.537 (<0.001)	3.159 (0.017)	-0.255 (1.000)	
Face2Face	-0.527 (1.000)	1.586 (1.000)	4.210 (<0.001)	4.076 (<0.001)
Interviews	0.475 (1.000)	-0.097 (1.000)	-1.506 (1.000)	1.408 (1.000)
TalkShows	-2.852 (0.046)	-3.872 (0.001)	-6.401 (<0.001)	6.300 (<0.001)
US: Kruskal-Wallis chi-squared = 91.8274, df = 6, *p* < 0.001				
FBGroups	2.901 (0.039)			
FBStatus	2.863 (0.044)	-0.201 (1.000)		
Tweets	3.966 (0.001)	0.904 (1.000)	1.169 (1.000)	
Face2Face	-2.455 (0.148)	4.751 (<0.001)	4.763 (<0.001)	5.631 (<0.001)
Interviews	-3.239 (0.013)	-4.511 (<0.001)	-4.472 (<0.001)	4.948 (<0.001)
TalkShows	-3.511 (0.005)	-5.788 (<0.001)	-5.838 (<0.001)	6.698 (<0.001)

5.1.7.5 Interactive registers on dimension 5.

The general trend in figure 5.18 indicates that U.S. online interactive and spoken registers are more narrative as compared to the Pakistani data. The only significant differences are between Facebook groups and face-to-face conversations. More in-group variation is also observable in most U.S. registers, as shown by the spread of whiskers.
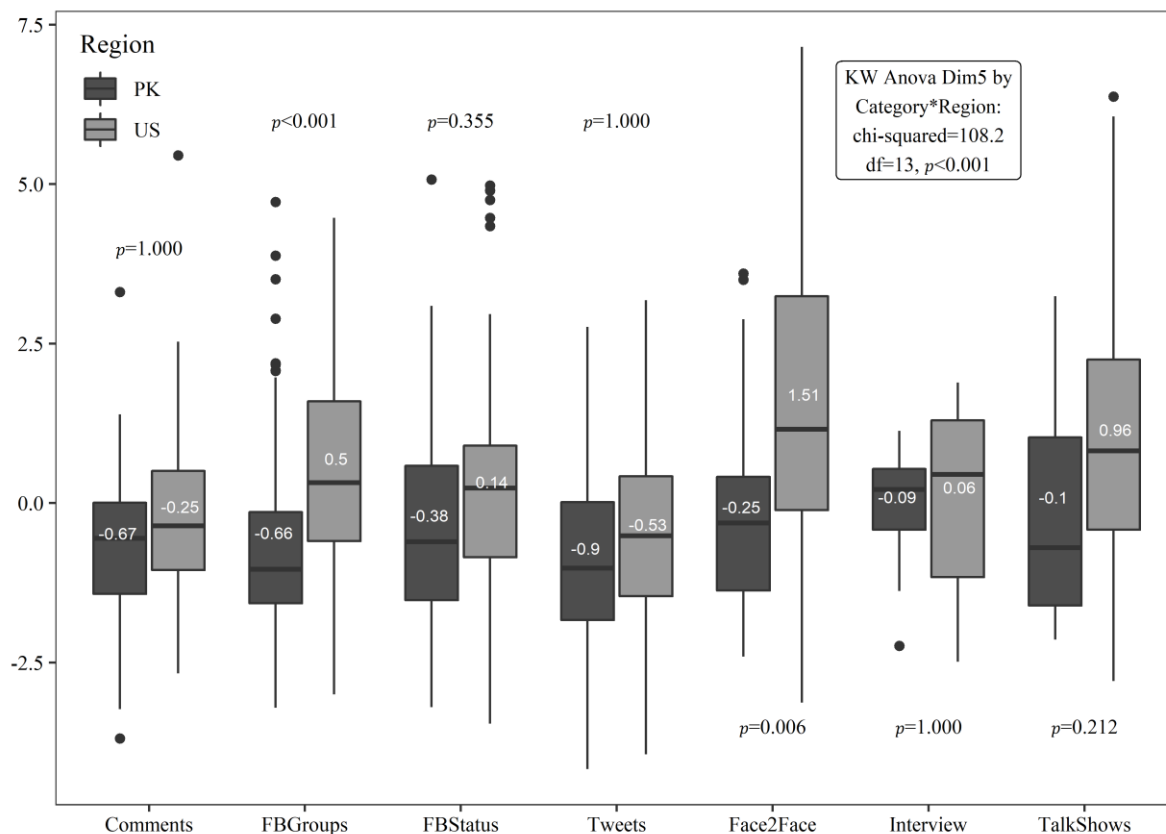
Figure 5.18 Group comparisons: Online interactive registers versus spoken conversations on dimension 5 Narrative Focus

Table 5.17 confirms the above observation that this dimension is not much relevant to Pakistani interactive registers, as the post-hoc comparisons also indicate. U.S. face-to-face conversations and talk shows, on the other hand, are slightly more narrative as compared to some online interactive registers.

Table 5.17 Online interactive registers versus spoken conversations: Region-wise group comparisons on dimension 5 Narrative Focus

	Comments	FBGroups	FBStatus	Tweets
PK: Kruskal-Wallis chi-squared = 14.5398, df = 6, *p* = 0.02				
FBGroups	1.046 (1.000)			
FBStatus	-0.583 (1.000)	-1.644 (1.000)		
Tweets	1.658 (1.000)	0.578 (1.000)	2.285 (0.234)	
Face2Face	-1.174 (1.000)	2.002 (0.475)	0.724 (1.000)	2.493 (0.133)
Interviews	-1.477 (1.000)	-1.925 (0.569)	-1.233 (1.000)	2.176 (0.31)
TalkShows	-1.033 (1.000)	-1.827 (0.71)	-0.600 (1.000)	2.293 (0.229)
US: Kruskal-Wallis chi-squared = 48.9379, df = 6, *p* < 0.001				
FBGroups	**-2.929 (0.036)**			
FBStatus	-1.891 (0.616)	1.173 (1.000)		
Tweets	0.906 (1.000)	3.857 (0.001)	2.859 (0.045)	
Face2Face	-4.805 (<0.001)	2.355 (0.195)	3.398 (0.007)	5.577 (<0.001)
Interviews	-0.814 (1.000)	0.492 (1.000)	-0.016 (1.000)	1.204 (1.000)
TalkShows	-3.721 (0.002)	-1.290 (1.000)	-2.294 (0.229)	4.482 (<0.001)

5.1.7.6 Interactive registers: section summary.

This subsection provides a summary of group comparisons between online interactive registers and spoken conversations. Significant regional differences are only found in Facebook groups, status updates, and tweets. Pakistani Facebook groups are less conversational, and they are abstract evaluative and non-narrative. Pakistani status updates are also less oral and they are abstract evaluative. Tweets also show a significant inclination towards abstract evaluative informational discourse. In general, Pakistani online interactive registers have higher scores on dimensions 3 and 4, while the scores on dimensions 1, 2, and 5 are lower.
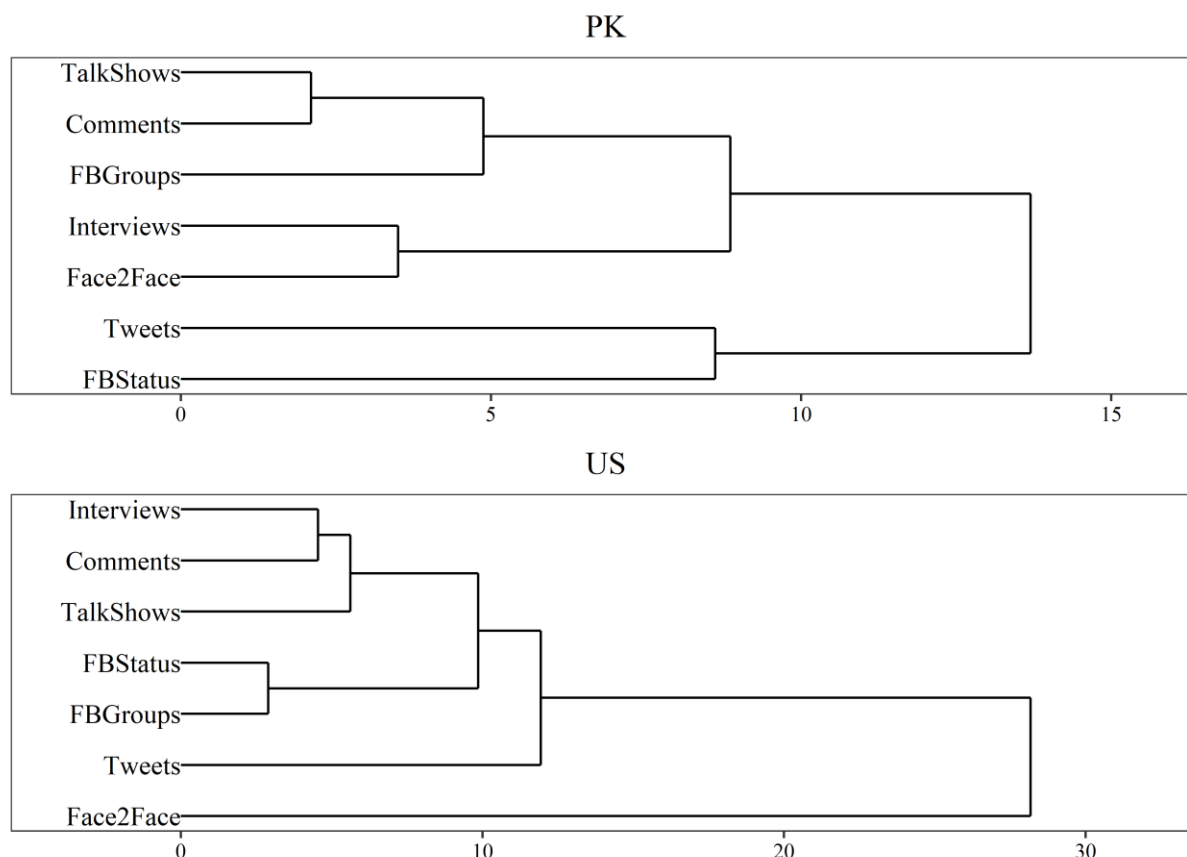
Figure 5.19 Summary of comparisons between online and offline interactive registers in both regional varieties using hierarchical cluster analyses of mean dimension scores

Figure 5.19 provides a summary of group comparisons between online interactive and spoken conversations within each regional variety. The hierarchical cluster analysis shows how registers are similar to each other, and hence how different they are from each other. Pakistani interactive registers form two unequal clusters. The upper cluster consists of five interactive registers, and is further subdivided in two clusters. The upper three registers are generally less oral and conversational, and incline towards abstract evaluative discourse. The lower two registers, i.e. face-to-face conversations and interviews, are comparatively more interactive and conversational. Facebook status updates and tweets have been put in one cluster, but the distance in the branching nodes show that these registers are less similar to each other. In fact, it could be

argued that there are four clusters in the Pakistani data, where Facebook status updates and tweets make up the third and fourth clusters.

No clear-cut clustering is visible in U.S. English, as can be observed in the lower half of figure 5.19. The registers in the first sub-cluster, i.e. interviews, talk shows, and comments, are generally less oral and personal. Facebook groups and status updates are quite similar on dimensions 1, 3, and 5. Twitter, like in Pakistani English, is quite distinct. Face-to-face conversations are added to the cluster tree at the end, confirming that they differ significantly from most online interactive registers. They are highly oral, fragmented, non-abstract, and narrative.

The cluster patterning of interviews and face-to-face conversations is different in both regional varieties. Interviews has only 10 texts, and differences can be attributed to the small sample and the differences in the format of interviews (radio interviews in the case of the Pakistani data as opposed to TV interviews in the case of the U.S. data). The comparison of face-to-face conversations, on the other hand, shows that native speakers use face-to-face conversations for very different functional purposes, which results in linguistic differences. Pakistani face-to-face conversations, on the contrary, are comparatively similar to other interactive registers like Facebook group discussions, talk shows, and comments. This shows that Pakistani English is limited in terms of linguistic and functional diversity, as compared to U.S. English.

5.1.8 Gender differences in online registers

The possibility of variation in various online registers based on the gender of the writers is explored in this section. Table 5.18 provides an overview of text distribution based on the gender of the writers in both regional varieties, and register categories. As the table shows, not

all online registers have been included in this analysis. Facebook groups and comments are discarded, because they have discussion threads produced by multiple users, so the text cannot be distinguished based on a single writer. Individual blogs are subdivided into single-writer and multi-writer blogs, as multi-writer blogs could include male and female writers in a single text. The gender of the authors was confirmed by looking at the meta information available on the respective blogs, or in some cases the writer names provided indications of their gender. The texts with unclear gender of the writer have not been included. Lastly, the division of texts based on gender is not equal, as it can be observed in the table that the texts produced by male writers are approximately double as compared to female writers.

Table 5.18 Distribution of texts in online registers of Pakistani and U.S. English based on the gender of the user

Category	Pakistani English		U.S. English	
	Female	**Male**	**Female**	**Male**
BlogsSW	31	29	23	28
BlogsMW	28	103	81	91
BlogsNews	67	109	116	165
BlogsNM	34	23	38	58
BlogsTech	40	181	34	105
FBStatus	42	62	45	63
Tweets	41	74	44	58
Total	283	581	381	568

The following subsections provide comparisons of gender differences on dimensions 1 and 3. The other dimensions are not included, because they showed mixed or no trends. Figures 5, 6, and 7 in Appendix I provide these details.

5.1.8.1 Gender differences in online registers on dimension 1 Oral versus Literate Discourse.

Figure 5.20 outlines mean dimension scores of each gender in Pakistani and U.S. English. As the graph reveals, there are mixed trends or no differences between both genders in most registers of U.S. origin. Pakistani English, on the other hand, shows that, apart from new media blogs, female writers produce slightly more oral texts, as compared to their male counterparts. However, this is just a trend and the differences are not statistically significant.
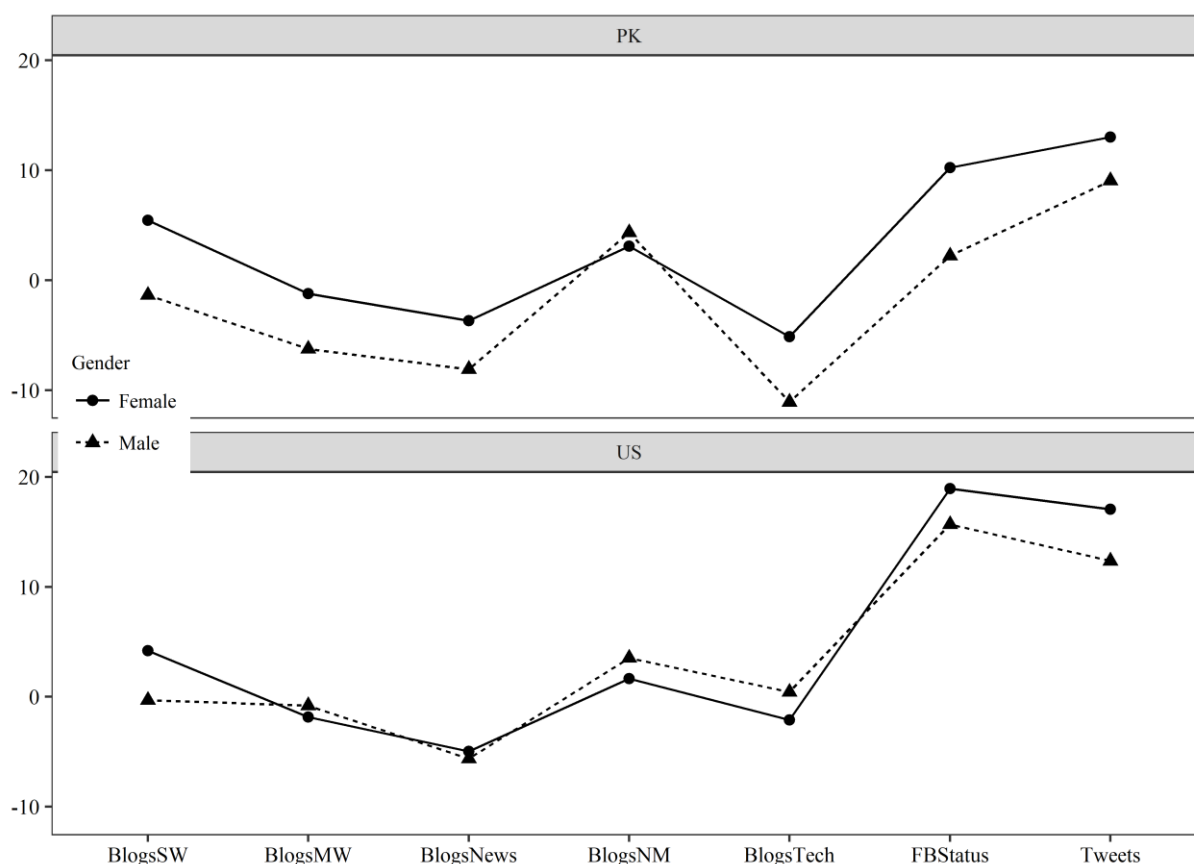
Figure 5.20 Online texts produced by male and female users on dimension 1 Oral versus Literate Discourse (mean dimension scores)

5.1.8.2 Gender differences in online registers on dimension 3 Abstract Evaluative versus Non-Abstract Information.

Gender differences in most registers of Pakistani and U.S. origin are mixed, with no clear trend, as Figure 5.21 explains. However, three online registers in Pakistani English, i.e. single-writer blogs, multi-writer blogs, and news blogs, show a consistent trend, where male writers incline towards abstract evaluative informational discourse, and female writers have comparatively lower scores. The difference is most visible in single-writer blogs, which is also the only register with significant differences between both genders.
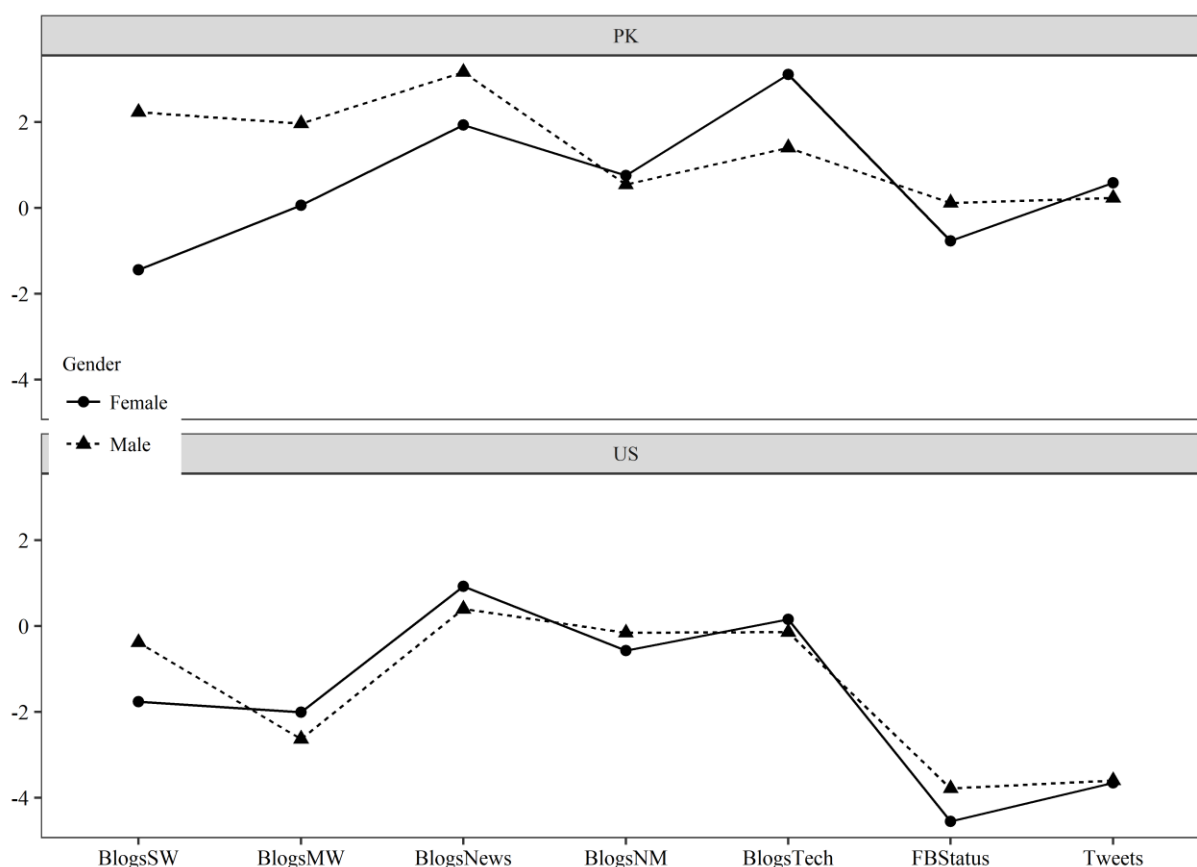
Figure 5.21 Online texts produced by male and female users on dimension 3 Abstract Evaluative versus Non-Abstract Information (mean dimension scores)

5.1.8.3 Gender differences: section summary.

Gender differences between the texts produced by male and female users of Pakistani and U.S. origin are mostly nonsignificant, with no clear-cut trends. However, male users of Pakistani origin write slightly less oral, or more literate texts. Additionally, if the findings of dimension 1 and 3 are combined, male Pakistani individual and news blog writers produce slightly more informational and abstract evaluative texts, as compared to their female counterparts.

5.1.9 Case study: chronological differences in technology blogs

The register of technology blogs is a professional register, which focuses on technology related topics. It has been observed in subsection 5.1.6 that technology blogs show quite distinctive regional traits.

Table 5.19 Year-wise distribution of texts with word count in technology blogs: ProPakistani (Pakistani English) and TechCrunch (U.S. English)

Year	Pakistani English		U.S. English	
	Texts	**Words**	**Texts**	**Words**
2005	-	-	113	34,672
2006	-	-	261	65,088
2007	-	-	288	73,989
2008	181	64,729	353	104,198
2009	319	116,488	360	134,001
2010	343	120,883	62	30,214
2011	360	126,962	360	142,203
2012	338	110,606	360	197,833
2013	360	122,462	360	186,248
2014	360	131,674	360	199,084
2015	360	136,815	360	196,291
2016	360	141,558	360	217,811
2017	360	134,756	360	173,254

Year	Pakistani English		U.S. English	
2018	60	22,869	49	21,846
Total	3401	1,229,802	4006	1,776,732

The aim of this section is to explore year-wise linguistic variation in technology blogs to find out if there is a shift going on. Two technology blogs were selected from both regional varieties for this purpose. ProPakistani is one of the oldest technology blogs from Pakistan, which has been online since 2008. Similarly, TechCrunch has been selected from U.S. English, which started in 2005. Table 5.19 provides an overview of the data, which was especially collected for this case study in early 2018. On average 360 texts were collected using a stratified and opportunistic sampling approach. The texts were sampled in such a way that around 30 texts belong to each month. However, as it can be observed from years 2005, 2008, 2010, and 2018, the number of texts is not stable across years. There are not many texts available in the years 2005 and 2008, as the blogs started in these particular years. Only a couple of months from the year 2018 were included in the data, which is the reason for the small number of texts in this year. Lastly, only 62 texts were collected from TechCrunch in 2010, which was the result of a technical bug in the webpage downloading software. This data did not go through the usual process of review and correction, as the main part of the data used in previous sections.

The data was tagged using the Biber tagger, and the dimension scores were calculated as explained in subsection 5.1.4. The following subsections discuss the year over year trend in both technology blogs on the previously identified dimensions. Only dimensions 1 and 4 are included in this section, while the graphs for other dimensions are included in Appendix I.

5.1.9.1 Chronological differences in technology blogs on dimension 1 Oral versus Literate Discourse.
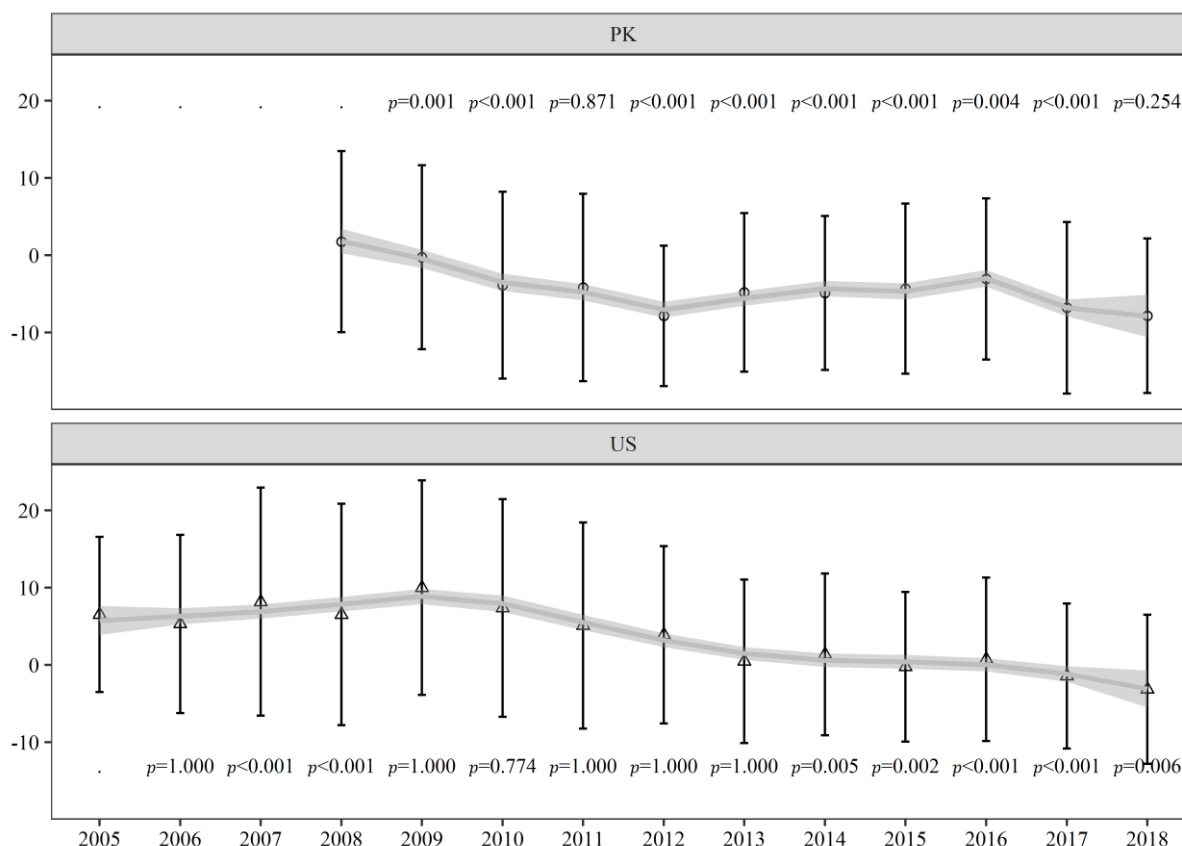
Figure 5.22 Chronological trend of technology blogs on dimension 1 Oral versus Literate Discourse (mean dimension scores and standard deviation as error bars)[31]

Figure 5.22 shows the yearly trend of both technology blogs on dimension 1. The graph confirms that Pakistani technology blogs are more literate as compared to their U.S. counterparts. Additionally, the regression line reveals that there is a steady decrease in dimension scores over the years. The mean dimension scores from year 2009 onwards in the Pakistani data are

---

[31] *P* values are based on post-hoc Dunn's test comparing each year to the first year. Regression lines have been calculated using Generalised Additive Model gam() function in R graphics library ggplot2 (Wickham, 2009).

significantly different from the starting year 2008, except years 2011 and 2018. A number of years in U.S. data are also significantly different from the starting year 2005, especially years from 2014-2018. The graph shows an increase in literate informational discourse over the years, which in turn points towards a stylistic shift going on in the technology blogs of both varieties.

Table 5.20 presents four text samples from both regional varieties taken from the very first year and the year 2018. Since the year over year trend shows a progression from oral to literate discourse, the text samples also contrast between oral versus literate. Though not all texts in the respective years have such a high positive or negative score, the mean dimension scores and standard deviations do show that the number of texts with high positive scores decreases over time in both varieties.

Table 5.20 Sample texts from technology blogs with high scores on dimension 1 Oral versus Literate Discourse taken from years 2005, 2008, and 2018

Text Samples
Buying a mobile phone **is** not an **easy** task. Sometimes **you** **make** mistakes that you **regret** later on, as **you try** to **get** a mobile phone but **fail** to judge **if it's** worth buying. By the way, just buying **what is popular is** not always a **good** idea. **I** have made this guide for **you** guys, so **you can consider** these things and **buy** the **best** phone in market. **This will** surely help **you** to **buy** a mobile that **you will never** regret later on. Price: Teens **don't earn** their money. But **love** to **have** a fully loaded media mobile phone. Teens **should consider** their budget. Not **everyone can buy** an expensive mobile phone. **See** you budget range and then **list** out the mobile phones available in that range. *File# PTC_207, Year: 2008, Variety: Pakistani English, Dimension Score: 35.31* Waxmail **is** a great productity tool that **allows you** to attach mp3 audio files to an email. **We** first wrote about Waxmail late last month **when they** released their Outlook product. Waxmail just announced that they now **have** support for Outlook Express. **I'm hoping ɸ** they **include** other email applications over time - like Mail and Entourage for the Mac. A lot of people overlooked Waxmail as just another audio recording tool. But **I've** found that **I use it more** and **more** often. There **is** just **something** very cool about recording a quick sound file to accompany certain emails. **It's** sometimes **easier** to say **what you think** sometimes rather than write **it**. *File# PTC_207, Year: 2005, Variety: U.S. English, Dimension Score: 31.73* A **notification issued by the Pakistan Electronic Media Regulatory Authority's (PEMRA) Operation Wing** today (Wednesday) has barred **all** the **licensed** channels and networks **from** promoting Valentines Day **in** Pakistan.

Text Samples
The directive **has been issued by** **Islamabad High Court** following a petition (NO.541/2017) **against** the **promotion** **of** **Valentines Day** which **was started by** an individual named **Mr.** **Abdul Waheed**. The **notification** **has been directed** to **all** **Pakistani** **satellite** TV channels, FM radio licensees, and distribution service licensees.   The petitioner's, **to** their claim, **said** that **Valentines Day** **should not be celebrated** **in** **Pakistan** because it is not part **of** **Muslim** **tradition**.   *File#* PTC_11490*, Year: 2018, Variety: Pakistani English, Dimension Score: -27.74*   **Samsung** vice chairman and the company's heir apparent **Jay Y. Lee** **has left** prison **after** a South Korean high court **suspended** his **bribery** sentence.   **Lee**, whose father is Samsung's chairman, **was previously sentenced** **to** **five** years **after** being found guilty **of** bribery, **embezzlement**, capital flight and **perjury** charges. His sentence **was** **reduced** **to** 2.5 years today **on** appeal, and he **was allowed** to leave **on** a **four-year** **probation**, according to **Bloomberg**.   **Lee** **was first arrested** last February and he **had served** close **to** **one** year **in** custody. **His** absence **hasn't impacted** **Samsung's** business, which **has posted** a series **of** **record** financial returns **over** the **past** year.   *File# PTC_207, Year: 2018, Variety: U.S. English, Dimension Score: -22.87*

The text from Pakistani English on the positive side is a cell phone purchase guide written by the founding author of ProPakistani. The text contains linguistic features like 2nd person pronouns, 1st person pronouns, present tense, mental verbs, activity verbs, demonstrative pronouns etc., which are typical features of interactive and personalised discourse. In contrast, the text taken from the year 2018 is a press release, which is highly impersonal and only focuses on conveying information using passive voice, post nominal passive modifiers, prepositions, and proper nouns. The same is the case with the first text taken from TechCrunch in the year 2005, where the writer provides a review of a software product. The text has a personal touch due to the presence of 1st and 2nd person pronouns, with additional features like present tense, predicative adjectives, *WH* clauses, and pronoun *it*. Similar to the Pakistani text, the text sample from 2018 is a news report with an abundance of prepositional phrases, perfective aspect, proper nouns, and passive voice. These text samples show how blogs have evolved from the informal,

conversational, and oral style of blogs to the more impersonal news reports-like style in the last 10 or so years.

5.1.9.2 Chronological differences in technology blogs on dimension 4 Reporting Style.
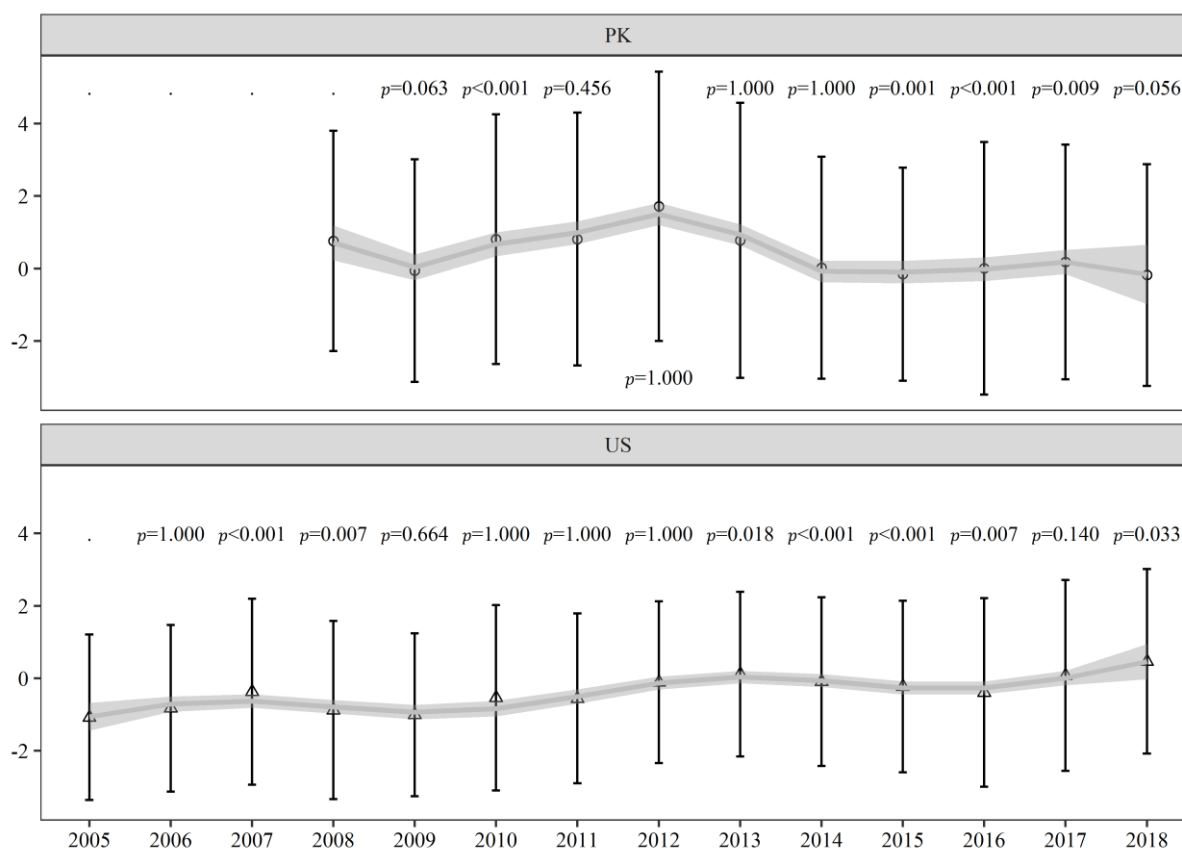
Figure 5.23 Chronological trend of technology blogs on dimension 4 Reporting Style (mean dimension scores and standard deviation as error bars)

Figure 5.23 shows the year over year trend of both technology blogs on dimension 4. As it is clear from the upper half of the figure, the Pakistani technology blog has a haphazard trend regarding reporting style over the years. However, the U.S. technology blog shows a steady increase over the years in linguistic features loaded on this dimension, i.e. communication verbs with and without *that* clauses, *that* clauses controlled by communication verbs, and *that* clauses

controlled by general verbs. Though this dimension is not very strong, apparently texts reporting what is said by someone else are increasing in number over the years.

