# The Emoji Sentiment Lexicon: Analysing Consumer Emotions in Social Media Communication

**Lisa Holthoff**
University of Duisburg-Essen

Cite as:
Holthoff Lisa (2020), The Emoji Sentiment Lexicon: Analysing Consumer Emotions in Social Media Communication. *Proceedings of the European Marketing Academy*, 49th, (63029)

Paper from the 49th Annual EMAC Conference, Budapest, May 26-29, 2020.

EUROPEAN MARKETING ACADEMY

**The Emoji Sentiment Lexicon:**
**Analysing Consumer Emotions in Social Media Communication**


**Abstract:**

Due to the ongoing digitalization of communication, many companies are increasingly interested in tools that allow them to analyze their consumers' brand-related online messages. Current sentiment analysis tools are able to analyze the emotional tone of written text automatically. However, since consumers often use emojis to express emotional subtext in social media messages, I propose the additional integration of emojis into these tools. Study 1, an online experiment with 1,519 participants, confirms that the inclusion of emojis increases the correct interpretation of brand-related tweets (with regard to polarity as well as emotion category). In study 2, a qualitative study with 1,157 participants, I develop a categorical emoji lexicon for unsupervised sentiment analyses. Finally, I validate the resulting lexicon in a quantitative study with 1,926 participants. The lexicon improves and simplifies the identification of emotional consumer reactions in social media communication.



*Keywords: Categorical Sentiment Analysis, Social Media Communication, Emoji Lexicon*

*Track: Digital Marketing & Social Media*

## 1. Introduction

The way we communicate is changing rapidly and the importance of social media communication is growing constantly. Digitalization and innovative channels of communication like Twitter, Instagram, or Snapchat simplify the sharing of information. Simultaneously, social media platforms influence the way in which language is used. Text messages become shorter and feature neologisms, slang, and abbreviations (Ghiassi, Skinner, and Zimbra, 2013; Go, Bhayani, and Huang, 2009; Kiritchenko, Zhu, and Mohammad, 2014; Mostafa, 2013). Users often express tone of voice and volume with punctuation and capitalization. In addition, since it is difficult to display gestures and facial expressions online, they increasingly use emoticons and emojis to substitute paralanguage and other nonverbal communication markers. For this reason, emojis are currently understood as one of the fastest-growing forms of communication in history. This is, for example, reflected in *Oxford Dictionaries*' decision to pick a pictograph as *Word of the Year 2015*. They chose the 'face with tears of joy' emoji 😂 because it expressed best the year's ethos, mood, and preoccupations (Oxford Dictionaries, 2015). On twitter alone, the 'Face with Tears of Joy' was used in more than 2.6 billion tweets since 2013 (Rothenberg, 2016).

Due to these changes in communication, many companies are increasingly interested in tools that allow them to analyze consumers' online messages in order to understand attitude and mood towards their brands. However, while consumers increasingly use emojis to express their emotions and the little symbols are thus an essential part of emotional subtext, most sentiment analysis tools only consider written text. I propose the additional integration of emojis into these tools and study 1 confirms that an inclusion of emojis enhances the correct interpretation of social media messages. In study 2, I develop and validate a categorical emoji lexicon for unsupervised sentiment analyses based on qualitative as well as quantitative data. This lexicon improves and simplifies the identification of emotional consumer reactions in social media communication.

## 2. Theoretical Foundation: Social Media, Emojis, and Emotion Measurement

Emojis are pictographs intended to express emotions, ideas, or activities. The little symbols emerged during the 1990s in Japan and quickly became a global phenomenon (Novak, Smailović, Sluban, and Mozetič, 2015). To represent them universally, the first emojis were added to Unicode in 2009 and today the computing industry standard includes 1,809 basic emojis (Unicode, 2019). It is assumed that emojis act as nonverbal explaining and bonding components that enable more expressive electronic communication and enhance message recipients' apprehension (Huang, Yen, and Zhang, 2008; Novak et al., 2015). In this function, 92% of the online population use emojis frequently or occasionally (Emoji Research Team, 2015). Already in March 2015, every second post on Instagram contained emojis (Dimson, 2015), in 2018 more than 700 million emojis were used in Facebook posts every day (Emojipedia, 2018), and on twitter alone, the 'Face with Tears of Joy' was used in more than 2.6 billion tweets since 2013 (Rothenberg, 2016). Gradually, all social media platforms recognized the trend towards the use of emojis in electronic communication and they even started celebrating World Emoji Day on July 17.

