Meaning Making Through Minimal Linguistic Forms in Computer-Mediated Communication

Muhammad Shaban Rafi

Abstract
The purpose of this study was to investigate the linguistic forms, which commonly constitute meanings in the digital environment. The data were sampled from 200 Bachelor of Science (BS) students (who had Urdu as their primary language of communication and English as one of the academic languages or the most prestigious second language) of five universities situated in Lahore, Pakistan. The procedure for analysis was conceived within much related theoretical work on text analysis. The study reveals that cyber-language is organized through patterns of use, which can be broadly classified into minimal linguistic forms constituting a meaning-making resource. In addition, the expression of syntactic mood, and discourse roles the participants technically assume tend to contribute to the theory of meaning in the digital environment. It is hoped that the study would make some contribution to the growing literature on multilingual computer-mediated communication (CMC).

Keywords
computer-mediated communication, meaning, morphemic reduction, syntactic reduction, mood

Introduction
Computer-mediated communication (CMC) is proliferating in the lives of most people today. The ubiquity of CMC invites interesting debate on the phenomenon of meaning making or success of communication in the digital environment. As a result, communication theorists (Baron, 2008; Crystal, 2001, 2006; Herring, 1996; Thurlow, Lengel, & Tomic, 2004) have been investing considerable time to investigate the emergence of digital meaning. This scenario seems to be shaping if not determining many aspects of our real life, as has been argued by Turkle (2011). Herring, Stein, and Virtanen (2013) further assert that the Internet mirrors co-construction of the self in an ad hoc manner. The transformation of “the self” from real to virtual began with the invention of electronically mediated communication, such as through the telegram, telephone, fax, and so on, phenomenon that is almost a century and half old. However, the virtual self has attracted much attention only in recent years. This transformation manifests an evolution of meaning-making self, which is expressed through such mediated communication.

One assumes that there is a good reason to believe that the expression of meaning is fundamental to language. The present study hypothesizes that certain mechanisms, primarily morphemic and syntactic reduction, are one of the ways used to achieve this end. We also claim that the use of such minimal mechanisms, more obviously in the context of CMC, is not random, but rather, it is necessitated by the expression of meaning under the special circumstances in question. It may sound as though we hold deterministic attitude toward the relationship between form and meaning, but our intension is simply to claim the priority of meaning over form.

Much work by the researchers of CMC (Danet & Herring, 2007; Segerstad, 2002; Tagg, 2009; Thurlow & Poff, 2013, as well as those referred previously) shows a split approach, either taking formal reduction or minimization to be an act of linguistic transgression or considering this kind of variation to be an inherent property of language, as had always been evident in the nature of historically older forms of reduced language, for instance, diary register (Weir, 2012, and the references cited therein). Crystal (2006) asserts that CMC heralds novel manifestations of this property mirroring linguistic behavior, in ways primarily different, from the traditional modes of communication as has been noted by Bodomo (2009). Rafi (2010), arguing on the same lines, states that Internet users are habitually and increasingly customizing language to capture their experiences and to express their e-identity through various linguistic innovations.

While discussing grammatical features of English, Ko (1996) finds that electronically governed communication typically involves relatively short or shortened words. While

---

1University of Management and Technology, Lahore, Pakistan

Corresponding Author:
Muhammad Shaban Rafi, University of Management and Technology, C-II, Johar Town, Lahore 10033, Pakistan.
Email: shabanrafi@hotmail.com
explaining these features, he shows that the use of present tense, coordinations, adverbials, demonstrating pronouns, and intensifiers is used more frequently in real-time or online communication than in any other type of discourse. However, grammatical forms such as prepositional phrases, relative clauses, and perhaps subordinate clause in general are comparatively infrequent in CMC discourse. Furthermore, he claims that discourse markers and hedges are, likewise, less common and the patterns of turn taking are far from neat and orderly. He also mentions this proportionately the huge role that icons play in CMC. Kalmann and Gergle (2009, 2010) assert that repetition of punctuation and letters indicates the stretching of a word, emulating a stretched out syllable like in spoken conversation. Furthermore, they explain that these repetitions tend to communicate tempo, pitch, prosody, and other paralinguistic elements for achieving visual emphasis. Reporting similar results, Herring (2011) additionally points out the prevalence of “nonstandard” typography and orthography practices. Danet and Herring (2007) note that research on other languages has exhibited similar tendencies.

Probably, it is no longer the case that the Internet communication is predominantly in English. Bodomo (2009) asserts that bilingualism/multilingualism has now become the norm in CMC. As the Internet has been increasingly becoming multilingual, researchers have explored new patterns of use and language combination in bilingualism/multilingualism communities (e.g., Androutsopoulos, 2007; Axelsson, Abelin, & Schroeder, 2007; Barasa, 2010; Durham, 2007; Paolillo, 2007; Sargeant, Tagg, & Ngampramuan, 2012; Warschauer, El Said, & Zohry, 2007). The Urdu spoken in Pakistan, largely adopting the linguistic ecology of English (Rafi, 2013), is case in point. Thus, the use of Urdu on the Internet is always embedded within a larger Anglophone context. Not infrequently involving switches between the two languages or substitution in writing of an English letter like b for the whole corresponding semihomophonous Urdu word more specifically which means “also.” The substitution noted above does not lead to any change in meaning; it does, though, just help the users make meaning by means of minimal linguistic forms.

In general, the present study intends to investigate forms that are reduced or have unique configurations, and which are exploited to project the same range of meaning as corresponding full forms. In section “Method,” the methodology and theoretical as well as the linguistic backgrounds are outlined. The remaining sections except the concluding one present an analysis of the data.

Method

Much like researchers dealing with any kind of data, researchers who deal with digital data are often confronted with a variety of non-trivial questions. Most of them relate to the size and representativeness of data samples, data processing techniques, limitations of genres, kind and amount of necessary contextual information, and ethical issues, such as anonymity and privacy protection. Keeping such questions in mind, this section describes the procedures of data collection, ethical considerations of collecting and handling data, data analysis, and the theoretical underpinnings.