5.1.9.3 Chronological differences in technology blogs: section summary.

Only dimensions 1 and 4 have been found relevant to the year over year evolution of technology blogs in both varieties. The trend on dimension 1 shows that technology blogs started as an extension of personal and other blog types, which adopted the same informal and personal style. However, over the years this blogging style is decreasing steadily. The trend on dimension 4, at least for the U.S. technology blog, confirms that the focus is shifting to conveying information about someone else or what is communicated by others.

5.1.10 A summary of findings of MD analysis

Table 5.21 Variance explained (R^2 values) by three categorical variables on five dimensions

	Oral versus Literate Discourse	Conversational Style	Abstract Eval. vs Non-Abstract Info.	Reporting Style	Narrative Focus
Category	0.492	0.311	0.236	0.13	0.136
Region	0.027	0.070	0.067	0.000	0.014
Internet	0.062	0.002	0.062	0.057	0.021

The following summary points can be listed as findings of MD analysis:

- Out of five dimensions, the first three dimensions are moderately strong, while the last two dimensions are less relevant.

- As per table 5.21, register or situational variation is much more important as compared to regional variation.

- Dimensions 1, 2, and 3 are relevant to regional variation. Pakistani registers are generally less oral, less conversational, and more abstract evaluative.

- Dimensions 1, 3, and 4 are somewhat relevant to the online versus offline distinction.

- Blogs are generally more oral, less abstract evaluative, and less reporting as compared to news reports and columns.

- Pakistani technology blogs, news blogs, Facebook groups, status updates, and tweets generally show significant differences as compared to their U.S. counterparts.

- Technology blogs in Pakistani English and news blogs in U.S. English are more like columns and news reports in their respective regional variety.

- Online interactive registers are generally less oral, less conversational, less reporting, and non-narrative as compared to spoken conversations.

- Comments, interviews, and talk shows are quite similar.

- Regional differences in relation to face-to-face conversations and online interactive registers show that Pakistani interactive registers have limited functional range and linguistic diversity.

- The texts produced by Pakistani male individual and news blog writers are more literate and abstract evaluative as compared to the female writers.

- The chronological evolution shows that technology blogs are becoming more like columns and news reports, moving away from the informal and conversational style associated with blogs.

5.2 Finding Text Types using Cluster Analysis

This part of the chapter analyses the dimension scores data obtained after the EFA to put the texts from online registers into similar groupings. Another dimensionality reduction

technique called Cluster Analysis (CA) has been used in this process. The following subsections provide the necessary background, the procedure, and the results.

5.2.1 The rationale behind text types

As explained in <u>subsection 2.3.1.4</u>, Biber and colleagues in various studies (e.g. Biber, 1989; Biber, 2004; Grieve et al., 2010) distinguish between register and text type. The former, in their opinion, is a linguistic variety defined solely based on situational characteristics, while the latter is defined on the basis of linguistic characteristics. In the present case, Facebook groups, status updates, and various types of blogs are registers, because they are defined by extra linguistic and situational characteristics. However, each register can include a range of texts with a different distribution of linguistic features, as it has been shown by the box plots in the detailed group comparisons in the previous part of this chapter. Hence, highly oral texts can exist in, for example, Facebook groups as well as in spoken conversations. Such texts with similar linguistic traits belong to the same text type. The linguistic traits here refer to the dimension scores allocated to each text after conducting an EFA, which in turn is based on individual lexico-grammatical and semantic features. In practical terms, the statistical technique of CA is applied to cluster individual texts based on their dimension scores, irrespective of their register category. The resulting groupings/ clusters are referred to as text types.

5.2.2 Prerequisites of CA: clustering method, cluster-ability, and number of clusters

Generally, three types of clustering techniques are identified (Popat and Emmanuel, 2014): partitional clustering, hierarchical clustering, and density-based clustering. Previous studies of text types, e.g. Biber (2004), have used non-hierarchical methods, because theoretically no hierarchical structure was expected. The same assumption also holds for this particular dataset. The most widely used partitional clustering techniques available in R (R Core

Team, 2013) are k-means and Partitioning Around Mediods (PAM) clustering. The former is a very popular clustering method, but it is sensitive to outliers (Popat and Emmanuel, 2014, p. 811). PAM, on the other hand, uses actual data points in the given dataset to start the clustering process, which makes it robust against outliers (Popat and Emmanuel, 2014). As a result, Partitioning Around Mediods was selected as the clustering method.

The next step was to find out the cluster-ability of the data, which was performed using Hopkins statistic available in clustertrend package (YiLan and RuTong, 2015) in R. According to Kassambara (2017, p. 124), the null hypothesis in this test statistic is that the data is uniformly distributed, and hence not cluster-able. A value lower than 0.5, or nearer to 0 indicates that the null hypothesis can be rejected. This value for the present dataset was 0.1724581, which is well below the threshold of 0.5. Kassambara (2017) as well as Grieve (2018) also recommend the use of visual methods to confirm cluster-ability. This issue is discussed again in the next section, where the results of the CA are presented.

Table 5.22 Number of clusters as recommended by NbClust()

Number of clusters	Frequency of clustering methods
0	2
1	1
2	9
3	4
4	5
8	1
9	3

The third step in performing the CA was to determine the number of clusters to extract from the dataset. Kassambara (2017) recommends the use of NbClust() function (Charrad Ghazzali, Boiteau, and Niknafs, 2014), which employs 30 different methods to estimate the best number of clusters in the given dataset. As table 5.2 reveals, nine methods have recommended

two as the best number of clusters, while five methods have recommended this number to be four. The second value, i.e. four clusters, was adopted due to better interpretability of the resulting clusters. Finally, the clusters were extracted using pam() function in cluster package (Maechler, Rousseeuw, Struyf, Hubert, and Hornik, 2017) in R.

5.2.3 Results and limitations of CA
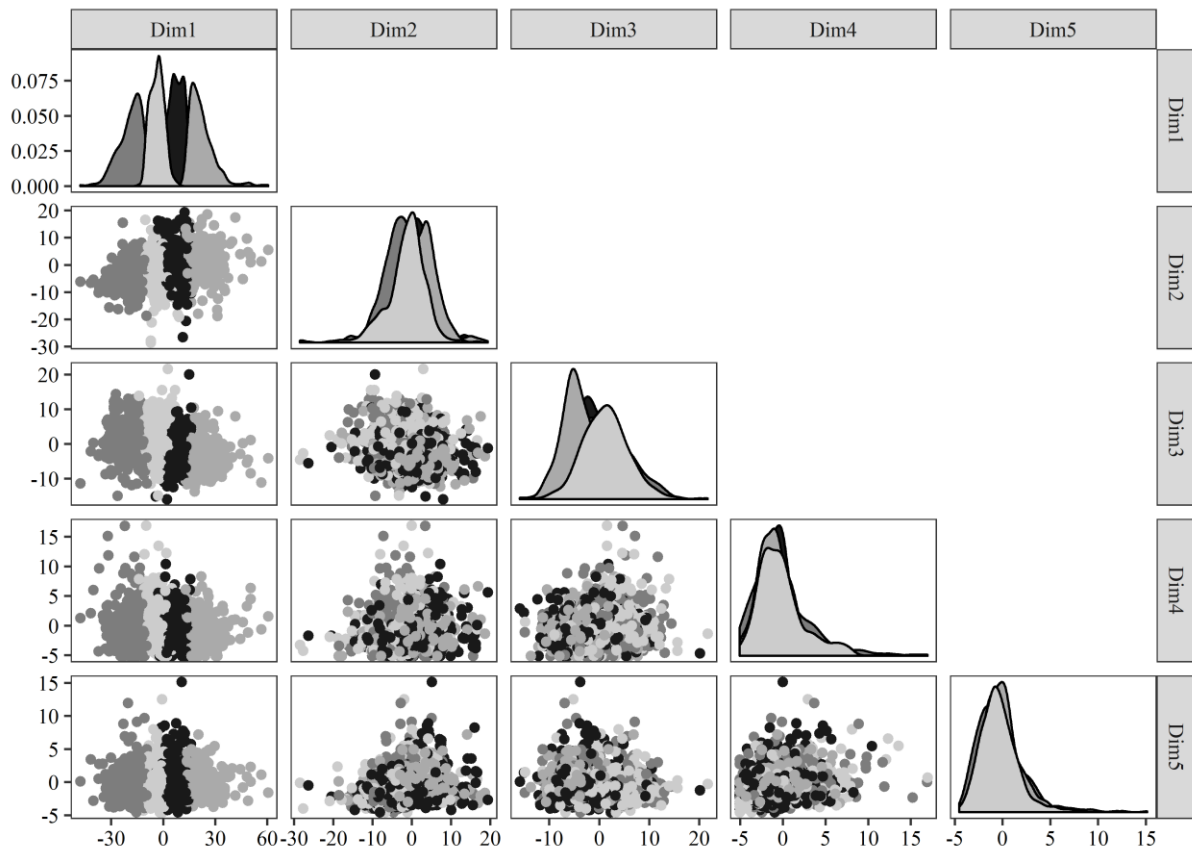
Figure 5.24 Dimension-wise visual description of 4 clusters

A visual depiction of the resulting clusters is provided as a scatterplot matrix in figure 5.24. Each dimension is plotted against the other four dimensions to show the distinction between the four clusters. The figure exhibits that dimension 1 is the most relevant and divisive measure in defining and distinguishing the clusters. The data shows mixed results on the other four dimensions. Additionally, the data points are generally very close together, which indicates the presence of a continuum instead of distinctive text groupings.

A summary of the cluster analysis results is provided in table 5.23. Cluster 1 is the biggest cluster with the lowest median within cluster distance. Column 3 provides the average silhouette width of each cluster, which is a measure of cluster strength. As per Kassambara (2017, p. 141), a silhouette coefficient that is nearer to 1 indicates a very good cluster. The silhouette values in the table also confirm that the resulting clusters are rather loose groupings of texts, with a large number of 'peripheral texts' instead of 'core texts' (Biber, 1989, p. 16).

Table 5.23 Summary of the cluster analysis

Cluster Number	Cluster Size	Average Silhouette Widths	Maximum within Cluster Distance	Median within Cluster Distance
1	511	0.275	45.97	9.21
2	314	0.305	42.06	11.55
3	346	0.283	48.90	10.51
4	391	0.271	45.56	9.55

5.2.4 Interpretation of cluster results: Text Types

Table 5.24 Cluster mediods

Cluster	Dim1	Dim2	Dim3	Dim4	Dim5
1 Unmarked Oral	7.9	1.43	-1.83	-0.82	-0.46
2 Literate Nominal	-17.82	-3.325	-0.06	-1.05	-0.845
3 Oral Non-abstract	20.27	2.505	-4.42	-1.03	-0.15
4 Unmarked Abstract Evaluative	-3.35	-0.9	1.32	-0.62	-0.59

Four texts included in the dataset along with their dimension scores are shown in table 5.24. Each text represents the centre point of the respective cluster. The first three dimensions are the most relevant, while the last two dimensions do not show much variation in relation to different clusters. Clusters 2 and 3 are the most distinctive ones, with very high dimension scores on both sides of dimension 1. Additionally, the text scores on dimensions 2 and 3 are also taken into consideration in naming these two clusters. A positive score on dimension 3 only manifests on cluster 4 with other dimensions having non-distinctive scores. Hence, this cluster is named as 'Unmarked Abstract Evaluative'. Lastly, cluster 1 is also an unmarked grouping of texts, which are generally moderately oral. However, the inclusion of almost 50% of the texts from online registers in this cluster shows that this is the most common text type.

Table 5.25 provides text samples from cluster 1, which has generally moderately oral texts. The samples, which are taken from news and technology blog comments, contain a mix of oral and literate features, *do* as pro-verb, 1ˢᵗ person pronouns, 2ⁿᵈ person pronouns, prepositional phrases, predicative adjectives, nouns, and adverbs etc.

Table 5.25 Text samples from both regional varieties on Cluster 1 Unmarked Oral

Text Samples
[…] Nice thoughts, do we have enough evidence that animals are not eating seed cake made of gmo crops seed. If the author could show in a video this impact then it would have strong impact on people. By now, its all talks. We as a nation need to be more scientific in approach rather than just talks without any evidence <#>
Make sure to make a video of animals not preferring GMO seed cake. To be honest, take two different seed types, GMO and non GMO. Use two different extractor to for oil extraction, so that you may not mix the treatment. Then let the cake cool down. Offer two different cakes to animal and label it. See the response. Make video of all this. 2nd our people are very intelligent and know how to adulterate the foods 'milawat', . You never know that cake which animals are not eating, might have something mixed by producer.. But in that case too it needs evidence. […]
Text# PK_CN_5_6.txt, Register: Comments News, Variety: Pakistani English, Distance from Cluster Centre: 3.39
Wow. The writing in this article. <#>
Get a copy editor on this one, LMAO <#>

Text Samples
chill out y'all @XXX and @xxx :\ <#>

Constine was in a hurry brother chill laugh out loud

<#> I'm a Spotify subscriber but I'm going to check out Tidal now. I canceled my Apple Music subscription right before it renewed after the 3-month free trial period ended. I wasn't sold on their UI. I mean it would be cool for Apple to completely control my life but not just yet.

<#> Let's make some bets on what Apple will pay for it? 150 million.

<#> 2 billion

<#> Yeah I think I'm way off. Probably in the 100's of millions of not hitting there billion dollar mark. We shall see […]

Text# CT_US_3_5.txt, Register: Comments Tech, Variety: U.S. English, Distance from Cluster Centre: 1.93

Text samples from cluster 2, which has the smallest number of texts, are given in table 5.26. The texts exemplify a highly literate and nominal discourse with features like proper nouns, prepositional phrases, passive voice, and perfective aspect.

Table 5.26 Text samples from both regional varieties on Cluster 2 Literate Nominal

Text Samples

Akhuwat is implementing Chief Minister Self-Employment Scheme #Gilgit Baltistan. This is the largest interest free micro lending program in GB which has served more than 20,000 families which constitute 20% population of Gilgit Baltistan. We are grateful to the Chief Minister and the Chief Secretary for their trust and support. Credit for initiative and continuation of this program goes to Mr. Sajjad Salim, Mr. Younis Dagha and Mr. Sikander Sultan, former Chief Secretaries of Gilgit Baltistan.

<#>

Zakat for clothes is giving one old pair to a needy person when you buy a new pair. Let's distribute. It's destined to come back.

<#>

Spanning across generations, Akhuwat's message of compassion and solidarity will resonate long after each of us has played their part in the movement. Our hopes rest with the young ones. […]

Text# 532001.txt, Register: FB status, Variety: Pakistani English, Distance from Cluster Centre: 4.01

Here's a short break from carping, and another sampling of sparkling prose from recent Times stories.

Business Day, 2/4:

Playboy Puts On (Some) Clothes for Newly Redesigned Issue

Paradoxical as it may sound, Playboy has undergone major cosmetic surgery and emerged from the operating room looking more natural. …

Text Samples
In short, the new Playboy, which will appear on newsstands as early as this weekend, has ditched its jauntily illicit aura and become a slightly saucier version of a lot of other magazines, like Esquire and GQ. But the March issue retains elements of the original DNA, including a lengthy interview (with the MSNBC host Rachel Maddow) and a long essay by a famous writer […] *Text# 613070.txt, Register: BlogsNews, Variety: U.S. English, Distance from Cluster Centre: 0.87*

The text samples from cluster 3 in table 5.27 are more personalised and interactive as compared to the text samples in cluster 1. The use of 1st and 2nd person pronouns, amplifiers (e.g. *totally*), predicative adjectives, contractions, linking adverbials (*so*), and mental verbs etc. can be spotted.

Table 5.27 Text samples from cluster 3 Oral Non-abstract

Text Samples
[…] Congratulations dear Mishi for your award. You truly deserve it. And thank you for passing it down to me(blush) I`m humbled by your kind words. Have a lovely day. I shall do a post about this soon :) <#> @Pandora..oh com'and you totally deserved it girl;-) Congrats to you too;p <#> @ankita thank you dear...keep visiting more then:) <#> It's kind of hard to keep up with your speed, yet, I feel more than happy to go through your posts almost on daily basis. Brilliance is a thing of beauty to me. You are too creative to be ignored. You are a home maker, socially inclined, creative writer -----and what not---list is already going on and on. […] *Text# CI-002.txt, Register: Comments Ind., Variety: Pakistani English, Distance from Cluster Centre: 4.41* <#> Thow back to all the middle school music with Nikki Speranzo Joy Awobue Amanda Low Hahaha. Battle field...... Oh no...... <#> Alexis Chai *listening to tove llo* "I wanna fuck her voice!!!" The majority of what I hear about if not supernatural from my roommate........ <#> So Adam Fuchs has concluded that I'm the main girl in elf because she said go away to buddy lml <#> So first week at Darrow finished. Its pretty good!!! Hopefully its stay that way hahah

Text Samples
<#> so my laptop was for some reason not uploading my vid so i just took pictures of it instead i would like thanks devin for nominating me and i would like to challenge drew, adam, tori, lucy and devin to do it in 24 hours or pay 100 to charity. *Text# 632015.txt, Register: FB status, Variety: U.S. English, Distance from Cluster Centre: 0*

Table 5.28 Text samples from cluster 4 Unmarked Abstract Evaluative

Text Samples
[…] Kudos to XX and Mr XX for their valuable comments. Ms Farahnaz seems to adopt a line of least resistance for starters. Couldn't care less who is at the helm of affairs, as long as I get my daily bread and my regular business in an environment of bumpy or blissful path. Isn't that a disconnect with that Social Contract the civilized world had been vociferating all along. No man is an island to himself. We live in a society, a community, a plurality, a Nationality and we cannot remain isolated with the individual or collective problems. We choose the rulers, just as we choose our goals and our destiny. It is totally a wrong concept that once a government has been elected, it remains unfettered, uncontrolled and uninhibited. In a democratic country, any government that fails to deliver, automatically loose their mandate to govern. The performance of a government is almost akin to that of a beehive under a glass shell. […] *Text# PK_CN_2_4.txt, Register: Comments News, Variety: Pakistani English, Distance from Cluster Centre: 0* This clip of Republican presidential candidate Chris Christie discussing his religion and his sex life is an entertaining demonstration of the embarrassing difficulty of going public on the subject of God and sex, sex and religion. The guy in the white shirt in the foreground appears to be wishing this weren't happening (which suggests he's a Christie supporter.) I see it as a performance almost worthy of Donald Trump. I don't quite get why he felt he needed to address this subject, but then neither do I understand why I feel compelled to write about it. I do think a lot about how to write these posts in a non-tacky manner that doesn't cause cringing. […] *Text# BI_US_55.txt, Register: Blogs Ind., Variety: U.S. English, Distance from Cluster Centre: 6.82*

The text sample taken from Pakistani news blog comments contains vocabulary like nominalisations and various types of abstract nouns loaded on the positive side of dimension 3, as table 5.28 shows. The text sample from U.S. individual blogs, though talking about religion and politics, is rather less abstract evaluative as compared to its Pakistani counterpart.

5.2.5 Distribution of text types across online registers
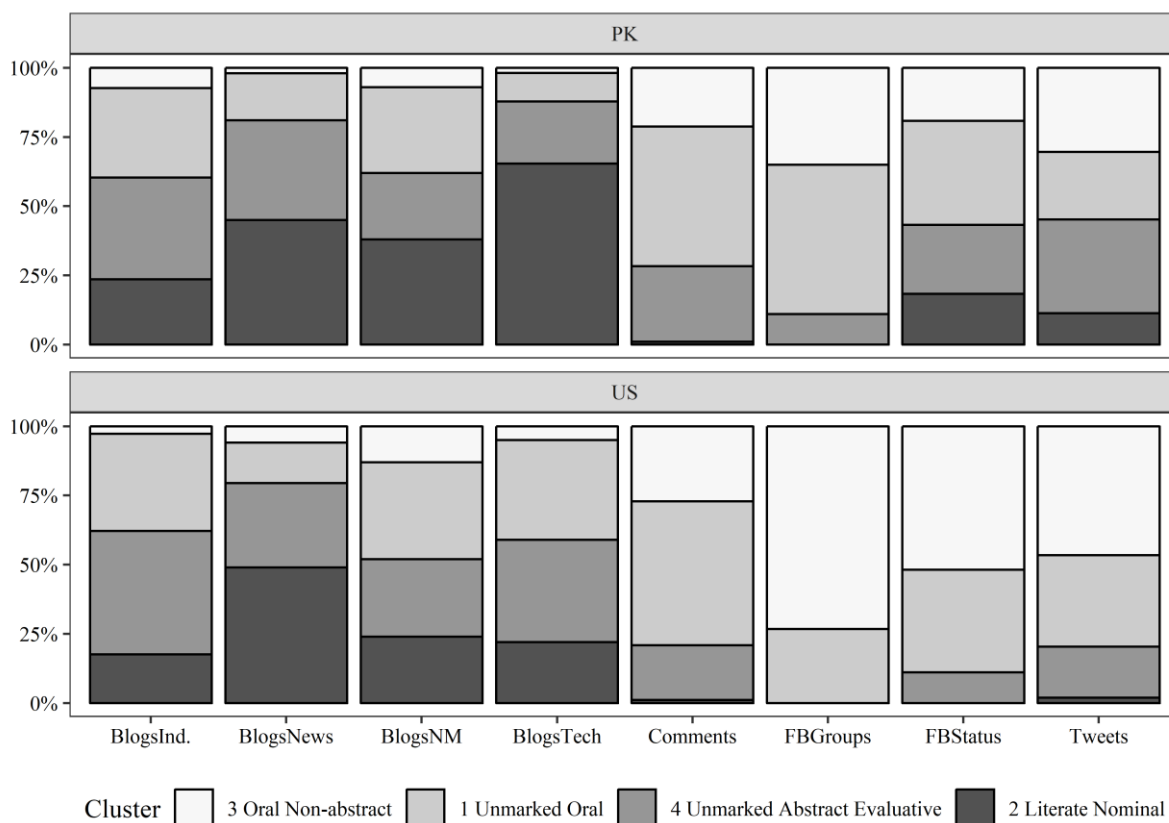
Figure 5.25 Distribution of text types across online registers in both regional varieties

The percentage distribution of the four text types is described in figure 5.25 using a stacked bar plot. The legend is arranged from highly oral texts on the left to highly nominal texts on the right. Hence, the darker the colour, the more literate the text type is. As it can be seen, oral non-abstract types of texts are less common in Pakistani English, with the exception of individual blogs, where a very small number of blogs are highly oral and non-abstract. Two registers that clearly stand out are technology blogs and Facebook groups. Pakistani technology blogs include an overwhelming majority of texts belonging to the literate nominal text type. Similarly, the abstract evaluative and unmarked oral text types are more common in Pakistani Facebook groups, as compared to their U.S. counterparts, where the majority of texts belong to

the oral non-abstract type. Lastly, interactive online registers have a higher percentage of texts from text types 3 and 1, as compared to different types of blogs.

5.2.6 Text Types: summary

The discussion in previous subsections has shown that the boundaries defined by cluster analysis in this dataset are rather blurred, which indicates the presence of tendencies instead of distinctive text groupings. Despite these limitations, the results of the text types analysis have confirmed some of the findings regarding the variation noted in the summary section (5.1.10) of MD analysis, i.e. the texts in Pakistani English registers are less oral, more literate, and abstract evaluative. Additionally, oral non-abstract text types are more common in online interactive registers.

5.3 Regional Variation Using Canonical Discriminant Analysis

5.3.1 Canonical discriminant analysis

Canonical discriminant analysis (CDA) is a statistical technique that is used after a MANOVA – provided that the test results are significant – to find out "the linear combination(s) of the dependent variables that best separates (or discriminates) the groups" (Field, Miles, and Field, 2013, p. 719) or independent variables. Egbert and Biber (2016) note the use of discriminant analysis in studying register variation, though it has not been as widespread as EFA. According to the authors, despite fundamental differences in the statistical basis of these two statistical methods, the results should be similar because the use – and hence the co-occurrence of – linguistic features in a situational variety is motivated by communicative or discourse functions. It has been observed in parts 5.1 and 5.2 that Pakistani registers are less oral, more literate, more abstract evaluative, and less conversational as compared to their U.S. counterparts. CDA has the ability to maximally distinguish the given groups of texts (independent variables)

based on linguistic features (dependent variables). Hence, the aim of this last part of the chapter is twofold:

- to identify the group of lexico-grammatical and semantic features that maximally distinguish Pakistani English from U.S. English;

- to try and interpret the discriminating group functionally and find out the general communicative function differentiating both regional varieties.

5.3.2 Prerequisites and performing CDA

The same dataset was used in the present analysis, with the addition of a couple of additional linguistic features like definite and indefinite articles, and *to* infinitives. The second step was to eliminate the linguistic features that were nonsignificant predictors of group differences, which in this case was Pakistani English versus U.S. English. Egbert and Biber (2016, p. 13) used a stepwise discriminant analysis to eliminate such variables. The same was performed here using greedy.wilks() function in klaR package (Weihs, Ligges, Luebke, and Raabe, 2005) in R. The said function performs a forward variable selection using Wilk's lambda criterion. This automatic procedure was performed on the initial set of 100 linguistic features, which resulted in 44 linguistic features that significantly discriminated both regional varieties. Afterwards, CDA was performed using candisc package (Friendly and Fox, 2017). The function candisc() provides lists of features like EFA, which are called discriminants. The number of discriminants is equal to the number of text groups provided as input -1. Since there were only two groups, i.e. PK and US, only one discriminant or set of linguistic features along with their loadings was produced. The output also includes variance explained or R^2, and weighted factors scores like EFA. CDA was performed two times, one of them including all online and offline registers, and the other only online registers. Since the results of both analyses were more or less

similar, only the discriminant resulting after comparing online as well as offline registers is included here. The following subsections provide the details of the discriminant and its functional interpretation.

5.3.3 CDA results and interpretation

The CDA performed on online as well as offline registers resulted in 14 linguistic features that are listed in table 5.29 along with their loadings and examples. As the feature loadings show, the minimum limit of feature inclusion was reduced to +/-0.20 to help include more features and interpret them easily.

The majority of features lie on the positive side, and they also have higher factor loadings. Emphatics and adverbs of place have been observed on the oral and non-abstract sides of dimensions 1 and 3 in section 5.1. Contractions and *that* deletion are associated with informal language. *That* complement clauses and clause level coordinating conjunctions are associated with clausal as opposed to phrasal discourse, as noted by Biber (2014). Lastly, *WH* relative clauses as opposed to *that* relative clauses "are often considered to be more literate and appropriate to careful language" (Biber et al., 1999, p. 612). Though these features do not include typical pronominal features like 1st and 2nd person pronouns, which are associated with highly interactive discourse, the overall functional orientation of these features points towards oral/ informal discourse.

Table 5.29 Linguistic features with their loadings on discriminant 1 differentiating Pakistani and U.S. online and offline registers

Feature	Loading	Example
coordinating conjunctions – clause level	0.56	and, or, but It is funny *and* it is ironic.
contractions	0.52	isn't, amn't, aren't, he's
indefinite articles	0.45	a, an
adverbs of place	0.32	above, beside, outdoors…

Feature	Loading	Example
emphatics	0.29	a lot, for sure, really
that deletion	0.26	He thinks ϕ the glass is empty
that clauses controlled by verbs of likelihood	0.25	assume, believe, doubt, gather… + *that* clause
stranded prepositions	0.22	I find it so difficult to get *at*.
that relative clauses	0.21	The *boy that* is standing there, is…
type token ratio	0.21	
WH relative clauses	-0.22	The *book which* I gave you, provides…
coordinating conjunctions – phrase level	-0.23	They are *open-minded and progressive*.
agentless passives	-0.24	The snake *is killed*.
nominalisations	-0.38	ending in *-tion, -ment, -ness, -ity*

Note: variance explained is 0.47

The negative side only includes four linguistic features, which are related to nouns or informational discourse. Nominalisations have been found in previous sections to co-occur with passive voice and different types of abstract nouns on dimensions 1 and 3. Biber (1986) associates them with abstract information, as opposed to situated information. The opposition between *WH* versus *that* relative clauses has already been discussed in the previous paragraph. Agentless passives and phrase level coordinating conjunctions strengthen the assumption of literate and informational discourse. Hence, the positive and negative sides of discriminant 1 can be labelled as 'Oral/ Informal versus Literate/ Formal Language'.

5.3.4 Distribution of online and offline registers on discriminant 1 of CDA
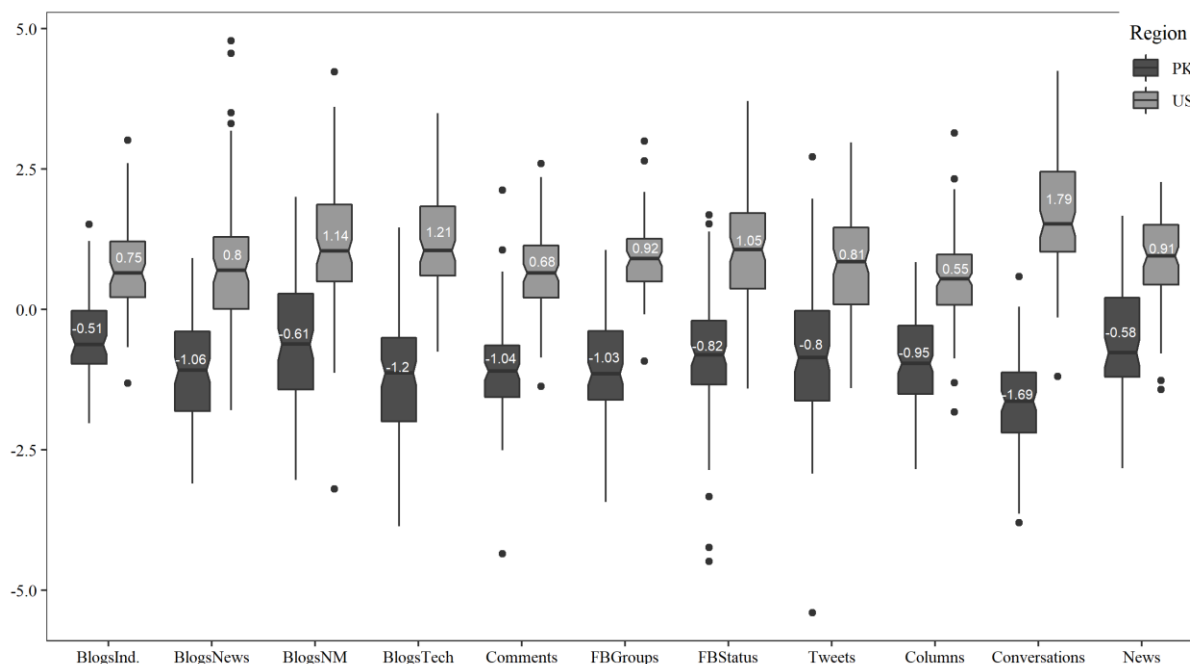
Figure 5.26 Regional variation in online and offline registers of Pakistani and U.S. English on CDA discriminant 1 Oral/ Informal versus Literate/ Formal Language

Weighted factor scores of each register category in both regional varieties are shown in figure 5.26 as box plots. The notches, i.e. the bend in boxes in the middle, show confidence intervals. The differences in each pair of registers in both regional varieties are statistically significant, because the notches do not overlap. Technology blogs, news blogs, Facebook groups, and status updates are among the online registers with high differences in factor scores between both regional varieties. The highest difference, though, is in spoken conversations, which shows how formal and literate Pakistani spoken conversations are in comparison to their U.S. counterparts.

5.3.5 Summary of CDA

It has been observed in the summary subsection 5.1.10 that regional division accounts for a very small fraction of variance in the data as compared to the situational division of the data.

Nonetheless, CDA confirms the results of MD analysis and text type analysis, i.e. Pakistani English registers resort to more literate discourse.

5.4 Conclusion and Summary of Findings

The data has been analysed using three quantitative statistical techniques, namely MD analysis, CA, and CDA. MD analysis has been the most thorough and detailed of all three. The following five dimensions of variation have been identified:

1. Oral versus Literate Discourse

2. Conversational Style

3. Abstract Evaluative versus Non-Abstract Information

4. Reporting Style

5. Narrative Focus

Cluster Analysis identified four loose groupings of texts as text types, namely:

1. Unmarked Oral

2. Literate Nominal

3. Oral Non-abstract

4. Unmarked Abstract Evaluative

Lastly, CDA identified a general dimension of variation explaining the regional differences between Pakistani and U.S. English, namely Oral/ Informal versus Literate/ Formal Language.

All three analyses have shown that literate, formal, and abstract evaluative informational discourse is the most common functional trait of Pakistani English registers, which indicates that conveying abstract and other types of information is the most common communicative function for Pakistani English users. Technology and news blogs, and Facebook groups and status updates have shown significant linguistic differences in various dimensions. Additionally, comparisons of

online and similar offline registers, comparisons of male and female writers, and year over year

comparisons of technology blogs on various dimensions have shown linguistic variation and

regional trends. The findings of this chapter are examined in chapter 7 in the light of the

situational description of online registers of Pakistani English to try and explain the reasons for

the linguistic variation.

Chapter 6 Functions of Code Switching in Online Registers of Pakistani English

6.1 Background

Code switching is the juxtaposition of two different grammatical subsystems in the same speech exchange (Gumperz, 1982, p. 59). Herring (2011) notes that synchronous CMC has been "explicitly compared to speech and conversation" (p. 2) by a number of scholars, while asynchronous CMC has been considered closer to writing instead of speech. Online interactive registers included in this study, e.g. comments, tweets, and Facebook groups, are asynchronous in nature. Nevertheless, they share some characteristics of spoken conversations, e.g. orality, as has been observed in chapter 5. Some types of blogs, e.g. individual and news media, have also been found to be oral and informal as compared to similar offline registers like opinion columns and news reports. Previously, code switching has been observed and extensively studied in online registers, e.g. bilingual blogs (Montes-Alcalá, 2007; San, 2009), discussion forums (McLellan, 2005), tweets (Novianti, 2013), and Facebook interactions (Halim and Maros, 2014; Henry and Ho, 2016). However, such studies in the Pakistani context have largely focused on code switching (Parveen and Aslam, 2013) and language fusion (Rafi, 2013; Rafi, 2017) in Facebook interactions by Pakistani internet users. The present chapter can fill this gap by examining this phenomenon using a larger dataset and a variety of registers. Hence, the aim of this chapter is to identify instances of code switching in online registers of Pakistani English and to try and ascertain the functions that they perform in the English texts.

The following subsections discuss the relevant aspects of code switching, the data used, the method of analysis, and the results of the analysis.

6.1.1 Types of code switching

Code switching has been viewed as a "dynamic social phenomenon", and as "a language phenomenon that can inform linguistic theory" (Mahootian, 2006, p. 517). Two main types of code switching have been identified in previous literature (p. 512): code switching which occurs at sentence or clause boundaries has been labelled as inter-sentential, while code switching within sentence/ clause or phrase is labelled as intra-sentential code switching. Both of these types view the phenomenon in terms of the structure of phrases, clauses, and sentences. However, the concept of "tag switching" (Mahootian, 2006) incorporates the structural and functional aspects into one. The utterances identified as tag switches are mainly discourse level items, e.g. tag questions like *you know*.

