

Arabic Emoji Sentiment Lexicon (Arab-ESL): A Comparison between Arabic and European Emoji Sentiment Lexicons

Shatha Ali A. Hakami

University of Birmingham &
Jazan University
United Kingdom / Saudi Arabia
sahakami@jazanu.edu.sa

Robert Hendley

University of Birmingham,
School of Computer Science
United Kingdom
r.j.hendley@cs.bham.ac.uk

Phillip Smith

University of Birmingham
School of Computer Science
United Kingdom
p.smith.7@cs.bham.ac.uk

Abstract

Emoji (the popular digital pictograms) are sometimes seen as a new kind of artificial and universally usable and consistent writing code. In spite of their assumed universality, there is some evidence that the sense of an emoji, specifically in regard to sentiment, may change from language to language and culture to culture. This paper investigates whether contextual emoji sentiment analysis is consistent across Arabic and European languages. To conduct this investigation, we, first, created the Arabic emoji sentiment lexicon (Arab-ESL). Then, we exploited an existing European emoji sentiment lexicon to compare the sentiment conveyed in each of the two families of language and culture (Arabic and European). The results show that the pairwise correlation between the two lexicons is consistent for emoji that represent, for instance, hearts, facial expressions, and body language. However, for a subset of emoji (those that represent objects, nature, symbols, and some human activities), there are large differences in the sentiment conveyed. More interestingly, an extremely high level of inconsistency has been shown with food emoji.

1 Introduction

Textual digital communication has become habitual these days among Internet users. However, it has been argued that in such communication, the vital nonverbal cues are missing, which can potentially lead to ambiguity and misunderstanding (Kiesler et al., 1984). To address this issue, users tend to employ many surrogates, including non-standard punctuation, emoticons, and emoji.

Evans (2017) defined emoji as a form of developed punctuation (the way of encoding nonverbal prosody cues in writing systems) that supplements written language to facilitate the writers articulating their emotions in text-based communication. Practically, emoji are actual icons in Unicode Transformation Format (UTF) that are a successor to

emoticons (i.e., ASCII character based) with more sophisticated rendering. These icons can represent facial expressions, body language, food, animals, places, and natural objects like flowers and trees. It has been observed that including emoticons, as well as emoji, in text not only helps the receivers to infer some contextual information, but it also eases understanding of the expressed sentiment (Dresner and Herring, 2010; Skovholt et al., 2014).

Sentiment analysis, is the computational study of people’s opinions, sentiments, emotions, and attitudes. It is one of the most active research areas in natural language processing (NLP) and is also extensively studied in data mining, web mining, and text mining (Ambady et al., 2000; Chen Yuet Wei, 2012). Usually, it is a one dimensional measure from negative to positive and often it is quantized to just three values: negative, neutral or positive. Sentiment analysis has become an important tool in classifying and interpreting text.

As digital language is increasingly used across platforms (Ai et al., 2017; Danesi, 2016), emoji are studied as an integral aspect of sentiment analysis in various research domains. This includes marketing (Ge and Gretzel, 2018), law (Goldman, 2018), healthcare (Willoughby and Liu, 2018), food-related (Vidal et al., 2016) and addictive-substances-related contents (Tran et al., 2018).

The basis of many sentiment analysis approaches is the sentiment lexicon, with words and phrases classified as conveying positive or negative sentiments. Several general-purpose word-based and phrase-based lexicons of subjectivity and sentiment have been constructed. Emoji, likewise, can and should be viewed as an integral feature in the sentiment analysis process. However, most sentiment-analysis research has focused on emoji in languages other than Arabic and, consequently, most of the resources developed (such as emoji sentiment lexicons and emoji-specific corpora) are non-Arabic. Thus, an analysis of Arabic emoji sentiment us-

age, provided as a result of this study, would be a valuable resource for research in Arabic language analysis.

In this work, we seek to investigate the differences in sentiment interpretation of emoji expressed in informal texts between two cultures (Arabic and European), through the following:

- **Constructing** an Arabic sentiment lexicon of 1034 emoji, extracted from 144,196 tweets in existing Arabic datasets. The lexicon is made available for open-access for research use¹.
- **Comparing** the sentiment conveyed by a subset of emoji in the Arabic lexicon with the corresponding ones in a standard European emoji lexicon.

The rest of this paper is organized as follows. Section 2 reviews related work upon which we build; Section 3 presents the study’s design; Section 4 presents the results analysis and discussion. Finally, in Section 5 we draw conclusions from this work along with highlighting its limitations as well as some recommendations for future work.

2 Related Work

The usage patterns of emoji have been studied from several viewpoints such as relationships with specific topics (Zhao et al., 2018), seasons (Barbieri et al., 2018b), countries and regions (Ljubešić and Fišer, 2016) as well as languages (Cohn et al., 2018). Here, we review some of the existing research that analyses emoji across cultures, and as sentiment features.

2.1 Emoji Across Cultures

Emoji usage can be understood as “visible acts of meaning” (Miller et al., 2017). Visible acts of meaning, as defined by Bavelas and Chovil (2000), are analogically encoded symbols that are sensitive to a sender-receiver relationship, and that are fully integrated with the accompanying words. This involves the sender-receiver cultural background as an aspect that might affect emoji-text sentiment analysis.

For example, Gao and VanderLaan (2020) presented a study suggesting that Eastern and Western cultures are different in their use of mouth versus eye cues when interpreting emotions. According to

the study, the norm in Western cultures is to display the overt emotion while in Eastern cultures, the norm is to present more subtle emotion. Westerners interpret facial emotional expressions through the mouth region. Conversely, Eastern cultures focus more on the eyes. The researchers also found that such differences extend to written and paralinguistic signals such as emoji and, consequently, this has implications for digital communication.

