Tweet with a Smile and a Unicorn:
the use of emoji on Twitter in the Netherlands and England

Maximilian Roele
Graduation Thesis
Media Technology MSc program, Leiden University
November 2017
Thesis advisors: Prof. Max van Duijn, Prof. Janelle Ward

Abstract

With the immense growth of social media, emoji have also spread all over the world. Currently, there are more than a thousand of these colourful pictographs showing faces, creatures and objects. However, not all emoji see regular use, while others are extremely popular. This exploratory study attempts to answer the following question: how and why are emoji used on Twitter in the Netherlands and England? Through a quantitative analysis of two million tweets, current emoji usage in the two countries is described. A following qualitative study consisting of sixteen semi-structured interviews provides insight into why individuals use emoji and how they select emoji. Three important themes that influence the number of emoji used and the selection of emoji were found: the individual's purpose on Twitter, the perceived functionality of emoji and the individual's selection criteria for emoji. Overall, emoji play an important role in online communication and seem more complex than their light-hearted appearance suggests.

Keywords: emoji, emotion, social media, computer-mediated communication, Twitter

1. Introduction

Social media has grown immensely in the past decade, reaching individuals all over the world and playing an increasingly important role in the daily lives of many. Twitter had over 313 million monthly active users in 2016 (Twitter, 2016), while Facebook had 1.28 billion users logging in every day (CNN, 2017). On these and many other social media platforms, text is the predominant way through which people communicate with each other. This text is often accompanied by emoticons, typographic symbols such as ':)', or emoji, more complex pictographs of faces, creatures and objects that are sometimes seen as the successors to emoticons, such as '😊' (Pavalanathan & Eisenstein, 2016).

Emoji were first introduced in the Japanese mobile phone market in 1999 but gained global popularity when Apple introduced emoji on the iPhone in 2010. Since then, they have been brought to various social media platforms and mobile applications, with new emoji being added on a regular basis (Novak, Smailović, Sluban & Mozetič, 2015: 2). However, much research in online communication thus far has focused on emoticons or other aspects of messages such as lexical markers. Scholarly attention has only recently begun to shift towards emoji in fields such as

sentiment analysis (Suttles & Ide, 2013: 122-123). As a result, not much has been written in a scholarly context about why emoji are used thus far.

The purpose of this study is to contribute to the body of work on emoji and investigate how and why they are used on Twitter in two Western European countries. First, theory on nonverbal cues and emotion in face-to-face communication and computer-mediated communication (CMC) will be discussed. Second, the quantitative methods used to gather and analyse data will be described. This is followed by a description and discussion of the results. Based on the conclusions from this analysis, interviews will be conducted to add more in-depth insights. Finally, an answer to the following research question will be formulated: how and why are emoji used on Twitter in the Netherlands and England?

1.1 Non-verbal cues in communication

Smiling, hand-waving and a nod every now and then: these are some common ingredients for a daily conversation. Whenever we interact with others in face-to-face communication, we use a host of nonverbal cues such as body language and facial expressions to convey messages and intentions more accurately, avoid misunderstandings when joking, and to show how we feel (Matsumoto, Frank & Hwang, 2013: 4-7). Nonverbal cues can stimulate social bonding by signalling feelings of intimacy, affection and attraction, or negative feelings such as dominance and dislike (Burgoon & Hale, 1988). In addition, nonverbal cues and especially facial expressions appear to be important in the creation and maintenance of personal relationships (Ekman, 1992: 177). Thus, nonverbal cues seem to play a significant role not only in daily conversation, but in social behaviour as a whole.

However, Twitter and other forms of computer-mediated communication (CMC) are often said to lack this nonverbal information, an observation that is central to many CMC theories. Walther (2011) distinguished several theories that discuss this issue. Cues-Filtered-Out theories argue that this lack of nonverbal cues makes it impossible to form an impression of someone's personality and other personal characteristics such as gender, which then makes it hard to build meaningful relationships through CMC (Walther, 2011: 445-454). Meanwhile, Hyperpersonal CMC agrees that there is little nonverbal information to be found online, but claims that impression formation is still possible. Instead, CMC users rely on writing style, language and other cues provided by the platform they are using to create an impression of the person they are interacting with (Walther, 2011: 455-465). The use of emoticons and emoji may be a way to return nonverbal information to online computer-mediated communication and improve both understanding and impression formation.

In previous research focusing on emoticons, it was shown that emoticons can convey emotions or moods between sender and recipients and are seen as explicit or direct references to the emotions they represent (Derks, Fischer & Bos, 2008: 3). It has even been claimed that emoticons function as 'nonverbal surrogates' that directly replace facial expressions while communicating online (Derks, Bos & Grumbkow, 2007: 843). Emoticons can also clarify ambiguous messages or stress specific sentiments in the text it accompanies (Ganster, Eimlre & Krämer, 2012: 226). All put together, emoticons appear to have similar functions in online communication as nonverbal cues do in face-to-face communication. This all may also be true for emoji, which are now slowly replacing the use of the simpler emoticons like ':)' (Pavalanathan & Eisenstein, 2016). However, while it is likely that emoji can function similarly to emoticons, more research is needed before such conclusions can be drawn. Furthermore, as emoji are not limited to only facial expressions, their functionality may be more expansive than that of emoticons.

1.2 Use and interpretation of emoji

A few works have set out to investigate what individuals use emoji for, how emoji are used and how emoji are interpreted. Kelly and Watts (2015) conducted twenty interviews and found that emoji were used to maintain a connection with a conversational partner, for playful interaction, and for creating shared, unique meanings that were only understood within a specific relationship. These functions were said to make emoji important for building and maintaining relationships, similar to nonverbal cues in face-to-face communication. However, the focus of their study was on the use of technology in every-day communication and inclusion of emoji as a major topic occurred only in later interviews. By addressing emoji specifically in this study, more uses for emoji and other motivations behind their use may be uncovered.

Culture may also influence which emoji are used. This may be the result of modulation through display rules: a social group's norms about when, where and how emotions are expressed (Ekman & Friesen, 1975: 20). In the field of emotion science, the overarching Western, individualistic culture and the Eastern collectivist culture are often compared (Matsumoto, Yoo & Fontaine, 2008; Matsumoto et al., 1998). For example, Matsumoto (1991) found that Western individualistic culture encourages individuals to show intense emotions like extreme joy and is more accepting of expressions of negative emotions, compared to Eastern collectivist culture. This difference may be true for emoji as well. Lu et al. (2016) conducted a cross-cultural study of emoji and used Hofstede's[1] (2017) cultural dimensions framework to separate and investigate 102 countries. They found that

[1] Hofstede's cultural framework was first developed between 1967 and 1973 to describe the effects of a society's culture on its members' values. Its dimensions include power distance, individualism, uncertainty avoidance, masculinity and indulgence. At the time of writing, it contains scores for 102 different countries.

individualist and long-term oriented cultures are more likely to show positive emotions through positive emoji and use far less negative emoji when compared to other cultures. Some cultures were overall less likely to use emoji to show emotions, both positive and negative.