This development is not limited to private communication. Social media platforms offer an easy, fast, and cheap possibility to spread opinions, experiences, or sentiments towards brands. Thereby, individual consumers can strongly influence the image of and mood towards a brand or product (Jansen, Zhang, Sobel, and Chowdury, 2009). According to the idea of competitive intelligence, social media monitoring and instant identification of mood alterations are essential to gain strategic competitive advantages (Ghiassi et al., 2013; Kaplan & Haenlein, 2011). As a result, a growing number of companies have an increasing demand for emotion measurement methods that help them understand consumer sentiment in digital communication.

Such methods of emotion measurement can be divided into self-report measures and autonomic measures. Self-report measures comprise verbal self-reports, visual self-reports with picture scales, and moment-to-moment ratings. Until now, autonomic measures consider facial expressions (e.g. smiles), physiological reactions (e.g. sweat), and neuroscientific methods (e.g. brain imaging) (Poels & Dewitte, 2006). I propose measurement of digital expression of emotions as a new sub-category of autonomic emotion measurement. Emojis are increasingly used to display facial expressions like smiles or frowns as well as physiological reactions like blushing or sweating and should be treated accordingly.

Due to the phenomenon's novelty, research on emojis in online communication is still in its early stages. Only few studies explore the sentiment of social media communication such as tweets including emojis (Novak et al., 2015; Vidal, Ares, and Jaeger, 2016) and, to my best knowledge, none of these focus on brand- or marketing-related tweets. A slightly larger number of studies propose more general sentiment analyses containing emojis, but all of them focus on the polarity (i.e. positive vs. negative sentiment) of emojis (e.g. Go et al., 2009; Kiritchenko et al., 2014; Kouloumpis, Wilson, and Moore, 2011; Mohammad, Kiritchenko, and Zhu 2013; Novak et al., 2015; Thelwall, Buckley, and Paltoglou, 2011). However, to capture the complex nature of customer reactions, it might not be sufficient to divide reactions into positive, negative, and neutral sentiment. Thus, Yamamoto, Kumamoto, and Nadamoto (2015) propose a multidimensional sentiment calculation method for emoticons with the four dimensions emphasis, assuagement, conversion, and addition. However, the results of this study are subject to limitations since the experiment used to extract the sentiment of emoticons was conducted with only 10 participants. In addition, following the idea of basic and secondary emotions (Plutchik, 2003), I rather propose emotion categories like anger, sadness, or joy to reflect consumer sentiment adequately.

## 3. Empirical Investigation: The Emoji Sentiment Lexicon

### *3.1 Study 1 (pre-study): Relevance of emojis for sentiment analyses of brand-related tweets*

Consumers increasingly use emojis to express their emotions and the little symbols are thus an essential part of emotional subtext. Study 1 is an online experiment that aims at showing that an inclusion of emojis increases the correct interpretation of brand-related tweets (with regard to polarity as well as emotion).

The experimental design comprises 72 brand-related tweets, 60 containing one emoji and 12 containing two emojis. These tweets were drafted by 11 different Twitter-users and each of these tweets was based on a brand-related consumer scenario (e.g. *Imagine that you are at Starbucks and you are angry because you have to wait in line for too long.* or *You have not been to McDonald's for a while and you are happy about the improved quality.*). In five online experiments (1a-1e), participants were asked to evaluate between 10 and 15 tweets (randomly assigned and in random order) regarding the expressed polarity and emotion. Polarity was assessed as rather positive, neutral, or rather negative. The emotion had to be picked from a drop-down menu containing 14 emotions: seven emotions particularly relevant in brand communication as well as Ekman's seven basic emotions (Ekman & Cordaro, 2011). In order to identify differences in comprehension reliably, each tweet was randomly shown either with or without emojis. Table 1 gives an overview of the five experiments:

| Study | Number of tweets | Number of participants | Number of evaluated tweets | Gender | Mean age (SD) |
|---|---|---|---|---|---|
| 1a | 27 (15 randomly drawn) | 641 (328-377 per tweet) | 9,615 | 69.1% female | 28.15 (9.37) |
| 1b | 12 | 221 | 2,652 | 53.8% female | 37.74 (13.74) |
| 1c | 13 | 232 | 3,016 | 59.9% female | 27.34 (7.33) |
| 1d | 10 | 224 | 2,240 | 54.5% female | 25.65 (7.84) |
| 1e | 10 | 201 | 2,010 | 62.7% female | 27.77 (8.35) |
| Total | 72 | 1,519 | 22,533 | 62.5% female | 29.00 (10.23) |

Table 1. Overview of Experimental Design and Recruited Participants (study 1a-1e)

