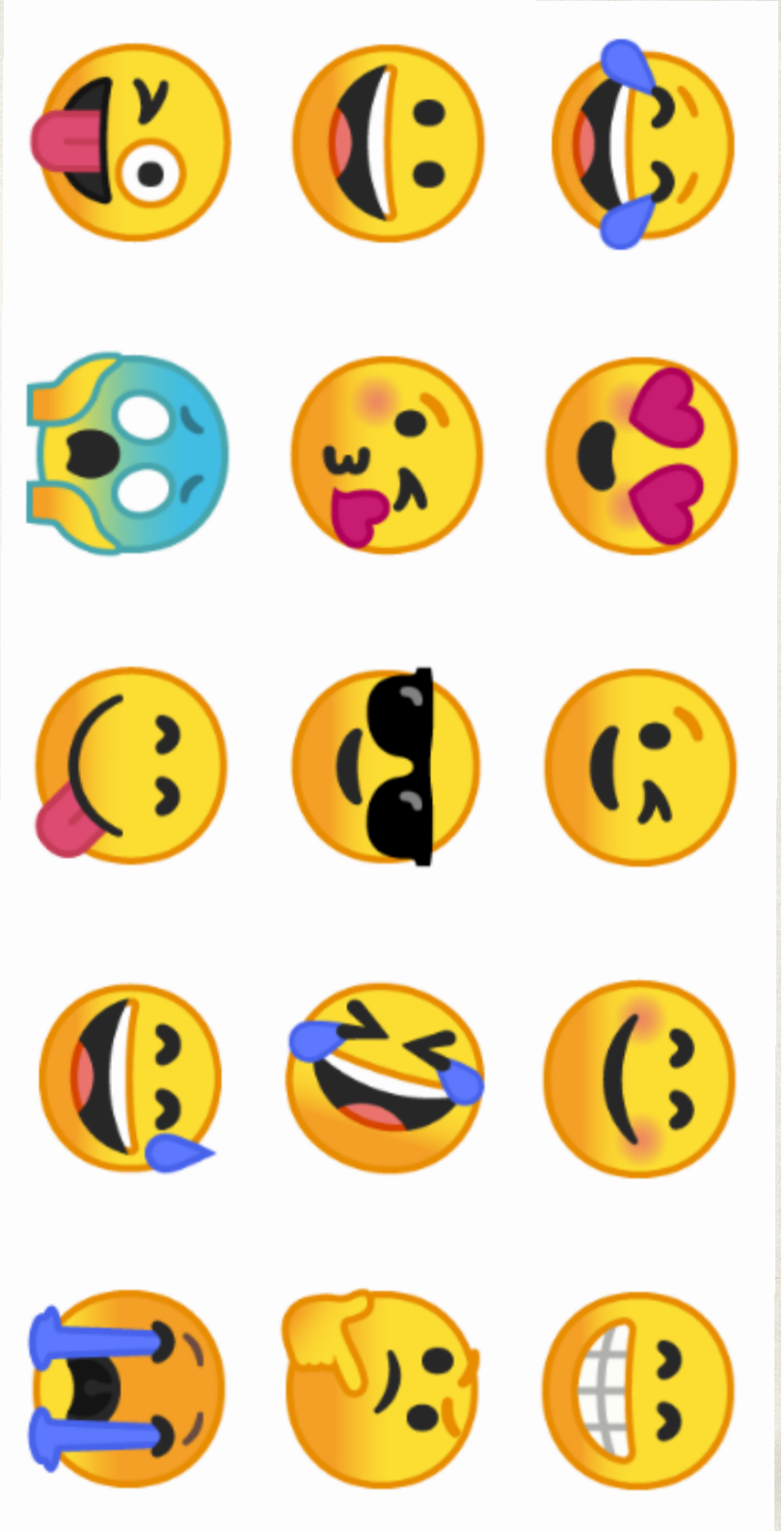
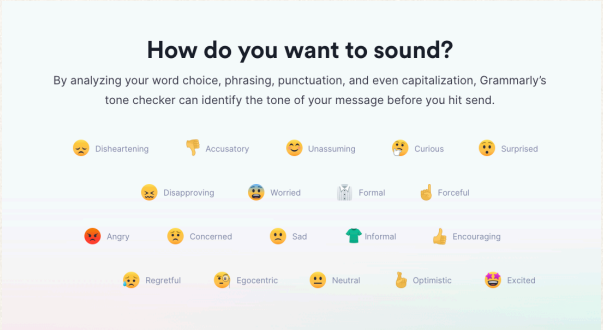
Emoji, emoticons, reactions in social media

