Melnyk, Lidiia (2021) Hate speech targets in COVID-19 related comments on Ukrainian news websites. Journal of Computer-Assisted Linguistic Research 5: 47-75. ISSN: 2530-9455.

Hate speech targets in COVID-19 related comments on Ukrainian news websites

Lidiia Melnyk
Friedrich Schiller University Jena, Germany
Corresponding author: lidiia.melnyk@uni-jena.de

Received: 22 July 2021 / Accepted: 29 October 2021 / Published: 22 December 2021

Abstract
The research focuses on hate speech in the comments section of Ukrainian news websites. Restricted to solely COVID-19 related comments, it seeks to analyze the development of hate speech rates throughout the pandemic. Using a semi-automated machine-learning-aided approach, the paper identifies hate speech in the comments and defines its main targets. The research shows that a crisis like the COVID-19 pandemic can strengthen existing negative stereotypes and gives rise to new forms of stigmatization against social and ethnic groups.

Keywords: hate speech, FastText, cluster

1. INTRODUCTION

The global outbreak of COVID-19 in 2020 caused significant emotional, social, and financial damage around the globe, which led to increased levels of fear and anxiety (Zandifar et al. 2020). Intense negative emotions manifested themselves in a growing number of micro-aggression incidents, verbal and physical abuse, and online harassment (Ziems et al. 2020). In a critical situation like the pandemic, people extensively consume online content and excessively turn to the Web for communication and opinion sharing.

In my research, I am seeking to analyse the rate of hate speech in COVID-19 related comments on Ukrainian news websites. I am focusing on the groups that became targets of hate speech during the pandemic and identifying if the new targets of hate speech emerged in response to it. Since Ukrainian society is bilingual, I am looking at the Russian and Ukrainian comments separately and then identifying similarities in the development of hate speech for the two languages and two news websites.

Online reader comments are "a new form of interactivity that could provide a larger public forum and a greater level of civic participation" (Erjavec 2014, 452). Cammaerts (2009) and Erjavec and Covacic (2012) point at the increasing level of hate speech in the news website

DOI: HTTPS://DOI.ORG/10.4995/JCLR.2021.15966
PUBLISHER: UNIVERSITAT POLITECNICA DE VALENCIA.

comments, which motivates the need for hate speech research of the online comments section. Offensive language, mobbing, and hate speech are frequent in online comments. Leaving an abusive comment in the comment section 'bears no immediate threat of sanctions, and users feel free to express open criticism or insult' (Dordevic 2020, n.a.). Though the primary goal of the comment section is to provide an opportunity for discussion, it becomes a platform for derogatory language and hate-igniting comments. The comment section turns into another medium for hate speech propagation. It increases negative stereotyping and creates new forms of harmful stigmatisation (Cotik et al. 2020).

Online hate is contagious and results in more hate speech comments (Ziems et al. 2020). Furthermore, the response to hateful comments is not sufficient as the percentage of hate speech is significantly higher than counter hate (Ziems et al. 2020), which necessitates the timely recognition of hate speech and appropriate reactions to it to prevent further micro-aggressions and insults. Though there is already a significant research body on one target group of hate speech, for example, anti-Asian hate during the pandemic (Ziems et al. 2020; Fan et al. 2020; Gee et al. 2020), it is crucial to identify all possible groups that might suffer from COVID-19 related stigmatisation.

The Ukrainian media has been actively using hate speech from 2006-2007 (Morhun 2016). In 2016, the Ukrainian Parliament passed a law, introducing 'Prohibition on the promotion of the exclusivity, superiority or inferiority of persons on the grounds of their religious beliefs, ideology, belonging to a particular nation, physical or property status, social origin' (Verkhovna Rada of Ukraine 2014, 904). After the annexation of Crimea, the Russian-Ukrainian conflict was consistently getting extensive media coverage in the Ukrainian news. It resulted in both the Russian and Ukrainian groups becoming targets of hate speech, as hate speech is an integral part of information warfare in the Russian-Ukrainian conflict and shows strong similarities with military propaganda (Postic 2018). Since the COVID-19 pandemic caused many financial, economic and social problems in Ukrainian society, the targets of hate speech might change and get more diverse.

Hate speech is a violent, offensive act of speech that intends to harm a specific group of people, in most cases, mentally. Depending on the effect hate speech might have on its target, there are two types of harm: constitutive and consequential damages (Maitra and McGowan 2012). Constitutive harm of hate speech is being enforced during speaking, whereas consequential harm occurs as a result of it. The immediate constitutive harm can result in psychological distress, a risk to a person's self-esteem (Gelber and MacNamara 2016), leading to psychological and mental distress. Consequential harm contributes to negative stereotyping, promotes discrimination (Maitra and McGowan 2012), and instigates real-life violent actions against the targeted group of people.

Hate speech and hate crime have previously been studied predominantly in the legal domain. The definition of it was crucial for the legal standpoint on the necessity of its criminalisation (Kennedy et al. 2018). The term ' hate speech' was coined by legal scholars from the United States in the 1980s, referring to how the legal systems fought against the harmful racist comments (Brown 2017, 424). From the legal perspective, it serves as an umbrella term for the discriminatory events against a specific group of people, whereby the concept itself is not new (Brown 2017, 426) but readjusted to the modern legislation. In its Recommendation against Racism and Intolerance, No 15 of 8 December 2015, the Council of Europe (2015) defines hate speech as 'the advocacy, promotion or excitement, in any form, of the denigration, hatred or vilification of a person or group of persons, as well as any harassment, insult, negative stereotyping, stigmatisation or threat of such a person or group of persons...' (3). Besides the outlined forms of hate speech, it can also be expressed implicitly through the approval or justification of crimes against humanity or genocide and the glorification of individuals involved in these crimes. The European Commission emphasises the necessity to define hate speech in its implicit and explicit expressions to protect persons or groups of persons that could become a target of it and suffer any physical, mental, emotional, or any other type of harm as a consequence.

There are many different perspectives on what constitutes hate speech. In my research, I consider abusive language aimed to insult, harm, or promote violence against a specific group of people that all share common characteristics (gender, race, ethnic affiliation, age group) to be hate speech. I distinguish between hate speech and vulgar language or insults, where personal insults not based on negative group stereotyping and without a reference to a group membership are not considered hate speech. Hate speech should always have a group marker.

Hate speech in the comments section on Ukrainian news websites correlates with the level of social acceptance of the governmental response to the pandemic. The rate of hate speech is dependent on the source of news, but not on the language of the comment. Even though the news on COVID-19 causes the creation of new hate speech targets, the comments section remains an important part of the Ukrainian-Russian information war, strengthening already existing negative stereotypes.

2. DATA COLLECTION

2.1 Data Sampling

For my data sampling, I selected an unbiased representative source with the ability to filter the meta topic and obtain historical data. I decided to scrape the comments from COVID-19 dedicated sections of the unmoderated online forums of famous news websites. In this way, the author of the comment is already primed by COVID-19 related news and is likely to build his argumentation around the news.

I selected two media sources to provide a comprehensive overview of the societal response to COVID-19. The websites meet the following selection criteria:

- It is a national-level, not a local media source
- It is possible to sort out the news on the ' COVID-19' meta topic
- The source enables commenting and does not require social media identification so that users can create an account under a made-up name to prevent any pressure or fear of being recognised

- The comments can be scraped with a web scraper.

Censor.net and Korrespondent.net met the selection criteria. Each website has its specific HTML structure, therefore, I wrote a separate web scraper for each of them. The scrapers were programmed in Python and were based on the BeautifulSoup library and Urllib handling module. The scrapers went through all the COVID-19 related news, opened each link to the news message and went through the comments section. The scrapers retrieved text of the comments, date of comments, reply to comments if existent, and the name of the author. The author's name will not be published or used in the research.

Censor.net reports 250,000 unique daily visitors, which accumulated to over 7,000,000 unique visitors every month. Censor.net is more beloved by the male audience (63% of the readership) than by the female one (37%) (Censor.net n.d.). The target audience of the media source is an average Ukrainian aged 25-30, who is interested in the domestic and foreign politics of Ukraine, most likely living in a big Ukrainian city. Readers aged between 25 and 54 years old represent 71% of the website visitors.

Founded in 2000, Korrespondent.net covers events that are happening in Ukraine and the world. Korrespondent.net targets a middle-class Ukrainian audience, small and medium-sized business owners, with a 77% male readers ratio. The source attracts roughly 15,000,000 unique readers monthly, which accounts for 145,000,000 views (UMH n.d.).

2.2 Corpus metadata

The corpus consists of 107 thousand unique comments split into 470 thousand sentences. Out of the entire corpus, 25% stem from Korrespondent.net and 75% - from Censor.net.

Retrieved comments range between January 2020, the first mention of the new virus, and June 2021, the last update of the data. The comments cover the entire development of COVID-19 pandemic and, given the diversity of the sources and different target groups of their readers, can be considered representative of the general opinions in Ukrainian society. Korrespondent.net had a few hundred mentions of the coronavirus from 2013-2016 sorted out, whereas Censor.net data from January and February 2020 was not available.