From another perspective, Guntuku et al. (2019) compared emoji use across cultures in terms of frequency, context, and topic associations. This includes potential mapping of emoji use differences with previously identified cultural differences in users’ expression about diverse concepts such as death, money, emotions and family. They also investigated the relative correspondence of validated psycho-linguistic categories with Ekman’s emotions (Ekman, 1992). The study considered two Eastern countries (China and Japan) and three Western countries (USA, UK and Canada). Their analysis revealed recognizable normative and culture specific patterns. Also, it revealed the ways in which emoji can be used for cross-cultural communication. For example, the emoji categories of emotion ‘Anger’ (e.g., 😡, 😠, 😤, 😡) in the study was tied to the basic emotions of anger which had been found by Ekman (1992) to be universally expressed and recognized facially across cultures. By contrast, the study showed that the emoji for rice bowl (🍚) and ramen (🍜) dominated the East, while meat-related emoji (e.g., 🍗, 🍖) are the majority in the West, meaning that such emoji seemingly are cultural-specific.

There are few studies of the use of emoji in Arabic culture. An Arabic socio-linguistic investigation on the use of emoji was applied to Omani, WhatsApp textual contents by Al Rashdi (2018). This work performed a qualitative analysis of selected texts with emoji using theories and methods of interactional sociolinguistics. The results of the work demonstrated that Arab users (mainly Omanis) utilize emoji not only as indicators of their emotions, but also as what Gumperz (1982) calls “contextualization cues”. In line with Gumperz’s theory, the study showed that emoji use in Arabic culture is as indicators of celebration, other message approvals, or a signal of task fulfillment. Also, Arabs use emoji as a response to thanks/compliment expressions; linking devices, and openings/closings of conversations.

¹<https://github.com/ShathaHakami/Arabic-Emoji-Sentiment-Lexicon-Version-1.0>

Another recent investigation of the phenomenon of emoji as a sentiment indicator within Arabic text was done by [Hakami et al. \(2020\)](#). In this work, the researchers undertook an empirical sentiment analysis along with a “Coding and Counting” approach ([Herring et al., 2004](#)). They concluded that an emoji in Arabic context, and perhaps in other cultural contexts as well, can be a true sentiment indicator, a multi-sentiment indicator, an ambiguous sentiment indicator, or a No-sentiment indicator.

2.2 Emoji as Sentiment Features

Many researchers have dedicated their efforts to exploiting emoji as a textual feature for sentiment analysis. One of the key elements of the use of emoji in sentiment analysis is the emoji sentiment lexicon. These have been constructed for different languages. Emoji Sentiment Ranking (ESR) is the first emoji sentiment lexicon with 751 emoji ([Kralj Novak et al., 2015](#)). The lexicon was constructed from over 70,000 tweets in 13 European languages, including English, Spanish, Polish, Russian, a union of tweets in Serbian, Croatian and Bosnian; Hungarian, German, Swedish, Slovak, Slovenian, Portuguese, Bulgarian and Albanian. This corpus was annotated for sentiment by 83 human annotators, each of whom was a native speaker of at least one of the languages. The sentiment of the emoji was computed from the sentiment of the tweets in which they occur and reflects the actual use of emoji in a context. The researchers of the work observed no significant differences in the emoji sentiment rankings between the 13 languages. Consequently, they considered the constructed lexicon a generic European language-independent resource for automated sentiment analysis. We use this lexicon for our investigation, and we refer to as European Emoji Sentiment Lexicon (Euro-ESL).

The number of emoji included in the European lexicon is smaller than that of the current set of emoji. Expanding the lexicon manually requires time and effort to reconstruct the labeled dataset. This encouraged [Kimura and Katsurai \(2017\)](#) to present a simple and efficient method for automatically constructing an emoji sentiment lexicon with arbitrary sentiment categories. The method extracted sentiment words from WordNet-Affect ([Strapparava et al., 2004](#)) and calculates the co-occurrence frequency between the sentiment words and each emoji. Based on the ratio of the number of occurrences of each emoji among the sentiment

categories, each emoji is assigned a multidimensional vector whose elements indicate the strength of the corresponding sentiment. In experiments conducted on a collection of tweets, they were able to show a high correlation between the European lexicon and their lexicon for three sentiment categories.

Compared to other languages, even less research has been done on emoji for sentiment in the Arabic language. Nonetheless, some significant recent work does exist. One of the earliest works on analysing Arabic emoji usage for emotional content was done by [Hussien et al. \(2016\)](#). This study addressed the problem of emotion detection in Arabic tweets using emoji. The aim of the work was to show that training a sentiment classifier on an automatically annotated tweet (using emoji) provides more accurate results than training the same classifier on a manually annotated tweet.

Other work aims at exploring the impact of combining emoji based features with various forms of textual features on the sentiment classification of dialectical Arabic tweets ([Al-Azani and El-Alfy, 2018a](#)). The study concluded that simpler models can be constructed with much better results when emoji are merged with a Word2vec embedding model and the selection of the most relevant subset of features as input to the classifier. Different work by the same researchers adopted the Euro-ESL lexicon, mentioned earlier, to evaluate emoji as nonverbal features for sentiment analysis in Arabic texts ([Al-Azani and El-Alfy, 2018b](#)). Several machine learning algorithms were evaluated on the suggested features. The experimental results demonstrated that emoji-based features alone can be a very effective means for detecting sentiment polarity.

Similarly, [Abdellaoui and Zrigui \(2018\)](#) used ten subjective emoji (from the Euro-ESL) along with the Arabic word sentiment lexicon Ar-SeLn ([Badaro et al., 2014](#)) to construct and annotate a large-scale dataset for Arabic sentiment analysis. Their process used a dataset of 6 million Arabic tweets with a vocabulary of 602,721 distinct entities. They named their dataset TEAD and released a subset of it for public use.