A computational linguistics perspective



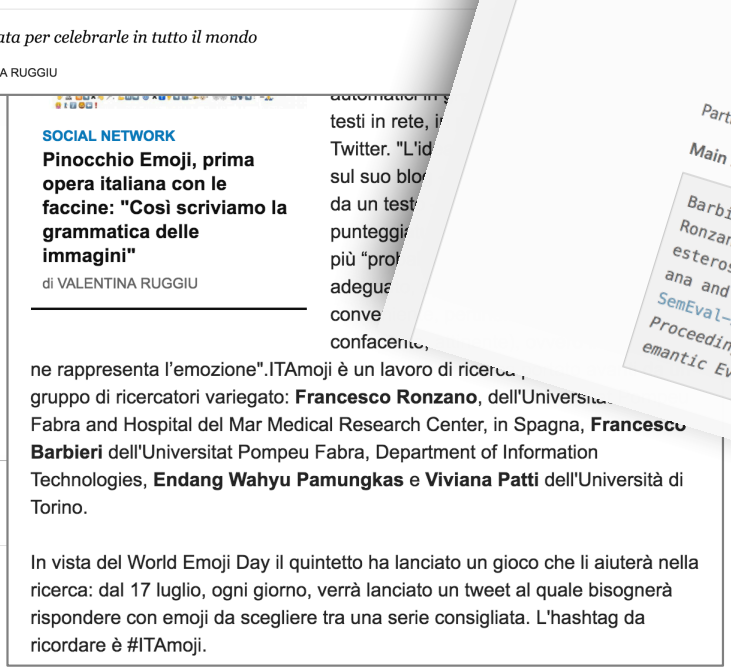
Emojis in Social Media

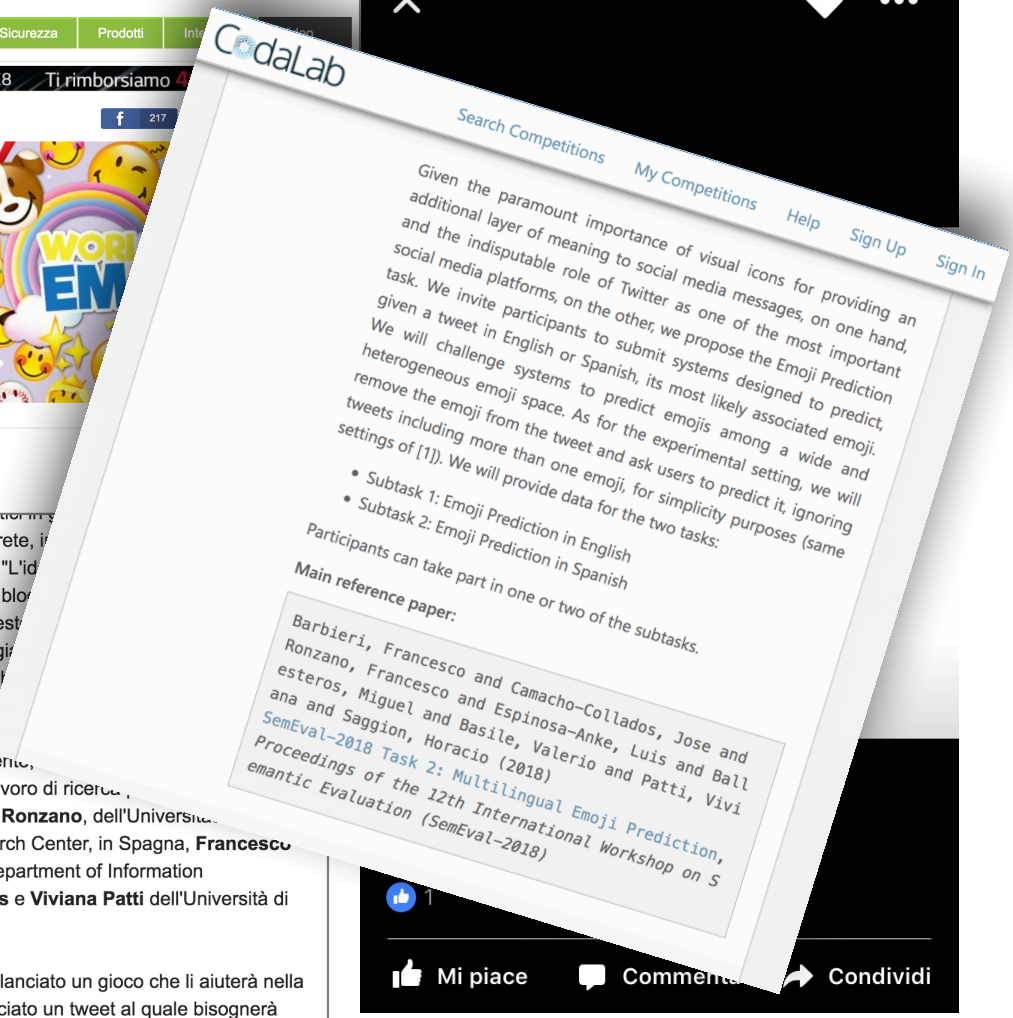
- The correct interpretation of the semantics of Emojis can benefit several tasks like:
  - Sentiment analysis, Emotion detection (recognition)
  - Human-computer interaction settings: how can we teach an artificial agent to correctly interpret and recognize emojis' use in spontaneous conversation
    - Chatbots conversation's tone (generation)
  - Very advanced tools supporting digital writing (emails) (generation)



Giulia Donato, Patrizia Paggio: Classifying the Informative Behaviour of Emoji in Microblogs. LREC 2018.

#ITAmoji che passione!





Expressive signals: emoticons & emoji

- Emojis are ideograms which are naturally combined with plain text to visually complement or condense the meaning of a message
- Emoticons vs emoji

Emoji

From Wikipedia, the free encyclopedia

Emoji (絵文字?, Japanese pronunciation: [emodʑi]) are ideograms and smileys used in electr
like ASCII emoticons or kaomoji, exist in various genres, including facial expressions, cor

Originating on Japanese mobile phones in the late 1990s, emoji have become increasing
which was followed by similar adoption by Android and other mobile operating systems.[
(Lion).[4] Microsoft added monochrome Unicode emoji coverage to the Segoe UI Symbol
Segoe UI Emoji font.[5]

Originally meaning pictograph, the word *emoji* comes from Japanese *e* (絵, "picture") + *m*
"emotion" and "emoticon" is just a coincidence.[6]

In the examples below, all emoji in body text and tables will be supplied by the default br
devices running different operating syst *that shown right*

Contents [hide]

Semantics?
Challenging!

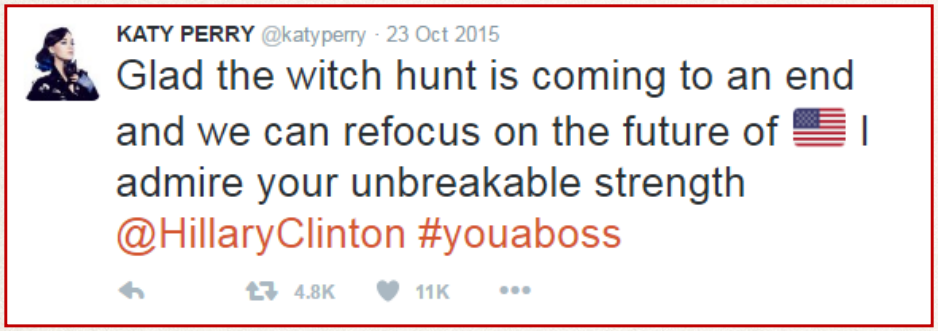
ò.ò	o_0 oppure 0_o o.0 oppure 0.o 0o oppure o0	Confusione, diffidenza
>_<	>.< è.é oppure è_é	Fastidio, seccatura
U_U	u_u oppure (U_U) U.U oppure u.u oppure (U.U) oppure U.u oppure UoU oppure ù.ù / V_V oppure v_v oppure (V_V) V.V oppure v.v oppure (V.V) oppure ˇ-ˇ	Superiorità (in modo ironico)

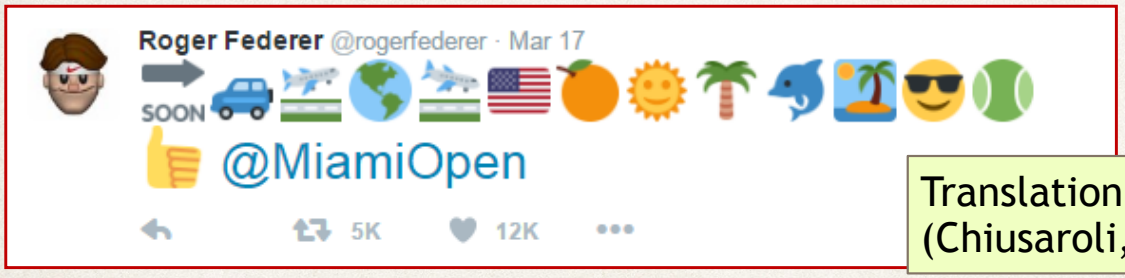
Emojis in Social Media

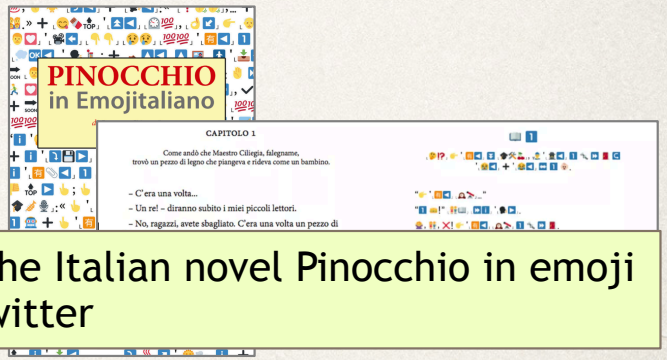
- Emojis and emoticons are non-verbal elements widely used in social media texts, which we often combine with words in our messages
- Reasons/functions
  - Clarifying our communicative intentions
  - Expressing emotions (especially emoji faces)
  - Visually complementing or enriching expressiveness of a short text message
  - Making the tone of an online conversation more empathetic
  - Playing, having fun

Emojis in Social Media

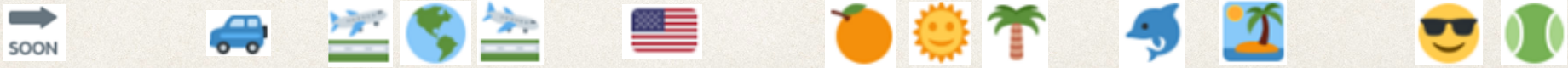
Emojis are **pictograms** extensively exploited
to visually complement and enrich the expressiveness
of short text messages in Social Media







Translation project of the Italian novel Pinocchio in emoji (Chiusaroli, 2017) on Twitter



In a few days I'll travel by plane to America; I'll enjoy the sun and the sea and I'll play tennis

Emojis in Social Media

- Computational linguistics studies the semantics of emojis and the relationship between words and emojis, which can provide fundamental information in the construction of tools for the automatic analysis and processing of content generated by users on the social web
  - Emphasizing what is already expressed in the text
  - Conveying a meaning or an emotion that cannot be recognized from words alone.
  - Redundancy

Giulia Donato, Patrizia Paggio: Classifying the Informative Behaviour of Emoji in Microblogs. LREC 2018.

