UNIVERSITY OF TARTU
Institute of Computer Science
Data Science Curriculum

Lisanna Lehes

Understanding Public and Leaders' Opinion about Russo-Ukrainian War through Social Media Platforms: An Estonian Case Study

Master's Thesis (15 ECTS)

Supervisor(s): Somnath Banerjee PhD,
Rajesh Sharma PhD

Tartu 2023

Understanding Public and Leaders’ Opinion about Russo-Ukrainian War through Social Media Platforms: An Estonian Case Study

Abstract:

With its fast-paced communication, social media is a valuable data source for studying how people express their opinions and views. Focusing on the Estonian context, this study aimed to analyze the Public and Leaders’ (politicians) opinions regarding the Russo-Ukrainian war through social media platforms.

Two datasets were collected: Facebook posts for Leaders and Twitter tweets for the Public. To study the content shared on social media platforms, topic modeling and sentiment analysis techniques were used to gain insights into the sentiments, key topics, and discourse patterns surrounding the discussions on war.

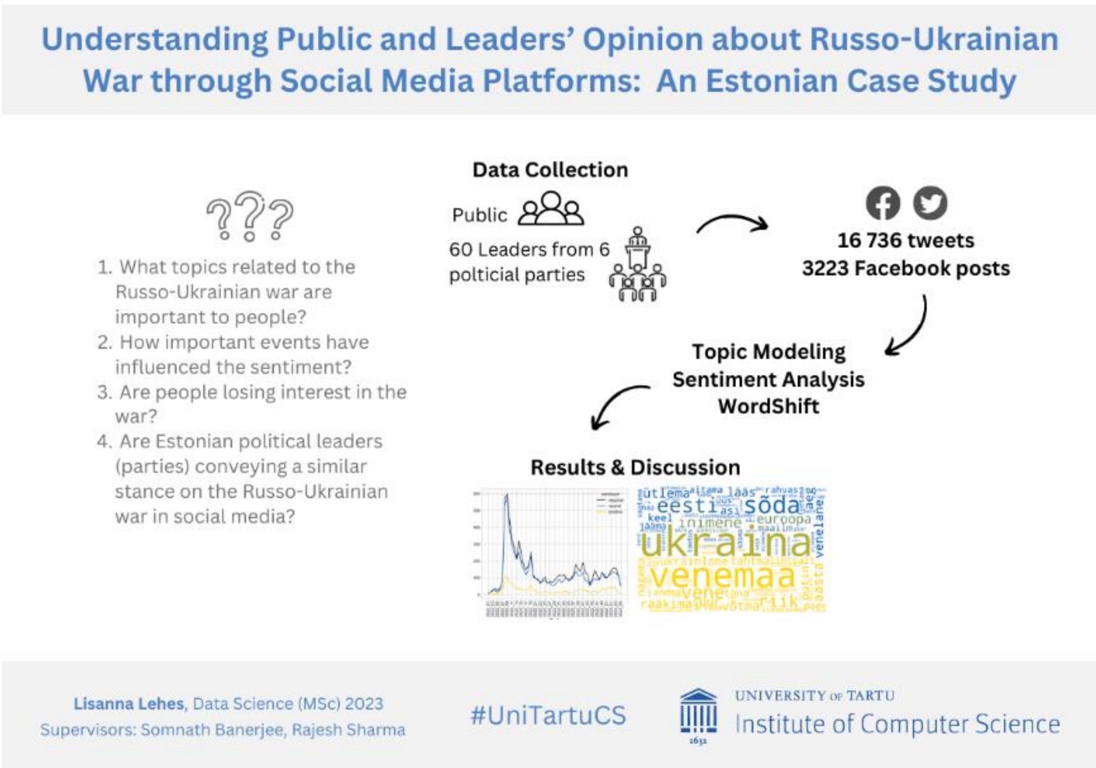
Suggesting a shared interest between the Public and Leaders, the results revealed overlapping topics of interest, including war, energy security, economy, sports, education, and news. However, distinct differences emerged, where Leaders mainly focused on discussing general themes such as domestic and foreign politics, while the public engaged in a more diverse range of topics. Public sentiment leaned towards negative, while the Leaders’ leaned towards neutral with variance among parties.

We conclude that analyzing social media data allowed us to focus on two different perspectives, political Leaders and the Public. The analyses provided valuable insights into the critical issues discussed by both groups and the impact of war and particular events on their posting activity.

Keywords:

Social Media, Sentiment Analysis, Topic Modeling, Text Analysis, Ukraine, Russia, War, Estonia

CERCS: P170 Computer science, numerical analysis, systems, control



Vene-Ukraina sõja hoiakute analüüs Eesti avalikkuse ja poliitikute sotsiaalmeedia põhjal

Lühikokkuvõte:

Erinevatel sotsiaalmeedia platvormidel luuakse igapäevaselt suurel hulgal erilaadseid andmeid, mis on väärtuslikuks allikaks inimeste seisukohtade ja meelestatuse uurimisel. Magistritöö eesmärk oli analüüsida nii avalikkuse kui ka poliitiliste liidrite hoiakuid Vene-Ukraina sõja osas ning seejuures oli uuritav kontekst esitatud läbi eestikeelse sotsiaalmeedia.

Magistritöö läbiviimiseks loodi kaks andmestikku, kus kasutati Facebooki postitusi liidrite, ja Twitteri postitusi avalikkuse analüüsimiseks. Nimetatud sotsiaalmeedia platvormidel jagatud sisu uurimiseks kasutati teemade modelleerimist ja meelestatuse analüüsi. Leitud sisu andis ülevaate sõjateemaliste arutelude üldisest diskursusest, seal esinevatest teemadest ja meelestatusest.

Tulemused näitavad, et nii avalikkust kui ka liidreid kõnetavad teemad olid sõda, energia julgeolek, majandus, sport, haridus ja sõja esitamine uudistes. Erinevusena tuli välja, et poliitiliste liidrite postituste teemad keskendusid rohkem sise- ja välispoliitikale, samas kui avalikkus käsitles mitmekesisemaid teemasid. Avalikkuse postituste meelestatus oli Vene-Ukraina sõja teemal pigem negatiivne, liidrid olid aga oma postitustes neutraalsemad. Märgata oli ka teatavat meelestatuse erinevust erakondade lõikes.

Valitud sotsiaalmeedia platvormide analüüs võimaldas keskenduda kahele erinevale vaatenurgale – avalikkuse ja poliitiliste liidrite vaatele. Analüüs andis väärtusliku ülevaate mõlema rühma jaoks olulistest teemadest, ning kuidas sõda ja sellega kaasnevad sündmused mõjutasid postitamise aktiivsust.

Võtmesõnad:

Sotsiaalmeedia, Meelestatuse analüüs, Teemade modelleerimine, Tekstianalüüs, Ukraina, Venemaa, Sõda, Eesti

CERCS: P170 - arvutiteadus, arvuline analüüs, süsteemid, kontroll

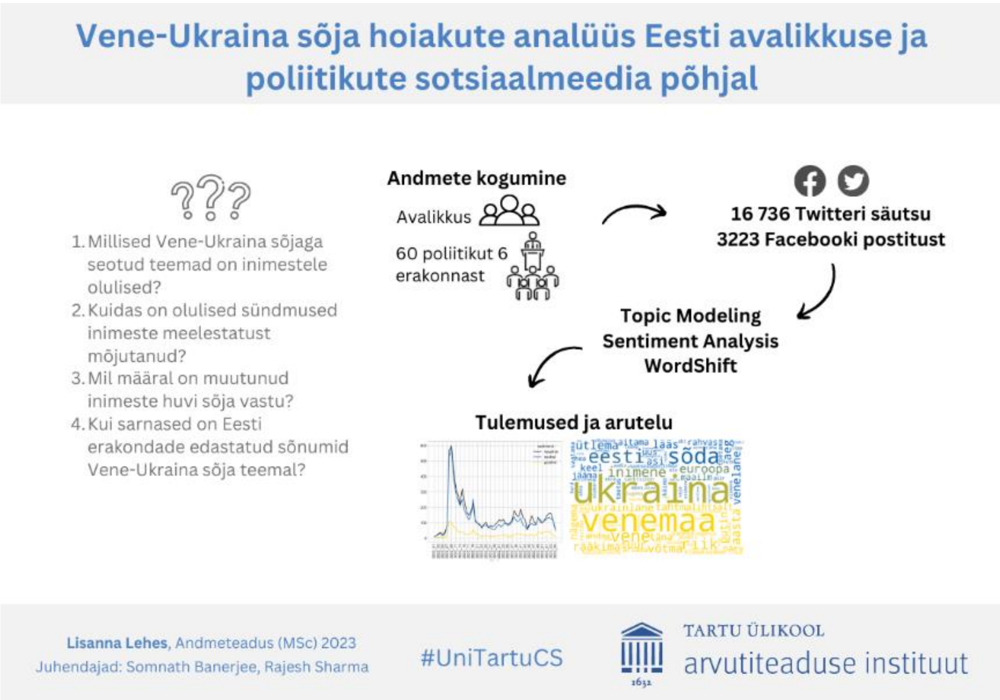


Table of Contents

1 Introduction

Social media platforms have been widely used worldwide to describe feelings and express opinions on certain events. It is a popular communication tool for both – the Public to interact and express their views and the Leaders, who can reach out to their audience and followers quickly. Depending on the demography, many social media platforms exist, like Twitter, Facebook, Instagram, Reddit, Telegram, and many others. These social media platforms allow us to learn about ongoing issues, conflicts, and wars, like the current war in Ukraine. Each platform has unique features and functions shaping the discourse.

In previous studies, social media platforms have been utilized in many ways to study certain events or societal issues. For instance, to understand the refugee crisis [1,2] or the pandemic [3,4]. In recent work, NLP techniques like topic modeling and sentiment analysis have been used to study the Russo-Ukrainian war [5–7]. Although previous studies employ social media analysis on the Russian invasion, this thesis explores the Estonian perception of the ongoing Russo-Ukrainian war in two groups, specifically the Leaders and the Public. A more novel approach of using two popular social media platforms is used. In the context of this thesis, we consider politicians as Leaders. The subsequent research questions are:

1. What topics related to the Russo-Ukrainian war are important to people?
2. How important events have influenced the sentiment?
3. Are people losing interest in the war?
4. Are Estonian political leaders (parties) conveying a similar stance on the Russo-Ukrainian war in social media?

This thesis studies Leaders' and the Public's opinions about the Russo-Ukrainian War. We focus on local demography by studying the Estonian political leaders across six political parties and the Public. For studying Leaders, we relied on Facebook, a more popular platform among politicians. We collected 3223 Facebook posts related to the war across six political parties for the Leaders. For the Public, we relied on the data from Twitter, and 16 736 tweets related to the war were collected. The data for both was collected for an approximately one-year timeframe.

The thesis is divided into multiple sections. Firstly, a paragraph discussing the related work. Paragraph 3 covers the data collection and data processing principles. In paragraph 4, the employed methodology is described, specifically how the topics were extracted, and an overview of the utilized sentiment analysis tools is given. Section 5 introduces the results, beginning with Leaders' topics and sentiments, followed by the Public's topics and sentiments, and finally, both are compared. In section 6, the results are discussed and compared with the previous research, and the research questions are answered. The section also covers the limitations and potential concepts for future work. Lastly, the overall conclusion of the work is presented.

2 Related Work

This section focuses on using social media data for research. In connection with the current thesis topic, which aims to understand the war in Ukraine through Estonian social media data, existing research on the topic is introduced. The related work provides an understanding of the current state of knowledge and identify potential areas for further research.

2.1 Social Media Analytics

Social media allows users to express their opinions and interact with each other at exceptional speeds. Therefore, social media is integral to daily communication between people and for public figures to convey their opinions and views on various topics. Additionally, social media is an important data source for researchers to analyze the discourse on various issues. Thus, social media analytics refers to gathering data from social media platforms and analyzing it to understand specific topics or problems [8], like the war in Ukraine. The most common methods for understanding the content of social media are, for example, sentiment analysis and topic modeling.

Due to the amount of data generated in social media, it is impossible to read all of the posts or tweets, so sentiment analysis makes it easy by providing a polarity assessment on these kinds of texts. Sentiment Analysis or opinion mining is a Natural Language Processing (NLP) task to extract text sentiments and impressions regarding various topics [9,10]. A sentiment analysis task can be considered a text classification problem [11] since the texts can be classified as positive, negative, or neutral. The method can be applied in many domains. Although, in the context of this thesis, sentiment analysis is used to study the sentiment of (political) leaders and the general public towards the Russian invasion of Ukraine and how these sentiments have changed over time.

In this work, the analysis is done with Estonian social media texts. For Estonian language sentiment analysis, there is currently one dataset [12] that consists of annotated news extracts. Regarding analyzing Estonian texts and their sentiments, [13] has used a weakly-supervised text classification paradigm to classify the sentiment of Estonian texts. [12] proposed a lexical approach as well as trained machine learning models. [14] examined Estonian sentiment analysis on an entity level by using corporate entity mention detection and sentiment analysis in the Estonian economic news domain. [15] created a language resource enabling sentiment analysis using Estonian Wordnet by joining the valence scores from [12] with the information from Estonian Wordnet.

Topic modeling identifies patterns in the text data by helping to identify what events or topics a document is discussing [16]. Topics are represented by a set of keywords that describe the topic. In analyzing social media data, topic modeling can be used to determine important topics and themes in user-generated content, such as tweets or Facebook posts [16].

2.2 Social Media Analytics in the Context of Conflicts

Utilizing social media can provide valuable insights into the attitudes and opinions of people involved and into the political contexts in which these conflicts occur. For example, Twitter data has previously been used to analyze the sentiment regarding the Syrian refugee crisis [1]. In connection with this, media discourse has been analyzed to understand the sentiment on migration in Sweden [2]. Sentiment analysis techniques have been used to understand the perception of ISIS in different countries [17]. Social media content has been examined concerning natural disasters to identify perceptions and behaviors among affected populations [18–20]. Sentiment analysis and topic modeling has been used to study COVID-19 and vaccines [3,4]. As previous examples demonstrate, social media is a powerful tool for understanding various conflicts and events.

Some initial analyses have already been performed and published in the context of the war in Ukraine. For example, [5] analyzed public sentiment toward economic sanctions employed against Russia. Looking at almost one million Facebook posts from 108 countries, an index was constructed to describe whether and how people's attitudes align with their government votes in the United Nations. According to the results, the most popular themes were military operations, political leaders and organizations, economic sanctions, humanitarian concerns, energy and oil prices, and the roles of media in the war. Moreover, the trading relationship with Russia played an important role in shaping the people's sentiments about the war. Whereas, [6] analyzed Ukrainian Telegram users and obtained the following key topics: attacks, particular cities and events, sanctions, children, and refugees. Additionally, a negative structural sentiment over time was found that corresponds well to the overall negative event, which is war. [21] analyzed several news sources from Europe, the US, Ukraine, and Russia, from where topics like gas and oil, army casualties, civilian deaths, protests, and more were obtained. [7] focused on Twitter's English language data sentiment and emotion analyses of the Russian invasion of Ukraine. Again, there were more negative sentiments, with sadness being the most prominent emotion expressed. Furthermore, *war* was the most occurring word, and the results indicated that the public sentiment leaned against the aggressor country.

In contrast, more than 6 million public posts on VK (Russia's largest social media platform) have been analyzed. The results revealed that the invasion was mainly discussed in connection with the losses in the Russian army, and references to World War II were common. Users who posted about the war were older than average VK users and followed pro-regime and pro-war content more frequently [22].

Finally, multiple datasets [23–25] for analyzing the war have been created on the topic by using specific keywords to filter the data from Twitter.

In the context of conflicts and the ongoing Russo-Ukrainian war, both sides use social media in the information war. For example, [26] tried to detect Russian propaganda using SVM and BERT, and found that the propaganda style seemingly mimics international codes of conduct for journalism, adapts to each target language, and is country-specific. Another aspect is memes in the sense of hate speech, [27] found that memes serve as one of the formats through which Russophone Telegram channels propagate hate speech. Additionally, the coverage of the war can vary greatly depending on the region or a news outlet. For example, it has been found that while the Western press outlets have focused on the military and humanitarian aspects of the war, Russian media have focused on the supposed justifications for the "special military operation," and Chinese news has concentrated on the conflict's diplomatic and economic consequences [28].

2.3 Leaders and the Public

Understanding the Public's and Leaders' perceptions can provide valuable information about many issues. Regarding Leaders, these are the people who have a strong influence, as they have large followings and are considered to be more experienced and able to shape the minds and thoughts of other people in society [29,30]. For example, [31] focused on identifying the opinion leaders in the education, economics, and politics domain, then analyzed the sentiment and topics appearing in their posts. They found that opinion leaders significantly influence their followers and can change their sentiments through social media posts.

In this thesis, we consider Estonian politicians as Leaders. During times of war, politicians represent the country by meeting with allies and drafting foreign policies. The outcome of their work and stances taken are shared on social media. To illustrate, the research on newspapers in Spain and Italy showed that the key players in the COVID-19 pandemic were politicians [32]. For example, concerning COVID-19, studies using data from Twitter have been done to analyze Leaders and their reactions to the pandemic [33,34]. Considering political leaders is crucial in gaining insights into issues like war and other national or international conflicts.

