SemEval 2018 Task 2: Multilingual Emoji Prediction

Francesco Barbieri$^\diamondsuit$ **Jose Camacho-Collados**♣   **Francesco Ronzano**$^\heartsuit$
Luis Espinosa-Anke♣   **Miguel Ballesteros**♠   **Valerio Basile**♥
Viviana Patti♥   **Horacio Saggion**$^\diamondsuit$

$^\diamondsuit$ Large Scale Text Understanding Systems Lab, TALN. UPF. Barcelona, Spain
♣School of Computer Science and Informatics, Cardiff University
♠IBM Research, U.S
$^\heartsuit$ Integrative Biomedical Informatics Group, GRIB, IMIM-UPF, Barcelona, Spain
♥Dipartimento di Informatica, University of Turin, Italy
$^{\diamondsuit\heartsuit}$`{name.surname}@upf.edu`, ♠`miguel.ballesteros@ibm.com`,
♣`{espinosa-ankel,camachocolladosj}@cardiff.ac.uk`,
♥`{patti,basile}@di.unito.it`

Abstract

This paper describes the results of the first shared task on Multilingual Emoji Prediction, organized as part of SemEval 2018. Given the text of a tweet, the task consists of predicting the most likely emoji to be used along such tweet. Two subtasks were proposed, one for English and one for Spanish, and participants were allowed to submit a system run to one or both subtasks. In total, 49 teams participated in the English subtask and 22 teams submitted a system run to the Spanish subtask. Evaluation was carried out *emoji-wise*, and the final ranking was based on macro F-Score. Data and further information about this task can be found at `https://competitions.codalab.org/competitions/17344`.

1 Introduction

Emojis are small ideograms depicting objects, people, and scenes (Cappallo et al., 2015). Emojis are one of the main components of a novel way of communication emerging from the advent of social media. They complement (usually) short text messages with a visual enhancement which is, as of now, a *de-facto* standard for online communication (Barbieri et al., 2017). Figure 1 shows an example of a social media message displaying an emoji.

Sometimes I think I wanna change the world... and I forget it just starts with changing me. ❤

Figure 1: Message from Twitter including a single red heart emoji.

Emojis[1] can be considered somehow an evolution of character-based *emoticons* (Pavalanathan and Eisenstein, 2015), and currently they represent a widespread and pervasive global communication device largely adopted by almost any social media service and instant messaging platforms.

Any system targeting the task of modeling social media communication is expected to tackle the usage of emojis. In fact, their semantic load is sufficiently rich that oversimplifying them to *sentiment carriers* or *boosters* would be to neglect the semantic richness of these ideograms, which in addition to mood (😟) include in their vocabulary references to food (🍘), sports (🎿), scenery (🏖), etc[2]. In general, however, effectively predicting the emoji associated with a piece of content may help to improve different NLP tasks (Novak et al., 2015), such as information retrieval, generation of emoji-enriched social media content, suggestion of emojis when writing text messages or sharing pictures online. Given that emojis may also mislead humans (Barbieri et al., 2017; Miller et al., 2017), the automated prediction of emojis may help to achieve better language understanding. As a consequence, by modeling the semantics of emojis, we can improve highly-subjective tasks like sentiment analysis, emotion recognition and irony detection (Felbo et al., 2017).

In this context, Barbieri et al. (2017) introduced the task of emoji prediction in Twitter by training several models based on bidirectional Long Short-Term Memory networks (LSTMs) (Graves, 2012), and showing they can outperform humans in solv-

---

[1]`https://unicode.org/emoji/charts/full-emoji-list.html`
[2]`https://unicode.org/emoji/charts/emoji-ordering.html`

24

ing the same task. These promising results motivated us to propose the first shared task on Multilingual Emoji Prediction. Following the experimental setting proposed by Barbieri et al. (2017), the task consists of predicting most likely emoji associated of a given text-only Twitter message. Only tweets with a single emoji are included in the task datasets (trial, train and test sets), so that the challenge can be cast as a single label classification problem.

In this paper, we first motivate and describe the main elements of this shared task (Section 2 and 3). Then, we cover the dataset compilation, curation and release process (Section 4). In Section 5 we detail the evaluation metrics and describe the overall results obtained by participating systems. Finally, we wrap this task description paper up with the main conclusions drawn from the organization of this challenge, as well as outlining potential avenues for future work, in Section 6.

2   Related Work

Modeling the semantics of emojis, and their applications thereof, is a relatively novel research problem with direct applications in any social media task. By explicitly modeling emojis as self-containing semantic units, the goal is to alleviate the lack of an associated grammar. This context, which makes it difficult to encode a clear and univocous single meaning for each emoji, has given rise to work considering emojis as function words or even affective markers (Na'aman et al., 2017), potentially affecting the overall semantics of longer utterances like sentences (Monti et al., 2016; Donato and Paggio, 2017).