Emojis in Social Media

- Relationship between words and emojis
  - Redundant: the emoji of interest repeats the information present in the text or that its meaning is implied by the text.
  - Non-redundant: captures cases in which the emoji adds information that is neither explicitly present nor implied in the text.
  - Non-Redundant + POS: refers to a specific kind of redundant use, indicates that the emoji is used with a syntactic function (and can be labeled with its own POS), thus replacing a word.

1. *Redundant*

   "We'll always have Beer. I'll see to it. I got your back on that one. 🍺"

2. *Non-Redundant*

   "I wish you were here ✈️"

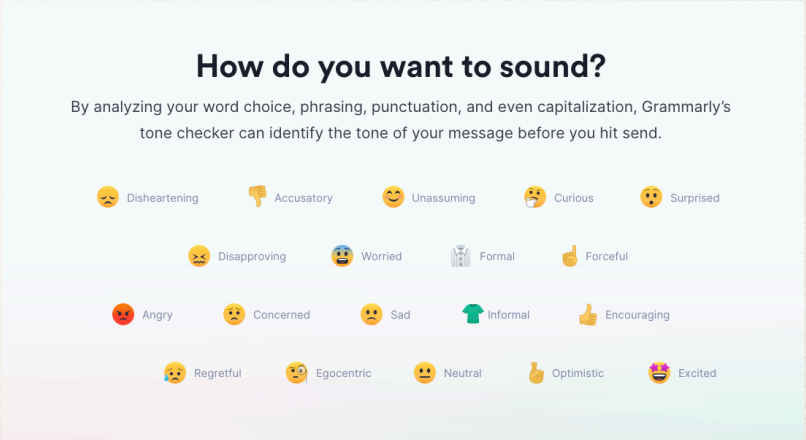
3. *Non-Redundant + POS*

   "Thank you so so so so much ily Here's a 🍕 as a thank you gift x"

Giulia Donato, Patrizia Paggio: Classifying the Informative Behaviour of Emoji in Microblogs. LREC 2018.

Emojis in Social Media

- The correct interpretation of the semantics of Emojis can benefit several tasks like:
  - Sentiment analysis, Emotion detection (recognition)
  - Human-computer interaction settings: how can we teach an artificial agent to correctly interpret and recognize emojis' use in spontaneous conversation
    - Chatbots conversation's tone (generation)
  - Very advanced tools supporting digital writing (emails) (generation)

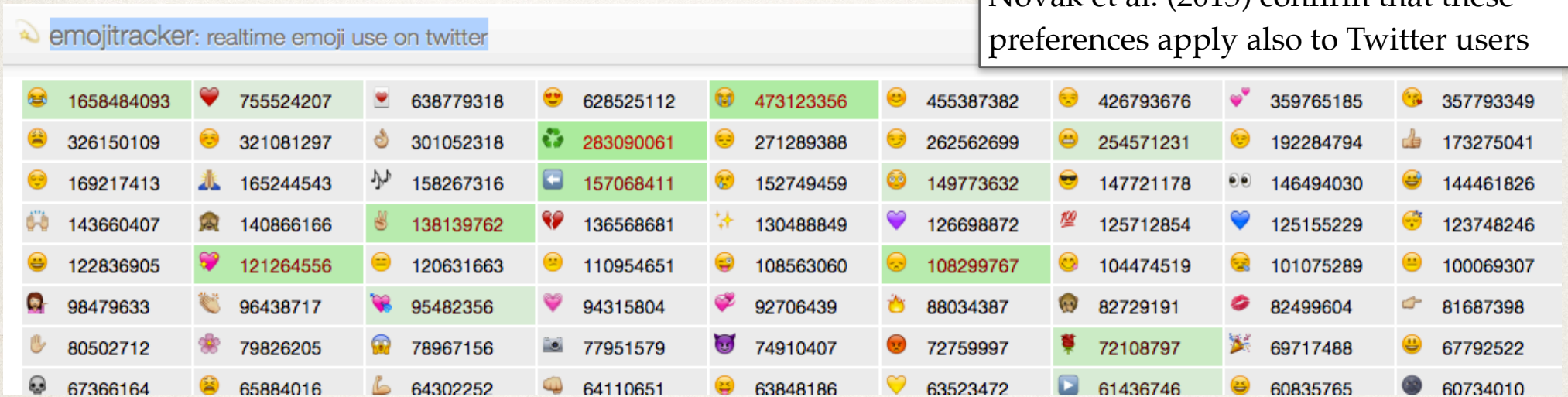


Giulia Donato, Patrizia Paggio: Classifying the Informative Behaviour of Emoji in Microblogs. LREC 2018.

Emojitracker: realtime emoji use on Twitter

- http://www.emojitracker.com/

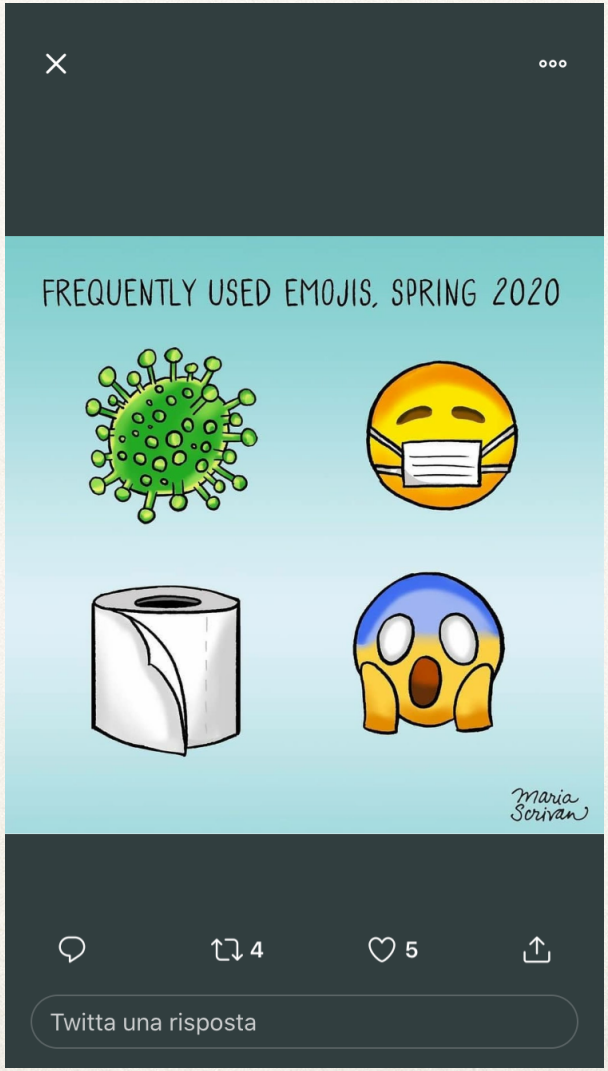
According to (Swiftkey, 2015)
the top used emoji categories are the ones that include the happy and the sad faces. Novak et al. (2015) confirm that these preferences apply also to Twitter users



