Language of Text Messages: A Corpus Based Linguistic Analysis of SMS in Pakistan

Researcher:

Malik Naseer Hussain

Supervisor:

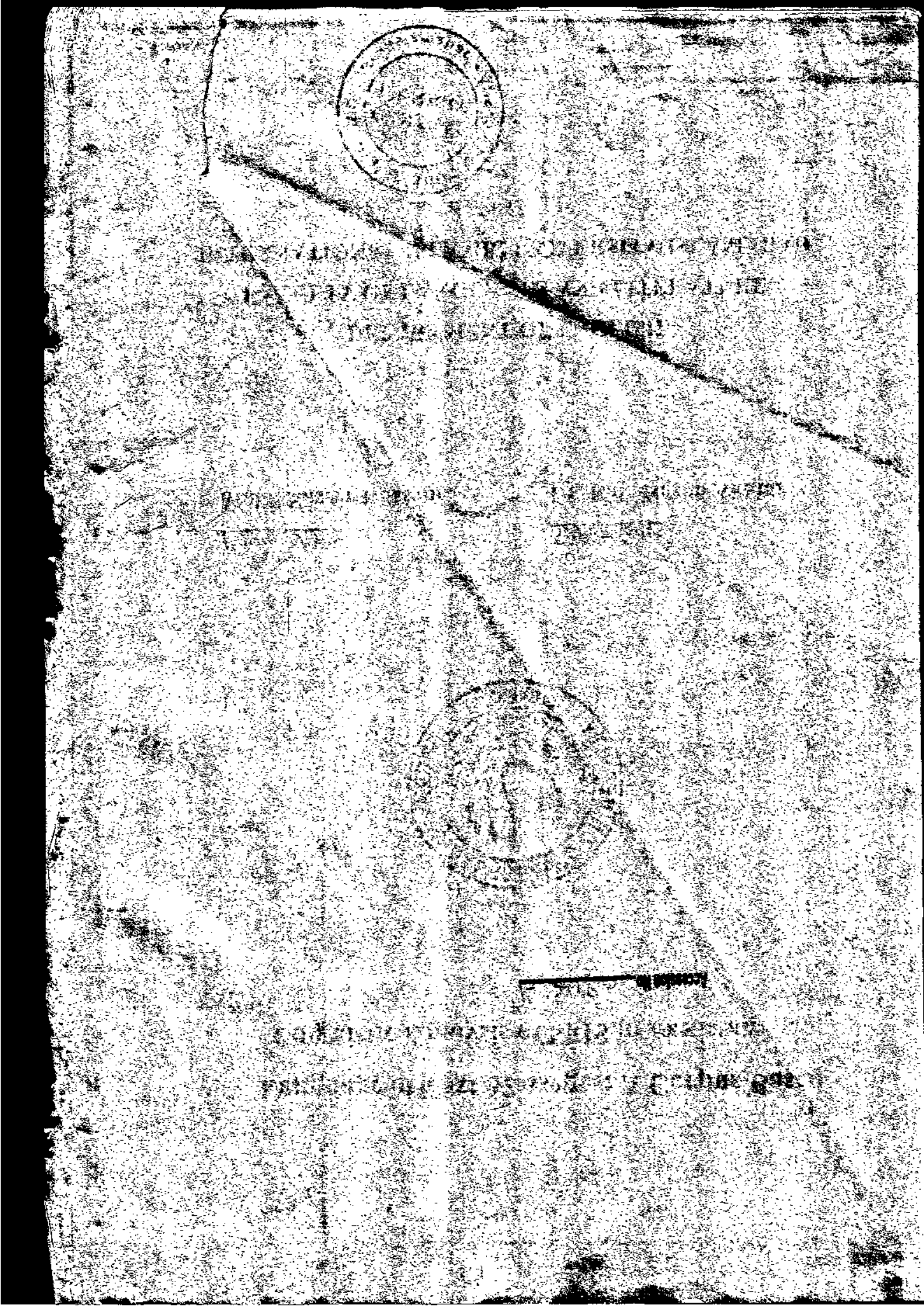
Dr. Raja Nasim Akhtar

DEPARTMENT OF ENGLISH

FACULTY OF LANGUAGES AND LITERATURE

INTERNATIONAL ISLAMIC UNIVERSITY ISLAMABAD

2013



I Linguistics
II Spelling - Standard language
III Theses

Language of Text Messages: A Corpus Based Linguistic Analysis of SMS in Pakistan

By

Malik Naseer Hussain
Reg. No. 19-FLL/PhDEng/F-07

A dissertation submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in English (Linguistics)

DEPARTMENT OF ENGLISH
FACULTY OF LANGUAGES AND LITERATURE
INTERNATIONAL ISLAMIC UNIVERSITY ISLAMABAD
2013

اقْرَأْ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ

Read! In the name of thy Lord Who createth

IN THE NAME OF ALLAH, THE MOST GRACIOUS, THE MOST MERCIFUL

قُلْ إِنَّ صَلَاتِي وَنُسُكِي وَمَحْيَايَ وَمَمَاتِي لِلَّهِ رَبِّ الْعَالَمِينَ

Say: Truly, my worship, and my sacrifice, and my living, and my dying are for Allah, Lord of the Worlds. (Al-Qur'an, 6, 162)

Acceptance by the *Viva Voce* Committee

Language of Text Messages: A Corpus Based Linguistic Analysis of SMS in Pakistan

Name of Student: Malik Naseer Hussain

Registration No: 19-FLL/ PhDENG/F07

Accepted by the department of English, Faculty of Languages & Literature, International Islamic University, Islamabad, in partial fulfillment of the requirement for the Doctor of Philosophy degree in English.

***Viva Voce* Committee**

External Examiner
Dr. Riaz Hassan
Dean, Faculty of Social Science, Air University, E-9 Islamabad

Dr. Munawar Iqbal Ahmad
Acting Dean
Faculty of Languages & Literature

External Examiner
Dr. Wasima Shehzad
Professor (Linguistics), Department of Humanities, Air University, E-9, Islamabad

Dr. Munawar Iqbal Ahmad
Chairman
Department of English

Internal Examiner
Dr.Ayaz Afsar
Associate Professor
Department of English, IIUI

Supervisor
Dr. Raja Nasim Akhtar
Professor, Department of English, AJ&K University, Muzaffarabad.

07 October, 2013

FOREIGN EVALUATION

As per HEC policy, the thesis has been evaluated by the following foreign experts from technologically/academically advanced countries.

1.

Prof. Dr. Marina Bondi

Professor, Department of Linguistics and Cultural Studies

Dean, Faculty of Humanities

University of Modena & Reggio Emilia, Italy

Email: marina.bondi@unimore.it

2.

Dr. Paul Thompson

Director, Centre for Corpus Research

Arts Building, University of Birmingham

Edgbaston, Birmingham

B15 2TT, UK

Email: p.thompson@bham.ac.uk

DECLARATION

I, Malik Naseer Hussain, Registration No. 19-FLL/PhDEng/F-07, a student of PhD in English Linguistics at International Islamic University Islamabad do hereby solemnly declare that the dissertation submitted by me in partial fulfilment of the requirements for the degree of Doctor of Philosophy in English (Linguistics) is my original work, except where otherwise acknowledged in the dissertation, and has not been submitted earlier, and shall not be submitted by me in future for obtaining any degree from this or any other university.

Malik Naseer Hussain

Dated: 7th October 2013

ACKNOWLEDGEMENTS

First and foremost, all thanks and gratitude to **Allah Almighty** for giving me strength and opportunities to complete my PhD degree. After Allah, I pay homage to the last Prophet of Allah, **Muhammad (PBUH),** whose Sunnah is to thank people. He says, "لاَ يَشْكُرُ اللَّهَ مَنْ لاَ يَشْكُرُ النَّاسَ / He does not thank Allah who does not thank people" Sunan-Abu-Dawud, 4811.

In this regard, I owe thanks to a large number of people and institutions. The first among them is **Higher Education Commission,** for offering me the merit based scholarship for PhD studies. Next, I owe debts to my research supervisors **Prof. Dr. Raja Nasim Akhtar** (my PhD supervisor) and **Prof. Dr. Ayaz Afsar** (my MPhil supervisor) for their support and guidance. I owe a great debt to the Chairman, Department of English, IIUI, **Prof. Dr. Munawar Iqbal Ahmed,** for his unflinching support in every administrative matter.

There are many teachers to whom I thank for their guidance at different phases of my academic career. My special thanks are to my IIUI teachers, Prof. Dr. Ahsan Wahga, Prof. Dr. Qabil Khan, Prof. Dr. Maqsood Alam Bukhari, Prof. Dr. Muhammad Rasheed, Prof. Dr. Nabi Baksh Jumani, Prof. Dr. Safeer Akhtar, Prof. Dr. Safeer Awan; my teachers at AIOU, Prof. Dr. Rehana Masrur, Prof. Dr. Naveed Sultana; and my teachers at Govt. Postgraduate College Attock, Prof. Ashiq Hussain, Prof. Altaf Baig, Prof. Zaheer Ahmed, Prof. Mahmood Khattak, Prof. Shakeel Ahmed, and Prof. Sharif Zaid.

Without good friends, one cannot easily accomplish a long journey of studies. So I wish to thank all my colleagues and friends whose support and good wishes remained with me during my educational career. I especially thank my PhD class fellows Ibrar Anver Sheikh, Khalid Mahmood, Tassadaq Hussain, Muhammad Nawaz, Abid Qureshi, Aleem Shakir, Riaz Dasti; my university fellows Muhammad Zamurd, M. Azhar Khan, Mudassar Nazir, Abdul Qayyum, Hafiz Zahid, Muslim Khan, Muhammad Saleem, Ameer Sultan; my friends Abdul Qayyum Sahar, Umer Farooq, Tariq Mehmood Sheikh, Najam-us-Saqib, Yasir Mahmmood, Atif Haseeb, Muhammad Younas Khan, Arshad Mahmood Nashad, Shuhaib Anwer, Masood Anwer, Faheem Qureshi, Amanat Khan, Muhammad Farooq, Abdul Qadoos; my colleagues Abdul Majid Sajjad, Kamran Jamil, M. Khizer, Qaiser Baig, Khurram Baig, Khawar Mehmmood Khokhar, Ashiq Abbasi, Zuhair Wasti, Kamal Pasha, Sajeela Shabir, and Munawar Kashif.

In the family, the biggest unpaid debt on me is of my father (May his soul rest in peace). What I am today, it is just because of him. People have many dreams in their lives but I was the only dream of his life. So I cannot pay his debt. In addition to him, there is yet a whole chain of my relatives to whom I owe thanks and apologies. I apologize to my family as I could not give them much time. I wish my heartfelt thanks and recognition to my parents, my wife, my sons Bilal Haider, Awais Haider, my daughters Arooj Fatima, Tehreem Fatima, my kin Malik Ibrar Hussain, Malik Zaheer Hussain, and Malik Sajjad Hussain.

I also thank all those who provided data for this study, or helped in collecting the data. Thanks to all whose works I used/consulted. My special thanks are to Prof. Laurence Anthony for providing AntConc, the corpus linguistics software, free of cost. While acknowledging people I cannot overlook to mention some places and institutes which have played an unseen role in my growth. I especially mention my native town Padhrar in District Khushab, F. G. Boys Jamshaid High School Attock Cantt, Govt. Postgraduate College Attock, and International Islamic University Islamabad. Their impacts have been very great on my personality

In brief, the list of thanks and gratitude is long but the conventional space available for acknowledgements is short, so without naming further, at the end I thank all unnamed and unspecified friends, colleagues, relatives, students, and institutes for the part they have played in my life and career.

ABSTRACT

The thesis presents a corpus based empirical analysis of the language used in SMS text messaging in Pakistan. The study is descriptive in nature and examines types, causes, and patterns/principles of various linguistic adaptations made in text messages. It also provides historical insights, and discusses linguistic-cum-educational impacts and implications of these adaptations. Primarily, the study is based on the linguistic analysis of an SMS corpus of 5000 interpersonal text messages collected from Pakistani texters. For triangulation purposes, it also examines the metalinguistic perceptions of 500 texters who also provided their personal text messages for the study. The study explores linguistic adaptations in six major categories, i.e. lexical, syntactic, punctuation, space, code, and script adaptations in text messages. It was found in the study that linguistic adaptations in text messages are mostly made under certain principles/patterns. Most intentional adaptations in text messages are caused by three major factors that are to be *economical* in the use of time and effort, to be *creative/innovative* in developing new language patterns, and to be *rapid* in SMS communication. Some unintentional adaptations are caused by the *careless attitude* or *poor language command* of texters. In addition, many punctuation adaptations are specifically made for *paralinguistic purposes*. Code alterations are made in the *bilingual settings* of texters. Lastly, the Roman script is preferred because most Pakistani texters are not adept in typing the Arabic/Urdu script, and so they use the Roman script for both Urdu and English. In the historical perspective, these adaptations are not completely new in nature because their traces are found in history or in other modes of communication. Moreover, these adaptations have various linguistic/educational impacts, and these impacts lead to certain implications for the conventional standards of languages and their teaching.

CONTENTS

LIST OF TABLES

Note on Style Manual

The study follows the "*Publication Manual of the American Psychological Association*" (2010) for in-text citations, references, and style purposes.

CHAPTER 1

INTRODUCTION

This is the introductory chapter of the study. It starts with the background on which the foundation of the study is laid. It includes some primary issues of the study like statement of the problem, objectives and research questions, significance of the study, and limitations and delimitations of the study. It also provides a glimpse of the research methodology and a brief account of the organization of the dissertation into various chapters.

1.1 Background of the Study

Mobile phone has rapidly spread over the world faster than any other communication technology, and is now widely used over the globe. Historically passing from different phases of mechanical telegraph to electric telegraph, and telephone to radio telegraphy, finally, the concept of mobile phone was perceived in communication technology. The concept was initiated in 1947 by Douglas H. Ring and William R. Young at Bell Telephone Laboratories in the USA (Ring, 1947). The first mobile phone call from a handheld set was made on April 3, 1973 by a Motorola engineer Martin Cooper to his rival Joel Angel, the research head at AT&T [American Telephone and Telegraph] Bell Labs. But it was not until 1980s that mobile phones were actually used as a medium of mass communication (Agar, 2004; Klemens, 2010; Mercer, 2006).

Mobile phone text messaging, commonly known as Short Message Service [SMS], was initiated in early 1990s by cellular services but it did not become fashionable until about 1998 (Crystal, 2004). Originally, it was "intended as a way for mobile providers to share alerts and other service-oriented information with their networks of users. It was conceived, then, as a method of business communication, and it was

imagined not as a back-and-forth process" among texters (Zurhellen, 2011, p. 638). In the case of Pakistan, text messaging is "a comparatively new service offered by some of the mobile companies" (Pakistan Telecommunication Authority [PTA], 2005, p. 44), initiated in 2003. Amazingly, within a few years from its inception, it has become a common means of interaction and communication among the young people in Pakistan. A tremendous increase has been seen in its usage in Pakistan during the last few years (Pakistan Telecommunication Authority, 2008, 2010).

After the inception of SMS at the level of masses, this mode of communication has developed its own variety of language. The SMS variety of language has its own unique features which differ in many ways from any standard language. Crystal (2004) asserts that factors like technical restriction of 160 characters per message and the special shape of mobile keypads have motivated the use of new forms of language, such as the use of short forms which include acronyms and abbreviations. In this regard, various terms are used to describe the language of text messages like *textese* (Sutherland, 2002), *txt* (Shortis, 2007), *texting* (Crystal, 2008b), and *SMS language* (Hard af Segerstad, 2002).

The language of text messages is also closely related to the language used in Computer Mediated Communication [CMC]. CMC has developed as a specific register/variety of language under the impact of computer and information technologies. It is divided into two types that are synchronous CMC and asynchronous CMC. According to Crystal (2004), in synchronous mode of CMC, communication takes place in the real time such as Instant Messaging [IM] and Internet Relay Chat [IRC]. In asynchronous mode of CMC, communication takes place in postponed time such as Email. Asynchronous mode of CMC does not require interlocutors to be online and

available at the same time in order to access and respond messages (Goggin, 2004; Hard af Segerstad, 2002).

With the emergence and evolution of Computer Mediated Communication, a new form of language emerged which is writing in nature but also carries features of speech. Hence, the emergence and evolution of CMC has initiated a real dichotomy for researchers as whether to consider CMC as a written or spoken form of language. While Baron (2000, p.248) gives a metaphor to CMC as "an emerging language centaur, part speech, part writing", Crystal (2004) suggests that CMC is more than just a hybrid of speech and writing. Crystal assumes that although CMC displays properties of both mediums, it holds features that neither of the two mediums have. Crystal suggests that it must be seen as a new variety of communication and he calls it a third medium.

SMS mode of communication shares many features with Computer Mediated Communication. Although IM, E-mail, blogs, and SMS are all broadly categorised under the umbrella term CMC, yet SMS uses mobile phone technology instead of computers or the Internet. As SMS is a mobile phone based mode of communication, it is abundantly used by the literate people of Pakistan. The other modes of CMC mentioned above are the Internet and computer based, so they are not in the easy access of most Pakistani population. Moreover, the mobile phone service in Pakistan is very economical, so it is in the reach of every Pakistani. When the concept of mobile phone service comes to SMS, its SMS packages are so tempting that people prefer texting instead of calling.

Hence, with the expansion of SMS during the recent years, the importance of the variety of language used in SMS has increased manifold. This penetration of SMS language to masses has also initiated researchers to analyse and explore many features of this new medium of communication.

For the reasons given above, in many countries of the world, research interest is increasing day by day in the social, cultural, and communicative aspects of mobile text messages in different sociocultural contexts. Academic research into the medium ranges from technological to ethnographic, its transformation of youth culture, and its impact on people's lives (Ling, 2008; Kasesniemi & Rautianen, 2002; Oksman & Turtianen, 2004). Various gender based differences have also been found in different settings (Ling, 2008). Other, more common areas of focus, have been the use of text messages in business and commerce (Bamba & Barnes, 2007), in political campaigning (Prete, 2007) and in media broadcasting (Enli, 2007).

But when this discussion comes to the pure linguistic research on SMS, the research in this field plays a relatively lesser role than the other related fields. The scholarly interest in the linguistic analysis of text messaging remained very slow during the first few years of the use of Short Message Service, but in the recent years this area has also got momentum in many parts of the world. Nevertheless, in Pakistan, the linguistic analysis of SMS continues to be a less examined area of research.

Overall, the available linguistic research in this field is largely limited to abbreviations and spelling variations. These features, although of interest, only touch the surface of a true linguistic understanding of text messages. Moreover, most available studies are based largely on interviews, diaries and observation. The actual SMS data based studies are confined to a limited amount of data, for example Hard af Segerstad (2002), and Thurlow and Brown (2003) employ 1152 and 544 text messages respectively. Moreover, SMS language is mostly unexplored through corpus linguistics.

Tagg (2009) conducted a large-scale corpus based study on about 11000 SMS, but that study is in the specific background of a monolingual English speaking country. In this regard, the present study is different in nature because it is based on the data

collected from bilingual speakers, whereas most of the earlier studies in the field were conducted in monolingual societies, especially in English speaking countries.

In the perspective of Pakistan, with an exception of a few small scale studies which are remotely related to the linguistic analysis of text messages (Aslam, Ahmad, & Sajid, 2011; Janjua, 2010; Rafi, 2008, 2010), the research on linguistic analysis of SMS is still needed. Even, the linguistic research in other modes of CMC is also overlooked in Pakistan, although it has become a very compelling field of linguistic inquiry in the world in the recent years. By now, no PhD level or any large scale linguistic study in any area of CMC has been conducted in Pakistan.

Pakistan is a bilingual/multilingual country where English and Urdu are frequently code-mixed and code-switched in the educated circles. Constitutionally, English is the only official language whereas Urdu is the national language of the country. In this bilingual setting, Urdu-English code-mixed language used in text messages in Pakistan is different from the SMS language used in text messages in monolingual countries.

Hence, it is this gap in the bilingual setting of Pakistan, the lack of thorough linguistic analysis of text messages based on a large corpus that motivated the researcher to conduct this study. The study aimed to fill this gap in the field of mobile text messages to increase the linguistic understanding of this prevalent and significant mode of communication in the context of Pakistan.

1.2 Statement of the Problem

The present study explores the linguistic features of the language used in text messages in Pakistan. The study examines SMS language in a large corpus of 5000 text messages from multiple aspects. The register specific language used in text messages is different in many ways from the formal language used for academic and official

purposes in Pakistan. So the purpose of the study is to examine, identify and describe those linguistic adaptations that characterise SMS language as a register specific variety of language. The term "linguistic adaptations" here refers to the *language modifications* made in the language used in text messages.

In the bilingual setting of Pakistan, the purpose of the study is to examine and describe Urdu-English code-mixed language used in text messages. The study investigates various issues of both English and Urdu used in text messages. The English language here refers to standard British English which is followed in Pakistan for official and academic purposes, whereas the Urdu language refers to standard Urdu used for academic purposes. Pakistani texters frequently adapt both the languages in text messages to make them suitable to the technology-based medium of SMS communication.

Keeping in view this background, the study examines *linguistic adaptations* made in text messages in six major categories, i.e. *lexical*, *syntactic*, *punctuation*, *space*, *code*, and *script* adaptations. More specifically, the study aims at exploring and describing *types*, *causes*, and *principles/patterns* of linguistic adaptations made in text messages. The study also provides *historical insights* into linguistic adaptations of text messages, and examines the *linguistic-cum-educational impacts and implications* of these adaptations. In this regard, the study focuses on the following objectives and research questions.

1.3 Objectives and Research Questions

In the light of the statement of the problem, the study aims at exploring linguistic adaptations made in text messages in Pakistan. For their continued use, most of these adaptations have become a feature of SMS language. To examine these adaptations, the study has been designed to achieve following research objectives:

1) To identify, categorize, and describe various linguistic adaptations made in text messages.
2) To sort out causes of various linguistic adaptations made in text messages.
3) To describe and elaborate principles/patterns of these linguistic adaptations.
4) To illustrate the adaptations in the historical perspective.
5) To highlight linguistic-cum-educational impacts of the adaptations, and the implications of these impacts.

Based on these objectives, following research questions have been formulated. These research questions provide a foundation to the study. Among them, the first one is the primary research question, whereas the next four questions are the secondary questions to further explore and support the primary question of the study.

1) What are the major types/categories of linguistic adaptations that characterise SMS language as a register specific variety of language?
2) Why are these linguistic adaptations made in text messages in the context of Pakistan?
3) How are various linguistic adaptations made in text messages in Pakistan?
4) Whether these adaptations are new in nature, or we have traces of such adaptations in history or in other modes of communication?
5) Do these adaptations have impacts on the conventional linguistic and educational standards, and what are the future implications of these impacts?

1.4 Significance of the Study

Primarily, the significance of the study is linked with the tremendous use of SMS mode of communication the world over. Only within Pakistan, the significance of texting is evidenced by the growing numbers of SMS traffic. According to the "Annual Report 2008" of Pakistan Telecommunication Authority (2008, p. 41):

SMS business grew very fast in last two years. Last year almost all cellular mobile operators offered lucrative SMS deals increasing the SMSs many folds. During the year 2007-08, cellular mobile operators generated more than 25 billion SMS, which shows about 200% growth in this segment. During last two years the segment witnessed tremendous growth.... The operators are targeting different segments of the customers and offering concessions and packages to these customers. Interestingly, the SMS seems more attractive to operators for revenue generation where they offered lowest tariffs during the quarter.

The statistics, presented in PTA "Annual Report 2008", highlight the tremendous increase in the use of SMS per year.

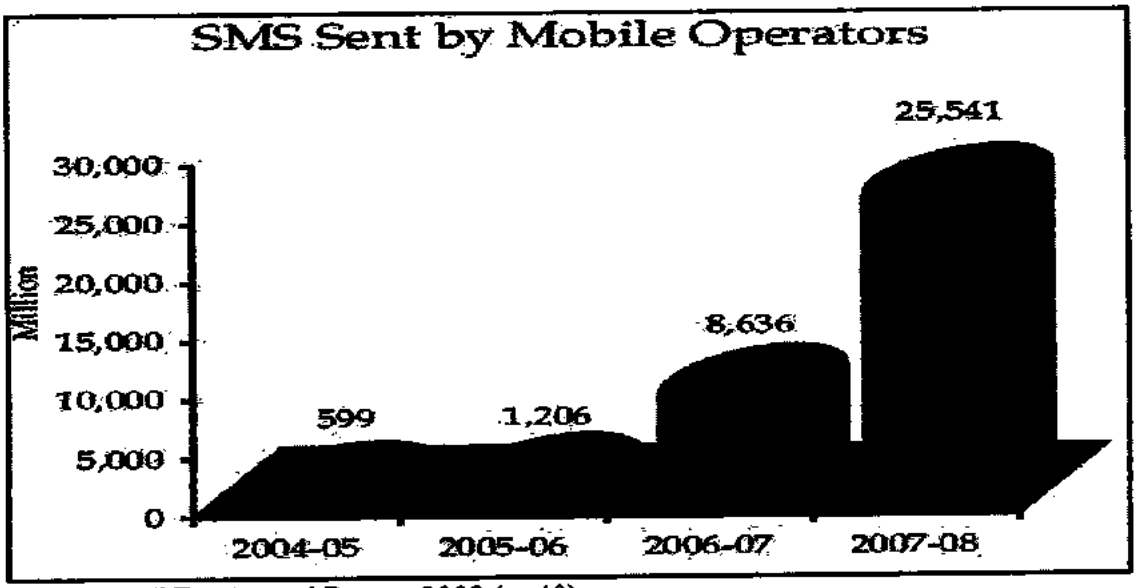


Figure 1.1 PTA Annual Report 2008 (p. 40)

The figure shows that the number of SMS sent per year has increased from 599 million (2004-05) to more than 25 billion (2007-2008) within the span of three years. Another study on SMS traffic in Pakistan (Pakistan Telecommunication Authority, 2010, p. 7) further highlights the tremendous explosion taking place in SMS use in Pakistan.

Table 1.1
SMS Traffic in Pakistan in Year 2009

Time Wise Traffic of SMS	Number of SMS
SMS traffic in year 2009	151.6 billion

SMS traffic per month	18.6 billion
SMS traffic per day	621.2 million
SMS traffic per second	7190
SMS traffic per subscriber per month	128

Pakistan Telecommunication Authority (2010) claims that in Pakistan SMS "traffic growth is tremendous between the Y2008 and Y2009 and expected to grow with the same rate in the upcoming years, due to its low cost as compared to voice tariffs and bundle packages are being offered by the operators" (p. 8).

On the basis of these statistics, it could be argued that an understanding of the language used through this mass medium of communication is very much necessary. Texting is very common both in the personal and professional communication of people in Pakistan. People can now update one another about the current news through SMS, can remind of mutual commitments and appointments, and can entertain themselves through quotes, jokes, poetic verses and anecdotes. At the professional level, many organisations in Pakistan are using text messages as a very effective communication tool. Educational institutes are using SMS technology to keep parents updated about the academic performance of their children. National Database and Registration Authority [NADRA] has started its SMS service of providing people's personal data for various purposes by sending the Computerised National Identity Card [CNIC] number to certain codes. The major services covered by NADRA are like car chassis number, identity of a person, voter registration, and number of mobile SIMs registered in one's name etcetera. Mobile phone companies use this service for various promotions and services. Banks are using SMS to keep their customers updated about their accounts. Mail delivery services are using SMS technology to keep their customers updated about their consignments. People use SMS to get appointments from doctors in hospitals. Even the business companies are using SMS technology to advertise their products. Hence, SMS

technology has become a very dominant communication tool in different spheres of the public and private life. At the same time, the public debate regarding the effect of texting on language and on children's literacy continues.

Bodomo (2010) states that mobile phone based SMS communication has not received as much attention as email and other CMC modes of communication. However, mobile phone is indeed more powerful mode of communication than computers and the Internet. About the significance of SMS language and the studies on SMS language, he further states:

> Along with other kinds of computer, the Internet or other digital technology mediated communication, SMS seems to be causing a silent evolution or even a revolution with regards to the linguistic and communication habits of people all over the world. This is especially so among the youth where one can safely say that more than 80% of people between the ages of 12 and 25 frequently use SMS as a mode of communication with their peers. (P. 111)

This enormous impact of SMS mode of communication, on the public, demands researchers to turn their attention to this area of research. It also demands them to try to answer the questions related to the impacts of text messages on various issues like language and education. Regarding these insights into the significance of the linguistic analysis of text messages, this study focuses on the linguistic as well as educational significance of SMS language.

1.4.1 Linguistic Significance

In the recent years, teachers and media critics have claimed that language is deteriorating under the influence of CMC and SMS modes of communication (Bodomo, 2010; Crystal, 2008b). In this regard, the present study accounts for various linguistic adaptations made by texters in text messaging. The study compares the register specific

adaptations of SMS language with standard British English and standard Urdu in the written mode. Some people regard the unique linguistic features of SMS as aliens to the standard language and treat them as a proof of language deterioration without in-depth investigation. The present study shows whether SMS language is deterioration of the standard language or a technology-based register specific variety of language in the perspective of language change and variation.

1.4.2 Educational Significance

In the past two decades, educationists, researchers and teachers have paid much attention to the use of CMC as an aid in language teaching (Greenfield, 2003; Warschauer, 1997). However, as some researchers declare that CMC language has already evolved as a new form of language and literacy (Bodomo and Lee, 2002; Crystal, 2004), the educational and pedagogical implications of this language are a prompt concern for language teachers. It is important for language teachers to realize and understand language in the new scenario, with the new forms of expression. The new forms of the English language imply for language teachers to be able to incorporate authentic elements into teaching. Teachers may have to consider incorporating the authentic and most frequent SMS language elements into their teaching. Language teachers may not only teach students how to use a language, but also show students how a language is used in the society. In this way, this study offers both prescriptive and descriptive insights into the English language from linguistic as well as educational perspective.

1.5 Limitations and Delimitations

Due to the limitations of time, the sample size of the present study is delimited to 5000 text messages (SMS). Still, being the personal as well as specialized corpus, 5000

SMS all collected from the real users is not a small number considering various studies in the field where the number of SMS is mostly lesser than one thousand.

For the questionnaire survey, 500 texters (who also provided the first 2500 SMS in the corpus) are delimited to two big cities of Rawalpindi and Islamabad, as these two cities were in the easy approach of the researcher. Moreover, the people belonging to different regions, ethnic groups and social classes of Pakistan live in these two cities. Bodomo (2010) states that young people are the most frequent users of text messages. In this study too, text messages collected for SMS-Corpus mostly belong to the youth.

In addition, the corpus is delimited to the real users only and no data from websites has been taken. Furthermore, text messages sent through mobile services as publicity stunts have been excluded because 1) they use controlled language, and 2) junk messages are sent to all users by the mobile service providers. In this regard, all respondents of the questionnaire survey were asked to provide only those text messages which they sent or received during their interpersonal communication.

Moreover, the present study is delimited to SMS communication only, as SMS is the most frequently used technology-based mode of communication in Pakistan. All other modes of CMC are confined only to a limited number of people, whereas SMS has become the most popular source of communication among the masses in Pakistan. Therefore, other technology-based modes of communication like instant messages, e-mails, blogs, or multimedia message service [MMS] etcetera have not been examined in the study. Hence, this study is based on the linguistic analysis of mobile phone based text messages only.

1.6 An Overview of Research Methodology

Data collection and data analysis are two very important steps in any empirical study. Therefore, this overview of the research methodology covers these two vital

issues. The detailed research methodology is given in Chapter Three, and so, this section offers only a cursory view of the methodology adopted in the study.

1.6.1 Data Collection

The data used in this study is of two types. One type of data is the metalinguistic data of the perception of texters about their use of SMS language. This metalinguistic data is mainly based on texters' opinions about the language used in text messages. The data has been collected through both close-ended and open-ended questions. The tool used for the collection of metalinguistic data is a questionnaire comprising 13 questions (see Part-A of Appendix-A).

The other type of data is based on a corpus of 5000 SMS, referred as SMS-Corpus in the study. SMS-Corpus includes personally transmitted SMS among Pakistani texters during their interpersonal communication. These 5000 text messages were collected through two tools. Tool one is the above mentioned questionnaire. Through this tool, 500 texters were asked to provide their actual interpersonal SMS in the given place of the questionnaire. Through this tool, 2500 SMS were collected (see part-B of Appendix-A). The second tool for the collection of text messages was the *Forward Message* feature of mobile phone sets. Regarding the use of this tool, the researcher asked 50 of his direct friends and students to forward about 20 to 50 SMS from their Inbox/Outbox to the researcher's mobile phone. In this way, 1500 hundred SMS were collected, with an average of 30 SMS from each respondent. Moreover, the researcher requested two of his friends to preserve both Inbox/Outbox SMS in their mobile phones for a few months, until the total number crossed 1000 SMS. Consequently, the remaining 1000 SMS were collected. In these two steps of *forwarded messages*, 2500 text messages were collected from 52 respondents, and the total number of 5000 SMS was accomplished (see section 3.2.2 for details).

1.6.2 Data Analysis

The present study is largely descriptive and aims at describing various linguistic adaptations made in text messages. Hence, in the light of research questions and scheme of the study, the data has been divided into two broader categories, i.e. *metalinguistic data* based on the perception of texters about SMS language, and *linguistic data* based on SMS-Corpus of 5000 text messages about the linguistic adaptations made in text messages.

The analysis of *metalinguistic data* has been presented in Chapter Four of the study. Results have been tabulated and described in the same chapter. In this regard, the questionnaire based data has been statistically analysed by using *MS Excel* programme of *Microsoft Office-2007* (see section 3.5 for details).

The analysis of *linguistic adaptations* made in SMS-Corpus of 5000 text messages has been mainly analysed through the concordance software "AntConc" of corpus linguistics. The results of the linguistic analysis have been tabulated, described and discussed in Chapters 5 to 8 of the study. Moreover, the results of both types of metalinguistic and linguistic analyses have been triangulated, and interlinked with each other at various points in the dissertation (see section 3.4 for details).

1.7 Organization of the Dissertation

The study is divided into nine chapters. Chapter 1 is the introductory chapter which starts with the background on which the foundation of the study is laid. It includes some primary issues of the study like statement of the problem, objectives and research questions, significance of the study, and limitations and delimitations of the study. It also offers a quick look on the research methodology and a brief account of the organization of the dissertation into various chapters.

Chapter 2 deals with the review of literature. It starts with a detailed introduction to corpus linguistics and then moves towards the assumptions of language change and variation. It also throws light on the nature of CMC and SMS modes of communication. Then, the review of key studies on SMS language is offered. Finally, the chapter throws light on SMS studies conducted in Pakistan.

Chapter 3 presents research methodology and data description. It comprises areas like research design and construction of SMS-Corpus, description of SMS-Corpus data, analysis of SMS-Corpus, collection and analysis of metalinguistic data, and description of questionnaire respondents.

Chapter 4 is based on the questionnaire data, and deals with perceptions of texters about SMS variety of language. The chapter does not concern the use of SMS language itself, rather reflections of texters on SMS language. The analysis is based on responses of 500 texters collected through a questionnaire. Hence, this chapter deals with the metalinguistic awareness of texters about SMS language.

Chapter 5 presents the analysis of lexical adaptations in text messages. The chapter has been mainly divided into four sections. The first section gives an overall introduction to lexical adaptations in text messages. The next two sections discuss morphological and phonological principles and patterns of lexical shortenings. Section four presents an overall perspective of spelling and letter-case variations in text messages. The chapter also presents critical discussions on all areas of lexical adaptations covered in the chapter.

Chapter 6 addresses the issue of syntactic adaptations made in text messages by texters. The chapter covers three types of syntactic/grammatical adaptations; i.e. 1) omissions/ellipses of words, 2) reduplications/repetitions of words, and 3) deviant uses

of grammar. The chapter discusses patterns and principles of these adaptations and also offers possible reasons of these adaptations in the light of key works on the topic.

Chapter 7 investigates punctuation and space adaptations through examples from SMS-Corpus. In the chapter, the adaptations of both types have been analysed for their deviated uses in text messages, whereas punctuations have additionally been analysed for their creative uses in text messages.

Chapter 8 deals with code and script adaptations in the bilingual setting of Pakistan. It includes code-switchings and code-mixings in text messages between Urdu and English. It also addresses the issue of the excessive use of the Roman script/alphabet for the Urdu language in text messages. Moreover, the chapter presents the corpus based Romanization scheme for Urdu.

Chapter 9 is the concluding chapter of the study. It includes the two key aspects of the study that are findings and conclusions. Both the key concerns of the study have been grouped in the last chapter. Findings of the study have been arranged addressing all five research questions of the study one by one. Conclusions, on the other hand, address individual adaptations one by one to highlight major discoveries of the research.

CHAPTER 2

LITERATURE REVIEW

This chapter starts with a detailed introduction to corpus linguistics which offers methodological foundations to the study. It accommodates assumptions of language change and variation which provide theoretical foundations to this study. It also throws light on the nature of CMC and SMS modes of technologically influenced communication. Then, the review of literature on SMS language is offered. Finally, the chapter throws light on the studies conducted on SMS in Pakistan.

2.1 Corpus Linguistics

Corpus linguistics is a linguistic branch and paradigm within linguistics, such as sociolinguistics or psycholinguistics (Tognini-Bonelli, 2001). However, seeing the linguistic analyses that are conducted under this paradigm, it becomes obvious that corpus linguistics can be seen as a methodology used for the analysis of linguistic data, i.e. "a methodological basis for pursuing linguistic research" (Leech, 1992, P. 105), than a separate paradigm within linguistics. Thus, corpus linguistics is a linguistic branch as well as a research methodology. As a linguistic methodology, it is a scientific process of studying language based on samples of corpora or real world texts (Meyer, 2004).

Etymologically speaking, the word *corpus* is a Latin term that means *body* in English. In modern linguistics, a corpus (plural corpora or corpuses) is a large and structured set of texts which can be stored and processed electronically. Although corpora analyses have been manually done in the past, yet corpora are now mostly processed by computers. Modern day machine readable corpora are used in multiple analyses using both qualitative and quantitative approaches. They are used for statistical analysis, checking occurrences, hypothesis testing or validating linguistic rules. A corpus

may contain texts in a single language or in multiple languages, i.e. monolingual and multilingual corpora (McEnery & Wilson, 2001; Meyer, 2004).

2.1.1 Historical Development of Corpus Linguistics

A milestone in modern corpus linguistics was the publication of a book titled *Computational Analysis of Present-Day American English* by Kucera and Francis (1967). Their book was based on the analysis of *Brown Corpus,* a million word corpus of current American English developed in early 1960s. They applied various computational techniques to analyze corpus data utilizing different elements of linguistics, sociology, psychology, language teaching, and statistics. Shortly thereafter, a Boston publisher named Houghton Mifflin Harcourt approached Kucera and Francis to use Brown corpus for their new *American Heritage Dictionary* that was the first dictionary which incorporated corpus linguistics in lexicography (Dash, 2005; Leech, Hundt, Mair, & Smith, 2009).

Other publishers followed this approach for designing their dictionaries. The British publisher Collins used the *Bank of English* corpus to compile *Collins COBUILD English Language Dictionary* for users learning English as a foreign language. In the same manner, the *Survey of English Usage* corpus was used in the development of the *Comprehensive Grammar of English* in 1985. Following Brown Corpus, a number of similar corpora were developed at different times and places, for example *Australian Corpus of English, International Corpus of English,* and *British National Corpus* (Kennedy, 1998).

The first computerized corpus of transcribed spoken language was developed in 1971 by the *Montreal French Project* having one million words. In addition to these corpora of living languages, computerized corpora have also been made of collections of texts in ancient languages. An example of such corpora is the *Quranic Arabic Corpus* of

the Qur'an developed in 2009. This is an annotated corpus of the language of the Qur'an which includes POS tagging, syntactic analysis, and morphological segmentation. A detailed view of collections and data archives of the world corpora are available at David Lee's Web page: http://www.uow.edu.au/~dlee/CBLLinks.htm

2.1.2 Theoretical Issues of Corpus Linguistics

In early 1960s, at the time of the development of Brown Corpus that was the first computer based corpus, generative grammar dominated linguistics. At that time, all linguistic theories and methodologies were subjected to adhere with the principles of generative grammar. Hence, corpus linguistics was not welcomed at that time; rather it was bitterly criticized by linguists like Chomsky. As a result, the efforts of Kucera and Francis, the pioneers of corpus linguistics, to create Brown Corpus were not accepted in a friendly way by many linguists (Meyer, 2004).

The main reason of rejection of corpus linguistics in the start was the authority of Chomskyan theories. Hence, less emphasis was placed on empirical observations. With the authority of his works, linguistics remained away from empiricism for many years. Following de Saussure, he made a distinction between two approaches of language viz. the theory of language system and the theory of language use. These two approaches were termed as *competence* and *performance* respectively (Meyer, 2004).

Chomsky, the founder of generative grammar, was more interested in competence than performance. He was against corpus linguistics as this approach was foremost based on actual performance data than competence. Nonetheless, corpus linguistics research continued in spite of early criticisms, and it even strengthened due to technological advances in computer software. It is due to the advancements in computer technology and the modern information technology that texts of several million words in

length can be processed (Sinclair, 1991). Nowadays, there are linguists with valid reasons that speak in favour of using corpora in linguistic analysis (Biber, 1995).

At present, the reasons of the rejection of approaches based on Chomsky's theories are that they do not cope with vast areas in the language study like register variations. In the study of language, registers and variations probability play an important role in selecting certain combinations of meaning with certain frequencies. However, the bitter criticism of corpus data arising from the Chomskyan tradition has led corpus linguists to remedy the drawbacks of corpus data like its representativeness.

2.1.3 Methodological Issues of Corpus Linguistics

As corpus linguistics is a linguistic methodology, thus there are a number of methodological issues to be addressed for better understanding and clarification of corpus linguistics as a linguistic methodology.

2.1.3.1 Corpus Driven versus Corpus Based Approaches

In corpus linguistics methodology, corpus-driven and corpus-based are two different approaches to analyse corpus data. The corpus-driven approach is a method in which already existing corpus data becomes the empirical basis to reach an assumption or theory without any prior assumptions and expectations. The corpus-driven linguists aim to build theory from scratch. They make claims or assumptions exclusively on the basis of observed corpus data (Elewa, 2004; Tognini-Bonelli, 2001).

On the other hand, the corpus-based approach is the methodology that uses corpus data to support or prove an existing assumption or theory, or to retrieve illustrative samples for a theory. It is a method that investigates corpus data to test, confirm, or improve linguistic theories and assumptions. Hence, the evidence of corpus data is used to support or examine an already existing theory rather than as a determining factor of a theory or assumption. Mostly, this approach uses corpus data to support a

theory or a claim rather to challenge it. Normally, this approach does not challenge the pre-existing theories and cannot render unexpected results, yet it is used to extend, elaborate, or improve some classic assumptions (Elewa, 2004; Tognini-Bonelli, 2001).

Although, having established that both corpus-driven and corpus-based approaches have differences, yet "the sharp distinction forced between the corpus-based vs. corpus-driven approaches to linguistics is in reality fuzzy" (Xiao, 2008, p. 996). This means that the distinction between the two is overstated and it leads to deception rather than clarity. Hence, the practical purposes of corpus linguistics can be achieved through the marriage between two approaches instead of dividing them into two opposite poles that cannot join together.

2.1.3.2 Corpus Design and Sampling

The selection of corpora for analysis is not done haphazardly, because a corpus is considered as a planned and structured collection of texts. Therefore, prior to the process of collecting a corpus there should be theoretical research to specify what type of data, time period, language variety, size of sample, and corpus design is needed (McEnery & Wilson, 2001; Meyer, 2004; Sinclair, 1991).

Indeed, the nature of corpus design depends on the purpose of the corpus such as a basis for a dictionary; creation of a word frequency list; study of some linguistic phenomenon; study of language change; or anything else. It is due to the diversity of purposes of corpora that there is no consensus among corpus linguists as to the procedures or the selection criteria to be followed in corpus designs. Yet, one of the first considerations in constructing a corpus is to specify for whom and for what purposes the corpus is designed viz. for personal research, or to serve as a public resource (Elewa, 2004; Kennedy, 1998; Sinclair, 1991).

Once, the type, purpose and content of a corpus is decided, the next step is to select and sample actual texts to become the body of the corpus. In this regard, any selection of texts is considered a sample, irrespective of being representative or not, but a corpus must be representative in order to be appropriately used as the basis for generalizations concerning a language as a whole. However, we have to bear in mind, that there may be a corpus that is designed to represent not the language as a whole but one particular genre (Elewa, 2004; Kennedy, 1998; Sinclair, 1991).

2.1.3.3 Corpus Annotation

As a research methodology, corpus linguistics has generated a number of research methods. In this perspective, Wallis and Nelson (2001) introduced a 3A model of corpus analysis that is 1) *Annotation, 2) Abstraction,* and 3) *Analysis.* Among the given terms, *annotation* is the most known and important term and a method in corpus linguistics. Annotation consists of the application of a scheme to texts. It may include structural markup such as tagging, parsing, and numerous other representations.

The most known annotation of a corpus is part of speech tagging, commonly known as *POS tagging*. In POS tagging, the information about each word's POS category (e.g., noun, verb, adjective, adverb, etc.) is added to the corpus in the form of tags. Another common feature of annotated corpus is *lemmatization* that is indicating the lemma (base) form of each word. For the concerns of meaning, corpora can be semantically annotated. Such type of annotation is call *semantic tagging*. A number of smaller corpora are also parsed (syntactically tagged) through tree diagrams to determine their grammatical structures. These corpora are normally termed *Tree Banks* or *Parsed Corpora* (Gries, 2009; McEnery & Wilson, 2001).

Although corpus annotation allows some sophisticated analyses to be carried out, however, "sociolinguists should not feel that they need to be compelled to be fascinated

by corpus annotation" (Baker, 2010, p. 16). The purpose of annotation of corpora is to do some additional linguistic research. An added advantage of developing an annotated corpus is that some future users can easily perform more experiments on annotated corpus. Therefore, linguists with other interests and differing perspectives than the originators can exploit this work. But it all depends upon the nature of the study whether some type of annotation to the corpus is needed, as annotated corpora is not mandatory for all types of studies. Many studies are conducted only through concordance tools while no annotation is added with corpora. Hence, corpus annotation sometimes serves as an added advantage but it is not a binding for all corpora.

2.1.3.4 Corpus Analysis Tasks

Once a corpus is prepared in electronic form, the next step is to process it by computer for its use in linguistic research. Since most corpora are incredibly large, it is not justified to search a corpus without the help of some corpus processing software (Elewa, 2004). Hence, we need computer based tools to analyze electronic data. There are a number of tools designed for such a purpose. Many corpus linguists like Baker (2010), Biber, Conrad and Reppen (1998), Kennedy (1998), McEnery and Wilson (2001), and Meyer (2004) give their views of such tools and processes to analyze a corpus. There are different processes and computer techniques that can be used in analyzing corpus data. An overview of mostly used processes and techniques is given here.

- **Frequency/Word Lists**

Baker (2010) considers that frequency is the foundation of corpus linguistics. He asserts that "frequency refers to the number of times something occurs in a corpus" (p. 19). Frequency lists can be made by identifying every word form in the text, counting identical forms and classifying them according to a particular order, i.e. alphabetically,

or according to their frequency. This can be done either in descending or ascending order. Listing of words according to their frequencies shows us the number of occurrences of single words in the text. Thus, by analyzing a list, one can get an idea of what further information would be worth acquiring. Moreover, one can make guesses about the structure of the text, and can focus on analysis.

- **Concordances**

Baker (2010) states that, "a concordance is a table of all of the occurrences of a linguistic item in a corpus, presented within their linguistic contexts" (p. 21). Different to word frequency lists, the concordance of words is the representation of words within their contextual environment. This can give extra information about the nature and behaviour of words. This process is generally termed KWIC (key word in context) in corpus linguistics. In a concordance program, the searched word can be highlighted by putting it in the centre of each line, with a space on each side. The arrangement of each searched word is alphabetical according to the left-hand or the right-hand context.

- **Collocation**

In addition to KWIC and word frequency lists, most corpus linguistics tools also offer the option of collocation. Baker (2010) states that collocation is a way of demonstrating frequent relationships between words. According to him, "if two words collocate, then they have a tendency to occur near or next to each other in naturally occurring language use" (p.24). In addition to simple collocations of related words, if the corpus analysis package is a bit more sophisticated, it may also provide with lists of collocates based on some statistical tests.

2.1.4 Corpus Analysis Software/Programmes

In empirical linguistic research, corpus analysis has become an essential method of linguistic inquiry, starting from lexical semantics and grammars to sociolinguistics,

psycholinguistics, and discourse pragmatics. Apart from more computationally advanced and varied tasks associated with corpora, such as lemmatization, tagging, and parsing, corpus-linguistic research revolves around a number of basic but essential procedures. These include frequency lists, concordances and collocations (Wiechmann & Fuhs, 2006).

Among these elementary tasks, the creation of concordances, that is formatted display of all occurrences of a particular type in a corpus, may be considered an important and fundamental task. Thus some reliable, fast, and user-friendly search-and-retrieval software is of great value to any researcher working in this field (Wiechmann & Fuhs, 2006). Concordance packages are mainly concerned with the development of frequency lists, concordances and collocations. Hence, an introduction to this very commonly used corpus analysis tool is given here.

Wiechmann and Fuhs (2006) discuss ten programmes that are very commonly used. According to them "the first three are commercial solutions, the remaining seven are freeware" (p. 109). These programmes are 1) MonoConc Pro, 2) WordSmith Tools, 3) Concordance by R. J. C. Watt, 4) Multi Language Corpus Tool by Scott Piao, 5) ConcApp by Chris Greaves, 6) AntConc by Laurence Anthony, 7) Aconcorde by Andrew Roberts, 8) Simple Concordance Program by Alan Reed, 9) Concordancer for Windows [WConcord], and 10) TextSTAT by Matthias Hüning.

David Lee provided an exhaustive list of more than 200 corpus tools concerning different tasks on his web link: http://www.uow.edu.au/~dlee/software.htm
In his list, the first two entries of programmes are 1) Concordancers, Search Engines, Text-analysis Tools, and 2) Web-based Concordancers. In his categorization, the tools that are free and more useful are at the top of the list. As his list is very exhaustive, thus only the first 10 tools are given here in the same order as given on his web page. These

tools are 1) *AntConc by Laurence Anthony*, 2) *WordSmith Tools*, 3) *MonoConc Pro*, 4) Dexter by Gregory Garretson, 5) *Aconcorde by Andrew Roberts*, 6) Qwick, 7) *Concapp by Chris Greaves*, 8) Conc by John Thomson / SIL, 9) Multiconcord, and 10) Conc by Mario Saraceni.

In a comparison of both the lists, it is interesting to note that five programmes (written in italics in the second list) are common in both lists. Paul Baker (2010, p. 8) also provides a list of seven popular corpus tools. In his list, the first three tools, namely 1) WordSmith Tools, 2) AntConc, and 3) MonoConc Pro are also found in the above two lists given by David Lee, and Wiechmann & Fuhs. Hence this comparison of the corpus analysis tools shows that various reviewers of the corpus linguistics tools agree on the basic and the most essential concordance tools given above. This agreement of reviewers on concordance tools increases the worth of these basic concordance software packages.

