Sentiment of Emojis

Petra Kralj Novak

DEPARTMENT OF
KNOWLEDGE
TECHNOLOGIES
Jožef Stefan Institute

4th Conference on
CMC and Social Media Corpora for the Humanities

Outline

- EMOJI FACTS
- OPINION MINING
- EMOJI SENTIMENT

Emoticon

- Facial expression composed of keyboard characters
- Emotion + icon (English)

O:-)	Angel
X-(	Angry
~:0	Baby
:-D	Big Grin
:-#	Braces
</3	Broken Heart
=^.^=	Cat
*<:o)	Clown
O.o	Confused
B-)	Cool
:'(	Crying
…	

Emoji

- Pictographs (pictorial symbols): things, emotions, activities, …
- Presented in a colorful form
- E 'picture' + *moji* 'letter, character' (Japanese)

Unicode® Emoji

Version	Emoji count	Year
1.0.0	76 (+76)	1991
3	78 (+2)	1999
3.2	86 (+8)	2002
4	94 (+8)	2003
4.1	108 (+14)	2005
5.1	112 (+4)	2008
5.2	139 (+27)	2009
6	855 (+716)	2010
6.1	868 (+13)	2012
7	972 (+104)	2014
8	1,013 (+41)	2015
9	1,085 (+72)	2016

- Single Unicode characters emoji counts
- Many more emojis are composed of sequences of two or more characters (e.g. regional indicators in pairs represent ≈ flags)

Emojis were first defined in Unicode 6.0, and pre-6.0 characters were only defined as emoji in 6.0 or later.

167 new Emoji
Unicode 9.0.0 (July 2016)

№	Code	Brow.	Chart	Apple	Googᵈ	Twtr.	One	FBM	Wind.	Sams.	GMail	SB	DCM	KDDI	Name	Date	Keywords
1	U+1F923	☐		—				—			—	—	—	—	rolling on the floor laughing	2016×	face | floor | laugh | rolling
№	Code	Brow.	Chart	Apple	Googᵈ	Twtr.	One	FBM	Wind.	Sams.	GMail	SB	DCM	KDDI	Name	Date	Keywords
2	U+1F924	☐		—				—			—	—	—	—	drooling face	2016×	drooling | face
№	Code	Brow.	Chart	Apple	Googᵈ	Twtr.	One	FBM	Wind.	Sams.	GMail	SB	DCM	KDDI	Name	Date	Keywords
3	U+1F920	☐		—				—			—	—	—	—	cowboy hat face	2016×	cowboy | cowgirl | face | hat
4	U+1F921	☐		—				—			—	—	—	—	clown face	2016×	clown | face
5	U+1F925	☐		—				—			—	—	—	—	lying face	2016×	face | lie | pinocchio
№	Code	Brow.	Chart	Apple	Googᵈ	Twtr.	One	FBM	Wind.	Sams.	GMail	SB	DCM	KDDI	Name	Date	Keywords
6	U+1F922	☐		—				—			—	—	—	—	nauseated face	2016×	face | nauseated | vomit
7	U+1F927	☐		—				—			—	—	—	—	sneezing face	2016×	face | gesundheit | sneeze
№	Code	Brow.	Chart	Apple	Googᵈ	Twtr.	One	FBM	Wind.	Sams.	GMail	SB	DCM	KDDI	Name	Date	Keywords
8	U+1F936	☐		—				—			—	—	—	—	Mrs. Claus	2016×	celebration | Christmas | claus | mother | Mrs.
9	U+1F936 U+1F3FB	☐ ☐		—				—			—	—	—	—	Mrs. Claus: light skin tone	2016×	celebration | Christmas | claus | light skin tone | mother | Mrs.
10	U+1F936 U+1F3FC	☐ ☐		—				—			—	—	—	—	Mrs. Claus: medium-light skin tone	2016×	celebration | Christmas | claus | medium-light skin tone | mother | Mrs.
11	U+1F936 U+1F3FD	☐ ☐		—				—			—	—	—	—	Mrs. Claus: medium skin tone	2016×	celebration | Christmas | claus | medium skin tone | mother | Mrs.
12	U+1F936 U+1F3FE	☐ ☐		—				—			—	—	—	—	Mrs. Claus: medium-dark skin tone	2016×	celebration | Christmas | claus | medium-dark skin tone | mother | Mrs.
13	U+1F936 U+1F3FF	☐ ☐		—				—			—	—	—	—	Mrs. Claus: dark skin tone	2016×	celebration | Christmas | claus | dark skin tone | mother | Mrs.
14	U+1F934	☐		—				—			—	—	—	—	prince	2016×	prince
15	U+1F934 U+1F3FB	☐ ☐		—				—			—	—	—	—	prince: light skin tone	2016×	light skin tone | prince
16	U+1F934 U+1F3FC	☐ ☐		—				—			—	—	—	—	prince: medium-light skin tone	2016×	medium-light skin tone | prince
17	U+1F934 U+1F3FD	☐ ☐		—				—			—	—	—	—	prince: medium skin tone	2016×	medium skin tone | prince
18	U+1F934 U+1F3FE	☐ ☐		—				—			—	—	—	—	prince: medium-dark skin tone	2016×	medium-dark skin tone | prince
19	U+1F934 U+1F3FF	☐ ☐		—				—			—	—	—	—	prince: dark skin tone	2016×	dark skin tone | prince

Sequence of Emojis

👤👥👤👤 **Family: Man, Woman, Boy, Boy**

A family with a father, mother, and two sons.