In the data analysis of study 1a-1e, I compared the participants' understanding of each tweet (with and without emojis) with the underlying scenario to determine the percentage of correct classifications. I then conducted group comparisons and chi-squared tests to demonstrate the relevance of emojis for a valid analysis of emotional tone in social media messages. The chi-squared tests show a significantly better understanding of the underlying emotion for 66.13% of tweets (<.005; <.001: 54.84%) as well as a significantly better assessment of polarity for 59.68% of tweets (<.005; <.001: 50%). Overall, participants correctly assessed the emotional tone (polarity as well as specific emotion) of tweets containing emojis significantly (.000) more often than the tone of tweets without emojis. I was able to replicate these results in every study.

To extend the results, I controlled for effects of the number of included emojis (one, two, and three) per tweet in a sixth experiment (study 1f) with 12 tweets and 246 participants ($m_{age}$ 28.40, SD 12.13). The understanding of polarity significantly correlated with the number of emojis in only one of the 12 tweets, the understanding of emotion in only two tweets. Therefore, study 1f indicates that the number of emojis does not significantly influence the understanding of the emotional tone of a tweet. In summary, I conclude that emojis are an essential part of emotional subtext in tweets. Especially when analyzing distinct emotions, emojis should be integrated into sentiment analyses.

### *3.2 Study 2: Emoji lexicon for sentiment analyses*

After verifying that it is important to include emojis when mining for emotions in social media messages, I aim at developing a sentiment analysis tool that considers emojis in addition to text. Sentiment analyses or opinion mining are automated analyses of verbal expressions of mood and belong to the field of text mining and natural language processing (Liu, 2015). Sentiment analyses can be divided into unsupervised lexicon-based approaches and supervised automated approaches with machine learning algorithms (Ortigosa, Martín, and Carro, 2014). Lexicon-based approaches are more flexible, but only consider emotions included in an underlying lexicon, while automated analyses are more precise, but depend on a large corpus of labelled training data and are thus time-consuming and domain-specific. I chose the lexicon-based approach in order to create a tool that is applicable in various marketing domains without depending on domain-specific training data.

In study 2, I developed an emoji lexicon for lexicon-based sentiment analyses based on a qualitative study (2a) with 1,157 participants. I validated it with a second, quantitative study

(2b) with 1,926 participants. The lexicon simplifies the identification and interpretation of emotional consumer reactions in social media communication.

Since emojis are almost exclusively used when communicating online and since I aim developing a method to analyze emotions in consumers' social media communication, my target group consists of individuals that are active in the online world. Therefore, I decided to collect data on the web and to recruit participants primarily via social media platforms. As existing studies concentrated on the polarity of emojis, I could not build upon prior research to match emojis and the emotions expressed with them. Hence, study 2a is an exploratory analysis that is qualitative in nature. Since I am interested in the expression of emotions, my research focuses on emojis. In an online questionnaire, participants had to indicate how familiar they are with emojis in general as well as how frequently and where they use them. Afterwards, I asked respondents in an open question format to state the emotion they would most likely express with a given emoji. I included all emojis that express emotions and that appeared in at least 45 million tweets since 2013 (Rothenberg, 2016). Each respondent was randomly assigned eight of the 31 relevant emojis. Subsequently, I coded all answers in a consistent format. In total, the sample of study 2a consists of 1,157 completed questionnaires. One third of participants are students, one third employees, and one third others; 58.8% of respondents are female. Respondents are between 12 and 82 years old with a mean age of 34.3 ($SD = 13.88$). Of all participants, 98.4% were familiar with emojis, 76.6% use them several times a day, and another 10.4% several times a week. Most respondents use emojis regularly in messengers (94.4%) and on social media platforms (63.9%).

Study 2a reveals that only 22.6% of emojis are clearly associated with a specific emotion. The 'Sleeping Face' 😴 for example expresses fatigue for 93.8% of respondents and boredom for the rest. Most emojis, however, are more ambiguous. Approximately half (51.6%) do at least consistently reveal positive, negative, or neutral emotions. For instance, the 'Face Blowing a Kiss' 😘 is understood as sign of positive emotions by 97.9% of consumers and displays love, gratitude, affection, or warmth, while the 'Sleepy Face' 😪 is used to express negative emotions by 93.6% of consumers and stands for sadness, illness, disappointment, or exhaustion. The understanding of the remaining 25.8% of emojis is even more heterogeneous. For example, 23.8% of participants interpreted the 'Smirking Face' 😏 negatively, 25.7% neutrally, and 50.5% positively. The emotions most often matched to the 'Smirking Face' reach from mischievousness, amusement, and self-confidence to doubt and discontent.