2.2 Language Change and Variation

One of the main characteristics of language is change and variation. The study of language change and variation is one of the favourite topics of historical linguistics, dialectology and sociolinguistics. Baker (2010, p. 5) differentiates between language variation and change as "*synchronic variation* and *diachronic variation*" respectively. According to him, language change is a diachronic/historical process whereas language variation is a synchronic process. To further highlight both these processes of language change and variation, the related issues are given in the following sections.

2.2.1 Language Change

The change of any human language over time is a natural process. All living languages change over time and if a language does not change through time, it is likely to become obsolete, or extinct in terms of its usage (Fromkin, Rodman, & Hyams, 2003).

In the words of Lehmann (1973, p. 1) "languages change constantly, a fact that becomes most noticeable when we read texts a few centuries old". He supports this assumption by making references to the language used by Shakespeare in his plays. According to him various kinds of change are noticeable in almost any passage of Shakespeare. He categorizes historical changes in any language in three main categories, i.e. 1) sound change or change in phonological systems, 2) syntactic or morphological change, and 3) semantic change or change in the lexicon.

Languages also change over time under the impact of *language contact* of two or more languages or language dialects. This type of change is specific to bilingual or multilingual societies. In this regard, Bauer (2002, p. 8) states:

> Typically, in the colonial situation, a lot of speakers of many different dialects come face to face, and in the short term the result is a period of diversity where everyone is accommodating to everyone else. During this period, speakers may not be aware of any trends or emerging patterns. Gradually, however, order emerges from the chaos, the trends become clearer and a new mixed dialect is formed.

The new mixed dialect usually keeps features of the various dialects involved in the formation of the new mixed dialect. But in the most of cases, the form used by the majority of speakers survives in the new mixed dialect.

In the specific context of British English in the colonial period, when "English speakers left Britain, they started to meet various kinds of entities and actions which were not familiar to them" (Bauer, 2002, p. 84). Gradually, they started to borrow or coin words for unfamiliar things from other languages. Later on, these words became part of the colonial Englishes, and then of standard British English.

In the context of Urdu-English language contact in Pakistan, the issue is of historical nature. Pakistan and India both remained a British colony till the independence of Indian subcontinent into two independent countries India and Pakistan in 1947. This historical position of Pakistan as a former British colony has provided a historical ground for Urdu-English language contact.

In the historical perspective, from the day of the independence of Pakistan, Urdu is the only national language of Pakistan, whereas English enjoys the status of the single official language of the country. Although it was decided in the 1973 Constitution of Islamic Republic of Pakistan that the English language would gradually be replaced with the Urdu language as the official language, but this change could not be achieved despite many efforts of the promoters of Urdu. Hence, the historical as well as the modern global impact of English in Pakistan has resulted in the continuous language contact of Urdu and English, and this contact leads to code switching and code mixing. Rasul (2008) states that a new code "Urdish" is the outcome of the hybridization of Urdu and English in Pakistan.

Moreover, this Urdu-English language contact has also influenced the writing script for Urdu. Urdu is traditionally and academically written in the Urdu (Perso-Arabic) script, which in turn is the modification of the Arabic script. Yet, the history of Romanization of the Urdu language goes back to pre-Pakistan British era when Romanized Urdu was used by the British to use the Urdu language for their colonial purposes. In this regard, John Gilchrist (1796) Romanization system is a pioneering work in this field. At present, the abundant use of the Internet and mobile phone technologies by masses has revived Urdu Romanization in Pakistan. Today, the abundant use of the Internet and mobile phone technologies has once again led people in Pakistan to use the Roman script for Urdu.

2.2.2 Language Variation

Biber (1995) is of the view that variation is inherent in human language. A single speaker may use different linguistic forms on different occasions, and different speakers of a language may say the same thing in many ways. These variations in language occur due to a number of factors. These factors are like the purpose of communication, the relationship between the speaker and the hearer, the production circumstances, various demographic affiliations, the influence of one language on another, the introduction of new technologies, the spread of certain changes through an entire speech community, and the use of a language in specific situations and professions (Bodomo & Lee, 2002; Fromkin et al., 2003).

Hence, languages are categorized into different varieties on the bases of geographical, social, professional, gender, functional and situational differences. In this regard, language varieties can be broadly divided into two kinds, i.e. "*registers,* referring to situationally defined varieties, and *dialects,* referring to varieties associated with different groups of speakers" (Biber, 1995, p.1).

Yule (2006) states that variation according to use in specific situations is studied in terms of register. In this way, language like any other social activity should be appropriate to the speaker and the situation in which it is used. For this reason, men and women speech is different in different communities. In the same way to give a sports commentary in the language of the Bible, in a religious sermon, or in legal language would be a bad mistake or a joke (Trudgill, 1995). Language in this way varies not only in terms of speakers but also in terms of social settings, contexts and situations in which it is used.

Under the global needs of a common language, English has become an international language and a *lingua franca* of the world. Therefore, English has diverged

into more regional and social varieties than many other languages of the world. According to the classification given by Biber (1995), these regional and social varieties can be termed as dialects and registers of English respectively. Regarding the differences found in these varieties, Bauer (2002, p. 100) states that there are significant differences between the Englishes spoken in England and in other parts of the world. These differences of the English language are so significant that if they move further apart then eventually it will not be appropriate to consider them as varieties of the same language.

According to Bauer (2002), if we look back into the European history, Latin is an example of such variations. Latin changed so much in the course of a millennium that it was no longer called Latin in the places where it was used. Its various dialects became so different that speakers from different regions could not even understand one another. Over the course of time, Latin was divided into French, Italian, Portuguese, Spanish, and Romanian. So in a few hundred years, we may also say that people speak British, American, Australian, and South African instead of saying that they speak English.

Everything we know about language history suggests that varieties of any language diverge when left to themselves. As an example, there is ample evidence that local, regional and national varieties of English continue to diverge. At the present time, the impact of modern information and communication technologies has powerfully influenced languages over the world to change and vary, and English is at the forefront of these threats to languages. These impacts have divided many languages into situational and technology specific varieties and registers. This issue is further elaborated in the next section.

2.2.3 Impacts of Technology on Language

Among the different reasons of language change and variation, the emergence and growth of computer, the Internet, and mobile phone technologies is a dominant

factor of language change in the modern era. At the forefront, the impact of computer technology has stimulated language change and variation in many ways. In comparison with the language in print medium, the language used in CMC mode is different in style due to the conditions of speed and space (Ross, 2006). The impact of CMC technology on language change and variation has been confirmed by Danet and Herring (2007, p.28) in the following words: "Features of CMC technology can shape language use". In the words of Ross (2006), computer mediated language is generally distinguished by the use of contracted forms, elision of sentence subject, and modified spellings.

Regarding this impact of technology on language change and variation, a theoretical framework of language change and variation was proposed by Bodomo and Lee (2002), and also given in Bodomo (2010). It is labeled as *Technology-conditioned approach to Language Change and Use* [TeLCU]. The framework emphasizes the causal relationship between the emergence of new information communication technologies [ICTs] and the development of new forms of language and literacy. The framework highlights that with the emergence of new technologies, new concepts and ideas may emerge in due course. In order to express the new ideas emerging along with the new media and tools, new forms of language and literacy are often necessary. Consequently, the ways in which people use their language may face changes due to the new media of communication (Bosco, 2007).

Differing views such as Labov (2001) may suggest that the introduction of new media of communication may not always impose a prevalent effect on language, and the evidence is that the influences of media like television and radio on language are not significant ones. However, the technologies with which the above framework deals are Information and Communications Technologies instead of those old-fashioned ones. One of the products of the TeLCU model is new forms of language which include e-

terminologies, acronyms and abbreviations used in cyber talk, and mobile phone language (Bosco, 2007).

2.3 Computer Mediated Communication (CMC)

Computer Mediated Communication is an outcome of the impact of computer technology towards the change and variation of written language over time. This new form of written language is the result of information technology used through computer networks. The term CMC was first invented by Hiltz and Turoff (1978). They referred to CMC as the human communication using computer technology. According to Herring (2001), "computer mediated discourse is the communication produced when human beings interact with one another by transmitting messages via networked Computers" (p. 612). Many alternative terms such as cyber communication, electronic communication, and online communication have also been used to describe CMC (Bosco, 2007).

Hard af Segerstad (2002) divides CMC into a*synchronous* and *synchronous* modes. Moreover, she includes mobile phones based communication in CMC besides computers and the Internet. She asserts that CMC is human language mediated by computers and other modern technologies like the Internet and mobile phones. According to her, typical examples of asynchronous mode of CMC include electronic mail (E-mail) and short message service (SMS) while synchronous ones are web chat and instant messaging (IM) such as MSN and Yahoo messengers. In this regard, a brief introduction to both synchronous and asynchronous modes of CMC is given in the following sections.

2.3.1 Synchronous CMC

Hard af Segerstad (2002) states that "like spoken interaction, synchronous CMC requires its interlocutors to be online simultaneously" (p. 59). This definition does not mean that the message sender and the receiver have to be present physically in the same

room. Rather, synchronous CMC makes written communication become interactive written discourse. Communication relies mainly on text and graphic means. Examples of synchronous CMC are various forms of real time chat such as web chat and internet relay chat (IRC) in which a group of people may interact with one another in written communication. Users enter the conversation by typing their messages which are displayed in the chat window to all users who have logged in. Another well known form of synchronous CMC is instant messaging (IM) such as MSN and Yahoo messengers. IM is primarily a synchronous genre like IRC but it can also be asynchronous as many internet messengers have a function which enables users to store their chats in the computer and read messages sent to them afterward (Yan, 2008).

Crystal (2004) also criticizes different forms of synchronous CMC as they are not fully synchronous such as spoken face to face interaction. He points out that chat rooms, IRC and instant messaging are too constrained by the slow response time to be a good analogy of speech. Herring (1999) also criticizes the synchronicity of all modes of CMC as they all lack simultaneous feedback of the spoken conversation. In synchronous CMC users do not see or hear their interlocutors and hence cannot access the non-verbal responses. According to Allwood (2000), spoken conversational interaction shows that simultaneous feedback plays an important role in signaling listenership, timing turn taking effectively and maintaining continuous interaction. Consequently, the absence of simultaneous feedback results in overlaps and discontinuity in the turn sequences.

2.3.2 Asynchronous CMC

As stated by Hard af Segerstad (2002), "text-based asynchronous CMC is communication that does not require participants to be online and available at the same time or place in order for communication to take place successfully" (p.58). This mode of CMC gives the sender enough time to plan and edit the message, as messages are

composed off-line. E-mail (internet based) and SMS (mobile phone based) messages are the typical examples of asynchronous CMC.

Both E-mail and SMS do not require the sender and the receiver of the message to be online simultaneously, when the message is sent. Hence, asynchronous CMC is a delayed time activity in principle, though the receiver may read and respond the message instantly in many cases. In principle, asynchronous CMC is not a real time activity and is close to traditional postal communication. In this sense, it is identical with the written communication. But on the other hand, as the communication technologies are very easily available to people round the clock, hence different modes of asynchronous CMC such as E-mail and SMS sometimes seem identical to synchronous CMC. This shows that rapid developments and changes in communication technologies and human behaviour have also abridged the gulf between asynchronous CMC and synchronous CMC (Crystal, 2004). In this regard, Zurhellen (2011) states that although text messages are asynchronous in nature, yet "it would not be incorrect to understand a text message, at least metaphorically, as a kind of call to which the receiver must respond or risk disturbing the discourse expectations (p. 639). In the light of above comments, it seems that the clear distinction between synchronous and asynchronous mode has blurred and the boundaries between them have squeezed under the influence of modern communication technologies. Therefore, some communication modes like SMS can be termed both asynchronous as well as synchronous modes of communication.

2.3.3 Linguistic Features of CMC

In the recent years, the topic of CMC has become one of the major areas in the field of Linguistics, and the popularity of CMC among the masses has generated a large amount of CMC data for research purposes (Bosco, 2007). Although the history of CMC from the perspective of linguistics does not seem to be long, yet since 1990s the number

of studies has been notably increasing (e.g., Baron, 1998; Crystal, 2004; Hard af Segerstad, 2002; Herring, 1999, 2001).

The major studies of the past in textual CMC are holistic in nature. They cover both asynchronous and synchronous modes of CMC. Moreover, they are based on rather limited corpora in comparison to the individual studies in the particular fields of CMC. In these studies the data are usually collected from CMC systems which are accessible to the public. In this regard, Crystal (2004) presents various interesting findings on the language use of different modes of CMC. In the first place, he declares the computer mediated language as *netspeak*, and states that it is a written language but has been pulled in the direction of speech. Yet, it is identical to neither of the two, and carries the properties of both. Hence, it is a third medium.

Crystal (2004) discusses a number of significant research issues regarding the language used in CMC. Different kinds of internet scenarios are introduced including both asynchronous and synchronous modes of CMC such as e-mail, chat groups, virtual words and also the World Wide Web (www). He provides a comprehensive overview across different modes of CMC, covering their linguistic features in CMC setting. He concludes the impact of the Internet or CMC on language in these words:

> What is truly remarkable is that so many people have learned so quickly to adapt their language to meet the demands of the new situations, and to exploit the potential of the new medium so creatively to form new areas of expression. It has all happened within a few decades. The human linguistic faculty seems to be in good shape, I conclude. (Crystal, 2004, p. 242)

Hard af Segerstad (2002) also offers a comprehensive study on the use and adaptation of written language to the conditions of CMC. Her study covers both asynchronous and synchronous modes of CMC. In asynchronous mode, she analyses Email and SMS while

in synchronous mode her focus is on Web Chat and Instant Messaging. As the main contribution of her study, she "challenges the popular view that language is deteriorating and that it is deteriorating at a faster rate because of CMC" (p.263). She declares that "language is neither deteriorating nor are modern users less able to use it successfully now than in the past" (p.263). According to her, the new modes of communication have fostered new norms of expression and developed innovative ways of using written language. She concludes this issue in the following words:

> Language use in CMC has been proven to be adapted to suit the particular conditions according to level of synchronicity, means of expression and situation in the respective communicative setting. Demands from time pressure and limitations in messages size, among several factors, constrain language use in CMC; communicators employ creative devices in an attempt to convey more meaning than normative written language does as a rule. Use and adaptation of language in the modes of CMC that have been investigated in this dissertation is evidence of and exemplifies human ability to adapt in general, and the uniquely human ability to adapt linguistic behaviour specifically. (Hard af Segerstad, 2002, p. 263)

Hard af Segerstad (2002) challenges a common assertion that CMC is a genuinely new medium. She maintains that the features that different modes of CMC show are not genuinely new, rather they are just different variants adapted to fit the situation. Still, the language use in these scenarios is significantly distinctive.

In Pakistan, Computer Mediated Communication has been an ignored area of study. Only recently, a study has been conducted by Aslam, Ahmad and Sajid (2011), which they conducted on the orthographic features of Instant Messages (IM) in Pakistan. They covered the areas like the message length, virtual phonology, consonants change,

mono-lettered words, letter-number homophones, digit-word homophones, final 'g' clipping, virtual orthography, vowel clipping, acronyms and initials. Overall, they investigated instant messages only from orthographic/spelling point of view, pondering on the short forms of words. Regarding the acceptability of these short forms, they concluded that time would decide whether or not these orthographic adaptations become regular feature of English orthography.

2.4 Short Message Service (SMS)

Communication through SMS or mobile text messages is mostly thought to be one mode of Computer Mediated Communication. This technology was first developed in Europe in early 1990s (Crystal, 2004). The SMS protocol was developed as part of Group Special Mobile (GSM) networks. GSM technology was introduced to establish a uniform mobile telephone system for Europe. With the passage of time the term GSM came to mean the Global System for Mobile communications and this name is prevalent among the existing generation of mobile users.

According to Bodomo (2010), SMS was commercially introduced in 1995. He states that "the first SMS message was a Christmas greeting sent out in Britain in 1992" (p. 112). Hence, within two decades of its origin, it has proven its position as one of the major mass media of communication, "with an estimation of over one billion messages exchanged per day around the world" (p. 112).

Text messages sent through GSM technology of mobile phones are used as a mean of communication in much of the world including Pakistan, and at present mobile industry is one of the most rapidly growing industries of Pakistan. Text messages are created by tapping digits on the keypad of mobile phones. These number strokes correspond to the desired letters of the Roman alphabet. The modern *feature phones* and

smart phones also provide the full QWERTY keypads with reduced key strokes on keypads.

Regarding the language of text messages, it resembles with the language used in the Internet based CMC modes like instant messaging, email, and blogs etcetera as it is short, casual and full of abbreviations. Viewed from the pure technical point of view, SMS is not computer mediated communication as it was actually designed to be managed through mobile phones using satellite technology and not through computer networks.

However, over the time many digital technologies have become interchangeable platforms for transmitting and receiving data. As an example, emails can be received and sent through mobile phones; whereas SMS can be received and sent through computer networks. Baron (2003) states that it is possible that as platforms become increasingly interchangeable, the kind of written language appearing in email, instant messaging, and SMS will tend to become more homogeneous, short, informal, spoken like, and full of abbreviations and condensed syntax. Therefore the language used in both computer and mobile phone mediated modes of communication is similar in nature. The language used in these modes is short and creative, and is adapted to fit these modes.

2.4.1 Studies on Linguistic Features of SMS

The development of communication technologies has contributed to the appearance of new forms in the written language like email, instant messages and SMS. They are a constant source of new codes and abbreviations. Each of these means of communication has its own specifications but the common point in all of them is that their users adopt new behaviors in their written language. Hence, new abbreviations and non-standard spelling forms are developed. This phenomenon has come under the inquiry of sociolinguists and they have started to describe the ways through which the

language is adapted in different situations using fewer words and fewer characters (Baron, 2003).

In this regard, CMC and SMS modes of communication are closely related to each other as the language used in both modes is similar in nature. Most of the linguistic features of both varieties of technologically influenced modes resemble to each other. Moreover, the studies conducted on the linguistic aspects of both modes also resemble and present similar results. Both the modes of communication use adapted language that suits to the conditions and demands of these technologies.

Still, the specific linguistic features of SMS language differentiate it from many other varieties of written language. Its specific lexical, syntactic, typographical, and graphical features differentiate it from other written varieties of language. Moreover, its feature of hybridizing written-spoken forms brings this written variety very close to the spoken form of language. In the case of bilingual societies, the frequent code alterations make it a very specific written variety.

In this context, SMS language has already established as a situational variety of written language with its own specific mobile platform (Baron, 2008; Bodomo, 2010; Crystal, 2008b). It is very closely related to CMC, yet it is different from many other situational varieties and registers of written language due to its specific linguistic features. The research conducted in this field establishes that SMS language is a distinct situational variety of written language. Many terms like genre, style, and register etcetera can be assigned to SMS language but this study mostly uses two terms, i.e. language variety as adopted by Crystal (2004), and language register.

Regarding the specific features of SMS language, most text messages are not written in the standard written form, instead users try to describe written words what they want their readers to hear in their SMS voices. Through the new written conventions of

SMS, texters have developed a written form of intonation that serves the purpose of spoken discourse. In this way, SMS language has invented its own unique style (Ling, 1998).

The limited space of the mobile screen and small size of the mobile keypad initially motivated texters to condense as much meaning as possible in the short messages. Doring (2002) believes that abbreviations and acronyms fulfil a collective identity function. Moreover, these adaptations require a special shared knowledge to understand the language and consequently to use it. The use of these personalised short forms is an indicator of group affiliation and group identity. The language specific to texters often does not correspond to standard language and the mass media. Therefore, SMS language is sometimes termed as weird, the new language, SMS lingo, or the secret code of youth.

Crystal (2008b) refutes these disapproving terms and maintains that "various features of text messages are not weird. They are under certain rules, and these rules are centuries old" (p. 27). Moreover, the adaptations made in SMS language are not any standards in themselves. Only those adaptations will remain in longer use that will be consistently used by a large number of people, and will survive the test of time. In this regard, Crystal (2003, p. 7) states:

> Language has no independent existence, living in some sort of mystical space apart from the people who speak it. Language exists only in the brains and mouths and ears and hands and eyes of its users. When they succeed, on the international stage, their language succeeds. When they fail, their language fails.

As these discussions on the specific features of SMS language are increasing, studies on the linguistic aspects of SMS are also increasing day by day in the world. Hard af Segerstad (2005) presents an overview of such studies in different parts of the world. She

states that "of the studies that have been published on text messaging, most have centred on the use and function of texting and only a small proportion on language use in SMS" (p. 36). Her statement shows that in the first few years after the inception of text messages, no proper attention was paid to the linguistic aspects of this service. Yet, she presents a summary of the various linguistic features of text messages through the available studies.

In this regard, linguistic features of text messages can be classified under various broader categories like lexical, syntactic, punctuation, space, code and script adaptations. Among them, lexical adaptations are the most noticed feature. Hard af Segerstad (2005) states that in many cases SMS spellings imitate the phonetic value of speech to save time and effort by the economical use of keystrokes. Sometimes, though, unconventional spellings result in the same or more number of keystrokes as the normative spellings. This aspect of SMS spellings proves that the economy principle is not absolute and that rational behaviour is more at work than the principle of economy. The omission of vowels makes this mode of writing like the consonant writing. The intended meaning therefore seems to come through even without vowels. Moreover, SMS language is hybridization of written and spoken language. It includes colloquial expressions and assimilations, and frequently uses the lower-case.

Crystal (2008b), on the other hand, considers logographic elements like "b" for "be", "2" for "to", and "@" for "at" etcetera, as the most noticeable features of text orthography. He states that "logograms in texting may be used alone, or in combination" as "b4" for "before" (p. 38). Thurlow and Poff (2011) state that text messages have many novel or unorthodox linguistic forms such as shortenings, contractions, G-clippings and other clippings like dropping final letter, acronyms, initialism, letter/number homophones, misspellings and typos, and non-conventional spellings. One of the best

examples of unconventional spellings is the use of consonant clusters, which rely on the premise that consonants usually have more semantic detail/value than vowels. In addition, many of the nonconventional spellings found in texting are widespread and pre-date the mobile phone, in any case.

After lexical adaptations, syntactic adaptations are another frequently discussed feature of SMS language. Hard af Segerstad (2005) discusses syntactic reductions and states that they are caused by the medium in which they are used. She elaborates that the common syntactic reductions are the deletion of subject pronouns, prepositions, articles, possessive pronouns, copula, auxiliary, and modal verbs.

In addition to lexical and syntactic adaptations, the unconventional uses of punctuation marks are another feature of text messages. Baron (2008) discusses different aspects of punctuations in texting and states that seventy-one percent of text messages have no punctuation mark at the end of the text message. In the same way she gives the examples of text messages where the conventional punctuation marks are replaced with deviated punctuation marks. Bosco (2007), on the other hand, exposes that punctuations in text messaging have a high degree of creativity. She discusses the specific uses of punctuations in text messages that are 1) excessive use of ellipsis, 2) excessive use of exclamations, and 3) omission of periods and apostrophes.

A feature of text messages that is confined to the bilingual societies is of code adaptations. Bilingual speakers employ their linguistic resources more effectively and flexibly when they have different language choices to convey their messages. English borrowings and code-switchings in text messages are made in most bilingual societies due to the expansion in technical vocabulary in English which lacks equivalents in other languages.

In this regard, Thurlow and Poff (2011) report a few studies conducted on code adaptations in text messages in multilingual cultures, predominantly, English in contact with some local language of a bilingual country. They state that South African texters blend English with isiXhosa by writing English nouns with isiXhosa prefixes. Finnish teenagers mix Finnish with a medley of foreign language words and expressions, drawing suitable expressions from any language mastered by the writer. Greek texters use graphemes in alphabet-choice and code-switches. In the setting of Kuwait, Haggan (2007) found that Kuwaiti texters use a mixture of Arabic and English in their text messages. In contrast, Nigerian texters avoid Nigerian words in their text messages, preferring Standard British English (Chiluwa, 2008). The use of code-mixed language in text messages also entails adaptations of writing scripts in text messages. For the texters of non-Roman languages, the use of Roman script is also a preferred choice for code-mixed text messages.

Regarding the metalinguistic perceptions of SMS language critics, Thurlow and Poff (2011) state that ideological debate revolves around the negative impacts of SMS linguistic adaptations on language and literacy. They present an overall view of the issues addressed by researchers. They state that while a few scholars insist that texting has a negative influence on standard writing, spelling and grammar, most empirical studies maintain that texting does not pose a threat to the conventional norms of teaching and learning. These scholars usually argue that, although there may be some diffusion of texting style into formal writing, texters recognize that language is context specific.

In gist, most studies on the linguistic aspects of text messages address individual issues, and they leave many aspects uncovered. Those studies are very rare which cover multiple aspects of adaptations. Most studies in the field are only concerned with orthographic features, and other linguistic features are rarely touched in many studies. In

this regard, as termed by Biber (1995, p. 18), this is a multi-dimensional study covering "multiple linguistic parameters of variation" in text messages.

Moreover, most studies of the past are observation based, and real SMS data has been very rarely analysed in most studies. Even, the real data based empirical studies are based on limited data. Moreover, most studies of the past have not employed modern tools of corpus linguistics for the analysis of data. Judged from these aspects, the present study fills gaps of earlier studies of the past in the field. Moreover, if the present study is judged in the context of Pakistan, it is a valuable contribution in a very important but overlooked field of enquiry among modern communication technologies.

2.4.2 Studies on SMS in Pakistan

In Pakistan, the research on SMS in general as well as from linguistic aspects is very rare. A couple of available studies are small scale, limited in scope, and in some specific perspective. In this regard, Muhammad Shaban Rafi's (2008) article is confined to the analysis of SMS from the gender perspective. His study is based on 100 text messages, and the perceptions of 25 males and 25 females. His study examined "the assumption that a great motor of SMS lives among females whose lexical and morpho-syntactic choices are different from males." (p. 3). He concluded that a significant difference is found between linguistic patterns of male and female texters. He states that females are more skillful in writing long and lexically complex messages than their male counterparts. The study also concludes that SMS language is a novice intelligible language, which is leaving backwash effects on formal English and media language in Pakistan.

Fouzia Janjua (2010), on the other hand, studied the effects of mobile phone text messages on the spellings of university students. She did not actually analyse text messages, instead she analysed the data of 60 e-mails, 90 lecture notes, and 50 scraps

from Orkut and Facebook to examine the effects of SMS language on these modes of written language. This study is indeed contradictory with the established theories of the field. The key studies on CMC reveal that SMS language is one mode of CMC (Baron, 2008; Bodomo, 2010; Crystal, 2004, 2008b), and CMC was established even before the advent of SMS. So instead of considering the possibility that CMC mode may have impact on the formation of SMS language, Janjua asserts that shortened language of CMC like e-mails, Orkut, and Facebook is due to the impact of SMS language. In this way, her basic thesis is contrary to the established theories in the field, and she has generated certain questions regarding the established theories in the field.

Javid, Malik and Gujjar (2011) also conducted a small scale study on the mobile phone culture and its psychological impacts on students' learning. Their study was based on a small sample consisting of 390 students of "The Islamia University of Bahawalpur, Bahawalnagar Campus". Yet, their study is not actually concerning the area of present study because neither it concerns SMS language, nor the linguistic analysis. It is rather concerned with psychological impacts of mobile phone culture on the students' learning.

In short, the available studies in Pakistan, in the related fields of the present study are actually a bit away from the concerns of this study. Moreover, they are small scale studies, based on smaller data and are limited to their specific issues. On the other hand, the present study is mainly concerned with the linguistic analysis of text messages (SMS). It is based on a larger corpus of 5000 text messages, and on a larger metalinguistic data of 500 texters. Moreover, the present study uses modern corpus tools to analyse SMS-Corpus of 5000 actual interpersonal text messages of Pakistani texters. An important contribution of this study is that it addresses multiple issues and areas of inquiry of SMS language.

CHAPTER 3

METHODOLOGYAND DATA DESCRIPTION

This chapter deals with the methodological issues of the study. It comprises issues like research design of the study, details about the construction of SMS-Corpus, description of corpus data, procedures adopted for the analysis of SMS-Corpus, and the collection and analysis of metalinguistic data. Finally, it provides descriptions about gender, age, education, social status, and institutional affiliation of questionnaire respondents.

3.1 Research Design of the Study

A clear understanding about the research design is an important step towards the success of any research project. The research design of any study provides the theoretical and methodological foundations to the study. In this regard, the present study is a corpus based linguistic analysis of SMS text messages. The study is empirical and descriptive in nature. It is a *mixed-methods* study that integrates both *qualitative* and *quantitative* methods of analysis.

From theoretical point of view, the study is led by the impact of modern information technologies on human languages. In this regard, it is supported with a theoretical framework, i.e. *Technology-conditioned approach to Language Change and Use* (TeLCU), proposed by Bodomo and Lee (Bodomo & Lee, 2002; Bodomo, 2010). The framework emphasizes the causal relationship between the emergence of new information communication technologies (ICTs) and the development of new forms of language and literacy. Consequently, people may face changes in their languages due to the uses of new modes of communication (Bosco, 2007).

From methodological point of view, the study is broadly divided into two types of analyses, i.e. *linguistic* and *metalinguistic* analyses of the language used in text messages in Pakistan. The major concern of the study is the *linguistic analysis* of SMS-Corpus of 5000 text messages (Chapters 5 to 8). The *metalinguistic analysis* of the perception of texters about *linguistic adaptations* made in text messages is based on Part-A of the questionnaire (Chapter 4).

In the study, the term "adaptations" refers to the register specific changes made in the language used in text messages. The language used in text messages is a technology-based specific variety/register of language with its own specific features. In the bilingual setting of Pakistan, the language used in text messages is Urdu-English code-mixed language. In this context, the purpose of the study is to explore and describe the register specific Urdu-English code-mixed language of text messages.

3.2 Construction of SMS-Corpus

The construction of a personal corpus is a lengthy, time-consuming, and laborious job. It may involve many steps like the decision about its design, collection of data, its compilation, resolving ethical issues, and then to develop its final electronic copy for computer processing. In the study, all these issues regarding the construction of SMS-Corpus of 5000 text messages were tackled in the following ways.

3.2.1 Design of the Corpus

The very first step towards the development of a corpus is the decision about its design. A decisive factor in the design of any corpus is the purpose of the corpus. Corpora may be of many kinds like the reference corpora vs. the specialized corpora, and the public corpora vs. the private corpora. If the study is limited to highlight only one genre or a family of genres, the specialized corpus is designed. In the same way whether

the corpus is meant for personal or public use also defines the design and size of the corpus (Elewa, 2004).

Regarding the size, Paul Baker (2010) states that all corpora need not consist of millions of words. A corpus can be small in size if it does not have the grand goal of being a general reference corpus. McEnery et al. (2006) stress upon the creation of a problem oriented corpus to answer specific research questions. In fact, the size of any corpus depends on two factors, i.e. practical considerations and the focus of the research. Moreover, the specialized corpora need not be larger like general reference corpora as their aim is to represent smaller sections of language in comparison to the general reference corpora (Tagg, 2009).

In this context, the present study deals with a single genre, i.e. SMS text messages, and the corpus required for this study is not a larger general reference corpus. Therefore, SMS-Corpus designed for this study is a specialized personal corpus of 5000 SMS collected from Pakistani texters.

Moreover, it is an unannotated corpus. As SMS language at present has no standard orthography, so there is no standard annotation system like *part of speech [POS] tagging*, *semantic tagging*, or *syntactic parsing* available yet. Moreover, SMS-Corpus designed for the study may not be criticized for missing POS or semantic tagging because no standard spellings of this variety of language do exist yet. Additionally, all required investigations in the study are comfortably possible on the unannotated corpus, with the help of corpus analysis tools.

3.2.2 Collection of SMS

A very practical approach to the collection of social data is the utilization of social networks (Tagliamonte, 2006). According to this approach one can access the data by utilizing one's social contacts. In this scenario, an important element is the utilization

of friend-to-friend chain to reach the social data. One's friends, colleagues, relatives, students can play an intermediary role in the community of human beings to reach the desired data.

Therefore, the *Social Networking* approach was adopted for the collection of text messages. The researcher requested his friends, colleagues and students to provide their own text messages, and to play an intermediary role to collect text messages from a wide range of texters. No doubt, interpersonal text messages contain matters of private nature and may not be easily obtained without any social networking. Crystal (2008b) states that he also encountered hurdles when he sought to obtain private text messages for his well acclaimed book *Txting: The Gr8 Db8.* He states that "I have asked many people, and most give me a knee-jerk negative reaction.... It was as if I had asked them for a window into their most intimate world" (p. 103). Interestingly, in the case of the present study, by employing the social networking technique, the researcher had not to face much difficulty in collecting 5000 private interpersonal text messages of Pakistani texters.

In this regard, two tools of data collection were used. The first tool is the conventional "Questionnaire", whereas the second tool is "Forward Message" tool of text messages in mobile phones. The detail of the collection of 5000 text messages through both tools is given as follows.

3.2.2.1 Collection of SMS through Questionnaire Tool

The first step towards this tool was the preparation of a valid questionnaire. Before the distribution of the actual questionnaire, a pilot questionnaire was prepared and distributed among 20 respondents of different age groups and qualifications. The pilot survey was carried out in order to foresee possible shortcomings of the first questionnaire. So, in the light of the feedback of the respondents, the first pilot

questionnaire was modified. It was again distributed to 10 respondents. Finally, with their suggestions and personal judgment, the actual one-page (legal size, back-to-back) questionnaire was prepared. The questionnaire was divided into two parts (Part-A and Part-B). In Part-A, Questions 1 to 5 were about name, age, gender, qualification, etc., and Questions 6 to 13 were concerning the metalinguistic awareness of each respondent about SMS language. In Part-B, each respondent was requested to provide his/her six personal SMS, meeting certain criteria given in the guidelines of the questionnaire (see Appendix-A for details).

The second step towards this tool was the distribution of questionnaires among 1000 respondents. It was planned at the outset of the study that 2500 SMS would be collected through the *Questionnaire* tool and remaining 2500 SMS would be collected through mobile phone based *Forward Message* tool. Following this procedure, 1000 questionnaires were distributed with the foresight that a few hundred questionnaires would not come back. Another foresight was that all returned questionnaires would not be completely filled, and some might not be legible to include in the corpus. So, this was preplanned that only 500 questionnaires, meeting the maximum criteria given in the questionnaire, would be used for the first 2500 SMS of the corpus. Consequently, 1000 questionnaires were distributed among 1000 respondents using the social network of friends, colleagues and students of the researcher.

Using the *convenient sampling* technique, all the questionnaires were distributed among the respondents living in Rawalpindi and Islamabad (known as twin-cities in Pakistan). Personal contacts were used through friend-to-friend chain to obtain the personal/private text messages of respondents. The city population of twin-cities (one being the capital of Pakistan) belongs to almost all regions and ethnic groups of Pakistan. Therefore, text messages collected from the texters of these two cities was expected to be

representative of the population of Pakistan as these two cities are the hub of the varied population of Pakistan. It was tried to distribute the questionnaires among the male and female texters equally but the acceptance as well as the return rate of the questionnaires from females remained low (see Table 3.2).

When the questionnaires were received back, most of the things happened as expected. Out of 1000 distributed questionnaires, 764 were received back. Out of the 764 returned questionnaires, 500 questionnaires were selected that fulfilled the maximum criteria given in the questionnaire. From these 500 questionnaires, with an average of five SMS from each questionnaire were assigned serial numbers (1 to 2500) to make an electronic copy of the text. In this way, the target number of 2500 SMS was accomplished through this tool. It is important to mention here that these 500 texters, who provided 2500 SMS, also provided the metalinguistic data about the language of text messages that has been analysed in Chapter 4.

3.2.2.2 Collection of SMS through Forward Message Tool

The second tool used to collect the remaining 2500 SMS was the *forward message* feature of mobile phone. An extra advantage of this tool is that it reduces the chance of typographical errors which may occur in the case of typing the questionnaire-based text messages. However, this method was not used to collect all 5000 text messages, as this method was feasible to collect the text messages but not for the metalinguistic data needed to triangulate the study.

To utilize this tool, the researcher arranged with 50 of his direct friends and students that each would forward him (the researcher) about 20 to 50 SMS. These SMS could either be from *inbox* or *outbox* of their mobile phones. In this way, 1500 SMS were collected with an average of 30 SMS from each respondent. The same tool was utilized in another way. The researcher requested two of his university friends that they

would not delete any SMS from their mobile phones until each had accumulated at least 500 SMS in his mobile phone. Hence, through this way the remaining 1000 SMS were collected. In these two steps of *forward message* tool, 2500 text messages were collected from total 52 respondents.

3.2.3 Compilation of SMS-Corpus

After the collection of 5000 text messages, the next step in this study was the compilation of the corpus. For the computer processing of a corpus, through the corpus analyzing tools, it must be in machine readable (electronic) form. So, the compilation of 5000 SMS to prepare an electronic copy was carried out in two steps.

In the first step, the questionnaire-based 2500 handwritten SMS were assigned individual identifications (IDs) in the form of serial numbers (1 to 2500) on all 500 questionnaires. Then all 2500 SMS were carefully typed by the researcher himself, so that the chances of typographical errors could be curtailed. As SMS spellings are different from the standard spellings, so *Spelling & Grammar* check and *AutoCorrect* options of MS-Word file were turned off. All 2500 SMS were given IDs/serial numbers in the corpus too, from 1 to 2500.

In the second step, the forward message based 2500 forwarded SMS of friends, colleagues, and students were transferred from the mobile phone of the researcher to MS-Word file of the computer. These 2500 text messages collected through 52 respondents were transferred through *Bluetooth* device from *Nokia N70* mobile phone to the computer by using *Nokia PC Suite* software. Although this software facilitates in transferring the mobile phone SMS to a computer file through *copy* and *paste* method, yet one cannot transfer them in a bulk. Therefore, the researcher had to repeat copy-paste procedure for 2500 times. These forwarded SMS were given their IDs/serial numbers from 2501 to 5000 in the corpus file. Hence, these 2500 forwarded SMS also became the

part of previous 2500 SMS collected through the questionnaire, and the target corpus of 5000 SMS was accomplished.

Before finalizing the compilation process of SMS-Corpus of 5000 text messages, the corpus was examined, filtered and purified from unwanted material. To comply with the guidelines given in Part-B of the questionnaire, the repeated/duplicated messages from the same sources were removed from the corpus. Similarly, some publicity messages of mobile phone companies were removed and replaced with the interpersonal text messages.

The final electronic copy of this personalized corpus of 5000 text messages was assigned a specific name, i.e. "SMS-Corpus" to identify it in the study. Hence, the term "SMS-Corpus" has been repeatedly and consistently used in the study to refer to this personalized corpus of the researcher.

3.2.4 Ethical Considerations in Corpus Construction

Any research, that uses the personal/private data of respondents, needs to address ethical issues of the privacy of data. In this regard, in the questionnaire it was stated that the data provided by the respondents will be used for research purposes only, and identities of the respondents will not be disclosed at any stage.

The second step of ethical considerations concerns the actual-user based corpus of private one-to-one SMS communication of texters. Although the Internet is replete with the public online SMS services, yet this study did not directly acquire any SMS from online services. So this is purely an interpersonal SMS corpus of Pakistani texters collected by the researcher.

For the purpose of anonymisation, various personal identities of texters within text messages were replaced with some common codes. Similar to Tagg (2009), personal details of texters like their names, phone numbers, personal addresses, emails,

registration numbers, bank account numbers, birth dates, class/institution IDs of students, and personal website addresses were replaced. These personal details were replaced with common terms, i.e. NAME, PLACE, PHONE-NUMBER, E-MAIL, ADDRESS, BIRTH-DATE, PLACE, CLASS, INSTITUTE, REG-NUMBER, AC-NUMBER and WEBSITE, all in capital letters.

This anonymisation was made to ensure that the individuals could neither be identified nor contacted. There were certain problems in the anonymisation of personal details of texters. For example, all individual names could not be automatically anonymised because it is difficult to decide whether a name in an SMS is a public figure or a personal name of an individual. Hence, *Find* and *Replace* tools of MS Word programme were cautiously used. Somewhere these electronic tools were used, and in some cases manual changes were made. In short, it was ensured up to the maximum that the personal details of texters remain anonymised.

3.3 Description of SMS-Corpus Data

Before going into the details regarding the linguistic analysis of SMS-Corpus, this section presents an overall statistical view of SMS-Corpus. It provides the size of SMS-Corpus in various categories. The data has been analysed through AntConc software of corpus linguistics. The statistical details are given in the following table.

Table 3.1
A view of Corpus Data (5000 SMS)

Word Types/Tokens Definition	**Number of Word Types**	**Number of Word Tokens**
Letters based	14318	66811
Numbers (digits) based	155	2100
Punctuation based	362	10862
Symbols based	30	337
All four categories combined/activated	**18592**	**67444**

The table shows the total number of word types and word tokens from different perspectives. It shows that the letter based word types and word tokens are the most frequent words in the corpus. The reason is self evident because any Romanized language mainly depends on letters. Punctuation based expressions are the second category of frequent words as SMS language uses punctuation marks in abundance.

Overall, combining all four categories of word types/tokens, the total number of word types in the corpus is 18592, whereas the total number of words (word tokens) in this specialized personal corpus is 67444. These statistics do not include 5000 serial numbers (IDs) of text messages that were assigned to text messages in the corpus to identify them. For all analytical purposes through AntConc software, serial numbers were removed from SMS-Corpus.

3.4 Analysis of SMS-Corpus

From the analytical point of view, the study is divided into two types of analyses, i.e. *linguistic* and *metalinguistic* analyses. The major concern of the study is the *linguistic analysis* of SMS-Corpus of 5000 text messages. The analysis of both types has been conducted separately in different ways, and the results have been presented, described and discussed in Chapters 4 to 8 of the study. The metalinguistic analyses are presented in Chapter 4, whereas the linguistic analyses of SMS-Corpus are given in chapters 5 to 8 of the study. In this section, the detail of the procedure adopted to analyse SMS-Corpus is discussed as shown below.

SMS-Corpus was analysed both qualitatively and quantitatively. The results, descriptions, and discussions on various linguistic adaptations made in SMS text messages have been presented in the concerned chapters. For the quantitative analysis, the concordance software "AntConc 3.2.2.1w" developed by Laurence Anthony (2011) was used.

3.4.1 Description of Corpus Analysis Tools

In corpus linguistics, many tools and software programmes have been developed to process electronic data (see section 2.1.4). Nowadays corpora can be processed in a number of ways. Electronic corpora can be tokenized, annotated with POS and semantic tagging, syntactically parsed through tree-banks, and analysed with number of concordance tools. In many cases, information about the lemma (base) form is added with word tokens (Ludeling & Kyto, 2008).

The most used tools in this field of corpus processing and analysis are concordance tools. In Chapter 2 of this study, an overview of such tools has been given (see section 2.1.4). Based on that review, the software selected for the corpus analysis of this study is *AntConc* concordance software.

Like many other tools in corpus linguistics, AntConc does not process corpus files developed in MS-Word format. The types of files that can be processed through AntConc are 1) txt, 2) html, 3) htm, and 4) xml files. For this reason the file format of SMS-Corpus was changed into text (*.txt) file so that it can be processed through AntConc corpus software.

Various versions of AntConc software are available free of cost at Laurence Anthony's webpage.[1] The software version used for the corpus processing of this study is *AntConc 3.2.2.1w* (windows based version released on May 1, 2011). The analysis of this study was started in year 2011, and at that time *AntConc 3.2.2.1w* was the latest edition of the software. Therefore, this version has been consistently used throughout the span of study. The software is developed by Laurence Anthony of Waseda University, Tokyo, Japan. The first programme of this series was released in April 2002, and since

[1] http://www.antlab.sci.waseda.ac.jp/antconc_index.html

then more than 30 windows-based versions of this programme have been developed by the author. The author of this corpus toolkit is a Professor at the *Center for English Language Education in Science and Engineering, Waseda University, Japan.*

Laurence Anthony introduced the AntConc software in 2002 as a comparatively simple concordance programme for the use of students in classrooms. As time passed, he improved the programme and now it has become a very useful text analysis tool, very successfully competing with some high valued paid tools in the field of corpus linguistics (see section 2.1.4). The author keeps on improving the programme and removing its bugs. An enormous value of this programme is that it is freely available for research purposes. A screenshot of *AntConc 3.2.2.1w* in its launched form is given as shown below.

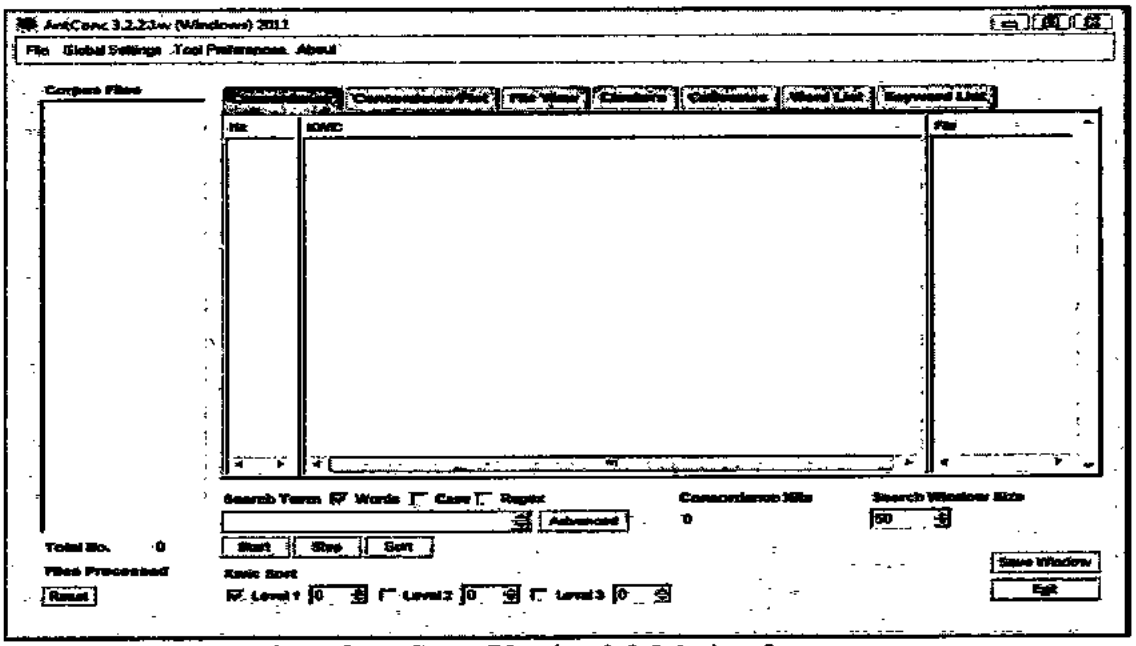


Figure 3.1 A screenshot of AntConc (Version 3.2.2.1w) software

AntConc is a complete corpus toolkit. It contains seven major tools and each tool can be launched either by clicking on its name tag in the toolbar window, or by just clicking the function keys F1 to F7 on the keyboard. A brief overview of each tool is given as follows.

- ***Concordance.*** This is the first tool shown on the AntConc toolbar panel. This tool can show search results in a *Key Word in Context* (KWIC) format. This tool shows how words and phrases are commonly used in a corpus.
- ***Concordance Plot.*** This is the second tool shown on the AntConc toolbar panel. This tool can show the concordance results in the form of a barcode plot. This tool shows the concordance position in a bar chart.
- ***File View.*** This is the third tool shown on the AntConc toolbar panel. This tool can show the search results in the original files. This tool shows the results in the actual file locations, generated in other tools of AntConc.
- ***Clusters (N-Grams).*** This is the fourth tool shown on the AntConc toolbar panel. This tool can show clusters (lexical bundles) based on the search condition. Its N-gram tool can scan the entire corpus for N length clusters and shows the common/regular expressions in a corpus.
- ***Collocates.*** This is the fifth tool shown on the AntConc toolbar panel. This tool can show collocates of a search term. This tool investigates non-sequential patterns in a corpus.
- ***Word List.*** This is the sixth tool shown on the AntConc toolbar panel. This tool can show counts/frequencies of all the words in the corpus and presents them in an ordered list. This tool very quickly finds the most frequent words in a corpus.
- ***Keyword List.*** This is the last tool shown on the AntConc toolbar panel. This tool can show unusually frequent or infrequent words in the given corpus, in comparison with the words in a reference corpus.

In addition to these seven tools of corpus analysis given in the toolbar panel, AntConc also facilitates by modifying definitions of the programme through its *Global Settings* and *Tool Preferences* options. Hence, AntConc is not a static programme and it

can be used in multiple ways by changing its settings. A detailed working procedure of the programme is given in its Readme file available at Laurence Anthony's webpage[2].

3.4.2 Procedure of SMS-Corpus Analysis

The analysis of SMS-Corpus for this study has been mainly conducted with the help of AntConc software (version 3.2.2.1w). The corpus file was saved as a plain text (*.txt) for AntConc analysis. After opening the file in AntConc, multiple ways were adopted to extract the required data for the study. Out of seven tools of AntConc, five tools were repeatedly used for the analysis. These are *Concordance, File View, Clusters, Collocates*, and *Word List. Word list* and *Concordance* tools were used more than other three tools. AntConc tools were used to facilitate the data searches. The decisions about *what to search, how to search,* and *how to arrange the extracted data,* were all manually done by the researcher. In the *Global Settings* of AntConc, the four token/word classes, i.e. *Letter, Number, Punctuation*, and *Symbol* were frequently replaced with one another. Sometimes one category was used in isolation and sometimes accompanied with some other category (details are appended with the respective tables). Hence, AntConc tools were used according to the different needs of the study, and the details are given in the concerned sections/tables.

The analysis has been focused to explore the register specific Urdu-English code-mixed language used in the text messages of Pakistani texters. The analysis explores this register specific code-mixed variety of language from various aspects.The study mainly examines the six major types of linguistic adaptations made in text messages (see Chapters 5 to 8). These types are *lexical, syntactic, punctuation, space, code,* and *script* adaptations. *Adaptations* here refer to the register specific changes/modifications made in the language used in text messages, and these adaptations define SMS language as a

[2]http://www.antlab.sci.waseda.ac.jp/software/README_AntConc3.2.2.1.pdf

language variety/register. The analysed data has been divided into four major areas/sections, and has been described and discussed in Chapters 5 to 8.

Firstly, the corpus analysis has been started from examining *lexical adaptations* in text messages. The analysis is further divided into three parts. The first part concerns *morphological adaptations* (shortenings) in text messages. These shortenings are based on shortenings of words by reductions of letters. These fall into three categories namely *initialism, contractions*, and *clippings*. The second part deals with *phonological shortenings*. Phonological adaptations are based on the shortenings of words by replacing words or parts of words with the letter, number, or symbol homophones. The third part deals with *spelling and case variations* in the corpus. All three areas of lexical adaptations were analysed with the help of AntConc tools. The results and further procedures of analysis regarding lexical adaptations have been discussed in the concerned sections (see Chapter 5).

Secondly, SMS-Corpus has been analysed for *syntactic adaptations* in text messages. Syntactic adaptations have been further divided into three categories that are *omissions/ellipses* of words, *reduplications* of words, and the use of *deviant grammar* in text messages. *Omissions/ellipses* have been further analysed as omissions of *personal pronouns, auxiliaries, copula verbs, articles, prepositions, conjunctions*, and *infinitives*. The second category, i.e. *word reduplications* have been further analysed in three areas that are *exact, rhyming* (consonant changing), and *ablaut* (vowel changing) reduplications. The third category of *grammar deviations* have been analysed in three areas that are deviations of *tense* (verbs), *word class*, and *word order* in sentences (see Chapter 6).