6.1.2 Conversational code switching

Blom and Gumperz (1972) view code switching as a social phenomenon. They find two types of code switching. Situational code switching is related to diglossia, i.e. situational factors, like the place of speech exchange, decide the language choice of the speaker. Metaphoric code switching, on the other hand, "relates to particular kinds of topics or subject matters rather than to change in social situation" (Blom and Gumperz, 1972, p. 425). The same type of code switching is later described as "conversational code switching" in Gumperz (1982, p. 61). The relationship between language and social context is much more complex in this type of code switching. The speakers apparently rely on abstract social norms that they share with the audience "to communicate metaphoric information about how they intend their words to be understood" (p. 61). The metaphoric and stylistic motivations, hence, can be studied in terms of conversational functions behind these code-switched passages.

6.1.3 Functions of conversational code switching

Gumperz (1982, p. 75) analyses three different multilingual situations to identify six conversational functions of code switching, which are briefly described here:

- Quotation: the speakers provide quotations in the form of direct or indirect speech.

- Addressee specification: a particular addressee is specified whom the message is directed to.

- Interjections: the code switching serves as a sentence filler.

- Reiteration: the code is switched to elaborate or emphasise the message.

- Message qualification: such switches consist of "sentence and verb complements or predicates following a copula" (p. 79).

- Personalisation versus objectivization: the specification of a function is relatively complex and difficult in this group. It includes such functions like the speaker's distance from the message, involvement with the message, "whether the statement is personal opinion or knowledge" etc. (p. 80). Grosjean (1982, p. 152), as cited in Mahootian (2006, p. 516), identifies additional discourse functions with regard to this category, e.g. marking group identity, emphasising solidarity etc.

Zentella (1997, p. 92) also finds code switching to be an important conversational strategy in the speech of Puerto Rican children in New York. She identifies three main functions of code switching, namely footing, clarification/ emphasis, and crutch-like code switching. Some of her sub-categorisations of code switching functions overlap with those of Gumperz (1982), e.g. quotations, and the subcategories of clarification like translations, apposition and/or apposition bracket etc. In the case of crutch-like code switching, Zentella (1997, pp. 98-99) notes that it is a way for bilingual speakers to "keep on speaking by depending on a translated

synonym as a stand-in". In her opinion, such instances occur partly because the speakers are not consciously aware and code switching has become 'such an effortless process' for them. Montes-Alcalá (2007) calls such instances of code switching 'sentence fillers'.

6.1.4 Functions of code switching in online registers

A number of studies have applied Gumperz (1982), Zentella (1997), and other similar taxonomies to study the functions and reasons of code switching in online registers. Montes-Alcalá (2007) mainly draws her seven functions of code switching from Zentella. H, and notes that they are purely linguistic. Er categories of lexical items (family terms and kinship terminology), tags (sentence fillers and discourse related items), and free (code switching with multiple functions and stylistic reasons) have been found helpful in the present analysis. San (2009) uses the taxonomy proposed by Li (1996; 2000) to study English and Chinese code switching in blogs. Halim and Maro (2014) examine the reasons of code switching in Malay-English Facebook interactions using a hybrid taxonomy developed from Zentella, Montes-Alcalá, and San. Novianti (2013) and Parveen & Aslam (2013) have studied tweets and Pakistani Facebook interactions respectively using similar taxonomies. Both of these studies find that lexical need or lack of appropriate words in the target language are the main reasons for code switching.

6.1.5 Developing a taxonomy for the present analysis

There is a considerable overlap among the taxonomies or categorisations of functions/ reasons of code switching applied to online registers, as it has been noted in the previous section. The present analysis has mainly used the taxonomies of Gumperz (1982) and Montes-Alcalá (2007), as they have been found the most relevant. Though Gumperz differentiates between conversational code switching and lexical borrowing, Montes-Alcalá and other studies discussed

above consider 'lexical need' as one of the reasons of code switching. In their study on Pakistani English newspapers, Baumgardner, Kennedy, and Shamim (1993) identify 54 different groups of lexical items adopted from Urdu and other regional languages into Pakistani English. Hence, this study identifies 7 categories of code switching, namely Addressee Specification, Emphatic, Free, Lexical, Message Qualification, Quotation, and Tags. A description of the data, the coding process, and the taxonomy along with examples are presented in the following sections.

6.2 Data Collection and Coding

A brief overview of the data from online registers of Pakistani English used for this analysis has been provided in table 6.1. The number of words is slightly higher as compared to the data used in chapter 5 due to additional texts in technology and news blogs. As this data was primarily collected for MD analysis, the stretches of language (tweets/ comments/ Facebook replies) which were predominantly in Urdu or other regional languages were simply removed. Column 5 of the table shows that these omissions primarily occurred in Facebook groups, where participants switched to Urdu or used a fusion of both languages. Hence, the data included in this analysis mainly consists of English with phrase and clause level switching to Urdu or – in rare cases – other regional languages. The texts were manually scanned and tagged (<indig></indig>) for this phenomenon. Additionally, a separate list of frequently occurring single- and multi-word items related to the categories of tags and lexical was developed during the manual scanning process. Later, the concordances were extracted using the tags and the word list. The concordance lines were then organised in an excel sheet and coded.

Table 6.1 The description of the data used for code switching analysis

Registers	Words	Blogs/Users	Switches	Omissions
BlogsSW	290,191	60	241	0
BlogsMW	104,244	8	76	0

Registers	Words	Blogs/Users	Switches	Omissions
BlogsNews	138,115	8	81	0
BlogsNM	42,341	4	141	0
BlogsTech	98,215	4	7	0
Comments	334,703	84	643	307
FBGroups	160,069	11 Groups	175	1,204
FBStatus	67,272	104 Users	191	0
Tweets	58,202	115 Users	256	195
Total	1,293,352	--	1811	0

6.3 Analysis of Code Switching with Examples

The current section presents the results of the code switching analysis along with examples. Table 6.2 exhibits that the highest number of code switches occurs in interactive online registers like comments, Facebook status updates, and tweets. New media blogs have the highest percentage of code switches in blogs (141 code switches with just 42,000 words). Table 6.3 shows the percentages of each type of code switching in different registers. The following subsections elaborate on different code switching categories with examples.

Table 6.2 Frequency of code switching instances in online registers

	Tags	Lexical	Quotation	Free	Emphatic	Addressee Specification	Message Qualif.	Total
BlogsSW	72	86	65	13	5	0	0	241
BlogsMW	12	19	34	6	5	0	0	76
BlogsNews	16	27	32	5	0	0	1	81
BlogsNM	8	76	38	14	2	0	3	141
BlogsTech	0	1	6	0	0	0	0	7
Comments	315	110	55	65	74	17	7	643
FBGroups	134	8	4	18	6	2	3	175
FBStatus	95	39	20	23	13	1	0	191
Tweets	89	22	13	64	30	35	3	256
Total	741	388	267	208	135	55	17	1,811

Table 6.3 Percentage of code switching instances in online registers

	Tags	Lexical	Quotation	Free	Emphatic	Addressee Specification	Message Qualif.
BlogsSW	29.88	35.68	26.97	5.39	2.07	0.00	0.00
BlogsMW	15.79	25.00	44.74	7.89	6.58	0.00	0.00
BlogsNews	19.75	33.33	39.51	6.17	0.00	0.00	1.23
BlogsNM	5.67	53.90	26.95	9.93	1.42	0.00	2.13
BlogsTech	0.00	14.29	85.71	0.00	0.00	0.00	0.00
Comments	48.99	17.11	8.55	10.11	11.51	2.64	1.09
FBGroups	76.57	4.57	2.29	10.29	3.43	1.14	1.71
FBStatus	49.74	20.42	10.47	12.04	6.81	0.52	0.00
Tweets	34.77	8.59	5.08	25.00	11.72	13.67	1.17
Total	40.92	21.42	14.74	11.49	7.45	3.04	0.94

Note: Percentages are based on row totals.

6.3.1 Tags

Tags is the most productive code switching category in the data (41%). The majority of code switches in comments (49%) and Facebook groups (76%) belong to this category. Previous literature has shown that this category includes code switches like linguistic crutches or sentence fillers, interjections, tag questions, and idiomatic phrases that are used without translating to the target language. Examples of idiomatic expressions (1, 2), tag questions (3), and interjections (4, 5) are provided below.

(1) I have long abandoned doing <indig>"behas baraye behas"</indig> (argument for the sake of argument) with people. I only just pray that Allah show us(me and all humanity) the right path and give us... (Comments)

(2) ...but we want only PM's Resign cause this time PM is most eligible to be <indig>"Bali Da Bakra"</indig> (scape goat). (FB Status)

(3) oh no...another seven hours or more jam this week....amazing...<indig>han na..</indig>? (Isn't it?) (Comments)

(4) <indig>Wah ji,</indig> (Great) thanks for the headsup! :) (Comments)

(5) <indig>Wah</indig> (wow),, Ghani gives an opportunity to peep into the mind of a

genius how he... (Tweets)

Another set of tags is related to religious expressions borrowed from Arabic in Pakistani

languages. Mahboob (2009, p. 183) considers such items like *Insha-Allah* (God willing) lexical

borrowings performing pragmatic functions. In the present data it has been found that such

religious expressions mainly function at the discourse level without affecting the overall meaning

of the message being communicated. The following examples show some of the instances of

these religious expressions, including conversations starters and their abbreviations.

(6) <indig>Alhamdulillah</indig> (praise be to God) I have retrieved my Facebook

Account. I am extremely grateful to my ... (Facebook status updates)

(7) <indig>AOA,</indig> (abbreviation of Islamic greetings, peace be upon you) sir

Mohsin i was your student and i also want to inquire that a ... (Facebook groups)

(8) Hello & <indig>Assalam o Alaikum</indig> (Islamic greetings, peace be upon you)

TechJuice, WoW WoW Thank You SoOo much, I can't believe... (Comments)

(9) I myself probably will be travelling from Isb <indig>IA</indig> (abbreviation of God

willing) on 2nd. (Facebook groups)

(10)   We are a nation and will <indig>insha ALLAH</indig> (God willing) remain a

proud, mentally, ethically, morally and religiously glued... (Single-writer Blogs)

(11)   <indig>Ma Sha Allah</indig> (God willing)...... comprehensive details you

provided Xxxxx Xxxxxx thnkx (Facebook groups)

(12)   However, the truth is that Hazrat Ayesha<indig>(RA)</indig> (abbreviation of

Arabic phrase may God be pleased with her) was 17 at the time of her marriage.

(Comments)

(13)   <indig>Salam</indig> (Greetings) to all dear Members! one of my blogs is rated

3rd in google search! (Comments)

(14)   The Prophet <indig>(SAW)</indig> (abbreviation of Arabic phrase peace be

upon him) instead of saying anything to the women turned the face of the

companion...(Facebook status updates)

A third set of tags includes such discourse level items that can be categorised as discourse

particles. Lange (2009) has noted the use of Hindi word *yaar* (dude/ friend) in Indian English

conversations as a discourse particle. The same also occurs in Pakistani online registers, as

examples (19) and (20) show. Additionally, different types of politeness markers and honorifics

have also been found in interactive online registers. As examples (15-18) show, sometimes they

are not translatable.

(15)   Zohair <indig>bhai</indig> (brother) can we update the android version of this

set ? (Comments)

(16)   I would love to invite you Shilpa <indig>jee</indig> (honorific) to my city

Lahore too as here in Lahore, Pakistan there are countless... (New media Blogs)

(17)   And you, Barrister <indig>Saab</indig> (honorific), should understand that best;

after all we have the same thing...(Comments)

(18)   floydian <indig>sahab</indig> (honorific) u shud really write for dawn. <#>

agreed! you should! ... (Comments)

(19)   Xxxxxxxx Xxxxxxx <indig>yar</indig> (dude/ man) it's just a joke. Don't turn it

into a deep reality show or something man. Just a game bro. (Facebook groups)

(20)   … Ayesha said, " <indig>Yaar!</indig> (dude/ man) I don't know why girls

choose to wear only red color on their wedding ... (Single-writer Blogs)

6.3.2 Lexical

This is the second most productive category (21%). The code switches in comments (17%), single writer (35%) and new media blogs (54%) are mainly from this category. It has also been labelled as 'lexical need' in previous literature, as Montes-Alcalá (2007, p. 167) notes. The code switches are generally one-word items that are either nouns or adjectives. Noun phrases, however, also occur in very rare cases. Montes-Alcalá finds kinship and family terminology in her corpus. Baumgardner et al. (1993) also note the occurrence of kinship terminology, items related to eating and drinking, religious concepts, and Urdu words acting as adjectives in Pakistani English newspapers. Examples (21-25) show code switches for family and kinship terminology in the data.

(21)   You owe a lot to your <indig>Ami ji</indig> (mother), please offer five times prayers and pray hard for her all the time... (Single-writer Blogs)

(22)   'When I was in grade 4 I guess, I went to live at my <indig>nani's</indig> (maternal grandmother's) place with my family. (New media Blogs)

(23)   he lived with my <indig>chacha</indig> (paternal uncle younger than father) and would take him along for a long walk... (Single-writer Blogs)

(24)   I'd request you all to please remember my <indig>dada</indig> (paternal grandfather) in your prayers. (Single-writer Blogs)

(25)   But by people you look up to. Like your <indig>mama</indig> (maternal uncle) and <indig>mami</indig> (maternal uncle's wife). You may find the wrong role model there. (Single-writer Blogs)

Examples (26-30) show single word and multi-word items related to edibles and wearables.

(26)   This is possible because one of the popular ingredients

<indig>"Lawng"</indig> (clove) that is set on fire. (New media Blogs)

(27)   Sweet Seneca @ Xxxxxxx Let's be honest, our love for <indig>chai</indig> (tea)

is one that cannot be compared see you making so many (New media Blogs)

(28)   I was also utterly astounded to see you making so many <indig>chapattis</indig>

(breads) single-handedly on a regular basis ... (New media Blogs)

(29)   Hate to break it to you Zara, but that's a <indig>lungi</indig> (skirt). (New media

Blogs)

(30)   BL Building or the OGDCL building) on Jinnah Avenue (Blue Area), be on the

look out for these yummy <indig>garma garam danay</indig> (fresh cooked corn

seeds). (News Blogs)

Various types of religious terms can be found in examples (31-37).

(31)   A nice <indig>nikkah</indig> (marriage vows) ceremony and then you're off to

<indig>Haveli</indig> (manor) for a delicious acne National or brunch. (New media

Blogs)

(32)   Some people even started <indig>Ijtamai dua</indig> (collective prayer) for other

girls to soon receive a <indig>rishta</indig> (match). (New media Blogs)

(33)   Sins that were considered taboo, sins that were considered <indig>gunah e

kabira</indig> (bigger sins), as unthinkable have become such a normal part of our

society. (Facebook status updates)

(34)   Then the <indig>eeman</indig> (faith) is weak =p (Comments)

(35)   ...to seize the opportunity from the projected phenomenal growth of the

<indig>halal</indig> (religiously permitted) markets. (Single-writer Blogs)

(36)   The <indig>haram</indig> (religiously forbidden) money made from

starving/looting poor consumers in Pakistan, was used... (Comments)

(37)   I was fearful of sounding like a <indig>mullah</indig> (Muslim cleric) if I went

on too long. (Single-writer Blogs)

Lastly, examples (38-41) show adjectives of Urdu origin used in English utterances.

(38)   Although out of all of them around the world, <indig>desi</indig> (indigenous)

parents reactions are the funniest. (New media Blogs)

(39)   Instead, we were warned - nay, ordered - to keep all <indig>besharam</indig>

(indecent) period-talk to ourselves. (New media Blogs)

(40)   What happens next: The propaganda machinery of the Army via the

'<indig>bikau</indig> (sold) mainstream media' goes into overdrive, ... (News Blogs)

(41)   ...go anywhere in the world and you'd see men checking out women, in the

<indig>gora</indig> (white) parts of the world... (Comments)

6.3.3 Quotation

This is the third most frequent function of code switching with 14.74% of switches

belonging to this category. Most blogs and comments have higher percentages of switches in this

category. Technology blogs have 6 out of 7 switches in this category. The data shows that, apart

from switching for direct and indirect speech, code switching is used for quoting poetic verses,

film and song names, and other proper nouns. Examples (42-43) show code switching used for

quoting direct speech. Example (44) is a poetic verse quoted by the user. Lastly, code switching

to provide titles of movies (45), songs (46), and books (47) also performs the same function.

(42)   I remember when Rubeena had come to our house and was begging Naheed and saying, <indig>"Bajeeeee, saaday naal ik kumaytee dal lo na,</indig> (Sister, please start a committee with us).[32] (Multi-writer Blogs)

(43)   Someone said: <indig>"Zor zor se forever alone wali feeling aarai hai"</indig> (Having a very strong feeling of being forever alone). (New media Blogs)

(44)   Let me quote it here for you ..its something like.. <indig>"Aj Bohat Din Baad Suni hay Barish Ki Awaz Aj Bohat Din Baad Kisi Munzir Nay Rusta Roka hay...</indig> (after so many days heard the sound of rain today, after so many days a scenery has stopped me) (Single-writer Blogs)

(45)   (Fell in love with Big B in <indig>"God tussi great ho"</indig> (God you are great) :P). (Comments)

(46)   Reminds me of <indig>'Wo kaghaz ki kashti wo barish ka paani'</indig> (that paper boat and the rain water) by jagjit singh... (Comments)

(47)   According to this thought-provoking Islamic book <indig>"Maut Ka Manzar"</indig> (Scene of Death) I once read... (Single-writer Blogs)

6.3.4 Free

Free switching is the fourth most productive reason for code switching with 11.49% of switches. A large percentage of switches in tweets (25%) belongs to this category. Such types of code switches are difficult to categorise in any other functional category. As has been noted by Montes-Alcalá (2007), either there is more than one reason for code switching, or the process is

---

[32] It is a group activity to save money, in practice in various parts of Pakistan, especially in the urban areas. The participants agree to provide a certain amount of money each month, or biweekly. Each month the collected amount goes to one person, whose name is decided using a lottery or by mutual consent. This process continues until each participant receives their collective amount.

purely for stylistic purposes. In examples (48-49) the code switches appear to be for stylistic reasons. Examples (50-52) might involve more than one function of code switching, including lexical need, emphasis, and tag switching. Lastly, examples (53-55) include one-word code switches with no apparent reason deducible.

(48)   She got <indig_n>Nikkahfied</indig_n> (married) over the weekend in a family ceremony that took place in Lahore. (New media Blogs)

(49)   ... Who doesn't want to get married in this beauty: <indig>Haye, qabool hai</indig> (ohh, I do) AF. (New media Blogs)

(50)   When the most "poor" person one has met is the <indig>"kaam wali maasi"</indig> (house maid) in one's home. (Comments)

(51)   After looking at the <indig>"Zinda dilan e Lahore"</indig> (motivated people of Lahore) - I have nothing left to say except <indig>"Lor Lor Ae"</indig> (Lahore is Lahore)... (Single-writer Blogs)

(52)   The story revolves around three <indig>'halat key sataye hue'</indig> (affected by circumstances) individuals who just can't find a solution to their miseries. (News Blogs)

(53)   @XxxxxXxxxx noon league <indig>aur</indig> (more) ashamed... no way brother (Tweets)

(54)   you can stop being shallow now, you're still the class 8 <indig>ka</indig> ('s) nerd. (Single-writer Blogs)

(55)   @XxxxXxxXxxXxxxx wait, we were friends?! <indig>Kab</indig>??? (when?). (Tweets)

6.3.5 Emphatic

The use of code switching for emphasis makes up just over 7.45% of the data. This type of code switching includes literal translations that serve the purpose of repetition, or the language is switched to explain, elaborate, and in some cases strengthen the argument by continuing the message in the other language. Examples (56-57) show literal or approximate translations to re-iterate the same message. Examples (58-60) show that the same messages continue in the other language, but the code-switched stretches are the continuation of the previous communication instead of translation. In the case of (59), the code switching also seems to appeal to the emotions of the reader to emphasise the writer's point of view. The commenter in (61) switches language to Urdu to emphasise and makes a clear request to the addressee for a reply. Lastly, in examples (62-63) the function of code switching seems to be elaboration and/or explanation.

(56)   @xxxxxxxxxxxxx @XxxXxxx Khan feels good by deceiving ppl....<indig>jhoot pa jhoot</indig> (lie over another lie). (Tweets)

(57)   ...and at the end all I can say is: Your thoughts, <indig>aapki soch</indig> (your thought), your brain's working, your creativity, your writing is GREAT! (Comments)

(58)   <indig>Bhuhuat alaa!</indig> (great!) I love it when you kick some ass, kung fu shtyle :) (Comments)

(59)   <indig>Allah ka waasta hai AIK ho jao</indig> (For God's sake please become one). Please unite, come to common terms, tolerate ... (Facebook status updates)

(60)   After all those years of complaining about "NOBODY LOVES ME..NOBODY CARES FOR ME..<indig>HAI ALLAH may mur kion nai jati!!</indig>" (oh God why don't I die?) This feels good, trust me ;p (Single-writer Blogs)

(61)   which is the best between huawei y300 and g510 plz <indig>zohair bhai answer

zaroor dena</indig> (Zohair brother please do answer). (Comments)

(62)   @9Xxxxx Im so bored. <indig>Koi achee movie suggest krdou</indig> (Suggest

me a good movie). :P Have you seen speed? :P (Tweets)

(63)   He did not come for Pakistan but Sharifistan. For them <indig>Pakistan Lahore

say shuru ho k Lahore pay hi khatam ho jata hai</indig> (For them Pakistan starts at

Lahore and ends there as well). (Tweets)

6.3.6 Addressee Specification

Addressee specification is the second least productive category of code switching, which

mainly occurs in blog comments and tweets in this data. The user mainly switches to identify a

particular addressee. Social media websites like Twitter have a function called @ mention, which

allows users to tag their addressees using their name or alias. The same can happen in comments

and other online registers, or simply by writing the name of the addressee followed by switching

to Urdu or another regional language. Examples (64-65) show that the users identify the

addressees by switching to Urdu and then continue the discussion in English.

(64)   @Xxxx_xxxx_xxx <indig>kidr busy ho?</indig> (where are you busy?) Why

you don't come online much. (Tweets)

(65)   @ xxxxxx <indig>Yaar banday ko pata nai hu tu bongi mut mara karoo</indig>

(dude, if one does not know one should not speak out of stupidity) Americans won't

build that fragile and open building in there dreams... (Comments)

6.3.7 Message Qualification

This is the least productive category of code switching with under 1% of overall switches. As the examples (66-68) show, the code switches either function as predicates of the English verb, or function as sentence qualifiers.

(66)   I guess <indig>is pa ppc k under qazaf b lg sakti hy</indig> (an Islamic punishment can also be applied under Pakistan penal code) am I right or am I right (Tweets)

(67)   ... i just wanna say that <indig>poora ka poora pakistan ab khatam honey wala hey bus!!</indig> (all of the Pakistan is just going to end). (Comments)

(68)   I don't know why some people worry so much about <indig>kay shadi kay baad kya hoein ga</indig> (that what would happen after marriage). (Comments)

6.4 Discussion and Conclusion

The analysis of the data using the function of code switching complements the results of chapter 5 in some ways. The frequency of code switching instances in various registers approximately follows the oral versus literate pattern in the first dimension of MD analysis, i.e. interactive registers and less literate blog types (single- and multi-writer blogs, and new media blogs) have more instances of code switching. The under-usage of code switching in Facebook groups can be attributed to the omissions (table 6.1), which were the result of Urdu or a fusion of Urdu and English utterances. The over-usage of code switching in new media blogs seems to be a deliberate strategy to connect with the audience by creating informal and culturally grounded texts. A look at the most important function of code switching, i.e. tags, also confirms that the most frequently code-switched items – e.g. tag questions, sentence fillers, religious expressions like *Insha Allah*, discourse particles like *yaar*, politeness markers like *ji*, and honorifics like

sahab – are mostly related to spoken discourse in Urdu and other local languages. Other reasons for code switching, e.g. lexical need, quotation, free, and emphatic also show the multilingual competence of Pakistani users of English and the appropriation of the English language to communicate local ideas and concepts. It could be concluded that though these communicative exchanges in English are mostly for informational purposes and less personal as compared to similar exchanges in a native variety of English, the use of code switching adds a level of orality and interactivity to these texts, which cannot be captured by a lexico-grammatical analysis of English only.

Chapter 7 Discussion and Conclusion

7.1 Introduction

The aim of last chapter is to synthesise the results of previous chapters – i.e. four, five, and six – and discuss them from two aspects. The first aspect is to explain the results of the linguistic analysis in the light of situational characteristics for each register of Pakistani English and the variety as a whole. The second aspect is to link these results and interpretations to the theory and scholarship of World Englishes. Lastly, this chapter concludes the thesis by highlighting advantages and limitations of the study, and possible future research directions.

7.2 Synthesising and Linking Results

Two types of analyses have been performed in the previous three chapters. The situational description of the registers under study has provided the context of usage. The linguistic analysis in chapter 5 has identified their linguistic and functional characteristics in a comparative way. Additionally, the analysis of code switching instances has shown how multilingual resources have been utilised in online registers. The present section is the third step in register analysis, as Biber and Conrad (2009, p. 6) see it. In other words, this section aims to arrive at functional links between the situational and linguistic analyses conducted previously. Table 7.1 collects very brief summaries of findings related to each register from previous chapters. A summary of answers to research questions 1-9 is partially reiterated below, while the answer to the research question 10 is addressed in the next section with the help of table 7.1:[33]

---

[33] Chapter 1 Introduction lists all research questions and guides to the relevant sections for their answers in section 1.3.

1. The main differences in online versus offline registers of Pakistani English are related to accessibility (ease of access for internet-based registers), general age group (relatively younger internet users), and social background (more diverse groups of internet users). The situational differences between the online registers of both regional varieties are related to topics, communicative purposes, and the characteristics of the participants (the U.S. English being more diverse in all of these aspects).

2. The MD analysis resulted in five dimensions of variation: Oral versus Literate Discourse, Conversational Style, Abstract Evaluative versus Non-Abstract Information, Reporting Style, and Narrative Focus. The first three dimensions are moderately strong, while the last two dimensions are weak.

3. Dimensions 1, 3, and 4 are somewhat related to the distinction between online versus offline registers. The comparison of blogs and similar registers (opinion columns and news reports) shows that blogs are more oral, less abstract evaluative, and less reporting. The comparison of online interactive registers and spoken conversations shows that the former registers are generally less oral, less conversational, less reporting, and non-narrative as compared to the latter.

4. Regional variation is mainly visible on dimensions 1, 2, and 3, where Pakistani registers are generally less oral, less conversational, and more abstract evaluative as compared to their U.S. counterparts.

5. There are no clear-cut differences based on gender in the U.S. data. However, the findings from dimensions 1 and 3 show that male Pakistani bloggers of individual and news blog categories produce slightly more informational and abstract evaluative texts in comparison to their female counterparts.

6. Only two dimensions (i.e. 1 and 4) show identifiable chronological trends. The findings of dimension 1 indicate that the informal blogging style has been decreasing over the years in both regional varieties. The findings of dimension 4 show an increase in reporting style in the U.S. English technology blog.

7. The CA identified four text types in online registers of both regional varieties, namely: Unmarked Oral, Literate Nominal, Oral Non-abstract, and Unmarked Abstract Evaluative. The most prominent differences are in Facebook groups (Oral Non-abstract being the most frequent text type in the U.S. data) and technology blogs (Literate Nominal being the most frequent text type in the Pakistani data). The results confirm the findings of the MD analysis, i.e. the texts in Pakistani English registers are less oral, more literate, and abstract evaluative.

8. The results of the CDA show that the general communicative purpose differentiating both regional varieties can be labelled as 'Oral/ Informal versus Literate/ Formal Language', where the Pakistani data is inclined towards the literate/ formal side.

9. Seven functional types of code switching have been identified, namely: Tags, Lexical, Quotation, Free, Emphatic, Addressee Specification, and Message Qualification. Out of these seven categories, the first three make up about 70% of the code switching occurrences.

Table 7.1 A brief summary of results

Register	Characteristics (Pakistani English)	Characteristics (U.S. Counterpart)
Individual Blogs	**Situational:** Writers = voluntary urban (lower) middle and upper class; Audience: limited urban class; Topics: self and surroundings, literature, social issues, religion, beauty etc.; Communicative purpose: narrate, comment, opine, how-to, creative writing etc.; shorter life-span **Linguistic:** Narrative (.) **(Dis)similar registers:** N = New-media Blogs; F = News, Columns **Main text types:** Unmarked Abstract Evaluative (37%), Unmarked Oral (32%) **Main CS Functions:** Lexical (33%), Quotation (31%), Tags (26%) **Other:** Male bloggers write comparatively less oral texts	**Situational:** Writers and audience = diverse; Topics: self, law, politics, beauty, music etc.; Communicative purpose: comment, opine, explain/ how-to etc.; longer life-span **Linguistic:** Non-abstract Info. (-), Non-Reporting Style (-) **(Dis)similar registers:** N = Tech Blogs; F = News, Columns **Main text types:** Unmarked Abstract Evaluative (45%), Unmarked Oral (35%) **Other:** --
News Blogs	**Situational:** Writers = voluntary and professional (urban lower) middle class and other strata; Audience: general audience of English newspapers; Topics: social issues, politics, literature etc.; Communicative purpose: comment, opine, creative writing, (tech/ film) review etc.; blog sections are within newspaper websites **Linguistic:** Literate (--), Abstract Evaluative Info. (+.), Narrative (+) **(Dis)similar registers:** N = NM Blogs; F = News **Main text types:** Literate Nominal (45%), Unmarked Abstract Evaluative (36%) **Main CS Functions:** Lexical (33%), Quotation (39%) **Other:** Male bloggers write comparatively more literate and abstract evaluative informational texts	**Situational:** Writers = voluntary and professional; Audience: diverse; Topics: religion, word play, economics, law, sports etc.; Communicative purpose: narrate, comment, opine, report, explain, review etc.; mostly separate subdomains and topical blogs **Linguistic:** Literate (--), Reporting Style** (+.), Narrative (+) **(Dis)similar registers:** N = Columns; F = NM Blogs **Main text types:** Literate Nominal (49%), Unmarked Abstract Evaluative (30%) **Other:** --
New-media Blogs	**Situational:** Writers = social media/ tech savvy, professional, urban middle-class; Audience: urban & other young audience reachable through social	**Situational:** Writers = professional; Audience: diverse and international; Topics: viral stories on social media, celebrity gossip, technology, social

Register	Characteristics (Pakistani English)	Characteristics (U.S. Counterpart)
	media; Topics: viral stories on social media, celebrity gossip, social issues, religion etc.; Communicative purpose: informal comment, report, entertain etc.; interaction with audience through social media; ads **Linguistic:** Literate (-), Narrative (+) **(Dis)similar registers:** N = Ind. Blogs; F = News, Columns **Main text types:** Literate Nominal (38%), Unmarked Oral (31%) **Main CS Functions:** Lexical (54%), Quotation (27%)	issues etc.; Communicative purpose: informal comment, report, entertain etc.; interaction with audience through social media; ads **Linguistic:** Conversational Style (++), Narrative (++) **(Dis)similar registers:** N = Ind. Blogs, Tech Blogs; F = News, Columns **Main text types:** Unmarked Oral (35%), Unmarked Abstract Evaluative (28%)
Tech. Blogs	**Situational:** Writers = professional, tech (/social media) savvy, journalist, (urban lower) middle class; Audience: young and old audience connected to the internet, tech professionals, bloggers etc.; Topics: technology, science, gadgets, internet, tech companies; Communicative purpose: report, review/ describe, guide, how-to, explain; updates through social media; ads **Linguistic:** Literate*** (---.), Non-Conversational Style*** (---), Abstract Evaluative Info. (+), Non-narrative (---.) **(Dis)similar registers:** N = Columns, News Blogs; F = Ind. Blogs **Main text types:** Literate Nominal (65%), Unmarked Abstract Evaluative (22%) **Main CS Functions:** Quotation (86%), total = 7 **Other:** Chronological increase in literate style	**Situational:** Writers = professional, tech (/social media) savvy, journalist; Audience: diverse and international; Topics: technology, science, gadgets, internet, tech companies; Communicative purpose: report, review/ describe, guide, how-to, explain; updates through social media; ads **Linguistic:** Conversational Style (+.), Non-narrative (---) **(Dis)similar registers:** N = Ind. Blogs, NM Blogs; F = News, Columns **Main text types:** Unmarked Abstract Evaluative (37%), Unmarked Oral (36%) **Other:** Chronological increase in literate style (dimension 1) and increase in reporting style (dimension 4)
Comments	**Situational:** Participants: young people from big and small cities, students, entrepreneurs, tech. enthusiasts, bloggers, English newspaper readers etc.; Topics: personal, social issues, religion, education, tech, cooking etc.; Communicative purpose: interact, respond, debate, ask (for help/ tips) **Linguistic:** Oral (+.), Non-narrative (-)	**Situational:** Participants: diverse possibly international; Topics: religion, politics, law, cooking, tech etc.; Communicative purpose: interact, respond, debate etc. **Linguistic:** Oral (++), Conversational Style* (++)

Register	Characteristics (Pakistani English)	Characteristics (U.S. Counterpart)
	(Dis)similar registers: N = Talk Shows; F = Tweets, Face2Face **Main text types:** Unmarked Oral (51%), Unmarked Abstract Evaluative (27%) **Main CS Functions:** Tags (49%), Lexical (17%); highest no. of CS instances	**(Dis)similar registers:** N = Interview, Talk Shows; F = Face2Face **Main text types:** Unmarked Oral (52%), Oral Non-abstract (27%)
Facebook Groups	**Situational:** Participants: young people from big and small cities, PhD scholars, students, entrepreneurs, tech enthusiasts, pet owners, English teachers etc.; Topics: foreign study, admissions, job ads, general help, pets, technology, start-ups; Communicative purpose: ask (for help), discuss, inform, advertise (jobs etc.); Topic specific groups with sometimes tens of thousands of participants **Linguistic:** Oral** (++.), Non-narrative*** (-) **(Dis)similar registers:** N = Face2Face, Talk Shows; F = FBStatus **Main text types:** Unmarked Oral (54%), Oral Non-abstract (35%) **Main CS Functions:** Tags (76%)	**Situational:** Participants: diverse localised communities with similar interests; Topics: study, sports, politics, community, food, pets; Communicative purpose: ask (for help), discuss, inform; localised public groups with a few hundred participants **Linguistic:** Oral (++++), Conversational Style*** (++), Non-abstract Info.*** (---.), Non-Reporting Style* (-.) **(Dis)similar registers:** N = FBStatus; F = Face2Face **Main text types:** Oral Non-abstract (73%), Unmarked Oral (27%)
Facebook Status updates	**Situational:** Participants: generally young entrepreneurs, tech enthusiasts, bloggers, English teachers, famous personalities; Topics: job ads, business, advice, religion, social issues; Communicative purpose: advertise, advise, interact **Linguistic:** Non-Reporting Style (-.) **(Dis)similar registers:** N = Comments; F = Face2Face **Main text types:** Unmarked Oral (38%), Unmarked Abstract Evaluative (25%) **Main CS Functions:** Tags (49%), Lexical (20%)	**Situational:** Participants: young and old; Topics: social issues, politics, self etc.; Communicative purpose: interact, express etc. **Linguistic:** Oral*** (+++.), Conversational Style** (+), Non-abstract Info.*** (---.), Non-Reporting Style (-.) **(Dis)similar registers:** N = FBGroups; F = Face2Face **Main text types:** Oral Non-abstract (52%), Unmarked Oral (37%)

Register	Characteristics (Pakistani English)	Characteristics (U.S. Counterpart)
Tweets	**Situational:** Participants: generally young students, political/social activists, bloggers, journalists, expats; Topics: self, politics, social issues, religion, education etc.; Communicative purpose: advertise, react, inform **Linguistic:** Literate (-.), Non-Conversational Style (------), Non-Reporting Style (-.), Non-narrative (-.) **(Dis)similar registers:** N = FBGroups, FBStatus; F = Interview **Main text types:** Unmarked Abstract Evaluative (34%), Oral Non-abstract (30%) **Main CS Functions:** Tags (34%), Free (25%); second highest no. of CS instances	**Situational:** Participants: diverse; Topics: self, politics, social issues etc.; Communicative purpose: inform, express, interact **Linguistic:** Oral (++.), Non-Conversational Style (---), Non-abstract Info.*** (---), Non-Reporting Style (--) **(Dis)similar registers:** N = FBStatus; F = Face2Face **Main text types:** Oral Non-abstract (47%), Unmarked Oral (33%)

Note: (.) = group mean minimum ±0.125 standard deviations away from the overall mean; +/- = group mean ±0.25 standard deviations away from the overall mean; ++/-- = group mean ±0.50 standard deviations away from the overall mean; ++++/---- = group mean ±1.0 standard deviations away from the overall mean;[34] Significant differences between regional varieties: * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$; N = Nearest, F = Farthest; CS: Code switching

7.2.1 Synthesising results: individual registers

The following subsections generally respond to research question 10 for individual registers.[35]

7.2.1.1 Individual blogs.

The situational analysis shows that individual blogs got popularity at the end of the 2000s among a small mostly urban community of youngsters that were born in the last couple of decades of the 20th century (i.e. millennials). The linguistic analysis reveals that there are no

---

[34] Adapted from Egbert (2014, p. 144)
[35] How can the linguistic variation be explained in terms of communicative functions, situational differences, and usage?

distinct characteristics on any dimension associated with these blogs (apart from them being slightly narrative). They have more in group variation on dimension 1 'Oral versus Literate Discourse' (cf. subsection 5.1.6.1). Additionally, two main text types can be attributed to thematic and diary type blogs identified in previous research. The gender differences show that male writers incline towards thematic blog writing (i.e. more literate), while female writers mostly write diary type blogs (i.e. more oral). Additionally, a shorter lifespan and the presence of creative writing (poetry and short stories etc.) leads to conclude that these blogs functioned as personal publishing platforms for these bloggers in English. The blog community thrived for a few years, but eventually most blogs were abandoned after the year 2012. Two factors appear to be the most plausible in this regard (though there can be many others): the popularity of social media websites like Facebook and Twitter in the country after 2012 and the introduction of blog sections on newspaper and news channel websites around the same time.

