

and its meanings: An emoji study

CSS Lab Holiday Paper Series 2021

Anna Di Natale, Emma Fraxanet, Max Pellert, Jana Lasser,
Hannah Metzler, Alina Herderich, Segun Aroyehun, Apeksha Shetty,
David Garcia

December 29, 2021

Non-verbal communication is not easy, but Italians are somehow better at it. See this video of a little Italian girl as an example: <https://www.youtube.com/watch?v=Z5wAWyqDrnc>. In this paper of the Computational Social Science Lab Holiday Paper Series, we set out to learn about non-verbal, non-facial communication online through the use of one of the most iconic Italian expressions: . We study its meaning in multilingual Twitter datasets and its association with sentiment to develop a tool to help using hand emoji in online communication. We further study the role of skin color in hand emoji in Reddit and the change in their frequency of use since the beginning of the COVID-19 pandemic, as an example of an *emojical study* focusing on various aspects of the use of .

Emojis offer the possibility of new non-verbal expression online. Without having to rely on pictures or videos, we can encode our facial expression and hand gestures in text. Over recent years, the frequency of emojis in online communication has been growing steadily, for example with tweets containing emojis reaching a 20% rate in 2020 [1]. As new emojis are added to include offline gestures, some cultures use them with different meanings as others. Furthermore, gestures can have a meaning in a culture and be adopted by another culture through online text communication, creating a novel kind of online assimilation of non-verbal signs.

One of the newest and most confusing emojis is , a typical Italian gesture but with other meanings and uses in other languages. The pinched-fingers emoji, as it was named by the Unicode Consortium when debuted in March 2020, was supposed to graphically represent the exasperated expression “what do you want?”, since this is the common meaning of this hand gesture when made by Italians. Its creators submitted a proposal [2] supporting the need for such an emoji, basing their arguments on the uniqueness and completeness of such a gesture, together with its high frequency of usage in the large group of Italian-speakers and people with Italian cultural affinity or ancestry. Finally, their proposal was accepted in 2020, together with another hundred new emojis. However, a conversation around other possible meanings of this gesture rooted in other cultures and languages pointed out that it can also mean “wait a minute”

in Israel and in the Arab world or represent an iconic gesture of the famous K-pop star, Yuri.

Here¹, we study the meanings of 🙌 by analyzing Twitter data to identify the contexts in which it appears. We consider various languages to detect the drivers of its use in the events of recent years. We continue by studying the sentiment associated with 🙌 and design a NLP tool to help non-native-hand-waving speakers to learn how to express themselves in emoji hand language. We then analyze how the use of general hand emojis has changed during the pandemic and how skin modifiers of hand emojis are associated with negative votes to Reddit comments.

1 The meanings of 🙌

Using the brand new historical search endpoint of the the Twitter API v2, we retrieved an exhaustive dataset of tweets in English that contain 🙌 up to 2021-11-28, ignoring retweets. To obtain a set of topics that can give a general intuition of the different uses and contexts of the emoji, we applied topic modelling with the Mallet Latent Dirichlet Allocation (LDA) algorithm in Python’s Gensim package wrapper [7].

We preprocessed the text by removing stopwords, detecting bigrams, and studying the most and least frequent words. Since the quality of preprocessing tools differs across languages, we do this first exploration only with English data. Additionally, we assume that English, in contrast to other languages, can be seen as an approximation to the different uses of the emoji in other languages when the speakers of a language communicate online with general audiences by tweeting in English. As a consequence, English Twitter text might also contain the meaning which Italians associate with 🙌.

Another important parameter in the use of the topic model is the number of topics. We explore a large range of values and compute the coherence score for each number of topics, obtaining a maximum coherence score for a model with 24 topics. For the obtained topics, reported in Figure 1, we group and describe the diverse contexts in which we encounter the “unique” (or not so much) Italian emoji. We also take into account examples of tweets that represent each topic, and compare with a smaller model with 15 topics that summarizes better some groups of topics.