Thirdly, SMS-Corpus has been analysed for *punctuation* and *space* adaptations in text messages. This analysis has been further divided into three categories. In the first

category, *omissions, repetitions*, and *substitutions* of punctuations have been analysed. In the second category, the analysis has been made to examine the use of punctuation marks as *emoticons* and *artistic devices* in text messages. In the third category, the *deviant/variant* uses of *spaces* have been explored for their *omissions* as well as *unnecessary additions* (see Chapter 7).

Lastly, the analysis has been made for the adaptations of *linguistic code* and *writing script*. These two adaptations have been further analyzed as Urdu-English *code-mixing*, *code-switching*, and the use of *Roman script for Urdu*. In the first part of this analysis, code-mixing between Urdu and English has been examined in *words, phrases,* and *clauses* of one language into the other. Code-mixing has been further analysed in two areas, i.e. English to Urdu code-mixing (English embedded in Urdu), and Urdu to English code-mixing (Urdu embedded in English). In the second part, the inter-sentential *code-switching* has been analysed at three levels, i.e. *Urdu to English, English to Urdu,* and *two-way in both contact languages*. In the third part, the analysis has been conducted on the *variations of Roman spellings* for the Urdu language (see Chapter 8).

3.5 Collection and Analysis of Metalinguistic Data

The metalinguistic data about the perception of texters was collected through Part-A of the questionnaire (see section 3.2.2.1). The questionnaire used in the study was divided into two parts. Part-A of the questionnaire was about the personal details of respondents (Questions 1 to 5), and their metalinguistic awareness about the linguistic adaptations of text messages (Questions 6 to 13). In Part-B, respondents were requested to provide actual SMS from their mobile phones in the given space of the questionnaire. The text messages provided in Part-B became the part of SMS-Corpus, whereas as Part-A of the questionnaire is described here.

Part-A of the questionnaire consisted of 13 questions both open-ended and close-ended. Questions 1 to 5 were about the personal details of respondents (see section 3.6), whereas Questions 6 to 13 were concerning the metalinguistic awareness of texters about SMS language (see Chapter 4). These questions were statistically analysed by using *MS Excel* programme of *Microsoft Office-2007*. To enter the data of 500 questionnaires in MS Excel, the data was coded in *letters* and *numbers* accordingly. The detail of the analytical procedure for each question of the questionnaire through MS Excel is given here.

- ***Question 1*** and ***2*** were the *optional* questions about the *name*, *phone number*, and *model of the mobile set.* Since they were the personal and optional questions, so most of the respondents did not respond to these questions. Therefore, Questions 1 and 2 have not been analysed in the study.
- ***Question 3*** was about the *gender* and *age* of the respondents. For the quantitative analysis of gender ratio, all 500 responses were entered in the coded form, i.e. "M" for males and "F" for females. At the end, all male and female respondents were counted through *AutoSum* tool of MS Excel (see Table 3.2). Regarding age, all responses were entered in years ranging from 11 to 70 years. *Not given* responses were coded as "NG". At the end, all responses were tabulated in 7 age groups in multiples of 10. The total respondents in each age group were calculated through AutoSum tool (see Table 3.3).
- ***Question 4*** was about the educational background of respondents. Responses were entered in *numbers* and *letters*. Academic qualification in years was coded in numbers, except PhD which was coded as "PHD". *Not given* responses were coded as "NG". The calculations were made through AutoSum tool. At the end,

the responses were tabulated in 8 education groups in multiples of two (see Table 3.4).

- ***Question 5*** was in two parts. The first part was about the social status of the respondents as being *students*, *professionals*, and *unemployed*. The data was entered as "ST" for students, "PRO" for professionals, "UN" for the unemployed, and "NG" for no response to this question. The calculated data was tabulated in four groups (see Table 3.5). The second part was about the institutional/ organizational affiliation of the respondents. The data was entered by short codes of 25 institutions/organizations. For the non-availability of data "NG" code was entered. The analysed of data is presented in Table 3.6.
- ***Question 6*** was the first question in the category of metalinguistic awareness of texters about their use of SMS language. This question examined the metalinguistic awareness of texters whether they think that SMS language is a new variety of written language. For MS Excel analysis, this close-ended question was coded as "Y" for yes, "N" for no, and "U" for uncertain responses (see Table 4.1).
- ***Question 7*** was about the impacts of SMS language on the written language of texters. The question was close-ended with four options. The responses were coded and entered in MS Excel as "P" for positive, "N" for negative, "B" for both negative and positive, and "NG" for no response (see Table 4.2).
- ***Question 8*** was based on the opinions of texters whether SMS variety of language should be incorporated in pedagogy. This close-ended question was coded and entered in MS Excel as "Y" for yes, "N" for no, "U" for uncertain, and "NG" for no response (see Table 4.3).

- ***Question 9*** was about the metalinguistic awareness of texters about the reasons and factors of the telegraphic style of SMS language. The question was a mixed close-ended open-ended question. The first three options (A to C) were close-ended and respondents could mark on any relevant option. Each marked/ticked response of three options was coded as "Y" for MS Excel analysis (see Table 4.4). The fourth option (D) of the question was open-ended as respondents were asked to specify some other reasons of the telegraphic style of SMS language. The responses were broadly grouped and coded into six types (see Table 4.5).
- ***Question 10*** was about the communicative functions of text messages. This question was also a mixture of close-ended and open-ended options. The first seven options (A to G) of the question were close-ended and respondents could mark on all relevant options. Each marked/ticked response of the first seven options was coded as "Y" for MS Excel analysis (see Table 4.6). The eighth option (H) was open-ended, and respondents were free to specify any other communicative function of text messages. The responses were broadly grouped and coded into 11 types (see Table 4.7).
- ***Question 11*** was an entire open-ended question about the factors of phonological adaptations in written text messages. The responses were a combination of multiple factors of phonological adaptations, and respondents offered up to four factors in a single response. Hence, their responses were coded in 14 categories for the statistical analysis in MS Excel (see Table 4.8 & 4.9).
- ***Question 12*** was also an entire open-ended question about the factors of code and script adaptations in SMS Language. The responses were coded and statistically analysed in 11 categories (see Table 4.10).

- ***Question 13*** was the last question in the questionnaire. It was a close-ended question about the sociocultural and ethical impacts of text messages. The responses were entered in MS Excel as "Y" for yes, "N" for no, "U" for uncertain, and "NG" for no response (see Table 4.11).

3.6 Description of Questionnaire Respondents

In corpus linguistics, the concern about any kind of metadata is seriously taken. In this regard, the first 2500 text messages in SMS-Corpus were collected from the 500 respondents of the questionnaire, so the data about the gender, age, education, social status, and institutional affiliation of the 500 respondents was statistically analysed through MS Excel and has been described in this section.

3.6.1 Gender Ratio of Respondents

The gender wise ratio of the 500 respondents of the questionnaire is shown below.

Table 3.2
Gender Ratio of Respondents

Gender	Number of Respondents	Percentage
Male	320	64.00%
Female	180	36.00%
TOTAL	500	100.00%

The table shows that among the 500 respondents, 320 (64%) were male and 180 (36%) were female respondents. The main reason of low female ratio was the reluctance of female respondents to provide their personal SMS data.

3.6.2 Age Group of Respondents

The age wise detail of the 500 respondents is given in the following table.

Table 3.3
Age Group of Respondents

Age group	Number of Respondents	Percentage
11-20	293	58.60%
21-30	170	34.00%
31-40	11	2.20%

41-50	5	1.00%
51-60	2	0.40%
61-70	2	0.40%
Age not given	17	3.40%
TOTAL	500	100.00%

The table shows that a big chunk of respondents belonged to the first age group of 11-20 years followed by 21-30 years of age. These two groups count more than 90% of all respondents. In the existing national and international scenario of texters, it is not a surprise that the majority of texters are the youth who are also the school, college and university students and so they belong to this age group. Bodomo (2010, p.111) states that "more than 80% of people between the ages of 12 and 25 frequently use SMS as a mode of communication with their peers". In the case of present study, this figure is a bit higher for the age group 11 to 30, as the age limit is also increased from both the lower and the higher sides. Moreover, the sampling of the present study for the data collection was convenience based, and in convenience based sampling students were the most eager respondents. This is also the fact that students mostly belong to this age group.

Crystal (2008b, p.89) also supports this phenomenon of the excessive use of text messages by teens on the bases of many studies. He states that "the conventional wisdom is that it's a teenage thing; and indeed, the surveys do show a huge bias in that direction."

3.6.3 Educational Background of Respondents

The educational qualification of 500 respondents is shown below.

Table 3.4
Educational Background of Respondents

Completed Education in Years	Number of Respondents	Percentage
Below 10	12	2.40%
10	149	29.80%
12	159	31.60%
14	141	28.20%
16	30	6.00%
18	1	0.20%
PhD	1	0.20%

Education not given	7	1.60%
TOTAL	500	100.00%

The above table shows that three major groups are of those respondents who have completed education of 10, 12 and 14 years. This implies that most of the respondents are college and university students. This picture is more vivid in the next section, and it also corresponds to the justifications given in section 3.6.2 about the excessive use of text messages by the youth who are college or university students.

3.6.4 Social Status of Respondents

The status ratio of all 500 respondents as being student, professional or unemployed is given in the following table.

Table 3.5
Social Status of Respondents

Status	Number of Respondents	Percentage
Students	440	88.00%
Professionals	47	9.40%
Unemployed	8	1.60%
Status not given	5	1.00%
TOTAL	500	100.00%

Interestingly, this table makes it crystal clear that the biggest chunk of SMS respondents in this study is of students followed by professionals. The justification of the higher number of students is given in section 3.6.2.

3.6.5 Institutional/Organizational Affiliation of Respondents

It is clear from the description of the metadata of previous sections that most of the respondents are students, so their institutional affiliations is given in this section. The institutional/organizational affiliation of all 500 respondents is given in this table.

Table 3.6
Institutional/Organizational Affiliation of Respondents

	Institution/Organization	Number	Percentage
1)	AIR University Islamabad	3	0.60%
2)	Allama Iqbal Open University Islamabad	3	0.60%

3)	Army Public Schools and Colleges (Rawalpindi)	8	1.60%
4)	Bahria University Islamabad	3	0.60%
5)	COMSATS Institute of Information Technology Islamabad	59	11.80%
6)	FAST National University Islamabad	4	0.80%
7)	Fatima Jinnah Women University Rawalpindi	3	0.60%
8)	Fauji Foundation Schools and Colleges (Rawalpindi)	6	1.20%
9)	Fazaia Schools and Colleges (Rawalpindi & Islamabad)	94	18.80%
10)	Federal Govt. Schools and Colleges (Rawp. & Islamabad)	23	4.60%
11)	Federal Urdu University Islamabad	4	0.80%
12)	Foundation University Rawalpindi	2	0.40%
13)	Institutes in cities other than Rawalpindi & Islamabad	4	0.80%
14)	International Islamic University Islamabad	119	23.80%
15)	IQRA University Islamabad	3	0.60%
16)	National University of Modern Languages Islamabad	4	0.80%
17)	National University of Science & Technology Islamabad	52	10.40%
18)	Private Schools and Colleges (Rawalpindi & Islamabad)	19	3.80%
19)	Professionals in government jobs	12	2.40%
20)	Professionals in private jobs	16	3.20%
21)	Punjab Government Schools and Colleges (Rawalpindi)	21	4.20%
22)	Quaid-e-Azam University Islamabad	3	0.60%
23)	Rawalpindi Medical College Rawalpindi	2	0.40%
24)	Riphah International University Rawalpindi	3	0.60%
25)	SKANS School of Accountancy Islamabad	1	0.20%
26)	Institution/Organization not given	29	5.80%
	TOTAL	500	100.00%

Table 3.6 shows the complete picture of institutional/organizational affiliation of all 500 respondents. Although, in actual, the respondents belonged to about 100 individual institutes/organizations, yet for the sake of brevity and simplification most of the institutes were categorized under the broader categories. For example, schools and colleges were categorized under some generic/chains names. Only the institutional affiliation with universities or degree awarding institutes was given independently. Even after using classification/categorization of institutes, institutional affiliation of 500 respondents can be analysed through the 26 categories given in the table.

As the model of convenience sampling was adopted in this study, so a big chunk of the respondents belonged to the parent organization/university of the researcher where he is enrolled for his PhD. This is why the highest number of respondents among all

categories is affiliated with International Islamic University Islamabad. Similarly, the next higher categories are those where the researcher could reach easily by using the social networking approach under the model of convenience sampling.

CHAPTER 4

METALINGUISTIC AWARENESS OF TEXTERS

The chapter is based on the questionnaire data, and deals with views and comments of texters about the language used in text messages in Pakistan. The chapter does not concern SMS language itself, rather perceptions and reflections of texters on SMS language. The analysis is based on the responses of 500 texters who also provided 2500 text messages for SMS-Corpus. Hence, this chapter deals with the metalinguistic awareness of texters about the register specific variety of language used in text messages. In this regard, various features of SMS language have been analysed from the perspective of texters themselves.

4.1 Introduction to Metalinguistic Awareness

The questionnaire based metalinguistic data of the study actually concerns the metalinguistic awareness of texters about the language used in text messages in Pakistan. As the term *metalinguistic awareness* lacks a uniform and stable definition in the available literature, so the term is defined first. Moreover, *metalinguistic awareness* is sometimes referred as *metalinguistic ability* or *metalinguistic knowledge*, so the term itself needs clarification to serve the purpose of this study.

In this regard, *metalinguistics* is a branch of linguistics that deals with language in the context of culture and society. According to *A Dictionary of Linguistics and Phonetics* (Crystal, 2008a), *metalinguistics* refers to the overall relation of the linguistic system to other systems of behaviour in the associated culture. For the specific definitions of metalinguistic awareness, following definitions are offered.

- Cazden (1972, p. 303) used the term *metalinguistic awareness* as "the ability to reflect upon language as well as comprehend and produce it".

- According to the French psycholinguist Jean Emile Gombert (1992), metalinguistic awareness is the ability to reflect on a language.
- *Longman Dictionary of Language Teaching and Applied Linguistics* (Richards & Schmidt, 2002, p. 329) defines metalinguistic knowledge as the "knowledge of the forms, structure and other aspects of a language, which a learner arrives at through reflecting on and analyzing the language".

As metacognition is defined *cognition about cognition,* metalinguistics is the *cognition about language.* In this sense, "metalinguistic awareness refers to a specific cognitive skill that generally accounts for an expressed ability to contemplate language and to understand and use an array of linguistic constructs, rules, norms, and patterns" (Bryant T. Jensen, 2008, p. 551).

Now, the questionnaire based analysis of the metalinguistic awareness of the texters of the study is offered. This chapter deals with *Questions 6 to 13* of the questionnaire (see Appendix-A). The analysis of the metalinguistic awareness of texters is started from Question 6 of the questionnaire.

4.2 SMS Language as a Register Specific Variety of Language

This section determines the metalinguistic awareness of texters whether they consider SMS language as a register specific variety of language. In this regard, the question was raised to 500 texters of the study whether they think that SMS language is emerging as a new variety of written language. In this close-ended question, they were offered three options, i.e. YES, NO and UNCERTAIN to respond. The results of their responses are shown below.

Table 4.1
SMS Language as a Register Specific Variety of Language

Response	Number of Respondents	Percentage
Yes	417	83.40%
No	49	9.80%

Uncertain	34	6.80%
TOTAL	500	100.00%

Table 4.1 shows that reflecting on SMS language, 417 respondents agreed that the register specific language used in text messages is emerging as a new variety of written language. The results show that a majority of texters, that is, 83.4% of the sample, perceive that they are using an Urdu-English code-mixed language in text messages that is different from English or Urdu languages they use in the academic writing. Therefore, texters themselves seem to be well aware of the fact that the variety of language they use in text messages is a new trend in the written modes of language.

4.3 Orthographic Impacts of SMS Language

In this section, on the basis of 500 responses to Question 7 of the questionnaire, the metalinguistic awareness of texters on the impacts of SMS language on the written language of texters is analysed. In this regard, the respondents were asked to respond on how SMS language was affecting their written language. They were offered four options to this close-ended question. The results of their responses are as follows.

Table 4.2
Orthographic Impact of SMS Language

Response	Number of Respondents	Percentage
Positive	38	7.60%
Negative	174	34.80%
Both positive & negative	204	40.80%
No effect	72	14.40%
Response not given	12	2.40%
TOTAL	500	100.00%

Table 4.2 shows that concerning the impacts of SMS language on the written language of texters, the majority of texters are divided into two opinions, that is, *negative* impacts, *and both positive and negative* impacts. Overall, the results show a mixture of opinions. Only 7.60% respondents stated that SMS language was affecting their written

language positively, in comparison to 34.80% respondents who perceived that SMS language had negative impact on their language. On the other hand, 40.80% stated that SMS language had both positive and negative impacts on their language, whereas 14.40% asserted that text messages had no effect on their language.

In brief, these mixed results do not go either in favour or against SMS language for its orthographic impacts on the written language of texters. These results show that from orthographic point of view, the metalinguistic awareness of texters is yet not established and texters themselves are divided in making a clear opinion on the effects of SMS language on the standard language. These divided results also imply that texters perceive the orthographic impacts of SMS language on the written language in different perspectives.

4.4 Incorporation of SMS Language in Pedagogy

This section deals with the application of SMS features of language in pedagogy, based on results of the responses of Question 8 of the questionnaire. Nowadays, there is a great concern of linguists and educationists about the impacts of SMS language on the language learning of students. For this reason, to determine the metalinguistic awareness of texters on this issue, they were asked whether SMS language should be incorporated in teaching and learning process. They were given three options, that is, YES, NO and UNCERTAIN in a close-ended question. The results of their responses are given below.

Table 4.3
Incorporation of SMS Language in Pedagogy

Response	Number of Respondents	Percentage
Yes	129	25.80%
No	316	63.20%
Uncertain	49	9.80%
Response not given	6	1.20%
TOTAL	500	100.00%

Table 4.3 shows that the majority of texters, that is 63.20%, perceive that SMS language should not be incorporated in the pedagogical process, whereas about one quarter, that is 25.80%, are in favour that this telegraphic variety of language can be introduced in teaching/learning process. On the whole, the results indicate that the majority of texters do not favour the incorporation of SMS language in pedagogy. These results show that texters themselves perceive that SMS language is a situational variety and it should be used in appropriate situations and contexts.

4.5 Factors of Telegraphic Style of SMS Language

On the basis of Question 9 of the questionnaire, the metalinguistic awareness of texters about reasons and factors of the telegraphic style of SMS language was analysed. They were asked to respond about the reasons and factors of the short abbreviated style of SMS language. Their opinion was sought in a mixed close-ended open-ended question. The first three options of the question were close-ended and respondents could mark on any relevant option. The fourth option was open-ended as respondents were asked to specify some other reasons of the telegraphic style of SMS language. The results of this question are given in the following two tables.

Table 4.4
Factors of Telegraphic Style of SMS Language

Response	Number out of 500	Percentage
A)Small keypad of mobile phones	88	17.60%
B)For time saving	386	77.20%
C)Understanding is more important than accuracy	293	58.60%

The results of Table 4.4 indicate that in response to this question, respondents were not bound to mark on a single option. They were free to mark all relevant options which they thought were the cause of short telegraphic style of SMS language. The results of their responses show that out of 500 respondents only 88 (17.60%) reflected that small keypad of their mobile phone was real cause of the telegraphic language.

Whereas a majority of respondents 386 (77.20%) reflected that the element of time is the main cause of telegraphic style of SMS language. Similarly, 293 (58.60%) respondents thought that in text messages understanding is more important than accuracy so this telegraphic style is adopted in SMS communication.

Fourth option to Question 9 of the questionnaire (see Appendix-A) was an open-ended option. In this option, respondents were asked to provide any other factor that they thought was the cause of short telegraphic style of SMS language. In response, following six categories were offered by the respondents in addition to three factors already given in Table 4.4. The results of their responses are given in the next table.

Table 4.5
Respondents Specified Reasons/Factors of Telegraphic Style

Response	Number out of 500	Percentage
1) Brevity/To be precise	2	0.40%
2) Ease in writing	12	2.40%
3) Poor English of texters	1	0.20%
4) Rate of SMS/Money saving	2	0.40%
5) Secrecy/ To use coded-words	1	0.20%
6) Style/Trend/Fun	10	2.00%
TOTAL	28	5.60%

All the responses given in Table 4.5 are specified by the respondents of the questionnaire. As this was an additional and open-ended option, therefore only 28 out of 500 respondents responded. Yet, the additional reflections offered by the respondents convey some other important aspects and causes of the telegraphic style of SMS language. For example, brevity/precision of expression is an important factor in telegraphic style of SMS language. Ease in writing, that is using less effort, also matters a lot in writing text messages, and so twelve respondents offered this reason from themselves. Similarly, the poor command on the English language, rate/charges of SMS, secret/coded language of texters may also be among the reasons and factors of adopting telegraphic style in text messages. Lastly, 10 (2.00%) respondents reflected that the short

telegraphic text messages are the style/trend nowadays, and they use it for fun. In short, all these reflections show that texters are aware of the metalinguistic elements about SMS language.

4.6 Communicative Functions of Text Messages

The analysis of communicative functions of text messages is based on the results of Question 10 of the questionnaire. Like Question 9 of the questionnaire, respondents were offered a mixture of close-ended and open-ended options. The first seven options (A to G) of the question were close-ended and respondents could mark on all relevant options. The eighth option (H) was open-ended, and respondents were free to specify any other communicative function of text messages. So in the following two tables, the results of metalinguistic awareness of texters about the communicative functions of text messages are presented as follows.

Table 4.6
Communicative Functions of Text Messages

Response	Number out of 500 for each category	Percentage
A)Chatting /Sharing jokes with friends	378	75.60%
B)Requesting for favours	152	30.40%
C)Responding requests and favours	194	38.80%
D)Information sharing	266	53.20%
E)Love making/Romance	145	29.00%
F)Teaching/Learning purposes	100	20.00%
G)Conveying good wishes and prayers	313	62.60%

The results given in Table 4.6 indicate that all seven communicative functions given in the questionnaire options were endorsed by the respondents. As the respondents were free to mark as many options as they deemed true, so they marked on those relevant options that they perceived to be true. As a result, there was not a single communicative function of text messages which got less than 100 (20.00%) responses out of 500. The results show that three communicative functions (i.e., A, D and G) each secured more

than 50% agreement of 500 respondents. The remaining four options (i.e., B, C, E, and F) received below than 50% favour of the respondents. Overall, a majority of respondents reflected that the options given in Table 4.6 are the communicative functions of text messages.

Eighth option (H) of Question 10 of the questionnaire was open-ended. Here, respondents were free to provide any other communicative function of text messages that was not given in options A to G of Question 10. In response to this open-ended option, respondents suggested many other communicative functions of text messages not covered in the questionnaire options A to G. These communicative functions of text messages offered by respondents are categorized into 11 categories in the following table.

Table 4.7
Respondents Specified Communicative Functions of Text Messages

Response	Number out of 500	Percentage
1) Acquiring/Sharing balance	1	0.20%
2) Emergency communication	3	0.60%
3) Enquiring about health	2	0.40%
4) Keeping contact	8	1.60%
5) Making new friends	1	0.20%
6) Money making	1	0.20%
7) Sharing poetry	1	0.20%
8) Sharing quotes	2	0.40%
9) Teasing friends/others	4	0.80%
10) Using SMS as a pastime	3	0.60%
11) Using SMS at workplace	1	0.20%
TOTAL	27	5.40%

The communicative functions given in Table 4.7 are all specified by the individual respondents of the questionnaire. This being an open-ended option in the questionnaire, only 27 out of 500 respondents, responded to this option. Still, the reflections offered by the few respondents are very valuable to comment on the social,

cultural and ethical trends of Pakistani texters, especially youngsters as they are the most frequent users of SMS.

4.7 Factors of Phonological Adaptations in SMS Language

Incorporation of the features of spoken mode of language into text messages is a common feature of SMS language. To analyse the metalinguistic awareness of Pakistani Texters about this aspect of SMS language, an open-ended question was included in the questionnaire (see Question 11 in Appendix-A). The respondents were asked about the reasons and factors of their mixture of phonological features in SMS language. As the responses were sought in an open-ended question, the responses were a combination of multiple factors of phonological adaptations in SMS Language. Therefore, their responses have been analysed in two tables.

Table 4.8
Factors of Phonological Adaptations (as combinations of responses)

	Variables/Factors	Number	Percentage
1)	Effort Saving (ES)	12	2.40%
2)	Style/Trend (S)	34	6.80%
3)	Space Saving (SS)	15	3.00%
4)	Style + Time Saving (STS)	20	4.00%
5)	Style+Time+Space Saving (STSS)	2	0.40%
6)	Time+Effort+Money Saving (TEMS)	1	0.20%
7)	Time+Effort+Money+Space Saving (TEMSS)	1	0.20%
8)	Time+Effort Saving (TES)	20	4.00%
9)	Time+Money Saving (TMS)	4	0.80%
10)	Time Saving (TS)	297	59.40%
11)	Time+Space Saving (TSS)	28	5.60%
12)	Understanding Important (UI)	11	2.20%
13)	Irrelevant Response (IR)	7	1.40%
14)	Response Not Given (NG)	48	9.60%
TOTAL		500	100.00%

The results of Table 4.8 indicate that responses given in the table are combinations of multiple factors/variables as given by the respondents. To reduce the above mixed data, from 14 mixed categories into eight individual categories, the

combinations of different factors have been individually counted in the next table. Therefore, the comments have been given after the next table. In the next table, all combined variables of the previous table have been individually counted to classify them into individual factors. Therefore, the next table presents a clear picture of individual factors of phonological adaptations in SMS language.

Table 4.9
Factors of Phonological Adaptations (as individual responses)

Individual Factors	Individual Factors Added	Number	Percentage
1) Effort saving (ES)	12+1+1+20	34	6.80%
2) Money saving (MS)	1+1+4	6	1.20%
3) Style/Trend (S)	34+20+2	56	11.20%
4) Space saving (SS)	15+2+1+28	46	9.20%
5) Time saving (TS)	20+2+1+1+20+4+297+28	373	74.60%
6) Understanding important (UI)	11	11	2.20%
7) Irrelevant response (IR)	7	7	1.40%
8) Response not given (NG)	48	48	9.60%

Table 4.9 shows a clear picture of the individual factors which are the most prominent causes of phonological adaptations in SMS language. The results show that most prominent factor of phonological adaptations in SMS language is the time saving factor. As time is perhaps the most important commodity in human lives, therefore 373 (74.60%) out of 500 respondents reflected in the favour of this factor as being the main cause of incorporating features of spoken language into SMS language.

4.8 Factors of Code and Script Adaptations in SMS Language

Code-switching/mixing between Urdu and English languages is a common feature of Pakistani speakers. In contrast, this feature is comparatively rare in the written mode. But, when we discuss SMS language, this written mode is replete with code-switching between the two languages. Therefore, to analyse the metalinguistic awareness of Pakistani texters about code adaptations, Question 12 was included in the questionnaire (see Appendix-A). The respondents were asked about the reasons and

factors for which they use Urdu-English code-mixed language, and Romanized Urdu, in their text messages.

In this regard, the responses of texters were categorized into 11 categories as given in the next table. Here it is important to mention that in response to this open-ended question, texters themselves were to provide factors of code-switching, therefore 77 (15.40%) respondents did not give any response to this question. The results are shown below.

Table 4.10
Respondents Specified Factors of Code and Script Adaptations

Response	Number of Respondents	Percentage
1) Ease in writing	103	20.60%
2) Ease in writing and understanding	18	3.60%
3) Fluency in writing and understanding	2	0.40%
4) As a habit	6	1.20%
5) Love with Urdu	15	3.00%
6) Roman keypad	9	1.80%
7) Style/trend/fun	27	5.40%
8) Time saving	24	4.80%
9) Understanding of others	213	42.60%
10) Irrelevant answer	6	1.20%
11) Response not given	77	15.40%
TOTAL	500	100.00%

The above table shows many possible reasons and factors of code and script adaptations in text messages as reflected by the respondents of the questionnaire. The table shows the factors for which texters mix English and Urdu in their text messages. The results show that the two most prominent reasons of code-switching reflected by texters are 1 and 9 of the above table. The factor of *Understanding of Others* is favoured by 213 (42.60%) respondents, followed with the factor of *Ease in Writing* favoured by 103 (20.60%) respondents. It shows that texters feel ease in composing text messages in the Roman script. Similarly, they think that text messages written in Romanized Urdu

can be easily understood by their recipients. These factors also imply that most Pakistani texters feel ease in writing and understanding messages written in Romanized Urdu.

4.9 Sociocultural Impacts of Text Messages

There is a general perception in Pakistani society and especially among the elderly people that SMS communication is negatively affecting the social, cultural, ethical and religious values of the society. This perception is so strong in Pakistan that the researcher of this study was advised by the Board of Advanced Studies and Research of his university that *while making linguistic analysis of SMS in Pakistan the researcher should correlate the same with the ethical and cultural values.* Therefore, the last question (Question 13) of the questionnaire was based on the same issue whether texters think that text messages conflict with the social, cultural, ethical and religious values of Pakistani society. This was a close-ended question with three options, that is, YES, NO and UNCERTAIN. The results of the responses are offered as shown below.

Table 4.11
Sociocultural Negative Impacts of Text Messages

Response	Number of Respondents	Percentage
Yes	315	63.00%
No	93	18.60%
Uncertain	83	16.60%
Response not given	9	1.80%
TOTAL	500	100.00%

Table 4.11 shows that a majority of texters, that is, 315 out of 500 reflect that SMS communication has negative impacts on the social, cultural, ethical and religious values of our society. The responses opposite to the majority opinion were only 93 (18.60%). These results correspond to the general opinion of the people of Pakistan about the social, cultural, ethical and religious impacts of SMS mode of communication.

CHAPTER 5

LEXICAL ADAPTATIONS

It is a common observation of linguists, educationists, media critics and general masses that SMS orthography is replete with short forms of words. In this regard, this chapter deals with the analysis of lexical adaptations made in text messages at the word or lexical level. The chapter has been mainly divided into four sections. The first section gives an introduction to lexical adaptations in text messages. Sections two and three cover the principles and patterns of morphological and shortenings. Section four examines spelling and letter-case variations in text messages. The chapter presents critical discussions on the most frequent lexical/orthographic adaptations in the light of results generated with the help of AntConc software. Moreover, examples of text messages from SMS-Corpus for various adaptations have been given in section 5.4.

5.1 Introduction to Lexical Adaptations

It is an established fact that all living languages adapt themselves and respond to the needs of their users (Rumsiene, 2006). Halliday & Hasan (1985), in their theory of the relationship of language and society, claimed that every language has to adapt its functions to the needs of the society. They form the idea that a living language may not at any point of its development become stable, where it does not accept further changes. It is therefore highlighted that the phenomenon of adaptations to the needs of the society is an indispensable feature of any language. This is true because languages exist only through societies in which they develop and grow. Consequently, language users modify languages according to their own needs, and the needs of related technologies to these languages (Rumsiene, 2006).

Linguistic adaptations in text messages to the needs of texters are therefore based on the general phenomenon of language adaptation and change. Linguistic adaptations in text messages are of different kinds, but this chapter deals only with lexical adaptations. Regarding lexical adaptations in text messages, it is pertinent to mention that the most notable and visible adaptations are of short forms of words, commonly known as *shortenings* or *abbreviations*. The SMS variety of language is full of short forms of words, and the terms *shortenings* or *abbreviations* are the generic names mostly used for such short forms of words.

The problem with these two terms, that is *shortenings* and *abbreviations,* is that linguists are not agreed on their uniform definitions and their classifications under certain heads. Hence, for the varying definitions of these terms, first, these terms are clarified here. In a broader sense, both the terms have the same meaning and are interchangeable, yet there are the examples in the available literature where *abbreviations* have been categorized as a subcategory of *shortenings* (Bosco, 2007), and oppositely, *shortenings* as a subcategory of *abbreviations* (Crystal, 2008b). This type of paradoxical classification of these terms is confusing. Therefore in the study, the broader term of the two, that is *shortenings,* has been used to refer to all types of lexical short forms. Furthermore, to remove the ambiguity of definitions of subcategories of *shortenings*, all confusing terms used for short forms in the chapter have been defined and explained through examples.

Coming back to *lexical shortenings*, it is stated that they are the most frequent feature of SMS text messages. These shortenings are of different kinds, and are caused by various factors. The variant kinds of *shortenings* can be classified under two umbrella terms, i.e. 1) *morphological shortenings*, and 2) *phonological shortenings* (Bodomo, 2010; Crystal, 2004).

Of these two categories, shortenings through morphological adaptations are the most common type of all lexical *shortenings*, either in SMS or in the standard language. They are the most frequent phenomenon of SMS word formation. Therefore, the analysis is started from the shortenings made through morphological adaptations.

5.2 Morphological Shortenings

The process of morphological adaptations in text messages involves *letter reductions* from words for the purpose of brevity, where a single letter or a group of letters are omitted from a word. The phenomenon of letter reductions involves the dropping of letters from words by using certain processes. These processes of letter reductions are divided into certain subcategories. In the case of letter reductions in text messages, these processes have been subcategorized under three heads, i.e. 1) *initialism*, 2) *contraction*, and 3) *clipping*. The said subcategories, as earlier stated, face the difficulty that they do not have established agreed-upon definitions by linguists, so they have been defined and explained through examples.

5.2.1 Initialism (Alphabetism and Acronymism)

Crystal (2008b), an authority on SMS and CMC communication, states that *initialism* involves the dropping of all letters from words except initial letters. The term *initialism* can be further divided into *alphabetism* and *acronymism* on the bases of difference of pronunciation of shortened words through this process. In the case of *alphabetism*, all letters in the term are pronounced as separate alphabets like AIOU, IIUI, UN, and USA. On the other hand, *acronyms* are pronounced as single words like NUML, NUST, NATO, and UNESCO etc.

In this study, both the processes of *alphabetism* and *acronymism* have been analysed under their umbrella term *initialism*, and the results have been tabulated in the following table.

Table 5.1
10 Most Frequent Initials in SMS-Corpus

Initials	Frequency	Target Expression
Aoa (75), AOA (38), AoA (20), aoa (7)	140	*Assalam-o-Alaikum* (peace be upon you)
pm (46), PM (7)	53	Post meridiem (after midday)
sms (36), SMS (12), SmS (5)	53	Short Message Service
am (25), AM (10)	35	Ante meridiem (before midday) [not counted for verb form "be"]
ATI (11), ati (9)	20	*Anjman Tulba Islam* (a student organization in Pakistan)
S.A.W.W (8), SAW (7), S.A.W (1), S.A.W. (1), S=A=W=W (1), S-A-W-W (1)	19	*Sallallahu Alaihi (Wa'alehi) Wasallam/* (Peace be upon him) [used with prophet Muhammad (PBUH)]
tc (13), Tc (3), TC (2)	18	Take care
isi (13), Isi (2)	15	Inter Services Intelligence
nust (5), NUST (3), Nust (1)	9	National University of Science and Technology
hec (4), HEC (2)	6	Higher Education Commission

Note. Letter token class activated in AntConc

The results generated with the help of AntConc, in Table 5.1, show that words from both English and Urdu languages are adapted as *initials*. Out of the 10 most frequent *initials* in SMS-Corpus given in the table, three are of the Urdu language. These are AOA (the most frequent initials in SMS-Corpus), ATI, and SAW/SAWW. This aspect of results shows that the phenomenon of *initialism* is not limited to the English language only.

The results also show that both the categories of *initialism* that is *alphabetism* (e.g., AOA, PM, SMS) and *acronyms* (e.g., NUST) are used in text messages. But the analysis of the results shows that the majority of *initials* in the table are pronounced as *alphabets* and not as *acronyms*, with the exception of HEC (pronounced in both ways) and NUST. This implies that their pronunciation as *alphabets* is more common than as *acronyms*.

Another aspect of the pronunciation of *initials* is that one cannot claim about certain words whether they fall under the category of *alphabetism* or *acronymism*. Here,

it is important to mention that if a word made up of initial letters can be pronounced in both ways, it is not at once decided by all speakers whether to pronounce it as an *alphabet* or an *acronym*. The pronunciation of such *initials* sometimes takes time to be finally determined by the speakers whether to pronounce them as *alphabets* or *acronyms*. One such example from the results is of *initial* letters "HEC". These *initial* letters are pronounced in both ways in Pakistan, i.e. as *acronym* /hek/, and as *alphabets* /eɪʧ i: si:/. Regarding the pronunciation of the initials of this term, there is no linguistic restriction for the term to be pronounced in one way or the other. So in this case, time will determine its stable pronunciation based on the frequency of its use, which of the two pronunciations remains in fashion. At present, according to the personal observation of the researcher, it is mostly used as *alphabets,* that is, /eɪʧ i: si:/. This complication of the pronunciation of *initials* as *alphabets* or *acronyms* also suggests that the bifurcation of these two terms seems redundant in some cases, and may be avoided if it is not indispensable.

The striking point of SMS *initialism* is that everyday expressions like "*assalam-o-alaikum"* (peace be upon you), and "take care", are frequently used as *initials* to save time and effort of texters.

5.2.2 Contractions

Contractions are an important feature of text messages. They involve morphological or letter reductions at the middle of words or phrases. In its nature, a contraction is the shortened form of a single word, two words, or very rarely multiple words, by dropping middle letters (Thurlow & Brown, 2003). The *single-word* and *two-word* contractions are common types, and involve omissions of letters at their middle positions. The omitted letters are normally replaced by an apostrophe, but the apostrophe

may also be omitted in text messages. A view of the 20 most common contractions used in SMS-Corpus is given in the following table.

Table 5.2
20 Most Frequent Contractions in SMS-Corpus

Contractions	Frequency	Target Expression
I m (73), i m (19), im (13), I'm (3), I M (2), Im (1), IM (1)	112	I am/I'm
dont (29), Dont (23), dnt (12), dn (11), don't (10), Don't (4), Dn (1), Dnt (1), don,t (1)	92	Do not/don't
its (52), Its (23), It's (5), itz (5), it's (2), IT'S (1), IT's (1), ITS (1), Itz (1), iTs (1)	92	It is /It's
Dr. (32), dr (14), Dr (4), dr. (7), DR. (3)	60	Doctor
nt (45), Nt (7),	52	Not
wht (17), wt (10), Wht (9), Wt (2)	38	What
Ill (25), I ll (5), i'll (3), I'll (2),	35	I will/I'll
Abt	28	About
bt (21), Bt (4)	25	But
frm (21), Frm (1)	22	From
Hw (11), hw (10)	21	How
clg (11), colg (9), Colg (1)	20	College
wl (19)	19	Will
can't (10), cant (7),	17	Cannot/can't
didnt (11), didn't (3)	14	did not/didn't
thats (8), Thats (4), that's (2)	14	That is/that's
cme (8), Cme (3)	11	Come
brng (9), Brng (1)	10	Bring
cmng (10)	10	Coming
wen (9), Wen (1)	10	When

Note. All four token classes activated in AntConc

The results in Table 5.2 show that both *single-word* and *two-word* contractions are found in SMS-Corpus. Out of twenty most frequent contractions in SMS-Corpus given in the table, seven are *two-word* contractions, whereas 13 are *single-word* contractions. These results of the *single- word* and *two-word* contractions highlight that both *single-word* and *two-word* contractions are commonly found in text messages. In the *single-word* contractions in SMS any letter from anywhere, from a word, can be omitted.

When *two-word* contractions of text messages are compared with that of the standard language *two-word* contractions, it is note-worthy that their occurrence in text

messages is much more than that of the standard language. In this scenario, it is further highlighted that the standard language is divided into formal and informal categories. Although, informal categories of standard language, like letters and dialogues, use *two-word* contractions more frequently, yet their frequency in text messages is much greater.

The results also indicate another feature of SMS contractions that the use of apostrophe in contractions is very rare, irrespective of the fact whether they are *single-word* or two-*word* contractions. This apostrophe-free feature of SMS contractions is very rare in the contractions of the standard language.

5.2.3 Clippings

In the shortenings of words, clipping is the process of reduction of a word to one of its parts. In this process letters are dropped from words, at their 1) beginning, 2) end, or 3) both beginning and end. This process can be further clarified through following categorization.

1. Back clipping (ending part of a word is dropped)
2. Fore-clipping (beginning part of a word is dropped)
3. Middle clipping (both beginning and ending parts of a word are dropped)

According to Brinton and Brinton (2010), clipping is the result of deliberately dropping part of a word, either the end or the beginning, or less often both, while retaining the same meaning and same word class. Likewise, Crystal (2008a) states that clipped forms are the reductions of longer forms of words, usually removing the end of the word like *ad* from *advertisement*, sometimes dropping the beginning like *plane* from *airplane*, and both beginning and ending together like *flu* from *influenza*.

In the light of the given definitions, the results of the most frequent clippings from the corpus are given below.

Table 5.3
10 Most Frequent Clippings in SMS-Corpus

Clippings	Frequency	Target Expression
u (879), U (137)	1016	You (first two letters *yo* clipped)
r (314), R (80)	394	Are (first letter *a* and last letter *e* clipped)
b (250), B (68)	318	Be (last letter *e* clipped from verb "be")
ur (227), Ur (22), UR (10)	259	Your (first two letters *yo* clipped)
m (181), M (58)	239	Am (first letter *a* clipped)
n (157), nd (43), N (30), ND (2), Nd (1)	233	And (first letter *a* and last *d* clipped)
ok (109), Ok (92), OK (8), oka (3), Oka (1)	213	Okay (last letters *ay* and *y* clipped)
wil (97), Wil (4), WIL (2)	103	Will (last letter *l* clipped)
uni (60), univ (12), Uni (5), UNI (2)	79	University (last letters *versity* and *ersity* clipped)
hav (33), Hav (4)	37	Have (last letter *e* clipped)

Note. Letter token class activated in AntConc

The analysis of the results in Table 5.3 shows that out of 10 most frequent clipped words, eight are monosyllabic words except "okay" and "university". This phenomenon of clipping of short/monosyllabic words in text messages is very different from clipping of the standard language where longer/multisyllabic words are usually clipped (Crystal, 2008b; Yule, 2006). The purpose of clipping of monosyllabic words seems obvious that SMS orthography leads to the ultimate shortenings of words, where texters tend to strip off words till the bare minimal recognizable forms, which is not the case in the standard language.

Regarding the frequency of three categories of clipping that is *back clipping, fore-clipping*, and *middle clipping*, Crystal (2008a) states that *back clipping* is the most frequent form of clippings in the standard language, followed by *fore-clipping* and *middle clipping* respectively. When this phenomenon is observed in the results of Table 5.3, the phenomenon seems also valid for SMS clippings. The results show that five most frequent clipped words (B, OK, WIL, UNI, HAV) fall under the category of *back clipping,* four (U, UR, M, ND) under *fore-clipping,* and only two (R, N) under the category of *middle clipping.* These results strengthen Crystal's (2008a) views about the

frequency of three categories of clippings both in the standard language as well as in the language of text messages.

A very important finding from these results is that in most cases only vowel letters have been clipped. Out of the ten most frequent clipped words in the corpus, eight involve *vowel clippings*, leaving only two words "will" and "university" where consonant letters have been clipped. Hence, the results in this table foreground that vowel dropping is a frequent phenomenon in SMS lexical shortenings. This aspect of findings corresponds to Crystal (2008b, p. 26), where he states that texters intuit the basic principle of *information theory* that "consonants carry much more information than vowels". Here, it is further clarified that the English people are unaccustomed to vowel-free writing, while in many languages of the word like Arabic and Urdu, short vowels are not written. If needed, their function is performed through *diacritics*. This shows that vowels carry much less information than consonants and so vowels can be easily omitted from words, without the loss of lexical or semantic information.

In the case of the clipping made in the word "will", where only the last letter "l" has been clipped, the results correspond to Crystal's (2008b, p. 46) theory that "final consonants are often dropped too". The justification, he offers, is that the last consonants are usually silent and so their clipping does not cause loss of any information. He further states that double medial consonants can also be reduced to a single consonant because this process also does not involve loss of any lexical information.

5.2.4 Discussion on Morphological Shortenings

Regarding the *initialism* in text messages, Crystal (2008b) claims that initialism in itself is not a new phenomenon, and people have been initializing common phrases for centuries. He claims that the phenomenon of initialism is so common in the English language that in some cases people forget whether the words like CD, DVD and AIDS

are initials or actual words. In some cases people may even be unable to tell what the initials originally stood for, as is the case with LASER (light amplification by the stimulated emission of radiation). He also provides a history of initialism in the English language.

> The Latin initialism pm (post meridiem 'after midday') is first recorded in English in 1666; NB (nota bene 'note well'), in 1673. IOU is known from 1618. In the nineteenth century we find RIP ('rest in peace') and ND ('no date'). In the early twentieth century we have AWOL ('absent without leave'), NBG ('no bloody good'), and SWALK ('sealed with a loving kiss'). The mid-century produced such forms as ETA ('estimated time of arrival'), SNAFU ('situation normal, all fouled/fucked up'), AKA ('also known as'), and the famous TTFN ('ta-ta for now'), used by the Cockney charlady Mrs Mopp in the BBC radio series ITMA ('It's That Man Again'). (p. 43)

In his well acclaimed encyclopedia, "The *Cambridge Encyclopedia of the English Language*", Crystal (1995, p.120) gives some more examples of the historical use of initial language. He cites a writer of New York "*Evening Tatler*" who termed "the initial language ...a species of spoken shorthand" in 1839. This historic information gives an idea that whatever hype surrounds the uses of initialism in SMS, is baseless. Actually, this phenomenon is not new in language. The only difference is that the frequency of its use has increased manifold in SMS and CMC.

Furthermore, the comparative analysis of the frequency of *alphabetism* and *acronymism* shows that the process of *alphabetism* is found much more than *acronymism* in SMS communication. Another very important aspect of *initialism* noted in text messages in Pakistan is that *sentence initialism* like "HRU" (How are you?), with the

highest frequency of five in the category of sentence initialism in SMS-Corpus, is very rare in comparison to the initialism found for phrases.

While the *two-word* contractions are mainly discussed in the related literature e.g., "I'm", "don't", "can't", "it's", etc., the *single-word* contractions can also be seen in the available literature. Field (2009, pp. 70-71) discusses both the *single-word* and *two-word* contractions. She says that when two words are combined into one, an apostrophe is placed where the letter or letters are omitted. Some single words are contracted by using the first and last letters only. These are contractions that do not usually need a full stop at the end, e.g., "Dr", "Mr", "Mrs", etc. Very rarely *three-word* contractions are also seen in the informal use like "don't've" for "do not have", however in SMS-Corpus no three-word contraction was found.

Regarding the criticism and that adaptations like clipping are making SMS language a new form of language, Crystal (2008b, pp. 50-51) defuses the situation by saying that this is not a new innovation in text messages. He gives many examples like the days of week (mon, tues, sat), names of months (jan, feb, dec), educational subjects (bio, chem, lang), and many other e.g., uni(versity), bro(ther), min(ute), approx(imately), doc(tor), etc(etera), gov(ernment), and so on. Moreover, he states:

> English has abbreviated words in this way ever since it began to be written down, and all of the above have long histories. Words like exam, vet, fridge, cox, and bus are so familiar that they have effectively become new words. Yet it's worth noting that when some of these abbreviated forms first came into use, they also attracted criticism. (p. 51)

Therefore the criticism that such forms/clippings are totally new forms, and that they are only connected to SMS orthography, is baseless.

5.3 Phonological Shortenings

Both phonological and morphological shortenings are the most frequent features of SMS orthography. Phonological shortenings (adaptations) in text messages involve the representation of words, or parts of words, by using letters, numbers, and symbols due to *phonological/phonetic approximation.* In this process, sounds of single letters, numbers, and symbols resemble the pronunciation of words or parts of words, and so they are used instead of actual spellings of words. Historically, such processes are known as *logographs* or *logograms* which involve the use of symbols, digits, and letters as words (Bodomo, 2010; Crystal, 2008b).

In the analysis of SMS-Corpus, all three categories of letter, number, and symbol homophones have been analysed and categorised, separately as well as combined, for their abundant use in text messages. In the following table, the results of AntConc analysis of SMS-Corpus for *letter homophones/logographs* are presented.

5.3.1 Letter Homophones/Logographs

Of all three categories of letter, number, and symbol homophones used for words or parts of words, letter homophones are most frequent in text messages. In this process, letters/alphabets of the English language serve as *logographs* to represent words, or parts of words. They replace actual spellings where sounds of words or parts of words correspond to the sounds of letters. So, to save time and effort of typing actual spellings, they are used instead, and this has become a common norm among texters, the world over. Now, the results of this category from SMS-Corpus are presented in the next table.

Table 5.4
10 Most Frequent Letter Homophones in SMS-Corpus

Letter Homophones	Frequency	Target Expression
u (879), U (137)	1016	You
k (849), K (111)	960	/kei/ (Urdu postposition for *in, of, after*) /keih/ (Urdu conjunction for *that*)
r (314), R (80)	394	Are
b (250), B (68)	318	Be (English verb form *be*)

		/b^{h}i:/ (Urdu word meaning *also*)
ur (227), Ur (22), UR (10)	259	Your/u+r (two letter homophones)
m (181), M (58)	239	Am
n (157), N (30)	187	And
g (44), G (31)	75	/dʒi:/ (words *yes, sir, madam* in Urdu)
Q (36), q (29)	65	/kjũ:/ (word *why* in Urdu)
h (27), H (23)	50	How
		/hɛ/ and /hẽ/ (verb form *be* in Urdu)

Note. Letter token class activated in AntConc

The results in Table 5.4 show the abundant use of letter homophones for words. In this connection, Bodomo (2010) and Crystal (2008b) state that letter based logographic writing is an important feature of text messages. This phenomenon of SMS language also reminds us of the earlier logographic writings of the ancient civilizations. Historically speaking, this phenomenon is not new in nature, but when observed in the modern formal written language it is rarely found. So, the abundant use of letter logographs in text messages is a specific feature of the register specific language used in SMS and CMC modes of communication.

The results of the study show that this phenomenon is not only limited to the English language, because Urdu words have been equally replaced with letter homophones of the actual words. The table shows that, if letter "U" has been most frequently used for the approximation of English "YOU", the letter "K" has been most frequently used for Urdu "/keɪ/" and "/keɪh/". This trend shows that the Urdu language is also in no exception for the use of letter homophones for words. In SMS-Corpus, nine letters "U, K, R, B, M, N, G, Q, and H" have been mostly used for their sound/pronunciation approximation with words, or parts of words.

5.3.2 Number Homophones /Logographs

In text messages, number or digit homophones are another most frequent category of phonological approximation. Like letter homophones, number homophones are also used in both cases that is, for full words as well as for parts of words. The most

frequent occurrences for this category are also tabulated separately, in the following table.