The polysemy of emoji has been explored userwise (Miller et al., 2017), location-wise, specifically in countries (Barbieri et al., 2016b) and cities (Barbieri et al., 2016a), gender-wise, time-wise (Barbieri et al., 2018b; Chen et al., 2017), and even device-wise, due to the fact that emojis may have different pictorial characteristics (and therefore, different interpretations), depending on the device (e.g., Iphone, Android, Samsung, etc.) or *app* (Whatsapp, Twitter, Facebook, and so forth)[3] (Tigwell and Flatla, 2016; Miller et al., 2016).

---

[3] The image that represents the same emoji can vary, e.g., for the emoji U+1F40F, the following are over different renderings by platform in Unicode v11 (up to April 2018): Apple 🐏, Google 🐑, Twitter 🐑, EmojiOne 🐏, Facebook 🐑, Samsung 🐑, Windows 🐑.

Today, modeling emoji semantics via vector representations is a well defined avenue of work. Contributions in this respect include models trained on Twitter data (Barbieri et al., 2016c), Twitter data together with the official unicode description (Eisner et al., 2016), or using text from a popular keyboard app Ai et al. (2017). In the latter contribution it is argued that emojis used in an affective context are more likely to become popular, and in general, the most important factor for an emoji to become popular is to have a clear meaning. In fact, the area of emoji vector evaluation has also experienced a significant growth as of recent. For instance, Wijeratne et al. (2017a) propose a platform for exploring emoji semantics. Further studies on evaluating emoji semantics may now be carried out by leveraging two recently introduced datasets with pairwise emoji similarity, with human annotations, namely EmoTwi50 (Barbieri et al., 2016c) and EmoSim508 (Wijeratne et al., 2017b). In the application avenue, emoji similarity has been studied for proposing efficient keyboard emoji organization, essentially for placing similar emojis close in the keyboard (Pohl et al., 2017).

An aspect related with emoji semantic modeling in which awareness is increasing dramatically is the inherent bias existing in these representations. For example, Barbieri and Camacho-Collados (2018) show that emoji modifiers can affect the semantics of emojis (they looked specifically into skin tones and gender). This recent line of research has also been explored in Robertson et al. (2018) who argue, for example, that users with darker-skinned profile photos employ skin modifiers more often than users with lighter-skinned profile photos, and that the vast majority of skin tone usage matches the color of a user's profile photo.

The application of well defined emoji representations in extrinsic tasks is, an open area of research. A natural application, however, lies in the context of sentiment analysis. This has fostered research, for example, in creating sentiment lexicons for emojis (Novak et al., 2015; Kimura and Katsurai, 2017; Rodrigues et al., 2018), or in studying how emojis may be used to retrieve tweets with specific emotional content (Wood and Ruder, 2016). Moreover, Hu et al. (2017) study how emojis affect the sentiment of a text message, and show that not all emojis have the same impact. Finally, the fact that emojis carry sentiment

and emotion information is verified in the study by Felbo et al. (2017), where an emoji prediction classifier is used as pre-trained system, and then is fine-tuned for predicting sentiment, emotions and irony.

The last item to be covered in this review involves multimodality. Recently, emojis have been also studied from a prism where visual signals are incorporated, taking advantage of existing social media platforms like Instagram, with a strong focus on visual content. Recent contributions show that the usage of emojis depends on both textual and visual content, but seem to agree in that, in general, textual information is more relevant for the task of emoji prediction (Cappallo et al., 2015, 2018; Barbieri et al., 2018a).

3 Task Description

Given a text message including an emoji, the emoji prediction task consists of predicting that emoji by relying exclusively on the textual content of that message. In particular, in this task we focused on the one emoji occurring inside tweets, thus relying on Twitter data.

> Last hike in our awesome camping weekend! 😁

Figure 2: Example of tweet with an emoji at the end, considered in the emoji prediction task.

The task is divided into two subtasks respectively dealing with the prediction of the emoji associated to English and Spanish tweets. The motivation for providing a multilingual setting stems from previous findings about the idiosyncrasy of use of emojis across languages (Barbieri et al., 2016b) (see Figure 3): one emoji may be used with completely different meanings depending not only on the language of the speaker, but also on regional dialects (Barbieri et al., 2016a).

For each subtask we selected the tweets that included one of the twenty emojis that occur most frequently in the Twitter data we collected (Table 1). Therefore, the task can be viewed as a multi-label classification problem with twenty labels.