The most recent Arabic study for emotion analysis of textual content with emoji has been done by [Hussien et al. \(2019\)](#). The study proposed a distant supervised learning approach where the training sentences are automatically annotated based on the

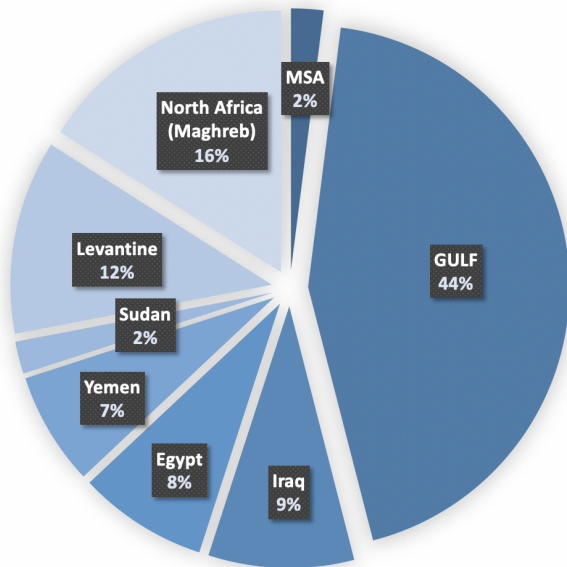


Figure 1: A regional-based distribution of the Arabic dialects in the collected dataset. This distribution includes the modern standard Arabic (MSA).

emoji they contain. The study’s authors experimentally showed that training classifiers on cheap, large and possibly erroneous data annotated using their approach leads to more accurate results compared with training the same classifiers on the more expensive, much smaller and error-free manually annotated training data.

3 Study Design

Kralj Novak et al. (2015) conducted a study, which considered context-sensitivity when analyzing the sentiment of emoji and texts, described above (i.e., in subsection 2.2). We followed their approach, with some additional steps, to construct our Arabic emoji sentiment lexicon (Arab-ESL) as follows:

3.1 Data Collection

Before we collected our data, some criteria were defined. Our consideration was on data that is from a social media platform, written in the Arabic language, multi-dialect, multi-aspect and, more importantly, containing emoji. With such criteria, collecting, cleaning and preparing a great deal of raw data for sentiment annotation in a short time is impossible. Therefore, we targeted 14 different public datasets of Arabic social media (all are tweets from the Twitter platform). These contain 144,196 tweets that meet our criteria. This data collection procedure resulted in a dataset that is diverse in its Arabic dialects. They are from all of the

Arabic regions: Levantine (Syria, Lebanon, Jordan, and Palestine), Iraq, Gulf (Saudi Arabia, Kuwait, Qatar, Bahrain, UAE, and Oman), Yemen, Egypt, Sudan, and Maghrib (Libya, Algeria, Tunisia, and Morocco). CAMEL Tools (Obeid et al., 2020) were used to detect and classify these dialects in our dataset based on their Arabic regions as illustrated in Figure 1.

Some of the targeted datasets were constructed for sentiment classification tasks or other related tasks, such as hate speech detection. Therefore, these datasets have been manually annotated with sentiment labels, emotional labels, or other related labels like *hate* and *offensive*. For some other datasets (i.e., AT SAD (Train & Test), Arabic Twitter Data for Sentiment, and TEAD), they were available for public use with automatically predicted sentiment labels. This meant that we had to re-annotate these datasets with sentiment labels after testing their reliability. The details of this procedure are explained in the following subsection. Table 1 summarises all the details of the considered tweets from the targeted datasets.

3.2 Data Preparation

We applied our data preparations based on the work in (El-Beltagy et al., 2017). For general preparation, we removed URLs, all non-emoji symbols (including #, @, and emoticons), punctuation marks, numbers and repeated letters. For Arabic text specific preparation, we unified the letters that appear in different forms, and removed diacritics and unknown characters (e.g., Quranic symbols).

In addition, we normalized all the Emoji ZWJ sequences into their base emoji. An emoji ZWJ sequence is a combination of multiple emoji which display as a single emoji on supporting platforms. These sequences are joined with a Zero Width Joiner (ZWJ) character, which is an invisible Unicode character that joins two or more other characters together in sequence to create a new emoji. For example, the *Woman with Veil: Medium Skin Tone* emoji (👩🏫) is a ZWJ sequence combining *Person With Veil* (👩), *Medium Skin Tone* (🏫), *Zero Width Joiner* (invisible) and *Female Sign* (♀). For such cases, we considered the base emoji only, which is the *Person With Veil* (👩).

In general, this initial data preparation procedure aimed to normalise the data into a coherent form so that it can be handled easily for sentiment analysis by both human and machine.

Dataset Name / Sentiment Annotation Process	Dataset Reference	# of Tweets Containing Emoji
ATSAD (Train & Test) / (Re-annotated) Automatic	(Abu Kwaik et al., 2020)	77,211
Arabic Twitter Data for Sentiment / (Re-annotated) Automatic	(Fathi, 2019)	42,832
TEAD / (Re-annotated) Automatic	(Abdellaoui and Zrigui, 2018)	11,950
ArSAS / Manual	(Elmadany AbdelRahim and Magdy, 2018)	4,266
ATSAD (Gold) / Manual	(Abu Kwaik et al., 2020)	3,775
SS2030 / Manual	(Alyami and Olatunji, 2020)	1,061
SemEval-2018 (Task1 & Task2) / Manual	(Mohammad et al., 2018) (Barbieri et al., 2018a)	594
ArSenTD-Lev / Manual	(Baly et al., 2019)	389
SemEval-2017 (Task4) / Manual	(Rosenthal et al., 2017)	263
Syria Tweets / Manual	(Salameh et al., 2015)	64
L-HSAB / Manual	(Mulki et al., 2019)	23
Arabic Floods Detection / (Re-annotated) Manual	(Alharbi and Lee, 2019)	768
DART / (Re-annotated) Manual	(Alsarsour et al., 2018)	470
ArSAS / (Re-annotated) Manual	(Elmadany AbdelRahim and Magdy, 2018)	285
Arabic-Tweets-vs-Dialects / (Re-annotated) Manual	(Abdelaal, 2018)	129
L-HSAB / (Re-annotated) Manual	(Mulki et al., 2019)	42
Total		144,196

Table 1: Summary of the datasets used.

