Text Messages: A Computer-Mediated Discourse Analysis

Waheed M. A. Altohami
Department of English
College of Science and Humanities in Al-Kharj
Prince Sattam Bin Abdulaziz University (PSAU), Al-Kharj, Saudi Arabia

Abstract—This study explores the discourse of text messages from a microlinguistic perspective by means of concordance analysis. It aims at sorting the dominant phonological, lexical and grammatical features that mark texting as a peculiar asynchronous mode of computer-mediated communication. Also, it investigates how technology reshapes texters' linguistic habits as long as spatio-temporal constraints are imposed. The study goes beyond the description of linguistic features as it takes at its core the explanation of the functions performed by each of these features. Findings showed that most of the phonological, lexical and grammatical features of the discourse of text messages were consciously employed to save space and to speed up communication. Furthermore, the study demonstrated that though the discourse of text messages is space-bound and visually decontextualized, it proved to be cohesive, adaptable and interactive in order to perform common language functions such as greetings, expressing attitudes, congratulations, showing involvement, asking for information and demonstrating social solidarity. Finally, based on textual evidence, findings showed that texters created a set of orthographical surrogates to recompense the absence of verbal and para-verbal cues due to specific technological affordances.

Keywords—Text messages; computer-mediated communication; discourse; technological affordances

I. INTRODUCTION

Communication refers to the exchange of thoughts, opinions or information through spoken, written or signed forms of language. Electronic gadgets have developed a distinct medium of communication, known as computer-mediated communication (CMC for short), with peculiar micro- and macrolinguistic features. CMC is generally viewed as a computer-based human-to-human communication in particular <u>contexts</u> shaping and reshaping <u>media</u> for various purposes [1]. That is, "the more people chat, text, email, blog … etc. on the Internet from different cultures, the more homogenised language may become" [2]. In addition to the spoken and written forms of language, such developing electronic medium is claimed to constitute a third modality [3]. In an electronic environment, an individual is surmised to be linguistically and communicatively competent if he manages to translate much of the written text and emotext (i.e. informal language codes used by communicators to mark up their messages and convey affective and socio-emotional information) into sounds. Simultaneously, he is expected to able to render his talk into peculiar orthographical forms. In so

doing, he is acknowledged as a member of an online discourse community.

Indeed, the way technologies affect language has motivated considerable speculation and inspired scholars in a plethora of research areas, including linguistics. It is reported that one influential feature of CMC environments is the evolution of a new discursive world with a new cultural context yielding peculiar discourse properties [4]. Although each of CMC modes, including text messages, blogs, wikis, and social media platforms, manages to communicate the intended message, they differ –both qualitatively and quantitatively –in respect of their dominant linguistic features. Hence, the present study seeks to provide a qualitative microlinguistic analysis of the discourse of text messages on the phonological, lexical and grammatical levels. Each distinct linguistic feature is paradigmatically assumed to communicate a specific discourse function. It is noteworthy that the present study focuses on English texting as English is undoubtedly the lingua franca of all modes of online communication [3, 5, 6]. Moreover, as claimed in [7], though there is a growing academic interest in texting, a few in-depth studies has been conducted to describe the language used. Existing work on the discourse of text messages is confined to superficial description of the amalgamation of the common features of spoken and written languages that finally helped with the emergence of a new medium of communication. One more limitation of available literature on the language of text messages is that most studies do not tackle the communicative functions that the gamut of subtle linguistic features of texting perform. Hence, the present study is planned to fill in this gap by offering a comprehensive description of the linguistic features of text messages with special reference to the communicative functions they generally perform in everyday interaction among adult texters.

Any analysis of online behavior that is grounded in empirical, textual observations is known as computer-mediated discourse analysis (henceforth CMDA) [8]. A good question in CMD research should be empirically answerable and textually motivated. Also, it should be ostensibly interesting to the research community, and therefore it is advised to be motivated by a hypothesis, and open-ended [8]. In view of these guidelines, the questions of the current study can be stated as follows: (1) what are the key phonological, lexical and grammatical features of the discourse of text messages? And (2) what are the functions performed by each of these features. Answers to these questions shall fulfill a three-fold objective. First, it demonstrates that despite differences in language

production conditions, synchronicity level and the medium used, the discourse of text messages displays certain linguistic features that would render it as a distinct mode of communication. Second, it explains that CMC differs from face-to-face (FTF) communication, not only because of lack of non-verbal cues, but also as a result of ever-changing technologies. Finally, it analyzes the functions fulfilled by the gamut of the linguistic features remarkable of the discourse of text messages.

Though the present study assumes that the discourse of text messages would display particular linguistic features due to the technological affordances imposed on the creation of a text message, it is also hypothesized that such features would be similar to those marking other CMC modes. Furthermore, it is assumed that the same discourse function would be fulfilled by different linguistic features on the phonological, lexical and grammatical levels.

The reminder of this paper is divided as follows. Section II is a survey of related research on the discourse of text messages. Section III offers the theoretical preliminaries upon which the current study is based. Section IV explains the methodology in terms of the procedures of data collection and data analysis. Section V is a qualitative analysis of the data. Section VI discusses findings and offers insights for further research.

II. Literature Review

Digital technology has remarkably caused drastic changes and raised doubts concerning the language of human-human and human-computer communications. However, descriptions of the technologies, functionalities and frameworks of CMC are likely to be outdated by the time they are published due the evolutionary nature of most CMC modes. The rapid growth of the number of text messages all over the globe generated a considerable number of linguistic issues. More specifically, the discourse features of such mode are ostensibly under-researched. Most of the studies that were concerned with the discourse of text messages adopted different microlinguistic and macrolinguistic approaches [7, 9-16]. That is, they analyzed the discourse of text messages from the phonological level up to the pragmatic level. However, as the following review would demonstrate, none of these studies was particularly concerned with the functions that each linguistic feature performs in the course of texting.