Twitter datasets were shared among participants by providing a list of tweet IDs[4] or directly the

---

> It's flipping hot out here! 🔥

> Iniciamos el nuevo año con ilusión! 🔥

Figure 3: Example of distinct use of the fire emoji across languages: the first tweet (English) comments on the torrid weather, while the second one (Spanish) exploits the same emoji to wish an happy new year ('We start the new year with enthusiasm!').

English

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
❤️	😍	😂	💕	🔥	😊	😎	✨	💙	😘	📷	🇺🇸	☀️	💜	😉	💯	😁	🎄	📷	😜

Spanish

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
❤️	😍	😂	💕	😊	😘	💪	😉	😋	🇪🇸	😎	💙	💜	😛	💞	✨	🎶	😘	😁	-

Table 1: The 20 most frequent emojis of each language (due to a data processing issue we only considered 19 emojis in the Spanish task).

text of each tweet. The last approach was adopted to share the test sets (more details are provided in Section 4).

4 Task Data

The data for the task consists of a list of tweets associated with a given emoji (i.e. label). As explained in the previous section, the dataset includes tweets that contain one and only one emoji, of the 20 most frequent emojis. We split the data in trial[5], training and test data. The quantity of tweets per set is displayed in Table 2.

The tweets were retrieved with the Twitter APIs and geolocalized in United States and Spain for subtasks 1 and 2, respectively. As for the trial and training data, the tweets were gathered from October 2015 to February 2017, whereas for the test data we decided to gather the tweets corresponding to the last months until the evaluation period started (from May 2017 to Jan 2018). This would prevent participants from gathering these tweets before-hand and also would enable us to test the emoji prediction task on a more realistic setting, as the test data is subsequent to the training data.

	Trial	Training	Test
English	50,000	500,000	50,000
Spanish	10,000	100,000	10,000

Table 2: Number of tweets for trial, training and test for each of the subtasks.

5 Evaluation

This section introduces the overall evaluation setting of this shared task. We first describe briefly the evaluation metrics used and then provide a succinct description of the baseline system.

5.1 Evaluation Metrics

As this was a single label classification problem, the classic *precision* (Prec.), *recall* (Recall), *f-score* (F1) and *accuracy* (Acc.) were used as official evaluation metrics. Note that because of the skewed distribution of the label set we opted for macro average over all labels.

5.2 Baseline

The baseline system for this task was a classifier based on FastText[6] (Joulin et al., 2017). Given a set of N documents, the loss that the model attempts to minimize is the negative log-likelihood over the labels (in our case, the emojis):

$$ loss = -\frac{1}{N} \sum_{N}^{n=1} e_n \log(softmax(BA_{x_n})) $$

where e_n is the emoji included in the n-th Twitter post, represented as hot vector, and used as label. Hyperparameters were set as default[7].

5.3 Participant Systems

Due to the overwhelming number of participants, we cannot describe all systems.[8] We do, however,

---

[6]github.com/facebookresearch/fastText
[7]https://github.com/facebookresearch/fastText#full-documentation
[8]This is the list of systems that ranked below the baseline in either of the subtasks: #TeamINF (Ribeiro and Silva, 2018), CENNLP (J R et al., 2018), DUTH (Effrosynidis et al., 2018), ECNU (Lu et al., 2018), EICA (Xie and Song, 2018), EPUTION (Zhou et al., 2018), LIS (Guibon et al., 2018), Manchester Metropolitan (Gerber and Shardlow, 2018), Peperomia (Chen et al., 2018), PickleTeam! (Groot et al., 2018), Shi (Shiyun et al., 2018), SyntNN (Zanzotto and Santilli, 2018), TAJJEB (Basile and Lino, 2018), The Dabblers (Alexa et al., 2018), THU_NGN (Wu et al., 2018), Tweety (Kopev et al., 2018), UMDSub (Wang and Pedersen, 2018), YNU-HPCC (Wang et al., 2018). Note that some participants did not submit a final paper but they are included in the results table.

briefly mention the main features of some significant systems ranked above the baseline in either of the subtasks.

- **Tübingen-Oslo** (Çöltekin and Rama, 2018). This supervised system consists of an SVM classifier with bag-of-n-grams features (both characters and words). *Tübingen-Oslo* is the top performing system in both tasks.

- **NTUA-SLP** (Baziotis et al., 2018). This system uses a Bi-LSTM with attention, and pre-trained word2vec vectors. They used external resources for associating each tweet with information on emotions, concreteness, familiarity, and others. They only participated in the English subtask but they classified second (according to the F1 score) with the highest recall.