3.3 Manual Sentiment Annotation

As mentioned above, most of the targeted datasets (10 out of 13 datasets) were previously manually annotated with inconsistent labels. Therefore, unifying the annotation labels was necessary, and applied as follows.

Syria Tweets, SemEval-2017 (Task4), ArSenTD-Lev, ArSAS, SS2030, and ATSAD (Gold) datasets were constructed for a sentiment analysis task. In this case, we only unified their labels to be *negative*, *neutral*, or *positive*. The ArSAS dataset had a group of tweets labelled as *Mixed* which we kept for later manual re-annotation with sentiments.

On the other hand, labels in SemEval-2018 and L-HSAB datasets were not explicitly for sentiment. The L-HSAB dataset is intended to be for an offensive/hate-speech detection task. In this dataset, we replaced the labels *Abusive* and *Hate* with the label *negative*. However, we kept the tweets that were labeled as *Normal* for later manual

re-annotation with sentiment. In SemEval-2018, the labels were for the emotions rather than sentiments. Thus, the negative labels (i.e., *Anger*, *Fear*, and *Sadness*) were replaced by the label *negative*; and the positive label (i.e., *Joy*) was replaced by the word *positive*.

Finally, the DART, Arabic-Tweets-vs-Dialects and Arabic Floods Detection datasets were created for natural language processing tasks other than sentiment analysis. The DART and Arabic-Tweets-vs-Dialects datasets were for an Arabic-dialects specification task while the Arabic Floods Detection dataset was for a disaster detection task. Along with the tweets in the ArSAS dataset that were labelled as *Mixed* and in the L-HSAB that were labelled as *Normal*; we formed a collection of tweets including Arabic Floods Detection, DART and Arabic-Tweets-vs-Dialects datasets to be manually re-annotated with sentiment. This annotation was done, independently, by four native Arabic

speaking annotators, two males and two females. Table 1 specifies the number of tweets from each dataset that have been re-annotated.

3.4 Machine Sentiment Annotation

Sentiment annotation by a machine needs some pre-processing steps. In our case, besides the dataset preparation mentioned above, the following steps were applied.

First, all of the emoji in the collected dataset were extracted, and all of their official Unicode English names were found. We used the Python library *emoji*. Then, the emoji’s English names were translated into the Arabic language in two steps: using the *Google Translate* API for Python, followed by manual Arabic-English translation checking by a linguist. Lastly, each emoji symbol within a text was replaced with its Arabic name. This was to ensure that the machine will not neglect any token (i.e., Arabic word) in the provided textual context and that it will classify its sentiment appropriately.

For automatic annotation of our Arabic textual dataset with sentiment, two recently released Arabic machine sentiment annotators were tested: Mazajak and CAMEL Tools.

Mazajak (Abu Farha and Magdy, 2019) is the first online Arabic sentiment analyser, it is based on a deep learning model built on a convolutional neural network (CNN) followed by a long short-term memory (LSTM). It achieves state-of-the-art results on many Arabic dialect datasets including SemEval 2017 and ASTD. This analyser provides different functionalities for Arabic sentiment analysis including two modes for raw text processing: the batch mode and the online API. We used the API mode.

CAMEL Tools (Obeid et al., 2020), on the other hand, is a collection of open-source tools for Arabic NLP in Python. It provides utilities for many NLP tasks, including dialect identification and sentiment analysis. The sentiment analyser in this system was built using large pre-trained language models (i.e., HuggingFace’s Transformers, mBERT, and AraBERT); and various Arabic dialectic datasets for fine-tuning and evaluation.

3.5 Validity and Reliability Tests for Sentiment Annotations

Regarding human annotation, we assumed that all of the tweets extracted from the already-existing datasets were reliably annotated. However, as described in subsection 3.3 above, it was necessary to

re-annotate a subset of that data that was not labeled with sentiment. For that, we used the majority-voting approach between four human annotators (two males, and two females) as the procedure for this labelling reliability assurance. Two annotators, at least, had to be in agreement on one sentiment label for a tweet. In cases where two annotators disagreed on a specific sentiment, the annotation from a third annotator was considered to determine the decision.

To test the reliability of the machine annotation, we first sampled a random set of tweets which were already manually annotated. Then we used the accuracy metric between the manual and the automatic annotations in two ways. The first accuracy test was conducted between the labels from the manually annotated tweets and the labels of the same tweets annotated by Mazajak. The test resulted in an *accuracy* = 0.71. The second test was between the labels of the same manually annotated tweets and their sentiment annotation results from CAMEL Tools. This test resulted in an *accuracy* = 0.74. Although the two tests results are close, CAMEL Tools out-performed Mazajak by 0.03. Therefore, we used the sentiment labels assigned by CAMEL Tools.