Thurlow and Brown [9] conducted a discursive analysis of 544 text messages written by teenagers. Serving the sociolinguistic Grecian maxims of (a) brevity and speed, (b) paralinguistic restitution and (c) phonological approximation, messages were both linguistically unmarked, but communicatively adept. Findings showed that the average length of text messages was about 14 words (with an average of 65 characters). The most frequent linguistic forms highlighted in the data included (a) shortenings, contractions, and clippings, (b) acronyms and initialisms, (c) letter/number homophones, (d) typos and non-conventional spellings, and (e) accent stylizations.

Bush [10] was much concerned with the way abbreviations in text messaging change written discourse. The study classified such abbreviations into nine separate semantic categories: homophonic single grapheme (e.g. 'n' for 'and'), numeric (e.g. 'a3' for 'anytime, anyplace, and anywhere'), aphesis/abbreviation (e.g. '@coll' for 'at college'), dropping vowel (e.g. 'fwd' for 'forward'), acronyms (e.g. 'gol' for 'giggling out loud'), alphanumeric (e.g. 'dem&' for 'demand'), grapheme changes (e.g. 'ezi' for 'easy'), abbreviated phrase (e.g. 'Ic**wenuxme' for 'I see stars when you kiss me') and contractions (e.g. 'wassup' for 'what's up').

Bieswanger [11] conducted a contrastive analysis of different shortening strategies in English and German, represented by 201 and 387 text messages respectively. The analysis was much concerned with lexical reductions that were listed into six classes: initialisms, clippings, contractions, letter-number homophones, phonetic spellings and word-value character. Findings showed that English generally use more shortenings, contractions and phonetic spellings than German.

Ling and Baron [12] compared text messages and instant messages produced by American college female students. The corpus consisted of 191 messages, comprising 1473 words. The analysis targeted message length, emoticons, lexical shortenings and sentential punctuation. Text messages averaged 7.7 words. The salient linguistic properties were emoticons, acronyms (e.g. 'ttyl' for 'talk to you later'), abbreviations (e.g. 'R' for 'are'), and vowel deletion (e.g. 'ovr' for 'over').

Rafi [13] tested the morpho-syntactic and lexical choices made by males and females Pakistani texters. Findings highlighted the use of abbreviations (e.g. 'y' for 'why') and shortening (e.g. 'n' for 'and'). At lexical level, texters used meaningful condensed forms (e.g. 'intro' for 'introduction'). Also, there was a significant gender-based difference in the number of words per a text message, the complexity of the messages and the use of punctuation. Similarly, Plester and Wood [14] focused on preteen British children's use of text messaging, particularly the use of abbreviations. Findings showed that the most frequent linguistic features were rebus spelling, letter/number homophones, phonological reductions, symbols and accent stylization.

By means of WordSmith software, Tagg [7] used a corpus of 11.000 text messages to explore their defining linguistic features that differentiate them from spoken interaction. Findings showed that text messages are marked by colloquial and regiolectal contractions, spelling variation, misspellings and respellings, speech-like ellipsis, excessive use of headers and tails, visual morphemes, and puns and word playing. Unlike Tagg [7], Thurlow and Poff [15] were much pragmatically oriented as they approached text messaging as a pragmatic phenomenon. They adapted the Gricean maxims of quantity, quality, manner, and clarity to the designation of three maxims peculiar to the language of texting: brevity and speed, paralinguistic restitution, and phonological approximation. The maxim of brevity and speed was manifested in the abbreviation of lexical items and the minimal use of capitalization and standard punctuation. Likewise, the maxims of paralinguistic restitution, and phonological approximation highlighted the absence of prosodic features such as stress and intonation.

McSweeney [16] investigated the functions performed by text messages exchanged by bilingual communicators. Based on a large corpus of text messages, the study highlighted how texters change the conventions of orthography into emoji and digital stickers. The most crucial functions performed by text messages were to express specific societal relationships, underlie politeness markers, express aspects of identity, and highlight implied information. In addition to these functions, the study offered an all-inclusive account of the linguistic features of text messages, including abbreviations, emojis and emoticons, differentiating between their literal meaning and pragmatic meaning. For instance, while 'lol' stands for 'laugh out loud', it is meant to belittle, mock, or ridicule an addressee.

To recap then, it is obvious that most of the studies that approached the discourse of text messages focused more on the lexical features rather than other microlinguistic features. Also, the functions of these features were not clearly stated. Hence, the present study sets out to fill in this gap by offering a full account of the phonological, lexical and grammatical features of the discourse of text messages and highlighting their functions that are unquestionably situationally motivated.

III. THEORETICAL PRELIMINARIES

A. Computer-Mediated Communication

The term 'computer-mediated communication' (CMC) emerged in the 1980s as an umbrella term covering a range of modes used for communicating online, including e-mail, chat systems, video conferencing, instant messages, blogs, wikis, etc. Unlike FTF communication, CMC lacks common non-verbal cues such as facial expressions, eye contact and gestures. Also, the auditory channel is deactivated and feedback is either absent or lagging. It has been likened to speech, and to writing, and considered to be both and neither simultaneously.

CMC modes are claimed to support scholarly activities and research, personal and group communication, discussion, play and learning, dissemination and retrieval of information [17]. Also, they engage users in purposeful exchanges of information, opinions and interests with other humans, and offer second language learners a chance to enhance their learning experience [4]. Moreover, CMC modes productively augment human identity and community [18], and enable health professionals and care teams to document and disseminate information instantly and effectively [19]. Finally, they are claimed to provide learning environments that support constructivist approaches to learning [20], and to represent an effective instructional tool for pragmatic instruction [21].