Data Collection

The first stage of data collection consisted of the selection of a sample of appropriate participants. The sample included both male and female students who were between 18 and 24 years of age and were registered in the Bachelor of Science (BS) program in five different universities situated in Lahore, Pakistan. A pilot study suggested that this cohort was indeed suitable for the investigation under discussion (see, for example, Ilyas & Khushi, 2012; Leppänen, 2007; Rafi, 2008, 2010, 2013). This study assumes that CMC is typical of this age group almost anywhere in the world.

According to Grinter and Eldridge (2001), young people bring with them to college a well-developed practice of e-styles. As social networks found relatively young people its potential users, the present study assumes that they know how to accommodate and appropriate the English language in CMC. Jørgensen (2001, as cited in Leppänen, 2007 ) argues that youth can take language in their own possession and typically active in generating more general language change. Researchers (e.g., Leppänen, 2007; Paolillo, 2011; Peuronen, 2008) assert that there is always association between being young and using new patterns of English along with the native language.

Facebook being popular forum for conversations among young people was assumed to be a good source of the data (cf. Boyd & Ellison, 2007; Raskin, 2006; Sengupta & Rusli, 2012). The data collection was confined to Facebook wall, which gives easy access to enormous subject pool, without violating anyone’s privacy. Furthermore, it is characterized by the use of a broad range of traits at the very informal end of the linguistic spectrum like in a natural setting, which is the primary focus of the present study.

Sample. The sample was drawn from the five private (by and large not funded by the State) institutions of higher learning listed in Table 1. The reason to choose these institutions was to make the sample as representative of BS students as possible. In terms of academic, linguistic, and cultural background, the sample was homogeneous. All the participants were from BS program but strictly speaking in different disciplines (e.g., Business of Administration, Engineering, Computer Sciences, and English) and academic years (e.g., covering first year to fourth year). As the participants were between 18 and 24 years of age, they might be considered young people. Youth is defined here as a chronological age—number of years since birth (cf. Leppänen, 2007). Also, communication on Facebook was more or less within the same group. Notwithstanding the possibility of this forum to
connect people from different linguistic backgrounds, the participants in this study were evidently Urdu/English bilinguals. It should be added that, strictly speaking, the first language of most of them was Punjabi but Urdu was their primary language of communication, both at home and elsewhere, and English was their most important academic language and the most prestigious second language. They acquired Punjabi, however, through informal contact with, for example, lower or middle lower working class—who speaks it maybe to mark solidarity and identity. They used to meet with each other both in offline and online contexts. Moreover, their communication was not only limited to Pakistani contexts but it could be referred to international settings also. The participants linked and shared activities covering information exchange, debate, problem solving, exchanging picture, jokes, and video related to varying themes, such as greeting, politics, showbiz, sports, and sex. Peuronen (2008) argues that these activities “best describe their being” (p. 104). Their communication in online context comprised new creative ways. Importantly, they identified themselves mostly either through Urdu or English or both. The data were mainly in Romanized Urdu, English, or a mixture of Urdu and English. The corpus used was simply in multiple languages, all written in the same (Roman) script. Tone used in the message threads was typically informal; however, a number of snippets seemed to be formal too. A lot of words that the participants used were clearly from their specific academic context.

Five volunteers who were also students at these institutions were engaged for help during the process of data collection. Each one of them coordinated the data gathering process for one of the five institutions. The researcher shared with them the purpose and ethical boundaries of the study. Each of them managed to add on average 375 students over a period of 2 months. Thus, the researcher had access to all the students through these volunteers. The data from only 50 participants from each institution were analyzed on the basis of the quantity of their posting output. Table 1 shows the distribution of the sample over the five institutions. The study investigated linguistic postings of each participant transmitted over the period of a week. The logic behind collecting the whole week’s data was to observe most of the linguistic features that the participants used in their communication. Each posting consisted of usually more than one utterance. Thus, the data collected were naturalistic and observational, with minimum interference from the researcher. However, the participants were asked questions to seek clarifications where it was absolutely necessary.

**Nature of the data.** As shown in Table 2, the data were collected from 200 participants whose average age was 21 years. The data indicate that each participant spends on average 6.25 hr on Facebook per week and hosts on average 150 followers, thus creating the likelihood of feedback on his or her posting. Of 2,516 linguistic postings, 588 are in Romanized Urdu; 1,135 are in English; and the remaining 793 postings are a blend of both Urdu and English. This shows an increasing trend of English along with mixed Urdu and English in digital discourse. As is evident from Table 2, the participants have shown less and marginal need for Urdu as a medium of communication. The most plausible reason behind the marginal use of Urdu is perhaps the importance given to English in academia. Importantly, English is their most prestigious second language. They come across relatively fewer opportunities to use written Urdu in their academic context, for example, examination, project, assignment, and so on. On average, each posting consisted of 137 words ranging between 1 and 612 words across the data. As many as 27,476 words are accumulated for the analysis.

Much like face-to-face conversation, communication between the participants reflected physical approximation. The illustrations in [1] confirm that the distant or even near future marking system is marginal in the context of CMC. Overall, the data included the postings, which demonstrated participants’ relationships regarding how close they were with their followers.

**Table 1.** Distribution of Sample.

| Institution                                      | Number of participants |
|-------------------------------------------------|------------------------|
| COMSATS Institute of Information and Technology | 50                     |
| National University of Computer and Emerging Sciences | 21                    |
| Superior University                             | 50                     |
| University of Lahore                            | 50                     |
| University of Management and Technology         | 29                     |
| Total                                           | 200                    |

**Table 2.** Demographics and Nature of Data.

| I. Total number of participants                  | 200        |
|-------------------------------------------------|------------|
| a. Average age                                  | 21 years   |
| b. Average time spent on Facebook in a week     | 6.25 hr    |
| c. Average number of followers of each participant | 150        |
| II. Total number of linguistic postings in the data |           |
| a. Roman Urdu                                   | 588        |
| b. English                                      | 1,135      |
| c. Mixed Roman Urdu and English                 | 793        |
| III. Total number of words in the data          | 27,476     |
| a. Range of words in a posting                   | 1-612      |
| b. Average number of words in a posting          | 137        |

[1] a. <tum kahn gum ho bhai . . . ?> (Where are you brother?)
b. <Guys kia plan hai iss Sunday ka ??> (Guys, what is plan on this Sunday?)
c. <wat an awsome dance . . . !!!> (What a dance it is!)
d. <girl . . . is an innocent nd beautiful creature . . which only deserves to be luvd . . . ♥> (A girl is an innocent and beautiful being who deserves only to be loved.)
e. <Gud Morng To all my chintoo mintoo frndzzZZzzzz:)>(Good morning to all my dear friends.)
f. <2mrw morning em going near xxx on xxx mangni . . . yahuuuuuuu> (I am going to attend xxx’s engagement near xxx tomorrow.)