- **EmoNLP** (Liu, 2018). This system is based on a Gradient Boosting Regression Tree Approach combined with a Bi-LSTM on character and word ngrams. It is complemented with several lexicons as well as learning sentiment specific word embeddings.

- **UMDuluth-CS8761** (Beaulieu and Asamoah Owusu, 2018) This supervised system combines an SVM with a bag-of-words approach for extracting salient features. This is one of the most competitive systems with the highest precision in English and the third best result in Spanish.

- **Hatching Chick** (Coster et al., 2018). This system builds an SVM classifier (with gradient descent optimization) on words and character ngrams. They obtained the second best result in the Spanish subtask, but their English system performed worse than the baseline.

- **TAJJEB** (Basile and Lino, 2018). This system made use of an SVM classifier over wide variety of features such as tf-idf, part-of-speech tags and bigrams. The system was competitive on both languages, outperforming the baseline on the Spanish dataset.

- **Duluth UROP** (Jin and Pedersen, 2018). This system consists of a soft voting ensemble approach combining different machine learning algorithms (Naïve Bayes, Lo-

gistic Regression, Random Forests, etc.). Infrequent classes are oversampled using the SMOTE algorithm. As for features, they use both unigrams and bigrams.

English			Spanish		
Emo	**F1**	**%**	**Emo**	**F1**	**%**
❤️	87.8	21.6	❤️	69.6	21.4
😍	37.8	9.7	😍	37.3	14.1
😂	47.1	9.1	😂	53.4	15
💕	26.9	5.2	💕	8.5	3.5
🔥	55.5	7.4	😊	14.9	5.1
😊	16.2	3.2	😘	26.9	4
😎	22.6	4	💪	39.8	3.1
✨	36.2	5.5	😉	16.3	4.5
💙	24	3.1	👌	13	1.8
😘	22.2	2.4	🇪🇸	49.9	4.2
📷	40	2.9	😎	14.7	3.4
🇺🇸	64.7	3.9	💙	14.2	4.1
☀️	63.7	2.5	💜	6.8	2.4
💜	17.1	2.2	😜	7.7	2.7
😉	13	2.6	💞	5.6	0.9
💯	29.2	2.5	✨	20	4.2
😄	14.3	2.3	🎶	23.7	2.1
🎄	73.6	3.1	💕	8.6	1.3
📸	38.4	4.8	😄	5.1	2.1
😜	9	2	-	-	-

Table 4: Best F1 measure (among all the teams) for each emoji in English (20) and Spanish (19). We also report the relative frequency percentage of each emoji in the test set.

5.4 Results

Each system was evaluated according to its capacity to perform well across all emojis under consideration. As mentioned, and due to the skewed distribution of the label set, we evaluated each participating system according to Macro F-Score (F1).

The overall results are provided in Table 3, and already several interesting conclusions can be drawn from them. For instance, it is noteworthy the fact that the best systems for both subtasks are more than 10 points apart (English better), which suggests that a *one-size-fits-all* model may be suboptimal for this task, and that indeed the

particularities of each individual language should be taken into consideration for best performance. The most precise systems were *EmoNLP* and *Tübingen-Oslo*, whereas the highest Recall was obtained by *NTUA-SLP* and again *Tübingen-Oslo* (English and Spanish respectively, in both cases). Clearly, the *Tübingen-Oslo* system shows a fine balance between precision and recall, perhaps due to its little preprocessing, fine-tuning and reliance on external libraries. It seems reasonable to assume, thus, that combining word and ngram embeddings as features, with SVMs and NN classifiers, provides a robust and high performing architecture for emoji prediction, with the added value of being resource/knowledge agnostic.

5.5 Analysis

This evaluation is finally complemented with the overall emoji-wise performance across all systems (Table 4). The lexical notion of *near synonymy* seems to clearly apply to emojis as well, as we can clearly see a worse performance on those emojis which are pictorically similar (e.g., the photo camera with and without flash, or the expected confusion between least frequent hearts and the red heart, which accounts for over 20% of the whole label set in the test data).

Finally, emojis with several interpretations and less frequent seem to be much more difficult to predict (e.g., the face 😜 in the English and Spanish dataset, and 😄 in the Spanish dataset). Zhou et al. (2018) showed in their system description paper how exploiting user-specific features may provide significance performance boosts.[9] This additional user-specific information may clearly help in these difficult cases which proved to be hard for all systems.

6 Conclusions

In this paper we have described the SemEval 2018 shared task in multilingual emoji prediction. The task, consisting in predicting the most likely emoji given the text of a tweet, was well received, with almost 50 system runs submitted to the English subtask and more than 20 to the Spanish subtask. One of the main conclusions that can be drawn is that the baseline we used (FastText) was highly competitive, with only 6 and 5 system runs performing better in English and Spanish.