I confirmed these findings in an additional study. Study 2b was a confirmatory quantitative analysis of the emotions expressed using specific emojis. Of the 1,926 respondents, 73% were female, with a mean age of 29.1 ($SD = 12.78$). In an online questionnaire, the participants saw emojis and had to choose the emotion that they felt was the best match from a drop-down menu. The emotions contained in this menu were mentioned by at least 5% of participants in study 2a. When analyzing the results, I classified all emojis into three categories: (1) nine emojis had a 'clear emotional categorization' with more than 80% consistency of polarity and more than 50% frequency of the main emotion, (2) another 16 emojis had a 'clear polar categorization' with more than 80% consistency of polarity, and (3) six emojis had an 'unclear categorization'.

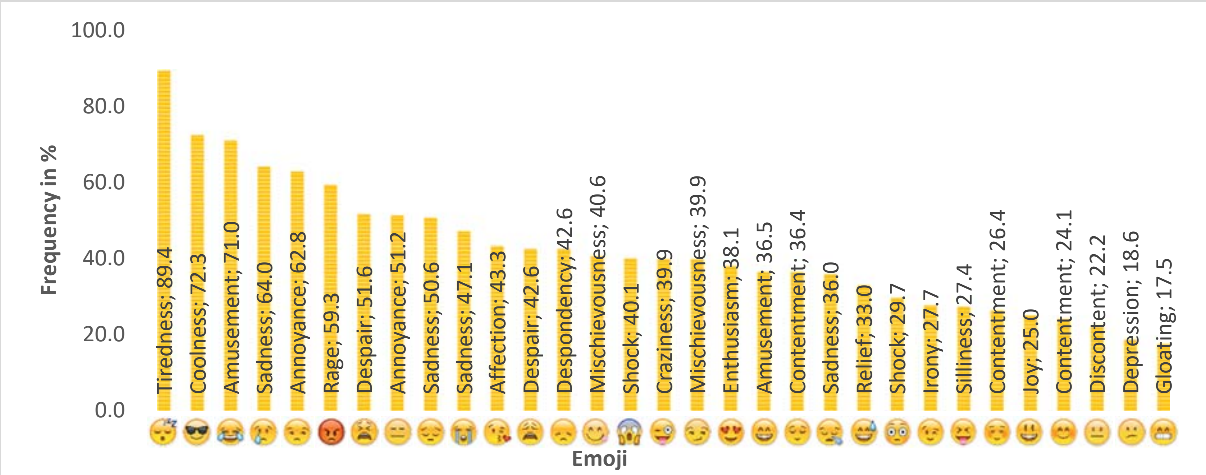

Figure 1. Most Frequently Named Emotion per Emoji

## 4. Results

The exploratory qualitative analysis in study 2a revealed that only 22.6% of emojis are clearly associated with a specific emotion in an open question format. Approximately half (51.6%) do at least consistently reveal positive, negative, or neutral emotions. The understanding of the remaining 25.8% of emojis is even less homogeneous. On the basis of the results from study 2a, I confirmed my findings in a second study. Study 2b is a confirmatory quantitative analysis of the emotions expressed through the use of specific emojis. By combining the results from study 2a and 2b, I developed a representative categorical sentiment lexicon for emojis, which can be used to identify specific consumer emotions in electronic communication. Due to the detected ambiguity concerning the emotional meaning of 71% of emojis, I propose a combined analysis of visual (emojis) and verbal (text) stimuli to factor in contextual influences and thus fully comprehend the emotional tone of electronic consumer communication. Understanding not only the polarity of consumer sentiment, but identifying emotion categories that are regularly expressed with emojis like anger, sadness, joy, or interest will enable companies to react more appropriately to electronic consumer communication. In order to display the variance in emotion interpretations, each emoji should not only correspond to the most frequently named emotion, but sentiment analyses should proportionally consider all alternative interpretations. The 'Sleeping Face' 😴 for example should include the label fatigue (93.8%) as well as the label boredom (6,2%). Such hierarchy of affective labels was already implemented successfully in verbal sentiment lexica like the WordNet-Affect lexicon (Strapparava & Valitutti, 2004).

