
Abstract - This research focuses on the interrelation between news content on COVID-19 of three largest online news sites in Latvia (delfi.lv, apollo.lv, tvnet.lv) and the audience reaction to the news in the Latvian and Russian channels during the state of emergency. By using a tool for audience behaviour analysis, the Index of the Internet Aggressiveness (IIA), for analysis of audience comments, the study aims to uncover how and whether news about COVID-19 affect the level of audience aggressiveness. The study employs two data collection methods: news content analysis and IIA data analysis, in which ten index peaks are selected in each of the two emergency periods (spring 2020, fall and winter 2020/21). The study data consists of content analysis of 400 news items and analysis of ~80,000 comments, identifying the level of aggressiveness, the number and structure of comment keywords. The results show that the level of public aggressiveness is only partially formed by the attitude towards COVID-19 news: less than half of the most aggressively commented news is devoted to information about COVID-19. An increase in the level of aggressiveness of the audience of online news sites can be observed at the end of 2020 and at the beginning of 2021 when it is higher than over the course of 2020.

IIA is an online comment analysis platform, which analyses user-generated comments on news on online news sites according to pre-selected keywords, allowing to grasp the dynamics of commenters’ verbal aggressiveness. In addition, IIA exploits a machine learned classifier to recognize not only potentially aggressive keywords but also to analyse the entire comments. In January 2021, the IIA data set consists of ~24.89 million comments (~611.97 million words) added to ~1.34 million news articles.

Keywords - big data; COVID – 19; internet users’ comments; online news; verbal aggressiveness

I. INTRODUCTION

Latvia, just like other countries, conveyed information related to the COVID-19 pandemic in 2020 and 2021 – restrictions, officials’, and experts’ recommendations on health preservation – to the society through media. Whereas the residents’ mobility opportunities are reduced the role of Internet media in communication is soaring. In Latvia, the largest news portals – Delfi, Tvnet and Apollo, with over half a million unique users per day [1] – constantly provide updates on the COVID-19 pandemic and give the opportunity to react to the news in comments. This study is based on the assumption that society’s attitude is partly expressed in the comments section of the COVID-19 news and it can be measured with the help of The Index of the Internet Aggressiveness (IIA).

Given that the pandemic caused many inconveniences, increased insecurity in various groups of the society and affected the general mood of the society the aim of this study is to find out whether the information provided by online media about COVID-19 affects the level of users’ aggressiveness on news sites’ comment sections. Using data from the original digital research tool IIA, this study analyses the data on COVID-19 news content and...
audience attitudes expressed in comments, focusing on the level of aggressiveness towards COVID-19 news.

In Latvia, the pandemic is characterized by two periods of state of emergency (spring 2020 and autumn 2020/winter 2021). Each of them is different in the severity of the pandemic: in the first stage Latvia had a very low incidence of the new coronavirus disease and high public confidence in information sources (media, government, experts, etc.) [2], while in the second stage the epidemiological situation in Latvia was much more serious – the health care system was in danger of collapsing [3], the organization of vaccination was delayed [4], while the level of public confidence in the government and the media decreased rapidly [5]. By mid-February 2021, Latvia ranked the 6th in the European Union and the European Economic Area in terms of cumulative infection rates [6]. Hypothesizing that the attitude of Internet news commenters conditionally characterizes the public mood, but the course of the pandemic depends on the attitude of the population towards the information related to COVID-19, the following research questions are defined in this study:

1) How does the news about COVID-19 in the two emergency phases affect the level of aggressiveness in the online comments recorded by the IIA?
2) How does the content of COVID-19 news affect the proportion of aggressive keywords included in the content of comments in Latvian and Russian languages?

To answer the research questions, a study design was developed, which analysed the data automatically provided by the IIA and the news content analysis data, comparing the number of comments, the number of keywords describing aggressiveness and the content of the news in two languages and over two emergency periods. The data were evaluated in the context of theoretical findings on verbal aggressiveness in Internet comments.