Table 5.5
10 Most Frequent Number Homophones in SMS-Corpus

Number Homophones	Frequency	Target Expression
2	397	To, Too, Two, Second (as in 2nd)
4	253	/f/ (as in 4rm), -fore (as in b4), For, Four, Fourth (as in 4th)
1	201	One, First (as in 1st)
3	138	Three, Third (as in 3rd)
8	128	/t/ (as in w8/wait), Eight, Eighth (as in 8th)
20	70	Twenty
9	69	/n/ (as in f9/fine, 9t/night), Nine, Ninth (as in 9th)
10	67	Ten, Tenth (as in 10th)
5	67	Five, Fifth (as in 5th)
30	59	Thirty

Note. Number token class activated in AntConc

In the case of number/digit homophones, it is pertinent to mention that all numbers, in their very nature, are already logographs/symbols of their corresponding numbers written in spellings/words like "1" for "ONE", and "2" for "TWO" etcetera. So, when AntConc analysis was carried on SMS-Corpus to find out number homophones, it was revealed that digits have been mostly used to replace corresponding numbers in words, both in Urdu and English. Furthermore, when the list of 10 most frequent number homophones in Table 5.5 was analysed, it showed six digits/numbers which were only used for numbers and not for words. Therefore, such redundant numbers in the table have been shaded dark as they do not refer to any significant findings.

The significant numbers/digits used in SMS-Corpus as number homophones are "2", "4", "8", and "9". Only these four numbers have been frequently used as number homophones or phonetic approximants of words, or parts of words. Another important finding from the results is that numbers/digits have been very rarely used as homophones or phonetic approximants for Urdu words in SMS-Corpus.

5.3.3 Symbol Homophones/Logographs

Symbols or graphical signs are an important part of the language used in text messages. Some symbols like "&", "%", and "#" have been historically used as *logographs/logograms* in the English language too. As logographs they serve as words, and are used for saving time and effort. Texters also use them for innovation in their messages, and to make their messages creative and appealing. In the following table, the results of AntConc analysis have been presented for various uses of symbols in text messages.

Table 5.6
10 Most Frequent Symbol Logographs in SMS-Corpus

Symbol Logographs	Frequency	Target Expression
&	198	And
%	27	Percent
#	26	Number
@	25	Used for *At*, and letter *A*
»	23	Used as *number* sign
§	16	Used for symbol *&* and letter *S*
$	12	Used for letter *S* [*NB. Regex deactivated*]
£	12	Used for letters *L* and *E* (as in £and and ON£)
«	8	Used as number sign
¥	4	Used for letter *Y* (as in ¥our)

Note. Symbol and Punctuation token classes activated in AntConc (Regex activated)

Table 5.6 shows the use of those symbols which have been mostly used in the corpus as words, or parts of words. When symbols are used in this way to represent words or parts of words, they are called logographs. The use of logographs is not a new development of texters. Signs and symbols have been used as words since long in many languages of the world. Chinese, Japanese and many other languages still use symbols for words and in these languages such symbols are termed as characters. Hence, the point is that signs and symbols used as words or parts of words in text messages is not a novel thing. Even in the alphabetic languages like English they have been used since long. The

only difference is that texters have increased their use manifold, and so such symbols sometimes seem to be striking in text messages for their abundant use.

In Table 5.6, 10 most frequent symbols have been shown which are used to replace words, parts of words, or letters. Here, an important point of discussion is that in SMS-Corpus, only four symbols "&", "%", "#", and "@" have been used as symbol logographs/homophones of words. The remaining six symbols given in the table have been used in unusual ways. They have not been used for those words or purposes for which they are conventionally used. For example the symbols "§", "$", "£", and "¥" are conventionally used as logographs/signs for "Section", "Dollar", "Pound", and "Yen/Yuan" respectively, but in SMS-Corpus they have been used in different ways. Likewise, symbols/signs "»" and "«" are named *guillemets*, and are conventionally used as "angle quotes/speech marks" but in SMS-Corpus they have been used as number signs. The reasons seem obvious that texters used them only to beautify their text messages, to differ with others in style.

Crystal (2008b) considers symbols as one of the most noticeable feature of text messages. He asserts that these typographic symbols are used to represent words, parts of words, or even sometimes noises associated with actions like "X" for "kiss". He states that "When graphic units are used in this way, they are technically known as logograms or logographs – or, in the case of some languages (such as Chinese) characters" (p. 37).

5.3.4 Mixed Use of Letter, Number and Symbol Homophones

Sometimes, letter, number, and symbol homophones/logographs are mixed together in words, where one part of a word is occupied by one category of homophones and other part of the word by another category of homophones. The ideal homophones/logographs are those where single letters, numbers, and symbols serve as

words or parts of words. Some examples of their combined/mixed use like "f9" for "fine", "w8" for "wait", and "b4" for "before" can be noted in the following table.

Table 5.7
10 Most Frequent Mixed Homophones in SMS-Corpus

Letter, Number and Symbol Homophones	Frequency	Target Expression
4m (18), 4rm (7)	25	From
w8 (20), wa8 (2)	22	Wait
2morrow (9), 2moro (5), 2mrw (4), 2morow (3)	21	Tomorrow
2day (17), 2dy (1)	18	Today
f9 (9)	16	Fine
w8ing (9), w8ng (4), w8g (2)	15	Waiting
gr8(10), g8 (2)	12	Great
b4 (11)	11	Before
ni8 (4), 9i8 (1)	5	Night
b+ (4)	4	B plus (B+)

Note. All four token classes activated in AntConc (Case deactivated)

The results of table 5.7 show, as pointed out in the discussion of section 5.2.2, that only four number homophones, "2, 4, 8, 9" have been used as homophones in letter-number homophones category. Hence, only these four numbers are significant as homophones of words, or parts of words.

5.3.5 Discussion on Phonological Shortenings

On the whole, the results of phonological adaptations show that SMS orthography is gradually blurring the established boundaries of speech and writing. In this regard, Crystal (2004) analysed the online *computer mediated communication* and coined the term "Netspeak" for this type of spoken-cum-written language. He remarks that "Netspeak is better seen as written language which has been pulled some way in the direction of speech than as spoken language which has been written down" (p. 47).

Now the question arises whether the phonological approximation of letters, numbers, and symbols is a new phenomenon developed by SMS or CMC users. This question has been answered by Crystal (2008b), while he declares that "there is actually nothing novel at all about such text messages as c u l8r . . . . They are part of the

European ludic linguistic tradition, and doubtless analogues can be found in all languages which have been written down" (p. 41).

In a broader perspective, the opinion of Crystal (2008b) is hard to challenge, yet the case of the Urdu language in using numbers/digits as phonetic approximants for words is a bit different. In this regard, the results of Table 5.5 and Table 5.7 show that number/digit homophones have not been used for Urdu words, or parts of words. But, in the case of letter homophones, Crystal's opinion is valid for the Urdu language, and has been proved from the data of the study. Regarding the historical tradition of uses of letters and numbers as logographs or phonetic approximants of letters, he asserts that generations of human beings have used such trends of language in the past. So, we should not be taken aback when we come across such uses of language in text messages.

5.4 Spelling and Case Variations

Variations in spellings and letter cases are a commonly observed phenomenon of SMS language. Some people call these trends as variations and some as deviations. Some critics term these variations as non-standard and unconventional forms. To study such issues about spelling and punctuation patterns of text messages, the analysis has been carried out on SMS-Corpus, and the results of both categories are offered in separate sections.

5.4.1 Spelling Variations

Variations in spellings are a commonly discussed issue concerning text messages. Linguists give different names to these spellings like deviant, variant, non-standard, and unconventional spellings (Crystal, 2008b; Tagg, 2009). There is a general complaint against texters that they invent and use spellings in their own individual ways. Therefore, many words can be seen in text messages with many different spellings. To explore and

discuss this issue of text messages, 30 most frequent words with variant spellings in the corpus have been tabulated in the following table.

Table 5.8
30 Most Frequent Spelling Variations in SMS-Corpus

Variant Spellings	Frequency	Target Expression
to (1039), 2 (219)	1258	To
u (764), you (323), yo (4), yu (3)	1094	You
and (263), & (203), n (168), nd (34)	668	And
is (503), iz (21), z (14)	537	Is
r (353), are (178)	531	Are
the (454), da (41), d (36)	531	The
of (413), ov (15), f (3)	431	Of
ur (258), your (173)	431	Your
b (261), be (162)	423	Be
plz (250), pls (48), please (64), plzzz (4), plx (3)	369	Please
for (229), 4(137)	366	For
will (181), wil (99), wl (19), vl (9), vil (2), vill (2)	312	Will
m (185), am (117)	302	Am
at (209), @ (25), 8 (3)	234	At
good (172), gud (52), gd (9)	233	Good
have (138), hav (36), hv (29), hve (2)	205	Have
what (134), wht (40), wat (14), wt (12)	200	What
with (148), wid (22), wd (6), vd (2)	178	With
not (124), nt (50)	174	Not
come (122), cum (17), com (16), cm (8), kum (2)	165	Come
ok (155), okay (7)	162	Ok/okay
but (130), bt (25)	155	But
where (111), whr (15), whre (6), w (4)	136	Where
that (111),dat (18), tht (6),	135	That
this (123), ths (6), dis (4), ds (1)	134	This
how (89), hw (21), h (15)	125	How
from (71), frm (22), 4m (18), 4rm (8)	119	From
coming (72), comng (11), cmng (11), cuming (7), cming (5), kming (2), cumng (2)	110	Coming
after (79), aftr (18)	97	After
msg (69), message (22)	91	Message

Note. All four token classes activated in AntConc (Case deactivated)

The analysis of Table 5.8 shows that in most cases, variant/deviant spellings have been generated by shortening standard spellings of words. There are only a few instances where deviant words are same in length to the actual standard spellings. Most importantly, there is not a single example where deviant/variant spellings are more in length (number of letters) than actual letters of standard spellings. This proves the

prevalent common view about SMS language that it mainly rests on actual spellings and shortens redundant spellings.

An important aspect of variations and deviations of spellings in text messages is that they are made under certain rules. This is not appropriate to say that no rules are followed. There are instances where unusual spellings are seen but these instances are just a negligible part of the whole body of spellings in text messages. Moreover, such unusual minor instances are because of speed/carelessness in typing, and are just a small part in comparison. In the majority of cases, spelling adaptations can be justified under certain rules. For example, two very important aspects proved through the results of Table 5.8 are that most of the possible shortenings are made either through morphological reductions, or by phonetic/phonological reductions.

To further elaborate these trends of shortenings, 20 actual examples from SMS-Corpus have been given in two categories in the following tables. The first table presents examples of those variations/deviations which are intentionally employed by texters as they fall under certain rules.

Table 5.9
Examples of Variant/Deviant Spellings (Intentional)

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex-1[12]	*wazup!*	What is up?
Ex-2[77]	*wht abt ur* study	What about your study?
Ex-3[117]	*Pls* wait	Please wait.
Ex-4[163]	w8ing 4 u	I am waiting for you.
Ex-5[190]	*Cal pls*	Call me, please.
Ex-6[209]	*Congrats* for getting job.	Congratulations for getting job.
Ex-7[796]	*H r u?*	How are you?
Ex-8[929]	*I lv u*	I love you.
Ex-9[952]	*Cm 4 T.*	Come for tea.
Ex-10[1618]	*Misin u soooo* much!	I am missing you so much!

The table shows deviations of spellings in italics. A look at spelling deviations shows that these deviations have been made very artistically and purposefully, and they account for certain rules like morphological or phonological shortenings as discussed in

the previous sections of this chapter. There are shortenings through initialism, contraction, clipping, phonological approximation, and consonant spelling etcetera. The purpose of such deviations is to save time and effort. Therefore, such purposeful deviations fall under certain rules and patterns of SMS language and are used by most texters.

Contrary to the intentional deviations given in the previous table, the next table presents those deviations which are caused by carelessness or lack of language command of texters. These deviations have been termed *unintentional* because they cannot be attributed to certain patterns/principles of SMS orthography. They do not even fall under the category of those spellings which are frequently used by texters under certain patterns/principles of SMS spellings.

Table 5.10
10 Examples of Variant/Deviant Spellings (Unintentional)

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex-11[286]	… *theyre* is no option	… there is no option.
Ex-12[338]	Your pic was on the front *oage*.	Your picture was on the front page.
Ex-13[432]	*Prayewrs* change things, … *thiknk* of wat ALLLAH can do 4 u!	Prayers change things, … think of what Allah can do for you!
Ex-14[748]	work hard *fpr* the success	Work hard for the success.
Ex-15[1399]	when you are *cominh* home??	When are you coming home?
Ex-16[1667]	*thankx* we enjaoyed a lot! :D	Thanks we enjoyed a lot! :D
Ex-17[1671]	Congratulations on *yopur* brothers wedding!	Congratulations on your brother's wedding!
Ex-18[3654]	*WHWRE*	Where?
Ex-19[4885]	who *si* at home	Who is at home?
Ex-20[4999]	Can u collect *2morrnw* to be safe?	Can you collect tomorrow to be safe?

The spelling deviations given in italics in the above table are unintentional in nature. They do not fall under any recognized principles of SMS language. Such unintentional and carelessly made deviations are even not supported by the supporters of SMS orthography. Instead, they fall under the category of typos. They can be attributed to elements like the slip of thumb, hurry, careless attitude, or lack of language command, but not to any set principles/patterns of SMS orthography.

5.4.2 Case Variations

Like variations/deviations in spellings, variations/deviations in letter cases are another commonly observed issue of text messages. To explore this issue, the analysis of SMS-Corpus was carried out through AntConc, and the results with the discussion on the issue are presented in following table and its follow up discussion.

Table 5.11
20 Most Frequent Case Variations in SMS-Corpus

ALL UPPER CASE	all lower case	Upper and lower case combined	Fre-quency	Target Expression
TO (20)	to (971)	To (54), tO (2)	1047	To
OF (26)	of (384)	Of (6)	416	Of
SIR (16)	sir (175)	Sir (136)	327	Sir
FOR (15)	for (210)	For (8)	233	For
ALLAH (120)	allah (5)	Allah (171)	296	*Allah*
PLZ (16)	plz (141)	Plz (88)	245	Please
YOUR (17)	your (139)	Your (17)	173	Your
GOOD (26)	good (42)	Good(96), GooD (2), gOOd(1)	166	Good
AOA (38)	aoa (7)	Aoa (75), AoA (20)	140	*AOA*
EID (44)	eid (12)	Eid (40), EiD (2)	98	*Eid*
MORNING(16)	morning(37)	Morning (25), MorninG (2)	80	Morning
DUA (22)	dua (13)	Dua (21), dUa (1), DuA (1),	58	*Dua*
QURAN (27)	quran (6)	Quran (36)	69	*Quran*
MUBARAK(20)	mubarak (12)	Mubarak (23), MubArak (1)	56	*Mubarak*
SMS (12)	sms (36)	SmS (5)	53	SMS
SAW (3)	saw (5)	(NIL)	8	*SAW*
AMEEN (15)	ameen (2)	Ameen (16)	33	*Ameen*
GOD (13)	(NIL)	God (16)	29	God
JAZAKALLAH (13)	(NIL)	JazakAllah (2), Jazakallah (4), JazaKAllah (1), JazakALLAH(1)	21	*Jazakallah*
E-MAIL (15)	e-mail (3)	(NIL)	18	E-mail

Note. Letter token class activated in AntConc (Case activated)

Table 5.11 clearly shows that like variations of spellings, variations in the upper and lower cases of letters/words can also be noted in text messages. The results have been divided into three categories that is "all upper case", "all lower case", and "upper and lower case combined". As far as "all upper" and "all lower" case words are concerned, they are normally due to ease of typing, that means if screen-window is set at

lower case, texters will use all lower case, and if it is set at upper case they will use all upper case. Sometimes, upper case is used to emphasize on words.

Moreover, "all upper", "all lower", and "upper and lower mixed" cases are the part of linguistic literary tradition about which Crystal (2008b) also refers. The noticeable significant point is that in the "upper and lower combined" category, there are variant creative innovations. For creative and stylish purposes, texters employ mixed upper and lower cases in unusual but appealing ways.

To further highlight these patterns of letter case variations in text messages, the next three tables present examples of case variations in actual sentences from SMS-Corpus. These three categories are "all upper", "all lower", and "upper and lower mixed" cases of letters as given in the next tables.

Table 5.12
Examples of Variant/Deviant Cases (ALL UPPER)

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex-21[115]	W R U P	Where are you? Please.
Ex-22[153]	I LOVE YOU	I love you.
Ex-23[186]	K……T.C	Okay, take care.
Ex-24[732]	EID GREETING 2 U AND UR FAMILY.	Eid greetings to you and your family.
Ex-25[2057]	U R WELCOME.	You are welcome.
Ex-26[2377]	FOOD IS READY	Food is ready.
Ex-27[2378]	PL CALL ME	Please call me.
Ex-28[2379]	AIRCRAFT IS BACK	The aircraft is back.
Ex-29[2381]	HOW MUCH DO I NEEDE TO PAY	How much do I need to pay?
Ex-30[2382]	SIR ROOM IS CONFIRMED	Sir, the room is confirmed.

The above table shows letter cases in "all upper" category. All the spellings of words in these examples are upper cases of letters. The next table shows the examples of all lower cases of letters.

Table 5.13
Examples of Variant/Deviant Cases (all lower)

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex-31[14]	hv u done qstions of o.d.e?	Have you done the questions of O.D.E?
Ex-32[19]	how are you?	How are you?

Ex-33[20]	where are you?	Where are you?
Ex-34[24]	okay.	Okay.
Ex-35[139]	when are you coming?	When are you coming?
Ex-36[140]	what is the time?	What is the time?
Ex-37[147]	who are you?	Who are you?
Ex-38[277]	do me a favour.	Do me a favour.
Ex-39[279]	I am in physics department	I am in Physics Department.
Ex-40[474]	we in front of bba department.	We are in the front of BBA Department.

This table shows that all spellings in the given examples of sentences are in lower letter cases. A careful look on the table shows that beginning letters of sentences, beginning letters of proper nouns, and even acronyms are in lower cases of letters. These deviations are normally attributed to the time and effort saving technique of texters, as in this way they save themselves from some extra keystrokes. The next table shows examples of sentences from the corpus where both lower and upper cases of letters have been mixed in variant ways.

Table 5.14
Examples of Variant/Deviant Cases (Upper-Lower Mixed)

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex-41[155]	hello! How r you	Hello! How are you?
Ex-42[167]	GIRL FRIEND MSNG u.	Your girl friend is missing you.
Ex-43[235]	we r sitting in Manag. dept	We are sitting in Management Department.
Ex-44[347]	NuTHiN Much, U telloo :)	Nothing much, you tell :)
Ex-45[349]	DAMN iTs BoooriNgg	Damn! It's boring.
Ex-46[356]	I have received the MONey.	I have received the money.
Ex-47[732]	When Will U Come. WAITING	When will you come? I am waiting.
Ex-48[822]	c u at 6PM	See you at 6pm.
Ex-49[949]	Going 2 Room	I am going to the room.
Ex-50[1515]	BEST OF LUCK FOR tomorrow's quiz!	Best of luck for the tomorrow's quiz!

The above table presents examples of actual text messages where both upper and lower case letters have been used in deviant ways. The mixed use of upper and lower case letters in deviant ways in text messages is mostly attributed to the artistic reasons of adaptations.

5.4.3 Discussion on Spelling and Case Variations

In morphological reductions, words are shortened by clipping/omitting some part of the word; while in phonetic/phonological reductions, actual words or parts of words are replaced by letter or number homophones which act as sound approximant of actual words or parts of words. In result, the analysis shows that texters aim at conveying the sense of the word. Texters aim at conveying the message to their recipients in a concise way by saving time and effort. The same has been stated by texters in metalinguistic analysis about spelling deviations and telegraphic uses of SMS language in Chapter 4 (see section 4.5)

Here, an important point is highlighted. The spelling deviations in text messages are of two types, that are unintentional/unconscious and intentional /conscious deviations. All unintentional/unconscious deviations occur only because of the lack of concentration and they are termed as typographical errors. But the important point is that when corpus based analysis is made on any SMS corpus, it shows that such individual instances are just a few. The majority of deviant spellings are made intentionally and they follow certain rules/patterns. Furthermore, the frequency of individual spellings also determines intentional and unintentional spellings. Intentional variants of spellings are frequently found in the corpus. In this regard, this study shows that if certain spellings are in a large number in SMS-Corpus, by a large number of texters, they follow certain rules/patterns. All such rule governed spellings of words are so large in number that it seems probable that linguists will be ultimately compelled to accept them as part of dictionaries.

This trend of deviant spellings has been termed as *"typoglycemia"* on the Internet. Google search for this word shows numerous entries on many websites. An internet based dictionary defines this word as, "the mind's ability to decipher a mis-

spelled word if the first and last letters of the word are correct".[3] The typoglycemic element of texting portrays that texters are prone to misspell, both unconsciously and deliberately. But Crystal (2008b) defends texters by stating that they mostly manipulate spellings instead of committing mistakes. He supports the use of deviant spellings in his well acclaimed book "*Txtng: The Gr8 Db8*", under the cover of English literary tradition:

> Several of these nonstandard spellings are so much part of English literary tradition that they have been given entries in the Oxford English Dictionary. Cos is there from 1828, wot from 1829, luv from 1898, thanx from 1936, and ya from 1941. (p. 49)

As far as unintentional deviations in spellings are concerned, they may be termed as typographical errors, and so they cannot be supported to be maintained and promoted. Yet, this point is also noticeable that SMS text messages are a register specific variety of language which does not much concentrate on the accuracy of spellings but on the understanding of messages. Therefore, texters do not care for the correction and proof reading of spellings.

The variations/deviations in capitalization are also purposefully made by texters. Mostly, texters use deviant cases where they feel the need for it. Sometimes, they want to create some feeling, or emphasize/undermine something; so they employ deviant cases as a tool. Those cases are very few where case deviations occur due to negligence, lack of knowledge or concentration. Hence, if the variations/deviations in cases of words are purposeful, and are used for creative purposes, critics have weak grounds to criticize. On the other hand, if the deviations are unintentional and are lapses on the part of texters then texters themselves are providing grounds for criticism.

[3] http://www.urbandictionary.com/define.php?term=typoglycemia

Crystal (2008b) believes that variations/deviations are acceptable if they are appropriate to the situations in which they are used. Even a famous grammarian F. R. Palmer (1983, p.16), in his well acclaimed book "*Grammar*" maintains that "some forms of language are acceptable in certain situations . . . ." Furthermore, he throws light on the standard and variant forms of language in the following words:

> It is wrong, then, to consider the dialect form as a corrupt form of the standard. Indeed, it is always wrong to consider dialects as corrupt forms. They are not corrupt, but different, forms, of the language. It may well be that they are not acceptable for many purposes, in the speech of educated people, in the mass media, etc., but this is wholly a matter of social convention, not of linguistic inferiority. This is in no way to deny the importance of social conventions. We break the conventions at our peril, we are dubbed 'ignorant', we fail to get the job we hoped for; but we ought not to provide pseudo-linguistic grounds to justify the conventions. (p.26)

The views of Crystal (2008b) and Palmer (1983) maintain that variant forms of language, whether spellings or cases, are acceptable if they are appropriate to the situations in which they are used. Conventions and standards are right at their own place, but variations/deviations cannot be banned everywhere, otherwise languages will cease to develop. Moreover, the critics of SMS orthography should wait and let these SMS spelling variations/deviations pass the test of time. The filter of time will allow only those trends of spellings to exist which are used by a majority of people, and if they pass this filter test then prescriptive grammarians cannot possibly stop these adaptations made at the level of masses.

CHAPTER 6

SYNTACTIC ADAPTATIONS

The chapter deals with the syntactic adaptations made in text messages. The chapter covers three types of syntactic/grammatical adaptations that are 1) omissions/ellipses of words, 2) reduplications/repetitions of words, and 3) deviant grammar. All three categories of syntactic adaptations have been further classified into various subcategories. The chapter discusses all the patterns and principles of these adaptations and also offers possible reasons for these adaptations in the light of key works in the area. As frequencies of syntactic adaptations in SMS-Corpus are not as abundant as of lexical adaptations, so instead of word/sentence frequencies, examples of sentences have been given in tables for various syntactic adaptations.

6.1 Introduction to Syntactic Adaptations

Modern communication technologies like the Internet and mobile phones have significantly influenced human languages (Baron, 2008; Bodomo, 2010; Crystal, 2004, 2008b; Hard af Segerstad, 2002). The mobile phone based SMS technology is the latest way of fast communication that has spread over the globe. In Pakistan, it is the most economical mode of communication among the youth. In the early studies on SMS language, it was thought that the technical restrictions like the small keypad, cost restrictions like heavy rates of SMS, and space restrictions like 160 characters, are the main motivations for the short forms of language used in text messages (Crystal, 2004; Hard af Segerstad, 2002, Thurlow & Brown, 2003). But the present explorations in the field show that what was believed in the past is not the whole story. Now most of the said reasons no more exist, for example, no heavy rates, and no restriction of 160 characters, yet the same shortened language. The small keypad of mobile phones is also

not the big cause for the short forms of language in text messages. If this had been the main cause, internet users, with full size keyboard, would not use the shortened language. Moreover, modern mobile sets also cater for the limitations of small keypad, by offering touch screen and QWERTY keypads, yet the short forms of language are still in fashion. Zurhellen (2011, p. 638) also carries these opinions while he states:

> Although these shortcomings have been mitigated by improvements to the networks through which messages are sent and by revisions to the keypad that made it resemble a computer keyboard rather than that of a telephone, many users continue to employ a kind of texting shorthand that is efficient, innovative, and playful.

Hence, the more justified reason is that the short forms of language on the Internet and mobile phones are now more motivated by style and speed, than restrictions. Everyone wants to save time and effort, and wishes to adopt the modern style too. For these new trends, SMS communication allows for the use of syntactic and lexical short forms, which save time and effort, as texters have to touch the keypad for lesser time and strokes.

In this regard, Ong'onda, Matu, and Oloo (2011) define that text messaging is the technological mediated discourse that "pursues simple sentences structure for communication" (p. 2). They summarise that syntactic adaptations are found in text messages in the following categories: "omission of pronouns and auxiliary verbs, omission of objects, omission of articles, omission of *to* infinitive, grammatical agreements, contractions and different word orders" (p. 4). They link the syntactic variations in text messages with "the technical attributes of the mobile phone and shared background or context of the situation" (p. 4).

Regarding the syntactic adaptations in text messages, Bosco (2007) states that in text messages simple sentences or even incomplete sentences without subjects are expected. Moreover, articles may be occasionally omitted while omission of personal pronouns, especially the first and second pronoun *I, we* and *you*, may be frequent. Thurlow and Poff (2011) state that text messages contain complex sentence structures in the form of multiple clauses. Harf af Segerstad (2002) states that the omission of auxiliary verbs, personal subject pronouns, and function words are common syntactic adaptations in text messages.

The above syntactic adaptations of text messages can also be compared with MSN messenger texts as stated by Bodomo (2010) that "MSN texts are full of common grammatical errors such as omission of subjects, inappropriate tense forms, sentence structure errors" (p. 72). Baron (n.d.) states that in late 1990s she surveyed the literature on email, bulletin boards, and computer conferencing, and concluded that computer mediated language was a mixed modality that resembled speech. It contained many first and second person pronouns, present tense and contractions, informal style. She also concluded that online language was rude and obscene. Moreover, the messages composed and transmitted online looked like writing, and participants used a wide range of vocabulary and complex syntax. All the conclusions drawn by Baron about the computer mediated online language are no different from the language used in text messages at syntactic level. This chapter shows such features of SMS text messages at the syntactic level.

6.2 Word Ellipses/Omissions

In linguistics, an ellipsis refers to the omission of a word, or words, from a sentence. Bodomo (2010) states that word ellipses are the omissions in discourse, and they are represented with dots in standard writing. But in SMS communication, as a

usual practice, word ellipses occur frequently but no one uses dots to represent them. The abundant use of ellipses in text messages demands for the inquiry of this phenomenon in text messages. In this regard, the analysis of various word omissions/ellipses in SMS-Corpus has been presented in the following sections.

6.2.1 Omissions of Personal Pronouns

Omissions of personal pronouns, which act as subjects and objects of sentences, are the most frequently occurring word ellipses in text messages. In this regard, a few selected examples from the corpus are presented here for analysis.

Table 6.1
Examples of Omissions of Personal Pronouns

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex- 51[17]	M w8ing	*I* am waiting.
Ex-52[119]	cal u later.	*I/we* call you later.
Ex-53[190]	Cal pls	Call *me/we*, please.
Ex-54[196]	Sent u an email.	*I* sent you an email.
Ex-55[555]	m in uni	*I* am in the university.
Ex-56[689]	m going, plz pick @ 12	*I* am going. Please pick *me* at 12 o'clock.
Ex-57[746]	can come to play	Can *you* come to play?
Ex-58[1025]	M studying.	*I* am studying.
Ex-59[1048]	Am f9	*I* am fine.
Ex-60[1374]	I will cme in eve becuze there iz clas.	*I* will come in the evening because there is *my* class.

Note. Ellipses represented in Italics

Table 6.1 shows that the first person pronoun "I", in the subjective case, has been omitted the most of time. Results show that "I", in the subjective case, is the most omitted pronoun in text messages. Other personal pronouns, like "You" and "We" may also be omitted but comparatively in fewer numbers. In the same way, sometimes, first and second person pronouns in the objective case are also omitted. Overall, SMS language is a spoken like language that omits first and second person pronouns, mostly in the subjective case. Hard af Segerstad (2005) states that the omission of subject pronoun is a feature of the spoken informal interaction, and is not associated with

traditional writing. Yet, SMS language being a spoken-like variety of the written language incorporates many features of the spoken language in it.

6.2.2 Omissions of Auxiliary and Copula Verbs

Text messages also omit auxiliary (including modal verbs) and copula verbs very frequently. Auxiliary verbs are the helping verbs that give further semantic or syntactic information about the main/principal verb. In English, these are the 'be" form of verb (is, am, are, was, were, been, being), "dummy verb" (do, does, did), and "have" form (has, have, had). Modal verbs which show modality like (shall, will, may, might, can, could, should, etc.), are also considered auxiliary verbs as they help main verbs.

Copula verbs are the linking verbs which link the subject with its predicate. In English, the most common copula verbs are forms of "be" verb (is, am, are, was, were, been, and being). Some other rarely used copula verbs are like (feel, seem, become, look, taste, smell, sound, appear, get, stay, keep, remain), and so on.

Regarding the omission of these types of auxiliary and copula verbs, various examples from SMS-Corpus are presented in the next table.

Table 6.2
Examples of Omissions of Auxiliary and Copula Verbs

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex-61[474]	we in front of bba department.	We *are* in the front of BBA Department.
Ex-62[513]	Hi. Plz call me bak It urgent	Hi. Please call me back. It *is* urgent.
Ex-63[1264]	How are u and whn u come.	How are u and when *will* you come?
Ex-64[1542]	u there???	*Are* you there?
Ex-65[1560]	Bz?	*Are* you busy?
Ex-66[1603]	Every one waiting 4 u	Everyone *is* waiting for you.
Ex-67[1670]	meeting at 9:00am in conference room	The meeting *is* at 9:00am in the conference room.
Ex-68[1834]	I F9.	I *am* fine.
Ex-69[4769]	sir your phone not responding	Sir your phone *is* not responding.
Ex-70[4837]	clear now	It *is* clear now.

Note. Ellipses represented in Italics

Table 6.2 shows that both auxiliary (helping) verbs and copula (linking) verbs are omitted in text messages. Significantly, among all auxiliary, modal, and copula verbs, the verb form "be", in three present tense forms (i.e., is, am, are) is omitted the most. This feature of SMS communication is also shared with the spoken language. Hard af Segerstad (2005) justifies these omissions by stating that leaving out copula, auxiliary or modal verbs may save several keystrokes of texters.

6.2.3 Omissions of Articles

Omitting articles (a, an, the) is another feature of the language used in text messages. Their omission is presented through examples from the corpus.

Table 6.3
Examples of Omissions of Articles

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex-71[276]	I am infront of library	I am in *the* front of *the* library.
Ex-72[474]	we in front of bba department.	We are in *the* front of BBA Department.
Ex-73[558]	m in meeting	I am in *the* meeting.
Ex-74[1140]	I m in uni	I am in *the* university.
Ex-75[1200]	vr in parking	We are in *the* parking.
Ex-76[1353]	wt is dt fr Exams.	What is *the* date for examinations?
Ex-77[1357]	R u cumming to college?	Are you coming to *the* college?
Ex-78[1374]	I will cme in eve becuze there iz clas.	I will come in *the* evening because there is my class.
Ex-79[1641]	I m in office	I am in *the* office.
Ex-80[1670]	meeting at 9:00am in conference room	The meeting is at 9:00am in *the* conference room.

Note. Ellipses represented in Italics

Table 6.3 shows omissions of articles in text messages. In SMS-Corpus, they are the third most frequent category of omissions in text messages. Among the three articles (i.e., a, an, the), the most frequent ellipses occur in omissions of the definite article "the".

6.2.4 Omissions of Prepositions, Conjunctions, and Infinitives

This section deals with some less frequent omissions of words like prepositions, conjunctions and infinitives. These classes of words are not omitted as much, as the words discussed in the previous three sections, yet their omissions in text messages

cannot be completely ignored. Hence their omissions in text messages are presented in these three categories through examples from the corpus.

Table 6.4
Examples of Omissions of Prepositions, Conjunctions, & Infinitives

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex-81[228]	R u sure its egle.	Are you sure *that* it is an eagle?
Ex-82[402]	I'll be home in acouple ofmins.	I will be *at* home in a couple of minutes.
Ex-83[888]	comng w8 10 mnt	I am coming. Wait *for* 10 minutes.
Ex-84[900]	I think u r bzy.	I think *that* you are busy.
Ex-85[984]	I am worried Miss you	I am worried *and* miss you.
Ex-86[1021]	I dnt listen u.	I do not listen *to* you.
Ex-87[1876]	oks...M comi pick u...	Okay, I am coming *to* pick you.
Ex-88[3922]	W8 2 mint	Wait *for* two minutes.
Ex-89[4661]	Leter 4m FPSC u qualified 4 intrvw .	There is a letter from FPSC *and* you are qualified for the interview.
Ex-90[4751]	Tell the VP I m onm leaqve	Tell the VP *that* I am on leave.

Note. Ellipses represented in Italics

Table 6.4 shows three word classes that have been omitted in the examples taken from SMS-Corpus. Omissions of conjunctions have been shown in five examples (SMS-No 228, 900, 984, 4661, 4751), prepositions in four examples (SMS-No 402, 888, 1021, 3922), and "to" infinitive in one SMS (1876). The results show that conjunctions and prepositions are two word classes/parts of speech that are more frequently omitted than infinitives.

6.2.5 Combined Multiple Word Ellipses/Omissions

In addition to the individual omissions of different word classes made in isolation, there are many instances where various word classes are combined in a single sentence. In such text messages, texters use words very economically and pragmatically, as they share the background situations with their addressees, and hence no semantic/pragmatic information is lost.

Table 6.5
Combined Multiple Word Ellipses/Omissions

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Normalized Response (Guessed)
Ex-91[96]	Library	*I am in the* library.

Ex-92[180]	In Department	*I am* in *the* department
Ex-93[184]	café.	*I am in the* café.
Ex-94[385]	Eating?	*Are you* eating?
Ex-95[1303]	witnin 5 mints.....	*I will reach* within five minutes.
Ex-96[3976]	In room?	*Are you* in *the* room?
Ex-97[4270]	Reached	*I have* reached.
Ex-98[4384]	Nt yet	*I have* not yet *(main verb).*
Ex-99[4385]	Just took	*I* just took *(noun/object).*
Ex-100[4573]	By 8	*I will (main verb)* by 8 *o'clock.*

Note. Ellipses represented in Italics

Table 6.5 shows that in all examples more than one word was omitted. Mostly, one or two words have been written to communicate the whole message. The omitted words are more than the written words. Hence, these combined omissions of multiple words are the instances of telegraphic and pragmatic uses of text messages, by employing the syntactic adaptations in a very artistic and adept way. Although, most of words in these text messages are missing, but their recipients can easily understand messages because of the shared knowledge of texters.

The given examples show that any of the three main parts of a simple sentence (i.e., subject, verb, and object) can be easily omitted by texters. Moreover, at the word level, nouns, pronouns, auxiliary verbs, copula verbs, main verbs, articles, and prepositions have been variously omitted in the above examples (91 to 100), as combined multiple ellipses.

6.2.6 Discussion on Word Ellipses/Omissions

In the adaptations of word ellipses, results were examined in two categories, that is single word ellipses and multiple word ellipses. The related discussion on both categories is offered here. In the single word ellipses (examples 51-90), the personal pronoun "I", in the subjective case, is the most observed case of ellipses. The same has also been advocated in Frehner (2008), that the most common of all word ellipses is the pronoun "I" used as the subject of interactive sentences. Other pronouns which are

normally omitted in text messages are "you" and "we" in the subjective case. These first and second person pronouns are also omitted in the objective and genitive cases, but comparatively in the lesser number. Third person pronouns are rarely omitted.

The second major category of single word ellipses are auxiliary (including modal verbs) and copula verbs. Among them, the present tense forms of verb "be" (i.e., is, am, are), are the most frequent ellipses of all auxiliary/copula verbs.

The third major category of single word ellipses are articles (i.e., a, an, the) in text messages. Omissions of the definite article "the" in text messages in Pakistan is comparatively much higher than they are in the countries where English is used as a native language. The main reason is that the regional languages of Pakistan like Punjabi, Sindhi, Pashto, Balochi, and the national language Urdu, do not use the definite article "the" in the sense as it is used in English. So, Pakistani texters omit the definite article "the" more frequently than it is omitted by the native English speakers.

In addition to the above ellipses, there are many instances of ellipses of conjunctions and prepositions in text messages. In the same way, ellipses of the infinitive marker, "to", can also be seen in text messages.

In the category of multiple word ellipses/omissions, words are omitted in groups. This is the most telegraphic way of text messages, where texters try to convey the messages just in single words, or in two, three words. Hence, the omitted words are normally more than the typed/written words. These combined omissions of various words can be divided into two categories that are omissions at the syntactic level and at the word level. At the syntactic level, all three main parts of a simple sentence (subject, verb, & object) have been omitted in the given examples.

Moreover, in the omissions at the word level, various parts of speech have been collectively omitted in the given examples. The results show that nouns, pronouns,

auxiliary verbs, copula verbs, main verbs, articles, and prepositions have been omitted from the examples given in Table 6.5. This trend of ellipses in text messages shows that texters can omit any word from their text messages, because they communicate through a medium where they share the background knowledge with their recipients, and do not fear for the loss of information.

Overall, the results of this study, regarding the word ellipses in text messages, are not much different from many other studies conducted in different parts of the world (Baron, 2008, Bodomo, 2010; Bosco, 2007; Crystal, 2008b; Frehner, 2008; Hard af Segerstad, 2002; Ong'onda, Matu, & Oloo, 2011; Thurlow & Brown, 2003). In this regard, Frehner (2008) states that syntactic word ellipses in text messages occur in such a large amount that they surpass the word ellipses in any other type of written communication, and even the oral communication.

Hard af Segerstad (2005) terms these syntactic omissions as grammatical reductions and states that these are found in mobile text messaging in order to save time, effort and space. She asserts that the written language in text messages shows the characteristics of informal spoken interaction. As most messages are transmitted between people who know each other, so SMS senders can rely on the receivers' ability to infer meanings even from the grammatically shortened sentences. This ability of message receivers allows senders to omit syntactic elements from their text messages. In this way, SMS texting does not need to be as explicit as is the traditional writing.

6.3 Word Reduplications

In contrast to word ellipses, the phenomenon of word reduplications is opposite in nature. While word ellipses are used for telegraphic purposes, oppositely, word reduplications or triplications are used for emphasis and creative purposes.

Word reduplications are normally used in two or three word phrases, in which exact words or parts of words are repeated. In other words, it is a standard morphological process in linguistics, in which a word is repeated exactly, or with a partial change. It is found in a wide range of languages with the variant frequency in different languages. In the Urdu language, word reduplications are found abundantly. This phenomenon is mostly used in the spoken mode but text messages, being a mixture of written spoken mode, also incorporate word reduplications. Hence, the analysis has been conducted for both English and Urdu in SMS-Corpus.

The term triplication is used for repeating a word three times. If words are repeated more than three times, this process can be termed as word multiplications. As defined by Brinton and Brinton (2010, p. 100), from the technical aspect, there are three categories of reduplications, that is "exact reduplication", "rhyming reduplication", and "ablaut reduplication". The following sections show the category wise trends of word reduplications in text messages.

6.3.1 Exact Reduplications/Triplications

In exact reduplications/triplications, complete/full words are repeated without any change like the baby talk. In the English language, their examples are like "bye, bye", "very, very", and "no, no". In SMS-Corpus, exact reduplications/triplications were found in both Urdu and English text messages, so exact reduplications/triplications have been separately analysed for both the languages. Table 6.6 and Table 6.7 separately present examples for both English and Urdu from SMS-Corpus.

Table 6.6
Examples of Exact Reduplications/Triplications in English

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Normalized SMS
Ex-101[573]	Send me balance Papa *plz plz plzzz*	Send me balance. Papa, *please! please! please!*
Ex-102[1055]	Happy Birthday baji nd *many many* happy returns of the day	Happy birthday baji (elder sister) and *many many* happy returns of the day!
Ex-103[1579]	Happy Birthday! *Many many*	Happy birthday! *Many many* happy

	happy returns of the day!	returns of the day!
Ex-104[1585]	Hahaha! I wana tel u smthn..m vry *hapy hapy* thez dayz.	*Hahaha!* I want to tell you something. I am very *happy happy* these days.
Ex-105[1596]	*Many Many* Happy Returns of the day	*Many many* happy returns of the day!
Ex-106[1764]	Happy b.day sw8 sis may u have *many many* mores.	Happy birthday sweet sister! May you have *many many* more.
Ex-107[2075]	Gud nite n tk cr. C u by by.	Good night and take care. See you, b*ye, bye.*
Ex-108[2099]	I'm *waiting waiting* 4 U	I am *waiting waiting* for you.
Ex-109[2153]	Special Frnds are *very very very* hard to find.	Special friends are *very very very* hard to find.
Ex-110[2234]	God bless U *ever, ever* n *ever*	God bless you *ever, ever,* and *ever.*

Note. Reduplications/triplications represented in Italics

Table 6.6 shows that "please-please", "many-many", "happy-happy", "bye-bye", "waiting-waiting", "very-very", and "ever-ever" are different word examples which have been used for reduplication/triplication. In these examples, the most frequent pair is "many-many" as shown in the table. In SMS-Corpus, the most frequent reduplication is of "many-many" (*seven times*), followed by "plz-plz" (*six times*). Other reduplicated words normally occurred *once or twice*. These figures show the limited number of exact reduplications in text messages. On the other hand, some words like "plz", "very" and "ever" have also been used as word triplications.

The purpose of these reduplications/triplications is to have emphasis in talk. The use of word reduplications and triplications in British English, mostly in the spoken form, is not new. They are used in the spoken mode, and in the informal settings even in the written mode of language. Therefore, in text messages, their use is not astonishing because text messages are informal as well as speech based mode of communication.

Table 6.7
Examples of Exact Reduplications/Triplications in Urdu

Ex-No[SMS-No]	Actual SMS (Urdu)	Transliteration (IPA)	Translation (English)
Ex-111[06]	Uth joo app or ab parh lu...*plz plz*	ʊtʰ ʤɑ:ɔ ɑ:p ɔr əb pəɽh lo ... *pli:z! pli:z!*	You get up and study now ... please! Please!
Ex-112[94]	apna id send kro aur btao *kn kn* se manual send krne ha	əpnɑ: ɑ:i: di: send kəro ɔr bəʈɑ:ɔ *kɔn kɔn* se mənju:al send kərne hɛ̃.	Send your ID and tell which manuals are to be sent to you.

Ex-113[212]	*Pagal! Pagal! Pagal!*	*pa:gəl! pa:gəl! pa:gəl!*	Mad! Mad! Mad!
Ex-114[263]	yar aj ks k ski class ha	jɑ:r ɑ:d͡ʒ *kɪs kɪs* ki: klɑ:s hɛ?	Dear whose class is today?
Ex-115[329]	*Acha Acha* main samngh gya.	*ət͡ʃʰa: ət͡ʃʰa:* mɛ̃ səməd͡ʒʱ gɪjɑ:	Ok! Ok! I understood.
Ex-116[575]	Roti banaooo! Bhuk lge hay *Jaldi Jaldi*!	roti: bənɑ:o! bʰu:k ləgi: hɛ. *d͡ʒəldi: d͡ʒəldi:!*	Prepare bread! I am hungry. Hurry up! Hurry up!
Ex-117[579]	*Jago Jago* subah hoe.	*d͡ʒa:go d͡ʒa:go* sʊbhɑ: hɔi:.	Get up! Get up! It is morning.
Ex-118[735]	*AAJA AAJA* RAY.	*a:d͡ʒa: a:d͡ʒa:* re.	You come, come.
Ex-119[918]	ji ji ji wesay hi kesay hi?	*d͡ʒi: d͡ʒi: d͡ʒi:* vɛse hi: kɛse hi:?	Yes, yes, yes. Without any reason, how without any reason?
Ex-120[4895]	Meri taraf se aap ko or ap ke ghar walo ko *Achi achhi sweet sweet* Eid Mubarak ho.	meri: ɻərəf se ɑ:p ko ɔr ɑ:p ke gʱər vɑ:lõ ko *ət͡ʃi: ət͡ʃi: swi:t swi:t* i:ɖ mʊbɑ:rək ho.	I wish you and your family the nice, nice, sweet, sweet Eid Greetings.

Note. Reduplications/triplications represented in Italics (see transliteration)

Table 6.7 shows word reduplications/triplications in the Urdu language. In the spoken mode, word reduplications/triplications are more common in Urdu than they are in English, but in SMS-Corpus their frequency was not found more than maximum *three or four times* for each word given in the examples.

In Urdu, words of English are also used as reduplicated words like "plz-plz", and "sweet-sweet" in the above examples. The impact of English is so pervasive on the lexicon of Urdu that many English words are used in Urdu both in the formal as well as in the informal settings. There are numerous borrowed/loan words in Urdu that have been adopted from English, and are frequently used in the informal spoken Urdu. Hence, the borrowed words from English into Urdu are seen in text messages too, as text messages are informal as well as speech based mode of communication.

6.3.2 Rhyming Reduplications (Consonant Changing)

Rhyming reduplication is the process of repeating a word with the replacement of a consonant sound. Normally, the replaced consonant is the first consonant of the word. The words like "walkie-talkie", "super-duper" and"hanky-panky" are the typical

examples of rhyming reduplication in British English. Rhyming reduplications are frequently used in the Urdu language in the informal spoken mode. Some examples have also been found in SMS-Corpus from both English and Urdu languages as presented in the following table.

Table 6.8
Examples of Rhyming Reduplications

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Transliteration (IPA) for Urdu SMS	Normalized/Translated SMS in English
Ex-121[546]	Jab apny waladain ko koi chez paish kro to *aisy jaisy* ghulam apne Aqa k samne hath phailata ha	d͡ʒəb əpne vɑ:lɖɛn ko koi: t͡ʃi:z peʃ kro ʈo *ese d͡ʒese* ɣʊlɑ:m əpne ɑ:qɑ: ke sɑ:mne hɑ:ʈʰ pʰelɑ:ʈɑ: hɛ	When you offer anything to your parents, do as a servant begs to his master.
Ex-122[918]	ji ji ji *wesay hi kesay hi?*	d͡ʒi: d͡ʒi: d͡ʒi: *vɛse hi: kɛse hi:*?	Yes, yes, yes. Without any reason, how without any reason?
Ex-123[2023]	me *sooperdooper*	.....	I am *sooper dooper*.
Ex-124[2171]	If you chouse Allah as your Instructor pilot your plane will *never ever crash.*	.....	If you choose Allah as your instructor pilot, your plane will *never ever* crash.
Ex-125[3239]	Zindgi me 2 logo, ka boht khyaal rakhna 1 wo jino ne tumari JEET k liey sub kuch hara ho. (Bap) 2 wo jino ne her *dukh sukh* me sirf tumien pukara ho. (Maa)	zɪnɖəgi: mẽ do logõ, kɑ: bhoʈ xɪɑ:l rəkʰnɑ: 1) vo d͡ʒɪno ne ʈumhɑ:ri: d͡ʒi:ʈ ke li:e səb kʊt͡ʃʰ hɑ:rɑ: ho. (bɑ:p) 2) vo d͡ʒɪno ne hər *ɖʊkʰ sʊkʰ mẽ* sɪrf ʈumhẽ pʊkɑ:rɑ: ho. (mɑ̃:)	Take much care of two persons, 1) Who lost everything for your victory (Father), 2) Who called for you only, in every joy and sorrow (Mother).

Note. Reduplications represented in Italics

Table 6.8 shows examples of rhyming reduplication in both English and Urdu in SMS-Corpus. In English text messages, *two pairs* "sooper-dooper" and "never-ever" have been used. Both the reduplications have been used by replacing the first consonant of the word. For Urdu, *three pairs* "aisy-jaisy", "wesay hi-kesay hi", and "dukh-sukh" have been found. All three pairs in Urdu have been reduplicated by replacing the first consonant sound. Hence, the phenomenon of rhyming reduplication in SMS-Corpus occurred by replacing the first consonant sound in both languages.

6.3.3 Ablaut Reduplications (Vowel Changing)

Ablaut reduplication is the process of repeating a word with the replacement of a vowel sound. Normally, the replaced vowel is somewhere in the middle of the word. The common examples of ablaut reduplication in English are "chit-chat", "ding-dong", and "zig-zag" etc. In Urdu, the common examples are like "theek-thaak", "fit-faat", and "shopping-shupping" etc.

Table 6.9
Examples of Ablaut Reduplications

Ex-No[SMS-No]	Actual SMS (Urdu)	Transliteration (IPA)	Translation (English)
Ex-126[1269]	the*ek thalk* aap sunain pa munir	*tʰi:k tʰa:k* ɑ:p sʊnɑ:ẽ pɑ: mʊni:r	I am fine. You tell, brother Munir.
Ex-127[1273]	*Fit Fat*	*fit fa:t*	I am extremely fine.
Ex-128[2417]	Hi. *Thek thak*.... Bs baithi hui thi...	hɑ:ɪ. *tʰi:k tʰa:k*....bəs bet^{h}i: hʊi: ʈ^{h}i:....	Hi! I am fine. I was just resting.

Note. Reduplications represented in Italics

Table 6.9 shows three examples of ablaut reduplications. Each was found only *single time* in SMS-Corpus. All three are from the Urdu language, as no example could be found for the English language in SMS-Corpus. Among all three categories of reduplications, that is exact, rhyming, and ablaut, this is the least occurring category in SMS-Corpus for both languages.

6.3.4 Discussion on Word Reduplications/Triplications

Among all three categories of reduplications/triplications, exact reduplication is most common both in Urdu and English in SMS-Corpus. As stated in the opening paragraph of the chapter, exact frequencies for word reduplications have not been given in tables, instead examples of sentences from SMS-Corpus have been tabulated. Still, their frequencies have been mentioned in the follow-up discussion after tables.