7.2.1.2 News blogs.

News blogs appear to be the successors of thematic individual blogs. There are a number of situational and linguistic findings that support this belief. The situational characteristics of news blogs are very similar to individual blogs, but at a larger scale – more diverse kinds of writers, wider audience, and topics and communicative purposes similar to those of thematic blogs. Additionally, individual blog writers already sent their posts to newspaper blog sections. The prominent linguistic dimensions and main text types for news blogs also show that the discourse produced in these blogs focuses on commenting about surroundings instead of self. News blogs lie somewhere between opinion columns (i.e. less literate and abstract evaluative) and individual blogs (i.e. more literate). In comparison, U.S. news blogs are much more similar to columns and news reports (in terms of reporting style and narration). News blogs sections on

Pakistani newspapers are continuously updated, while some U.S. newspapers like *USA Today* might be closing their blog sections, as noted in subsection 4.3.2.

To sum up, Pakistani news blogs appear to be less planned and researched but still edited essays and pieces of creative writing that function as a venue for young and less experienced writers to show their writing skills to a wider audience.

7.2.1.3   New media blogs.

Situationally new media blogs differ from the previous two types of blogs in at least three ways: i.e. professional writers, focus on viral stories, and communicative purposes like informal comment. They are the youngest of four blogs sub-types and mostly try to appeal to a younger audience, including teenage internet and social media users from the country. Though Pakistani new media blogs try to copy the style of their U.S. counterparts, they are comparatively less oral and conversational in terms of linguistic characteristics. Although code switching is present in other blog types as well, it seems that here its overall function is to make blog posts more informal and culturally relevant to the audience. The presence of ads exhibits that the main purpose of these blogs is to generate money by producing interesting content to attract readers. For various reasons English tabloid magazines have remained virtually non-existent in Pakistan. It could be argued that these blogs are digital equivalents of tabloids or celebrity gossip magazines in digital form.

7.2.1.4 Technology blogs.

Technology blogs are similar to new media blogs, because they are written professionally, display ads, and have their own websites. Pakistani and U.S. technology blogs are similar in terms of situational characteristics like topics and communicative purposes. However, they are different in terms of blog writers and audience, which are local in the case of Pakistan. In terms

of linguistic characteristics, Pakistani technology blogs follow a literate and informational style that is nearer to opinion columns and news reports instead of a blog-like informal style followed by their U.S. counterparts. The chronological analysis of the Pakistani technology blog (ProPakistani) reveals that it has developed from a thematic blog on technology to a news media website-like outlet that produces content on technology and related topics. To conclude, Pakistani technology blogs can be considered digital magazines and/or newspapers that focus on technology related news reports, articles, and other content in English.

7.2.1.5 Comments.

Comments show mixed linguistic and situational characteristics. Pakistani comments are non-narrative, which shows that the commenters are concerned with the issues under discussion in the post content. It is also related to the communicative purpose of asking for help, e.g. in technology news blog comments, which happens in the here and now. In general, comments follow the regional trend of blogs in terms of situational characteristics (e.g. a younger audience and participants) and linguistic characteristics (e.g. less oral and more abstract evaluative informational discourse). Previous research has shown that comments are opinion oriented. The linguistic markers of expressing opinion appear to be scattered in the first three dimensions (Oral versus Literate Discourse, Conversational Style, and Abstract Evaluative versus Non-abstract Information) of MD analysis, which include elements of stance marking and evaluation. Pakistani comments include another layer of orality with a high amount of code switching instances. One of the most frequent functions is their usage as tags, which are mostly conversational or discourse level elements like conversation starters, greetings, tag questions, and discourse particles (cf. subsection 6.3.1). This phenomenon gives a local flavour to Pakistani blog comments in English. To conclude, owing to their similarity to talk shows (and interviews

in case of U.S. English), comments can be considered as a discussion-oriented register like broadcast discussions, but on the internet.

7.2.1.6 Facebook groups.

As compared to comments, Facebook groups show clear links between situational and linguistic characteristics. Moreover, the regional differences in both types of characteristics are also very visible (cf. prominent regional differences in table 7.1). Like in the previously discussed registers, the participants of Pakistani Facebook groups are younger educated professionals and students, and thus less diverse (in terms of age groups) as compared to their U.S. counterparts. The other situational characteristic that distinguishes both regional varieties is the topics discussed, e.g. higher education and business (Pakistani English) versus politics and community issues (U.S. English). The posts asking for help in relation to business and education are related to present and future, which makes Pakistani English Facebook groups non-narrative. Additionally, though they are oral, they at the same time contain abstract evaluative information in some cases due to posts like job ads. The similarity between face-to-face conversations and Facebook groups shows that both registers deal with similar content and communicative purposes, which is not the case with their U.S. English equivalents. Lastly, like in comments an additional layer of orality is introduced by code switching, which in the majority belongs to the category of tags (cf. subsection 6.3.1). It can be concluded that the use of English in most interactive registers of Pakistani English (i.e. Facebook groups and face-to-face conversations) is restricted to certain non-personal domains.

7.2.1.7 Facebook status updates.

Facebook status updates also have clear differences in terms of situational characteristics that result in linguistic characteristics and sometimes significant regional differences, as table 7.1

shows. The communicative purposes and topics of status updates are restricted to business and social issues etc. The exclusion of communicative purposes like expression of the self or feelings (as is present in the case of U.S. status updates) is probably due to the multilingual backgrounds of Pakistani users. Local languages like Urdu are most probable candidates to express personal feelings instead of English, which is limited to certain domains only. Based on these characteristics, it appears that Facebook status updates are more likely to be used for professional purposes in Pakistani English as compared to more personal purposes in U.S. English.

7.2.1.8 Tweets.

Tweets are linguistically different from other online registers. As it has been noted in the review of previous research on tweets in chapter 2, the restriction on message length is one of the main reasons of these differences. The situational differences between Pakistani and U.S. tweets are similar to those of Facebook status updates and Facebook groups. Resultantly, the linguistic characteristics also show a similar trend, i.e. in general tweets are literate, non-conversational (with long lists of @ mentions and hashtags), and non-narrative in Pakistani English. It seems that tweets are also mostly restricted to professional domains in case of Pakistani users of English. Lastly, a subset of tweets by students and younger participants includes code switching into local languages like Urdu and Pashto, which makes them somewhat oral and interactive.

7.2.2 Synthesising results: the variety

To synthesise the results for the variety as a whole, they need to be looked at from three different aspects:

Table 7.2 Summary of situational and linguistic differences

Aspect	Situational Differences	Linguistic Differences
Online versus offline registers in Pakistani English	Online registers are more accessible); online text producers are relatively younger; and relatively less known (as compared to text producers of opinion columns, talk shows, and interviews)	Blogs: Online discourse is less formal, i.e. less literate and abstract informational in varying degrees depending on the blog type. Interactive registers: Similarity in terms of orality, conversational style, and abstract evaluative informational discourse.
Regional comparison of online versus offline registers	PK: Participants of interactive registers are generally peers, friends, professional colleagues, or simply strangers. U.S.: Participants of face-to-face conversations are likely to be close relatives, while generally not the case for online interactive registers	Blogs: Technology blogs are very similar to opinion columns and news reports, the same case with news blogs in U.S. English Interactive registers: U.S. face-to-face conversations are highly different from interactive online registers, while the opposite is true in the case of Pakistani English.
Regional comparison of online registers	PK: Generally younger, urban class text producers; non-personal and professional domains of language use U.S.: Text producers of diverse backgrounds and age groups; Possibility of the presence of personal as well as professional domains	PK: Texts are generally less oral, less conversational, and contain abstract evaluative informational discourse U.S.: Texts are oral, conversational, and contain non-abstract information, more likely to be narrative.

The above summary leads to the following conclusions regarding Pakistani English online registers. The situations where English can be used as well as the user base have expanded with the advent of online registers. However, the domains where English can be used remain (semi-)public, professional, and non-personal ones that are similar to existing offline registers. The linguistic analysis has revealed that more oral and informal language is being produced (e.g. blogs versus opinion columns). At the same time, the comparison with U.S. English shows that a

formal/ literate type of discourse is the general norm. It can further be concluded that Pakistani English – in comparison to U.S. English – remains restricted and limited in terms of situational as well as linguistic/ functional characteristics. Lastly, as it has also been noted in previous subsections, the multilingual nature of the Pakistani context is an important factor in the above-mentioned usage patterns of English.

7.3 Connecting with World Englishes Theory and Scholarship

This section relates the findings of this study with the most relevant models of World Englishes, discussing them in an approximately chronological order. Afterwards, the nature of communicative interactions on the internet and the issue of register is discussed by relating it to more recent discussions on theorisation in World Englishes.

7.3.1 Models of World Englishes and present findings

Kachru's (1985) Three Circle model is the oldest and most relevant for the variety in a historical context. Kachru's observations regarding outer circle varieties can also be related to internet registers of Pakistani English, particularly in two ways. Firstly, the domain of education appears in the form of discussions and information exchange regarding study abroad in Pakistani Facebook groups. Secondly, the decades old 'nativised literary tradition' of producing short stories, essays, and novels in English continues in the era of the internet as well. Specifically, individual blogs and later on news blogs are also used by Pakistani users of English to produce literature in the form of poetry and short stories. Moreover, voluntary forms of blog writing (individual and news blogs) can also be considered a venue for Pakistani writers to showcase their (creative) writing skills. The English language on the internet appears to be firmly rooted in historical and traditional domains of usage (e.g. as noted by Haque, 1983). Hence, it could be argued that the status of Pakistani English as an outer circle variety remains unchanged.

Schneider's (2007) Dynamic Model is the natural choice to look at the developmental profile of the variety. Though Schneider (2007) does not address Pakistani English as a case study for the model, Indian English that is historically a sister variety of Pakistani English does get discussed. He notes that Indian English has been in the third phase, i.e. 'nativisation', since the start of the 20th century. If the same stage is supposed for Pakistani English, the results of chapter 6 – i.e. code switching instances related to discourse level items – could be presented as possible evidence of nativisation. Localised address forms, greetings, tag questions, and religious expressions like *InshaAllah* (God willing) in online English interactions show that local conventions have been created and are being used. They also indicate a Pakistani and Islamic identity adopted by Pakistani users of English. However, the model is only partially applicable to these findings for certain reasons. For example, the settler strand, one of the basic assumptions of the model, is not present for Pakistani English in the 21st century, especially for online registers. Moreover, the issue of globalisation is not addressed by this model.

The concept of 'extra- and intra-territorial forces' form the EIF model (Buschfeld et al., 2018), however, appears to be suitable to explain the reasons and (de)motivations to communicate in English on the internet. Buschfeld et al.'s (2018) observations regarding globalisation in both intra- and extra-territorial aspects are also applicable to this context. Globalisation and globalised connectivity are extra-territorial forces, while embracing globalisation and the provision of accessibility to the internet by the government are intra-territorial forces. Pakistani expats working abroad tend to prefer English while talking to people back home on social media. The influence of a foreign environment on language choice can be considered as an extra-territorial force that is applicable to Pakistani online registers, e.g.

Facebook groups, personal and news blogs, blog comments etc. The education policy, specifically the widening role of English in school curricula, is also an intra-territorial force.

The idea of 'territorial-ness' in the context of the internet soon becomes problematic, because it is not always that straightforward to identify the 'territory' of a factor. Buschfeld et al. (2018, p. 38) also note that in the context of the internet the concept of 'extra- and intra-territorial forces' can be reformulated as 'external' and 'internal' factors or forces. For example, economics (producing content to show ads in English and earn money in the case of technology and new media blogs) is one of the biggest motivations for Pakistani internet users to use English online (e.g. technology and new media blogs). The content can be consumed by local users (which is the case most of the times), but it can also reach the international audience. In both cases, it will generate advertisement money for the corresponding bloggers. It could be considered an intra-territorial force or an extra-territorial force depending on the place where the content is being consumed or the place of the advertisement agency that is paying for ads. However, this can be very easily considered a factor or force that is external to the community itself and that motivates the users to engage in English communication.

The accessibility to the internet (and online registers) that helps users gain access to and engage in text production in voluntary blog types (i.e. individual and news blogs) is an external factor. The corresponding internal factor is the motivation of the users to become a part of the community by writing blog posts in English that bring no monetary reward and their wish to express their creative writing skills.

Meer and Deuber's (forthcoming) concept of 'translocal forces' is also applicable to the various motivations of using English described above, especially those that cannot be associated with a territory. Hence, the tendency of expats residing abroad to use English on social media

while addressing people back home, the motivation of professional bloggers to generate English content to earn advertisement money, and the motivation to use English by the students aspiring to go abroad can be labelled as translocal forces, which motivate the use of English and are not associable to specific borders.

7.3.2 Localised interactions in a global space

The next relevant issue is to look at the nature of online interactions, i.e. the extent to which they are 'globalised'. Generally speaking, discussions related to online English communication tend to assume that these communicative situations are international or have international participants (Seidlhofer, 2009; Buschfeld et al., 2018, p. 38; Deshors and Gilquin, 2018, p. 283). The communicative interactions analysed in this study, however, are nearer to those analysed by Sharma (2012) and van Rooy and Kruger (2018), because they are localised interactions in the global space of the internet. It is true that there are no clear-cut boundaries on the internet as it would be the case for a nation-state. Resultantly, there could be participants in these communicative situations that do not come from Pakistan (commenters on news blogs, or participants in Facebook group discussions). Moreover, Pakistani expats writing blogs, commenting on blog posts, or interacting with other Pakistanis on social media point towards the globalised nature of these communicative interactions. Despite the presence of such cases, these communicative situations are like tiny islands in the seemingly endless space of the internet, where the vast majority of participants share national, ethnic, linguistic, and religious affiliations.

The above entails that there is a complex relationship between identity construction, linguistic repertoire, and situational factors (topics like technology, business, education; monetary considerations in the case of professional blogging; the desire to express oneself in English in voluntary blog writing; the sheer presence of English on the internet; expats residing

abroad etc.). Pakistani users interacting in English on the internet become a part of the global whole that communicates in English. At the same time, they colour their communication in a very local way by using linguistic resources (i.e. code switching) from local languages (similarly as van Rooy and Kruger, 2018 observe for South African users). English is one of many choices in their linguistic repertoire, but the choice of English over other local languages (including the local lingua franca, Urdu) is determined by a number of situational factors, a few of which have been listed above. At the same time, it is the situational factors that also decide where English will not be used, e.g. in more personal domains. In this way, the situational factors help establish localised identities in a global space by determining the choice and kinds of uses associable to English.

7.3.3 The case for communicative situation and MD framework

As it has been noted in section 2.1, the need to consider 'communicative event' or 'communicative situation' in the process of theorisation in World Englishes has recently been emphasised (e.g. Deshors and Gilquin, 2018). This study has interpreted and implemented this idea in the form of register analysis, as Biber and colleagues see it (e.g. Biber and Conrad, 2009). The results of MD analysis show that register or communicative situation can predict linguistic variation much more strongly as compared to region (cf. table 5.21). Additionally, the statistical interaction between register and regional variation has largely remained significant, which has shown that the relationship between the two types of variations is a complex one. These conclusions regarding the importance of register and situational variation in the study of World Englishes reiterate the observations made in previous MD studies like Bohmann (2017; forthcoming).

The MD analysis framework is certainly not the only way to study communicative situation, but it does provide a systematic way to analyse situational factors and quantify linguistic variation occurring due to these factors. In this process, MD analysis complements traditional variationist studies that reduce register to one of many extra-linguistic factors and consider it stylistic variation (Szmrecsanyi, 2019). To exemplify from the present study, the results of MD analysis do not offer any quantitative or other type of evidence of linguistic or structural innovations happening in Pakistani English on the internet. However, they do indicate the communicative functions attached to the majority of lexico-grammatical and semantic features found in Pakistani English as discussed in subsection 7.2.2. Hence, these findings highlight the importance of the study of communicative situation – especially using a quantitative method like MD analysis – for a better theorisation of variation in the English language.

7.3.4 The nature of digital Englishes

In the light of the above discussion, it is not implausible to consider digital varieties of English as extensions of so-called non-digital varieties of English. The internet does not have boundaries like nation-states, so digital Englishes should be described in terms of particular communicative situations (Deshors and Gilquin, 2018) or communities of practice (Seidlhofer, 2009). However, it should be noted that there exist localised communities of practice (e.g. as described in the present study; van Rooy and Kruger, 2018) as well as global or international communities (e.g. as discussed in Buschfeld et al.,2018). This is because the internet is a network of localised and globalised (sub-)sub-networks, so everyone connected to the internet is not connected to everyone else. The communicative interactions on the internet, hence, can reflect the sociolinguistic realities (e.g. communicative purposes, situational parameters, and linguistic

structures) of the good old non-digital Englishes. This implicates that at the practical level data corresponding to nation-state-based varieties of English can be collected from the internet, e.g. the GloWbE corpus (Davies and Fuchs, 2015). At the theoretical level, it means that traditional models of English can still be applicable to digital Englishes, e.g. Kachru's Three Circle model as discussed in subsection 7.3.1. Lastly, it also entails that any theorisation and model making effort for digital Englishes will be incomplete without taking into account both aspects, i.e. the globalised as well as the localised nature of these varieties of English.

7.4 Advantages of the Study

The study has combined the analysis of linguistic structures with that of situational characteristics to inform about various register categories of Pakistani English online. In this process, the study has demonstrated how Pakistani English functions differently from U.S. English in similar communicative situations. By focusing on register as a central issue, it has also exhibited the complex and multidimensional nature of language variation that would otherwise remain unnoticed.

As discussed in section 7.3, it has been shown that the communicative interactions on the internet can be localised (among people of similar backgrounds) as well as globalised (among people of different backgrounds) in nature. Both types of interactions need to be considered to arrive at a better understanding of English on the internet.

In terms of method, the study has joined a very small group of previous studies to demonstrate the suitability of the MD analysis framework – which combines linguistic and situational analysis in a systematic way – to conduct research in the paradigm of World Englishes.

7.5 Limitations of the Study

The following limitations should be taken into account while interpreting the results of this study:

- The strength of the model used for the MD analysis (i.e. variance explained) exhibits that this analysis only partially explains the linguistic phenomena involved in the data under study. The selection of features, for example, could be expanded to increase the reliability of the linguistic model. Similarly, the limitation of cluster analysis noted in subsection 5.2.3 should also be taken into consideration.

- The type of analysis performed here using the MD framework follows the tradition of text linguistics, as explained in chapter 2. The results and interpretations of this type of analysis are different from corpus-based variationist studies that are generally conducted in the paradigm of World Englishes. As a result, this study cannot provide insights, for example, into linguistic structures and innovations prevalent in Pakistani English, because those structures normally do not have a functional significance and are much less frequent.

- The texts produced by Pakistani users of English with their multilingual backgrounds certainly did not consist of pure English communication all the time. As it has been noted in previous chapters, only English texts or predominantly English utterances were selected. If the texts and utterances consisting of a fusion of English and local languages were to remain in the data, the analysis of code switching instances could be much more rich and insightful.

7.6 Recommendations for Future Research

There are many ways to expand and complement this study with other types of analyses.

The range of offline registers included in this study could be expanded to get deeper insights regarding the status of online registers in the overall spectrum of registers of Pakistani English. Variationist analyses could be conducted on similar data to inform about those aspects of language variation that could not be covered by register analysis alone.

The study of internet registers from the South Asian region as a whole could be another potentially fruitful future research direction. Until now, Indian English is the most studied variety from this region (e.g. Mukherjee and Hoffman, 2006). Not many studies include Pakistani and Sri Lankan Englishes largely due to the partial availability or simply unavailability of corpora (Hussain, 2016; Mukherjee, Schilk, and Bernaisch, 2010). The same is the case for Bangladeshi English. The studies that have included most of the South Asian varieties of English (Koch and Bernaisch, 2013; Bernaisch, Gries, and Mukherjee, 2014) utilise data collected from newspaper websites (Mukherjee and Schilk, 2012). Alternatively, the GloWbE corpus (Davies and Fuchs, 2015), which consists of blogs and newspaper related registers, is also used when varieties other than Indian English are studied from the South Asian region. A regional level study with a range of internet-based registers – similar to the present one – can, thus, shed light on how English is being used on the internet by South Asian users of English.

References

Abbas, S. (1993). The power of English in Pakistan. *World Englishes*, *12*(2), 147–156. https://doi.org/10.1111/j.1467-971X.1993.tb00017.x

Abidi, S. (2017, November 3). Student mobility trends from Pakistan. *The News*. Retrieved from https://www.thenews.com.pk/print/241585-Student-mobility-trends-from-Pakistan

Ahmad, M., Mahmood, M. A., Mahmood, M. I., & Siddique, A. R. (2019). Use of modal verbs as stance markers in Pakistani English newspaper editorials. *Online Journal of Communication and Media Technologies*, *9*(1). https://doi.org/10.29333/ojcmt/5722

Ahmad, S. (2016). *Linguistic variation across Pakistani print media: A multidimensional analysis* (PhD Thesis). Govt. College University Faisalabad, Faisalabad.

Ahmad, S., & Ali, S. (2014). Impact of Urduised English on Pakistani English Fiction. *Journal of Research (Humanities)*, 61–75.

Ahmad, S., & Mahmood, M. A. (2015a). Comparing explicit features of Pakistani press reportage with British press reportage: A multi-dimensional analysis. *NUML Journal of Critical Inquiry*, *13*(II), 1–31.

Ahmad, S., & Mahmood, M. A. (2015b). Linguistic variation among sub-categories of press reportage in Pakistani print media: A multidimensional analysis. *Journal of Social Sciences*, *6*(2), 1–16.

Ali, S. (2013). Modal verbs in Pakistani English: Case study of *Can* and *Could* a corpus-based comparative analysis. *International Journal of African and Asian Studies*, *1*, 42–46.

Ali, S., & Ahmad, S. (2016). Discourse style variation among the leading novelists of Pakistani fiction in English: A multidimensional analysis. *ELF Annual Research Journal*, *18*, 227–246.

Alvi, U. (2013). *Linguistic analysis of Pakistani newspaper editorials: A multidimensional approach* (PhD Thesis). Govt. College University Faisalabad, Faisalabad.

Alvi, U., Mehmood, M. A., & Rasool, S. (2016). A multidimensional analysis of Pakistani press editorials. *The Dialogue*, *XI*(3), 270–284.

Anwar, B. (2009). Urdu-English code switching: The use of Urdu phrases and clauses in Pakistani English (a non-native variety). *International Journal of Language Studies*, *3*(4), 409–424.

Archer, D., Wilson, A., & Rayson, P. (2002). Introduction to the USAS category system. *UCREL Working Paper, Lancaster University*. Retrieved from http://ucrel.lancs.ac.uk/usas/usas_guide.pdf

Asghar, S. A., Mahmood, M. A., & Asghar, Z. M. (2018). A multidimensional analysis of Pakistani legal English. *International Journal of English Linguistics*, *8*(5), 215–229.

Azher, M., Mehmood, M. A., & Shah, S. I. (2018). Linguistic variation across research sections of Pakistan academic writing: A multidimensional analysis. *International Journal of English Linguistics*, *8*(1), 30. https://doi.org/10.5539/ijel.v8n1p30

Aziz, M. S., & Mahmood, A. (2012). Ditransitive verbs: A corpus based comparison of Pakistani and British written English. *Wulfenia Journal*, *19*(11), 275–286.

Baumgardner, Robert J. (1995). Pakistani English: Acceptability and the norm. *World Englishes*, *14*(2), 261–271.

Baumgardner, Robert Jackson. (1987). Utilizing Pakistani newspaper English to teach grammar. *World Englishes*, *6*(3), 241–252.

Baumgardner, Robert Jackson. (1992). 'To Shariat or not to Shariat?': Bilingual functional shifts in Pakistani English. *World Englishes*, *11*(2–3), 129–140.

Baumgardner, Robert Jackson. (1993). *The English Language in Pakistan*. Karachi: Oxford

  University Press.

Baumgardner, Robert Jackson. (1996). Innovation in Pakistani English political lexis. In Robert

  Jackson Baumgardner (Ed.), *South Asian English: Structure, use, and users* (pp. 174–88).

  Urbana and Chicago: University of Illinois Press.

Baumgardner, Robert Jackson, Kennedy, A. E., & Shamim, F. (1993). The Urduization of

  English in Pakistan. In *The English language in Pakistan* (pp. 83–203). Karachi: Oxford

  University Press.

Bernaisch, T., Gries, S. T., & Mukherjee, J. (2014). The dative alternation in South Asian

  English(es): Modelling predictors and predicting prototypes. *English World-Wide*, *35*(1),

  7–31. https://doi.org/10.1075/eww.35.1.02ber

Biber, D. (1986). Spoken and written text on dimensions in English: Resolving the contradictory

  findings. *Language*, *62*, 384– 414.

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

Biber, D. (1989). A typology of English texts. *Linguistics*, *27*(1), 3–44.

  https://doi.org/10.1515/ling.1989.27.1.3

Biber, D. (1994). An analytical framework for register studies. In D. Biber & E. Finegan (Eds.),

  Sociolinguistic perspectives on register (pp. 31–56). New York: Oxford University Press.

Biber, D. (2004). Conversation text types: A multi-dimensional analysis. *Le Poids Des Mots:*

  Proc. of the 7th International Conference on the Statistical Analysis of Textual Data,

  Louvain: Presses Universitaires de Louvain, 15–34.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*

  (Vol. 23). Amsterdam/Philadelphia: Benjamins.

Biber, D. (2012). Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*, *8*(1). https://doi.org/10.1515/cllt-2012-0002

Biber, D. (2014). Using multi-dimensional analysis to explore cross-linguistic universals of register variation. *Languages in Contrast*, *14*(1), 7–34. https://doi.org/10.1075/lic.14.1.02bib

Biber, D. (2019). Text-linguistic approaches to register variation. *Register Studies*, *1*(1), 42–75. https://doi.org/10.1075/rs.18007.bib

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

Biber, D., & Egbert, J. (2016). Register variation on the searchable web: A multi-dimensional analysis. *Journal of English Linguistics*, *44*(2), 1–44. https://doi.org/10.1177/0075424216628955

Biber, D., & Egbert, J. (2018). *Register variation online* (1st ed.). Cambridge: Cambridge University Press.

Biber, D., Egbert, J., & Davies, M. (2015). Exploring the composition of the searchable web: A corpus-based taxonomy of web registers. *Corpora*, *10*(1), 11–45. https://doi.org/10.3366/cor.2015.0065

Biber, D., Egbert, J., Gray, B., Oppliger, R., & Szmrecsanyi, B. (2016). Variationist versus text-linguistic approaches to grammatical change in English: Nominal modifiers of head nouns. In M. Kyto & P. Pahta (Eds.), *The Cambridge handbook of English historical linguistics* (pp. 351–375). https://doi.org/10.1017/cbo9781139600231.022

Biber, D., & Gray, B. (2013). Identifying multi-dimensional patterns of variation across registers. In M. Krug & J. Schlüter (Eds.), *Research methods in language variation and change* (pp. 402–420). New York: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Essex: Longman.

Biber, D., & Jones, J. K. (2009). Quantitative methods in corpus linguistics. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (pp. 1286–1304). Berlin: De Gruyter.

Biber, D., & Kurjian, J. (2007). Towards a taxonomy of web registers and text types: A multi-dimensional analysis. In M. Hundt, N. Nesselhauf, & C. Biewer (Eds.), *Corpus linguistics and the web* (pp. 109–131). Amsterdam/New York: Brill Rodopi.

Bieswanger, M. (2016). Electronically-mediated Englishes: Synchronicity revisited. In L. Squires (Ed.), *English in computer-mediated communication: Variation, representation, and change* (pp. 281–300). Berlin/Boston: De Gruyter.

Blom, J.-P., & Gumperz, J. J. (1972). Social meaning in linguistic structures: Code-switching in Northern Norway. In J. J. Gumperz & D. H. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 407–434). New York: Holt, Rinehart and Winston.

Bohmann, A. (forthcoming). The role of register in structuring variation in World Englishes. In B. Schneider & T. Heyd (Eds.), *Bloomsbury World Englishes*. London: Bloomsbury.

Bohmann, A. (2016). Language change because Twitter? Factors motivating innovative uses of *because* across the English-speaking Twittersphere. In L. Squires (Ed.), *English in computer-mediated communication: Variation, representation, and change* (Vol. 93, pp. 149–178). Berlin/Boston: De Gruyter.

Bohmann, A. (2017). *Variation in English world-wide: Varieties and genres in a quantitative perspective* (PhD Thesis, The University of Texas at Austin). Retrieved from https://repositories.lib.utexas.edu/handle/2152/47343

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.

Buschfeld, S., & Kautzsch, A. (2017). Towards an integrated approach to postcolonial and non-postcolonial Englishes. *World Englishes*, *36*(1), 104–126. https://doi.org/10.1111/weng.12203

Buschfeld, S., Kautzsch, A., & Schneider, E. W. (2018). From colonial dynamism to current transnationalism: A unified view on postcolonial and non-postcolonial Englishes. In S. C. Deshors (Ed.), *Modeling World Englishes: Assessing the interplay of emancipation and globalization of ESL varieties* (pp. 15–44). Amsterdam: Benjamins.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, *61*(6), 1–36.

Clarke, I., & Grieve, J. (2017). Dimensions of abusive language on Twitter. *Proceedings of the First Workshop on Abusive Language Online*, 1–10.

Coats, S. (2016). Grammatical feature frequencies of English on Twitter in Finland. In L. Squires (Ed.), *English in computer-mediated communication: Variation, representation, and change* (Vol. 93, pp. 179– 209). Berlin/Boston: De Gruyter.

Collot, M., & Belmore, N. (1996). Electronic language: A new variety of English. In S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives* (Vol. 39, pp. 13–28). Amsterdam/Philadelphia: Benjamins.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four

recommendations for getting the most from your analysis. *Practical Assessment,
Research & Evaluation*, *10*(7), 1–9.

Dąbrowska, M. (2013). *Variation in language: Faces of Facebook English*. Frankfurt: De

Gruyter.

Davies, M. (2008-). *The corpus of contemporary American English (COCA): 560 million words,
1990-present.* Retrieved from https://corpus.byu.edu/coca/

Davies, M., & Fuchs, R. (2015). Expanding horizons in the study of World Englishes with the

1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide*,
36(1), 1–28.

Deshors, S. C., & Gilquin, G. (2018). Modeling World Englishes in the 21st century new

reflections on model making. In S. C. Deshors (Ed.), *Modeling World Englishes:
Assessing the interplay of emancipation and globalization of ESL varieties* (pp. 281–
294). Amsterdam: Benjamins.

Dinno, A. (2017). *dunn.test: Dunn's test of multiple comparisons using rank sums*. Retrieved

from https://CRAN.R-project.org/package=dunn.test

Distefano, C., Zhu, M., & Mîndrilă, D. (2009). Understanding and using factor scores:

Considerations for the applied researcher. *Practical Assessment, Research & Evaluation,
14*(20), 1–11.

Diwersy, S., Evert, S., & Neumann, S. (2014). A weakly supervised multivariate approach to the

study of language variation. In B. Szmrecsanyi & B. Wälchli (Eds.), *Aggregating
dialectology, typology, and register analysis: Linguistic variation in text and speech* (pp.
174–204). Berlin: De Gruyter.

Du Bois, J. W., Chafe, W. L., Meyer, C., Thompson, S. A., & Martey, N. (2000). *Santa Barbara corpus of spoken American English*. [CD-ROM]. Philadelphia: Linguistic Data Consortium.

Egbert, J. (2014). *Reader perceptions of linguistic variation in published academic writing* (PhD Thesis). Northern Arizona University.

Egbert, J., & Biber, D. (2016). Do all roads lead to Rome?: Modeling register variation with factor analysis and discriminant analysis. *Corpus Linguistics and Linguistic Theory, 14*(2), 1–41. https://doi.org/10.1515/cllt-2016-0016

Egbert, J., Biber, D., & Davies, M. (2015). Developing a bottom-up, user-based method of web register classification. *Journal of the Association for Information Science and Technology*, *66*(9), 1817–1831. https://doi.org/10.1002/asi.23308

Egbert, J., & Staples, S. (2019). Doing multidimensional analysis in SPSS, SAS and R. In T. Berber-Sardinha & M. V. Pinto (Eds.), *Multidimensional analysis research methods and current issues* (pp. 125–144). London: Bloomsbury.

Ehret, K., & Taboada, M. (2018). *Are online comments like conversations? A multi-dimensional exploration of online news comments*. Conference presentation presented at the New York. New York.