II. LITERATURE REVIEW

User comments are one of the most popular forms of audience participation [7]. Scholars have had different views about the importance of interactivity in the process of news consumption [8], [9]. Distribution, content, and interactivity opportunities of online news audience comments have developed studies that evaluate the influence of Internet media user activities on the development of the public sphere [10], [11], as well as the quality of discussions and the involvement of the public at large in discussing important issues. Data from these studies have shown that an opportunity to speak or to ask whether the anonymity of users facilitates verbal aggression. Of importance is the idea that the psychological effect of online anonymity [18] encourages people to engage in self-disclosure at a far higher level than is the case with face-to-face communications.

III. METHODS

The research method was developed in several phases. In order to design the IIA platform (http://barometrs.korpuss.lv), 3,500 potentially “aggressive” word forms were selected from a sample comment corpus of 65,000 words, establishing a stock of 900 basic words in all forms and spelling variations. Text normalization [19] was conducted to handle transliteration and to obtain a more homogenous dataset. Keywords in the study were grouped into 11 categories: discriminatory words, swear words, injustice, calls of violence, aggressive activities, individuals symbolically associated with aggression, associations with aggressiveness, military terms related to aggression, denunciation, aggressiveness-related emotions, and treason. Each category was assigned with a specific weight depending on the level of aggressiveness: highly, moderately and less aggressive words. Next, a formula was devised to calculate the level of verbal aggressiveness on each news portal during any given period of time. The IIA formula takes into account the number of potentially aggressive keyword occurrences per day multiplied with the pre-calculated weights of keywords. The weighted sum of keyword occurrences is divided by the total number of words in user-generated comments per day.

In a later phase of development, a machine-learning experiment was conducted. First, a representative dataset of 10,000 comments was created, where each comment was human-annotated as aggressive or non-aggressive. Each comment was independently classified by at least two annotators; a third annotator intervened if the two made opposite choices. By using this dataset, an automatic binary classifier was trained. As the evaluation shows [19], the accuracy of the automatic classifier is 72.2% which is rather close to the inter-annotator agreement of 78%. The evaluation also shows that the whole-comment classifier produces a trendline that correlates to the trendline produced by analysing isolated aggressive keywords. This confirmed the hypothesis that the use of potentially aggressive words in a comment suggests that the entire comment is aggressive.

Further improvements to the IIA platform made it possible to identify the top news articles which attracted comments that led to the peaks of the IIA, determining the most aggressive comments as well.

To determine the average value of aggressiveness the data points of the trend graphs are calculated as follows. First, the total relative frequency of the potentially
aggressive keywords is calculated for each date by dividing the total count of keyword occurrences with the total word count for that date. This, however, results in a rather noisy graph with a lot of local peaks in both directions. Therefore, second, a smoothened trend graph is derived by calculating a moving average value for each date. The window for the moving average is 7 days, taking into account 3 days before the current date and 3 days after.

This research involved two methods – the IIA data, with 40 peaks (ten peaks within each study period both in the Latvian and the Russian language news), and content analysis of 400 news reports in Latvian and Russian. The study used IIA data on the level of aggressiveness, number of comments, and the keyword proportion under the most aggressively commented news during the COVID-19 pandemic in Latvia. Top 10 most aggressively commented news items in each peak were selected for the study according to IIA data in each of 10 peaks in the two stages of the research: from March 12, 2020 to June 10 (hereinafter stage I) and November 9, 2020 to February 9, 2021 (stage II).

Alongside IIA data, the study used content analysis following a codebook consisting of 10 code groups. News and comments were analysed using 45 categories through which topics of news in various portals and news sources, number of comments and the number of aggressive keywords were identified. Separate categories were created to evaluate COVID-19 related news.

IV. RESULTS

Stage I of the pandemic in Latvia covers the period from March 12, 2020, when the Latvian government announced the first emergency situation due to COVID-19, up to June 10, 2020, when the emergency was lifted, leaving certain restrictions in place. Stage II of the pandemic in Latvia is defined in this study as a three-month period starting November 9, 2020, when the Latvian government declared the second state of emergency due to deteriorating epidemiological situation, until February 8, 2021. Altogether ~ 663.5 thousand comments in Latvian and 5584.9 comments in Russian were analysed.

The IIA data show that the average aggressiveness increased in the second stage of the pandemic (see average aggressiveness value in Fig. 1) in comments in Latvian (from 3.4 to 3.6) while decreased slightly (from 1.9 to 1.8) in comments in Russian. IIA average value over more than a year, from January 2020 to February 2021, was similar: 3.5 in Latvian and 2.0 in Russian.

