

## Analysing emojis in context in a corpus of Twitter data

Andrew Kehoe & Matt Gee (Birmingham City University)

Social media has become an increasingly multimodal space. Highfield & Leaver (2016: 47) explore the growing significance of emojis, images, animated GIFs and videos, noting that “visual social media content can highlight affect, political views, reactions, key information, and scenes of importance”. Emojis in particular have received increasing attention from a variety of perspectives (e.g. Dresner & Herring 2010, Ljubešić & Fišer 2016, Miller et al. 2016, Wijeratne et al. 2016, Maíz-Arévalo 2017).

In this paper we carry out a large-scale data-driven corpus pragmatic analysis of emoji use on Twitter. Our corpus contains a random selection of 40 million tweets in English and German downloaded between February and May 2018 using the Twitter API. Table 1 shows the proportion of tweets in our random sample containing multimodal elements.

Language	Total	Images	Videos	Animated GIFs	Emojis
English	38,902,024	8,625,473 (22%)	2,259,007 (6%)	530,617 (1%)	6,381,274 (16%)
German	493,183	68,205 (14%)	8,083 (2%)	3,953 (1%)	65,528 (13%)

Table 1. Frequency of multimodal elements in a random sample of tweets

Our focus in this study is on the 6 million tweets containing at least one emoji (a considerably higher proportion than those containing videos or animated GIFs). Table 2 shows the most frequent emojis found in our random sample, using the official character names as defined in the Unicode standard.

English		German	
Emoji	% of Tweets	Emoji	% of Tweets
😂 Face With Tears of Joy	4.42	😂 Face With Tears of Joy	2.49
😭 Loudly Crying Face	1.85	❤️ Red Heart	1.36
❤️ Red Heart	1.51	😍 Smiling Face With Heart-Eyes	0.89
🔥 Fire	1.31	🔥 Fire	0.57
😍 Smiling Face With Heart-Eyes	1.16	👍 Thumbs Up	0.56
💕 Two Hearts	0.61	😊 Smiling Face With Smiling Eyes	0.53
😄 Rolling on the Floor Laughing	0.51	😉 Winking Face	0.51
👏 Clapping Hands	0.49	🤔 Thinking Face	0.43
🙏 Folded Hands	0.47	😘 Face Blowing a Kiss	0.39
✨ Sparkles	0.46	😭 Loudly Crying Face	0.39

Table 2. Most frequent emojis in a random sample of tweets

Of particular interest to us are those emojis with more than one possible use. In our previous corpus pragmatic studies we used collocational analyses to search for apologies in a large corpus of blog posts and comments (Lutzky & Kehoe 2017) and to examine the use of swearwords in the same corpus (Lutzky & Kehoe 2016), finding collocational analyses to be a useful method for disambiguating different uses of the same word. We now apply the same technique to the disambiguation of emojis. For example, the ‘folded hands’ emoji can be interpreted as thanking, pleading, praying or giving a high five, and we find collocates reflecting all of these uses. We are also able to examine collocation between emojis. We find that in the vast majority of cases emojis tends to co-occur with themselves (e.g. ‘face with tears of joy’ collocates with itself 4.9 million times) but, interestingly, we also find collocation between semantically-related emojis (e.g. ‘smiling face with heart eyes’ and ‘red heart’).

Throughout the paper we use specific examples to illustrate how the words and other emojis collocating with a particular emoji offer clues about its intended meaning and illocutionary force, and we describe some of the challenges we faced, relating to collocational span and the frequent repetition of emojis within single tweets.

## References

- Dresner, E., Herring, S.C., 2010. Functions of the Nonverbal in CMC: Emoticons and Illocutionary Force. *Communication Theory* 20, 249–268.
- Highfield, T., Leaver, T., 2016. Instagrammatics and digital methods: studying visual social media, from selfies and GIFs to memes and emoji. *Communication Research and Practice* 2, 47–62.
- Ljubešić, N., Fišer, D., 2016. A Global Analysis of Emoji Usage, in: *Proceedings of the 10th Web as Corpus Workshop*.
- Lutzky, U. and Kehoe, A., 2016. “Your blog is (the) shit”: a corpus linguistic approach to the identification of swearing in computer mediated communication. *International Journal of Corpus Linguistics* 21(2), pp.165-191.
- Lutzky, U. and Kehoe, A., 2017. “I apologise for my poor blogging”: Searching for apologies in the Birmingham Blog Corpus. *Corpus Pragmatics* 1(1), pp.37-56.
- Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., Hecht, B., 2016. “Blissfully Happy” or “Ready to Fight”: Varying Interpretations of Emoji, in: *Proceedings of the Tenth International AAAI Conference on Web and Social Media*. Presented at the ICWSM.
- Maíz-Arévalo, C., 2017. Emoticons in transactional and interactional exchanges: social networking chitchat versus working negotiation. *ODISEA. Revista de estudios ingleses*.
- Wijeratne, S., Balasuriya, L., Sheth, A., Doran, D., 2016. EmojiNet: Building a Machine Readable Sense Inventory for Emoji. *Proceedings of the International Workshop on Social Informatics*.