Comments are both in Russian and Ukrainian languages, but the language distribution varies depending on the source:

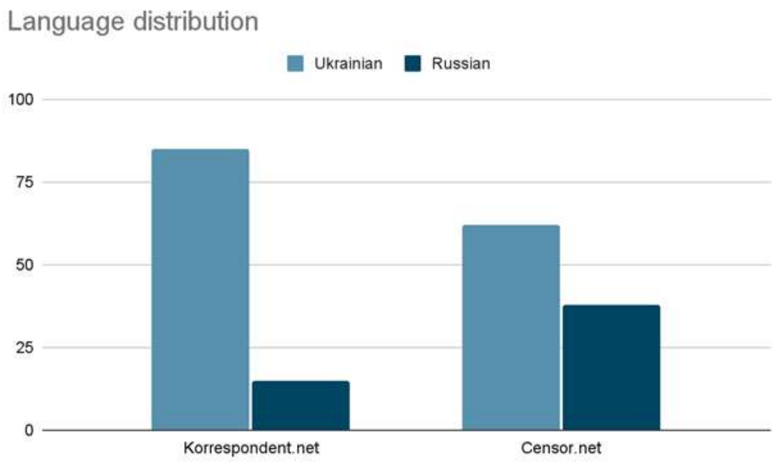

FIGURE 1. LANGUAGE DISTRIBUTION

2.3 Annotation guidelines

The research relies on an unsupervised machine learning model from FastText. For text categorisation purposes, the uploaded training data included only two possible labels: HATE (if the annotator considers the comment to represent hate speech) or NO (for all the other cases).

The students of Lviv Polytechnic University carried out the complete annotation under the mentorship of Olena Levchenko. Ten students volunteered to assist with the annotations. They were divided into pairs. Each pair worked simultaneously on a different set of annotations, equally divided into both Russian and Ukrainian. Annotated comments were uploaded to the public repository on GitHub and can be found in open access[1].

The annotators followed the FastText friendly labelling:

- ' __LABEL__NO' for all the comments that do not have hate speech in them;
- ' __LABEL__HATE' for data points that can be characterised as hate speech.

The presence of any group reference or marker was crucial for distinguishing between hate speech and personal insults. As long as they do not rely on harmful stereotyping, personal insults are not considered hate speech. Hate speech comments insult a person's human dignity by expressing a call for violence, dehumanisation, negative stereotyping, and downplaying historically motivated attacks (Kennedy et al. 2018). Complaints about or insults of the government or political parties were not considered hate speech for current research.

Call for violence is 'the use of vulgar, humiliating, and discriminating words, phrases, and sentences to harm, attack, and label others, victims of which can be vulnerable groups in the society' (Wang 2018, 18). In 'čym bil'še kacapiv zdoxne tym krašče' ('The more kacaps die, the better'), the author is implicitly calling for violence through wishing death to the

[1] Melnyk, Lidiia. "Ru_ukr_hate_speech_comments: Annotated Russian and Ukrainian Comments for Hate Speech Identification." GitHub. Accessed October 11, 2021. https://github.com/LidiiaMelnyk95/Ru_Ukr_hate_speech_comments.

representatives of the Russian nation, who are being referred to through hate-based ethnic slurs.

Dehumanisation can be expressed through the comparison of a group to an animal or denial of human rights and characteristics. For example, ' Rjus'ky svyni zbyrayut'sja kupyty 2 mil'jony doz vakcyny Pfayzer' (' Russian pigs intend to purchase 2 million doses of Pfizer vaccine'). The sentence is considered hate speech as it dehumanises the Russians.

Negative stereotyping is harmful as it might intensify negative expectations of the conduct of a group (Ramos-Oliveira and Pankalla 2019). For example, by saying that 'Kytays'ki liky zvučyt' nadto zahrozlyvo' ('Chinese medicine sounds too threatening'), a person refers to the negative stereotype that products created in China have low quality. Therefore, medicine from China might be dangerous.

Downplaying historical attacks is also hate speech as it uses historical and cultural context to voice negative sentiment towards a group, empower or refer to hateful ideology, and disrespect the national tragedy of marginalized groups. A comment 'vam li ne znat' što lutšeye mesto observacii eto koncentratsionnyy lager' ('Don't you know that a concentration camp is the best place for observation') downplays the Holocaust in a mockingly sarcastic way.

Comments like 'Zelenskyi pod šumok klyančit po miru deneg štob po-bystromu potom s3,14dit' ('Zeleskyi is begging for money around the world to steal it later quickly') or 'Zelenskyi ko vsemu yešče i zaigravšiysya posredstvennyy akteriško tyažko bol' noy narcissizmom' ('Moreover, Zelenskyj is a mediocre actor who is seriously ill with narcissism') target one person and are not considered hate speech.

2.4 Inter-annotator agreement

Kennedy et al. (2018) outline a high level of inter-annotator disagreement regarding hate speech annotation due to personal subjective 'differences in understanding of the definition of hate speech, interpretations of the annotated texts, or evaluating harms done to certain groups.' To estimate the level of inter-annotator agreement, I computed the Cohen's Kappa agreement coefficient and Matthews correlation coefficient to balance out possible overoptimistic results from Cohen's Kappa (Chicco, Warrens and Jurman 2021). The coefficients were computed separately for the Russian and Ukrainian languages.

Comments	Cohen's Kappa Score	Matthews corrcoef score
Ukrainian comments	0.431	0.433
Russian comments	0.41	0.404

TABLE 1. INTER-ANNOTATOR AGREEMENT SCORES

The annotators have reached a slightly lower agreement with the Russian comments. Cohen's Kappa scores range between -1 and 1, with 1 being perfect agreement, 0 - random agreement, and -1 standing for complete disagreement. The scores for both annotations represent moderate agreement (Hallgren 2012), which is plausible for the highly subjective topic of hate speech. The third annotator went through the comments the annotators disagreed on and decided whether it is hate speech or not.

3. DATA PROCESSING

Data processing consisted of the following steps:

Step 1. Removal of comments with unusual activity
Step 2. Language identification and division of comments into the Russian and Ukrainian comments
Step 3. Comments splitting
Step 4. FastText classification of the Ukrainian and Russian comments
Step 5. Part-of-speech tagging and lemmatisation of the comments
Step 6. Word2vec vectorization of the Ukrainian and Russian data
Step 7. K-means clustering of the Ukrainian and Russian data

TABLE 2. DATA PROCESSING PROCEDURE

3.1 Bots, trolls, fake accounts

Both sources represent unmoderated online forums, which makes them an easy target of bots and trolls. The use of fake accounts and bots is a well-established phenomenon in the Ukrainian media space in the so-called 'information war' (Sopilko 2016) between pro-Ukrainian and pro-Russian Internet users.' The main purposes of the utilisation of computational propaganda tools include not only manipulating public opinion but often discrediting opponents and defending the interests of different business and political groups' (Zhdanova and Orlova 2017, 18). Bots, trolls, and fake accounts are the most commonly used automated, manual or semi-automated tools to influence public opinion in the Ukrainian media space (Zhdanova and Orlova 2017), with fake accounts being the most frequently used tool.

There exists a solid research base on depicting bots on such social media as Twitter or Facebook (Liu 2019; Mazza et al. 2019). Widely used libraries such as Botometer rely on the analysis of user profiles and user activity, the information which is being provided by the social media API (Application Programming Interface). Unfortunately, there was no user-specific information from Censor.net and Korrespondent.net. Nevertheless, it was still possible to identify the atypical behaviour of some users. I identified the most active commenters for both sources, republishing the same comment multiple times under different news articles within a short time. Carrying out the complete classification of the users into bots/fake accounts and people would require building a separate machine learning model and preparing an annotated training corpus (Skowronski 2019). That would go beyond the focus of my research. Therefore, to ensure that accounts demonstrating unusual activity are not influencing the findings, I kept only the original comment and removed the duplicates.

3.2 Language identification model

The peculiarity of the contemporary language situation in Ukraine is the co-existence of Ukrainian and Russian and different forms of bilingualism (Zbyr 2015). Distinguishing between the two languages is necessary for the text classification model and semantic vectorization to function properly.

I used the Multinomial Naive Bayes (MNB) classifier as it provided the most precise results compared to such libraries tested as FastText, Langid, Langdetect, Textblob, Polyglot, and Google Translator libraries. MNB considers all attributes (i.e., features) to be ' independent of each other given the context of the class, and it ignores all dependencies among attributes' (Jiang et al. 2016, 346). I used the Python script to develop the classifier based on the Scikit-Learn module (Sklearn).

All the comments were considered to belong to either Russian or Ukrainian, depending on the probability of the MNB classifier model. The measures to evaluate the MNB model include precision, recall and f-1 score. Precision is the ratio *tp / (tp + fp)*, with *tp* being the number of true positives and *fp* - the number of false positives. Recall is calculated as *tp / (tp + fn)*, where *fn* stands for false negatives. F-1 score is defined as a weighted harmonic mean of the precision and recall with 1 being the best value and 0 - the worst.

The character-based MNB classifier approach proved to be relatively accurate with the following precision and f-1 scores:

language	precision	recall	f-1 score	support
Russian	0.94	0.99	0.97	6966
Ukrainian	0.99	0.93	0.96	6852

TABLE 3. MNB PERFORMANCE IN LANGUAGE IDENTIFICATION

The Sklearn classification report also provides support values for the MNB model, support is the number of occurrences of each class in a one-dimensional array-like matrix of ground truth target values.