During stage I of the study (Fig. 1) the total level of aggressiveness in comments of both languages was similar dynamically, with a low level from mid-March to end of April, as well as end of May 2020, and increasing by beginning of May and June.

During stage II of research (Fig. 2) the dynamics of aggressiveness are similar in both language environments. Lower aggressiveness levels were observed by the end of November 2020 and beginning of January 2021, short periods of higher aggressiveness were observed at the end of November, beginning of December, New Year period, second half of January and an upward tendency beginning of February.

Further data analysis reveals the proportion of aggressive keywords in each language and the number of keywords in comments on all news and news on COVID-19 in the specified period.

During stage I of the pandemic, 100 analysed news stories contained 14 381 aggressive comments in Latvian and 9819 comments in Russian (Table 1). In Latvian comments 10 068 keywords were identified, revealing aggressive communication according to IIA, whereas in Russian comments this figure was 5204. On average, each news story received comments, containing 101 aggressive keywords in Latvian and 52 in Russian. During the second state of emergency, a number of aggressive comments has increased. Under the 100 analysed news stories in Latvian, 16 380 comments were found, and in Russian the figure was 13 787. However, a number of aggressive keywords changed only slightly. In the Latvian comments, 10 699 keywords pertaining to IIA aggressive communication were found, and in Russian – 5541. An average of 106 keywords was found under each news story in Latvian and 55 aggressive keywords under each news story in Russian.

Comparing the level of aggressiveness related to COVID-19 pandemic and the aggressiveness expressed on other topics, it was observed that during stage I of the pandemic 3963 aggressive keywords (39% of all keywords) were included in the comments on COVID-19 directly or indirectly related news in Latvian and 1632 (31%) – in Russian (Table 1).
TABLE 1. GENERAL DATA ON AGGRESSIVENESS LEVEL AND NEWS COMMENTS STRUCTURE (STAGES I AND II)

<table>
<thead>
<tr>
<th></th>
<th>Stage I</th>
<th>Stage II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LV</td>
<td>RU</td>
</tr>
<tr>
<td>Average value of aggressiveness</td>
<td>3.4</td>
<td>1.9</td>
</tr>
<tr>
<td>Total number of keywords (100 news pieces)</td>
<td>10 606 5204</td>
<td>10 699 5541</td>
</tr>
<tr>
<td>Total number of keywords (COVID-19 news)</td>
<td>3963 1632</td>
<td>5068 2176</td>
</tr>
<tr>
<td>Share of COVID-19 related keywords (%)</td>
<td>39% 31%</td>
<td>47% 39%</td>
</tr>
<tr>
<td>Average number of keywords (per news item)</td>
<td>101 52</td>
<td>106 55</td>
</tr>
<tr>
<td>Number of aggressive comments (10 days)</td>
<td>14 381</td>
<td>9819 16380</td>
</tr>
</tbody>
</table>

At stage II of the pandemic the proportion of aggressive keywords related to COVID-19 news has increased: 5068 aggressive keywords (47%) were found on COVID-19 news in Latvian, and 2176 in Russian (39%). In both stages of the pandemic, a share of aggressive keywords in comments on COVID-19 news is 8% higher in comments in Latvian than in Russian.

Comparing COVID-19 related news and other news (Table 2) in the selected sample it was found that in Latvian slightly more than half of the news, while in Russian – less than half of the news was related to the COVID-19 pandemic. In stage I of the pandemic 52 news items in Latvian and 47 in Russian were related to the pandemic (accordingly, 11 and 18 of them indirectly). In stage II of the pandemic the same amount of news items in Latvian (52) and a slightly lesser amount of news in Russian (43) were related to the pandemic (accordingly, 12 and 2 of those related indirectly).

