

Word Shortening Strategies: Egyptian vs. Non-Egyptian English Tweets

Iman Mohamed Mahfouz¹

¹ Arab Academy for Science, Technology and Maritime Transport, Egypt

Correspondence: Iman Mohamed Mahfouz, Arab Academy for Science, Technology and Maritime Transport, Egypt. E-mail: imahfouz@aast.edu

Received: June 6, 2018 Accepted: June 26, 2018 Online Published: August 22, 2018

doi:10.5539/ells.v8n3p27 URL: <https://doi.org/10.5539/ells.v8n3p27>

Abstract

The language of Computer-mediated Communication (CMC) is known to deviate from standard language in many ways dictated by the characteristics of the medium in order to achieve brevity, speed as well as innovation. Together with the intrinsic features of CMC in general, the character limitation imposed by the popular social media platform, Twitter has triggered the use of a number of linguistic devices including shortening strategies in addition to unconventional spelling and grammar. Using two parallel corpora of English tweets written by Egyptians and non-Egyptians on a similar hashtag, the study attempts to compare the shortening strategies used in both datasets. A taxonomy for orthographic and morphological shortening strategies was adapted from Thurlow and Brown (2003) and Denby (2010) with particular focus on message length, punctuation, clipping, abbreviations, contractions, alphanumeric homophones and accent stylization. Given the scarcity of linguistic studies conducted on Egyptian tweets despite the vast amount of data they offer, the study compares the findings about tweets written by Egyptians in English as a foreign language to previous studies. The findings suggest that Egyptians tend to omit punctuation more frequently, whereas non-Egyptians favor abbreviations, contractions and clipped forms. The results also indicate that Twitter may be shifting towards longer messages while at the same time increasingly employing more shortening strategies. The study also reveals that character limitation is not the only factor shaping language use on Twitter since not all linguistic choices are governed by brevity of communication.

Keywords: character limitation, Egyptian tweets, shortening strategies, Twitter language

1. Introduction

Computer-mediated communication (CMC) is “the communication produced when human beings interact with one another by transmitting messages via networked computers” (Herring, 2001, p. 621). Owing to the ubiquity of the Internet in the 21st century, social networks have offered the unmatched opportunity to linguistically examine and analyze huge amounts of data from user interactions. Also referred to as *netspeak*, *netlingo* and *weblish*, the language of CMC can deviate from standard language conventions since it is affected by the brevity, speed, and creativity of this medium (Verheijen, 2015).

Launched as a micro-blogging service in 2006, Twitter imposed a 140-character count limit per tweet. This posed a challenging task to users since they had to write as concisely as possible, which often meant that they had to sacrifice adhering to some English language conventions. Users were thus obliged to communicate their ideas within this limitation, resulting in some frustration on their part. According to Gonzales (2012), 9% of English tweets reach the maximum limit, given that tweets written in English account for almost 40% of global tweets. In November 2017, however, Twitter decided to double the number of characters allowed for tweets with the aim of allowing more flexibility of self-expression. Nevertheless, some have viewed this as defeating the purpose of this platform since they think that “the beautiful brevity of Twitter is its bread and butter” (Kovac, 2017, par. 8). Arab users are no exception as they believe that other platforms, such as Facebook for instance may be more suited to more wordy posts (Khalife, 2017).

Character limitation, together with the intrinsic features of CMC in general have triggered the use of a number of linguistic devices including abbreviations and unconventional spelling and grammar. Twitter language has thus been characterized by being brief and compact and hence comparable to SMS (short message service) and IM (instant messaging) (Denby, 2010). Besides serving the purpose of character limitation, these shortening

strategies also result in an innovative use of language (Marko, 2016). As White (2015) maintains, “CMC discourse is typically described as using a simplified language which has the effect of making communication more efficient” (p. 72). This has been attacked by some purists who believe that it degrades the language. A case in point is Chomsky who claims that Twitter leads to “very shallow communication” (Jetton & Chomsky, 2011, question 37).

Opposing views, by contrast, challenge this popular view that language is deteriorating because of increased use of CMC, since it promotes creative and innovative language adaptation (Hård af Segerstad, 2002; Denby, 2010). The extent to which the language of CMC deviates from standard language depends on a number of factors, including individual user characteristics, such as age, gender, ethnic background, and familiarity with CMC, as well as discourse topic and recipient of the message (Verheijen, 2015). Hård af Segerstad (2002) holds that “specialized use of short forms is thus an indicator of belonging in a community and is a component of the group identity” (p. 214).

The number of active Twitter users in the Arab region is around 10.8 million, with Saudi Arabia taking the lead (29%), followed by Egypt (18%) (Mohammed Bin Rashid School of Government, 2017). Collectively, the Arab world generates 27.4 million tweets per day. The sweeping majority of Egyptian tweets are written in Arabic (90%), whereas only the remaining 10% of Egyptian tweets are written in English, which constitutes the principal foreign language of the general population in Egypt (#Egypt Twitter users, 2017).

