Varying linguistic purposes of emoji in (Twitter) context

Noa Na'aman, Hannah Provenza, Orion Montoya
Brandeis University
{nnaaman,hprovenza,obm}@brandeis.edu

Abstract

Research into emoji in textual communication has, thus far, focused on high-frequency usages and the ambiguity of interpretations. Investigation of emoji uses across a wide range of uses can divide them into different linguistic functions: function and content words, or multimodal affective markers. We report on an annotation task on English Twitter data with the goal of classifying emoji usage by these categories, and on the effectiveness of a classifier trained on these annotations. We find that PART OF IT IS STRAIGHTFORWARD, but ANOTHER PART OF IT IS COMPLICATED.

1 Background

Early work on Twitter emoticons (Schnoebelen, 2012) pre-dated the wide spread of Unicode emoji on mobile and desktop devices. Schnoebelen studied the Recent work (Miller et al., 2016) has explored the

2 Annotation task

2.1 Guidelines

Although recognizing the presence of emoji characters is trivial and unambiguous, the linguistic distinctions we sought to annotate were ambiguous and seemed prone to disagreement. Therefore in the guidelines we structured the annotation process in such a way as to minimize cognitive load and lead the annotators to intuitive and natural decisions. Our task was motivated partially by the observation that emoji are used in contexts that make them graphical replacements that map to the existing lexicon, and are therefore straightforward to interpret. The guidelines build on this observation and dictate a process that aims to take advantage of such uses. The process was a flow that

presented annotators with a few simple questions at each step, to determine whether to assign a label or to move on to the next step.

Our guidelines gave a cursory background about emoji and their uses in social media, assuming no particular familiarity with the range of creative uses of emoji. In hindsight we realized that we assumed that the annotators would have a fair degree of familiarity with Twitter. The short-message social platform has many distinctive cultural and communicative codes of its own, not to mention subcultures, and continuously evolving trends combined with a long memory. As two of the authors are active and engaged users of Twitter, we unfortunately took it for granted that our annotators would be able to decipher emoji in contexts that required knowledge of InterNet language and Twitter norms.

The task included:

- Identifying each emoji in the tweet

- Deciding whether multiple contiguous emoji should be considered separately or as a group

- Choosing the right tag for the emoji (or sequence)

- Providing a translation or interpretation for each tagged span.

Eliciting an interpretation serves two main goals. First, as a coercive prompt for the user to bias them toward a linguistic interpretation. A replaceable phrase that fits with the grammar of the sentence is a different proposition than a marker that amounts to a standalone utterance such as "I am laughing" or "I am sad". Secondly, one of the eventual applications of annotated corpus may be emoji-sense disambiguation (ESD), and mapping to a lexicalized expression would be useful grounding for future ESD tasks. The text field was

very helpful during the adjudication process, clarifying the annotators' judgments and understanding of the task (when done correctly).

For each tweet, annotators were asked to do a first read without annotating anything, to get a sense of the general message of the tweet and to think about the relationship between the emoji and the text. After the first reading, they are asked to determine whether the emoji is serving as punctuation or a function word; then if it is a content word; and if it is neither of those, then to examine it as a multimodal emoji. A key test, in our opinion, was asking annotators to simulate reading the message of the tweet aloud to another person. If a listeners comprension of the core message seemed to require a word or phrase to be spoken in place of an emoji, then that would be a compelling sign that it should be tagged as function or content.

Multimodal was the category assigned to uses that failed the first two tests. We provided some guidance for deciding between 'topic', 'attitude' or 'gesture' as sub-types of the multimodal category. During adjudication it became obvious that we left this part too open to interpretation. We encountered many cases where we had trouble deciding which of two disagreeing annotators was 'right' about choosing one type over another. A clearer typology of multimodal emojis, and, if possible, a more deterministic procedure for labeling emoji with these subtypes, are the major desiderata for any future iteration of this project.

2.1.1 Later amendments to the guidelines

After the first batch of annotation work, in response to concerns and suggestions of our annotators, we made three major changes to the guidelines:

- Added the option of tagging a tweet as "out of scope," since some uses did not fit in our task description. One major example was an entire movie plot retold as a sequence of emoji characters. This is out of scope because here the emoji do not appear in a context that puts them in relation to other text, which is the phenomenon we sought to examine.

- Added co-reference tagging for topic-marker emoji that had overt co-referring words in the tweet. These are the only cases where we tagged non-emoji character spans.

- Added the option of tagging sequences of emoji that were split up by non-emoji text but should still be treated as one unit. This wasn't used at all in the following annotation batches, because the tweet format that made it necessary did not appear in the tweet sets for those batches.

2.2 Data collection and filtering

Tweets were pulled from the public Twitter streaming API using the `tweepy` library. The collected tweets were automatically filtered to include only tweets with characters from the Emoji Unicode ranges (i.e. generally `U+1FXXX`, `U+26XX–U+27BF`); only tweets labeled as being in English; to exclude tweets with embedded images or links (more below).

Tweets with links and images were excluded from consideration to reduce time investment and cognitive load for annotators. Our early explorations found frequent cases where emoji were tweeted to show a reaction to an attached image or linked page (especially a blog post or news story) and that these tended toward ambiguous interpretations akin to those found by Miller et al. A given tweet's `U+1F62D 'loudly crying face'` might be showing true sympathy, or sarcastically saying "cry my a river." The amount of annotator effort necessary for an annotator to determine this in context would require an understanding of the tweeter's past opinions and their stance on the parties involved in the story they linked. These are very interesting questions for future research, but we determined them to be out of scope for the present research, which focuses on high-level, coarse-grained distinctions.