Novak et al. (2015) created the first emoji sentiment lexicon and described the use of emoji, investigating 70.000 tweets with emoji in thirteen European languages. They scored tweets according to a positive-neutral-negative sentiment scale and assigned these scores to the emoji contained within the tweets. This means that the scores were based on the context and sentiment of the tweet, not the emoji itself. Novak et al. (2015) found relatively high inter-annotator accuracy for the sentiment labels, as well as high and significant correlations for the labelled sentiments across countries. This may be because they used the relatively simple labelling of positive-neutral-negative, instead of more complex sentiments such as anxiety, humour, and so on. In terms of the emoji use, they found that 4% of all tweets included emoji and that the most commonly used emoji were overwhelmingly positive.

Other researchers have often struggled to find agreement when classifying emoji with specific emotions (Novak et al., 2015). Miller et al. (2016) found that 'anthropomorphic emoji' showing facial expressions are subject to greatly differing interpretations, with individuals describing different emotions or messages. There is even greater misconception when individuals have to interpret cross-platform emoji: emoji can look different across mobile platforms and (models of) devices. Researchers can thus opt to use natural language processing tools to automatically assign sentiment, based on various pre-set markers such as the use of specific words or hash tags (Lu et al., 2016). However, while useful, it is questionable how accurate such tools are when assigning relatively complex sentiments such as 'anxiety' to very short texts and ambiguous tweets or emoji.

1.3 Research question

This study will try to fill the gaps of knowledge on emoji described above and focus on how and why emoji are used. The first factor that may play a role is self-representation, or how individuals present themselves online. Often, an online profile represents an idealized version of the user in order to appear more appealing to others (Ellison, Heino & Gibbs, 2006: 418). Lu et al. (2016) found that negative emotions are often not expressed through negative emoji and Novak et al. (2015) reported that the great majority of emoji that are used represent positive sentiments. Twitter users may be hesitant to show negativity as this might affect how others view their online presence, which in turn lowers the use of negative emoji as a whole. Display rules may also influence this: Twitter users may be using emoji that are considered socially appropriate, which could limit the set of emoji they choose from to 'appropriate' ones.

The second factor takes a more linguistic approach. Peirce's study on signs in language may help to explain why certain kinds of emoji are used more often than others. Peirce identified three categories of signs: the icon, index and symbol. An icon displays the object it signifies through visual similarity. An index represents something through its connection with the object it represents: smoke, for example, signifies there is fire. The symbol is a sign that has no clear resemblance or connection to what it signifies and its meaning must be learned (Merrell, 2001: 31). An icon can be easily understood through its visual similarity, though over time it can develop into a more complex index or symbol as more abstract concepts or interpretations are added to it through social interactions within a particular linguistic community (Stokoe, 2001: 43). It seems likely that emoji that are easily understandable or share visual similarity to familiar expressions are more popular than emoji that are abstract. Furthermore, emoji may become symbolic and gain unique meanings within communities or even within specific relationships, as found by Kelly and Watts (2015).

This is an exploratory study that remains open to other factors that might influence emoji use. By using both quantitative and qualitative methods to explore the possible explanations above and possibly identify other factors that may be important, this study will attempt to answer the following research question: how and why are emoji used on Twitter in the Netherlands and England?

2. Method: Twitter data

In the following section, the methods that will be used for Twitter data gathering and analysis will be elaborated on. The qualitative section will be discussed after the initial data analysis as it draws from the conclusions made in this analysis.

2.1 Data mining tweets

Two million tweets from England (in English) and the Netherlands (in Dutch) were gathered over a period of 1.5 weeks in July 2017. These countries were selected in order to collect a larger sample of Western tweets in two different languages, which may provide more insights than data in one language. Only geocoded tweets were used in order to ensure that the tweets were posted by individuals in the areas that were part of this study. A language filter was insufficient for this purpose as both Dutch and English are spoken in various countries. A downside is that Twitter users must explicitly allow Twitter to embed their location in tweets, which means that geocoded tweets represent only a small sample of the total number of tweets. While this may have introduced a selection bias, it was accepted due to the large sample size and because geocoding was to conduct this study.

While gathering tweets, the data was processed to remove information that was irrelevant for this study such as a user's Twitter description, display name and the number of replies. Data that could be used to identify specific Twitter users was removed, with the exception of tweet text and usernames. Tweet text was removed from the data after the quantitative analysis had concluded. Usernames were only stored temporarily so that individuals could randomly be selected for interviews and were deleted immediately after the interviews had been concluded.

Sentiment scores were calculated and assigned to tweets, based on the emoji contained in the tweets. The sentiment scores are based on the emoji sentiment lexicon by Novak et al. (2015). They designated sentiment as follows: positive as 1, neutral as 0 and negative as -1. The final assigned sentiment score of an emoji consisted of the average of all scores. In this study, we use the range of 1 to 0.11 as positive, 0.1 to -0.1 as neutral and -0.11 to -1 as negative sentiment scores. The sentiment lexicon was chosen for use in this study because of the large amount of data it is based on and the high accuracy during its coding process. However, the lexicon does not include scores for all emoji found in the data.

3. Results and data analysis

This section first describes current emoji usage, their frequency, the sentiment scores associated with the used emoji and the use of emoji combinations. Based on the findings, questions are formulated for the following qualitative analysis.

3.1 General emoji usage in Dutch and English tweets

Of all the English tweets, 17.2% or 175.684 included emoji, while 9.7% or 97.784 of Dutch tweets did. Interestingly, both of these percentages are higher than the 4% found in Novak et al. (2015)'s study and might point towards a general increase in emoji usage within the period of a few years. Both the Dutch and the English tend to place their emoji towards the end of their tweets. In Dutch tweets, emoji first appeared on average at 81.4% of a tweet's length, and at 79.7% in English tweets.

Amount of emoji used. In total, Dutch Twitter users used 163.447 emoji, while the English used 364.213 emoji. This results in an average of 1.67 emoji in Dutch and 2.07 emoji in English per tweet. Table 1 shows how often a certain amount of emoji was used in a single tweet. The majority of tweets in both languages only include a single emoji; a large number of tweets include either two or three but rarely more than that. Overall, Dutch Twitter users appear to use less emoji both in general and per tweet compared to English users.

Table 1. Amount of emoji in tweet, the number of tweets and percentage of total tweets with emoji

Tokens	Dutch (%) (*N* = 97784)	English (%) (*N* = 175684)
1	63887 (65.3%)	91020 (51.8%)
2	18652 (19.1%)	40461 (23.0%)
3	8632 (8.8%)	22480 (12.8%)
4	3339 (3.4%)	10099 (5.7%)
5	1333 (1.4%)	4280 (2.4%)
6	873 (0.9%)	3094 (1.8%)
7	335 (0.3%)	1277 (0.7%)
8	226 (0.2%)	896 (0.5%)

Variety in emoji usage. The Dutch used 960 different emoji while the English used 1016. At first glance, this suggests great variety, yet only few emoji actually see regular use. The top 10 most popular emoji, which will be discussed in-depth in the following section, represent 38.5% of the total amount of used emoji in Dutch tweets and 35.9% in English tweets. Most of the remaining emoji were used less than 0.1% or even 0.01% of the total amount of emoji. Rarely used emoji tend to represent clothing, vehicles (), animals (), specific activities () or abstract symbols.