TABLE 2. STRUCTURE OF NEWS CONTENT (1ST AND 2ND PERIOD)

<table>
<thead>
<tr>
<th></th>
<th>Stage I</th>
<th>Stage II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LV</td>
<td>RU</td>
</tr>
<tr>
<td>News on COVID-19</td>
<td>41 29</td>
<td>40 41</td>
</tr>
<tr>
<td>News, indirectly related to Covid-19</td>
<td>11 18</td>
<td>12 2</td>
</tr>
<tr>
<td>Other news</td>
<td>48 53</td>
<td>48 57</td>
</tr>
</tbody>
</table>

Indirectly related news include such news where the pandemic is not the main focus of the journalist yet reveal its presence, for example, news on a State Security Service official explaining spy activity curbed by the pandemic or news on anti-racism solidarity protests in the shadow of COVID-19 restrictions on gathering.

Analysing the aggressively commented news in the Latvian language, three main topics emerge: international events, crime, and national politics (Table 3, 4). Other themes, like economic and social processes, war and conflict or culture, appear on the IIA peak less frequently. In stage I of the pandemic, international news in Latvian were predominantly related to the killing by police of George Floyd, a 46-year-old black man, in the United States (US) and the consequent anti-racism protests in the US and worldwide. In aggressively commented international news in Russian, anti-racism protests also emerge as a major topic though it is equally important to such topics as the criminal trial in the Netherlands on the crushed Malaysia Airlines Flight MH17 in the Eastern Ukraine and political persecution of anti-regime activists in Russia. International news in stage II of the pandemic (Table 4) were mostly concerned of former US President Donald Trump’s pre-election events and Joe Biden’s election as the US President. Many commenters’ attention was drawn by news about protests in Belarus regarding the Lukashenko oppressive regime. Other aggressively commented news in this category is news related to events in Russia, especially revolving around the poisoning of Alexey Navalny. News in these two topics was commented aggressively twice as often in Russian as in Latvian – 35 news items in Russian and 18 in Latvian, respectively.

Among national politics news, in stage I of the pandemic the most aggressively commented topics in Latvian concerned the news on constructing the new State Security Service building. In Russian, aggressive comments evolved around news on Latvia’s national day on the 4th of May and celebrations of the Victory Day on the 9th of May. The news on the controversial issue of the end of the Second World War was widely commented by both Latvian and Russian audiences under the topic of international conflicts and war. In stage II of the pandemic, the most aggressive comments on news about national politics were found under materials that regard a decision made by Constitutional Court on November 12 granting the rights to parental leave to a same-sex couple 9 (Table 4).

COVID-19 related news in the first period account for 52% all news in Latvian and 47% of news in Russian; in the second period they amount to the same 52% of all in Latvian and 43% in Russian (Table 3, 4). In stage I of the pandemic, most of the aggressively commented news in Latvian was related to the COVID-19 statistics in Latvia and abroad, and to the government’s stance on the pandemic and combating of it. Slightly less aggressiveness was provoked by news describing COVID-19 restrictions.

<table>
<thead>
<tr>
<th>General news topics</th>
<th>LV</th>
<th>RU</th>
<th>3.4</th>
<th>1.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>National politics</td>
<td>5</td>
<td>12</td>
<td>COVID-19 news</td>
<td>LV</td>
</tr>
<tr>
<td>COVID-19 statistics in Latvia</td>
<td>12</td>
<td>6</td>
<td>COVID-19 news</td>
<td>LV</td>
</tr>
<tr>
<td>COVID-19 statistics in other countries</td>
<td>9</td>
<td>3</td>
<td>COVID-19 news</td>
<td>LV</td>
</tr>
<tr>
<td>Government stance</td>
<td>7</td>
<td>4</td>
<td>COVID-19 news</td>
<td>LV</td>
</tr>
</tbody>
</table>
In Russian, the aggression was expressed towards news about restrictions, though in majority of cases restrictions were related to the topic of the news indirectly (for example, in news about 4th and 9th of May celebrations in Latvia involving a reference about COVID-19 restrictions on gathering). To a lesser extent, aggression evolved around news about COVID-19 statistics in Latvia and topics related to “other” themes such as Russian disinformation on Latvia as a COVID-19 failure story and comments of Bill Gates on his involvement in spread of COVID-19 and chipping.

In stage II of the pandemic, the most aggressively commented news in Latvian was similar to those in the first period: the news where government officials express their position on pandemic-related issues, news explaining restrictions and COVID-19 statistics. In addition, in stage II, vaccination emerges as an issue. News on protests against restrictions and the impact of the pandemic on Latvian economy have not been commented aggressively.