The Public often expresses concern and discusses about various issues via social media platforms. According to [35], the concept of public opinion, through which individuals can be heard by political elites and each other, is vital to democracy. Public opinion mining through social media is extensively researched, and some examples have already been covered in paragraph 2.2.

This section presented a review relevant to the use of social media in research. Previous research on the Russian invasion of Ukraine has primarily focused on English, Ukrainian, or Russian language data. Although the availability of data in the mentioned languages is much higher than in Estonian, it is essential to gain insights into the Estonian perception of the war, considering Estonia’s vital role as an ally to Ukraine. Therefore, this thesis complements the previous research by applying social media analytics techniques to Estonian language data and adds the comparison between two groups - the Public and the Leaders.

3 Data

This section describes the data acquisition process from the two popular social media platforms, Facebook[1] and Twitter[2].

Twitter is a social media platform that enables users to voice their opinions and emotions through short posts called tweets. Tweets are limited to 280 characters or less and can include text, images, and links. Users can also retweet or like other users' tweets and create and follow hashtags on specific topics or events. It is a popular platform for people, politicians, or public figures to connect with their followers.

Facebook is a social media platform with one of the largest user bases that allow users to connect with people, share photos and videos, and join groups and communities. In addition to individual profiles, Facebook also includes public pages for public figures, which allow them to share updates and engage with their audience. Another aspect of Facebook is that it allows for writing longer posts.

The need to use two social media platforms arises because Twitter can be used to collect data about the Public. At the same time, most politicians prefer to use Facebook. Using only Twitter would result in an unequal representation of parties (Leaders), as most politicians do not use Twitter. Therefore, Facebook is the chosen platform for analyzing leaders. On the other hand, we selected Twitter for collecting the public data for analysis.

3.1 Data Collection

The war is an ongoing event, and due to the limitations in the size of available Estonian data, as much as possible was collected. Even though the war started at the end of February 2022, we collected data before that to measure the war's impact on social media activity. Therefore, at least a year's worth of data was collected from both social media platforms. Public's data was collected from January 2022 until January 2023, and the Leaders' data was collected from January 2022 until March 2023. The differences in Public and Leaders' timelines are due to Twitter closing its free API for research. For Twitter, an academic research API account was used, which no longer exists since 09.02.2023[3].

Leaders dataset

The Leaders dataset will be a curated list of politicians from each party based on their followers on their Facebook public page. Leaders are from different parties that are currently relevant in Estonia. These parties are the Estonian Reform Party (Eesti Reformierakond, REF), Estonian Centre Party (Eesti Keskerakond, KESK), Estonia 200 (Eesti 200, E200), Social Democratic Party (Sotsiaaldemokraatlik Erakond, SDE), Conservative People's Party of Estonia (Eesti Konservatiivne Rahvaerakond, EKRE), Isamaa. The Estonian Greens (Eestimaa Rohelised) and The Right (Parempoolsed) were not collected due to insufficient amount of public pages and posts to consider in the current analysis.

The Riigikogu (parliament) elections 2023 candidates list[4] was used to find leaders. From there, we checked the Leaders' Facebook pages and noted the number of followers. Finally, ten people from each party were chosen based on the number of followers they had. The follower count was chosen as the indicator because it directly reflects the popularity and influence of a

[1] https://www.facebook.com/
[2] https://twitter.com/
[3] https://twitter.com/TwitterDev/status/1621026986784337922
[4] https://rk2023.valimised.ee/en/candidates

politician. Furthermore, it was readily available for comparison. The precondition was having a public page, not a personal page, that is visible to everyone without an account.

Facebook data was collected manually without the help of crawlers to avoid violating Facebook's terms of service. Additionally, crawlers can raise concerns about privacy or data protection, which can be avoided by manually and responsibly collecting the data. Although Facebook API exists, it has become very limited in its function and currently can only be used to get information about self-owned pages.

A politician's page was opened in the Mozilla Firefox browser to get the data. The network tab under Developer Tools was used to monitor the network traffic. In addition, the network requests were filtered to only show GraphQL queries (a popular query language developed by Facebook). After that, manual scrolling was done until the beginning of 2022 was reached, and then the HAR file with all the data was downloaded.

Public dataset

The data for the Public was collected from Twitter using Twitter API for Academic Research[5]. Twitter API allows us to search tweets by multiple query parameters. Public opinion is found using relevant keywords to find tweets and posts related to the Ukrainian war.

3.2 Legal basis

As the results of the thesis required data processing from social media, it is necessary to assess the legitimate interest in collecting and processing such data. For this purpose, a Specialist of Data Protection at the University of Tartu and the Estonian Data Protection Inspectorate were consulted.

The research was carried out with depersonalized data. Therefore, this data is not considered personal data within the meaning of the legislation and does not need to be processed per the rules on protecting personal data.

For analysis, social media posts were gathered from various personalized accounts. This means that before data was depersonalized, personal data was processed, and legitimate interest needed to be assessed.

In the context of this thesis, a private research, the legal basis for processing data is a legitimate interest. To assess if legitimate interest applies, a three-step assessment has been carried out following the instructions provided by Data Protection Inspectorate[6]:

1. The data processing is carried out for the purposes of a research project concerning the analysis of social media content, i.e., data collection is necessary to achieve the objective.
2. The data is collected from publicly available sources. The results are presented in an aggregated way, i.e., the individual cannot be identified. Once the data has been processed, any personalized data will be deleted.
3. The collected data refers to those posts which have been publicly made available by the owner of those posts and are necessary for social media research. Although the data has been made public in the past, no individual's posts will be republished. Only the data needed to complete the research is collected, which will be deleted once the work is completed. There is no expected impact on the data subjects as the results are not presented at the individual level.

[5] https://developer.twitter.com/en/products/twitter-api/academic-research

[6] https://www.aki.ee/sites/default/files/dokumendid/oigustatud_huvi_juhend_aki_26.05.2020.pdf

The work complies with the general obligations laid down in Article 5 of GDPR[7], including the principles of lawfulness, fairness, transparency, purpose limitation, accuracy, data minimization, storage limitation, integrity and confidentiality, and accountability.

3.3 Data Processing

Leaders

The HAR files from Facebook were read in as JSON files and then parsed in a Jupyter notebook using the *Haralayzer package*[8]. Then, entries were extracted and again converted into JSON for easier parsing. The data structure varied, and many edge cases had to be considered when parsing the data. For example, on Facebook it is possible to do multiple types of posts – posts with video, photos, gifs, 3D pictures, events, shared posts, timeline posts, and many more. Furthermore, the data is structured so that it is easy to generate the web page but not to extract analyzable data. That meant we had to identify long key chains and create individual logic to extract the text content and timestamps from the JSON. The end result was a Pandas dataframe containing the politician's name, the party they belong to, the date of the post, and the post's text content.

Facebook posts were collected from ten people from each party between January 2022 and March 2023. The official party Facebook page was always included in these ten accounts. The limitation of 10 politicians per party was set because of two reasons. The main reason is time constraints. Collecting data manually takes a significant amount of time. Scrolling one profile can take up to an hour, depending on the posting frequency. Also, since we are taking the top 10 most followed person pages, then the following pages that come after have a smaller following and also usually post less frequently. In addition, many politicians use their personal accounts instead of politician public pages to voice their opinions, restricting data collection.

Before any preprocessing or filtering, the dataset had 14 829 posts. A distribution of posts for each party is depicted in Figure 1. Notably, for most parties, the distribution of posts at starting point is quite balanced.

Filtering based on keywords was performed to get the posts related to Ukraine and the war. Figure 2 shows the distribution of posts after preprocessing and filtering based on specific keywords. The keywords were selected manually, so to get posts related to the war, e.g., "ukraina" (*Ukraine*), "ukrainlane" (*Ukrainian*), "põgenik" (*refugee*), "sõjapõgenik" (*war refugee*), "konflikt" (*conflict*), "nato" (NATO, The North Atlantic Treaty Organization), "lääneriigid" (*the West*), "agressioon" (*aggression*), "venemaa" (*Russia*), "venelane" (*Russian*), "Putin," "erioperatsioon" (*special operation*), "sanktsioon" (*sanction*), "sõjakuritegu" (*war crime*).

Although initially, the post distribution was balanced, in Figure 2, it can be observed that the distribution changes considerably after filtering to only Ukraine and war-related posts. This is expected as some parties voiced more opinions about the war and others less. Potential limitations wise, in the case of showing overall results from the leaders, the data would lean more towards the frequent posters. However, since the goal is to analyze Leaders' Facebook posts, it is evident that there is already an inherent imbalance between them. Balancing this (e.g., using oversampling techniques such as text augmentation) would mean skewing the reality. Additionally, since we are comparing parties' overall sentiment and topics, it does not matter how many posts they have.

[7] https://eur-lex.europa.eu/eli/reg/2016/679/oj

[8] https://haralyzer.readthedocs.io/en/latest/

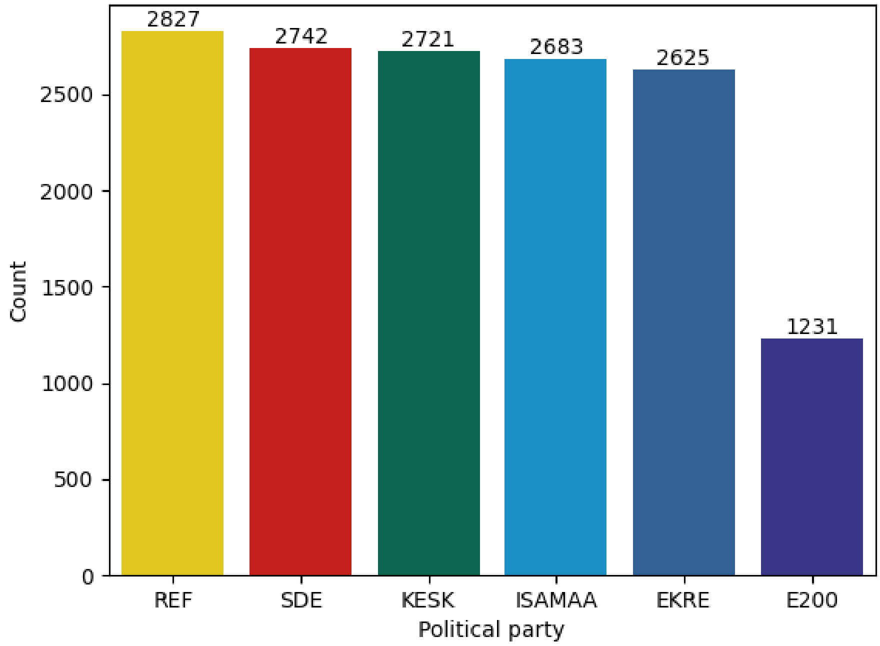


Figure 1. Facebook posts per party before preprocessing.

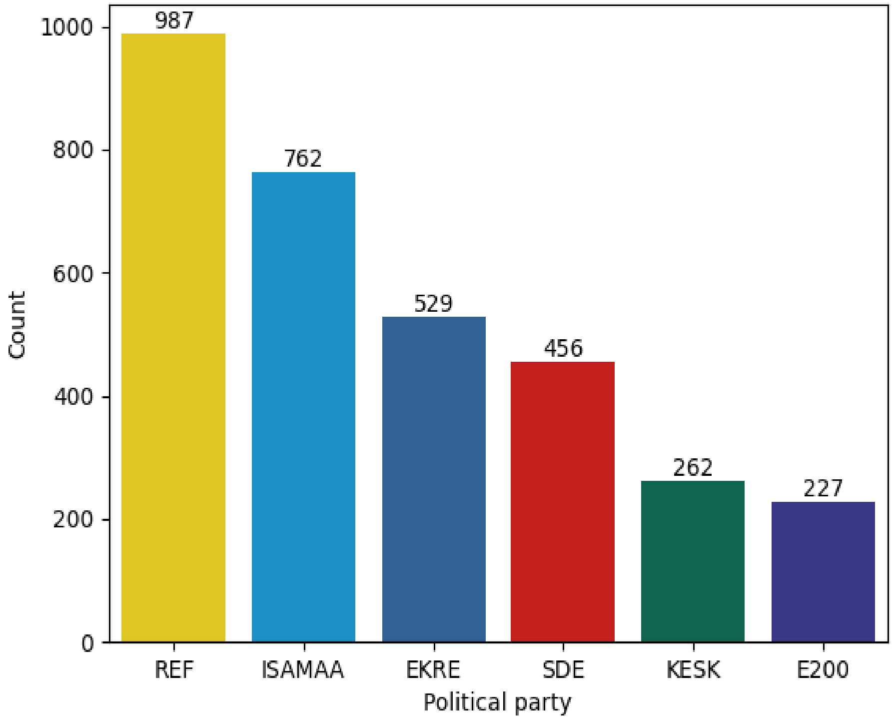


Figure 2. Facebook posts per party after preprocessing and filtering.

Although BERTopic (methods are discussed in section 4) does not need preprocessing, some noise in the data was removed. Preprocessing was done by lower-casing the posts and removing all emojis, symbols, URLs, over spaces, and text written in Russian, Ukrainian, or English to keep only Estonian posts.

Thus, after preprocessing and filtering, the sample of Leaders is composed of 3223 posts made on the politicians' public Facebook pages. It means that about ~22% of posts by the top 10 followed politicians from each party were related to the Ukrainian war. Although it is fair to note that many of the posts from the initial 14 829 are related to the elections held in March of 2023, so the posting activity was much higher than usual.

The distribution of how each of the ten accounts contributed to the preprocessed and filtered final result can be seen in Figure 3. In the figure, only the parties' official accounts are marked, so not to mention any politician's names. For example, for E200, it is evident that the most frequent poster is the official party account and that the top three posters make up over 90% of the posts. This fact again adds to the notion that the parties and their politicians' social media behavior varies greatly.

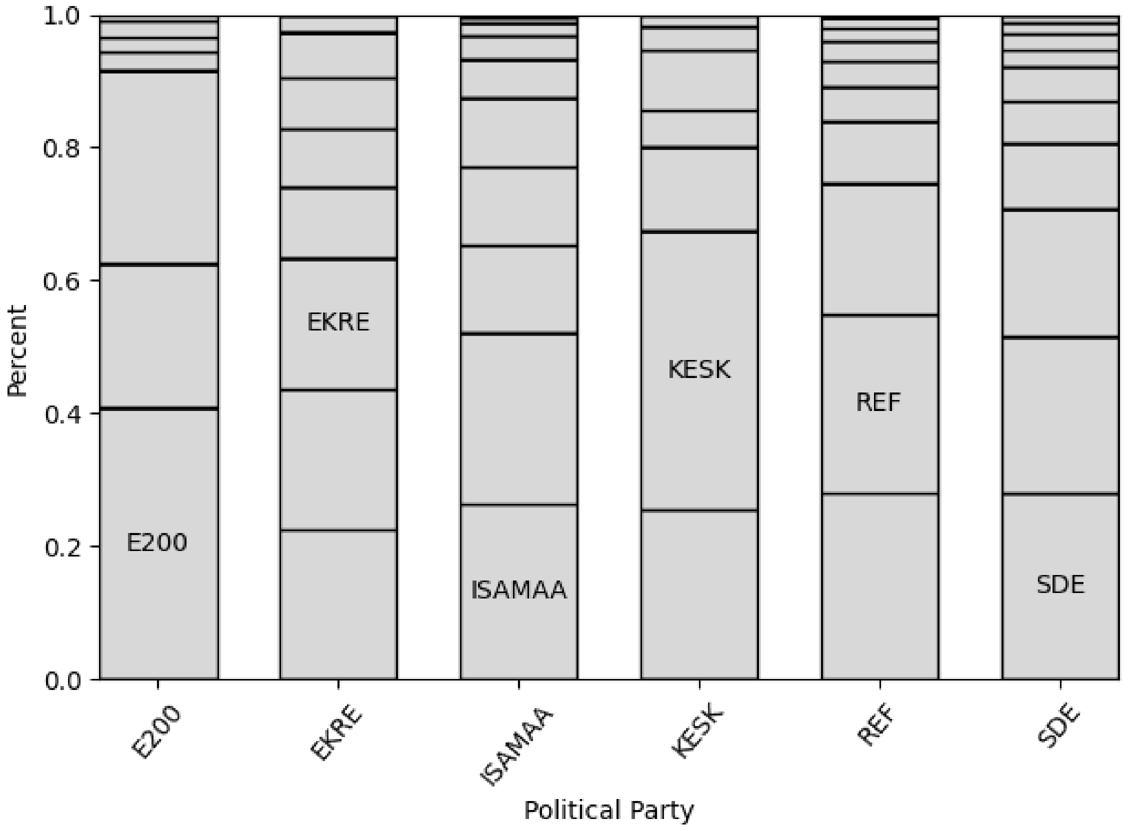


Figure 3. Distribution of leaders' posts per party

The frequency of leaders' posts on Ukraine and the war can be seen in Figure 4. As expected, most posts were made at the beginning of the war, i.e., in March of 2022. The posting activity on the topic decreases significantly during the summer months and picks up again in the autumn of 2022. Another increase in posting activity is in February 2023, which is probably related to the fact that it has been one year since the war in Ukraine started, as well as the incoming elections in Estonia[9]. A drop can be seen in March 2023, which could be caused by the fact that the elections had just ended, and posting activity fell.