The most common classification of CMC modes is based on synchronicity. On the one hand, synchronous CMC is like spoken interaction [21] as it is produced when communication occurs between two or more users who join an ongoing conversation in real time, e.g. chat systems. Newhagen and Rafaeli [22] distinguishes six special features for synchronous CMC: multimodality, hypertextuality, packet switching, synchronicity and interactivity. On the other hand, asynchronous CMC is not simultaneous as there is a time gap between message transmission and reception so that users can read and re-read them to give a more thoughtful answer as in the case of text messaging that is mostly based on one-to-one,

text-based, two-way transmission of messages. Synchronous and asynchronous CMC modes cannot be sharply distinguished, and therefore should be distributed on a continuum ranging from the highly synchronous to the highly asynchronous [6].

The following section discusses how the properties and technological affordances characteristic of CMC modes are realized in the mode of text messaging in addition to the restrictions affecting language production via such mode.

B. Text Messages

Text messages (also known 'SMS' and 'texting') are one of the most popular and pervasive applications of mobile phones. Text messaging has seen phenomenal growth since the late 1990s. What makes it ubiquitous is that it is cheap, personal and unobtrusive. However, one should mark the interplay between what technology affords and what communicators bring to technology itself [15]. Commenting on the problems that a person might encounter when text messaging, it is claimed that such problems are related to (a) understanding evolving language as text messages have their specialized language, (b) determining intent from content, and (c) mis-addressing messages [23].

Text messages perform various functions. They foster all kinds of social relationships, e.g. sympathizing, requesting for a call, reflecting the time of the day, sharing interests, exchanging personal news and gossip, greeting, offering help and advice, grooming, and circulating jokes, riddles and remarks [24]. Also, they are used to arrange and adjust times to talk, coordinate with friends and family, gossip and chat [23], demonstrate one's involvement in a social network, and vote in electoral campaigns [15, 25].

The language of text messages is referred to as 'textese' and 'texting'. Textese or texting is approached as a multidisciplinary field of CMC that explores mobile and computer-mediated media and their impact on communication, social practices, and information dissemination and exchange [7]. Later, the term 'celllinguistics' evolved to refer to this area [24]. Such language used in texting is found to include extensive use of abbreviations and acronyms, elision of vowels, use of emoticons, and incomplete syntactic structures, minimal use of punctuation marks, respellings and untraditional openings and closures. Much of these linguistic features are ascribed to the technical constraints of texting as well as the existence of common background knowledge among texters.

C. Computer-Mediated Discourse Analysis

Discourse simply refers to any stretch of spoken or written language with particular features performing a set of functions that shape social practices. It covers both linguistic and non-linguistic meaning-making. Electronic gadgets are claimed to alter discourse conventions, adding a new electronic element to language [26]. Studies on the reproduced language of CMC use the term 'computer-mediated discourse' (henceforth CMD) to refer to language used in all CMC modes. CMD is empirically viewed as an interdisciplinary approach in the area of CMC that takes at its core the analysis of the language used in electronic environments by means of discourse analysis methods [27]. It received different labels such as 'netspeak' [3] and 'SMS speak' [28].

Such discourse is viewed as a fusion of speech and writing constituting a language variety or a register of its own [29, 30]. Unlike spoken language, CMD lacks simultaneous feedback, and non-segmental phonology is absent. Also, multiple interactions can be carried on simultaneously. Unlike written language, CMD is characterized by its dynamicity and hypertextuality [31]. Still, CMD is perceived as a new medium rather that a new variety of English [3]. As argued in [29], among the situational features that explain the linguistic manifestations of CMD are (a) the degree of participants' common background knowledge and interests, (b) the purpose of communication, (c) the roles played by an addressor, and addressee and an audience in a communicative situation, and (d) the speaker-text relationship.

The introduction of internet-based technologies caused the evolution of new linguistic practices that are the focus of the so-called "computer-mediated discourse analysis" (CMDA for short). Any analysis of online behavior that is grounded in empirical, textual observations is counted as computer-mediated discourse analysis [8]. CMDA can be employed to investigate micro-level linguistic phenomena such word formation processes, lexical choice, sentence structure, and code switching. It also studies macro-level linguistic phenomena such as coherence, cohesion, gender equity, ideology and identity as expressed through discourse [8]. CMDA assumes that (a) discourse displays concurrent language patterns produced either consciously or unconsciously, (b) discourse involves speaker's choices that help to gain insights into linguistic as well as non-linguistic phenomena, (c) CMD is subject to the technological affordances imposed by different CMC modes [8]. To recap, CMDA is an approach that relies on diverse theories and methods to explore the discourse properties displayed in different CMC modes.

IV. METHODOLOGY

A. Data

One common problem with building a corpus of text messages is that texters are always reluctant to share their messages for academic research. Therefore, it is highly recommended to collect data from different sources [32]. Therefore, data in the present study are collected from two main sources. The first source is part of Tagg's corpus of text messages [7]. I got a permission from her to use the corpus for academic research purposes, publishing substantial amounts of the corpus. The second source is different websites publishing samples of text messages. The final corpus consisted of around 15.000 words standing for 12.000 text messages created mainly by British speakers. The largest part of data has been collected between 2004 and 2007 from adult volunteers who shared their personal – rather than business – text messages. All personal information such as names, addresses, etc. are removed for the sake of texters' privacy, and they are replaced with the capital form [PRIVATE].

B. Procedure of Analysis

As mentioned earlier, the research method adopted in the present study is basically qualitative. Since the language of text messages is principally intended to mimic spoken language, the present study adapts Carter and McCarthy's model of spoken grammar [33] and Biber's Multi-Dimensional, Multi-Feature Model [34] which are markedly based on data-driven approaches. That is, the present study seeks to identify the linguistic features marking the discourse of text messages on the phonological, lexical and grammatical levels as well as to explore the communicative functions performed via these features. To achieve this objective, a three-fold procedure of analysis is followed. First, the corpus is closely investigated and the dominant linguistics features are identified and sorted through investigating the concordance lines generated by AntConc corpus analysis toolkit (3.5.8) [35]. The list of key words displayed by AntConc act as guidelines toward the investigation of the peculiar discourse shaped and reshaped by texters. These concordance lines help with identifying the frequency of the different linguistic features as they are shown across the body of the corpus (cf. Fig. 1). Second, representative, contextually-relevant examples are given and described. The representativeness of the illustrated examples is based on the rough frequency of each linguistic feature. Finally, the communicative function performed by each linguistic feature is explained in relation to other features. For this reason three key, mostly-cited references in grammar and lexis [33, 36, 37] are consulted and co-referenced.