3.6 Arabic Emoji Sentiment Lexicon (Arab-ESL) Construction

We applied the Emoji Sentiment Ranking (ESR) model, that was proposed for Euro-ESL by (Kralj Novak et al., 2015), to construct an Arabic emoji sentiment lexicon (Arab-ESL) using our prepared data. In this model, the sentiment of the emoji was computed from the sentiment of texts labelled by human annotators. Sentiment labels take one of three values: negative, neutral and positive. Each sentiment label, c , is a discrete three-valued variable:

$$c \in \{-1, 0, +1\}$$

This variable represents the order of the sentiment values and the distances between them. An emoji could occur in several tweets each of which are labelled with sentiment. A discrete distribution:

$$N(c), \sum N(c) = N, c \in \{-1, 0, +1\}$$

captures the sentiment distribution for the set of relevant tweets. N denotes the number of all the occurrences of the object in the tweets, and $N(c)$ are the occurrences in tweets with the sentiment label c . We considered the multiple occurrence of

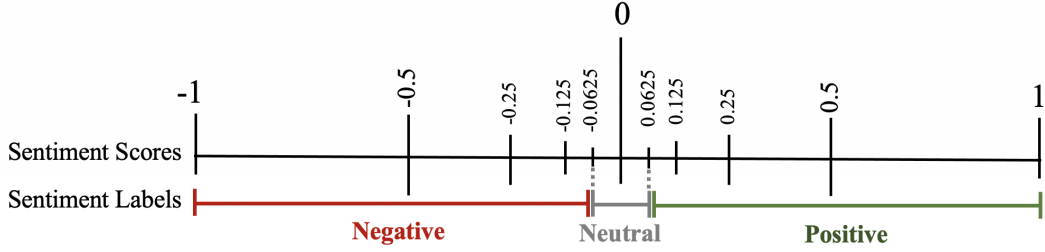


Figure 2: A scale for determine sentiment labels from sentiment scores.

Emoji	Emoji Name	Emoji Class	Occurrence (N)	Sentiment Score (\bar{S})	Sentiment Label (L)
😂	Tears of Joy	Facial Expression	25,908	0.272426	positive
💔	Broken Heart	Heart	18,564	-0.934066	negative
❤️	Red Heart	Heart	15,876	0.560658	positive
😭	Loudly Crying Face	Facial Expression	12,318	-0.677789	negative
😊	Smiling Face with Heart-Eyes	Facial Expression	6,815	0.869552	positive
🌹	Rose	Nature	6,173	0.766402	positive
😊	Slightly Smiling Face	Facial Expression	5,717	0.226692	positive
💕	Two Hearts	Heart	5,573	0.487888	positive
😄	Grinning Face with Sweat	Facial Expression	5,499	0.533552	positive
💙	Blue Heart	Heart	5,444	0.616091	positive

Table 2: The top ten emoji in the Arabic Emoji Sentiment Lexicon (Arab-ESL).

an emoji in a single tweet. From the above we formed a discrete probability distribution:

$$(p_-, p_0, p_+), \sum_c P(c) = 1$$

The components of the distribution (i.e., p_- , p_0 , and p_+) denote the negativity, neutrality, and positivity of the emoji, respectively. Then, we estimated the probabilities from relative frequencies:

$$P(c) = \frac{N(c)}{N}$$

This estimation gives a good approximation for large samples. However, we used the Laplace estimate (Good, 1965) when $N \leq 900$, as it is recommended when estimating the probability of small samples.

Then, the sentiment score \bar{S} of the emoji was computed as the mean of the distribution:

$$\bar{S} = (-1 \cdot P(-)) + (0 \cdot P(0)) + (+1 \cdot P(+))$$

Furthermore, finding sentiment labels of emoji in both lexicons helps test the agreement between the two, which we will explain later. Therefore, three

scaled-groups of sentiment scores have been classified under three sentiment labels L . Figure 2 demonstrates the scale that we have used to determine the thresholds in the sentiment scores when determining the sentiment labels in our lexicon. Emoji with sentiment score i , where $-1 \leq i < -0.0625$, was classified as negative. On the other hand, emoji with sentiment score i , where $1 \geq i > 0.0625$, was classified as positive. Lastly, an emoji was classified as neutral when its sentiment score i was in the range where $-0.0625 \leq i \leq 0.0625$. Part of the resulting emoji sentiment lexicon (10 out of 1034 emoji) is shown in Table 2.

3.7 Correlation and Agreement Tests between Euro-ESL and Arab-ESL

Because of the richness and diversity of the emoji, it is difficult to hypothesize, a priori, how specific emoji may differ between cultures. Therefore, we aim to make this investigation depend on the sentiment scores' correlation rather than causality. For that, a correlation measurement between the Euro-ESL and the Arab-ESL was utilized. The metric was the Pearson correlation coefficient and the focus was on the sentiment scores of 479 common emoji in the two lexicons.

Emoji Category	# of Emoji	Pearson	<i>p</i> -value	Cohen’s κ
Object	106	0.414	0.101	0.1092
Nature	87	0.391	0.000	0.1430
Facial Expression	79	0.747	0.269	0.4379
Symbol	79	0.275	0.254	0.0581
Food	46	-0.159	0.289	0.0160
Body Language	38	0.583	0.000	0.2438
Human Activity	19	0.275	0.254	-0.0556
Heart	16	0.834	0.584	1.0000

Table 3: Correlation and agreement tests results of the common emoji in Arab-ESL and Euro-ESL under eight different categories. The green color indicates the highest correlation/agreement and the red color the lowest.






Heart Emoji		Arab-ESL		Euro-ESL	
Emoji	Name	Score	Label	Score	Label
	Broken Heart	-0.934066	negative	-0.120846	negative
	Red Heart	0.560658	positive	0.746087	positive
	Two Hearts	0.487888	positive	0.632917	positive
	Heart with Arrow	0.240646	positive	0.683417	positive
	Sparkling Heart	0.862897	positive	0.713381	positive

Table 4: Example heart emoji sentiment scores and labels in both Arab-ESL and Euro-ESL.