Types of exchanges. Interactivity is a defining characteristics of CMC, which is organized around topics or threads. Barnes (2003) defines that “a thread is a chain of interrelated messages that respond to each other” (p. 20). A linguistic posting on the wall can be generally described as fragments, phrases, and clauses. The postings on the wall evolved in a coherent whole around topics of varying themes. Mostly, a topic was initiated by a single person followed by his or her followers’ comments, which was either closed with words of thanks or reflection from the originator or left without proper closing. The exchanges on the wall can be classified into a three-part structure of Initiation, Response, and Reflection (IRR). This IRR structure can be seen in the following example:

[2]
A: <newly made my own pic>
B: <its awsome bro>
A: <thanks>

Unlike IRR structure, instances of dialogue structures, two-part adjacency pairs (e.g., question–answer), and non-linear flow of conversation were also prevalent. The dialogue structure and adjacency pair can be seen in [3] and [4]. The postings in the data were extrovert in nature and each provided a framework within which the next was formulated.

[3]
A: <sick of formatting now . . . (>  
B: <m sick of dissertation now>  
C: <Ahhhh!!>  
D: <aww!!! my poor baby>  
E: <O hoo . . :)>  
A: <there is nothing more killing than formatting more than 50 times line spacing . . . margins . . . pagination . . . ToC, LoT, LoF . . . wot not and wot not . . . there is nothing more mechanical and boring than this . . . > :(  
F: <do not be fed up now . . . its a continuous process of accomplishing task.>  
A: <have been taking breaks and delaying it . . . first deadline is over now wanna get it over and forever>

A: <i hate this document now . . . dont wanna even open it up>
[4]  
A: <wt r u doing now a days?>  
B: <BS social sciences 4m xxx.>  
A: <how abt xxx?>  
B: <he/she too with me.>

Although Facebook is thought to be an asynchronous technology and it often transmits messages in near-real time, many participants were observed replying instantly to postings, rendering the technologically asynchronous medium effectively synchronous. Thurlow and Poff (2013) argue that with the convergence of new (and old) media, technological boundaries and generic distinctiveness of instant messaging, texting, and emailing are becoming blurred. Notable examples of this are found in micro-blogging (e.g., Twitter and status updates on Facebook), the multi-functionality of smart-phones (e.g., BlackBerry) and to some extent, Apple’s iPhone. This study assumes that dividing Facebook conversations on the wall between synchronous and asynchronous modes may be misleading because the speed of communication mostly inhibits us to draw a clear line between synchronous and asynchronous communication. As there is a range of possibilities, we may consider the participants’ conversation on the wall as falling between synchronous and asynchronous modes. But strictly speaking, this study considered Facebook communication on the wall analogous to asynchronous communication.

The participants used to grid a kind of coherent discourse because most of their postings were in response to a single individual’s reflections. It was found that communication on the wall usually occurred between one-to-one or one-to-many—a kind of interpersonal communication. Barnes (2003) reinforces the fact that the Internet can be used to distribute messages in any number of directions, that is, one-to-one or one-to-many.

Ethical considerations. Ensuring that pre-existing ethical standards were properly met was crucial. In this study, we adhered to guidelines suggested by Mann and Stewart (2000) while collecting data.

The participants were informed about the nature of the study. They were given assurances regarding confidentiality, security of information, and unauthorized eavesdropping; that is, information that might identify the participants, places, institutions, and times was never to be disclosed. The participants were identified by means of cryptonyms in reporting research. Access to the database was restricted to the participants and the researcher. However, the researcher could not forbid the use of racist and sexist language, and other contentious and provocative material. Mann and Stewart (2000) argue that Internet research does not have to conform to these restrictions.
**Data Analysis**

To attempt the research question, minimal linguistic forms and what they communicate in terms of mood were analyzed. Jørgensen, Karrehøk, Madsen, and Møller (2011) remark that linguistic features are best suited as the basis for the analysis of language in polylanguaging context. These forms were classified into lexical and syntactic tiers. Word reduction was further subclassified into abbreviations (e.g., *gf* for “girlfriend”), clippings (e.g., *pic* for “picture”), and logograms (e.g., *somed* for “someone”). For clarity, logograms were further subclassified into phonetic spelling (*u* for “you”), lexo-numeric (*gr8* for great), digito-lexeome (*2morow* for “tomorrow”), and digit word semi-homophone (*4* for “for”). The purpose of these classifications and subclassifications was to tabulate precisely the lexical features for measure of frequency. Frequency was calculated to gauge how many times a particular feature occurred and to suggest its permanence in CMC. The frequency of occurrence was determined if at least a word was repeated twice within a conversation and a minimum of 5 times in the whole data.

What CMC-unique structures are, and how they uncover functional roles, I examined (a) types of structures, (b) deletion of grammatical features, and (c) syntactic expression of mood. Types of structures were further classified into fragmented structure, simple structure, compound structure, and complex structure. To analyze whether or not the proportions of structure type are different across the languages, I calculated chi-square test of independence. As the characteristic of linguistic reduction was assumed as one of the key features across the corpus, grammatical properties, which were commonly deleted, were measured. This measure provided us with further insight to assess how far the participants compromised on structural rules when communicating in digital environment. Furthermore, a careful analysis of structures constituting mood (e.g., declarative, interrogative, imperative, and exclamatory) was examined to uncover functional roles and negotiation common to Facebook.