---

[9]The use of user-specific data was not allowed by the main competition regulations and therefore none of the systems in the final ranking made use of it.

ENGLISH					SPANISH				
Team	F1	Prec.	Recall	Acc.	Team	F1	Prec.	Recall	Acc.
Tübingen-Oslo	**35.99**	36.55	36.22	47.09	Tübingen-Oslo	**22.36**	**23.49**	**22.80**	**37.27**
NTUA-SLP	35.36	34.53	**38.00**	44.74	Hatching Chick	18.73	20.66	19.16	37.23
hgsgnlp	34.02	35	33.57	45.55	UMDuluth-CS8761	18.18	19.02	18.6	34.83
EmoNLP	33.67	39.43	33.7	**47.46**	TAJJEB	17.08	18.99	20.36	25.13
ECNU	33.35	35.17	33.11	46.3	Duluth UROP	16.75	17.11	18.1	28.51
UMDuluth-CS8761	31.83	**39.80**	31.37	45.73	BASELINE	16.72	16.84	17.52	31.63
BASELINE	30.98	30.34	33	42.56	Nova	16.7	17.2	17.07	26.50
THU_NGN	30.25	31.85	29.81	42.18	ECNU	16.41	16.91	16.48	30.82
TAJJEB	30.13	29.91	33.02	38.09	MMU - Computing	16.34	17.83	16.4	28.92
EmojiIt	29.5	35.17	29.91	39.21	PickleTeam!	15.86	17.57	16.76	29.70
Reborn	29.24	33.67	28.94	42.43	ART @ Tor Vergata	14.91	15.81	15.51	30.68
freeze	29.13	31.54	29.23	37.14	CENNLP	14.68	16.32	16.2	34.85
csy	28.93	31.12	29	36.85	YNU-HPCC	14.25	17.51	15.98	31.19
Nova	27.89	28.49	28.2	34.83	Amrita_CEN_NLP1	12.13	12.46	12.41	21.64
Sheffield	27.18	28.57	26.61	37.69	erai	11.36	12.72	11.39	23.38
YNU-HPCC	26.89	26.97	29.71	32.53	Lips Eggplant	10.89	15.78	10.62	23.88
mboyanov	26.77	32.82	27.42	36.79	thelonewolf190694	10.87	11.13	12.55	27.04
kaka manData	26.59	30	26.97	36.34	The Dabblers	9.2	17.28	9.92	27.72
Duluth UROP	26.59	27.18	27.87	33.8	LIS	8.81	15.16	10.14	28.53
CENNLP	26.45	31.62	26.87	41.18	jogonba2	7.99	17.81	9.85	29.99
UMDSub	25.99	33.01	26.71	41	hjpwhu	3.9	7.46	6.81	13.81
THU_HCSI	25.83	32.38	25.9	35.34					
Peperomia	25.68	28.98	26.04	35.34					
MMU - Computing	24.98	28.94	25.04	34.59					
NoEmotionsAttached	23.3	25.27	24.47	32.76					
PickleTeam!	22.86	26.17	24.37	34.09					
Reborn	21.97	26.52	22.06	30.64					
PALM_gzy	21.97	26.52	22.06	30.64					
#TeamINF	21.5	26.21	20.84	31.59					
Hatching Chick	21.44	25.97	21.48	36.52					
CORAL	21.35	32.82	22.48	34.05					
Meisele	20.02	25.74	19.54	30.71					
erai	19.96	22.1	19.62	28.36					
SBIG	19.44	25.41	16.12	19.84					
The Dabblers	18.92	25.02	18.96	30.45					
ART @ Tor Vergata	18.39	24.49	17.25	29.45					
Amrita_CEN_NLP1	17.96	19.47	17.75	24.41					
Lips Eggplant	17.69	21.81	17.19	26.81					
XSSX	16.45	31.56	16.77	30.99					
Kno.e.sis	14.42	18.72	18.49	18.99					
thelonewolf190694	14.21	13.66	17.35	30.7					
LIS	13.53	25.58	14.14	29.42					
uaic2018	11.06	13.65	11.24	19.61					
jogonba2	8.52	24.16	9.51	25.6					
SBIG2	6.44	18.76	8.49	12.64					
alsu_wh	3.73	4.38	5.06	9.83					
Innovating world	3.09	18.47	5.73	22.74					
hjpwhu	2.04	2.63	3.22	3.92					

Table 3: Ranking of the participating systems by precision, recall, F1 and accuracy for the English track and the Spanish track. Those above the horizontal line ranked above the task baseline.