The supervised MNB Classifier model was trained specifically on the news taken from Censor.net. The training sets consisted of 6,000 identical news in Russian and Ukrainian, referring to coronavirus topics.

The minimum threshold value was 70%. I excluded non-Cyrillic comments. In case the probability of adherence was lower than 70%, the comment was manually reviewed and assigned to a specific language.

3.3 Hate speech identification with FastText

FastText is an open-source library to train supervised and unsupervised models for text representation and text classification. Joulin et al. (2016) introduced FastText as a word classification and text classification approach in 2016. Simple in use and training, FastText has proven to be equal in accuracy to deep learning classifiers. It has previously been used for sentiment analysis and hate-speech detection (Pratiwi, Budi and Alfina 2018). 'It is based on n-gram features, dimensionality reduction, and a fast approximation of the softmax classifier (Joulin et al. 2016). We show that a few key ingredients, namely feature pruning, quantisation, hashing, and re-training, allow us to produce text classification models with tiny size, often less than 100kB when trained on several popular datasets, without noticeably sacrificing accuracy or speed' (Joulin et al. 2017). The adaptability and lightweight of the library paired with a good precision score were the deciding factors for choosing it as a classifier.

Hate speech is not equally distributed through an entire comment. One hate speech sentence might result in the whole comment being classified as hate speech, which could negatively affect the model's precision.

The Ukrainian and Russian models were downloaded from the FastText repository and run with a programmed Python script. The Python script used followed the guidelines suggested by FastText developers[2]. The examples used to train the models consisted of the annotated comments described above. I tested the FastText classification model on both comments consisting of multiple sentences (the model trained on the full annotated comments) and sentences separately with the model trained on sentences. Comments splitting resulted in a 6% higher precision for both Russian and Ukrainian models.

Measure	Russian comments full	Ukrainian comments full	Russian comments split into sentences	Ukrainian comments split into sentences
Recall	0.87	0.85	0.93	0.91
Precision	0.87	0.86	0.93	0.90

[2] "Text Classification · Fasttext". n.a.. Fasttext.Cc. https://fasttext.cc/docs/en/supervised-tutorial.html.

f-1 score	0.87	0.86	0.93	0.91

TABLE 4. FASTTEXT MODEL PERFORMANCE

The eyeball inspection of the hits also proved that splitting comments into sentences provides more accurate results.

The hate speech rate in the comments was assumed to depend on the development of the pandemic in Ukraine. The first cases of COVID-19 in Ukraine occurred in March 2020. A strict lockdown took place between mid-March and mid-June 2020. After that, the government introduced an adaptive lockdown framework with restrictive measures being introduced only in the regions with high infection rates. In November 2020, the 'weekend lockdown' was introduced, with all stores and gastronomy closing over the weekend to prevent big indoor gatherings. A second strict lockdown took place between the 8th and 24th of January 2021, returning to the adaptive lockdown framework afterwards. The first vaccinations took place in February 2021. The last effort to introduce lockdown and curfew restrictions in Ukraine took place in April 2021.

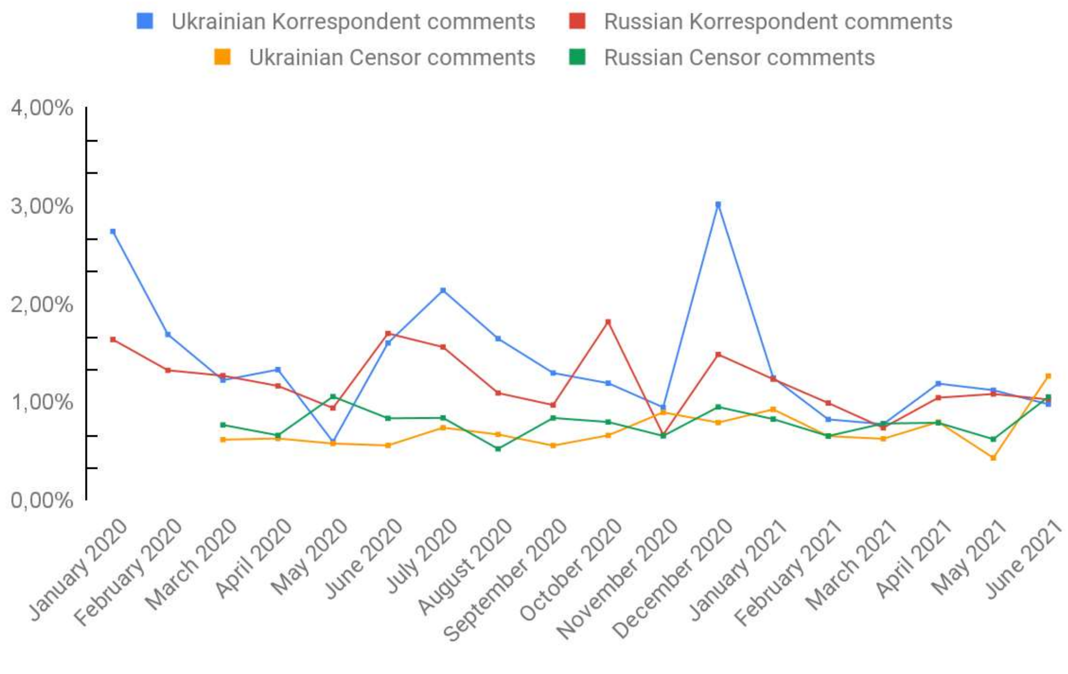

FIGURE 2. HATE SPEECH DYNAMIC BETWEEN 01.2020 and 06.2021

The average percentage of hate speech for all the sources combined is 1%. In Korrespondent.net, the rate of hate speech in Russian comments reached 1.17% on average, whereas for Ukrainian comments, it is slightly higher and is on average 1.3%. Censor.net, with the majority of comments being written in Russian, displayed the following hate speech statistics:

- a relatively low average percentage of Ukrainian hate speech compared to Korrespondent.net by 0.7%
- 0.78% of all Russian sentences were classified as hate speech.

Therefore, the share of hate speech in the totality of comments seems to be more dependent on the source or news the source is publishing (Dordevic 2020) than on the language. The share of hate speech in Censor.net for both Ukrainian and Russian comments was almost the same, it was similar for Korrespondent.net.

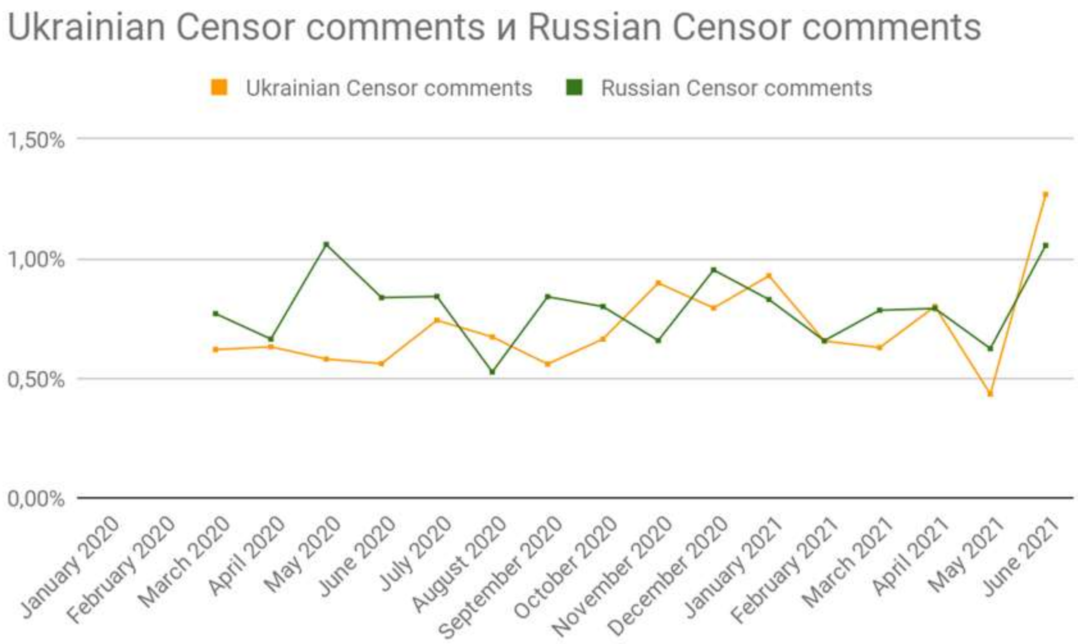

FIGURE 3. HATE SPEECH DYNAMICS ON CENSOR.NET

Some differences can be observed between the spread of hate speech in Russian and Ukrainian comments. The hate speech percentage of Russian comments peaked in April 2020, with the other upward tendencies in September 2020, November 2020, and June 2021. The hate speech rate of Ukrainian comments has not demonstrated much dynamics until a sudden hike in June/July 2020 and the increases in hate speech percentage in October 2020 and December 2020.