The Family: Man, Woman, Boy, Boy emoji is a sequence of the 👨 Man, 👩 Woman, 👦 Boy and 👦 Boy emojis. These are combined using a zero width joiner between each character and display as a single emoji on supported platforms.
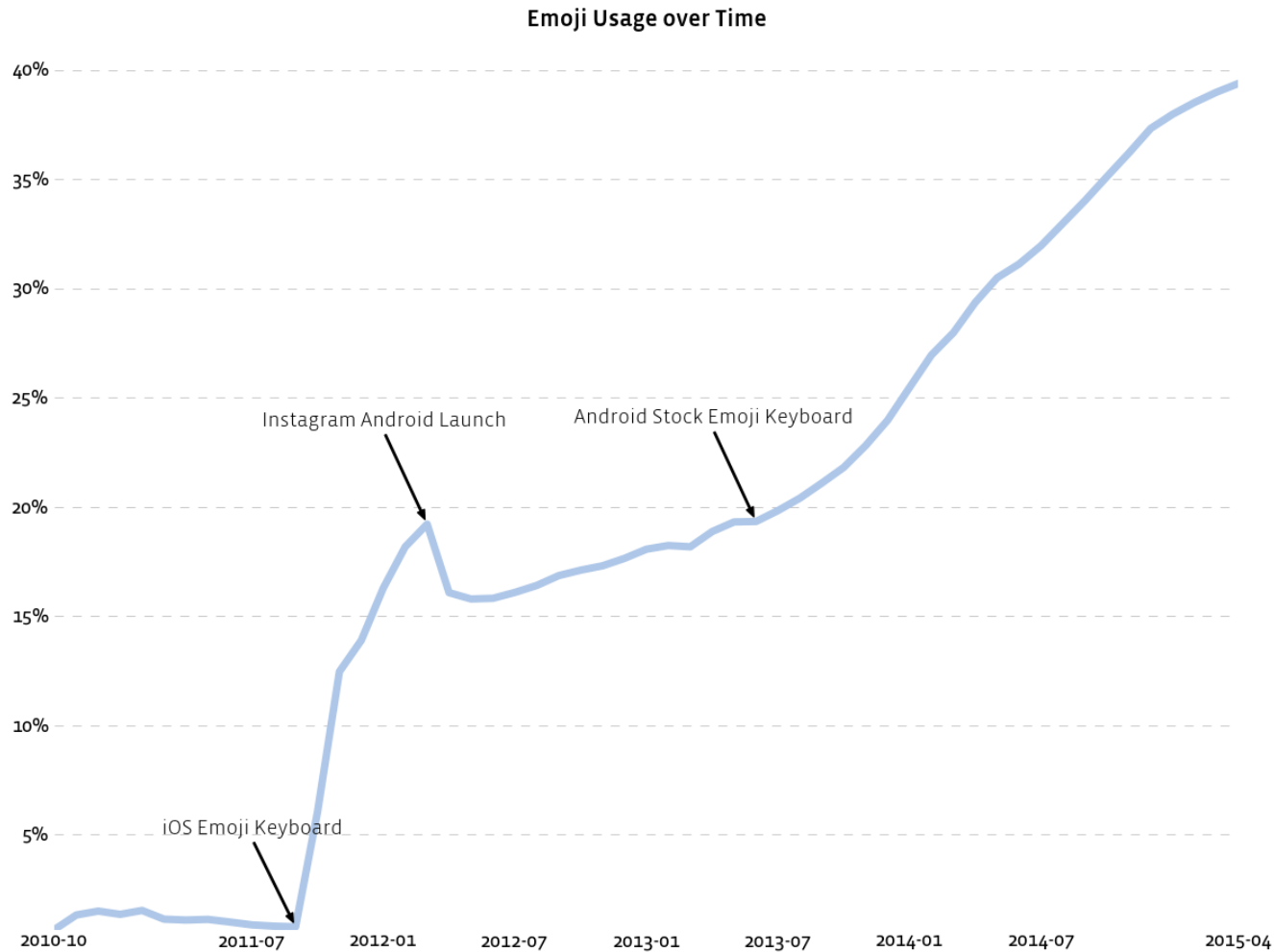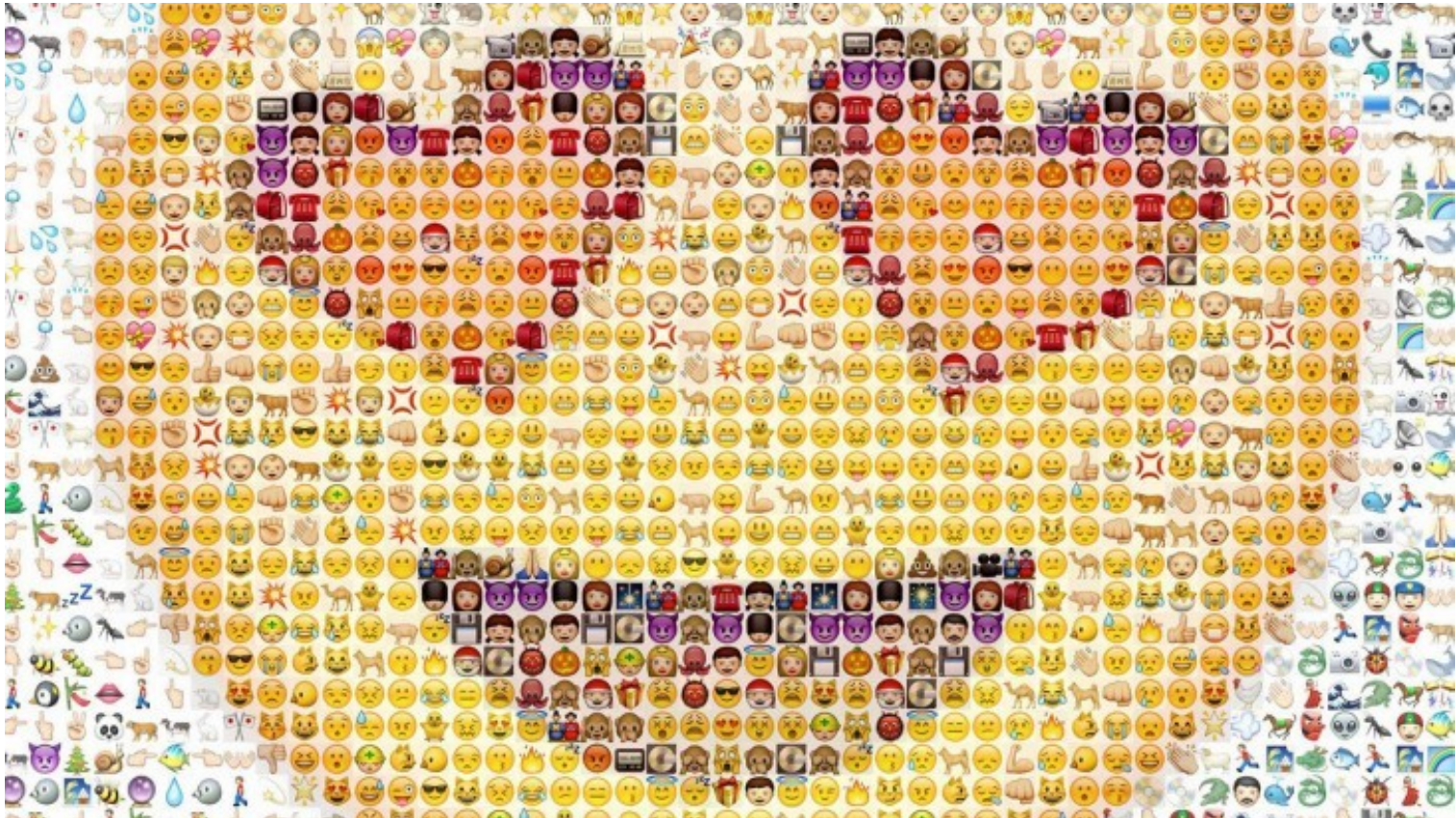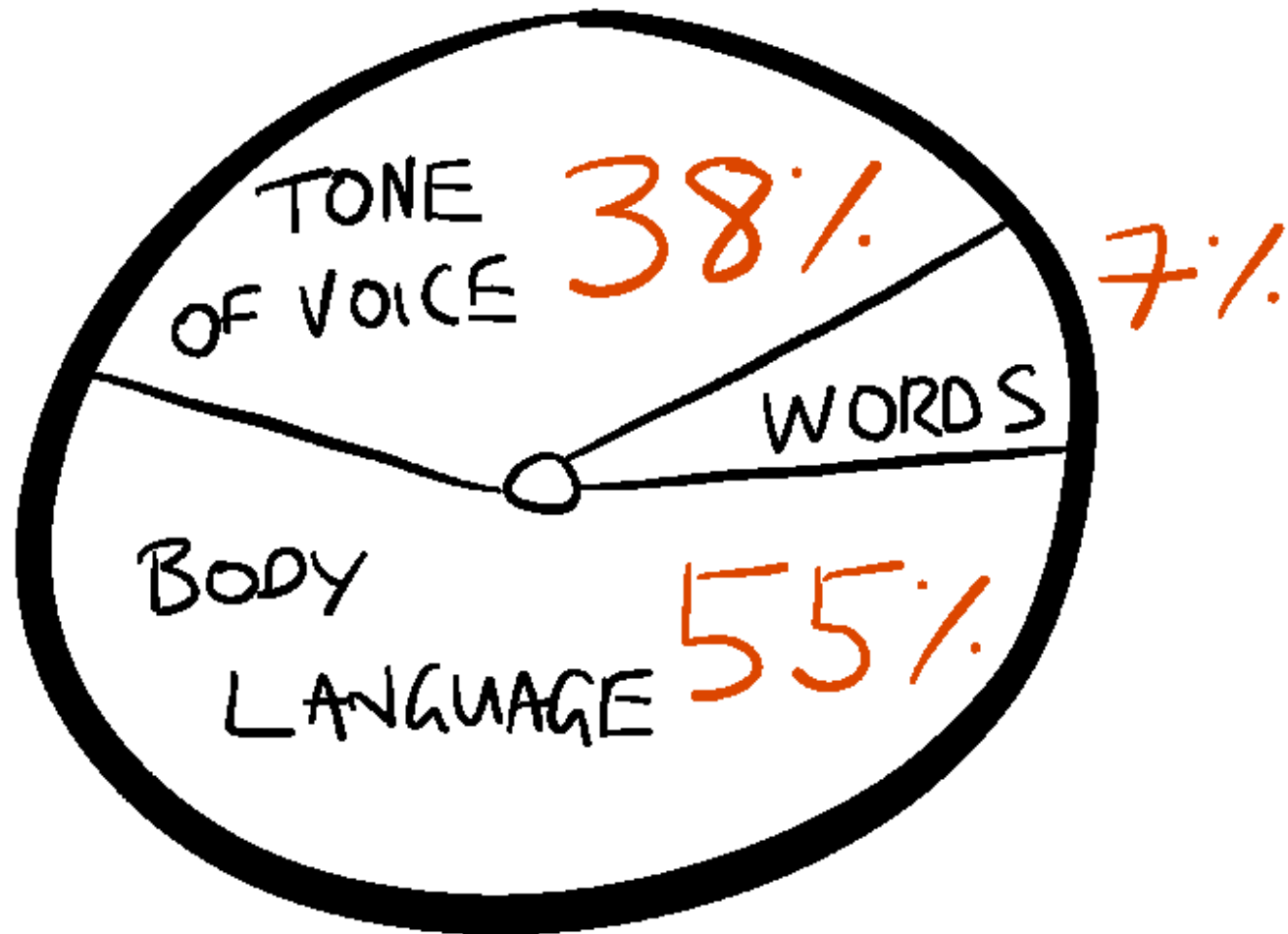
Unicode® Emoji

Version	Emoji count	Year
1.0.0	76 (+76)	1991
3	78 (+2)	1999
3.2	86 (+8)	2002
4	94 (+8)	2003
4.1	108 (+14)	2005
5.1	112 (+4)	2008
5.2	139 (+27)	2009
6	855 (+716)	2010
6.1	868 (+13)	2012
7	972 (+104)	2014
8	1,013 (+41)	2015
9	1,085 (+72)	2016

Emojis were first defined in Unicode 6.0, and pre-6.0 characters were only defined as emoji in 6.0 or later.

Emoji Popularity: Instagram

Emoji Usage over Time

Why do we ❤️ Emoji?

Oxford Dictionaries named 😂 (Face With Tears of Joy) its 2015 Word of the year.

WHEN COMMUNICATING FEELINGS AND ATTITUDES ...

TONE OF VOICE 38%

WORDS 7%

BODY LANGUAGE 55%

Mehrabian, Albert (1971). *Silent Messages* (1st ed.). Belmont, CA: Wadsworth. ISBN 0-534-00910-7.

Emoticons in Mind

When we look at a smiley face, the same parts of the brain are activated as when we look at a real human face.