Table 2 shows the number of different emoji used inside a single tweet. As can be seen, the majority of tweets include only one kind of emoji. In general, when multiple emoji were used in a tweet, they were either repetitions of the same emoji or a combination of two different ones. Combinations consisting of more than two different emoji were not used often. The combinations that were used most often will be discussed more in-depth in section 3.3.

Table 2. Amount of different emoji types in a tweet, number of tweets and percentage of tweets with emoji.

Types	Dutch (%)(*N* = 97784)	English (%)(*N* = 175684)
1	74277 (76.0%)	119686 (68.1%)
2	15939 (16.3%)	35116 (20.0%)
3	4995 (5.1%)	13031 (7.4%)
4	1597 (1.6%)	4681 (2.7%)
5	583 (0.6%)	1744 (1.0%)
6	190 (0.2%)	688 (0.4%)
7	94 (0.1%)	307 (0.2%)
8	39 (0.1%)	184 (0.1%)

Emoji sentiment. In table 3, an overview is shown of the amount of positive, neutral and negative emoji that were used. Both Dutch and English Twitter users overwhelmingly use positively-scored emoji, with only a small proportion of emoji that were scored as neutral or negative. There are a number of tweets for which no sentiment score could be calculated: this is the case when a tweet only included emoji which were not assigned a sentiment score. These emoji were introduced after

the creation of Novak et al.'s (2015) sentiment lexicon. In addition, some symbols like the copyright sign are technically counted as emoji but are also not included in the lexicon. A list of emoji without a sentiment score (used at least 100 times) can be found in appendix I. This is unlikely to have influenced the results too much: the majority of emoji without a sentiment score were rarely used.

Table 3. The frequency of sentiment scores in tweets and the percentage of tweets with emoji.

Sentiment	Dutch (%)(*N* = 97784)	English (%) (*N* = 175684)
Negative	4795 (4.9%)	12622 (7.2%)
Neutral	4496 (4.6%)	13470 (7.7%)
Positive	80386 (82.2%)	134190 (76.4%)
Missing	8016 (8.2%)	15035 (8.6%)

A visual representation of the distribution of emoji sentiment scores can be seen in figure 1. For Dutch emoji, the average sentiment score is .393. The sentiment in English tweets is slightly less positive at .348. In both languages, the majority of emoji had positive or very positive sentiment scores as seen by the histograms slanting towards positivity. In addition, there are large peaks that indicate a high prevalence of a specific sentiment score. These peaks can be explained by the popularity of specific emoji or combinations of emoji with similar sentiment scores.

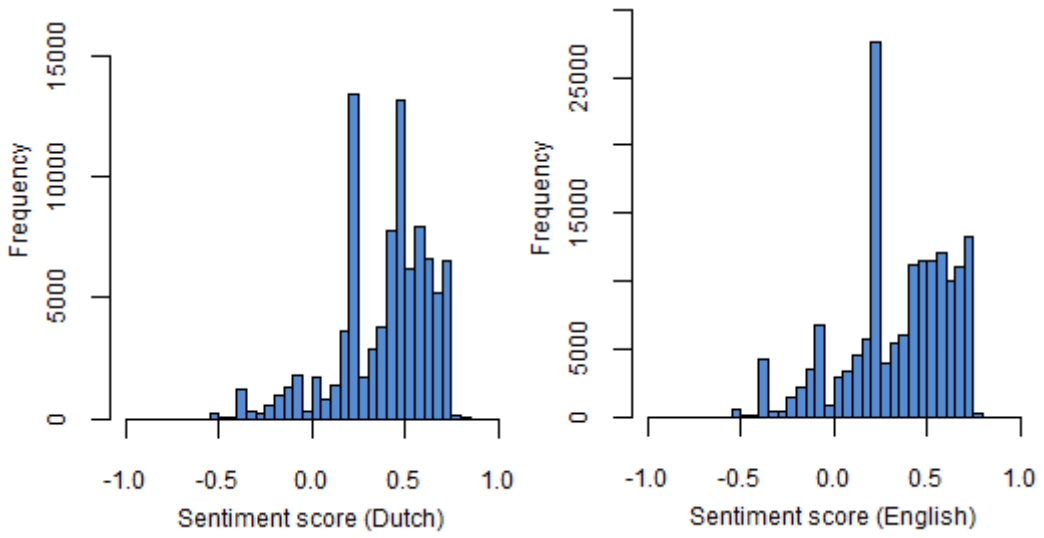

Figure 1. Left: Dutch (*N* = 89769). Right: English (*N* = 160650). Distribution of emoji sentiment in tweets.

3.2 Commonly used emoji

Figures 2 and 3 provide an overview of the top 10 emoji for Dutch and English, how frequently these emoji were used and what their associated sentiment scores are. The majority of emoji are on the positive side and six of them are present in both lists with slight differences in the frequency of use. In the Dutch list, one emoji lacks a sentiment score as it is not included in Novak et al.'s (2015) emoji sentiment lexicon, though it is likely that it is also perceived as a positive emoji. It is interesting to

note that the Dutch top 10 includes absolutely no neutrally or negatively scored emoji, while the English list includes both 😭 and 😩. In addition, the English use various hand gesture emoji.

The 😂 is the undisputable leader of emoji and is used more often than any other emoji. It is responsible for 13.5% and 15.6% of the total amount of emoji used in Dutch and English tweets respectively. Its high frequency explains the high prevalence of its sentiment score (0.221) for both Dutch and English in figure 1. However, the frequency with which these popular emoji are used drops swiftly as one goes down the list. The tenth emoji in the list only makes up 1.6% and 1.3% of the total amount of emoji used in Dutch and English tweets respectively.

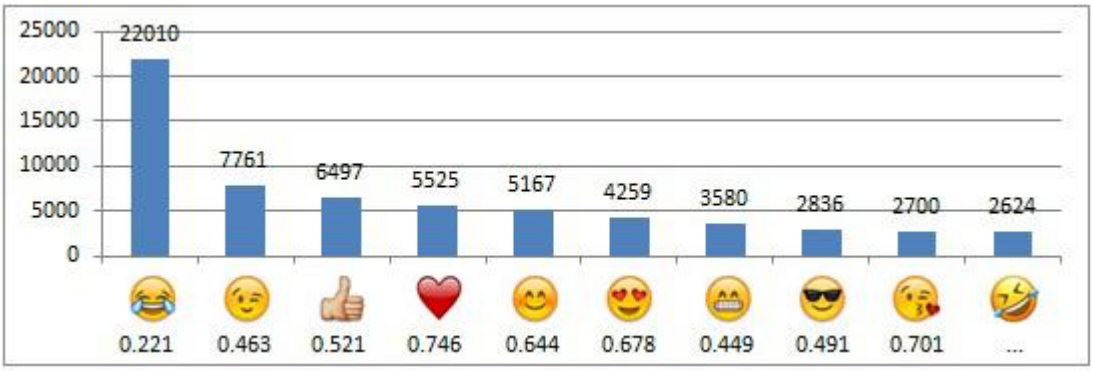

Figure 2. Dutch top 10 emoji, frequency and associated sentiment scores (left to right is top 1-10).