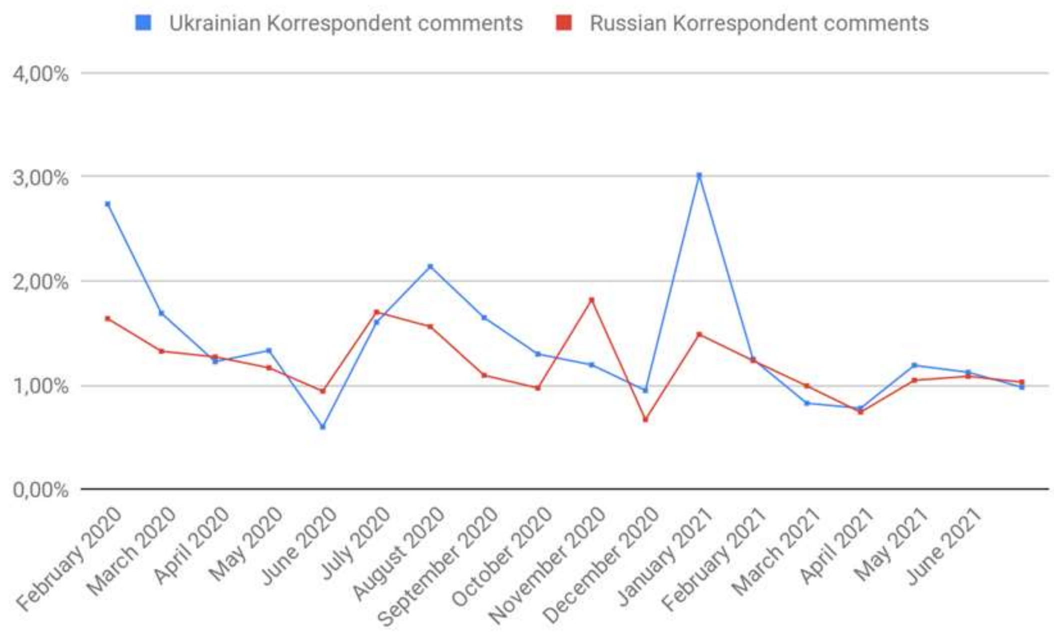

FIGURE 4. HATE SPEECH DYNAMICS ON KORRESPONDENT.NET

Korrespondent.net has demonstrated a much higher share of hate speech in its comments. The users left an overwhelming majority of comments in Ukrainian in Spring 2020, with the user activity gradually losing its dynamics throughout the research period. The same phenomenon could also be applied to the comments in Russian, with the user activity being at its absolute peak in the time between February and June 2020.

The share of Ukrainian hate speech on Korrespondent.net appears to be more dynamic than on Censor.net, with the absolute highest hate speech rate dating in December 2020. Another significant peak was evident in March/June 2020 and at the very beginning of the pandemic in January 2020.

The dynamics of the hate speech rate in Russian comments does not fully coincide with the Ukrainian. Thus, after the initial high start in January 2020, it demonstrated drastic growth in April/May 2020, with the next all-time high peak in September 2020, which was not present in the Ukrainian corpus. Then, the subsequent increase was experienced in December 2020, but its percentage is significantly lower than the Ukrainian counterpart.

3.4 Word embedding

To vectorise my data, I needed to normalise and lemmatise it. After the initial preprocessing and data cleaning, I carried out part-of-speech-tagging with the help of Stanza, an open-source Python natural language processing toolkit (Qi et al. 2020). Stanza was pre-trained by its creators on both Russian and Ukrainian datasets. I kept only verbs, nouns, proper names, adjectives, adverbs, and numbers. I also conducted a Symspell-based spell-check and spelling corrector to standardise the data and get it ready for lemmatisation.

Word2vec vectorisation is the next step of research needed to get vector representation of lemmas to group them by their semantic similarity later (Sivakumar et al. 2020). It is commonly applied for analogy and semantic analysis of words and performs well with all sizes of data sets. Therefore, it was also used for further classification of sentences labeled as hate speech. I got the numerical representation of lemmas in vector space with skip-gram-based Word2vec optimised with negative sampling. Based on past appearances, Word2vec can make highly accurate guesses on the context of individual words. It would place words from hate speech comments used in a similar context close to each other.

Word2vec was preferred over other vectorisation models (e.g. FastText) due to its proven efficacy for semantic tasks. Both FastText and Word2Vec are based on the skip-gram models and both display good results in the vectorisation of the Slavic languages. FastText vectorisation is highly efficient in working with new data. FastText is also frequently recommended due to the simplicity and high speed of its training (Cothenet 2020). Nevertheless, the focus of the model centred around the interpretability of regular words and the result was not reliant on the identification of uncommon lexemes, in which case Word2vec is recommended over FastText (Kriazhova 2019). Moreover, for the vectorisation of the Russian language Word2vec shows slightly higher precision (Kriazhova 2019).

3.5 K-Means Clustering

The data size was reduced to comments classified as hate speech. My next step involved K-means clustering, unsupervised categorisation of words into semantically similar groups (Singh, Tiwari and Garg 2011), as it has a proven record of performing well with small data sets. The clustering was performed with Python script using the Sklearn module. To increase the density and accuracy of clustering, I improved the K-means clustering method with the expectation-maximisation algorithm.

K-means clustering was carried out with the results provided below. The number of clusters was estimated using the silhouette method (Yuan and Yang 2019). The silhouette method for k-means clustering aims at maximising the silhouette score with the correct number of clustering being reached when the silhouette score is the greatest. I manually tested different ranges of K for clustering with the help of the Sklearn metrics module in Python. The highest possible silhouette score for my data was achieved at five clusters.

The silhouette score scope falls between [-1;1], where 1 indicates that the data point is compact within its clusters and far from the other clusters. A silhouette score of 0 denotes overlapping clusters, whereas -1 indicates bad clustering performance. The silhouette coefficient for each data point is calculated as follows:

$$s(\boldsymbol{o}) = \frac{b(\boldsymbol{o}) - a(\boldsymbol{o})}{\max\{a(\boldsymbol{o}), b(\boldsymbol{o})\}}$$

- ***s(o)*** stands for the silhouette coefficient for ***o*** data point;
- ***a(o)*** is the average distance between ***o*** and the other data points in the cluster

- ***b(o)*** is defined as the minimum average distance between the current datapoint ***o*** and the other clusters, to which the datapoint in question does not belong.

Source	Silhouette Score
Ukrainian Censor.net	0.45
Russian Censor.net	0.58
Ukrainian Korrespondent.net	0.52
Russian Korrespondent.net	0.62

TABLE 5. K-MEANS CLUSTERING PERFORMANCE

The best K-means silhouette score results were achieved for the Ukrainian comments of Censor.net and the Russian comments of Korrespondent.net. The silhouette score is above 0 for all the clusters, which verifies the plausibility of the clustering method.

3.6 Censor.net Clusters

Hate speech comments from Censor.net were divided into 5 clusters, with the fifth cluster being the smallest and the most distant one.

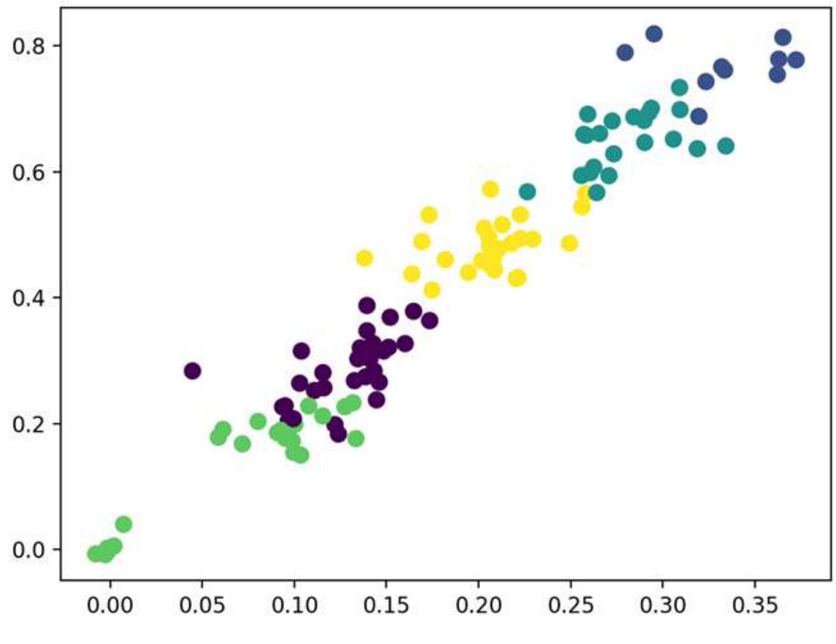

FIGURE 5. 2-D REPRESENTATION OF THE UKRAINIAN CLUSTERS FROM CENSOR.NET

Cluster 73%	Cluster VACCINE	Cluster WORLD	Cluster RUSSIA	luster Cluster UKRAINE
73 %	vakcyna	ukrajinskij	73 %	ukrajina
baran	zbroja	spravžnij	moskovskyj	naselennja
baranovirus	kazaty	cydity	kacap	idiot
vybory	pravda	svit	kacapskyj	kytaj
idijot	robyty	koronavirus	moskal'	narod
krajina	vakcynacija	kytajskyj	rosija	vakcynuvaty
holosuvaty		kytajka		ukrajinec'
daun		zaraza		debil
maska		zelenyj		zemlja
zebiloty		zebobik		holova
zebily		vlada		zakon
znaty		virus		miscevyj
stado		vynnyj		namordnyk
lox		vereščaty		
loxdaun		velykyj		
loxtorat		bilyj		
luhandon		kytajec'		

TABLE 6. TRANSLITERATED UKRAINIAN CLUSTERS FROM CENSOR.NET

The 73% cluster is actively targeting the Ukrainian population and, in particular, people who voted for the current president. They are frequently referred to as '73%' (as 73% of the voters supported Zelenskyi during the elections), 'zebily' (created through the blending of the swearing word 'debily,' Ukrainian for fool/imbecile, and prefix Ze from Zelenski), 'zebilota' (same as previously explained), 'daun' (Ukrainian for mentally disabled person), 'loxtorat' (created through the blending of a swearing word ' lox' (loser/douche) and ' elektorat' (electorate).