- Petra Kralj Novak, Jasmina Smailovic, Borut Sluban, and Igor Mozetic. 2015. Sentiment of emojis. PloSone, 10(12):e0144296

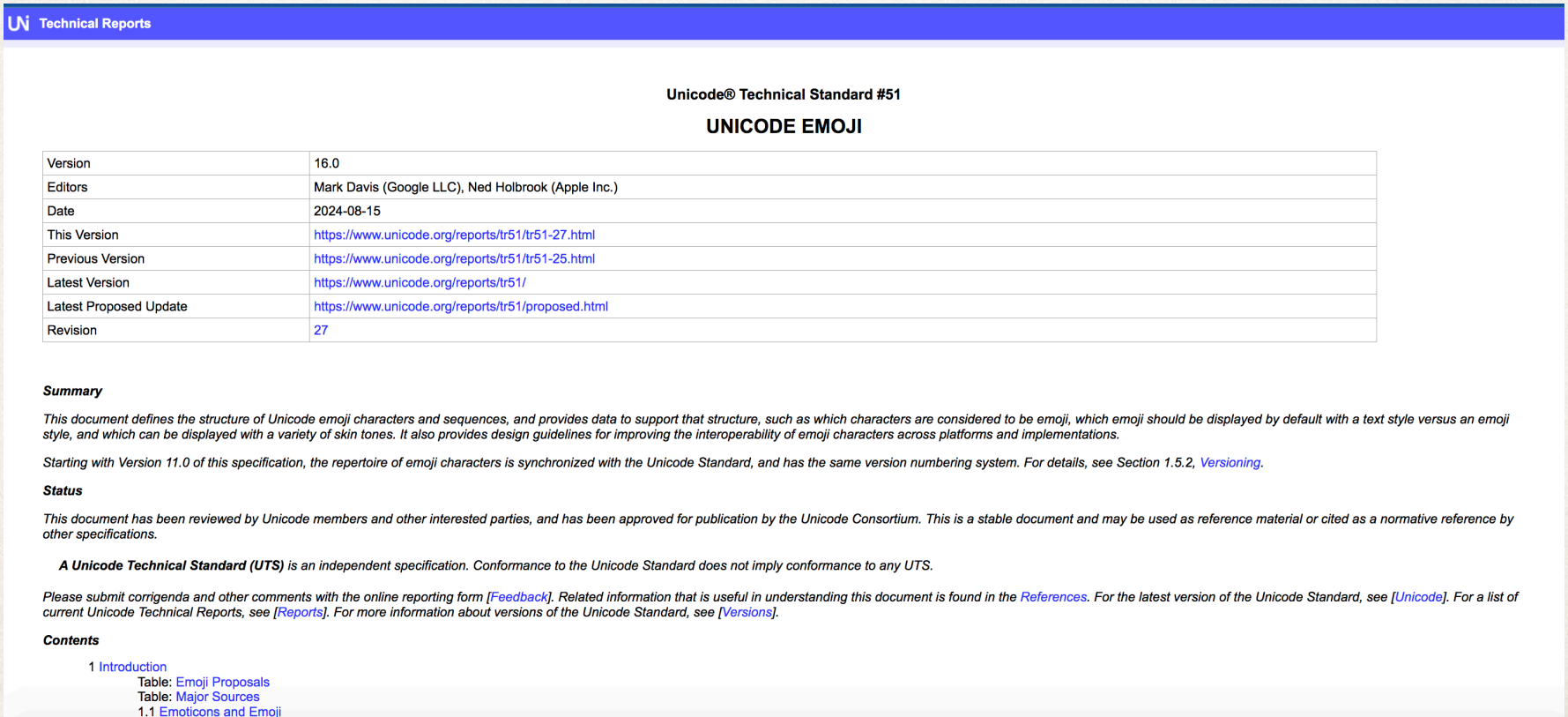
  http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0144296

Emoji e COVID-19



Unicode emoji!

- http://unicode.org/reports/tr51/
- Interoperability of emoji characters among platforms and emoji modifiers



Technical Reports

Unicode® Technical Standard #51

UNICODE EMOJI

Version	16.0
Editors	Mark Davis (Google LLC), Ned Holbrook (Apple Inc.)
Date	2024-08-15
This Version	https://www.unicode.org/reports/tr51/tr51-27.html
Previous Version	https://www.unicode.org/reports/tr51/tr51-25.html
Latest Version	https://www.unicode.org/reports/tr51/
Latest Proposed Update	https://www.unicode.org/reports/tr51/proposed.html
Revision	27

Summary

This document defines the structure of Unicode emoji characters and sequences, and provides data to support that structure, such as which characters are considered to be emoji, which emoji should be displayed by default with a text style versus an emoji style, and which can be displayed with a variety of skin tones. It also provides design guidelines for improving the interoperability of emoji characters across platforms and implementations.

Starting with Version 11.0 of this specification, the repertoire of emoji characters is synchronized with the Unicode Standard, and has the same version numbering system. For details, see Section 1.5.2, Versioning.

Status

This document has been reviewed by Unicode members and other interested parties, and has been approved for publication by the Unicode Consortium. This is a stable document and may be used as reference material or cited as a normative reference by other specifications.

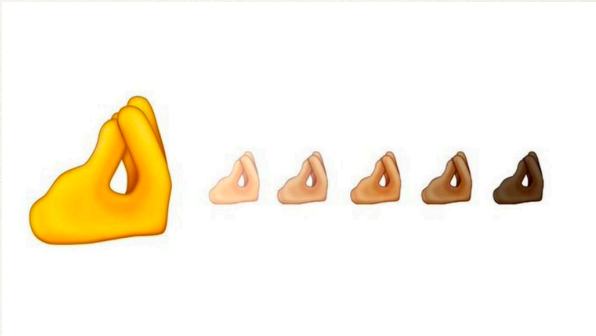
***A Unicode Technical Standard (UTS)** is an independent specification. Conformance to the Unicode Standard does not imply conformance to any UTS.*

Please submit corrigenda and other comments with the online reporting form [Feedback]. Related information that is useful in understanding this document is found in the References. For the latest version of the Unicode Standard, see [Unicode]. For a list of current Unicode Technical Reports, see [Reports]. For more information about versions of the Unicode Standard, see [Versions].

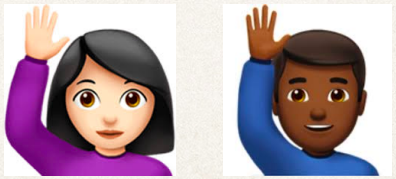
Contents

Emoji modifiers

- Emoji modifiers are features that provide more precise information of a given emoji.
- A hand-based emoji can have different skin colors: light, medium-light, medium, medium-dark, or dark.



- This information has been recently added in the official encoding of emojis https://unicode.org/reports/tr51/#Emoji_Modifiers_Table
- At the same time, some emojis like a person rising a hand could be displayed as a woman or a man



- Possible studies on role of **gender** and **skin color** in social media communication.