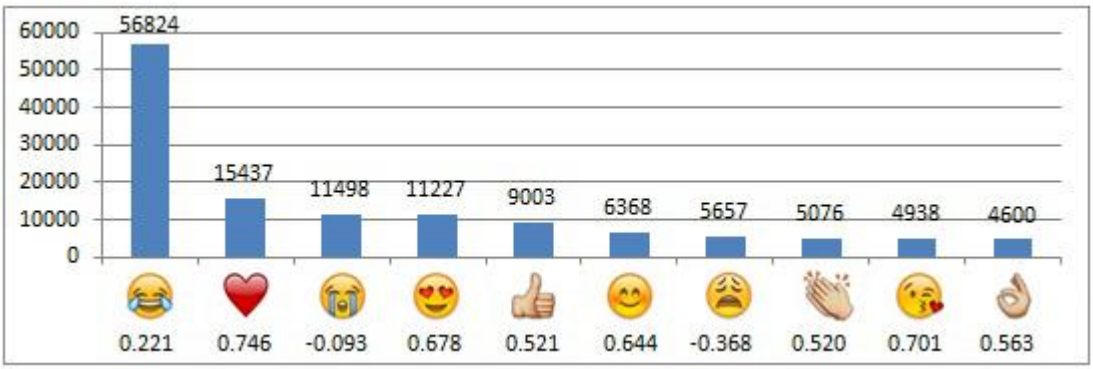

Figure 3. English top 10 emoji, frequency and associated sentiment scores (left to right is top 1-10).

The frequency of use for emoji beyond the top 10 is clustered together and continues to drop. The use of certain emoji may be influenced by events or seasons, like Christmas tree emoji during the Christmas period. As our data was gathered during the summer, the 😎 was often used in Dutch tweets about sunny holiday destinations. This increase in use may have pushed it to the top 10 as the difference in frequency is very small for most emoji. This also suggests that a fixed top 10 of emoji may not exist as emoji can easily rise or decline in popularity, resulting in different top emoji at different times. For future studies, a longer period of time from which data is gathered may provide results that are not so influenced by such events.

All the popular emoji refer to emotions, facial expressions and gestures. Interestingly, the most popular Dutch emoji include several variations of a smiling or happy face, while the English list is more varied and includes different expressions and hand gestures. It appears emoji are used as a means to transfer such nonverbal cues to a platform where users cannot see each other and thus have to rely on these emoji to communicate this information.

3.3 Combinations of emoji

Combinations of emoji appear quite often. 34.7% of Dutch tweets and 48.2% of English tweets included two or more emoji. A total of 10053 different combinations were found in Dutch tweets and 22717 different combinations in English tweets, though most combinations only occurred once[2]. The latter suggests that, while there is a lot of variety, there are only few combinations that are more widely used by Twitter users. Instead, users may be choosing highly specific combinations that suit a particular feeling or situation.

An overview of the most common emoji combinations is seen in figures 4 and 5 for Dutch and English tweets, of which five combinations are present for both languages. These combinations do not have to consist of different emoji: in fact, most popular 'combinations' consist of the same emoji repeated several times, instead of a mix of different emoji. There is a large difference in the frequency of use between the top combination and all that follow after it: the 😂 emoji is far ahead and is often used multiple times within a single tweet. Furthermore, the frequency with which many combinations are used does not differ much from other combinations and, as with single emoji, specific events like Christmas may affect the use of emoji and thus the frequency with which certain combinations occur.

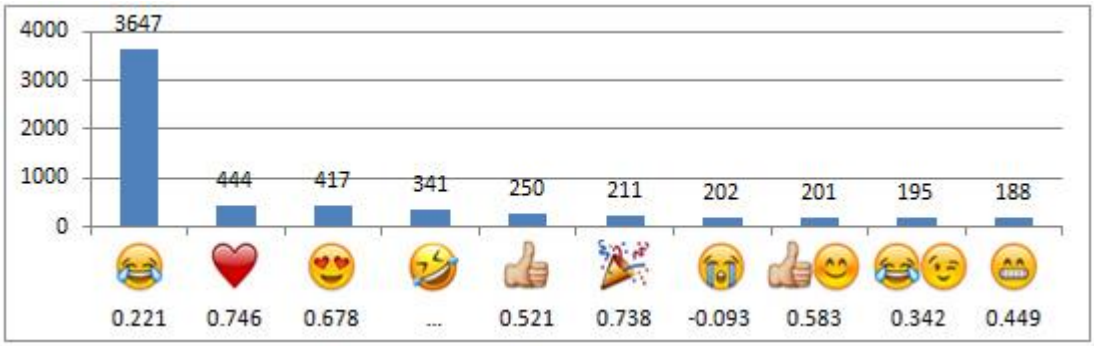

Figure 4. Dutch top 10 combinations, frequency and associated sentiment scores (averaged).

[2] Combinations of emoji with modifiers are not included. Modifiers themselves are not visible: they only change the appearance (such as the skin colour in case of the Fitzpatrick modifier) of the emoji they accompany. If a combination consisted of multiple emoji and modifiers, only the emoji were preserved.

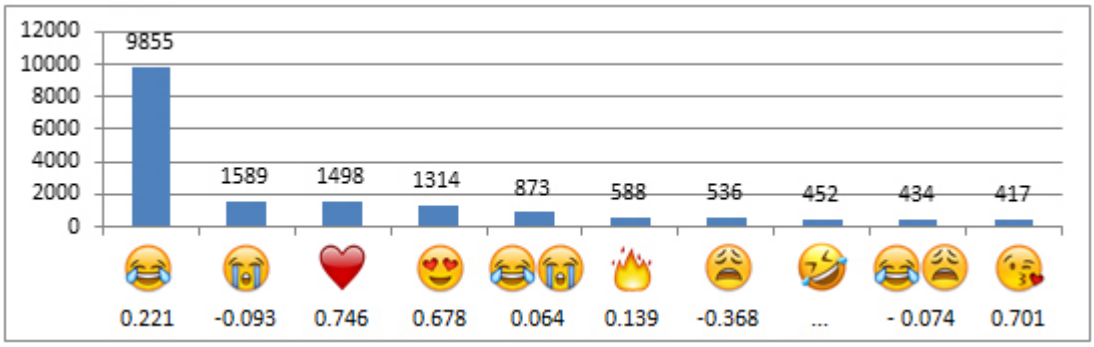

Figure 5. English top 10 combinations, frequency and associated sentiment scores (averaged).

Interesting emoji in the lists are the 🎉 in the Dutch list, as well as the 🔥 and the two seemingly contradictory combinations of 😂😭 and 😂😩 the English list. The 🎉 is rated as highly positive by Novak et al. (2015) and is likely used several times in one message to emphasize the festive or happy sentiment it conveys – much like the highly popular 😂 emoji. The 🔥 emoji appears to also be used in a very positive way in English tweets: it is often used in reference to music, events and fashion being 'lit', 'fire' or awesome. Its relatively low sentiment score may be explained by the fact that Novak et al. (2015) aggregated data from different European languages, which may lack the saying of something being 'fire'. Dutch is likely one such language: Dutch tweets tended to use the 🔥 more literally, referring to hot things such as the weather and spicy food. This shows a small linguistic difference that also resulted in different emoji usage between the two countries.