In terms of participating systems, and according to the post-participation survey the participants completed, we can see a high prevalence of neural approaches, with only 9 systems opting for more traditional linear models (6 SVMs, 3 Random Forests). Among the chosen neural architectures, LSTMs and CNNs are by far the preferred ones. It is noteworthy, however, the excellent performance of SVMs as used in the best performing system on both English and Spanish datasets.

This task has set the foundations for upcoming work on modeling emoji semantics, first, by providing a standardized testbed for emoji prediction in two languages, and second, by providing a comprehensive evaluation with a wide range of ideas, which we hope are of use for future research. Emojis, undoubtedly, are becoming increasingly important in understanding social media communication and in human-computer interaction, and thus we believe the problem of modeling emoji semantics can be further extended as follows. (1) Leveraging multimodal information (e.g., associated images (Barbieri et al., 2018a)); (2) incorporating more and more diverse languages (one step in this direction will be the re-run of this task for Italian at the Evalita 2018 evaluation campaign[10]); and (3) considering individual and communicative contexts for overall performance improvements.

Acknowledgments

References

Wei Ai, Xuan Lu, Xuanzhe Liu, Ning Wang, Gang Huang, and Qiaozhu Mei. 2017. Untangling emoji popularity through semantic embeddings. In *ICWSM*, pages 2–11.

Larisa Alexa, Alina Lorent, Daniela Gifu, and Diana Trandabat. 2018. The dabblers at semeval-2018 task 2: Multilingual emoji prediction. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 402–406, New Orleans, Louisiana. Association for Computational Linguistics.

Francesco Barbieri, Miguel Ballesteros, Francesco Ronzano, and Horacio Saggion. 2018a. Multimodal emoji prediction. In *Proceedings of NAACL: Short Papers*, New Orleans, US. Association for Computational Linguistics.

Francesco Barbieri, Miguel Ballesteros, and Horacio Saggion. 2017. Are emojis predictable? In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 105–111, Valencia, Spain. Association for Computational Linguistics.

Francesco Barbieri and Jose Camacho-Collados. 2018. How Gender and Skin Tone Modifiers Affect Emoji Semantics in Twitter. In *Proceedings of the 7th Joint Conference on Lexical and Computational Semantics (*SEM 2018)*, New Orleans, LA, United States.

Francesco Barbieri, Luis Espinosa-Anke, and Horacio Saggion. 2016a. Revealing patterns of Twitter emoji usage in Barcelona and Madrid. *Frontiers in Artificial Intelligence and Applications. 2016;(Artificial Intelligence Research and Development) 288: 239-44.*

Francesco Barbieri, German Kruszewski, Francesco Ronzano, and Horacio Saggion. 2016b. How cosmopolitan are emojis?: Exploring emojis usage and meaning over different languages with distributional semantics. In *Proceedings of the 2016 ACM on Multimedia Conference*, pages 531–535. ACM.

Francesco Barbieri, Luis Marujo, William Brendel, Pradeep Karuturim, and Horacio Saggion. 2018b. Exploring Emoji Usage and Prediction Through a Temporal Variation Lens. In *1st International Workshop on Emoji Understanding and Applications in Social Media (at ICWSM 2018)*.

Francesco Barbieri, Francesco Ronzano, and Horacio Saggion. 2016c. What does this emoji mean? a vector space skip-gram model for Twitter emojis. In *Proc. of LREC 2016*.

Angelo Basile and Kenny W. Lino. 2018. Tajjeb at semeval-2018 task 2: Traditional approaches just do the job with emoji prediction. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 467–473, New Orleans, Louisiana. Association for Computational Linguistics.

Christos Baziotis, Athanasiou Nikolaos, Athanasia Kolovou, Georgios Paraskevopoulos, Nikolaos Ellinas, and Alexandros Potamianos. 2018. Ntua-slp at semeval-2018 task 2: Predicting emojis using rnns with context-aware attention. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 435–441, New Orleans, Louisiana. Association for Computational Linguistics.

Jonathan Beaulieu and Dennis Asamoah Owusu. 2018. Umduluth-cs8761 at semeval-2018 task 2: Emojis: Too many choices? In *Proceedings of The*

12th International Workshop on Semantic Evaluation, pages 397–401, New Orleans, Louisiana. Association for Computational Linguistics.

Spencer Cappallo, Thomas Mensink, and Cees GM Snoek. 2015. Image2emoji: Zero-shot emoji prediction for visual media. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 1311–1314. ACM.

Spencer Cappallo, Stacey Svetlichnaya, Pierre Garrigues, Thomas Mensink, and Cees GM Snoek. 2018. The new modality: Emoji challenges in prediction, anticipation, and retrieval. *arXiv preprint arXiv:1801.10253*.