In addition to analysis of quantitative data, I also drew on the message threads to elaborate and support my verdict concerning the research questions, which helped bring triangulation to increase the credibility and validity of the results. These snippets were in English, Romanized Urdu, and mixed English and Urdu. They were demonstrated within mathematical symbols (such as < >) along with their transliteration in the parentheses. In relation to paralinguistic features, I considered their visual aspects only as a meaning-making resource in a message thread. Thus, the analysis was backed up by a large corpus covering both quantitative and qualitative data sets, which assisted in attempting the research questions with more extensive analysis, with far more participants and more rigorous sampling procedures.

**Theoretical underpinning.** The procedure for analysis was conceived within much related theoretical work on text analysis. Sinclair (1991, as cited in Stubbs, 1996) argues that intuitive judgments are particularly untrustworthy with respect to frequency and distribution of different forms and meanings of words, and to the interaction of lexis, grammar, and meaning. In the present study, against the reliance on invented data, example utterances were explicitly taken from the data. The accountability to data was thus taken into account. Unlike contemporary studies, which are based on a small data set, a large corpus based study was planned to address the underlying research question. The actual language text duly recorded was the main concern in the present study. In addition, interpretation of cyber-linguistic features was based on the whole data rather than relying consistently on a few examples.

There is reason to believe that a few linguistic features of a text are distributed evenly throughout. To encompass the whole data, the study recorded the frequency of commonly occurring minimal lexical as well as syntactic features across the corpus to draw emerging patterns. One of the main uses of the corpus is to identify what is central and typical in the language (Sinclair, 1991, as cited in Stubbs, 1996). Thus, linguistic patterns that reside in a language can best be judged when they are studied comparatively across the text corpora.

As noted above, the data were accumulated from five groups. The purpose behind the selection of different cohorts was to have a comparative standpoint whether linguistic features were spread across the board or concentrated in a particular cohort. Overall, the data were organized into lexical and syntactic tiers. Stubbs (1996) argues that the most powerful interpretation emerges if comparisons of texts across corpora are combined with the analysis of the organization of individual texts. Another approach would have been to view linguistic features in several different media genres (e.g., email, social networking, and text messaging), though compatible with this principle, seems to motivate relatively different linguistic choices (see, for example, Baron, 2008; Crystal, 2001, 2006; Thurlow & Poff, 2013) on formal and/or informal use of the linguistic spectrum. Whereas the primary focus of this study was to collect naturalistic and observational data that were limited to Facebook communication on the wall.

Every language is governed by linguistic selections and restrictions. For example, both Urdu and English pursue linguistic conventions to transform and transmit a structure. We can, of course, say how are you? or *Ap ka kia hal hai?* But we may not say you are how or *Ap hai kia hal ka.* Hence, grammatical structure restricts the lexis that occurs in it, and conversely any lexical item can be specified in terms of the structures (Stubbs, 1996). The study presupposed that the linguistic flexibility that young people exercised in the mediated communication was based on the system of language more or less; however, it was the meaning that governed their apparently peculiar linguistic choices.
Semantic analysis of corpus has been of main concern in the Firthian school of thought. As Firth (1957, as cited in Stubbs, 1996) puts it, “You shall know a word by the company it keeps” (p. 11). Firth’s notion of form and meaning in context was extended by Halliday (2002, 2003, 2004, 2005, 2007) in systemic linguistics. Systemic linguistics adopts a descriptive approach to language investigation that answers the questions: What is language? and How does language work? In the same way, the present study gives importance to the description of the characteristics of cyber-linguistic features. It may sound as though we hold deterministic attitude toward the relationship between form and meaning, but our intention is simply to claim the priority of meaning over form.

**Minimal Linguistic Forms**

**Word Reduction**

Linguistic reduction was the most common feature in the data. The participants mostly reduced linguistic forms from Urdu and English. However, it was English which acquired the impact of linguistic reduction the most. They minimized, seemingly either through conscious or subconscious efforts, words by compounding a number and lexeme or morpheme, for example, *some1* for someone, *gr8* for great, *b4* for before, and so on. Words that are composed of two syllables in which one syllable is substituted with a digit and the other remains constant may be called *lexo-numeric*. Conversely, the participants also customized words, may be labeled *digito-lexeme*, by substituting segment or segments of a word with a digit at the onset, for example, *2morow* for tomorrow, *4get* for forget, and so on. Among other categories of reduction, *digit word homophones* were prevalent, which were formed by replacing a full word with a digit, for example, 1 for “one,” 2 for “to” or “too,” 4 for “for,” and so on. As many as 1,106 words were tabulated in these categories, which can be grouped under the basic term, namely, logogram. Thus, the logograms, in many ways, were the most frequently occurring features in cyber-communication of young Pakistani students.

The initial ambiguity that might have been as a result of minimal linguistic forms was seemingly compensated, if not altogether but mostly, through extralinguistic features. Apart from this, linguistic reduction can further be explained within the system of language. Furthermore, segmental reduction in English and Urdu is systematic, which indicates supposedly uniform patterns for the success of communication.

Reduction in the segments of English words can be characterized by phonological and morphological properties. As shown in Table 3, the derived forms conform to fixed phonological patterns, which are characterized by templates in which “C” stands for consonant sound and “V” represents vowel sound. Moreover, bold letters, for example, “C” stands for the sound that was picked up while committing the reduction in the original word. As is evident from the given template, vowel sound or vowel-like sound is more susceptible to reduction. Vowel sounds seem to behave like the fillers in consonant sounds. We may speculate that English words can be read without vowels exerting a little effort. Conversely, there is no instance of reduction in consonant sounds; however, the consonant sounds which were reduced are primarily from consonant clusters. Hence, the consonant sound was used to cover the rest of the segment in a word. Moreover, diphthongs and weak vowel sounds, especially /ə/ to show a sound occurring at various segmental positions in a word, were occasionally truncated. As highlighted in the given template, it is the consonant sound in a word which the primary carrier of a meaning is. This can be compared with early Semitic languages such as Arabic and Hebrew in which the words’ roots were isolated sets of consonants, for example, in Arabic the root word meaning “write” had the form *k-t-b*.