[9] Riigikogu elections were held on the 5th March of 2023

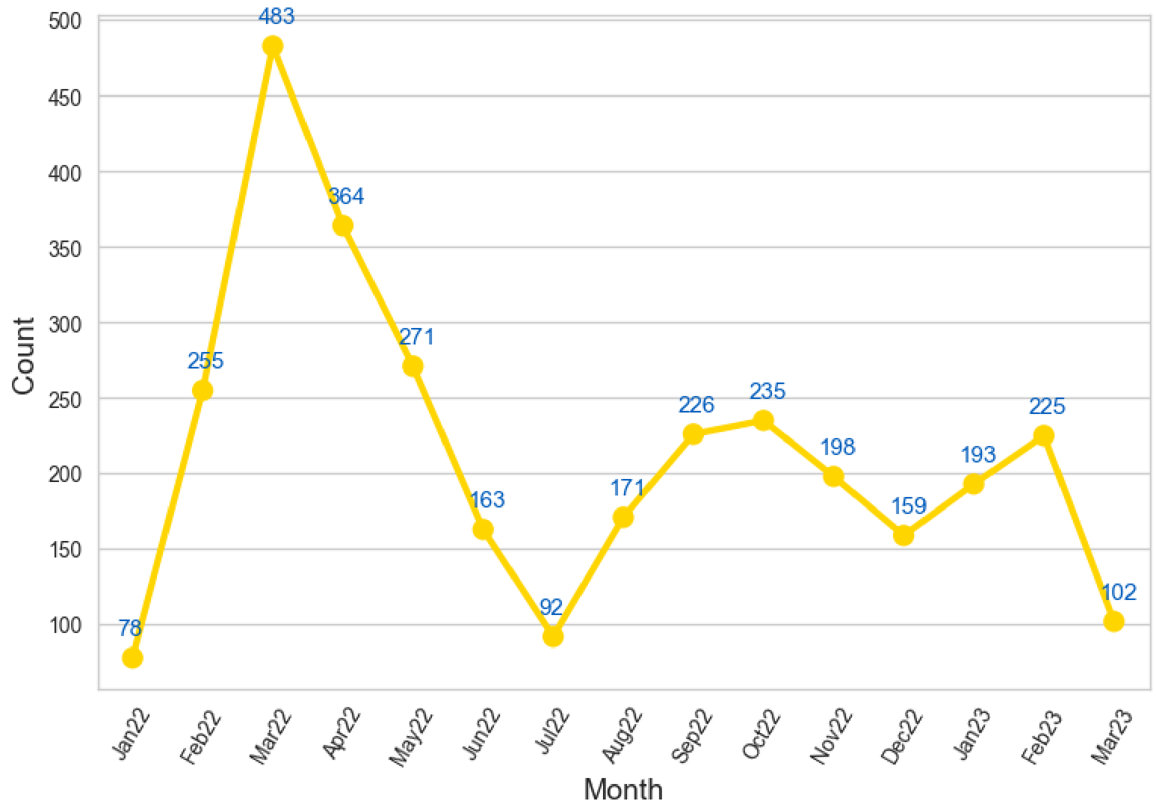


Figure 4. Frequency of Leaders' Facebook posts related to the War

Additionally, a list of politicians' names was created to later remove the Leaders' tweets from the Public's dataframe. For that, the names of politicians from different levels were collected. Political leaders include past presidents, European Parliament members from the last two periods[10][11], the last five governments[12], the last two Parliaments (Riigikogu)[13][14], and heads of municipalities (mayors/heads of municipalities)[15].

Public

For public data, we defined the time interval where we wanted to get the data and used different keywords and combinations of keywords. These keywords can be found in Appendix I. We collected the data for each tweet for later preprocessing – tweet text, tweet id, timestamp, author id, tweet language, and if it is a retweet.

Firstly, a script was written to get the Leaders' Twitter usernames based on their names to filter leaders', i.e., politicians' tweets out from the Public's dataset. Twitter has an API where it is possible to search users based on their names. Of course, there is a limitation: if the politician has not added their name and posts under a pseudonym, we could not filter them out. After fetching the usernames, manual checking was performed to delete the cases where the name and username did not match. From the initial 72 465 tweets, after leaving out politicians' tweets, the dataset consists of 48 805 tweets. For the Public, even though when fetching the

[10] https://www.europarl.europa.eu/estonia/et/eesti-saadikud/2014-2019.html

[11] https://www.europarl.europa.eu/estonia/et/eesti-saadikud/eesti-saadikud-2019-2024.html

[12] https://valitsus.ee/peaminister-ministrid/varasemad-valitsused#vabariigi-valitsuse-

[13] https://www.riigikogu.ee/riigikogu/koosseis/riigikogu-liikmed/

[14] https://www.riigikogu.ee/tutvustus-ja-ajalugu/riigikogu-ajalugu/

[15] https://omavalitsus.fin.ee/omavalitsus/

tweets they were filtered according to language and specific keywords, the tweets collected through the API were still quite noisy, and many tweets were discarded from the final dataset. For example, the tweet language detection was not always perfect, and some keywords were treated ambiguously, e.g., *soda* vs. sõda (*war* in Estonian). The preprocessing included lowercasing and removing all tweets that were not in the Estonian language (e.g., tweets in Finnish, English, Thai, Albanian, and Russian) and removing URLs, emojis, mentions, and retweeted tweets.

The cleaned dataset consists of 16 736 tweets in Estonian that were collected from January 2022 until January 2023. Figure 5 shows how tweets distribute in our timeline.

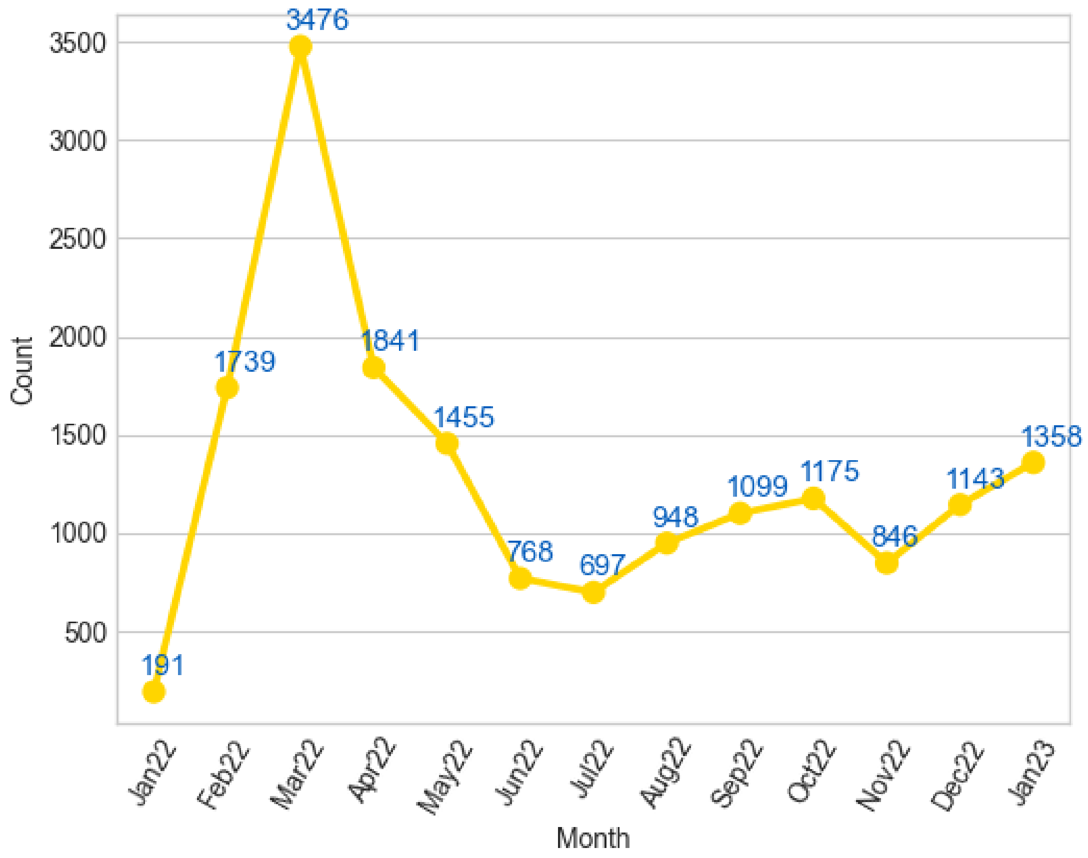


Figure 5. Frequency of Public's tweets related to the war

4 Methods

Topic modeling and sentiment analysis techniques were employed to answer the research questions. One of the goals is to compare what topics related to the war are important to different parties. Similarly, comparisons between Public and Leaders posts are made to see if the important topics overlap and if there is a sentiment difference between the two.

4.1 Topic Modelling

BERTopic [36] was employed as a topic modeling approach as it is a state-of-the-art model that performs well on informal short texts like social media posts [37]. BERTopic is a topic modeling technique that leverages BERT embeddings and a class-based TF-IDF to create dense clusters allowing for easily interpretable topics while keeping essential words in the topic descriptions [38].

The BERTopic modeling process starts with transforming input documents into numerical representations using embedding models [35]. Here, we used a Sentence Transformer from a list of selected models[16]; specifically, *paraphrase-multilingual-MiniLM-L12-v2* is used since it supports over 100 languages, including Estonian. Wikipedia corpus for each language (excluding user and talk pages) was taken as the training data for each language[17]. Although through Flair, Hugging Face Transformers and an Estonian language-based EstBERT could also be used [35], then in initial comparison, the multilingual model created more coherent and interesting topics.

The topic model works modularly, meaning the sub-models can be adjusted accordingly:

- **Vectorizer model**. In our case, scikit-learn's CountVectorizer was used to remove the Estonian stopwords [39] while tokenizing the topics. The CountVectorizer processes the documents after they are clustered, which means that it can be used to optimize the topic representations and avoid stop words appearing in the topic representations, as they do not provide any value to the topic[18].
- **c-TF-IDF model**. In BERTopic, ordinary TF-IDF is adjusted to work on a topic level instead of a document to represent the topics from bag-of-words matrix accurately. c-TF-IDF extracts the topic words and considers what makes the documents in one cluster different from those in another [40]. In our case, the c-TF-IDF model also deals with reducing frequent words. These are words that tend to appear quite often but are not considered as stop words, e.g., in the context of this work, these words would be 'Estonia,' 'Ukraine,' 'Russia,' etc. because either of these words appears in almost all of the texts.
- **UMAP model**. UMAP is used as a default in BERTopic to reduce dimensionality. E.g., with UMAP and its *n_neighbours* hyperparameter, the appearance of micro clusters can be prevented.
- **Clustering model**. By default, BERTopic uses HDBSCAN to extract topics. While it is very capable, it also generates an outlier class (-1 topic) that can be difficult to interpret. Additionally, we did not want to lose any data by filtering out the texts with the outlier class, so the clustering method was changed to K-means because it does not generate outliers.

While BERTopic itself reduces the number of topics so that similar topics are clustered together, the ability to manually reduce topics by merging them was also used. The number of

---

[16] https://www.sbert.net/docs/pretrained_models.html
[17] https://github.com/google-research/bert/blob/master/multilingual.md#list-of-languages
[18] https://github.com/MaartenGr/BERTopic/issues/40

topics was reduced because initially, some small clusters were created that often overlapped with each other and, due to the size, would not be helpful for further analysis.

4.2 Sentiment Analysis

For performing sentiment analysis in the Estonian language, there is not a wide selection of tools to choose from. In this work, sentiment analysis was performed by using three different methods because no single method was ideal, and creating a Sentiment Analysis model trained on social media data (informal texts) was out of the scope of this thesis. In the end, majority voting was applied to get the final result. Both public data from Twitter and Leaders' data from Facebook were run through these different methods.

Firstly, tartuNLP/EstBERT [41] and Estonian Valence Corpus [12] were used to pre-train a BERT model for sentiment analysis. Estonian Valence corpus consists of 4085 news extracts from Postimees Daily. All documents in the corpus are labeled with both sentiment and rubric classes. The four sentiment labels are positive, neutral, negative, and ambiguous. Texts with ambiguous labels were not considered because there was no need for a fourth classification label, and retaining them has been shown to lower the classification accuracy [12]. The model reached an accuracy of 0.745, similar to the result that the EstBERT team obtained and listed on the Huggingface[19]. This pre-trained model was used as a sentiment analysis model with the Leaders and Public datasets.

Second, the *valence model*[20] created at the Institute of the Estonian Language under the project 'Statistical Models of the Emotionality of Speech and Written Text' was used to identify the texts' positivity, negativity, and neutrality. The classifier is based on Naïve Bayes and has an accuracy of 73.6 [12]. The code is initially established to work in a server environment, e.g., a website with the valence detector here, but it can also be used as a standalone script that outputs an HTML file. In our case, we created a text file with all the tweets or Facebook posts and gave it as input to the script. The created HTML file was then parsed using the *BeautifulSoup package*[21] to extract each tweet's positive, negative, and neutral values.

Thirdly, a lexicon-based approach was done by using the R package Syuzhet [42]. The package comes with built-in dictionaries and the ability to use custom dictionaries. To get the sentiment, we implemented NRC Emotion Lexicon [43] which is also available in Estonian. The lexicon is a list of words for positive and negative sentiments and their associations with eight emotion classes (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust). Although, in this work, only the negative and positive sentiments were used. To get the sentiments, the Syuzhet package function *get_sentiment* was used, which outputted the sentiment scores. Values under zero were considered negative, values over zero positive, and values equal to zero neutral. Using this method comes with a limitation since the terms in the lexicon have been achieved by translating the English terms into various languages using *Google Translate*[22].

Finally, majority voting was used to get each post's final sentiment label based on each method's predictions. To achieve that, we constrained the output of each method to one of the three classes – negative, neutral, or positive. The final result was considered neutral if all three methods predicted a post or tweet differently. The majority vote results for both the Public and Leaders can be seen in Table 1.

[19] https://huggingface.co/tartuNLP/EstBERT
[20] https://github.com/EKT1/valence
[21] https://beautiful-soup-4.readthedocs.io/en/latest/
[22] https://translate.google.com/

Table 1. Majority voting results for sentiment analysis

	Negative	Neutral	Positive
Leaders	719	1597	899
Public	8135	7121	1480

4.3 Additional tools

Grammarly[23] and Reverso[24] were used to improve this thesis's spelling, grammar, and overall delivery.

[23] https://www.grammarly.com/
[24] https://www.reverso.net/spell-checker/english-spelling-grammar/

5 Results

This chapter presents the findings from topic modeling and sentiment analysis, including the top topics identified and the main sentiments expressed within both of the datasets – the Public and the Leaders. The codebase can be accessed from the author's GitHub repository[25].

5.1 Leaders

Topic Modelling Results

Table 2 shows the result of BERTopic modeling in the form of identified keywords and a short description of the topics. According to the table, 11 topics arise across our time period (from January 2022 to March 2023).

Table 2. Topics for Leaders

Topic	Description	Frequency
War	Identifying keywords: *Ukraine, Russia, Russian, Estonia, war, Putin*	1627
Allies	Identifying keywords: *Europe, parliament, union, Baltics, to Ukraine*	422
Politics	Identifying keywords: *elections, citizen, Estonia, to rescue, ekre*	242
Defence	Identifying keywords: *national defence, defence league, defence force, national security, population defence*	199
History	Identifying keywords: *Estonian War of Independence, republic, anniversary, peace, years*	144
Energy Security	Identifying keywords: *electricity, gas, energy war, energy, car*	139
Economy	Identifying keywords: *additional budget, euro, million, government, state budget*	127
Sports	Identifying keywords: *Belarus, athletes, independence day*	110
Education	Identifying keywords: *school, culture, language, children, education, to study, teacher*	105
News	Identifying keywords: *interview, effect, to talk about, radio*	66
Environment	Identifying keywords: *agriculture, environment, climate, circular economy, rural life*	40

We obtained the most posts with the topic named *war*. This topic includes various themes ranging from military and war to humanitarian aid and domestic politics.

The topic named *allies* focuses on foreign politics, also about different meetings with representatives of countries or decisions from the European Parliament, and giving additional support to Ukraine. It also talks about increasing defence in Europe and helping refugees.

[25] https://github.com/LisannaL/Thesis_MSc/

The *defence* topic consists primarily of posts related to what investments need to be made to ensure the safety of Estonian citizens (National Defence, Defence budget). Frequent topics are also related to cybersecurity and cyber attacks, NATO, and how domestic political problems can help Russia to divide society. Additionally, some posts criticize giving weapons to Ukraine (Estonia will not have enough left).

Politics covers different parties' opinions about various Ukraine-related topics, e.g., refugees, giving support (weapons), and how their election program plans to change things. Also, worries about immigration and internal security.

The *history* topic is related to Estonian Independence Day, the same day as when the Russo-Ukrainian war started (24th of February). Also, posts about patriotism, deportation during Soviet times, and how that is connected to the actions of Russia in modern times.

Energy Security topic covers posts about the problems in the energy markets – dependency on the Russian electricity grid and worries about the winter. The topic also contains posts about the Universal electricity plan, how to conserve energy, and criticism about adopting new energy sources.

The *economy* topic concerns fuel and gas taxes, increasing defence costs, the need for investments, and creating additional state budgets to cover costs. A common topic is also inflation.

Sports topic talks mainly about whether Russians and Belarussians should be allowed to compete in National sports events like the Olympics.

The *education* topic talks about moving towards Estonian language-based education. Also, how Ukrainian refugee children adapt to the Estonian education system.

News topic contains mainly posts where either a news article or show is referenced in connection to how the politician gave an interview or voiced an opinion.

Like the energy topic, the environment topic discusses energy security, food security, and our production output. In the context of energy security, discussion about the usage of fossil fuels is common.

When looking at the Top 3 topics for each month in Table 3, the most popular topics are almost always *war* or *allies*. This could have occurred for two reasons – the initial filtering contained keywords related to War and Ukraine. Furthermore, due to the nature of the dataset and thesis topic, *war,* and *allies* are the most common talking points. Though, through time, some change can be seen. For example, in May of 2022, the *economy* becomes an important topic which might be due to the price rises for food and household goods. The *history* topic is relevant for certain times of the year. In this case, it is February when the Estonian independence anniversary also takes place.

Table 3. Top 3 topics per each month

Jan22	War	Allies	History
Feb22	War	Allies	History
Mar22	War	Allies	Defence
Apr22	War	Allies	Defence
May22	War	Allies	Economy
Jun22	War	Allies	Defence
Jul22	War	Allies	Energy Security
Aug22	War	Politics	Sports
Sep22	War	Sports	Politics
Oct22	War	Allies	Energy Security
Nov22	War	Allies	Defence
Dec22	War	Allies	Defence
Jan23	War	Politics	Allies
Feb23	War	History	Allies
Mar23	War	Allies	Sports

As expected the most prevalent topics are related to the ongoing war. Topics *war* and *allies* being the biggest ones, but we can also see a grouping of other topics emerging that indicate other important issues to the Leaders. Topics like *politics*, *energy security*, *economy*, *history*, and *defence* also frequent the top 3 keywords leaderboard. This gives us additional insight into the issues that politicians find important to discuss.

Sentiment Analysis Results

As seen in Figure 4, the maximum posting activity was in March of 2022. In addition to that, we can observe the sentiments and their frequencies in Figure 6. What is most notable is that the majority of Leaders' posts are neutral. Although, through our defined timeline, there are occasional spikes of posts that can be related to some of the events[26] happening.

Another interesting thing to see is that the Leaders started posting more in Autumn, which is probably related to the *energy security* issues as the winter is approaching, but as well as because Russia started the partial mobilization. Other upticks are in the beginning of 2023, which could possibly be related to the approaching elections in Estonia, and the fact that it has been one year since the war started.

[26] https://edition.cnn.com/interactive/2023/02/europe/russia-ukraine-war-timeline/

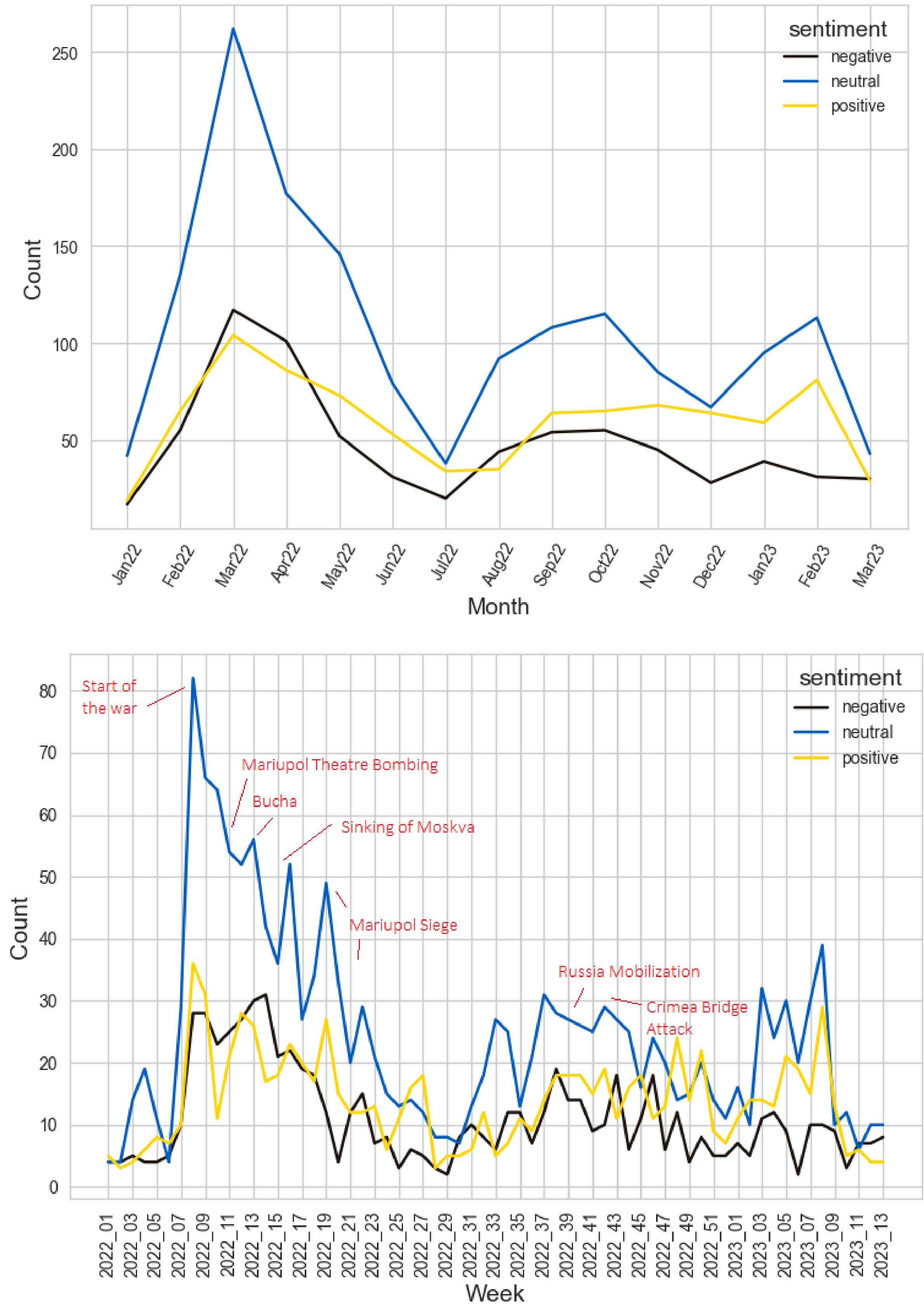


Figure 6. Leaders' sentiments through time, monthly, and weekly intervals

Figure 7 provides insight into topics and their sentiments. We notice that topics with the highest proportion of negative sentiments are *politics*, *war*, and *energy security*. In contrast, the ones with more neutral or positive sentiments are *news*, *education*, and *allies*.

Negative posts in *politics* topic are often about politicians criticizing other politicians. For example, the posts are often about who is at fault because of the current economic situation (the current coalition blames the current opposition that was in power before, opposition blames the coalition because they have previously been in power for a long time). Additionally, negative sentiments are expressed towards using symbols that justify Russia's military actions.

Also, *war* and *energy security* topics have a higher negative sentiment because many of these topics bring uncertainty to people's lives. *Energy security* might have a more negative sentiment, because it directly affected people negatively by rising prices and the uncertainty it brought. Similarly, the *war* topic is more negative, as people are against it and Russia's actions in Ukraine.

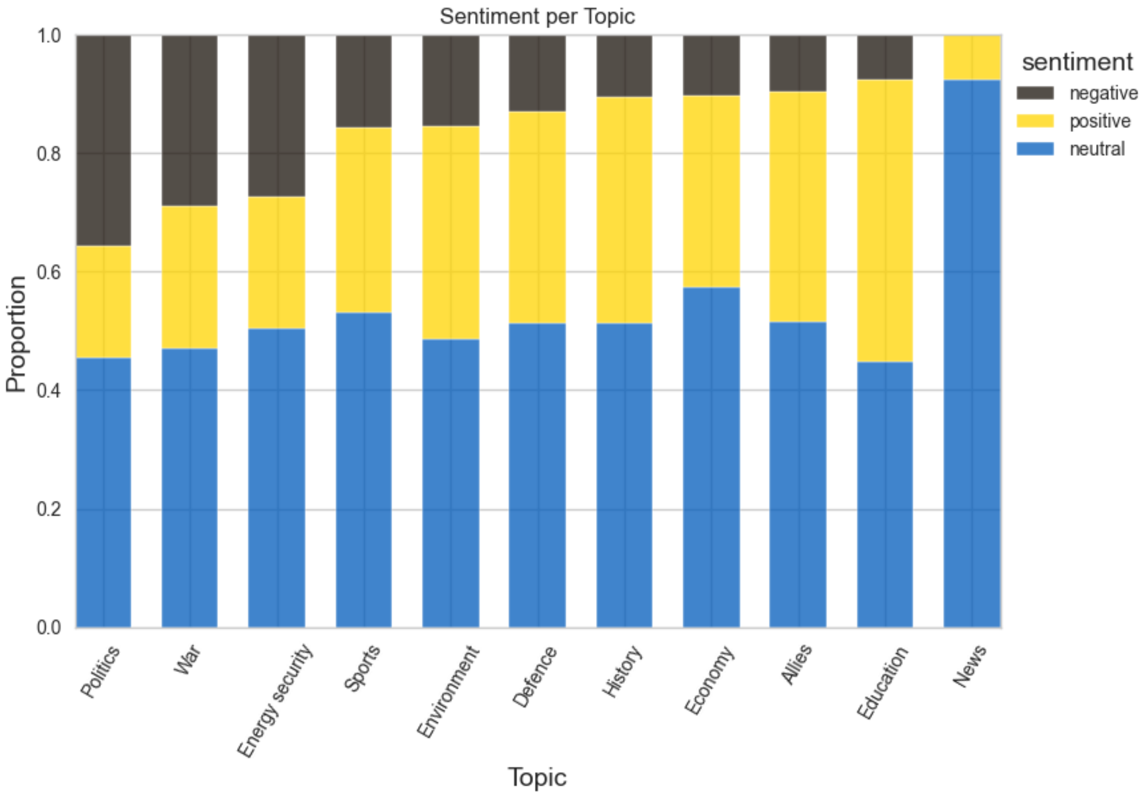


Figure 7. Topics and their sentiments present in Leaders' dataset

Figure 8 shows how each topic has changed over time. For instance, the *energy security* topic gains some traction at the beginning of autumn, probably because of the upcoming heating season. Another interesting example is *sports*, where the number of posts increased in the August of 2022 and January of 2023, although categorized as neutral, where criticism against ROC's decision to allow Russians and Belarusians to compete under a neutral flag was expressed.

For the topic of *allies*, many of the posts were written at the beginning months of the war, which is understandable; since then, it has been essential to organize all sorts of support for Ukraine.

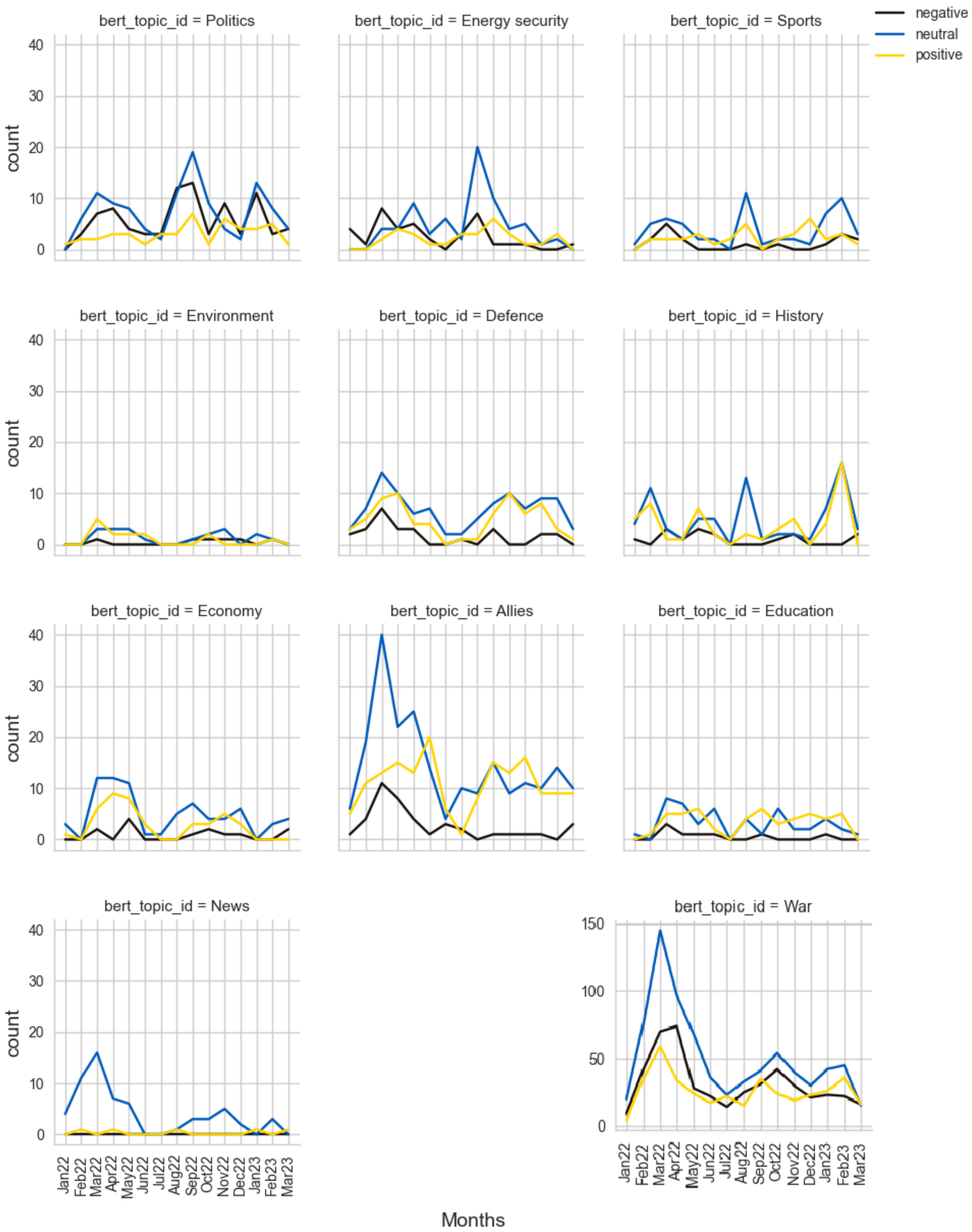


Figure 8. Leaders' topics and their sentiments through time

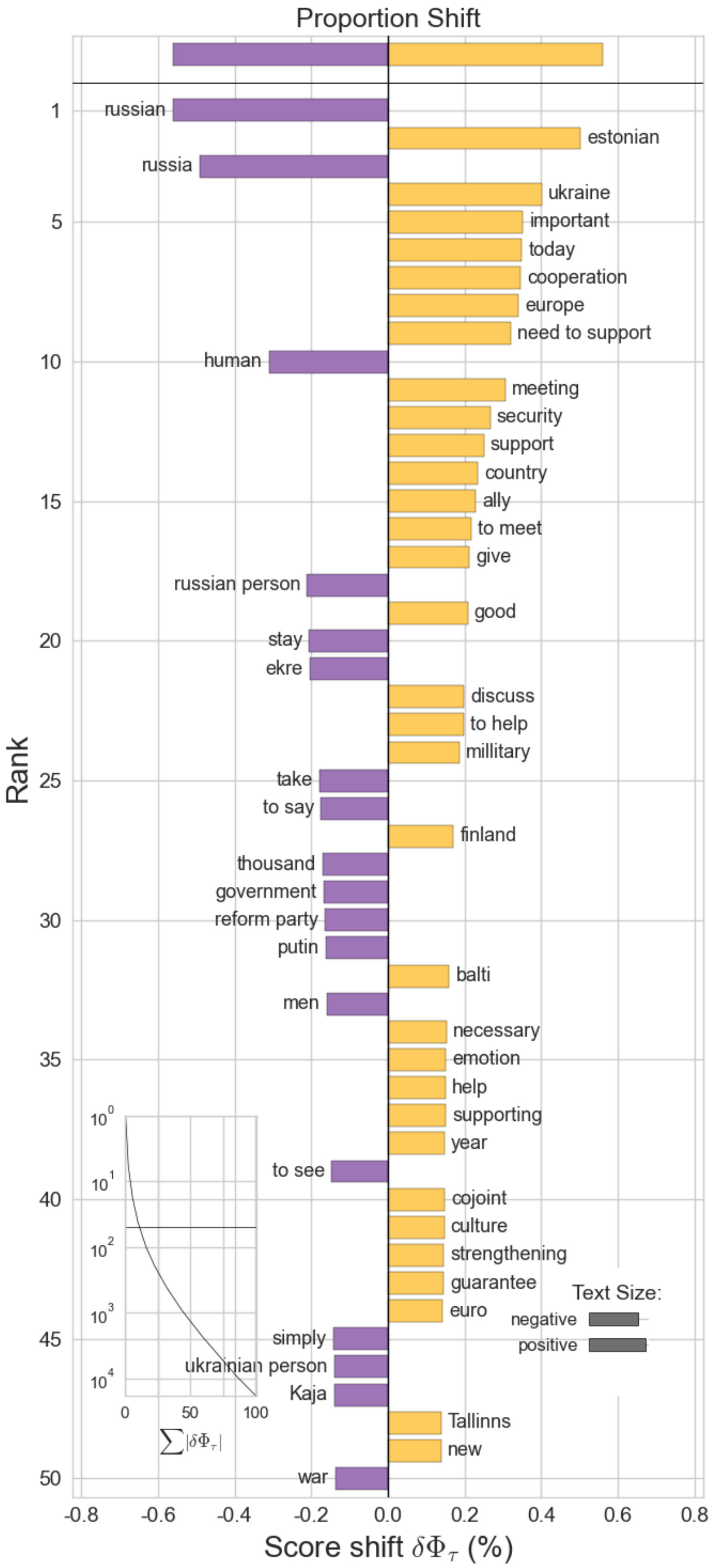


Figure 9. Proportion Shift of negative and positive sentiment datasets[27]

[27] Here, for plotting, Estonian words are translated to English.

As expected, overall in the corpus, the word ukraina (*Ukraine*) is the most occurring, reflecting the current situation. Furthermore, the second and third most frequent words are eesti (*Estonia*) and venemaa (*Russia*). Given the filtering criteria, it makes sense that these three words are the most common as the war is between Ukraine and Russia, and Estonian political leaders post about the war. The *Shifterator* Python package [44] further compares negative and positive sentiment posts. Shifterator enables one to make pairwise comparisons between texts by comparing how different words contribute to each text.

Relative word frequency-based proportion shifts were used when comparing negative and positive sentiment posts. Figure 9 shows the proportion shift. If the difference is positive, the word is relatively more common in the positive text and vice versa for the negative[28].

Based on Figure 9, the words *Russia*, *Russian*, and *Putin* are more common in posts with negative sentiments. The same applies to Estonian domestic politics-related words like *EKRE* (a party), *Reform party*, and *government*. Words that are more common in texts with positive sentiment focus on cooperation, support, and the willingness to help Ukraine.

Overall, politicians may be mindful of their public image and seek to maintain a relatively neutral perception in their published social media posts. Therefore, they may avoid expressing strong emotions to appeal to potential supporters.

Political Parties Comparison

Figure 10 shows the sentiment distributions of political party posts. Interestingly, while most parties have a similar proportion of neutral posts, making up about half of the posts, the positive and negative sentiment proportions vary slightly. The majority of neutral posts suggest that political parties may be hesitant to take strong stances on social media or may be trying to avoid controversy by staying neutral. However, some parties stand out, with more positive or negative sentiments than others. For example, EKRE stands out with having the most negative posts, and KESK stands out with having the most positive posts.

A higher percentage of posts with negative sentiment could come from creating posts that handle topics related to the war that are inherently negative, like war crimes. However, it could also come from posts meant to create controversy with polarizing issues or parties' criticism of each other.

In the context of war, a higher number of positive posts could be related to posts that give hope - like donations, success in war, and allies response. However, it could also come from downplaying the seriousness of war.

[28] https://shifterator.readthedocs.io/en/latest/cookbook/frequency_shifts.html

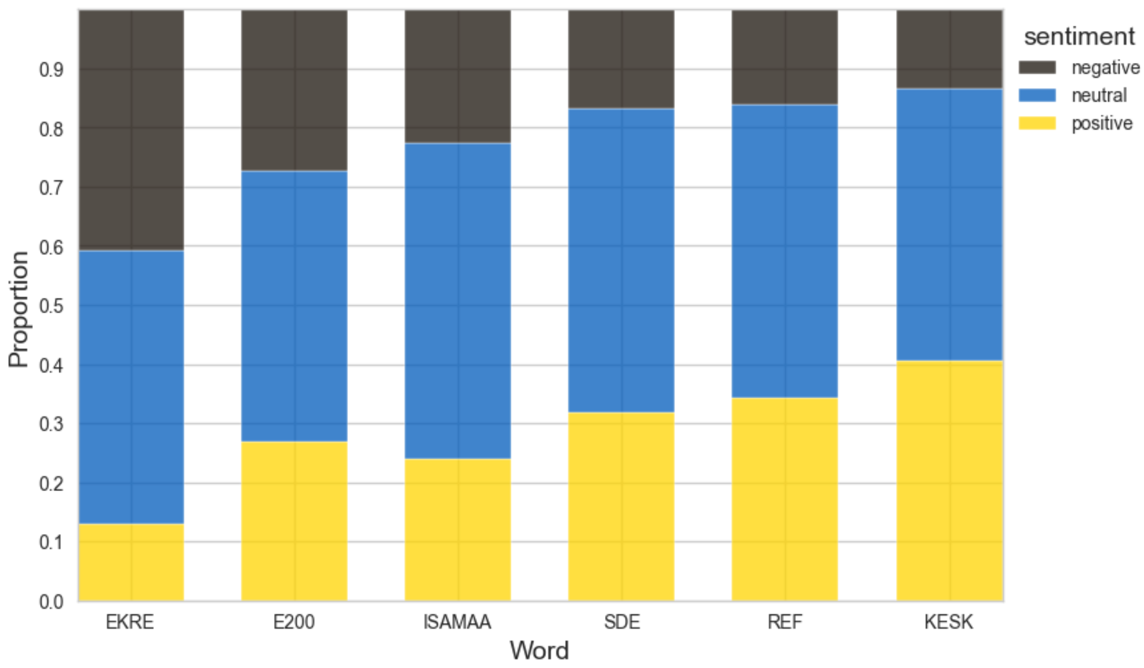


Figure 10. Sentiment proportions per party

As seen in Figure 11, E200 and EKRE, showed some increase in the number of posts when the war started. However, there was a decline after that, and the number of posts stayed relatively stable throughout the year. Isamaa, REF, and SDE show multiple spikes throughout the year, indicating that they reacted to some events, while KESK was primarily stable, with a slight decline during the summer months.

Figure 11. Parties' sentiments through time

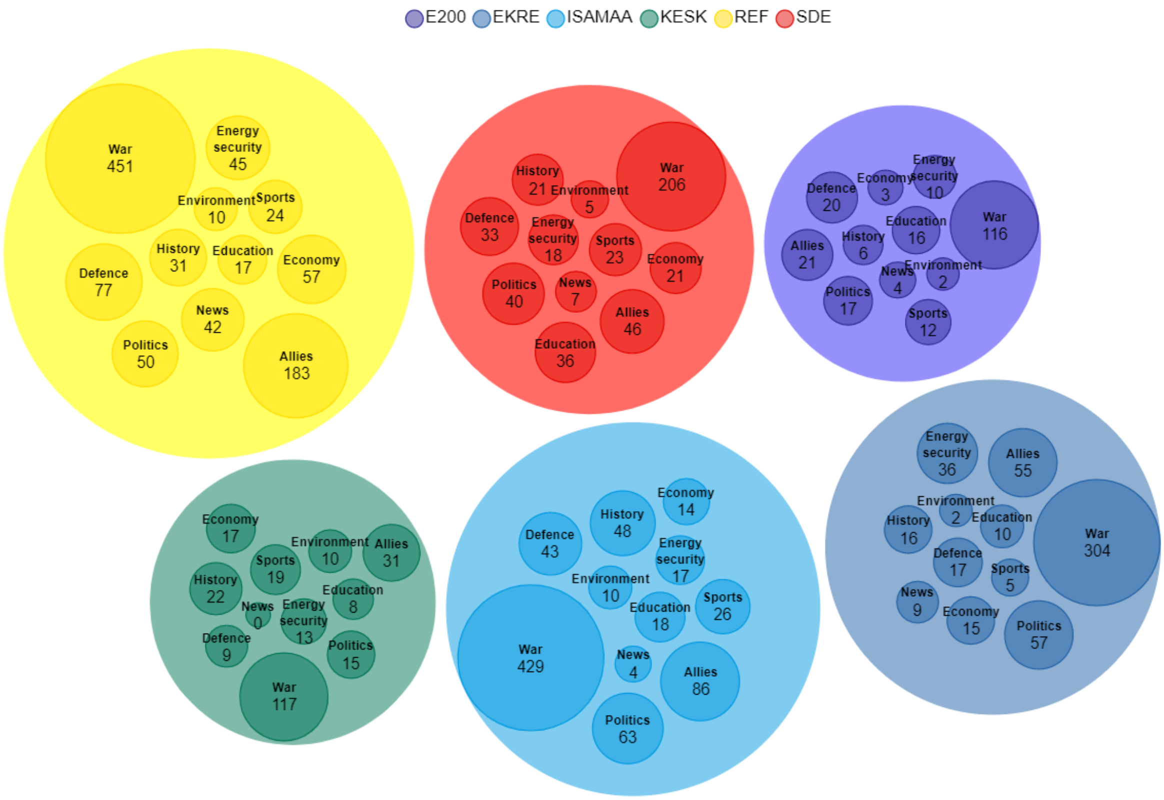


Figure 12. Topics allocation per parties

The results of topic modeling show (Figure 12) that while *war* and *allies* are the most popular topics for all the parties, they still differ regarding other issues. For example, REF and E200 most discussed topics are related to *war*, *allies*, and *defence*. SDE, EKRE, and Isamaa most discussed topics are *war*, *allies*, and *politics*, while KESK has *war*, *allies*, and *history*.

Based on the dataset creation rules, parties have quite a significant overlap. However, the similarity could also be attributed to the fact that when important topics arise in society, all parties usually state their opinions and stance. What is also interesting from the chart above is the proportion of topics – we observe how *allies* and *war* make up a higher percentage of posts for REF and Isamaa. Comparing the proportions of topics between parties, the data shows that, for example, how *energy security* is an essential topic for EKRE, *sports* for KESK, *history* for Isamaa, and *education* and *politics* for SDE.

In conclusion, we noticed that all parties' activity spiked at the start of the war but quickly declined afterward. For most parties, after the decline, activity stayed stable, then Isamaa, REF, and SDE had multiple spikes throughout the year. As for the topic distribution, the top 2 were always the same - *war* and *allies*. Where parties differed was the proportion of the main topics and also what were the other important topics, indicating that parties had slightly different values.

5.2 Public

Topic Modeling Results

Table 4 shows the result of BERTopic modeling in the form of identified keywords. According to the table, 15 topics arise across our defined time period.

Table 4. Topics for Public

Topic	Description	Frequency
Russia	Identifying keywords: *Putin, Russia, war, Russian*	5235
Estonia	Identifying keywords: *Estonia, Europe, Ukraine, Russian*	4248
War	Identifying keywords: *war, West, Russian, in Ukraine, USA, Russia*	2710
News/Social Media	Identifying keywords: *Twitter, news, media, account*	1021
Combat	Identifying keywords: *missile, cities, airspace, to die*	660
Refugees	Identifying keywords: *refugee, refugees*[29]	427
Energy Security	Identifying keywords: *gas, Europe, fossil fuel, electricity, oil*	409
Donations	Identifying keywords: *money, donations, to donate, for support*	385
Eurovision	Identifying keywords: *Eurovision, song, to win*	355
Education	Identifying keywords: *school, children, language*	304
Nuclear	Identifying keywords: *NATO, nuclear plant, nuclear weapon*	299
Women	Identifying keywords: *women, woman, mother*	240
Sanctions	Identifying keywords: *sanction, SWIFT, EL*	188
Flag & Colors	Identifying keywords: *flag, color, colored, yellow*	138
Sports	Identifying keywords: *athletes, FIFA, Belarus, ROC, competition*	117

The topic, *Russia*, discusses the ongoing war in Ukraine, Putin, and Russia's involvement.

The topic of *Estonia*, similarly to *Russia*, focuses on the ongoing war. Additionally, it focuses on the war in the context of Estonian domestic politics.

The topic of *war* is related to the ongoing war in Ukraine and the involvement of Russia and Ukraine's allies. The topic is focused on the military conflict and the political and military dynamics of the conflict in Ukraine.

The *news/social media* topic involves news made on the Ukraine-War topic and matters discussed on social media platforms like Twitter.

The *combat* talks about the need to give more military support to Ukraine. Also, about the constant military actions of Russia and ensuing war crimes.

[29] In Estonian, there are multiple synonyms for the word 'refugee' and the topic consists mainly of these.

The *refugees* topic focuses on people that escaped the war, and also about helping refugees in Estonia and those that have been misplaced in Ukraine.

Energy security covers tweets about energy and its sources and our geopolitical situation, which means how Europe as a region heavily depends on energy imports and how things have changed with the war.

The *donations* topic discusses different ways to support Ukraine and calls for action.

The *Eurovision* topic is related to the Eurovision that took place in 2022. It generated many posts about Ukraine winning and Russia being banned from participating.

The *education* topic discusses moving towards Estonian language-based education and how Ukraine children are adopting the Estonian education system.

The *nuclear* topic contains posts related to the nuclear threats by Russia and the occupation of the nuclear plant in Ukraine.

Women's topic contains posts about Ukrainian women providing support on the frontlines, but also about war crimes towards women, the need for abortions, and mothers having to escape the war.

The *sanctions* topic contains tweets about European sanctions against Russia and their effectiveness. A common topic is also removing Russian banks from the SWIFT.

The *flag and colors* topic contains posts related to Ukraine symbols and wearing them in public.

A *sports* topic similar to *Eurovision* contains tweets about banning Russian and Belarussian athletes from the Olympics.

The top 3 topics for every month were always the same in the following order - *Russia*, *Estonia*, and *war*. The composition of the top topics makes sense, as they have the most tweets and follow the theme of the dataset. To get more context, Table 5 shows how the top 4-6 topics change over time. The most popular topics are *news/social media*, *combat*, and *energy security*. Interestingly, at the start of the war, it can be observed that *refugees* and *women* topic are popular. As the year goes on, starting from June, *energy security* becomes a more prominent topic as autumn and winter are nearing. Although *combat* is a popular topic throughout the year, we see it becoming more popular during July and August. This could be because of the Ukrainian counterattacks starting. It is also worth pointing out that the *Eurovision* topic became more popular in May. In December, a *sports* topic rose to the top 6 because of the International Olympic Committees' signals regarding allowing Russian and Belarussian athletes to compete.

Table 5. Top topics per each month

Jan22	Nuclear	News/Social Media	Energy Security
Feb22	News/Social Media	Combat	Donations
Mar22	News/Social Media	Combat	Refugees
Apr22	News/Social Media	Refugees	Women
May22	Eurovision	News/Social Media	Combat
Jun22	News/Social Media	Combat	Energy Security
Jul22	Combat	News/Social Media	Energy Security
Aug22	Combat	News/Social Media	Refugees
Sep22	News/Social Media	Energy Security	Combat
Oct22	News/Social Media	Combat	Energy Security
Nov22	News/Social Media	Combat	Energy Security
Dec22	News/Social Media	Combat	Energy Security
Jan23	News/Social Media	Combat	Sports

Topic Modeling results provide insight into the most popular topics related to the Ukraine-Russia war discussed by the Public. The analysis revealed 15 topics, where *Russia*, *Estonia*, and *war* were the most discussed. Although throughout the year, other topics also become prominent, for example, *energy security* and *combat*.

Sentiment Analysis Results

As shown in Figure 4, similarly to Leaders' the Public's highest posting activity was in March of 2022. We depict the sentiments with their frequencies in Figure 13. Most of the Public's posts are negative or neutral (see Table 1). Again similar to Leaders, there are occasional spikes of tweets that can relate to some of the events[30] happening. Rather negative spikes in the 14th week and the 19th week may correspond to the Butcha massacres being discovered and the end of the Mariupol Siege accordingly.

[30] https://edition.cnn.com/interactive/2023/02/europe/russia-ukraine-war-timeline/

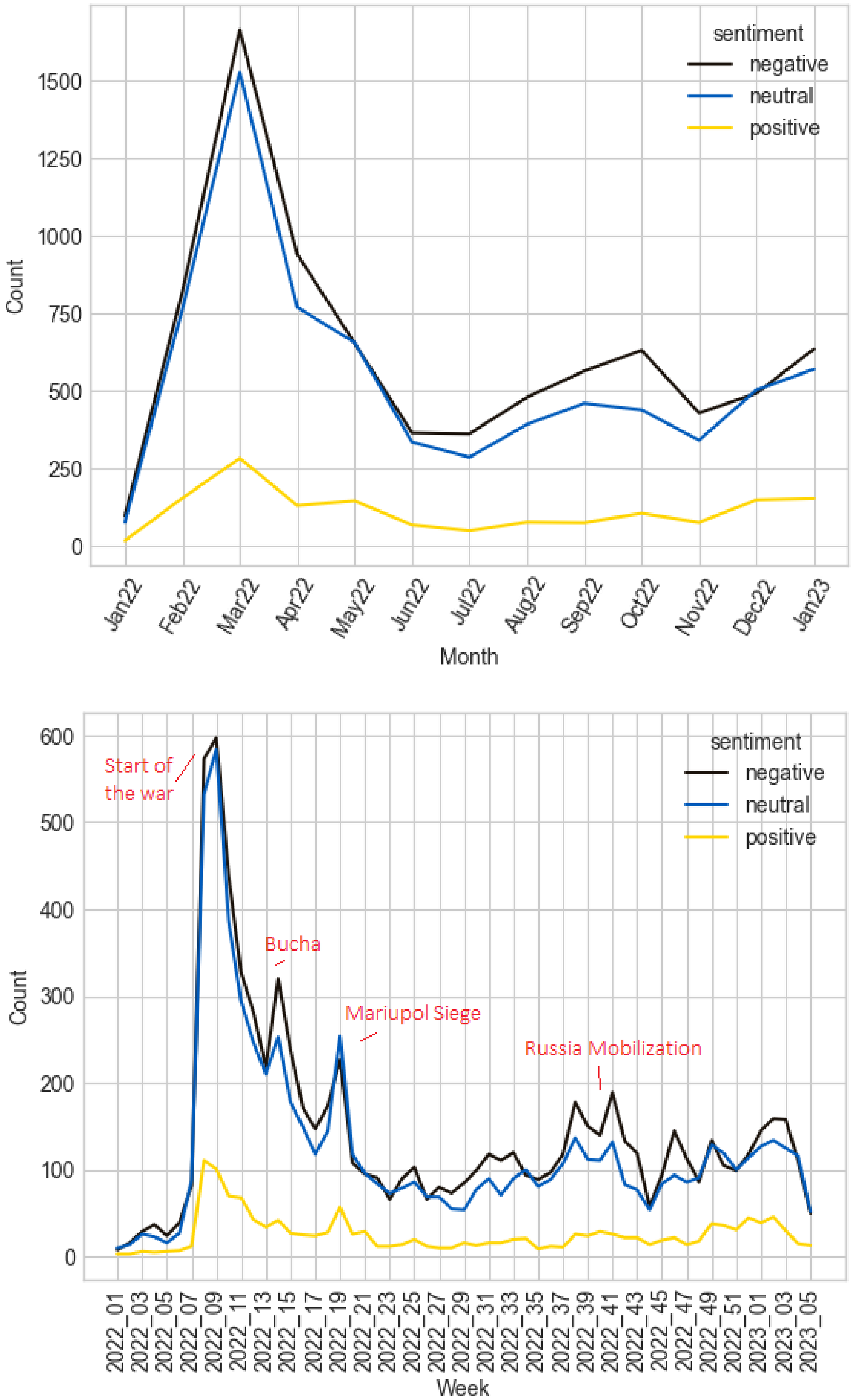


Figure 13. Public's sentiments through time, monthly, and weekly intervals

Figure 14 provides insight into topics and their sentiments. The results show that we have about eight topics where more than half of the tweets are negative. The topics with the highest negative sentiments are *combat*, *nuclear*, *women*, *news/social media*, and *energy security*. In contrast, the ones with more neutral or positive sentiments are *donations*, *Eurovision*, and *education*.

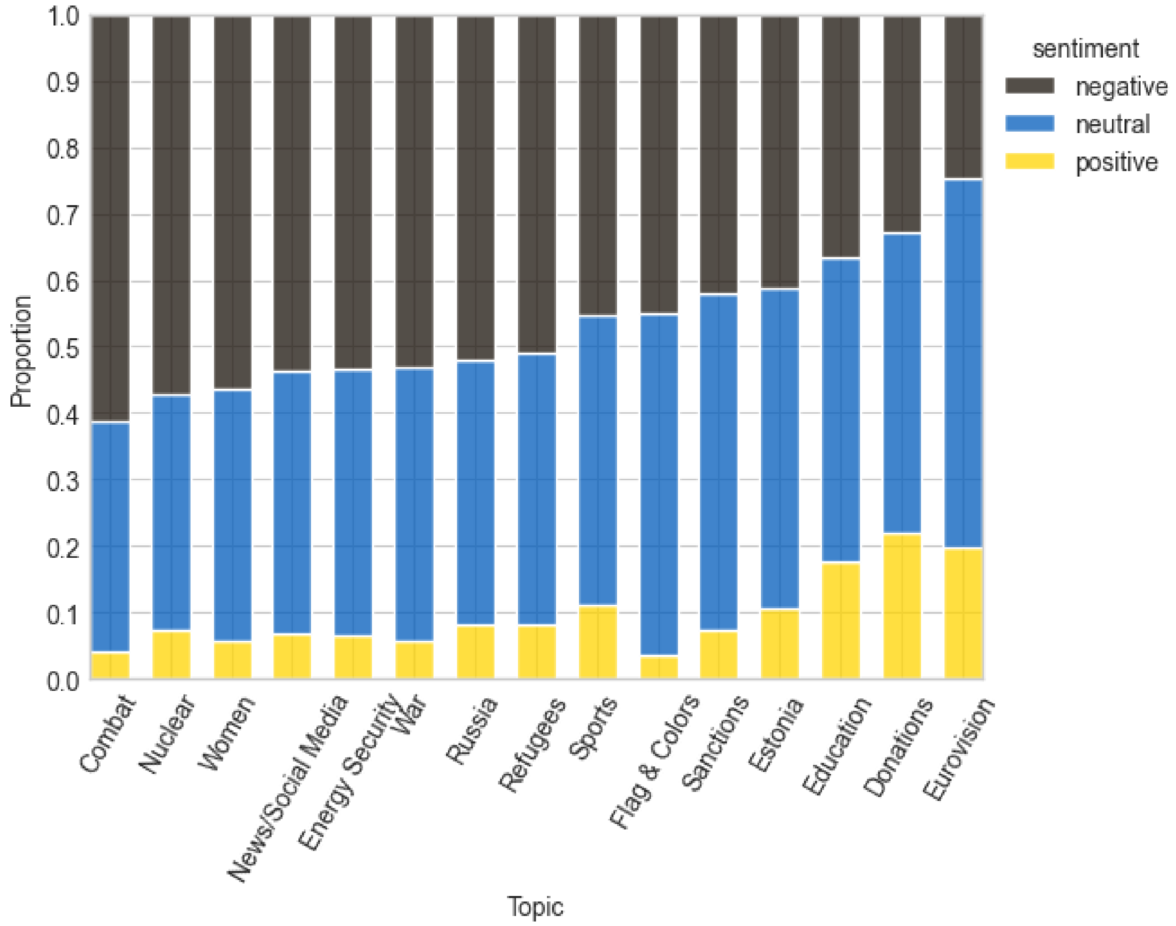


Figure 14. Topics and their sentiments present in the Public's dataset

Negative posts on a topic focusing on *combat* indicate Russia and Ukraine's forces clashing on the battlefield, e.g., n-grams like "vene vägi" (*Russian military*), "venemaa rünnak" (*Russian attack*) and "ukraina sõdur" (*Ukraine soldier*) appear often. Furthermore, there are references to weapons, missiles, and drones used in military operations. Additionally, mentions of controversial events (e.g., bio lab) also appear, potentially relating to the controversy (now already debunked) surrounding Ukraine and its development of biological weapons[31]. Moreover, concerns about Estonian airspace are mentioned.

The negative posts on the *nuclear* topic are focused on the nuclear power plant and nuclear weapons. The ngrams "venemaa tuumarelv" (*Russian nuclear weapon*) and "ukraina tuumajaam" (*Ukrainian nuclear power plant*) appear often, and this suggests that the discussion revolves around the potential danger posed by these nuclear facilities. For example, potential environmental risks associated with the nuclear power plant and Russia's threats about using nuclear weapons.

[31] https://en.wikipedia.org/wiki/Ukraine_bioweapons_conspiracy_theory

Negative posts on a topic focusing on *women* revolve around Ukrainian women and the Russian military. There are mentions of rape, indicating that sexual violence is a significant issue that people talk about. Furthermore, women are also spoken about with a negative undertone, with prostitution and immigration being mentioned.

The texts classified as negative under the *news/social media* topic discuss how the media report and interpret the conflict. For instance, the role of Russian propaganda in shaping public opinion and, consequently, promoting Russian interests in the conflict.

The negative sentiment in the *energy security* topic focuses on Russia's dominance in the gas market and the dependency of Europe on Russian gas supplies. The texts also suggest that Russia has used its energy resources as a political tool to pressure neighboring countries like the Baltics. Therefore, there are concerns about the reliability of Russia's gas supplies and the need to diversify energy sources or reduce dependence on Russian resources.

On the contrary, the positive posts in *donations* focus on supporting Ukraine and different ways to do that. For example, people donate money and other resources to assist Ukraine through non-profit organizations or fundraising. The positive posts on the *Eurovision* topic express happiness about Ukraine winning the Eurovision of 2022. For *education*, there are posts about how to help refugees learn Estonian and how to support them to be able to continue their education.

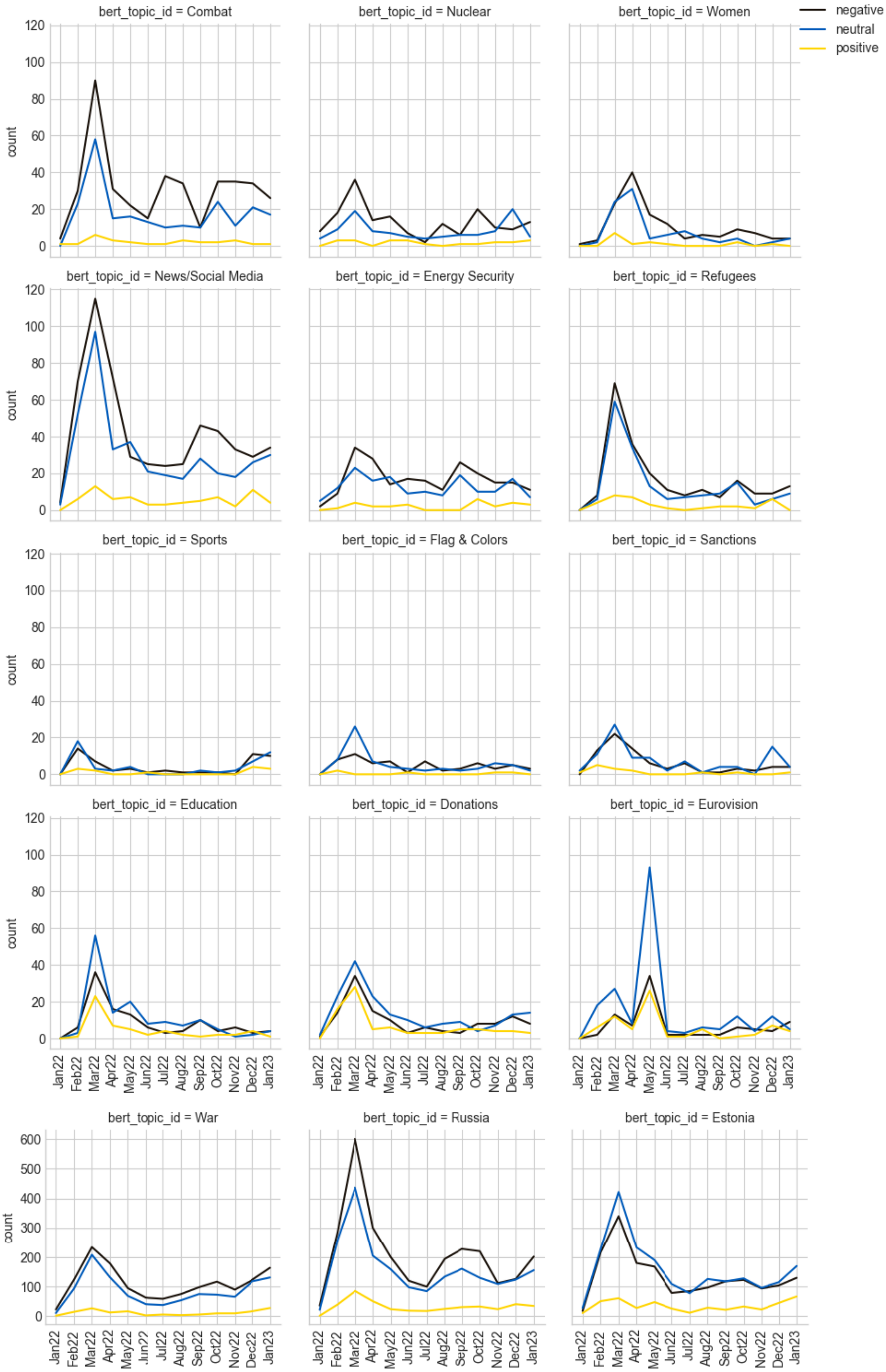


Figure 15. Public's topics and their sentiments through time

Figure 15 shows how each of the topics has changed through time. For instance, the topic of *combat* spikes multiple times throughout the year. Firstly, the beginning of the war (March of 2022). Secondly, in July of 2022, probably due to the destruction of Russia's ammunition warehouse in Nova Kakhovka[32]. Thirdly, in October 2022, the tweets discussed the Crimean Bridge explosion[33] and cyberattacks against countries supporting Ukraine.

Women and *refugees* topics are the most popular at the beginning of the war. The same can be said for *education* and *donations*.

For instance, similarly to Leaders (see Figure 8), the *energy security* topic gains some traction at the beginning of autumn, probably because of winter and the upcoming heating season. Moreover, the NordStream gas pipeline explosion is being discussed.

The *Russia* topic generally contains various tweets related to the war. An interesting thing to notice from the posting activity is the initial uptick of tweets during the start of the war. Additionally, there is more significant interest from August to October of 2022. The additional posting activity could be attributed to important topics discussed in the news, like the Ukraine counteroffensive in Kherson and Kharkiv, respectively, in August and September and the Crimea bridge attack in October[34].

Estonia topic contains tweets related to Estonian politics, Estonian Independence Day, learning the Estonian language, and Estonians' reaction to Ukrainians and Russians living in Estonia. Similarly to previous topics, there was an initial uptick during the start of the war that can be attributed to the Estonian Independence Day and also due to the uncertainty the war causes.

Logically, most posts on the *Eurovision* topic were written when the Eurovision of 2022 took place. However, before May, there was a discussion revolving around Russia being excluded. After, discussions on Ukrainian music and musicians performing in charity events to support Ukraine.

Figure 16 shows the word frequency-based proportion shift between negative and positive sentiment subsets of the Public. Likewise to Leaders, the words relating to the conflict and difficult political situation are *Russia*, *Russian*, *war*, and *Putin*, *which* are more common in Public's posts with negative sentiments. Also, a domestic policy-related word like *EKRE* (a party) is mentioned relatively more in the negative sentiment subset. Words more common in texts with a positive sentiment are associated with a sense of collaboration and support, such as *Ukraine*, *Estonia*, *Europe*, *support*, *help*, and *cooperation*.

---

[32] https://www.reuters.com/world/europe/ukraine-prepares-fresh-russian-assault-west-braces-worsening-energy-crisis-2022-07-12/

[33] https://edition.cnn.com/2022/10/08/europe/crimea-bridge-explosion-intl-hnk/index.html

[34] https://www.weforum.org/agenda/2023/02/ukraine-war-timeline-one-year/

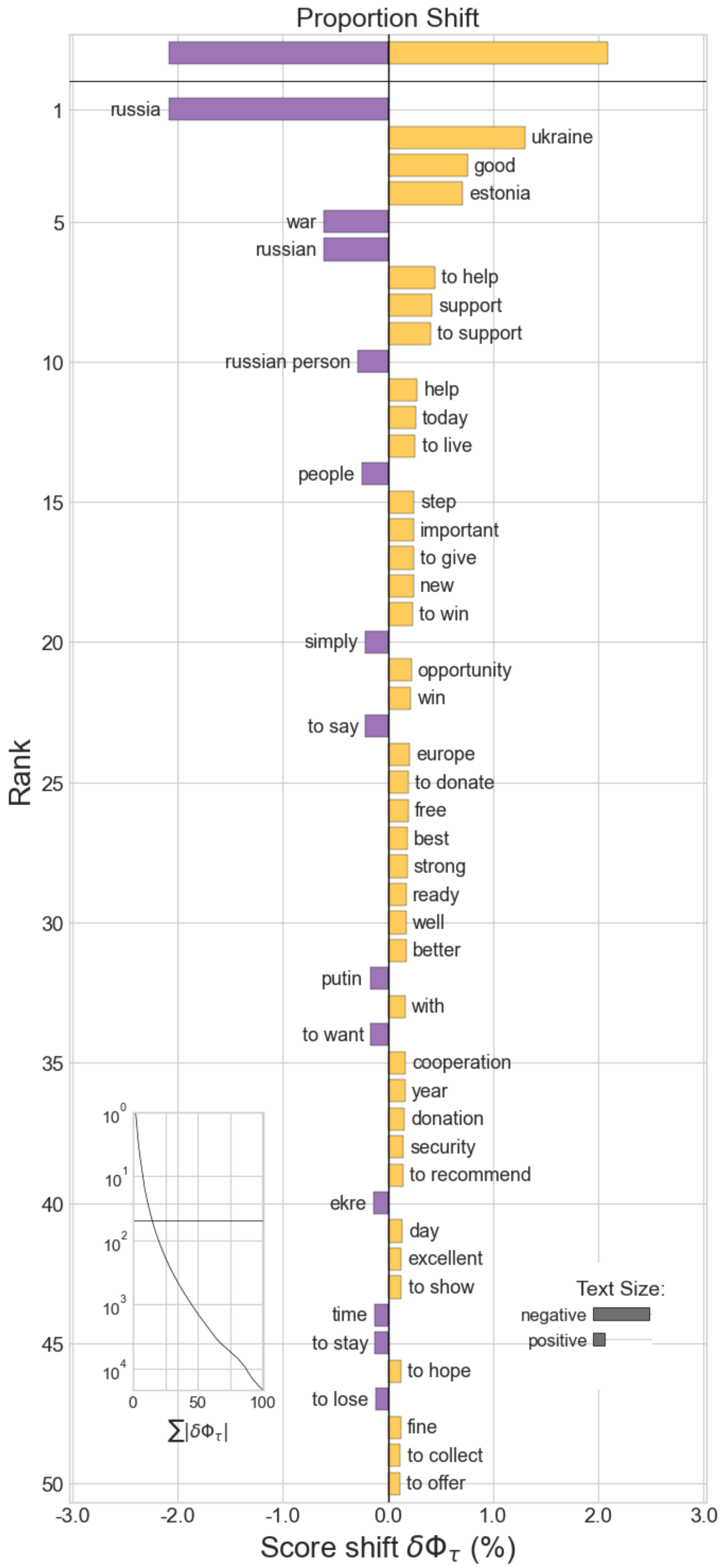


Figure 16. Proportion Shift of negative and positive sentiment datasets[35]

[35] Here, for plotting, Estonian words are translated into English.