The first three topics are related to Italian culture: Topic 1 is a direct representation of Italian culture and language and its relation to hand gestures and emojis. Topic 2 and Topic 3 are related to food and sports, specifically to Italian gastronomy and sport events in which Italy was heavily involved (e.g., they won the UEFA Euro 2020). Secondly, Topic 4 shows a strong connection to politics, which could be linked to complaints about regulations and political decisions. The next broad group of topics (5-9) is related to the concept of quality of different objects (art, movies, music...). The fact that the expression

¹This paper is the result of an unofficial workshop we titled “how not to suck handling emojis in our research”.

	Interpretation of the topic	Terms per Topic
0	Italian culture	italian , call, big, emoji, fan, kid, learn, speak , family, proud , brother, group , grow, young, lady, respect , exist, wife, app, culture, teach , son, parent, language, english, french , member, marry, husband, accent, spanish, gesture , describe, hand_gesture
1	Food and drinks (especially italian)	word, free, eat, cheese, food , offer, spread, pizza , wallet, sauce, pasta, cook, chicken, serve, dinner, spicy, delicious, bread, fry, breakfast
2		put, top, add, bit, nice, sweet, drink , fav, coffee, water , lil, order , extra, soft, mix, tea, pop, cake, cold, spice
3	Sports (Euro 2020)	game, win, point, team, season , hold, kiss, power, goal, pass, ball, match, player, stand, football, club , chance, score, euro, league
4	Politics	world, country, create , base, city, build, human, bear, important, lead, wrong, opinion, form, job, covid, rule, issue, state, company, control
5	Quality regarding art, books, shows, movies, series and music	nice, amazing, style, set, piece, super, pick, detail, awesome, touch, dude, design, draw , special, congrat, card, level, remind, sick, step
6		perfect, hit, fine, line, quality, hard, picture , agree, high, number, shoot , scream, summer, spot, view, mood, simple, hell, rain, shot
7		end, read, character, write, story , episode, book, finish, idea, honestly, act, easy, hard, start, entire, show, anime, drama, season, ship
8		show, movie , amazing, perfection, absolutely, favorite, scene, moment, chef_kiss, fight, question, performance, favourite, absolute, incredible, classic, answer, film, concept, iconic
9		song, music, listen , album, hear, sound, voice, stream, part, vocal, version, playlist, release, track, masterpiece, dance, vibe, favorite, fave, banger
10	Crypto currencies and money	buy, money , pay, big, sell, small, worth, spend, low, bag, price, deal, month, rich, high, run, amount, brand, bitcoin, crypto
11	Spam / Viral	follow, tweet, post, twitter, find, content, stuff, account, reply, thread, photo, meme, funny, moot, retweet, main, follower, comment, tag
12	Fandoms, K-pop	girl, boy, baby, bring, deserve, bt, body, world, kill, energy, imagine, vibe, king, pure, round, moon, child, rise, jungkook, gorgeous
13		fuck, hot, ass, guy, fucking, pic, damn, fire, sexy, date, send, vibe, single, dog, beat, dude, era, shut, daddy, namjoon
14		hair, color, black, light, side, red, blue, cut, white, short, flavor, room, green, clean, suit, dark, skin, gold, pink, jimin
15		pretty, true, take_chance, broken_heart, chefs_kiss, bestie, cool, eye, pull, mwah, talent, ur, cat, layout, literally, manifest, chef_kiss, bias
16	Negative emotions	stop, bad, lol, cry, break, literally, heart, bitch, shit, vote, joke, brain, laugh, block, forever, mad, hurt, anymore, miss, fuck
17	School deadlines and stress	night, start, morning, sleep, hour, long, tomorrow, tonight, stay, school, class, run, minute, late, wake, test, home, luck, pain, final
18-24	Non-Interpretable	Group of topics with high occurrences of family, daily actions and other very general and broad selection of words

Figure 1: Words and interpretation of the topics obtained from the Mallet LDA model. The order of the words is directly related with their relevance to the topic.

the “chef’s kiss” appears in these topics hints that 🍷 may express appreciation for good quality and excellence in these tweets. Next, we surprisingly find a topic related to cryptocurrency, in which 🍷 could mean holding a investment in cryptocurrency for future profits. We further observe a topic that we interpret as spam-related tweets, which make use of a large number of emojis to attract attention. An example of a fictional tweet that combines these two topics would be the following: “*We just minted an NFT of 🍷 and put it for sale for the affordable price of 100 ETH!! Buy it now here: <https://bit.ly/3q4IEUQ>*”².