Among the three given categories of reduplications, exact reduplication is the only category that has also been used in *triplication* in SMS-Corpus; the other two categories were only used in reduplication. Moreover, exact reduplication is also known

as "true" reduplication, as both the words are meaningful words. The other two categories are known as *quasi* reduplications, because the second word, in most of the cases, does not have its own meaning, and is used to mean "etcetera".

The purpose of exact reduplication/triplication in text messages is to ensure emphasis in talk. Hence, reduplicated words are used for emphasis and stress. In other words, it is a kind of foregrounding in text messages. Although, to employ reduplications, texters have to press more keys, yet they use them for style, fun, and creative purposes. Hence, the rule of economy, that is saving time and effort, is over ruled in the case of reduplication of words.

After exact reduplication, the rhyming reduplication is the second frequent category of reduplications in SMS-Corpus. In rhyming reduplication, the first word usually gives the actual meaning, while the repeated word is used to broaden the meaning of the first word and means "etcetera". Rhyming reduplications are formed by replacing the first consonant of the repeated word. In English, the word pairs like "walkie-talkie", "super-duper", "hanky-panky", "razzle-dazzle" etc. are the classic examples of rhyming reduplications.

In Urdu, "Sh-reduplication" and "V-reduplication" are very common formations of rhyming reduplications. The typical examples are "khana-shana/khana-vana" (meal etcetera), "roti-shoti/roti-voti" (bread/meal etcetera), "chai-shai/chai-vai" (tea etcetera), and "thanda-shanda/thanda-vanda" (cold drink etcetera).

The third category of word reduplications is ablaut reduplication. This is a less common category of the three in SMS-Corpus. In this type, the vowel sound of the second word is replaced at the middle of the word. The typical examples of ablaut reduplications in English are like "chit-chat", "criss-cross", "ding-dong", "ping-pong",

and "zig-zag" etcetera. In Urdu the examples are like "shopping-shupping" (shopping etcetera), "theek-thaak" (completely fine), and "fit-faat" (completely healthy) etcetera.

The process of reduplication is indeed not a new phenomenon in languages. In British English, the reduplicated words like 'riff-raff" and "willy-nilly" existed even centuries ago. Moreover, reduplications of words are found in most of the languages of the world. Afrikaans "krap-krap-krap" (to scratch forcefully), Turkish "tabak-mabak" (plates/dishes etcetera), Napalese "khana-sana" (meal etcetera), and Persian "dava-mava" (argument etcetera) are a few examples of reduplications in some other languages. Therefore, the phenomenon of reduplication is neither new, nor specific to a certain language, or a specific register of a language. It is a common phenomenon of most of the languages of the world, in the informal and spoken settings.

This phenomenon of reduplication/triplication is common in text messages as this written mode of communication is actually more tilted towards speech than writing. Moreover, SMS is an informal mode of communication, so it accommodates all kinds of reduplications.

6.4 Grammar Deviations

There are various approaches of describing the grammar of a language. According to one approach, the grammar of a language is "a set of *rules* which specify all the possible grammatical structures of the language. In this approach, a clear distinction is usually made between grammatical (sometimes called *well-formed*) sentences and ungrammatical sentences" (Lock, 1996, p. 1). Greenbaum and Nelson (2002, p. 1) also describe grammar as a "set of rules that allow us to combine words in our language into larger units." They term grammar in this sense as *syntax* too.

Regarding the grammar of English, Greenbaum and Nelson (2002, p. 1) state that there are many grammars of English owing to its many national and register specific

varieties. Furthermore, even a single variety may have more than one grammar based on various grammatical approaches it follows, for example the *formal* and *functional grammars*, and the *prescriptive* and *descriptive grammars*. These grammars differ in the ways in which they set out rules for a language variety. Whatever rules a grammar approach supports, it is evident that certain underlying rules specify the very structure of a language. In this regard, Greenbaum and Nelson state:

> Some combinations of words are possible in English and others are not. As a speaker of English, you can judge that *Home computers are now much cheaper* is a possible English sentence whereas *Home computers now much are cheaper* is not, because you know that *much* is wrongly positioned in the second example. Your ability to recognize such distinctions is evidence that in some sense you know the rules of grammar even if you have never studied any grammar. (p. 1)

Hence, we operate rules whenever we speak or write a language, and the compliance of a sentence to these rules indicates that a sentence is grammatical. In this sense, the language used in text messages may have its own specific rules of grammar, but at present there exist no standard defined rules of grammar for SMS language. Therefore, SMS language analysts normally compare the grammar adaptations made in text messages with some standard variety of the English language. In this regard, the present study analyses the grammar deviations made in text messages in Pakistan on the rules of the *formal grammar* of standard *British English*, as it is the variety of language that is mostly followed in Pakistan.

The analysis concerns three most observed areas of grammar adaptations in text messages, which are *deviant tense, word class/form,* and *word order* in sentences. Some other deviations like orthographical deviations (spellings or punctuations) have not been analysed in this section because they do not conventionally fall under the

grammatical/syntactic category. Moreover, the deviations of spellings and punctuations have been analysed in chapter 5 and 7 respectively.

6.4.1 Deviant Tense

In the formal grammar of standard British English, the deviations of tenses are considered very seriously. With the deviation of tense, the whole sense is distorted, so the deviations of tenses/verbs are the most censured deviations by the critics of SMS language. In the following table, a few instances of such deviations from SMS-Corpus have been presented to analyse grammar.

Table 6.10
Examples of Deviant Tense (Verbs)

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex-129[606]	I have *send* u the C.V. via corier.	I have *sent* the C.V. via courier.
Ex-130[634]	*whts* u doing nw a days,	What *are* you doing nowadays?
Ex-131[751]	*I come* to play today	I *will* come to play today.
Ex-132[1380]	What's going on & what work *had* u dne.	What is going on and what work *have* you done?
Ex-133[1406]	You girls are like trains one goes and after a while another *come.*	You girls are like trains, one goes and after a while another *comes.*
Ex-134[1459]	my best wishes *is* always with u.	My best wishes *are* always with you.
Ex-135[1601]	Everything is pre-written, But, with prayers it can be *change*	Everything is pre-written but with prayers it can be *changed.*
Ex-136[2091]	I'm *wa8* 4 u plz come soon	I am *waiting* for you. Please come soon.
Ex-137[2810]	Hv you *went* back to lahore	Have you *gone* back to Lahore?
Ex-138[4897]	He is very cheap, I did not *knew*	He is very cheap; I did not *know.*

Note. Deviations represented in Italics

In the examples given in Table 6.10, there are two types of adaptations in the use of tense, that is the deviations in the standard uses of verbs, and their omissions where needed. In one example there is an instance, in example-131, of the omission of an auxiliary verb. In all other examples, main verbs or auxiliary verbs have been used in the deviant way. Overall, with these deviations or omissions of verbs, meanings of sentences have been changed or distorted, and in formal grammar of the English language such deviations of verbs/tenses do not fit on the conventional standards. Therefore, formal

grammarians of the English language term such sentences as ungrammatical, although they may be accepted in the register specific language variety of text messages.

6.4.2 Deviant Word Class/Form

In text messages, it has been observed that texters do not care about the deviant use of word class/form. Texters replace the conventional words with deviant words, either intentionally or unintentionally. Such deviations of words lead to distorted meanings and are considered ungrammatical in the formal grammar of the English language. Examples of such deviant word forms have been given in the next table from SMS-Corpus.

Table 6.11
Examples of Deviant Word Class/Form

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex-139[32]	Hello, its NAME …	Hello, it's NAME …
Ex-140[78]	*me* cuming to uni.	*I am* coming to the university
Ex-141[266]	*Me* F9 and w8ing for *electric.*	*I am* fine and waiting for the *electricity.*
Ex-142[267]	*Me* going to uni.	*I am* going to the university.
Ex-143[522]	*Me* back	*I am* back.
Ex-144[622]	I *me* cming w8 for 10 mintes.	I *am* coming. Wait for ten minutes.
Ex-145[649]	I *real* Misssss Yoouuu	I *really* miss you.
Ex-146[983]	class will be *at* Monday *on* 11:30	The class will be *on* Monday *at* 11:30.
Ex-147[1020]	I like *u* joke bt u r nt gud by	I like *your* joke but you are not good. Bye.
Ex-148[1424]	*whose* there?	*Who is* there?

Note. Deviations represented in Italics

Table 6.11 shows that various words have been unconventionally used where they do not fit grammatically. Mostly, the first person subjective pronoun "I" has been replaced with the first person objective pronoun "me". Although, people frequently use "me" instead of "I" in the informal speech, yet this use is considered ungrammatical in formal grammar, and is admonished by the formal grammarians. These uses of deviant word forms are either due to the carless attitude towards standards, or the lack of command of texters on the rules of formal English. Whatever the reason, these

unconventional deviations in word forms are termed as ungrammatical by formal grammarians.

6.4.3 Deviant Word Order

Another category of ungrammatical nature in text messages is the wrong order of words in sentences. There is a standard sentence structure in every formal language that defines the order of words in a sentence to make a sense in that language. Yet, this aspect of word order in text messages is also compromised either intentionally or unintentionally, and examples of such uses have been presented in the following table.

Table 6.12
Examples of Deviant Word Order

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex-149[13]	ven *u vill* cum 2 univ.	When *will you* come to the university?
Ex-150[747]	where *you are*	Where *are you?*
Ex-151[1120]	wn *u wl* cm	When *will you* come?
Ex-152[1387]	when *you will* be home???	When *will you* be at home?
Ex-153[1399]	when *you are* cominh home??	When *are you* coming home?
Ex-154[1454]	I need some balance Now plz give me *30. Rs*	I need some balance now. Please give me *Rs. 30.*
Ex-155[1760]	Whn *v will* rch da Hotel?	When *will we* reach the hostel?

Note. Deviations represented in Italics

In Table 6.12, most of the examples are of deviations of the word order between the subject and auxiliary verb in the interrogative constructions. It is a common principle of the English language, that in the positive construction of a sentence, an auxiliary verb is used after the subject while in the interrogative constructions auxiliary verbs precede the subjects. Contrary to this principle, some texters do not change this word order in the interrogative construction and hence deviate from the rules of formal language.

6.4.4 Discussion on Grammar Deviations

In the study, three types of grammar deviations in the English language, used in text messages, were analysed. These are the deviations of tense, word class/form, and word order in the sentence structure of text messages.

David Crystal (2008a) states that in linguistics, the conformity of a sentence to the rules defined by a specific grammar of a language indicates that the sentence is grammatical. An asterisk is normally used with an ungrammatical sentence that shows that the sentence is incapable of being accounted for by the rules of a grammar. But in practice, the decision whether a sentence is grammatical or ungrammatical may cause difficulty. The evident reason for this difficulty is the multiple approaches of defining the grammar of a language.

In this scenario, the present analysis has been made by comparing the grammar deviations made in text messages with the rules of *formal grammar* of standard *British English* as it is the variety of language that is mostly followed in Pakistan. The *formal* approach of grammar has been chosen as, in contrary to functional grammar, it emphasizes the rules of a language instead of its communicative functions.

Based on the perspective of formal grammar of standard British English, the examples given in the three categories of grammar deviations in text messages are ungrammatical. Most of the deviations discussed in the grammatical category actually occur due to the lack of texters' command on the rules of the English language, besides their carelessness about the language concern. Hence, the deviations of grammar do not normally occur due to creative reasons, or for the sake of the economical use of language, as is the case with some other deviations. For these reasons, grammatical deviations are the serious concern of the critics of SMS language.

A notable point about the deviations of grammar in text messages is that they do not occur in as high frequency as is the case with the spelling deviations. Therefore, although they are a very serious concern for the critics of SMS language, yet it is due to their low frequency that they are not as noticeable as the spelling deviations. Moreover, these register specific deviations of grammar in text messages are also found in the

computer mediated communication on the Internet, and we cannot only blame text messages for such deviations. Bodomo (2010) states that deviations of grammar are also found in the online communication, and hence they are a common feature of CMC and SMS language.

CHAPTER 7

PUNCTUATION AND SPACE ADAPTATIONS

Punctuations and spaces are two very closely related adaptations in text messages. The chapter investigates them through examples from SMS-Corpus. After a brief introduction to the both types in the first section, punctuations have been discussed in the next two sections, whereas adaptations of spaces have been analysed in the last section. Punctuations have been discussed in two sections as their adaptations in text messages are more than the adaptations of spaces. Both of them have been analysed for their deviant uses in text messages, whereas punctuations have additionally been analysed for their paralinguistic uses in text messages.

7.1 Introduction to Punctuation and Space Adaptations

In many dictionary definitions, punctuations are defined as the graphic marks/non-alphabetic signs that clarify meaning in written sentences or represent spoken sentences in writing. These symbols clarify both grammatical and semantic aspects of the text (Bussmann, 1996; Richards & Schmidt, 2002). They serve to divide a piece of writing into sentences, phrases, and words, etcetera. Punctuations include a wide range of marks like commas, colons, semicolons, dashes, periods, question marks, exclamations, and ellipses.

In the light of the rules of the formal language, punctuation marks are rarely used in text messages in conventional ways. There are many adaptations in the use of punctuations in text messages. Therefore, the unconventional uses of punctuations in text messages are one of major issues in the field (Bodomo, 2010; Bosco, 2007; Hard af Segerstad, 2005). In text messages, punctuation marks are used in many different ways from the conventional standards of punctuations. They may be omitted or excessively

used. Texters also manipulate them for creativity and innovations in text messages. They use punctuation marks for many paralinguistic and artistic purposes. Consequently, the creation of various types of emoticons (emotion icons) is a common practice among texters over the world.

Many studies on the unconventional uses of punctuation marks have also been conducted in text messages. In this regard, Baron (2008) investigated different aspects of sentence punctuations in texting and instant messaging. In her study, she found that seventy-one percent of text messages had no punctuation mark at the end of the text message. In the same way, she gave examples of sentences where conventional punctuation marks were replaced with deviated punctuation marks.

Bosco (2007) conducted a study on the language of text messages in Hong Kong, and found that the use of punctuations in text messaging had a high degree of creativity. She discussed the significant differences between the uses of punctuations in SMS versus the traditional written language in three categories, i.e. 1) excessive use of ellipsis, 2) excessive use of exclamations, and 3) omission of periods and apostrophes.

Regarding the second concern of the chapter, that is space adaptations, they are defined here. In their very nature, both *spaces* and *punctuations* are used to divide a piece of writing into various parts. In this regard, punctuations are the graphic marks which divide a piece of writing into sentences, phrases, and words; whereas spaces are the blank areas to separate a written script into sentences, phrases, and words (Palmer, 1983; Richards & Schmidt, 2002). Since both perform similar functions in language, therefore both have been grouped together in this chapter.

7.2 Deviant Uses of Punctuations

There are various deviant uses of punctuations in text messages. These deviations can be classified in different categories. In this study, the unconventional uses of

punctuation marks have been classified into three broader categories. These are 1) omissions of punctuations, 2) repetitions of punctuations, and 3) substitutions of punctuations. The analysis of each category is offered through examples from the corpus in the following sections.

7.2.1 Omissions of Punctuations

In text messages, it is a common observation that punctuation marks are omitted where they are conventionally deemed necessary. There are various examples of text messages in SMS-Corpus where no punctuation mark was used at all. To explore various patterns of omissions of punctuation marks in text messages, a few examples from SMS-corpus have been given in the following table.

Table 7.1
Examples of Omissions of Punctuations

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex- 156[17]	M w8ing	I am waiting.
Ex- 157[23]	I ll be there after some time	I'll be there after some time.
Ex- 158[141]	I miss you	I miss you!
Ex- 159[155]	hello! How r you	Hello! How are you?
Ex- 160[396]	Im out now. Comming or not	I'm out now. Are you Coming or not?
Ex- 161[650]	Ok bye tc gud ni8	Okay! Bye! Take care. Good night.
Ex- 162[776]	Are u fre 2day	Are you free today?
Ex- 163[778]	OK. I ll be w8ng 4 u	Ok. I'll be waiting for you.
Ex- 164[819]	I m going take my register with u	I am going. Take my register with you.
Ex- 165[2809]	Where r u inform me wen u r free	Where are you? Inform me. When are you free?

Table 7.1 shows that in the given examples, punctuations have been very rarely used. In the examples, four very crucial punctuation marks that are period/full stop (.), question mark (?), exclamation mark (!), and apostrophe (') have been frequently omitted. Hence, it is a common practice in text messages that punctuation marks are omitted where they are conventionally necessary. Yet, there are reasons for omitting punctuations marks in text messages.

In text messages, punctuations are mainly omitted because their absence does not cause the loss of any information in short messages. So texters consider it the loss of time and effort to use punctuations marks. For texters understanding is more important than any language standards. Moreover, the medium of SMS allows them for this liberty because most of the time text messages are used for informal communication. Hence, texters omit punctuation marks for ease of typing. By omitting punctuation marks they save time and effort, and hence concentrate only on the communicative function of the message.

7.2.2 Repetitions of Punctuations

To the one hand, texters omit punctuations to save time and effort, but at the same time texters use punctuation marks repetitively in some text messages. To study this phenomenon of repetitions of punctuations in text messages, examples from SMS-Corpus have been presented in the following table.

Table 7.2
Examples of Repetitions of Punctuations

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex- 166[27]	okay…	Okay.
Ex- 167[56]	Hi, Hw r u? what happened to u lastnight???	Hi, how are you? What happened to you last night?
Ex-168 [146]	Hey! I saw you in uni.. can we talk??	Hey! I saw you in the university. Can we talk?
Ex- 169[233]	Gud Luck!!...	Good luck!
Ex- 170[478]	Hi, hw r u? wats up???	Hi, how are you? What's up?
Ex- 171[1387]	when you will be home???	When will you be at home?
Ex- 172[1547]	I Missss……u!	I miss you!
Ex- 173[1563]	wt???	What?
Ex- 174[1588]	Party without me….noway!!!	Party without me, no way!
Ex- 175[2211]	Hey…. How r u!!!!????	Hey! How are you?

Table 7.2 shows examples where same punctuation marks have been repetitively used by deviating from the rules of formal language. There are examples of the repetitive uses of dots (.), questions marks (?), and exclamation marks (!). All the said punctuation marks have been excessively used in these text messages than is normally required.

Here, the principles of saving time and effort have been overruled, as examples do not show the economical use of punctuations. For the repetitive uses of punctuations, texters have to press keys of mobile phones many times, and this practice takes more time and effort. Therefore, the repetitive use of punctuation marks is intentionally employed by texters to convey their messages emphatically and powerfully.

The repetitive use of same punctuation marks is so common in text messages that in SMS-Corpus numerous instances of such uses were found. In this regard, the frequency list of their use in SMS-Corpus is given as follows.

Table 7.3
10 Most Frequent Repetitions of Punctuations (AntConc Analysis)

Punctuation Marks	Frequency	Target Expression
..	204	extra emphasis, unfinished
...	101	extra emphasis, unfinished
! !	53	extra emphasis, much surprised
? ?	48	extra emphasis on question
! ! !	40	extra emphasis, much surprised
....	30	extra emphasis, unfinished
? ? ?	28	extra emphasis on question
,,	17	extra pause
" " "	14	creative/stylish use
.....	11	extra emphasis, unfinished

Note. Punctuation token class activated in AntConc

Table 7.3 shows that same punctuations have been repetitively used, and the purpose of such repetitions is to convey some extra meanings. The most frequent repetitive uses, as noted in Table 7.2 too, were of three punctuation marks, i.e. dots (.), questions marks (?), and exclamation marks (!).

7.2.3 Substitutions of Punctuations

In addition to the omissions and repetitions of punctuation marks, another category of deviant uses of punctuation marks in text messages is the substitution or replacement of conventional punctuation marks with deviant/unconventional punctuation

marks. The analysis of this category is offered in the following table, through the examples from SMS-Corpus.

Table 7.4
Examples of Substitutions of Punctuations

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex-176[13]	ven u vill cum 2 univ.	When will you come to the university?
Ex-177[27]	wht r u doing…	What are you doing?
Ex-178[74]	I am w8ing for u?	I am waiting for you.
Ex-179[75]	I, m f9.	I'm fine.
Ex-180[99]	Where shall v meet….	Where shall we meet?
Ex-181[161]	Hi – dude h.r.u	Hi! Dude, how are you?
Ex-182[220]	we all lov u?	We all love you.
Ex-183[221]	Go to hell?	Go to hell.
Ex-184[228]	R u sure its egle.	Are you sure that it's eagle?
Ex-185[933]	H r u.	How are you?

Table 7.4 shows 10 examples of the replacement of the appropriate punctuations with the deviant/inappropriate punctuation marks. In the examples, question marks have been replaced with periods (examples 176, 177, 180, 184,185), periods with the questions marks (examples 178, 182, 183), an apostrophe with a comma (example 179), and an exclamation mark with a dash (example 181). Such deviations neither fit in the category of economical uses, nor in the creative uses of text messages. Therefore these are the deviations, which occur, either due to the lack of knowledge of the rules of punctuations, or due to the careless attitude of texters.

7.2.4 Discussion on Deviant Punctuations

Punctuations are the graphic marks in writing to clarify meaning, and to separate words into sentences, clauses, and phrases (Richards & Schmidt, 2002). They are the signals to readers to clarify grammatical and semantic aspects of the text (Bussmann, 1996). When people speak, they can pause, stop, or change their tone of voice to make meanings clear and understandable. People cannot do this in writing without using punctuation marks. Hence, punctuation marks are the signs that indicate the structure of written language. But when we study the written language of text messages, here, the

conventional standards of punctuations are not found. Therefore, this section addresses the deviations of punctuations made in text messages.

The section deals with three types of deviant uses of punctuations in text messages. These are 1) omissions of punctuations, 2) repetitions of punctuations, and 3) substitutions of punctuations. The analysis of each category has been made through examples from SMS-corpus, and the deviations of each category can be justified.

The deviations of "omissions of punctuations" occur due to the principle of economy. Texters normally omit punctuation marks to save time and effort, and hence they use punctuations economically. Hard af Segerstad (2005, p. 41) states that "by omitting punctuation a user saves the time and effort it takes to type those characters (period, comma, etc)". She further states that omitting the punctuation marks saves keystrokes, which could be important when the message size is restricted by technical limitations, and therefore the strategy of omitting punctuations is used for saving time and effort.

The deviations in the second category, which is "repetitions of punctuations", occur due to the creative reasons. Texters use excessive punctuations to emphasize their point, or to create something novice in their text messages. In this process, they have to press keys many times, yet they accept this ordeal for the creative touch to their text messages.

Bodomo (2010) states that same punctuation marks are repetitively used to emphasize something more powerfully, and to stress on the same thing. Repetitive use of periods/dots (.) is employed where texters have left something unsaid (ellipsis). They are sometimes used to emphasize something that is meant but not said. Exclamation marks (!) are used to emphasize, and show the interest and curiosity of the texters. Repetitive

question marks (?) are used to emphasize the question or enquiry. Commas (,) are repetitively used to show some extra pause.

The third category of deviations, which is "substitutions of punctuations", occurs due to the careless attitude of texters. These unusual deviations in the third category cannot be attributed to any recognized or justifiable reasons except the carelessness of texters. Sometimes these deviations occur due to the lack of linguistic knowledge of texters. Whatever are the reasons, the deviations in this category are mostly reproached by the advocates of standards in language.

7.3 Creative Uses of Punctuations

In a formal language, punctuation/graphic marks are conventionally used for linguistic purposes. Their linguistic purpose is to clarify meaning in written sentences or to represent spoken sentences in writing. They clarify the grammatical and semantic aspects of written sentences, and separate them into clauses, phrases, and words (Bussmann, 1996; Richards & Schmidt, 2002). But, in text messages, punctuation marks can be used for both linguistic and creative purposes.

Crystal (2008b) differentiates between the uses of graphic marks for linguistic and paralinguistic purposes. He states that graphic symbols are used in two ways, i.e. as *pictographs* and *logographs*. When graphic symbols are used as pictographs, their *visual shape* conveys emotions; whereas in logographs, their *pronunciation* conveys words. In this regard, this section deals with the pictographic uses of punctuation marks in text messages where their visual shape is used for creative purposes.

Here, it is also highlighted that punctuation marks, and some other graphic symbols, are used in text messages by pressing the "*" button/key on most feature phones. Although, for the creative purposes, texters have to press keys repetitively, yet they use the excessive punctuation marks for the paralinguistic purposes like emoticons,

and many artistic creations. Therefore, the principles of economy and speed are overruled by texters when they have concern for creative uses of text messages.

In this study, for the analysis of the creative uses of punctuations, their use has been divided into two categories. They have been analysed in two parts, i.e. 1) punctuations as paralinguistic devices, and 2) punctuations as artistic devices. The analysis of both categories is given below.

7.3.1 Punctuations as Paralinguistic Devices/Emoticons

Paralinguistics is a subfield of linguistics that deals with the nonverbal aspects of communication in any human language, such as body language and wordless expressions. To convey the nonverbal messages in a language, various paralinguistic devices can be used like smiling, gestures, facial expressions, laughing, breathing, murmuring, whispering, crying, coughing, and even tone and pitch of voice (Bussmann, 1996).

In a spoken language, paralinguistic devices are those methods that can be used to convey feelings and emotions through expressions, and not through words. Paralanguage is used to convey emotions that may not be conveyed through words/verbal expressions. Paralinguistic devices play an important role in human communication. Theoretically, there are no utterances in human speech which are without paralinguistic devices because speech requires the presence of such devices as additional aids.

In a written language, such emotions are conveyed through a special kind of expression markers known as *emoticons* (emotion-icons). In the actual sense, emoticons are emotion markers like (☺ for happy) or (☹ for sad), but punctuation marks can also be used to manually develop them. When such pictures are used as facial expressions to convey emotions, they are historically termed as *Pictograms* or *pictographs.*

About the first use of emoticons, Naomi S. Baron (n.d.)[4] states that on September 19, 1982, Scott E. Fahlman sent the first message carrying two emoticons to an online bulletin board at Carnegie Mellon University. He used a smiley " :-) " and a frowny " :-(" , to convey the happy and sad emotions respectively. Baron also traces the earlier uses of emoticons in 1963, when a graphic artist Harvey R. Ball created a yellow button based smiling face for the State Mutual Life Assurance Company of America. Regarding their use in text messages, the following table presents the use of punctuation marks as emoticons in SMS-Corpus.

Table 7.5
10 Most Frequent Punctuations as Emoticons (AntConc Analysis)

Punctuation Marks	Frequency	Target Emotion
:-) /☺	48	happy, emoticon of happiness (smiley)
:)	30	happy, emoticon of happiness
;-)	25	happy, blinking eye
. !	21	surprised
. . !	21	much surprised
:-D	18	happy, grinning
;)	14	happy, winking
:-P	12	happy, tongue out
:-(/ ☹	9	unhappy, emoticon of sorrow (frowny)
('.')	8	normal face, neither happy nor sad

Note. Punctuation token class activated in AntConc

The table shows the most frequent punctuations used as paralinguistic devices in SMS-Corpus. The table shows the most frequent instances where two or more than two punctuation marks have been collectively used to create emoticons. With each instance of punctuation marks as emoticons, frequencies as well as respective emotions have been given in the table.

The important point is that in face to face communication emotions are conveyed through facial expressions or physical gestures, whereas in text messages, punctuation

[4] http://www.american.edu/cas/lfs/faculty-docs/upload/Baron_Emoticons-1-the-Myth-of.pdf

marks are used for paralinguistic meanings. Here, the use of punctuation marks as emoticons is further examined through the examples from SMS-Corpus.

Table 7.6

Examples of Punctuations as Emoticons

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Translation/Normalization
Ex-186[753]	Tarasta hai Dil Teri awaz k liye....	The heart is restless for your voice,
	Teray Mohabat bhare alfaz k liye...	For your love-filled words,
	Hum o9r kuch nahi mangte tum se bas	I do not ask for anything more just
	(..)	.
	</I/>	.
	!!	.
	Insan ban jao Khuda k liye:-)	Be a human for God's sake :-)
Ex-187[2337]	Jahapanah.....	My Lord!
	Tusi Great HOO	You are great
	Tohfa Qabool Karo	Accept the gift
	. .	.
	[:.]	.
	,-/\\	.
	}}``poo!	Poo!
	-?-?	.
Ex-188[2513]	Tanoo ik gal dassan?	May I tell you one thing?
	('_')	.
	Edhr koi hega te nai.?	Does anyone see here?
	(': ')	.
	Nai..?	No?
	odhr?	There?
	(':')	.
	Nai.?	No?
	Lo fer suno,	Then listen,
	*	.
	*	.
	*	.
	"Tussi bary yad aa raye o" :-(	I miss you a lot :-(
Ex-189[2579]	,*,('.')<)(> I Miss !! U	I miss you!
	,*,('.') !)(? Do u miss !! me?	Do you miss me?
	,*,(-_-) No..?	No?
	!)(> ¡/ Allah Poochega Tum sey ?	God will make you answer.
Ex-190[2581]	('.')>	.
	<\"_/"_	.
	<('.') /"/>	.
	_/"_ ('.')>	.
	<\"_/"_	.
	<('.') /"/>	.
	_/"\	.
	Hat jaaaaoo aj Faraz ko "ROZAY" se chakker aarahy hain ...:-	Be away, today Faraz feels giddy for observing fast.
Ex-191[2610]	Gud morng	Good morning

	.*...*. ,, ,,, ,, *("._.")* (") (")„„(")(")„„ I Pray U be Safe U be Happy Every Minute Every Hour Everyday of ur life; GOD BLESS U!Amen!	. I pray You to be safe You to be happy Every minute Every hour Every day of your life God bless you! Amen!
Ex-192[2849]	.-""-. (';') („) („) Hell0 G gal Suno... . . . . Suno v Kan Edhar Kr0.. .-""-. (")('; ') '(„)(„) Main Kya Ji "gud morning	. . Hello Dear! Listen To Me, ... Please listen Bring the ear to this side, . . Dear, I say, "good morning".
Ex-193[3070]	SALAM G('.') (',')D M(';')Rn!nG,	Peace be upon you, GOOD Morning!

Table 7.6 shows different examples of text messages from SMS-Corpus. The examples show that punctuation marks have been used as emoticons/paralinguistic devices. To highlight the emotions attached with each SMS, a brief description is also given here.

In example-186, the body of a man has been created with punctuations marks. In example -187, a man has been created who is releasing his stool. In example-188, a face has been shown in three settings/moves, which are "looking straight", "looking right", and "looking left" at three different remarks. In example-189, a face has been shown in four moves, that are "looking left" (first and second move), "looking starlight" (third move), and "distorted face" (fourth move). In example-190, the head of a famous poet of Pakistan, named Faraz, has been shown pounding with the giddiness for observing fast. In example-191, a man has been shown praying with the raised hands. In example-192, a

man has turned his both ears to the left side to listen to the texter. Example-193 contains three facial emoticons to wish good morning.

On the whole, in all the given examples, emoticons have been created with the help of punctuation marks. Moreover, all the given text messages have been created with great artistic quality and all of them convey some emotions that add the emotional meanings to the written words. Besides, the use of emoticons makes text messages interesting too.

7.3.2 Punctuations as Artistic Devices

Another category of creative uses of punctuation marks is the utilization of punctuations as artistic devices. In this category, only those text messages have been included which employ punctuation/graphic marks for artistic purposes. Those messages have not been included which use punctuation marks for emoticons like facial expressions. In the following table, examples from SMS-Corpus are given for this category.

Table 7.7
Examples of Punctuations as Artistic Devices

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Translation/Normalization
Ex-194[2563]	w <(_1_)l Se Dro w <(_1_)l Ko Pukaro w <(_1_)l Se Maango w <(_1_)l ko yaad rkho w <(_1_)l se umeed Rkho Hamesha kamiyab raho gay.	. Fear from Allah, . Seek help from Allah, . Demand from Allah, . Remember Allah, . Keep hope from Allah, You will always succeed.
Ex-195[2630]	I ", ,"+m," ", "+ wishing + ""+.+"U ÷ , ,+a"+.+"HAPPY +" +,.+""+.+""+. "* EID *." "*Mubbarak*" ",,"	I am wishing you a happy . EID Greetings!
Ex-196[2677]	*.*__Π__*.*.*. .*/___/Δ*.*.*	. .

	.l_Ω_[l_l]..*	.
	Wo Ghar	Those houses
	Zamin pr Sitaron ki tara Chamkte hen.	glitter like stars on Earth,
	Jin me	in which
	QURAN para jata hy.Aye apny ghr ko Quran se abad karen.	the Quran is recited, so adorn your house with the Quran.
Ex-197[2795]	/\ (* /\	.
	|||| _.-'-._ _.-'-._ |||	.
	||||_(_____) (_____) |||| *	.
	ZilHAAJ ka Chand Mubark ho.	I wish you happy Zil-Haj Moon.
Ex-198[2954]	;*Frè§h*; ;*Ã!R*;	Fresh Air,
	;*DröP§*; öf ;*DEW*;	Dew Drops,
	Á ÑèW §h!Nè öf ;§üN*;	A New Shine of the Sun,
	;*HäpP!ñé§§*;	Happiness,
	&	&
	";*£öVè*;"	Love,
	ÖpéÑs ;*¥öüR*; ;*DööR*;& §ä¥	Open Your Door, & Say
	¤,GØØD,¤	GOOD
	;*MØRÑIÑG*:-..	MORNING!
Ex-199[2978]	"BEGINING OF 20**"	BEGINNING OF 2010
	|""|___|""|.¤.	.
	*| APPY¤	HAPPY
	|__|""""|__|.¤.	.
	FIRST SWEET	FIRST SWEET
	. DAY .	DAY
	.__..OF*.__.*.*	OF
	*.__.*20**.*.__.*	2010
	GOOD MORNING	GOOD MORNING
Ex-200[4459]	|\ (*	.
	|||| _.-'-._ _.-'-._	.
	||||_(_____) (_____)	.
	Har Ibteda Se Pehle,	Before every start,
	Har Inteha K Baad	After every end,
	ZaatE NABI Buland Hy,	The entity of the prophet is great,
	ZaatE KHUDA K Ba	After the entity of God
	,,,,*,,*,,*,,A"	A
	,,,,*,,*,,S"	S
	,,,,*,,A"	A
	,,,,L"	L
	*,,A"	A
	M"	M
	*,,A"	A
	,,,,L"	L
	,,,,*,,i"	I
	,,,,*,,*,,K"	K
	,,,,*,,*,,*,,u"	U
	,,,,*,,*,,*,,*,,M"	M (peace be upon you)

Table 7.7 shows the use of punctuation marks for artistic purposes, excluding emoticons. It shows that various symbols and punctuation marks have been used in text messages to create artistic patterns.

In example-194, the word Allah (اللہ) has been created in the Urdu/Arabic script by using punctuation marks, digit "1", and Roman alphabet "w". This is a very creative use of punctuation marks to develop a word of Urdu/Arabic script. In examples 195, 198, and 199, eid greetings, good morning, and happy new-year have been wished amidst the beautiful patterns of punctuation marks. In example-196, a house has been adorned among glittering stars with the help of symbols and punctuation marks. In examples 197 and 200, two very attractive images of mosques have been formed by using punctuation marks.

Overall, artistic images and patterns created through punctuation marks/symbols are very beautiful and attractive. Moreover, these artistic patterns add further meanings and beauty to the concerned text messages. Zurhellen (2011) also appreciates such innovations in text messages and states that the kind of linguistic play found in text messaging can be viewed as a source of pleasure for texters.

7.3.3 Discussion on Creative Punctuations

In the historical perspective, the first scientific study of nonverbal communication was of Darwin (2009), in his 1872/1890 classis book "The Expression of the Emotions in Man and Animals". He argued that all mammals show their emotions on their faces. Many years later Tomkins (1962, 1963, & 1991) conducted his classic studies (1962-1991) on human emotions. Another contribution towards nonverbal communication was of Birdwhistell (1985). He stated that all body movements have meanings and all forms of paralanguage have a grammar that can be analyzed in similar terms to a spoken

language. He estimated that "no more than 30 to 35 percent of the social meaning of a conversation or an interaction is carried by the words" (p. 158).

Besides these studies, Borg (2010, p.17) states about a 1971 study conducted by Professor Albert Mehrabian of the University of California, Los Angeles that human communication consists of 93 percent body language and paralinguistic devices whereas only 7 percent of communication consists of words. The ratio of nonverbal and verbal communication is further divided into 55% visual, 38% vocal, and 7% word based communication. Nelson and Quick (2008) also present the related figures with a slight difference that "most of the meaning in a message (an estimated 65 to 90 percent) is conveyed through nonverbal communication" (p.187).

Moreover, the division between verbal and nonverbal (linguistic and paralinguistic) devices of language also applies to written language. When this aspect of human communication comes to written language, the textual representation of a writer's mood or facial expression is made through the emotion-icons (emoticons). This element of nonverbal communication through paralinguistic devices becomes significant in those informal writings which need to convey emotions. In this regard, text messages are an appropriate specimen of such emotion driven communication. Therefore, text messages are replete with non-verbal devices to convey emotions; and punctuation marks are tools to construct paralinguistic meanings in a written language.

Regarding the use of punctuation marks for nonverbal purposes in this study, the analysis was made in two categories, i.e. punctuations as emoticons, and punctuations as artistic devices. In both the categories, it was observed that various creative uses of graphic marks add extra nonverbal meanings to words. Moreover, it was noted that in both the categories, punctuation marks were very effectively and innovatively used by texters.

The analysis shows that texters frequently use punctuation marks, or some other graphic symbols to convey non-verbal meanings. Texters convey feelings of happiness, sadness, surprise, emphasis, and much more through punctuation marks. They wink eyes and smile through punctuations marks in text messages. They create various artistic patterns for fun. Hence, repetitive and combined uses of various punctuation marks serve many purposes to texters. Texters employ these punctuations as *paralinguistic* and *artistic* devices to convey feelings and non-verbal meanings.

Bodomo (2010) states that the combinations of question marks (?) and exclamation marks (!) are employed to "show surprises and doubts at the same time" (p. 48). He asserts that there seem no fixed rules to be applicable for using punctuations in such informal communications. He speculates different functions performed by the non-standard ways of using punctuations like "the ease of communication", "to indicate hesitations and thoughts", "to express emotion and mood", "to indicate incompleteness of sentences", and "to show informality and familiarity in informal situations" (p. 48).

Punctuation marks are used in repetitions and combinations for the purposes of creativity or for conveying emotions. When texters wish to convey the emotions like happiness, sorrow, or surprise in text messages, they employ punctuation marks.

Hard af Segerstad (2005) also states that emoticons are used to enrich alphabetic writing by conveying moods and emotions that are normally expressed with extra-linguistic cues such as facial expressions and tone of voice in spoken interaction. The experienced communicator seems to know that some messages may need additional information to clarify word-only communication. For the extra advantages of emoticons and artistic patterns, the economy principle does not always hold, and people spend more time and effort to insert excessive graphic marks to enrich their messages with the additional connotations of meanings.

7.4 Deviant Uses of Spaces

In writing, a space is a blank ( ) area where there is no written content. Spaces are a very important feature of alphabetic writing systems. About their importance in writing, Palmer (1983) states, "The placing of the spaces is as much a part of our education and is as well established as spelling and punctuation, perhaps more so than the latter" (p. 42).

Different languages have different rules of spacing that suit their respective writing scripts. English applies the insertion of one space to separate words, but all languages do not follow this practice. In the English language, written materials use spaces as word separators. As a principle, SMS language is also not an exception to this rule, because if spaces are not used in writing, meanings remain distorted. But it has been noted that in text messages there are instances where space deviations exist. These deviations can be classified under two categories, i.e. 1) omissions of spaces, and 2) additions of spaces. These phenomena of under-spacing and over-spacing have been analysed through examples from SMS-Corpus.

7.4.1 Omissions of Spaces (Under-Spacing)

Omissions of spaces in text messages are an observable phenomenon. They have been explored in SMS-corpus, and a few examples from SMS-Corpus are presented in the following table.

Table 7.8
Examples of Omissions of Spaces

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex-201[12]	wazup!	*What is up?*
Ex-202[276]	I am infront of library	I am *in front* of the library.
Ex-203[278]	Shakespear said: New frends may bepoems but old frends are alphabets.	Shakespeare said: New friends may *be poems* but old friends are alphabets.
Ex-204[336]	I will havechanged my # from tonight.	I will *have changed* my number from tonight.
Ex-205[402]	I'll be home in acouple ofmins.	I'll be at home in *a couple of minutes.*
Ex-206[672]	Gudnyt	*Good night.*

Ex-207[1200]	vr in parking	*We are* in the parking.
Ex-208[1352]	Hv agr8 holiday.	Have *a great* holiday.
Ex-209[1590]	I'm js gonna reach in amin.	I am just going to reach in *a minute*.
Ex-210[1668]	going to Risalpur got amatch.	I am going to Risalpur. I have got *a match*.

Table 7.8 shows various examples of space-less text messages. The insertion of a space at word boundaries is conventional to the English language. If a blank space is not inserted after each word, words mix up and meanings and identities of words are distorted. Nevertheless, all ten examples in the table show that various words have not been separated from neighbouring words with spaces. These omissions of spaces are conventionally unacceptable in the perspective of a formal language.

7.4.2 Additions of Spaces (Over-Spacing)

The other category of deviations in spacing is the additions of spaces where they are not needed. This phenomenon of excessive insertions of spaces (over-spacing) has been analysed in SMS-Corpus. The examples are given in the next table.

Table 7.9
Examples of Additions of Spaces

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Standard English (Normalized SMS)
Ex-211[44]	Everyone is incomplete without some thing, what do u think.	Everyone is incomplete without *something*, what do you think?
Ex-212[608]	"Worst reality" What is the heaviest burden of life????	"Worst reality": What is the heaviest burden of life?
	An Empty pocket.	An empty pocket.
Ex-213[1236]	Take care of your self.	Take care of *yourself*.
Ex-214[1603]	Every one waiting 4 u	*Everyone* is waiting for you.
Ex-215[4710]	you were no t there	You were *not* there.

In Table 7.9, there are two types of over-spacing that are, within the word boundaries (examples 211, 213, 214, and 215), and within discourse boundary (example 212). Conventionally, singles are not broken into parts until they are complete. Same is the case with the compound words which are conventionally written as one word. But in

the given examples we see that both single and compound words have been broken with spaces within word boundaries. Conventionally, these divisions/breaks in words with spaces, within word boundaries are ungrammatical.

In the same way, within a discourse, sentences are conventionally separated from each other with single or two spaces. To include more than two spaces to separate sentences is ungrammatical. Individual discourses (paragraphs) can be separated from each other with line spacing, but semantically connected sentences are not separated with line spacing. Nevertheless, we see in text messages that sentences are separated from each other with line spacing and it is a common practice in text messages.

7.4.3 Discussion on Deviant Spaces

By now, different adaptations in the thesis have been variously attributed to three types of reasons. These reasons are 1) to be economical in the use of time and effort, 2) to be artistic in the innovative patterns, and 3) to be careless for the purposes of speed. Economical causes are active when texters want to save time and effort; artistic causes have grounds when texters use extra time and effort to create something artistic; and the carelessness/haste occurs when texters struggle for speed. Of all three major causes of adaptations in text messages, the last one is most criticised one because these deviations mostly occur unintentionally and cannot be attributed even to the known patterns and principles of SMS adaptations.

The analysis of deviant uses of spaces in the study shows that deviations in this area also fall under the third category of causes. So the deviations of spaces in text messages even do not fit to the frequent patterns maintained by the medium in which they have been used. They have been used inappropriately, and Crystal (2008b) states that deviations can be accepted and tolerated only if they are at least appropriate to the medium in which they are used.

Regarding the causes of space adaptations in text messages, Hard af Segerstad (2005) attributes the reasons of space omitting to time and effort saving. She states that "similar to the strategy to omit punctuation, by omitting space between words the user saves keystrokes as well as time and effort" (p. 1). She further states that in some cases it is not necessary to save space, but omitting it renders a personal tone to the messages of individual texters. She does not attribute these deviations to the carless attitude of texters.

CHAPTER 8

CODE AND SCRIPT ADAPTATIONS

The chapter deals with code and script adaptations in the bilingual/multilingual setting of Pakistan. It includes code-switchings and code-mixings in text messages between Urdu and English. It also addresses the use of the Roman script/alphabet for the Urdu language in text messages. Moreover, the chapter presents the corpus based Romanization scheme for Urdu on the bases of Romanization patterns found in SMS-Corpus.

8.1 Introduction to Code and Script Adaptations

In a bilingual/multilingual society, it is a very common practice that speakers code switch from one contact language to the other. This phenomenon in text messages has been overlooked by researchers because most available studies in the field were conducted in monolingual societies, mostly in English speaking countries. Crystal (2008b) expects that the phenomenon of code adaptations may be a major characteristic of text messages if it is investigated empirically. He states that "as most parts of the world are bilingual, I would expect code-mixing to develop to be a major characteristic of texting; but we will need larger samples of data before this can be demonstrated" (p.130). In this regard, the present study will be a good addition in the field of code adaptations in text messages for bilingual societies.

As Pakistan is a bilingual/multilingual country, so swinging between Urdu and English is a very common phenomenon in speech. Even in written modes, like newspapers, one can find many instances of code-mixing (Rasul, 2009). Moreover, the mode of SMS is a hybrid of speech and writing (Baron, 2008; Crystal, 2008b), so one can easily expect code-switching/mixing in text messages in a bilingual society.

Thurlow and Poff (2011) report a few studies conducted on code adaptations in text messages in multilingual cultures. In the most of cases, English is in contact with some local language of a bilingual country. They state that South African texters blend English with isiXhosa by writing English nouns with isiXhosa prefixes. Finnish teenagers mix Finnish with a medley of foreign language words and expressions, drawing suitable expressions from any language mastered by the writer. Greek texters use graphemes in alphabet-choice and code-switches.

The phenomenon of code adaptations in the context of Pakistan is of historical nature. The historical position of the region as a former British colony has provided a firm ground for Urdu-English language contact. Urdu as the single national language and English as the single official language remain in close contact to each other in both formal and informal settings. Moreover, historical as well as modern global impacts of English in Pakistan have resulted in the continuous language contact of Urdu and English, and this contact leads to code-switching and code-mixing. Rasul (2008) states that due to the hybridization of Urdu and English, a new code "Urdish" has emerged in Pakistan.

In any study of code adaptations, two very closely related concepts which need to be understood are code-mixing and code-switching. Different experts interpret them in slightly different ways. These differences sometimes create confusions for the people who are not specialised in the field. Therefore, main categorizations of code adaptations are introduced here, especially focussing on code-mixing and code-switching.

In this regard, Poplack (1980) divides language switching into three categories, i.e. 1) Tag switching, 2) Intra-sentential switching, and 3) Inter-sentential switching. Tag switching involves the insertion of 'tags" from the guest (embedded) language into the host (frame/matrix) language. Intra-sentential switching occurs within the

clause/sentence boundary, like the insertion of words and phrases of one language to the other. In inter-sentential code-switching both languages swing from one to the other, that means, some sentences from one and some from the other language are used. This switching occurs outside the clause/sentence boundary.

Rasul (2006) in her study of Urdu-English hybridization explores code-mixing of English into Urdu. She takes Urdu as the matrix language, and examines the embedding of English in it. She divides code-mixing into three categories, i.e. 1) Insertion, 2) Hybridization, and 3) Synthesis. She divides insertions into four sub-categories that are word, phrase, clause, and sentence insertions. She divides hybridization also into four categories that are single/compound words and noun/verb phrases. Her third category synthesis is based on insertions of abbreviations, greetings, names, repetitions, tags, and fillers etcetera.

In short, different studies define and categorise code alterations in their specific perspectives, yet the broader division is usually made into two categories, i.e. *inter-sentential* and *intra-sentential* code alterations. In this regard, the *Encyclopedia of Bilingual Education* states, "Code switching is usually divided into two types: *intersentential,* switching between sentences, or *intrasentential,* switching within sentences.... many European researchers reserve the label *code switching* for the first type and use *code mixing* for the second" (Troike, 2008, p. 143).

Therefore, for the operational purposes of this study, all code adaptations of SMS-Corpus have been divided into two broader categories that are *code-mixing* and *code-switching*. In simple words, code-mixing is a hybrid form of two contact languages where words/phrases are inserted from the guest (embedded) language into the host (frame) language within the clause/sentence boundary; whereas code-switching is a swing between two contact languages on/outside the sentence boundary.

In the setting of a bilingual country, the other related concept with code adaptations is script adaptations in text messages. In written language, when two languages are frequently code-switched, it is practically not possible that half utterances should be written in one script and half in the other. Urdu is traditionally written in the Arabic/Perso-Arabic script, which is also sometimes called the Urdu script as it is the modified form of both Arabic and Perso-Arabic scripts. But texters in Pakistan mostly use the Roman script in which English is conventionally written. The main reason of adopting Roman script for both the languages is that Pakistani texters feel easy in this script. Most of the mobile sets do not offer Urdu keypad, and even if it is available texters do not use it for their own convenience.

The result of the abundant use of the Roman script for Urdu in text messages is that there is the revival of Roman Urdu that was once a need in the British colonial period for official purposes. As texting has become a common medium of communication in Pakistan among the educated people, so the revival of the Roman script as a rival script to the Urdu script cannot be overlooked. Moreover, majority of people in Pakistan write text messages in the Urdu language, but use an unconventional script for it. This phenomenon is of serious nature for those who have concern for the Urdu script as the sole script of the Urdu language. To study the concerns, problems, and principles related to these issues, this study examines SMS-Corpus in the following sections.

8.2 Code-mixing

Code-mixing of Urdu and English is a very common language phenomenon in Pakistan. In this regard, a major study of Urdu-English hybridization was conducted by Rasul (2006). She analysed the data of 10 TV programmes of the Urdu language (talk shows and discussion panels) of "six hours and 38 minutes" duration. In her analysis,

she took Urdu as the matrix language, and examined the embedding of English in it. She divided the code-mixed language into three categories, i.e. 1) Insertion, 2) Hybridization, and 3) Synthesis. Among the four sub-categories of "insertion", the most frequent insertions are the word insertions (1783), followed by phrase insertions (979), clause insertions (99), and sentence insertions (53), hence making the total 2914 insertions (p. 358). In "hybridization", all four categories of single/compound words and noun/verb phrases make the total of 2041 instances (p. 293). Her third category "synthesis" is based on insertion of abbreviations, greetings, names, repetitions, tags, and fillers etcetera. But this category is indeed "overlapping with the data presented in the previous two categories" (p. 359).

Regarding the operational definition of code-mixing, Myers-Scotton (2006) regards code-mixing as lexical borrowing. According to him, when words and phrases are borrowed from one language to the other without any cognitive process, this phenomenon is called "lexical borrowing". For him, borrowing is a one-way process from a more-prestigious language to a less-prestigious language. He divides it into two main categories that are cultural and core borrowing. In "cultural borrowing", new words enter the recipient language from the donor language, whereas in the "core borrowing" same (duplicate) words from the donor language are borrowed whose alternatives already exist in the recipient language.

In the educated circles of Pakistan, in spoken language Urdu is mostly used as the host (frame) language and English as the guest (embedded) language. On the other hand, in text messages both Urdu and English act as "frame" and "embedded" languages. Hence, in text messages, code-mixing is a two way process and both the ways have been analysed in two separate sections. The analysis and discussion on both types of Urdu-English code-mixing in text messages is given below.