Specifically, we, first, extracted the common emoji between the two lexicons, where their frequency N is greater than or equal to 5. This resulted in 479 emoji in total. Then, we sorted the emoji in the Arab-ESL based on their frequency N in descending order. Accordingly, we sorted the corresponding emoji in the Euro-ESL based on their order in the Arab-ESL. This sorting process eased testing the correlation of the same emoji in the two lexicons. Moreover, we applied this test on eight emoji categories, independently, using the correlation measure in each. The emoji categories are: *Facial Expressions*, *Body Language*, *Human Activity*, *Hearts*, *Nature*, *Food*, *Object* and *Symbol*. Also, we disregarded two categories, *Place* (with seven emoji) and *Flag* (with two emoji), due to the small numbers of emoji. The test’s results for each emoji category are tabulated in Table 3.

Agreement between the labels refers to the degree of concordance between the two sets of labels. The statistical Cohen’s Kappa agreement tests (McHugh, 2012) was used to assess the variability of emoji sentiment labels in Euro-ESL and Arab-ESL. Again, we applied this agreement test on the eight emoji categories that were mentioned in the correlation testing process. The results of this agreement test under each emoji category are shown in Table 3.

4 Results Analysis and Discussion

Here, we analyze our data patterns from an abductive perspective (Haig, 2018) to form preliminary cultural emoji sentiment theories between European and Arabic languages. As is shown in Table 3, across the two lexicons, the Pearson correlation for *Heart* and *Facial Expression* emoji categories are 0.834, and 0.747, respectively. Likewise, the sentiment labels agreement tests for the *Heart* and *Facial Expression emoji* are the highest (i.e., almost perfect for *Heart* and moderate for *Facial Expression*) among other emoji categories. This indicates a strong consistency and detects a normal pattern in these types of emoji that are common across the two cultures.

Indeed, many sociological theories, proposing the universality of basic emotions (Ekman, 1992), point out the tendency of emotional emoji to show high levels of similarity even between distinct cultures, like West and East (Guntuku et al., 2019). Thus, emoji categories of *Facial Expression*, *Heart*, and *Body Language* would be expected to be more convergent compared to other categories. Table 4 exemplifies such matches between the two cultures for sentiment scores and labels, for heart emoji.

On the other hand, the lowest correlations in emoji sentiment between the two lexicons occur in the *Food* (-0.159), *Symbol* (0.275), and *Human*






Food Emoji		Arab-ESL		Euro-ESL	
Emoji	Name	Score	Label	Score	Label
	Birthday Cake	0.733333	positive	0.612745	positive
	Honey Pot	0.212766	positive	0.045455	neutral
	Fish Cake with Swirl	0.575758	positive	-0.593750	negative
	Beer Mug	-0.206897	negative	0.492537	positive
	Baby Bottle	-0.058824	neutral	0.454545	positive

Table 5: Example food emoji sentiment scores and labels in both Arab-ESL and Euro-ESL.

Activities (0.275) categories. Sentiment label agreement tests showed similar results with the same emoji categories. Also, the Cohen’s Kappa agreement coefficient resulted in $\kappa = 0.0160$ for *Food* emoji, $\kappa = 0.0581$ for *Symbol* emoji, and $\kappa = -0.0556$ for *Human Activity* emoji.

In fact, this is not surprising. Cultures are often instantiated in cuisines representing identities, religions, and dietary preferences (Van Gelder, 2014; Jallad, 2008; Van den Berghe, 1984). Besides, symbols are often representative figurative characters of many specific cultural values (Aaker et al., 2001). Moreover, many human activities such as praying, working, or playing are affected by culture (Smith et al., 2013). As an example, Table 5 illustrates the differences in emoji sentiment scores and labels under the *Food* category between Arab-ESL and Euro-ESL.

5 Conclusion, Limitations and Future Work

Investigating emoji sentiment indications between cultures is one of the primary applications of an emoji sentiment lexicon. In this work, we exploited most of the existing Arabic resources (covering all Arabic dialects) to contribute an Arabic emoji sentiment lexicon (Arab-ESL). As an application of the constructed lexicon, we compared the Arab-ESL with the already-existing similar European one. The results shows that cultures might share similar emoji sentiment indications. This occurs with emoji that represent common human behaviours, such as facial expressions, and body language, or basic emotions such as love and sadness. However, there are other emoji where their sentiments might be affected by a cultural-specific aspect, such as food, symbols, and human activities. These conclusions can be drawn from the results of employing two statistical methods: correlation coefficient and agreement coefficient.

We are aware that there are variations, both geographically and *within* a specific culture (e.g., Arabic or European) that makes it very difficult to completely rely on the results of such an investigation. However, we intend in the future to make the resulting emoji sentiment lexicon more fine grain for further, focused and detailed analytical studies of emoji within the Arabic language.

Another limitation is that the sentiment dimensions are more than just a one dimensional-scale (i.e., negative to positive), and that this should be explored in the future. The emoji’s eloquence allows them to be assigned more fine-grained sentiment labels like very negative and slightly negative, or more detailed emotional labels like anger, happiness, and sadness. In the future, this might help for an additional structuring of the emoji that can be obtained from correlations between their sentiments or emotions. For example, deriving different forms of facial expressions expressing happiness.

Also, the tests on sentiment classification were only based on simple statistics of sentiment scores of emoji in texts. Future work can combine emoji sentiment scores within Arabic semantics for more advanced sentiment analysis. Besides, an emoji’s textual context is crucial in determining the role of the emoji as a modifier of the meaning. More investigation should be applied with Arabic content.

Finally, considering the promise of emoji in NLP application tasks, studies should discover the contribution of emoji in multi-modal and cross-lingual sentiment analysis along with transfer learning and deep learning tasks. Moreover, it will be interesting to observe how the use of emoji by Arabic users is growing, and whether their textual communication is increasingly being replaced by this pictorial language. Also, the correlation between sentiment and meaning of emoji evolves over time. It might be important to explore the change in the meaning of controversial emoji, and how they are affected by the corresponding social processes.