The 73% cluster also includes lemmas used to condemn people who follow the restrictions and any actions aimed at imposing lockdown to stop the spread of the virus. Thus, lockdown was frequently referred to as 'loxdaun' (lockdown for idiots) and coronavirus as 'baranovirus' (compound word for goat virus) and such lemmas as 'baran' (goat) and 'stado' (herd). Therefore, the cluster of hate speech comments targets the Ukrainian population, who voted for the current president as it is to blame for the restrictive measures put by the government on its voters or, in other words, herd.

The VACCINE cluster and RUSSIA clusters are both relatively small. The second one includes such lemmas as vaccination, vaccine, weapon, and truth. The topic of vaccination was widely present in the Ukrainian news. Therefore, the cluster points to the presence of vaccination discussion in society. The RUSSIA cluster contains few lemmas referring to Russia and Russians, including 'kacap,' ' kacaps' kyj,' 'moskovs' kyj'' with ' kazap' being an example of negative stereotyping against ethnic groups.

Lemmas such as 'vlada' (government), 'zelenyj' (the green),' zebobiky' (rude way of referring to the government), ' koronavirus' (coronavirus), ' naselennja' (population) represent the WORLD cluster. In the Ukrainian media space, ' the green' is being used in connection to the current government due to the literal interpretation of the president's surname. Other lemmas present in the cluster include China, the Chinese, and contagion, which shows that the Chinese population also became the target of hate speech. 'Namordnyk' (muzzle), the insulting way of referring to the face mask, was also found in this cluster.

The UKRAINE cluster is, similarly to the first one, focusing on the Ukrainian situation in particular. It mentions Ukraine, Ukrainian, country, population, territory, and local. It also contains many vulgar words such 'zebil,' 'douche' and 'Idiot.' The majority of lemmas found in the hate speech comments were targeting the Ukrainian community, especially government supporters, including mentions of other countries and ethnic groups such as the Russian and Chinese populations.

Clustering of the Russian comments from Censor.net achieved the best Silhouette score performance regarding density and separation of the clusters.

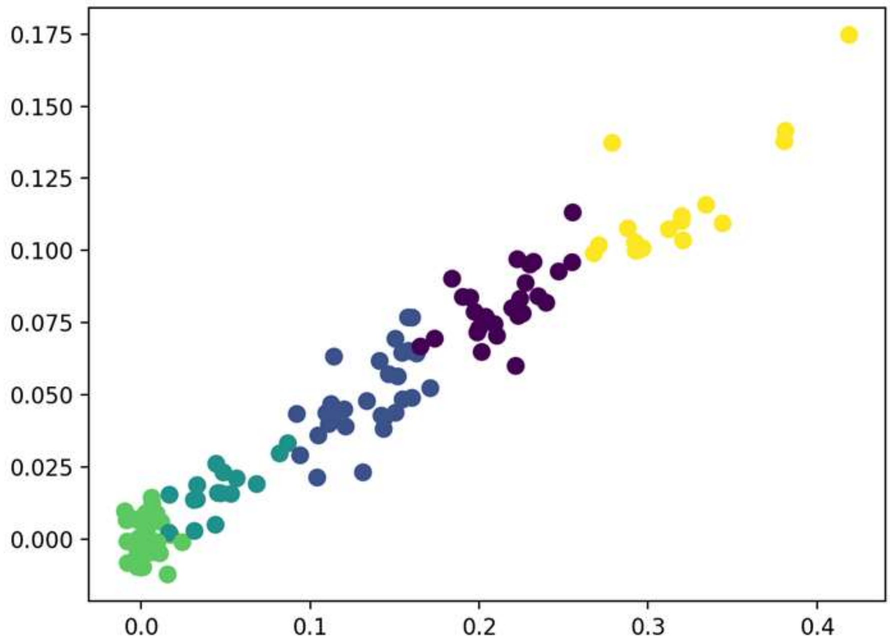

FIGURE 6. 2-D REPRESENTATION OF THE RUSSIAN CLUSTERS ON CENSOR.NET

The vector representation of lemmas in the first (bright green cluster) is the densest and closest to each other, whereas the fifth (yellow one) cluster displays some outlier positions.

Cluster INNER LIFE	Cluster RUSSIA	Cluster INTERNATIONAL	Cluster NEIGHBORS	Cluster 73%
cerkov'	virus	73%	lox	73%
umnyj	zelenskyj	baranovirus	pandemija	polnyj
umeret'	maska	kitaj	pol' ša	ploxoj
ukrainskij	period	vakcinirovat'	smertnost'	ostal' noj
tupoj	pravo	vozdux	rumynija	nedelja
stepanov	real' nyj	vrač	zakryt'	naselenie
resul' tat	rost	granica	granica	million
problema	rf	evropa		kovid
porjadok	rynok	zabolet'		zebuin
porošenko	test	italija		zebil

loxdaun	ukraina	idiot		ze
lečit'	sputnik	karantin		zarplata
korona	putin	kiev		zarazit'
zelenyj	raška	kitaez		debil
žyd	pokupat'	koronavirus		verit'
		namordnik		vakcyna
		privesti		baran
		kitajskij		president

TABLE 7. TRANSLITERATED RUSSIAN CLUSTERS FROM CENSOR.NET

The INNER LIFE cluster includes mostly terms not related to the COVID-19 topic, even though it mentions coronavirus and 'lohdown.' It also mentions Stepanow, current health minister of Ukraine, and Poroshenko, the former president. The cluster includes an ethnic anti-sionist slur 'žyd.'

The RUSSIA cluster contains such lemmas as ' rf' (Russian Federation), ' Sputnik' (Russian vaccine), 'putin,' ' raška' (mixture of the English word Russia and Ukrainian suffix -k- to mark belittlement and disdain), ' virus' (virus) and 'zelenskyj.' The cluster centers around the suggestion of the Ukrainian president to purchase the Russian vaccine.

The INTERNATIONAL cluster contains mentions of China, Europe, and Italy, which were all coronavirus hotspots. It also mentions vaccinations ('varkcinirovat') and quarantine ('karantin'). It contains insulting referrals to coronavirus as 'baranovirus' (goat virus), which was also present in the Ukrainian clusters, and ethnical slurs such as 'kitaez' (insulting way to say Chinese).

The NEIGHBORS cluster includes mentions of the neighboring countries on the western border, including ' pol' ša' (Poland) and 'rumynija' (Romania). The cluster talks about the neighbors' experience in battling the pandemic, mentioning border ('granica') closures and mortality rates.

Similarly to the Ukrainian comments, the supporters of the current president become one of the main hate speech targets in the Russian comments on Censor.net. The group referral to his voters is ' 73%', ' zebily' (formed through the fusion of 'Zelenskyj' and 'debil,' douche/idiot), ' zebujiny' (created through the fusion of ' Zelenski' and 'babujin,' baboon). The cluster includes lemmas for population, millions, and the president.

3.7 Korrespondent.net Clusters

There were only 11 thousand sentences written in Ukrainian on Korrespondent.net. Therefore, the sample of hate speech comments is relatively small but performs well, keeping in mind its Silhouette (0.52) scores.

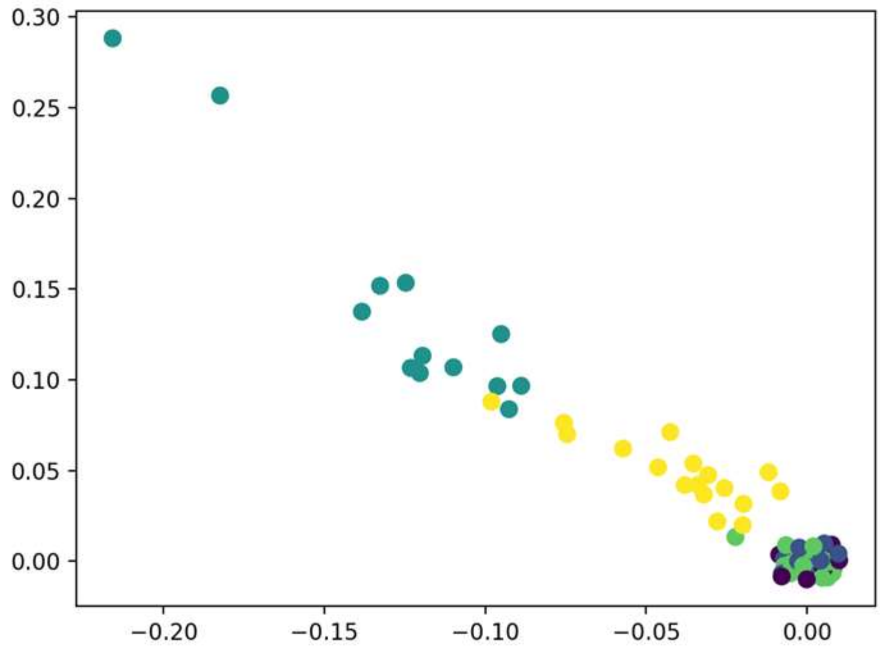

FIGURE 7. 2-D REPRESENTATION OF THE UKRAINIAN CLUSTERS FROM KORRESPONDENT.NET

As shown in the visualisation, the three clusters (green, dark blue, and dark violet) are located very closely to each other and are overlapping, which could be due to the small size of the data set. In contrast, vector representation of the light blue cluster hints at some anomalies at its representation as two of the data points are not connected to the other vectors.