Jing Chen, Dechuan Yang, Xilian Li, Wei Chen, and Tengjiao Wang. 2018. Peperomia at semeval-2018 task 2: Vector similarity based approach for emoji prediction. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 425–429, New Orleans, Louisiana. Association for Computational Linguistics.

Zhenpeng Chen, Xuan Lu, Sheng Shen, Wei Ai, Xuanzhe Liu, and Qiaozhu Mei. 2017. Through a gender lens: An empirical study of emoji usage over large-scale android users. *arXiv preprint arXiv:1705.05546*.

Çağrı Çöltekin and Taraka Rama. 2018. Tübingen-oslo at semeval-2018 task 2: Svms perform better than rnns in emoji prediction. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 32–36, New Orleans, Louisiana. Association for Computational Linguistics.

Jol Coster, Reinder Gerard van Dalen, and Nathalie Adrinne Jacqueline Stierman. 2018. Hatching chick at semeval-2018 task 2: Multilingual emoji prediction. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 442–445, New Orleans, Louisiana. Association for Computational Linguistics.

Giulia Donato and Patrizia Paggio. 2017. Investigating redundancy in emoji use: Study on a Twitter based corpus. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 118–126.

Dimitrios Effrosynidis, Georgios Peikos, Symeon Symeonidis, and Avi Arampatzis. 2018. Duth at semeval-2018 task 2: Emoji prediction in tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 463–466, New Orleans, Louisiana. Association for Computational Linguistics.

Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bošnjak, and Sebastian Riedel. 2016. emoji2vec: Learning emoji representations from their description. *arXiv preprint arXiv:1609.08359*.

Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *Proc. of EMNLP 2017*.

Luciano Gerber and Matthew Shardlow. 2018. Manchester metropolitan at semeval-2018 task 2: Random forest with an ensemble of features for predicting emoji in tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 488–493, New Orleans, Louisiana. Association for Computational Linguistics.

Alex Graves. 2012. *Supervised Sequence Labelling with Recurrent Neural Networks*, volume 385 of *Studies in Computational Intelligence*. Springer.

Daphne Groot, Rémon Kruizinga, Hennie Veldthuis, Simon de Wit, and Hessel Haagsma. 2018. Pickleteam! at semeval-2018 task 2: English and spanish emoji prediction from tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 451–455, New Orleans, Louisiana. Association for Computational Linguistics.

Gaël Guibon, Magalie Ochs, and Patrice Bellot. 2018. Lis at semeval-2018 task 2: Mixing word embeddings and bag of features for multilingual emoji prediction. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 499–503, New Orleans, Louisiana. Association for Computational Linguistics.

Tianran Hu, Han Guo, Hao Sun, Thuy-vy Thi Nguyen, and Jiebo Luo. 2017. Spice up Your Chat: The Intentions and Sentiment Effects of Using Emoji. *Proc. of ICWSM 2017*.

Naveen J R, Hariharan V, Barathi Ganesh H. B., Anand Kumar M, and Soman K P. 2018. Cennlp@semeval-2018 task 2: Enhanced distributed representation of text using target classes for emoji prediction representation. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 483–487, New Orleans, Louisiana. Association for Computational Linguistics.

Shuning Jin and Ted Pedersen. 2018. Duluth urop at semeval-2018 task 2: Multilingual emoji prediction with ensemble learning and oversampling. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 479–482, New Orleans, Louisiana. Association for Computational Linguistics.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of tricks for efficient text classification. In *European Chapter of the Association for Computational Linguistics*, Valencia, Spain.

Mayu Kimura and Marie Katsurai. 2017. Automatic construction of an emoji sentiment lexicon. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pages 1033–1036. ACM.

Daniel Kopev, Atanas Atanasov, Dimitrina Zlatkova, Momchil Hardalov, Ivan Koychev, Ivelina Nikolova, and Galia Angelova. 2018. Tweety at semeval-2018 task 2: Predicting emojis using hierarchical attention neural networks and support vector machine. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 494–498, New Orleans, Louisiana. Association for Computational Linguistics.

Man Liu. 2018. Emonlp at semeval-2018 task 2: English emoji prediction with gradient boosting regression tree method and bidirectional lstm. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 387–391, New Orleans, Louisiana. Association for Computational Linguistics.

Xingwu Lu, Xin Mao, Man Lan, and Yuanbin Wu. 2018. Ecnu at semeval-2018 task 2: Leverage traditional nlp features and neural networks methods to address twitter emoji prediction task. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 430–434, New Orleans, Louisiana. Association for Computational Linguistics.

Hannah Miller, Daniel Kluver, Jacob Thebault-Spieker, Loren Terveen, and Brent Hecht. 2017. Understanding emoji ambiguity in context: The role of text in emoji-related miscommunication. In *11th International Conference on Web and Social Media, ICWSM 2017*. AAAI Press.

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. *Proc. of ICWSM16*.

Johanna Monti, Federico Sangati, Francesca Chiusaroli, Martin Benjamin, and Sina Mansour. 2016. Emojitalianobot and emojiworldbot - new online tools and digital environments for translation into emoji. In *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016), Napoli, Italy, December 5-7, 2016.*, volume 1749 of *CEUR Workshop Proceedings*.

Noa Na'aman, Hannah Provenza, and Orion Montoya. 2017. Varying linguistic purposes of emoji in (Twitter) context. In *Proceedings of ACL 2017, Student Research Workshop*, pages 136–141.

Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. 2015. Sentiment of emojis. *PloS one*, 10(12):e0144296.

Umashanthi Pavalanathan and Jacob Eisenstein. 2015. Emoticons vs. emojis on Twitter: A causal inference approach. *arXiv preprint arXiv:1510.08480*.

Henning Pohl, Christian Domin, and Michael Rohs. 2017. Beyond just text: Semantic emoji similarity modeling to support expressive communication. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 24(1):6.

Alison Ribeiro and Ndia Silva. 2018. #teaminf at semeval-2018 task 2: Emoji prediction in tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 412–415, New Orleans, Louisiana. Association for Computational Linguistics.

Alexander Robertson, Walid Magdy, and Sharon Goldwater. 2018. Self-Representation on Twitter Using Emoji Skin Color Modifiers. *Proc. of ICWSM 2018*.

David Rodrigues, Marília Prada, Rui Gaspar, Margarida V Garrido, and Diniz Lopes. 2018. Lisbon emoji and emoticon database (leed): Norms for emoji and emoticons in seven evaluative dimensions. *Behavior research methods*, pages 392–405.

Chen Shiyun, Wang Maoquan, and He Liang. 2018. Shi at semeval-2018 task 2: An effective attention-based recurrent neural network model for emoji prediction with characters gated words. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 420–424, New Orleans, Louisiana. Association for Computational Linguistics.

Garreth W Tigwell and David R Flatla. 2016. Oh that's what you meant!: reducing emoji misunderstanding. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 859–866. ACM.

Nan Wang, Jin Wang, and Xuejie Zhang. 2018. Ynu-hpcc at semeval-2018 task 2: Multi-ensemble bi-gru model with attention mechanism for multilingual emoji prediction. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 456–462, New Orleans, Louisiana. Association for Computational Linguistics.

Zhenduo Wang and Ted Pedersen. 2018. Umdsub at semeval-2018 task 2: Multilingual emoji prediction multi-channel convolutional neural network on subword embedding. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 392–396, New Orleans, Louisiana. Association for Computational Linguistics.

Sanjaya Wijeratne, Lakshika Balasuriya, Amit Sheth, and Derek Doran. 2017a. Emojinet: An open service and api for emoji sense discovery. *International AAAI Conference on Web and Social Media (ICWSM 2017). Montreal, Canada*.

Sanjaya Wijeratne, Lakshika Balasuriya, Amit Sheth, and Derek Doran. 2017b. A semantics-based measure of emoji similarity. *International Conference on Web Intelligence (Web Intelligence 2017). Leipzig, Germany*.

Ian Wood and Sebastian Ruder. 2016. Emoji as emotion tags for tweets. *Emotion and Sentiment Analysis Workshop, LREC*.

Chuhan Wu, Fangzhao Wu, Sixing Wu, Zhigang Yuan, Junxin Liu, and Yongfeng Huang. 2018. Thu_ngn at semeval-2018 task 2: Residual cnn-lstm network with attention for english emoji prediction. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 407–411, New Orleans, Louisiana. Association for Computational Linguistics.

Yufei Xie and Qingqing Song. 2018. Eica team at semeval-2018 task 2: Semantic and metadata-based features for multilingual emoji prediction. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 416–419, New Orleans, Louisiana. Association for Computational Linguistics.

Fabio Massimo Zanzotto and Andrea Santilli. 2018. Syntnn at semeval-2018 task 2: is syntax useful for emoji prediction? embedding syntactic trees in multi layer perceptrons. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 474–478, New Orleans, Louisiana. Association for Computational Linguistics.

Liyuan Zhou, Qiongkai Xu, Hanna Suominen, and Tom Gedeon. 2018. Epution at semeval-2018 task 2: Emoji prediction with user adaption. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 446–450, New Orleans, Louisiana. Association for Computational Linguistics.