Nonverbal Communication and Emojis Usage in Arabic Tweets: A Cross-Cultural Study

Miramar Etman, Seham Elkareh

Phonetics and Linguistics Department, Faculty of Arts, Alexandria, Egypt
Email: miramar.etman@alexu.edu.eg, seham.elkareh@alexu.edu.eg

Abstract

People spend most of their time communicating their thoughts, ideas, attitudes, and emotions on social media platforms like Twitter, however, an important mode of communication as the nonverbal component which requires visual and audible cues is not allowed due to the nature of these text-based platforms. The aim of this research is to discover the alternative ways Arabs use across different dialects to compensate for the absence of the nonverbal component. To be able to discover that the researchers collected a corpus of tweets written in the Arabic language by using python through the Twitter application programming interface (API). The results can be summed up as follows: emojis helped Arabs to communicate their facial expressions and the top used emoji across the different dialects was Face with Tears of Joy, it was also apparent that the top used emojis reflected the universal emotions, regarding the usage of hand gestures, Egyptian dialect came in the first place and Emirati dialect in the second place. Prosodic features such as the tone and loudness of the voice are expressed by the mean of character repetition, Punctuation usage across the Arabic dialects was limited, and Lebanese seemed to use them the most, Arabs tend to replace punctuation marks with emojis, finally, Arabs used vocal expressions like Interjections to communicate their affective state.

Keywords

Nonverbal, Twitter, Cross-Cultural, Emojis, Gestures, Paralinguistic, Prosodic, Interjections

1. Introduction

Non-verbal communication is a system involving an assortment of features often used together to support expression [1]. The field of examining nonverbal

communication has begun with the publication of Charles Darwin book The Expression of the Emotions in Man and Animals in 1872 [2] and it is considered as the spark that influenced many researchers to study nonverbal behavior and after that, Ekman and Friesen presented a classification for five types of communicative movements which reflected communicative functions [3]. According to researchers in [4], we can classify nonverbal communication into three main categories: Kinesics, Haptics, and Proxemics. Each of these aspects is further divided into other types. Kinesics includes the use of gestures, head movements and posture, eye contact, and facial expressions, Proxemics includes any aspect that relates to space and distance and how it influences our communication, and Haptics includes any kind of communication by touch. Other researchers as [1] include in addition to the previously stated types Vocalics or paralinguistics like Tone of voice, volume, speed, and pitch and Appearance like hairstyle and clothing into their classification for nonverbal communication.

Nonverbal communication is an integral part of our daily face-to-face communication. Most of the studies suggest that nearly 60 - 70 percent of our communication accounts for the non-verbal component [5], other researchers suggest the Albert Mehrabian's 7-38-55 percent communication rule which proposes that only 7 percent of our communication is verbal, and 93 percent is nonverbal; 38 percent vocal, and 55 percent facial [6]. From the previously stated percentages, it is apparent how humans cannot communicate ideas, feelings, and attitudes efficiently without relying on nonverbal cues, however, on Text-based computer-mediated communication as social media platforms communicators only communicate through text which poses a challenge for them to express the nonverbal cues while delivering their messages. One of the situations that influenced the researchers to study the representations of the nonverbal cues on Twitter is the misunderstanding that occurs between people while texting, for example, people could complain that one of the texters is rising his/her voice when one character is repeated inside the word or when an emoji does not fit the context although the verbal message is written correctly and that reflects the importance of the nonverbal cues in written messages and how it is more strong than the verbally written messages.

The following research tries to answer the following questions. How Arabs across different dialects compensate for the absence of nonverbal cues on Twitter? Does Twitter limit and restrict the use of nonverbal cues? How the nonverbal components as facial expressions, paralanguage, and prosodic features are expressed by the means of emojis, character repetition, punctuations, and interjections?

2. Related Work

Few works had been done on examining nonverbal cues across text-based computer-mediated communication especially on social networking platforms like Facebook and Twitter, however, in the following years with the increasing num-

ber of users on online platforms and after the COVID-19 pandemic which directed people towards a more digitalized world, this number will increase. This section will present the previous work examined nonverbal communication on Twitter.

Park and El Mimouni study [7] is considered one of the most important studies that focused on cross-cultural nonverbal communication across Twitter, the main goal of this study is to examine how users are drawn from three different languages and cultural boundaries manage the lack of contextual cues through an analysis of Arabic, English, and Korean tweets. They examined linguistic contractions, (un)conventional shorthand, prosodic features, punctuations, Language-specific features, and emoticons and showed by examples how each of these categories are expressed across the three languages.