The combination of 😂😭 and 😂😩 includes emoji that are positively and neutrally or negatively rated on sentiment by Novak et al. (2015). In the full context of the tweets, these combinations make more sense and appear to be quite literal in their use. 😂😭 can be used to show how someone is crying with laughter over something hilarious. 😂😩 is often used to laugh at something that is both amusing yet also irritating, as this tweet from the data shows: "when people say 'I love you' after being together for 5 minutes 😂😩".

The combinations provide us with two insights. First, that emoji usage may depend on language. If an emoji can be used as a direct reference to a common saying like the 🔥 emoji, it may be used more often. Likewise, if an emoji does not have a direct equivalence in language, it may be less popular. Second, the majority of combinations again are visually similar to many facial expressions or gestures, supporting the idea that emoji function as surrogates for nonverbal communication online.

3.4 Discussion: data analysis

The current use of emoji can be described as follows: most common emoji and combinations represent nonverbal cues, there is limited variety in the emoji used, and the majority of emoji are positive. It may be that the use of emoji is similar to the use of icons in language described by Peirce (Merrell, 2001). Twitter users seem to converge on a small set of emoji that are easy to understand and steer clear from more symbolic emoji that can be unclear in their meaning. Negative emoji are rarely used. These findings raise several questions that may be answered during the interviews: how do Twitter users choose specific emoji? What are the requirements an emoji has to fulfil for them? Why are negative emoji so rarely used?

Another factor that may play a role is language. Languages can have words or concepts that are unique to that language and emoji may become symbols for these within specific linguistic communities, like the 🔥 emoji. This fits with the idea of emoji as icons, indices or symbols which gain meaning through social interactions in these groups. However, an in-depth content and topic analysis would likely be a more suitable method than interviews to investigate this finding. In this study, content analysis and the comparison of emoji use across countries were not the main focus and was mostly done to provide clarity or examples.

Unfortunately, the emoji sentiment lexicon may provide sentiment scores that do not accurately reflect the sentiment that emoji represent. As the scores are averaged, an emoji that sees both positive and negative use can receive a 'neutral' score. This score can reflect how an emoji is used overall within a larger dataset, but not the sentiment of emoji within a single tweet or the variety of sentiments that one emoji can convey. In this study, the lexicon was still useful to find how sentiment was distributed in the dataset and the Twitter data itself provided a way to reflect upon the real sentiment that emoji in specific tweets carried. Future studies may wish to rely on other methods to extract the sentiment of tweets and emoji.

4. Method: qualitative interviews

In order to garner new insights on the use of emoji and answer the question of how and why emoji are used on Twitter, sixteen semi-structured interviews were conducted with Twitter users from the Netherlands and England who use emoji.

4.1 Participants

In order to gather participants for the interviews, two random samples of 40 Twitter users were collected from both the Dutch and English datasets using SPSS. The sample size was larger than the

total amount of interviewees sought in order to compensate for accounts that could not be contacted. The remaining users were invited to participate through Twitter direct messages and were informed they had been randomly selected from a dataset of tweets. After inviting 41 randomly selected individuals, a total of 5 Dutch and 6 English interviewees were found. Due to a low response rate on Twitter, snowball sampling was used in order to gather a larger and possibly more varied set of interviewees. Interviewees were asked if they had friends who used emoji on Twitter and if they were willing to participate in an interview. Three Dutch and two English participants were recruited this way. There were five male and three female Dutch participants; ages ranged between 18 and 26 with a mean of 23. There were four male and female English interviewees, with ages between 18 and 30 and a mean of 25. Before commencing the interview, all interviewees were informed that their answers would be anonymized, that they could end the interview at any moment and that there was no compensation.

4.2 Data collection

Sixteen semi-structured interviews were conducted in October 2017 through instant messaging on Twitter, Whatsapp and Discord. This provided benefits such as allowing interviewees to spend time gathering thoughts and work on their message, looking at previous messages, and being able to use emoji directly (Voida et al., 2004). Interviews lasted 45 to 60 minutes on average. Some interviews were conducted over the span of a few days, but most participants completed the interview in one sitting. The chat logs were copied and used as the transcripts. All of the interviewees' names, locations and other information that could be used to identify someone were altered or deleted from the transcripts to preserve anonymity.

There were seven main questions; probing questions were asked in order to better understand answers, to have interviewees expand on a short answer or to continue with an interesting phenomenon that was mentioned. The questions were mainly focused on the use of Twitter, motivations behind the use of emoji in general and the selection of specific emoji. The questions were formulated based upon existing literature, the gathered data and the questions raised in section 3.4.

4.3 Data coding

Every interview was coded through the open-axial coding process, based on grounded theory (Corbin & Strauss, 2008: 62-74; Benaquisto, 2008: 84-88) and similar to Kelly and Watts' (2015) work. In the open codes, answers were summarized and interesting as well as recurring answers were highlighted. Some examples of these open codes are: 'emoji show feelings', 'emoji represents

mood' and 'emoji shows my expression'. These open codes were then reduced through axial coding: similar codes that represented a topic were categorized into overarching axial codes. The open codes mentioned prior would for example all fit under the axial code 'express emotion'. Finally, through selective coding, main themes that arose from the axial codes were identified. Three main themes that captured the essence of the interviewees' answers and which seem to impact emoji usage are as follows: the *purpose of Twitter*, *emoji functionality* and *emoji selection criteria*.

5. Interview analysis

The analysis will first begin with a description of the interviewees' purpose on Twitter and how they use the platform. Secondly, the focus of the analysis will shift to how interviewees adopt and currently use emoji on Twitter, while making some comparisons to the results from the quantitative study. Thirdly, the functionality of emoji as perceived by the interviewees will be described. Finally, the process through which interviewees select specific emoji to use in their tweets will be discussed. Quotes are used to illustrate the findings through the interviewees' own words (Creswell & Poth, 2017: 245).

5.1 Purpose of Twitter

When interviewees were asked about what they mainly used Twitter for, three broad categories emerged from their answers: personal, semi-professional, and professional use. Nine interviewees use Twitter exclusively for personal reasons but the activities they described greatly varied and include: sharing opinions, tweeting to celebrities, tweeting about the news, personal entertainment, using Twitter as an anonymous yet public diary, and sharing one's favourite lyrics.

Most interviewees who use Twitter in a more personal sense indicated that they often use emoji in their tweets with at least one emoji every 5 to 10 tweets. In general, emoji were described as a way to express personality and liveliness in tweets, or as one interviewee wrote: 'adding spices to what would otherwise be bland text'. The number of emoji they use was said to also be influenced by their relationship with the individual they are tweeting to: many interviewees said they use more emoji when tweeting to people that they are closer with and less emoji with more distant individuals such as celebrities.