Fig. 1.   Sample Concordance Lines for 'r'.

V. DATA ANALYSIS

This section analyzes the data of text messages on the phonological, lexical and grammatical levels. These three levels would give an integrative outline of the linguistic features marking the discourse of text messages from a microlinguistic perspective.

A. Phonological Features

Texters tend to find correspondences between phonemes and graphemes in order to show how language sounds. Though texting is totally performed in writing, the following features are identified.

1) Phonetic spelling: In Standard English, phonemes do not always correspond to graphemes. However, data shows many one-to-one correspondences between phonemes and graphemes. Such phenomenon is referred to as 'phonetic spelling'. Phonetic spelling fits the informal nature of CMC modes, reduces the time used for standard spelling, and conveys other auditory information. Letters and numerical digits substitute sounds within a word based on homophony, e.g. '2morrow' (tomorrow) and 'm8' (mate). Furthermore, a single letter or number can be used to replace an entire word, e.g. 'b' (be), 'r' (are) and 'u' (you). Similarly, certain letters and sounds are replaced by others. For instance, <v> replaces <th> as in '2gever' (together); <d> replaces <th> as in 'dis' (this); <i> replaces <igh> as in 'rite' (right); <f> replaces <th> as in

'fink' (think); <f> replaces <ph> as in 'fone' (phone); and <f> replaces <gh> as in 'enough'. While phonetic spellings are used creatively to speed up the course of interaction and save space, sound replacements are meant to reflect some local informal pronunciations (e.g. Th-fronting) as a way of showing identity.

2) Consonantal writing: Consonantal writing refers to the omission of vowels from words, thereby forming consonant clusters or telegraphic messages. Texters tend to delete vowels in order to save space, i.e. the 160-charater limit in each message, e.g. 'thx' (thanks), 'ppl' (people), 'tmrw' (tomorrow), 'bbq' (barbeque), etc.

3) Stress and intonation: Due to lack of verbal cues in text messages, texters tend to mark the stress and intonation of certain chunks of information through two main devices. The first strategy is the use of parenthetical brackets, e.g. Have fun and keep me texted (updated jargon), that's fine (but let me know if it changes and u can't). The second strategy is the use of capital (reduplicated) letters. Capital letters are used to draw attention to the importance of the content of the message, e.g. HOPE URE OK? TAKE CARE & I'LL SPEAK 2U SOON LOTS OF LOVE; PPL FROM WRK WILL B THERE. LOVE PETE.

B. Lexical Features

The lexical features of the discourse of text messages are explored with reference to lexical density, lexical classes, shorthand techniques, and emoticons.

1) Lexical density: The lexical density of a text can be identified through type/token ratio, mean word length and mean sentence length. It is maintained that type/token ratio is a measure of the lexical richness of a text [38]. The type/token ratio of the corpus is 21.38% which is quite high. This suggests careful word choice and more precise presentation of informational content [34, 39]. Furthermore, it is stated that the average word length in English is 5 characters [37]. Based on Microsoft Word readability statistics, the mean word length in the corpus is 3.7 characters, i.e. it is relatively short. This is justified by the use of many abbreviations, acronyms and shorthand techniques. With regard to the mean sentence length, is calculated by dividing the total number of words by the total number of sentences. The corpus shows that the mean sentence length is 8.41 characters. This is justified by the use of simple syntactic structures and modifiers. Also, the word limit for each text message force texters to use very short sentences.

2) Lexical classes: The most common lexical classes in the corpus are conjuncts, downtoners, hedges, amplifiers, emphatics and discourse markers. First, conjuncts are adverbs that add information to a sentence. They mark logical relations between clauses, relating sentence to other parts of discourse, e.g. 'never' (also spelt neva), 'rather', 'however', 'otherwise' and 'instead'. Second, downtoners (e.g. 'nearly', 'slightly', 'pretty', 'almost', etc.) are degree adverbs that indicate a degree of probability or uncertainty regarding the provided information

in order to scale down the effect of the modified item [36, 37]. Also, downtoners facilitate cooperation among the communicators while forming a face-saving act on the part of the hearer [40], e.g. 'we are in one i like to share only if you knew what i meant nice', 'R u coming pub We would look slightly right of centre to u'). Third, hedges, a subclass of adverbs, are mitigating devices which mark propositions as probable or uncertain. They distinguish facts from personal opinions. Also, they are not used to make attitudinal stance. The most frequent hedges in the corpus are about, I think, sort of, and approximately as in 'well i didt come ere for ages n i wuz chief about 2 weeks ago', 'i jus nearly had a heart attack racist rap', and 'Ill call u 2mrw at ninish'.

Fourth, amplifiers are degree adverbs which boost the force of verbs, increasing their intensity. They denote emphasis, certainty and reliability of propositions. They can communicate both interpersonal and ideational information. The most common amplifiers in the corpus are 'so', 'too' (also spelt '2'), 'really', 'more', and 'very' as in 'Got to go hospital soon, really nervous, wish me luck', 'I luv u soo much u don't understand how special u r 2 me ring u 2morrow luv u'. Fifth, emphatics mark the presence of certainty and convey emotional and personal view towards the underlined proposition. Emphatics mark involvement with the topic [34]. The most frequent emphatics are 'just', 'really', 'surely', 'exactly' and 'indeed' as in 'no idea really, pickle toes and Oops just got your text'. Finally, discourse markers are syntactically-independent lexemes attached to the clause. They are claimed to facilitate ongoing interaction, mark topic shifts, indicate the speaker's attitude, make discourse more coherent, and structure the relationship among participants in a conversation [33, 37]. The most frequent discourse markers in the corpus are 'okay', 'then', 'well', 'now' and 'yeah' as in 'Ok I can go Friday', 'Mm i'm in london now and both of me are having our first drink', 'Yeah ok, i'm just watching desperate housewives'.

3) Shorthand techniques: Shorthand techniques refer to the ways in which words, phrases and sentences are shortened. The corpus shows four basic shorthand techniques: abbreviations, clips, compounds and contractions. All of these techniques are intended to speed up interaction and save space. First, abbreviations are found to stand for individual words as a letter might be homophonous to a whole word (e.g. 'c' for 'see' and 'y' for 'why'). Also, texters tend to omit vowels as in 'gs' for 'guess' and 'lrg' for 'large' and syllables as in 'fav' for 'favourite' and 'prob' for 'probably'. Similarly, apostrophe is mostly omitted as in 'cant', 'aint' and 'hes', and double letters are reduced as in 'gota' for 'gotta' and 'til' for 'till'. Some abbreviations are meant to stand for compound words as in 'NY' for 'New York' and 'neway' for 'anyway', phrases as in 'aka' for 'also known as' and 'rsvp' for 'repondez s'il vous plait', and whole sentences 'ta' for 'thanks a lot' and 'ams' for 'ask me something'.

Second, clipping refers to dropping letters from any part of a word. The most interesting examples in the corpus are 'cos' for 'because', 'THO' for 'though' and 'v' for 'have'. These clips imitate natural speech, thereby creating real environment for

interaction, especially those that imitate non-rhotic /r/ as in 'neva' and 'betta' and G-clipping as in 'sumthin' and 'darlin'. Third, compounds in the corpus are not traditional as texters tend to remove spaces among words and contract them to save more space. The result is structures such as 'whaddya' for 'what do you do', 'chockablock' for 'chock-a-block', and 'WATERSHD' for 'water shed'. Still, they are not frequent as texters usually type at slower rate. Finally, contractions refer to dropping letters, usually vowels, from the middle of words. They may be standard or colloquial. Standard contractions such as 'doesn't' and 'he's' are not common. Colloquial contractions communicate emotions of affection and therefore they establish tone [41]. Based on Weber [42], contractions in the corpus include (a) progressive verbs as in 'clickin' for 'clicking' and 'gunna' for 'going to', (b) auxiliary verbs as in 's' for 'does' and 'r' for 'are', (c) modals as in 'wud' for 'would', (d) catenative verbs as in 'gotta' for 'got to', and (e) personal pronouns as in 'em' for 'them' and 'u' for 'you'.

4) Emoticons: Emoticons (or emotional icons) are textual expressions or graphical images formed from ordinary typographical symbols or ASCII characters. They are used to represent facial expressions, tone, intonation, voice inflection or feelings in CMC. They are surrogates of verbal, non-verbal, paraverbal cues in CMD. Metz [43] mentions four forms of emoticons: (a) verbalizing physical cues (e.g. 'hehe' for laughing), (b) using asterisks to describe physical actions (e.g. '*Lady Marion arrives in the room with a flair and arrogance fitting of her status*'), (c) stressing an utterance using capital letters (e.g. 'HEY EVERYBODY'), and (d) smileys what are characters used to communicate a feeling, and they may be graphic, symbolic (e.g. **:-)** for 'smile'), or written between two angle brackets (e.g. <s>).

In the corpus, emoticons are rarely used to verbalize physical action such as laughing as in 'hehehe'. However, asterisks are never used. Furthermore, most of the emoticons used in the corpus are symbolic. The most frequent smileys are :) 'happy', ;-) 'wink', :('sad', B-) 'cool' and =/ 'confused'. Capital letters are used to attract the attention of the addressee and they are generally used to mark physical actions especially laughter (e.g. 'LOL' and 'HAHA'), astonishment (e.g. 'WHO?'), excitement (e.g. 'EXCITING') and sadness (e.g. 'DAMMMN').

C. Grammatical Features

This section focuses on the most crucial grammatical features marking the corpus under investigation as well as their communicative functions based on the context in which they are used.

1) Tense and aspect markers: A close reading into the corpus shows that texters mostly focus on present actions and therefore most of the verbs are used in the present tense to report events (e.g. 'he writes the mail now'), plan future events (e.g. 'the group is meeting nxt week'), and comment on actions in progress (e.g. 'the game is gettin tough'). The past participle form is rarely used to state the completion of an even (e.g. 'he has bought the dragon'). The various time and aspect markers in the corpus show lexical diversity and mark informational discourse. Furthermore, most of the structures are used with

first and second pronouns which indicate involvement in the discourse.

2) Adverbials: Adverbials carry much of the informational load of an utterance, indicating the place, time, manner and frequency of an action. They are markers of involvement in discourse [39]. Place adverbials indicate distance, direction, or position as in 'what's new around here?' and 'ur m8s agen after u bein racist towards dem'. Based on the common classification of the functions of time adverbials [37], the time adverbials in the corpus are used to tell time (e.g. 'now', 'early', 'later', etc.), express duration (e.g. 'for'), convey frequency (e.g. 'again', 'once', etc.), and mark temporal relationships (e.g. 'before', 'after', 'while', etc.). With regard to manner adverbials, they are used to reflect on how different actions are performed. The corpus shows that they are basically used to modify verbs rather than adjectives and other adverbs (e.g. 'I was perfectly nice to you', 'things which are literally and physically not possible'). Finally, frequency adverbials are used to describe how often an action happens. The corpus show that they are often used after verb to be (e.g. 'im always bumpin in 2 ppl i kno'), before verbs (e.g. 'they only have 1 grandchild down here') and before nouns and adjectives (e.g. I surrender to every word u whisper).

3) Nominal forms: Nominal forms refer to the way nouns are formed, and they are related to information packing in an utterance. The corpus shows that nouns in the discourse of text messages are formed through nominalization, gerunds and nouns. Most of nominal forms are formed with suffixes, e.g. -er, -tion, -ence, and -ness. Nominalizations tend to pack much information in few words while expanding textual units [39]. Similarly, nominalization renders discourse more objective, summarizes propositional content, and contributes to the informational load of a text [44]. Like nominalizations, gerunds and present participle forms integrate information as in 'Hey j! r u feeling any better, hope So hunny', 'did you have any luck catching PRIVATE?' and 'PRIVATE has just rung me wanting your number'. Indeed, nouns are the most frequent grammatical category in the corpus as they function as referential specifiers referring to places, times, tools, abstract ideas, professions and people.

4) Adjectives: Adjectives expand information and limit its scope. The corpus shows three categories of adjectives: inflectional (e.g. 'All the best this afternoon'), periphrastic (e.g. 'Try that hotel. It is more convenient') and basic (attributive and predicative). Basic adjectives are the most frequent as attributive adjectives are descriptive (e.g. 'I have just come over all hot, sweaty, shaky and feel faint!') while predicative adjectives are emotive (e.g. 'anything that is warm').

5) Pronouns: Pronouns are used when the entities referred to are situationally prominent and when the reference is general [37]. Though they economize speech, they offer accurate specification. The corpus shows the use of personal, demonstrative, and indefinite pronouns. Personal pronouns mark a great deal of involvement among chatters. Accordingly, the absences of personal pronouns indicate

impersonality. They are classified into first person, second person and third person pronouns. First and second person pronouns are used excessively to show involvement in discourse, while third person pronouns show involvement with the topic rather than the addressee. It is noteworthy that the pronoun 'it' is rarely used since the discourse of text messages is not continuous. One more device that supports cohesion in the discourse of text messages is the use of demonstrative pronouns (e.g. 'this', 'that', 'these' and 'those') with some dialectical variations (e.g. 'dat' and 'dis'). Most of the demonstrative pronouns are used with nominal referent, thereby participating in clarifying their referents as in 'PRIVATE, is this your phone still?'. Finally, indefinite pronouns (e.g. 'all', 'much', 'everthin', 'every1', 'nething', etc.) are sometimes used in case a texter does not want to make himself/herself explicit as in 'hope all is well', and 'Another successful night, was a big hit'.

6) Questions: Questions generally shows interest and involvement in discourse. One key function of text messages is to ask for information on a variety of topics. Wh-questions mark a high degree of interpersonal interaction and personal involvement [34]. Texters in the corpus under investigation use different forms of questions as in 'Y did you volunteer?' and 'WHO IS THIS SI THEN?' Unlike Wh-questions, yes-no questions request a truth value based on elements already mentioned in the questions. Yes-no questions open with the operator (an auxiliary or a modal) followed by the subject. By answering 'yes' or 'no', the addressee supplies a truth value by opting for one of the already specified elements. However, the addressee may also supply additional information. The following statistics represent the frequency of yes-no questions in both corpora. In the discourse of text messages, they speed up the course of interaction as they offer quick short answers as in 'do you still have a spare big chill?' and 'can you please give PRIVATE and family my condolences tomorrow?'. The last form of questions is alternative question (also known as 'OR questions'). Like yes-no questions, they provide quick short answers by opting for one alternative as in 'shall we meet at 9 or 10 2mrw?'

7) Passives: Passive structures foregrounds the content rather than the agent. However, the agent may be deleted (i.e., agentless passives) or confirmed (i.e., by-passives). Agentless passive syntactic structures render the information offered in a text message as static and abstract. The corpus displays two basic passive structures: be-passive (be + past participle) and get-passives (get + past participle). Be-passive structures simply describe a state as in 'these tools were invented by PRIVATE' and 'their visas have been denied and they cant reenter'). Whereas get-passive describe the process of getting into a new state as in 'PRIVATE get drunk n make friends'.

8) Modals and semi-modals: Modals are indicative of language users' attitudes toward topics. They are generally classified into pure modals and semi-modals. Pure modals keep the same form and aspect, and they are negated with 'not'. Also, they are followed by the base form of the verb.

Based on the semantic notion impeded within pure modals, the corpus displays five categories of modals: (a) obligation and necessity modals (e.g. 'must', 'shouldn't', 'shud', etc.) as in 'I should get my card tonight', (b) ability and possibility modals (e.g. 'can', 'cud', 'couldn't', etc.) as in 'IF WE MEET WE CAN GO 2 MY HOUSE', (c) epistemic possibility modals (e.g. 'mite', 'may', etc.) as in 'PRIVATE may be l8 if he goes at 7', (d) volition and prediction modals (e.g. 'will' and 'shall') as in 'U WILL SWITCH YOUR FONE ON DAMMIT', and (e) hypothetical modals (e.g. 'wud', 'would' and 'wouldn't') as in 'we would drop in 2nt'. In general, these modals are used to offer more details about the meaning of the main verb and to communicate certain attitudes and social relations. Unlike modals, semi-modals are multi-word constructions that function like modals, e.g. 'ought to', 'be going to', 'need to' and 'be able to'. Their form and aspect change. They are basically intended to show informal style, particularly when contracted forms such as 'gonna' are used as in 'PRIVATE is going to ask big PRIVATE if i can go on 3 month trial'.

VI. RESULTS

Based on the results of the studies previously reviewed herein, it becomes clear that computer-mediated discourse has emerged as a result of adaptation to the technological constraints posed by different technological affordances. Such technological affordances arise from the material characteristics framed through the properties of the electronic environment offered by various CMC modes as well as users' cultural and individual knowledge that is employed to perform interpersonal functions. Furthermore, these affordances are spatio-temporal in the first place. As an asynchronous CMC mode, text messages are produced at a slow rate and hence they carry much of the properties of planned speech. Space also is another crucial affordance as texters have to abide by 160 characters only per message, thereby rendering the discourse of text messages markedly economic. That is, reductions in general save time and space and speed up the communication process which pertaining to the common properties of both spoken and written languages. Though it is interactive and revisable, the discourse of text messages is not time-bound. Rather, it is space-bound and visually decontextualized. In short, all of these factors and constraints underlie the peculiar linguistic properties of text messages discourse on the phonological, lexical and grammatical levels. The corpus tool, represented by AntConc corpus analysis toolkit (3.5.8) [35], helped to compare and sort such features.