As noted above, the participants mostly reduced English forms; however, they simply substituted bi-syllabic Urdu words with the whole corresponding mono-syllabic English homophones or semi-homophones—may be considered a creative way of reducing morphemic properties of Urdu primarily in Romanized script. While substituting Urdu forms with its English counterparts, the overextended sound had no or less homophonous correspondence in some instances. It is evident that the substitution does not cover aspirated sound because this requires use of superscript *h* that the participants simply avoided.

| Base form | Template | Derived form |
|-----------|----------|--------------|
| Am        | VC       | m            |
| And       | VCCVCC   | n/nd         |
| Be        | CV       | b            |
| Because   | CVCVCC   | bcz          |
| But       | CV       | bt           |
| Comment   | CVCVCC   | cmnt         |
| Good      | CV       | gd           |
| Hospital  | CVCCVCCVC| hsptl        |
| Love      | CV       | lv           |
| Now       | CV       | nw           |
| Please    | CCVC     | plz          |
| tension   | CVCCVC   | trshn        |
| Thanks    | CVCCC    | thnx         |
| That      | CV       | tht          |
| Was       | CV       | ws/wz        |
| What      | CV       | wt           |
Table 4. Substitution of Urdu Alphabets with Corresponding English Phoneme.

| Urdu alphabets | English phoneme |
|----------------|-----------------|
| ا، اغ | /a/ |
| ث، س، ص | /s/ |
| د، ڈ | /d/ |
| ر، ڑ | /r/ |
| ز، ز، ض | /z/ |
| ت، ث | /t/ |
| ق، ک | /k/ or /q/ |
| غ، گ | /g/ |
| ح، ه | /h/ |

[c], pronounced as /s/ is a homophone of which means “also” or “like” or “of” or “for”
[i], pronounced as /ai/ is a homophone of which means “coming”
[g], pronounced as /dʒi:/ is a homophone of which means “yes”
[k], pronounced as /ket/ is a homophone of which means “that”
[q], pronounced as /kjuːt/ is a homophone which means “why”

The logic behind these substitutions is that all the above-mentioned Urdu bi-syllabic words have the same or nearly the same sound in English alphabets. It seems as a matter of ease the participants replaced bi-syllabic forms with their mono-syllabic counterparts in their conversations. Similarly, they also substituted the following English words with their counterparts in Urdu.

[a], is pronounced as /əː/ is a homophone of which means “come”
[gay], is pronounced as /get/ is a homophone of which means “went” or “past forms of will (would)”
[hi], is pronounced as /hɪ/ is a homophone of which is an intensifier
[her], is pronounced as /hɛr/ is a homophone of which means “every”
[key], is pronounced as /ki/ is a homophone of which means “what”
[log], is pronounced as /lɒɡ/ is a homophone of which means “people”
[may], is pronounced as /meɪ/ is a homophone of which means “in”
[or], is pronounced as /ɔː(r)/ is a homophone of which means “more”
[pass], is pronounced as /pæs/ is a homophone of which means “near”
[pay], is pronounced as /peɪ/ is a homophone of which means “on”
[say], is pronounced as /seɪ/ is a homophone of which means “from”
[such], is pronounced as /ʃʌtʃ/ is a homophone of which means “true”

Table 4 shows that another very interesting phenomenon, may be called reduction, is the substitution of Urdu graphemes with English phonemes as has been noted by Rafi (2013) and Ahmad (2011). There are 38 alphabets in Urdu but in Romanized Urdu 66% of them are reduced to 24%. Around 58% of Urdu alphabets find seemingly semi-homophonous corresponding letters in English. Mostly involved substitutions, between Urdu alphabets, for example, /q/, /k or q/, and /g/, respectively. However, switches between two languages in writing Urdu letters, for example, /h/ and /h/ for the whole semi-homophonous English phoneme, for example, /s/, /z/, /t/, and /h/, respectively, were not infrequent. Surprisingly, the participants usually reduced English vowel sounds; however, when Romanizing Urdu, they preferred to minimize mostly if not always Urdu consonant sounds. More or less, these instances seem to be a reflection of linguistic accommodation, which can be used to support the verdict that CMC is a new discourse. Trudgill (2003) defines accommodation as “the process whereby participants in a conversation (usually face-to-face interaction) converge their accent, dialect, or other language characteristics according to the language of other participant(s)” (p. 3). Although convergence of Urdu graphemes with their English counterparts is apparently due to Roman transliteration of the Urdu language, this leads us to reveal that Urdu alphabets are reduced more or less to the size of English which are supposedly compensated with normal key stocks—to avoid complex application (shift, alt, and shift & alt) that these graphemes might have required. There is a fair chance that this trend may continue and consolidate with the present keyboard features (see, for example, Sperlich, 2005). Investigation of technological limitations and their impact on languages have access to Internet can be an interesting study to gauge how widespread the phenomenon is! However, investigation of this dimension is beyond the scope of the present study.

The second most noticeable lexical feature is the reduction of words to their initial letters. They are known as abbreviations. Abbreviations are commonly used and an accepted trend of writing the initials of words to make a new word that is not essentially pronounceable. However, there are instances of abbreviations, which obey phonetic property, for example, AFAP for as far as possible, SOB for son of bitch, Yo for your own, and so on. Abbreviations are usually spelled in capital letters; however, there are cases where the strings of words were abbreviated in lower case letters also. Like reduction in the base forms, abbreviations also involve loss of material. The principle of orthography is, however, of central importance in abbreviations, for example, dp for digital picture, gf for girlfriend, np no problem, and so on. As mentioned above, in limited cases, phonetic properties were also applied to derive abbreviations known as acronyms, for example, asap for as soon as possible, lol for lot of laughter, afap for as far as possible, and so on.
There are words that were derived from the first part of the base word. This process is labeled as clipping. Clippings appear as a mixed bag of forms reduced from base forms, which express familiarity with the denotation of the derivative. Thus, pic was used typically by the participants to refer to digital image and bro was part of their vocabulary to show probably an intimate relationship. There are clippings, for example, add/addy for address, cos for because, fav for favorite, grats for congratulations, del for delete, dif for different, jus for just, moro for tomorrow, rehi for hello again, uni for university, and web for website, which are characteristic features of cyber-communication.