Churches O, Nicholls M, Thiessen M, Kohler M, Keage H (2014) **Emoticons in mind: An event-related potential study**. Social Neuroscience 9(12)doi:10.1080/17470919.2013.873737
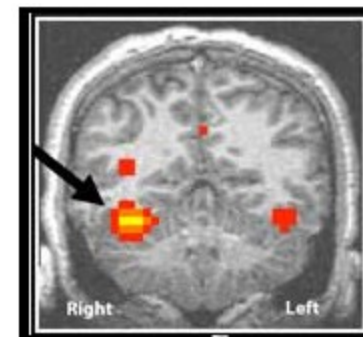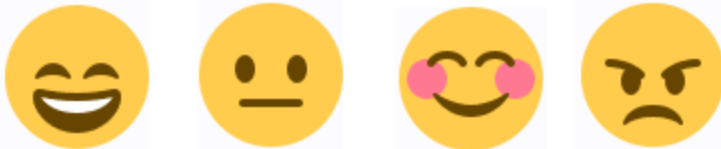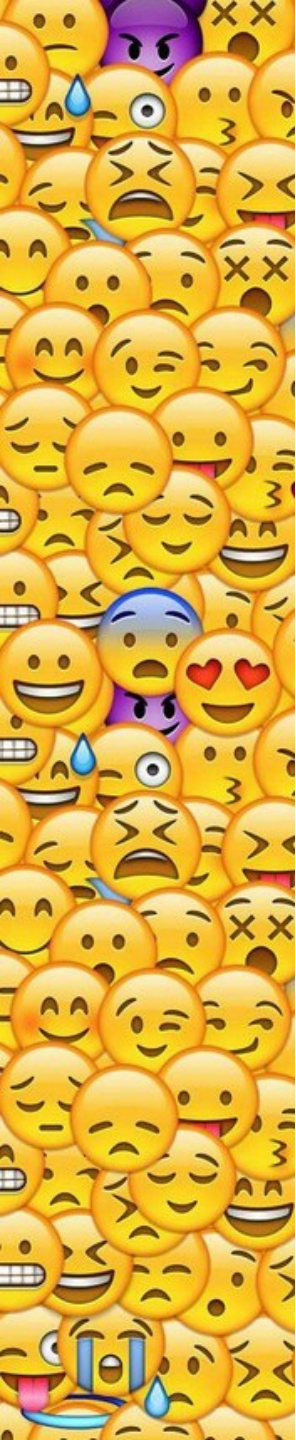
Outline

- **EMOJI FACTS**

- **OPINION MINING**

- **EMOJI SENTIMENT**

Opinion Mining
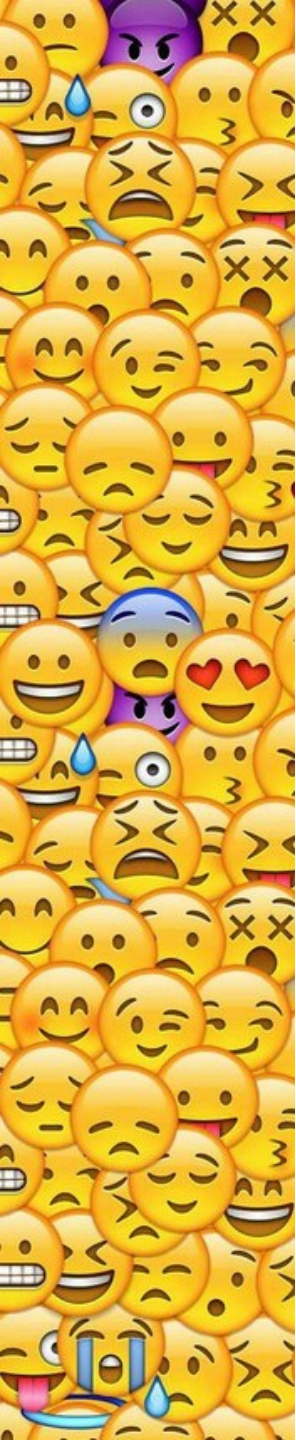
- Opinion mining (or sentiment analysis) is the computational study of people's opinions, sentiments, emotions, and attitudes.
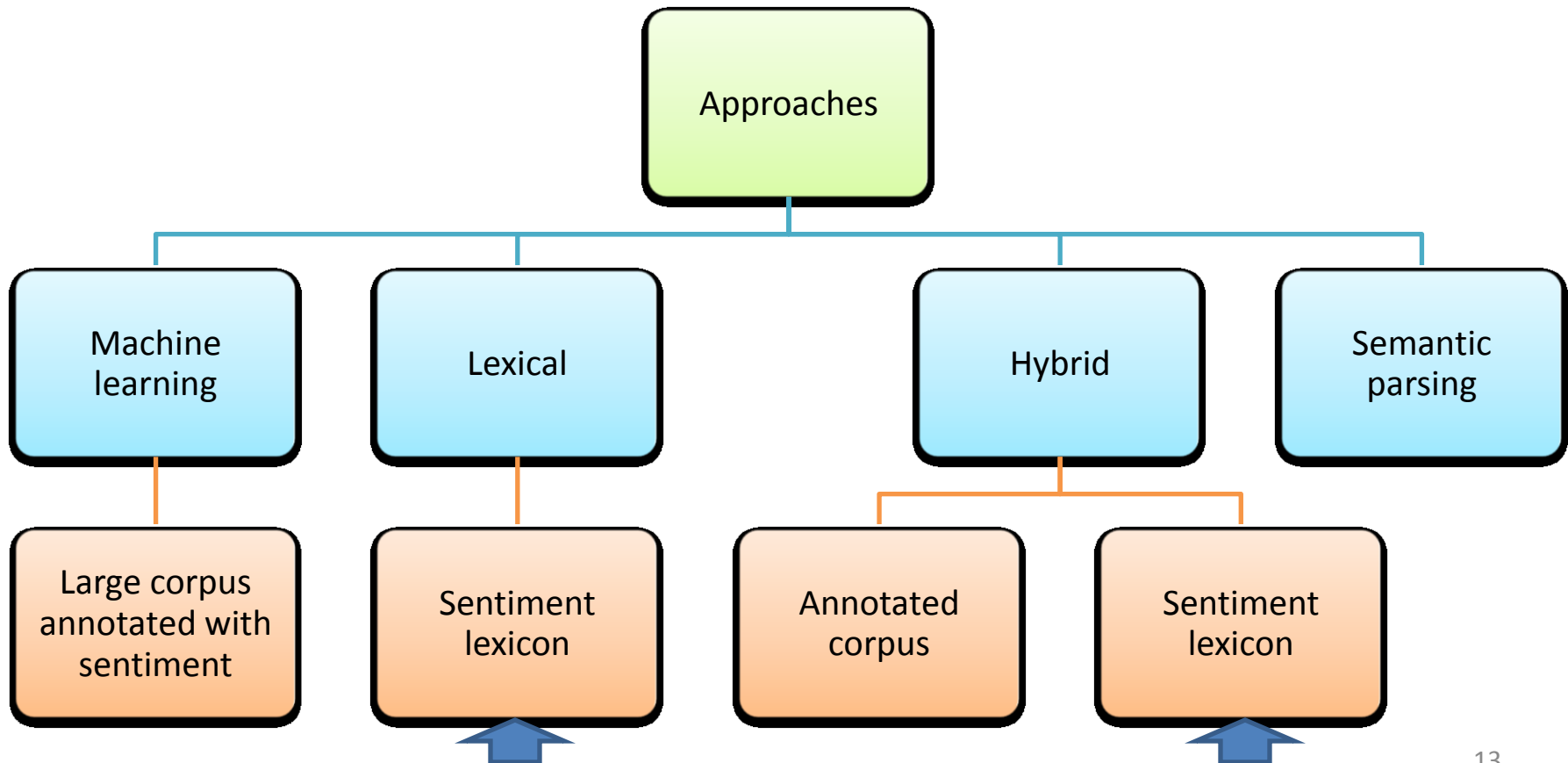
Sentiment Lexicons

- Lists of words associated to scores of sentiment polarity
- Well developed for English
  - Bing Liu's Opinion Lexicon
  - MPQA Subjectivity Lexicon
  - Harvard General Inquirer
  - **SentiWordNet** assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity.

superb#1 brilliant#1

of surpassing excellence; "a brilliant performance"; "a superb actor"

P: 0.875 O: 0.125 N: 0

Baccianella S, Esuli A, Sebastiani F. "**SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining**." LREC. Vol. 10. 2010.

Outline

- EMOJI FACTS
- OPINION MINING
- EMOJI SENTIMENT

Kralj Novak P, Smailović J, Sluban B, Mozetič I (2015) **Sentiment of Emojis**. PLoS ONE 10(12): e0144296. doi:10.1371/journal.pone.0144296
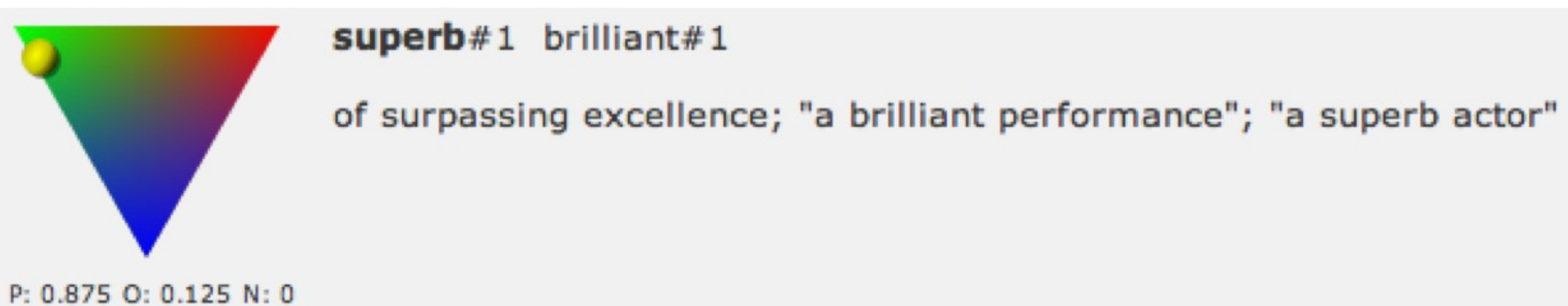
Why Sentiment of Emojis?