References

- Jennifer Lynn Aaker, Veronica Benet-Martinez, and Jordi Garolera. 2001. Consumption symbols as carriers of culture: A study of Japanese and Spanish brand personality constructs. *Journal of personality and social psychology*, 81(3):492.
- Ahmad Abdelaal. 2018. Arabic-Tweets-vs-Dialects. <https://www.kaggle.com/ahmed9914/arabic-tweets-vs-dialects>. Online; accessed 24 October 2020.
- Houssein Abdellaoui and Mounir Zrigui. 2018. Using Tweets and Emojis to Build TEAD: an Arabic Dataset for Sentiment Analysis. *Computación y Sistemas*, 22:777 – 786.
- Ibrahim Abu Farha and Walid Magdy. 2019. Mazajak: An online Arabic sentiment analyser. In *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, pages 192–198, Florence, Italy. Association for Computational Linguistics.
- Kathrein Abu Kwaik, Stergios Chatzikyriakidis, Simon Dobnik, Motaz Saad, and Richard Johansson. 2020. An Arabic tweets sentiment analysis dataset (AT-SAD) using distant supervision and self training. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, pages 1–8, Marseille, France. European Language Resource Association.
- Wei Ai, Xuan Lu, Xuanzhe Liu, Ning Wang, Gang Huang, and Qiaozhu Mei. 2017. Untangling emoji popularity through semantic embeddings. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11.
- Sadam Al-Azani and El-Sayed M El-Alfy. 2018a. Combining emojis with Arabic textual features for sentiment classification. In *2018 9th International Conference on Information and Communication Systems (ICICS)*, pages 139–144. IEEE.
- Sadam Al-Azani and El-Sayed M El-Alfy. 2018b. Emoji-based sentiment analysis of Arabic microblogs using machine learning. In *2018 21st Saudi Computer Society National Computer Conference (NCC)*, pages 1–6. IEEE.
- Fathiya Al Rashdi. 2018. Functions of emojis in WhatsApp interaction among Omanis. *Discourse, Context & Media*, 26:117–126.
- Alaa Alharbi and Mark Lee. 2019. Crisis detection from Arabic tweets. In *Proceedings of the 3rd Workshop on Arabic Corpus Linguistics*, pages 72–79.
- Israa Alsarsour, Esraa Mohamed, Reem Suwaileh, and Tamer Elsayed. 2018. DART: A large dataset of dialectal Arabic tweets. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Sarah N. Alyami and Sunday O. Olatunji. 2020. Application of support vector machine for Arabic sentiment classification using Twitter-based dataset. *Journal of Information & Knowledge Management*, 19(01):2040018.
- Nalini Ambady, Frank J. Bernieri, and Jennifer A. Richeson. 2000. Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In *Advances in experimental social psychology*, volume 32, pages 201–271. Elsevier.
- Gilbert Badaro, Ramy Baly, Hazem Hajj, Nizar Habash, and Wassim El-Hajj. 2014. A large scale Arabic sentiment lexicon for Arabic opinion mining. In *Proceedings of the EMNLP 2014 workshop on Arabic natural language processing (ANLP)*, pages 165–173.
- Ramy Baly, Alaa Khaddaj, Hazem M. Hajj, Wassim El-Hajj, and Khaled Bashir Shaban. 2019. Arsentdlev: A multi-topic corpus for target-based sentiment analysis in Arabic Levantine tweets. *CoRR*, abs/1906.01830.
- Francesco Barbieri, Jose Camacho-Collados, Francesco Ronzano, Luis Espinosa-Anke, Miguel Ballesteros, Valerio Basile, Viviana Patti, and Horacio Saggion. 2018a. SemEval 2018 task 2: Multilingual emoji prediction. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 24–33, New Orleans, Louisiana. Association for Computational Linguistics.
- Francesco Barbieri, Luis Marujo, Pradeep Karuturi, William Brendel, and Horacio Saggion. 2018b. Exploring emoji usage and prediction through a temporal variation lens. *arXiv preprint arXiv:1805.00731*.
- Janet Beavin Bavelas and Nicole Chovil. 2000. Visible acts of meaning: An integrated message model of language in face-to-face dialogue. *Journal of Language and Social Psychology*, 19(2):163–194.
- Pierre L. Van den Berghe. 1984. Ethnic cuisine: Culture in nature. *Ethnic and Racial Studies*, 7(3):387–397.
- Amanda Chen Yuet Wei. 2012. *Emoticons and the non-verbal communication: With reference to Facebook*. Ph.D. thesis, Christ University.
- Neil Cohn, Tim Roijackers, Robin Schaap, and Jan Engelen. 2018. Are emojis a poor substitute for words? sentence processing with emoji substitutions. In *CogSci*.
- Marcel Danesi. 2016. *The semiotics of emoji: The rise of visual language in the age of the internet*. Bloomsbury Publishing.
- Eli Dresner and Susan C. Herring. 2010. Functions of the nonverbal in cmc: Emoticons and illocutionary force. *Communication theory*, 20(3):249–268.
- Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.