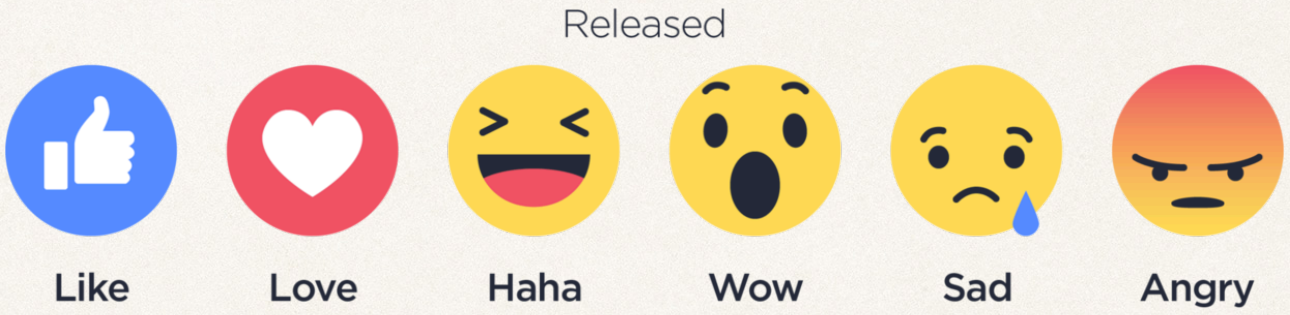
Francesco Barbieri, José Camacho-Collados:
*How Gender and Skin Tone Modifiers Affect Emoji Semantics in Twitter. *SEM@NAACL-HLT 2018: 101-106*

Recent evolutions: Memoji!

- Memoji — derived from the word “me” — extended Animoji features to avatars that you can create to look very much like you or someone you know
- Personalization (thanks for the hint during the question time!)

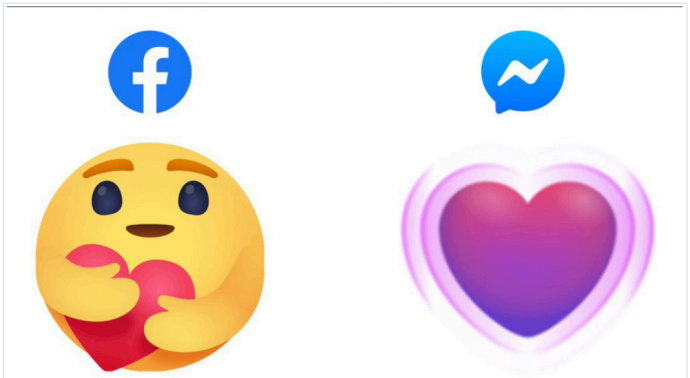
Espressive signals: Facebook emoji reactions

- Perspectives on sentiment: author of the post / community



Need a hug? Facebook releases new 'care' emoji reactions during COVID-19 outbreak

Saturday, May 2, 2020 11:04PM



Emojis in Social Media

- Computational linguistics studies the semantics of emojis and the relationship between words and emojis, which can provide fundamental information in the construction of tools for the automatic analysis and processing of content generated by users on the social web

- ITAmoji: a project for the construction of automatic tools for predicting emojis to be associated with texts on the net, in particular on Twitter.

- Idea: it is possible to deduce from a text - given the words, the syntax, the punctuation, ... - the most "probable" emoji or face to be associated, that is, the most appropriate convenient, pertinent,

  - …in some way the sign that represents its emotion.

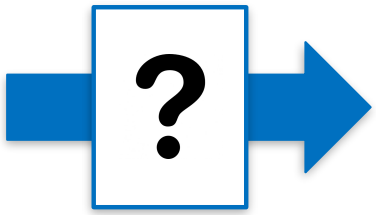
  - Not always easy!

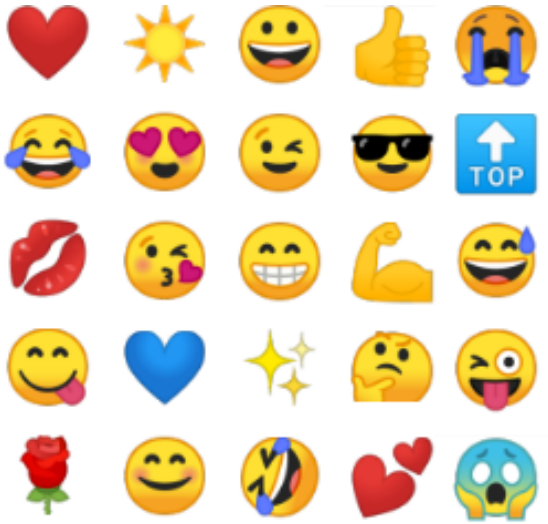
Francesco Ronzano, Francesco Barbieri, Endang Wahyu Pamungkas, Viviana Patti, Francesca Chiusaroli: "Overview of the EVALITA 2018 Italian Emoji Prediction (ITAMoji) Task". Proc. of EVALITA 2018.

Italian Emoji Prediction Task

ITAmoji

Given a Tweet, can we predict the emoji used by its author?





...by considering:

- **Tweets with Italian text messages**
- **a wide and heterogeneous Emoji space**

Emoji Prediction Task@Semeval18

- Twin Multilingual Emoji Prediction Task, organized in the context of SemEval-2018 in order to challenge the research community to automatically model the semantics of emojis inTwitter:
  - English, Spanish
- Successful! 49 teams (English subtask); 22 teams Spanish subtask.

Innamorato sempre di più 😍 [URL]

- **ITAmoji**:
  - Widening the setting for **cross-language comparisons** for emoji prediction in Twitter
  - Experimenting with **novel metrics** to better assess the **quality of the automatic predictions**
  - How **humans** perform when they are requested to identify the most likely emoji(s) to associate to the text of an Italian tweet?

NEW NEW NEW

Francesco Barbieri, José Camacho-Collados, Francesco Ronzano, Luis Espinosa Anke, Miguel Ballesteros, Valerio Basile, Viviana Patti, Horacio Saggion: **SemEval 2018 Task 2: Multilingual Emoji Prediction.** SemEval@NAACL-HLT 2018: 24-33

Italian Emoji Prediction Task

ITAmoji

Dataset

From the following two Twitter datasets:

Geo-localized

Italian tweets geo-localized in Italy ar
retrieved from October 2015 to February 2018 by the Twitter Streaming API

Newspaper followers

Tweets posted by the followers of th
top-10 most popular Italian newspaper's accounts

March 2017, Federazione Italiana Editori Giornali

We selected all the Tweets including only one emoji (eventually repeated), chosen among the following set of 25 emojis:

and, by preserving the relative Emoji frequencies, we randomly selected 275,000 tweets (62% from Geo-localized collection, 38% from Newspaper followers collection)

Italian Emoji Prediction Task

ITAmoji

Dataset

- **Training set: 250,000 tweets**

  Released on: 29th May 2018

- **Test set: 25,000 tweets**

  Released on: 3rd September 2018

  Results due on 9th September 2018

Same proportion of Tweets per Emoji (samples per label) in the training and test datasets

```
{"tid":"TWEET_ID",
"uid":"USER_ID",
"created_at":"CREATION_DATE",
"text_no_emoji":"TWEET_TEXT_WITHOUT_EMOJI",
"label":"EMOJI_LABEL"}
```