8.2.1 English Embedded in Urdu

The most frequently occurring category of code-mixing in text messages in Pakistan is the mixing of English into Urdu. This occurs at word, phrase, and clause levels within the sentence boundaries, and therefore it is also termed as intra-sentential code-mixing. Urdu, the national language of Pakistan, acts as the frame language and English, the official language of Pakistan, acts as the embedded language. The insertions/mixings are made at the levels of single words, phrases, and clauses. For the extensive analysis of different code-mixed items, examples from SMS-Corpus have been categorised in two separate tables covering "single words", and "phrases plus clauses".

Table 8.1
Examples of Single Word Insertions (English-into-Urdu)

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Category of Code-mixing	Normalized/Translated SMS in English
Ex-216[21]	Aaj *class* hai?	noun	Is there class today?
Ex-217[57]	*Sir* g, *busy* to nea	noun, adjective	Dear sir, are you busy?
Ex-218[223]	*Hi* yar aaj main *late* aaon ga *uinv*!	interjection, adverb, noun	Hi! Dear, today I will come late to the university.
Ex-219[554]	*Hi!* Kahan ho???	interjection	Hi! Where are you?
Ex-220[723]	Main phar rahi the. Phir *suddenly* tmhare yad any lagi.	adverb	I was studying then suddenly you came in mind.
Ex-221[3827]	Mai b namaz parrlon *but* kaam *complete* karain abi	conjunction, verb	I am also going to offer prayer but complete the work, now.
Ex-222[4039]	*Wow*.kab se join karin gay	interjection	Wow! When will you join?
Ex-223[4063]	Aajain *but* aapko *address* nai pata	conjunction, noun	Come but you don't know the address.
Ex-224[4202]	Kitna *write* kar lya hy	verb	How much have you written?
Ex-225[4809]	Assalam-o-alaikum…*Any* narazgi?	adjective	Peace be upon you…. Any resentment?

Note. Code-mixed words represented in Italics

Table 8.1 shows examples of single word insertions from English into Urdu. The table shows that single words of English have been embedded into Urdu in six categories of word classes that are nouns, adjectives, verbs, adverbs, conjunctions, and interjections. Hence, in the broader sense only two word classes (parts of speech) have been left over that are pronouns and prepositions. When these results are compared with Rasul (2006),

she includes three categories that are nouns, adjectives, and adverbs. Abbas, Aslam, and Rana (2011) include five categories that are nouns, adjectives, verbs, adverbs, and conjunctions.

In addition to the single word insertions (mixing/embedding), the insertions of phrases and clauses are also commonly found in text messages. Their analysis is offered through the next table.

Table 8.2

Examples of Phrase & Clause Insertions (English-into-Urdu)

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Category of Code-mixing	Normalized/Translated SMS in English
Ex-226[238]	Bht *tough questions* hein	noun phrase	The questions are very tough.
Ex-227[246]	Ap phr *tym waste* kr rhe ho.	verb phrase	You are again wasting time.
Ex-228[664]	*My name is Khan* ka part 2 aaney wala hai.	dep. clause (noun clause)	Part-2 of "My Name Is Khan" is about to release.
Ex-229[1828]	Salam bhai…ye tu mghe nahe patra *if u r sleeping*	dependent clause	Salam brother, I don't know if you are sleeping.
Ex-230[2121]	Yar aik request ha tum se *if u dnt mind?*	dependent clause	Dear I have a request to you if you don't mind.
Ex-231[4410]	ek kiss kar lana *on the hand.*	prep. phrase	Please give a kiss on the hand.
Ex-232[4668]	*The great* HAZARA Zindabad	adjectival phrase	Long live the great Hazara!
Ex-233[4791]	*Reg no* batao	noun phrase	Tell your registration number.
Ex-234[4902]	ab so jao *GUD NIGH*	adj. phrase	Now go to sleep, good night.
Ex-235[4997]	sir main *examination i/c room* main hun	noun phrase	Sir I am in the examination in-charge room.

Note. Code-mixed phrases/clauses represented in Italics

Table 8.2 shows examples of phrase and clause insertions from English into Urdu. The table shows that at the phrase level four categories of noun, verb, adjectival, and prepositional phrases have been code-mixed from English into Urdu. Hence, at the broader level, only the adverbial phrases have been left over out of the five main classes of phrases. When these results are compared with Rasul (2006), she presents three categories that are noun, verb, and prepositional phrases. Therefore, adjectival phrase is an extra category of code-mixing found in text messages.

At the level of clause insertions, it is noteworthy that only dependent (sub-ordinate) clauses have been code-mixed in Urdu from English. Moreover, this code-mixing of clauses occurs at the clause boundary, mostly in the presence of some complementizer (subordinating conjunction) like "if", "that", and "and" etcetera.

8.2.2 Urdu Embedded in English

The way English is embedded in Urdu in text messages, Urdu is also embedded in English sentences. This also occurs at three levels that is, at word, phrase and clause level. In this category of Urdu to English code-mixing, English acts as the frame (host) language and Urdu as embedded (guest) language. It means that grammatical structure is provided by the English language instead of the Urdu language, overruling the common view that Urdu always provides the grammatical structure to the English language in the code-mixing process.

The analysis in this category has also been divided into two categories that are "single words" and "phrases plus clauses", for the better understanding of the process.

Table 8.3
Examples of Single Word Insertions (Urdu-into-English)

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Category of Code-mixing	Normalized/Translated SMS in English
Ex-236[80]	*JANI* I hope u wil fine	noun	Dear I hope you will be fine.
Ex-237[175]	mama loves you *beta!*	noun	Mama loves you son!
Ex-238[176]	*Kia?* Dunt u/	interjection	Kia! Don't you?
Ex-239[361]	*Aur* How was your day today?	conjunction	And how was your day today?
Ex-240[364]	*Oye* kindly bring the trip photos.	interjection	Hello! Kindly bring the trip photos.
Ex-241[428]	Congrats *yar* on ur result.	noun	Dear congratulations on your result.
Ex-242[458]	*api* r u free?	noun	Sister, are you free?
Ex-243[2046]	Plz make ! cup of tea 4 *abu*	noun	Please make a cup of tea for father.
Ex-244[2089]	plz brng my Phy key 4rm bazar	noun	Please bring my Physics key from the market.
Ex-245[4230]	Just to see you and chat with u about *dunidari* things.	adjective	Just to see you and chat with you about worldly affairs.

Note. Code-mixed words represented in Italics

Table 8.3 shows examples of single word insertions from the Urdu language into English. The table shows that single words of Urdu have been code-mixed into English in four categories of word classes that are nouns, adjectives, conjunctions, and interjections. In the broader sense, four word classes (parts of speech) have been left over that are pronouns, prepositions, verbs and adverbs. Now the analysis is presented in two other categories that are phrases and clauses.

Table 8.4
Examples of Phrase & Clause Insertions (Urdu-into-English)

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Category of Code-mixing	Normalized/Translated SMS in English
Ex-246[81]	*Abu g* when u send money.	noun phrase	Dear father, when will you send money?
Ex-247[446]	*Assalamo alaikum* Sir, will u plz tel me my grade?	noun phrase	Peace be upon you sir. Will you please tell me my grade?
Ex-248[646]	*Kai khail hai* go to Murree.	dep. clause (noun clause)	what is your opinion to go to Murree?
Ex-249[649]	I real Misssss Yoouuu *Dil Se……*	postpositional phrase	I really miss you from heart.
Ex-250[736]	I lve u *mari jan*	noun phrase	I love you my dear.
Ex-251[1007]	I'm extremely sory *agar ap ko meri koi bt b buri lagi to*	dependent clause	I am extremely sorry if you felt hurt on any of my comments.
Ex-252[2269]	maom z v8ng nd *abu b tumhara pouch rhe hain.*	dependent clause	Mom is waiting and father is also inquiring about you.
Ex-253[2624]	Happy *eid mubarik* to u and u family.	noun phrase (greeting)	Happy Eid greetings to you and your family.
Ex-254[3610]	I remember *pictures mai koi b etna pyara nai tha!*	dependent clause	I remember that there was no one so cute in the pictures.
Ex-255[4752]	Sir I m very sorry I could tell hgim yesterday, *InsahAllah* I will tell him today	noun phrase	Sir I am very sorry. I could tell him yesterday, God willing I will tell him today.

Note. Code-mixed phrases/clauses represented in Italics

Table 8.4 shows examples of phrases and clauses of Urdu embedded in English. The table shows that at the phrase level noun phrases are the only insertions found in abundance. One example of postpositional phrase (i.e., prepositional phrase in English) was also found in the corpus. Hence, the results show that Urdu into English code-mixing is not as rich as is English into Urdu code-mixing.

At the clause level insertions, like English into Urdu code-mixing, only the dependent (sub-ordinate) clauses have been code-mixed in English from Urdu. Moreover, the code-mixing of clauses also occurs at the clause boundary, mostly in the presence of a complementizer (subordinating conjunction) like "agar (if)" in Example-251[1007], and "and" in Example-252[2269]. These two examples also show that the complementiser can be either from English or Urdu.

8.2.3 Discussion on Code-mixing

Ho (2007) defines that code-mixing is the change of one language to another within the same utterance, and it is a common phenomenon in societies in which two or more languages are used. He further states that studies of code-mixing enhance our understanding of the nature, processes and constraints of languages. In this context, this study offers insights into the code-mixing in text messages.

The results of the study show that code-mixing in text messages is not a one-way process, instead a two-way traffic. It means that both Urdu and English act in both ways. In some instances, Urdu acts as the frame language whereas in others English performs the role of the frame language. This two-way process encompasses all three levels of code-mixing that is at single-word, phrase and clause levels.

The difference between the two is that English-into-Urdu code-mixing is more rich than Urdu-into-English code-mixing. This can be seen in the results given in this section. At the single-word level, English-into-Urdu code-mixing involves words in six categories that nouns, adjectives, verbs, adverbs, conjunctions and interjections. When this comes to Urdu-into-English code-mixing, this process decreases to only four categories that are nouns, adjectives, conjunctions, and interjections.

The difference at the phrase level is also noteworthy because English-into-Urdu code-mixing involves four types of phrases that are noun, adjectival, verb, and

prepositional phrases, leaving only adverbial phrases, in the five major categories of phrases. On the other side, phrase level Urdu-into-English code-mixing is mostly limited to noun phrases. At the clause level, both "English-into-Urdu" and "Urdu-into-English" code-mixings involve dependent clauses. On the whole, English-into-Urdu code-mixing is richer than Urdu-into-English code-mixing on the bases of single-word and phrase code-mixings.

When the present study is compared with Rasul (2006), the results mostly match with each other. Both Rasul and this study conclude that single-word code-mixing is more common than phrases; and phrase code-mixing is more common than clauses. At the word level, nouns are the most frequent code-mixed elements whereas at phrase level noun phrases are the most common code-mixed items.

As far as the broader differences of both studies are concerned, Rasul (2006) takes spoken data whereas this study is based on a written mode, or hybrid of two modes (her work on written mode is covered in her 2009 study). Moreover, she takes all types of insertions (i.e., word, phrase, clause and sentence) as code-mixing, whereas this study categorises inter-sentential (sentence) switching as code-switching. This difference occurs due to the reason that her study initially defines that her data is based on conversations of TV programmes in the Urdu language. Hence, in her study, Urdu is the frame language because all TV programmes actually were meant to be in the Urdu language. On the other hand, this study is based on text messages, where texters can switch from one language to the other any time, so no language can be constantly termed as the frame language. In text messages as well as in general terms, the frame language is the language that provides the grammatical structure, and therefore the frame language is determined on sentence to sentence bases. So this study takes inter-sentential (outside sentence boundary) switching as code-switching and not as code-mixing. Another

difference between her code-mixing and the code-mixing of this study is that her work covers one way English-into-Urdu code-mixing, whereas this is a two way study that covers both English-into-Urdu and Urdu-into-English code-mixing.

When this study is compared with Abbas, Aslam, and Rana (2011), and Rasul (2006, 2009) about the comparative frequency of word, phrase, clause, and sentence level code alterations, all studies converge at one point that single lexical items (i.e., single words) are the most common and frequent elements in all four kinds. Phrases are at second, clauses at third, and sentences at fourth level, as far as the comparative frequencies of code-mixing/switching are concerned. Moreover, among the single word code-mixing, nouns are the most frequent code-mixed items concluded by all the above mentioned studies.

When these results are seen in the light of Myers-Scotton (2006), he regards the code-mixing of words and phrases as lexical borrowing. For him, borrowing is a one-way process from the more-prestigious language to the less-prestigious language. But the present study shows that lexical borrowing is a two-way process as shown in sections 8.2.1 and 8.2.2. Both English-into-Urdu and Urdu-into-English borrowings do exist in SMS-Corpus. Hence, Myers-Scotton's theoretical framework of one-way borrowing from the more-prestigious language to the less-prestigious language does not fit in the case of text messages at least.

This study also presents interpretations regarding Myers-Scotton's (2006) two categories of borrowings that are "cultural borrowing" (of new words), and "core borrowing" (of duplicate words) from the donor language to the recipient language. In SMS-Corpus "core borrowing" has been found more than "cultural borrowing". In the case of both English-into-Urdu and Urdu-into-English borrowings, mostly duplicate

words have been code-mixed. Therefore, it is implied that duplicate words are code-mixed in text messages mainly due to either "economical" or "artistic" reasons.

8.3 Code-switching

Myers-Scotton (2006) states that code-switching is the borrowing of grammatical structures that involves cognitive process. He divides the contact languages into "matrix (host) language", and "embedded (guest) language". He further divides code-switching into two types that are "classic" and "composite" code-switching. In the classic code-switching, main grammatical (morpho-syntactic) frame is offered by the matrix language, whereas in the composite code-switching both languages participate equally. After this brief introduction, the analysis regarding issues and patterns of code-switching in text messages is given as follows.

8.3.1 Code-switching from Urdu to English

Inter-sentential code-switching from Urdu to English is very common in text messages. This phenomenon is further elaborated through examples from SMS-Corpus.

Table 8.5

Examples of Code-Switching from Urdu to English

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Normalized/Translated SMS in English
Ex-256[43]	Yar tm khelne a skte ho.??? *I am waiting*	Dear can you come for playing? I am waiting.
Ex-257[171]	Ap kidher ho *whats happening.*	Where are you? What is happening?
Ex-258[254]	cafe ponch, me cmng in 10 min.	Reach at café. I'll come in 10 minutes.
Ex-259[340]	yr! Hum khana khanaey ja rahey hain. *come to my room now.*	Dear, we are going to have meal. Come to my room now.
Ex-260[366]	Ma theek hoo, *How are you?*	I am fine. How are you?
Ex-261[566]	Mai 12:30 tak free ho jao gi. *Pick me from centre.*	I will be free by 12:30. Pick me from the centre.
Ex-262[585]	Me aj late ho jaon gi. *So, don't wory.*	I will be late today. So don't worry.
Ex-263[643]	Jani mery number pr 30 Rs ka load krwa do. *I love u so much*	Dear send easy-load of Rs. 30 on my number. I love you so much.
Ex-264[4843]	yar who mein soch rai hoon tahira ko bol hiu don....! *wat do u say!*	Dear I think, I should tell Tahira. What do you say?
Ex-265[4893]	itni bari baat nahi hey, *don't make mess of it*	This is not a big issue. Don't make fuss of it.

Note. Code-switched part represented in Italics

Table 8.5 shows examples in which code-switching has been applied in one-way i.e., from Urdu to English. In all examples code-switching has been applied on/outside the sentence boundary. In all examples the first utterance is in Urdu, whereas the following utterance is in the English language. Therefore this type has been termed as Urdu to English code-switching. Now, in the following section, examples are offered in the opposite category that is English to Urdu code-switching.

8.3.2 Code-switching from English to Urdu

This category of code-switching is also common in text messages but less common to the previous category. The analysis through examples from SMS-Corpus is shown below.

Table 8.6
Examples of Code-Switching from English to Urdu

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Normalized/Translated SMS in English
Ex-266[100]	Ok w8 *m viva dy k ata hn....*	Okay wait. I come after taking viva.
Ex-267[105]	ok *theak*	Okay. It is fine.
Ex-268[123]	Tigr, m waiting for ur reply, *aana ha ke nae bta.*	Tiger, I am waiting for your reply. Will you come or not, inform.
Ex-269[304]	Em f9, *tum kaisay ho, kia ho raha ha?*	I am fine. How are you? What's going on?
Ex-270[620]	Me cming in 10 mintes, *main abi ghr ce nikla hn.*	I am coming in 10 minutes. I have just come out of home.
Ex-271[3821]	Leave this *ye btaain practice karli hy?*	Leave this. I have practised these things.
Ex-272[4204]	W8 *dadi ko dawai daydon pehly fir juice p k phon karti hon*	Wait. Let me give medicine to father first then I call you after drinking juice.
Ex-273[4294]	sorry i an back in mzd. *kal ap ne call nai ke?*	Sorry I am back in mzd. Yesterday you did not call me.
Ex-274[4446]	Good night.I think so u r busey with frinds.ok. *ALLAH HAFIZ.*	Good night. I think so you are busy with friends. Okay. May God protect you!
Ex-275[4842]	Plz bring my notes and copy *Aaj mat bhool jana*	Please bring my notes and copy. Do not forget today.

Note. Code-switched part represented in Italics

Table 8.6 shows examples in which code-switching has been made from English to Urdu. Like the previous section, the code-switching has been applied outside sentence boundaries. In all examples, first utterances are in English, whereas following utterances

have code-switched to Urdu. Hence this type has been labelled as English to Urdu code-switching.

8.3.3 Two-Way Urdu-English/English-Urdu Code-switching

Now, after presenting two types of one-way switching, this section concerns two-way code-switching, either from Urdu to English, or from English to Urdu, and then back way once again.

Table 8.7
Examples of Two-Way Code-switching

Ex-No[SMS-No]	Actual SMS from SMS-Corpus	Normalized/Translated SMS in English
Ex-276[18]	AOA. I'm sory us waqt sir k pas thi	Peace be upon you. I'm sorry. I was with the teacher at that time.
Ex-277[151]	kdhr ho? y is ur num goin busy? kis k sath ho	Where are you? Why is your number going busy? Who is with you?
Ex-278[303]	Ha, bro kidr ho? Em' w8ing in cafe.	Hi! Brother, where are you? I'm waiting in the café.
Ex-279[426]	TODAY IS IMPRESS DAY". Meri koi c ek aisi khobi bato Jis se ap bot impress hn. Reply must I'm waiting.	Today is Impress Day. Tell me one of my qualities that you are impressed of. You must reply. I am waiting.
Ex-280[3881]	Ok sorry sorry pehlay btana tha na kaam karain ok good night	Okay! Sorry, sorry! You should have told earlier. Do your work. Okay! Good night.

Note. Code-switched part represented in Italics

Although, two-way code-switching is less common in SMS-Corpus, yet examples show that this kind does exist in text messages. This is noteworthy that two-way code-switching moves back and fro in single messages. Hence, such frequent two-way switches seem to be a very interesting phenomenon keeping in mind the small size of text messages.

8.3.4 Discussion on Code-switching

In linguistics, code adaptations are the practices of language swing between two contact languages, and in bilingual societies this is a very common practice in conversation. It is the simultaneous use of two or more languages in conversation. Actually, code-switching and code-mixing are the spoken language phenomena, yet these processes also travel to the written modes of communication as well.

In its very nature, code-switching is the integration of two languages on/outside sentence boundaries. This is also termed as inter-sentential switching. Although the two contact languages are switched on/outside sentence boundaries, yet they must exist within the single discourse. Examples given in the tables show that all three possible categories of Urdu-to-English, English-to-Urdu, and two-way code-switching do exist in text messages. This aspect shows that cultural dominance of the English language, as discussed by Myers-Scotton (2006), is not the only cause of code-switching in text messages.

The perception analysis of texters' metalinguistic awareness about code adaptation shows that code adaptations occur due to two main reasons (see Table 4.10). These are texters' own ease in writing text messages (20.6%), and understanding of the recipients of text messages (42.6%). To have style and fun in text messages through code alterations is the third significant reason explored through perception analysis of 500 texters. These results overrule the cause of cultural dominance and the prestige influences of one language over the other. If this had been the case, code-switching in text messages would not have occurred in two ways in both Urdu and English.

As already discussed in the introduction to this section, Myers-Scotton (2006) divides code-switching into two types, i.e. "classic" and "composite" code-switching. Classic code-switching is a one way process, from the less prestigious language to the more prestigious language, where grammatical frame is provided by the more prestigious language. On the other hand, in composite code-switching both languages participate equally. In this regard, examples of code-switching from SMS-Corpus show that code-switching in text messages is a two way process, i.e. English to Urdu and Urdu to English. Therefore, code-switching in text messages falls under the category of composite code-switching.

8.4 Romanization of Urdu

Romanization is the process of writing a language in the Roman script, while the conventional writing script of that language is someone else. The process of Romanization is further divided into transliteration and transcription. Transliteration deals with the written text, whereas transcription with the speech. Each of the two processes has its own set of rules. In transliteration, Romanization attempts to transform the written language into Roman alphabets by one-to-one mapping of characters, whereas in transcription, all sounds of a spoken language are transcribed on phonetic bases. Regarding transcription, besides the Roman alphabet/script, International Phonetic Alphabet [IPA] is an ideal sound based system of transcribing all the languages of the world.

Urdu, the national language of Pakistan, is traditionally and academically written in Urdu (Perso-Arabic) script, which in turn is the modification of the Arabic script. The history of Romanization of the Urdu language goes back to pre-Pakistan British era when this process was used by the British to use the Urdu language for colonial purposes. In this regard, John Gilchrist's (1796) Romanization system is a pioneering work in this field. At present, the abundant use of the Internet and mobile phone technologies by masses has revived the use of Urdu Romanization in Pakistan.

8.4.1 Variations of Spellings in Roman Urdu

To study the corpus based Romanization patterns adopted for Urdu in text messages, the study offers the variant spellings of the most frequent Urdu words from SMS-Corpus. The results and the interpretations in this regard are offered below.

Table 8.8
50 Most Frequent Spelling Variations in Roman Urdu (AntConc analysis)

Variant Roman Spellings of Urdu Words	FRQ	Target Urdu Words in IPA
k(959), ki (548), ka (462), ke (91), kay (30), keh (21), key (6), ky (4), kaa(2)	2123	/ke/, /ki:/, /kɑ:/, /kɪh/

hy (450), hai (444), ha (290), he (194), hay (159), hey (66), h (48), hei (4), hae (3)	1658 /hɛ/
me (484), mai (197), main (195), ma (176), may (124)	1176 /mẽ/, /mɛ̃/
to (1046), tu (91), too (20), tou (9), toe (2)	1168 /ʈo/, /ʈu:/, /ʈũ:/
a (637), aa (63)	1000 /ɑ:/
the (469), tha (123), thi (40), they (36), thay (13)	681 /ʈʰi:/, /ʈʰe/, /ʈʰɑ:/
b (293), be (168), bhi (69), bi (21), bhe (11)	562 /bʰi:/
se (311), sy (133), sa (90), say (86), si (16), sey (10), sea (3), ce (3)	652 /se/, /si:/, /sɑ:/
ho (517), hu (21)	538 /ho/
is (518), es (20)	538 /ɪs/
or (383), aur (140), aor (6),	529 /ɔr/
nai (255), nahi(109), nhi(58), nahe(24), nae(23), nhe(23), nahee (2)	494 /nəhi:/, /nəhĩ:/
ko (489)	489 /ko/
2 (278), do (138)	416 /ɖo/
hain (332), han (81)	413 /hɛ̃/
ap (174), aap (172), app (17)	363 /ɑ:p/
kr (178), kar (134), ker (21)	333 /kər/
kya (145), kia (142), kiya (19), kea (12), keya (2)	320 /kɪjɑ:/, /ki:jɑ:/
na (296), nah (4),	300 /nɑ:/, /nəh/, /nɑ̃:/
ga (175), gi (80), ge (43)	298 /gɑ:/, /gi:/, /ge/
Allah (296),	296 /ələh/
aoa (140), Assalam-o-ala'kum (83) [*variant spellings= ikum (71)+ ekum (7)+ykum (5)*], salam (65)	288 /e o e/, /əsəlɑ:m o əlʔi:kʊm/, /səlɑ:m/
1 (167), ek (65), aik (45),eik (3), ak (3)	283 /ek/
ne (150), ny (103), nay (17), ney (6)	276 /ne/
meri (80), mera (57), mere (56), mairi (18), mairy (13), mery (10), mari (7), mri (5), merey (3), maira (2), mary (2), mry (2)	255 /meri:/, /merɑ:/, /mere/
ye (199), yeh (30)	229 /jɪh/
e (211)	211 /e/, /i:/
hon (96), hun (51), hn (41), hoon (17), huu(1)	206 /hũ:/
us (181), os (8)	189 /ʊs/
yar (139), yr (34), yaar (15)	188 /jɑ:r/
pr (95), per (47), par (43)	185 /pər/
4 (175)	175 /t͡ʃɑ:r/
o (171), oh (26)	171 /o/, /oh/
aj (98), aaj (51), ajj (7)	156 /ɑ:d͡ʒ/
wo (140), woh (11)	151 /vʊh/
jo (140)	140 /d͡ʒo/
sb (78), sab (40), sahb (13)	131 /səb/, /sɑ:hb/
kal (119), kl (9)	128 /kəl/
tum (83), tm (38)	121 /ʈʊm/
koi (112), koe (2),	114 /koʔi:/
hum (73), hm (24), ham (13)	110 /həm/
3 (108)	108 /ʈi:n/
raha (71), rha (30), reha (3)	104 /rəhɑ:/
pe (71), pay (15), pa (13)	99 /pe/, /peh/
eid (98)	98 /i:ɖ/
g (70), ji (14), je (7), gee (4), jee (2)	97 /d͡ʒi:/
ghar (75), ghr (19)	94 /gʰər/
ya (91)	91 /jɑ:/

kb (44), kab (38), kub (5)	87	/kəb/
tak (60), tk (24), tuk (2)	86	/ʈək/

Note. Letter and number token classes activated in AntConc (Case deactivated)

Table 8.8 shows variations of spellings in Roman Urdu. In the table, 50 most frequent spelling variations in Roman Urdu have been presented. When these variations in Roman Urdu spellings are compared with the spelling variations of the English language in text messages, both languages employ certain patterns of phonological and punctuation adaptations, and paralinguistic devices. But it is noteworthy that Urdu rarely uses number and symbol homophones which are otherwise common in the English language (see section 5.2.4)

When we study the complications of Roman spelling patterns for the Urdu language in text messages, the table shows that variant Roman spellings have been used to represent a single word of Urdu. Similarly, one spelling combination can represent more than one Urdu word. This variation in spelling occurs because the Roman script is not the conventional script of the Urdu language, and texters intuitively develop spellings for Urdu words. This is also noteworthy that the problem only concerns with vowel sounds (both pure and nasal). An analytical look on variant Roman spellings in the table shows that consonant letters do not vary from texter to texter, and remain consistent. The problem is only with vowel letters, because one single vowel letter can represent various vowel sounds, and one single vowel sound can be represented with various vowel letters.

Moreover, some texters use consonantal style of writing and therefore do not use vowel letters in the most of cases. According to the principle of information theory (Crystal, 2008b, p.26), "consonants carry much more information than vowels" in the pronunciation of words. This is the reason that three short vowels in the Urdu (Perso-Arabic) script, i.e. "/ə/", "/ɪ/", and "/ʊ/" are respectively represented through diacritics (*zair, zabar, paish*) only, or omitted altogether. In this way, in consonantal writing, short

vowels are perceived through the linguistic background knowledge of the concerning language and this vowel-less writing does not pose any difficulty for native speakers. Historically, this vowel free or consonantal writing has not been common for the English language, yet for languages based on the Arabic script like Arabic, Urdu, and Persian, this is a common phenomenon. Moreover, in the present times consonantal writing is even used for the English language in SMS and CMC modes of writing (see section 5.1.3).

In text messages, spelling adaptations of Romanized Urdu are not similar to spelling adaptations of English. Zurhellen (2011) states that many regular features of text messages written in English may not be the case for some other languages. He states that text messages in some languages may not contain "abbreviated material, non-standard spellings or paralinguistic restitutions" (p. 639). As bilingual texters simultaneously communicate in two different languages, so they may use different patterns than used by monolingual texters. If Romanized Urdu text messages are compared on these lines with the text messages of English, the results show that Urdu text messages employ minimum morphological clippings. Moreover, Urdu text messages contain very few number and symbol homophones.

On the whole, the corpus based variant Roman spelling patterns for Urdu have been observed and analysed from different perspectives. Based on results of the analysis, the study also includes Romanization scheme for Urdu. Hence, in the next section, the corpus based Romanization scheme for Urdu is presented.

8.4.2 Romanization Scheme for Urdu

On the basis of Romanization patterns followed by texters in SMS-Corpus, this study offers Romanization scheme for Urdu. This scheme is not intuitive in nature, rather based on the corpus of 5000 SMS collected from Pakistani texters.

Table 8.9
Roman Characters Used for Urdu in SMS-Corpus

S.No.	IPA for Urdu	Characters of Urdu	Roman Characters Used for Urdu in the Corpus
01	/b/	ب	b
02	/p/	پ	p
03	/ʈ/, /t/	ٹ - ط - ت	t
04	/s/	ص - س - ث	s
05	/d͡ʒ/	ج	j
06	/t͡ʃ/, /t͡ʃʰ/	چھ – چ	ch
07	/h/, /ʰ/	ھ – ہ - ح	h
08	/x/, /kʰ/	کھ – خ	kh
09	/ɖ/, /d/	ڈ – د	d
10	/z/	ظ - ض - ز - ذ	z
11	/ɾ/, /ɽ/	ڑ – ر	r
12	/ʒ/, /j/	ی – ژ	y
13	/ʃ/	ش	sh
14	/ɣ/, /gʰ/	گھ - غ	gh
15	/f/	ف	f
16	/q/	ق	q
17	/k/	ک	k
18	/g/	گ	g
19	/l/	ل	l
20	/m/	م	m
21	/n/, /õ/	ں – ن	n
22	/ʋ/	و	w
23	/bʰ/	بھ	bh
24	/pʰ/	پھ	ph
25	/ʈʰ/, /tʰ/	ٹھ – تھ	th
26	/d͡ʒʰ/	جھ	jh
27	/ɖʰ/, /dʰ/	ڈھ – دھ	dh
28	/ə/, /ɪ/, /ʊ/, /ɑ:/, /e/, /ɛ/, /i:/, /o/, /u:/, /ɔ/, /ʔ/	ا - آ - ع - ؤ - ہ - ئ - ے	a, e, i, o, u, y, w

Table 8.9 shows Romanization patterns for 50 characters (including 10 aspirated/murmured characters) of the Urdu language. Two characters (و), and (ی) that act as vowels as well as consonants have been given twice in the table, but their second time usage has been darkened and not counted. In this way, Romanization scheme has been offered for 50 (45 consonant) characters of the Urdu language. Moreover, the corresponding IPA scheme for Urdu characters has also been given, based

on the available studies in the field (e.g., Hussain, n.d.; Ijaz, 2007; Ijaz & Hussain, n.d.; Kabir & Saleem, n.d.).

A significant point about Romanization scheme for Urdu is that consonant characters do not pose any complication. In SMS-Corpus, only 27 patterns of Roman consonant characters (18 single characters and 9 digraphs) were found. These results indicate that specific Roman consonant characters have been consistently used for Urdu characters in the corpus. The results show that texters face no problem about the use of Roman characters for Urdu consonant characters. They consistently use same Roman consonant letters or letter combinations (digraphs) for Urdu consonant characters/alphabets as shown in the table.

On the other hand, regarding the use of Roman vowel letters for Urdu vowel sounds/characters, no consistent scheme can be offered based on the results of the corpus. Texters use Roman vowel letters for Urdu vowel letters/sounds intuitively and without consistency. This non-consistency poses complications, as texters perceive vowel letters and sounds differently, and hence a consistent scheme of vowel letters/sounds is not practically possible. A focused corpus study on the vowel patterns may offer some insights into this area, yet to offer a single consistent Romanization scheme for vowels is not practically possible. If a scheme is offered but is not actually followed by the majority of people then it is safe to abstain from such theoretical propositions. Moreover, missing vowel letters and using only consonantal spellings is also very common in text messages. This phenomenon also poses complications in defining any Romanization scheme for vowel characters/sounds.

Coming back to the use of Roman consonant letters for Urdu, it was found in SMS-Corpus that three Roman letters were not used for Urdu. These letters are "C", "V", and "X". Letter "C" was used in "ch" combination for /t͡ʃ/ (چ) sound, but not as a single

letter in isolation. Instead of "V", texters used "W" for all /v/ (و) sounds. Similarly, "X" was not used for Romanization of Urdu in the corpus.

8.4.3 Discussion on Romanization of Urdu

Urdu is conventionally written in the Arabic/Urdu script in "Nastaliq" font of Persian origin. It was occasionally written in the Roman script during the British colonial period of Indian subcontinent. Today,the use of the Roman script for Urdu has again revived in SMS and CMC modes under the impact of modern communication technologies. The impact of text messages is so pervasive that the youth of Pakistan has developed its own style and conventions of Roman spellings for Urdu. On the one hand they do not sacrifice their national language, while on the other hand they abstain from the labour of typing the Arabic/Urdu script on the small keypads of mobile phones. Hence, the Roman script is gaining more popularity among the youth in Pakistan day by day.

When the use of the Roman script for Urdu is seen in the broader perspective, it poses advantages as well as threats. The modern world has become a globalized world, and languages also need to adopt new ways to get global identity. In this regard, Robinson (2009, p. 40) highlights the "continued global dominance of the Roman alphabet" over the world and states that "South Asia in practice treats the alphabet used to write English as a universal script". In this perspective, the Roman alphabet is the global script, so languages written in this script may get more global recognition. This is the era of technology, and the Roman script has become the preferred script of modern technological devices. Therefore, to move with the pace of the modern technological development, it also is needed that one can communicate in that script. So it is a plus point for a language to be written and read in a script that is the common script of the modern developed world.

Another very significant aspect of the Roman script for Urdu is discussed here. Urdu is the twin sister to Hindi, as both are the by-products of the Hindustani language, but the geo-political issues have distanced both the languages. At present, both the languages are conventionally written in different scripts, i.e. Urdu in the Arabic/Urdu script, while Hindi in the Devanagri script. In a broader perspective, both the languages are a single language, and are mutually intelligible in the spoken form but the divide of script has made them alien to the each other. This artificial divide has minimised on the Internet as speakers of both Pakistan and India frequently use the Roman script for Hindi and Urdu on the Internet. Hence, this artificial divide has been removed at least on the Internet and speakers of both the languages are able to understand each other on the Internet.

As far as threats of the Roman script for the Urdu script are concerned, they are ideological in nature. The Arabic/Urdu script is the conventional script of Urdu and is an identity for Urdu. Languages are recognised through their scripts, so if the conventional script of Urdu is gradually replaced with the Roman script, it is like losing its identity. In the same way, the Arabic script is the gateway to the Qura'n for Muslims. If the coming generations gradually lose their grip on the conventional script of their religious books, they may also lose the first-hand knowledge of the Qur'an. One may get the same knowledge from translations of the Qur'an, but translations/interpretations are not an alternative to the first-hand knowledge. It is important to mention the Islamic belief that the Qur'an is *the Word of God* whereas translations are words of men. So words of men cannot be equal to *the Word of God*. Therefore, if the Roman script keeps promoting, the Arabic script must also be retained. We may adopt the writing script of the developed world, but not at the cost of our conventional religious script.

The practices of the Pakistani youth on the Internet and in text messages in adopting the Roman script for Urdu pose many complications. The youth is using the Roman script for Urdu without any set norms and conventions. The Roman script is being used by the youth intuitively. This has created many variations in the Roman spellings of Urdu words. These variations mostly concern vowel sounds, and hence there is the need that an effort should be made to develop some corpus based vowel patterns of the standard Roman spellings for Urdu. So if the Roman script is inevitable for Urdu due to its global impacts, efforts should be made to make this process systematic to enjoy it rather to groan on it.

CHAPTER 9

FINDINGS AND CONCLUSIONS

This is the concluding chapter of the study. It includes two key aspects of the study that are findings and conclusions. The significance and worth of any empirical study mainly depends on its findings and conclusions. In this regard, both the vital concerns of the study have been grouped in the last chapter. Findings of the study have been arranged addressing all five research questions of the study one by one. Conclusions, on the other hand, address the sum-up of all the categories of linguistic adaptations addressed in the study.

9.1 Findings of the Study

The study was based on five research questions regarding the language of text messages in the context of Pakistan. These questions are given as follows.

1) What are the major *categories/types of adaptations* in SMS language?
2) Why are the specific adaptations made? (*causes of adaptations*)
3) How are these adaptations made? (*principles/patterns of adaptations*)
4) Are these adaptations new in nature? (*adaptations in historical perspective*)
5) Are these adaptations challenging the conventional standards? (*linguistic and educational impacts/implications of adaptations*)

On the bases of analyses, results, and discussions of individual sections/chapters, this section offers key findings of the study in the light of research questions. In short, this section answers the research questions of the study.

9.1.1 Types of Adaptations

SMS language is replete with various types of adaptations at various levels. This section deals with all those categories and subcategories of adaptations that were

explored in the study. The study divided all adaptations into four major groups, i.e. 1) lexical adaptations, 2) syntactic adaptations, 3) punctuation & space adaptations, and 4) code & script adaptations. All these types/categories of adaptations have been analysed and discussed in chapters 5 to 8 of the study. In this regard, the description of these types of adaptations is given below, starting from the first category, i.e. lexical adaptations.

9.1.1.1 *Types of Lexical Adaptations*

In the light of the available literature, and the analysis of text messages, all possible lexical adaptations have been classified into certain categories. In this regard, the study has divided lexical adaptations into three broader categories.

- ***Morphological shortenings.*** Morphological adaptations are based on shortenings of words by reductions of letters. These shortenings are of three types, namely *initialism, contractions*, and *clippings*.
- ***Phonological shortenings.*** Phonological adaptations are based on shortenings of words by replacing words or parts of words with letter, number, or symbol homophones/logographs. The logographic homophones replace words or parts of words in text messages because of their phonetic/phonological approximation with the actual words, or parts of words.
- ***Spelling and case variations.*** Text messages are adapted in spellings and capitalizations through deviations in conventional norms. Therefore, even a single word is written in text messages with variant spelling and case forms.

9.1.1.2 *Types of Syntactic Adaptations*

The second major category in which adaptations are made in text messages are syntactic adaptations. In the study, syntactic adaptations have been further divided into following three categories.

- ***Word ellipses/omissions.*** Ellipses refer to the omissions of words from sentences. In the study, they have been further divided into omissions of personal pronouns, auxiliaries, copula verbs, articles, prepositions, conjunctions, and infinitives. These omissions are made in isolation as well as in combinations in text messages.
- ***Word reduplications.*** Word reduplications are normally used in the form of two or three word phrases, in which exact words or parts of words are repeated. They have been further divided into three types that are exact, rhyming, and ablaut reduplications.
- ***Grammar deviations.*** This is the third category of syntactic adaptations in text messages. They have been further divided into deviations of tense, word class, and word order in sentences.

9.1.1.3 *Types of Punctuation and Space Adaptations*

The third major category of adaptations in the study comprises punctuations and spaces. These adaptations have been further divided as follows.

- ***Deviant uses of punctuations.*** Deviations in punctuations have been further subcategorised into omissions, repetitions, and substitutions of punctuations.
- ***Creative uses of punctuations.*** In this category, punctuations function as paralinguistic and artistic devices in text messages.
- ***Deviant uses of spaces.*** Like punctuations, spaces are also used in deviant ways like omitted (under used) and unnecessarily added (over used).

9.1.1.4 *Types of Code and Script Adaptations.*

In the context of Pakistan, the bilingual texters of Urdu and English make code and script adaptations in text messages. These two adaptations have been further classified as follows.

- ***Code-mixing.*** Code-mixing is embedding of words, phrases, and clauses of one language into the other. Code-mixing has been further divided into two types, i.e. English to Urdu code-mixing, and Urdu to English code-mixing.
- ***Code-switching.*** Code-switching involves the swing from one to the other language on sentence boundaries and is also termed as inter-sentential switching. In text messages, this occurs at three levels that are Urdu to English, English to Urdu, and two-way in both contact languages (within single discourses).
- ***Romanization of Urdu.*** Pakistani texters mostly write Urdu text messages in Romanised Urdu, and hence adapt the Roman script for their national/local language. In this regard, the analysis was conducted on variations of Roman spellings for the Urdu language.

9.1.2 Causes of Adaptations

The second research question of the study was about exploring the causes of adaptations. In this regard, the study is based on the perception that adaptations in text messages are caused by various factors. Therefore, all major causes of linguistic adaptations in text messages were determined on the bases of metalinguistic as well as linguistic analysis. The study found that most of intentional adaptations in text messages are caused by three dominant factors that are 1) economical use of time and effort, 2) creative/innovative purposes, and 3) to be rapid in SMS communication. Sometimes, the careless attitude of texters about the conventional standards of language, or the poor command on language rules becomes the cause of many unintentional adaptations.

Besides these general causes of adaptations, some adaptations have their specific causes like some punctuation adaptations are made for paralinguistic purposes. Code adaptations are made because Pakistani bilingual texters frequently swing between Urdu and English. Similarly, script adaptations are made because the most of Pakistani texters

are not adept in typing the Arabic/Urdu script and hence they prefer to use the Roman script for Urdu.

After this cursory view of the major causes of linguistic adaptations in text messages, the category wise causes of all adaptations are given below.

9.1.2.1 Causes of Lexical Adaptations

Many studies in the field show that texters are tilted to use lexical adaptations (Baron, 2008; Bodomo, 2010; Crystal, 2008b; Hard af Segerstad, 2002; Tagg, 2009). Hence, the extensive use of lexical adaptations is not without reasons. To address this issue, reasons of lexical adaptations have been offered here.

The first category of lexical adaptations in the study is of morphological shortenings. Morphological adaptations like initialism, contractions, and clippings are mainly done to save time and effort of texters. Only a few people believe that the small keypad of mobile phones is a major cause of the telegraphic style in text messages. Texters know that SMS is an informal medium of communication and there is no need to follow conventions and standards. They think that if their recipients can understand them, this is enough for them. Hence, through morphological adaptations, they clip/cut parts of words so that they have to write minimum to convey maximum.

In phonological adaptations, letter, number, and symbol homophones of words act as logographs. The purpose here is also the same that is saving time and effort. A few texters also stated that they use phonological adaptations for style and space saving. Hence, texters employ logographs due to the approximation of their sound/pronunciation with actual words, or parts of words.

Most variations of spellings and capitalization in SMS orthography occur to save time and effort, while in some cases texters want to adopt a different style for the purpose of creativity.

9.1.2.2 Causes of Syntactic Adaptations

In the study, all three categories of syntactic adaptations have been attributed to three different causes. Word ellipses occur due to economical reasons. When texters want to save time and effort, they use the telegraphic style. They shorten sentences by omitting less important words, and keeping only key words which are needed for the transmission of information. When the question comes of the artistic use of language they overrule the first rule. So in adaptations of reduplications, texters embellish text messages with repeated words.

Regarding the grammatical deviations in text messages, both the causes of economical and creative/innovative uses of language do not fit the situation. Most of these deviations occur due to lack of texters' command on rules of language, besides their careless and speedy manner in typing text messages.

9.1.2.3 Causes of Punctuation and Space Adaptations

In the study, the first kind of punctuation adaptations is of omissions. In this regard, texters omit punctuations as they want to save time and effort. Texters believe that the absence of punctuation marks does not alter any meaning, and no information is lost due to their absence. The second kind of punctuation adaptations is of repetitions. These repetitive adaptations are usually made to emphasize something. The third category of deviations is the substitution of punctuations. In this category, texters use deviant punctuation marks instead of conventional punctuation marks. The substituted/altered use of punctuations occurs due to the careless attitude of texters. Sometimes these deviations of substitutions occur due to the lack of linguistic knowledge of texters. In addition to these causes, punctuations are also used in deviant ways for creative purposes like paralinguistic and artistic devices.

As far as space adaptations are concerned, they are made in two categories that are omissions and additions of spaces. This study suggests that both deviated types of spaces occur due to the carless attitude of texters. Mostly, they occur unintentionally and these unintentional uses of deviant spaces cannot be attributed to economical or creative reasons.

9.1.2.4 Causes of Code and Script Adaptations

Pakistan is a multilingual/bilingual country where English is the official language, while Urdu is the national language of the country. Both the contact languages are used in speech as well is writing. In speech, it is a universal phenomenon that when speakers are adept in both contact languages, they code-mix and code-witch from one language to the other in different situations.

Although, code-mixing and code-switching are actually the spoken language phenomena yet they can penetrate to writing. In the case of text messages, the possibility of code adaptations is much more than any other written mode because text messages are the hybrid of speech and writing. Therefore, code adaptations occur in Pakistani text messages due to the bilingual background of texters.

When the issue comes to script adaptations, Urdu and English both have their own conventional scripts, which are Arabic/Urdu and Roman respectively. Many mobile phone sets provide Urdu alphabets/characters, but the majority of Pakistani texters are not adept in typing of the Arabic/Urdu script. So they prefer to write Urdu text messages in the Roman script. Moreover, they code mix/switch in both languages very frequently, and hence they prefer to use only the Roman script for both languages.

9.1.3 Principles/Patterns of Adaptations

The third research question of the study was about exploring the principles and patterns of adaptations made in text messages. In this regard, many linguists and

researchers in the field of CMC and SMS are of view that linguistic adaptations in text messages are not haphazard, and are made under certain principles and patterns (Baron, 2008; Bodomo, 2010; Crystal, 2004, 2008b; Hard af Segerstad, 2002; Ling, 2005, 2008; Tagg, 2009; Thurlow & Poff, 2011). Therefore, this study strived to explore principles and patterns of various linguistic adaptations in text messages. All such principles/patterns have been discussed in detail in the relevant sections of the thesis, and the key findings are given here.

9.1.3.1Principles/Patterns of Lexical Adaptations

Lexical adaptations in the study have been divided into three categories, and principles/patterns of these adaptations are given here.

- ***Morphological shortenings.*** For the analysis of morphological adaptations, all morphological shortenings were categorised into three types that are initialism, contractions, and clippings.
  - *Initialism.* Shortenings of initialism are made by dropping all letters/alphabets from words in phrases and sentences, leaving only initial letters. Adaptations of *initialism* are made in two ways that are *alphabetism* and *acronymism*. The classification is based on the pronunciation of words. If shortened words are pronounced as words like UNESCO or NATO, they are called acronyms, and if they are pronounced as alphabets like UN or USA, they are named alphabets. In SMS-Corpus, initials for phrases were mostly found, but for sentences they were rare.
  - *Contractions*. Two types of contractions are frequently made in text messages, which are single-word, and two-word contractions. In the corpus, it was found that the apostrophe mark was rarely used, which is otherwise an essential component of the conventional way of contractions.

- *Clippings.* Three types of clippings are mostly made in text messages that are *back clipping* (where ending part of a word is dropped), *fore-clipping* (where beginning part of a word is dropped), and *middle clipping* (where both beginning and ending parts of a word are dropped). Results of SMS-Corpus supported the opinion of Crystal (2008a), as back clipping was at its maximum, followed by fore-clipping and middle clipping respectively. Another aspect of clippings was also observed in the light of Crystal (2008b) that in the most of clippings vowels are dropped much more than consonants, because vowels carry much less lexical, semantic, and phonological information than consonants.

➢ ***Phonological shortenings.*** Phonological shortenings are made when *letter*, *number*, and *symbol* homophones/logographs replace words or parts of words due to the resemblance of their sounds/pronunciation with words or parts of words.

- *Letter logographs/homophones.* Out of three categories of phonological shortenings, letter logographs are mostly used in text messages. Letters like "u", "k", "r", "b", "m", and "n" have been mostly used in SMS-Corpus to replace words or parts of words, because of their pronunciation/sound approximation with words or parts of words.
- *Number logographs/homophones.* The second major occurrence of phonological shortenings in text messages is of number homophones or logographs. In SMS-Corpus, four number logographs/homophones of words, or parts of words were found. These are numbers "2", "4", "8", and "9". These four numbers have been used in various ways like "2 for too, and to", "4 for b4/before", "8 for w8/wait", "9 for 9t/night" and so on. Other numbers

are used as logographs of their corresponding numbers only like "1" for "one".

- *Symbol logographs/homophones.* Symbol/sign logographs are the least used category of all three phonological shortenings in SMS-Corpus. Moreover, there are only four symbol logographs in SMS-Corpus, which have been used in their real sense. These are "&", "%", "#", and "@" symbols and they have been used in their conventional sense as symbol logographs for words "and", "percent", "number", and "at" respectively. The other six symbol logographs "§", "$", "£", "¥", "»", and "«" are conventionally used as logographs, but in SMS-Corpus they have been used for other creative purposes, and not as logographs (see section 5.3.3). However, as a rule, all 10 symbol logographs shown in results, can replace their corresponding words, being homophones of those words.

➢ ***Spelling and case variations.*** Findings related to variations/ deviations of spellings and letter cases are given as follows.

- *Variations of spellings.* Results of variations of spellings show that the most of variations of spellings are made through two processes, which are morphological and phonological shortenings. Morphological shortenings are made through initialism, contraction, and clipping, while phonological shortenings are made through letter, number, and symbol homophones. All these variations are deliberately made by texters. There are only a few negligible instances in SMS-Corpus where inexplicable spellings have been used due to typographical mistakes, or poor spellings of texters.
- *Variations of letter cases.* In SMS-Corpus, three types of letter cases have been used, that are "all upper", all lower", and "upper and lower mixed"

cases. As far as "all upper" and "all lower" case words are concerned, they are normally due to the ease of typing, that means if screen-window is set at lower case, texters will use "all lower" case, and if it is set at upper case they will use "all upper" case. Sometimes, upper case is used to emphasize on some words. For creative and stylish purposes, texters employ "mixed upper and lower" cases in unusual but appealing ways.

Hence, findings of this chapter show that lexical adaptations in text messages are made under certain rules and patterns. Typographical issues are also involved in some of lexical adaptations, but their frequency is very low, in comparison to the rule governed adaptations.

9.1.3.2 Principles/Patterns of Syntactic Adaptations

The second major category of adaptations in text messages covered in the study is syntactic adaptations. Findings regarding principles of various syntactic adaptations are given in this section.

- ***Word ellipses/omissions.*** Word ellipses have been further sub-categorized into certain types. These types with their underlying principles are discussed as follows.
  - *Omissions of personal pronouns.* Results show that in omissions of personal pronouns, the first person pronoun "I", in subjective case, was omitted the most of time. The other two personal pronouns, "You" and "We" were also omitted but their frequency was low. In the same way, first and second person pronouns in objective case were also omitted with very low frequency. Overall, SMS language is a spoken like language that omits first and second personal pronouns, mostly in the subjective case.