Data on aggressively commented COVID-19 news in Russian show that commenters’ behaviour partly coincides with the Latvian commenters. Most frequently aggressive comments are addressed towards the government representatives’ opinions and stance on the pandemic, and news on COVID-19 statistics in Latvia.

The biggest difference between Latvian and Russian-speaking users was in the attitude towards protests against the pandemic restrictions on December 12, 2020, in Riga. It is arguably related to the fact that one of the protests, held by the Freedom Monument, was organized by Aleksey Roslikov, a popular Russian-speaking politician in Latvia and a member of association ‘Stability – Yes!’. Commenter aggressiveness was also observed by news related to disinformation, for example, on activist Valentin Jermeyev and Marina Kornatovska, physician assistant of Riga East University Hospital, spreading false news on COVID-19 patient treatment, and the news about them both being detained.

V. CONCLUSIONS

The study demonstrates that the fluctuations in the level of aggressiveness in Latvian and Russian language environments are directly linked to actual real-life events reported in the news. COVID-19 news published in both stages represented on average half of the level of aggression of internet commenters, but they have not significantly affected the overall level of aggressiveness of users of news portals. As the average level of aggression recorded by IIA in the rest of the year was similar in both languages, it can be concluded that while the pandemic determined the attitudes and difficulties expressed by internet commenters to a certain extent, the changes which the public faced had not raised the overall level of aggressiveness. Only during stage II of the study, the real-life events in Latvia and the news about them start showing signs of a slight increase in aggression in the Latvian comment environment, while the level of aggression has decreased slightly in the Russian comment environment. Overall, the highest level of aggression was provoked by news about events other than the pandemic – international politics, Latvian national politics and crime.

Analysing the number of aggressive comments, as well as the number of aggressive keywords in the comments of both language portals in both stages, the study concludes that the number of news comments is similar (with bigger difference between Latvian and Russian comments during stage I of the pandemic), but the number of aggressive keywords in Latvian and Russian comments differ.
Namely, Russian-published news commenters use fewer aggressive keywords. There are also fewer aggressive keywords in Russian compared to Latvian in the comments written on COVID-19 news. This implies that Latvian news commenters use more aggressive words in their comments, and, compared with Russian news commenters, have sharper attitudes towards COVID-19 news and have expressed more dissatisfaction and concern. In general, however, the level of aggressiveness of COVID-19 news comments has been lower than in other topics’ news comments in both languages. In stage I of the pandemic, 39% of aggressive comments are expressed in 52% of COVID-19 related news in Latvian and 31% of comments go to 47% of COVID-19 news. In stage II of the pandemic, 47% of aggressive keywords are dedicated to COVID-19 in 52% of COVID-19 related news, whereas in Russian comments 39% of aggressive keywords are directed towards 43% of COVID-19 news.

Similar news themes have attracted the attention and aggressive comments in both Latvian and Russian-language portals. In each language, however, there were different actual topics that have caused higher aggressiveness. Although there is a linguistically divided language portals. In each language, however, there were aggressive com ments in both Latvian and Russian -keywords are directed towards 43% of COVID-19 news. whereas in Russian comments 39% of aggressive keywords are dedicated to COVID -19 in 52% of COVID -19 related news in Latvian and 31% of comments go to 47% of COVID -19 news. This implies that Latvian news commenters use more aggressive words in their comments, and, compared with Russian news commenters, have sharper attitudes towards COVID-19 news and have expressed more dissatisfaction and concern. In general, however, the level of aggressiveness of COVID-19 news comments has been lower than in other topics’ news comments in both languages. In stage I of the pandemic, 39% of aggressive comments are expressed in 52% of COVID-19 related news in Latvian and 31% of comments go to 47% of COVID-19 news. In stage II of the pandemic, 47% of aggressive keywords are dedicated to COVID-19 in 52% of COVID-19 related news, whereas in Russian comments 39% of aggressive keywords are directed towards 43% of COVID-19 news.

ACKNOWLEDGEMENTS

This study was supported by the Ministry of Education and Science, Republic of Latvia, as part of the project “Life with COVID-19: Evaluation of overcoming the coronavirus crisis in Latvia and recommendations for societal resilience in the future” [grant number VPP-COVID-2020/1-0013].

REFERENCES