On the phonological level, although the discourse of text messages is basically communicated in a written form, it is found it contains various phonological surrogates that compensate segmental and suprasegmental phonological features. The most common phonological features are phonetic substitution, sound replacement, capital letters and letter-number homophones. Texters substitute letters in irregular standard spelling for those corresponding to particular sounds and even syllables, e.g. '2' for 'two', 'to', 'too' and and '2gether'. Consonants are sometimes replaced by others, especially 'th' that is replaced by 'f' as in 'fink', or 'd' as in 'da' or 'v' as in 'wiv'. Also, the deletion of many vowels in text messages render

most of them as telegraphic. Capital letters are used to mark intonation and stress shift as in 'I will do it 2DAY'. Finally, letter-number homophones are used to save space and speed up communication.

Obviously, the corpus is lexically rich. Still, the type/token ratio of the corpus proved to be not high. Similarly, the average word length is below the average word length in Standard English. These features are typical of informal writing and interactive discourse. The most frequent lexical classes in the corpus are conjuncts, downtoners, hedges, and discourse markers. All of these classes are used to communicate various attitudes such as (un)certainty, seriousness, authenticity and involvement. For spatio-temporal reasons, shorthand techniques abound in the discourse of text messages. Abbreviations are created by apostrophe omission, clipping, double letter reduction, letter(s) omission, and syllable omission. They are found to stand for individual words, compound words, phrases and whole sentences. Equally important, emoticons are employed to offer non-textual information and to regulate interaction.

Similarly, texters tend to use peculiar grammatical features that best communicate their attitudes, stances and intentions. Almost all tense and aspect markers, including the base form, the present and past participles as well as the infinitive, present and past forms, are used in the corpus. This shows highly packed information despite the short texture of text messages. Such information cover the time, place, manner, and frequency of events, and they are highlighted via frequent adverbials. Also, different nominal and adjectival forms are used to expand and elaborate on the propositional content of messages. In addition, pronouns are used to mark personal relationships, show involvement in the discourse, and render the discourse of text messages referentially cohesive. Similarly, questions, particularly wh-questions, are employed to maintain interaction and trigger additional information. Finally, modals and semi-modals are used in the first place to encode texters' attitudes and to attach an evaluative tone to the message.

VII. Conclusion

The present study went beyond the superficial description of the linguistic features of the discourse of text messages. It took as it core the communicative functions performed by each linguistic feature taking into consideration the linguistic context of the whole utterance. The study hopefully contributes to our knowledge about the way technological devices motivate specific language behavior. The findings would help with enhancing mobile learning strategies and applications by offering content peppered with the linguistic features that texters usually use in their casual communications. Further future research is recommended to apply both quantitative and qualitative methods to explore the frequency of different linguistic features particular to the discourse of text messages and to identify the pragmatic as well as the cognitive basis for paradigmatic variations in such discourse. Also, further research would investigate the role played by gender in the formulation of the discourse of text messages.

References

[1] J. December. (1997) Notes on defining of computer-mediated communication. CMC Magazine. Available: https://www.december.com/cmc/mag/1997/jan/december.html.

[2] T. Neuage, "Conversational analysis of chatroom talk," Doctor of Philosophy, University of South Australia, Australia, 2003.

[3] D. Crystal, Language and the Internet. Cambridge: Cambridge University Press, 2001.

[4] M. R. Salaberry, "A theoretical foundation for the development of pedagogical tasks in computer mediated communication," CALICO Journal, vol. 14, no. 1, pp. 5-34, 1996.

[5] S. Hjarvard, "The globalization of language: How the media contribute to the spread of English and the emergence of medialects," Nordic Review, vol. 1, no. 2, pp. 75-97, 2004.

[6] N. S. Baron, Always on - Language in an online and mobile world. New York: Oxford University Press, 2008.

[7] C. Tagg, "A corpus linguistic study of SMS text messaging," Doctor of Philosophy The University of Birmingham, The United Kingdom, 2009.

[8] S. C. Herring, "Computer-mediated discourse analysis: An approach to researching online behavior," in Designing for virtual communities in the service of learning, S. A. Barab, Kling, R., Gray, J. H. , Ed. New York: Cambridge University Press, 2004, pp. 338-376.

[9] C. Thurlow, Brown, A. , "Generation txt? The sociolinguistics of young people's text-messaging," Discourse Analysis Online, vol. 1, no. 1, pp. 1-27, 2003.

[10] C. Bush, "Language beyond the text: txt msgs 4 a new gnr8n," Journal of Media and Culture, vol. 3, no. 2, 2005.

[11] M. Bieswanger, "2 abbrevi8 or not 2 abbrevi8: A contrastive analysis of different shortening strategies in English and German text messages," Texas linguistics forum, vol. 50, Available: http://nl.ijs.si/janes/wp-content/uploads/2014/09/bieswanger07.pdf

[12] R. Ling, Baron, N. S. , "Text messaging and IM: Linguistic comparison of American college data," Journal of Language and Social Psychology, vol. 26, no. 3, pp. 291-298, 2007.

[13] M. S. Rafi, "SMS text analysis: Language, gender and current practices," presented at the The 26th Annual TESOL France Colloquium, France, 2008. Available: https://www.tesol-france.org/uploaded_files/files/Coll07-Shaban-Rafi.pdf.

[14] B. Plester, Wood, C. , "Exploring relationships between traditional and new media literacies: British preteen texters at school," Journal of Computer-Mediated Communication, vol. 14, no. 14, pp. 1108–1129, 2009.

[15] C. Thurlow, Poff, M., "Text messaging," in Handbook of the pragmatics of CMC, S. C. Herring, Stein, D., Virtanen, T. , Ed. Berlin and New York: Mouton de Gruyter, 2011.

[16] M. A. McSweeney, The pragmatics of text messaging: Making meaning in messages. New York: Routledge, 2018.

[17] J. December, "Units of analysis for Internet communication," Journal of Computer-mediated Communication, vol. 1, no. 4, 1996.

[18] S. Turkle, "Life on the screen: Identity in the age of the internet," London: Weidenfeld and Nicolson, 1996.

[19] K. Vroman, Kovacich, J. , "Computer-mediated interdisciplinary teams: Theory and reality," Journal of Interprofessional Care, vol. 16, no. 2, pp. 159-170, 2002.

[20] A. Romiszowski, Mason, R. , "Computer-mediated communication," in Handbook of research on educational communications and technology, D. H. Jonassen, Ed. Mahwah, NJ: Lawrence Erlbaum Associates, 2004, pp. 397-431.

[21] J. M. Sykes, "Synchronous CMC and pragmatic development: Effects of oral and written chat," CALICO Journal, vol. 22, no. 3, pp. 399-431, 2005.

[22] R. Newhagen J. E., S. , "Why communication researchers should study the internet: A dialogue," Journal of Communication, vol. 46, no. 1, pp. 4-13, 1996.

[23] R. E. Grinter, Eldridge, M. A. , "y do tngrs luv 2 txt msg?," in Proceedings of the seventh European Conference on Computer-Supported Cooperative Work (ECSCW) Bonn, Germany, 2001, pp. 219-238: Dordrecht, Netherlands: Kluwer Academic Publishers.

[24] D. Crystal, Txtng: The gr8 db8. New York: Oxford University Press, 2008.

[25] G. Goggin, "Mobile phone culture and the love of text messaging," presented at the The Annual Meeting of the Australian and New Zealand Communication Association, Christchurch, New Zealand, 2005.

[26] M. N. Lamy, & Hampel, R. , Online communication in language learning and teaching. New York: Palgrave Macmillan, 2007.

[27] S. C. Herring, "Computer-mediated discourse," in Handbook of discourse analysis, D. Tannen, Schiffrin, D., Hamilton, H. , Ed. Oxford: Blackwell Publishers, 2001, pp. 612-634.

[28] K. Freudenberg, "Investigating the impact of SMS speak on the written work of English first language and English second language high school learners," Master's Degree, Stellenbosch University, South Africa, 2009.

[29] M. Collot, Belmore, N. , "Electronic language. A new variety of English," in Computer-mediated communication. Linguistic, social and cross-cultural perspectives, S. C. Herring, Ed. Amsterdam: John Benjamins 1996, pp. 13-28.

[30] S. Nilsson, "A brief overview of the linguistic attributes of the blogosphere," presented at the BlogTalk - A European conference on weblogs: Web-based publishing, communication and collaboration tools for professional and private use, Vienna, Austria, 2003.

[31] D. Crystal, "The scope of Internet linguistics," presented at the The American Association for the Advancement of Science Meeting (AAAS), Washington, DC., 2005. Available: http://www.davidcrystal.com/?fileid=-4113.

[32] Y. How, Kan, M. , "Optimizing predictive text entry for short message service on mobile phones," in Proceedings of the 11th Conference on Human Computer Interfaces International (HCII 05), Las Vegas, 2005.

[33] R. Carter, McCarthy, M. , Cambridge grammar of English: A comprehensive guide. Cambridge: Cambridge University Press, 2006.

[34] D. Biber, Variation across speech and writing. Cambridge: Cambridge University Press, 1988.

[35] L. Anthony, "AntConc corpus analysis toolkit (3.5.8)," (3.5.8) ed, 2017.

[36] R. Quirk, Greenbaum, S., Leech, G., Svartvik, J. , A comprehensive grammar of English language. London: Pearson Longman, 1985.

[37] D. Biber, Johansson, S., Leech, G., Conrad, S., Finegan, E. , Longman grammar of spoken and written English. London: Longman, 1999.

[38] S. Granger, Wynne, M. , "Optimising measures of lexical variation in EFL learner corpora," in Corpora Galore, J. Kirk, Ed. Amsterdam and Atlanta: Rodopi, 1999, pp. 249-257.

[39] W. Chafe, Danielewicz, J. , "Properties of spoken and written language," in Comprehending oral and written language, R. Horowitz, Samuels, J. S. , Ed. New York: Academic Press, 1987, pp. 83-113.

[40] M. Stubbs, Discourse analysis: The sociolinguistic analysis of natural language. Chicago: The Chicago University Press, 1983.

[41] J. Androutsopolous, "Introduction: Sociolinguistics and computer-mediated communication," Journal of Sociolinguistics, vol. 10, no. 4, pp. 419-438, 2006.

[42] R. Weber, "Variation in spelling and the special case of colloquial contractions," Visible Language, vol. 20, no. 4, pp. 413-426, 1986.

[43] J. M. Metz, "Computer-mediated communication: Literature review of a new context," Interpersonal Computing and Technology: An Electronic Journal for the 21st Century, vol. 2, no. 2, pp. 31-49, 2009.

[44] I. Vázquez, "A corpus-based approach to the distribution of nominalization in academic discourse," in Corpus linguistics: Applications for the study of English, A. M. Hornero, Luzón, M. J., Murillo, S. , Ed. International Academic Publishers: Bern, 2008, pp. 397-415.

AUTHOR'S PROFILE

Correspondence: Department of English, College of Sciences and Humanities in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia. E-mail: w.altohami@psau.edu.sa.

Waheed M. A. Altohami is assistant professor of linguistics at Department of English, College of Science and Humanities in Al-Kharj, Prince Sattam bin Abdulaziz University, Saudi Arabia. He also worked as a lecturer in linguistics, Department of Foreign Languages, Faculty of Education, Mansoura University, Egypt. His research interests are discourse analysis, semantics, corpus linguistics and translation.