Despite the widespread adoption of Twitter internationally, little research has investigated the differences among users of different languages. In fact, Hu, Talamadupula and Kambhampati (2013) describe the characteristics of Twitter language as an under-researched area. In prior research, the natural tendency has been to assume that the behaviors of English-speaking users also apply to other language users (Hong, Convertino, & Chi, 2011). It would thus be interesting to examine language use on Twitter by Egyptians as non-native speakers of English. The significance of the present study lies in the fact that very few studies have been conducted on tweets written by Egyptians. The present paper hence examines Egyptian tweets written in English in order to determine to what extent they employ word shortening strategies and how they are affected by character limitation in this medium using a taxonomy of strategies adapted from Thurlow and Brown (2003) and Denby (2010). This is compared to a parallel sample of tweets written by non-Egyptians on the same topic. The study also compares these findings to those reported about tweets written by native speakers.

2. Literature Review

Crystal (2001) maintains that new technologies have immediate linguistic consequences. The example he gives is the 160-character limit for SMS which has given rise to even more concise abbreviations than those used on the Internet. New spelling conventions which usually reflect pronunciation have also emerged online which are not taxed for being non-standard. Texts are thus shaped by limitations of space, as well as the constraints of typing and reading messages on different types of devices. The physical limitations, though, are not enough to account for the complex patterns and group variations found in CMC. The increasing popularity of Twitter and the ease of accessing user data have motivated a considerable number of scholars to tackle the content and linguistic analysis of Twitter data.

By applying Grice’s (1975) maxim of quantity which states that one should not make their contribution more informative than is required, we may expect CMC to be as brief as possible. By way of analogy, Thurlow and Brown (2003) developed three sociolinguistic maxims for CMC: (1) brevity and speed; (2) paralinguistic restitution; and, (3) phonological approximation (p. 15). The model they suggest for analysis is pioneering and can be useful to apply in different media. Rua (2005) also conducted a comparative study of shortening devices in text messages in English, French and Spanish and found that English favors initialization to letter reduction.

Several other scholars have attempted to identify the strategies used to reduce language in texting and CMC. In his key study, Denby (2010) explored linguistic innovation and the effect of character limitation on Twitter messages based on Ling and Baron’s model (2007) which was used to compare text messaging to IM. His findings indicate that character limitation does not significantly affect twitter language since twitter shows a strong tendency to avoid features such as abbreviation, initialisms and the omission of standard punctuation. He concludes that Twitter language surprisingly possesses more formal linguistic traits than it has been believed to (see also Maity, Ghuku, Upmanyu, & Mukherjee, 2016).

White (2013) examined language economization in chatting by non-native speakers of English as evidence of learner autonomy with particular reference to reduction and ellipsis. According to White (2015), the fact that non-native speakers seem to follow the same reduction patterns of native speakers suggests that reduction is a cross-cultural and universal process. He distinguishes different types of reduction, including syntactic,

morphological and orthographic reduction. Syntactic processes include the deletion of subjects, articles and modal auxiliaries as well as ellipsis. However, he mainly focuses on morphological and orthographic reduction, of which he discusses three main types: clipping, phonetic respelling and homophone respelling. Syntactic reduction, of which pronoun ellipsis or pro-dropping is the most common, can save no more than 4 characters by pronoun elimination, whereas using abbreviations can act as a more effective economization strategy (Hård af Segerstad, 2002). In general, higher frequency items, especially function words, are more likely to be subject to reduction processes (White, 2015).

Riggs (2012) also lists several ways that users can employ to shorten their tweets. These include using abbreviations, removing vowels and prepositions, as well as using symbols, such as + or & instead of *and*. Maity et al. (2016) conducted a quantitative study of the evolution of a number of sociolinguistic aspects on Twitter. They found that the average length of words on Twitter is decreasing and that users are employing more short forms to communicate.

In the Arab world, very few studies have been conducted on the language of Twitter. In fact, Khedr (2014) points to the scarcity of linguistic research on Arabic CMC in general. Most studies have been concerned with the role played by Twitter in political uprising (Kavanaugh, Yang, Li, Sheetz, & Fox, 2011), sentiment analysis and information retrieval (Darwish, Magdy, & Mourad, 2012; Alhumoud et al., 2015) or the use of Twitter in language teaching and learning (Allam, Elyas, Bajnaid, & Rajab, 2017). Those who considered Twitter in terms of language use, however, focused on its negative effect on the Arabic language and were merely speculations that lacked sound theoretical foundation.

Hence, from a theoretical perspective, the present research relies primarily on the premises of Thurlow and Brown (2003) for the brevity maxim of CMC which states that communication online is governed by brevity and speed, thus triggering the use of different forms of abbreviations in addition to the reduction of punctuation. The study also makes use of Denby's (2010) account of the characteristics of twitter language with specific reference to character limitation. White's (2015) notion of the universality of shortening devices in CMC is also relevant since the study examines tweets by non-native speakers of English compared to those produced by native speakers.