Emoticons in opinion mining

- As features for sentiment models
- As labels to generate samples of negative and positive texts
- Sentiment conveyed by emoticons is domain and topic independent

Emojis are replacing emoticons in social media

B-)   →   😎

Dataset

- Over 1.6 million tweets in 13 European languages (Apr. 2013 – Feb. 2015)
- Labeled by 83 native speakers: negative (-1) < neutral (O) < positive (+1)
- 4% of our tweets contain emojis (69,546)

Language	No. of tweets	No. of annotators
Albanian	53,005	13
Bulgarian	67,169	18
English	103,034	9
German	109,130	5
Hungarian	68,505	1
Polish	223,574	8
Portuguese	157,393	1
Russian	107,773	1
Ser/Cro/Bos	215,657	13
Slovak	70,425	1
Slovenian	133,935	7
Spanish	275,588	5
Swedish	58,547	1
Total	1,643,735	83

Mozetič I, Grčar M, Smailović J (2016) **Multilingual Twitter Sentiment Classification: The Role of Human Annotators**. PLoS ONE 11(5): e0155036. doi:10.1371/journal.pone.0155036

Overlap with emojitracker

Emojitracker	Tweets with emojis ~4 billion	Different emojis used 845	Pearson correlation /	Spearman rank correlation /
Emoji Sent. Rank.				
$N \geq 1$	69,673	969 (721)	0.945*	0.897*
Emoji Sent. Rank.				
$N \geq 5$	69,546	**751** (608)	0.944*	0.898*

Sentiment Score

🌷 **BOUQUET**

Discrete distribution	negative: 3 neutral: 17 positive: 75

95 occurrences

Discrete probability distribution	negativity $p_- = 0.04$ neutrality $p_0 = 0.18$ positivity $p_+ = 0.78$

Sentiment Score	Sentiment score = 0.74

Sentiment Score

Discrete distribution

Laplace estimate to estimate the probability

$$p_c = \frac{N(c) + 1}{N + k} \ , \ \text{(for large } N: \ p_c \approx \frac{N(c)}{N})$$

Discrete probability distribution

Sentiment score is the mean of the discrete probability distribution

$$\overline{s} = -1 \cdot p_- + 0 \cdot p_0 + 1 \cdot p_+ = p_+ - p_-$$
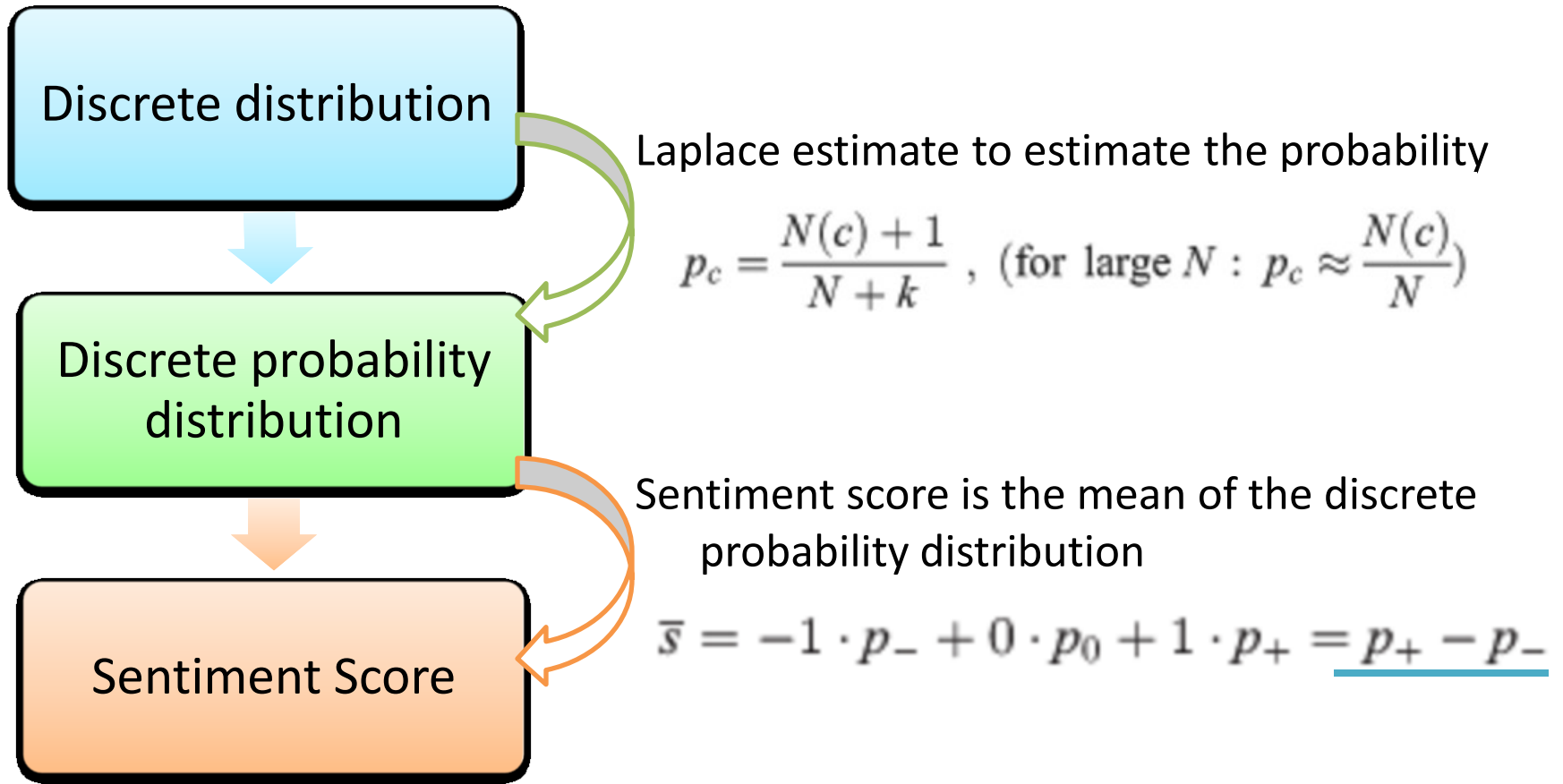
Sentiment Score

Sentiment Bar

🌷 BOUQUET

- Negative: 0.04
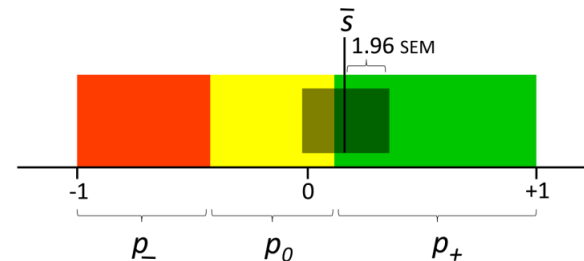- Neutral: 0.18
- Positive: 0.78
- Sentiment Score: 0.74
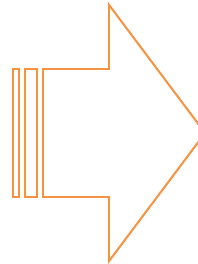
THUMBS DOWN SIGN

FLUSHED FACE

$\bar{s}$
1.96 SEM

CHOCOLATE BAR

-1   0   +1

p_- p_0 p_+

Emoji Sentiment Lexicon

http://kt.ijs.si/data/Emoji_sentiment_ranking/

Char	Image [twemoji]	Unicode codepoint	Occurrences [5...max]	Position [0...1]	Neg [0...1]	Neut [0...1]	Pos [0...1]	Sentiment score [-1...+1]	Sentiment bar (c.i. 95%)	Unicode name	Unicode block
😊	😂	0x1f602	14622	0.805	0.247	0.285	0.468	0.221		FACE WITH TEARS OF JOY	Emoticons
❤	❤️	0x2764	8050	0.747	0.044	0.166	0.790	0.746		HEAVY BLACK HEART	Dingbats
♥	❤️	0x2665	7144	0.754	0.035	0.272	0.693	0.657		BLACK HEART SUIT	Miscellaneous Symbols
😍	😍	0x1f60d	6359	0.765	0.052	0.219	0.729	0.678		SMILING FACE WITH HEART-SHAPED EYES	Emoticons
😭	😭	0x1f62d	5526	0.803	0.436	0.220	0.343	-0.093		LOUDLY CRYING FACE	Emoticons
😘	😘	0x1f618	3648	0.854	0.053	0.193	0.754	0.701		FACE THROWING A KISS	Emoticons
😊	😊	0x1f60a	3186	0.813	0.060	0.237	0.704	0.644		SMILING FACE WITH SMILING EYES	Emoticons
👌	👌	0x1f44c	2925	0.805	0.094	0.249	0.657	0.563		OK HAND SIGN	Miscellaneous Symbols and Pictographs
💕	💕	0x1f495	2400	0.766	0.042	0.285	0.674	0.632		TWO HEARTS	Miscellaneous Symbols and Pictographs
👏	👏	0x1f44f	2336	0.787	0.104	0.271	0.624	0.520		CLAPPING HANDS SIGN	Miscellaneous Symbols and Pictographs

22

Language Independent?