Tantawi and Rosson [8] investigated the paralinguistic function of emojis on Twitter, 1600 tweets were examined. The analysis conducted on the collected tweets was divided into two main aspects; the first aspect is topic analysis to find out the most common topics on Twitter and the second aspect focused on finding out the emojis function, the tweets they analyzed contained three paralinguistic features for emojis; attitude, gesture, and topic. Their results showed that emojis are primarily used to signal attitudes and emotions. They also discussed implications for the design of Twitter and other text-based communication tools.

Álvarez and Muñoz study [9] has twofold goals; the first one is to describe tweets on Twitter from a discourse contrastive point of view and to study several numbers of distinguishing features of language use that are specific to this social network in Spanish and English [9]. One of the most important findings of the study is how the character limitation that Twitter force upon their users did not cause communication failures on the contrary they were able to communicate themselves very well.

3. Methodology of the Study

To reach the research goals the following stages have been applied: the first stage was concerned with data collection, the second stage focused on data cleaning and normalization, the third stage is the data analysis which has been done by using python scripts to get the corpus pure structure information and measurements for the nonverbal devices categories which are presented in Table 1. In

Table 1. Nonverbal devices categories.

Category	Description
Emojis	graphical symbols oricons used in text-based communication
Paralanguage and Prosodic Features	Intonation, Tone of the voice, stress and paues
Interjection	Purely emotive words used to signal feelings and emotions

addition to that, a manual analysis was done to investigate the nonverbal devices inside the corpus for each Arabic dialect.

3.1. Data Sample

The data were collected by using Python through Twitter API, since the main study purpose is to examine the nonverbal cues and usage of emojis, the researchers collected the tweets by using 91 emojis that represent facial expressions and 20 hand gestures emojis. For each emoji, 100 tweets were collected but for very few numbers of emojis, less than 100 tweets were collected as usage of those emojis is not widespread. Totally more than 10,000 were obtained. The Arabic language was specified in the search query so only Arabic tweets were collected.

3.2. Data Cleaning and Normalization

The collected corpus from the API had duplicated tweets and sometimes Urdu language tweets were obtained due to the borrowed letters the language has from Arabic alphabet, as a result those duplicated and mistakenly collected tweets were removed.

Regarding the tweets which have mentioned location there was a challenging task since Twitter users could use multiple alternative names to refer to the same country or use the city which they live without referring to the country, this kind of unbalancing data would make further analysis problematic and difficult to work on. The researchers decided to normalize the location for all tweets, so each country has only one name shape which refers to it, for example, "egy", "Egypt, Alex" and "Alexandria" are normalized to "Egypt". All the data cleaning and normalization were done by using simple techniques in Excel spreadsheets like removing duplicates for removing the repeated tweets.

4. Data Analysis and Results

After data cleaning and normalization, the analysis was performed on two main levels; the first level focused on examining the emojis usage in Arabic language and across different dialects and that was analyzed with the help of custom python scripts. The second level focused on describing what devices twitter users use to compensate for the absence of nonverbal cues with 280 characters restriction for each tweet. For dialectal analysis, only dialects with more than 100 tweets were subjected to further analysis. Table 1 summarizes the nonverbal device categories examined in our study; some of these categories were derived from previous literature analysis.

4.1. Corpus Pure Structure Information

Corpus pure structure information like the number of tweets for each dialect, the average number of characters, and the number of words were obtained to gain insights about whether Twitter users are able to communicate efficiently with 280-character limitation or they need more characters to deliver their messages.

Table 2. Corpus pure structure information.

Total Number of tweets	Average number of characters per tweet	Average number of words per tweet
8726	69	8

Table 3. Pure structure information for each dialect.

Dialect	Total Number of tweets	Average number of characters per tweet	Average number of words per tweet
Saudi	1281	72	8
Egyptian	631	70	8
Kuwaiti	248	63	8
Emirati	112	67	8
Iraqi	109	67	8
Lebanese	104	76	9

Table 2 above presents the pure structure information for the whole corpus after data cleaning and normalization and Table 3 Shows pure structure information among each dialect. From the results above we may infer that character limitation which Twitter imposes upon their users is not problematic for Arabic native speakers as the average tweet length is 69 characters with a total of 8 words. It also seems that Saudi and Lebanese dialects use more characters than other dialects. We can relate the reason behind the highest number of characters per tweet that Lebanese use with the fact that they use punctuation marks more than other dialects, this is apparent in Table 8 below.