- Samhaa R El-Beltagy, Mona El Kalamawy, and Abu Bakr Soliman. 2017. Niletmrg at semeval-2017 task 4: Arabic sentiment analysis. *arXiv preprint arXiv:1710.08458*.
- Hamdy Mubarak Elmadany AbdelRahim and Walid Magdy. 2018. Arsas: An arabic speech-act and sentiment corpus of tweets. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France. European Language Resources Association (ELRA).
- Vyvyan Evans. 2017. *The emoji code: The linguistics behind smiley faces and scaredy cats*. Picador USA.
- Mohammed Fathi. 2019. Arabic Twitter Data for Sentiment. <https://www.kaggle.com/muhammedfathi/arabic-twitter-data-for-sentiment>. Online; accessed 24 October 2020.
- Boting Gao and Doug P VanderLaan. 2020. Cultural influences on perceptions of emotions depicted in emojis. *Cyberpsychology, Behavior, and Social Networking*, 23(8):567–570.
- Jing Ge and Ulrike Gretzel. 2018. Emoji rhetoric: a social media influencer perspective. *Journal of Marketing Management*, 34(15-16):1272–1295.
- Eric Goldman. 2018. Emojis and the law. *Wash. L. Rev.*, 93:1227.
- Irving John Good. 1965. The estimation of probabilities: An essay on modern bayesian methods, vol. 30.
- John J Gumperz. 1982. *Discourse strategies*. 1. Cambridge University Press.
- Sharath Chandra Guntuku, Mingyang Li, Louis Tay, and Lyle H Ungar. 2019. Studying cultural differences in emoji usage across the east and the west. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 226–235.
- Brian D Haig. 2018. An abductive theory of scientific method. In *Method matters in psychology*, pages 35–64. Springer.
- Shatha Hakami, Robert Hendley, and Phillip Smith. 2020. Emoji as sentiment indicators: An investigative case study in arabic text. In *The Sixth International Conference on Human and Social Analytics*, pages 26–32. IARIA.
- Susan C Herring, S Barab, R Kling, and J Gray. 2004. An approach to researching online behavior. *Designing for virtual communities in the service of learning*, 338:338–376.
- Wegdan Hussien, Mahmoud Al-Ayyoub, Yahya Tash-toush, and Mohammed Al-Kabi. 2019. On the use of emojis to train emotion classifiers. *arXiv preprint arXiv:1902.08906*.
- Wegdan A Hussien, Yahya M Tashtoush, Mahmoud Al-Ayyoub, and Mohammed N Al-Kabi. 2016. Are emoticons good enough to train emotion classifiers of arabic tweets? In *2016 7th International Conference on Computer Science and Information Technology (CSIT)*, pages 1–6. IEEE.
- Nader Al Jallad. 2008. The concepts of al-halal and al-haram in the arab-muslim culture: a translational and lexicographical study. *Language design: journal of theoretical and experimental linguistics*, 10:077–86.
- Sara Kiesler, Jane Siegel, and Timothy W McGuire. 1984. Social psychological aspects of computer-mediated communication. *American psychologist*, 39(10):1123.
- Mayu Kimura and Marie Katsurai. 2017. Automatic construction of an emoji sentiment lexicon. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pages 1033–1036.
- Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. 2015. Sentiment of emojis. *PLoS one*, 10(12):e0144296.
- Nikola Ljubešić and Darja Fišer. 2016. A global analysis of emoji usage. In *Proceedings of the 10th Web as Corpus Workshop*, pages 82–89.
- Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, 22(3):276–282.
- Hannah Miller, Daniel Kluver, Jacob Thebault-Spieker, Loren Terveen, and Brent Hecht. 2017. Understanding emoji ambiguity in context: The role of text in emoji-related miscommunication. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11.
- Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. *SemEval-2018 task 1: Affect in tweets*. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 1–17, New Orleans, Louisiana. Association for Computational Linguistics.
- Hala Mulki, Hatem Haddad, Chedi Bechikh Ali, and Halima Alshabani. 2019. *L-HSAB: A Levantine Twitter dataset for hate speech and abusive language*. In *Proceedings of the Third Workshop on Abusive Language Online*, pages 111–118, Florence, Italy. Association for Computational Linguistics.
- Ossama Obeid, Nasser Zalmout, Salam Khalifa, Dima Taji, Mai Oudah, Bashar Alhafni, Go Inoue, Fadhl Eryani, Alexander Erdmann, and Nizar Habash. 2020. Camel tools: An open source python toolkit for arabic natural language processing. In *Proceedings of the 12th language resources and evaluation conference*, pages 7022–7032.

- Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. [SemEval-2017 task 4: Sentiment analysis in Twitter](#). In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 502–518, Vancouver, Canada. Association for Computational Linguistics.
- Mohammad Salameh, Saif Mohammad, and Svetlana Kiritchenko. 2015. [Sentiment after translation: A case-study on Arabic social media posts](#). In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 767–777, Denver, Colorado. Association for Computational Linguistics.
- Karianne Skovholt, Anette Grønning, and Anne Kankaanranta. 2014. The communicative functions of emoticons in workplace e-mails:-. *Journal of Computer-Mediated Communication*, 19(4):780–797.
- Peter B Smith, Ronald Fischer, Vivian L Vignoles, and Michael Harris Bond. 2013. *Understanding social psychology across cultures: Engaging with others in a changing world*. Sage.
- Carlo Strapparava, Alessandro Valitutti, et al. 2004. Wordnet affect: an affective extension of wordnet. In *Lrec*, volume 4, page 40. Citeseer.
- Tuan Tran, Dong Nguyen, Anh Nguyen, and Erik Golenz. 2018. Sentiment analysis of emoji-based reactions on marijuana-related topical posts on facebook. In *IEEE International Conference on Communications (ICC)*.
- Geert Jan Van Gelder. 2014. *Of Dishes and Discourse: Classical Arabic Literary Representations of Food*. Routledge.
- Leticia Vidal, Gastón Ares, and Sara R Jaeger. 2016. Use of emoticon and emoji in tweets for food-related emotional expression. *Food Quality and Preference*, 49:119–128.
- Jessica Fitts Willoughby and Shuang Liu. 2018. Do pictures help tell the story? an experimental test of narrative and emojis in a health text message intervention. *Computers in Human Behavior*, 79:75–82.
- Peijun Zhao, Jia Jia, Yongsheng An, Jie Liang, Lexing Xie, and Jiebo Luo. 2018. Analyzing and predicting emoji usages in social media. In *Companion Proceedings of the The Web Conference 2018*, pages 327–334.