> *"I find it pretty important, to add things [emoji] like that. If someone never uses emoji or something at all, it feels off. It's just a part of your personality really. Just like how you express yourself in real life. Not using emoji, it feels like someone who doesn't look at you when you talk in a face-to-face conversation." – Rick (NL)*

Four of the interviewees use Twitter for both personal and professional purposes, tweeting about daily life and sharing opinions while also discussing their own work or business. Three interviewees described their Twitter use as mainly professional, though what they perceived as professional differed: two participants used Twitter to publicly communicate with companies and their customer service, while one interviewee used the platform exclusively to discuss his own products.

> *"I mainly use it to contact companies directly. They seem to be more responsive when you mention them in a Tweet haha. I think it is because others can see what they are mentioned in and they are scared of it having a negative influence on them." – Mara (ENG)*

Those who use Twitter for professional reasons said they use fewer emoji as more of their tweets are serious and using emoji then feels inappropriate or irrelevant to the topic. They also limit the number of emoji they use as they consider using too many emoji as unprofessional. This is a sentiment shared by almost all interviewees, regardless of their main purpose on Twitter, and many indicated they tend to reduce or avoid the use of emoji in serious or professional tweets.

5.2 Emoji adoption, interpretation and usage on Twitter

However, how do Twitter users learn the meaning and use of emoji? Almost all interviewees said they learnt the interpretation and situations in which to use emoji through interaction within their own social circle and through online communities. This interaction was not limited to Twitter: for many, it occurred over many years and over different platforms. Older interviewees described using MSN's smileys in their teenage years use emoji similarly to how they used smileys then. Others said they learnt how to use and interpret emoji through communication on Facebook, Whatsapp and online communities such as Discord, then carried over this knowledge to Twitter. This is illustrated by the following quote:

> *"Actually, I think the use of emojis is not platform dependent. Emojis are just part of someone's expressive vocabulary and some people have preferences for specific emoji, just like with filler words. And the emojis are pretty similar across different media." – Marieke (NL)*

All interviewees indicated they also base their interpretation of an emoji on its appearance. Often, the meaning of an emoji is clear due to its visual resemblance to real facial expressions, gestures, objects and well-known symbols such as the heart of love. This is reminiscent of Peirce's 'icon' that represents something through visual resemblance (Merrell, 2001). However, there are also emoji which are not as straightforward, or emoji that were said to have a secondary or more abstract meaning. For these, interviewees said that the meaning is learnt through interaction with others who use this emoji or by deducing the meaning of an emoji from the tweet's context.

New meanings can also be introduced to emoji: five interviewees described how emoji took on seemingly unrelated, funny meanings that only a specific friend or group of friends could understand. Interestingly, this also suggests that the meaning and use of emoji does not only rely on one's culture, as was hypothesized, or one's language, as found in the quantitative analysis. Social communication between much smaller groups of people such as subcultures, specific groups of friends and family, or even two individuals also impact emoji usage. Overall, this process seems to be close to Peirce's theory on signs, where an icon can develop into a more abstract symbol with a deeper meaning through social interaction (Stokoe, 2001).

> *"I learnt the meaning [of emoji] from others.* 😩 *I didn't know this one until I saw others use it when they find something sexy as f***. Others are pretty self-explanatory, kissing a heart, sunglasses being cool, crying of laughter, thumbs up, really universal signs." – Colin (NL)*

The emoji that almost all interviewees use closely resembles the list of popular emoji for their particular country. When asked to share their most commonly used emoji on Twitter, almost everyone listed the following emoji: 😂❤️😭😉😍😊😎👍👏 . Usually, only two or three of an interviewee's most commonly used emoji were different from the most popular ones. Which emoji interviewees used instead was said to depend on the emoji that are used within their social circle, as well as personal taste and their own interpretation of emoji. The popular 😉 emoji for example evoked a strong, negative reaction in two interviewees (one Dutch, one English) as they perceived it as a mocking emoji that accompanies fake compliments, as well as an emoji that comes across as desperate when used in flirty tweets – thus, they would never use this emoji themselves.

For combinations of emoji, most interviewees indicated that they often repeat the same emoji, with 😂😂 being the most commonly mentioned one. Combining different emoji is less common and only seven interviewees indicated they did this. This kind of combination was said to be very specific to the situation and what interviewees wanted to express. This explains why the quantitative study showed a high variety in combinations, but only few combinations that were used frequently.

> *"For example I do* 👀✍️ *when I want to let them know I'm writing down what someone says, that's funnier than when you really write that out." – Rick (NL)*

Combinations of different emoji also appear to serve a different purpose than repeated tokens. All interviewees agreed that repeating emoji emphasized something, while mixing different emoji had a variety of uses: conveying more complex emotions that a single emoji or words could not, sending stand-alone messages such as 👀✍️, or to add embellishments without further meaning like 🦄🌈

to a tweet. This agreement suggests some kind of common, learnt understanding of how emoji combinations are to be used.

> *"Combinations that use different emojis send a different message than repeating the same one. Combinations tell more of a story and repeating one is more for emphasis." – Amy (ENG)*

5.3 Emoji functionality

The convergence towards relatively similar emoji on the one hand and the avoidance of many emoji and combinations on the other hand, may in part be explained by what interviewees perceive to be the functionality of emoji: expressing emotion, modulating the tone of a tweet, clarifying tweets and adding liveliness to tweets.

Expressing emotion. Almost every single interviewee considers expressing emotion as the main purpose of emoji. Emoji are seen as a way to accurately express what someone is feeling and five interviewees said that the emoji they use even reflect their facial expression at the time they use them - albeit in an exaggerated manner. This relates to the finding in the quantitative section that most popular emoji represent nonverbal cues. Furthermore, the use of emoji that appear more exaggerated in sentiment may be due to the way less strong emoji like 🙂 are interpreted: five interviewees considered them to come across as insincere or 'too dull' to add anything meaningful. This function of emoji might also explain why there is a limited variety in emoji usage: several interviewees said that they felt they could express most of their emotions with a set of ten to twenty emoji. There may be a fixed, small subset of emoji that is capable of accurately representing most emotions that people want to express online and are thus used more often than any other emoji.

> *"It's not like I'm always smiling in front of my pc/phone but it's comparable when I use emoji. I do it when the feeling comes up sort of. After years of chatting, it's automatic." – Rick (NL)*

Modulating tweets. Emoji and combinations of emoji were said to be used to modulate the sentiment expressed in a tweet: they soften or emphasize a sentiment. Most often, this is used to express the intensity of emotion more accurately by, for example, using multiple of the same emoji to emphasize happiness. Several other reasons were also mentioned during the interviews, such as using positive emoji to appear more genuine or friendly, using emoji to add nuance or complexity to a sentiment expressed in a tweet's text, or to soften the severity of rants and jokes at the expense of others. One interviewee, Michael, explains that he mixes the sentiment as expressed by the text of a tweet with that of an emoji in order to express something more nuanced or complex:

"I generally use emoji when an emotion I'm trying to convey is complex. I'll search for an emoji that I feel supports it. It is more common that I'll use an emoji in conjunction with a tweet that is melancholic/perplexed or tentatively excited, than one that is simply happy or sad. I don't feel the need to include an emoji to express something I am happy about, but it is useful in support when trying to express something more nuanced, like cautious optimism." – Michael (ENG)

Clarifying tweets. In cases where interviewees felt that a tweet might be misunderstood, eight interviewees indicated they use emoji to clarify that a tweet is, for example, a joke or sarcasm. The 😂 was commonly mentioned as an indicator for sarcasm or irony. Those who did not perceive clarification as a function of emoji either expected their audience to understand their tweets without the use of an emoji, or preferred to make their tweets as clear as possible through text only.