**5. Practical Implications and Outlook**

The emoji lexicon for sentiment analyses provides an innovative tool for companies to analyze electronic word-of-mouth on social media platforms instantly, to monitor emotional brand image constantly, and to compare public mood towards their brand with mood towards their competitors. Social media data is up-to-the-minute and has the potential to show consumers' immediate reactions to specific events, marketing activities, or even scandals. In addition, the created emoji lexicon can improve companies' social media communication by revealing possible emoji interpretations, minimizing misunderstandings due to a 'wrong' use of emojis, and optimizing emoji targeting (e.g. via the corresponding twitter tool). Understanding not only the polarity of consumer sentiment, but identifying emotion categories that are regularly expressed with emojis like anger, sadness, or joy will enable companies to react more appropriately to electronic consumer communication.

One of the main limitations of the presented lexicon is the fact that the presented studies were conducted only in German. A replication of the results in other languages could give insights into intercultural generalizability. In addition, the current lexicon includes only 31 frequently used emojis. Future research should extend the number of emojis.

**References**

Dimson, T. (May 01, 2015). *Emojineering part 1: Machine learning for emoji trends*. Retrieved from http://instagram-engineering.tumblr.com/post/117889701472/ emojineering-part-1-machine-learning-for-emoji. (Last accessed: November 28, 2019).

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, 3(4), 364-370.

Emoji Research Team (September, 2015). *2015 Emoji report*. Retrieved from https://cdn.emogi.com/docs/reports/2015_emoji_report.pdf. (Last accessed: November 28, 2019).

Emojipedia (2018). *World Emoji Day Statistics*. Retrieved from https://worldemojiday.com/statistics/?fbclid=IwAR2vECl8dqKL5jAkjLFNmIwVSa3Sqt3H6EJpAl1c41oZwbtDhPk8TrK4RUI. (Last accessed: December 3, 2019).

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*, 40(16), 6266-6282.

Go, A., Bhayani, R., & Huang, L. (2009), *Twitter sentiment classification using distant supervision*. Technical report, Stanford Digital Library Technologies Project.

Huang, A. H., Yen, D. C., & Zhang, X. (2008). Exploring the potential effects of emoticons. *Information & Management*, 45(7), 466-473.

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11), 2169-2188.

Kaplan, A. M., & Haenlein, M. (2011). The early bird catches the news: Nine things you should know about micro-blogging. *Business Horizons*, 54(2), 105-113.

Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50 (May), 723-762.

Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the OMG!. In L. Adamic, R. Baeza-Yates, & S. Counts (eds.), *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 538-541). Menlo Park, AAAI.

Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. New York: Cambridge University Press.

Mohammad, S M., Kiritchenko, S., & Zhu, X. (2013). NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. In S. Manandhar, & D. Yuret (eds.), *Second Joint Conference on Lexical and Computational Semantics (*SEM), Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)* (pp.321-327). Atlanta, GA, Association for Computational Linguistics.

Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, 40(10), 4241-4251.

Novak, P. K., Smailović, J., Sluban, B., & Mozetič, I (2015). Sentiment of emojis. *PloS one*, 10 (12), e0144296.

Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*, 31, 527-541.

Oxford Dictionaries (November 17, 2015), *Word of the year 2015*. Retrieved from https://languages.oup.com/press/news/2019/7/5/WOTY. (Last accessed: November 28, 2019).

Plutchik, R. (2003). *Emotions and life: Perspectives from psychology, biology, and evolution*, Washington, DC: APA.

Poels, K., & Dewitte, S. (2006). How to capture the heart? Reviewing 20 years of emotion measurement in advertising. *Journal of Advertising Research*, 46(1), 18-37.

Rothenberg, M. (2016). *Emojitracker: Realtime emoji use on twitter*. Retrieved from http://emojitracker.com/. (Last accessed: November 28, 2019).

Strapparava, C., & Valitutti, A. (2004). WordNet-Affect: An affective extension of WordNet. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, & R. Silva (eds.), *Proceedings of the 4th International Conference on Language Resources and Evaluation* (pp. 1083-1086). Lisbon, LREC.

Thelwall, M., Buckley, K. & Paltoglou, G. (2011). Sentiment in twitter events. *Journal of the American Society for Information Science and Technology*, 62 (February), 406-418.

Unicode, Inc. (October 21, 2019). *Full Emoji List, v13.0β*. Retrieved from https://unicode.org/emoji/charts-13.0/full-emoji-list.html. (Last accessed: December 2, 2019).

Vidal, L., Ares, G., & Jaeger, S. R. (2016). Use of emoticon and emoji in tweets for food-related emotional expression. *Food Quality and Preference*, 49, 119-128.

Yamamoto, Y., Kumamoto, T, & Nadamoto, A. (2015). Multidimensional sentiment calculation method for twitter based on emojis. *International Journal of Pervasive Computing and Communications*, 11 (2), 212-232.