Redundant/duplicate tweets were filtered by comparing tweet texts after removal of hashtags and @mentions; this left only a small number of cloned duplicates.

2.3 Inter-annotator agreement

We anticipated strong agreement on the marking of emoji spans. But annotators were encouraged to group multiple emoji in a single span if they were a semantic/functional unit. This exposed some differences of opinion, with one annotator tagging a pair of characters together, another separately. So the overall Krippendorff α for annotations was around 0.9.

The assignment of category labels was a bit trickier, especially given the subcategory labels. See table 1.

Dataset	# taters	span rem	total	mm	content
Week 1	4	78	0.2071	0.4251	0.1311
Week 2	4	49	0.8743	0.7158	0.8531
Week 3A	2	11	0.9096	0.4616	0.792
Week 3B	2	6	0.7436	0.3905	1.0
Week 4A	2	3	0.8789	0.4838	0.7435
Week 4B	2	1	0.3954	0.5078	1.0
Total/mean	4	150	0.6681	0.4974	0.7533

Table 1: Fleiss's κ scores and other annotation numbers

3 Machine-learning task

Our experiment was to train a sequence tagger to assign the correct linguistic-function label to an emoji character. Our annotators had assigned labels and subtypes, but

3.1 Data sparseness

For reasons described above, we had a small and arbitrary sample of emoji usage available to study. Although we tagged 775 spans in 567 tweets, we only saw 300 distinct emoji, and 135 of them occurred only once. This drove us away from

We felt that needed more complex features than we could get with an HMM, and we needed to treat features independently. So we went with Conditional Random Fields (Lafferty et al., 2001).

3.2 Feature engineering

We used CRFSuite (Okazaki, 2007) and, after experimenting with the different algorithms available, found that the averaged perceptron algorithm (**?**) gave the best results. See table 2

A standard HMM would have treated observations of the 135 nonce emoji as entirely unique and incomparable. The best thing we could likely have done with an HMM is shown on the final line of table 2: just using a tuple of `< is_emoji, POS, emoji_class >`, ignoring the individual character, yields performance that is competitive with the early iterations of the CRF feature set.

4 Future directions

We clearly need more data. 89 examples of content and functional uses of emoji is not enough to reliably model the behavior of these characters. Since our classification is mostly binary. More annotation may yield much richer models of the variety, and will help get a better handle on the range of emoji polysemy. Clustering of contents based on observed features may help induce more empirically valid subtypes than the ones defined by our specification.

This project was motivated by a curiosity, and a hypothesis, about emoji semantics. Anglophone Twitter users use emoji in their tweets for a wide range of purposes, and a given emoji character means different things in different contexts. It seems inevitable that emoji will become subject to lexicographical description. Modern lexicography has evolved from millennia of attention to words and their combinations. Every emoji linguist notes the fascinating range of pragmatic and multi-modal effects that emoji can have in electronic communication. If these effects are to be given lexicographical treatment and categorization, they must also be organized into categories that are not part of the typical range of classes used to talk about words.

We have mentioned the notion of emoji-sense disambiguation (ESD). ESD would require an empirical inventory of emoji senses, presumably from an empirical lexicon of this sort.

Even our small sample has shown a number of characters that are genuinely used both as content words and as topical or gestural cues. Sometimes emoji of the "Earth globe" (U+1F30D–U+1F30F) are used to mean 'world', and sometimes they simply decorate a tweet that is expressing an environmental message. Sometimes these three globes—one showing Earth centered on Europe/Africa, one showing Earth centered on the Americas, and one centered on Asia/Australia—are used to denote those specific regions, or sometimes they are concatenated, either for emphasis ('the whole entire world') or to indicate the passing of time. There are a number of flower emoji, and sometimes they are used to decorate a message about flowers themselves, and sometimes they add sentiment to a message—and, just as in culture

Table 2: Performance of feature iterations

feature	F1 word	F1 mm	P cont	R cont	F1 cont	Macro-avg F1
character	0.9721	0.7481	0.3571	0.3333	0.3448	0.8441
prev +emo?	0.9914	0.8649	0.4286	0.4000	0.4000	0.8783
prev +POS	0.9914	0.8784	0.5000	0.4667	0.4828	0.8921
prev +position	0.9914	0.8844	0.4667	0.4667	0.4667	0.9028
prev +contexty	0.9914	0.8831	0.6250	0.3333	0.4348	0.8848
prev +emo_class (best?)	0.9914	**0.8933**	**0.7273**	0.5333	0.6154	**0.9168**
best - character	0.9906	0.8514	0.6429	**0.6000**	0.6207	0.9090
best - contexty	0.9922	0.8750	0.4706	0.5333	0.5000	0.8945
emo?+POS+emo_class	0.9914	0.8421	0.6000	0.4000	0.4800	0.8855

away from keyboards, a rose `U+1F339` conveys a different sentiment than a sunflower `U+1F33B`.

References

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 282–289, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Hannah Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht. 2016. "Blissfully happy" or "ready to fight": Varying interpretations of emoji. In *Proceedings of the Tenth International Conference on Web and Social Media, ICWSM 2016, Cologne, Germany, May 17–20, 2016*. Association for the Advancement of Artificial Intelligence, May.

Naoaki Okazaki. 2007. Crfsuite: a fast implementation of conditional random fields (crfs).

Tyler Schnoebelen. 2012. Do you smile with your nose? Stylistic variation in Twitter emoticons. In *University of Pennsylvania Working Papers in Linguistics*, volume 18, pages 117–125. University of Pennsylvania.