Create emoji sentiment lexicon for each language separately

⬇

Order emojis by their sentiment score

⬇

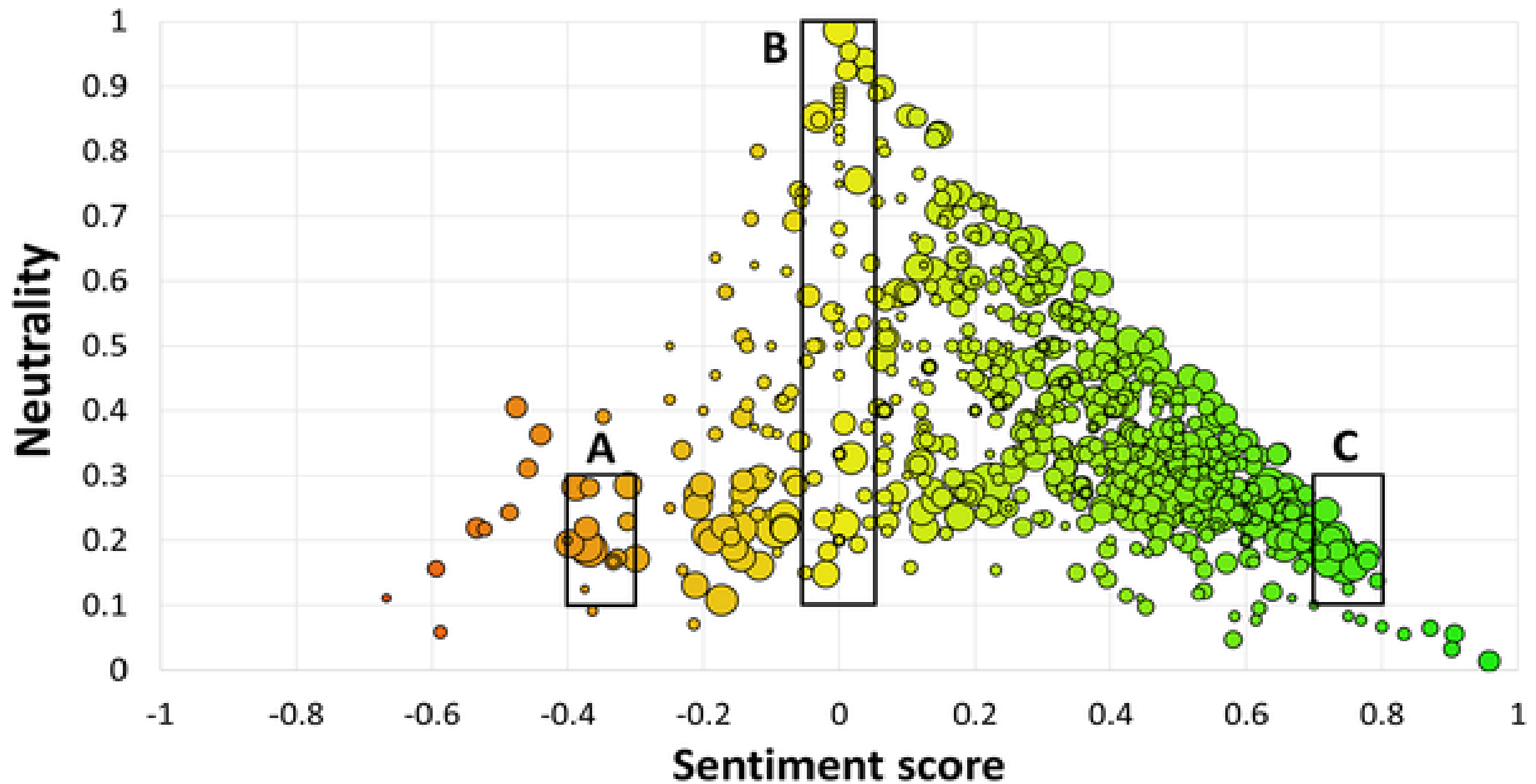Compare the ordered list with Emoji Sentiment Lexicon

Emoji Sentiment in Different Languages

Emoji Sent. Rank.	Tweets with emojis 69,546	Different emojis used 751	Pearson correlation /	Spearman rank correlation /
English	19,819	511	0.834*	0.819*
Spanish	22,063	448	0.552*	0.573*
Polish	8,112	253	0.810*	0.783*
Russian	5,007	221	0.777*	0.756*
Hungarian	2,324	176	0.588*	0.612*
German	3,062	142	0.782*	0.783*
Swedish	2,797	139	0.702*	0.674*
Ser/Cro/Bos	2,096	123	0.708*	0.615*
Slovak	1,526	108	0.620*	0.499*
Slovenian	996	66	0.526*	0.541*
Portuguese	796	56	0.410*	0.429*
Bulgarian	607	36	0.557*	0.443*
Albanian	341	19	0.363 *	0.416 *

doi:10.1371/journal.pone.0144296.t005

- Pearson's correlation and Spearman's rank correlation are relatively high, and significant for all the languages, except Albanian.