Evert, S. (2018). *A multivariate approach to linguistic variation*. Retrieved from http://www.stefan-evert.de/SIGIL/sigil_R/materials/07_multivar_talk.handout.pdf

Express 24/7 bows out amidst revenue drop. (2011, November 29). *The Express Tribune*. Retrieved from https://tribune.com.pk/story/299263/express-247-bows-out-amidst-revenue-drop/

Farooq, M. (2019, February 5). Active social media users in Pakistan grow by 5.7%: Report. *Pakistan Today*. Retrieved from https://profit.pakistantoday.com.pk/2019/02/05/active-social-media-users-in-pakistan-grow-by-5-7-report/

Ferguson, C. A. (1994). Dialect, register, and genre: Working assumptions about conventionalization. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 15–30). New York: Oxford University Press.

Field, A., Miles, J., & Field, Z. (2013). *Discovering statistics using R*. Retrieved from https://books.google.de/books?id=mMcJrgEACAAJ

Friendly, M., & Fox, J. (2017). *candisc: Visualizing generalized canonical discriminant and canonical correlation analysis*. Retrieved from https://CRAN.R-project.org/package=candisc

Friginal, E., Waugh, O., & Titak, A. (2018). Linguistic variation in Facebook and Twitter posts. In E. Friginal (Ed.), *Studies in corpus-based sociolinguistics* (pp. 342–362). New York: Routledge.

Gillani, M., & Mahmood, R. (2014). Politeness strategies in Pakistani Business English letters: A study of opening and closing strategies. *International Journal of Linguistics*, *4*, 13–23.

Gray, B., & Egbert, J. (2019). Editorial: Register and register variation. *Register Studies*, *1*(1), 1–9. https://doi.org/10.1075/rs.00001.edi

Gries, S. Th., & Bernaisch, T. (2016). Exploring epicentres empirically: Focus on South Asian Englishes. *English World-Wide*, *37*(1), 1–25. https://doi.org/10.1075/eww.37.1.01gri

Grieve, J. (2016). *Regional Variation in Written American English*. https://doi.org/10.1017/cbo9781139506137

Grieve, J. (2018). Visualisation is necessary to decide if clustering is appropriate in corpus analysis. *ICAME39*.

Grieve, J., Biber, D., Friginal, E., & Nekrasova, T. (2010). Variation among blogs: A multi-dimensional analysis. In A. Mehler, S. Sharoff, & M. Santini (Eds.), *Genres on the Web* (pp. 303–322). London/New York: Springer.

Grosjean, F. (1982). *Life with Two Languages*. Retrieved from https://www.ebook.de/de/product/3239172/francois_grosjean_life_with_two_languages.html

Gumperz, John J. (1982). *Discourse strategies* (Vol. 1). Cambridge: Cambridge University Press.

Halim, N. S., & Maros, M. (2014). The functions of code-switching in Facebook interactions. *Procedia - Social and Behavioral Sciences*, *118*, 126–133. https://doi.org/10.1016/j.sbspro.2014.02.017

Hall, R. A. (1962). The life cycle of pidgin languages. *Lingua*, *11*, 151–156. https://doi.org/10.1016/0024-3841(62)90021-9

Halliday, M. A. (1978). *Language as social semiotics*. London: Edward Arnold.

Halliday, M., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Haque, A. R. (1983). The position and status of English in Pakistan. *World Englishes*, *2*(1), 6–9.

Hardy, J. A., & Friginal, E. (2012). Filipino and American online communication and linguistic variation. *World Englishes*, *31*(2), 143–161.

Henry, A., & Ho, D. G. E. (2016). Code-switching in Bruneian online retail transactions. *World Englishes*, *35*(4), 554–570. https://doi.org/10.1111/weng.12226

Herring, S. C. (Ed.). (1996). *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives* (Vol. 39). Amsterdam/Philadelphia: Benjamins.

Herring, S. C. (2007). A faceted classification scheme for computer-mediated discourse. *Language@Internet*, *4*(1). Retrieved from http://nbn-resolving.de/urn:nbn:de:0009-7-7611

Herring, S. C. (2011). Computer-mediated conversation part II: Introduction and overview. *Language@ Internet*, *8*(2), 1–12.

Herring, S. C., & Paolillo, J. C. (2006). Gender and genre variation in weblogs. *Journal of Sociolinguistics*, *10*(4), 439–459. https://doi.org/10.1111/j.1467-9841.2006.00287.x

Herring, S. C., Scheidt, L. A., Bonus, S., & Wright, E. (2004). Bridging the gap: A genre analysis of weblogs. *Proceedings of the 37th Hawaii International Conference on System Sciences - 2004*, 1–11. https://doi.org/10.1109/HICSS.2004.1265271

Hinrichs, L. (2016). Modular repertoires in English-using social networks: A study of language choice in the networks of adult Facebook users. In L. Squires (Ed.), *English in computer-mediated communication: Variation, representation, and change* (pp. 17–42). Berlin/Boston: De Gruyter.

Hundt, M., Hoffmann, S., & Mukherjee, J. (2012). The hypothetical subjunctive in South Asian Englishes: Local developments in the use of a global construction. *English World-Wide*, *33*(2), 147–164. https://doi.org/10.1075/eww.33.2.02hun

Hussain, Z. (2016). *A multidimensional comparative analysis of Pakistani English* (PhD Thesis). Govt. College University Faisalabad.

Hussain, Z., & Mahmood, M. A. (2014). Invariant tag questions in Pakistani English: A comparison with native and other non-native Englishes. *Asian Englishes*, *16*(3), 229–238. https://doi.org/10.1080/13488678.2014.951465

Hymes, D. (1974). *Foundations in sociolinguistics: An ethnographic approach*. Routledge.

Ijaz, M. T., Mahmood, M. A., & Ameer, A. (2014). A corpus based study of the errors committed by Pakistani learners of English at graduation level. *Journal of Education and Practice*, *5*(24), 159–162.

Iqbal, A. (2018). *Pakistani Englishes: Syntactic variations* (1st ed.). https://doi.org/10.4324/9780429489914

ISPAK. (2014). *History of Internet in Pakistan*. Retrieved from http://www.ispak.pk/internet_pakistan.php

Jabeen, F., Mahmood, M. A., & Rasheed, S. (2011). An attitudinal study of Pakistani English. *Interdisciplinary Journal of Contemporary Research in Business*, *3*(5), 109–119.

Jabeen, F., Rai, M. A., & Arif, S. (2011). A corpus based study of discourse markers in British and Pakistani speech. *International Journal of Language Studies*, *5*(4), 69–86.

Jameel, I., Mahmood, M. A., Hussain, Z., & Shakir, A. (2014). A corpus-based analysis of linking adverbials in Pakistani English. *International Journal of Linguistics*, *6*(3), 133–140.

Jonsson, E. (2015). *Conversational writing: A multidimensional study of synchronous and supersynchronous computer-mediated communication*. Frankfurt: Peter Lang.

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11–30). Cambridge: Cambridge University Press.

Kachru, B. B. (1988). The sacred cows of English. *English Today*, *4*(04), 3. https://doi.org/10.1017/S0266078400000973

Kachru, B. B. (1992). Models for non-native Englishes. In B. B. Kachru (Ed.), *The other tongue* (2nd ed., pp. 48–74). Urbana and Chicago: University of Illinois Press.

Kassambara, A. (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning* (Vol. 1). STHDA.

Khan, H. I. (2012). The evolution of Pakistani English (PakE) as a legitimate variety of English. *International Journal of Applied Linguistics and English Literature*, *1*(5), 90–99.

Khan, H. I. (2013). *An investigation of two universities' postgraduate students and their teachers' perceptions of policy and practice of English medium of instruction (EMI) in Pakistani universities* (PhD Thesis). University of Glasgow, Glasgow.

Khan, M. (2014, March 18). English as medium of instruction at primary schools welcomed. *Dawn*. Retrieved from https://www.dawn.com/news/1093911

Khushi, Q. (2011). *The forms and functions of English for army personnel in Pakistan with a specific reference to Pakistan Military Academy, Kakul* (PhD Thesis). Bahauddin Zakariya University, Multan.

Kirk, J., & Nelson, G. (2018). The International Corpus of English project: A progress report. *World Englishes*, *37*(4), 697–716. https://doi.org/10.1111/weng.12350

Kirkpatrick, A. (2007). *World Englishes implications for international communication and English language teaching*. Cambridge: Cambridge University Press.

Kirkpatrick, A. (2014). World Englishes. In Constant Leung & Brian V Street (Eds.), *The Routledge companion to English studies* (pp. 33–45). New York: Routledge.

Koch, C., & Bernaisch, T. (2013). Verb complementation in South Asian English(es): The range and frequency of "new" ditransitives. In G. Andersen & K. Bech (Eds.), *English corpus*

linguistics: Variation in time, space and genre selected papers from ICAME 32 (pp. 69–89). Amsterdam: Rodopi.

Kopaczyk, J., & Tyrkkö, J. (2018). Blogging around the world: Universal and localised patterns in online Englishes. In J. Kopaczyk & J. Tyrkkö (Eds.), *Applications of pattern-driven methods in corpus linguistics* (pp. 277–310). Amsterdam: Benjamins.

Kortmann, B., & Szmrecsanyi, B. (2004). Global synopsis: Morphological and syntactic variation in English. In B. Kortmann, E. W. Schneider, K. Burridge, R. Mesthrie, & C. Upton (Eds.), *A handbook of varieties of English*: *Vol. 2: Morphology and syntax* (pp. 1142–1202). Berlin, New York: De Gruyter.

Krishnamurthy, S. (2002). The multidimensionality of blog conversations: The virtual enactment of September 11. *Internet Research 3.0, 2002*.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Lange, C. (2009). 'Where's the Party yaar!' Discourse particles in Indian English. In T. Hoffmann & L. Siebers (Eds.), *World Englishes–problems, properties and prospects. Selected papers from the 13th IAWE conference* (pp. 207–22). Amsterdam/Philadelphia: John Benjamins.

Lange, C. (2016). The 'intrusive *as*'-construction in South Asian varieties of English. *World Englishes*, *35*(1), 133–146. https://doi.org/10.1111/weng.12173

Lee, C. K. (2011). Micro-blogging and status updates on Facebook: Texts and practices. In C. Thurlow & K. Mroczek (Eds.), *Digital discourse: Language in the new media* (pp. 110–128). Oxford: Oxford University Press.

Levshina, N. (2014). Geographic variation of quite + ADJ in twenty national varieties of English: A pilot study. *Yearbook of the German Cognitive Linguistics Association*, *2*(1). https://doi.org/10.1515/gcla-2014-0008

Li, D. C. (1996). *Issues in bilingualism and biculturalism: A Hong Kong case study* (Vol. 21). Peter Lang.

Li, D. C. S. (2000). Cantonese-English code-switching research in Hong Kong: A Y2K review. *World Englishes*, *19*(3), 305–322. https://doi.org/10.1111/1467-971x.00181

Liimatta, A. (2016). *Exploring register variation on Reddit: A multi-dimensional study* (PhD Thesis). University of Helsinki, Helsinki.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2017). *cluster: Cluster analysis basics and extensions*. Retrieved from https://cran.r-project.org/web/packages/cluster/cluster.pdf

Mahboob, A. (2014a). Language variation and education: A focus on Pakistan. In S. Buschfeld, T. Hoffmann, M. Huber, & A. Kautzsch (Eds.), *The evolution of Englishes the Dynamic Model and beyond* (pp. 267–280). Amsterdam/Philadelphia: Benjamins.

Mahboob, A. (2004). Pakistani English: Morphology and syntax. In B. Kortmann & E. W. Schneider (Eds.), *A handbook of varieties of English* (Vol. 2, pp. 1045–1057). Berlin, New York: De Gruyter.

Mahboob, A. (2014b). *The future of English in Pakistan* (pp. 7–26). Retrieved from Strengthening Participatory Organization website: http://www.spopk.org/wp-content/uploads/2018/09/dp_volume_i.pdf

Mahboob, A. (2003). The English Language in Pakistan a brief overview of its history and linguistics. *Pakistan Journal of Language*, *4*(1), 1–48.

Mahboob, A. (2009). English as an Islamic language: A case study of Pakistani English. *World Englishes*, *28*(2), 175–189. https://doi.org/10.1111/j.1467-971x.2009.01583.x

Mahboob, A. (2013). Pakistani English. In B. Kortmann & K. Lunkenheimer (Eds.), *World atlas of varieties of English* (pp. 531–539). Berlin, New York: De Gruyter.

Mahboob, A. (2015). Identity management, language variation and English language textbooks: Focus on Pakistan. In D. N. Djenar, A. Mahboob, & K. Cruickshank (Eds.), *Language and identity across modes of communications* (Vol. 6, pp. 153–177). Berlin/Boston/Munich: De Gruyter.

Mahmood, A., Mahmood, R., & Saeed, T. (2011). The subjunctive in World Englishes: A study of British, American & Pakistani Englishes. *European Journal of Social Sciences*, *20*(3), 489–498.

Mahmood, Muhammad A. (2009). *A corpus based analysis of Pakistani English* (PhD Thesis). Bahauddin Zakariya University, Multan.

Mahmood, Muhammad Asim, & Ali, N. (2011). Circumstance adverbials in Pakistani English: A corpus-based comparative analysis. *International Journal of Academic Research*, *3*(5), 47–51.

Mahmood, Muhammad Asim, Hassan, S. U., Mahmood, R., & Arif, N. (2012). A comparative study of ditransitive verb complementation patterns in Pakistani and British Englishes. *US-China Foreign Language*, *10*(5), 1159– 1165.

Mahmood, R. (2009). *A lexico-grammatical study of noun phrase in Pakistani English* (PhD Thesis). Bahauddin Zakariya University, Multan.

Mahmood, R., Batool, A., Shah, S. K., & Parveen, S. (2013). A corpus driven comparative analysis of modal verbs in Pakistani and British English fictions. *Research on Humanities and Social Sciences*, *3*(11), 28–37.

Mahmood, R., Mahmood, A., & Saeed, T. (2011). Hyphenated compounds in Pakistani English. *European Journal of Scientific Research*, *52*(2), 252–264.

Mahmood, R., Obaid, M., & Shakir, A. (2014). A critical discourse analysis of figurative language in Pakistani English newspapers. *International Journal of Linguistics*, *6*(3), 210. https://doi.org/10.5296/ijl.v6i3.5412

Mahmood, R., & Shah, S. K. (2011). Culture-specific vocabulary: A study of Urdu nouns in Pakistani English. *Komunikacija i Kultura Online*, *II*(2), 72–86.

Mahootian, S. (2006). Code switching and mixing. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (2nd ed., pp. 511–527). https://doi.org/10.1016/b0-08-044854-2/01507-8

Mair, C. (2013). The World System of Englishes: Accounting for the transnational importance of mobile and mediated vernaculars. *English World-Wide*, *34*(3), 253–278. https://doi.org/10.1075/eww.34.3.01mai

Mair, C. (2015). Response to Davies and Fuchs. *English World-Wide*, *36*(1), 29–33. https://doi.org/10.1075/eww.36.1.02mai

Mair, C. (2016). Englishes beyond and between the three circles: World Englishes research in the age of globalization. In E. Seoane & C. Suárez-Gómez (Eds.), *World Englishes new theoretical and methodological considerations* (Vol. G57, pp. 17–36). https://doi.org/10.1075/veaw.g57.02mai

Manan, S. A., David, M. K., Dumanig, F. P., & Channa, L. A. (2017). The glocalization of

    English in the Pakistan linguistic landscape. *World Englishes*, *36*(4), 645–665.

    https://doi.org/10.1111/weng.12213

Matthiessen, C. M. I. M. (2019). Register in Systemic Functional Linguistics. *Register Studies*,

    1(1), 10–41. https://doi.org/10.1075/rs.18010.mat

McLellan, J. A. (2005). *Malay-English language alternation in two Brunei Darussalam on-line

    discussion forums* (PhD Thesis). Curtin University of Technology.

Meer, P., & Deuber, D. (forthcoming). Standard English in Trinidad, multinormativity, and

    translocality: Implications for the Dynamic Model and the Extra- and Intra-territorial

    Forces Model. In S. Buschfeld & A. Kautzsch (Eds.), *Modelling World Englishes: A joint

    approach towards postcolonial and non-postcolonial varieties*. Edinburgh: Edinburgh

    University Press.

Mesthrie, R., & Bhatt, R. M. (2008). *World Englishes. The study of new linguistic varieties (Key

    topics in sociolinguistics)*. Cambridge: Cambridge University Press.

Miller, C. R., & Shepherd, D. (2004). Blogging as social action: A genre analysis of the weblog.

    Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs, 1–21.

Moag, R. F. (1992). The life-cycle of non-native Englishes: A case study. In B. B. Kachru (Ed.),

    The other tongue (2nd ed., pp. 233–252). Urbana and Chicago: University of Illinois

    Press.

Moghees, A., Dar, S. A., Zaid, M. unisa, & Saeed, M. (2017). The reduction and simplification of

    English grammar during computer mediated communication in Pakistani context.

    Linguistics and Literature Review, *3*(2), 1–27. https://doi.org/10.29145/2017/llr/030201

Montes-Alcalá, C. (2007). Blogging in two languages: Code-switching in bilingual blogs. *Selected proceedings of the third workshop on Spanish sociolinguistics*, 162–170. Cascadilla Proceedings Project Somerville, MA.

Mukherjee, J. (2007). Steady states in the evolution of new Englishes: Present-day Indian English as an equilibrium. *Journal of English Linguistics*, *35*(2), 157–187. https://doi.org/10.1177/0075424207301888

Mukherjee, J., & Hoffmann, S. (2006). Describing verb-complementational profiles of New Englishes: A pilot study of Indian English. *English World-Wide*, *27*(2), 147–173. https://doi.org/10.1075/eww.27.2.03muk

Mukherjee, J., & Schilk, M. (2012). Exploring variation and change in new Englishes looking into the International Corpus of English (ICE) and beyond. In T. Nevalainen & E. C. Traugott (Eds.), *The Oxford handbook of the history of English* (pp. 189–199). Oxford: Oxford University Press.

Mukherjee, J., Schilk, M., & Bernaisch, T. (2010). Compiling the Sri Lankan component of ICE: Principles, problems, prospects. *ICAME Journal*, *34*, 64–77.

Narrowteq. (2016). *Darcy ripper (version 1.2.3)*. Retrieved from https://darcyripper.com/features/downloads/

Nelson, G. (2002). *ICE Markup Manual for Written Texts*. Retrieved from http://ice-corpora.net/ice/written.doc

Nelson, G. (2015). Response to Davies and Fuchs. *English World-Wide*, *36*(1), 38–40. https://doi.org/10.1075/eww.36.1.02nel

Neumann, S. (2014). *Contrastive register variation: A quantitative approach to the comparison of English and German*. Berlin: De Gruyter.

Nguyen, L. H., & Holmes, S. (2019). Ten quick tips for effective dimensionality reduction. *PLOS Computational Biology*, *15*(6), e1006907. https://doi.org/10.1371/journal.pcbi.1006907

Nini, A. (2014). *Multidimensional Analysis Tagger 1.2, Manual*. Retrieved from https://sites.google.com/site/multidimensionaltagger/versions

Novianti, W. (2013). The use of code switching in Twitter (a case study in English education department). *Passage*, *1*(2), 1–10.

Over 30,000 teachers to get English training. (2018, June 16). *The Express Tribune*. Retrieved from https://tribune.com.pk/story/1736139/1-30000-teachers-get-english-training/

Pakistan Bureau of Statistics. (2017a). *Newspapers and periodicals by language and province*. Retrieved from http://www.pbs.gov.pk/sites/default/files//tables/Newspapers%20and%20Periodicals%20by%20Language%2C%20Provinces_0.pdf

Pakistan Bureau of Statistics. (2017b). *Provisional summary results of census 2017*. Retrieved from http://www.pbs.gov.pk/sites/default/files/Population_Census_2017_Results_0.pdf

Pakistan Telecommunication Authority. (2018). *PTA Annual Report 2018*. Retrieved from https://www.pta.gov.pk/assets/media/pta_ann_rep_2018_121218.pdf

Parveen, S., & Aslam, S. (2013). A study on reasons for code-switching in Facebook by Pakistani Urdu English bilinguals. *Language in India*, *13*(11).

Parveen, S., & Mehmood, A. (2013). A study on attitudes towards varieties of spoken English in Pakistani context. *Language in India*, *13*(3), 652–667.

Pennycook, A. (2009). Plurilithic Englishes: Towards a 3D model. In K. Murata & J. Jenkins (Eds.), *Global Englishes in Asian contexts: Current and future debates* (pp. 194–207). New York: Palgrave.

Popat, S. K., & Emmanuel, M. (2014). Review and comparative study of clustering techniques. *International Journal of Computer Science and Information Technologies*, *5*(1), 805–812.

Puschmann, C. (2010). *The corporate blog as an emerging genre of computer-mediated communication: Features, constraints, discourse situation*. Göttingen: Universitätsverlag Göttingen.

Quirk, R. (1985). The English language in a global context. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 1–6). Cambridge: Cambridge University Press.

R Core Team. (2013). *R: A language and environment for statistical computing*. Retrieved from http://www.R-project.org/

Rafi, M. S. (2013). Urdu and English contact in an e-discourse: Changes and implications. *Gomal University Journal of Research*, *29*(2), 78–86.

Rafi, M. S. (2017). Bilingualism and identity construction in the digital discourse. *Journal of Multicultural Discourses*, *12*(3), 254–271. https://doi.org/10.1080/17447143.2017.1342649

Rahman, T. (1991). The use of words in Pakistani English. *English Today*, *7*(02), 32–38. https://doi.org/10.1017/S0266078400005514

Rahman, T. (2015). *Pakistani English*. Islamabad: National Institute of Pakistan Studies, Quaid-iAzam University.

Rahman, T. (2016). The development of English in Pakistan. In G. Leitner, A. Hashim, & H.-G. Wolf (Eds.), *Communicating with Asia* (pp. 13–27). https://doi.org/10.1017/CBO9781107477186.002

Raj, A. (2017, May 11). The case for Urdu as Pakistan's official language. *Herald*. Retrieved

    from https://herald.dawn.com/news/1153737

Rasheed, S., & Mahmood, M. A. (2014). The use of progressives in Pakistani learners' English:

    A corpus based study. *Journal of Education and Practice*, *5*(9), 128–133.

Rashid, A., Mahmood, M. A., & Ahmad, S. (2017). Linguistic variation across research sections:

    A multidimensional analysis of Pakistani academic journal articles. *Global Language

    Review*, *II*(I), 15–37. https://doi.org/10.31703/glr.2017(II-I).02

Revelle, W. (2017). *psych: Procedures for psychological, psychometric, and personality

    research*. Retrieved from https://CRAN.R-project.org/package=psych

Rofess, S., & Mahmood, M. A. (2015). An investigation of generic structures of Pakistani

    doctoral thesis acknowledgements. *Journal of Education and Practice*, *6*(28), 128–143.

Sadia, S., & Ghani, M. (2019). Modality in editorials of Pakistani English newspapers: A corpus

    based study. *International Journal of English Linguistics*, *9*(1), 144.

    https://doi.org/10.5539/ijel.v9n1p144

San, H. K. (2009). *Chinese-English code-switching in blogs by Macao young people* (Master's

    Thesis). The University of Edinburgh.

Sardinha, T. B. (2014). 25 years later: Comparing internet and pre-internet registers. In T. B.

    Sardinha & M. V. Pinto (Eds.), *Multi-dimensional analysis, 25 years on: A tribute to

    Douglas Biber* (Vol. 60, pp. 81–105). Amsterdam/Philadelphia: Benjamins.

Sarfraz, S. (2011). Error analysis of the written English essays of Pakistani undergraduate

    students: A case study. *Asian Transactions on Basic & Applied Sciences*, *1*(3), 29–51.

Schildhauer, P. (2016). *Personal weblog a linguistic history*. Frankfurt: Peter Lang.

Schilk, M., & Hammel, M. (2014). The progressive in South Asian and Southeast Asian varieties of English – mapping areal homogeneity and heterogeneity. In L. Vandelanotte, K. Davidse, & C. Gentens (Eds.), *Recent Advances in Corpus Linguistics* (pp. 147–171). https://doi.org/10.1163/9789401211130_009

Schneider, Edgar W. (2007). *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.

Schneider, Edgar Werner. (2003). The dynamics of new Englishes: From identity construction to dialect birth. *Language*, *79*(2), 233–281. https://doi.org/10.1353/lan.2003.0136

Schneider, Edgar Werner. (2014). New reflections on the evolutionary dynamics of world Englishes. *World Englishes*, *33*(1), 9–32. https://doi.org/10.1111/weng.12069

Schubert, C. (2016). Introduction: Current trends in register research. In C. Schubert & C. Sanchez-Stockhammer (Eds.), *Revisiting register in English* (pp. 1–15). Berlin: De Gruyter.

Seargeant, P., & Tagg, C. (2011). English on the internet and a 'post-varieties' approach to language. *World Englishes*, *30*(4), 496–514. https://doi.org/10.1111/j.1467-971X.2011.01730.x

Seidlhofer, B. (2009). Common ground and different realities: World Englishes and English as a lingua franca. *World Englishes*, *28*(2), 236–245. https://doi.org/10.1111/j.1467-971X.2009.01592.x

Shakir, A. (2013). *Linguistic variation across print advertisements in Pakistani media a multidimensional analysis* (PhD Thesis). International Islamic University, Islamabad.

Shakir, M., & Deuber, D. (2019). A multidimensional analysis of Pakistani and U.S. English

blogs and columns. *English World-Wide*, *40*(1), 1–23.

https://doi.org/10.1075/eww.00020.sha

Shamim, F. (2008). Trends, issues and challenges in English language education in Pakistan.

Asia Pacific Journal of Education, *28*(3), 235–249.

https://doi.org/10.1080/02188790802267324

Shamsie, M. (2017). Pakistani-English Writing. In *Oxford Research Encyclopedia of Literature*

(pp. 1–29). https://doi.org/10.1093/acrefore/9780190201098.013.69

Sharma, B. K. (2012). Beyond social networking: Performing global Englishes in Facebook by

college youth in Nepal. *Journal of Sociolinguistics*, *16*(4), 483–509.

https://doi.org/10.1111/j.1467-9841.2012.00544.x

Special Correspondent. (2010, May 15). DawnNews changes to Urdu. *The Express Tribune*.

Retrieved from https://tribune.com.pk/story/13380/dawnnews-changes-to-urdu/

Szmrecsanyi, B. (2019). Register in variationist linguistics. *Register Studies*, *1*(1), 76–99.

https://doi.org/10.1075/rs.18006.szm

Tabassum, M., Shah, S. K., & Bilal, M. (2013). A critical discourse analysis of the left and right

wing ideologies in Pakistani English newspaper editorials. *Journal of Education and

Practice*, *4*(13), 72–78.

Talaat, M. (2002). *The form and functions of English in Pakistan* (PhD Thesis). Bahauddin

Zakariya University, Multan.

Talaat, M., & Anwar, B. (2010). The impact of Urdu-English code-switching on Pakistani

English. *Kashmir Journal of Language Research*, *13*(1), 121–146.

Titak, A., & Roberson, A. (2013). Dimensions of web registers: An exploratory multi-dimensional comparison. *Corpora*, *8*(2), 235–260. https://doi.org/10.3366/cor.2013.0042

Tremayne, M. (Ed.). (2007). *Blogging, citizenship, and the future of media*. New York: Routledge.

van Rooy, B., & Kruger, H. (2018). Hybridity, globalisation and models of Englishes. In S. C. Deshors (Ed.), *Modeling World Englishes: Assessing the interplay of emancipation and globalization of ESL varieties* (pp. 77–108). Amsterdam: Benjamins.

Web Desk. (2013, January 28). PTV's English channel to launch tomorrow: Report. *The Express Tribune*. Retrieved from https://tribune.com.pk/story/499693/ptvs-english-channel-to-launch-tomorrow-report/

Weihs, C., Ligges, U., Luebke, K., & Raabe, N. (2005). KlaR Analyzing German business Cycles. In D. Baier, R. Decker, & L. Schmidt-Thieme (Eds.), *Data Analysis and Decision Support* (pp. 335–343). Berlin: Springer-Verlag.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. Retrieved from http://ggplot2.org

Xiao, R. (2009). Multidimensional analysis and the study of world Englishes. *World Englishes*, *28*(4), 421–450.

Yano, Y. (2001). World Englishes in 2000 and beyond. *World Englishes*, *20*(2), 119–132. https://doi.org/10.1111/1467-971X.00204

Yano, Y. (2009). The future of English: Beyond the Kachruvian three circle model? In K. Murata & J. Jenkins (Eds.), *Global Englishes in Asian contexts: Current and future debates* (pp. 209–225). New York: Palgrave.

YiLan, L., & RuTong, Z. (2015). *clustertend: Check the clustering tendency*. Retrieved from

  https://CRAN.R-project.org/package=clustertend

Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create*

  affiliation on the web (Vol. 6). New York: Bloomsbury.

Zentella, A. C. (1997). *Growing up bilingual: Puerto Rican children in New York*.

  Malden/Oxford: Blackwell.