Cluster CONFLICT	Cluster CRIMEA	Cluster PRO-RUSSIAN	Cluster RUSSIA	Cluster 73%
svynja	ukrajinec'	73%	73%	73%
ukrajina	haz	banjat	dohnuty	dohnyty
prypynennja	terytorija	šarovaryj	rjus' kyj	zebabujin
potribnyj	krym	lapti	janukovyč	zebily
luhandony	rubl'	dykun	ameryka	zebujin
luhandon	srsr	pravda	zahnyvaty	

loxdaun		transportuvannja	krajina	
luhanda		vakcyna	lapta	
		vatan	nezabarom	
		vatka	pohano	
		kyzjak	raš̓ka	
		morh	rašnjavyty	
		nafta	rosija	
		putin	staty	

TABLE 8. TRANSLITERATED UKRAINIAN CLUSTERS FROM KORRESPONDENT.NET

'Luhandon,' 'Luhanda' and ' Luhandony' are all highly abusive ethnic slurs created as a fusion of Luhansk, the eastern Ukrainian territory, and colloquial lemma for a condom ('Handon'). Other lemmas from the cluster, such as 'svynja' (pig) and Ukraine, make it possible to assume that the population of the Luhansk region has become a hate speech target.

The CRIMEA cluster does not contain any lemmas with a negative sentiment. It includes such words as 'krym' (Crimea), ' srsr,' ' terrytorija' (territory), 'rubl' (ruble). Presumably, the cluster revolves around the problematic situation around Crimea's separation from Ukraine.

Lemmas like ' Vatan' and 'vatka' possessed highly negative connotations and appeared in the Ukrainian media space around 2013th. Currently, the concept is being used to refer to the Ukrainian or Russian citizens who are actively supporting Russian politics and ideology. Together with 'lapta' (ethnic slur calling Russians after the name of their traditional shoes) and 'banjat' (ban), the modern Internet slang concentrated in the TROLLS cluster. The 73% cluster includes abusive names used to call the current president's supporters, such as 'zebily,' ' 73%,' 'zebujiny,' and 'zebabujiny.'

The distribution of the Russian Korrespondent.net clusters on a vector space displays some similarities with the Russian comments from Censor.net and Ukrainian comments from Korrespondent.net. The light blue cluster is the most loosely distributed on the vector space, with the other clusters having some connected data points.

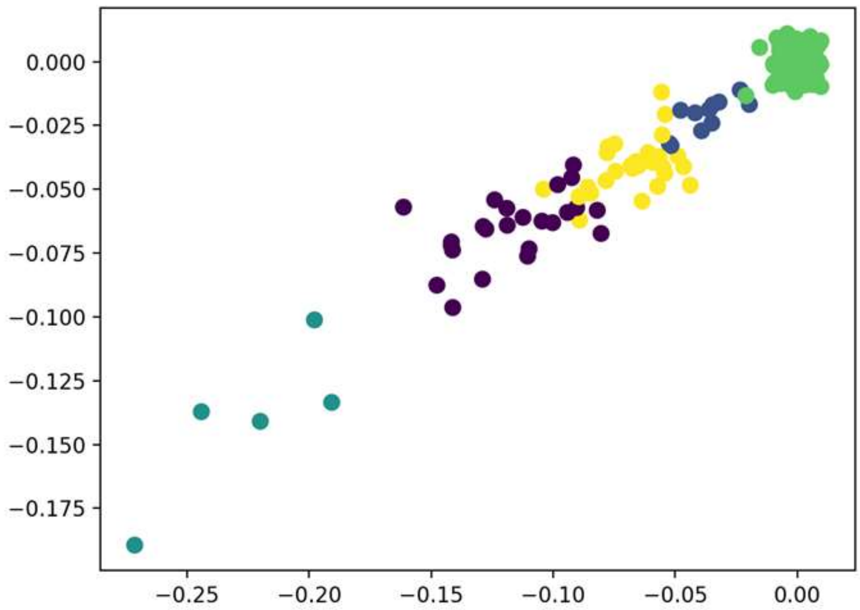

FIGURE 8. 2-D REPRESENTATION OF THE RUSSIAN CLUSTERS FROM KORRESPONDENT.NET

The light blue vectors are further away from each other than the vectors from the other four clusters.

Cluster CHINA	Cluster PRO-RUSSIA	Cluster PRO-UKRAINE	Cluster VIRUS	Cluster 73%
legkie	der' mo	xoxol	bol' noj	moskal'
ukrainskij	životnoe	test	virus	73%
ukraina	zarazit'	ukra	gonkong	tupoj
kitaec	debil	sdoxnut'	zabolet'	ukrainec
milliard	donbass	xrjukat'	granica	massovyj
special' no	vatan	ukronazik		vakcyna
zavesti	koronovirus	jutjub		vozvraščenije
kitaj	kreml'	xyxla		zebil
čajniz	kazah	ploxoj		škola
	naselenie	poezd		cena

	paraška	prijti		urod
	luhandon	ljuboj		pasport
	podderžat'	koronavirus		medik
	pozor	naved		informacija
	rossijskij	kitaez		
	rossijanin	italija		
	svin' ja	bystro		
	sraka	bandery		
	krysa			
	putinskij			

TABLE 9. TRANSLITERATED RUSSIAN CLUSTERS FROM KORRESPONDENT.NET

Like in the other three groups of comments reviewed, the 73% cluster was present in the Russian hate speech comments from Korrespondent.net. The cluster includes such lemmas as 'zebil' and '73%', which links it to the 73% voters group, 'pasport,' 'price' (passport), 'informacija' (information), and 'ukrainec' (Ukrainian). The cluster also contains vaccination-related lemmas.

The Russian comments from Korrespondent.net are the only part of the research data, which allows for the clear identification of the CHINA cluster with lemmas like 'kitaj' (China), 'kitaec' (Chinese), 'čajniz' (derogatory use of transliterated English word Chinese), 'legkie' (lungs) and 'ukraina' (Ukraine). Many hate speech comments here focused on targeting the Chinese population in the context of coronavirus spread.

The VIRUS cluster is the smallest one and consists of such lemmas as 'virus' (virus), 'bol'noj' (seak), 'zabolet'' (get seak), 'granica' (border), and 'gonkong' (hong kong). A single lemma 'hong kong' is not enough to suppose a group reference to the Hong Kong population in the cluster.

The PRO-RUSSIAN cluster in the Russian comments, similarly to the Ukrainian comments, targets people referred to as 'vatan' or 'vatnik,' who are avid supporters of the Russian regime. The cluster here also includes 'kreml' (Kremlin) and 'donbass' (shortening for Doneck basin, the pro-Russian Ukrainian region), also referred to through the ethnic slur 'Luhandon.' The cluster contains lemmas with relatively neutral connotations like 'rossija' (Russia), 'rossijanin' (Russian), 'Putin,' but also 'paraška' (which is an abusive derogatory way of referring to Russia through merging country name with the jargon word for the toilet).

With words like 'ukra' (a disdaining shortening for Ukrainians), 'xyxl,' and 'xoxol' (ethnic slurs referring to Ukrainians), combines the PRO-UKRAINIAN cluster hate speech comments targeted at the Ukrainian population. The cluster contains lemmas such as 'ukronazik' (compound word for Ukrainian nazis) and 'bandery' (proper name derivative stemming from the leader of the Ukrainian rebel army). Therefore, the PRO-UKRAINIAN cluster targets the opposite group than the PRO-RUSSIAN cluster.

4. DISCUSSION

The hate speech rate in the comments is dynamic and grows in response to socially challenging situations. Figure 2 shows that at the beginning of the pandemic, the rate of hate speech was already at its peak. It corresponds with the general tendency, which can be due to 'the virus spreading at accelerated speed within China and beyond' (Stechemesser, Wenz and Levermann 2020, n.a.). In February 2020, the percentage of COVID-19 related hate speech started decreasing due to China's imposing strict quarantine rules but peaked again with the further spread of cases around the globe.

Figure 2 reflects the upward tendency of hate speech rate for both Censor.net and Korrespondent.net between April to May 2021. In Ukraine, in this period, the government imposed the first restrictions, and the number of daily cases started snowballing. Due to the loosening of regulations and switching to the 'adaptive quarantine model,' the hate speech rates for both languages and sources reached their absolute lowest in June 2020.