- Emoji Sentiment Ranking can be considered a universal, language-independent resource, at least for European languages.
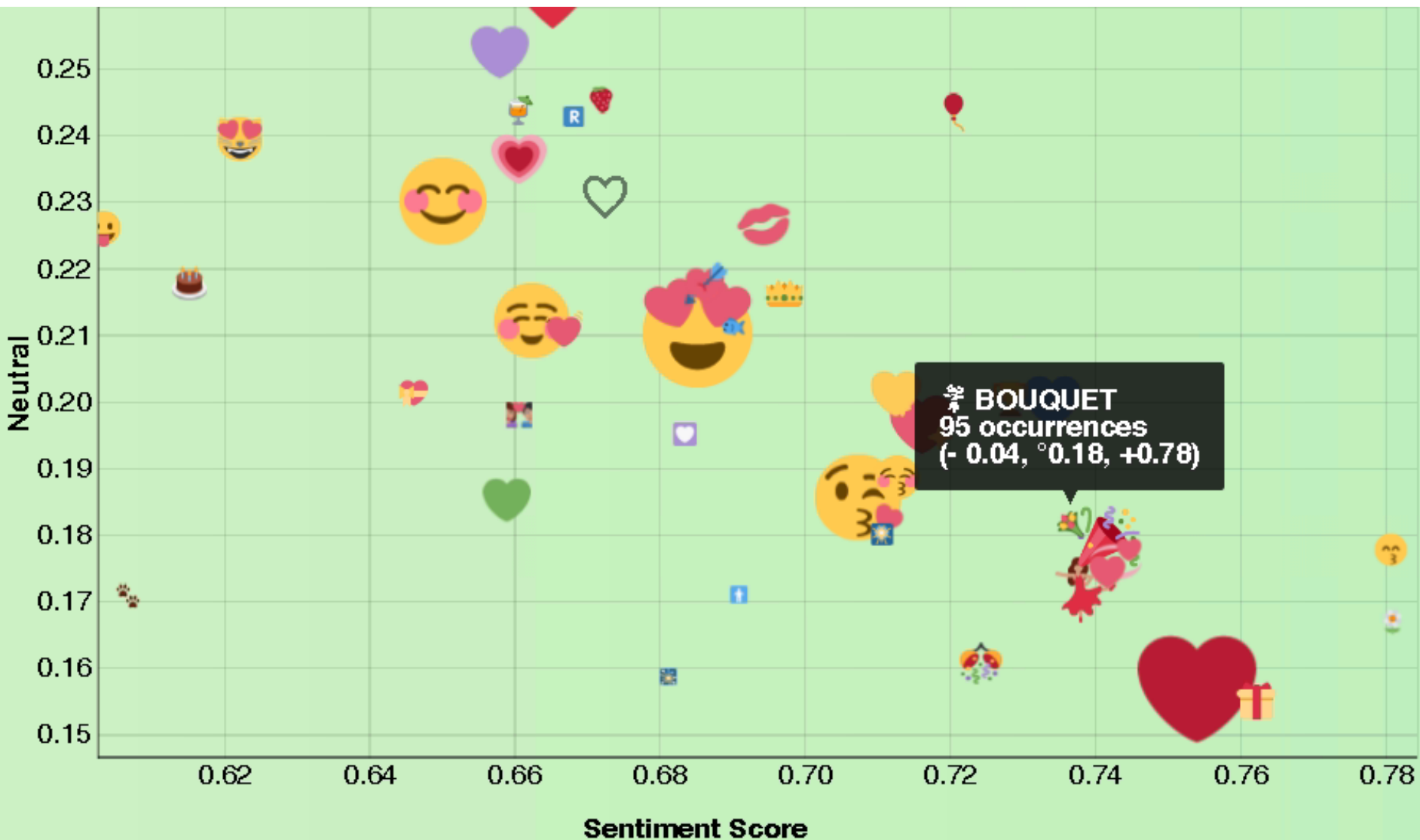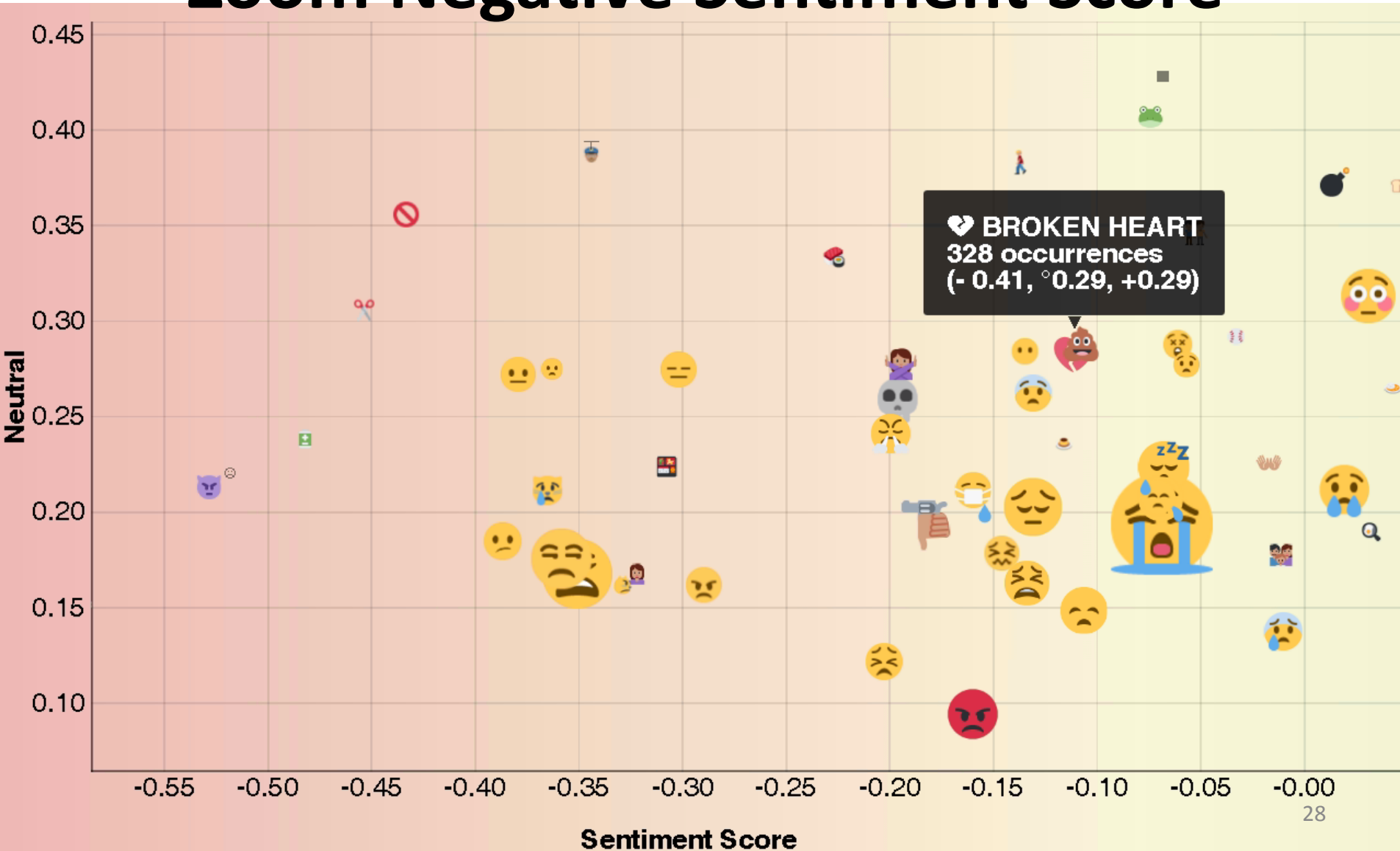
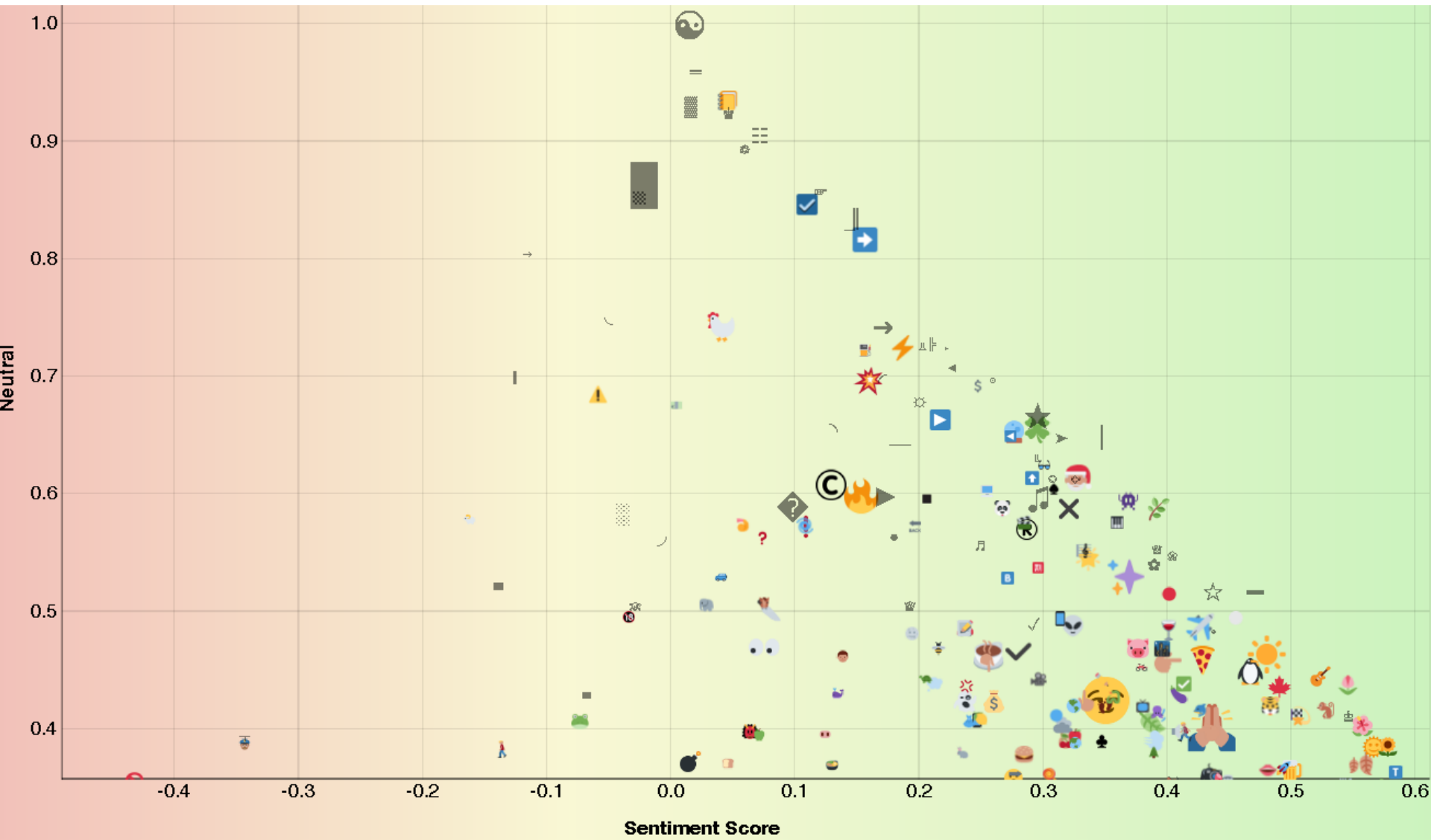# Emoji Sentiment Map