Appendix I Additional Figures
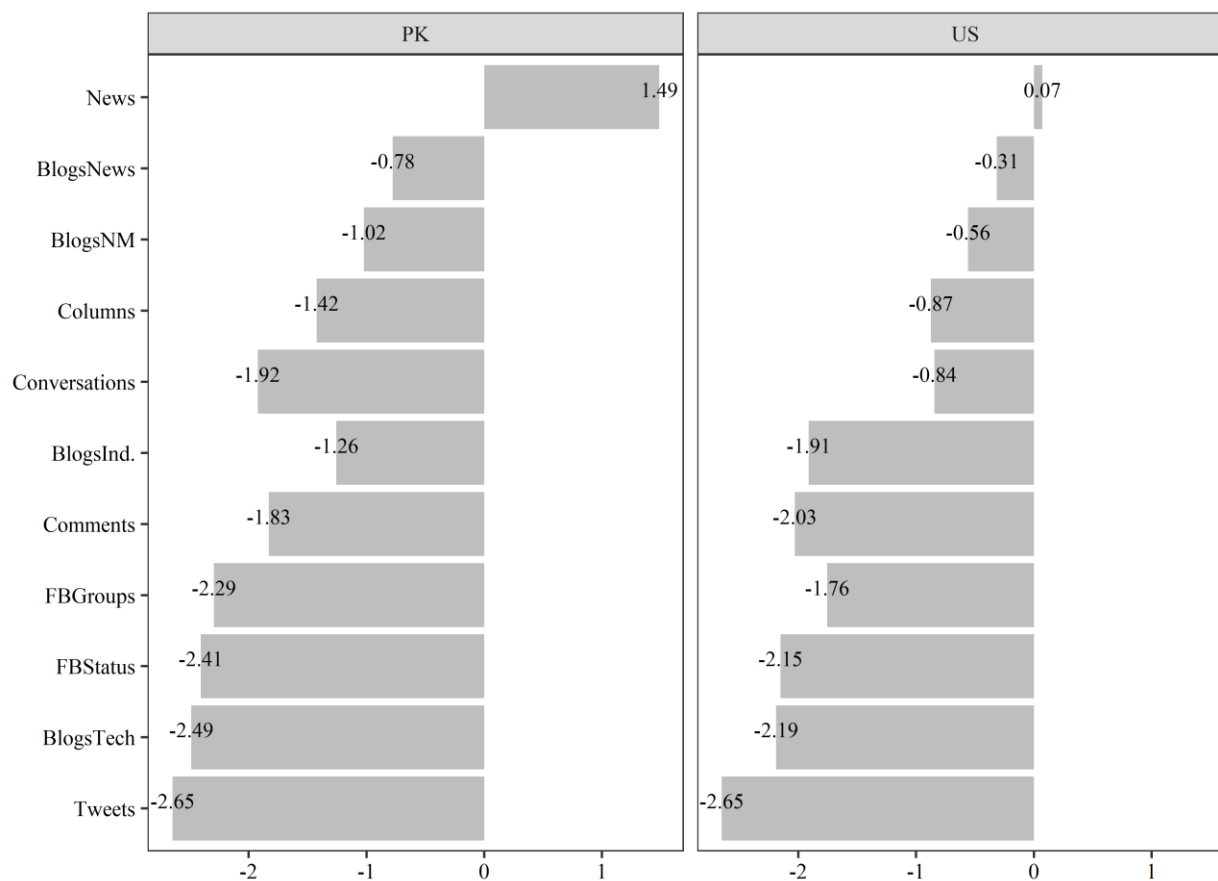
Figure 27 Online and offline registers on Biber's (1988) dimension 2 Narrative versus Non-narrative Discourse
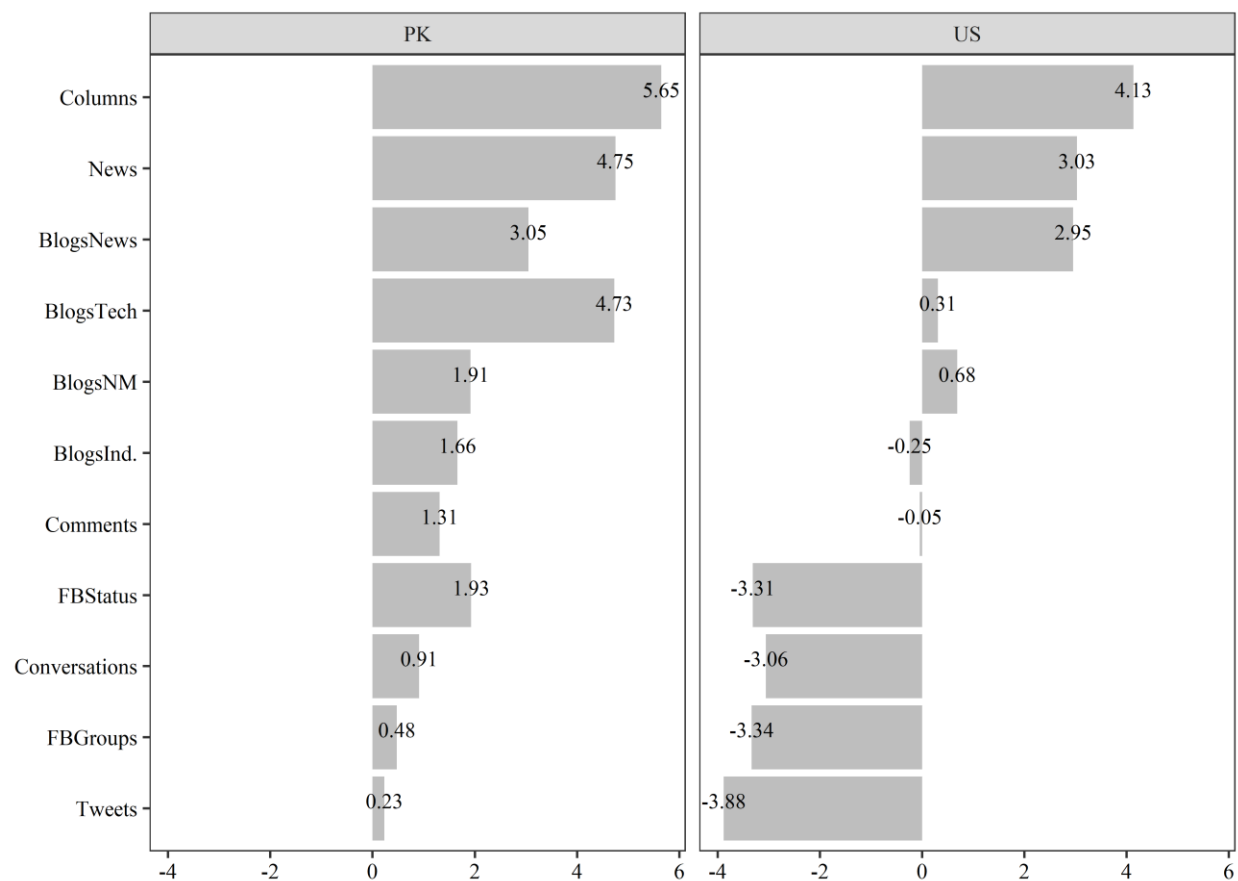
Figure 28 Online and offline registers on Biber's (1988) dimension 3 Elaborated versus Situation -Dependent Reference
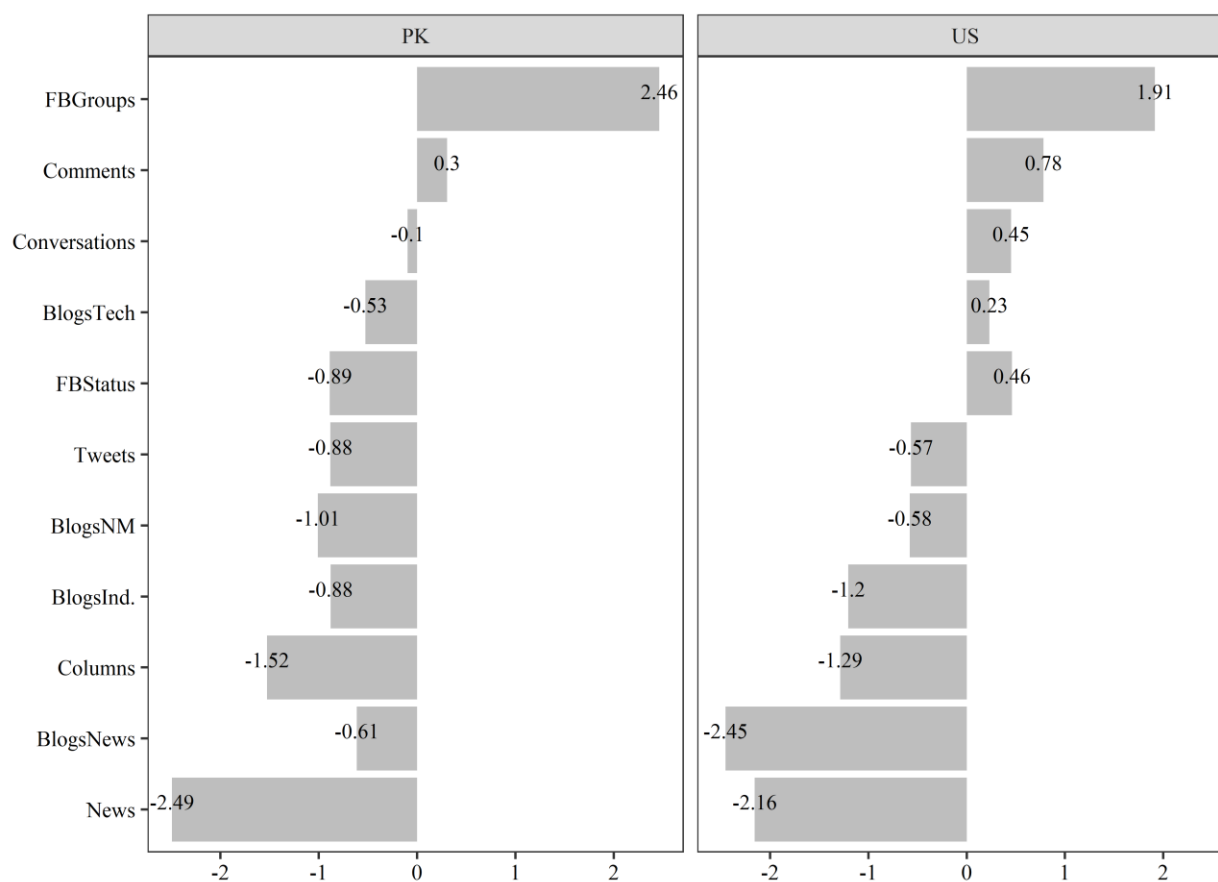
Figure 29 Online and offline registers on Biber's (1988) dimension 4 Over Expression of Persuasion
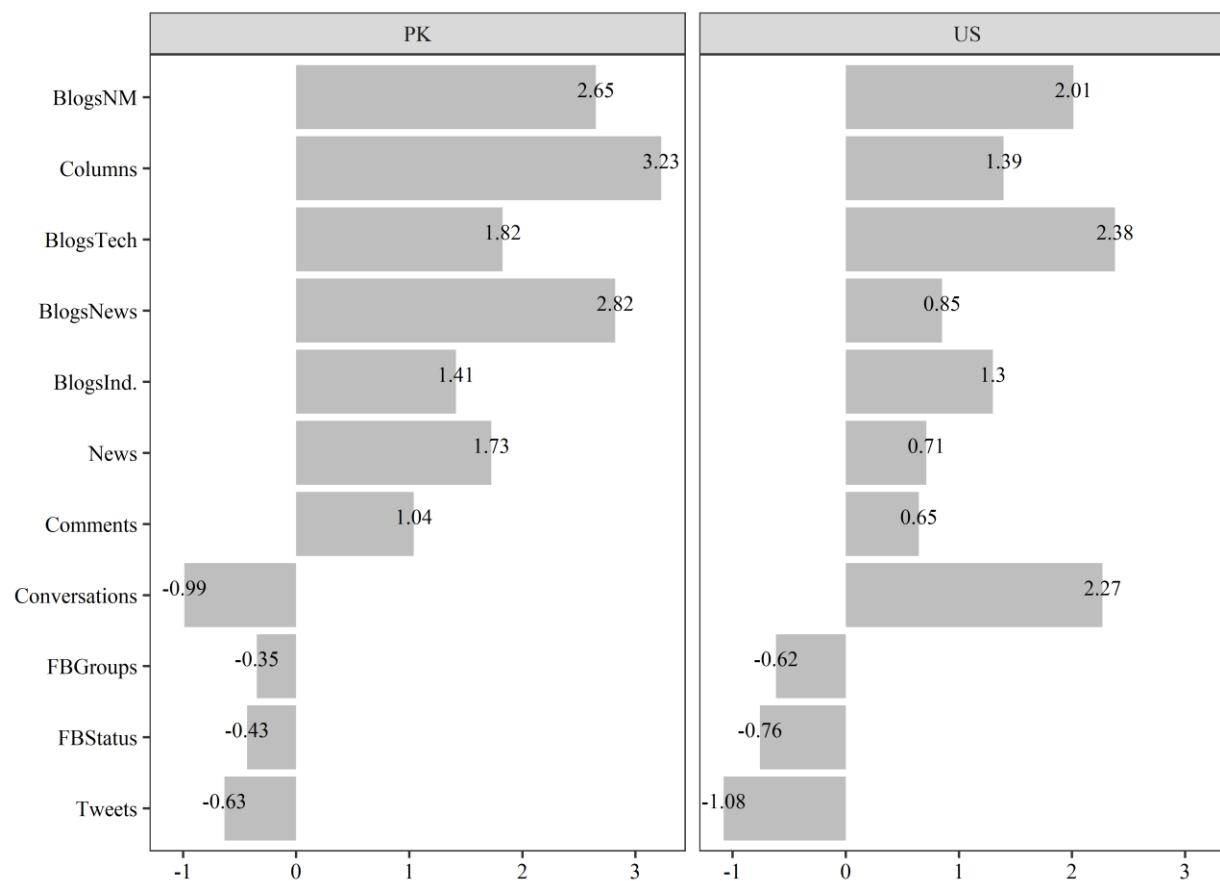
Figure 30 Online and offline registers on Biber's (1988) dimension 5 Abstract versus Non-Abstract Style
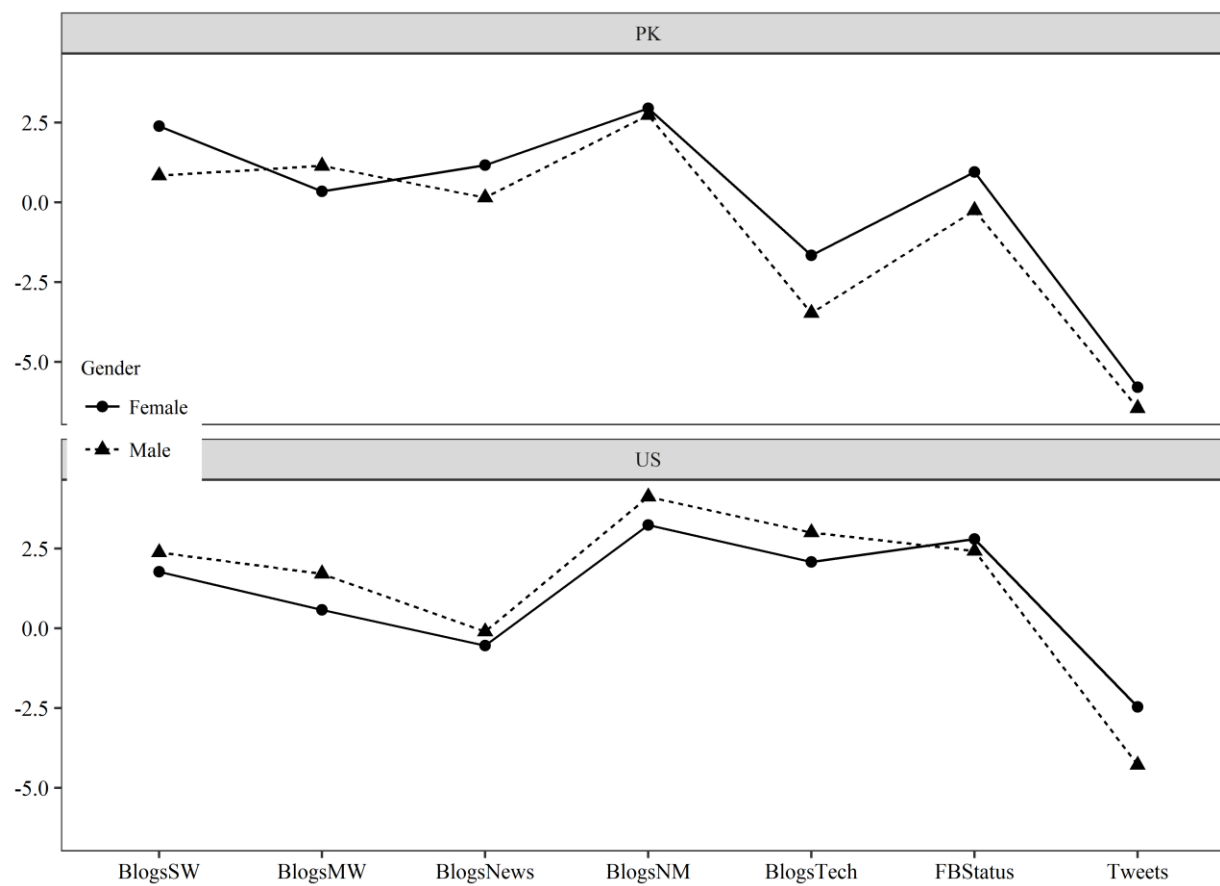
Figure 31 Online texts produced by male and female users on dimension 2 Conversational Style (mean dimension scores)
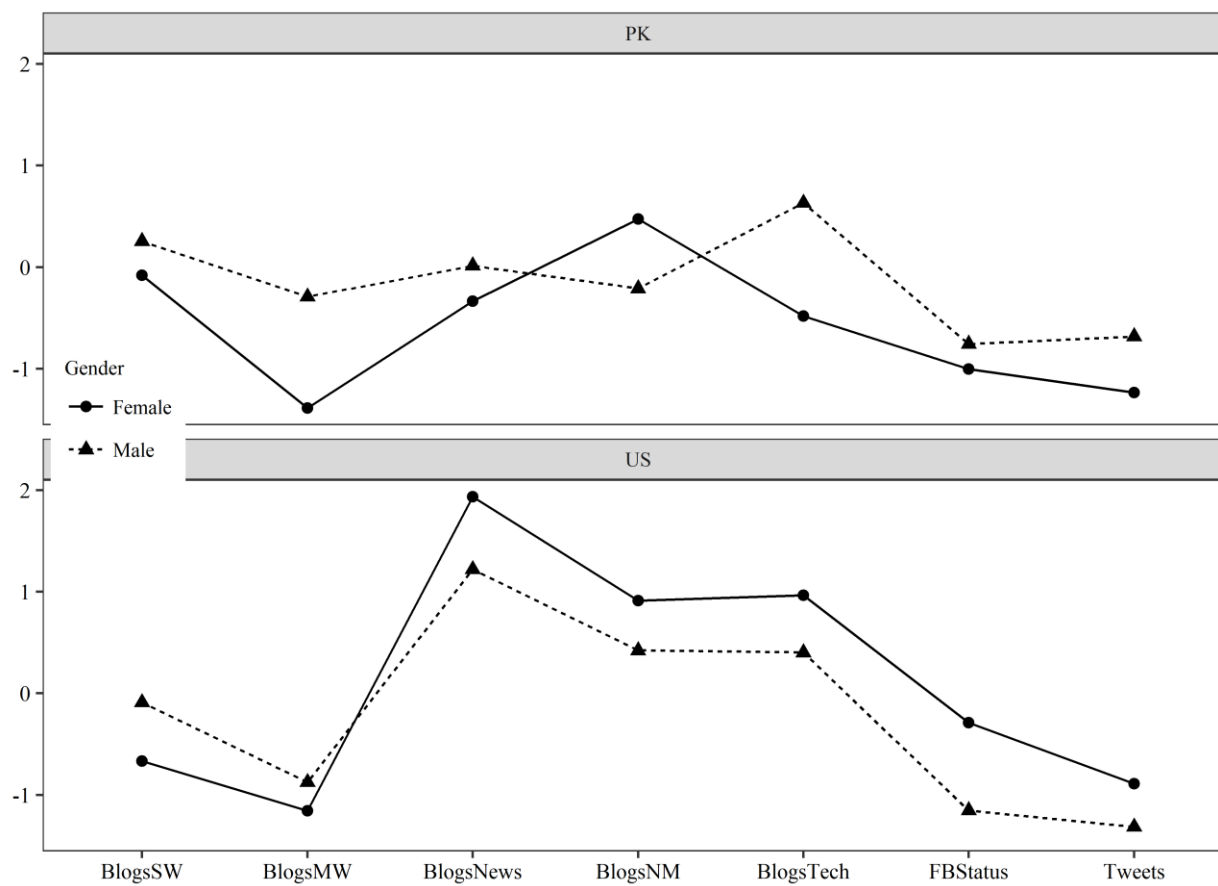
Figure 32 Online texts produced by male and female users on dimension 4 Reporting Style (mean dimension scores)
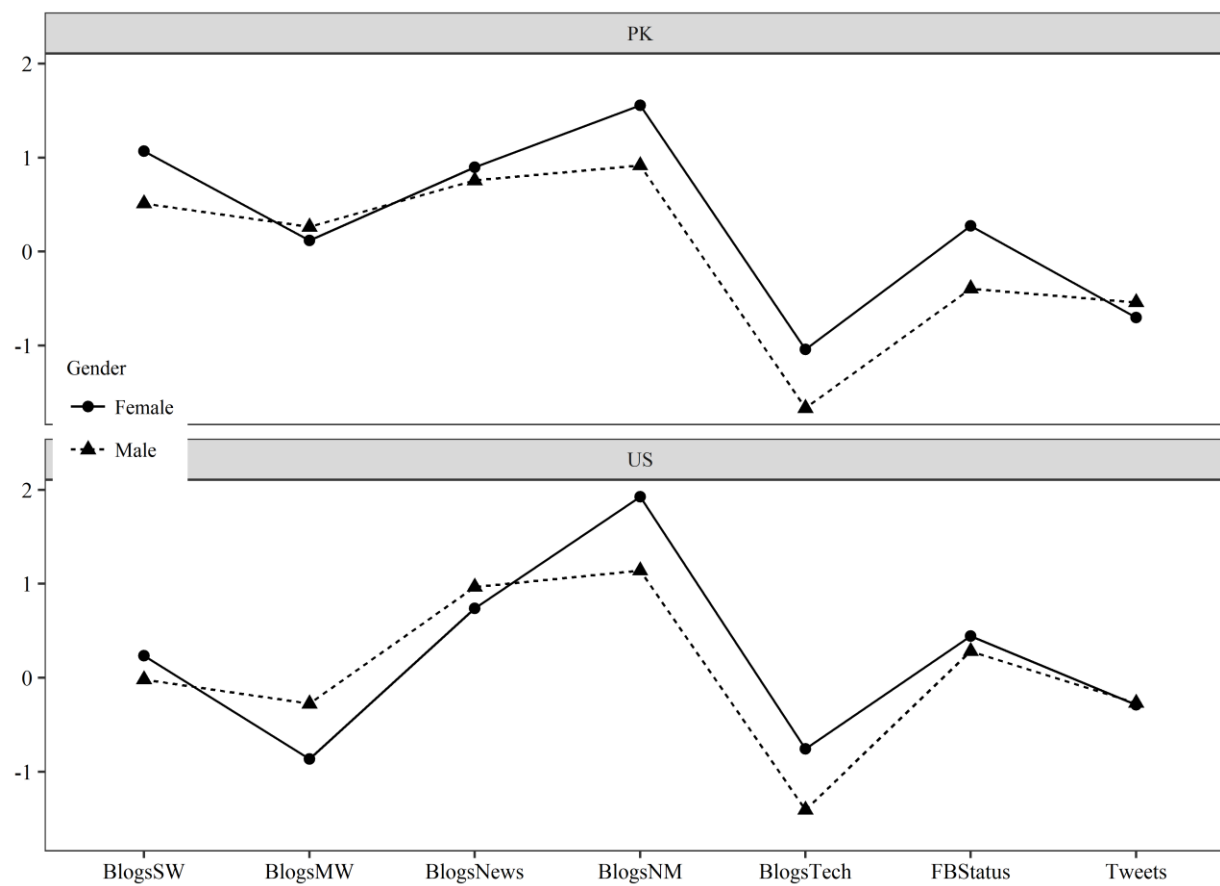
Figure 33 Online texts produced by male and female users on dimension 5 Narrative Focus (mean dimension scores)
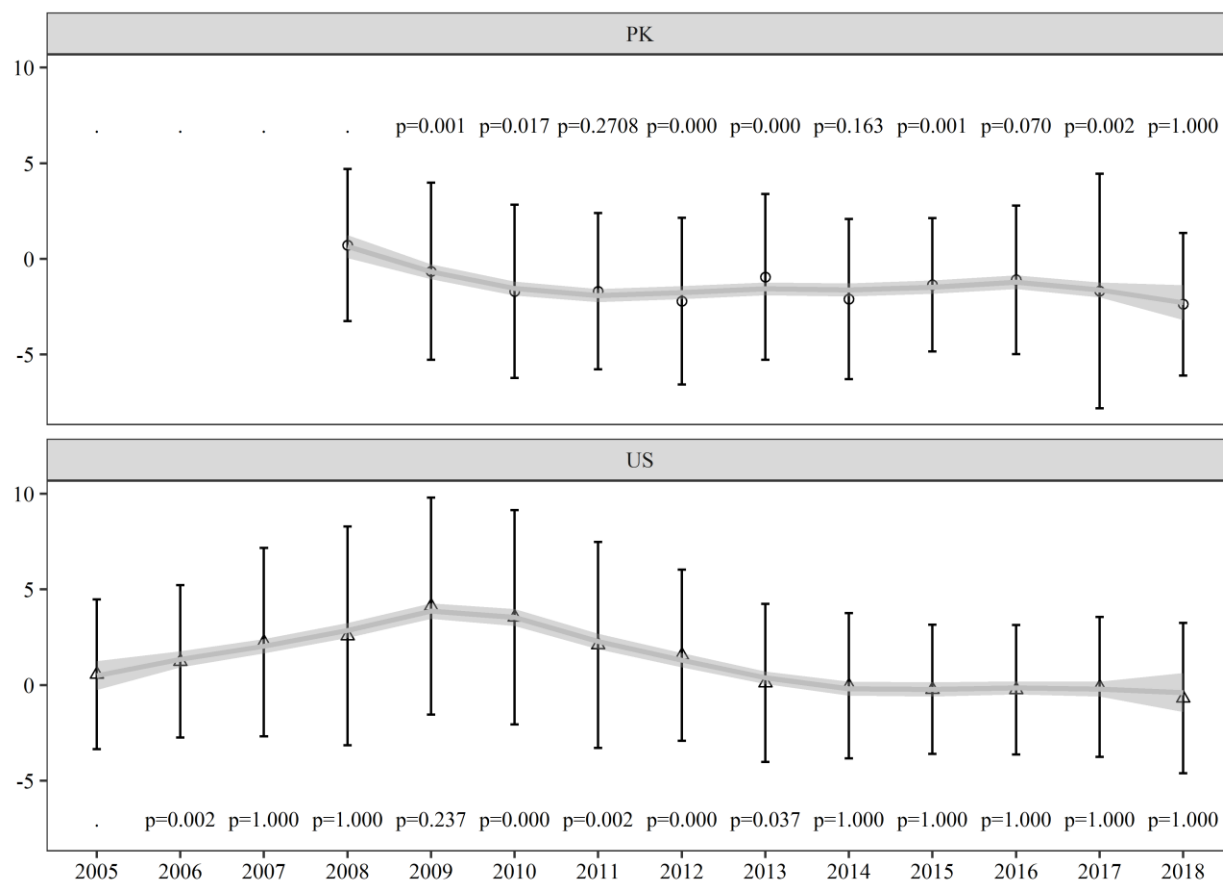
Figure 34 Chronological trend of technology blogs on dimension 2 Conversational Style (mean dimension scores and standard deviation as error bars)

Note: *P* values are based on post-hoc Dunn's test comparing each year to the first year.
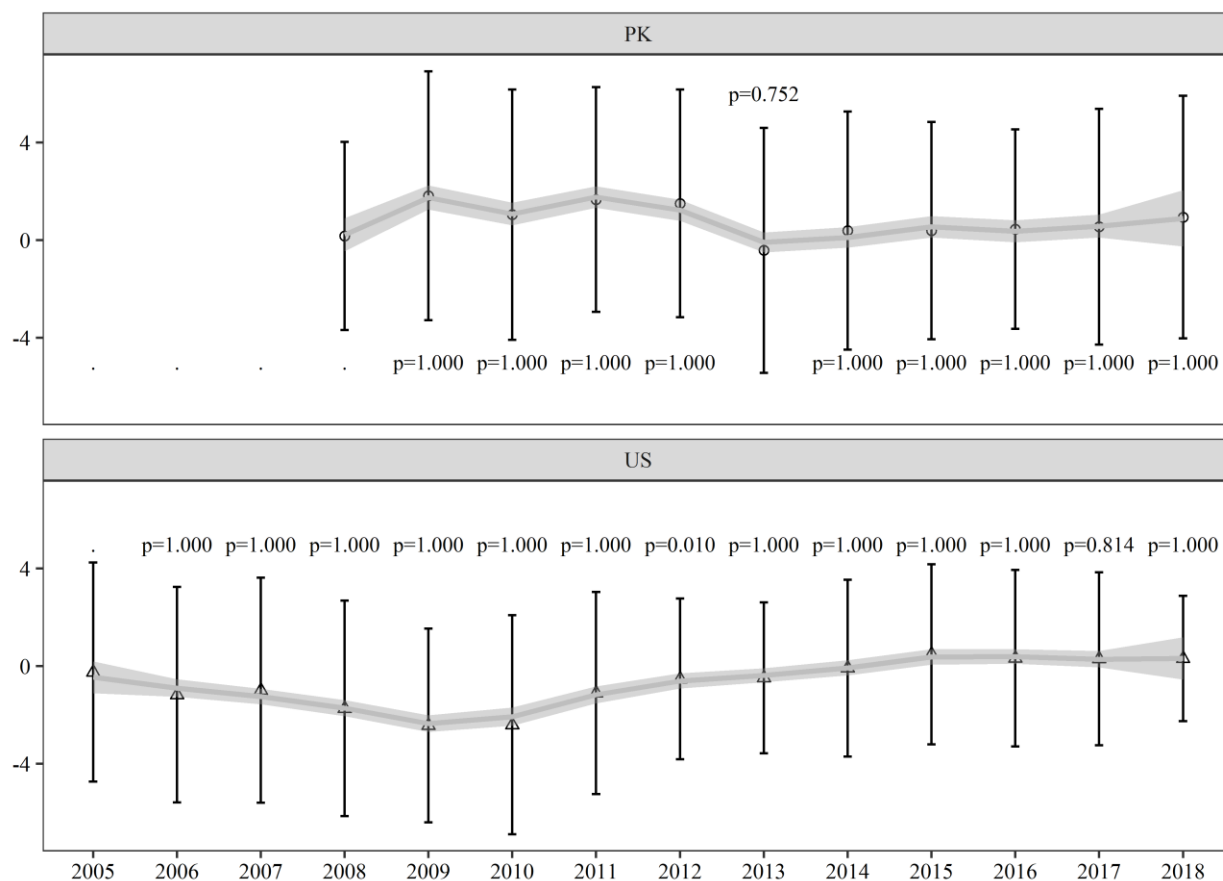
Figure 35 Chronological trend of technology blogs on dimension 3 Abstract Evaluative versus Non-abstract Information (mean dimension scores and standard deviation as error bars)

Note: *P* values are based on post-hoc Dunn's test comparing each year to the first year.
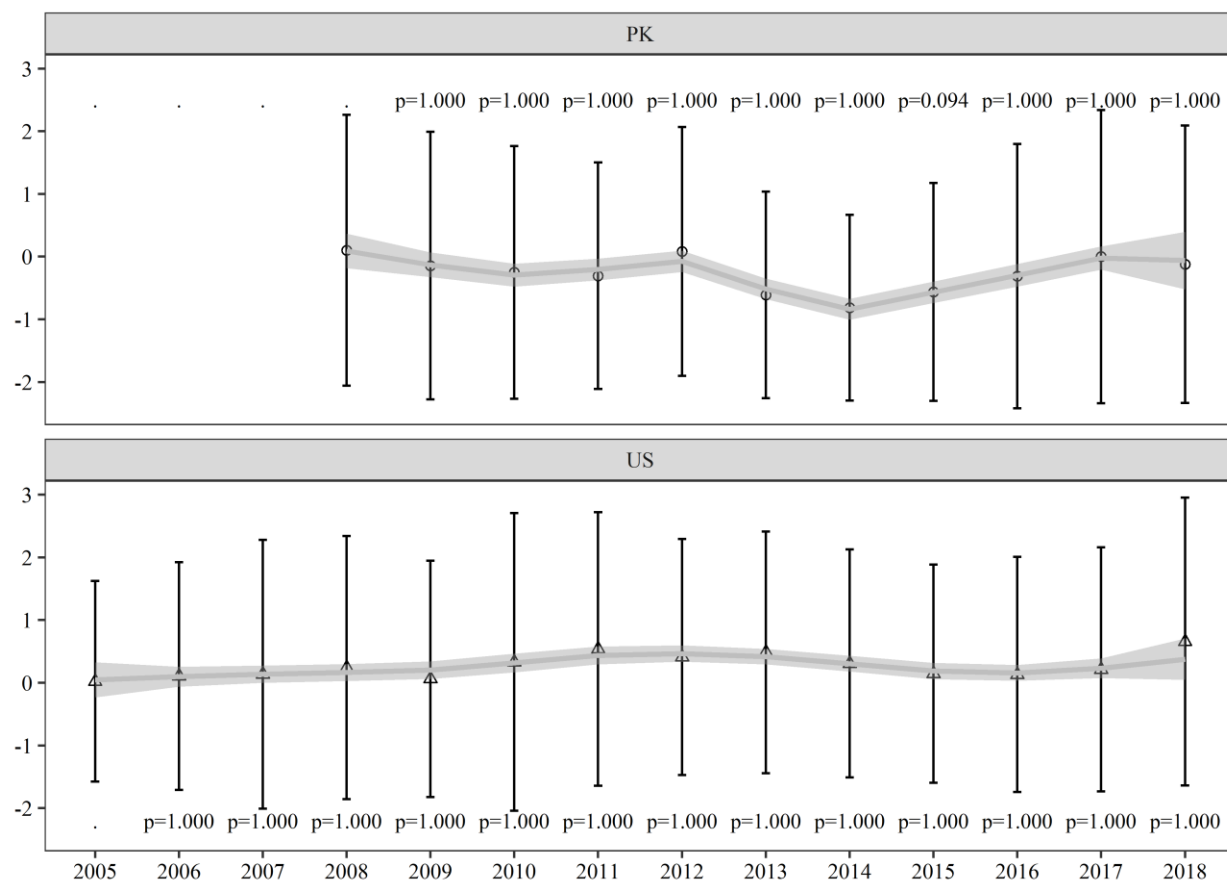
Figure 36 Chronological trend of technology blogs on dimension 5 Narrative Focus (mean dimension scores and standard deviation as error bars)

Note: *P* values are based on post-hoc Dunn's test comparing each year to the first year.

Appendix II Linguistic Features

Stance related lexico-grammatical features from new MD analysis (Biber, 2006, p. 92–93).

1. **Modal and semi-modal verbs** (See LGSWE, pp. 483ff.)
   a. possibility/permission/ability: can, could, may, might
   b. necessity/obligation: must, should, (had) better, have to, got to, ought to
   c. prediction/volition: will, would, shall, be going to

2. **Stance adverbs** (See LGSWE, pp. 557–558; 853–874)
   a. **Epistemic:**
      Certainty: actually, always, certainly, definitely, indeed, inevitably, in fact, never, of course, obviously, really, undoubtedly, without doubt, no doubt
      Likelihood: apparently, evidently, kind of, in most cases/instances, perhaps, possibly, predictably, probably, roughly, sort of, maybe
      Attitude: amazingly, astonishingly, conveniently, curiously, hopefully, even worse, fortunately, importantly, ironically, rightly, sadly, surprisingly, unfortunately
      Style: according to, confidentially, frankly, generally, honestly, mainly, technically, truthfully, typically, reportedly, primarily, usually

3. **Complement clauses controlled by stance verbs, adjectives, or nouns**
   a. **Stance complement clauses controlled by verbs**
      Stance verb + that-clause (See LGSWE, pp. 661–670)
      – **Epistemic verbs:**
      Certainty: conclude, demonstrate, determine, discover, find, know, learn, mean, notice, observe, prove, realize, recognize, remember, see, show, understand
      Likelihood: assume, believe, doubt, gather, guess, hypothesize, imagine, predict, presuppose, presume, reckon, seem, speculate, suppose, suspect, think
      – **Attitude verbs:** agree, anticipate, complain, concede, ensure, expect, fear, feel, forget, hope, mind, prefer, pretend, require, wish, worry
      – **Speech act and other communication verbs:** announce, argue, assert, claim, contend, declare, emphasize, explain, imply, insist, mention, promise, propose, recommend, remark, respond, say, state, suggest, tell
      Stance verb + to-clause (See LGSWE, pp. 693–715)
      – **Probability (likelihood) verbs:** appear, happen, seem, tend
      – **Cognition/perception verbs (likelihood):** assume, believe, consider, expect, find, forget, imagine, judge, know, learn, presume, pretend, remember, suppose
      – **Desire/intention/decision verbs:** agree, choose, decide, hate, hesitate, hope, intend, like, love, mean, need, plan, prefer, prepare, refuse, want, wish
      – **Verbs of causation/modality/effort:** allow, attempt, enable, encourage, fail, help, instruct, manage, oblige, order, permit, persuade, prompt, require, seek, try

– **Speech act and other communication verbs:** ask, claim, invite, promise, remind, request, be said, show, teach, tell, urge, warn

  b. Stance complement clauses controlled by adjectives

  Stance adjective + that-clause (See LGSWE, pp. 671–674; many of these occur with extraposed constructions)

  – **Epistemic adjectives:**

  Certainty: apparent, certain, clear, confident, convinced, correct, evident, false, impossible, inevitable, obvious, positive, right, sure, true, well-known

  Likelihood (extraposed): doubtful, likely, possible, probable, unlikely

  – **Attitude/emotion adjectives:** afraid, amazed, aware, concerned, disappointed, encouraged, glad, happy, hopeful, pleased, shocked, surprised, worried

  – **Evaluation adjectives:** amazing, appropriate, conceivable, crucial, essential, fortunate, imperative, inconceivable, incredible, interesting, lucky, necessary, nice, noteworthy, odd, ridiculous, strange, surprising, unacceptable, unfortunate

  Stance adjective + to-clause (See LGSWE, pp. 716–721; many of these occur with extraposed constructions)

  – **Epistemic (certainty/likelihood) adjectives:** apt, certain, due, guaranteed, liable, likely,

  prone, unlikely, sure

  – **Attitude/emotion adjectives:** afraid, ashamed, disappointed, embarrassed, glad, happy, pleased, proud, puzzled, relieved, sorry, surprised, worried

  – **Evaluation adjectives:** (in)appropriate, bad/worse, good/better/best, convenient, essential, important, interesting, necessary, nice, reasonable, silly, smart, stupid, surprising, useful, useless, unreasonable, wise, wrong

  – **Ability or willingness adjectives:** (un)able, anxious, careful, determined, eager, eligible, hesitant, inclined, obliged, prepared, ready, reluctant, (un)willing

  – **Ease or difficulty adjectives:** difficult, easier, easy, hard, (im)possible, tough

  c. Stance complement clauses controlled by nouns

  Stance noun + that-clause (See LGSWE, pp. 648–651)

  – **Epistemic nouns:**

  Certainty: assertion, conclusion, conviction, discovery, doubt, fact, knowledge, observation, principle, realization, result, statement

  Likelihood: assumption, belief, claim, contention, feeling, hypothesis, idea, implication, impression, notion, opinion, possibility, presumption, suggestion

  – **Attitude/perspective nouns:** grounds, hope, reason, view, thought

  – **Communication (non-factual) nouns:** comment, news, proposal, proposition, remark, report, requirement

  Stance noun + to-clause (See LGSWE, pp. 652–653)

agreement, decision, desire, failure, inclination, intention, obligation, opportunity, plan, promise, proposal, reluctance, responsibility, right, tendency, threat, wish, willingness

4. **Semantic classes of verbs** (Biber et al., 1999, pp. 361–371) adopted from Biber (2006, p. 246–247)

   a. **Activity:** 'primarily denote actions and events that could be associated with choice, and so take a subject with the semantic role of an agent' (LGSWE, pp. 361–362, 367–368, 370): buy, make, get, go, give, take, come, use, leave, show, try, work, move, follow, put, pay, bring, meet, play, run, hold, turn, send, sit, wait, walk, carry, lose, eat, watch, reach, add, produce, provide, pick, wear, open, win, catch, pass, shake, smile, stare, sell, spend, apply, form, obtain, arrange, beat, check, cover, divide, earn, extend, fix, hang, join, lie, obtain, pull, repeat, receive, save, share, smile, throw, visit, accompany, acquire, advance, behave, borrow, burn, clean, climb, combine, control, defend, deliver, dig, encounter, engage, evercise, expand, explore, reduce

