Journal of language works

Vol. 8, No. 1, 2023

Emojis as Graphical Hate Speech Markers

Sofiia Lazareva
Aarhus University
202203549@post.au.dk / slazareva10@gmail.com

Abstract

This article examines the functioning of emojis in a hate speech context. It investigates whether emojis are an independent mechanism of hate speech or if they serve as an auxiliary device. By this, I mean if emojis can constitute a hate speech message. A corpus containing hate speech from three different Russian social networks is analyzed: the Russian Toxic Comments corpus from Odnoklassniki and two self-collected corpora from Twitter and VK. The findings suggest that emojis do not function as a separate mechanism of hate speech, but are rather used to clarify the intended meaning of the message. The study also highlights that the usage of emojis is context-dependent, specifically in terms of whether the user is involved in a confrontation with another user.

1. Introduction

This article is dedicated to a sociolinguistic study of emojis as a hate speech marker. This means that emojis often accompany hateful messages in online discourse and is a feature among others that can potentially be used to detect them. To study emoji functioning in a hate speech context, I analyze a corpus containing data from Russian popular social media (VK, Odnoklassniki, and Twitter). I have chosen these social media sites since they are among the most popular in Russia. The usage of emojis is becoming more and more extensive. Data from social media suggest that every fifth tweet contains at least one emoji, and more than 700 billion emojis are used in Facebook posts every day (Kaiser, Grosz 2021: 1010). At the same time, there are only a few linguistic studies on how emoji function in different contexts and which purposes they serve (e.g., Cohn et al. 2019; Danesi 2016; McCulloch 2019). Learning how emojis function in the context of hate speech can contribute both to computer-mediated discourse studies and to automatic hate speech detection which primarily utilizes lexicon-based approaches. Emojis can for example be used to increase accuracy of automatic hate speech classifiers.

2. Literature review

2.1 Hate speech definitions

There is no clear linguistic definition of hate speech. Paz et al. define it as a willful public statement aiming at discrediting a group of people (Paz et al. 2020: 1). Warner and Hirschberg, on the other hand, describe hate speech as an insulting statement which uses stereotypes to spread hateful ideologies (Warner & Hirschberg 2012: 1). Powell at al. point out that hate speech does not only exist in the form of oral and written texts but also as videos, pictures, and sound (Powell et al. 2018: 10). Fortuna and Nunez expand it by including humor as a separate hate speech type (Fortuna & Nunes 2018: 5). Humor becomes essential for understanding online hate speech because, first of all, offensive humor is what many online communities spreading hate speech are based on, and secondly, hateful messages formulated as humorous utterances or media are harder to detect and

report. Apart from this, humor, just like all the other hate speech forms, is used to reinforce harmful stereotypes about the targeted groups and can also influence targeted individuals' mental state. Even though the definition of hate speech is still subject to debate, researchers agree on its degrading and dehumanizing functions (Powell et al. 2018: 10).

Turning to how hate speech is expressed in online settings, Lingam and Aripin (2017: 113) and Ruzaite (2018: 102) consider hate speech to be degrading comments, rumors, defamation, insults, dehumanizing statements, sarcasm, threats, harassment, trolling, and incitement to discrimination and violence. There are also specific forms of expression for different hate speech types. Racist hate speech is spread through stereotyping, racist humor, and creation of racist online communities (Castaño-Pulgarín et al. 2021: 4). A typical mechanism for sexist hate speech is public condemnation, for example slut shaming (Sundén & Paasonen 2018: 643). Hateful opinions arising from political disagreements are often distributed in the form of analytical comments where special rhetoric figures and pejoratives are used (Trajkova & Neshkovska 2018: 304).

In this paper, I take hate speech to mean all utterances matching the UN's definition: "...any kind of communication in speech, writing or behavior, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, color, descent, gender or other identity factor". Hate speech is a very complex phenomenon which is hard to reveal without making reference to the context and groups involved in the explanation.

2.2 Hate speech in Computer-Mediated Communication

Hate speech has become more prevalent with the common use of computer-mediated communication. Computer-mediated communication has several characteristic features which makes it especially prone to hate speech. López and López distinguish four of them:

- anonymity of all the participants;
- invisibility of the aggressor;
- an easy access to communities spreading hateful ideologies;
- instantaneousness of spreading (López & López 2017: 11).

Anonymity of all the participants and invisibility of the aggressor create an illusion of impunity which, combined with means of instantly spreading aggression and hateful ideology, provides a fruitful soil for promoting hate speech.

2.3 Hate speech strategies

Hate speech strategies should be categorized in order to identify hate speech on different levels. Van Dijk distinguishes the following strategies while studying ethnic minorities discrimination:

a. overgeneralization: this strategy is used when someone is trying to prove that a negative accident that has been subject to discussion is actually a systematic phenomenon
b. examples: this means appealing to anecdotal experience
c. corrections: the speaker corrects themselves to manipulate implications and the associations that an interlocutor has

d. emphasizing: rhetorical intensification and usage of expressive formulas aiming at attracting interlocutor's attention e.g., "it's terrible that…"
e. concessions are used when trying to create a positive self-representation and demonstrate tolerance and sympathy to later have an opportunity to overgeneralize if counterexamples are presented. A typical example is, for instance, "but not all of them are like that, of course"
f. repetition: this strategy is also used to draw interlocutor's attention
g. contrast is often based on a juxtaposition of two groups: "us" and "them". An example would be "the immigrants can simply live on welfare, and we have to work hard to provide for ourselves"
h. mitigation of original statements is similar to concessions in the sense that it is also used to create a positive self-representation, but in this case the goal is to prevent negative inferences
i. displacement is trying to present the speaker's opinion as common knowledge or someone else's opinion, e.g. "I don't mind it personally, but many others are offended by it"
j. avoiding the conversation in order not to reveal a lack of knowledge or not to admit the inconsistencies in speaker's arguments or inferences
k. presuppositions, implications, suggestions, and indirect formulations are used to avoid stating specific propositions (van Dijk 1983: 398-400).

There are also specific hate speech strategies used in mass media, namely delegitimization (i.e., denying legitimacy and excluding certain groups from society), marginalization (i.e., unequal access to mass media and misrepresentation), demonization and exploitation of harmful stereotypes about discriminated groups as the primary ones (Gladilin 2013: 145, 148; Jakubowicz 2006: 602).

Related to hate speech is the case of internet conflicts. Since hate speech is abundant in online conflicts, one may imagine that similar features between the two will arise. Shulginov and Tillabaeva's study on Internet conflicts indicates that the characteristic lexical features in conflicts are an extensive use of pejoratives and expressives, imperatives, and various rhetorical figures (such as making a distinction between the opposing sides). Syntactically aggression can be expressed through short messages consisting of one or two sentences (Tillabaeva & Shulginov 2020: 54-55).

2.4 Semantic functioning of emojis

Understanding how emojis function in terms of semantics can shed light on whether they can be used as a separate hate speech strategy or merely an instrument. By strategy, I mean whether emojis can constitute an independent, clear-cut hate speech message, and by instrument I mean whether emojis serve the purpose of clarifying the intent or tone of the message.

The approach that I have chosen is hybrid semantics. The main idea of hybrid semantics is that truth conditions of a sentence are different from its use conditions, so there are items that can express only truth or use conditions or both (Gutzmann 2015: 21-40). For example, the sentence "This jerk Kaplan was promoted" expresses both truth conditions because it contains information about Kaplan's promotion and use conditions because it expresses the speaker's negative attitude to Kaplan (Gutzmann 2015: 21).

Emojis are similar to what Gutzmann calls "isolated expletive expressions" which are, essentially, items expressing use conditions that do not require any syntactic arguments (Gutzmann 2015: 40). Kaiser and Grosz consider emojis from a similar perspective. They propose that there are two main types of emojis: face and non-face (action) emojis, and that the emoji type defines how it functions in terms of pragmatics. Face emojis in the beginning or at the end of a message describe the experiencer's attitude towards the referent mentioned in the message, and non-face (action) emojis function like first- and third-person pronouns (Kaiser & Grosz 2021: 1012, 1017). This is similar to Gutzmann's approach in the sense that face emojis also act like isolated expletive expressions since they only express use conditions (the attitude to the referent).

McCulloch and Gawne (2018: 2). suggest that emojis are to be viewed as beat gestures — the kind of gestures that accompany speech. Beat gestures, unlike symbolic gestures, do not express anything on their own and are rather used to highlight the tone of the utterance. Emojis, like beat gestures, often occur at the end or at the beginning of the message acting like a specific type of emotionally charged punctuation (Dürscheid & Siever 2017: 8; Na'aman et al. 2017: 140). What is more important, however, is the fact that emojis, like beat gestures, do not form grammars and do not have the same structure as written or oral speech and act "in relation to the written text" (McCulloch & Gawne 2018, 3).

2.5 Grammatical position of emojis

Speaking of grammatical functioning of emojis, the first thing worth mentioning is that they usually appear at the end or at the beginning of the sentence. This is consistent with the role of emojis as emotionally charged punctuation. Dürscheid and Siever's study of the Swiss What's App corpus indicates that emojis are used to emphasize the intonation and the tone of the sentence (Dürscheid & Siever 2017: 8). Na'aman et al. (2017, 140). propose a similar approach as they consider emojis to be paralinguistic emotional markers, and this conclusion is supported by Kralj Novak et al.'s Twitter corpus study: They found that the greater sentiment score an emoji has, the more likely it is to be placed at the end of the sentence (Kralj Novak et al. 2015: 11).

Cohn et al. found that nouns and adjectives are most likely to be replaced with emojis compared to other parts of speech (Cohn et al. 2019: 14). Expressions that only consist of emojis are usually combined on the associative basis. Another option is reduplications (Cohn et al. 2019: 9). McCulloch and Gawne have also found that the most common emoji n-grams are usually reduplications or conceptually linked sequences (McCulloch & Gawne 2018: 3).

Summing up, we have seen two different perspectives on emojis. On the one hand, they appear in sentence-initial or sentence-final positions and function as a special type of emotional or tonal markers that do not require syntactic arguments which makes them isolated expletive expressions. Emojis are likely to form sequences consisting of emojis combined on associative grounds or reduplicated emojis. On the other hand, they may in some cases be integrated in the sentence to substitute certain sentence constituents, namely noun and adjective phrases.

3. Materials and methods

3.1 Corpus

The corpus used for this study is compiled of hate speech utterances from different sources. Three social media subcorpora were used for this study:

a. Russian Toxic Comments (containing comments from "Odnoklassniki")
b. A self-collected Twitter corpus
c. A self-collected VK corpus.

These social media sites were chosen due to their popularity among Russian-speaking users. They also allow to create communities, including communities based on hateful ideologies, and enable users to share their opinion in a public forum. This is reinforced by the instantaneous spreading of all kinds of messages on these platforms and the fact that promotion algorithms that these platforms use prioritize posts which are actively being commented on.

Russian Toxic Comments is a dataset containing comments from a Russian social network "Odnoklassniki" which was gathered as part of an internal hate speech detection campaign and then marked up. "Odnoklassniki" is a popular Russian social network used primarily by elderly citizens. Each comment was assigned a tag: "normal", "insult", "threat", and "obscenity". The dataset was filtered using the *pandas* library for Python so that only comments containing the three last tags remained. The corpus was filtered again using the *spacy* and *spacymoji* libraries for Python in order to find comments containing emojis. The comments containing hate speech were selected manually based on the definition of hate speech and the usage of hate speech strategies described above. The final sample consisted of 266 comments.

The Twitter and VK subcorpora (VK is a major Russian social network used by most Russians) were collected in February-April 2022 using the Russian hate speech lexicon made by Somin from the HSE Laboratory of Linguistic Conflictology and Modern Communicative Practices (the results are not yet published; received through personal communication). I searched Twitter for tweets written in Russian and containing one of the 236 words in Somin's hate speech lexicon. The results were inspected manually once more to select comments which fit the hate speech definition and strategies described above. The Twitter subcorpus was gathered using the *rtweet* package for R. Retweets (without quotes) were excluded as only original posts, replies and quote retweets were considered informative enough. The corpus included a total of 201 tweets. The VK subcorpus was gathered using the *vk_api* library for Python. Since *vk_api* does not allow to parse all posts or comments containing a certain keyword and it is only possible to parse comments or posts from a particular page, four group pages were selected for the analysis: "Ne Poverish", "Borscht", "Lentach", and "Tsargrad". These groups were chosen due to the following reasons:

1. a high number of subscribers ("Ne Poverish" and "Borscht" have more than 7.000.000 subscribers each, "Lentach" and "Tsargrad" have approximately 2.500.000 and 500.000 subscribers respectively)

2. they all have a different focus and a different target audience: “Borscht” and “Ne Poverish” are entertainment-based, “Lentach” and “Tsargrad” are news aggregators with Lentach’s audience being younger liberal people and Tsargrad being oriented towards older and more conservative users
3. the comments are open in all these groups which, coupled with their size, leads to frequent appearance of conflicts. For example, Shulginov and Tillabaeva have already used comments from “Lentach” and “Borscht” to analyze conflict communication (Tillabaeva & Schulginov 2020: 47).

Comments and tweets containing emojis were detected using the *spacy* and *spacymoji* libraries for Python.

3.2 Emoji Sentiment Ranking

Emoji Sentiment Ranking is an emoji lexicon designed for automated sentiment analysis (Kralj Novak et al. 2015: 1). Each emoji is assigned the following parameters: occurrences in the initial corpus that was used for markup and features extraction, position in a sentence, sentiment polarity, and sentiment score.

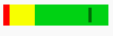
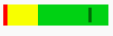
Char	Image [twemoji]	Unicode codepoint	Occurrences [5...max]	Position [0...1]	Neg [0...1]	Neut [0...1]	Pos [0...1]	Sentiment score [-1...+1]	Sentiment bar (c.i. 95%)	Unicode name	Unicode block
😂		0x1f602	14622	0.805	0.247	0.285	0.468	0.221		FACE WITH TEARS OF JOY	Emoticons
❤		0x2764	8050	0.747	0.044	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

Figure 1: Emoji Sentiment Ranking interface

The key parameter for this study was sentiment score which was computed on the basis of sentiments of tweets that contained the emoji (Kralj Novak et al. 2015: 3). The initial corpus included tweets in 13 European languages. Sentiment data for different languages was compared, and authors came to the conclusion that there is no statistically significant difference between them, that is, this lexicon can be considered universal (Kralj Novak et al. 2015: 12). In other words, the scores should be applicable also to a Russian corpus.

3.3 Markup of the corpus

The Twitter and VK subcorpora were marked up during the analysis. Two prototypical situations were distinguished:

1. [Conflict] the author of a tweet/comment is participating in a conflict with someone. The working definition of conflict in this case is: "any situation or behavior involving parties (individuals or groups) who are, or consider themselves to be, instrumentally, intellectually and/or emotionally opposed or simply feel antagonistic to each other" (Evans, Jeffries & O'Driscoll 2019: 2)
2. [Non-conflict] the author of a tweet/comment agrees with someone as a part of a discussion or the tweet/comment does not belong to a discussion

Table 1: Frequencies of different types of situations in the Twitter and VK subcorpora

	Conflict	Non-conflict
Twitter	100	101
VK	144	105

It will be shown later that there is a certain correlation between the situation and the pattern of emoji usage.The data from Odnoklassniki were not marked up because the original links were not preserved.

3.4 Methods

The results were interpreted in the critical discourse paradigm. Critical discourse analysis provides a wide variety of tools to analyze how ideologies are expressed and reproduced and how power hierarchies are constructed. It also concerns the injustice and discrimination issues which are connected to hate speech studies. Moreover, as it has been suggested above, emojis function as isolated expletive expressions, i.e., they are not integrated into the syntactic structure of a sentence. Some elements of corpus linguistics were also employed, namely, frequency analysis and statistical analysis. Frequency analysis was performed using the *pandas*, *spacy*, and *spacymoji* libraries for Python. Statistical analysis was performed using R.

3.5 Hypotheses about emoji usage in the corpus

The following hypotheses were proposed as a result of preliminary theoretical analysis:

1. Since hate speech performs dehumanizing and demeaning functions, I expect that the most frequent emojis will also perform similar functions and therefore have a negative sentiment score according to Emoji Sentiment Ranking (e.g., 😡)
2. Emojis accompanying hate speech utterances will be related to hateful and violent concepts, such as 😠, 👊, 🖕, 🐵, 🔫, 💩, ☠️, 🤮, on the account of them aiming at hurting a person or a group of people (Wiegand & Ruppenhoffer 2021: 371). Emojis also tend to function as graphical depictions of the sentence content, so if concepts related to hate or violence are

mentioned in a sentence, the emojis are likely to be related to the same concept (Dainas & Herring, 2017: 2187). For example, if a sentence contains an appeal to physical violence, emojis following that sentence may also depict concepts related to physical violence, such as 🔫 or 👊.

3. The corpus will contain emoji sequences that evoke dehumanizing associations by e.g., visually depicting stereotypes about target groups since, as I have proposed, emojis in hate speech utterances aim at demeaning and degrading individuals and groups of people. This is not necessarily related to hateful and violent concepts since an emoji sequence can be constructed in such a way that it simply depicts a stereotype.
4. Emojis are more likely to duplicate hate speech expressed lexically in a sentence than replace it because overall emojis are not likely to be embedded into the sentence structure, i.e. I do not expect emojis to function as grammatical arguments in a sentence (Cohn et al. 2019: 2).
5. Emojis depicting facial expressions and gestures will be the most frequent category of emojis in the corpus since they express the author's attitude to a referent (Kaiser & Grosz 2021: 1017).

4. SENTIMENT SCORES OF THE MOST FREQUENT EMOJIS

Table 2: The most frequent emojis and their sentiment scores

Emoji	n	Sentiment score
🤣	470	no data
😂	439	0,221
👎	192	-0,188
😁	66	0,449
😡	49	-0,173
👍	43	0,521
😆	41	0,409
😄, 😅	38	0,421; 0,178
💩	30	-0,116

As can be seen from Table 2, the most frequent emojis from the corpus generally have a positive sentiment score, except for 👎, 😡 and 💩. Moreover, laughing and smiling emojis are abundant in both conflict and agreeing situations.

4.1. Positive sentiment emojis in conflict and non-conflict situations

In conflict situations, I propose that laughing and smiling emojis serve as a means of establishing a hierarchy in the discourse. For example, (1) is an excerpt from a comment thread under a news report about coronavirus rates in Russia. One of the participants suggested sending all the unvaccinated people out of the country. Their surname, given that it is not a nickname, was common for Jewish people in Russia so another user replied with this comment containing several pejorative terms for Jewish people. The example is coded as ”conflict” because the author disagrees with the previous comment and confronts the interlocutor in an aggressive manner.

(1) [Имя], а давай тебя еврейско жидовская рожа мы отправим на твою родину израиль,там и будешь доживать с прививкой нашей и гордится что не жидовской привился,гнида жид порхатый😡😡😡 мало вас сук мочили видно🤣🤣🤣

[Name], let’s send you jewish kike to your motherland israel, that’s where you’ll live out your days with our vaccine and be proud that you weren’t vaccinated with a jewey one, lousy jewish scum 😡😡😡 seems that not enough of you bitches were whacked 🤣🤣🤣

The author of this comment is using hate speech by degrading their interlocutor and showing that they are insensitive towards their opponent’s experience as a Jewish person by using slurs and referring to the Holocaust. This comment consists of two sentences. The first one is finished off by a sequence of reduplicated ‘enraged face’ emojis (😡). The second sentence, as it has been mentioned before, ostensibly contains a reference to the Holocaust and also ends with a sequence of reduplicated 'rolling on the floor laughing’ emojis (🤣). This can serve as an instrument for restoring the hierarchical damage that the first participant inflicted since the author of the comment supposedly finds the Holocaust something to be laughed about. In non-conflict situations, laughing and smiling emojis mainly serve as a marker for dehumanizing humor which is another hate speech strategy (Fortuna & Nunes 2018: 5), and which is used in bonding between like-minded writers. The next example, (2), was taken from a comment thread under a news report saying that the Baltic states refused to pay for Russian gas in rubles, thus trying to drain Russian foreign exchange reserves.

(2) Гордые шпротоеды ещё и в рот возьмут за газ!🤣🤣🤣

Proud sprat-eaters will give head for the gas! 🤣🤣🤣

The author of the message uses an ethnophaulism for citizens of Baltic states, “sprat-eaters” (it is often used as a derogatory term and refers to a type of canned fish produced in Baltic states during the Soviet times), combined with a sexualized joke. Reduplicated laughing emojis act as an additional means for marking humor. The example is coded as non-conflict because the author of this comment supports another interlocutor in this comment thread in their assumption that citizens of the Baltic states in this situation are trying to compensate for the poor financial situation in their countries.

5. Emojis related to hateful and violent concepts

As for emojis related to hateful and violent concepts, they turned out to be quite common in the corpus (there were 74 instances of this pattern), although not as common as reduplicated sequences of laughing and smiling emojis. In most cases emojis depicting hateful and violent concepts are integrated into sequences of other emojis rather than just stand alone by themselves or reduplicated. (3) is a reply taken from a Twitter thread containing a rumor about Japanese authorities allegedly using chemical weapons in Ukraine.

(3) У япошек - опыт большой делать страшные опыты и гадости😩💩🤦‍♀️🤢

These Nippons have great experience of making awful experiments and playing nasty tricks 😩💩🤦‍♀️🤢

The first emoji in this sequence is 😩 (weary face) probably used to demonstrate frustration, followed by 💩 (pile of poo) which is associated with the topic of defecation. It is unlikely that it is used as an insult here, because otherwise it would be integrated differently into the sentence structure, so it may be assumed that the purpose of this emoji is to indicate disgust. The third emoji, 🤦‍♀️ (woman facepalming), represents a universal gesture for shame. The sequence ends with the 🤢 (nauseated face) emoji aimed at marking aversion once again. This comment was aimed at like-minded readers so the potential purpose of using emojis that express strong negative emotions was to show the members of the same group that the author of the message supports their aversion towards Japanese people.

Some of the emojis that are included in this category can be classified as culture-specific. For instance, Wiegand and Ruppenhofer propose that 🐵 (monkey face) emoji has racist connotations (Wiegand & Ruppenhofer 2018: 371). An example of an emoji that is specific to Russian hate speech discourse which is often found in the sample gathered for the present study is 🐷 (pig face) and other emojis depicting pigs. This may be related to the fact that there is a number of ethnophaulisms for Ukrainians that employ comparison to pigs: for example, “saloed” (salo-eater; it is a very common slur for Ukrainians based on the stereotype that Ukrainian people usually eat salo — salted pork fat), “khokhloschwayn” (“Ukrainian swine”), etc.

6. Emoji sequences evoking dehumanizing associations

Emoji sequences that evoke dehumanizing associations are the least frequent pattern of emoji usage in the corpus (there were only 13 instances). In cases where it was possible to establish context, it is apparent that this kind of emoji sequence is limited to non-conflict situations.

The reason for this might be that in such cases the authors of hateful messages pursue other purposes than asserting their dominance. They often try to use so-called "dog whistles" or demonstrate in-group affiliation. For instance, (4) is a comment left under a news report about the Russian military in the "Tsargrad" VK group.

(4) Парни вернитесь пожалуйста живыми, а вот Укропию, хохложопых- бандерлогов сравняйте под чистую🫦🐖🐷😈👹💩💩💩💩💩🤭🤮👊👎☝️ Россия-Сила, Россия-Победа, Россия-Вперед 🇷🇺💘💪👏👍🔥

Boys come back alive please, and Ukropia, Ukrainian-ass Bandar-logs — raze them to the ground 🫦🐖🐷😈👹💩💩💩💩💩🤭🤮👊👎☝️ Russia is Power, Russia is Victory, Go Russia 🇷🇺💘💪👏👍🔥

The first emoji sequence contains three pig emojis, presumably referring to ethnophaulisms for Ukrainians, followed by 😈 (smiling face with horns) and 👹 (ogre) which possibly indicate the author's evil intentions towards Ukrainians. The next block consists of the reduplicated 💩 (pile of poo) emojis which, as I have suggested above, may mark disgust, and also 🤭 (face with hand over mouth) and 🤮 (face with open mouth vomiting) emojis. I propose that there is a link between these two emojis because, although Emojipedia describes 🤭 (face with hand over mouth) emoji as being aimed at showing embarrassment or shock, this emoji also resembles a gesture that one can make before an attack of vomiting. The next emojis in the sequence are 👊 (fisted hand sign) which reflects some of the sentence contents, namely, the call for violence against Ukrainians, and 👎 (thumbs down), expressing disapproval of Ukrainians. The last emoji in the sequence is ☝️ presumably showing that the author of the message is sure about the significance of their words.

As it has been shown above, emoji sequences that evoke dehumanizing associations may convey a quite complex message. A possible explanation for associatively connected of emojis being the rarest pattern might be that emojis are limited in their sequencing abilities due to constraints in combinatoriality and and to the fact that they impose an unnatural information structure on visual communication (Cohn et al 2019: 16).

7. Emojis replacing or duplicating hate speech

In the examples we have observed so far, emojis occurred outside sentences, usually indicating the end of a sentence in a punctuation-like manner. Emojis that occur sentence-internally turned out to

be a relatively infrequent pattern (there were only two examples of this). There were 27 occurrences of emojis duplicating hate speech.

In all cases except one, emojis replaced adjectives and nouns, in accordance with Cohn et al's findings (Cohn et al, 2019: 14). The only case when an emoji replaces a predicate was found in the VK subcorpus.

(5) «[имя]»🐖((🧐А разве ВК не запрещен на украине, салоед?((🤷Что ж ты нарушаешь законы своей🐖 недостраны и пользуешься «вражеской» соцсетью,,, и спонсируешь нашу РФ🇷🇺,,, а, цыбулятник!?((🤷🧐

[name]🐖((🧐 Isn't VK forbidden in ukraine, salo-eater? ((🤷 Why are you breaking your 🐖 laws and using your enemy's social media,,, and sponsoring our Russian Federation🇷🇺,,, you onion-eater!?🤷🧐

This comment was targeted at a Ukrainian person who left a comment supporting the demolition of a monument commemorating Russian-Ukrainian friendship in Kyiv. In the first and in the last sentence, the pig emoji supposedly replaces a predicate noun "pig". I consider it to be an instance of hate speech since "pig" is not simply an insult here. The author of the comment consecutively uses the pig emoji (🐖) throughout the whole message as an ethnophaulism for Ukraine and Ukrainians.

The reason for the tendency to replicate hateful messages might be that emojis tend to serve as graphical representations of the sentence content (Dainas & Herring 2017: 2187). For example, sentence (6) contains a rainbow flag emoji (🏳️‍🌈) which presumably reduplicates "Gayrope" (concatenation of the words "gay" and "Europe" used as a pejorative term for Europe and European citizens).

(6) [имя], нацистская свинособака заткни свою пасть смердит аж в гейропе вонь стоит 🏳️‍🌈🖕

[name], shut your face nazi swine it stinks so hard you can smell it in gayrope 🏳️‍🌈🖕

When integrated into the sentence structure, emojis can convey non-redundant information, but the overall trend is for emojis to be placed at the end of the sentence where they usually perform the above described functions (Donato & Paggio 2017: 124; Cohn et al. 2019: 2).

8. Categories of the most frequent emojis

As it can be seen from Table 2, the most frequent emojis in all three subcorpora are face emojis or gestures. Face emojis either depict smiling or laughing faces or faces that express anger, confusion, or disgust. Gestures that are depicted by emojis can be described as universally understood, for example 👎 (thumbs down) or 🤦 (person facepalming). I believe that these emojis are rarely context-sensitive and have well-established meanings in the hate speech context.

Pig emoji (🐖), which is also among the most common in the Russian Toxic Comments corpus, can be added to amplify the insult, as it is shown in (6), and as a culture-specific emoji referring to Ukrainian people, as it has already been mentioned. Pig emoji can be used as an insult on its own, but I believe that in this case it can serve as an ethnophaulism since there is a significant layer of corresponding lexical units in Russian. It is generally hard to predict which emojis might be used as culture-specific because they depend very much on the context of the discussion and on the parties involved in it. For example, 🐵 (monkey face emoji) is considered to be racist in the Western context but in this sample there were no cases. Additionally, 🇷🇺 (Russian flag) and 🇺🇦 (Ukrainian flag) are also high on the frequency list in the Twitter and VK subcorpora respectively. The reason for that might be that the data were gathered in February-April 2022 when the Russian invasion of Ukraine began, so many of the discussions that got included in the sampling were dedicated to this topic. As it has been mentioned before, emojis often act as graphical representations of the sentence content so flags of different countries are used to visually depict them in political discussions.

Overall, the analysis showed that, contrary to my expectations, positive sentiment emojis are more prevalent in the corpus. They occur both in conflict and non-conflict situations and have slightly different functions. Most of them are positioned at the beginning or end of the sentence much like "emotionally charged punctuation". There are fewer examples of emoji sequences evoking dehumanizing associations or emojis related to hateful and violent concepts, i.e., sequences of emojis that may convey a message on their own. The most frequent category of emojis in the corpus were either face emojis or emojis depicting universally understood gestures which goes in accordance with viewing emojis as "isolated expletive expressions" since such emojis primarily convey the author's attitude to a referent (i.e., use-conditions).

9. Correlation between emoji patterns and situation

As it has been mentioned previously, while working on the corpus analysis, I have made a suggestion that the patterns of emoji usage correlate with the situation in which they are used:

1. smiling and laughing emojis are more likely to be used in conflict situations
2. non-conflict situations are usually associated with emojis that duplicate hate speech, emojis related to hateful and violent concepts, and sequences of emojis that evoke dehumanizing associations

I have used a chi-squared test to examine this idea because the data that I am operating is categorical. I have built contingency tables and performed the test using the chisq.test() function in R. The null hypothesis was that there would be no correlation between the pattern of emoji usage and the situation they were used in. As a result of the chi-squared test, the null hypothesis was rejected (p=7.672e-11). Observed counts can be seen in Table 3.

Table 3: Observed counts of different patterns

Pattern of emoji usage	Conflict situation	Non-conflict situation

Smiling and laughing emojis	217	112
Emojis related to hateful and violent concepts	13	42
Emojis evoking dehumanizing associations	1	7
Emojis duplicating hate speech	4	13

It can be seen from Table 3 that smiling and laughing emojis are more prevalent in conflict situations, while other patterns of emoji usage correlate with non-conflict situations. It may seem counterintuitive that smiling and laughing emojis are more prevalent in conflict situations because smiling and laughing emojis seem to be a sign of bonding or sympathy between different users but, as it can be seen, this is not the case. A possible explanation may be that, as it has been mentioned before, in conflict situations emojis can be used as a tool for asserting dominance, and smiling and laughing emojis are the easiest way to demonstrate insensitivity and contempt towards the opponent. In non-conflict situations, there is no immediate need to restore the hierarchical damage that has been inflicted by the previous opponent, so the patterns of emoji usage can become more creative, thus making emoji a tool for employing stereotypes, evoking associations, and also communicating disgust, fear, and other negative emotions more openly than in the situations of conflict.

10. Conclusion

In this article, I analyzed how emojis function in hate speech contexts using data from three popular Russian social networks: VK, Odnoklassniki, and Twitter. It has been shown that emojis or their sequences cannot be employed as a separate hate speech strategy, i.e., they do not seem to convey an independent message; they are rather a subsidiary mechanism used for various purposes: establishing a power hierarchy, bonding between like-minded readers, conveying unfavorable emotions, and marking dehumanizing humor. There are only a few examples in the corpus where emojis seem to trigger unfavorable associations but in all these cases they mostly reflect the contents of the message; they are not used as means of insult alone.

The use of different emoji patterns seems to be conditioned to different situations. Smiling and laughing emojis are prevalent in conflict situations and other patterns are mostly employed in non-conflict situations which suggests that these patterns help the authors of the messages pursue different goals and present themselves differently in different circumstances. However, this study can provide valuable insights into the functioning of online hate speech, in particular, the goals that the authors of hateful messages are trying to reach. As it can be seen from the examples analyzed in the article, the purpose is often not simply to degrade someone or a group of people, but to assert

dominance and show insensitivity, to gain status among the members of their group, or to connect with like-minded readers.

It was not possible to find emojis that are exclusively or predominantly used in hate speech. On the contrary, the most frequent emojis were the common smiling and laughing faces which are also used in other contexts and situations. However, it seems that there are some emojis that might be culture-specific in hate speech contexts (e.g., the pig emoji used to refer to Ukrainian people in discussions about Russian-Ukrainian conflict) so future studies aiming at constructing lexicons of emojis may focus on this issue.

References

Castaño-Pulgarín, S. A. et al. (2021). Internet, social media and online hate speech. systematic review. *Aggression and Violent Behavior*, *58*. https://doi.org/10.1016/j.avb.2021.101608.

Cohn, N., Engelen, J., & Schilperoord, J. (2019). The grammar of emoji? constraints on communicative pictorial sequencing. *Cognitive Research: Principles and Implications*, *4*(1). https://doi.org/10.1186/s41235-019-0177-0.

Danesi, M. (2016). The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet. Bloomsbury Academic.

Dainas, A., & Herring, S. (2017). "Nice Picture Comment!" Graphicons in Facebook Comment Threads. *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2185–2194.

Dürscheid, C., Siever, C. (2017) Beyond the Alphabet – Communication with Emojis.

Dokumentatsiya vk_api. s.a. https://vk-api.readthedocs.io/en/latest/# (retrieved 5 June 2023)

Donato, G., & Paggio, P. (2017). Investigating redundancy in emoji use: Study on a Twitter based corpus. *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 118–126. https://doi.org/10.18653/v1/w17-5216.

Fortuna, P., & Nunes, S. (2018). A survey on automatic detection of hate speech in text. *ACM Computing Surveys*, *51*(4), 1–30. https://doi.org/10.1145/3232676.

Gladilin, A. (2013) Yazik Vrazhdi v Traditsionnikh i Novikh Media (Hate Speech in Traditional and New Media). V*estnik Chelyabinskogo gosudarstvennogo universiteta*, vol. 21, 144–153.

Jakubowicz, A. (2006) How Do Media Marginalize Groups. *Encyclopedia of Language and Linguistics*, Elsevier, pp. 1009–1023.

Evans, M., Jeffries, L., & O'Driscoll, J. (Eds.). (2019). *The Routledge Handbook of Language in Conflict* (1st ed.). London: Routledge. https://doi.org/10.4324/9780429058011.

Kaiser, E., & Grosz, P. G. (2021). Anaphoricity in emoji: An experimental investigation of face and non-face emoji. *Proceedings of the Linguistic Society of America*, *6*(1), 1009-1023. https://doi.org/10.3765/plsa.v6i1.5067.

Kearney, M.W. (2019). rtweet: Collecting and analyzing Twitter data. *Journal of Open Source Software,* 4(42). 1829. doi:10.21105/joss.01829

Kralj Novak et al. (2015). Sentiment of Emojis. *PLOS ONE*, *10*(12). https://doi.org/10.1371/journal.pone.0144296.

Lingam, R. A., & Aripin, N. (2017). Comments on fire! classifying flaming comments on YouTube videos in Malaysia. *Jurnal Komunikasi, Malaysian Journal of Communication*, *33*(4), 104–118. https://doi.org/10.17576/jkmjc-2017-3304-07.

López, C. A., and López, R. M. (2017) Hate Speech in the Online Setting. *Online Hate Speech in the European Union: A Discourse-Analytic Perspective*, Springer Open, Cham, 10–12.

McCulloch, G. (2019). Because internet: Understanding the new rules of language. New York: Riverhead Books.

McCulloch, G., Gawne, L. (2018) Emoji Grammar as Beat Gestures. Proceedings of the 1st International Workshop on Emoji Understanding and Applications in Social Media (Emoji2018).

McKinney, W. et al. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference,* 445. 51–56.

Na'aman, N., Provenza, H., & Montoya, O. (2017). Varying linguistic purposes of emoji in (Twitter) context. *Proceedings of ACL 2017, Student Research Workshop*. https://doi.org/10.18653/v1/p17-3022.

Paz, M. A., Montero-Díaz, J., & Moreno-Delgado, A. (2020). Hate speech: A systematized review. *SAGE Open*, *10*(4). https://doi.org/10.1177/2158244020973022.

Powell, A., Scott, A. J., & Henry, N. (2020). Digital harassment and abuse: Experiences of sexuality and gender minority adults. *European Journal of Criminology*, *17*(2), 199–223. https://doi.org/10.1177/1477370818788006.

Ruzaite, J. (2018). In search of hate speech in lithuanian public discourse: A corpus-assisted analysis of online comments. *Lodz Papers in Pragmatics*, *14*(1), 93–116. https://doi.org/10.1515/lpp-2018-0005.

Spacy.io. s.a. https://spacy.io/ (retrieved 5 June 2023)

Spacymoji. s.a. https://spacy.io/universe/project/spacymoji (retrieved 5 June 2023)

Sundén, J., & Paasonen, S. (2018). Shameless hags and tolerance whores: Feminist resistance and the affective circuits of online hate. *Feminist Media Studies*, *18*(4), 643–656. https://doi.org/10.1080/14680777.2018.1447427.

Tillabaeva, A. A., & Shulginov, V. A. (2020). Speech behaviour of internet users in conflict communication. *Slovo.ru: Baltic Accent*, *11*(4), 45–57. https://doi.org/10.5922/2225-5346-2020-4-4.

Trajkova, Z., & Neshkovska, S. (2018). Online hate propaganda during election period: The case of macedonia. *Lodz Papers in Pragmatics*, *14*(2), 309–334. https://doi.org/10.1515/lpp-2018-0015

van Dijk, T. A. (1983). Cognitive and conversational strategies in the expression of ethnic prejudice. *Text - Interdisciplinary Journal for the Study of Discourse*, *3*(4), 375–404. https://doi.org/10.1515/text.1.1983.3.4.375.

Warner, W. & Hirschberg, J. (2012). Detecting Hate Speech on the World Wide Web. *Proceedings of the Second Workshop on Language in Social Media*, 19–26.

Wiegand, M., & Ruppenhofer, J. (2021). Exploiting emojis for abusive language detection. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. 369-380. https://doi.org/10.18653/v1/2021.eacl-main.28.