Emoji Sentiment Map
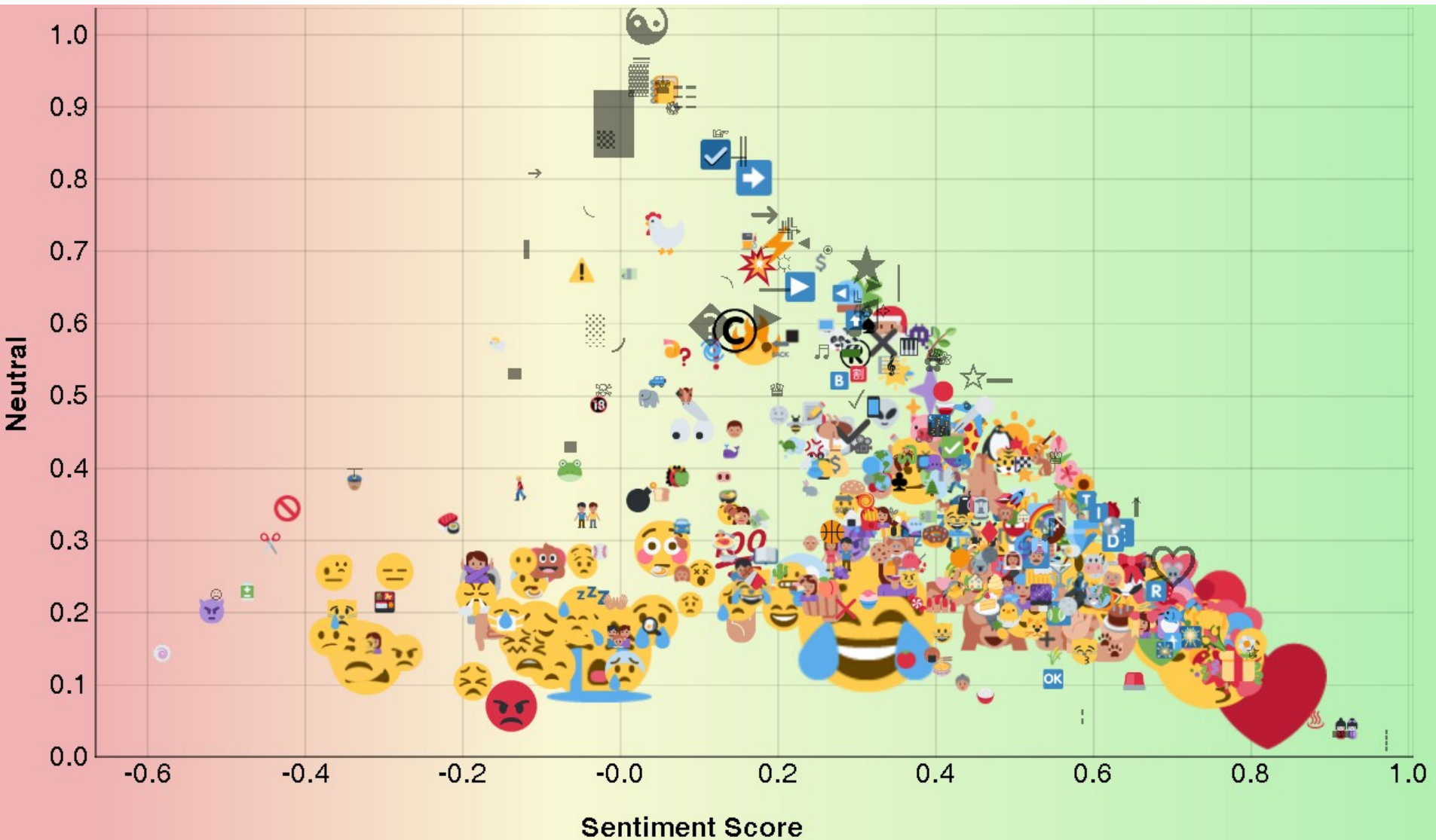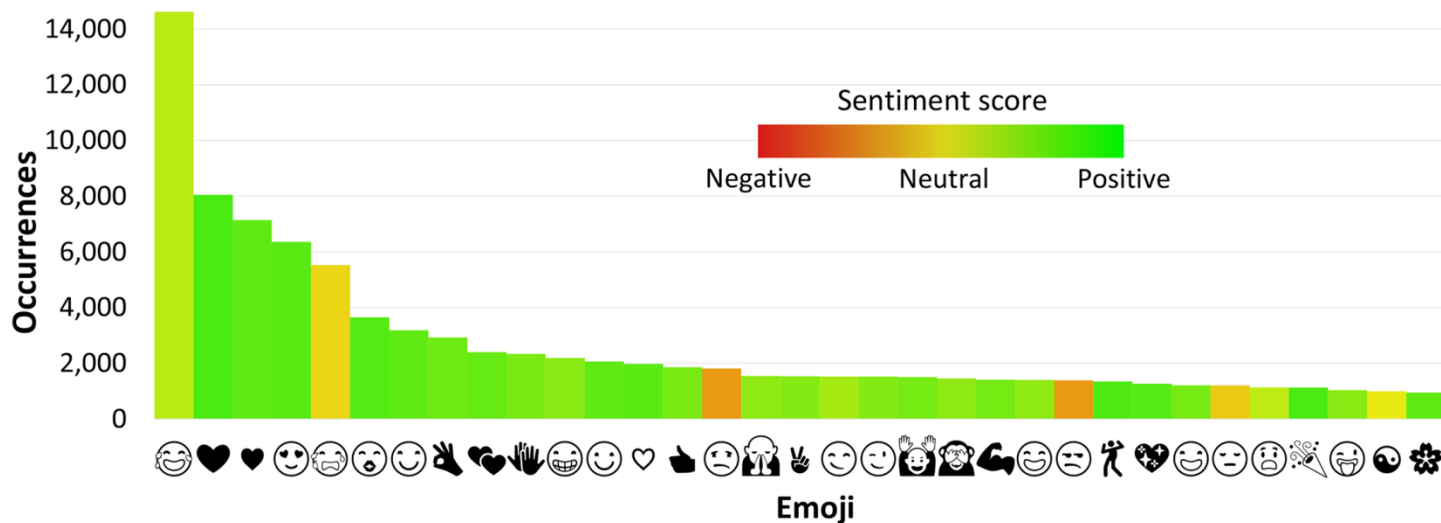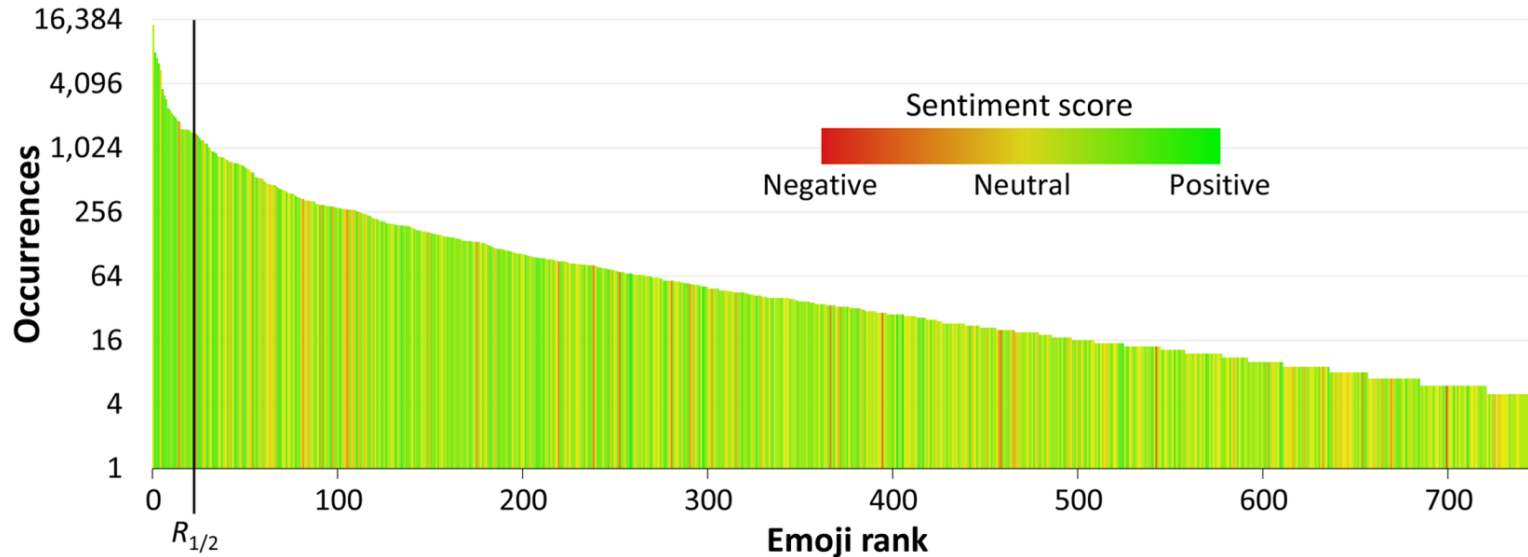
Emoji Sentiment Map:
Zoom Positive Sentiment Score

Neutral (y-axis): 0.15, 0.16, 0.17, 0.18, 0.19, 0.20, 0.21, 0.22, 0.23, 0.24, 0.25

Sentiment Score (x-axis): 0.62, 0.64, 0.66, 0.68, 0.70, 0.72, 0.74, 0.76, 0.78

💐 BOUQUET
95 occurrences
(- 0.04, °0.18, +0.78)

Emoji Sentiment Map:
Zoom Negative Sentiment Score

BROKEN HEART
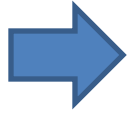328 occurrences
(- 0.41, °0.29, +0.29)

Emoji Sentiment Map:
Zoom Neutral Sentiment Score

Emoji Occurrence and Sentiment

Tweets With and Without Emojis

Sentiment

- Does the sentiment distribution differ between tweets with and without emojis?
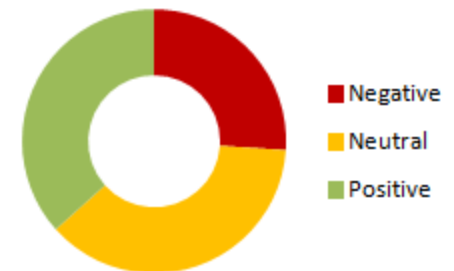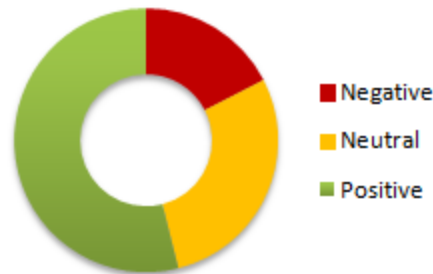
Annotator agreement

- Do annotators agree on the sentiment of the tweet any better if tweet contains emojis?

Tweets With and Without Emojis: Sentiment

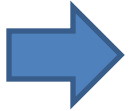| Sentiment	Tweets with emojis	Tweets without emojis
Negative	12,156 (17,5%)	410,301 (26,1%)
Neutral	19,938 (28,6%)	587,337 (37,3%)
Positive	37,579 (53,9%)	576,424 (36,6%)
Total	69,673	1,574,062
Mean	+0.365	+0.106
SD, SEM	0.762, 0.0029	0.785, 0.0006

doi:10.1371/journal.pone.0144296.t002

- Tweets with emojis are more positive.