3. Data and Methodology

A total of two hundred tweets were collected over a period of four days from May 22nd till May 25th, 2018. One hundred tweets were gathered from each of two hashtag searches conducted on the famous Egyptian football player Mohamed Salah who plays for the English Liverpool club. The first hashtag *#محمد صلاح* is in Arabic and the second is in English *#Mo Salah*, designating the same player in order to neutralize the effect of the topic of discussion. The hashtags were chosen since they were trending on the four days preceding the Champions League Final 2018. To collect the first dataset with the Arabic hashtag, the search filter was limited to tweets posted "near you" in order to ensure that these tweets were written in Egypt. For the second dataset, gathered using the English hashtag, on the other hand, the search was not filtered by location in order to yield world tweets. In both searches the language filter was limited to tweets written in English. Tweets were gathered at random since the first 100 tweets generated by the above search were selected for both hashtags. User characteristics were not taken into consideration since not all Twitter users mention personal information such as age and gender in their profiles (it is generally known that personal information about online users is usually unavailable or unreliable). The focus here was thus on tweets posted by Egyptians in general as opposed to those posted by non-Egyptians.

A taxonomy of word shortening strategies was adapted from Thurlow and Brown (2003) and Denby (2010). Only features related to brevity were selected while others pertaining to CMC in general but not involving brevity, such as logograms or unconventional use of capitalization, for instance, were overlooked. The current study focuses on the following shortening strategies:

1. Message length (word and character count and mean, single-word transmissions)
2. Punctuation (sentence-final and transmission-final)
3. Clipping
4. Abbreviations (acronyms and initialisms)
5. Contractions (with or without apostrophes)
6. Alphanumeric homophones
7. Accent stylization

For corpus annotation, the selected shortening strategies were identified and classified manually. After initial data coding, all the data was checked to ensure that no instances of shortening were overlooked and to filter out any possible misclassifications. The quantitative results presented concern the number of instances encountered of each shortening strategy in both datasets. The following main research questions are addressed:

- What are the word shortening strategies used by Egyptians in tweets they write in English?
- How does this compare to tweets written by non-Egyptians as well as to previous findings?
- To what extent does character limitation affect language use in tweets written by Egyptians in English with respect to brevity of communication?

4. Data Analysis

In this section, the two datasets are compared and contrasted in terms of the use of selected shortening strategies, namely message length, omission of sentence-final and transmission-final punctuation, clipping, abbreviations including acronyms and initialisms, contractions with or without apostrophes, alphanumeric homophones as well as accent stylization.

4.1 Message Length

Message length was calculated using the standard Microsoft Word ‘*Word Count*’ option for both the number of words and characters. The total number of words in the first dataset is 1322 amounting to a total of 5755 characters and a mean of 58 characters per tweet (see table 1). The second dataset, on the other hand, is composed of 2186 words and 9689 characters yielding a mean of approximately 97 characters per tweet. None of the tweets in the entire sample consists of single-word transmissions, but the shortest tweets occur in the first dataset with four tweets consisting of two words only, as opposed to a minimum of four words encountered in two tweets in the second dataset.

Table 1. Message length in the two datasets

Message length	Dataset 1	Dataset 2
Number of characters	5755	9689
Number of words	1322	2186
Minimum words per tweet	2 (4 occurrences)	4 (2 occurrences)
Average characters per tweet	58	97
Average words per tweet	13.2	21.9

4.2 Punctuation

According to Crystal (2001), punctuation not only conveys a great deal about grammatical structure, but also compensates for the prosody and paralinguistic features of speech which are absent in written communication. It is, however, rather sparse and often absent in CMC. By omitting punctuation, the user saves time, effort and keystrokes (Hård af Segerstad, 2002). Lately, it has been noticed that the full stop is becoming extinct in CMC and is used only when the writer wants to add “an emotional charge to what’s being said” (Crystal, 2016, par. 1).

Punctuation was examined at the end of posts (transmission-final) and at the end of sentences (sentence-final), since several posts consisted of multiple sentences. Most of the tweets in the first dataset (69%) lack final punctuation, whereas only 46% of the tweets in the second dataset have no punctuation mark at the end. This includes both statements and questions.

Extract 1: Do you love Mo Salah? - WHO DOESN'T (dataset 1)

Extract 2: Mo Salah has won the PFA' Fans' Player of the Year award (dataset 2)

It is worth mentioning that tweets that end in emojis (small digital pictures representing things or feelings) never have punctuation at the end, suggesting that the emoji somehow compensates for the punctuation mark.

Extract 3: 5 days to go 😊

Extract 4: It looks good and maybe he did see it but if he says anything everybody will be pestering him so wouldn't take offence. Nice job 👍

Punctuation within posts is more critical to comprehension than at the end (Ling & Baron, 2007). In the first dataset, punctuation within tweets is sometimes skipped, though to a lesser extent, resulting in run-on sentences. Sentence-final punctuation within tweets is omitted in 11% of the tweets in the first dataset compared to only 3%

in the second dataset.

Extract 5: Get well prepared It will be a very difficult game for you

Extract 6: Congratulations momo for the BBC African footballer of the year u deserve it

Tweets in the second dataset, on the other hand, mostly preserve punctuation within, even if they skip it at the end. Some even abound with the use of standard punctuation marks despite their conversational style.

Extract 7: Date night tuesday. Mo salah: A Football Fairytale Documentary. Channel 4 10 pm

Extract 8: Klopp: "Mo Salah, world class but not every day. Sadio Mane, world class, but not everyday. Roberto Firmino, world class, pretty much every day. "Bang on the money. One of the best, and most underrated (still) players in Europe. It's about time he had his own song".

4.3 Clipping

In addition to traditional back-clipping which involves deleting letters at the end of a word, Bieswanger (2006) also distinguishes initial clipping in which letters are deleted at the beginning of a word and mid-clipping in which letters are deleted in the middle. Only a few cases of clipping or letter omission were found in the sample. The most obvious of which is *mo* for Mohamed (which is the name commonly used to refer to the famous footballer). The most frequent clipped word in the sample is *because* (6 instances) which is clipped in four different ways: *coz*, *cuz*, *cos*, and *cause*. Only four other cases of clipping were encountered in the first dataset, *uni* (university), *CHL* (channel), *mil* (million) and *mins* (minutes) each used only once. By contrast, several cases of clipping were found in the second dataset as shown in table 2. The majority of clipped words fall into the category of back-clippings (e.g., *physio*, *Fri* and *ads*), while only a few are mid-clippings (e.g., *ppl* and *f'ing*). It is noticed that most of the clipped forms in the sample are standard and commonly used, whereas only few are non-standard (e.g., *cos*).

Table 2. Clipped forms in dataset 2

Clipped form	Full form	Clipped form	Full form
stats	statistics	ppl	people
St	Saint	vs (x2)	versus
physio (x3)	physiotherapist	ad (x3)	advertisements
Fri, Sat	Friday, Saturday	f'ing	fucking

4.4 Abbreviations

Bieswanger (2006) defines abbreviations as “shortenings that consist of the first letter (or letters) of a combination of more than one word” (p. 4). According to White (2015), the term abbreviation encompasses both initialisms (e.g., OMG) and acronyms (e.g., FIFA) which are pronounced as a single word not as separate letters. The selected tweets contain several acronyms and initialisms. In accordance with the hashtags under study, a large number of these belong to the field of football as shown in table 3, with the abbreviation *LFC* (Liverpool Football Club) as the most frequent in the sample (a total of 9 instances). The second dataset also comprises other miscellaneous initialisms including *NYC* (New York City), *p* (pence) and *m* (million), whereas the initialism *MS* is used in the first dataset to refer to *Mohamed Salah*.

Table 3. Football related abbreviations in the two datasets

Dataset 1		Dataset 2	
Abbreviation	Full form	Abbreviation	Full form
	FC		Football Club
	LFC		Liverpool Football Club
ECL	European Champion Clubs	PFA	Professional Footballers Association
PL	Premier League	CR7	Cristiano Ronaldo 7 (shirt number)
FA	Football Association	UEFA	Union of European Football Associations
FC	Football Club	EPL	English Premier League
MS	Mohamed Salah	FPL	Fantasy Premier League
		UCL	UEFA Champions League
		NSC	National Sports Complex
		MOM	Man of the match
		MCM	Michael Cormer München

In the above table, only UEFA is an acronym, whereas all the others are considered initialisms. Very few common CMC abbreviations were found, namely *OMG* (Oh my God) and *LOL* (laughing out loud), of which the latter was used in both datasets. Only two twitter-related abbreviations were found, which are *RT* (retweet) and *S/O* (shouting out) with the latter occurring only in the second dataset. The abbreviation *isa* (Insha'Allah transliterated from Arabic, meaning God willing) is used in the first dataset which was written by Egyptians.

4.5 Contractions

Contractions are usually used in informal speech and writing, and are “shorter to type than full forms, especially when omitting the apostrophe” particularly since apostrophes require four keystrokes on mobile phones (Ling & Baron, 2007, p. 8). The sample abounds with contractions in both datasets. The percentage of contractions was calculated in terms of the number of contractions used compared to the total number of potential contractions (Ling & Baron, 2007). The first dataset comprises a total of 22 cases of contractions out of 48 potential ones (46%), whereas the second comprises 55 contractions out of 85 potential ones (65%). The non-standard contraction *y'all* (‘you all’ which is common in Southern American dialects) is used in the second dataset.

Extract 9: Because you're the reason for our happiness, you'll never walk alone our Egyptian king (dataset 1)

Extract 10: Can't believe @MoSalah isn't England captain... That's a joke!!! (dataset 2)

By contrast, several grammatical full forms are used where contractions could have been possible.

*Extract 11: Success **does not** only mean an excellence in study... (dataset 1)*

*Extract 12: He **had been** complying since May 16, but has stopped according to Ruben Pons, the Spanish physiotherapist who treats him. After the final **he will** take it up again. (dataset 2)*

Only very few contractions are used without an apostrophe (9% of the contractions in the first dataset and 5.5% in the second). These include *dont* which is used once in the first dataset and twice in the second. The only two other instances of contractions used without an apostrophe are *its* in the first dataset (non-native speakers usually confuse this contraction with the possessive pronoun with the same spelling) and *thats* in the second dataset, each occurring only once.

Extract 13: i dont know what happen (dataset 1)

Extract 14: Dont sell lovren thats salah friend (dataset 2)

4.6 Alphanumeric Homophones

Alphanumeric homophones are among the most salient features of CMC. These refer to the use of a letter or number to represent the phonetic sequence that constitutes its realization in spoken language. The pronunciation of the letters/numbers is identical with parts of words, enabling them to replace a letter or letter sequences. There are three main types: letter homophone (*u* meaning *you*), number homophone (*4* meaning *for*) and a combination of letter and number homophones (*b4* meaning *before*) (White, 2015). In addition to serving as a shortening strategy, they also help develop a creative way of writing (Marko, 2016).

Only four alphanumeric homophones were encountered in the sample, namely *u* (you), *ur* (your), *z* (the) and *be4* (before). Of these, *u* and *z* are single-letter homophones since an entire word is substituted by one letter. It is interesting that the letter *z* is used instead of the word *the* in the first dataset, as this coincides with a common pronunciation error by Egyptian learners of English who tend to replace the voiced dental fricative with the

alveolar one. The ampersand logogram ‘&’ representing the conjunction *and* is used seven times in the second dataset but only once in the first. No other instances of alphanumeric homophones were found. Table 3 shows the percentage of use for the alphanumeric homophones compared to the total number of occurrences of the word in the sample.

Table 4. Use of alphanumeric homophones in the two datasets

Alphanumeric homophone vs. full form	Dataset 1	Dataset 2
u vs. you	24%	20%
ur vs. your	17%	0%
be4 vs. before	100%	0%
z vs. the	0.02%	0%
& vs. and	6.7%	26.9%

4.7 Accent Stylizations

This spoken-like spelling is also known as *phonetic respelling* or *phonological approximation* since it refers to spelling which imitates the phonetic value of speech. It is as though people “write it as if saying it to establish a more informal register which in turn helps to do the kind of small-talk and solidary bonding they desire” (Thurlow & Brown, 2003). Accent stylization sometimes takes fewer keystrokes to type, whereas some accent stylizations take just as many or more keystrokes than the conventionally spelt word (Hård af Segerstad, 2002). Phonetic respellings are different from clippings in that they contain at least one character that is not part of the standard spelling of the word in question (Bieswanger, 2006).

The data shows only three instances of accent stylizations in the first dataset as opposed to five cases in the second. The only case of g-dropping occurs with the word *fuckin* in the second dataset. In all cases the phonetic respelling is shorter than the original form except for *yeah* which is one keystroke more than *yes*.

Table 5. Accent stylization in the two datasets

Dataset 1		Dataset 2	
Accent stylization	Original form	Accent stylization	Original form
yeah	yes	gonna	going to
gotta	got to	fuckin	fucking
luv	love	ya (x2)	your
		fella	fellow

5. Findings and Discussion

This section presents a comparison between the two datasets, as well as a comparison between the findings of the present study and previous studies where applicable.

The data shows that tweets written by non-Egyptians are considerably longer and wordier than those written by Egyptians and are composed of more complex sentences (see figure 1). Egyptian users of Twitter tend to omit more punctuation marks both at the end of sentences and tweets than non-Egyptians who display more standard use of punctuation. It is not clear, however, whether this should be viewed as a shortening strategy or a language error since it is not unusual for learners to produce run-on sentences. Punctuation omission constitutes the largest portion of shortening strategies in both datasets and in the overall total.



Figure 1. Comparison of shortening strategies between the two datasets

Most cases of clipping fall under traditional back-clipping. This strategy is more frequent in the second dataset which displays a wider variety of clipped forms, suggesting higher language proficiency and awareness of word shortening strategies on the users' part. The majority of clipped forms are nouns, although the conjunction *because* is the most frequently clipped form in the sample and is clipped in a number of different ways.

The same applies to abbreviations which are more frequent in the second dataset. Football-related abbreviations are much more frequent (75%) than common CMC abbreviations (25%), suggesting that the topic of discussion may have a stronger effect than the medium in the type of abbreviations used. The second dataset also displays a higher tendency for contractions, which constitutes the second most frequent shortening device in the sample. Contractions without apostrophes are very rare in the sample in general. Although alphanumeric homophones and accent stylization are among the main characteristics of CMC, they are rather scarce in both datasets. Only function words are shortened using alphanumeric homophones in the data, whereas relatively more instances of accent stylization appear in the second dataset.

Table 6 summarizes the use of shortening strategies in the two datasets. The total frequency of shortening strategies is standardized per 1000 words since the two samples differ in word count. The overall frequency of shortenings per message in each of the corpora was calculated by dividing the overall number of shortening tokens by the number of text messages analyzed (Bieswanger, 2006). Although the total number of shortening strategies is higher in the second dataset, the first dataset displays a higher tendency for shortening when adjusted according to the word count. However, the second dataset shows a slightly higher overall frequency of shortening per message.

Table 6. Total use of shortening strategies in the two datasets

Shortening strategy	Dataset 1	Dataset 2
Message length mean in characters	58	97
Message length mean in words	13.2	21.9
Omission of transmission-final punctuation	69	46
Omission of sentence-final punctuation	11	3
Clipping	8	17
Abbreviations	13	35
Contractions	22	55
Contractions without apostrophe	2	3
Alphanumeric homophones	9	4
Accent stylization	3	5
Total shortening strategies	137	168
Shortening strategies adjusted to 1000 words	103.6	76.9
Overall frequency of shortening per message	1.37	1.68

It has been noticed that several words in the sample violate the maxim of brevity in a number of various ways. One of these is expressive lengthening which is a common feature of CMC (e.g., *LOOOOOL*, *GOAAAL*, *OMGG*, *MOOO Salah*). According to Hård af Segerstad (2002), "the economy principle is not absolute ... and may be overruled by other principles, for example the social, emotional effect of language play" (p. 213). Denby (2010)

also believes that “it is most probably unwise to absolutely conclude that ... shortenings are a result of the imposed character limitation on Twitter messages” (p. 27).

Punctuation marks are sometimes also repeated (e.g., ??, !!!, ...). Another case in point is using numbers spelt out in letters where they could have been written in numbers to minimize the number of characters (e.g., *two*, *three*). The use of white spaces also counts in the number of characters and was sometimes used where it was unnecessary or even erroneous, especially in the first dataset (e.g., *every one*). Moreover, some words were repeated for emphasis or effect.

Extract 15: Love you Love you Love you Love you Love you (dataset 1)

Extract 16: Allez Allez Allez (dataset 2)

The exact classification of particular shortening processes is rather controversial (White, 2015), since some processes may overlap or the same lexical item may display the use of more than one shortening strategy. For instance, *coz* may be classified either as clipping or accent stylization and *fuckin* may be classified as clipping or as g-dropping.

Table 7 compares the findings of the present study to Denby’s (2010) findings where applicable. The message length of tweets in the first dataset is very close to Denby’s (2010) finding for both words and characters, as opposed to that in the second dataset which shows a much higher message length. This implies that a change towards longer tweets may have been triggered by the recent doubling of character limit imposed by Twitter (see section 1). It is surprising, however, that this change was not reflected in tweets written by Egyptians.

Table 7. Comparison of findings to Denby (2010)

	Dataset 1	Dataset 2	Denby (2010) (adjusted to 100 messages)
Message length			
Message length mean in words	13.2	21.9	13.79
Message length mean in characters	58.6	96.9	65.21
Single-word transmissions	None	None	1.7
Punctuation			
Sentence-final	89%	97%	81%
Transmission-final	31%	54%	70%
Contractions			
Percentage of potential	46%	65%	63%
Standard apostrophe use	91%	94.5%	97%
Abbreviations			
	13	35	4
Clipping			
	8	17	1.9
Alphanumeric homophones			
	9	4	0

As regards punctuation, the second dataset displays the highest adherence to sentence final punctuation (97%) among the three samples. Tweets written by Egyptians fall short of the others where transmission-final punctuation is concerned (31%). In general, the sample of the present study points to more frequent omission of transmission-final punctuation than reported in the literature, which may suggest a trend towards increased punctuation omission at the end of tweets, as opposed to that within messages which is more keenly maintained in the sample of the present study.

Overall, Egyptians employed less contractions than non-Egyptians whose use of contractions was found to be comparable to a large extent to Denby (2010). Egyptian tweets also relatively displayed the least adherence to standard apostrophe use among all other tweets. It is surprising that both datasets of the present study reveal a significantly higher frequency of word shortening strategies including abbreviations, clipping and alphanumeric homophones than reported by Denby (2010) which may again indicate that this medium of CMC is increasingly favoring word shortening strategies.

6. Conclusion

Using two parallel samples of English tweets written by Egyptians and non-Egyptians on a similar hashtag, the study attempted to compare the strategies used for word shortening in both datasets. Several orthographic and morphological shortening strategies were observed in the data with particular focus on message length, punctuation omission, clipping, abbreviations, contractions, alphanumeric homophones and accent stylization

though with varying frequencies.

The relatively limited sample and scope of the present study may not render the findings generalizable to other larger corpora, let alone to Twitter as a medium of CMC. However, they can point to general patterns. The fact that Egyptians made use of several shortening strategies suggests that these strategies are cross-cultural and universal. However, some differences were detected between the two datasets indicating differences in preferences for shortening strategies between Egyptian and non-Egyptian users. Egyptians as non-native speakers were found to write shorter messages that contain less contractions, abbreviations and clipped forms which may require more language proficiency and awareness. Their rather frequent use of punctuation omission may not be merely viewed as a shortening strategy since it may reflect a linguistic error. It can thus be argued that non-Egyptians display better mastering of shortening strategies which is also confirmed by the higher score for the overall frequency of shortening per message in the second dataset.

The results of the present study display some variation from the literature in a number of aspects. The fact that average message length is higher than is commonly reported is probably due to the recent doubling of Twitter character-limit count. The higher frequency of omitted full-stops at the end of tweets may point to a trend towards favoring punctuation omission in this CMC medium. Similarly, the relative abundance of word shortening strategies, particularly clipping, abbreviations and alphanumeric homophones may also suggest that Twitter users are increasingly adopting these CMC devices.

Despite the role played by character limitation in shaping language use on Twitter, the results of the present study confirm Denby's (2010) findings that the linguistic features of this medium may be more inclined towards standard written language in comparison to other forms of CMC which possess more informal and non-standard features. This is confirmed by the several standard abbreviations and contractions, as well as the relative scarcity of alphanumeric homophones and accent stylization compared to other CMC mediums such as SMS and IM. It may have been expected that, if character limitation did indeed have a substantial effect on linguistic choices in tweets, features such as abbreviations, clipping and alphanumeric homophones would be prevalent; in fact, this was not the case, with these features only appearing in relatively few instances across the data.

Character limitation is, therefore, not the only factor that shapes the language of this microblogging medium. This has become all the more true especially after raising the character limit. The findings of the present study hence suggest that although shortening strategies are used in Twitter messages, tweets do not always adhere to Thurlow and Brown's (2003) brevity maxim. Moreover, the motivation for employing these strategies is not restricted to the need to shorten messages, but also to sound witty and innovative or to signal solidarity with a group of participants or online community.

Further studies are needed to investigate other linguistic features of Twitter discourse, as well as similar features in other forms of CMC, particularly Facebook, Instagram and chatting. It would be interesting to find out how the use of shortening strategies correlates with individual characteristics or user profiles, which was beyond the scope of the present research, especially across age and gender, or whether shortening is affected by the topic under discussion. Future studies are also needed to examine other types of shortening, especially from a syntactic perspective. Furthermore, it may be useful to conduct a contrastive study to compare the shortening strategies used by Egyptians in Arabic tweets, both standard and colloquial, to those used in English as a foreign language, especially that Arabic is known not to lend itself easily to abbreviations.

References

- Alhumoud, S., Altuwajri, M., Albuhaire, T., & Alohaideb, W. (2015). Survey on Arabic sentiment analysis in Twitter. *World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering*, 9(1). Retrieved from scholar.waset.org/1307-6892/10000687
- Allam, M., Elyas, T., Bajnaid, A. & Rajab, H. (2017). Using Twitter as an ELT tool in the Saudi EFL context. *International Journal of Linguistics*, 9(5). <https://doi.org/10.5296/ijl.v9i5.11813>
- Bieswanger, M. (2006). 2 abbrevi8 or not 2 abbrevi8: A contrastive analysis of different space and time-saving strategies in English and German text messages. *Grüneburgplatz*, 1(142). Retrieved from <https://goo.gl/gVRoKD>
- Crystal, D. (2001). *Language and the Internet*. UK: Cambridge University Press. <https://doi.org/10.1017/CBO9781139164771>
- Crystal, D. (2016, June 11). On the reported death of the full-stop / period. *DC Blog*. Retrieved from <http://david-crystal.blogspot.com/2016/06/on-reported-death-of-full-stop-period.html>