Overall, it seems that while sentiments of Leaders' posts were primarily neutral, then for Public posts, the sentiment leaned toward negative. This, in turn, can be attributed to the fact that while politicians should keep a more neutral tone, the public does not have such a constraint. Additionally, the Twitter platform is ideal for sharing short emotional tweets. The Ukraine war is expected to generate uncertainty and emotions in Estonians.

Keyword Analysis

In this section, we are looking into whether we can pick out any trends from keywords and see how long a keyword stays relevant for some events. Figure 17 shows the top three keywords change over time. After the initial spike (start of the war), there was an additional spike on the 19th week of the year. This corresponds with the end of the Mariupol Siege (Azovstal steel plant)[36]. After that, the posting frequency stays relatively the same, with minor spikes at the end of the year. For example, in the 38th week, Russian mobilization was a major topic[36]. The spike at the beginning of 2023 could possibly be because of the announced additional military aid for Ukraine[36].

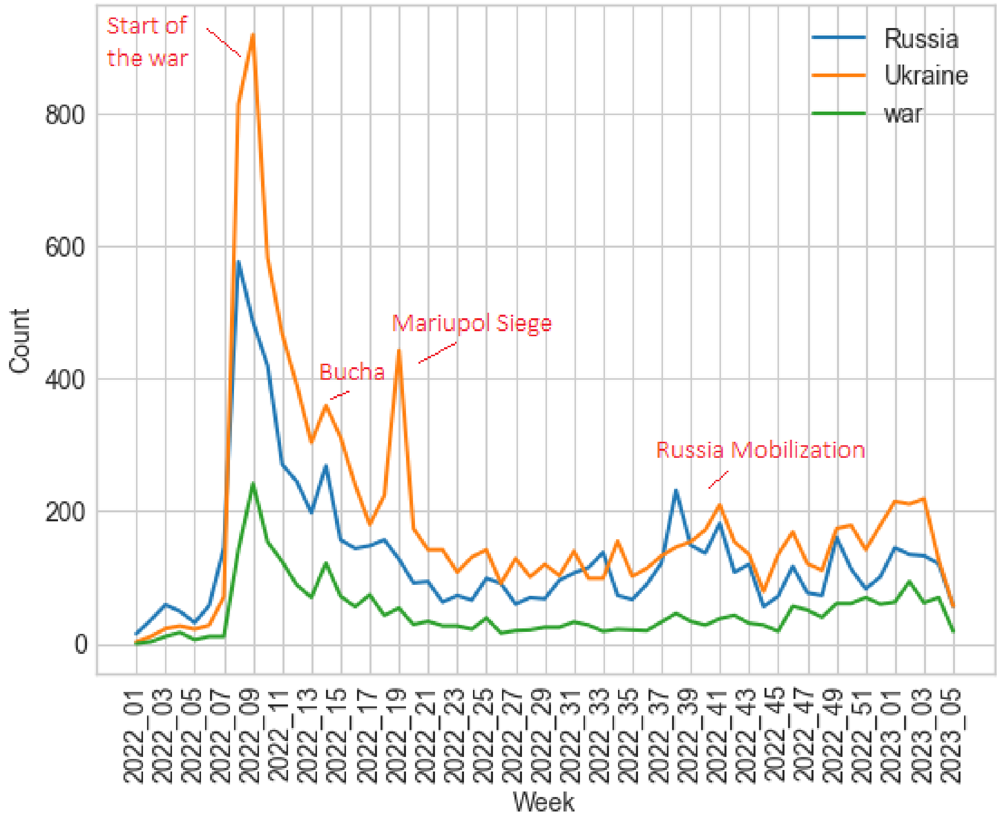


Figure 17. Change of keywords *Russia*, *Ukraine* and *war*

Keywords for Figure 18 were chosen so that we could see how keywords related to events that were important to people rise and fall out of memories. For example, on the 38th week, Russia announced a partial mobilization, and the results show that the keyword quickly spiked and fell after a couple of weeks. Similarly, the results indicate a *war crime* spike in the 13th week, corresponding to Bucha's uncovered massacres. We can also see that *sanctions* were a significant issue in the first weeks of the war, but we can also see an increase in their mentions

[36] https://edition.cnn.com/2023/01/25/europe/german-tanks-ukraine-intl/index.html

after the Bucha massacres were recovered. However, afterward the rest of the year, the activity for these keywords is relatively small, with a small spike in the 50th week of the year.

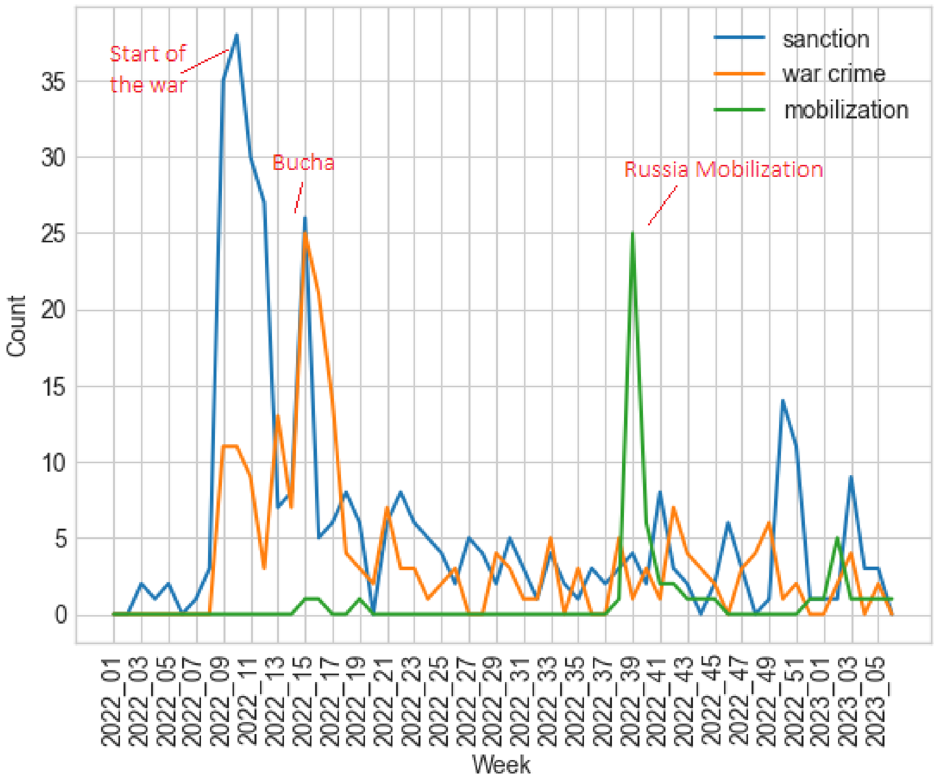


Figure 18. Change of keywords *sanction*, *war crime*, and *mobilization*

The spikes in keyword activity correspond to important events during the war, such as the Mariupol Siege and the Bucha massacres. The results indicate how the corresponding keywords quickly rise, and though some stay relevant for longer, the activity still decreases after a few weeks and falls out of people's attention.

5.3 Comparisons between Leaders and the Public

Figure 19 shows the proportion shift between the Leaders and the Public. For Leaders, it can be seen that compared to the Public, there is a more significant focus on domestic and foreign politics. Keywords related to domestic politics (*government*, *parliament*, *prime minister*, *political party*, *Reform Party*, *Isamaa*, *chairman*) and keywords related to foreign politics (*ally*, *union*, *member*, *international*, *Europe*) show up more in Leaders' posts than in Public posts.

The reason could be that they reflect on their daily activities related to foreign policy and also talk about domestic policy. At the same time, this is perhaps not a topic that ordinary people post that much. Keywords related to political parties can also be seen, which may indicate talks about the actions of own party or other parties. One topic that still comes up concerns national defense, which probably stems from the geopolitical situation of Estonia.

In contrast, words like *Ukraine*, *Russia*, *west*, and *war* appear more in public posts than in Leaders' posts. Although post size plays a role here, more usage of these words may occur because Twitter posts contain more emotion.

For Leaders, it is important to convey foreign and domestic-related posts to the Public through their Facebook posts, which might also help explain the difference between them. Another reason for this difference can come from the fact that the average Facebook post is multiple times bigger than an average Twitter post. This means that although the datasets contain a similar amount of words, there are multiple small Twitter posts mentioning Ukraine.

In conclusion, the analysis found that there is a difference between the Leaders’ and Public’s posts. Although this difference is expected and could be attributed to the leaders’ everyday activities, and their need to inform the Public of their political agenda, and at the same time, the Public may be more emotional in their posts to the ongoing events.

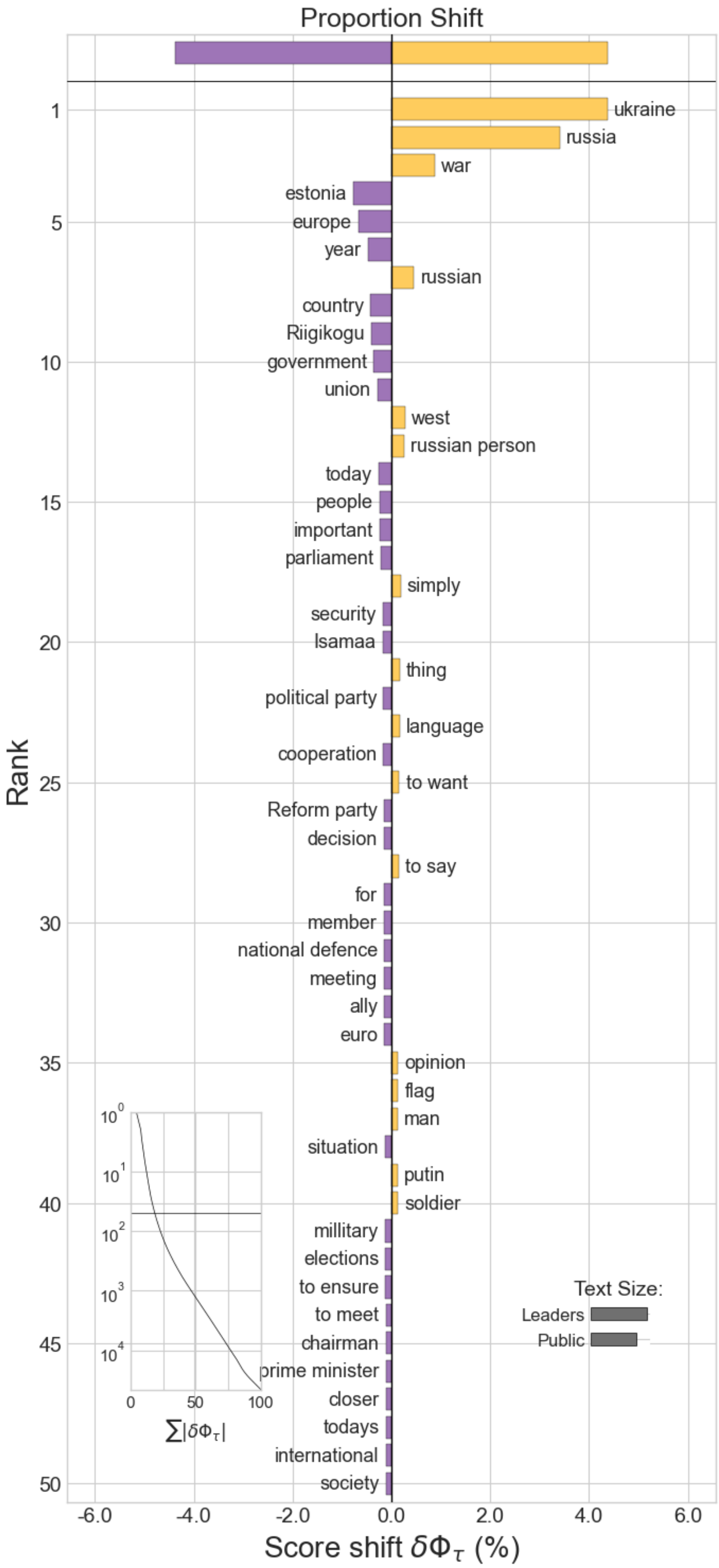


Figure 19. Proportion Shift of Leaders and Public[37]

[37] Here, for plotting, Estonian words are translated into English.

6 Discussion

This section connects the results with the proposed research questions. In addition, limitations are discussed and ideas for future work are proposed.

RQ1 - What topics related to the Ukrainian war are important to people?

According to topic modeling results, *war*, *energy security*, *economy*, *sports*, *education*, and *news* [5] were the overlapping topics that both the Leaders and the Public discussed (see Table 2 and Table 4). The overlap could be explained by the fact that these are the most important topics to both groups, that generated the most discussion. Moreover, while Leaders discussed the issues related to the *economy*, i.e., concerns about fuel and gas taxes, increasing defence costs, the need for investments, and creating additional state budgets to cover costs. The Public discussed a similar topic, *sanctions* and their effectiveness against Russia. This is supported by the fact that topics like *energy security* [5,13] and *sanctions* [5,6] have also appeared in the previous research.

Furthermore, topics like *allies* and *politics*, *defence*, *history*, and *environment* were obtained for Leaders. The *politics* topic has also appeared in an analysis concerning Facebook data on Public sentiment toward economic sanctions in the Russo–Ukrainian war [5]. In contrast, topics like *Russia*, *Estonia*, *combat*, *refugees*, *donations*, *Eurovision*, *nuclear (news)*, *women*, *flag & colors* were obtained for the Public. The following topics have also appeared in the previous research: combat [5,6,13], *refugees* [5,6]. The Public appears to focus more on diverse topics, such as *Eurovision*, *donations*, *nuclear* issues, and *refugees*. In contrast, the Leaders' discussions tend to revolve around more general topics concerning domestic and foreign politics-related themes (see Figure 19).. The Leaders may not express their opinions on all narrower topics, like gathering donations or about the Eurovision song contest.

RQ2 - How important events have influenced the sentiment?

When evaluating the overall sentiment, a clear difference occurred between the Leaders and the Public. The Leaders' posts' sentiments were primarily neutral, and for the Public, the sentiment leaned toward negative (see Figure 6 and Figure 13). Notably, 22% of Leaders' posts were negative, while 49% of the Public's posts were negative (see Table 1). For the Public, this finding aligns with prior research where negative sentiment dominated [6,7]. The negativity in the Public's sentiments might come from the fact that they are not official statements. However, Leaders' statements are official and represent a country's stance, so they have to communicate their thoughts more diplomatically. When considering both the Public's and the Leaders' positive and negative sentiment subsets, there were many similarities in what words made up the positive dataset and which one the negative (see Figure 9 and Figure 16).

To sum up, the overall posting trends for the Leaders and the Public are similar. Both group's activity spikes at the start of the war and then declines until the beginning of summer (see Figure 6 and Figure 13). Through our defined timeline, there are occasional spikes of posts that are related to some of the events happening. At the beginning of the war, many war-defining events happened in a short span of time, in February and March, events like Snake Island, Antonov Airport Battle, the attack on the Mariupol maternity hospital, and the Mariupol Theatre bombing. Therefore, the posting activity was extra high. Significant events have influenced the number of posts generated. Interestingly, there was a significant drop in Leaders' posting activity in July, which could be attributed to Riigikogu having a collective summer vacation.

RQ3 - Are people losing interest in the war?

When considering the number of posts, it could be observed that after the first months, the posting activity of war-related content faded but picked up again once a critical event happened. During the first three months of the war (Feb - Apr), the Public posted 7056 posts, and in the following nine months (May - Jan), they posted 9489 posts related to war. For the Leaders, the post count for the first three months is 1102, and for the following nine months, it is 1808 (see Figure 4 and Figure 5). The fact that the Public did 42% of posts and the Leaders 34% of posts related to the war during the first three months of the war indicates that after the initial shock other everyday topics took over.

For the Public, this result could indicate that the start of war was a significant topic for people, which throw them to share and find information on social media. The subsequent fall could also be attributed to the abundance of information narrowing the collective attention span [45], and soon the discussions on war got less attention.

The analysis of posting activity revealed fluctuations in the amount of war-related content over time. However, a decline in activity was followed by an increase in the number of posts triggered by significant events. The same could be observed by looking at specific keywords (e.g., *war crime*, *sanction*, *mobilization*) or the top three keywords (*Ukraine*, *Russia*, *war*), where the spikes in keyword activity correspond to important events during the war, such as the ending of Mariupol Siege and the Bucha massacres being discovered (see Figure 17 and Figure 18).

Although there was a decrease in posting activity after the start of the war, it still remained a relevant topic that people discussed, with major events creating subsequent spikes, and indicating that the topic remains relevant, and people are still interested in the war.

RQ4 - Are Estonian political leaders (parties) conveying a similar stance on the Ukrainian war in social media?

Most posts are considered neutral when comparing the political parties' sentiments (see Figure 10). Although, we observed that EKRE had the most posts classified as negative (40%) and the least amount as positive (~13%). In contrast, KESK had the most posts classified as positive (40%) and the least amount classified as negative (~13%). This difference in sentiment proportion could be attributed to the different party views and who their voters are.

When comparing parties' posting activity, the least amount of posts about the war was done by E200 and KESK, with 227 and 262 posts, respectively (see Figure 2). REF and ISAMAA did the most amount of posts, with 982 and 762 posts, respectively. The difference in posts can indicate how invested each party is in the Russo-Ukrainian war and how important they believe this topic to be to their voters.

While all the parties generated more posts at the start of the war and subsequently had a decline in war-related posts, we can still see that while some parties activity mainly was constant, then REF, ISAMAA, and SDE had multiple spikes throughout the year, that could indicate that they were voicing their opinion during important events in the war (see Figure 11).

Overall, while some political parties reacted according to the events that happened in the war, others posted more at the beginning of the war and then declined and stayed constant for the rest of the year. The topic modeling results showed that while *war* and *allies* were the most popular topics for all the parties, they still differed regarding other issues (see Figure 12), again indicating that parties have different views and values.

6.1 Limitations

Although the thesis analyzes various topics, the results could be more accurate with a better dataset. Due to the small size of the Estonian population and the activity in Social media, two different platforms were used. In our case, there was no practical way and legal basis to collect public opinion from Facebook, and there were not enough politicians present and posting on Twitter. Using one platform that is used the same way would make the results more comparable. Even though two platforms' data was used to enable the analysis, the amount of data could still be larger since, for example, we only got about 3000 posts from 60 politicians. These improvements could also help in creating and finding more meaningful topics.

It is also worth mentioning that although we consider Twitter data as Public data, we should not consider it an accurate representation of the population as only a tiny percentage of the population is active there. Nevertheless, in the case of this thesis, the focus is on social media analysis, and no further segmentation was done. For collecting regular peoples' opinions, Twitter was the best option before closing their free Academic API.

6.2 Future work

In future analysis, a more extensive dataset could be created. In this thesis, ten politicians were collected from each party. However, in future works, the definition of Leaders could be expanded to more politicians and also to other important thought leaders in society, like journalists and influencers with big followings. The Public's data could also be increased by extending the initial querying keywords. Furthermore, other social platforms could be assessed and considered for data collection (Reddit, Telegram, etc.).

The same approach could be used, for example, to conduct analysis in general on a particular social media platform (like Twitter). In extension, we could also use this process to compare different countries on how their Leaders' and the Public's sentiments and topics differ.

Another way to approach this subject is to consider the information or news fatigue that accompanies social media use. This could further be used to study how the abundance of information affects the people's use of social media and following war-related content.

Conclusion

This thesis explored the Estonian social media perception of the ongoing Russo-Ukrainian war in two groups, specifically the Leaders and the Public. A more novel approach was used by including two different popular social media platforms (Twitter and Facebook). Within the framework of this thesis, politicians were considered Leaders. To our knowledge, this topic has not been analyzed in the Estonian language before.

Two Estonian language datasets related to the war were collected, one for Leaders and one for the Public, consisting of 3223 Facebook posts and 16736 tweets, respectively. The Public and the Leaders were analyzed through approximately a one-year period starting from the war.

This thesis used topic modeling and sentiment analysis to analyze the text data. Topic modeling was used to identify and extract the underlying themes and topics discussed in social media about the war in Ukraine, expressed by both the Leaders and the Public. Sentiment analysis was also carried out to assess the overall sentiment associated with these topics.

The results revealed that the Leaders and the Public were both vocal about expressing their views on topics related to war, energy security, economy, sports, education, and news. We observed a peak in social media posts when specific events happened on the battlefield, for example, Butcha, Mariupol Siege, and mobilization in Russia. We observed that while Leaders were overall talking about more general topics concerning domestic and foreign politics (allies, politics, and defence), the general public had many diverse discussions ranging from Eurovision to nuclear issues and refugees. The findings suggested that the Leaders' posts were primarily neutral, while the sentiment leaned toward negative for the Public.

In conclusion, this thesis has explored Estonian social media in the context of understanding the Russo-Ukrainian war. Analyzing social media data allowed us to focus on two perspectives, political Leaders and the Public. Valuable insights have been gained regarding the important topics to the two groups and how the war and certain events affected their sentiments and posting activity. Although we consider social media a valuable tool for analyzing important topics and sentiments, the small size of Estonian social media activity was a limitation, highlighting the need for further research in larger contexts.

References

[1] Öztürk N, Ayvaz S. Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. Telematics and Informatics 2018;35:136–47. https://doi.org/10.1016/j.tele.2017.10.006.

[2] Yantseva V. Migration Discourse in Sweden: Frames and Sentiments in Mainstream and Social Media. Social Media + Society 2020;6:2056305120981059. https://doi.org/10.1177/2056305120981059.

[3] Melton CA, Olusanya OA, Ammar N, Shaban-Nejad A. Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. Journal of Infection and Public Health 2021;14:1505–12. https://doi.org/10.1016/j.jiph.2021.08.010.

[4] Yin H, Song X, Yang S, Li J. Sentiment analysis and topic modeling for COVID-19 vaccine discussions. World Wide Web 2022;25:1067–83. https://doi.org/10.1007/s11280-022-01029-y.

[5] Ngo VM, Huynh TLD, Nguyen PV, Nguyen HH. Public sentiment towards economic sanctions in the Russia–Ukraine war. Scottish Journal of Political Economy 2022;69:564–73. https://doi.org/10.1111/sjpe.12331.

[6] Maathuis C, Kerkhof I. The first two months in the war in Ukraine through topic modeling and sentiment analysis. Regional Science Policy & Practice 2023;15:56–74. https://doi.org/10.1111/rsp3.12632.

[7] Garcia MB, Cunanan-Yabut A. Public Sentiment and Emotion Analyses of Twitter Data on the 2022 Russian Invasion of Ukraine. 2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), 2022, p. 242–7. https://doi.org/10.1109/ICITACEE55701.2022.9924136.

[8] Lee I. Social media analytics for enterprises: Typology, methods, and processes. Business Horizons 2018;61:199–210. https://doi.org/10.1016/j.bushor.2017.11.002.

[9] Liu B. Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge: Cambridge University Press; 2015. https://doi.org/10.1017/CBO9781139084789.

[10] Wankhade M, Rao ACS, Kulkarni C. A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev 2022;55:5731–80. https://doi.org/10.1007/s10462-022-10144-1.

[11] Schuller B, Mousa AE-D, Vryniotis V. Sentiment analysis and opinion mining: on optimal parameters and performances. WIREs Data Mining and Knowledge Discovery 2015;5:255–63. https://doi.org/10.1002/widm.1159.

[12] Pajupuu H, Altrov R, Pajupuu J. Identifying Polarity in Different Text Types. Folklore: Electronic Journal of Folklore 2016;64:125–42. https://doi.org/10.7592/FEJF2016.64.polarity.

[13] Pung A. Weakly-Supervised Text Classification for Estonian Sentiment Analysis. Master - Computer Science. University of Tartu, 2022.

[14] Uustalu SK. Automated detection and sentiment analysis of registered entity mentions in Estonian language news media. Tallinn University of Technology, 2019.

[15] Jaanimäe G. Eesti Wordnet ja meelestatuse analüüs. Thesis. Tartu Ülikool, 2018.

[16] Vayansky I, Kumar SAP. A review of topic modeling methods. Information Systems 2020;94:101582. https://doi.org/10.1016/j.is.2020.101582.

[17] Mansour S. Social Media Analysis of User's Responses to Terrorism Using Sentiment Analysis and Text Mining. Procedia Computer Science 2018;140:95–103. https://doi.org/10.1016/j.procs.2018.10.297.

[18] Chen S, Mao J, Li G, Ma C, Cao Y. Uncovering sentiment and retweet patterns of disaster-related tweets from a spatiotemporal perspective – A case study of Hurricane Harvey. Telematics and Informatics 2020;47:101326. https://doi.org/10.1016/j.tele.2019.101326.
[19] Han X, Wang J. Using Social Media to Mine and Analyze Public Sentiment during a Disaster: A Case Study of the 2018 Shouguang City Flood in China. ISPRS International Journal of Geo-Information 2019;8:185. https://doi.org/10.3390/ijgi8040185.
[20] Xu J, Qiang Y. Analysing Information Diffusion in Natural Hazards using Retweets - a Case Study of 2018 Winter Storm Diego. Annals of GIS 2022;28:213–27. https://doi.org/10.1080/19475683.2021.1954086.
[21] Ivan Hladkyi. Russian Invasion of Ukraine – Topical Evaluation of World News Sources with Machine Learning. Master - Computer Science. 2022.
[22] Urman A, Makhortykh M. My war is your special operation: Engagement with pro- and anti-regime framing of the war in Ukraine on Russian social media 2022. https://doi.org/10.31219/osf.io/67snk.
[23] Chen E, Ferrara E. Tweets in Time of Conflict: A Public Dataset Tracking the Twitter Discourse on the War Between Ukraine and Russia 2023. https://doi.org/10.48550/arXiv.2203.07488.
[24] Pohl J, Seiler MV, Assenmacher D, Grimme C. A Twitter Streaming Dataset collected before and after the Onset of the War between Russia and Ukraine in 2022 2022. https://doi.org/10.2139/ssrn.4066543.
[25] Shevtsov A, Tzagkarakis C, Antonakaki D, Pratikakis P, Ioannidis S. Twitter Dataset on the Russo-Ukrainian War 2022. https://doi.org/10.48550/arXiv.2204.08530.
[26] Solopova V, Popescu O-I, Benzmüller C, Landgraf T. Automated Multilingual Detection of Pro-Kremlin Propaganda in Newspapers and Telegram Posts. Datenbank Spektrum 2023. https://doi.org/10.1007/s13222-023-00437-2.
[27] Makhortykh M, González Aguilar J. IS IT FINE? INTERNET MEMES AND HATE SPEECH ON TELEGRAM IN RELATION TO RUSSIA’S WAR IN UKRAINE. Hate Speech on Social Media: A Global Approach, 2023.
[28] Hanley HWA, Kumar D, Durumeric Z. “A Special Operation”: A Quantitative Approach to Dissecting and Comparing Different Media Ecosystems’ Coverage of the Russo-Ukrainian War 2022. https://doi.org/10.48550/arXiv.2210.03016.
[29] Oueslati W, Arrami S, Dhouioui Z, Massaabi M. Opinion leaders’ detection in dynamic social networks. Concurrency and Computation: Practice and Experience 2021;33:e5692. https://doi.org/10.1002/cpe.5692.
[30] Aleahmad A, Karisani P, Rahgozar M, Oroumchian F. OLFinder: Finding opinion leaders in online social networks. Journal of Information Science 2016;42:659–74. https://doi.org/10.1177/0165551515605217.
[31] AlKhatib M, El Barachi M, AleAhmad A, Oroumchian F, Shaalan K. A sentiment reporting framework for major city events: Case study on the China-United States trade war. Journal of Cleaner Production 2020;264:121426. https://doi.org/10.1016/j.jclepro.2020.121426.
[32] Tejedor S, Cervi L, Tusa F, Portales M, Zabotina M. Information on the COVID-19 Pandemic in Daily Newspapers’ Front Pages: Case Study of Spain and Italy. International Journal of Environmental Research and Public Health 2020;17:6330. https://doi.org/10.3390/ijerph17176330.
[33] Haman M. The use of Twitter by state leaders and its impact on the public during the COVID-19 pandemic. Heliyon 2020;6. https://doi.org/10.1016/j.heliyon.2020.e05540.

[34] Manfredi-Sánchez J-L, Amado-Suárez A, Waisbord S. Presidential Twitter in the Face of COVID-19: Between Populism and Pop Politics. Comunicar: Media Education Research Journal 2021;29:79–90.
[35] McGregor SC. Social media as public opinion: How journalists use social media to represent public opinion. Journalism 2019;20:1070–86. https://doi.org/10.1177/1464884919845458.
[36] Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure 2022. https://doi.org/10.48550/arXiv.2203.05794.
[37] de Groot M, Aliannejadi M, Haas MR. Experiments on Generalizability of BERTopic on Multi-Domain Short Text 2022. https://doi.org/10.48550/arXiv.2212.08459.
[38] Grootendorst M. Interactive Topic Modeling with BERTopic. Medium 2021. https://towardsdatascience.com/interactive-topic-modeling-with-bertopic-1ea55e7d73d8 (accessed May 7, 2023).
[39] Uiboaed K. Eesti keele stoppsõnad / Estonian stop words 2018. https://doi.org/10.15155/re-48.
[40] Grootendorst MP. c-TF-IDF - BERTopic n.d. https://maartengr.github.io/BERTopic/getting_started/ctfidf/ctfidf.html (accessed May 7, 2023).
[41] Tanvir H, Kittask C, Eiche S, Sirts K. EstBERT: A Pretrained Language-Specific BERT for Estonian. Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa), Reykjavik, Iceland (Online): Linköping University Electronic Press, Sweden; 2021, p. 11–9.
[42] Matthew L. Jockers. mjockers/syuzhet: An R package for the extraction of sentiment and sentiment-based plot arcs from text 2015.
[43] Mohammad SM, Turney PD. Crowdsourcing a Word–Emotion Association Lexicon. Computational Intelligence 2013;29:436–65. https://doi.org/10.1111/j.1467-8640.2012.00460.x.
[44] Gallagher RJ, Reing K, Kale D, Ver Steeg G. Anchored Correlation Explanation: Topic Modeling with Minimal Domain Knowledge. Transactions of the Association for Computational Linguistics 2017;5:529–42. https://doi.org/10.1162/tacl_a_00078.
[45] Lorenz-Spreen P, Mønsted BM, Hövel P, Lehmann S. Accelerating dynamics of collective attention. Nat Commun 2019;10:1759. https://doi.org/10.1038/s41467-019-09311-w.

Appendix

I. Public dataset keywords

Keywords related to	Keywords and combinations (in Estonian)
War	sõda, konflikt, agressioon, ukraina sõda, ukraina sõja, ukraina sõjas, ukraina sõjast, ukraina sõjaks, ukraiana sõjalt, ukraina sõjal, ukraina sõjale,ukraina sõjani, ukraina sõjaga ukraina venemaa, ukraina konflikt, ukraina aggressioon, venemaa agressioon, sõjakuritegu
Ukraine	ukraina, ua, ukrainlane, ukraina inimene, ukraina inimese, ukraina inimeste, ukraina inimestel, ukraina inimestega, ukraina inimestest, ukraina inimest, ukraina inimesele, ukraina inimesena, ukraina inimesega, ukraina putin
Refugees	põgenik, sõjapõgenik, pagulane, migrant, ukraina põgenik, ukraina põgeniku, ukraina põgenikku, ukraina põgenikule, ukraina põgenikust, ukraina põgenikuks, ukraina põgenikuna, ukraina pagulane, ukraina migrant, ukraina sõjapõgenik, ukraina sõjapõgeniku,
Russia	venemaa, agressor, venemaa agressioon, venemaa agressor, venemaa agressiooni, venemaa putin, venemaa sõda, venemaa erioperatsioon, venemaa putini, venemaa putinil, venemaa sanktsioon, venemaa ründama, venemaa ründamine, venemaa agressor, venemaa ründas, venemaa rünnak, venemaa konflikt, putini sõda, sõda putin
Allies	lääs, lääneriigid, nato, ukraina abi, ukraina toetus, ukraina koostöö, ukraina toetamine, ukraina liitlane, ukraina sanktsioon, ukraina ründama, ukraina ründas, ukraina rünnak, ukraina julgeolek, lääs ukraina, immigratsioon ukraina, ukraina agressor, ukraina elu eesti, varjupaiga ukraina, varjupaik ukraina
Estonia	eesti ukraina, eesti venemaa, eesti putin

II. License

Non-exclusive license to reproduce thesis and make thesis public

I, **Lisanna Lehes**,

(*author's name*)

1. herewith grant the University of Tartu a free permit (non-exclusive license) to:
   1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and
   1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

of my thesis

Understanding Public and Leaders' Opinion about Russo-Ukrainian War through Social Media Platforms: An Estonian Case Study,

(title of thesis)

supervised by Somnath Banerjee (PhD) and Rajesh Sharma (PhD),

(supervisor's name)

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive license does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, **09.05.2023**