Adding liveliness. Eight interviewees mentioned that they use emoji to add liveliness and fun to their tweets, without necessarily conveying any additional meaning or emotion. Emoji then serve as embellishments that make tweets look more aesthetically pleasing and 'more fun to read' according to interviewees. This function might lead to more varied choices in emoji. Some examples of emoji that interviewees used for this included emoji like 🌹 or 🌟 and combinations such as 🦄🌈.

5.4 Emoji selection criteria

In this section, the focus will shift towards the selection of specific emoji. During the interviews, several 'criteria' that had to be fulfilled by an emoji arose from interviewee's answers: they must be *relevant*, *accurate*, *clear* and *efficient*. However, all interviewees said they do not think extensively about which emoji they use; they saw it as a subconscious, automatic process where they simply 'know' which emoji fits best and fulfils these criteria within seconds at most.

Relevance. Relevance appears to be the most important factor in deciding which emoji to use. Interviewees tended to describe an emoji as relevant when it fit the topic of a tweet and if it matched the sentiment expressed in it. An example one interviewee gave of this was the use of 🎉 in a congratulatory tweet, as it matches both the topic and its positive sentiment. However, whether or not an emoji counts as 'relevant' appears to be highly personal: some interviewees can struggle to find relevant emoji that fit their requirements while others can always think of something fitting as long as it is tangentially related to the tweet's topic. This appears to be strongly influenced by an interviewee's own interpretation of emoji.

This factor may also partially explain the lack of use for negative emoji. Nine interviewees said they avoid posting negative tweets as they feel a need to filter themselves. As a result, they rarely or never use negative emoji because there are hardly any tweets in which they are relevant.

Furthermore, most interviewees said they find emoji to generally be irrelevant in serious or professional tweets. They consider emoji to be too cartoony and using them could downplay a serious tweet. Considering the finding from the quantitative study that most emoji represent less than 0.1% of the total emoji used, it is likely that most emoji represent objects and expressions that are not relevant for the majority of tweets.

Accuracy. When selecting an emoji to represent an emotion or expression, interviewees found it important to choose one that they felt most closely resembled what they wanted to express – this resemblance was often visual. This again depends on personal interpretations of emoji, which can differ greatly among individuals as was also found in prior studies such as Miller et al. (2016). Interviewees indicated that emoji, which at first glance may seem similar in appearance and in sentiment, have nuanced differences in interpretations. For example, one interviewee used 😄 to show general happiness, and ☺️ to show she was feeling 'very glad about something'. However, as stated before, it seems there are only a limited number of emotions that are expressed through a limited subset of emoji.

Accuracy also affects the (dis)use of negative emoji. While some interviewees avoid posting negative tweets and thus negative emoji as was hypothesized, this does not account for all cases: those who did post negative tweets indicated that they found it difficult to express negative emotions through emoji. Three interviewees could not find any negative emoji that accurately conveyed the feeling they wanted to express and felt that only using text to convey such a feeling was a better option.

> *"I would only show positive emotions. The heart. I find it hard to express negative emotions with emoji. I can show negative emotions without issues in real life and in text too, but not emoji. Because I think I can do that better in real life. None of them [emoji] fit." – Evert (NL)*

Clarity. Seven interviewees tried to limit themselves to more easily understandable emoji as they did not want their readers to misunderstand tweets. They tend to use popular emoji as they expect these to be understood by a wider audience due to their widespread use and their resemblance to real-life expressions and gestures. They avoid combinations as they consider these 'harder to understand'. In general, interviewees said they adjusted their emoji use depending on who they were tweeting to, as Lily describes:

> *"i use lots of combinations like 😎🔫, its just doing finger guns and being cool. i use combos with my mates because we understand ahaha. [...] some people just dont get it like it'll come across as something else to some people. So i tend to use standard emojis unless it's someone i know who gets it. Just the most popular ones. mostly no combos." – Lily (ENG)*

Efficiency. This factor relates to Twitter's user interface and the ease with which Twitter users can select emoji, though its role seems to be less important than the other factors. Many interviewees fall back on their 'most used' or 'recently used' emoji as it saves them the time and effort of having to search through various tabs and hundreds of emoji. This limits the variety of emoji used, as few individuals are willing to take the effort to pick an emoji they have not used before – it is not efficient to select such an emoji. One interviewee stated that Twitter's user interface was the sole reason he did not use as many emoji as he did on other social media platforms and he found it difficult to quickly select the ones he wanted to use.

> *"when I post something on twitter it's more often text because I don't put effort into picking my 'mood' or something. It has to do with the ease of use for me. On Skype I know how to type most [emoji] and on Whatsapp I have my list of recently used, but twitter and fb don't invite the use of emoji (in my opinion)" – Carlo (NL)*

Two factors were only mentioned in passing by a couple of interviewees: appearance and personality. A few interviewees said they do not use emoji they consider ugly or disturbing, like 🤡. Other emoji were avoided because interviewees felt they did not fit their personality – Jasper, for example, would never use a 'sassy' 💅 simply because it does not fit him as a person.

5.5 Discussion: interviews

During the interviews, many functions and criteria for emoji were identified. It is difficult to generalize the use of emoji because the sample size is likely too small to be representative of all Twitter users, and even then there were already great differences between interviewees. However, it seems that for most interviewees, all of the functions and criteria play a role but with different priorities. Expressing emotion and modulating tweets were often seen as the most important functions overall, with accuracy and relevance being important in the selection.

Overall, the adoption of emoji seems to be a circular process: individuals can both learn and introduce new meanings to emoji within social groups, specific relationships and across social media platforms. Future studies could investigate emoji use within specific subcultures or other smaller social circles to build upon this finding. These smaller social groups may play a bigger role in emoji usage that may be overshadowed by the focus on emoji usage by nationality or culture in many studies.

In the future, interviews conducted with a larger and likely more representative group may help to identify more functions for emoji and selection criteria, and to further investigate the emoji

adoption process in more detail. Furthermore, as this study only focused on those who did use emoji, it may be interesting to involve non-users of emoji to find out their motivations behind avoiding emoji and how they perceive emoji. As many of the interviewees found emoji to be important for the expression of emotion online, investigating non-users may provide insights into other forms of emotional expression in online communication as well.