4.2. Emojis

Emojis are graphic representations of facial expressions, body language and hand gestures, emojis are one of the most important nonverbal devices that text-based communicators use to signal emotions and attitudes. People sometimes use emojis and emoticons interchangeably, however, the two terms have different implications; Emoticons refer to a series of text characters (punctuation or symbols) that are utilized to textually form a gesture or facial expression [10] while emojis refers to the graphical icons that appear on keyboards especially on mobile texting applications and are used directly without the need to use any textual characters. Emojis help to feel the mood of a chat and the tone of a relationship. With emoji experience, people may infer about the things that are not expressed concretely [11]. In this study only emojis were examined for Arabic native speakers across dialects as they are more widespread over the social media platforms and more convenient for the study purpose. Emojis on Twitter always counted as two characters. Table 4 and Table 5 show the results for emojis usage analysis for the whole Arabic tweets and across dialects, respectively. Regarding the whole corpus it is apparent that Arab Twitter users use 😂 Face with Tears of Joy most of the time in their tweets and that reflects the importance and

Table 4. Emojis usage in Arabic tweets.

Number of emojis	Average number of emoji per tweet	Percent of emoji to number of characters	Most used emojis faces and gestures with their count				
23,418	3	5%	😂 1787	😭 1141	😅 382	🤣 356	😩 304
			😳 260	😫 255	😔 215	😶 213	😁 208

Table 5. Emojis usage for each dialect.

Dialect	Number of emojis	Average number of emoji per tweet	Average percent of emoji to number of characters	Most used emojis faces and gestures with their count				
Saudi	3337	3	4%	😂 245	😭 108	🤣 60	😅 54	😩 51
				🙂 49	😔 47	😖 42	😏 37	😬 35
Egyptian	1643	3	4%	😂 226	🤣 40	😅 39	👇 29	✊ 24
				😋 24	😓 21	🙁 21	👆 21	😐 20
Kuwaiti	541	2	4%	😂 48	😍 16	😅 13	😒 13	😭 11
				🤣 11	😌 8	😏 8	😙 7	🤔 7
Emirati	292	3	5%	😂 21	😅 18	🤣 14	😏 13	😩 11
				😳 7	😖 6	😄 5	😔 5	✊ 4
Iraqi	241	2	4%	😂 21	😒 13	🙂 9	😭 8	😐 5
				🙂 5	😅 5	😁 4	🤨 4	😌 4
Lebanese	218	2	4%	😂 29	🤣 9	🙂 7	😟 7	😋 6
				🤓 6	🙏 6	😌 5	😶 5	👇 5

the need of communicating their facial expressions while laughing. The second-place emoji differs across dialects but in the whole collected corpus the 😭 Loudly Crying Face emoji came in the second place. It can also be noted that the top emojis represent universal emotions. Hand gestures usage was prevalent in the Egyptian dialect and that could be noticed in the Egyptian face-to-face conversations where they use many gestures to communicate their message. Emirati and Lebanese top used emojis also had hand gestures. Saudi, Egyptian and Emirati use on average 3 emojis per tweet while Kuwaiti, Iraqi and Lebanese use 2 emojis per tweet. It is also worth mentioning that the percent of emojis number to the tweet characters across the whole corpus is 5% and that is not a low percent.

4.3. Paralanguage and Prosodic Features

One of the most important nonverbal cues after facial expressions that have the

power to change the meaning of words and reflect the speaker's emotional state, whether an utterance is a statement, question, or command and whether the emphasis or focus on a specific idea is the nonverbal cues of the voice as tone pitch and accent. When Twitter users wanted to emphasize and show the intensification of their emotions, they used character repetition to compensate for the absence of prosodic features. In prosodic terms the repetition of letter characters would seem to correspond to the duration feature and the capitalization to the feature loudness [12]. Since Arabic lacks the feature of capitalization both duration and loudness were expressed by the means of character repetition. Character repetition happen in vowels more than consonants and it also happens with emojis. Table 6 shows examples from Arabic dialects and as it is clear no dialect lacked the feature of character repetition to express nonverbal prosodic features.