- Darwish, K., Magdy, M., & Mourad, A. (2012, October). Language processing for Arabic microblog retrieval. *Proceedings of the 21st ACM International Conference on Information and Knowledge Management* (pp. 2427-2430). Retrieved from <http://www.qcri.com/app/media/2044>
- Denby, L. (2010). *The Language of Twitter: Linguistic innovation and character limitation in short messaging* (Master Dissertation). University of Leeds, Department of Linguistics and Phonetics. Retrieved from <https://goo.gl/WCTGBd>
- #Egypt Twitter users – Digital boom (2017). [PowerPoint slides]. Retrieved from <https://tinyurl.com/y9qtzpd7>
- Gonzales, J. P. (2012). Conventions of the English language in the Twitter verse. *Prized Writing*. Retrieved from <https://goo.gl/Xmpuwk>
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech acts* (pp. 41-58). New York: Academic Press.
- Hård af Segerstad, Y. (2002). *Use and adaptation of written language to the conditions of Computer-Mediated Communication* (Doctoral dissertation). Department of Linguistics, Göteborg University, Sweden. Retrieved from <https://goo.gl/UA4K8m>
- Herring, S. C. (2001). Computer-mediated discourse. In D. Tannen, D. Schiffrin, & H. Hamilton (Eds.), *Handbook of discourse analysis* (pp. 612-634). Oxford: Blackwell. Retrieved from <http://ella.slis.indiana.edu/~herring/cmd.pdf>
- Hong, L., Convertino, G., & Chi, E. H. (2011). Language matters in Twitter: A large scale study. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. Retrieved from <https://goo.gl/ghiwTo>
- Hu, Y., Talamadupula, K., & Kambhampati, S. (2013). Dude, srsly?: The surprisingly formal nature of Twitter's language. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*. Retrieved from <https://goo.gl/TP7ot6>
- Jetton, J. (Interviewer) & Chomsky, N. (Interviewee). (2011, March). *The secret of Noam: A Chomsky interview*. [Interview Transcript]. Retrieved from <https://goo.gl/deg2kZ>
- Kavanaugh, A., Yang, S., Li, L., Sheetz, S., & Fox, E. (2011). Microblogging in crisis situations: Mass protests in Iran, Tunisia, Egypt. *Workshop in Conjunction with the ACM Conference on Human Factors in Computing Systems (CHI '11)*. Retrieved from <http://hdl.handle.net/10919/52858>
- Khalife, L. (2017, Nov.). Arabs are really hating Twitter's 280 character update. *Step Feed*. Retrieved from <https://stepfeed.com/arabs-are-really-hating-twitter-s-280-character-update-3476>
- Khedr, M. Z. (2014, Nov.). *The status quo of the Arabic language in the domain of social media on the Internet*. Retrieved from www.majma.org.jo/res/data/seasons/32/32-7.doc
- Kovac, K. (2017, Sept.). Twitter unveils new 280-character count feature. *Sixth City Marketing*. Retrieved from <https://goo.gl/HhbHLd>
- Ling, R., & Baron, N. (2007). Text messaging and IM linguistic comparison of American college data. *Journal of Language and Social Psychology*, 26(3), 291-298. <https://doi.org/10.1177/0261927X06303480>
- Maity, S. K., Ghuku, B., Upmanyu, A., & Mukherjee, A. (2016). Out of vocabulary words decrease, running text prevail and hashtags coalesce: Twitter as an evolving sociolinguistic system. *System Sciences (HICSS), 2016 49th Hawaii International Conference*, 1681-1690. <https://doi.org/10.1109/HICSS.2016.212>
- Marko, D. (2016). The Use of alphanumeric symbols in Slovene tweets. *4th Conference on CMC and Social Media Corpora for the Humanities*. Retrieved from <https://goo.gl/bYsBKY>
- Mohammed Bin Rashid School of Government. (2017). *Arab social media report 7th edition: Social media and the Internet of things*. Retrieved from <https://goo.gl/MFsGqi>
- Riggs, N. (2012, April). *Twitter tips – 17 ways to shorten your tweets*. Retrieved from <https://www.nrmedia.biz/blog/twitter-tips-shorten-my-tweets>
- Rua, P. L. (2005). Shortening devices in text messaging: A multilingual approach. *Neophilologische Mitteilungen*, 106, 139-155. Retrieved from <https://goo.gl/JtR2Ga>
- Thurlow, C., & Brown, A. (2003). Generation Txt? The sociolinguistics of young people's text-messaging. *Discourse Analysis Online*, 1(30). Retrieved from <https://goo.gl/hk4SRT>

- Verheijen, L. (2015). Out-of-the-ordinary orthography: The use of textisms in Dutch youngsters' written computer-mediated communication. *Proceedings of the Second Postgraduate and Academic Researchers in Linguistics at York (PARLAY 2014) Conference*. Retrieved from <http://goo.gl/MFJUoX>
- White, J. R. (2013). Language economy in Computer-Mediated Communication: Learner autonomy in a community of practice. *Computer-Assisted Foreign Language Teaching and Learning: Technological Advances*.
- White, J. R. (2015). Processes and variations in language economisation. *Ampersand*, 2, 72-82. Retrieved from <http://dx.doi.org/10.1016/j.amper.2015.06.001>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).