- *Omissions of auxiliary and copula verbs.* Results show that both auxiliary (helping) verbs and copula (linking) verbs are omitted in text messages. Significantly, among all auxiliary, modal, and copula verbs, the verb form "be", in three present tense forms that is "is, am, are" is omitted the most.
- *Omissions of articles.* Among three articles "a, an, the", the most frequent ellipses occur in omissions of the definite article "the".
- *Omissions of prepositions, conjunctions, and infinitives.* Results show that conjunctions and prepositions are two other word classes that are frequently omitted in text messages. Infinitives are also omitted but in low frequency.
- *Combined multiple word ellipses/omissions.* In combined multiple word ellipses, text messages are communicated through single or two words only. In text messages, omitted words out number written words. Hence, combined omissions of multiple words are instances of telegraphic and pragmatic uses of text messages, by employing syntactic adaptations in artistic and adept ways. Results show that all three main parts of a simple sentence that is "subject", "verb", and "object" can be easily omitted by texters. Moreover, at the word level, "nouns"," pronouns", "auxiliary verbs", "copula verbs", "main verbs", "articles", and "prepositions" are frequently omitted as combined multiple ellipses.

➢ ***Word reduplications.*** Underlying principles of word reduplications are given below.

- *Exact reduplications/triplications.* Exact reduplications are used in both Urdu and English text messages. Pairs like "please-please", "many-many", "happy-happy", "bye-bye", "waiting-waiting", "very-very", and "ever-ever" are commonly used in English. In SMS-Corpus, the most frequent reduplication

is of "many-many" (seven times), followed by "plz-plz" (six times). Other reduplications normally occurred once or twice. Three words "plz", "very" and "ever" were even used in triplications. In Urdu, reduplications are more common than they are in English. In Urdu, words of English are also used as reduplicated words like "plz-plz", and "sweet-sweet".

- *Rhyming reduplications.* Rhyming reduplication is the process of repeating a word with the replacement of a consonant sound. Normally, the replaced consonant is the first consonant of the word. The results of the study show the rhyming reduplication in both English and Urdu. In English, two pairs "sooper-dooper" and "never-ever" were found. For Urdu, word pairs "aisy-jaisy", "wesay hi-kesay hi", and "dukh-sukh" were found.
- *Ablaut reduplications.* Ablaut reduplication is the process of repeating a word with the replacement of a vowel sound. Normally, the replaced vowel is somewhere in the middle of the word. The common examples of ablaut reduplication in the English language are "chit-chat", "ding-dong", and "zig-zag" etcetera but no example could be found for the English language in the corpus. In Urdu, common examples are like "theek-thaak", "fit-faat", and "shopping-shupping" etc. Among all three categories of reduplications, this is the least occurring category in both languages, so its occurrence in SMS-Corpus was also very low.

➢ ***Grammar deviations.*** Principles of various categories of grammar deviations are given as follows.

- *Deviant tense.* In results, two types of variations in the use of tenses have been shown, that are deviations in the conventional uses of tenses, and their omissions where needed. In the corpus, instances of deviations of main verbs

are more than auxiliary verbs. Overall, these deviations distort meanings of sentences, and in formal language such deviations of verbs/tenses do not fit on any conventional standards. Hence, such sentences with deviant verbs in text messages are termed as ungrammatical.

- *Deviant word class/form.* There are two types of word class deviations that are intentional and unintentional. A typical example of intentional uses is the first person subjective pronoun "I" that is replaced with the first person objective pronoun "me". The other uses of deviant word class in the corpus are either due to carless attitude, or the lack of the command of texters on rules of the English language. Whatever the reason, these unexpected deviations in words fall under the category of grammatical errors.
- *Deviant/variant word order.* There is a standard sentence structure in every language that defines the order of words in a sentence to make a sense in that language. Yet, this aspect of word order in text messages is compromised either intentionally or unintentionally. Most instances in the corpus in this category are of deviations of word order between the subject and the auxiliary verb in interrogative constructions.

9.1.3.3 Principles/Patterns of Punctuation and Space Adaptations

The third major category of adaptations studied in text messages is punctuations and space adaptations. Their findings regarding principles/patterns are given below.

- ***Deviant uses of punctuations.*** Deviant uses of punctuations have been further sub-categorized into three types as shown below.
  - *Omissions of punctuations.* In text messages, punctuation marks are omitted where they are conventionally necessary. There are various examples in the corpus where no punctuation mark has been used at all. In the corpus, four

very crucial punctuation marks that are period/full stop (.), question mark (?), exclamation mark (!), and apostrophe (') have been frequently omitted.

- *Repetitions of punctuations.* In text messages, same punctuation marks are repetitively used by deviating from the conventional uses of punctuations. There are examples of repetitive uses of dots (.), questions marks (?), and exclamation marks (!). Normally, the repetitive use of punctuation marks is intentionally employed by texters for emphasis.
- *Substitutions of punctuations.* In the corpus, there are many instances of deviant/inappropriate uses of punctuation marks. For example, question marks have been replaced with periods, periods with questions marks, apostrophes with commas, and exclamation marks with dashes.

➢ ***Creative uses of punctuations.*** Principles and patterns of creative uses of punctuations as paralinguistic and artistic devices are shown below.

- *Punctuations as paralinguistic devices.* Paralinguistic devices are those methods that can be used to convey feelings and emotions through expressions, and not through words. In the written language, such emotions are conveyed through expression markers known as *emoticons* (emotion-icons), like (☺ for happy) or (☹ for sad), but punctuation marks can also be used to manually develop them. In the corpus, various emoticons have been created through punctuation marks like the body of a man, a man releasing his stool, a face in three moves, a face in four moves, the head of famous Pakistani poet Faraz, a man praying with raised hands, a man with turned ears to the left side, and three facial emoticons to wish good morning (see Table 7.6).

- *Punctuations as artistic devices.* Another category of creative uses of punctuation marks is their use as artistic devices. In the corpus, various symbols and punctuation marks have been used to create artistic patterns to convey nonverbal meanings like the name of Allah (ﷻ), eid greetings, good morning, happy new-year, a house glittering with stars, and images of mosques (see Table 7.7). Overall, artistic patterns created through punctuation marks are very beautiful and attractive.

➢ ***Deviant uses of spaces.*** A space is a blank ( ) area where there is no written content. Different languages have different rules of spacing that suit their respective writing scripts. The Roman script for English applies the insertion of one space to separate words but in text messages deviations of spaces do exist in the following ways.

- *Omissions of spaces (under spacing).* The insertion of a space at word boundaries is conventional to the English language. Nevertheless, examples (in Table 7.8) show that various words have not been separated from neighbouring words with spaces. These omissions of spaces are grammatically unacceptable.
- *Additions of spaces (over spacing).* The excessive insertion of spaces is called over spacing. In examples, two types of over spacing have been shown that are "within word boundaries" and "within a discourse boundary". In text messages, both single and compound words are occasionally broken with a space within word boundaries. In the same way, sentences are separated from each other with the line spacing within discourses and it is a common practice in text messages (see Table 7.9).

9.1.3.4 Principles/Patterns of Code and Script Adaptations

The fourth major category of linguistic adaptations analysed in the study concerns code and script adaptations. Principles/patterns of these adaptations are given as shown below.

- ***Code-mixing.*** Code-mixing is the borrowing of words, phrases, and clauses from one contact language to the other. In text messages, both Urdu and English act as frame as well as guest languages. Overall, nouns and noun phrases are the most borrowed items in both categories.
  - *English code-mixed into Urdu.* The most occurring category of code-mixing in text messages in Pakistan is the mixing of English into Urdu. This occurs at word, phrase, and clause levels within sentence boundaries. Code-mixing of single words of English into Urdu occurs in six categories that are nouns, adjectives, verbs, adverbs, conjunctions, and interjections. Code-mixing of English phrases into Urdu occurs in four categories of noun, verb, adjectival, and prepositional phrases. At the clause level insertions, only the dependent (sub-ordinate) clauses have been code-mixed in Urdu from English. Moreover, code-mixing of clauses occurs at the clause boundary, mostly in the presence of some complementizer (subordinating conjunction) like "if", "that", "and" etcetera.
  - *Urdu code-mixed into English.* Code-mixing of Urdu into English occurs at three levels that are, word, phrase, and clause levels. Code-mixing of single words of Urdu into English occurs in four categories of word classes that are nouns, adjectives, conjunctions, and interjections. At the phrase level, noun phrases are found in abundance. One example of postpositional phrase (i.e., prepositional phrase in English) was also found in the corpus. Overall, Urdu

into English code-mixing is not as rich as is English into Urdu code-mixing. At the clause level, only the dependent (sub-ordinate) clauses are code-mixed into English from Urdu. Moreover, this code-mixing of clauses also occurs on the clause boundary, mostly in the presence of a complementizer (subordinating conjunction) like "agar (if)", "and" etcetera.

- ***Code-switching.*** In text messages, code-switching occurs in three ways that is from Urdu to English, English to Urdu, and two-way in both languages. Myers-Scotton (2006) divides code-switching into two types that are "classic" and "composite" code-switching. In the classic code-switching, main grammatical (morpho-syntactic) frame is offered by the matrix language, whereas in the composite code-switching both languages participate equally.
  - *Urdu to English code-switching.* Inter-sentential code adaptations from Urdu to English are very common in text messages. This phenomenon has been elaborated through examples of one-way switching that is from Urdu to English. In all examples, code-switching was applied on/outside the sentence boundary. The first utterance was in Urdu, whereas the following utterance was in the English language.
  - *English to Urdu code-switching.* This category of code-switching is less common to the previous category. Like the previous section, the code-switching of this type is applied outside sentence boundaries. In all examples, first utterances were in English, whereas the following utterances were in Urdu.
  - *Two-way code-switching.* This type involves two-way swing from one language to the other and then back way again. The two-way code-switching

is although less common in the corpus, yet examples show that this kind exists in text messages.

- ***Romanization of Urdu.*** Urdu, the national language of Pakistan, is traditionally and academically written in the Urdu script, which in turn is the modification of the Arabic script. On the other hand, Romanization is the process of representing a language in the Roman script, while the conventional writing script of that language is someone else. Today, the abundant use of the Internet and mobile phone technologies has led masses to use the Roman script for Urdu. In this regard, the present study explored the following issues of Roman Urdu.
  - *Variations of spellings in Roman Urdu.* To study the corpus based Romanization patterns adopted for Urdu in text messages, the study analysed 50 variant spellings of the most frequent Urdu words in SMS-Corpus. The results of the study show that variant Roman spellings have been used to represent any single word/sound of Urdu. Similarly, one spelling combination can represent more than one Urdu word/sound. Regarding consonants, texters consistently use specific consonant letters/digraphs for specific consonant sounds/phonemes of Urdu, and therefore consonants create no complications. The problem mainly concerns vowel sounds (both pure and nasal), because one single vowel letter can represent various vowel sounds, and one single vowel sound can be represented by various vowel letters. Moreover, some texters use the consonantal style of writing and therefore omit vowel letters in the most of cases.
  - *Romanization scheme for Urdu.* On the basis of the analysis, this study offers Romanization scheme for Urdu as found in SMS-Corpus. This scheme is not intuitive in nature because it is based on a corpus of 5000 SMS of Pakistani

texters. Table 9.1 shows Romanization patterns for total 50 Urdu characters (including 10 aspirated/murmured characters) of the Urdu language. Two characters (و) and (ی) used in two ways (vowels as well as consonants) have not been counted for their repeated use in the category of vowels. Moreover, the prevalent IPA scheme for Urdu characters/alphabets has also been given.

Table 9.1

Roman Characters Used for Urdu in SMS-Corpus (Table 8.9 condensed)

S. N.	Urdu IPA	Urdu Characters	Roman Characters	S. N.	Urdu IPA	Urdu Characters	Roman Characters
01	/b/	ب	b	15	/f/	ف	f
02	/p/	پ	p	16	/q/	ق	q
03	/ʈ/, /t/	ٹ - ط - ت	t	17	/k/	ک	k
04	/s/	ص - س - ث	s	18	/g/	گ	g
05	/d͡ʒ/	ج	j	19	/l/	ل	l
06	/t͡ʃ/, /t͡ʃʰ/	چ - چھ	ch	20	/m/	م	m
07	/h/, /ʰ/	ہ - ھ - ح	h	21	/n/, /õ/	ں - ن	n
08	/x/, /kʰ/	کھ - خ	kh	22	/ʋ/	و	w
09	/ɖ/, /d/	ڈ - د	d	23	/bʰ/	بھ	bh
10	/z/	ظ - ض - ز - ذ	z	24	/pʰ/	پھ	ph
11	/ɾ/, /ɽ/	ڑ - ر	r	25	/ʈʰ/, /tʰ/	ٹھ - تھ	th
12	/ʒ/, /j/	ی - ژ	y	26	/d͡ʒʱ/	جھ	jh
13	/ʃ/	ش	sh	27	/ɖʰ/, /dʰ/	ڈھ - دھ	dh
14	/ɣ/, /gʰ/	گھ - غ	gh	28	/ə/, /ɪ/, /ʊ/, /ɑ:/, /e/, /ɛ/, /i:/, /o/, /u:/, /ɔ/, /ʔ/	ا - آ - ع - ِ - ُ - ہ - ی - ے	a, e, i, o, u, y, w

The above given Romanization scheme for Urdu consonants is corpus based and is therefore practicable for use. On the other hand, no consistent Romanization patterns for Urdu vowel sounds/characters were found in the corpus. Texters use Roman vowel letters for Urdu vowel letters/sounds intuitively and without any consistency. This non-consistency poses complications, as texters perceive vowel letters/sounds differently, and hence a consistent scheme of vowel letters/sounds is not practically possible. Moreover, missing vowel letters and using consonantal spellings is also very common in text messages. This phenomenon also poses complications in defining any Romanization scheme for vowel characters/sounds.

Regarding the use of Roman consonant letters for Urdu, it was found in the corpus that there are three Roman letters that were not used for Urdu. These letters are "C", "V", and "X". Letter "C" is used in "ch" combination for /t͡ʃ/ (چ) sound, but not as a single letter in isolation. Instead of "V", texters used "W'' for all /ʊ/ (و) sounds. Similarly, "X" is not used for Romanization of Urdu in the corpus.

9.1.4 Adaptations in Historical Perspective

The fourth question of the study was to find out, in the light of key figures and key works in the field, whether adaptations of text messages are totally new forms of language, or we have traces of such forms in history. In response to this question, findings related to different adaptations have been accumulated here.

9.1.4.1 Lexical Adaptations in Historical Perspective

Many key figures in the field of linguistic analyses of text messages assert that literary history is full of such examples where traces of lexical adaptations, as found in text messages, can be seen (Baron, 2008; Bodomo, 2010; Crystal, 2008b). At the morphological level, shortenings like initialism (alphabetism plus acronymism), contractions, and clippings are not new forms of language. The only difference is that in the past these forms were rarely used in the formal language, whereas the informal language was not much noticeable. Now, these features of the informal language have also entered text messages, and their use has become abundant.

In the historical perspective, phonological shortenings like letter, number, and symbol homophones are not new forms of language. This question has been answered by Crystal (2008b, p.41), while he declares that "there is actually nothing novel at all about such text messages as c u l8r . . . . They are part of the European ludic linguistic tradition, and doubtless analogues can be found in all languages which have been written down". Regarding the historical tradition of uses of letters and numbers as logographs or

phonetic approximants of letters, he asserts that generations of human beings have used such trends of language in the past. So, we should not be taken aback when we come across such uses of language in text messages.

As far as spelling deviations in text messages are concerned, they are of two types, which are unintentional and intentional. All unintentional/unconscious deviations occur because of the lack of concentration and are termed as typographical errors. Crystal (2008b, p.49) defends intentional adaptations by saying that texters mostly manipulate spellings instead of committing mistakes. He supports the use of deviant spellings by stating that "several of these nonstandard spellings are so much part of English literary tradition that they have been given entries in the Oxford English Dictionary. *Cos* is there from 1828, *wot* from 1829, *luv* from 1898, *thanx* from 1936, and *ya* from 1941."

As far as unintentional deviations in spellings are concerned, Crystal (2008b) believes that variations/deviations are acceptable if they are appropriate to those situations in which they are used. Even a famous grammarian F. R. Palmer (1983, p.16), in his well acclaimed book "Grammar" accepts that "some forms of language are acceptable in certain situations . . . ." Furthermore, he says that "we break the conventions at our peril, we are dubbed 'ignorant', we fail to get the job we hoped for; but we ought not to provide pseudo-linguistic grounds to justify the conventions" (p.26).

Views of Crystal (2008b) and Palmer (1983) confirm that variant forms of language, whether spellings or cases, are acceptable if they are appropriate to the situation in which they are used. Conventions and standards are right at their own place, but variations/deviations cannot be banned everywhere, otherwise languages will cease to grow.

9.1.4.2 Syntactic Adaptations in Historical Perspective

Modern communication technologies like the Internet and mobile phones have significantly influenced human languages (Baron, 2008; Bodomo, 2010; Crystal, 2004, 2008; Hard af Segerstad, 2002). The mobile phone based SMS technology is the latest way that has influenced and modified conventions of languages over the globe. At the syntactic level, text messages have adapted SMS language in many ways. Regarding shortened sentences in text messages, Frehner (2008) states that syntactic word ellipses in text messages occur in such a large amount that they surpass word ellipses in any other type of written communication, and even the oral communication.

The process of reduplication found in syntactic adaptations is indeed not a new or modern phenomenon in languages. In the English language, reduplicated words like 'riff-raff' and "willy-nilly" existed even centuries ago. Moreover, reduplications of words are found in many languages of the world in the spoken language. Afrikaans "krap-krap-krap" (to scratch forcefully), Turkish "tabak-mabak" (plates/dishes etcetera), Napalese "khana-sana" (meal etcetera), and Persian "dava-mava" (argument etcetera) are a few examples of reduplications in some other languages. Therefore, the phenomenon of reduplication is neither new, nor specific to a certain language. It is a common phenomenon of many languages of the world, in informal and spoken settings. As text messages mostly use an informal and spoken like language, so this feature of the spoken language has entered the written mode of text messages.

Regarding grammatical deviations in text messages, Bodomo (2010) states that deviations of grammar are also found in online communication, and hence text messages share this feature with CMC mode.

9.1.4.3 Punctuation and Space Adaptations in Historical Perspective

Many studies on unconventional uses of punctuation marks have been conducted in text messages. In this regard, Baron (2008) investigated different aspects of sentence punctuations in texting and instant messaging. Bosco (2007) conducted a study on the language of text messages in Hong Kong, and found that punctuations in text messaging have a high degree of creativity. A very prominent feature of punctuations in text messages is their use as emoticons. About the first use of emoticons, Naomi S. Baron (n.d.) traces their uses in 1963, when a graphic artist Harvey R. Ball created a yellow button based smiling face for the State Mutual Life Assurance Company of America. She also states that on September 19, 1982, Scott E. Fahlman sent the first message carrying two emoticons to an online bulletin board at Carnegie Mellon University.

In this way, punctuations are used to add features of non-verbal communication in text messages. In the historical perspective, non-verbal communication is not new to human beings. The first scientific study of nonverbal communication was of Darwin (2009), in his 1872/1890 classis book "The Expression of the Emotions in Man and Animals". He argued that all mammals show their emotions on their faces. Many years later Tomkins (1962, 1963, & 1991) conducted his classic studies (1962-1991) on human emotions. Another contribution towards the nonverbal communication was of Birdwhistell (1985). He stated that all body movements have meanings and all forms of paralanguage have a grammar that can be analyzed in similar terms to the spoken language.

Hard af Segerstad (2005) also states that emoticons are used to enrich alphabetic writing by conveying moods or emotions that are normally expressed with extra-linguistic cues such as facial expressions and tone of voice in spoken interaction. The experienced communicator seems to know that some messages may need additional

information to clarify word-only communication. Regarding space adaptations in text messages, Hard af Segerstad states that "similar to the strategy to omit punctuation, by omitting space between words the user saves keystrokes as well as time and effort" (p. 1). She further states that in some cases it is not necessary to save space, but omitting it renders a personal tone to the messages of individual texters. Therefore, the adaptations of punctuations and spaces in text messages are not totally a new phenomenon, and they have traces in history for their variant uses.

9.1.4.4 Code and Script Adaptations in Historical Perspective

In bilingual societies, it is a very common practice that speakers code switch from one contact language to the other. As Pakistan is a bilingual/multilingual country, so swinging between Urdu and English is a very common phenomenon in speech. Even in written modes one can find instances of code-mixing (Rasul, 2009). Moreover, the mode of SMS is already a proven hybrid of speech and writing (Baron, 2008; Crystal, 2008b), so one can easily expect code-switching/mixing in text messages in a bilingual society.

In the context of Pakistan, the historical position of the region as a former British colony has provided a firm ground for Urdu-English language contact. Hence, historical as well as modern global impacts of English in Pakistan have resulted in the continuous language contact of Urdu and English, and this contact leads to code-switching and code-mixing. Rasul (2008) states that due to the hybridization of Urdu and English, a new code "Urdish" has emerged in Pakistan.

The other related concept with code adaptations is script adaptations in text messages. In the written language, when two languages are frequently code-switched, it is practically not possible that half utterances should be written in one script and half in the other. Urdu is traditionally written in the Urdu script which is a modified form of

Perso-Arabic script. But texters in Pakistan mostly use the Roman script for Urdu. The result of this continuous use of the Roman script in text messages is that there is the revival of Roman Urdu that was once a colonial need in the British colonial period for official purposes.

9.1.5 Impacts and Implications of Adaptations

The fifth and last research question of the study was to explore linguistic and educational impacts and implications of language adaptations found in text messages. To address this issue, three sources were utilized that are 1) available literature on the topic, 2) linguistic analysis of SMS-Corpus, and 3) metalinguistic analysis of texters' perceptions. Findings in this regard are offered in two categories, i.e. 1) linguistic impacts and implications, and 2) educational impacts and implications of adaptations.

9.1.5.1 Linguistic Impacts and Implications of Adaptations

This section deals with the first category that is linguistic impacts and implications of adaptations in text messages. It has been further divided into impacts and implications.

- ***Linguistic impacts of adaptations.*** Linguistic impacts of adaptations made in text messages deal with various issues. Among them, an important linguistic impact of adaptations of text messages is on the orthography of English. Since the advent of the print media in Britain, English spellings took centuries to reach present standard forms, but orthographic adaptations of text messages and the Internet have challenged these standards of orthography/spellings at the level of masses. Most of adaptations found in text messages are a part of the English literary history since centuries, but their use in the past was mostly limited to linguistic books and dictionaries. But mobile phones and the Internet have released this

genie of books from the bottle, and now it is not possible for anyone to push it back into the bottle.

So adaptations have influenced the English language in many ways. At the orthographic level, there are morphological shortenings like word initials, clippings, and contractions. The use of morphologically shortened words was very limited in the past but now texters use them frequently. The substitution of full words or parts of words with letter, number and symbol homophones is another impact of SMS adaptations on English orthography. Many languages of the world, both of present and past, are logographic in nature, for example Chinese and Japanese languages, but text messages have even pushed logographic elements into the English language. Moreover, the use of deviant forms of letter cases is also the impact of SMS adaptations on the English language.

At the syntactic level, SMS adaptations have influenced the English language by abundantly omitting essential words from sentences like the first person pronoun "I", auxiliary and copula verbs, articles, prepositions, conjunctions, and infinitives. Similarly, the use of reduplications and triplications for emphasis and artistic purposes is common in text messages, which were actually features of speech. Many deviations of grammar like deviated use of tenses, word class/form, and word order, which were strictly admonished by prescriptive grammarians, are now commonly used by texters.

Texters also adapt punctuations and spaces, and linguistic codes and scripts in bilingual settings, for various purposes. All these adaptations are impacts on the standard forms of human languages. Code adaptations have impacts on the written mode of language in a bilingual setting. Before the advent

of CMC and SMS modes of communication, code-mixing and cod-switching were never used in such abundance in any written mode. Similarly, today the use of the Roman script for many languages of the world is also the impact and a challenge of the Internet and text messages on conventional scripts of many languages of the world.

- ***Linguistic implications of adaptations.*** As there are linguistic impacts of SMS adaptations on language, similarly, certain linguistic implications were also inferred from these adaptations. Firstly, SMS adaptations have implications for the English language orthography. English orthography has been under certain standards since centuries, but now those standards have been challenged. Now we have reached a cross road of history, where we can neither accept these orthographic adaptations nor reject them. We cannot accept them because the standard orthography of the English language does not allow for these transgressions. We achieved and preserved these orthographic standards after the voyage of centuries. We cannot reject adaptations because they are the natural outcome of the modern uses of language under the influence of modern information technologies. Moreover, languages have to accept changes otherwise language forms found in standard books do not match actual language forms used by people. Therefore, the modern orthographic forms used in abundance must be given some space to exist, grow, and reform. We should not scold the youth for these adaptations, if they restrict their uses to the appropriate mediums like SMS and CMC modes (Crystal, 2008b). On the other hand, the youth have to wait till the time when the most used forms of register specific SMS/CMC spellings are given some weight in the formal language, and are accepted in language books and dictionaries.

Same is the case with other adaptations like syntactic, punctuation, code, and script adaptations. In these categories, there are two types of adaptations that are intentional and unintentional. Unintentional adaptations are actually the outcome of careless behaviour of texters. The deviations made by careless behaviours or due to the lack of command on language are actually very few in comparison. Such unintentional deviations fall under the category of errors and flaws of texters and even the descriptive linguists would not accept them because they do not count among the most frequently occurring adaptations. Hence, the intentional and the most frequently occurring adaptations which are made by the majority of texters under certain principles and patterns may not be scolded if they are used in mediums which are appropriate for their use.

Regarding code adaptations, it is pertinent to mention that they are the feature of the spoken language, and are used by a majority of people in bilingual societies. In written forms, their use was very limited in the past but SMS and CMC modes of communication have even allowed this transgression. Hence, code adaptations should also be accepted as a part of the written language in addition to the spoken language. Regarding the use of the Roman script for Urdu, it is not a new phenomenon for the Urdu language. Urdu used the Roman script in the British Raj in India, so why to abstain from it now, when many other languages of the world are also adopting it under the impact of modern communication technologies. Today, even those languages are using it, which have never used it in the past. This script has now become a global script, so we cannot and should not escape from this script. This is the script that has introduced us with the modern knowledge and developments. Mustafa Kamal of

Turkey adopted the Roman script in 1928, and that change opened many avenues of development for Turkey.

On the other hand, as far as the importance of the Arabic/Urdu script is concerned it cannot be denied. We should try preserving our own conventional script by bringing more and more world knowledge into the Urdu language that hosts this script. If the modern knowledge is brought into Urdu, it will become a motivation for us to learn and use the Urdu script. Empty slogans and emotional attachments are not enough to promote the Urdu script. Moreover, we should try to make the Urdu script technology-compliant and technology-friendly under Unicode standards, so that everyone can use it easily on the platforms of modern communication technologies. At present, efforts of computational linguists like Sarmad Hussain (n. d.) and Attash Durrani (2008) are a breakthrough in the field. National Language Authority should also play its role to modernise the Urdu script.

In a nutshell, it will be accurate to state that the Urdu script is our national, cultural and religious script and is a national asset. We cannot say goodbye to this script because our history, culture, and the most of all our religion in the form of "The Qur'an" are preserved in this script. But, on the other hand, we cannot also close our eyes to the Roman script because it is a window to the modern world, and is a move towards the global single script.

9.1.5.2Educational Impacts and Implications of Adaptations

This section deals with the second category of impacts and implications of adaptations, that are educational impacts and implications. Both the categories of impacts and implications are shown below.

- ***Educational impacts of adaptations.*** Educational impacts of text messages concern different areas. An important impact of SMS adaptations is on spellings of students. Students have started using morphologically and phonologically shortened words in their school, college and university tests, examinations and assignments. This is not only the case with Pakistan; rather the impact of SMS/CMC orthography on education is world over. This issue has posed problems for teachers, parents, and even students themselves as they receive reduced marks for these deviations (Baron, 2008; Bodomo, 2010; Bosco, 2007; Crystal, 2008b; Janjua, 2010).

Impacts on syntactic structures and grammar of students' writing are also noteworthy. Students use shortened sentences with omitted/missing words. Moreover, they use deviated grammar with deviated tenses and word forms. Sometimes, the word order in sentences is also deviated from the conventional word order. Most of these deviations in students' writings are under the influence of adaptations made on the Internet and in text messages. Although grammatical influences are not as pervasive as orthographic impacts yet we cannot ignore them.

Students also use deviated punctuation marks and hence this is also an enigma towards maintaining conventional standards in writing. Moreover, in some cases code and script adaptations have also influenced students' writings. As the researcher himself is a teacher of English, so it has been observed by him personally, and reported by his colleagues that in examination papers some students begin their answers in the English language but where they get stuck, they code switch to Roman Urdu. In some cases students also code mix Urdu words in the English language tests/assignments. On the other hand, it is a

common practice of students to use English words in the Urdu language assignments and papers. It has also been reported by language teachers, that students whose command is weak on the English language, they write Romanized Urdu in the English language papers, to fill the belly of the paper under the maxim that something is better than nothing. In this way, they employ the use of both code and script adaptations at the one and the same time in the English language tests. In short, impacts of CMC and SMS modes of communications on students' writings in particular, and on education and pedagogy in general, are increasing day by day. Therefore, educationists and linguists should ponder over these issues to tackle them skillfully and judicially.

- ***Educational implications of adaptations.*** Anything that has impacts on something, it also has implications. So educational impacts of SMS language lead to educational implications of adaptations made in text messages.

The first type of implications is related to students' spellings. Students' spellings are changing under impacts of adaptations made in CMC and SMS modes of communication. The significant issue is that SMS and CMC are the informal modes of communication, whereas students incorporate SMS spellings in the formal modes like academic writing or examinations. Therefore, it is yet to be decided whether CMC and SMS spellings get their place in academics or get banished. To answer these issues, discussions and researches are in progress in the world over, and it is hoped that some appropriate conclusions will be drawn in the near future.

Same is the case regarding the use of syntactic, punctuation, code and script adaptations for educational/academic purposes. These adaptations in various varieties of language are now a reality, and realities must be addressed. If

realities are ignored, they create further complications. Therefore, the use of adapted forms of language in academics demands that linguists and educationists draw some conclusions about the academic uses of these informal but pervasive linguistic adaptations.

9.2 Conclusions

After the presentation of findings in detail, this section offers brief conclusions drawn from the study. In findings, all five questions of the study were addressed one by one, and for each question, findings were presented from all areas of the study. Here in this section, instead of addressing research questions, all categories of linguistic adaptations are briefly concluded in the sequence as given in chapters 5 to 8.

The first major category of linguistic adaptations in the study is of lexical adaptations. Lexical adaptations are the most noticeable area of all linguistic adaptations in text messages. This category was analysed in three areas, which are morphological shortenings, phonological shortenings, and spellings and case variations. The study concludes that morphological shortenings of words are mostly made in three ways that are initialism, contractions and clippings. All three types of morphological shortenings are made by texters for economical purposes. Texters save their time and effort by using short forms of words. They employ word initials like "AoA", "SMS", 'SAW", etc., contractions like "I'm", "don't", "it's", and clipped words like "u", "r", 'ur" to save their time and effort.

Phonological shortenings are made by using letter, number, and symbol homophones of words (or parts of words) to substitute actual spellings of words. Texters mostly use letter homophones/logographs like "u", "k", "r", "b", number homophones/logographs like "2', "4", "8", "9", and symbol homophones/logographs for words like "&", "%", "#", "@" etcetera. In the historical perspective, these

logographs/homophones of words are not new forms of language. We have traces of such logographic languages both in the past and present. At present, logographic characters of Chinese and Japanese languages are examples of uses of logographs for words or parts of words.

Variations of spellings in text messages are made by using these morphological and phonological shortenings, and hence texters save their time and effort by avoiding conventional spellings of words. They also make spelling variations by using variant forms of letter-cases like all upper case, all lower case, and the variant mixture of both. Mostly they make lexical/orthographic adaptations for economical purposes, and sometimes for creative purposes. Less frequently, causes of lexical adaptations may be speed, carless attitude, or poor spellings of texters.

The second major category of linguistic adaptations in text messages is of syntactic adaptations. The study concludes that syntactic adaptations mostly occur in areas like word ellipses, word reduplications, and grammar deviations. Word ellipses/omissions occur in personal pronouns, auxiliary and copula verbs, articles, prepositions, conjunctions, and infinitives. These ellipses of words may occur in isolation or in chunks. In personal pronouns, the first person pronoun "I", in subjective case, is mostly omitted. Both auxiliary (helping) and copula (linking) verbs are omitted in text messages. Significantly, among all auxiliary, modal, and copula verbs, the verb form "be", in three present tense forms that is 'is, am, are" is omitted the most. In omissions of articles, the most frequent ellipses occur in omissions of the definite article "the". Moreover, any syntactic part of a simple sentence that is "subject", "verb", or "object" may be omitted by texters.

In contrast to word ellipses, syntactic adaptations in text messages are also made through reduplications and triplications of words for emphasis and creative purposes. In

this regard, texters employ all three types of reduplications that are exact, rhyming, and ablaut, in both Urdu and English. In exact reduplications, pairs like "please please", "many many", "happy happy", "bye bye", and "very very" etcetera were commonly used in English. In Urdu, pairs like "plz plz", "kn kn", "ks ks", "acha acha", and "achi achi" etcetera were used. In rhyming reduplications in English, word pairs like "sooper dooper" and "never ever" were found, whereas for Urdu, word pairs like "aisy jaisy", "wesay-hi kesay-hi", and "dukh sukh" were found. In ablaut reduplications, no example was found for English, whereas for Urdu, examples of "theek thaak", and "fit faat" were found. Among all three categories, exact reduplications are the most common whereas ablaut reduplications are the least common category of reduplications in both languages.

In syntactic adaptations, texters are also prone to use grammar deviations. They use tenses, word classes/forms, and word-order in deviant ways in SMS sentences. Regarding tense deviations, deviations of main verbs are more than auxiliary verbs. Among deviations of word classes/forms, a typical example is the first person subjective pronoun "I" that is replaced with the objective pronoun "me". In deviations of word order, the most of instances are of word order between subject and auxiliary verb in interrogative sentences. In grammar deviations, the most of deviations occur due to the careless attitude or the lack of grammatical rules of texters.

The third major category of linguistic adaptations analysed in the study is of punctuations and space adaptations. Adaptations of punctuations have been further divided into omissions, repetitions, and substitutions. Regarding omissions of punctuations, four very crucial punctuation marks that are period/full stop (.), question mark (?), exclamation mark (!), and apostrophe (') were most frequently omitted in the corpus. Texters omit them because their absence does not cause loss of any key information in short messages. In the category of repetitions, ellipses/dots (.), question

marks (?), and exclamation marks (!) were mostly repeated. The repetitive use of punctuation marks is intentionally employed by texters for emphasis or creative purposes. Sometimes, right punctuation marks are substituted with deviant/inappropriate punctuation marks. For example, in the corpus, question marks were replaced with periods, periods with questions marks, apostrophes with commas, and exclamation marks with dashes.

In text messages, punctuation marks are also used for creative purposes. They are used as paralinguistic and artistic devices. For paralinguistic purposes, they are used as emoticons (emotion icons). In speech, paralinguistic/non-verbal elements like gestures, tone, body language, and facial expressions are common ways to communicate emotions. In the written language, such emotions are conveyed through these emotion markers known as emoticons like (☺ for happy) or (☹ for sad). In text messages, punctuation marks are used to manually develop these emoticons. Punctuation marks are also used to create artistic patterns like the word Allah in Urdu (اللہ), eid-greetings, and good-mornings etcetera.

In text messages, spaces (blank areas) are also used in deviated ways like their inappropriate omissions or additions. The English language conventionally applies insertion of one space at word boundaries to separate words, but no insertion of spaces within single words. In text messages, these conventional rules of spaces are sometimes violated. In this regard, the study shows various examples of under-spaced and over-spaced elements in text messages (see section 7.4).

The fourth and last major category of linguistic adaptations analysed in the study is of code and script adaptations. Code adaptations are a natural phenomenon of bilingual societies both in speech and writing. Code adaptations involve two processes that are code-mixing and code-switching. Code-mixing is the borrowing/mixing of words,

phrases, and clauses of one contact language into the other within the sentence boundaries, whereas code-switching is swinging between two languages on/outside sentence boundaries. This study explored both processes of code adaptations in Urdu and English because both Urdu and English act as two dominant contact languages in Pakistan.

In Pakistan, code-mixing in text messages is a two way process, that is, from Urdu to English and English to Urdu. In some text messages, Urdu acts as the frame language whereas in some text messages English acts as the frame language. Among all three categories of code-mixing that is, at word, phrase and clause level, nouns and noun-phrases are the most code-mixed elements. At the clause level, dependent (sub-ordinate) clauses are the only code-mixed items in both Urdu-to-English and English-to-Urdu code-mixing. This code-mixing of dependent clauses occurs on clause boundaries, mostly in the presence of a complementizer (subordinating conjunction) like "if", "and", and "that" etcetera.

In text messages, code-switching occurs in three ways that is from Urdu to English, English to Urdu, and two-way in both languages. These inter-sentential code adaptations may occur from any contact language to the other. The two-way code-switching involves two-way swing, that is, from any one language to the other and then back way once again.

The phenomenon of script adaptations in text messages also concerns bilingual/multilingual societies. In bilingual/multilingual settings, two or more contact languages may conventionally be written in dissimilar writing scripts. In the context of Pakistan, Urdu, the national language of Pakistan, is conventionally written in the Arabic/Urdu script. On the other hand, the Roman script is conventionally used for English in Pakistan and the world over. But the abundant use of the Internet and mobile

phone technologies in Pakistan has also led masses to use the Roman script for Urdu instead of its conventional script.

In the light of the pervasive use of the Roman script for Urdu in Pakistan, the study analysed the corpus based Romanization patterns adopted for Urdu in text messages. The study analysed variant spellings of 50 most frequent Urdu words in SMS-Corpus. Regarding consonants, the study concludes that texters consistently use specific consonant letters/digraphs for specific consonant sounds/phonemes of Urdu, and therefore consonants are used in consistent patterns. Complications mainly occur in vowel sounds (both pure and nasal), because one single vowel letter can represent various vowel sounds, and one single vowel sound can be represented by various vowel letters. Moreover, some texters use consonantal spellings and omit vowel letters in the most of cases.

In this regard, on the bases of Romanization patterns found in the corpus, a Romanization scheme has been given for all 45 consonant characters/glyphs of Urdu. The Romanization scheme of consonants consists of only 27 letters/digraphs of the Roman alphabet (see Table 8.9/9.1). This scheme is not intuitive in nature, as it is based on Romanization patterns found in the corpus of 5000 text messages of Pakistani texters. Moreover, the prevalent IPA scheme for Urdu characters has also been added. The Romanization scheme presented for Urdu consonants is corpus based and is therefore practical for writing Urdu in the Roman script. Regarding the use of Roman consonant letters for Urdu, it was found that three Roman letters "C", "V", and "X" were not used for Urdu in SMS-Corpus.

Regarding the use of Roman letters for Urdu vowel sounds/characters, no consistent scheme could be offered based on results of the corpus. Texters use Roman vowel letters for Urdu vowel letters/sounds intuitively and without any consistency. This

non-consistency poses complications, because texters perceive vowel letters and sounds differently, and hence a consistent scheme of vowel letters/sounds is not practically possible. A minute and focused corpus study in the future may offer some insights into this area, but to offer a single consistent Romanization scheme for vowels will not be practical in itself.

After concluding all major types, and patterns/principles of various linguistic adaptations, the study also concludes major causes of linguistic adaptations in text messages. In this regard, the most of linguistic adaptations in text messages are mainly caused by three major factors. The first major cause is the economical concern of saving time and effort. Texters have to convey more by using less time and energy, so they adopt telegraphic style in text messages. Secondly, texters have to write in haste while doing some other activities as well. In speed, they aim at being rapid in typing text messages. Indeed, text messages are not a formal activity like academic writing. Texters mostly communicate with their colleagues, friends or family members in informal settings where formal language is not required. Hence, the economical and speed factors apply to all lexical adaptations of morphological and phonological shortenings, spelling and case variations, syntactic adaptations of word ellipses, omissions of punctuations and spaces, script adaptations of Roman Urdu, and sometimes even for code alterations.

The third major cause is the creative/innovative factor of inventing new patterns in text messages. This factor mostly applies where economical and speed factors are not the concerns of texters. This factor applies in lexical adaptations of spelling and case variations where repeated letters, symbols, and numbers are used, syntactic adaptations of word reduplications, various creative patterns of punctuations and spaces, and sometimes in code and script adaptations. These three major factors apply on intentional adaptations.

Besides intentional adaptations, some adaptations occur due unintentional adaptations. Among unintentional adaptations, a major factor is the careless attitude of texters about the conventional standards of languages. Moreover, many unintentional language adaptations/deviations in text messages occur due to the poor language proficiency of texters. These two factors mostly apply to the unintentional and non-consistent adaptations like odd spellings, typos, grammatical flaws and occasional omissions of within-word spaces etcetera. Such adaptations are non-consistent and weird in nature. They are mostly termed as deviations from conventional standards. These deviations are also the most criticised adaptations in text messages.

Besides these three general causes/factors of most adaptations, some adaptations have their specific factors like some punctuation adaptations are caused by paralinguistic purposes. Code adaptations are mainly caused by the bilingual setting of Pakistani texters who frequently swing between Urdu and English. Similarly, script adaptations are made because most Pakistani texters are not adept in typing the Urdu script and hence they prefer to use the Roman script.

Over all, various linguistic adaptations made in text messages can be attributed to specific causes, patterns, and principles. They are not completely new in the historical perspective because we have traces of such patterns of language use in history in one way or the other. They are not haphazard innovations of texters. Yet their use has abundantly increased under the impact of CMC and SMS modes of communication. This abundant and pervasive use and impact of these adaptations is challenging the educational and linguistic conventions and standards.

Still, one thing is obvious that in any language nothing is standard for ever, and nothing is permanent. Human languages always remain changing under certain influences, and the influence of modern communication technologies has become so

strong that it cannot be avoided and kept unnoticed for a longer time. Languages are belongings of their users, and language users are architects of these languages and language varieties. So when language architects decide to change any language, language variety, or language register, then prescriptive grammarians, formal linguists, conservative media, and traditional educationists cannot stop language change and variation. They can instead offer their insights to define future dimensions of those languages or language varieties which are under the powerful impact of modern communication technologies.

REFERENCES

Abbas, F., Aslam, S., & Rana, A. M. K. (2011). Code-mixing as a communicative strategy among the university level students in Pakistan. *Language in India, 11* (1 January 2011), 95-108. Retrieved from http://www.languageinindia.com/jan2011/codemixingfurrakh.pdf

Agar, J. (2004). *Constant touch: A global history of the mobile phone.* Cambridge, UK: Icon Books.

Allwood, J. (2000). An activity based approach to pragmatics. In H. Bunt, & B. Black (Eds.), *Abduction, belief and context in dialogue: Studies in computational pragmatics* (pp. 47-80). Amsterdam, Netherlands: John Benjamins.

American Psychological Association. (2010). *Publication manual of the American Psychological Association.* (6th ed.). Washington, DC: Author.

Anthony, L. (2011). *AntConc (Version 3.2.2.1w)* [Computer software]. Tokyo, Japan: Waseda University. Retrieved from http://www.antlab.sci.waseda.ac.jp

Aslam, R. F. M., Ahmad, A., & Sajid, M. A. (2011). A study of orthographic features of instant messaging in Pakistan: An empirical study. *Language in India, 11* (1 January, 2011), 192-205.

Baker, P. (2010). *Sociolinguistics and corpus linguistics.* Edinburgh, Texas: Edinburgh University Press.

Bamba, F., & Barnes, S. J. (2007). SMS advertising, permission and the consumer: A study. *Business Process Management Journal, 13*(6), 815-829.

Baron, N. S. (1998). Letters by phone or speech by other means: The linguistics of email. *Language & Communication, 18*, 133-170.

Baron, N. S. (2000). Alphabet *to email: How written English evolved and where it's heading.* London, UK: Routledge.

Baron, N. S. (2003). Language of the Internet. In A. Farghali (Ed.), *The Stanford handbook for language engineers* (pp. 59-127). Stanford, CA: CSLI.

Baron, N. S. (2008). *Always on: Language in an online and mobile world.* Madison Avenue, NY: Oxford University Press.

Baron, N. S. (n.d.). The myth of impoverished signal: Dispelling the spoken-language fallacy for emoticons in online communication. Retrieved from http://www.american.edu/cas/lfs/faculty-docs/upload/Baron_Emoticons-1-the-Myth-of.pdf

Bauer, L. (2002). *An introduction to international varieties of English.* Edinburgh, UK: Edinburgh University Press.

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

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

Birdwhistell, R. L. (1985). Kinesics and context: Essays on body motion communication. Pennsylvania: University of Pennsylvania.

Bodomo, A. B. (2010). *Computer-mediated communication for linguistics and literacy: Technology and natural language education.* Hershey, NY: IGI Global.

Bodomo, A. B., & Lee, C. K. M. (2002). Changing forms of language and literacy: Technobabble and mobile phone communication. *Literacy and Numeracy Studies: An International Journal in the Education and Training of Adults, 12*(1), 23-44.

Borg , J. (2010). Body *language: 7 easy lessons to master the silent language.* Upper Saddle River, NJ: Pearson Education.

Bosco, L. S. S. (2007). *SMS gener@tion: A study on the language of text messaging in Hong Kong* (Master's thesis, University of Hong Kong). doi: 10.5353/th_b3744163

Brinton, L. J., & Brinton, D. M. (2010). *The linguistic structure of modern English.* Amsterdam, Netherlands: John Benjamins.

Bussmann, H. (1996). *Routledge dictioanry of language and linguistics* (G. P. Trauth, & K. Kazzazi, Trans. & Eds.). London, UK: Routledge.

Cazden, C. B. (1972). *Child language and education.* New York, NY: Holt, Rinehart & Winston.

Chiluwa, I. (2008). Assessing the Nigerianness of SMS text messages in English. *English Today, 24*(1), 51-56.

Crystal, D. (Ed.). (1995). *The Cambridge encyclopaedia of the English language.* London, UK: Cambridge University Press.

Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge, UK: Cambridge University press.

Crystal, D. (2004). *Language and the Internet.* Cambridge, UK: Cambridge University Press.

Crystal, D. (Ed.). (2008a). *A dictionary of linguistics and phonetics* (6th ed.). Malden, MA: Blackwell.

Crystal, D. (2008b). *Txtng: The gr8 db8.* Oxford, UK: Oxford University Press.

Danet, B., & Herring, S. C. (2007). Introduction: Welcome to the multilingual internet. In B. Danet & S. C. Herring (Eds.), *The multilingual internet: Language, culture, and Communication online* (pp. 3-39). New York, NY: Oxford University Press.

Darwin, C. (2009). *The expression of the emotions in man and animals* (2nd ed.). New York, NY: Cambridge University Press.

Dash, N. S. (2005). *Corpus linguistics and language technology: With reference to Indian languages*. New Delhi, India: Mittal Publications.

Döring, N. (2002). Text msgs sent: Abbreviations and acronyms in the SMS communication. *Quarterly Magazine for the German Language, 112* (2), 97-114. Retrieved from http://www.nicola-doering.de/publications/sms-kurzformen-doering-2002.pdf

Durrani, A. (2008). *Urdu informatics*. Islamabad, Pakistan: National Language Authority.

Elewa, A.-H. (2004). *Collocation and synonymy in classical Arabic: A corpus-based study* (Doctoral dissertation, Centre for Computational Linguistics, University of Manchester, UK). Retrieved from http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.488171

Enli, S. (2007). Gate-keeping in the new media age. *Javnost-The Public, 2*, 47-61.

Field, M. (2009). *Improve your written English: Master the essentials of grammar, punctuation and spelling and write with greater confidence* (5th ed.). London, UK: Spring Hill.

Frehner, C. (2008). Email- SMS- MMS: The linguistic creativity of asynchronous discourse in the new media age. *Linguistic insights: Studies in language and communication*. Bern, Switzerland: Peter Lang.

Fromkin, V., Rodman, R., & Hyams, N. (2003). *An introduction to language* (7th ed.). Boston, MA: Wadsworth (Thomson).

Gilchrist, J. (1796). Grammar of the Hindoostanee language (Part third of volume first of a system of Hindoostanee philology). Calcutta, India: Chronicle Press.

Goggin, G. (2004). Mobile text. *Media/Culture: Journal of Media and Culture, 7*(1). Retrieved from http://journal.media-culture.org.au/0401/03-goggin.php

Gombert, J. E. (1992). *Metalinguistic development* (T. Pownall, Trans.). Chicago, IL: University of Chicago Press.

Greenbaum, S., & Nelson, G. (2002). *An introduction to English grammar* (2nd ed.). London: Longman/Pearson Education.

Greenfield, R. (2003, January). Collaborative e-mail exchange for teaching secondary ESL: A case study in Hong Kong. *Language Learning & Technology*, 7 (1), 46-70. Retrieved from http://llt.msu.edu/vol7num1/pdf/greenfield.pdf

Gries, S. T. (2009). *Quantitative corpus linguistics with R: A practical introduction.* New York, NY: Routledge.

Haggan, M. (2007). Text messaging in Kuwait: Is the medium the message? *Multilingua, 26*(4), 427-449.

Halliday, M.A.K., & Hasan, R. (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective.* Oxford, UK: Oxford University Press.

Hard af Segerstad, Y. (2002). *Use and adaptation of the written language to the conditions of computer-mediated communication* (Doctoral dissertation, University of Goteborg, Sweden). Retrieved from http://www.ling.gu.se/~ylvah/dokument/ylva_diss.pdf

Hard af Segerstad, Y. (2005). Language in SMS: A socio-linguistic view. In R. Harper, L. Palen, & A. Taylor (Eds.), *The inside text: Social, cultural and design perspectives on SMS* (pp. 33-51). Dordrecht, Netherlands: Springer.

Herring, S. C. (1999). Interactional coherence in CMC. *Journal of Computer-Mediated Communication, 4* (4). Retrieved from http://jcmc.indiana.edu/vol4/issue4/herring.html

Herring, S. C. (2001). Computer-mediated discourse. In D. Tannen, D. Schiffrin & H. Hamilton (Eds.), *Handbook of discourse analysis* (pp. 612-634). Oxford, UK: Blackwell.

Hiltz, S. R., & Turoff, M. (1978). *The network nation: Human communication via computer.* Reading, MA: Addison-Wesley Publishing.

Ho, J. W. Y. (2007). Code-mixing: Linguistic form and socio-cultural meaning. *The International Journal of Language Society and Culture*, Issue 21. Retrieved from http://www.educ.utas.edu.au/users/tle/JOURNAL/issues/2007/21-2.pdf

Hussain, S. (n.d.). *Letter-to-sound conversion for Urdu text-to-speech system.* Lahore, Pakistan: Center for Research in Urdu Language Processing, NUCES. Retrieved from http://acl.ldc.upenn.edu/W/W04/W04-1613.pdf

Ijaz, M. (2007). *Urdu letters to IPA and IPA to SAMPA*. Lahore, Pakistan: Center for Research in Urdu Language Processing, NUCES. Retrieved from http://www.crulp.org/Downloads/langproc/Urdu_IPA_to_Sampa.pdf

Ijaz, M., & Hussain, S. (n.d.)).*Corpus based Urdu lexicon development.* Lahore, Pakistan: Center for Research in Urdu Language Processing, NUCES. Retrieved from http://ling.uni-konstanz.de/pages/home/pargram_urdu/DAADlex/papers/ijaz_2007.pdf

Janjua, F. (2010). D efkt ov mobil fones on d spellings ov university students. *Kashmir Journal of language Research, 13 (2)*, 71-81.