Another surprising finding is that K-pop fandoms (or fan groups), such as Army (BTS fan group), dominate part of the usage of this emoji. Studying some of these tweets individually revealed that 🍷 can be used to replace the finger-heart gesture. The K-pop industry popularized this gesture in South Korea around the 2010s. It still does not have an official emoji, despite the strong presence of K-pop topics online. However, a new emoji for the finger-heart gesture will be released in the upcoming 2022 official emoji update. It would be interesting to study whether this context of 🍷 still appears a while after this release, or if its meaning finally returns its original one. Two further well-defined topics relate to negative emotions. The first includes insults in general and the second one is particularly linked to stress due to school or university deadlines, and possibly includes complaints. The other topics found in this classification were not easily interpretable, and dissolved into the topics described above when the model was set to 15 topics. We chose the version of 24 topics due to its higher coherence score and because we like it and this is a Holiday Paper Series article after all.

2 Differences across languages

Thanks to the new counts endpoint of the Twitter API v2, we can track the frequency of an emoji within a language over time. We calculated the daily count of tweets in six languages (English, German, Spanish, Italian, Japanese, and Korean) that contain 🍷, dividing the count over a baseline that we estimated with the daily count of tweets in those languages that contain at least one stopword. We excluded retweets in all these counts to measure frequency of expression and not in sharing text. We then smoothed the daily counts with a seven-day rolling average. Figure 2 shows the time series for Italian, English, and Korean with spikes highlighted. Using our access to Brandwatch, we generated word clouds including retweets³ on the peaks to get an idea of the words driving the increase in use of 🍷.

The first Italian spike is driven by users complaining about politics and the big brother TV show (‘gf’ stands for grande fratello, where Rosalinda is one participant). The second peak is related to the final match Italy vs England in the soccer Euro Cup. Here, the English hashtag #itscominghome referring

²We promise we will use the earnings of this NFT for future research.

³We know that wordclouds are the pie charts of text analysis, but we didn’t have time to make anything fancier. And we like them after all.

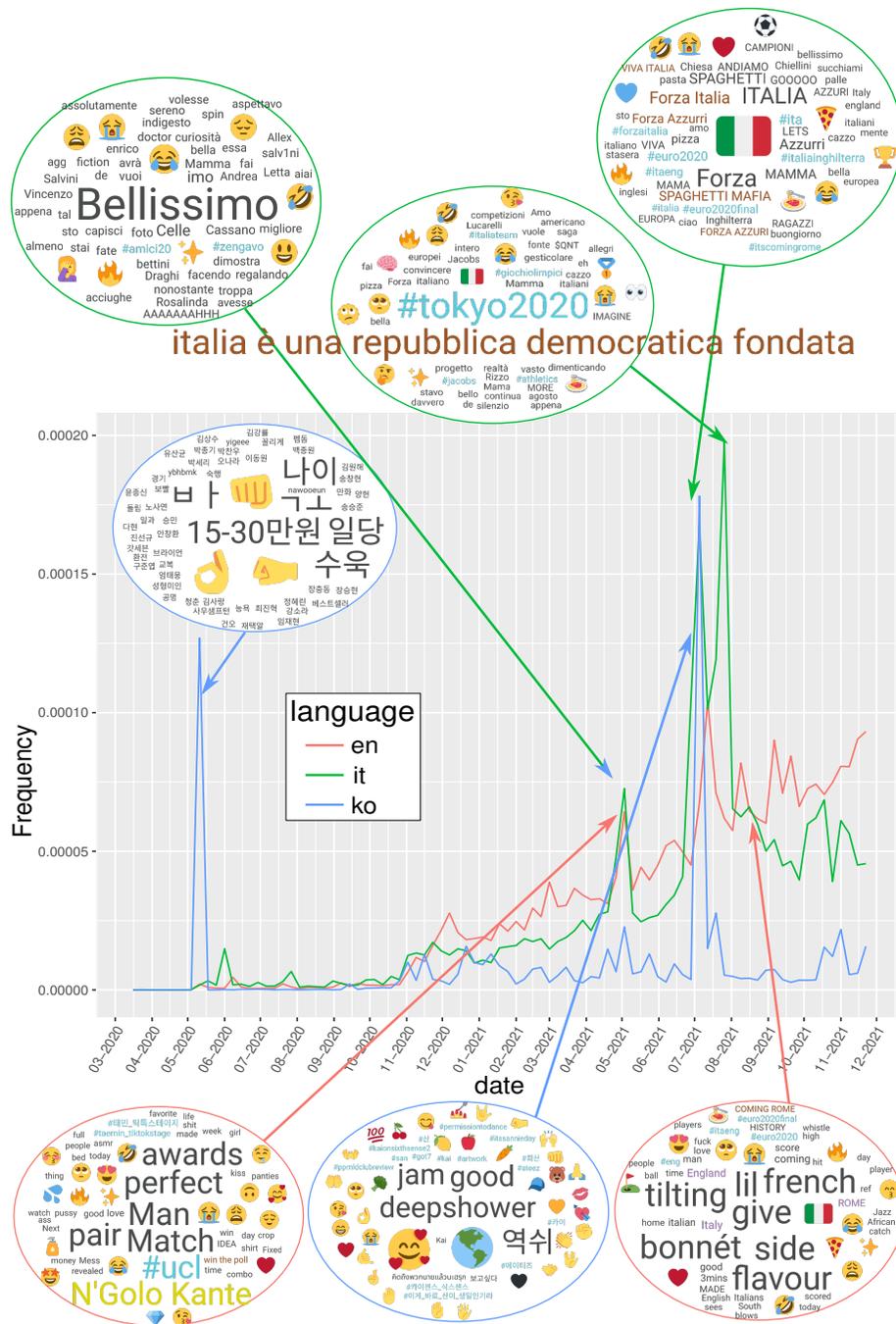


Figure 2: Frequency of usage of 🍷. Peaks are reported together with the relative wordclouds. The frequencies are plotted starting from March 2020 when the emoji was introduced.

to the cup was dubbed into #itscomingrome. Note how terms associated with Italians (both iconic behavior and stereotypes) are also used by Italian speakers (pizza, pasta, spaghetti, even mafia). The third spike is related to the Italian successes in the Olympic games, especially to the winner of the men’s 100m run, the Italian Jacobs. The enthusiasm about the Italian sportive successes, together with Italians making fun of themselves, is rendered in this popular tweet: <https://twitter.com/gocciodigin/status/1421835234023202818>. As we saw, the pinched fingers in Italian are used in tweets about the Euro 2020 football cup and the Olympics of Tokio 2020, showing its connection to national pride.

The first spike in English relates to a match of the UEFA Champions League (UCL). We didn’t manage to understand the association between this event and 🤞. The second spike in English also relates to the final match of Euro 2020, including the reference to #itscomingrome.

The first peak in Korean is driven by viral and spam-like tweets about money, using the meaning of 🤞 about money or wealth (as in holding a coin). The second Korean peak is linked to a song by the Korean DJ Deepshower, where 🤞 is used with the meaning of “quality”. In Korean, 🤞 is associated to many other hands emojis (see word clouds), while in Italian and English it is not.

A closer inspection of the wordclouds in Figure 2 shows different kinds of sentiment for different languages. To test this, we classified the tweets including the emoji at least once with a multi-lingual XLM-RoBERTa model fine-tuned for sentiment classification [3], which was trained on mostly on languages written in Latin characters. Figure 3 shows the fractions of tweets classified as positive, negative, and neutral across languages. We can see that the most positive language is Japanese, followed by Turkish and French. We identified many positive French tweets where 🤞 is related to quality, especially in the context of food. The most negative languages in the context of 🤞 are Spanish and Italian, referring to the meaning of exasperation that motivated the initial proposal of including the emoji in the Unicode standard.

3 Emoji tech: helping with emoji expression

For potential funding reasons, we also address the burning question of how Information and Communications Technologies can enable people to have better experiences with hand emojis in Italian by advising them on situationally-aware hand reactions. Our integrated, interactive, computationally-assisted approach is exclusively validated on synthetic, complimentary, native and non-native speaker examples ⁴. We came up with as many examples as we could without being distracted too much from other work tasks like planning an office Christmas party, which was then cancelled because of the restrictions related to the COVID-19 pandemic. This makes it a strong candidate for high performance on benchmark tasks (to be developed in a series of papers later) and of very little relevance outside of this stylized domain, as it is common in many

⁴Our buzzword sentence generator broke when producing this text

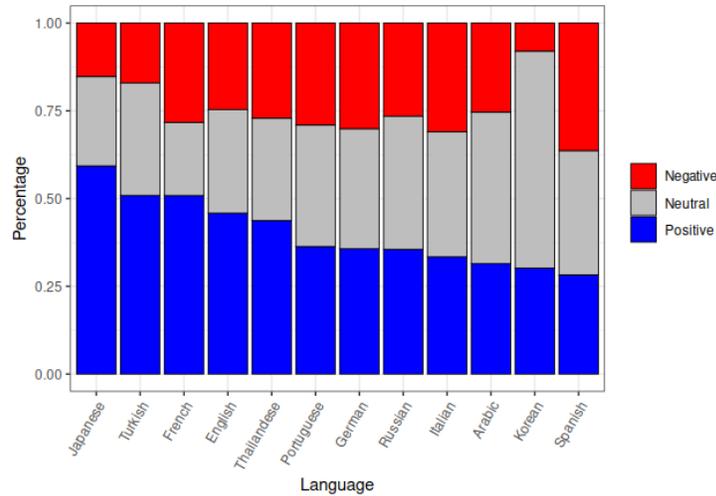


Figure 3: Distribution of sentiment of Tweets containing 🍌 across languages.

Machine Learning scenarios. In short, we substantially contribute to the current paradigm in NLP research but also contribute to the social sciences: Our approach is also grounded in deep theory, we promise.⁵

We use BERTino [8], a large-scale model pretrained on Italian text and fine-tune it on the FEEL-IT emotion classification dataset [5]. Our emotion classifier⁶ can detect 4 discrete emotions: joy, fear, anger and sadness. We combine the three negative emotions to one general negative class and map joy to positive. The preprocessing of the text cleaned it of many words that can not be legally passed as inputs to Italian language models because of blasphemy laws. BERTino-feel-it allows us to feed any Italian text into it and to recommend a fitting hand emoji to be added to the text. It produces the following (stereo)typical examples:

- “Odio il modo in cui la regione vicina prepara il piatto”/“I hate the way the neighboring region is preparing the dish”. Our classifier rightly sees the anger here and recommends the following emoji: 🍌
- “L’Italia è super campione del mondo in tutti gli sport”/“Italy is the super world champion in every sport”. The joy and national pride contained in this sentence is correctly translated into 🍌.
- “Non apriamo prima di mezz’ora, non importa cosa dice il cartello”/“We don’t open for another half hour, no matter what the sign says”. This example shows that our classifier is also able to capture fine-grained nuances of emotional expression in natural languages by choosing 🍌.

⁵On request, we also share Zotero libraries of 1000s of papers for intimidation purposes.

⁶The model is available from the authors on unreasonable request.

4 Hand and face emoji after mask mandates

During the COVID-19 pandemic, many countries have introduced several kinds of mask mandates, making masks a very usual sight in our everyday offline interactions. This reduces expressive power in facial non-verbal expression, making it harder to communicate through smiles or other signs that are partially or completely covered by masks. In our everyday life, we noticed that some of us use our hands more to communicate when using a mask, for example to express approval or disapproval. This is especially useful when masks muffle verbal expression or in very noisy environments. We thought that this could have an effect in how we choose emojis online: is it possible that hand emojis became more frequent over time compared to face emojis since the beginning of mask mandates during the COVID-19 pandemic?

Using the counts endpoint of the Twitter API v2, we counted the frequency of several hand emojis and facial expression emojis. We excluded the new emojis that appeared after the beginning of 2019 from these counts. Emojis related to the pandemic, like the face with a mask emoji 🤒, are included in the count of face emojis. This allowed us to calculate a frequency of hand emoji per face emoji on a weekly basis over time, in particular in the three years between December 2018 and December 2021. Figure 4 shows the weekly ratio of hand to face emojis, with a vertical line marking March 2020 as the moment when COVID-19 was declared a pandemic. You can see a change in the frequency of hand to face emojis, from values below 0.4 during 2019 to values closer to 0.45 in 2021. Note that we find this pattern considering all emojis, also the ones related to the pandemic, but the results are similar when excluding pandemic-related emojis like 🤒 and 🤧.

To test if the pandemic has increased the relative frequency of hand to face emojis on Twitter, we fitted a linear model of the trend of this ratio with data prior to January 2020 and extrapolated after the beginning of the pandemic. The blue line in the figure shows the prediction trend, which is substantially below the actual value of the ratio afterwards. This observation is consistent across languages too. We computed the model and measured the mean residual of the model in its extrapolation since March 2020 as a measure of the effect of the pandemic on this ratio. Table 1 reports the effect for the six languages we considered, both in terms of mean residual and as a percentage increase over the previous hand to face emoji ratio. You can see that the effect is heterogeneous across countries, with some experiencing large increases (e.g. English with 24% and Korean with 39%) but with Japan experiencing no significant increase. All increases were significant in t-tests except for Japanese, as in Japan the use of face masks in public was more frequent before the pandemic than in other countries.

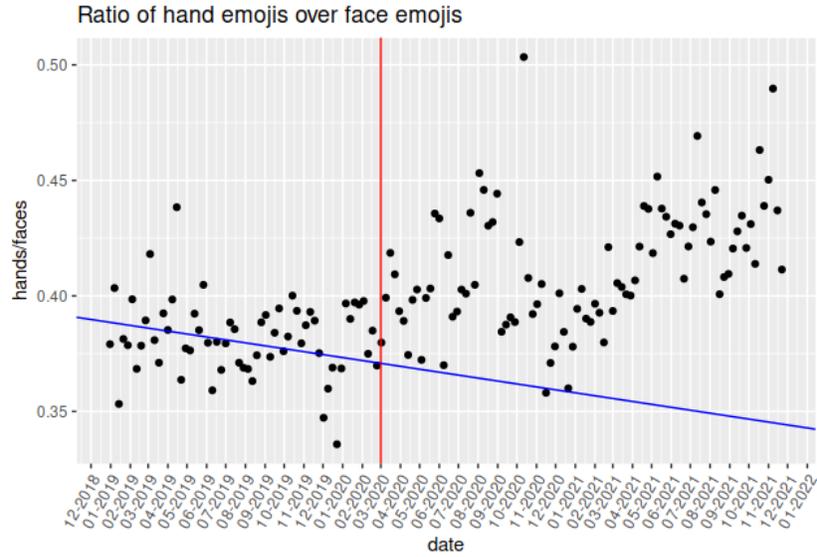


Figure 4: Ratio of hand emojis to face emojis on the 6 languages considered. The blue line indicates a linear model fitted on the data prior to January 2020. The red line indicates the moment when COVID-19 was declared a pandemic. Counts are aggregated on a weekly frequency.

Language	mean residual	percentage
Italian	$3.26 \cdot 10^{-4}$	15%
German	$5.60 \cdot 10^{-4}$	22%
Spanish	$4.63 \cdot 10^{-3}$	8.4%
Korean	$5.63 \cdot 10^{-3}$	39%
Japanese	$2.31 \cdot 10^{-4}$	0.2%
English	$4.53 \cdot 10^{-2}$	24%
All	$5.67 \cdot 10^{-2}$	14%

Table 1: Mean residuals of the prediction model for the frequency of hand emojis in 2019 computed on data from 2020 and 2021. P-values of the t.test are all $< 10^{-5}$ apart from Japan, which has $p = 0.8$.

5 Not all hands are equal on Reddit

Hand emojis can be adapted with skin color modifiers to match the skin of the person using them. Unlike other emojis, where only a small area is affected by the skin modifier, hand emojis change in color as they mainly show the skin of a hand. On Reddit, the subreddit `r/shitposting`⁷ had a running meme of how the darkest thumbs up emoji gets downvoted in comments in comparison to other skin colors of the same emoji⁸. Based on this observation, we test the *shitposting hypothesis*: comments on Reddit with dark skin emoji modifiers are more likely to have a negative score than comments with the same emoji but without the skin modifier or with light skin colors.

We analyzed the Reddit comments dataset of Pushshift.io⁹ [4] for one year between July 2020 and June 2021. We found 3,306,680 comments (in any language) with 👍 regardless of skin modifier. To test the shitposting hypothesis, we fitted a logistic regression model with a dependent variable that takes value one if the comment had a negative Reddit score (more negative votes than positive votes) or zero otherwise. As independent variables, we coded binary variables that take the value 1 if a skin color emoji is modifying the 👍 emoji in the comment. The result of the model can be seen on the left panel of Figure 5, where we show the predicted probabilities of a comment having a negative score for the cases of containing at least one emoji of each skin color. We find that the darkest skin tone is associated with approximately double probability of having a negative score than the other skin colors, which don't substantially differ in their skin probability. It appears that this result agrees with the hypothesis: a comment with 🖤 is substantially more downvoted than one with other modifiers.

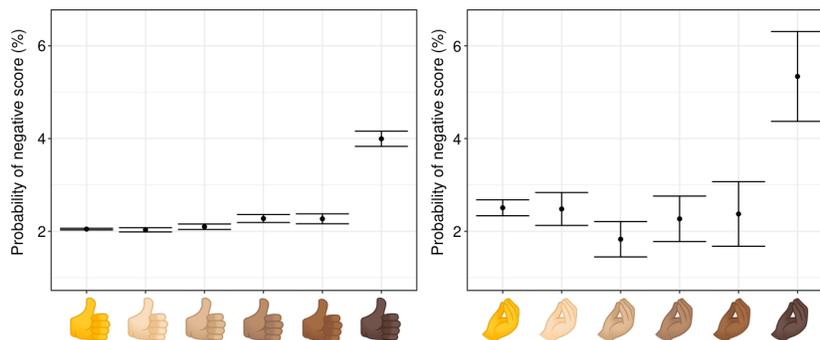


Figure 5: Predicted probability of a Reddit comment having a negative score in the logistic regression model of skin color modifiers for comments with 👍 (left) and with 👍 (right). Error bars show 95% Confidence Intervals.

⁷Visit this subreddit at your own risk. It is full of politically incorrect memes and can contain NSFW content.

⁸<https://www.reddit.com/r/shitposting/comments/r5peim/racism/>

⁹<https://files.pushshift.io/reddit/comments/>

We repeated the above analysis for comments in the dataset that contain 🍌, finding 49,025 comments in total. The results of the same model are shown on the right panel of Figure 5, revealing a similar pattern as for 🍌: the probability of receiving a negative score is similar across skin colors but jumps to nearly double the value for the darkest skin color.

To explore how generalizable this effect is across all emojis, not just hand emojis but also ones including faces, we repeated the analysis over a sample of 120 Million Reddit comments including all comments and coding the case in which there is no skin modifier at all. The result is shown in Figure 6. As in the cases of 🍌 and 🍌, the probability of a comment having a negative score is approximately double when it contains the darkest skin color modifier as opposed to not having any skin color modifier. This large dataset allows us to see some smaller differences between skin tones, where the second lightest skin tone has the same probability as without any skin color modifier and the other three have a slightly larger probability.

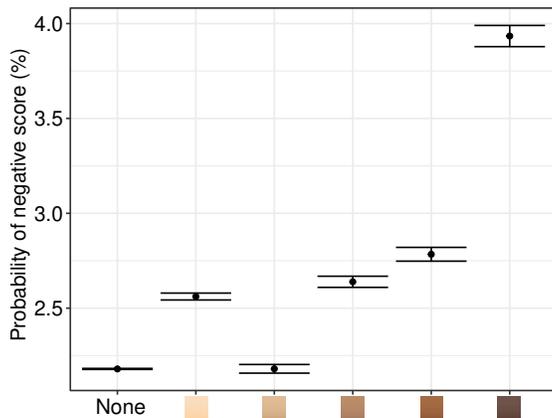


Figure 6: Predicted probability of a Reddit comment having a negative score in the logistic regression model of general skin color modifiers for all emoji, not just hands. Error bars show 95% Confidence Intervals.

Note that the lightest modifier, especially when applied to some face emoji, generates a figure with dark hair (for example 🧑), while the second lightest generates faces with blonde hair (for example 🧒). Previous research shows that the lightest modifier is used widely in East Asia [6]. Describing the origins of this effect requires a more serious research paper. Please take these insights as the result of a very initial exploration on an important topic that deserves a careful analysis. Future research should try to identify if it can be attributed to racism or it could be a response to some kind of blackface behavior, despite being found to be rare on Twitter in previous research [9]. We should especially test if the effect is mostly present on some subreddits and include a language model for the rest of the text that can serve as a set of control variables to isolate the effect of the skin color modifier alone.

6 To sum up

- We find that 🍷 can be used in various contexts, including Italian culture, K-Pop, and cryptocurrency spam.
- Language differences are clear too, with Japanese being the most positive language in the use of 🍷 and Spanish and Italian being the most negative.
- A combination of language models can be used to infer the hand emoji to add to a tweet as shown by our inspection of predictions for Italian.
- The fraction of hand to face emojis has increased since the beginning of the pandemic reflecting the need for hands in non-verbal expression during mask mandates, also online.
- Reddit comments with dark skin modifiers have approximately double the probability of getting a negative score, not just for 🍷 and 👍 but also for other uses of the skin color modifier.

Thanks for reading and all the best for 2022!

References

- [1] Emoji trends that defined 2020. <https://blog.emojipedia.org/emoji-trends-that-defined-2020>. Accessed: 2020-12-19.
- [2] Jennifer Lee Adriano Farano and Theo Schear. "what do you want?" pinched fingers emoji proposal. <https://www.unicode.org/L2/L2019/19159-pinched-fingers-emoji.pdf>. Accessed: 2020-12-19.
- [3] Francesco Barbieri, Luis Espinosa-Anke, and Jose Camacho-Collados. A Multilingual Language Model Toolkit for Twitter. In *arXiv preprint arXiv:2104.12250*, 2021.
- [4] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, volume 14, pages 830–839, 2020.
- [5] Federico Bianchi, Debora Nozza, and Dirk Hovy. FEEL-IT: Emotion and sentiment classification for the Italian language. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 76–83, Online, April 2021. Association for Computational Linguistics.
- [6] Steven Coats. Skin tone emoji and sentiment on twitter. *arXiv preprint arXiv:1805.00444*, 2018.

- [7] Andrew Kachites McCallum. Mallet: A machine learning for language toolkit. <http://mallet.cs.umass.edu>, 2002.
- [8] Matteo Muffo and Enrico Bertino. BERTino: An Italian DistilBERT model. In Felice Dell’Orletta, Johanna Monti, and Fabio Tamburini, editors, *Proceedings of the Seventh Italian Conference on Computational Linguistics CLiC-it 2020*, pages 317–322. Accademia University Press, 2020.
- [9] Alexander Robertson, Walid Magdy, and Sharon Goldwater. Self-representation on twitter using emoji skin color modifiers. In *Proceedings of the International AAI Conference on Web and Social Media*, volume 12, 2018.