   Phrasal activity verbs (LGSWE, pp. 409–410): come along, come on, come over, get out, get up, go ahead, go off, sit down, shut up, sit up, stand up, carry out, get back, get in, get off, look up, make up, pick up, put on, set up, take off, take on, take over, take up

   b. **Mental verbs:** 'denote a wide range of activities and states experienced by humans; they do not involve physical action and do not necessarily entail volition. Their subject often has a semantic role of a recipient' (LGSWE, pp. 362–363, 368–369, 370). Mental verbs include cognitive meanings (e.g., think, know), emotional meanings expressing various attitudes and desires (e.g., love, want), perception (e.g., see, taste), and receipt of communication (e.g., read, hear): see, know, think, find, want, mean, need, feel, like, hear, remember, believe, read, consider, suppose, listen, love, wonder, understand, expect, hope, assume, determine, agree, bear, care, choose, compare, decide, discover, doubt, enjoy, examine, face, forget, hate, identify, imagine, intend, learn, mind, miss, notice, plan, prefer, prove, realize, recall, recognize, regard, suffer, wish, worry, accept, afford, appreciate, approve, assess, blame, bother, calculate, conclude, celebrate, confirm, count, dare, deserve, detect, dismiss, distinguish, experience, fear, forgive, guess, ignore, impress, interpret, judge, justify, observe, perceive, predict, pretend, reckon, remind, satisfy, solve, study, suspect, trust

   c. **Communication verbs:** 'a special subcategory of activity verbs that involve communication activities (speaking, writing)' (LGSWE, pp. 362, 368, 370): say, tell, call, ask, write, talk, speak, thank, descibe, claim, offer, admit, announce, answer, argue, deny, discuss, encourage, explain, express, insist, mention, offer, propose, quote, reply, shout, sign, sing, state, teach, warn, accuse, acknowledge, address, advise, appeal, assure, challenge, complain, consult, convince, declare,

demand, emphasize, excuse, inform, invite, persuade, phone, pray, promise, question, recommend, remark, respond, specify, swear, threaten, urge, welcome, whisper, suggest

d. **Existence or relationship verbs:** These verbs 'report a state that exists between entities. Some of the most common verbs of existence or relationship are copular verbs' (e.g., be, seem, appear). Other verbs in this category 'report a particular state of existence (e.g., exist, live, stay) or a particular relationship between entities (e.g., contain, include, involve, represent)' (LGSWE, pp. 364, 369, 370–371): seem, look, stand, stay, live, appear, include, involve, contain, exist, indicate, concern, constitute, define, derive, illustrate, imply, lack, owe, own, possess, suit, vary, deserve, fit, matter, reflect, relate, remain, reveal, sound, tend, represent

e. **Occurrence verbs:** 'report events (typically physical events) that occur apart from any volitional activity. Often their subject has the semantic affected role' (LGSWE, pp. 364, 369, 370): become, happen, change, die, grow, develop, arise, emerge, fall, increase, last, rise, disappear, flow, shine, sink, slip, occur

f. **Facilitation or causation verbs:** 'indicate that some person or inanimate entity brings about a new state of affairs. These verbs often occur together with a nominalized direct object or complement clause following the verb phrase, which reports the action that was facilitated' (LGSWE, pp. 363, 369, 370): help, let, allow, affect, cause, enable, ensure, force, prevent, assist, guarantee, influence, permit, require

g. **Aspectual verbs:** 'characterize the stage of progress of some other event or activity, typically reported in a complement clause following the verb phrase' (LGSWE, pp. 364, 369, 371): start, keep, stop, begin, complete, end, finish, cease, continue

5. **Semantic classes of nouns adopted from Biber (2006, pp. 248–250)**

a. **Animate: humans or animals.** family, guy, individual, kid, man, manager, member, parent, teacher, child, people, person, student, woman, animal, applicant, author, baby, boy, client, consumer, critic, customer, doctor, employee, employer, father, female, friend, girl, god, historian, husband, American, Indian, instructor, king, leader, male, mother, owner, president, professor, researcher, scholar, speaker, species, supplier, undergraduate, user, wife, worker, writer, accountant, adult, adviser, agent, aide, ancestor, anthropologist, archaeologist, artist, artiste, assistant, associate, attorney, audience, auditor, bachelor, bird, boss, brother, buddha, buyer, candidate, cat, citizen, colleague, collector, competitor, counselor, daughter, deer, defendant, designer, developer, director, dog, driver, economist, engineer, executive, expert, farmer, feminist, freshman, eologist, hero, host, hunter, immigrant, infant, investor, jew, judge, lady, lawyer, learner, listener, maker, manufacturer, miller, minister, mom, monitor, monkey, neighbor, observer,

officer, official, participant, partner, patient, personnel, peer, physician, plaintiff, player, poet, police, processor, professional, provider, psychologist, resident, respondent, schizophrenic, scientist, secretary, server, shareholder, sikh, sister, slave, son, spouse, supervisor, theorist, tourist, victim, faculty, dean, engineer, reader, couple, graduate

b. **Cognitive: mental/cognitive processes or perceptions.** analysis, decision, experience, assessment, calculation, conclusion, consequence, consideration, evaluation, examination, expectation, observation, recognition, relation, understanding, hypothesis, ability, assumption, attention, attitude, belief, concentration, concern, consciousness, concept, fact, idea, knowledge, look, need, reason, sense, view, theory, desire, emotion, feeling, judgement, memory, notion, opinion, perception, perspective, possibility, probability, responsibility, thought

c. **Concrete: inanimate objects that can be touched.** tank, stick, target, strata, telephone, string, telescope, sugar, ticket, syllabus, tip, salt, tissue, screen, tooth, sculpture, sphere, seawater, spot, ship, steam, silica, steel, slide, stem, snow, sodium, mud, solid, mushroom, gift, muscle, glacier, tube, gun, nail, handbook, newspaper, handout, node, instrument, notice, knot, novel, lava, page, food, transcript, leg, eye, lemon, brain, magazine, device, magnet, oak, manual, package, marker, peak, match, pen, metal, pencil, block, pie, board, pipe, heart, load, paper, transistor, modem, book, mole, case, motor, computer, mound, dollar, mouth, hand, movie, flower, object, foot, table, frame, water, vessel, arm, visa, bar, grain, bed, hair, body, head, box, ice, car, item, card, journal, chain, key, chair, window, vehicle, leaf, copy, machine, document, mail, door, map, dot, phone, drug, picture, truck, piece, tape, note, liquid, wire, equipment, wood, fiber, plant, fig, resistor, film, sand, file, score, seat, belt, sediment, boat, seed, bone, soil, bubble, bud, water, bulb, portrait, bulletin, step, shell, stone, cake, tree, camera, video, face, wall, acid, alcohol, cap, aluminium, clay, artifact, clock, rain, clothing, asteroid, club, automobile, comet, award, sheet, bag, branch, ball, copper, banana, counter, band, cover, wheel, crop, drop, crystal, basin, cylinder, bell, desk, dinner, pole, button, pot, disk, pottery, drain, radio, drink, reactor, drawing, retina, dust, ridge, edge, ring, engine, ripple, plate, game, cent, post, envelope, rock, filter, root, finger, slope, fish, space, fruit, statue, furniture, textbook, gap, tool, gate, train, gel, deposit, chart, mixture

d. **Technical/concrete: tangible objects that are not normally perceived and/or cannot normally be touched.** cell, unit, gene, wave, ion, bacteria, electron, chromosome, element, cloud, sample, isotope, schedule, neuron, software, nuclei, solution, nucleus, atom, ray, margin, virus, mark, hydrogen, mineral, internet, molecule, mineral, organism, message, oxygen, paragraph, particle, sentence, play, star, poem, thesis, proton, unit, web, layer, center, matter, chapter, square,

data, circle, equation, compound, exam, letter, bill, page, component, statement, diagram, word, dna, angle, fire, carbon, formula, graph, iron, lead, jury, light, list

e. **Place: places, areas, or objects in a fixed location.** apartment, interior, bathroom, moon, bay, museum, bench, neighborhood, bookstore, opposite, border, orbit, cave, orbital, continent, outside, delta, parallel, desert, passage, estuary, pool, factory, prison, farm, restaurant, forest, sector, habitat, shaft, hell, shop, hemisphere, southwest, hill, station, hole, territory, horizon, road, bottom, store, boundary, stream, building, top, campus, valley, canyon, village, coast, city, county, country, court, earth, front, environment, district, field, floor, market, lake, office, land, organization, lecture, place, left, room, library, area, location, class, middle, classroom, mountain, ground, north, hall, ocean, park, planet, property, region, residence, river

f. **Quantity: nouns specifying a quantity, amount, or duration.** cycle, rate, date, second, frequency, section, future, semester, half, temperature, height, today, number, amount, week, age, day, century, part, energy, lot, heat, term, hour, time, month, mile, period, moment, morning, volume, per, weekend, percentage, weight, portion, minute, quantity, percent, quarter, length, ratio, measure, summer, meter, volt, voltage

g. **Group/institution: nouns that denote a group or institution.** airline, institute, colony, bank, flight, church, hotel, firm, hospital, household, college, institution, house, lab, laboratory, community, company, government, university, school, home, congress, committee

h. **Abstract/process: intangible, abstract concepts or processes.** action, activity, application, argument, development, education, effect, function, method, research, result, process, accounting, achievement, addition, administration, approach, arrangement, assignment, competition, construction, consumption, contribution, counseling, criticism, definition, discrimination, description, discussion, distribution, division, eruption, evolution, exchange, exercise, experiment, explanation, expression, formation, generation, graduation, management, marketing, marriage, mechanism, meeting, operation, orientation, performance, practice, presentation, procedure, production, progress, reaction, registration, regulation, revolution, selection, session, strategy, teaching, technique, tradition, training, transition, treatment, trial, act, agreement, attempt, attendance, birth, break, claim, comment, comparison, conflict, deal, death, debate, demand, answer, control, flow, service, work, test, use, war, change, question, study, talk, task, trade, transfer, admission, design, detail, dimension, direction, disorder, diversity, economy, emergency, emphasis, employment, equilibrium, equity, error, expense, facility, failure, fallacy, feature, format, freedom, fun, gender, goal, grammar, health, heat, help, identity, image, impact, importance, influence, input, labor, leadership, link, manner, math, matrix, meaning, music, network, objective,

opportunity, option, origin, output, past, pattern, phase, philosophy, plan, potential, prerequisite, presence, principle, success, profile, profit, proposal, psychology, quality, quiz, race, reality, religion, resource, respect, rest, return, risk, substance, scene, security, series, set, setting, sex, shape, share, show, sign, signal, sort, sound, spring, stage, standard, start, stimulus, strength, stress, style, support, survey, symbol, topic, track, trait, trouble, truth, variation, variety, velocity, version, whole, action, account, condition, culture, end, factor, grade, interest, issue, job, kind, language, law, level, life, model, name, nature, order, policy, position, power, pressure, relationship, requirement, role, rule, science, side, situation, skill, source, structure, subject, type, information, right, state, system, value, way, address, absence, advantage, aid, alternative, aspect, authority, axis, background, balance, base, beginning, benefit, bias, bond, capital, care, career, cause, characteristic, charge, check, choice, circuit, circumstance, climate, code, color, column, combination, complex, connection, constant, constraint, contact, content, contract, context, contrast, crime, criteria, cross, current, curriculum, curve, debt, density

Appendix III Spelling Corrections

Following is the list of spelling corrections that was generated during the editing and review process of the texts selected from online registers of both varieties. (Note: The list also contains regular expressions' notation.)

Original	Corrected	Original	Corrected
adress	address	hedntve	he'dn't've
aint	ain't	hedve	he'd've
airprt	airport	hes	he's
amnt	amn't	hesnt	he'sn't
apointment	appointment	hows	how's
arent	aren't	hru	how are you
aswell	as well	hv	have
b	be	Idnt	I'dn't
becuase	because	Ill	I'll
cant	can't	im	I'm
chckd	checked	Im	I'm
cn	can	isnt	isn't
couldnt	couldn't	itd	it'd
couldntve	couldn't've	itdnt	it'dn't
couldve	could've	itdntve	it'dn't've
coz	because	itdve	it'd've
didnt	didn't	itll	it'll
didnt	didn't	itsnt	it'sn't
doesnt	doesn't	Ive	I've
does'nt	doesn't	i've	I've
dont	don't	Ivent	I'ven't
Dont	Don't	lets	let's
dosnt	doesn't	maam	ma'am
earliar	earlier	meanz	means
exect	exact	mentnd	mentioned
gud	good	mightnt	mightn't
hadnt	hadn't	mightntve	mightn't've
hadntve	hadn't've	mightve	might've
hasnt	hasn't	msg	message
hav	have	mustnt	mustn't
havent	haven't	mustntve	mustn't've
havent	haven't	mustve	must've
hed	he'd	n	and
hednt	he'dn't	nd	and

Original	Corrected	Original	Corrected
neednt	needn't	somethingdnt	something'dn't
noone	no one	somethingdntve	something'dn't've
nooooo	no	somethingdve	something'd've
nwadays	now a days	somethingll	something'll
oclock	o'clock	somethings	something's
ofcourse	of course	so{2,}	so
oughtnt	oughtn't	thanx	thanks
pls	please	thatd	that'd
plz	please	thatll	that'll
ppl	people	thats	that's
ppr	proper	therednt	there'dn't
r	are	theredntve	there'dn't've
recieve	receive	theredve	there'd've
recieved	received	therere	there're
rply	reply	theres	there's
rt	right	theyd	they'd
senier	senior	theydnt	they'dn't
shant	shan't	theydntve	they'dn't've
shednt	she'dn't	theydve	they'd've
shedntve	she'dn't've	theydvent	they'd'ven't
shes	she's	theyll	they'll
shesnt	she'sn't	theyllntve	they'lln't've
shouldnt	shouldn't	theyllvent	they'll'ven't
shouldntve	shouldn't've	theyre	they're
shouldve	should've	theyve	they've
shud	should	theyvent	they'ven't
somebodyd	somebody'd	u	you
somebodydnt	somebody'dn't	upto	up to
somebodydntve	somebody'dn't've	ur	your
somebodydve	somebody'd've	urself	yourself
somebodyll	somebody'll	u've	you've
somebodys	somebody's	wasnt	wasn't
someoned	someone'd	wat	what
someonednt	someone'dn't	wats	what's
someonedntve	someone'dn't've	wednt	we'dn't
someonedve	someone'd've	wedntve	we'dn't've
someonell	someone'll	wedve	we'd've
someones	someone's	wel	well
somethingd	something'd		

Original	Corrected	Original	Corrected
wellntve	we'lln't've	atleast	at least
werent	weren't	cuz	because
weve	we've	thankyou	thank you
whatd	what'd	everythin	everything
whatll	what'll	becos	because
whatre	what're	lol	laugh out loud
whats	what's	wayyy	way
whatve	what've	reallllyy	really
whens	when's	Okie	Ok
whered	where'd	afta	after
wheres	where's	soo{2,}n	soon
whereve	where've	pikchaz	pictures
whod	who'd	wha{2,}t{2,}	what
whodve	who'd've	alone{2,}	alone
wholl	who'll	thingie	thing
whos	who's	How{2,}	How
whove	who've	how{2,}	how
whyd	why'd	werrd	weird
whyll	why'll	pretty{2,}	pretty
whyre	why're	lookie	look
whys	why's	shutup	shut up
wil	will	mileu	milieu
wont	won't	inspite	in spite
wontve	won't've	Btw	By the way
wory	worry	nevermind	never mind
wouldnt	wouldn't	readin	reading
wouldntve	wouldn't've	youl	you'll
wouldve	would've	ure	you're
You R	You are	curiousity	curiosity
youd	you'd	calss	class
youdve	you'd've	Ofcourse	Of course
youll	you'll	lierature	literature
youre	you're	intelleigence	intelligence
yourent	you'ren't	poractising	practising
youve	you've	dig{2,}	dig
youvent	you'ven't	agravates	aggravates
btw	by the way	behlaf	behalf
U	You	thier	their
abt	about	ecause	because

Original	Corrected	Original	Corrected
untill	until	juzz	just
undersatnd	understand	acc to	according to
atlest	at least	smthng	something
everthing	everything	avlbl	available
Hypocrit	Hypocrite	weneva	whenever
chosing	chossing	mstly	mostly
infact	in fact	espclly	especially
Cuz	Because	espcly	especially
supremecy	supremacy	ma{2,}n{2,}	man
freakin	freaking	acheivements	achievements
ive	I've	samestar	semester
puriest	purest	firstime	first time
y	why	realy	really
lmao	Laughing my ass off	theyr	they're
		qulaity	quality
srry	sorry	nevr	never
love{2,}	love	bcuz	because
mermories	memories	measurin	measuring
makin	making	alot	a lot
persuing	pursuing	confuct	conduct
tc	take care	lolz{2,}	laugh out loud
xoxo	hugs and kisses	tht	that
way{2,}	way	insted	instead
b/c	because	too{2,}	too
joinign	joining	thankooo	thank you
wiht	with	urs	yours
So{2,}	So	mins	minutes
heartidly	heartedly	tryed	tried
choce	choice	Thankyou	Thank you
proably	probably	peviously	previously
crickter	cricketer	yumm{2,}y{2,}	yummy
suwhere	somewhere	a lil	a little
obcession	obsession	fav	favourite
verifictaion	verification	thankz	thanks
wid	with	must ave	must've
wether	whether	bee{2,}n	been
pitty	pity	Cute{2,}	Cute
fav	favourite	thier	their
sista	sister	lolx	laugh out loud

Original	Corrected	Original	Corrected
swee{2,}t	sweet	beutiully	beautifully
nt	not	writen	written
dilmma	dilemma	kindda	kinda
realtionshps	relationships	doent	doesn't
preffer	prefer	stalkd	stalked
answerd	answered	persume	presume
thanko{2,}	thank you	thatz	that's
commenitng	commenting	behavr	behaviour
lucky{2,}	lucky	displayng	displaying
excercise	exercise	reall{2,}y	really
em	I'm	buny	bunny
yeh	yeah	noo{2,}	no
must ve	must've	understnad	understand
loadsheddin	loadshedding	governement	government
againt	against	Damnit	Dammit
porblems	problems	kno	know
hoep	hope	technqiues	techniques
tee{2,}ny	teeny	readin	reading
rox	rocks	writin	writing
cux	because	xpired	expired
Thanko{2,}	Thank you	reson	reason
teh	the	dimentia	dementia
probbaly	probably	enuff	enough
Thankz	thanks	superflous	superfluous
anaylzed	analyzed	BTW	By the way
pple	people	fullfil	fulfil
appearnces	appearances	shellter	shelter
belive	believe	illeterate	illiterate
considrng	considering	loo{2,}king	looking
thru	through	alterative	alternative
dnt	don't	folowing	following
lairz	liars	recipie	recipe
commentr	commenter	definately	definitely
realzed	realized	goo{2,}d{2,}	good
bt	but	recipies	recipes
messcre	massacre	rnt	aren't
spontanious	spontaneous	defintily	definitely
jus	just	sigh{2,}	sigh
lov{2,}e	love	Thankx	Thanks

Original	Corrected	Original	Corrected
favourite{2,}	favourite	yumm{2,}iest	yummiest
corriander	coriander	wo{2,}w	wow
greatful	grateful	loo{2,}ks	looks
thankx	thanks	yum{2,}	yum
yumm{2,}y{2,}	yummy	hvnt	haven't
qn	question	uv	you've
combi	combination	wud	would
wth	with	anywa	anyway
loo{2,}k{2,}s{2,}	looks	onez	ones
Awesoem	Awesome	chk	check
critisism	criticism	selecter	selector
more{2,}	more	thng	thing
Thanks{2,}	Thanks	dt	that
especally	especially	wn	when
writng	writing	knw	know
thmx	thanks	iz	is
thnx	thanks	wt	what
combinatio	combination	shuold	should
Lovley	Lovely	beatiful	beautiful
Yumm{2,}y	Yummy	lovey	lovely
vegitables	vegetables	meauremants	measurements
lentals	lentils	worng	wrong
thougth	thought	see{2,}	see
clearify	clarify	yumz{2,}	yum
temprature	temperature	defenitely	definitely
Plz	Please	benifits	benefits
heatin	heating	inisted	insisted
hapens	happens	cannt	cannot
lowr	lower	spoiing	spoiling
pefect	perfect	Pls	Please
hu{2,}ge{2,}	huge	Contect	Contact
doin	doing	havng	having
thnkng	thinking	frnd	friend
thanks{2,}	thanks	Congrats{2,}	Congrats
wich	which	wat{2,}	what
okk{2,}	ok	acros	across
v	we	waz	was
instaed	instead	shuld	should
dat	that	giv	give

Original	Corrected	Original	Corrected
thnk	think	changiing	changing
considerd	considered	honroed	honored
lvoe	love	exapmple	example
happ{2,}y	happy	exaggrate	exaggerate
Nyce	Nice	humnaity	humanity
awsome	awesome	belived	believed
i m	I am	mdae	made
greatt{2,}	great	continously	continuously
grt	great	litreature	literature
artical	article	campanion	companion
dn't	don't	specialy	specially
commisserate	commiserate	countaries	countries
belives	believes	disasterious	disastrous
raltions	relations	becoz	because
kickin	kicking	ahve	have
immediatly	immediately	outa	out of
graduly	gradually	jackin	jacking
tregic	tragic	guyz	guys
happending	happening	Atleast	At least
henious	heinous	Juz	Just
congoratss	congrats	Dnt	Don't
xx	kisses	agianst	against
negetivities	negativities	wll	will
thankyou	thank you	rememberance	remembrance
intrest	interest	accoding	according
jst	just	propsperity	prosperity
intresting	interesting	offensiv	offensive
droping	dropping	re{2,}ally	really
toliet	toilet	thinkin	thinking
expirience	experience	missin	missing
BTY	Back to you	givin	giving
luvd	loved	irrelvant	irrelevant
ds	this	sistah	sister
foriegn	foreign	definately	definitely
Foriegn	Foreign	Every1	Everyone
fovour	favour	Thts	That's
shoud	should	thm	them
thorugh	through	lotta	lot of
unneccessary	unnecessary	luv	love

Original	Corrected	Original	Corrected
wana	wanna	thingy	thing
masculline	masculine	ver{2,}y	very
intereting	interesting	channles	channels
true{2,}	true	diff	different
dts	that's	nyc	nice
oppurtunities	opportunities	forhead	forehead
Hypocrixy	Hypocrisy	very{2,}	very
Lol	Laugh out loud	misturizer	moisturizer
Well{2,}	Well	wil	will
well{2,}	well	everobody	everybody
Offcourse	Of course	stil	still
kinndda	kinda	definetly	definitely
hypocrtical	hypocritical	writting	writing
appologizing	apologizing	widly	widely
wd	with	apper	appear
delimma	dilemma	admorable	admirable
chnage	change	pleasent	pleasant
amazingg	amazing	definatly	definitely
super{2,}b	superb	appers	appears
ar	are	esthetic	aesthetic
bcaz	because	canves	canvas
goo{2,}d	good	extremly	extremely
fab	fabulous	complecated	complicated
everythi	everything	inorder	in order
stiched	stitched	lo{2,}ts of	lots of
followin	following	everytime	every time
thot	thought	writtings	writings
imo	in my opinion	usualy	usually
gna	gonna	homourously	humorously
gona	gonna	all{2,}	all
bleessed	blessed	fa{2,}r	far
differenece	difference	repitition	repetition
bloglovin	blogloving	actaully	actually
Loveing	Loving	burried	buried
glamarous	glamorous	insignicant	insignificant
this{2,}	this	politicans	politicians
Lovely{2,}	Lovely	thousnads	thousands
chek	check	presure	pressure
provdes	provides	congratz	congrats

Original	Corrected
achivement	achievement
kongrats	congrats
unbelieveably	unbelievably
exaclty	exactly
equiped	equipped
personel	personnel
infront	in front
somthing	something
registeration	registration
infornt	in front
benifit	benefit
intendng	intending
seperation	separation
computor	computer
existance	existence
vengence	vengeance
Doesnt	Doesn't
millitary	military
LOL	Laugh out loud
Thaanks	Thanks
youu	you
predertimed	predetermined
frm	from
wht	what
pleaze	please
metnion	mention
destablising	destabilising
bln	billion
rallys	rallies
discrimintation	discrimination
guaratee	guarantee
somoehow	somehow
unpleasent	unpleasant
cheif	chief
goverment	government
servents	servants
sucessful	successful
vioces	voices
totatly	totally

Original	Corrected
fasion	fashion
minstre	minister
siad	said
villify	vilify
heavey	heavy
dsclose	disclose
provential	provincial
indifferance	indifference
terraine	terrain
straving	starving
tentions	tensions
Anywayz	Anyways
bureauratic	bureaucratic
civlian	civillian
nee{2,}d	need
towrds	towards
awerness	awareness
alwyas	always
fantasitc	fantastic
compatable	compatible
buisness	business
socieity	society
desperatly	desperately
fuding	feuding
oppurtuinty	opportunity
perhpas	perhaps
leveleled	levelled
neccessarily	necessarily
inexpressable	inexpressible
rulz	rules
permenant	permanent
thts	that's
cumz	comes
ryt	right
Howcome	How come
lo{2,}v{2,}e{2,}	love
shd	should
plz{2,}	please
ambassader	ambassador

Original	Corrected	Original	Corrected
thnxx	thanks	monrchy	monarchy
awsum	awesome	Enterpreneurs	Entrepreneurs
stunningg	stunning	opressed	oppressed
Help{2,}	Help	traininig	training
thr	there	terroists	terrorists
leme know	lemme know	unusally	unusually
askd	asked	embarassing	embarrassing
sugest	suggest	miniscule	minuscule
thougt	thought	prejuidice	prejudice
destabilze	destabilize	falterring	faltering
alraedy	already	slauthered	slaughtered
propoganda	propaganda	faught	fought
includs	includes	predominently	predominantly
parrotting	parroting	cassually	casually
traversty	travesty	prfessional	professional
targetting	targeting	resposibility	responsibility
Everytime	Every time	illegaly	illegally
proscutors	prosecutors	bsic	basic
samposium	symposium	puting	putting
imagin	imagine	pursuade	persuade
heartly	heartily	diciples	disciples
criticzing	criticizing	whenevr	whenever
rigt	right	intelect	intellect
sufferring	suffering	depair	despair
becaaz	because	defination	definition
theiving	thieving	lenght	length
wehther	whether	intrested	interested
bcoz	because	propely	properly
wthout	without	missionery	missionary
sum1	someone	rememebr	remember
sumthing	something	intentially	intentionally
disuaded	dissuaded	withrew	withdrew
lanched	launched	gentl	gentle
incompetant	incompetent	dacades	decades
paracsitic	parasitic	govenment	government
de-stabilze	de-stabilize	frist	first
grevious	grievous	remmeber	remember
practsing	practising	ruuner	runner
religon	religion	avaiable	available

Original	Corrected	Original	Corrected
commisioner	commissioner	somethin	something
memebr	member	whast	what's
trafic	traffic	whn	when
sugesting	suggesting	stfu	shut the fuck up
discussin	discussing	lol\.	Laugh out loud
listning	listening	Lol\.	Laugh out loud
irrresponsible	irresponsible	nominies	nominees
seeign	seeing	satisfatory	satisfactory
higer	higher	suiteable	suitable
realtive	relative	minut	minute
achive	achieve	plx	please
admant	adamant	num	number
thast	that's	canvert	convert
misrable	miserable	villege	village
tryign	trying	personaly	personally
mintues	minutes	Personaly	Personally
gving	giving	improvment	improvement
sxtares	stares	usless	useless
inferority	inferiority	Insitutes	Insitutes
harased	harassed	Catagorized	Categorized
regualr	regular	subbcribers	subscribers
whne	when	un-necceserily	un-necessarily
vigilantee	vigilante	intouch	in touch
milion	million	gues	guess
dolar	dollar	Reviewd	Reviewed
freasks	freaks	lyk	like
repreiselss	reprisals	afforable	affordable
festivel	festival	messange	messenger
himselve	himself	spekaer	speaker
gogin	going	ov	of
sacfred	sacrificed	loudspkr	loudspeaker
descions	decisions	prblm	problems
westren	western	facng	facing
oppurtunity	opportunity	woried	worried
ptential	potential	sudnly	suddenly
potiental	potential	bafoons	buffoons
furustrsating	frustrating	shoudn't	shouldn't
crieteria	criteria	priior	prior
managment	management	infrom	inform

Original	Corrected	Original	Corrected
supermacy	supremacy	brothr	brother
releasee	released	phataaatic	pathetic
personaly	personally	fr	for
useable	usable	expercnd	experienced
satisifed	satisfied	marketng	marketing
performace	performance	startdgy	strategy
opreatar	operator	Addvertising	Advertising
Yaa{,2}ww{,2}nn{,2}	Yawn	servise	service
genious	genius	poo{2,}	poor
whr	where	comfirmed	confirmed
malacious	malicious	subscrier	subscriber
craxy	crazy	unimited	unlimited
transfr	transfer	penitated	penetrated
txts	texts	discrease	decrease
yu	you	recomendation	recommendation
lyng	lying	frenchise	franchise
Yu	You	seriel	serial
filteration	filtration	nunber	number
bycot	boycott	nmbers	numbers
bleive	believe	contect	contact
geninue	genuine	numbr	number
appourtunity	opportunity	liers	liars
contnsly	continuously	pacage	package
interfrng	interfering	experince	experience
dunt	don't	Budle	Bundle
agaqin	again	sepret	separate
watever	whatever	destoryed	destroyed
grivencies	grievances	opeartor	operator
So{2,}r{2,}y	Sorry	ragardless	regardless
authenticty	authenticity	mangement	management
befre	before	infomative	informative
ture	true	nuff	enough
majorty	majority	gd	good
Beacuse	Because	llike	like
economay	economy	FTW	For the win
astonised	astonished	especailly	especially
restructring	restructuring	everythng	everything
youngst	youngest	miss{2,}ing	missing
spcly	specially	goin	going

Original	Corrected	Original	Corrected
themseslvs	themselves	tbh	to be honest
purcahsing	purchasing	TBH	To be honest
Desposit	Deposit	cmpany	company
Cheq	Cheque	consistancy	consistency
reliabiliy	reliability	config	configuration
unlmited	unlimited	creat	create
provdie	provide	giude	guide
mignight	midnight	wtf	what the fuck
daectivation	deactivation	suppliments	supplements
workin	working	considred	considered
sbscribed	subscribed	everbody	everybody
manoply	monopoly	wor{2,}y{2,}	worry
compitions	competitions	img	image
subscribtion	subscription	intermadiate	intermediate
finanly	finally	noobs	newbies
brilient	brilliant	arround	around
dollors	dollars	checques	cheques
downloadin	downloading	prog	program
appiled	applied	recongnized	recognized
pakeges	packages	Recongnizes	Recognizes
unsub	unsubscribe	webiste	website
connectoin	connection	Bnk	Bank
coustomer	customer	accout	account
Aweso{2,}m{2,}e{2,}	Awesome	soultion	solution
budy	buddy	runing	running
abble	able	uptil	up till
tnk	thank	affilliates	affiliates
autmatically	automatically	tranafers	transfers
flope	flop	loctaion	location
dev	developer	furthur	further
poeple	people	commentry	commentary
tehy	they	monthyl	monthly
abt	about	recvd	received
devs	developers	minimu	minimum
calander	calender	creat	create
bro	brother	compaign	campaign
onyl	only	Noone	No one
aspx	as soon as possible	comapigns	campaigns
		mesage	message

Original	Corrected	Original	Corrected
interection	interaction	rurual	rural
equalvelent	equivalent	interere	interfere
domonted	dominated	Wud	Would
tentamount	tantamount	calaim	claim
Unfortunatly	Unfortunately	possibel	possible
reunificatiom	reunification	possibilites	possibilities
recoverred	recovered	commoness	commonness
econmies	economies	relegion	religion
did't	didn't	induviduals	individuals
neighours	neighbours	Thanku	Thank you
Acccording	According	papmer	pamper
welldone	well done	talkin	talking
awailable	available	Espacially	Especially
Princple	Principle	Deffinately	Definitely
occures	occurs	diffeence	difference
devine	divine	dosn't	doesn't
reigious	religious	thanku	thank you
miracleaus	miracles	Congrate	Congrats
influencials	influentials	sorrounded	surrounded
demoncracies	democracies	whtever	whatever
opionion	opinion	behnd	behind
assesment	assessment	Hollocast	Holocaust
awesum	awesome	billionares	billionaires
canabalism	cannibalism	genrate	generate
weaknesess	weaknesess	reluctent	reluctant
greatfull	grateful	possibe	possible
knowldge	knowledge	Hounarable	Honourable
diirecting	directing	seriouslyy	seriously
thhose	those	pic	picture
moraly	morally	captan	captain
childrens	children	expanditure	expenditure
ordianry	ordinary	incumbet	incumbent
attnetion	attention	conculsion	conclusion
advisor	adviser	initail	initial
veiws	views	infulenced	influenced
businiss	business	forieng	foreign
evey	every	abstard	bastard
replet	replete	rulling	ruling
successed	succeeded	exhorbitant	exorbitant

Original	Corrected
develuatn	devaluation
bllood	blood
learder	leader
devaluting	defaulting
adaquate	adequate
favourate	favourite
distabilize	destabilize
acceptibility	acceptability
intersting	interesting
strenth	strength
Ariticle	Article
evoporated	evaporated
habbits	habits
sattire	satire
corportations	corporations
emphatetic	empathetic
boundarys	boundaries
caliphat	caliphate
Caliphat	Caliphate
concience	conscience
cinemtaography	cinematography
lieklbae	likable
reflrcted	reflected
violance	violence
pedigaries	pedigrees
chanel	channel
exxxxxplosive	explosive
scholors	scholars
Wha{2,}t	What
wha{2,}t	what
convservative	conservative
openion	opinion
blam	blame
parsites	parasites
volture	vulture
scalvangers	scavengers
affluances	affluences
termandous	tremendous
virtu	virtue

Original	Corrected
puppits	puppets
reserch	research
converstation	conversation
langeuage	language
commentors	commenters
toatally	totally
bcz	because
forseen	foreseen
fee{2,}l	feel
diplometic	diplomatic
straigt	straight
frekin	freaking
Definitly	Definitely
knowledgeabe	knowledgeable
circusmstance	circumstance
Cries{2,}	Cries
Omg	Oh my God
Ure	You're
Wud	Would
tbspoon	tablespoon
litterly	literally
photograph{2,}	photograph
evaulate	evaluate
embasy	embassy
avoide	avoid
pasport	passport
Lolz+	Laugh out loud
heared	heard
Nd	And
arrnge	arrange
reqrd	required
wri{2,}tten	written
intranational	international
Bro	Brother
Hes	He's
Shes	She's
textd	texted
nex	next
arraive	arrive

Original	Corrected	Original	Corrected
opoitment	appointment	lols	laugh out loud
You ve	You've	Becouse	Because
I ve	I've	wheather	whether
scholarshipz	scholarships	Amercian	American
We ve	We've	forses	forces
we ve	we've	Clumsly	Clumsily
you ve	you've	Hystericaly	Hysterically
they ve	they've	miseryy	misery
They ve	They've	incidently	incidentally
relted	related	innappropriately	inappropriately
suvject	subject	enfused	infused
deta	data	extravagence	extravagance
workd	worked	riddiculously	ridiculously
crntly	currently	Wa{2,}y	Way
cancelation	cancellation	evidenciary	evidentiary
pofile	profile	Authorties	Authorities
exacptible	acceptable	beneficient	beneficent
appointmnt	appointment	advertize	advertise
aaply	apply	affaires	affairs
canceld	cancelled	benefitted	benefited
statmnt	statement	forgetten	forgotten
cource	course	awesome{2,}	awesome
lang	language	omg	oh my God
curriculm	curriculum	El{2,}evn	Eleven
iniate	initiate	effectees	affectees
Load-shadding	Load-shedding	i,ve	I've
Load-sheding	Load-shedding	yumm{2,}y	yummy
Vampiress	Vampires	ready{2,}	ready
vampiress	vampires	Lo{2,}ng	Long
Mangos	Mangoes	Choclate	Chocolate
mangos	mangoes	collegue	colleague
have'nt	haven't	bday	birthday
doenst	doesn't	Ingrediants	Ingredients
lil	little	tbspn	tablespoon
din't	didn't	itl	it'll
Peoplez	Peoples	Dunt	Don't
yes{2,}	yes	serach	search
DONT	DON'T	homogenous	homogeneous
DINT	DIDN'T	wudnt	wouldn't

Original	Corrected	Original	Corrected
Nah	No	Editon	Edition
nah	no	tbls	tablespoon
frenzily	frenziedly	clarifys	clarifies
scentences	sentences	grms	grams
ye{2,}ll	yell	tblsp	tablespoon
humourous	humorous	urselves	yourselves
tastlessness	tastelessness	Isnt	Isn't
giaganomous	magnanimous	Senstive	Sensitive
fave	favourite	Lovlies	Lovelies
abhore	abhor	lovlies	lovelies
describtion	description	Machiavellism	Machiavellianism
aything	anything	reparians	riparians
importatn	important	riverian	riverine
shleved	shelved	craz{2,}y	crazy
bookhelf	bookshelf	loads{2,}	loads
ddepends	depends	mangrooves	mangroves
conneting	connecting	rea{2,}l{2,}y{2,}	really
relgion	religion	penting	panting
initiatied	initiated	is'nt	isn't
synonamous	synonymous	please{2,}	please
goverenment	government	kee{2,}p	keep
increaes	increases	folowing{2,}	following
disect	dissect	rememeber	remember
disgaree	disagree	s{2,}o{2,}	so
arrving	arriving	simoutaneously	simultaneously
afterall	after all	legendry	legendary
probs	problems	traslate	translate
care{2,}	care	panle	panel
Afterall	After all	soldeir	soldier
Hello{2,}	Hello	wiill	wiill
feautring	featuring	leter	later
Gu{2,}y{2,}s{2,}	Guys	trajedy	tragedy
Also{2,}	Also	dailogues	dialogues
voilently	violently	Inspite	In spite
Similiarly	Similarly	defintely	definitely
seculer	secular	wraping	wrapping
spirtual	spiritual	subscirber	subscriber
Acttually	Actually	authnticity	authenticity
orignal	original		

Original	Corrected
vigiliance	vigilance
clamant	clement
overnite	overnight
dats	that's
catastrophy	catastrophe
widout	without
gallavanting	gallivanting
cudn't	couldn't
shudnt	shouldn't
siigh	sigh
yello{2,}w{2,}	yellow
hve	have
suggstions	suggestions
caliing	calling
housewifes	housewives
belie{2,}ving	believing
bizarro	bizarre
crosssed	crossed
Yes{2,}	Yes
Goo{2,}d	Good
kiddin	kidding
Alot	A lot
Excitment	Excitement
Untiil	Until
nee{2,}ds	needs
upgradable	upgradeable
tempation	temptation
that{2,}	that
fiunded	founded
aligment	alignment
meterological	meteorological
amphiteatre	amphitheatre
tital	title
demonstation	demonstration
fevereshly	feverishly
becaue	because
incomprable	incomparable
scence	scene
upheavel	upheaval

Original	Corrected
Februrary	February
provice	province
Soceity	Society
critcise	criticise
comendable	commendable
encourge	encourage
Fuckin	Fucking
lvl	level
yrs	years
fking	fucking
sittin	sitting
dem	them
sharin	sharing
srsly	seriously
sweepin	sweeping
passprt	passport
i ll	i'll
lookin	looking
frnds	friends
edu	education
simpl	simple
vich	which
momnt	moment
silenc	silence
presidnt	president
masculin	masculine
boys{2,}	boys
Coz	Because
pronouncin	pronouncing
comin	coming
dis	this
da	the
orgs	organisations
compitition	competition
cruntly	currently
terorism	terrorism
blv	believe
mak	make
billoions	billions

Original	Corrected	Original	Corrected
cabl	cable	administraton	administrating
miss{2,}	miss	day{2,}s	days
rtd	retired	benificent	beneficent
gvt	government	lmk	let me know
avl	available	LMK	Let me know
opnd	opened	imnpact	impact
omg{2,}	oh my God	havn't	haven't
4GET	FORGET	crackin	cracking
4get	forget	himslef	himself
RIP	Rest in peace	nothign	nothing
documnts	documents	Definately	Definitely
%age	percentage	IMO	in my opinion
intl	international	Lets	Let's
cousre	course	thx	thanks
befor	before	warnig	warning
precious{2,}	precious	Altho	Although
lovly	lovely	abiliy	ability
omgoodness	oh my goodness	apprhend	apprehend
subscrber	subscriber	pursueing	pursuing
aw{2,}e{2,}	awe	gettin	getting
contactor	contractor	maself	myself
droolin	drooling	owuld	would
singin'	singing	re{2,}a{2,}lly	really
Ho{2,}pe	Hope	WTF	What the fuck
prosperity{2,}	prosperity	IMHO	In my humble opinion
abundacne	abundance		
Kepp	Keep	imho	in my humble opinion
Hell{2,}o	Hello		
definintion	definition	fortunatley	fortunately
hypoctitical	hypocritical	callin	calling
retaleation	retaliation	Bo{2,}r{2,}i{2,}n{2,}g	Boring
rappin	rapping		
guarantee{2,}	guarantee	lo{2,}ng	long
tuff	tough	Anyhoo	Anyhow
str8	straight	fyi	for your information
STR8	STRAIGHT		
thnks	thanks	FYI	For your information
aproach	approach		
atttention	attention	af	as fuck

Original	Corrected	Original	Corrected
AF	As Fuck	Couldnt	Couldn't
shi{2,}t	shit	Youre	You're
tho	though	w/	with
tryna	trying to	you{2,}	you
Idk	I don't know	no{2,}	no
smh	shaking my head	pickin	picking
idc	I don't care	do{2}pe	dope
hmu	hit me up	Aa+nd	And
Lmao	Laughing my ass off	a{2,}nd	and
panickin	panicking	asap	As soon as possible
Hmu	Hit me up	ASAP	As soon as possible
smdh	shaking my damn head	feelin	feeling
ijs	I'm just saying	fukin	fuking
Gtfoh	Get the fuck out of here	pics	pictures
Tbh	To be honest	lmfao	Laughing my fucking ass off
idk	I don't know	Lmfao	Laughing my fucking ass off
OMG	Oh My God	funn{2,}y	funny
Omfg	Oh my fucking Gosh	gunna	gonna
seperatly	separately	probly	probably
pkgs	packages	Bday	Birthday
deff	definitely	Smh	Shaking my head
Luv	Love	Nothin	Nothing
Becuz	Because		
Havent	Haven't		
Lmao	Laughing my ass off		
Thx	Thanks		
Ty	Thank you		
Tfw	That feel when		
No{2,}	No		
OMFG	Oh my fucking Gosh		
rofl	Rolling On Floor Laughing		
wa{2,}y	way		

Appendix IV Sources

Following is the list of web sources (blogs, newspaper websites, Facebook, and Twitter links) from where the data was collected.

Pakistani Data

Single-writer blogs

1. http://alphaza.blogspot.com/
2. http://andweshout.blogspot.com
3. http://armageddonali.blogspot.com
4. http://bakefresh.blogspot.com/
5. http://beanbagtales.blogspot.com
6. http://blog-bilo.blogspot.com/
7. http://bytes0flife.blogspot.com
8. http://chocomallow.blogspot.com
9. http://clinicalhammer.blogspot.com/
10. http://commonr.blogspot.com
11. http://deeppurpleandblue.blogspot.com
12. http://diaryofanallpakistanigirl.blogspot.com
13. http://emaan-wahaj.blogspot.com
14. http://faysy.blogspot.com/
15. http://flat-tidings.blogspot.com/
16. http://humaira-anwar.blogspot.com/
17. http://illusionsandwords.blogspot.com
18. http://illuxon.blogspot.com/
19. http://kissmyroti.blogspot.com/
20. http://me-sugarlane.blogspot.com/
21. http://mhamzazafar.blogspot.com
22. http://mujtabachang.blogspot.com
23. http://myhumanistblog.blogspot.com
24. http://pakhaana.blogspot.com
25. http://peachypout.blogspot.com/
26. http://rehmat-yazdani.blogspot.com/
27. http://relationships-catgirl.blogspot.com/
28. http://roydonsblog.blogspot.com/
29. http://saadiam.blogspot.com
30. http://sarahinsouthkorea.blogspot.com/
31. http://scepticlife.blogspot.com/
32. http://thesunsetshow.blogspot.com/
33. http://thinkinglifeandyou.blogspot.com
34. http://think-islam.blogspot.com

35. http://vanishingfromtheworld.blogspot.com

36. http://weedmedia.blogspot.com

37. http://writersblocktgh.blogspot.com/

38. http://www.areejusman.com/

39. http://www.chaudhryjavediqbal.net

40. http://www.elzaa.com/

41. http://www.riazhaq.com

42. http://www.sarahassansblog.com/

43. http://zak-randomramblings.blogspot.com/

44. http://artkapakistan.wordpress.com/

45. http://ayesha5.wordpress.com

46. http://changinguppakistan.wordpress.com

47. http://fakihahassanrizvi.wordpress.com/

48. http://hadeelnaeem.wordpress.com

49. http://harisgulzar.wordpress.com

50. http://kalakawa.wordpress.com

51. http://mahrukhh.wordpress.com

52. http://nmafzal.com/

53. http://randomlyabstract.wordpress.com

54. http://sehartariq.wordpress.com/

55. http://tanzeel.wordpress.com/

56. http://teeth.com.pk/blog/

57. http://thesaproject.wordpress.com/

58. http://ukamkhan.wordpress.com

59. http://wasioabbasi.wordpress.com/

60. http://www.dholsipahi.com

Multi-writer blogs

1. http://cafepyala.blogspot.com/

2. http://cafeyouth.blogspot.com

3. http://fiverupees.com/

4. http://islamabad.metblogs.com/

5. http://karachi.metblogs.com/

6. http://lahore.metblogs.com/

7. http://pakistaniat.com/

8. https://www.pakpositive.com/pakistanibloggers/f2_p1.html

News blogs

1. http://blogs.aaj.tv/
2. https://blogs.arynews.tv/
3. http://blogs.dunyanews.tv/
4. https://www.samaa.tv/tag/samaa-blogs/
5. https://www.dawn.com/blogs
6. https://blogs.tribune.com.pk/
7. https://nation.com.pk/blogs
8. http://blogs.thenews.com.pk/blogs/

Technology blogs

1. https://propakistani.pk
2. https://thetech.pk/
3. https://www.techjuice.pk/
4. https://www.pakorbit.com
5.

New media blogs

1. http://www.parhlo.com/
2. http://www.mangobaaz.com/
3. https://trending.pk
4. https://images.dawn.com/

Facebook groups

1. https://www.facebook.com/groups/658084470964920/
2. https://www.facebook.com/groups/thepakistaniprobloggers/
3. https://www.facebook.com/groups/PakistanStartupJobs/
4. https://www.facebook.com/groups/PakistanStartups/
5. https://www.facebook.com/groups/197164200312640/
6. https://www.facebook.com/groups/pettalkpk/
7. https://www.facebook.com/groups/904054973009514/

8. https://www.facebook.com/groups/question.updates/

9. https://www.facebook.com/groups/10112282438/

10. https://www.facebook.com/groups/174980576174142/

11. https://www.facebook.com/groups/127580240922167/

Facebook status updates

1. https://www.facebook.com/muhammad.a.saqib

2. https://www.facebook.com/lubesQ

3. https://www.facebook.com/faiskap

4. https://www.facebook.com/hiraa.zubair?v=timeline

5. https://www.facebook.com/hassam.ahmed.9

6. https://www.facebook.com/aliabbaszaidi

7. https://www.facebook.com/mikaalzulfiqarofficial/

8. https://www.facebook.com/FerozeKhanOfficial

9. https://www.facebook.com/Official.UmairJaswal/

10. https://www.facebook.com/hamzaaliabbasi/

11. https://www.facebook.com/sahirlodhi

12. https://www.facebook.com/SheheryarMunawarOfficial

13. https://www.facebook.com/shehzadroypage/

14. https://www.facebook.com/noumaan.yaqoob

15. https://www.facebook.com/blessedAyesha

16. https://www.facebook.com/saad440

17. https://www.facebook.com/tafseer.ahmed

18. https://www.facebook.com/irfan.shahzad.2008

19. https://www.facebook.com/AbidHussainBeli?fref=nf

20. https://www.facebook.com/qudoos

21. https://www.facebook.com/arsalangenius

22. https://www.facebook.com/muhammadzohairchohan

23. https://www.facebook.com/khursheed.anwar.92

24. https://www.facebook.com/ghaziusman

25. https://www.facebook.com/ahmed.yaseen.90857

26. https://www.facebook.com/shazianawazsays

27. https://www.facebook.com/Zafar.tnw

28. https://www.facebook.com/wahibhaq

29. https://www.facebook.com/muhammad.gulfam.104

30. https://www.facebook.com/rashid.mahmood.94043

31. https://www.facebook.com/noor.u.qasmi

32. https://www.facebook.com/huma.habib.9210

33. https://www.facebook.com/oyeraza

34. https://www.facebook.com/Muzamil1991

35. https://www.facebook.com/mohsin.xia

36. https://www.facebook.com/MOHAMMAD.HAROON.ARAIN

37. https://www.facebook.com/usmangul85

38. https://www.facebook.com/wasio.abbasi

39. https://www.facebook.com/AdnanLive

40. https://www.facebook.com/jehan.ara

41. https://www.facebook.com/shamilaghyas

42. https://www.facebook.com/hareem.sumbul

43. https://www.facebook.com/rida.malik.5836

44. https://www.facebook.com/Maham1

45. https://www.facebook.com/its.a.hit

46. https://www.facebook.com/danishmughal

47. https://www.facebook.com/ayesha.siddiqa.351

48. https://www.facebook.com/rehan33

49. https://www.facebook.com/hassam.awan

50. https://www.facebook.com/tofiqpasha.mooraj

51. https://www.facebook.com/kamrankami55?

52. https://www.facebook.com/successbux

53. https://www.facebook.com/FarmanMayMon

54. https://www.facebook.com/maaz.cooder

55. https://www.facebook.com/Javaid.Omar

56. https://www.facebook.com/yousufrafi

57. https://www.facebook.com/fareed.gujjar.14

58. https://www.facebook.com/zubiarizvi

59. https://www.facebook.com/azharphysics

60. https://www.facebook.com/SMAsimQadri

61. https://www.facebook.com/affan.shahab

62. https://www.facebook.com/Mr.SHEHZAD

63. https://www.facebook.com/iahmedhams

64. https://www.facebook.com/ZaidPirwani

65. https://www.facebook.com/UmarMaj

66. https://www.facebook.com/SajidShah44

67. https://www.facebook.com/salluhee

68. https://www.facebook.com/amanullah.tanweer

69. https://www.facebook.com/invinciblesaad

70. https://www.facebook.com/sarahrehman

71. https://www.facebook.com/Nadi.Ramzan

72. https://www.facebook.com/matintraveller

73. https://www.facebook.com/ataazz

74. https://www.facebook.com/sameerahmedkhan.peace

75. https://www.facebook.com/rizshoukat

76. https://www.facebook.com/Ismail.Kodvavi

77. https://www.facebook.com/MJunaidIt

78. https://www.facebook.com/abs.itean

79. https://www.facebook.com/qudsiaehtram.ehtram

80. https://www.facebook.com/asma.inayat

81. https://www.facebook.com/madiha.mukarram

82. https://www.facebook.com/nargis.wahidali

83. https://www.facebook.com/aneelamirza

84. https://www.facebook.com/meher.inayat.3

85. https://www.facebook.com/sehrish.kazim

86. https://www.facebook.com/lubna.khawar

87. https://www.facebook.com/zehra.ilyas2

88. https://www.facebook.com/profile.php?id=513232132

89. https://www.facebook.com/saddaz

90. https://www.facebook.com/saman.wasti.1

91. https://www.facebook.com/yzma.bham

92. https://www.facebook.com/munira.mehdi

93. https://www.facebook.com/ayesha.junaid.54

94. https://www.facebook.com/Sara.bhurgari

95. https://www.facebook.com/kanwal.farooq.5

96. https://www.facebook.com/mahakhans

97. https://www.facebook.com/KishLovCake

98. https://www.facebook.com/Fatima.Muhammad.Taqvi

99. https://www.facebook.com/sarahperacha

100. https://www.facebook.com/mehreen.kandaan

101. https://www.facebook.com/urouge.iqbal

102. https://www.facebook.com/sabaeitizaz

103. https://www.facebook.com/AlyzehRS

104. https://www.facebook.com/Toobasyed17

Tweets

1. https://twitter.com/aadilkhalil

2. https://twitter.com/adnanejazkhan

3. https://twitter.com/afterteapls

4. https://twitter.com/ahsansamee

5. https://twitter.com/alishbanaeem2

6. https://twitter.com/allichkhan

7. https://twitter.com/almoestate

8. https://twitter.com/amnafazail

9. https://twitter.com/Areeshaxoxo

10. https://twitter.com/asifsomy

11. https://twitter.com/askchini

12. https://twitter.com/asmashirazi

13. https://twitter.com/azanamjad

14. https://twitter.com/baqirmisbah

15. https://twitter.com/cheenaagha

16. https://twitter.com/chnaveedasif

17. https://twitter.com/daauudfifty

18. https://twitter.com/daniaashahid

19. https://twitter.com/dearshahbazali

20. https://twitter.com/drkanchaa

21. https://twitter.com/drrahmadtufail

22. https://twitter.com/faizafaiyaz

23. https://twitter.com/faizanlakhani

24. https://twitter.com/hammadchauhan

25. https://twitter.com/haqnawazbutt

26. https://twitter.com/harrisalich

27. https://twitter.com/hasaankhawar

28. https://twitter.com/hasanrazaq

29. https://twitter.com/hijaziakhan

30. https://twitter.com/huzaifaahmadani

31. https://twitter.com/hyperalgesic

32. https://twitter.com/hzsays

33. https://twitter.com/iamafaqashfaq

34. https://twitter.com/iammudiqbal

35. https://twitter.com/ibrakhan

36. https://twitter.com/imhamzaofficial

37. https://twitter.com/imsumerapti

38. https://twitter.com/isuperbatmann

39. https://twitter.com/itsanuuu

40. https://twitter.com/itsimak

41. https://twitter.com/itssyedaa

42. https://twitter.com/itstehryym

43. https://twitter.com/jawairia_jiya

44. https://twitter.com/jawairianazir

45. https://twitter.com/jimmykhansial

46. https://twitter.com/kakhangandapur

47. https://twitter.com/kamranisbest

48. https://twitter.com/khanbwp

49. https://twitter.com/livedynamite

50. https://twitter.com/llifeless

51. https://twitter.com/LubnaSabir1

52. https://twitter.com/luqmankyz

53. https://twitter.com/maaaheeenn

54. https://twitter.com/mariasherazkhan

55. https://twitter.com/mmhemani

56. https://twitter.com/mnaeemkhaliq

57. https://twitter.com/munirhussaharal

58. https://twitter.com/mustahtaba

59. https://twitter.com/mwrao

60. https://twitter.com/mzsqadri

61. https://twitter.com/naeemaafzal

62. https://twitter.com/NageenYousaf

63. https://twitter.com/nailahgul

64. https://twitter.com/nailainayat

65. https://twitter.com/navidgeee

66. https://twitter.com/nazmamustafavi

67. https://twitter.com/nergisanees

68. https://twitter.com/nomiferoz

69. https://twitter.com/nylaawan

70. https://twitter.com/pleeezing

71. https://twitter.com/qaziyasar

72. https://twitter.com/rabbaneeya

73. https://twitter.com/rafiamkhan

74. https://twitter.com/ranasamofficial

75. https://twitter.com/ronietaz

76. https://twitter.com/saadarshadd

77. https://twitter.com/saadigeek

78. https://twitter.com/sabahatsays

79. https://twitter.com/Sadia_Mallick

80. https://twitter.com/saimilhan

81. https://twitter.com/sayyukhh

82. https://twitter.com/shizaahmeddaha

83. https://twitter.com/shtanveerakhtar

84. https://twitter.com/sohaibfurqan

85. https://twitter.com/sohebbandesha

86. https://twitter.com/sophiaghaffar

87. https://twitter.com/sqsaad

88. https://twitter.com/sundasnaeem

89. https://twitter.com/tariqmushtaqkh

90. https://twitter.com/tfarzeena

91. https://twitter.com/tubayaseenmalik

92. https://twitter.com/umarsahilpti

93. https://twitter.com/wahababbaxi

94. https://twitter.com/WareeshJawad

95. https://twitter.com/whereangelsdare

96. https://twitter.com/whybolt

97. https://twitter.com/xaidlodhi

98. https://twitter.com/xulfii

99. https://twitter.com/yassirsaleem

100. https://twitter.com/zobyrhashmi

101. https://twitter.com/zunairkh

102.   https://twitter.com/ptisaba

103.   https://twitter.com/ramshawar
         raich

104.   https://twitter.com/mehreenza
         hra

105.   https://twitter.com/baeghairat

106.   https://twitter.com/daniyyalal
         i

107.   https://twitter.com/muneebfar
         uq

108.   https://twitter.com/iffiviews

109.   https://twitter.com/jiyaawan

110.   https://twitter.com/kashifgilla
         ni

111.   https://twitter.com/mughalbh
         a

112.   https://twitter.com/proudofpk

113.   https://twitter.com/abdulraufa
         khtar

114.   https://twitter.com/areeshzub
         air

115.   https://twitter.com/burairali

U.S. English

Single-writer blogs

1. http://aboutracewriter.blogspot.com/
2. http://adventuresofjohnnynorthside.blogspot.com/
3. http://alaskareal.blogspot.com/
4. http://authentic-connecticut-republican.blogspot.com/
5. http://carolsviewofnewengland.blogspot.com/
6. http://dailyspiritualtools.blogspot.com/
7. http://frontporchindiana.blogspot.com/
8. http://iowageek.blogspot.com/
9. http://krystallitt.blogspot.com/
10. http://maineliving.blogspot.com/
11. http://ourlifeinidaho.blogspot.com/
12. http://peoriarocks.blogspot.com/
13. http://politakid.blogspot.com/
14. http://rbkaplan.blogspot.com/
15. http://redstatediaries.blogspot.com/

16. http://scribbles-corry.blogspot.com/

17. http://siriouslydelicious.blogspot.com/

18. http://stapletonkearns.blogspot.com/

19. http://thetruffle.blogspot.com/

20. http://unrulybehavior.blogspot.com/

21. http://upstateunderyournose.blogspot.com/

22. http://wahm-shelley.blogspot.com/

23. http://www.arizona-writer.com/

24. http://www.cynicalcook.com/

25. http://www.dreaminginblush.com/

26. http://www.kaintuckeean.com/

27. http://www.simpleweddingsblog.com/

28. http://www.southernrockiesnatureblog.com/

29. http://www.theamericanzombie.com/

30. http://www.walkingsaint.com/

31. http://bluegrassroots.org/

32. http://fiveaspects.net/blog/

33. http://jesseyancy.com/

34. http://lifestyledenver.com/

35. http://nmisscommentor.com/

36. http://onepoundago.com/blog/

37. http://redmassgroup.com/

38. http://shesintheglow.com/

39. http://templestark.com/

40. http://terry.ipearson.net/

41. http://tomburka.com/

42. http://utterlyboring.com/

43. http://wisdomweaverblog.com/

44. http://www.aguyonclematis.com/clematis-street-blog/

45. http://www.bamasteelmagnoliasbistro.com/

46. http://www.boundarywatersblog.com/

47. http://www.abetterwestvirginia.com/

48. http://www.despisingnone.com/

49. http://www.dirigoblue.com/

50. http://www.ericsiegmund.com/fireant/

51. http://www.kitchenkonfidence.com/

52. http://www.louisvillehotbytes.com/

53. http://www.paulandangela.net/

54. http://www.reardonreports.com/

55. http://www.peggypayne.com/blog/

56. http://www.umassonline.net/blog

57. https://inthegables.wordpress.com/

58. https://kieranhealy.org/blog/

59. https://littlebylisten.wordpress.com/

60. https://www.nwprogressive.org/weblog/

Multi-writer blogs

1. http://atlanta.metblogs.com/

2. http://boston.metblogs.com/

3. http://dc.metblogs.com/

4. http://denver.metblogs.com/

5. http://hawaii.metblogs.com/

6. http://nyc.metblogs.com/

7. http://phoenix.metblogs.com

8. http://pittsburgh.metblogs.com

9. http://portland.metblogs.com

10. http://sacramento.metblogs.com

11. http://seattle.metblogs.com

12. http://sf.metblogs.com/

13. http://www.dailyyonder.com/

14. http://www.themudflats.net/

News blogs

1. https://blogs.wsj.com

2. https://www.nytimes.com/interactive/blogs/directory.html

3. https://usatoday30.usatoday.com/blog-index.htm

4. https://www.chicagotribune.com/news/opinion/blogs/

Technology blogs

1. https://www.androidpolice.com/

2. https://www.engadget.com/

3. https://techcrunch.com/

4. https://www.wired.com/

New media blogs

1. http://www.viralnova.com/

2. http://mashable.com/

3. http://www.distractify.com/   4. http://www.upworthy.com/

Facebook groups

1. https://www.facebook.com/groups/113020160790/

2. https://www.facebook.com/groups/1639991656214998/

3. https://www.facebook.com/groups/394984244018281/

4. https://www.facebook.com/groups/565589170217332/

5. https://www.facebook.com/groups/laforbernie

6. https://www.facebook.com/groups/1840862356139476/

7. https://www.facebook.com/groups/150046741704357/

8. https://www.facebook.com/groups/baysidequeens/

9. https://www.facebook.com/groups/PokemonGOLAGroup/

10. https://www.facebook.com/groups/169919415878/

11. https://www.facebook.com/groups/webster.ny/

12. https://www.facebook.com/groups/carriebridgette/

Facebook status updates

1. https://www.facebook.com/profile.php?id=30825922

2. https://www.facebook.com/anthony.shindeldecker

3. https://www.facebook.com/james.begley.796774

4. https://www.facebook.com/heather.reaversotelo

5. https://www.facebook.com/brad.cox.982

6. https://www.facebook.com/Jason.kennedy.70

7. https://www.facebook.com/new.theatre

8. https://www.facebook.com/miranda.katz

9. https://www.facebook.com/ewierda

10. https://www.facebook.com/claudette.lacourse

11. https://www.facebook.com/kstotter

12. https://www.facebook.com/tedconstan

13. https://www.facebook.com/hellyeahimgay

14. https://www.facebook.com/jon.sadowski.9

15. https://www.facebook.com/elenawbrenda

16. https://www.facebook.com/lejla.subasic.927

17. https://www.facebook.com/pmkyle

18. https://www.facebook.com/taone.walker

19. https://www.facebook.com/joey.capestany

20. https://www.facebook.com/jafei.pollitt

21. https://www.facebook.com/joey.contino.16

22. https://www.facebook.com/joeylegittino

23. https://www.facebook.com/jimmy.kourafas

24. https://www.facebook.com/sellllli

25. https://www.facebook.com/profile.php?id=100009104607170

26. https://www.facebook.com/nancy.chalfant

27. https://www.facebook.com/kjc143

28. https://www.facebook.com/dawn.strohm.7

29. https://www.facebook.com/zdeangelis

30. https://www.facebook.com/kirsten.banks.75

31. https://www.facebook.com/thomas.steed1

32. https://www.facebook.com/natalia.beke

33. https://www.facebook.com/maranda.trexler.9

34. https://www.facebook.com/john.gaffney.351

35. https://www.facebook.com/tabatha.eschelbacher

36. https://www.facebook.com/bob.asbury.75

37. https://www.facebook.com/profile.php?id=100012362401714

38. https://www.facebook.com/alexis181818

39. https://www.facebook.com/vivian.lynn.5

40. https://www.facebook.com/kthlncole

41. https://www.facebook.com/kristin.waszkiewicz

42. https://www.facebook.com/sammi.blasio

43. https://www.facebook.com/tiffany.fisher.39982

44. https://www.facebook.com/stanton.paul.5

45. https://www.facebook.com/HouserPat

46. https://www.facebook.com/kevin.perkins.906

47. https://www.facebook.com/tommy.hyjynx

48. https://www.facebook.com/amanda.wrigglesworth.is.smarter.than.you

49. https://www.facebook.com/nalene.schiavoni

50. https://www.facebook.com/michael.fox.7106

51. https://www.facebook.com/kproblemo

52. https://www.facebook.com/profile.php?id=100001262595656

53. https://www.facebook.com/AlaskanWarrior

54. https://www.facebook.com/malcolm.jackson

55. https://www.facebook.com/JoAnnSantomero

56. https://www.facebook.com/D.L.Polonsky

57. https://www.facebook.com/karen.cobeen.9

58. https://www.facebook.com/richies.sox

59. https://www.facebook.com/eileen.valentinoflaxman

60. https://www.facebook.com/diana.sprout

61. https://www.facebook.com/jim.obrien.121

62. https://www.facebook.com/dewbabyr

63. https://www.facebook.com/ethelnalule1

64. https://www.facebook.com/chole00784

65. https://www.facebook.com/nikyla.boxley

66. https://www.facebook.com/joi.hardin

67. https://www.facebook.com/albert.martin.7739

68. https://www.facebook.com/robkorobkin

69. https://www.facebook.com/msMeshasocute

70. https://www.facebook.com/profile.php?id=100011043123358

71. https://www.facebook.com/qutie.boyd

72. https://www.facebook.com/CBP1206

73. https://www.facebook.com/SJD.LIFE

74. https://www.facebook.com/matt.craig.9081

75. https://www.facebook.com/danielle.barb

76. https://www.facebook.com/VasquezForSheriff

77. https://www.facebook.com/chrissy.alzapiedi

78. https://www.facebook.com/donald.cairns.16

79. https://www.facebook.com/profile.php?id=100004046654008

80. https://www.facebook.com/janine.lavecchia

81. https://www.facebook.com/jim.stadler.94

82. https://www.facebook.com/ibemikelee

83. https://www.facebook.com/paula.costanzoartopiades

84. https://www.facebook.com/frank.donato.524

85. https://www.facebook.com/jim.ryan.98

86. https://www.facebook.com/todd.mcnabney

87. https://www.facebook.com/barbara.sanelli

88. https://www.facebook.com/jason.smythe.9085

89. https://www.facebook.com/gino.romo

90. https://www.facebook.com/christopher.morell

91. https://www.facebook.com/jael.morales.338

92. https://www.facebook.com/erin.meehye.yi

93. https://www.facebook.com/LostHopes

94. https://www.facebook.com/ryan.balba

95. https://www.facebook.com/jeremiahborbe

96. https://www.facebook.com/kraziekajuro

97. https://www.facebook.com/alex.dasilva.355

98. https://www.facebook.com/mcmishkoff

99. https://www.facebook.com/adheresty

100.	https://www.facebook.com/lorraine.martinez.503

101.　　https://www.facebook.com/icefaerie

102.　　https://www.facebook.com/carlos.rosado.1690

103.　　https://www.facebook.com/zachary.palmer.5015

104.　　https://www.facebook.com/poppy.vest

105.　　https://www.facebook.com/one.clever.name

106.　　https://www.facebook.com/steven.s.dean.7

107.　　https://www.facebook.com/fabian.narez.1

108.　　https://www.facebook.com/patrick.grinage

Tweets

1. https://twitter.com/aarongordon26
2. https://twitter.com/abbyesmurdoc
3. https://twitter.com/acciomary
4. https://twitter.com/acehanna54
5. https://twitter.com/adam_maba
6. https://twitter.com/aishole
7. https://twitter.com/aliemaur
8. https://twitter.com/alukalani
9. https://twitter.com/amberjoys1
10. https://twitter.com/ambervannoy
11. https://twitter.com/amylippincott
12. https://twitter.com/anaisojeda1990
13. https://twitter.com/andrerivera801
14. https://twitter.com/angelaflournoy
15. https://twitter.com/anni_kaay
16. https://twitter.com/appstate311
17. https://twitter.com/asaleahh
18. https://twitter.com/bbrunson25
19. https://twitter.com/bcloritts
20. https://twitter.com/beauty_marlene
21. https://twitter.com/biggtwon
22. https://twitter.com/bivaughn
23. https://twitter.com/bobbye53
24. https://twitter.com/borninamess
25. https://twitter.com/brandcochran
26. https://twitter.com/brianafrapart
27. https://twitter.com/brigittybridge
28. https://twitter.com/caddguru
29. https://twitter.com/caedy

30. https://twitter.com/camhatsNYC

31. https://twitter.com/carolmwagner

32. https://twitter.com/chicosueno

33. https://twitter.com/chillwithreem

34. https://twitter.com/chrisandxbones

35. https://twitter.com/clubsodatwist

36. https://twitter.com/coorslightking

37. https://twitter.com/cperk54

38. https://twitter.com/cuprado

39. https://twitter.com/daaawwwnn_

40. https://twitter.com/dareber

41. https://twitter.com/derekmorley

42. https://twitter.com/doodlehedz

43. https://twitter.com/drinkwater2day

44. https://twitter.com/drtyronegray20

45. https://twitter.com/dwightali

46. https://twitter.com/emilyayay17

47. https://twitter.com/emjags

48. https://twitter.com/emmaenglish95

49. https://twitter.com/eton_rifles

50. https://twitter.com/evamatwil

51. https://twitter.com/evanpenn

52. https://twitter.com/fergnerduson

53. https://twitter.com/gabibartolucci

54. https://twitter.com/gkalkat

55. https://twitter.com/gropptimusprime

56. https://twitter.com/itsdominique

57. https://twitter.com/itsJonLevine

58. https://twitter.com/jdotdrew

59. https://twitter.com/jeffwitt

60. https://twitter.com/jfield1869

61. https://twitter.com/jordannmichaels

62. https://twitter.com/joselouis4077

63. https://twitter.com/kdfsmiley

64. https://twitter.com/kirstenm922

65. https://twitter.com/kqthekingbee

66. https://twitter.com/legnemannairb

67. https://twitter.com/mdstillwagon

68. https://twitter.com/miakoo

69. https://twitter.com/mittense

70. https://twitter.com/nclight86

71. https://twitter.com/nonchalantpeach

72. https://twitter.com/not_carlos

73. https://twitter.com/nursiemarcie

74. https://twitter.com/nvoll84

75. https://twitter.com/okuudere

76. https://twitter.com/omar1velous

77. https://twitter.com/pastorjmartin

78. https://twitter.com/peterhoneyman

79. https://twitter.com/rbigaugotham

80. https://twitter.com/renteriaboxing

81. https://twitter.com/rfreuden

82. https://twitter.com/rickyjenkins

83. https://twitter.com/roaztb33f

84. https://twitter.com/robinkaye

85. https://twitter.com/robprather

86. https://twitter.com/run10K31

87. https://twitter.com/ryansaline

88. https://twitter.com/sashajenn

89. https://twitter.com/sbattenbland

90. https://twitter.com/sir_melo3

91. https://twitter.com/slimt1213

92. https://twitter.com/spinning_will

93. https://twitter.com/ssummers3650

94. https://twitter.com/stephenstackzzz

95. https://twitter.com/stephyfierce

96. https://twitter.com/swmoore1967

97. https://twitter.com/synerjes

98. https://twitter.com/taralowerymua

99. https://twitter.com/themanmoe

100. https://twitter.com/thisisdking

101. https://twitter.com/thotpagne mami

102. https://twitter.com/wi_borzo

103. https://twitter.com/willyouare not

104. https://twitter.com/xoxo_shre ya

105. https://twitter.com/yungkilo_

106. https://twitter.com/zorinlynx

Appendix V Curriculum Vitae - Muhammad Shakir