The hate speech percentage starts growing for all the sources in the autumn months. Even though the 'adaptive quarantine' model had not been officially lifted, it did not cause many restrictions on Ukrainian social life. However, in November 2020, the government introduced 'the weekend lockdown,' which required closing all the entertainment facilities over the weekend. It brought much disagreement in society, which is reflected in the growing hate speech rate.

Hate speech in the comments was relatively high in December 2020 since the virus was easier to spread in the closed indoor room, where people tend to gather more in winter. Another possible reason for the hike in hate speech in the winter months might be the delayed vaccination or implementation of restrictions to prevent growing infection numbers.

The rate of hate speech in the comments peaked again in April 2020, which complies with the time of the last phase of restrictions in Ukraine. Kyiv, for example, was required to entirely close restaurants, shopping malls, fitness studios, and other entertainment facilities and restrict the use of public transportation. It could have contributed to the spike of hate speech in the comments.

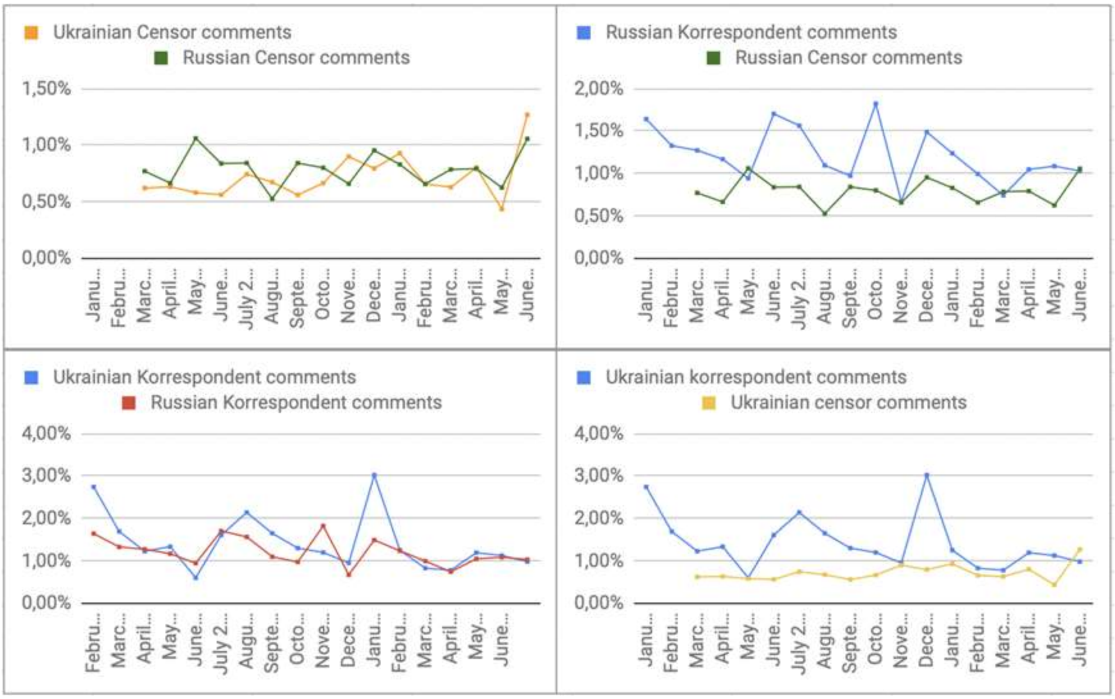

FIGURE 9. SOURCE-BASED AND LANGUAGE-BASED COMPARISON OF THE HATE SPEECH DYNAMICS

Hate speech appears to be more dependent on the website than on the language. Comments in the Russian language have similar dynamics between April-July 2020 and November 2020-January 2021, whereas the development of hate speech comments in the Ukrainian language has no evident crossing points.

Hate speech targets during the pandemic had both internal and external sources. The spread of the virus created many social, financial, and political challenges in Ukraine. The growing disagreement with the current politics of the government and its ability to solve the arising challenges resulted in negative stereotyping of a specific social group, namely the electorate of the current president. The so-called 73% cluster was present in all the four parts of hate speech comments analysed. Hate speech targeted at this group frequently was expressed in the form of dehumanisation through insulting comparisons of the group to baboons and goats and insults. There were also cases of abuses against human dignity by questioning the intellectual capacities of the group members, frequently referred to as "zebily".

Hate speech against the Russian population was present around all the comment groups. Anti-Russian sentiment has been dominant in the Ukrainian media as a part of information conflict for the hegemony in the Ukrainian media space (Isakova 2016). Russian citizens as well as Ukrainian migrants from Donetsk and Luhansk represent the leading target of hate speech in Ukraine (Isakova 2016). Korrespondent.net included in the cluster the Russians as the ethnic community and people, who might as well be Ukrainian, who are avid supporters of Russian politics and ideology. Hate speech is expressed through:

- abusive ethnic slurs like 'moskal' or 'kacap,'
- insults of human dignity ('Luhandon,' 'paraška')

- dehumanization (comparison to pigs and rats)

The Russian vaccine, 'Sputnik V,' is also frequently mentioned in hate speech comments. Given the widespread anti-Russian sentiment, it is likely that the negative stereotyping effect of hate speech influenced attitudes towards the Russian vaccine.

Whereas Censor.net hate speech clusters targeted such external groups as the neighbouring countries, China or the European Union, the hate speech of Korrespondent.net is more intrinsically focused. For example, Ukrainian hate speech comments of Korrespondent.net centred around the conflict in the East of Ukraine and Crimea. Hate speech against the population of Donetsk and Luhansk region insults human dignity through abusive references like 'Luhandon.' Naming the area 'Luhanda' hints at negative stereotyping not only against the region but also against an African country Uganda as the derogatory merge of ' Luhansk' and ' Uganda' is supposed to hint at the low level of economic and social development in the area.

Only the Russian comments from Korrespondent.net contained an anti-Ukrainian sentiment. This cluster of hate speech arises as a counterargument to the anti-Russian hate speech and is represented through ethnic slurs ('xoxol,' 'ukra') and negative stereotyping ('ukronazyk,' 'bandery').

As COVID-19 was spreading from China to other countries worldwide and came from the European countries to Ukraine, hate speech obtains external vectors. Stechemesser, Wenz and Levermann (2020) prove the growing anti-Chinese and anti-Asian sentiment due to the virus's origins. Both Russian and Ukrainian hate speech comments from Censor.net and Russian comments from Korrespondent.net are targeting the Chinese population and blaming it for causing the virus. INTERNATIONAL, CHINA, and WORLD clusters contain ethnic slurs like 'kitaez' or ' čainiz.' Hate speech explicitly targeted the Chinese population without generalising it into anti-Asian sentiment. Hate speech comments also contained mentions of other foreign countries like Poland or Romania.

5. CONCLUSIONS

Hate speech rate started at its peak with the first mentions of the COVID-19 and continued to be dynamic throughout the pandemic. It starts growing along with the introduction of the new restriction measures, but displays a downward trend during the loosening of restrictions or decreasing of infection rate. The hate speech rate does not fully coincide between languages and sources, but appears to be more source-dependent than language-dependent. Thus, for example, hate speech comments from Korrespondent.net peaked at similar times in both Ukrainian and Russian subgroups.

Hate speech on Ukrainian news websites strengthens existing negative stereotypes and gives rise to new forms of harmful stigmatisation (Cotik et al. 2020). The major target of hate speech is not COVID-related, but can be strengthened through it. Reflecting growing disagreement with the government, hate speech comments target the group of Ukrainians, who voted for and supported the current president. The aforementioned anti-Russian and anti-Ukrainian clusters as well as hate speech against Doneck and Luhansk populations are not new to the Ukrainian media space (Isakova 2016; Postic 2018) and have not explicitly emerged in response to the news about COVID-19. The spread of the virus resulted in the emergence of the new targets of hate speech in the comments under analysis.

REFERENCES

Brown, Alexander. 2017. "What is hate speech? Part 1: The myth of hate." *Law and Philosophy* 36, no. 4: 419-468.

Cammaerts, Bart. 2009. "Radical Pluralism and Free Speech in Online Public Spaces." *International Journal of Cultural Studies* 12, no. 6: 555–75. https://doi.org/10.1177/1367877909342479.

Chicco, Davide, Matthijs J. Warrens, and Giuseppe Jurman. 2021. "The Matthews Correlation Coefficient (MCC) is More Informative Than Cohen' s Kappa and Brier Score in Binary Classification Assessment." *IEEE Access* 9: 78368-78381.

Cothenet, Come. 2020. "Short Technical Information About Word2vec, Glove And FastText." Last modified May 25, 2020. https://towardsdatascience.com/short-technical-information-about-word2vec-glove-and-fasttext-d38e4f529ca8.

Cotik, Viviana, Natalia Demand, Franco M. Luque, Paula Miguel, Augustin Moro, Juan Manuel Perez, Pablo Serrati, Joaquin Zajac, and Demian Zayat. 2020. "A study of Hate Speech in Social Media during the COVID-19 outbreak." Paper presented at EMNLP 2020 Workshop NLP-COVID Submission, November 16-20.