6. Conclusion

How and why are emoji used on Twitter in the Netherlands and England? For both countries, most emoji that are used are positive, there is limited variety in emoji choices and there is a convergence towards a set of highly popular emoji, like 😂. The 'why' behind emoji usage can be explained by three main factors: the purpose of Twitter for an individual, the perceived emoji functionality and emoji selection criteria. These factors, as well as language, likely influence how many and which emoji are used. However, this all seems to be subject to personal preferences and prior experiences that individuals have had in online communication. Online communication across many online platforms – not just Twitter - and different social circles impacts the way someone perceives, interprets and uses emoji on Twitter; it is not only language or self-presentation that influences it. The use of emoji thus seems to depend on a more complex and social process than their light-hearted appearance suggests.

References

Benaquisto, L. (2008) “Codes and coding” in: L.M. Given (ed.) *The SAGE encyclopedia of qualitative research methods* (Thousand Oaks, CA: SAGE)

Burgoon, J.K. & Hale, J.L. (1988) “Nonverbal expectancy violations: model elaboration and application to immediacy behaviors” *Communication Monographs, 55,* 58-79

CNN (2017) “Facebook tops 1.9 billion monthly users” <http://money.cnn.com/2017/05/03/technology/facebook-earnings/index.html> Retrieved 16 June 2017

Corbin, J.M. & Strauss, A.L. (2008) *Basics of qualitative research: techniques and procedures for developing grounded theory* (Thousand Oaks, CA: SAGE)

Creswell, J.W. & Poth, C.N. (2017) *Qualitative inquiry and research design: choosing among five approaches* (Thousand Oaks, CA: SAGE)

Derks, D., Fischer, A.H., & Bos, A.E.R. (2008) “The role of emotion in computer-mediated communication: A review” *Computers in Human Behavior, 24*, 766-785

Derks, D., Bos, A.E.R., & von Grumbkow, J. (2007) “Emoticons and social interaction on the Internet: the importance of social context” *Computers in Human Behavior, 23*, 842-849

Ekman, P. (1992) “An argument for basic emotions” *Cognition and Emotion, 6*(3/4), 169-200

Ekman, P. & Friesen, W.V. (1975) *Unmasking the face: a guide to recognizing emotions from facial clues* (Englewood Cliffs, NJ: Prentice-Hall)

Ellison, N., Heino, R., & Gibbs, J. (2006) “Managing impressions online: Self-presentation processes in the online dating environment” *Journal of Computer-Mediated Communication, 11,* 415-441

Ganster, T., Eimler, S.C., & Krämer, N.C. (2012) ”Same same but different!? The differential influence of smilies and emoticons on person perception” *Cyberpsychology, Behavior, and Social Networking, 15(4)*, 226-230

Hofstede, G. (2017) “Geert Hofstede – Country comparison” <https://geert-hofstede.com/countries.html> Retrieved on 20 August 2017

Kelly, R. & Watts, L. (2015) “Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships” *Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances and Design,* Workshop

Lu, X., Ai, W., Liu, X., Li, Q., Wang, N., Huang, G. & Mei, Q. (2016) “Learning from the ubiquitous language: an empirical analysis of emoji usage of smartphone users” *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing,* 770-780

Matsumoto, D., Frank, M.G. & Hwang, H.S. (2013) *Nonverbal Communication: Science and Applications* (California: SAGE Publications)

Matsumoto, D., Yoo, S.H. & Fontaine, J. (2008) “Mapping expressive differences around the world: the relationship between emotional display rules and individualism versus collectivism” *Journal of Cross-Cultural Psychology, 39*, 55-74

Matsumoto, D., Takeuchi, S., Andayani, S., Kouznetsova, N. & Krupp, D. (1998) “The contribution of individualism-collectivism to cross-national differences in display rules” *Asian Journal of Social Psychology, 1*, 147-165

Matsumoto, D. (1991) “Cultural influences on facial expressions of emotion” *The Southern Communication Journal, 56(2)*, 128-137

Merrell, F. (2001) “Charles Sanders Peirce’s concept of the sign” In: Cobley, P. (ed.) The Routledge companion to semiotics and linguistics (pp. 28-39). London: Routledge

Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L. & Hecht, B. (2016) “”Blissfully Happy” or “Ready to Fight”: varying interpretations of emoji” *Proceedings of the Tenth International AAAI Conference on Web and Social Media 2016*, 259-268

Novak, P.K., Smailović, J., Sluban, B. & Mozetič, I. (2015) “Sentiment of emojis” *PLoS ONE, 10(12)*, 1-22

Pavalanathan, U. & Eisenstein, J. (2016) “Emoticons vs. Emojis on Twitter: a casual inference approach” *AAAI 2016 Spring Symposium on Observational Studies through Social Media and Other Human-Generated Content*

Stokoe, W.C. (2001) “The origins of language” In: Cobley, P. (ed.) The Routledge companion to semiotics and linguistics (pp. 40-51). London: Routledge

Suttles, J. & Ide, N. (2013) “Distant supervision for emotion classification with discrete binary values” In: Gelbukh, A. (eds.) Computational linguistics and intelligent text processing (CICLing 2013), (pp. 121-136). Heidelberg, Berlin: Springer

Twitter (2016) “Twitter Company” <https://about.twitter.com/company> Retrieved 16 June 2017

Voida, A., Mynatt, E.D., Erickson, T. & Kellogg, W.A. (2004) “Interviewing over instant messaging” *CHI ’04 Extended Abstracts on Human Factors in Computing Systems*, 1344-1347

Walther, J.B. (2011) “Theories of computer-mediated communication and interpersonal relations” In: M.L.Knapp & J.A. Daly (eds.), The Sage Handbook of Interpersonal Communication (4th edition), (pp. 443 – 479). Thousand Oaks, CA: Sage Publications

Appendix

Appendix I. Lists of emoji without sentiment score that were used at least 100 times per language, their frequency and percentage of total emoji.

Emoji (Dutch)	Frequency (%) (*N* = 97784)
- 1F923	2624 (2.7%)
- 1F914	2535 (2.6%)
- 1F644	1261 (1.3%)
- 1F917	1156 (1.2%)
- 1F643	626 (0.6%)
- 1F642	480 (0.5%)
- 1F913	359 (0.4%)
© - A9	242 (0.2%)
® - AE	218 (0.2%)
- 1F922	191 (0.2%)
- 1F641	174 (0.2%)

Emoji (English)	Frequency (%) (*N* = 175684)
- 1F914	4424 (2.5%)
- 1F644	4311 (2.5%)
- 1F923	3494 (2.0%)
- 1F643	2467 (1.4%)
- 1F917	1352 (0.8%)
- 1F642	1224 (0.7%)
- 1F919	861 (0.5%)
- 1F918	855 (0.5%)
- 1F37E	672 (0.4%)
- 1F942	604 (0.4%)
- 1F922	570 (0.3%)
- 1F91E	562 (0.3%)
- 1F5A4	461 (0.3%)
- 1F913	430 (0.2%)
- 1F4F8	373 (0.2%)
- 1F924	340 (0.2%)
- 1F641	321 (0.2%)
- 1F94A	296 (0.2%)
™ - 2122	260 (0.1%)
- 1F5E3	237 (0.1%)
- 1F915	230 (0.1%)
® - AE	201 (0.1%)
- 1F927	175 (0.1%)