Another unique way of segmental reduction is “g” omission in words ending with “g.” Perhaps, the participants thought it an additional sound while texting words ending with “g.” In native-English especially in some parts of the United Kingdom, for example, Norwich, Cardiff, and elsewhere, “g” dropping in colloquial speech is pervasive, as if there is simply an [n] on the end. This is indicative of some interesting facts such as migration of “g” dropping in the conversation of the participant that can be used as a correlate of foreign culture. This seems to explain that the so-called linguistic borders are melting down and spreading linguistic forms of the dominant language, which is English in our case, through new media communication.

**Types of Structures**

The participants used typically fragments and sentences (i.e., simple, compound, and complex) to carry out their conversation. Fragment can be referred to a segment of a sentence that may contain a single word or a string of words. The examples [4a-4k] indicate fragments consisting of a single word and a string of words. Three dots on either sides of the structure indicate a range of missing syntactic elements, for example, head word, verb (especially auxiliary verb), and complement. However, a sentence can be classified into a simple sentence (that contains a subject and a predicate), a compound sentence (that contains two independent clauses joined by a coordinator), and a complex sentence (which consists of an independent clause joined by one or more dependent clauses). The snippets [4l-4n] show simple structures consisting of a subject and a predicate. The example [4o and 4p] highlights types of compound and complex structures present in the corpus. Table 5 shows that the use of fragmented and simple structures is common. The data show 45% use of fragments followed by 44% use of simple structure across the corpus; however, compound and complex structures constituted relatively a low percentage. Chi-square test of independence shows p value (.03), which is less than .05; we can reject the null hypothesis, and say that types of structures and languages are related. I have further explained the association between types of structures and languages in the following section.

**Syntactic Reduction**

As discussed above, structural properties of English were found to be more susceptible to deletion. Nevertheless, the participants were relatively diligent about structural properties of Urdu. Figure 1 indicates syntactic constituents, which were regularly deleted. The structural properties of English, for example, pronoun and verb, were more frequently deleted than article, preposition, and inflection. Deletion of pronominal especially first person “I” was common. The participants frequently deleted “I” where it was preceded by an auxiliary “am.” They occasionally truncated “I am” to “am” or “m” or simply omitted “I,” for example, . . . will do asap . . . m having my exams!! . . . hope u understand!! . . . love you mwaaaah. Insertion of three dots signifies deletion of pronoun “I.” Similarly, they omitted auxiliary verbs, such as is, are, am, was, and were in their conversations. Given the structural deletion, there was omission of capitalization in the beginning of an utterance or in the case of proper nouns and addition of toggle case. As noted above, only a few

| Table 5. Frequency/Percentage of Types of Structures. |
|---------------------------------|-------|--------|-------|-------|
| Language          | Fragment | Simple | Compound | Complex |
| Urdu             | 231 (6%) | 554 (16%) | 64 (2%) | 43 (1%) |
| English          | 870 (25%) | 713 (20%) | 66 (2%) | 110 (3%) |
| Mixed Urdu and   | 493 (14%) | 283 (8%) | 53 (1%) | 68 (2%) |
| English          |        |        |         |      |
| Percentage       | 45     | 44     | 5       | 6     |

Note. Chi-square test of independence (p = .03 < .05).
instances of cohesive ties were found across the corpus. As a result of this, communication was more a reflection of short structures then supported by linguistic nexuses. The omission of linguistic features shows that the participants might have used their pragmatic knowledge to presuppose that the receiver would know how to map out deleted expressions in their utterances.

The snippets in [5] show that utterances, even though shortened, are coherent and meaningful. Apart from obligatory words, the participants omitted optional words by assuming that obligatory words might be sufficient to express meanings. However, the English language resists omission of obligatory strings of words because they hold meaning and omission may damage intelligibility. In my observation, the deictic expressions, main verb, and attributive forms seem sufficient for the projection of deleted string of words. There is reason to believe that even though their communication was not structurally rich, concurrent exchanges revealed somehow that the participants might have inferred the deleted string of words for the success of communication. It is logical to argue that instead of structure, meaning was assumed to be a primary source of carrying a message. How minimal linguistic forms construe syntactic expression of mood and overall relevant to the theory of meaning making is examined in the next section.

[5]

a. <yea I wid U> (Yes, I ... with you.)
b. <examz over . . . > (Exams ... over . . . )
c. <i think paper little bit short> (I think . . . . . . . short paper.)
d. <you so qute> (You ... so cute.)
e. <lyf goin gr8> (Life ... going great.)
f. <me cooking food for u> ( . . . cooking food for you.)
g. <me playing> ( . . . playing.)
h. <one of my favorite songs!> ( . . . one of my favorite songs!)
i. <awesome lyrics!!> ( . . . . awesome lyrics!)
j. <sometimes nothing to say> (Sometime . . . nothing to say.)
k. <more options?> ( . . . more options.)
l. <Sorry network problem> ( . . . . . . . network problem.)
m. <now feelIng bettEr after> (Now . . . feeling better.)
n. <not geTTing sleEp ... > ( . . . . . not feeling asleep.)
o. <BoRING LIFeeeE . . . !!!! > (just siTing alone ... : ( ( : ( . . . . . boring life . . . . just sitting alone.)
p. < . . . really missing u :(-:-) ( . . . . . really missing you.)
q. <this iz make by me> (This is make (past participle) by me.)
r. <she like it> (She like (inflectional “s”) it.)
s. <Please not disturb> (Please . . . not disturb.)
t. <u not understand my prob . . . > (You ... not understand my problem . . .)

Figure 2 reflects a classification of commonly deleted syntactic forms in the linguistic repertoire of Urdu/English bilinguals. The sign {Ø} in the figure nuances for the absence of a feature. As noted above, we did not find deletion in grammatical features of the Urdu language apart from superficial realization of Urdu graphemes. As shown in the figure, it is the English language that seems to give provision to deleting structural properties in digital discourse. It seems that the participants might have presupposed, of course to a certain extent, linguistic expressions which were deleted regularly. There is a fair reason to believe that the structural nativization of English is now read in the local context as supported by Michael (1993). In the next section, we will be concerned with the communicative intention of the types of structures that we have described in this section.