Tweets With and Without Emojis

Sentiment

- Does the sentiment distribution differ between tweets with and without emojis?

Annotator agreement

- Do annotators agree on the sentiment of the tweet any better if it contains emojis?

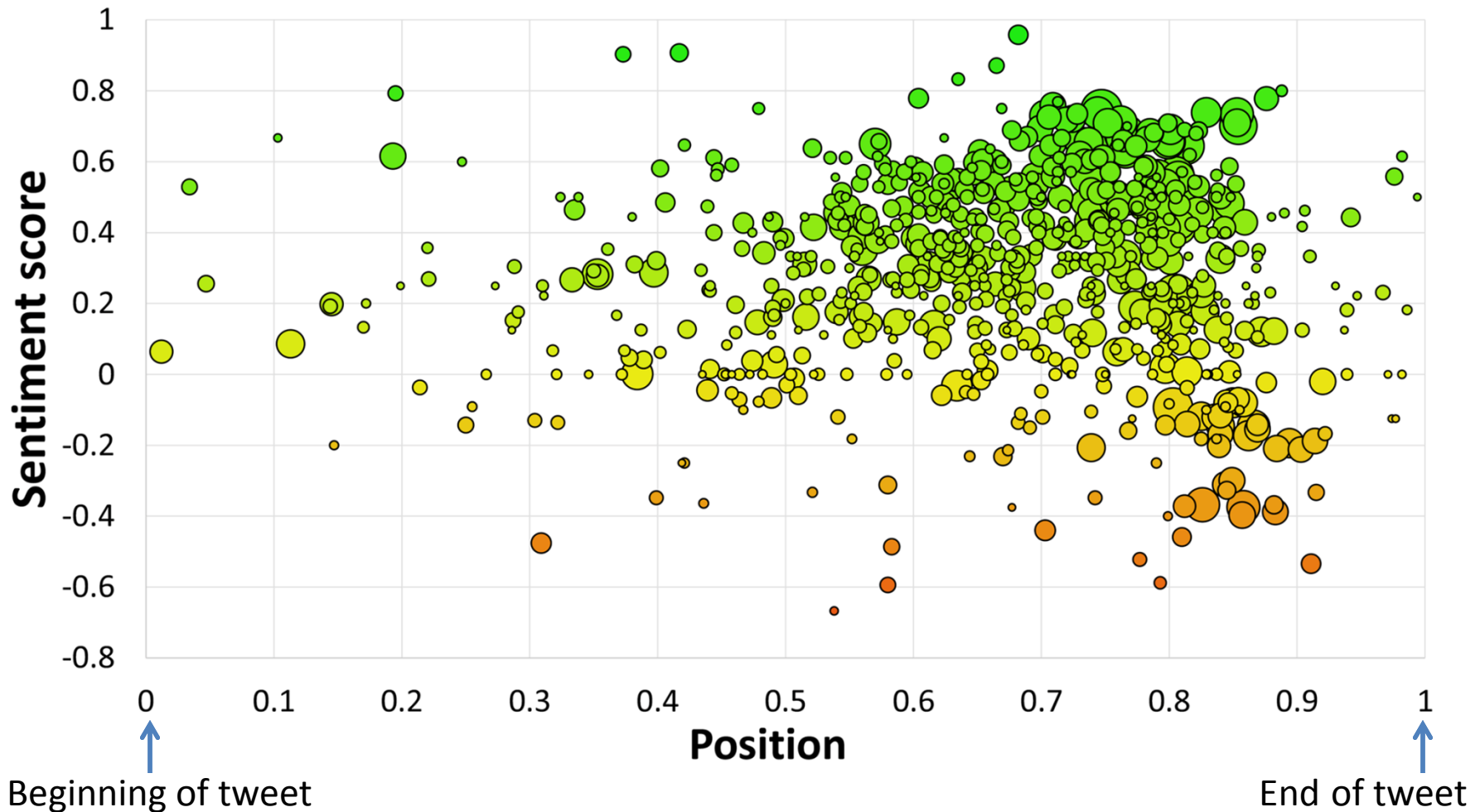# Tweets With and Without Emojis: Annotator Agreement

Agreement measure	Tweets with emojis	Tweets without emojis
Alpha	0.597	0.495
Accuracy	0.641	0.583
$\overline{F_1}(-, +)$	0.698	0.598
No. of tweets annotated twice	3,547	52,027

- The presence of emojis has a positive impact on the emotional perception of the tweets by humans.

Where in Tweet?

More emotionally loaded emojis, either negative or positive, tend to occur towards the end of the tweets.
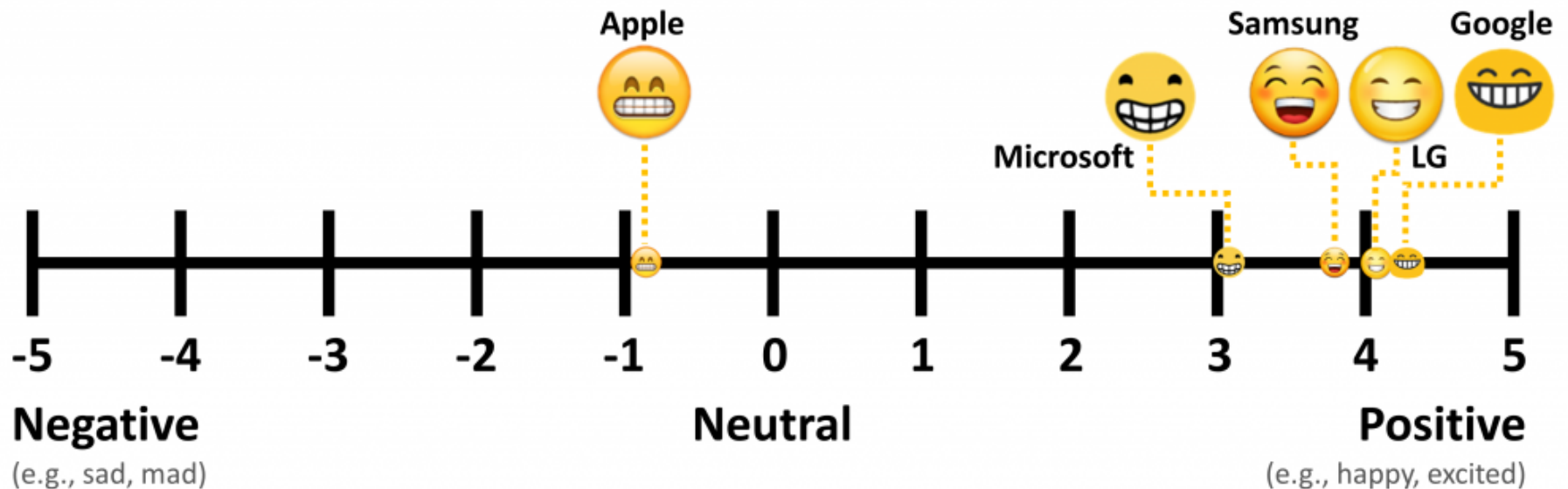
Applications

- Opinion mining and sentiment analysis
  - Automated sentiment classification
  - Automated dataset generation

- Fiverr.com is developing a system to automatically identify dissatisfaction, which is using our Emoji sentiment lexicon

fiverr
Anything For $5

Caution

Same Emoji + Different Smartphone Platform = Different Emotion

For example, if you send the Apple emoji to a Google Nexus, they'll see the Google emoji, and vice versa!

Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., and Hecht, B. (2016) **"Blissfully happy" or "ready to fight": Varying Interpretations of Emoji**. Proceedings of ICWSM 2016. Menlo Park, CA: AAAI Press.

Conclusion

- Emojis are ubiquitous
- Most of the most frequently used emojis are positive
- Tweets with emojis are more positive
- Annotators perceive sentiment better if emojis are present
- Emotionally loaded emojis tend to appear at the end of the tweet
- Emoji sentiment map
- Emoji sentiment lexicon
  - Language independent
  - Clarin.si
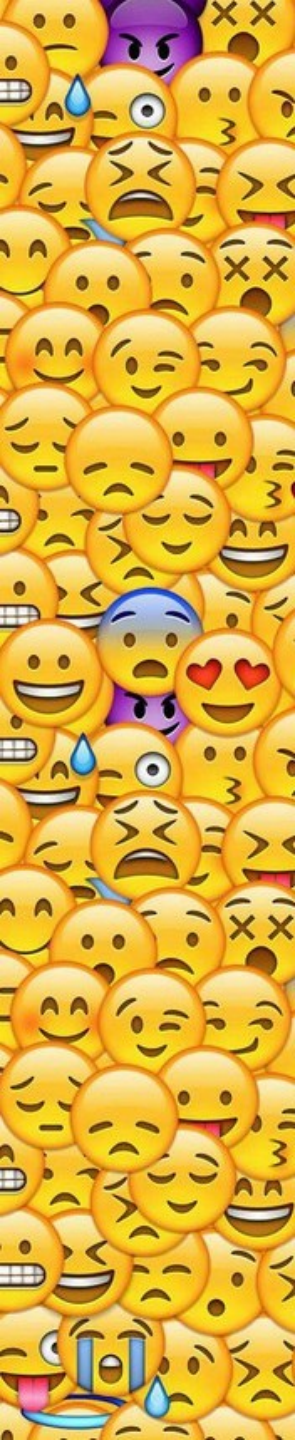
Thanks to coauthors

Jasmina Smailović, Borut Sluban and Igor Mozetič

THANK YOU.