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 [13]. Arabs usage for punctuation marks is very limited especially for the full stop, comma is the most used mark and the exclamation mark comes in second place followed by the question mark. It seems like emojis are replacing punctuation marks, for example, exclamation mark in some tweets is replaced by or accompanied with 🙁 Frowning Face with Open Mouth (e.g., 🙁 🙁 ياترا من يروح وقتنا معقول يجي احد احسن من بتس which translated as "I wonder as time passes is it possible that someone will come better than BTS" 🙁 🙁). Question mark is also replaced by or accompanied with 🤔 Thinking Face (e.g., 🤔 ليش التفرقة which translated as "why would you make a difference"). Lebanese dialect uses more punctuation marks than the other dialects and Emirati dialect nearly never uses them. But if we compared the Emirati usage for emojis we will find that it uses more emojis and that may replace the usage of punctuation marks. People feel very comfortable while communicating on social media platforms and they do not concern themselves with the grammatical structure of their sentences and that can be shown in the limited percent of punctuation marks per tweet and when it is used the main reason is to convey and express their emotions and attitudes. Table 7 and Table 8 show punctuation marks used in the whole corpus and among different dialects.

4.4. Interjections

Interjections are vocal expression of the emotions, Interjections are sound sequences, words, typical phrases, or clauses which can be realized as utterances signaled in speech by being produced with greater intensity, stress, and pitch, and as sentences in writing by an exclamation mark [14]. In Arabic, interjections are of two types: nouns of sounds (Asmaa' Al-Aswaat) and nouns of verbs (Asmaa' Al-Af < aal) [15]. Some researchers classify the interjections into three main classes which reflect the speaker's mental state or act. The classification includes three classes: emotive, volitive, and cognitive interjections (Wierzbicka 1992 as cited in [15]). Interjections used in the corpus mainly reflected the

Table 6. Character repetition examples from dialects.

Dialect	Example
Saudi	😫 🤭 🤭 ياااااااااارب
Egyptian	🤣 🤣 🤣 🤣 🤣 الجو روووووعة بس انتم عواااجيز 😷
Kuwaiti	" ألف ألف ألف مبروك 🤍 🎊 كم من الجماااااال يوازي كم شوقنا لك 😂 🔥 😂 😂 😂 الذهب يظل ذهب
Emirati	🤣 😻 😻 😻 😻 😻 مبرووووووك يا انتتي 🤍 💪 هيا اديهم دودوووووووووم 🔥 🔥 🔥
Iraqi	🙂 🙂 ههههههههههههه
Lebanese	" ياي يا جماعة يااااي !! بيحتار الواحد بالروح المعطاءة يلي عندك أو الروح التواقة للإصلاح! 😁

Table 7. Punctuation marks across the whole corpus.

Comma (،)	Question Mark (؟)	Exclamation Mark (!)	Period (.)	Average percent of punctuation to number of characters
702	527	628	135	0.2%

Table 8. Punctuation marks across different dialects.

Dialect	Comma (،)	Question Mark (؟)	Exclamation Mark (!)	Period (.)	Average percent of punctuation to number of characters
Saudi	109	74	108	27	0.3%
Egyptian	60	31	22	4	0.2%
Kuwaiti	23	15	30	5	0.3%
Emirati	2	3	6	1	0.1%
Iraqi	15	5	5	1	0.2%
Lebanese	14	10	19	0	0.4%

Table 9. Examples for tweets contain interjections found in the corpus.

Interjection	Example	Meaning
"واو"	واووو ماشاءالله 😍	Translated from english interjection "wow" which used as an exclamation of surprise
"يع"	حشرت كل الكلمات في بعضه ملاحظ 🤣 إمريضة ، تفشلين ،يع ،تثقفي؟ " اي هبد وخلاص 😂😂	Equivalent to English interjection "yuk" which used for expressing disgust and unpleasing.
"آخ"	آآآخ منك يا لقاح الكورونا😩الخوف قطاع العصب😳	Equivalent to English interjection "ouch" which used to express sudden pain and discomfort.
"صه"	اقول صه 🙆 🙀 .	Equivalent to English interjection "shh" which used to give someone order to be quiet

speaker's emotive state and some of the used interjections are transliteration from the English interjections. Table 9 shows examples for interjections from the collected corpus. Interjections helped the communicators to communicate

nonverbally without the need of using long written messages, so instead of writing "It is a disgusting thing" they can write "Yuk." instead and still the same message delivered.

5. Conclusion

Text-based computer-mediated communication social media platforms like Twitter do not limit and restrict the use of nonverbal cues, on the contrary, Twitter users find creative ways to show and express the nonverbal component whether by using emojis to represent their facial expressions, punctuation, and character repetition to reflect the paralinguistic prosodic features or even by using vocal expressions as Interjections to express their emotional state. Talking specifically about Arabs it seems that there is a little difference in using emojis, but when it comes to gestures significant differences were found. The 😂 Face with Tears of Joy came in the first place across all dialects as the most used emoji. Egyptians seem to use hand gestures more than other dialects, Emirati and Lebanese dialects use them less frequently. The number of emojis per tweet is nearly the same across dialects with an average of 3 emojis per tweet. Arabs rarely use punctuation marks, and it was apparent that emojis are replacing them. The top used punctuation mark was the comma followed by the exclamation mark. Lebanese dialect uses punctuation marks with a higher percentage than other dialects and that is reflected in their higher number of characters per tweet. Character repetition was pervasive across all dialects which reflects intensifying emotions and shows words that are pronounced louder than other words. Arabs use interjections to reflect their emotional state instead of writing a full sentence, only a vocal expression is written.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Tarjani, S. (2017) Non-Verbal Communication: A Significant Aspect of Proficient Occupation. *International Journal of Humanities and Social Science*, **22**, 69-72.

[2] Darwin, C. (1872) The Expression of the Emotions in Man and Animals. John Muray, London. https://doi.org/10.1037/10001-000

[3] Ekman, P. and Friesen, W.V. (1969) The Repertoire or Nonverbal Behavior: Categories, Origins, Usage, and Coding. *Semiotica*, **1**, 49-98. https://doi.org/10.1515/semi.1969.1.1.49

[4] Hans, A. and Hans, E. (2015) Kinesics, Haptics and Proxemics: Aspects of Non-Verbal Communication. *IOSR Journal of Humanities and Social Science*, **20**, 47-52.

[5] Madadzhe, R.N. (2011) The Analysis of the Impact of Nonverbal Communication Xitsonga Discourse. M.A. Thesis, University of Limpopo, Mankweng.

[6] Lapakko, D. (2007) Communication is 93% Nonverbal: An Urban Legend Proliferates. *Communication and Theater Association of Minnesota Journal*, **34**, 7-19.

[7] Park, J.R. and El Mimouni, H. (2020) Emoticons and Non-Verbal Communications across Arabic, English, and Korean Tweets. *Global Knowledge, Memory and Communication*, **69**, 579-595. https://doi.org/10.1108/GKMC-02-2020-0021

[8] Tantawi, Y. and Rosson, M. (2019) The Paralinguistic Function of Emojis in Twitter Communication. *International Journal of Information and Communication Engineering*, **13**, 271-276.

[9] Álvarez, I. and Muñoz, A. (2012) An Insight into Twitter: A Corpus Based Contrastive Study in English and Spanish. *Revista de Lingüística y Lenguas Aplicadas*, **7**, 37-50. https://doi.org/10.4995/rlyla.2012.1123

[10] Hakami, S.A. (2017) The Importance of Understanding Emoji: An Investigative Study. *Research Topics in HCI*, 1-20.

[11] Onursoy, S. (2018) Using Emoji: Self-Presentation and Different Meaning Creation Approaches. *Galatasaray University İletişim -ş-imDergisi*, **6**, 353-372.

[12] James, A. (2017) Prosody and Paralanguage in Speech and the Social Media: The Vocal and Graphic Realisation of Affective Meaning. *Linguistica*, **57**, 137-149. https://doi.org/10.4312/linguistica.57.1.137-149

[13] Mahfouz, I. (2018) Word Shortening Strategies: Egyptian vs. Non-Egyptian English Tweets. *English Language and Literature Studies*, **8**, 27-38. https://doi.org/10.5539/ells.v8n3p27

[14] Jovanović, V. (2004) The Form, Position and Meaning of Interjections in English. *Facta Universitatis: Series Linguistics and Literature*, **3**, 17-28.

[15] Asset, A., Lect, Alia, B. and Abdulla, Z. (2009) The Meanings of Interjections in English and Arabic. *Dalam Journal of the College of Arts University of Basra*, **2009**, 89-107.