**Syntactic Expression of Mood**

In attempting to express themselves, the participants did not only exchange utterances containing the set of rules, they also performed structural moods. In many ways, they juxtaposed linguistic features for the success of communication. For instance, in [6a], imperative structure was used along with the insertion of exclamatory tone to enquire. Similarly in [6b and 6c], declarative structure was used to enquire. In [6c], use of a
question mark at the boundary indicates that the participant had intention to enquire. In [6d], the structure is interrogative but the exclamation mark and comma seem to indicate the intention. Thus, the structure alone may not be sufficient to perceive the communicative function in digital discourse. The subsequent discursive practices they carried out uncover a success of their communication. They juxtaposed various linguistic and extralinguistic features to choose functional roles and to resolve a kind of ambiguity that may emerge as a result of apparently a flexible relationship between structure and function, for example, a statement can be read as a question or vice versa (see, for example, Sinclair & Coulthard, 1978).

![Example of text](image)

**Conclusion**

This study pursued a descriptive approach to analyze the linguistic repertoire of Urdu/English bilinguals to explain how CMC works. The findings reveal that the participants are prone to minimizing forms, apparently presupposed, in English and Urdu. The data indicate that they tend to delete optional words and segments more or less, which do not hide the base forms for the projection of missing elements. While mapping out minimal linguistic forms, we found unique configurations which were equally rich in their semantic manifestation.

Despite the fact that the participants committed structural irregularities in their communication, they adhered to grammatical rules and conversation principles to the extent that their communication was intelligible. The elements which were missed or omitted can be guessed at or perceived because, while communicating, they were not only concerned with just the structure of their message but with the meaning and context surrounding the message, as supported by Spears, Lea, and Postmes (2001). Perhaps meanings, more than receiver or mode, directed their linguistic choices as to how far they could manipulate the structure of English and Urdu. They marked moods with the types of structures we have discussed. Although we generally construe a mood both through the transformation of a structure along with insertion of punctuation marks, they did this simply by inserting punctuation marks and avoiding structural transformation. Apart from structure, the discourse roles the participants technically assumed tend to contribute to the theory of meaning in digital environment.

Unlike previous studies, which have revealed the ambivalent, if not weird, nature of cyber-linguistic features, this study helps define a broader classification of them, which supposedly project some unique configurations. The study reveals that even though Urdu and English are subject to morphological and syntactic reduction, their structural manipulation in the context of CMC does not obscure meaning. The study motivates future researchers to investigate linguistic reduction to reveal linguistic variation and change in the languages used for the Internet communication. This would further provide us with a foreground, so we can understand more about the linguistic behavior and social identity of e-communities.

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) received no financial support for the research and/or authorship of this article.

**References**

Ahmad, R. (2011). Urdu in Devanagari: Shifting orthographic practices and Muslim identity in Delhi. *Language in Society, 40*, 259-284.

Androustopoulos, J. (2007). Language choice and code switching in German-based diasporic web forums. In B. Danet & S. C. Herring (Eds.), *The multilingual internet: Language, culture, and communication online* (pp. 340-361). Oxford, UK: Oxford University Press.

Axelsson, A. S., Abelin, A., & Schroeder, R. (2007). Anyone speak Swedish? Tolerance for language shifting in graphical multiser virtual environments. In B. Danet & S. C. Herring (Eds.), *The multilingual internet: Language, culture, and communication online* (pp. 362-384). Oxford, UK: Oxford University Press.

Barasa, S. (2010). *Language, mobile phones and Internet: A study of SMS texting, email, IM and SNS Chats in computer mediated communication (CMC) in Kenya*. Janskerkhof, The Netherlands: LOT Publisher.

Barnes, S. (2003). *Computer mediated communication: Human-to-human communication across the Internet*. Boston, MA: Pearson Education.

Baron, N. S. (2008). *Always on: Language in an online and mobile world*. New York, NY: Oxford University Press.

Bodomo, A. B. (2009). *Computer-mediated communication for linguistics and literacy: Technology and natural language education*. Hershey, PA: IGI Global.

Boyd, D., & Ellison, N. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*, 10-25.

Crystal, D. (2001). *Language and the internet* (1st ed.). Cambridge, UK: Cambridge University Press.

Crystal, D. (2006). *Language and the internet* (2nd ed.). Cambridge, UK: Cambridge 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). Oxford, UK: Oxford University Press.

Durham, M. (2007). Language choice on a Swiss mailing list. In B. Danet & S. C. Herring (Eds.), *The multilingual internet: Language, culture, and communication online* (pp. 319-339). Oxford, UK: Oxford University Press.
Firth, J. R. (1957). A synopsis of linguistic theory. In Stubbs, M. (1996). Text and corpus analysis (p. 35). Oxford, UK: Blackwell.

Grinner, R., & Eldridge, M. (2001, September 16-20). Y do ingr luv 2 txt msg? Proceedings of the European Conference on Computer-Supported Cooperative Work, Bonn, Germany.

Halliday, M. (2002). Linguistic studies of text and discourse. London, England: Continuum Publisher.

Halliday, M. (2003). The language of early childhood. London, England: Continuum Publisher.

Halliday, M. (2004). The language of science. London, England: Continuum Publisher.

Halliday, M. (2005). Studies in English language. London, England: Continuum Publisher.

Halliday, M. (2007). Language and society. London, England: Continuum Publisher.

Herring, S. C. (1996). Computer-mediated communication: Linguistic, social and cross-cultural perspectives. Amsterdam, The Netherlands: John Benjamins.

Herring, S. C. (2011). Grammar and electronic communication. In C. Chapelle (Ed.), Encyclopedia of applied linguistics. Hoboken, NJ: Wiley-Blackwell. Retrieved from http://ella.slis.indiana.edu/~herring/e-grammar.2011.pdf

Herring, S. C., Stein, D., & Virtanen, T. (2013). Introduction to the pragmatics of computer-mediated communication. In S. C. Herring, D. Stein, & T. Virtanen (Eds.), Handbook of pragmatics of computer-mediated communication (pp. 3-31). Berlin, Germany: Mouton de Gruyter.

Ilyas, S., & Khushi, Q. (2012). Facebook status update: A speech act analysis. Academic Research International, 3, 500-507.

Jørgensen, J. N. (2001). Multi-variety code-switching in conversation 903 of the Koge Project. In S. Leppänen (2007). Youth language in media contexts: Insights into the functions of English in Finland. World Englishes, 26, 149-169.

Jørgensen, J. N., Karrebæk, M. S., Madsen, L. M., & Møller, J. S. (2011). Polylanguaging in superdiversity. Diversities, 13, 23-37.

Kalman, Y. M., & Gergle, D. (2009, November 12-15). Letter and punctuation mark repeats as cues in computer-mediated communication. Paper presented at the National Communication Association’s 95th Annual Convention, Chicago, IL.

Kalman, Y. M., & Gergle, D. (2010, September 12-14). CMC cues enrich lean online communication: The case of letter and punctuation mark repetitions. Paper presented at the 5th Mediterranean Conference on Information Systems, Tel-Aviv, Israel.

Ko, K. K. (1996). Structural characteristics of computer mediated language: A comparative analysis of interchange discourse. The Electronic Journal of Communication, 6, 13-22.

Leppänen, S. (2007). Youth language in media contexts: Insights into the functions of English in Finland. World Englishes, 26, 149-169.

Mann, C., & Stewart, F. (2000). Internet communication and qualitative research. London, England: SAGE.

Michael, H. (1993). The metaphysics of virtual reality. New York, NY: Oxford University Press.

Napoli, D. J., & Lee-Schoenfeld, V. (2010). Language matters. Oxford, UK: Oxford University Press.

Paolillo, J. C. (2007). How much multilingualism? Language diversity on the Internet. In B. Danet & S. C. Herring (Eds.), The multilingual internet: Language, culture, and communication online (pp. 319-339). Oxford, UK: Oxford University Press.

Paolillo, J. C. (2011). “Conversational” code switching on usenet and internet relay chat. Language@Internet, 8. Available from http://www.languageatinternet.org/

Peuronen, S.-R. (2008). Bilingual practices in an online community: Code-switching and language mixing in community and identity construction at www.godspeed.fi (Master’s thesis in English). Retrieved from https://jyx.jyu.fi/dspace/bitstream/handle/123456789/18390/urn_nbn_fi_jyu-200805061436.pdf?sequence=1

Rafi, M. S. (2008). SMS text analysis: Language, gender, and current practices. Online Journal of TESOL France. Available from www.tesol-france.org

Rafi, M. S. (2010). The sociolinguistics of SMS: Ways to identify gender boundaries. In T. Rotimi (Ed.), Handbook of research on discourse behavior and digital communication: Language structures and social interaction (pp. 104-111). New York, NY: IGI Publishers.

Rafi, M. S. (2013). Urdu and English in an e-discourse variation in the theme of linguistic hegemony. European Academic Research, 6, 1260-1275.

Rafi, M. S. (2013). Urdu and English contact in an e-discourse: Changes and implications. Gomal University Journal of Research, 29 (2), 78-86.

Raskin, R. (2006). Facebook faces its future. Young Consumers: Insight and Ideas for Responsible Marketers, 7, 56-58.

Seargeant, P., Tagg, C., & Ngampramuan, W. (2012). Language choice and addressivity strategies in Thai-English social network interactions. Journal of Sociolinguistics, 16, 510-531.

Segerstad, H. Y. (2002). Use and adaptation of written language to the condition of computer mediated communication. Göteborg, Sweden: Department of Linguistics, Goteborg University.

Sengupta, S., & Rusli, M. E. (2012, February 1). Personal data’s value? Facebook is set to find out. The New York Times. Available from www.nytime.com

Sinclair, J. McH. (1991). Corpus, concordance, collocation. In M. Stubbs (1996). Text and Corpus Analysis. Oxford, UK: Blackwell.

Sinclair, J. McH., & Coulthard, M. R. (1978). Towards an analysis of discourse. Oxford, UK: Oxford University Press.

Spears, R., Lea, M., & Postmes, T. (2001). Social psychological theories of computer-mediated communication. In W. P. Robinson & H. Giles (Eds.), New handbook of language and social psychology (pp. 601-623). Chichester, UK: John Wiley.

Sperlich, W. B. (2005). Will cyberforums save endangered languages? A Niuean case study. International Journal of the Sociology of Language, 172, 51-77.

Stubbs, M. (1996). Text and corpus analysis. Oxford, UK: Blackwell Publishers.

Tagg, C. (2009). A corpus linguistics study of SMS text messaging (Doctoral dissertation). Retrieved from etheses.bham.ac.uk/253/1/Tagg09PhD.pdf

Thurlow, C., Lengel, L., & Tomic, A. (2004). Thurlow, C., Lengel, L., & Tomic, A. (2004). A corpus linguistics study of SMS text messaging (Doctoral dissertation). Retrieved from etheses.bham.ac.uk/253/1/Tagg09PhD.pdf

Thurlow, C., Lengel, L., & Tomic, A. (2004). A corpus linguistics study of SMS text messaging (Doctoral dissertation). Retrieved from etheses.bham.ac.uk/253/1/Tagg09PhD.pdf

Thurlow, C., & Poff, M. (2013). Text messaging. In S. C. Herring, D. Stein, & T. Virtanen (Eds.), Handbook of pragmatics of computer-mediated communication (pp. 163-190). Berlin, Germany: Mouton de Gruyter.
Trudgill, P. (2003). *A glossary of sociolinguistics*. Oxford, UK: Oxford University Press.

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. New York, NY: Basic Books.

Warschauer, M., El Said, G. R., & Zohry, A. (2007). Language choice online: Globalization and identity in Egypt. In B. Danet & S. C. Herring (Eds.), *The multilingual internet: Language, culture, and communication online* (pp. 303-318). Oxford, UK: Oxford University Press.

Weir, A. (2012). Left edge deletion in English and subject omission in diaries. *English Language & Linguistics, 16*, 105-129.

**Author Biography**

Muhammad Shaban Rafi is a doctoral candidate in the Department of English Language and Literature, University of Management and Technology, Lahore, Pakistan. He teaches to postgraduate students. His research interests are language variation and change, sociology of cyber-communication, and language teaching.