Council of Europe. 2015. "On combating hate speech." OCRI general policy recommendation 15. Accessed July 10, 2021. https://rm.coe.int/ecri-general-policy-recommendation-no-15-on-combating-hate-speech/16808b5b01

Dordevic, Jasmina P. 2020. "The sociocognitive dimension of hate speech in readers' comments on Serbian news websites." *Discourse, Context and Media* 33. doi:10.1016/j.dcm.2019.100366.

Erjavec, Karmen, and Melita Poler Kovačič. 2012. ""You Don' t Understand, This is a New War!" Analysis of Hate Speech in News Web Sites' Comments." *Mass Communication and Society* 15: 899 - 920.

Erjavec, Karmen. 2014. "Readers of online news comments." *Annales.* Series historia et sociologia 24, no. 3: 451-462.

Fan, Lizhou, Huizi Yu, and Zhanyuan Yin. 2020. "Stigmatization in social media: Documenting and analyzing hate speech for COVID-19 on Twitter." *Proceedings of the Association for Information Science and Technology. Association for Information Science and Technology* 57. doi:10.1002/pra2.313.

Gee, Gilbert C., Marguerite J. Ro, and Anne W. Rimoin. 2020. "Seven reasons to care about racism and COVID-19 and seven things to do to stop it." *American Journal of Public Health* 110: 954-955.

Gelber, Katharine, and Luke McNamara. 2016. "Evidencing the harms of hate speech." *Social Identities* 22: 324 - 341.

Hallgren, Kevin A. 2012. "Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial." *Tutorials in Quantitative Methods for Psychology* 8, no. 1: 23–34. doi:10.20982/tqmp.08.1.p023.

Isakova, Tamara. 2016. "Hate speech as the problem of Ukrainian information space." *Strategic Priorities* 41, no. 4: 90-97.

Jiang, Liangxiao, Shasha Wang, Chaoquin Li, and Lungan Zhang. 2016. "Structure extended multinomial naive Bayes." *Information Sciences 329*: 346-356. doi:10.1016/j.ins.2015.09.037.

Joulin, Armand, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. "Bag of Tricks for Efficient Text Classification." *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics 2*: 427-431.

Joulin, Armand, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Herve Jegou, and Tomas Mikolov. 2016. "FastText.zip: Compressing text classification models." Paper presented at ICLR 2017. https://arxiv.org/abs/1612.0365.

Kennedy, Brendan, Mohammad Atari, Aida Mostafazadeh Davani, Leigh Yeh, Ali Omrani, Yehsong Kim, Kris Coombs, Shreya Havaldar, Gwenyth Portillo-Wightman, Elaine Gonzalez, Joseph Hoover, Aida Azatian, Alyzeh Hussain, Austin Lara, Gabriel Olmos, Adam Omary, Christina Park, Clarisa Wijaya, Xin Wang, Yong Zhang, and Morteza Dehghani. 2018. "The Gab Hate Corpus: A collection of 27k posts annotated for hate speech." *PsyArXiv.* July 18. doi:10.31234/osd.io/hqjxn.

Kriazhova, Anastasiia. 2019. "Методы оценивания семантической близости фраз для классификации текстовых сообщений" [Methods for estimating the semantic proximity of phrases for classification of text messages]. Paper presented at Mezhvuzovskaya nauchno-tekhnicheskaya konferentsiya studentov, aspirantov i molodykh spetsialistov im. YEV Armenskogo, 2019.

Liu, Xia. 2019. "A Big Data Approach to Examining Social Bots on Twitter." *Journal of Services Marketing* 33, no. 4: 369–79. doi:10.1108/jsm-02-2018-0049.

Maitra, Ishani, and Mary Kate McGowan. 2012. "Introduction and Overview." In *Speech and harm: Controversies over free speech,* edited by Ishani Maitra and Mary Kate McGowan, 1-24. Oxford University Press on Demand.

Mazza, Michele, Stefano Cresci, Marco Avvenuti, Walter Quattrociocchi, and Maurizio Tesconi. 2019. "RTbust." *Proceedings of the 10th ACM Conference on Web Science - WebSci '19:* 183-192. https://doi.org/10.1145/3292522.3326015.

Morhun, Anastasiia. 2016. "Війна Росії проти України: тенденції конфліктного дискурсу в сучасних українських ЗМІ." [The Russian war against Ukraine: Tendencies of the conflict discourse in the current Ukrainian media]. Naukoví zapiski Ínstitutu polítichnikh i yetnonatsíonal' nykh issledovaniy im. ÍF Kurasa NAN Ukraïni Наукові записки Інституту політичних і етнонаціональних досліджень ім. ІФ Кураса НАН України 3-4: 352-362.

Pratiwi, Nur Indah, Indra Budi, and Ika Alfina. 2018. "Hate Speech Detection on Indonesian Instagram Comments Using FastText Approach." Paper presented at 2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS). doi:10.1109/ICACSIS.2018.8618182.

Postic, Irena. 2018. "Hate speech and propaganda in Russian-Ukrainian conflict." PhD diss., Universität Wien.

Qi, Peng, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D.Manning. 2020. "Stanza: A Python Natural Language Processing Toolkit for Many Human Languages." Paper presented at 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. doi:10.18653/v1/2020.acl-demos.14.

Ramos-Oliveira, Diana, and Andrzej Pankalla. 2019. "Negative stereotypes: an analysis of Social Cognition in different ethnic groups. Social cognition of stereotypes." *Psicogente* 22: 1-15.

Singh, Vivek Kumar, Nisha Tiwari, and Shekhar Garg. 2011. "Document Clustering Using K-Means, Heuristic K-Means and Fuzzy C-Means." Paper presented at 2011 International Conference on Computational Intelligence and Communication Networks. doi:10.1109/cicn.2011.62.

Sivakumar, Soubraylu, Lakshmi Sarvani Videla, T Rajesh Kumar, J. Nagaraj, Shilpa Itnal, and D. Haritha. 2020."Review on Word2Vec Word Embedding Neural Net." Paper presented at 2020 International Conference on Smart Electronics and Communication (ICOSEC): 282-290. doi:10.1109/ICOSEC49089.2020.9215319.

Skowronski, Jason. 2019 "Identifying Trolls and Bots on Reddit with Machine Learning (Part 2)." Last modified July 30, 2019. https://towardsdatascience.com/identifying-trolls-and-bots-on-reddit-with-machine-learning-709da5970af1.

Sopilko, Iryna. 2016. "Тролі, боти та бот-мережі як загрози розвитку інформаційного суспільства" [Trolls, bots and bot-network as danger to the development of information society]. *Naukovi pratsi Natsional'noho aviatsiynoho universytetu.* Серія: Юридичний вісник «Повітряне і космічне право» 4, no. 41: 86-91.

Stechemesser, Annika, Leonie Wenz, and Anders Levermann. 2020. "Corona crisis fuels racially profiled hate in social media networks." *EClinicalMedicine* 23. doi:10.1016/j.eclinm.2020.100372.

Wang, Qingning. 2018. ”Linguistic violence and online political communications in China: The example of 鸡的屁 (Ji De Pi) as an ironic spoof of gross domestic product in online debates around environmental issues.” *Global Media and China* 3: 18-31.

Yuan, Chunhui, and Haitao Yang. 2019. “Research on K-value selection method of K-means clustering algorithm.” *J* 2.2: 226-235.

Zandifar, Atefeh, Rahim Badrfam, Nami Mohammadian Khonsari, Marzieh Assareh, Hossein Karim, Mehdi Azimzadeh, Mohammad Noori Sepehr, Ramin Tajbakhsh, Fatemeh Rahimi, Nima Ghanipour, Arash Agoushi, Saeed Hassani Gelsefid, Fateme Etemadi, and Mostafa Qorbani. 2020. ”COVID -19 and medical staff' s mental health in educational hospitals in Alborz province, Iran.” *Psychiatry and Clinical Neurosciences* 74, no. 9: 499-501. doi:10.1111/pcn.13098.

Zbyr, Iryna. 2015. ”Contemporary Language Issues in Ukraine: Bilingualism or Russification.” *Journal of Arts and Humanities* 4: 45-54.

Zhdanova, Mariia, and Dariya Orlova. 2017. ”Computational Propaganda in Ukraine: Caught between external threats and internal challenges.” Computational Propaganda Research Project, University of Oxford. Working Paper. http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Comprop-Ukraine.pdf.

Ziems, Caleb, Bing He, Sandeep Soni, Naren Ramakrishnan, Diyi Yang, Srijan Kumar, Georgia Institute of Technology, and Virginia Tech. 2020.”Racism is a Virus: Anti-Asian Hate and Counterhate in Social Media during the COVID-19 Crisis.” https://arxiv.org/abs/2005.12423v2.

Verkhovna Rada of Ukraine. 2014. “Про суспільне телебачення і радіомовлення”[On television and broadcasting]. Vidomosti Verchovnoyi Rady Ukrayiny *Відомості Верховної Ради України 27: 904.* https://zakon.rada.gov.ua/laws/show/1227-18#Text

Censor.net. n.d. “Реклама На Портале [Advertising on the Website].” Accessed July 22, 2021. https://censor.net/ru/reklama.

UMH. n.d. ”Korrespondent.net.” UMH MEDIA KIT. Accessed July 22, 2021. https://umhdigitalspecials.com/korrespondent.