-	Emoji label	Emoji symbol	Percentage of tweets in train / test dataset
1	red_heart	❤️	20.27
2	face_with_tears_of_joy	😂	19.86
3	smiling_face_with_heart_eyes	😍	9.45
4	winking_face	😉	5.35
5	smiling_face_with_smiling_eyes	😊	5.13
6	beaming_face_with_smiling_eyes	😁	4.11
7	grinning_face	😀	3.54
8	face_blowing_a_kiss	😘	3.33
9	smiling_face_with_sunglasses	😎	2.8
10	thumbs_up	👍	2.57
11	rolling_on_the_floor_laughing	🤣	2.18
12	thinking_face	🤔	2.16
13	blue_heart	💙	2.03
14	winking_face_with_tongue	😜	1.94
15	face_screaming_in_fear	😱	1.78
16	flexed_biceps	💪	1.67
17	face_savoring_food	😋	1.55
18	grinning_face_with_sweat	😅	1.52
19	loudly_crying_face	😭	1.49
20	TOP_arrow	🔝	1.39
21	two_hearts	💕	1.37
22	sun	☀️	1.28
23	kiss_mark	💋	1.12
24	sparkles	✨	1.07
25	rose	🌹	1.06

Italian Emoji Prediction Task

ITAmoji

Evaluation metrics

Top-prediction based metrics

- Macro F1: compute the F1 score for each label (emoji), and find their un-weighted mean (exploited to determine the final ranking of the participating teams)
- Micro F1: compute the F1 score globally by counting the total true positives, false negatives and false positives across all label (emojis)
- Weighted F1: compute the F1 score for each label (emoji), and find their average, weighted by support (the number of true instances for each label)

Rank-prediction based metrics:

Finer grained metrics to evaluate the prediction quality

- Coverage error: compute how far we need to go through the ranked scores of labels (emojis) to cover all true labels;
- Accuracy@n: is the accuracy value computed by considering as right predictions the ones in which the right

Only system runs that will provide the whole ordered set of predicted emojis are evaluated with respect to Accuracy@5/10/15/20 and Coverage Error, besides F-scores

Participating teams

12 runs submitted by 5 teams

All runs submitted provided the whole ordered set of predicted emojis for each Tweet of the test set →

top-prediction and rank-prediction based metrics computed for all runs

EMOJI PREDICTION APPROACHES OF PARTICIPATING TEAMS:

- **FBK_FLEXED_BICEPS** (3 runs): recurrent neural network architecture Bidirectional Long Short Term Memory (Bi-LSTM), together with user-timeline based features
- **GW2017** (3 runs): ensemble of two models, Bi-LSTM (word2vec models based on the time creation) and LightGBM (surfaces feature extracted from tweet text like number of words, number of characters)
- **CIML-UNIPI** (2 runs) ensemble composed of 13 models (12 based on TreeESNs and one on LSTM over characters
- **sentim** (3 runs) convolutional neural network (CNN) architecture which uses character embedding as input. 9 layers of residual dilated convolutions with skip connections are applied, followed by a ReLU activation
- **UNIBA** (1 run + 1 after-deadline run) ensemble classifier based on WEKA and scikit-learn. Several features are exploited by using micro-blogging based features, sentiment based features, and semantic based features

Results

Top-prediction based metrics (including baseline systems):

Rank	Team	Run Name	Macro F1	Micro F1	Weighted F1
1	FBK_FLEXED_BICEPS	base_ud_1f	36.53	47.67	46.98
2	FBK_FLEXED_BICEPS	base_ud_10f	35.63	47.62	46.58
3	FBK_FLEXED_BICEPS	base_tr_10f	29.21	42.35	39.57
4	GW2017	gw2017_p	23.29	40.09	37.81
5	GW2017	gw2017_e	22.21	42.19	36.90
6	CIML-UNIPI	run1	19.24	29.12	31.48
7	CIML-UNIPI	run2	18.80	37.63	34.101
-	**FastText baseline**		11.96	28.72	27.02
8	sentim	Sentim_Test_Run_3	10.62	29.43	23.24
9	sentim	Sentim_Test_Run_2	10.23	31.27	23.11
-	**Weighted random baseline**		3.94	10.36	10.36
10	GW2017	gw2017_pe	3.75	11.95	10.97
11	UNIBA	itamoji_uniba_run1	3.19	27.38	15.61
12	sentim	Sentim_Test_Run_1	1.95	6.48	3.99
-	**Majority baseline**		1.35	20.28	6.84

Teams runs ranked by Macro F1, the official ITAmoji evaluation metric

Results

Rank-prediction based metrics (including baseline systems):

Rank	Team	Run Name	Coverage Error	Accuracy@5 / 10 / 15 / 20
1	FBK_FLEXED_BICEPS	base_ud_1f	3.47	81.67 / 92.14 / 96.86 / 99.10
2	FBK_FLEXED_BICEPS	base_ud_10f	3.49	81.53 / 91.94 / 96.82 / 99.17
3	FBK_FLEXED_BICEPS	base_tr_10f	4.35	74.54 / 87.50 / 94.34 / 98.00
4	GW2017	gw2017_p	5.66	67.18 / 81.49 / 89.42 / 92.99
5	GW2017	gw2017_e	4.60	71.30 / 85.90 / 94.30 / 98.25
6	CIML-UNIPI	run1	5.43	64.60 / 83.02 / 93.00 / 98.01
7	CIML-UNIPI	run2	5.11	68.46 / 83.86 / 92.38 / 97.28
-	**FastText baseline**		7.23	59.07 / 74.22 / 82.58 / 88.89
8	sentim	Sentim_Test_Run_3	6.41	58.53 / 76.93 / 88.52 / 95.74
9	sentim	Sentim_Test_Run_2	6.33	57.60 / 77.17 / 89.70 / 96.41
-	**Weighted random baseline**		6.92	59.06 / 76.11 / 86.42 / 94.10
10	GW2017	gw2017_pe	13.49	27.93 / 43.04 / 56.00 / 66.27
11	UNIBA	itamoji_uniba_run1	6.70	58.78 / 75.97 / 86.36 / 93.53
12	sentim	Sentim_Test_Run_1	12.45	29.20 / 48.78 / 64.38 / 74.04
-	**Majority baseline**		6.63	60.07 / 76.43 / 86.51 / 94.12

Teams runs ranked by Macro F1, the official ITAmoji evaluation metric

Results

Overall considerations

- Prediction quality is affected by the **variability of the context of use that characterizes a specific emoji**, more than by the number of training samples

- The choice of an emoji strongly **depends on the preferences and writing style of each individual**, both representing relevant inputs to model in order to improve emoji prediction quality

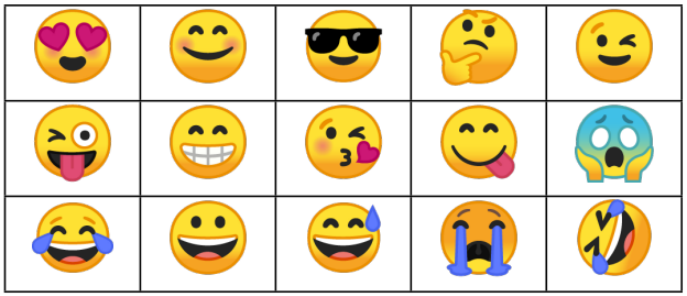
Emoji	Label	Macro F1	Num. Samples	% Samples
❤️	red heart	75.74	5069	20.28
😂	face with tears of joy	57.08	4966	19.86
💋	kiss mark	51.71	279	1.12
😋	face savoring food	48.34	387	1.55
🌹	rose	46.83	265	1.06
☀️	sun	44.69	319	1.28
😍	smiling face with heart eyes	42.93	2363	9.45
😘	face blowing a kiss	41.61	834	3.34
💙	blue heart	39.26	506	2.02
😊	smiling face with smiling eyes	38.92	1282	5.13
😀	grinning face	37.74	885	3.54
😉	winking face	34.98	1338	5.35
😁	beaming face with smiling eyes	34.47	1028	4.11
✨	sparkles	32.31	266	1.06
🤣	rolling on the floor laughing	31.79	546	2.18
👍	thumbs up	31.55	642	2.57
😎	smiling face with sunglasses	30.89	700	2.80
💪	flexed biceps	30.75	417	1.67
🤔	thinking face	29.06	541	2.16
💕	two hearts	27.48	341	1.36
😭	loudly crying face	25.62	373	1.49
🔝	top arrow	24.03	347	1.39
😅	grinning face with sweat	23.94	379	1.52
😜	winking face with tongue	23.66	483	1.93
😱	face screaming in fear	22.56	444	1.78

Table 6: Best F1 score for each emoji / label across all ITAmoji 2018 teams. The fourth and fifth columns respectively show, for each emoji, the number and percentage of test samples present in the test dataset.