Javid, M., Malik, M. A., & Gujjar, A. A. (2011). Mobile phone culture and its psychological impacts on students' learning at the university level. *Language in India, 11* (2 February 2011), 415-422. Retrieved from http://www.languageinindia.com/feb2011/mobilegujjar.pdf

Jensen, B. T. (2008). Metalinguistic awareness. In J. M. González (Ed.), *Encyclopedia of Bilingual Education* (pp. 551-554). Los Angelas: Sage.

Kabir, H., & Saleem, A. M. (n.d.). *Speech assessment methods phonetic alphabet (SAMPA) : Analysis of Urdu.* Lahore, Pakistan: Center for Research in Urdu Language Processing, NUCES. Retrieved from http://crulp.org/Publication/Crulp_report/CR02_02E.pdf

Kasesniemi, E.-L., & Rautiainen, P. (2002). Mobile culture of children and teenagers in Finland. In J. E. Katz & M. Aakhus (Eds.), *Perpetual contact: Mobile communication, private talk, public performance* (pp. 170-192). Cambridge, UK: Cambridge University Press.

Kennedy, G. D. (1998). *An introduction to corpus linguistics*. London, UK: Longman.

Klemens, G. (2010). *The cell phone: The history and technology of the gadget that changed the world.* Jefferson, NC: McFarland.

Kucera, H., & Francis, w. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Labov, W. (2001). *Principles of linguistic change: Social factors* (Vol. 2). Oxford, UK: Blackwell.

Leech, G. (1992). *Introducing English grammar.* London, UK: Penguin.

Leech, G., Hundt, M., Mair, C., & Smith, N. (2009). *Change in contemporary English: A grammatical study.* Cambridge, UK: Cambridge University Press.

Lehmann, W. P. (1992). Historical Linguistics (3rd ed.). London, UK: Routledge.

Ling, R. (1998). *The use of traditional fixed and mobile telephony for social networking among Norwegian parents.* R&D report 33/98. Kjeller, Norway: Telenor.

Ling, R. (2005). The sociolinguistics of SMS: An analysis of SMS use by a random sample of Norwegians. In R. S. Ling & P. E. Pedersen (Eds.), *Mobile*

Communications: Re-negotiation of the socialsphere (pp. 335 – 350). London,UK: Springer.

Ling, R. (2008). *New tech, new ties: How mobile communication is reshaping social cohesion.* London, UK: MIT Press.

Lock, G. (1996). *Functional English grammar: An introduction for second language teachers*. Cambridge, UK: Cambridge University Press.

Lüdeling, A., & Kytö, M. (2008). *Corpus linguistics: An international handbook* (Vol. 1). New York, NY : Walter de Gruyter.

McEnery, T., & Wilson, A. (2001). *Corpus Linguistics* (2nd ed.). Edinburgh, UK: Edinburgh University Press.

McEnery, T., Xiao, R., & Tono, Y. (2006) *Corpus-based language studies: An advanced resource book.* London, UK: Routledge.

Mercer, D. (2006). The telephone: The life story of a technology. London, UK: Greenwood.

Meyer, C. F. (2004). *English corpus linguistics: An introduction.* Cambridge, UK: Cambridge University Press.

Myers-Scotton, C. (2006). *Multiple voices: An introduction to bilingualism.* Malden, MA: Blackwell.

Nelson, D. L., & Quick, J. C. (2008), *Understanding organizational behaviour*. Mason, OH: Thomson South-Western.

Oksman, V., & Turtianen, J. (2004). Mobile communication as a social stage: Meanings of mobile communication in everyday life among teenagers in Finland. *New Media and Society, 6*(3), 319-339.

Ong'onda, N. A., Matu, P. M., & Oloo, P. A. (2011). Syntactic aspects in text messaging. *World Journal of English Language, 1*(1, April 2011). doi: 10.5430/wjel.v1n1p2

Pakistan Telecommunication Authority. (2005). *Annual report 2005*. Islamabad, Pakistan: Author.

Pakistan Telecommunication Authority. (2008). *Annual report 2008*. Islamabad, Pakistan: Author.

Pakistan Telecommunication Authority. (2010). *Annual report 2010*. Islamabad, Pakistan: Author.

Palmer, F. R. (1983). *Grammar* (2nd ed.). London, UK: Penguin.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a typology of code-switching. *Linguistics, 18,* 581-618.

Prete, M. I. (2007). M-politics: Credibility and effectiveness of mobile political communications. *Journal of Targeting, Measurement and Analysis for Marketing, 16*(1), 48.

Rafi, M. S. (2008). SMS text analysis: Language, gender and current practices. *TESOL France- Online Journal.* Retrieved from http://www.tesol-france.org/OnlineJournal.php

Rafi, M. S. (2010). The sociolinguistics of SMS: Ways to identify gender boundaries. *Handbook of Research on Discourse Behavior and Digital Communication: Language Structures and Social Interaction*. Pennsylvania, USA: IGI Publishers.

Rasul, S. (2006). *Language hybridization in Pakistan as socio-cultural phenomenon: An analysis of code-mixed linguistic patterns* (Doctoral dissertation, NUML, Islamabad, Pakistan). Retrieved from http://prr.hec.gov.pk/thesis/2426.pdf

Rasul, S. (2008). Language hybridization in Pakistani media: Redefining ideology . *Kashmir Journal of Language Research, 11*(1), 14-43.

Rasul, S. (2009). Code-mixing in Pakistani newspapers: A socio-linguistic analysis. *Kashmir Journal of Language Research, 12*(1), 25-49.

Richards, J. C., & Schmidt, R. (Eds.). (2002). *Longman dictionary of language teaching and applied linguistic* (3rd ed.). London, UK: Pearson Education.

Ring, D. H. (1947). *Mobile telephony: Wide area coverage* (Technical memoranda). Berkeley Heights, NJ: Bell Telephone Laboratories.

Robinson, A. (2009). *Writing and script: A very short introduction*. New York: Oxford University press.

Ross, N. J. (2006). Writing in the information age. *English Today, 87*(22), 39-45.

Rumsiene, G. (2006). Internet English: A technically based mode of language? *Studies About Languages, 9,* 56-63. Retrieved from http://www.kalbos.lt/zurnalai/09_numeris/08.pdf

Shortis, T. (2007). Revoicing txt: Spelling, vernacular orthography and unregimented writing. In S. Posteguillo, M. J. Esteve, & M.L. Gea (Eds.), *The texture of internet: Netlinguistics* (pp. 1-23). Cambridge, UK: Cambridge Scholar Press.

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford, UK: Oxford University Press.

Sutherland, J. (2002). Cn u txt? *The Guardian* (11November). Retrieved from http://www.guardian.co.uk/technology/2002/nov/11/mobilephones2

Tagg, C. (2009). *A corpus linguistics study of sms text messaging* (PhD dissertation, University of Birmingham). Retrieved from http://etheses.bham.ac.uk/view/year/2009.html

Tagliamonte, S. A. (2006). *Analysing sociolinguistic variation*. Cambridge, UK: Cambridge University Press.

Thurlow, C., & Brown, A. (2003). Generation txt? Exposing the sociolinguistics of young people's text-messaging. *Discourse Analysis Online, 1*(1). Retrieved from http://extra.shu.ac.uk/daol/articles/open/2002/003/thurlow2002003-paper.html

Thurlow, C., & Poff, M. (2011). Text-messaging. In S. C. Herring, D. Stein, & T. Virtanen (Eds.), *Handbook of the pragmatics of CMC*. Berlin, Germany: Mouton de Gruyter.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam, Netherlands: John Benjamins.

Tomkins, S. S. (1962). *Affect imagery consciousness: The positive affects* (Vol. I). London, UK: Tavistock.

Tomkins, S. S. (1963). *Affect imagery consciousness: The negative affects* (Vol. II). London, UK: Tavistock.

Tomkins, S. S. (1991). *Affect imagery consciousness: The negative affects-Anger and Fear* (Vol. III). New York, NY: Springer.

Tomkins, S. S., & Bertram, P. K. (1992). *Affect, imagery, consciousness* (Vol. IV, 1962-1962). New York, NY: Springer.

Troike, R. C. (2008). Code switching. In J. M. González (Ed.), *Encyclopedia of bilingual education* (pp. 551-554). Los Angeles: Sage.

Trudgill, P. (1995). *Sociolinguistics: An introduction to language and society* (3rd ed.). Harmondsworth, UK: Penguin Books.

Wallis, S. A., & Nelson, G. (2001). Knowledge discovery in grammatically analysed corpora. *Data Mining and Knowledge Discovery, 5*, 307-340.

Warschauer, M. (1997). Computer-mediated collaborative learning: Theory and practice. *The Modern Language Journal, 81* (4), 470-481.

Wiechmann, D. & Fuhs, S. (2006). Corpus linguistics resources: Concordancing software. *Corpus Linguistics and Linguistic Theory, 2*(1), 109-130. Retrieved from www.daniel-wiechmann.eu/downloads/prefinal%20draft.pdf

Xiao, R. (2008). Theory-driven corpus research: Using corpora to inform aspect theory. In A. Lüdeling, & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (pp. 987-1008). Berlin, Germany: Mouton de Gruyter.

Yan, L. P. (2008). *The impact of computer technology on language choice and CMC practice: A study of instant messaging in Hong Kong* (Master's thesis, University of Edinburgh, Scotland). Retrieved from http://www.era.lib.ed.ac.uk/bitstream/1842/2848/1/dissertation-s0790015.pdf

Yule, G. (2006). *The study of language* (3rd ed.). Cambridge, UK: Cambridge University Press.

Zurhellen, S. (2011). A misnomer of sizeable proportions: SMS and oral tradition. *Oral Tradition, 26*(2), 637-642. Retrieved from http://journal.oraltradition.org/files/articles/26ii/25_26.2.pdf

APPENDICES

APPENDIX-A

Questionnaire on SMS Language Used in Pakistan

ASSALAM-O-ALAIKUM

Dear Respondent,

I am a PhD Research Fellow at the Department of English, International Islamic University Islamabad. I am conducting research on the SMS Language used in Pakistan. In this regard, I need your valuable comments and some SMS data. I would be highly thankful to you for your cooperation in filling up this questionnaire. I assure you that the data provided by you will only be used for research purposes and your identity will not be disclosed at any stage.

With deep personal regards,
Yours faithfully,
Malik Naseer Hussain

[My contacts for your comments and queries]
Email: maliknaseerhussain@yahoo.com
Cell # **********

The questionnaire is divided into two parts. You are requested to fill in both parts.

PART-A

1. Your Name (***Optional***) _____ Phone No. ***(Optional)*** _____
2. Make(e.g., Nokia/LG/Samsung) & Model No. of your Mobile Set ***(Optional)***_____
3. **A)** Your Gender (Male/Female)_____ **B)** Your Age in Years_____
4. Your Completed Academic Qualification_____
5. Your Status (Please tick the relevant option) **Student / Professional / Unemployed**

For Students	Class _____ Degree Title _____ Institute_____
For Professionals	Profession _____ Job Title _____ Department_____

6. Do you think that SMS language is emerging as a new variety of written language? **Please tick one option.**

Yes	No	Uncertain

7. How is SMS language affecting your written language? **Please tick one option.**

Positively	Negatively	Both positively and negatively	No effects

8. Seeing the large use of SMS language by people, do you think that SMS language should be incorporated/introduced in teaching/learning? **Please tick one option.**

Yes	No	Uncertain

9. What according to you may be the reasons/factors of short telegraphic style of SMS language which make it different from standard/formal language? **Please tick ALL the relevant options.**
   - **A.** Small keyboard of mobile sets
   - **B.** For saving time
   - **C.** Accuracy is not important in SMS, understanding is important
   - **D.** Any other reason (Plz specify) _____
10. For which purposes/functions do you normally use/send SMS? **Please tick ALL the relevant options.**
   - **A.** Chatting with friends/sharing jokes
   - **B.** Requesting friends/relatives for favours
   - **C.** Responding to the requests/favours of friends/relatives
   - **D.** Information Sharing/exchanging
   - **E.** Love making/Romance
   - **F.** Teaching/learning purposes
   - **G.** Conveying good wishes/prayers to friends/relatives
   - **H.** Any other purpose (Plz specify) _____
11. Why do you sometime include features of spoken language in your SMS (e.g., *2* for *two, gr8* for *great, U* for *you, ur* for *your, c* for *see* etc.)? **Please specify reasons.** _____
12. Why do you use Romanized URDU words/sentences in your English SMS? **Please specify reasons.**

    _____
13. Do you think that some SMS conflict with our religious/ethical/cultural values? **Please Tick one option.**

Yes	No	Uncertain

PART-B

In this part of the Questionnaire you are requested to provide SIX actual SMS from your Mobile Phone in the given space. I hope it would not take your much time. You are requested to please consider the following guidelines before providing data.

- Please provide only one-to-one exchanges/personal communication.
- Publicity messages sent by Mobile companies should not be provided.
- Both Urdu and English SMS can be provided.
- Urdu SMS should only be given if they are written in English alphabets/letters.
- Both Sent (Outbox) and Received (Inbox) messages of any length can be given.
- SMS should be written in the given blocks as they are actually visible on the LCD/screen of your mobile phone. Messages should be given in their exact original form. **Please do not correct or change anything of your messages,** otherwise the purpose of the research will be lost.

SPACE FOR SMS DATA

- Please TICK before each SMS whether the SMS is **Received (Inbox)** or **Sent (Outbox)** message.
- Also mention the relationship of Sender/Receiver to YOU (e.g., **Friend/Brother/Sister/Husband/Wife/Teacher** etc)

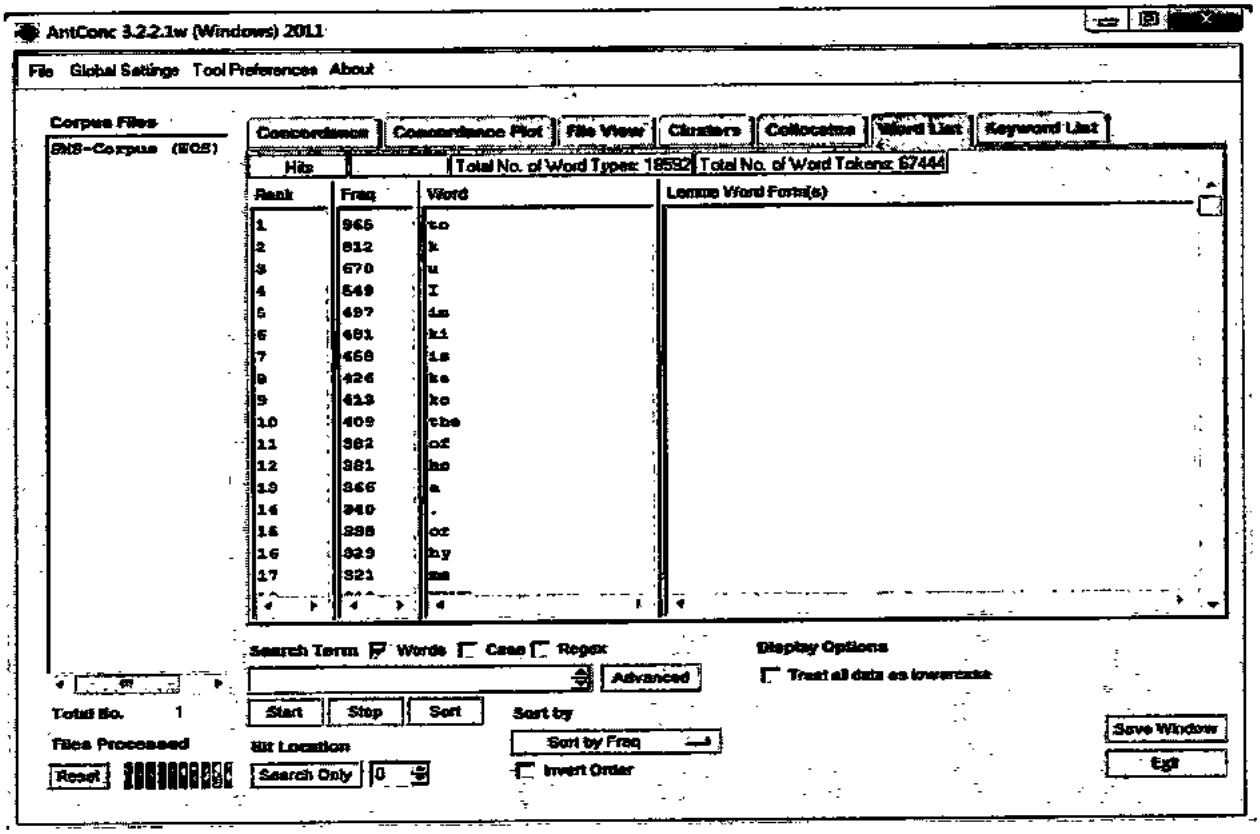
SMS # 1	SMS # 2	SMS # 3
Plz Tick: Received or Sent	Plz Tick: Received or Sent	Plz Tick: Received or Sent
Relation to You: _____	Relation to You: _____	Relation to You: _____
SMS # 4	SMS # 5	SMS # 6
Plz Tick: Received or Sent	Plz Tick: Received or Sent	Plz Tick: Received or Sent
Relation to You: _____	Relation to You: _____	Relation to You: _____

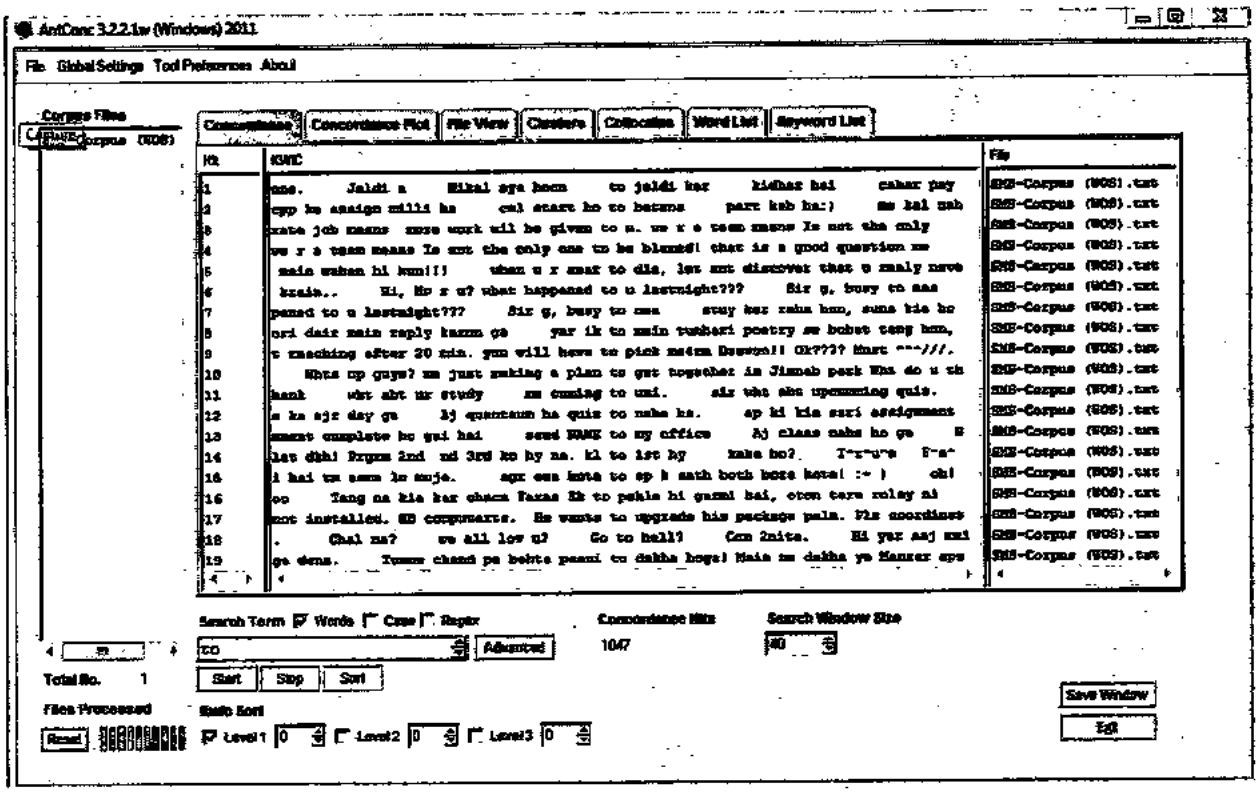
You are requested to please recheck whether you have filled up all parts of the Questionnaire.

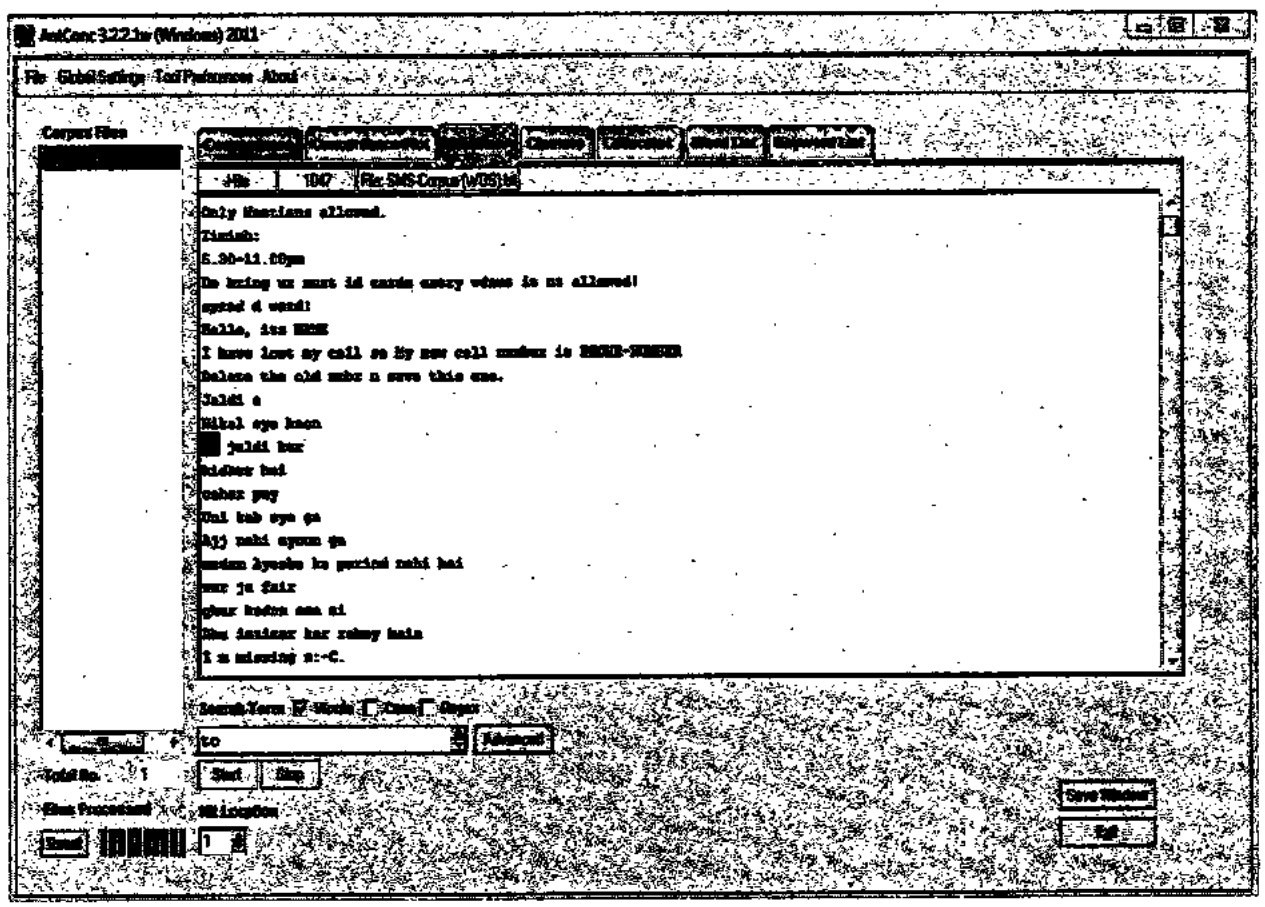
THANK YOU VERY MUCH FOR YOUR COOPERATION

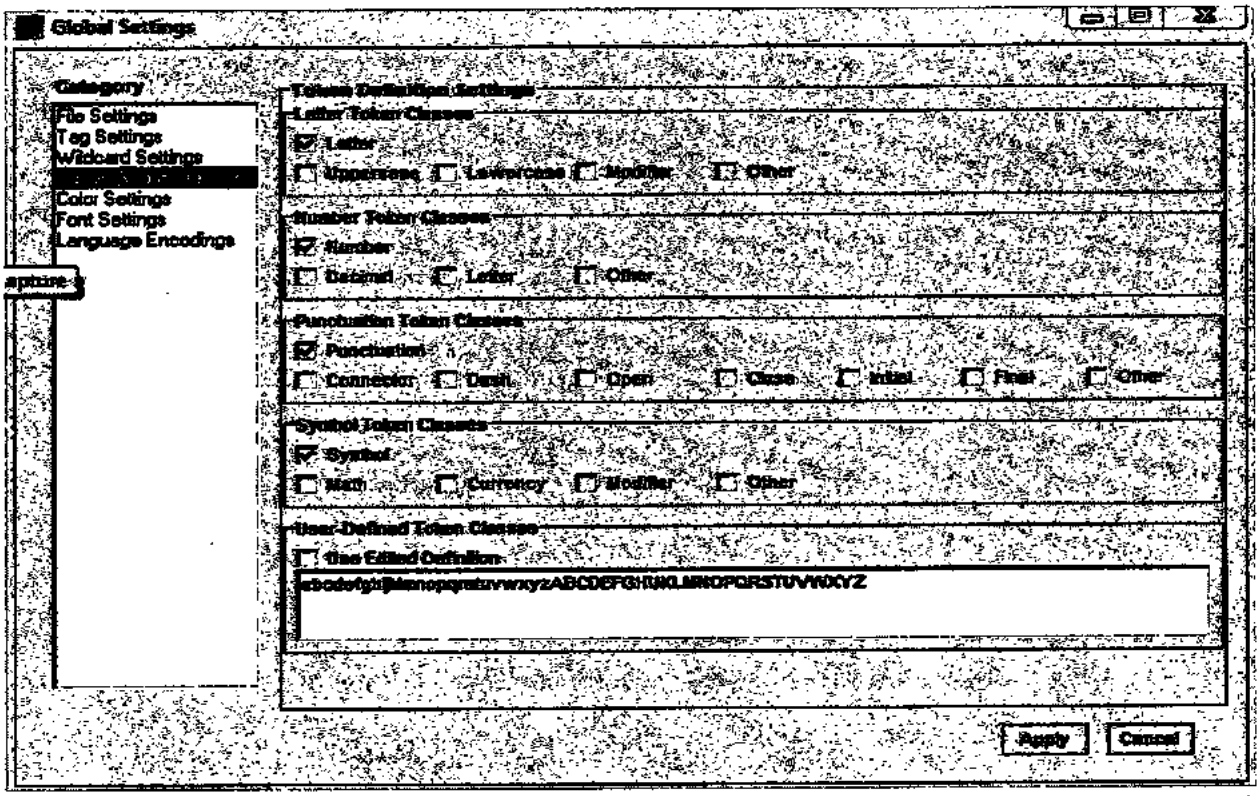
APPENDIX-B

Screenshots of AntConc Software in Its Launched Form








Global Settings
Category
File Settings
Tag Settings
Wildcard Settings
Color Settings
Font Settings
Language Encodings
Letter Token Classes
Letter
Uppercase
Lowercase
Modifier
Other
Number Token Classes
Number
Decimal
Letter
Other
Punctuation Token Classes
Punctuation
Connector
Dash
Open
Close
Initial
Final
Other
Symbol Token Classes
Symbol
Math
Currency
Modifier
Other
User-Defined Token Classes
Use Edited Definition
Apply
Cancel

APPENDIX-C

1000 Most Frequent Words in Corpus (Antconc Word List) Both English and Urdu

In Word List (*tool preferences*) option *__treat all data as lower case_* activated
Global setting: All four token classes activated
AntConc Statistics: Total No of word types: 15746
Total No of word tokens: 67444

RANK	FRQ	WORD	RANK	FRQ	WORD
1	1039	to	501	18	office
2	909	k	502	18	pak
3	764	u	503	18	sakta
4	753	i	504	18	sir.
5	542	in	505	18	sure
6	530	ki	506	18	zindgi
7	503	is	507	17	20**
8	470	ko	508	17	2day
9	456	ka	509	17	=
10	454	the	510	17	aftr
11	432	a	511	17	before
12	413	me	512	17	birthday
13	413	of	513	17	book
14	400	ho	514	17	chal
15	372	or	515	17	college
16	353	r	516	17	cum
17	349	hy	517	17	dain
18	344	name	518	17	eman
19	340	.	519	17	faraz
20	323	you	520	17	hal
21	302	se	521	17	hoa
22	280	hai	522	17	hoi
23	264	na	523	17	inshaallah
24	263	and	524	17	karta
25	261	b	525	17	kiya
26	258	sir	526	17	laga
27	258	ur	527	17	long
28	239	allah	528	17	mairi
29	233	nai	529	17	monday
30	232	hain	530	17	nay
31	229	for	531	17	rab
32	219	2	532	17	result
33	209	at	533	17	saath
34	208	plz	534	17	shadi
35	206	my	535	17	their
36	203	on	536	17	too
37	194	&	537	17	university

38	193	mai	538	17	up?
39	188	ye	539	17	w
40	187	main	540	17	wala
41	185	m	541	17	wali
42	184	ha	542	17	whats
43	178	are	543	17	wid
44	175	will	544	16	)
45	173	your	545	16	aapki
46	168	n	546	16	better
47	165	kr	547	16	com
48	164	he	548	16	dard
49	162	aap	549	16	dunya
50	162	be	550	16	f9
51	162	it	551	16	gunah
52	161	ma	552	16	hi!
53	161	us	553	16	israr
54	156	ap	554	16	itna
55	155	ok	555	16	keep
56	147	good	556	16	khana
57	143	ne	557	16	light
58	137	4	558	16	luck
59	134	aur	559	16	manzil
60	134	kya	560	16	message
61	133	wo	561	16	mgr
62	132	have	562	16	missing
63	132	kar	563	16	mje
64	132	so	564	16	mn
65	131	jo	565	16	month
66	130	but	566	16	muj
67	128	kia	567	16	mushkil
68	124	not	568	16	nikal
69	123	with	569	16	no.
70	122	come	570	16	pehly
71	118	do	571	16	person
72	117	am	572	16	point
73	117	sy	573	16	qaim
74	115	kal	574	16	said
75	115	this	575	16	si
76	114	no	576	16	students
77	113	may	577	16	wait
78	111	that	578	16	without
79	109	koi	579	16	zindagi
80	109	yar	580	15	20
81	108	all	581	15	;-)
82	108	when	582	15	app
83	103	hay	583	15	btao
84	103	u?	584	15	di
85	102	ga	585	15	die

86	101	where	586	15	e-mail
87	100	ny	587	15	gae
88	99	1	588	15	ha,
89	99	as	589	15	he.
90	99	wil	590	15	islam
91	98	can	591	15	jaga
92	98	we	592	15	karti
93	97	what	593	15	kehna
94	96	if	594	15	leave
95	94	hi	595	15	lecture
96	93	t	596	15	ov
97	92	aj	597	15	party
98	90	pr	598	15	pay
99	89	class	599	15	plz.
100	89	how	600	15	raat
101	89	ke	601	15	shukar
102	89	mein	602	15	sir,
103	89	nahi	603	15	sory
104	89	time	604	15	still
105	88	sa	605	15	story
106	87	tha	606	15	think
107	85	e	607	15	walon
108	84	ya	608	15	whr
109	82	hon	609	14	.!
110	78	o	610	14	8
111	77	after	611	14	abhi
112	77	its	612	14	ami
113	76	meri	613	14	asar
114	76	tum	614	14	ban
115	75	ab	615	14	band
116	75	ghar	616	14	campus
117	74	tu	617	14	dawat
118	72	by	618	14	day.
119	72	sb	619	14	don't
120	71	din	620	14	dont
121	71	say	621	14	dy
122	70	raha	622	14	enjoy
123	69	happy	623	14	feel
124	69	hum	624	14	hello
125	69	kuch	625	14	insan
126	69	send	626	14	inshallah
127	68	bhi	627	14	isi
128	68	pe	628	14	janab
129	68	today	629	14	jani
130	67	eid	630	14	jannat
131	67	from	631	14	jaon
132	66	han	632	14	jin
133	65	day	633	14	kabhi

134	65	kisi	634	14	kafi
135	65	life	635	14	kitna
136	64	there	636	14	la
137	64	was	637	14	man
138	62	call	638	14	mil
139	62	coming	639	14	mind
140	62	ek	640	14	news
141	62	phone-number	641	14	off
142	62	who	642	14	pick
143	61	3	643	14	rah
144	61	please	644	14	stay
145	61	u.	645	14	sub
146	60	just	646	14	subah
147	59	aa	647	14	thek
148	59	aoa	648	14	thy
149	59	liye	649	14	tomorrow.
150	58	!	650	14	u,
151	58	bring	651	14	uski
152	58	hai.	652	14	waisy
153	58	kaha	653	14	wat
154	57	un	654	14	would
155	55	tak	655	14	yaad
156	54	dil	656	14	z
157	54	jaldi	657	13	(institute)
158	54	very	658	13	2.
159	53	acha	659	13	;)
160	53	nhi	660	13	ai
161	52	don	661	13	aisa
162	52	pata	662	13	apny
163	52	q	663	13	ask
164	51	aaj	664	13	assignment
165	51	any	665	13	avoid
166	51	gi	666	13	ay
167	51	jis	667	13	bas
168	51	love	668	13	blessed
169	51	mera	669	13	bohat
170	51	mere	670	13	busy
171	51	msg	671	13	copy
172	50	kaam	672	13	dafa
173	50	morning	673	13	dnt
174	50	nt	674	13	dya
175	50	sath	675	13	eco
176	49	dear	676	13	ham
177	49	s	677	13	happiness
178	49	take	678	13	having
179	48	gai	679	13	hi,
180	48	ja	680	13	huwa
181	48	kam	681	13	im

182	47	baat	682	13	jamaat
183	47	uni	683	13	jang
184	46	bhai	684	13	khushi
185	46	room	685	13	kidr
186	45	g	686	13	kindly
187	45	hota	687	13	krta
188	44	aik	688	13	live
189	44	apni	689	13	mairy
190	44	de	690	13	means
191	44	ha.	691	13	mujh
192	44	kb	692	13	naam
193	44	many	693	13	out
194	44	pls	694	13	pa
195	44	quran	695	13	parh
196	44	reply	696	13	part
197	44	sms	697	13	pathan
198	43	dua	698	13	possible
199	43	get	699	13	program
200	43	hai?	700	13	read
201	43	has	701	13	someone
202	43	his	702	13	two
203	43	namaz	703	13	walk
204	43	per	704	13	y
205	42	dr.	705	13	year
206	42	go	706	13	zara
207	42	kis	707	12	*name
208	42	phir	708	12	15
209	41	agr	709	12	:-d
210	41	aoa.	710	12	??
211	41	apne	711	12	ada
212	41	back	712	12	apko
213	41	da	713	12	assalam
214	41	ga.	714	12	bje
215	41	jab	715	12	bta
216	41	kahan	716	12	check
217	41	krna	717	12	could
218	41	salam	718	12	dn
219	40	abi	719	12	doon
220	40	give	720	12	everything
221	40	jana	721	12	gain
222	40	karain	722	12	gi.
223	40	ll	723	12	haq
224	40	mubarak	724	12	hay,
225	40	one	725	12	hey!
226	40	tell	726	12	hn.
227	40	up	727	12	hun.
228	39	an	728	12	hyn
229	39	did	729	12	inform

230	39	jao	730	12	jaiy
231	39	par	731	12	kaise
232	38	always	732	12	kdr
233	38	gud	733	12	kea
234	38	hoti	734	12	mama
235	38	jihad	735	12	me,
236	38	meeting	736	12	men
237	38	now	737	12	mili
238	38	waqt	738	12	min
239	37	?	739	12	muhammad
240	37	best	740	12	nazar
241	37	jb	741	12	neend
242	37	kab	742	12	old
243	37	mujy	743	12	over
244	37	tm	744	12	own
245	37	v	745	12	pass
246	36	:-)	746	12	pehle
247	36	ana	747	12	programme
248	36	ata	748	12	rahain
249	36	d	749	12	sahib
250	36	hav	750	12	same
251	36	ho?	751	12	special
252	36	home	752	12	suna
253	36	make	753	12	talk
254	36	night	754	12	taraf
255	36	pray	755	12	thats
256	35	10	756	12	uni.
257	35	also	757	12	watch
258	35	gay	758	12	website
259	35	going	759	12	were
260	35	hey	760	12	wohi
261	35	lia	761	12	wt
262	35	like	762	12	yahan
263	35	need	763	12	yara
264	35	sab	764	11	!!
265	35	thanks	765	11	)*
266	35	they	766	11	11
267	35	tomorrow	767	11	3.
268	35	vo	768	11	30
269	35	waiting	769	11	:d
270	34	about	770	11	aao
271	34	late	771	11	again
272	34	me.	772	11	b4
273	34	nahin	773	11	being
274	34	nd	774	11	bhe
275	34	new	775	11	blast
276	34	oye	776	11	bro
277	34	rahe	777	11	chahye

278	34	want	778	11	clg
279	33	*	779	11	cme
280	33	ho.	780	11	comng
281	33	mujhe	781	11	diya
282	33	nice	782	11	done
283	33	pas	783	11	dreams
284	33	sirf	784	11	ever
285	33	usra	785	11	family.
286	32	him	786	11	fr
287	32	kro	787	11	gya
288	32	le	788	11	halat
289	32	log	789	11	hamara
290	32	our	790	11	hen.
291	32	thank	791	11	hmm
292	31	a.o.a	792	11	ho,
293	31	bht	793	11	hu
294	31	fine	794	11	isha
295	31	some	795	11	jawab
296	31	thie	796	11	ji
297	30	apnay	797	11	join
298	30	bat	798	11	kahin
299	30	har	799	11	kare
300	30	karna	800	11	khlaf
301	30	kay	801	11	khush
302	30	khud	802	11	lahore
303	30	only	803	11	lot
304	30	then	804	11	lye
305	30	thi	805	11	magar
306	30	try	806	11	makes
307	30	yad	807	11	march
308	29	..	808	11	morning.
309	29	hun	809	11	naseeb
310	29	hv	810	11	notes
311	29	hy.	811	11	p
312	29	hy?	812	11	phr
313	29	library	813	11	prayers
314	29	rahi	814	11	puri
315	29	see	815	11	reached
316	29	yeh	816	11	roz
317	28	5	817	11	search
318	28	abt	818	11	should
319	28	agar	819	11	silence
320	28	clas	820	11	subha
321	28	hazrat	821	11	success
322	28	hua	822	11	sun
323	28	know	823	11	surat
324	28	mat	824	11	sy,
325	28	people	825	11	tanzeem

326	28	remember	826	11	tara
327	28	smile	827	11	test
328	28	yr	828	11	tha,
329	27	every	829	11	theek
330	27	ge	830	11	unko
331	27	hain,	831	11	wa
332	27	jan	832	11	wapis
333	27	khan	833	11	wher
334	27	much	834	11	wishing
335	27	ok.	835	11	words
336	27	place	836	10	/
337	27	prayer	837	10	1.
338	27	rha	838	10	6
339	27	tera	839	10	@
340	27	yes	840	10	academy
341	26	:)	841	10	ah
342	26	aapko	842	10	andar
343	26	address	843	10	april
344	26	ao	844	10	baad
345	26	aya	845	10	banda
346	26	dekh	846	10	baqi
347	26	hain.	847	10	bhool
348	26	never	848	10	blood
349	26	rahay	849	10	bna
350	26	which	850	10	books
351	26	wht	851	10	brng
352	26	wish	852	10	came
353	25	abu	853	10	can't
354	25	apna	854	10	chand
355	25	bt	855	10	chor
356	25	bus	856	10	cmng
357	25	dost	857	10	colg
358	25	free	858	10	confirm
359	25	god	859	10	dimagh
360	25	hy!	860	10	doing
361	25	jaye	861	10	dun
362	25	meet	862	10	english
363	25	number	863	10	exam
364	25	pakistan	864	10	face
365	25	teri	865	10	gr8
366	24	c	866	10	hafiz
367	24	dia	867	10	hamari
368	24	fir	868	10	hate
369	24	forward	869	10	hath
370	24	gaya	870	10	hona
371	24	heart	871	10	hoty
372	24	it.	872	10	institute
373	24	kidhar	873	10	islami

374	24	next	874	10	iss
375	24	other	875	10	j
376	24	sorry	876	10	jany
377	24	sweet	877	10	jitna
378	23	(	878	10	k.
379	23	achi	879	10	karny
380	23	aple	880	10	kesi
381	23	bye	881	10	knw
382	23	data	882	10	koie
383	23	friend	883	10	krain
384	23	h	884	10	krty
385	23	hain?	885	10	laikin
386	23	kon	886	10	later
387	23	let	887	10	lekin
388	23	logon	888	10	liay
389	23	start	889	10	load
390	23	work	890	10	mery
391	22	bad	891	10	mubarik
392	22	bana	892	10	nam
393	22	bhai)	893	10	namaze
394	22	care	894	10	naraz
395	22	dena	895	10	net
396	22	frm	896	10	now.
397	22	great	897	10	phone
398	22	held	898	10	problem
399	22	help	899	10	problems
400	22	hm	900	10	pura
401	22	lo	901	10	raho
402	22	miss	902	10	rasul
403	22	most	903	10	really
404	22	nae	904	10	room1
405	22	nahe	905	10	salam.
406	22	paper	906	10	sey
407	22	tel	907	10	solve
408	22	than	908	10	soon
409	21	alaikum	909	10	study
410	21	because	910	10	survey
411	21	cal	911	10	ta
412	21	family	912	10	tarah
413	21	ha?	913	10	teacher
414	21	hw	914	10	tha.
415	21	iz	915	10	thay
416	21	kabi	916	10	thing
417	21	karo	917	10	time.
418	21	keh	918	10	u!
419	21	mean	919	10	umeed
420	21	name.	920	10	use
421	21	them	921	10	week

422	21	tk	922	10	wen
423	21	way	923	10	whn
424	20	:	924	10	world
425	20	attend	925	10	yaar
426	20	been	926	9	'
427	20	bi	927	9	(ameen)
428	20	bs	928	9	(name
429	20	gia	929	9	100
430	20	hn	930	9	2morrow
431	20	hope	931	9	2nd
432	20	jahan	932	9	4.
433	20	khuda	933	9	aapka
434	20	maghrib	934	9	aisay
435	20	nhe	935	9	amal
436	20	oh	936	9	aqal
437	20	pm	937	9	assalYam-o-alaikum
438	20	reach	938	9	assalamu
439	20	student	939	9	attendence
440	20	w8	940	9	available
441	20	why	941	9	awaz
442	20	you?	942	9	bacha
443	19	"	943	9	barish
444	19	,	944	9	bcoz
445	19	ali	945	9	becoz
446	19	beautiful	946	9	bilkul
447	19	bless	947	9	bomb
448	19	change	948	9	btaya
449	19	deta	949	9	chalain
450	19	es	950	9	coz
451	19	ghr	951	9	dair
452	19	had	952	9	death
453	19	hai,	953	9	dedicated
454	19	hen	954	9	deen
455	19	her	955	9	didn
456	19	hostel	956	9	doctor
457	19	hr	957	9	dude
458	19	jata	958	9	dukh
459	19	ker	959	9	dunia
460	19	khatam	960	9	even
461	19	last	961	9	farmaya
462	19	lay	962	9	father
463	19	maa	963	9	front
464	19	ni	964	9	ga,
465	19	right	965	9	gay.
466	19	she	966	9	gd
467	19	sunday	967	9	given
468	19	today.	968	9	hein
469	19	true	969	9	hom

470	19	wl	970	9	honay
471	19	you.	971	9	hoon
472	18	12	972	9	hostel2
473	18	4m	973	9	hostl
474	18	9	974	9	hstl
475	18	ahmad	975	9	hue
476	18	aoa,	976	9	hui
477	18	ati	977	9	hy,
478	18	aye	978	9	ik
479	18	bahir	979	9	islamabad
480	18	bar	980	9	it's
481	18	complete	981	9	jain
482	18	contact	982	9	jate
483	18	darse	983	9	kai
484	18	dat	984	9	karne
485	18	dr	985	9	karo,
486	18	friends	986	9	kha
487	18	got	987	9	lain
488	18	hay.	988	9	lazim
489	18	here	989	9	level
490	18	hoga	990	9	li
491	18	islamic	991	9	little
492	18	jazakallah	992	9	liya
493	18	job	993	9	ly
494	18	krtay	994	9	mail
495	18	mar	995	9	marks
496	18	masjid	996	9	marzi
497	18	matlab	997	9	mei
498	18	more	998	9	mobile
499	18	muje	999	9	mohsin
500	18	must	1000	9	muhabat

APPENDIX-D

300 Most Frequent Roman-Urdu Words in the Corpus

In Word List window option __*treat all data as lower case*_ activated
Global setting: Letter and number token classes activated
Words with common spellings in English and Roman-Urdu are not included in this list
AntConc Statistics for this analysis: Total No of word types: 11616
Total No of word tokens: 68275

S.NO	FRQ	ROMAN-URDU WORD	S.NO	FRQ	ROMAN-URDU WORD
01	959	k	151	32	bat
02	548	ki	152	32	dena
03	517	ho	153	32	yad
04	489	ko	154	31	bht
05	462	ka	155	31	har
06	450	hy	156	31	jan
07	444	hai	157	31	lo
08	332	hain	158	31	pak
09	311	se	159	31	rahi
10	296	Allah	160	30	aya
11	296	na	161	30	kay
12	290	ha	162	30	khud
13	255	nai	163	30	rha
14	199	ye	164	30	yeh
15	197	mai	165	29	abu
16	195	main	166	29	agar
17	178	kr	167	29	teri
18	176	ma	168	28	aapko
19	175	ga	169	28	apna
20	174	ap	170	28	hoga
21	172	aap	171	28	khan
22	159	hay	172	28	mat
23	150	ne	173	27	ahmad
24	145	kya	174	27	aye
25	142	kia	175	27	rahay
26	140	aoa	176	26	dekh
27	140	aur	177	26	dost
28	140	jo	178	26	fir
29	140	wo	179	26	islam
30	139	yar	180	26	jaye
31	134	kar	181	26	khuda
32	133	se	182	25	ali
33	123	tha	183	25	faraz
34	119	kal	184	25	kon
35	112	koi	185	24	achi
36	109	nahi	186	24	bana
37	103	ny	187	24	kidhar

38	98	aj	188	24	logon
39	98	eid	189	24	nahe
40	96	hon	190	24	tk
41	95	pr	191	23	bar
42	93	mein	192	23	gia
43	91	ke (91),	193	23	hoi
44	91	tu	194	23	matlab
45	91	ya	195	23	nae
46	90	sa	196	23	nhe
47	83	tum	197	22	dain
48	81	han	198	22	hr
49	80	ab	199	22	jahan
50	80	gi	200	22	ni
51	80	meri	201	22	ul
52	79	din	202	21	bi
53	78	sb	203	21	hu
54	75	ghar	204	21	jazakallah
55	74	bhai	205	21	kabi
56	73	hum	206	21	keh
57	71	pe	207	21	ker
58	71	raha	208	21	khatam
59	70	g	209	21	masjid
60	70	kuch	210	21	sakta
61	69	bhi	211	20	ati
62	69	quran	212	20	bs
63	68	kisi	213	20	darse
64	65	ek	214	20	deta
65	65	salam	215	20	es
66	64	q	216	20	hoa
67	63	aa	217	20	maghrib
68	62	kaha	218	20	muje
69	62	un	219	20	zindgi
70	61	liye	220	19	ay
71	59	dil	221	19	bahir
72	58	dua	222	19	gae
73	58	nhi	223	19	ghr
74	57	ja	224	19	inshaallah
75	57	jaldi	225	19	jata
76	57	mera	226	19	khana
77	56	gai	227	19	kiya
78	56	mere	228	19	lay
79	56	mubarak	229	19	mubarik
80	55	acha	230	19	pathan
81	54	gay	231	19	rab
82	54	pata	232	19	thy
83	52	hota	233	18	dunya
84	52	jis	234	18	eman
85	52	sath	235	18	hal

86	51	aaj	236	18	israr
87	51	hun	237	18	karta
88	50	de	238	18	krtay
89	50	kaam	239	18	la
90	50	namaz	240	18	laga
91	49	apni	241	18	mairi
92	48	jab	242	18	mar
93	48	jihad	243	18	saath
94	48	kam	244	18	saqi
95	47	baat	245	18	wali
96	47	hoti	246	17	app
97	47	jana	247	17	chal
98	47	krna	248	17	dya
99	47	per	249	17	farmaya
100	46	agr	250	17	gain
101	46	ana	251	17	hoon
102	45	aik	252	17	inshallah
103	45	kahan	253	17	itna
104	44	abi	254	17	jani
105	44	alaikum	255	17	kehna
106	44	kb	256	17	mje
107	43	apne	257	17	nay
108	43	ge	258	17	shadi
109	43	lia	259	17	suna
110	43	oye	260	17	wala
111	43	par	261	17	zindagi
112	42	jao	262	16	aapki
113	42	karain	263	16	ai
114	42	kis	264	16	ami
115	42	phir	265	16	dard
116	41	hn	266	16	dawat
117	40	ata	267	16	di
118	40	kro	268	16	gunah
119	40	sab	269	16	hafiz
120	40	thi	270	16	iqbal
121	39	karo	271	16	janab
122	39	rahe	272	16	jannat
123	38	jb	273	16	lain
124	38	kab	274	16	manzil
125	38	karna	275	16	mgr
126	38	mujy	276	16	mn
127	38	tm	277	16	muj
128	38	waqt	278	16	mushkil
129	37	usra	279	16	nikal
130	36	gaya	280	16	pehly
131	36	le	281	16	qaim
132	35	hazrat	282	16	si
133	35	hua	283	16	ta

APPENDIX-E

List of All Symbols Used in the Corpus (Total: 30)

Serial No.	Symbol	Frequency in Corpus
1.	=	69
2.	+	47
3.	>	40
4.	¤	38
5.	|	20
6.	<	17
7.	~	15
8.	§	14
9.	$	12
10.	||||	10
11.	£	10
12.	^	8
13.	=>	5
14.	>>	5
15.	¥	4
16.	═══	3
17.	$£	2
18.	<<	2
19.	═<	2
20.	>═	2
21.	¤¤	2
22.	¤¤¤	2
23.	<~	1
24.	═	1
25.	^^	1
26.	``	1
27.	~>	1
28.	~¤	1
29.	¤~	1
30.	§§	1

Note. Symbol token class activated in AntConc