

Nonstandard linguistic features of Slovene socially unacceptable discourse on Facebook

Kristina Pahor de Maiti, Faculty of Arts, University of Ljubljana, Slovenia

Darja Fišer, Faculty of Arts, University of Ljubljana, Slovenia, Jožef Stefan Institute, Ljubljana, Slovenia

Nikola Ljubešić, Jožef Stefan Institute, Ljubljana, Slovenia, Faculty of Computer and Information Science, University of Ljubljana, Slovenia

Abstract

Socially unacceptable discourse (SUD) is a topical issue in many areas of social science and humanities, but the analysis of its linguistic dimension is often neglected. This paper explores the surface linguistic features of SUD and aims to determine the level of linguistic standardness of SUD Facebook comments in comparison to its non-SUD counterpart. The analysis was conducted on a dataset extracted from the Slovene part of the FRENK corpus which contains SUD and non-SUD comments from Facebook pages of news outlets on the topics of migrants and LGBT issues. The dataset was manually annotated for nonstandard language features using a custom-built annotation schema. The results of the quantitative analysis show that SUD comments are statistically significantly less standard in comparison to non-SUD comments. Despite the high level of nonstandardness of both subsets of comments, the qualitative analysis proved that nonstandard features of SUD comments surpass the surface spelling deviations that are commonly found in computer-mediated communication.

Keywords: hate speech, socially unacceptable discourse, computer-mediated communication, linguistic standardness, Facebook

1 INTRODUCTION

Human communication has always included discriminatory discourse practices, but it rarely reached a broad public audience. This changed with the development of computer-mediated channels of communication, such as social media. A substantial increase in online hate speech against marginalized groups (immigrants and refugees, the LGBT community, Roma, religious groups, etc.) can now be observed especially in relation to certain (geo)political events (e.g., the EU refugee crisis beginning in 2015) (Motl and Bajt 2016).

In this paper, we use the concept of socially unacceptable discourse (SUD) which goes beyond the harshest forms of legally prosecutable hate speech as it addresses a broader range of practices and encompass a full spectrum of hateful, threatening, abusive and discriminatory speech as well as indecencies and insults. Several authors (Bajt 2016, Ivanou 2017, Tsesis 2002) have argued that not only hate speech but all forms of SUD pave the way towards social actions that can have an aggravating effect on individuals and social groups that are the target of SUD, as well as on the society as a whole. In light of this, it is clear that all manifestations of SUD have the potential to be harmful for the society and should also be treated as such.

Although there is a consensus on the negative effects of SUD on society (Bojarska 2019, Nielsen 2002) and some progress with regard to limiting its proliferation has been made (Jourová 2019),¹ its accurate and timely detection (ElSherief et al. 2018, Vidgen and Yasseri 2019, Zhang and Luo 2019) together with effective prevention of its negative influence (Ullmann and Tomalin 2019) remain a challenge. Moreover, despite its uncontested presence in contemporary society, SUD is still not thoroughly researched. As a societal phenomenon, its scholarly treatment extends over many different disciplines, which is why it necessitates an in-depth analysis in all of the relevant fields as well as a cross-disciplinary scope. For some time now, online hate has received a lot of attention from policy-makers and researchers from various scientific fields, but despite the fact that SUD is nearly always manifested through linguistic means, it is predominantly studied as a societal and psychological phenomenon (Waqas et al. 2019), and even though SUD as a research topic is not unknown to sociolinguistics and discourse studies, it has primarily been treated in the domain of critical discourse analysis, often leaving its linguistic characteristics neglected (Klein 2018). This chapter therefore focuses on a set of surface linguistic features with the aim to identify the general linguistic characteristics of hateful comments in comparison to non-hateful ones. Understanding that the language of computer-mediated communication (CMC language) is strongly characterized by a certain level of nonstandard writing (Crystal 2001), the main challenge of this study will be to distinguish the specific features of SUD from the general

¹ For a critique see Bukovská (2019).

characteristics of CMC language. The chapter is an extended version of a short conference paper (Pahor de Maiti et al. 2019) where we presented the initial results of the analysis of surface linguistic features of online hate speech. In this chapter, we elaborate on the presentation of related studies and the dataset preparation process as well as provide a more detailed interpretation of the results.

In the next sections, we will first present the results of related studies (section 2) and the study design (section 3). In section 4, we will move to the presentation and discussion of the results. The final section (section 5) includes an overview of the findings and describes possible next steps for future work.

2 RELATED WORK

2.1 Socially unacceptable discourse

By being a global space that enables immediate and participatory expression of ideas under the guise of anonymous or fake user identity, the internet has, among its many positive impacts, also evolved into a “breeding ground for the phenomenon of *cyberhate*” (López and López 2017) which is produced not only by members of extremist groups but also increasingly by regular internet users. Hate speech disseminates hatred or disqualification of an individual or group based on their race, skin colour, ethnicity, sex, disability, religion or sexual orientation (Nockleby 2000). However, Vehovar et al. (2012) point out that the definition of hate speech should be broadened and also include various forms of offensive speech, such as hurtful, derogatory or obscene comments about someone. In light of this, all such nuances of hate speech should be taken into account as research shows that even the most indirect inappropriate messages can incite harmful action (Baider et al. 2017, Muskat and Assimakopoulos 2017). In this chapter we prefer to apply the term socially unacceptable discourse (SUD) so as not to confuse it with the narrowly defined legally prosecutable hate speech (Banks 2011).

In Slovenia, research on SUD started in the 1990s and has been mainly confined to law and social sciences throughout the course of the early 2000s. From 2006 onwards, we have observed an intensification of research that goes beyond these fields and stretches into those of linguistics and computer science. Slovene linguists have researched SUD in traditional media (cf. Červ and Kalin Golob 2012), language reference books (cf. Gorjanc 2005) and translations (cf. Gorjanc 2012). Until now, however, not many analyses have been conducted on online SUD data and those that have (cf. Čeferin and Mežnar 2014, Jakopič 2013, Merljak 2011) mainly centred around the concept of hate speech, the legal implications of its use and editorial challenges for restricting SUD online. This shows

that the linguistic dimension of (online) SUD is still notably under-researched which motivated us to identify the language features of SUD in order to better understand, detect and prevent undesired communication patterns.

2.2 The language of computer-mediated communication

When investigating the language of online SUD, it is important to consider the general characteristics of CMC language in order to avoid misattribution of their general characteristics to SUD. CMC language is well known to include the use of unconventional spelling and often integrates informality and deviations from the norm on the level of grammar and punctuation (Crystal 2001). CMC language is also characterized by a number of features used by writers in order to overcome the technical limitations of both the communication platform and device used, as well as to fulfil the social need for a quick response (e.g., non-canonical text abbreviation strategies (Bieswanger 2013)), and to fill in for the limited affordances for nonverbal communication (e.g., emojis/emoticons and expressive punctuation (cf. Androutsopoulos 2011)).

Length and time restrictions are often tackled with shortening strategies, which may be realized on different levels of linguistic description, and are language-specific. For Slovene, Goli et al. (2016), who investigated Twitter messages, found that shortening is commonly used on the orthographic and lexical levels, as well as on the syntactic level, where it usually includes omissions of the auxiliary verb *to be*. Emoticons and emojis² are a popular feature of CMC language, and from a pragmatic standpoint they fulfil multiple communicative functions. Dresner and Herring (2010) found that emoticons and emojis can be used (1) as emotion indicators mimicking non-verbal communication, (2) as elements conveying non-emotional messages (e.g., a joke), and (3) as elements indicating the illocutionary force of the message (e.g., mitigating or reinforcing the content), especially when used in place of the final punctuation (Amaghloubeli 2012).

Due to certain communicative contexts, such as the instantaneous and informal nature of CMC, language of CMC also exhibits features that were traditionally attributed to speech (cf. Androutsopoulos 2011, Baron 2002). Zwitter Vitez and Fišer (2018) found that elements of modality, used to express opinion, judgement or certainty, represent the strongest connection between spoken discourse and language in Slovene online comments. A connection to spoken language has also been found by Goli et al. (2016) who argue that shortening on the orthographical level reflects the tendency to use phoneticized spelling.

2 An emoticon is a representation of a facial expression with a combination of keyboard characters (e.g., :->), whereas the an emoji represents a facial expression/emotion, or another notion/object in the form of a symbol, icon or a picture.

Beyond technical and communicative factors, research has already proved that socio-demographic factors, such as personality, social status, level of education and age, also play an important role in CMC and impact how people express themselves (Gill 2011, Hilte et al. 2018). Given that SUD is a highly subjective type of discourse, we can assume that personality traits significantly influence this type of discourse as well. The role of personality has been shown in relation to the use of shortening strategies. It was found that shortening is caused not only by the length limitations of some online platforms, but also exhibits the societal and emotional aspect of the message, conveying signs of authorship, community identity or illocutionary force (Goli et al. 2016). Similarly, emoticons/emojis reflect the author's profile on many different levels: they are not only a vehicle to express the emotions of the producer, but also fulfil other pragmatic functions, such as author's intentionality (Spina 2017).

2.3 Linguistic standardness

The notion of standardness defines the optimal form of the written language used for (mass) communication in the public sphere (Skubic 2005). However, private and public communication are closely intertwined on the internet, which causes difficulties when defining the necessary level of message formality, as CMC can be used for many purposes on the long axis from highly formal to fully profane contexts (Splichal 2017). The analysis of nonstandard writing is especially interesting for Slovene due to its highly prescriptive linguistic culture. In this chapter, we understand the notion of linguistic standardness in proportion to the level of author's compliance with the linguistic norm that is prescribed by the normative orthographic and grammar guides. The nonstandard writing practices of Slovene CMC language have been extensively studied in the JANES project³ and the analysis of Slovene tweets has shown that nonstandard writing (especially on the orthographic level) is common in informal communication on social media (Fišer et al. 2018).

2.4 The characteristics of socially unacceptable discourse

It has been shown that SUD is a frequent phenomenon in online comments on news portals and social media (cf. Lewandowska-Tomaszczyk 2017). The intensity of SUD in online comments actually reinforces proportionately to the discourse used in the initial news post (Bajt 2016), showing the great impact of media discourse on the discussions on online forums (cf. Gorjanc 2005)

³ The JANES project involved the building of a large corpus of Internet Slovene. See: <http://nl.ijs.si/janes/english/>.

and consequently the moral responsibility of the media in the publication and administration of news posts.

Given that SUD is more often expressed covertly than overtly, it is of key importance that analysis of SUD is not limited only to the level of lexis (e.g., presence of swear words). Muskat and Assimakopoulos (2017) showed that negative opinions are frequently expressed indirectly, which could be due to commenter's need for face-saving and the fear of appearing intolerant. In a similar vein, Červ and Kalin Golob (2012) found that authors of SUD, knowing that their message violates communicative norms, intentionally remove explicit negative evaluations and instead express the intended meaning through irony, rhetorical questions or punctuation marks. Furthermore, an interesting observation has been made by Zwitter Vitez and Fišer (2016), who found that negative comments actually have a more standard orthography, are longer and have a more complex syntax structure than positive comments. They argue that this is linked to the author's desire to present their negative opinion in a neutral linguistic form in order to be acceptable to a broader public (i.e., not focused on a specific socio-demographic community). This could also be explained by the author's tendency to appear as rational and competent as possible given that poorly written texts by the general public are often perceived as a sign of intellectual primitivity, lack of education and low social status, as well as an evidence of the lack or inability to think logically (Jessmer and Anderson 2001, Neustupný and Nekvapil 2003).

3 STUDY DESIGN

3.1 Research questions and hypotheses

The aim of our analysis was to investigate the length, lexical diversity and linguistic standardness of Facebook comments in order to establish whether any specific linguistic characteristic can be observed in hateful comments compared to the features that are typical of CMC in general. Therefore, our research questions and hypotheses for investigation are:

1. Research question 1: Comment length
 - 1.1. Hypothesis 1.1: SUD comments are shorter than non-SUD comments.
2. Research question 2: Lexical diversity
 - 2.1. Hypothesis 2.1: Vocabulary diversity is larger in non-SUD comments compared to SUD comments.
 - 2.2. Hypothesis 2.2: Non-SUD comments contain more emoticons and emojis than SUD comments.

3. Research question 3: Linguistic standardness
 - 3.1. Hypothesis 3.1: Punctuation to non-punctuation ratio is higher in SUD comments.
 - 3.2. Hypothesis 3.2: SUD comments are linguistically less standard than non-SUD comments.

3.2 Dataset

In this chapter, we used the Slovene part of the FRENK corpus which contains 6,545 and 4,571 comments about migrants and LGBT issues, respectively, that were posted in response to posts on the Facebook pages of the three Slovene mainstream news media with the most visited websites according to the Alexa service:⁴ 24ur.com,⁵ SiOL.net.Novice,⁶ Nova24TV⁷ (Ljubešić et al. 2019).

The FRENK corpus is annotated according to a project-specific two-level annotation schema that identifies the type (see Table 1) and target (see Table 2) of SUD. Comments that do not include socially unacceptable discourse are marked with the “Acceptable speech” label. The rest are assigned two-dimensional labels indicating the type of socially unacceptable discourse and its target. Comments targeting groups/individuals on the basis of their religion, gender, sexual orientation, ethnicity, race, etc. are annotated as “Background”. If SUD is aimed at individuals due to their particular group affiliation (professional or political affiliation, etc.), the “Other” category is selected. Comments without any specific target, but that nonetheless contain uncivil language, are marked as “Inappropriate speech”.

Table 1: Type of SUD in the FRENK corpus.

| | |
|--------------------|-------------------------------|
| Type of SUD | Background – violence |
| | Background – offensive speech |
| | Other – threat |
| | Other – offensive speech |
| | Inappropriate speech |

After the type of SUD is determined, the annotator identifies against whom the comment is directed.

⁴ <https://www.alexa.com/topsites/countries>

⁵ <https://www.facebook.com/24urcom>

⁶ <https://www.facebook.com/SiOL.net.Novice>

⁷ <https://www.facebook.com/Nova24TV>

Table 2: Targets of SUD in the FRENK corpus.

| | |
|----------------------|--|
| Target of SUD | Migrants/LGBT (comments targeting the specific social group) |
| | Related to migrants/LGBT (comments targeting individuals/groups that express a positive attitude towards the specific social group) |
| | Journalist/media (comments against the author of the news post or against the news media hosting the news post) |
| | Commenter (comments against authors of preceding comments to the same news post) |
| | Other (comments targeting individuals/groups that do not belong to previous categories or express negative attitude towards migrants/LGBT) |

For our analysis, we extracted all 520 comments in the FRENK corpus which contained elements of violence and threat regardless of their target for both topics (migrants: 417 comments; LGBT: 103 comments), and a randomized sample of 520 comments (respecting the same share per topic) labelled as “Acceptable speech”. The dataset was verticalized, morphosyntactically tagged and lemmatized with the ReLDI tagger (Ljubešić and Erjavec 2016).

3.3 Typology of nonstandard linguistic features

In order to observe the nonstandard linguistic features in SUD systematically, we developed an annotation schema for the tagging of nonstandard linguistic features in Facebook comments, based on the guidelines for normalizing CMC language (Čibej et al. 2016) and the Slovene Normative Orthography Guide (Toporišič 2007). Slovene has a strong prescriptive tradition that stretches far beyond orthography and also covers grammar and lexis. Our definition of standardness is based on the codified standard regarding spelling, lexis and grammar as set forth in the Normative Orthography Guide, but takes into account the specific communicative context of comments under investigation, especially when considering lexis (e.g., offensive words, typical for SUD comments on social media, were not annotated as nonstandard lexis). The annotation schema thus consists of five categories: orthography, lexis, morphology, syntax and word order (see Table 3). The subcategories describe the range of a certain category. They are only meant as support for a precise and reproducible annotation process and do not represent an additional annotation layer. In the initial phase, we tested our annotation schema on a smaller sample in order to consolidate the range of categories and set the categorization process (described in detail in section 3.4). The final annotation schema differed from its first version mainly in a higher number of subcategories (that were not anticipated from the related work on CMC writing practices).

Table 3: Annotation schema of nonstandard linguistic features in CMC language.

| Category | Description of a subcategory | Example <i>ex.</i> → <i>standard</i> /English/ |
|-----------------------|--|--|
| Orthography (O) | Incorrect use of lower/upper case | /.../ atlantika → <i>Atlantika</i> /Atlantic/ |
| | Incorrect punctuation and spacing | <i>Za vse tiste ki pravijo</i> → <i>Za vse tiste, ki pravijo</i> /For all those who say; missing comma/ <i>nebi</i> → <i>ne bi</i> /should not/ |
| | Typographical errors | <i>stslišče</i> → <i>stališče</i> /opinion/ |
| | Regional transformations of standard lexis | <i>kuj</i> → <i>takoj</i> (immediately) |
| | Character flooding | <i>BRAVOOOOOOOOOOOO</i> |
| | Omission of diacritics | <i>ce</i> → <i>če</i> (if) |
| | Alphanumeric words and creative spelling | <i>Kolikox</i> → <i>Kolikokrat</i> (How many times) |
| Phoneticized spelling | <i>nč</i> → <i>nič</i> (nothing) | |
| Lexis (L) | Content words from dialects & slang | <i>lih</i> → <i>ravno</i> (temporal just) |
| | Nonstandard abbreviations & acronyms | <i>not found</i> |
| | Words in foreign language | <i>gamad</i> → <i>golazen, mrčes</i> (vermin in Croatian) |
| | Semantically inappropriate words | <i>mogli</i> → <i>morali</i> (could instead of should) |
| Morphology (M) | Erroneous verb/noun affixes | <i>sprejom</i> → <i>sprejem</i> (with spray) |
| | Incorrect grammatical gender/number/aspect | <i>elektrosokerje</i> /.../ <i>ali so namenjena samo za nas</i> → <i>elektrosokerje</i> /.../ <i>ali so namenjeni samo za nas</i> (electroshocker ... or are they meant only for us) |
| Syntax (S) | Incorrect use of grammatical cases | <i>pod vzglavniku</i> → <i>pod vzglavnikom</i> (under the pillow) |
| | Incorrect use of definiteness | <i>taglavne</i> → <i>glavne</i> (the main; vernacular particle “ta” added) |
| | Syntactic ellipsis not justifiable by the context or clearly non-neutral | <i>pravijo da jih treba ubit</i> → <i>pravijo da jih je treba ubit</i> (they say that they need to be killed) |
| | Inappropriate parts of speech | <i>noben</i> (pronoun) → <i>nobeden</i> (noun) (no one) |
| | Nonstandard structures | <i>na vsake toliko kvatre</i> → <i>na vsake toliko</i> OR <i>na vsake kvatre</i> (from time to time; tautology) |
| Word order (W) | Nonstandard/non-neutral | <i>gamad voditeljska</i> → <i>voditeljska gamad</i> (vermin leaders) |

Our annotation schema does not strictly follow all the codification rules (e.g., we eliminated the subcategory for incorrect use of the supine),⁸ but has been designed to serve a dual purpose: (1) to establish whether there are any differences in the level of text standardness between SUD and non-SUD comments, and (2) to identify possible idiosyncrasies of SUD comments. In addition, although some features (e.g., marked word order) might not be traditionally classified as strictly nonstandard (e.g., because they simply fulfil a specific rhetorical function), they were still annotated as nonstandard due to their frequency observed during the annotation test phase.

3.4 Annotation

Manual annotation of the dataset was performed by one annotator who used the following reference guides: SSKJ – Dictionary of the Slovenian Standard Language;⁹ Pravopis – the normative orthography guide for Slovene;¹⁰ Slovene grammar (Toporišič 2004); Slovene lexicon SloLeks (Dobrovoljc et al. 2019);¹¹ Janes-Norm corpus (Erjavec et al. 2016). During the test annotation, we created internal guidelines in order to ensure consistency, which proved especially useful for disambiguation of borderline cases:

1. A token is defined as nonstandard if it can be assigned to at least one category in the typology.
2. The study uses a one-dimensional annotation schema, meaning that even if a token could be attributed to several subcategories of a single category, that category is annotated only once (e.g., *mas* → *imaš* (have) was labelled with the O-category label only once, despite the fact that the token is missing a diacritic (i.e., the first signal for the O-category label) and exhibits a phoneticized spelling (e.g., the second signal for the O-category label).
3. If the form of a token corresponds to multiple categories on the basis of multiple elements, all of the appropriate categories are indicated (e.g., *slovincov* → *Slovencev* (Slovenes) was labelled with the O- and M-category: O-category label for lower case used instead of the upper case, and M-category label for the incorrect suffix).

8 An impersonal verb form used predominantly with verbs of motion instead of which the infinitive is frequently misused, e.g., *go see* → *pojdi pogledat* (correct) vs. *pojdi pogledati* (incorrect). An explanation for this decision is given in section 3.4. Annotation.

9 Dictionary of the Slovenian Standard Language: <https://fran.si/130/sskj-slovar-slovenskega-knjiznega-jezika>

10 Normative orthography guide: <https://fran.si/>

11 Slovene morphological lexicon: <http://eng.slovenscina.eu/sloleks>

4. If the form of a token corresponds to multiple categories on the basis of the same element (e.g., missing the final letter), the O-category label is prioritized. Despite slightly skewing the final results (by boosting the O-category), this approach ensures consistency in the annotation process and still returns realistic results with regard to the level of overall non-standardness of SUD. This approach pertains especially to the following two types of examples:
 - 4.1. a token representing an infinitive verb which is missing its final letter (e.g., *treba /.../ kastrirat* → *treba/.../ kastrirati* (needs castrating)) which can be indicative of the incorrect use of supine (therefore the S-category) or of a typographical mistake (therefore the O-category). Tokens of this type were annotated with the O-category label which enabled us to refrain from automatically categorizing this feature as a grammatical mistake and rather assume it was an unintentional mistake.
 - 4.2. Similarly, the spelling variants of pronouns (e.g., *sma* → *sva* (we are)) possibly taking the O-category label (due to typographical mistake in an otherwise standard word) or the L-category label (due to it being a regional variant), were again placed under the O-category, so there was no need to differentiate between a typographical mistake and dialectal transformation.
5. We annotated any missing, redundant, excessively repeated or incorrect (combination of) punctuation marks. Spacing, however, was annotated only on the level of words (erroneously written together or apart, e.g., *nebo* → *ne bo* (will not)) and not on the level of punctuation marks due to the technical limitations that originate from the text verticalization process.
6. In case an emoticon/emoji appeared at the end of the sentence, it was treated as a final punctuation mark (cf. Pertot et al. 2016) mentions and emotional expressions – emojis and emoticons.
7. Word order and syntactic constructions deviating from the norm were indicated with only one W/S-category label at the position of the first token in the nonstandard phrase/sentence.
8. Since Slovene has a free word order and syntactic ellipsis can be interpreted in many ways, we annotated only clearly incorrect and non-neutral word order and omissions (e.g., missing auxiliary verb).
9. Comments written entirely in a foreign language were out of scope of this study.

4 RESULTS AND DISCUSSION

The following sections present the findings of our analysis based on the quantitative results that were gathered relating to comment length, lexical diversity and linguistic nonstandardness of SUD comments. Next, we present insights gathered from qualitative analysis regarding nonstandard orthographic and syntactic features, nonstandard word order, and the use of emoticons and emojis in SUD comments. Our discussion relates these findings to the research questions and hypotheses.

4.1 Basic statistics

Our dataset comprises a total of 19,091 tokens which are divided equally between SUD and non-SUD comments (see Table 4). For our analysis we extracted only the relevant tokens (18,103) and removed all irrelevant ones (988; comments written entirely in a foreign language).

Table 4: Structure of the dataset (number of tokens).

| | SUD | Non-SUD | Total |
|-------------|-------------|-------------|--------------|
| Nonstandard | 2,925 (30%) | 1,842 (22%) | 4,767 (26%) |
| Standard | 6,683 (70%) | 6,653 (78%) | 13,336 (74%) |
| Total | 9,608 | 8,495 | 18,103 |

4.2 Quantitative analysis

4.2.1 Comment length

The median for comment length, which was calculated by taking into account all tokens in the comments, was 12 tokens per comment for the SUD subset and 11 for the non-SUD subset. This result shows only a minor difference between the subsets, which is not statistically significant. Therefore, due to the similarity between both subsets, we can reject Hypothesis 1.1, which stated that SUD comments are shorter than non-SUD comments as a result of an immediate, emotional response to a newspaper article. Our findings also do not support the finding of Zwitter Vitez and Fišer (2016) that negative comments are longer in comparison to positive ones.

4.2.2 Lexical diversity

First, we calculated the type to token ratio (TTR) for each type of discourse over 100 random draws of 1,000 tokens. TTR is slightly higher for SUD comments (0.61) in comparison to non-SUD comments (0.58). Second, we calculated the content-to-function-word ratio for SUD comments, which was 1.32, and thus again slightly higher compared to that for non-SUD comments, which was 1.25.

These results show that vocabulary diversity is larger in SUD comments and therefore rejects Hypothesis 2.1, which stated the opposite. Assuming that SUD comments are more expressive and emotional, this outcome may not be surprising, given that people tend to use more colourful and creative language for emotionally-charged content. The lexical characteristics of hateful speech are addressed in detail by Franza et al. (2019), but a quick comparison of SUD and non-SUD nouns referring to a person in our dataset shows nine offensive nouns (*idiot, vermin, ...*) vs. one general noun (*human*). This limited but meaningful insight directly supports our finding on lexical diversity and the hypothesis about higher language expressivity.

Next, we calculated the relative frequency of emoticons and emojis, which was 0.005 for SUD comments and 0.009 for non-SUD comments. We also counted the number of different emoticons and emojis. SUD comments contained 24 different emoticons and emojis while non-SUD contained 34, 35% of which overlap with those found in SUD comments. To test whether the occurrence of emoticons and emojis in non-SUD comments in comparison to SUD comments is significantly higher, we ran an approximate randomization test with 1,000 iterations, obtaining a p-value of 0.0008. This means that the probability of obtaining the same or greater difference between the two types of comments randomly is below 0.001. Therefore, we can safely discard the null hypothesis that there is no difference between the usage of emoticons and emojis in SUD and non-SUD comments and confirm Hypothesis 2.2, which stated that there are more emoticons and emojis in non-SUD comments.

Less frequent use of emojis in SUD could be explained by the lack of available emoticons and emojis or the relative difficulty of accessing more specific emojis through the emoji keyboard (Bočková 2019) which can be perceived as too time-consuming during the creation of an emotionally-charged comment. In addition, the use of emojis could also be influenced by a communication strategy of the author, with which they try to achieve emotional detachment from the content of the comment. With the absence of emoticons/emojis, the emotional expressivity of the comment is lowered and the comment could be perceived as less emotional/more reasonable and thus more cogent, especially because – as argued

by Micciche (2007, as cited in Laffen and Fiorenza 2012) – emotions can be naively perceived as the opposite of reason. Both subsets contain more emojis than the more traditional emoticons but the overlap of the latter is bigger between the datasets. This is not unexpected, as different OS/application providers offer different sets of emojis (with greater expressiveness) in contrast to emoticons, which are limited to the keyboard characters. In addition, the user needs to possess some knowledge to be able to create emoticons, whereas emojis only need to be picked out of the proposed set. Moreover, it has been found that emojis are gradually substituting emoticons while fulfilling the same functions, which is generally message strengthening (Pavalanathan and Eisenstein 2015, Pertot et al. 2016).

4.2.3 Linguistic nonstandardness

By counting all punctuation marks (on the token level) versus all other tokens, we obtained a punctuation-to-non-punctuation ratio of 0.09 for SUD and 0.12 for non-SUD comments. This rejects Hypothesis 3.1, which assumed more punctuation marks in SUD comments due to a possibly higher expressiveness of such comments. The result is not surprising, as non-hateful comments are not necessarily all neutral. We suggest, therefore, that it would be useful to also take into account the sentiment of the comments in future research.

As Table 4 above showed, a total of 4,767 nonstandard elements were identified in the dataset. The share of nonstandard tokens in the non-SUD subset is 22%, whereas the SUD subset contains 30% of nonstandard tokens. In Table 5, the percentage of nonstandard features in SUD and non-SUD comments can be observed with regard to their type (e.g., spelling mistakes were put into the O-category (marked with O), i.e. nonstandard orthography). It should be noted that some tokens have been classified into more than one category (e.g., a token with incorrect spelling (O – Orthography) and grammatical case (S – Syntax)) which is why the final number of nonstandard tokens is higher than indicated in Table 4.

Table 5: Amount of nonstandard tokens per category in the dataset.

| | O | L | M | S | W |
|---------|-------------|-----------|---------|-----------|----------|
| SUD | 2,414 (82%) | 298 (10%) | 44 (2%) | 384 (13%) | 156 (5%) |
| Non-SUD | 1,659 (90%) | 104 (6%) | 17 (1%) | 87 (5%) | 44 (2%) |

Table 5 shows that by far the most prominent category in both types of discourse is Orthography, with slightly over 80% of the annotations in SUD and 90% in

non-SUD comments. The rest of the categories are much less frequent in both types of comments: Syntax represents 13% of the annotations in SUD comments and 5% in non-SUD, the Lexis category was assigned to 10% of SUD comments and 6% of non-SUD, Word order had a 5% share in SUD and 2% in non-SUD, and Morphology 2% in SUD and 1% in non-SUD.

The result of the chi-square test ($X^2(1, N = 18,103) = 178.4, p = 0.0001$) on the independence of the variables of linguistic standardness and the social acceptability of the comment showed that we can reject the null hypothesis on the independence of the variables and accept the alternative hypothesis that these two variables are actually dependent. Based on these results we can confirm Hypothesis 3.2, that SUD comments are more nonstandard than non-SUD comments. We should point out however that the study by Zwitter Vitez and Fišer (2016) to a certain extent contradicts our findings on SUD nonstandardness. Assuming that SUD comments are predominantly negative, then according to this earlier study (*ibid.*) the SUD comments should have been more standard than the non-SUD comments, for which we assume positive or neutral charge. Not knowing the actual sentiment distribution of our comments makes it difficult to reach a conclusion on this matter, but this certainly opens up an interesting next step in our future work.

While the prevalence of the O-category in both subsets was not unexpected, as “CMC language is prototypically known for the use of unconventional, non-standard spelling” (Verheijen et al. 2017), it is interesting that all the other categories were twice as frequent in the SUD-subset compared to non-SUD comments. This could indicate that nonstandard features in SUD comments are more profound and go deeper than the surface spelling deviations, which are typical of CMC language in general. Another interesting observation is the very low number of irrelevant comments (i.e., those written in languages other than Slovene) throughout the dataset ($\leq 1\%$), except in non-SUD migrants-related comments where the share of irrelevant tokens was 14%. While this is not the focus of our analysis, the fact that authors of non-hateful speech convey their message in different languages could indicate their closer connection with other cultures, which could also be a reason for their more inclusive stance.

4.3 Qualitative analysis

The quantitative analysis showed that differences between SUD and non-SUD comments exist on the level of lexical diversity and standardness, but not on the level of comment length. Following our aim to identify linguistic characteristics of SUD, we proceed with a qualitative analysis of certain aspects that stood out during the annotation process and analysis. In this section, we present the results

three items above (i.e., item 1: colloquial phrase formation,¹² item 2: interaction words, item 3: missing auxiliary verb) represent general CMC language characteristics rather than SUD-specific features, and suggest similarities between CMC in general and spoken language. In contrast, item 4 seems to highlight SUD-specific structures. The examples are especially interesting because they show how SUD comments are often formed as short, powerful calls to action or/and impersonal, infinitive structures which could indicate the author's desire to omit the performer of the action, be it themselves or their government. A similar observation was also reached in other studies (cf. Goli et al. 2016) which showed that shortening can be used to underline the specific illocutionary force of the message.

4.3.3 Nonstandard word order

Despite the fact that the syntax of Slovene language is characterized by free word order, certain placements of words are perceived by Slovene speakers as non-neutral. The analysis showed that such word order was used in more than 70% of the W-category tokens in both subsets. A closer look reveals that this is mainly due to the verbs that are placed at the end of the sentence. There were 45% of such cases in the SUD subset and 61% of such cases in the non-SUD subset (non-neutral position, i. e. final position of the verb in bold):

1. *take kot si ti bi jest na garmadi **zazgal*** (I would burn people like you at the stakes); *ste **se** drugače **obnašal*** (you behaved differently); *dam si jih **peljite*** (take them home); *v zemljo **se zabij*** (bugger off).

In the SUD comments, frequent non-neutral constructions of swear phrases with the noun preceding the adjective were also observed (non-neutral position of the adjective following the noun in bold):

2. *golazen **pedrska*** (fag vermin); *vlada **naša zablojena*** (our stupid government); *golazen **necloveska zblojena*** (stupid inhuman vermin); *vsiljivce **teroristične*** (terrorist intruders).

On the one hand, these examples once again point to the similarities of the studied comments with spoken discourse, while on the other they are indicative of the emphasizing role of non-neutral word order. Taking into account that in Slovene the most important information comes last, we could deduce that by placing a verb at the end the author puts the emphasis on the action, whereas in the case

¹² With the exception of the colloquial swear phrase formation which is by its nature typical of SUD.

of the adjective at the end of a sentence, the author probably wants to draw the reader's attention to the quality of the headword they modified and use it as a justification for the overall meaning of the comment.

4.3.4 Emoticons and emojis

Emoticons and emojis are not only emotion carriers, but are also used to convey the illocutionary force of the message (Dresner and Herring 2010). We first observed a well-known phenomenon of positive emoticons/emojis being used in SUD comments. The pragmatic function of this strategy is to weaken the illocutionary force of the message (Li and Yang 2018). For example:

3. *metek v glavo:)* (bullet to the head:); *noter bi jih zaprli pa se naj kurijo ;)* (let them be shut inside and burn ;); *lahko jim vržem samo ročno granato :P* (all I can do is to throw them a hand grenade :P).

Second, we analysed the overlapping emoticons/emojis, which were found to be of two kinds: thematic symbols (🔪) and facial expression symbols (:D). Negative symbols prevail in SUD comments (:-(, 🙄, 😏, 😞, etc.) while there are only three such symbols in non-SUD comments (:), 😊, :/). Although positive symbols are used in both subsets, non-SUD comments contain many more different symbols for happiness. We also noted the following subset-specific symbols: in SUD comments we found two symbols of physical and social power: 🦊 and 🐼, whereas in non-SUD comments we could observe symbols of love and peace: <3, :* and 🙌. These findings are not surprising per se as, in general, they underlie the content of the comments, as expected. However, the frequency, diversity, distribution and uniqueness of emoticons/emojis in a certain subset can be used as a good orientation point for further analysis of the underlying beliefs that are used as a basis for arguments in the comments.

5 CONCLUSION

This chapter looked into the basic surface linguistic features and the level of standardness of SUD comments in relation to non-SUD comments from the Slovene part of the FRENK corpus (Ljubešič et al. 2019). The analysis was based on a custom-built annotation schema which was used to manually annotate the dataset. The schema with its well-defined categories and dedicated guidelines is compact enough to ensure accurate annotations, and as such proved suitable for the scope of our analysis. In the future we suggest that the dataset should be

extended with multiple annotators to establish inter-annotator agreement. In addition, a multi-dimensional schema would be useful in order to obtain more detailed insights into the categories.

We conducted a quantitative analysis which revealed statistically significant differences between SUD and non-SUD comments on the level of lexical diversity and nonstandardness, but not on the level of comment length. The analysis showed that SUD comments are of similar length to non-SUD comments, but that SUD comments exhibit greater vocabulary diversity. Furthermore, we observed a lower frequency of emoticons/emojis and punctuation marks in SUD comments. The results also showed that SUD comments are indeed linguistically less standard than non-SUD comments, even though this feature did not prove as characteristic as initially expected, since the nonstandardness of non-SUD comments was also relatively high (30% vs. 22% respectively). SUD comments however exhibit a peculiar tendency towards nonstandard features (namely deviations in syntax and word order conventions) that surpass simple spelling errors (which are typical of CMC in general) and go deeper into the language structure.

In the qualitative analysis we further investigated the most striking features observed during the annotation process and quantitative analysis. The qualitative analysis showed that nonstandard punctuation marks are a characteristic feature of our entire dataset, but that expressive punctuation is more common in SUD comments. We also observed frequent non-neutral and informal syntactic structures and word order. On the one hand, this points to previously established similarities with spoken language, but on the other hand, close reading of those examples also showed that these elements have an emphasizing role in the comments and point to the tendency of the author to use them in order to justify or detach themselves from the content. Lastly, our analysis regarding the use of emoticons and emojis returned results that support findings from related studies. We found that emoticons and emojis are used to support the content and commonly also to weaken the illocutionary force of the message. The use of unique thematic emojis are indicative of the underlying beliefs held by the authors of the comments, and can be a good starting point for further sociolinguistic analysis of SUD.

With our analysis, we were able to address the surface linguistic features, which are an under-researched aspect of SUD, and thereby gain insights into the mechanics of the language used in SUD comments. We showed that differences between SUD and non-SUD exist in the surface linguistic structure, which is a promising finding that could assist machine learning and help better curb SUD online. However, our results can primarily serve as an orientation point due to the scope of the analysis, as indicated at the beginning of this section.

There are a number of areas following this study where we have been able to identify opportunities for future work. We envisage the examination of the variety of punctuation marks used and the characteristics of final punctuation to be one area. We also wish to further investigate the semantics of positive and unique thematic emoticons/emojis in SUD comments, and the semantic role of syntactic ellipsis and impersonal structures. Valuable insights could also be gained by investigating the syntactic complexity, vocabulary and argumentation strategies in SUD comments. Furthermore, it would be interesting to test if the identified SUD features help improve the automatic classification of SUD, as this would contribute to helping flag potential inflammatory discourse that may lead to aggravating effects. For comparative and validation purposes, we also intend to extend the analysis to the English subset of the FRENK corpus.

Acknowledgements

The work described in this chapter was funded by the Slovenian Research Agency within the national basic research project “Resources, methods, and tools for the understanding, identification, and classification of various forms of socially unacceptable discourse in the information society” (J7-8280, 2017–2020) and the Slovenian-Flemish bilateral basic research project “Linguistic landscape of hate speech on social media” (N06-0099, 2019–2023).

References

- Amaghlobeli, N. (2012): Linguistic features of typographic emoticons in SMS discourse. *Theory and Practice in Language Studies*, 2 (2): 348–354.
- Androutsopoulos, J. (2011): Language change and digital media: A review of conceptions and evidence. In N. Coupland and T. Kristiansen (eds.): *Standard Languages and Language Standards in a Changing Europe*: 145–161.
- Baider, F. H., Assimakopoulos, S., and Millar, S. L. (2017): Hate Speech in the EU and the C.O.N.T.A.C.T. Project. In F. H. Baider, S. Assimakopoulos, and S. L. Millar (eds.): *Online Hate Speech in the European Union – A Discourse-Analytic Perspective*: 1–6. Cham: Springer.
- Bajt, V. (2016): Anti-immigration hate speech in Slovenia. In N. Kogovšek Šalamon and V. Bajt (eds.): *RAZOR-WIRED. Reflections on Migration Movements through Slovenia in 2015*: 51–60. Available at: http://www.mirovni-institut.si/wp-content/uploads/2016/03/Razor_wired_publikacija_web.pdf (28 January 2020).
- Banks, J. (2011): European regulation of cross-border hate speech in cyberspace: The limits of legislation. *European Journal of Crime, Criminal Law and Criminal Justice*, 19 (1): 1–13. Available at: <https://doi.org/10.1163/157181711X553933>.

- Baron, N. S. (2002): *Alphabet to email: How written English evolved and where it's heading*. London: Routledge.
- Bieswanger, M. (2013): Micro-linguistic structural features of computer-mediated communication. In S. Herring, D. Stein, and T. Virtanen (eds.): *Pragmatics of computer-mediated communication*, vol 9: 463–485. Berlin: De Gruyter.
- Bočková, R. (2019): *The Use of Punctuation, Emoji and Emoticons in YouTube Abusive Comments*. Bachelor's thesis. Prague: Charles University, Faculty of Arts.
- Bojarska, K. (2019): *The dynamics of hate speech and counter speech in the social media – summary of scientific research*. Centre for the Internet and Human Rights (CIHR). Available at: https://cihr.eu/wp-content/uploads/2018/10/The-dynamics-of-hate-speech-and-counter-speech-in-the-social-media_English-1.pdf (28 January 2020).
- Bukovská, B. (2019): *The European Commission's Code of Conduct for Countering Illegal Hate Speech Online – An analysis of freedom of expression implications*. Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression. Available at: <https://www.ivir.nl/publicaties/download/Bukovska.pdf> (28 January 2020).
- Čeferin, R., and Mežnar, Š. (2014): Online hate-speech and anonymous internet comments: How to fight the legal battle in Slovenia? *Annales. Series Historia et Sociologia*, 24 (3): 477–488.
- Červ, G., and Kalin Golob, M. (2012): Sovražni govor kot tip vrednotenja: Jezikoslovni vidiki. *Teorija in Praksa*, 49 (1): 149–170.
- Čibej, J., Fišer, D., and Erjavec, T. (2016): Normalisation, tokenisation and sentence segmentation of Slovene tweets. In A. Utka, J. Vaičenonienė, and R. Butkienė (eds.): *Proceedings of Normalisation and Analysis of Social Media Texts (NormSoMe)*: 5–10.
- Crystal, D. (2001): *Language and the Internet*. Cambridge: Cambridge University Press.
- Dobrovoljc, K., et al. (2019): *Morphological lexicon Sloleks 2.0*. Slovenian language resource repository CLARIN.SI. Available at: <http://hdl.handle.net/11356/1230>.
- Dresner, E., and Herring, S. C. (2010): Functions of the nonverbal in CMC: Emoticons and illocutionary force. *Communication Theory*, 20 (3): 249–268.
- ElSherief, M., Kulkarni, V., Nguyen, D., Wang, W. Y., and Belding, E. (2018): Hate lingo: A target-based linguistic analysis of hate speech in social media. In *Proceedings of the Twelfth International AAI Conference on Web and Social Media*.
- Erjavec, T., Fišer, D., Čibej, J., and Arhar Holdt, Š. *CMC training corpus Janes-Norm 1.2*. Slovenian language resource repository CLARIN.SI. Available at: <http://hdl.handle.net/11356/1084>.
- Fišer, D., Miličević, M., and Ljubešić, N. (2018): Zapisovalne prakse v spletni slovenščini. In D. Fišer (ed.): *Viri, orodja in metode za analizo spletne slovenščine*: 124–139. Ljubljana: Znanstvena založba Filozofske fakultete.

- Franza, J., and Fišer, D. (2019): The lexical inventory of Slovene socially unacceptable discourse on Facebook. In J. Longhi and C. Marinica (eds.): *Proceedings of the 7th Conference on CMC and Social Media Corpora for the Humanities*: 43–47. Cergy-Pontoise University.
- Gill, A. (2011): Language and Personality in Computer-Mediated Communication. *Information Design Journal*, 19: 250–257.
- Goli, T., Osrajnik, E., and Fišer, D. (2016): Analiza krajsanja slovenskih sporočil na družbenem omrežju Twitter. In T. Erjavec and D. Fišer (eds.): *Zbornik konference Jezikovne tehnologije in digitalna humanistika*: 77–82. Ljubljana: Znanstvena založba Filozofske fakultete.
- Gorjanc, V. (2005): Diskurz slovenskih spletnih forumov–dokončen pokop strpnosti. *Večkulturnost v Slovenskem Jeziku, Literaturi in Kulturi*, 41: 20–29.
- Gorjanc, V. (2012): Ideologija heteronormativnosti, prevodna in jezikovna norma. In A. Bjelčević (ed.): *Ideologije v slovenskem jeziku, literaturi in kulturi*, vol. 48: 38–44. Ljubljana: Znanstvena založba Filozofske fakultete.
- Hilte, L., Vandekerckhove, R., and Daelemans, W. (2018): Adolescents' social background and non-standard writing in online communication. *Dutch Journal of Applied Linguistics*, 7 (1): 2–25.
- Ivanou, U. (2017): Stereotyping Vulnerable Groups. In S. Assimakopoulos, F. H. Baider, and S. Millar (eds.): *Online hate speech in the European Union: A discourse-analytic perspective*: 29–31. Dordrecht: Springer.
- Jakopič, K. (2013): Profesionalni razpihovalci sovražnega govora na spletu. *Medijska preža*, 44/45: 25.
- Jessmer, S. L., and Anderson, D. (2001): The effect of politeness and grammar on user perceptions of electronic mail. *North American Journal of Psychology*, 3 (2): 331–346.
- Jourová, V. (2019): *How the Code of Conduct helped countering illegal hate speech online*. EC Directorate-General for Justice and Consumers. Available at: https://ec.europa.eu/info/sites/info/files/hatespeech_infographic3_web.pdf (28 January 2020).
- Klein, G. B. (2018): Applied Linguistics to Identify and Contrast Racist 'Hate Speech': Cases from the English and Italian Language. *Applied Linguistics Research Journal*, 2 (3): 1–16.
- Laffen, A., and Fiorenza, B. (2012): Okay, My Rant is Over: The Language of Emotion in Computer-Mediated Communication. *COCOMP Computers and Composition*, 29 (4): 296–308.
- Lewandowska-Tomaszczyk, B. (2017): Incivility and confrontation in online conflict discourses. *Lodz Papers in Pragmatics*, 13 (2): 347–367.
- Li, L., and Yang, Y. (2018): Pragmatic functions of emoji in internet-based communication – A corpus-based study. *Asian-Pacific Journal of Second and Foreign Language Education*, 3 (1): 1–16.

- Ljubešić, N., and Erjavec, T. (2016): Corpus vs. Lexicon supervision in morphosyntactic tagging: The case of Slovene. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis (eds.): *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*: 1527–1531. European Language Resources Association (ELRA).
- Ljubešić, N., Fišer, D., and Erjavec, T. (2019): The FRENK Datasets of Socially Unacceptable Discourse in Slovene and English. In K. Ekštejn (ed.): *Text, Speech, and Dialogue*: 103–114. Dordrecht: Springer.
- López, C. A., and López, R. M. (2017): Hate Speech in the Online Setting. In S. Assimakopoulos, F. H. Baider, and S. Millar (eds.): *Online Hate Speech in the European Union – A Discourse-Analytic Perspective*: 10–12. Cham: Springer.
- Merljak, S. (2011): Samoregulacija spletnih medijev: Kodeks, moderiranje in celostna registracija uporabnikov. *Medijska preža*, 39: 25.
- Micciche, L. R. (2007): *Doing emotion: Rhetoric, writing, teaching. Language and the Internet*. Portsmouth, NH: Boynton/Cook Press.
- Motl, A., and Bajt, V. (2016): *Sovražni govor v Republiki Sloveniji: Pregled stanja*. Ljubljana: Mirovni inštitut.
- Muskat, R. V., and Assimakopoulos, S. (2017): The Implicit Dimension of Discriminatory Discourse. In S. Assimakopoulos, F. H. Baider, and S. Millar (eds.): *Online Hate Speech in the European Union – A Discourse-Analytic Perspective*: 42–46. Cham: Springer.
- Nielsen, L. B. (2002): Subtle, pervasive, harmful: Racist and sexist remarks in public as hate speech. *Journal of Social Issues*, 58 (2): 265–280.
- Neustupný, J. V., and Nekvapil, J. (2003): Language management in the Czech Republic. *Current Issues in Language Planning*, 4 (3–4): 181–366.
- Nockleby, J. T. (2000): Hate speech. *Encyclopedia of the American Constitution*, 3: 1277–1279.
- Osrajnik, E., Fišer, D., and Popič, D. (2015): Primerjava rabe ekspresivnih locil v tvitih slovenskih uporabnikov in uporabnic. In D. Fišer (ed.): *Zbornik konference Sloveščina na spletu in v novih medijih*: 50–56. Ljubljana: Znanstvena založba Filozofske fakultete.
- Pahor de Maiti, K., Fišer, D., and Ljubešić, N. (2019). How haters write: Analysis of nonstandard language in online hate speech. In J. Longhi and C. Marinica (eds.): *Proceedings of the 7th Conference on CMC and Social Media Corpora for the Humanities*: 37–42. Cergy-Pontoise University.
- Pavalanathan, U., and Eisenstein, J. (2015): Emoticons vs. Emojis on Twitter: A causal inference approach. *ArXiv Preprint ArXiv:1510.08480*. Available at: <https://arxiv.org/abs/1510.08480> (28 January 2020).
- Pertot, K., Petrovčič, M., and Strojan, N. (2016): #Analiza novih komunikacijskih elementov na družbenem omrežju @Twitter. In T. Erjavec and D. Fišer (eds.): *Zbornik konference Jezikovne tehnologije in digitalna humanistika*: 239–244. Ljubljana: Znanstvena založba Filozofske fakultete.

- Skubic, A. E. (2005): *Obrazi jezika*. Ljubljana: Študentska založba.
- Spina, S. (2017): Emoticons as multifunctional and pragmatic Resources: A corpus-based Study on Twitter. In *Media Corpora for the Humanities (Cmc-corpora17)*: 25–29.
- Splichal, S. (2017): Medijalizacija zasebnosti in javnosti: zla publiciteta, anonimnost in cenzura na internetu. In S. Splichal (ed.): *Zagovor javnosti: med svobodo izražanja in sovražnim govorom*: 9–29. Ljubljana: Slovenska akademija znanosti in umetnosti.
- Toporišič, J. (2004): *Slovenska slovnica*. Maribor: Obzorja.
- Toporišič, J. (2007): *Slovenski pravopis*. Ljubljana: Slovenska akademija znanosti in umetnosti.
- Tsesis, A. (2002): *Destructive messages: How hate speech paves the way for harmful social movements*. New York: NYU Press.
- Ullmann, S., and Tomalin, M. (2019): Quarantining online hate speech: Technical and ethical perspectives. *Ethics and Information Technology*: 1–12. Available at: <https://doi.org/10.1007/s10676-019-09516-z>.
- Vehovar, V., Motl, A., Mihelič, L., Berčič, B., and Petrovčič, A. (2012): Zaznava sovražnega govora na slovenskem spletu. *Teorija in Praksa*, 49 (1): 171–189.
- Verheijen, L., Fišer, D., and Beißwenger, M. (2017): WhatsApp with social media slang?: Youth language use in Dutch written computer-mediated communication. In D. Fišer and M. Beißwenger (eds.): *Investigating Computer-Mediated Communication: Corpus-Based Approaches to Language in the Digital World*: 72–101. Ljubljana: Ljubljana University Press.
- Vidgen, B., and Yasseri, T. (2019): Detecting weak and strong Islamophobic hate speech on social media. *Journal of Information Technology & Politics*, 17 (1): 66–78.
- Waqas, A., Salminen, J., Jung, S., Almerekhi, H., and Jansen, B. J. (2019): Mapping online hate: A scientometric analysis on research trends and hotspots in research on online hate. *PLOS One*, 14 (9).
- Zhang, Z., and Luo, L. (2019): Hate Speech Detection: A Solved Problem? The Challenging Case of Long Tail on Twitter. *Semantic Web*, 10 (5): 925–945.
- Zwitter Vitez, A., and Fišer, D. (2016): Linguistic analysis of emotions in online news comments an example of the Eurovision song contest. In D. Fišer and M. Beißwenger (eds.): *Proceedings of the 4th Conference on CMC and Social Media Corpora for the Humanities* (pp. 74–76). Znanstvena založba Filozofske fakultete. Available at: http://nl.ijs.si/janes/wp-content/uploads/2016/09/CMC-2016_Zwitter_Fiser_Linguistic-Analysis-of-Emotions.pdf (27 February 2020).
- Zwitter Vitez, A., and Fišer, D. (2018): Govorne prvine v nestandardni spletni slovenščini. In D. Fišer (ed.): *Viri, orodja in metode za analizo spletne slovenščine*: 254–272.