Human VS automated emoji predictions

...how humans perform when they are requested to identify the most likely emoji(s) to associate to the text of an Italian tweet?

By considering the set of **15 face-emojis** (included in the 25 emojis of ITAmoji):



we asked people to automatically predict the most likely emoji to associate to a subset of ITAmoji test tweets

Human VS automated emoji predictions

ITAmoji

...how humans perform when they are requested to identify the most likely emoji(s) to associate to the text of an Italian tweet?

figure eight

(formerly CrowdFlower)

- **1,005 tweets with one face-emojis from the ITAmoji test set, perfectly balanced across the 15 face emojis**
- **64 annotators provided 6,150 evaluations by spotting the 3 most likely face emojis to associate to the text of a tweet**

Scritture Brevi

#ITAmoji che passione

- **485 tweets posted on the Scritture Brevi Twitter acco (@FChiusaroli) : reply by specifying the most likely face emoji**
- **> 100 users with an average number of valid predictions/replies per tweet equal to 5.4**

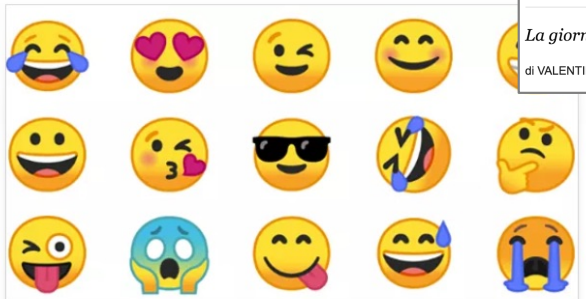
#ITAmoji che passione!

●○○○○ vodafone IT 00:06

Cerca

Francesca Chiusaroli
17 luglio ·

Parte nella giornata mondiale degli em
#ITAmoji, un gioco-test per EVALita 2018
#ITAmoji è un progetto di sentiment analy
con **Viviana Patti**, Francesco Ronzano,
Francesco Barbieri, Endang Wahyu
Pamungkas:



#ITAmoji che passione! | Scritture Brevi: il Blog
scritturebrevi.it

Scrivi un comment... GIF

Social Network

Home | News | Speciali | Mobile | Social Network | Sicurezza | Prodotti | Interattivi | Video

FESTEGGIA CON NOI! Acquista LG OLED TV E8 Ti rimborsiamo 400€ LG

World Emoji Day, così le faccine diventano una lingua

La giornata per celebrarle in tutto il mondo

di VALENTINA RUGGIU

SOCIAL NETWORK

Pinocchio Emoji, prima opera italiana con le faccine: "Così scriviamo la grammatica delle immagini"

di VALENTINA RUGGIU

testi in rete, in particolare in quelli di Twitter. "L'idea alla base - scrive Chiusaroli sul suo blog - è che sia possibile dedurre da un testo – date le parole, la sintassi, la punteggiatura e dunque il senso – l'emoji più "probabile" (cioè a dire, il più adatto, adeguato, appropriato, idoneo, conforme, conveniente, pertinente, opportuno, confacente, attinente), ovvero il segno che ne rappresenta l'emozione".ITAmoji è un lavoro di ricerca portato avanti da un gruppo di ricercatori variegato: **Francesco Ronzano**, dell'Universitat Pompeu Fabra and Hospital del Mar Medical Research Center, in Spagna, **Francesco Barbieri** dell'Universitat Pompeu Fabra, Department of Information Technologies, **Endang Wahyu Pamungkas** e **Viviana Patti** dell'Università di Torino.

In vista del World Emoji Day il quintetto ha lanciato un gioco che li aiuterà nella ricerca: dal 17 luglio, ogni giorno, verrà lanciato un tweet al quale bisognerà rispondere con emoji da scegliere tra una serie consigliata. L'hashtag da ricordare è #ITAmoji.

Francesca Chiusaroli
@FChiusaroli

And the winner is: #ITAmojiers (con grafia ji)! Sono gli appassionati annotatori del sentimento, i partecipanti a #ITAmoji, gli intrepidi costruttori del dizionario universale dei sentimenti in #emoji, con #scritturebrevi 👩🏻‍❤️‍👨🏻

Francesca Chiusaroli
28 LUGLIO

1

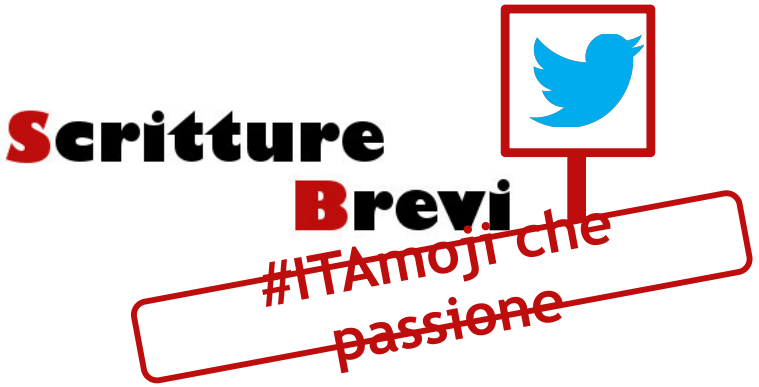
Mi piace Commenta Condividi

Human VS automated emoji predictions

Preliminary results

figure eight

(formerly CrowdFlower)



The emoji prediction performance of Figure Eight human annotators was better than 9 out of 12 systems submitted to ITAmoji (Macro F1: 24.74 - over 1,005 annotated tweets)

The emoji prediction performance of people from Scritture Brevi Twitter community was better than 8 out of 12 systems submitted to ITAmoji (Macro F1: 22.94 - over 485 annotated tweets)

Human VS automated emoji predictions

Preliminary results

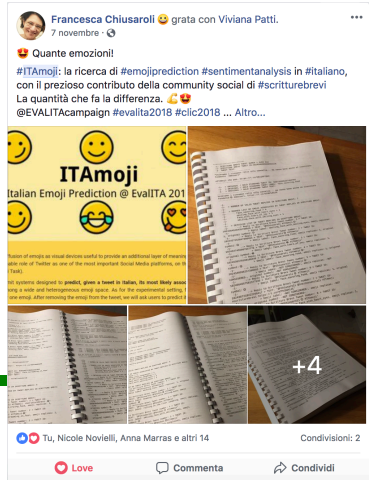
When we consider the 428 tweets annotated by Figure Height and Scritture Brevi:

Team	Run Name	Macro F1	Micro F1	Weighted F1
FBK_FLEXED_BICEPS	base_ud_1f	35.70	34.81	35.94
FBK_FLEXED_BICEPS	base_tr_10f	35.03	34.81	35.36
FBK_FLEXED_BICEPS	base_ud_10f	34.73	34.11	34.83
Figure Eight predictions		24.46	26.40	24.57
CIML-UNIPI	run1	24.03	25.00	23.65
Scritture Brevi predictions		22.94	24.06	22.99
GW2017	gw2017_p	20.40	23.13	19.97
GW2017	gw2017_e	20.33	22.66	19.83
CIML-UNIPI	run2	19.45	21.26	18.80
sentim	Sentim_Test_Run_2	12.17	15.19	11.59
sentim	Sentim_Test_Run_3	11.07	14.49	10.82
GW2017	gw2017_pe	5.01	7.48	5.02
UNIBA	itamoji_uniba_run1	2.95	7.47	2.84
sentim	Sentim_Test_Run_1	2.74	4.90	2.83

Figure Eight predictions constitute the third best performing approach, beating 9 out of 12 ITAmoji submissions
(only FBK systems that take into account emoji-preference of users obtain better scores)

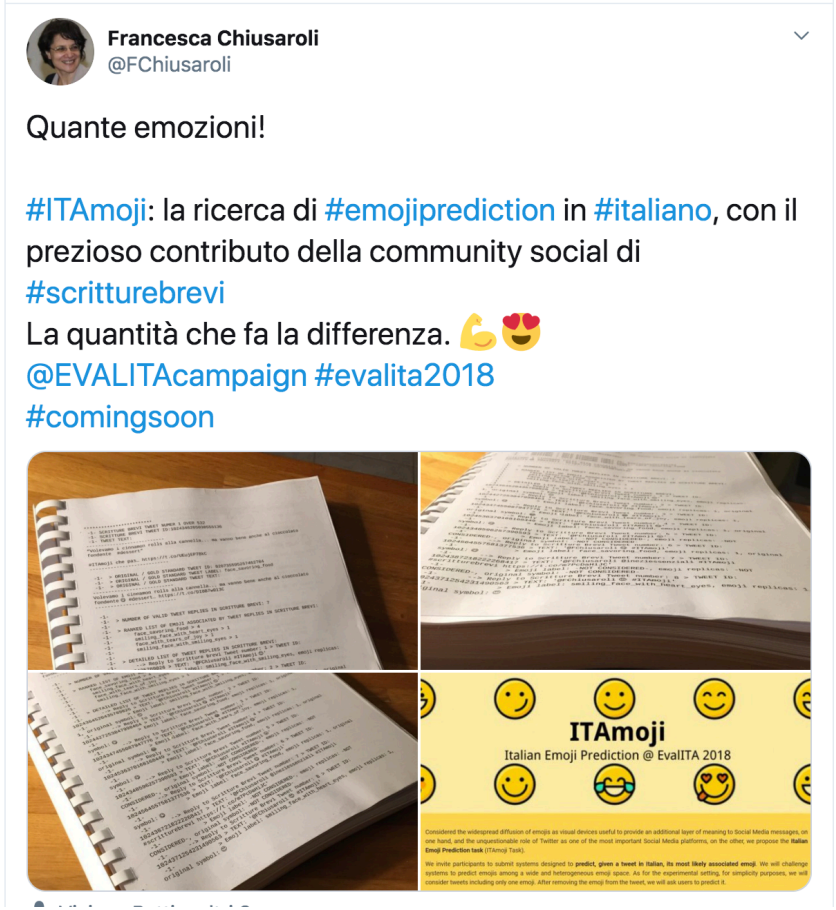
Lessons learned and next steps

- Better investigate the **effect of the different metrics considered to evaluate emoji prediction systems**
- **Cross-lingual comparison:**
  - Results in line with ones obtained in the twin shared task proposed for English and Spanish at Semeval 2018 with standard metrics **(emoji prediction is difficult!)**
  - **New experimental emoji-rank based metrics** in ITAmoji -> finer-grained evaluation of the systems' emoji prediction quality
  - **Computational models** are able to better capture the underlying semantics of emojis (similar experiment for English Barbieri et al 2017)
- **Qualitative analysis:** further explore the **comparison of human VS automated system WRT emoji prediction**, by analysing into more details the results of the Figure Eight and Scritture Brevi experiments
- Systematically evaluate the impact of **user-preferences and writing style** on emoji prediction
  - user-customized predictions?
- **Contaminations** with **author profiling**?



Lessons learned and next steps

- Qualitative analysis: further explore the comparison of human VS automated system WRT emoji prediction, by analysing into more details the results of the Figure Eight and Scritture Brevi experiments



EVALITA 2018

EVALUATION OF NLP AND SPEECH TOOLS FOR ITALIAN

Thanks for your attention!

ITAmoji - Italian Emoji Prediction

https://sites.google.com/view/itamoji

Francesco Ronzano Universitat Pompeu Fabra and Hospital del Mar Medical Research Center, Spain

Francesco Barbieri Universitat Pompeu Fabra, TALN, Department of Information Technologies, Spain

Endang Wahyu Pamungkas Università di Torino, Department of Computer Science, Italy

Viviana Patti Università di Torino, Department of Computer Science, Italy

Francesca Chiusaroli Università di Macerata, Department of Humanities, Italy

Interesting readings on emoji from the CL galaxy

- Petra Kralj Novak, Jasmina Smailovic, Borut Sluban, and Igor Mozetic. 2015. “Sentiment of emojis”. PloSone, 10(12):e0144296 http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0144296
- W17-1102: Ye Tian; Thiago Galery; Giulio Dulcinati; Emilia Molimpakis; Chao Sun “Facebook sentiment: Reactions and Emojis” https://aclweb.org/anthology/W/W17/W17-1102.pdf
- E17-2017: Francesco Barbieri; Miguel Ballesteros; Horacio Saggion “Are Emojis Predictable?” https://aclweb.org/anthology/E/E17/E17-2017.pdf
- Francesco Barbieri, José Camacho-Collados How Gender and Skin Tone Modifiers Affect Emoji Semantics in Twitter. *SEM@NAACL-HLT 2018: 101-106 https://www.aclweb.org/anthology/S18-2011/
- Donato, Giulia and Patrizia Paggio. “Classifying the Informative Behaviour of Emoji in Microblogs.” LREC (2018). https://www.aclweb.org/anthology/L18-1108/
- Francesco Barbieri, José Camacho-Collados, Francesco Ronzano, Luis Espinosa Anke, Miguel Ballesteros, Valerio Basile, Viviana Patti, Horacio Saggion: “SemEval-2018 Task 2, Multilingual Emoji Prediction” https://www.aclweb.org/anthology/S18-1003/ https://competitions.codalab.org/competitions/17344
- Francesco Ronzano, Francesco Barbieri, Endang Wahyu Pamungkas, Viviana Patti, Francesca Chiusaroli: “Overview of the EVALITA 2018 Italian Emoji Prediction (ITAMoji) Task” http://ceur-ws.org/Vol-2263/paper004.pdf https://sites.google.com/view/itamoji
- Monti, J., Chiusaroli, F., Sangati, F. (2021). Emojitaliano: A Social and Crowdsourcing Experiment of the Creation of a Visual International Language. In: Soares, M.M., Rosenzweig, E., Marcus, A. (eds) Design, User Experience, and Usability: UX Research and Design. HCII 2021. Lecture Notes in Computer Science(), vol 12779. Springer, Cham. https://doi.org/10.1007/978-3-030-78221-4_29: