Ατομική Διπλωματική Εργασία

# War of Words: Analysis of News Media Narratives of the Russo-Ukrainian Conflict

**Daniel Odysseas Katsiolas**

## UNIVERSITY OF CYPRUS

**DEPARTMENT OF COMPUTER SCIENCE**

**May 2024**

# UNIVERSITY OF CYPRUS

## DEPARTMENT OF COMPUTER SCIENCE


**War of Words: Analysis of News Media Narratives of the Russo-Ukrainian Conflict**


**Daniel Odysseas Katsiolas**


Supervising Professor
George Pallis

The Individual Thesis was submitted in partial fulfillment of the requirements for obtaining the Computer Science degree of the Department of Computer Science of the University of Cyprus

May 2024

## Acknowledgements

I would like to acknowledge and express my sincerest gratitude to Mr. Demetres Paschallides, for the excellent help and support that he provided to me through my time working on this thesis. His support and direction helped me greatly in realizing and exploring the variety of tools and methods that I applied to compile this thesis. With his direction and help I was able to utilize tools that I would have found difficult working on them by myself.

I would like to also acknowledge and express my gratitude to the GDELT project, for providing such an excellent and large database of information. Without the GDELT project's work it would have been beyond difficult in providing myself with the data and information that is core to my thesis.

I would like to acknowledge and express my gratitude to the developers of the CADE tool. This exceptional program, which tracks shifts in word usage over time, became a pivotal tool in my study for exploring and analyzing the collected news article data. Its capabilities significantly enhanced the depth and accuracy of my analysis, making it an indispensable component of this research.

## Summary

This thesis started with the express aim to explore and analyze the Russian and Ukrainian media's differences in how they view and present the current and ongoing Russo-Ukrainian war. The data that I have set to analyze and dissect, spans the first year of the conflict that being, from its very begging on the 24th of February 2022 until the 24th of February 2023. The reason why I chose to work on the first year is due to the practicality of working on such a large wealth of data. But at the same time the first year of the Russo-Ukrainian war stands as a very interesting period in our world history, given that throughout the first year alone, a plethora of world changing events occurred and a great deal of shifts and adaptations in the world stage happened within the first year of conflict.

To explore these media and narrative differences, I utilized and created a variety of programs and experiments, mostly belonging to the Natural Language Process field. The data that I gathered from my experiments, shows the differences in word contextual usage between the Russian and Ukrainian news media. The sentiment framing between particular entity mentions from both the Russian and Ukrainian news media. The temporal shifts in sentiment of the Russian and Ukrainian news media as whole and through particular entity mentions from both states' media over time. All these experiments, show an evident difference in how both sides of this conflict adapt to, present and confront, the reality of a major war that is happening between them.

# Contents

# Chapter 1

## Introduction



### 1.1 Background on the Russo-Ukraine war

The topic that this thesis covers is the Russo-Ukraine war. Currently one of the largest land wars that is being conducted in recent memory. The reasons for the start of the war are plenty and multifaceted. Leading Ukrainian and Russian sources differ greatly in the reason for the start of the war. Many can agree on the fact that this war has been simmering for a long time. Ever since the pro-Western Euromaidan protests that happened in Ukraine in 2014, which ousted then president of Ukraine Viktor Yanukovich[1]. This led to counterprotests by more Russian aligned Ukrainians, especially in Crimea. These events then gave way for Russia to invade the Crimea peninsula and annex it. This by many is considered, the true beginning of the Russo-Ukraine war.

Eventually the Ukrainians would form a more pro-European and pro-Western government led by Petro Poroshenko. His presidency though proved tumultuous, especially for many Russian aligned Ukrainians in the eastern regions of Ukraine, those being Donetsk and Luhansk. Many of these Ukrainians would then start their own protests in their regions demanding greater autonomy from Ukraine, until eventually in 2015 a proper insurrection would commence with many of these Ukrainians seizing their regional capitals and declaring regional autonomy or for the federalization of Ukraine[3]. Later, sources argue that Russia supported these separatist militias covertly, but Russia downplays its support[4]. These events would start the Ukrainian civil war, which would last for 8 years.

With the Ukrainian civil war on going, there would be a back and forth between Europe, Ukraine and Russia, those being the two Minsk Agreements which were attempts at resolving the Donetsk and Luhansk issues peacefully. But no headway was ever reached between the separatists and the Ukrainian government. As the years went by, the civil war simmered and hostility between Ukraine and Russia increased. In the year preceding the current war the question of Ukrainian ascension into NATO came into play[5], which can be considered one of the reasons Russia began its war.

The overarching reason for this war has many interpretations from both the Ukrainian and Russian sides. The Ukrainians believe that the overarching reason is due to what they perceive to be modern Russia's attempt at imperial expansion in a country that they consider theirs. The Russians believe the overarching reason for this war is due to Western and NATO expansion in what they perceive to be their country's strategic backyard[2].

### 1.2 Importance of news media shaping public perception

The news media, in all its forms, plays a crucial role in shaping and defining public perception of various topics and events. Every news outlet engages in narrative and information shaping to some extent. This tendency is inherent in media that is guided or controlled by specific interests. State-run media typically present the government's viewpoint, aligning their coverage with official perspectives. Conversely, private and independent media outlets often convey information through the lens of their political affiliations or the interests of their primary supporters.

The accessibility and reach of these media sources significantly influence how populations perceive current events and reality. For many individuals, news media is the primary source of information about the world. Therefore, understanding how media outlets present and shape information to align with their desired narratives is essential.

This dynamic is particularly evident in the context of the Russo-Ukrainian war, making it an ideal case study for examining how governments and media handle contentious issues such as war. The differing portrayals and narratives offer valuable insights into the power of media in framing public perception.

### 1.3 Objective of the Study

The objective of this thesis is to conduct an analysis of the differences in narrative and information presentation regarding the Russo-Ukrainian war between the two warring parties: Ukraine and the Russian Federation. By examining how each side portrays the conflict, this study aims to uncover the distinct methods and strategies employed in their media coverage. The analysis will explore the linguistic choices and framing techniques used by both Ukrainian and Russian media to shape public perception and influence international opinion. Through this investigation, the thesis seeks to provide deeper insights into the role of media in wartime propaganda and its impact on the broader information landscape.

### 1.4 Overview of Methodology

To achieve the objective of this thesis, a range of custom-built and pre-existing programming tools and techniques in the field of Natural Language Processing (NLP) were employed. NLP was chosen as the primary analytical framework because it provides the most effective means of processing large volumes of textual data and extracting meaningful insights. By leveraging NLP, this study can systematically analyze extensive corpora of media content from both Ukraine and the Russian Federation.

The methodologies used include advanced NLP models and techniques such as BERT (Bidirectional Encoder Representations from Transformers), which excels in understanding context and semantics in text. CADE (Compass-aligned Distributional Embeddings) was utilized to detect and analyze shifts in word usage and meaning over time, enabling a diachronic study of narrative changes. Additionally, Named Entity Recognition (NER) was applied to identify and categorize key entities mentioned in the texts, facilitating a deeper understanding of how different actors and elements are framed by the media on both sides.

These tools and methods were chosen for their ability to handle complex language patterns and provide nuanced analyses of the data, ultimately helping to reveal the divergent narrative strategies employed by Ukrainian and Russian media during the Russo-Ukrainian war.

# Chapter 2

## Related Works



### 2.1 Embeddings and Word2Vec

In this section, I will discuss embeddings and Word2Vec, two concepts that form the basis to my word and sentiment analysis of the news data I gathered.

Embeddings in Natural Language Processing (NLP) are a type of representation that converts high-dimensional categorical data, such as words, into vectors of real numbers in a generally lower-dimensional space. The goal of embeddings is to capture the semantic properties of the data so that similar data objects are represented by similar vectors. This similarity is usually determined by the context in which the data objects appear. In NLP, word embeddings transform words into dense vectors where each word is represented by a single vector in a predefined vector space.[6]

Word embeddings are learned from provided data and can capture various linguistic patterns. Semantic similarity ensures that words with similar meanings are closer in the vector space. Syntactic similarity groups words that share syntactic properties together.

Word2Vec is a popular and effective method for learning word embeddings from a large corpus of text. Developed by a team led by Tomas Mikolov at Google, Word2Vec offers two main architectures to compute vector representations of words. The Continuous Bag-of-Words (CBOW) model predicts a word given its context, which may be a single word or a group of words. The goal is to predict the probability of a word given the surrounding words, with the model learning embeddings that help it predict the target words from the context words. The Skip-Gram model works in the opposite way to CBOW. Given a target word, it predicts the surrounding context words. This model is particularly good at capturing a wide range of semantic and syntactic word relationships and tends to perform well on larger datasets.[6][7]

Both of these models use a simple neural network with one hidden layer, rather than deep learning, and they train using one of two methods. Hierarchical Softmax is an efficient way to compute the output from the softmax layer, while Negative Sampling simplifies the optimization problem for larger datasets by randomly sampling negative examples.[7]

## 2.2 Clustering

Clustering is another crucial concept that significantly aided my sentiment analysis. It is an unsupervised learning method used in machine learning and data analysis to group a set of objects or points such that objects or points within the same group, or “cluster”, are more similar to each other than to those in other groups, based on some measure, usually some form of distance. And so, this method finds and groups similar objects, thereby facilitating a better understanding of the data's structure without the need for prior labeling of groups. By employing clustering, I was able to reveal some patterns and relationships within the data, enhancing the accuracy and depth of my analysis.

Clustering has wide variety of applications ranging from statistical data analysis to pattern recognition, image analysis, information retrieval, and bioinformatics. Some common algorithms for clustering include:

- **k-means clustering**: Data is divided into k clusters, where each data point belongs to the cluster, whose centroid or some other representative point is closest to the data point.
- **Hierarchical clustering**: Is a family of clustering algorithms, that either have a bottom up approach (Agglomerative), or a top down approach (Divisive) in grouping objects or points.

In NLP, clustering is used to discover structures within text data, which can involve grouping similar words, documents, or other linguistic objects. For my analysis, I emphasized on word clustering.

Word clustering is all about grouping similar or related words based on semantics or syntactic properties. Word embedding techniques like Word2Vec, which I described in 2.1 transform words into continuous vectors space where semantically similar words are mapped to approximate points. We can then apply a clustering algorithm to these vectors to identify groups of synonyms or words that share similar contexts.

### 2.3 Other methods and tools

Other methods and tools that significantly contributed to my analysis included Named Entity Recognition (NER) and Compass-Aligned Distributional Embeddings (CADE). NER was instrumental in my effort to identify and classify some key entities within the text, which I later used for my semantic analysis. CADE, which employs Word2Vec to generate Temporal Word Embeddings with a Compass (TWEC), is the main modelling algorithm which I used for analyzing changes in word usage over time and between two distinct data sets. Together, these tools provided me with a strong framework for extracting and interpreting complex patterns from the data, thereby enriching the overall quality of my research.

Named Entity Recognition (NER) is a type of information extraction task, which identifies and classifies named entities in text into predefined categories. Some categories which are often present in NER platforms are names of persons, organizations, locations, expressions of time, geopolitical entities, products, etc.

NER is crucial for understanding the content of a document or a set of textual data by identifying key elements. It is widely used in various natural language processing applications such as machine translation, question answering, information retrieval, knowledge graphs, content recommendation, and in my case, I used it for sentiment analysis.

The NER system used in my analysis was SpaCy's implementation. SpaCy tokenizes the text in a non-destructive manner, preserving the original text and whitespace, which allows for accurate tokenization without losing any information. SpaCy then utilizes a statistical model built on the "thinc" library, which is optimized for deep learning tasks in NLP. This model has been trained on large, annotated corpora, enabling it to identify entities within the text efficiently. Once the entities are identified, SpaCy creates spans for these entities. Spans are sequences of tokens that together form the named entity.[8]

NER is overall a very important NLP technique that helped in structuring text data by identifying key elements and categorizing them into a set of predefined classes, which enabled me to do deeper text analysis.

CADE (Compass-Aligned Distributional Embeddings) is a method designed to address the challenge of semantic change over time in word embeddings. It was specifically developed to handle scenarios where the meaning of words may shift across different corpora, which might represent different time periods, genres, or domains. This technique is particularly useful in diachronic (historical) linguistics, where understanding how the meanings of words evolve over time is crucial.

CADE operates by aligning word embeddings from different corpora to a common space. The fundamental idea is to train word embeddings on separate corpora (e.g., texts from different time periods) and then align these embeddings in such a way that they are directly comparable. This alignment process is what CADE emphasizes, and it is referred to as "compass alignment."[9]

Here I will show a step by step breakdown on how CADE generally operates[9]:

1. **Embedding Generation**: First, generate word embeddings separately for each corpus using standard techniques like Word2Vec, GloVe, or FastText. Each corpus represents a different "slice" of data, potentially from different times or contexts.
2. **Compass Construction**: Identify or create a "compass" embedding space. This space acts as a reference or guide. Typically, the compass is built from either

one of the corpora or an amalgamation of multiple corpora. The idea is that the compass should represent a reliable and stable linguistic space against which other embeddings can be compared.

3. **Alignment Process**: Align the embeddings from each corpus to the compass embeddings. This involves transforming the embeddings from each corpus so that their geometric properties (such as distances and angles between vectors) match as closely as possible with those in the compass embedding. Techniques such as orthogonal Procrustes, which is often used for its ability to preserve the linguistic properties of the embeddings, can be employed here.
4. **Semantic Change Detection**: After alignment, the embeddings from different corpora can be directly compared to study semantic changes. For example, by observing how the position of a word's vector shifts over time or across corpora, with this one can infer changes in meaning, usage, or association.

CADE is particularly useful in fields where understanding the evolution and use of language is important. Like in Historical Linguistics, which is all about analyzing how the meanings of words have changed over historical periods. Cultural Studies which study the shifts in language that reflect cultural changes. And Social Sciences which is about analyzing communication patterns and changes in discourse over time.

CADE's implementation and design also overall carry plenty of advantages in analyzing the differences that two large textual data sets may have, these advantages are:

- **Robustness to Semantic Shift**: CADE is designed to robustly handle semantic shifts in language, making it a powerful tool for diachronic linguistic analysis.
- **Inter-Corpus Comparability**: By aligning embeddings to a common compass, CADE facilitates meaningful comparisons across different datasets or corpora.

CADE is an advanced method in the domain of NLP and computational linguistics, providing a nuanced approach to understanding how languages and meanings evolve over time. It stands as a testament to the adaptability of embedding techniques to specialized linguistic tasks, particularly those involving changes across different temporal and contextual datasets.

# Chapter 3

## Methodology



### 3.1 Selection of news sources

Initially, news selection was intended to be done automatically using a program with a built-in system for keyword extraction. This system employed BERT and an arbitrary textual source, which in my case was the Russo-Ukrainian War Wikipedia page. The program also required a country code to determine where to look for news data based on the generated keywords. However, this approach encountered significant issues. The major problem was that the program could not effectively gather links for news sites based on the country code, resulting in data primarily from third-party countries reporting on the war.

To overcome these challenges, I developed a different approach to data collection with substantial external assistance. This new method involved manually gathering a comprehensive collection of news sources from both Russia and Ukraine. I conducted an exhaustive search of the Internet to compile a variety of news sources and their links. After compiling this list, I cleaned up the URLs to their domain names. Then, I edited the Python code of my news data collection program to search for and source data exclusively from the curated list of links.

This revised approach proved to be highly effective, enabling the news data collection program to gather a substantial amount of news data from multiple sites on both sides of the war. This comprehensive dataset provided a robust foundation for my analysis.

### 3.2 Data collection

Data collection was carried out using a Python script, which I sourced from POLAR, a framework designed for modeling polarization and identifying polarizing topics in news articles. As detailed in section 3.1, I modified this script to better suit my data collection needs.

The script functions as an advanced news corpus collector, integrating with the GDELT database to retrieve, process, and analyze news articles over a specified date range based on specific criteria, including domain names. This integration allowed for efficient and targeted data collection, ensuring the relevance and accuracy of the gathered news articles for my research.

**Functionalities.**

Here I will list most of the important functionalities and their method names, explaining their use and function.

1. **Initialization:**
    - **NewsCorpusCollector**: This part initializes the script with parameters, such as output directory, date range (from and to) and a list of keywords. The keyword list in my modification stands unused due to me providing the script with a domain name list in its **collect_articles** method.
2. **GDELT Data Retrieval:**
    - **collect_archives**: This method downloads GDELT data files for the specified date range. The method also checks for existing files to avoid re-downloading and handles exceptions if download issues occur.
3. **Article Collection and Processing:**
    - **collect_articles**: Under its normal implementation it filters and collects articles based on actor country codes. But my variation of the script uses

a domain list to filter the downloaded GDELT CSV files. It processes the data frame to filter relevant articles, fetches their HTML content and performs further processing.

4. **Article Parsing and Storage:**
    - **retrieve_articles, pre_process_articles,** and related methods handle the downloading, parsing and preprocessing of articles. The articles are stored in structured JSON format.
    - **parse_html**: This method parses downloaded HTML files into structured data including the article text, title, publication date and images. It then saves this data in a JSON format for easier access and further processing.
5. **Utility Functions:**
    - There is also a variety of utility functions to aid in the processing of URLs and text data. These include methods for cleaning text, formatting titles and handling special characters.

Some key components and methods that the script uses, to better the process of downloading and storing the data are:

- **Parallel Processing**: The script utilizes Python's 'multiprocessing' to handle concurrent downloads and processing tasks to enhance performance.
- **Error Handling**: The script uses an extensive amount o 'try-except' blocks to manage and log errors during downloads and file operations, ensuring the program's robustness against common issues like network errors or missing data.
- **File Handling**: Data is saved and organized into structured directories for dumps, articles and processing outputs. It ensures data is easily accessible and systemically stored.

The only major change I made to the code was modifying the script input to utilize a domain name list, targeting the specific news sites from which I wanted to collect data.

I chose this script because it follows a structured approach, incorporating error handling and parallel processing, which ensures it can efficiently handle large volumes of data. This made it a valuable component in my data pipeline for analysis. Additionally, the script's ability to specify a date range was particularly helpful, as my thesis focuses on the entire first year of the Russo-Ukrainian war.

In describing the news data collection script, I mentioned the GDELT Project, the primary database used by the script to gather news data. GDELT, which stands for Global Database of Events, Language, and Tone, is an ambitious initiative aimed at constructing a comprehensive global database of societal-scale behaviors and beliefs across all countries. It is designed to monitor the world's broadcast, print, and web news from nearly every corner of the globe. GDELT provides structured data that reflects the complex interplay of global society and its myriad components. Its ability to parse and analyze vast quantities of information makes it an invaluable tool for accessing a wealth of collected data. This comprehensiveness and scale are why I chose GDELT as my data source, as it is one of the few large, major news data repositories available.

### 3.3 Experiment Implementation Whole Year

In section 2.3, I explained what CADE is and how it works. Here, I will describe the steps I took to preprocess the data and utilize it for analyzing a year's worth of news articles.

To preprocess the raw data gathered from the news data collection script, I created a Python script with the following steps:

1. **Collects JSON data:**
    - I built a script to retrieve the stored article JSON files. These JSON files were organized by date in subdirectories, facilitating the preprocessing of news data for specific time periods.
    - The script reads and appends the text sections of the stored article JSON files into a single string variable, as CADE requires single-line text files for model training.

2. **Text cleanup:**
    - **Remove URLs**: Initial tests revealed URL paths within the cleaned text, so I implemented a method to remove any URLs present.
    - **Remove XPaths**: Similar to URLs, XPaths appeared in the cleaned text and so I created a method to eliminate these structures.
    - **Removed Numbers and Symbols**: I added a text-cleaning step to eliminate stray numbers or special symbols.
    - **Preserving Text-Number Pairs**: While most numbers were removed, some, like weapon system names, needed to be retained. I developed a method to preserve text followed by numbers.
    - **Lowercasing, Uncontracting and Stop Word removal**: These steps were performed using simple Python scripts. Stop word removal was necessary as these common words provide no valuable information. Uncontracting words, such as converting "can't" to "cannot," was adopted from the news data collection script.
    - **Tokenization**: As a final measure, I tokenized the cleaned text to ensure it was converted into a single-line string with spaces separating the words, then saved it into a .txt file.

3. **Filtering:**
    - The filtering step involved checking if the name of the saved article JSON contained any unwanted keywords. This step ensured that articles from non-relevant news sites or opposition media were excluded, maintaining the purity of the text data in terms of their opinion and attitude towards the war.

By following these steps, I ensured that the data fed into CADE was properly cleaned, filtered, and formatted, which was crucial for accurate and meaningful analysis.

**CADE Experiments Implementation:**

Implementing CADE was straightforward, as I simply loaded its code into my runtime environment, allowing me to utilize its methods to build my models. However, due to recent Python updates and the fact that CADE had not been updated for three years, I encountered some package compatibility issues, which I quickly resolved.

To build my CADE models, I began by loading the pre-processed text data from both sides of the war. I concatenated these files into a single compass.txt file, which CADE uses extensively to create its models. I then trained CADE with the compass file, setting the dimensionality to $n = 64$. To create the models, I used two separate training slices from the cleaned text files of both sides of the war and then loaded both models with Word2Vec. This process resulted in two models corresponding to each side of the conflict.

With the CADE models completed, I turned to the initially overlooked keyword extractor from the news data collection script. This extractor utilizes BERT, a powerful model that processes words in relation to all other words in a sentence, thereby understanding the context of each word. I provided the keyword extractor with the Wikipedia page of the Russo-Ukrainian War as input data.

Using the keywords generated by the extractor, along with some words I manually selected for their potential in providing interesting results, I created a .txt file containing a large collection of words.

I then used these selected words for two experiments:

1. **Word Similarity Analysis:**
   Using Word2Vec's most_similar method, I gathered a large set of words that are "similar" to my chosen words. In this context, "similar" refers to words that have a close usage in the text on which the models were trained, rather than conventional similarity.

2. **Cosine Similarity Heatmap:**
   By mapping the chosen words in the vector spaces of both the Russian and Ukrainian models, I created a heatmap using the matplotlib and seaborn libraries. This heatmap visualizes the cosine similarity of the words between the two models, providing insights into the contextual differences and similarities in word usage between the two sides.

These steps allowed me to effectively utilize CADE and other tools to analyze the linguistic nuances and narrative differences in the media coverage from both Russia and Ukraine.

**Named Entity Framing Experiments Implementation:**

For the implementation of Named Entity Framing experiments, I had to revisit my preprocessing code and remove the conventional cleanup methods. This step was necessary to retain the news article data in its proper format for Named Entity Recognition (NER).

First, I modified the original preprocessing code to load the text portion of the articles from the JSON files into a string variable. This string variable was then processed using SpaCy's Named Entity Recognizer, which tagged the entire text of the articles. After tagging, I filtered out any entity mentions that had unwanted Named Entity labels, appending only the relevant entity mentions to a new list.

Next, I fed this list of entity mentions into an SBERT (Sentence BERT) model, a variation of BERT that specializes in sentence embeddings. The SBERT model generated embeddings for each entity mention.

These embeddings were then processed using a powerful clustering algorithm provided to me. This algorithm leverages parallel processing by splitting tasks into smaller chunks, thereby reducing overall computation time and efficiently handling large datasets.

With the clustered entity mentions, I ran a script to decompose each entity mention into its base words. I then used the CADE models to find the 500 most similar words for these base words. Subsequently, I filtered these 500 words to retain only adjectives, identified using SpaCy’s Part-of-Speech tagger. This process resulted in a Python dictionary containing each entity mention cluster and its top adjectives.

Next, I used TextBlob’s sentiment scorer to evaluate the sentiment of the adjectives. The sentiment scores were added to the Python dictionaries, which differentiated between positively and negatively scored adjectives.

As a result, I had two sets of files, one for Ukraine and one for Russia, each containing clustered entity mentions along with their adjectives and sentiment scores. By comparing these clusters, I identified similar entity mentions in both countries’ datasets. Finally, I visualized these comparisons using horizontal bar plots created with matplotlib, illustrating the sentiment scores for the corresponding entity mentions in both datasets.

This comprehensive approach allowed for a detailed analysis of how different entities are framed sentimentally in the media of Ukraine and Russia.

### 3.4 Experiment Implementation Every Month

Another set of experiments I conducted focused on exploring the temporal shift in sentiment for entity mentions.

The first step in analyzing temporal shifts was to modify my existing preprocessing algorithm. While the algorithm already had a date filter, I needed an automated way to gather and separate cleaned text files for each month for both countries' gathered article JSONs. To achieve this, I created a new loop that iterated based on a list containing the year and month in the format "202202, 202203," corresponding to the directories where my article JSONs were saved (e.g., "20220224"). The preprocessing script would collect the article text data for each day of the month, apply the existing cleanup methods, and append the cleaned text data. Before moving to the next month, the script saved the cleaned text for the month. This process was repeated for both countries.

This approach resulted in 12 text files containing monthly data for each country. Using these 12 sets, I created 12 compasses for CADE through simple loop concatenation.

Next, I adapted my existing NER clustering script to process data on a monthly basis. The modified script looped through each month, gathering daily text data and performing NER mapping for that month. The text was fed into the SBERT model, embeddings were generated, and the entity mentions were clustered. This resulted in 12 files containing different entity mention clusters for each month, for both Russia and Ukraine.

With these 12 files containing entity mention clusters, I loaded them into a script designed to build a separate CADE model for each month based on the compass and text files created earlier. Using these monthly CADE models, I found the top 500 most similar words with Word2Vec and filtered them to retain only adjectives using SpaCy's Part-of-Speech tagger. This script generated 12 files for each month for both countries, containing entity mention clusters and the adjectives similar to the words in those mentions.

These files were then processed through another script that looped through each month, scoring the adjectives of the clusters using TextBlob's sentiment scorer. Each month's entity mention clusters were saved with their adjective sentiment scores, separated into positive and negative categories. This resulted in another set of 12 files containing sentiment scores.

To facilitate analysis, I aggregated the positive and negative sentiments for each entity mention cluster for every month. This allowed me to plot the shifting sentiment of particular entity mention clusters over time using line plots that showed changes in sentiment each month.

Additionally, I created another set of files for each month that contained sentiment data built from the adjectives gathered from the initial CADE model trained on the entire year's data. This variation allowed me to illustrate the total sentiment shift for both countries throughout each month.

Finally, I conducted an experiment where I built 12 separate CADE models for both Ukraine and Russia. I then selected specific words to plot their cosine similarity across the 12 sets of Ukrainian and Russian CADE models. This experiment enabled me to visualize the change in cosine similarity of select words, highlighting the differences in word usage between the Ukrainian and Russian models over time.

# Chapter 4

## Narrative Comparisons and Entity Framing



### 4.1 Word contextual usage between Russian and Ukrainian news media

In this subchapter, I will analyze the two types of tests I conducted. The first test uses Word2Vec and CADE to examine word contextual usage. This test takes a list of words and generates the top 20 words that have a similar or closely related context within the model. Both the Ukrainian and Russian models were tested using the same list of words to identify any differences in high-proximity words between the models.

Through this test, I aimed to observe how Russia and Ukraine might treat different topics or words:

| Word: mig29 | |
|---|---|
| **UKR** | **RUS** |
| **f16: 0.89** | **jets: 0.87** |
| **fighter: 0.86** | **f16: 0.83** |
| **multirole: 0.80** | **downed: 0.67** |

Table 1: Shows the word mig29 and
some “similar” words with their distance.

Table 1 reveals a difference in the presentation of the "Mig29," the primary fighter jet in Ukraine's air force inventory. For Ukraine (indicated by the UKR label), the top similar words are “multirole” and “fighter,” reflecting the Ukrainian media’s portrayal of the aircraft's versatile combat roles. Additionally, "f16" appears as a top word, which correlates with Ukraine’s efforts in the latter half of 2022 to urge its Western allies for the supply of F-16 fighter jets.[10] This push is driven by Ukraine’s limited number of Mig29s and its comparative disadvantage in air power against Russia.

On the Russian side (indicated by the RUS label), the top words are somewhat similar to those identified in the Ukrainian context, such as “f16” and “jets.” This similarity suggests that Russian media also reported on Ukraine's efforts to modernize its air fleet. However, the word “downed” appears, indicating a different focus in Russian reports, which highlight instances of Ukrainian Mig29s being shot down.

| **Word: azov** | |
|---|---|
| **UKR** | **RUS** |
| **regiment: 0.89** | **battalion: 0.91** |
| **marines: 0.77** | **aidar: 0.84** |
| **sea: 0.65** | **nationalist: 0.73** |
| **palamar: 0.69** | **nazis: 0.66** |

Table 2: Shows the word azov and
some “similar” words with their distance.

Table 2 reveals the top words associated with "azov" from both sides of the conflict. The word "Azov" is commonly known as the name of the Azov Sea, a small shallow sea connected to the Black Sea, situated between Russia and Ukraine, and currently fully occupied by Russia. However, in both the Russian and Ukrainian models, the top similar words suggest a combat unit, reflecting the infamous Azov Brigade.

The Azov Brigade, widely recognized in the West as predominantly composed of Ukrainian nationalists, played a significant role in combat operations against Russian separatist militias in the Donbass region and later against the Russian army.[11]

As shown in Table 2, the Ukrainian model predominantly includes objective terms such as "regiment" and "marines," which likely refer to the 36th Marine Brigade of the Armed Forces of Ukraine, a unit that fought alongside the Azov Brigade during the

Mariupol siege.[12] Additionally, the word "palamar" appears, referring to Sviatoslav Palamar, the deputy commander of the Azov Brigade at the time.[13] The presence of the word "sea" also reflects the geographical aspect of the Azov Sea and probable references to the sea’s full occupation by Russia. This suggests that Ukrainian media generally maintain a neutral stance towards the word "azov," focusing on reporting the brigade’s defensive efforts during the Mariupol siege.

In contrast, the Russian model includes more loaded terms such as "nationalist" and "nazis." This reflects the Russian media’s emphasis on the neo-Nazi history of the Azov Brigade and its nationalistic character. The appearance of the word "aidar" refers to the Aidar Brigade, another unit with a similar nationalistic reputation, frequently reported negatively by Russian media. This negative framing underscores the Russian media's focus on the nationalist elements of these brigades and their long-standing involvement in fighting the Donbass separatists, who are regarded as Russian proxies.

This analysis highlights the distinct narrative strategies employed by Ukrainian and Russian media in framing the Azov Brigade, reflecting broader geopolitical narratives and propaganda efforts.

| **Words: donetsk** | |
|---|---|
| **UKR** | **RUS** |
| **luhansk: 0.74** | **lugansk: 0.82** |
| **kherson: 0.73** | **dpr: 0.77** |
| **kreminna: 0.64** | **shelled: 0.73** |
| **severodonetsk: 0.63** | **donbass: 0.68** |

Table 3: Shows the word donetsk and
some “similar” words with their distance.

Table 3 highlights the top similar words for "donetsk." Donetsk can refer to both Donetsk city and the Donetsk region, one of the breakaway regions of Ukraine that established militias and fought against the Ukrainian military for years before the current conflict.[15]

From Table 3, we also observe a clear difference in the top words associated with "donetsk" between the Ukrainian and Russian models.

In the Russian model, the top similar words include "lugansk," "dpr," "shelled," and "donbass." Lugansk is another city and region that rebelled against Ukraine. "DPR" stands for Donetsk People's Republic, which is how Russian media refer to the Donetsk region. The DPR declared itself an autonomous republic at the beginning of the Ukrainian civil war[14], with Russia recognizing its independence when the 2022 war began. The word "shelled" reflects the Russian media's frequent reports of Ukrainian army shelling of Donetsk city, a claim that Ukraine denies. "Donbass" is the overarching geographical name encompassing both Donetsk and Lugansk.

In contrast, the Ukrainian model primarily features names of other locations such as "luhansk," "kherson," "kreminna," and "severodonetsk." These are regions or towns near the frontline where significant fighting occurred. This suggests that Ukrainian media report on Donetsk as one of many contested areas in the ongoing conflict, without giving it special emphasis. This contrasts with the Russian media's treatment of the Donbass region, which is highlighted more prominently. Russian media likely underscore Donetsk and Lugansk to emphasize their narrative of liberating regions friendly to Russia from a nationalistic regime in Ukraine.

The differing spellings of Lugansk (Russian) and Luhansk (Ukrainian) also highlight a linguistic divergence. Following the onset of the war, Ukraine and Western media began rejecting the Russian spellings of many Ukrainian place names, opting instead for Ukrainian transliterations. This linguistic shift underscores the broader cultural and political separation between the narratives presented by Russian and Ukrainian media.

| Words: liberate | |
|---|---|
| **UKR** | **RUS** |
| **liberation: 0.76** | **denazify: 0.77** |
| **occupy: 0.74** | **occupying: 0.76** |
| **retake: 0.69** | **demilitarize: 0.75** |
| **regain: 0.68** | **donbass: 0.71** |

Table 4: Shows the word liberate and
some “similar” words with their distance.

Table 4 reveals significant differences in how the word "liberate" is likely framed in the media of both countries.

In the Ukrainian model, the top similar words to "liberate" include "liberation," "occupy," "retake," and "regain." The word "occupy" likely appears due to media reports referring to the act of retaking lost territory as occupying that territory. The presence of "liberation," "retake," and "regain" suggests that in Ukrainian media, "liberate" is closely associated with the process of reclaiming territory lost to Russia. This framing reflects Ukraine's view of liberation as synonymous with recovering its occupied land.

In contrast, the Russian model links "liberate" with more propagandistic terms such as "denazify" and "demilitarize." These terms reflect Russia's stated objectives in its conflict with Ukraine. The Russian media frequently describe the war as a special military operation aimed at protecting Russian minorities in Ukraine and demilitarizing a government they consider illegitimately established by a nationalist coup d'état. Additionally, "occupying" and "donbass" appear, indicating a focus on territorial control and the strategic importance of the Donbass region.

Here we can see the divergent narratives in Ukrainian and Russian media. Ukrainian media frame liberation in terms of reclaiming territory and national sovereignty, while Russian media frame it within the context of broader ideological and strategic goals.

| Words: aggressive | |
|---|---|
| **UKR** | **RUS** |
| **waging: 0.74** | **destructive: 0.78** |
| **justify: 0.72** | **irresponsible: 0.77** |
| **nature: 0.70** | **rhetoric: 0.75** |
| **hybrid: 0.68** | **refrain: 0.72** |

Table 5: Shows the word aggressive and
some “similar” words with their distance.

Table 5 highlights the stark differences in how both sides perceive and describe each other's actions in the war.

In the Ukrainian model, "aggressive" is associated with words such as "waging," "justify," "nature," and "hybrid." These associations suggest that Ukrainian media characterize Russia's actions as aggressive by nature and frequently justify their aggression. The term "waging" reflects the active conduct of war, while "justify" indicates the narrative that Russia attempts to rationalize its aggressive actions. "Nature" suggests an inherent aggressiveness attributed to Russia. The term "hybrid" likely refers to the concept of hybrid warfare, which Ukrainian media use to describe a broader and more insidious form of aggression by Russia, encompassing not just military actions but also cyber attacks and misinformation campaigns.[16]

In contrast, the Russian model links "aggressive" with terms like "destructive," "irresponsible," "rhetoric," and "refrain." The words "destructive" and "irresponsible" suggest that Russian media portray both Ukraine and Western actions during the conflict as aggressive and harmful. "Rhetoric" likely refers to what Russia perceives as aggressive language and positions from Western and European nations regarding the war. The word "refrain" indicates warnings from Russia, urging Ukraine or the West to avoid aggressive actions.

This analysis underscores the divergent narratives presented by Ukrainian and Russian media. Ukrainian media frame Russia's actions as inherently aggressive and unjustified, emphasizing a broader hybrid war. Meanwhile, Russian media depict the actions of Ukraine and its allies as the true sources of aggression, highlighting their destructive and irresponsible nature while criticizing the aggressive rhetoric of the West.

The second set of tests are heatmaps built from word matrices, which show the similarity in word usage between the Russian and Ukrainian CADE models. My analysis will emphasize the heatmap diagonals, as they indicate how similarly certain words are used between the models.

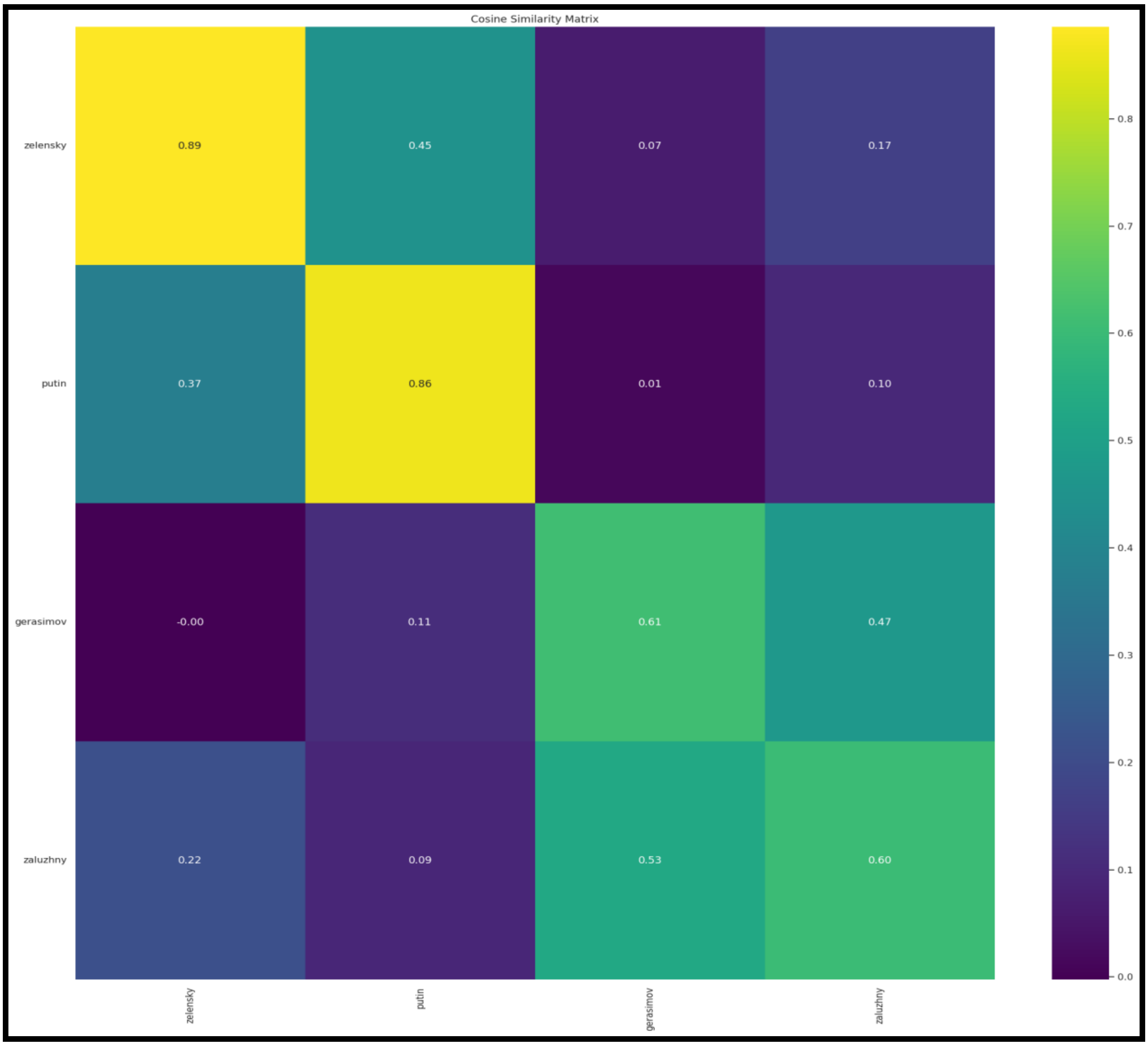


Figure 1: HeatMap showing how similarly the words “zelensky”, “putin”, “gerasimov” and “zaluzhny” are used between the Russian and Ukrainian models.

In Figure 1, we observe that the two models exhibit high similarity in the usage of the presidents' names, Volodymyr Zelensky and Vladimir Putin. This is likely because both names are central to media reporting in each country, leading to their vectors being mapped closely in the vector space.

However, the names "zaluzhny" and "gerasimov," referring to the highest-ranking military commanders of Ukraine and Russia respectively, show  a much lower similarity. This discrepancy likely reflects differing attitudes and emphasis in media coverage. If both Russian and Ukrainian media held similarly negative or positive attitudes towards their respective commanders, the similarity would likely be higher. The lower similarity suggests that Russian media may adopt a more reserved stance when discussing Ukrainian military leaders, while Ukrainian media, driven by the need for international support and being the defending nation, likely adopts a more negative tone towards Russian military leaders.

With the above analysis one can see how media narratives can shape the contextual usage of key figures in the conflict, reflecting broader national perspectives and propaganda efforts.

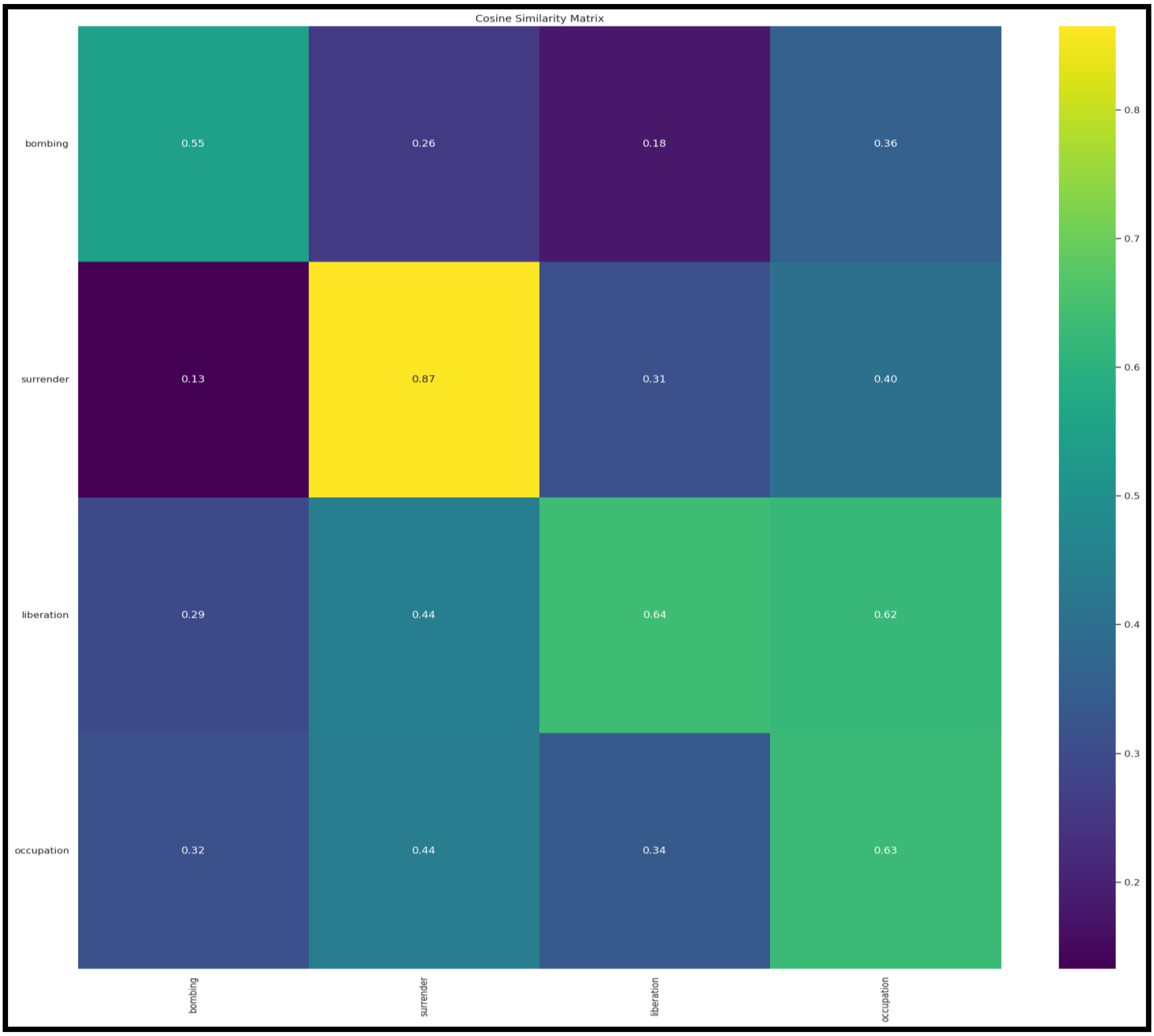


Figure 2: HeatMap showing how similarly the words “bomnbing”, “surrender”, “liberation” and “occupation” are used between the Russian and Ukrainian models.

Figure 2 reveals several interesting similarities and differences in word usage between the two models.

The word "surrender" scores a very high similarity between the two countries. This suggests that both nations have a similar attitude towards the concept of surrender, likely reflecting their desire for the opponent to surrender while being unwilling to do so themselves. The high similarity may also be influenced by significant battlefield events where soldiers from both sides had to surrender. For instance, Ukraine experienced a mass surrender during the Mariupol siege, while Russia surrendered a large swath of territory during the 2022 Kharkiv counteroffensive.[17][18]

The word "bombing" shows a low similarity, which is intriguing. One might expect both warring nations to have similar attitudes towards bombing. However, the low similarity likely reflects the differing capabilities and experiences of the two countries. Russia has the advantage in terms of bombardment, having conducted widespread bombings across Ukraine with a large arsenal of weapons. In contrast, Ukraine has carried out limited bombardments near the rear of Russian lines. These differences suggest that Ukrainian media report more frequently and negatively on bombings due to their greater impact on Ukraine, while Russian media report less frequently on successful, limited bombardments of their own territory.

The words "liberation" and "occupation" also show low similarity, similar to the findings in Table 4 of the word tests. This difference highlights the contrasting perspectives of Ukraine and Russia regarding these concepts. Ukraine views liberation as the reclaiming of its lost territories, whereas Russia frames its war as a liberation effort for Russian minorities in Ukraine's eastern regions. Both nations consider the other as occupying, but in different contexts. Ukraine treats "occupation" as the literal occupation of its lost territories by Russia. Conversely, Russia, having annexed some of its captured territories, views areas still held by Ukraine as occupied by Ukraine.[19]

### 4.2 Sentiment framing of entity mentions

In this section, I will present the sentiment framing of specific entity mentions that appear in both Russian and Ukrainian texts. The graphs provide insights into the general attitudes each country holds towards these entities. However, it is important to note that due to how TextBlob scores sentiment and how my implementation scores adjectives with high similarity to the component words of the entity mentions, the sentiment scores may not be entirely pure.

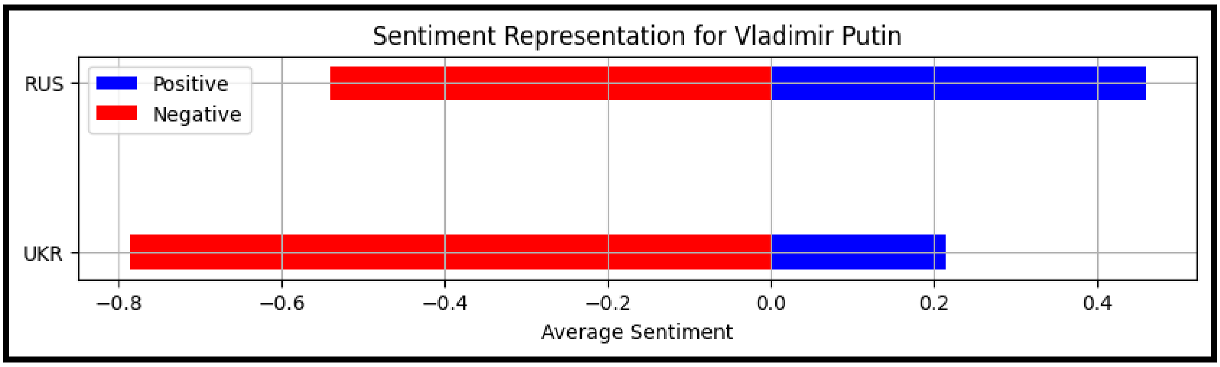


Figure 3: Sentiment Framing of the entity “Vladimir Putin”.

Figure 3 illustrates a stark difference in sentiment between the two countries regarding the president of the Russian Federation. The sentiment from Russian sources towards "Vladimir Putin" is almost neutral, indicating that Russian media likely reports on him in a more neutral and technical manner, without excessive praise.

On the other hand, the sentiment from Ukrainian sources is overwhelmingly negative, with very little positive sentiment. This suggests that Ukrainian media frequently derides and criticizes Vladimir Putin, actively pursuing a more negative portrayal of the Russian president.

This contrast highlights how media in both countries frame the president of the Russian Federation differently, reflecting their broader political and propaganda strategies. Ukrainian media likely aim to construct an image of a malevolent enemy, portraying Vladimir Putin in a highly negative light to reinforce their narrative of resistance and justify their defensive stance. In contrast, Russian media present their president in a more neutral and routine manner, likely to maintain an image of normalcy and stability in the face of international scrutiny.

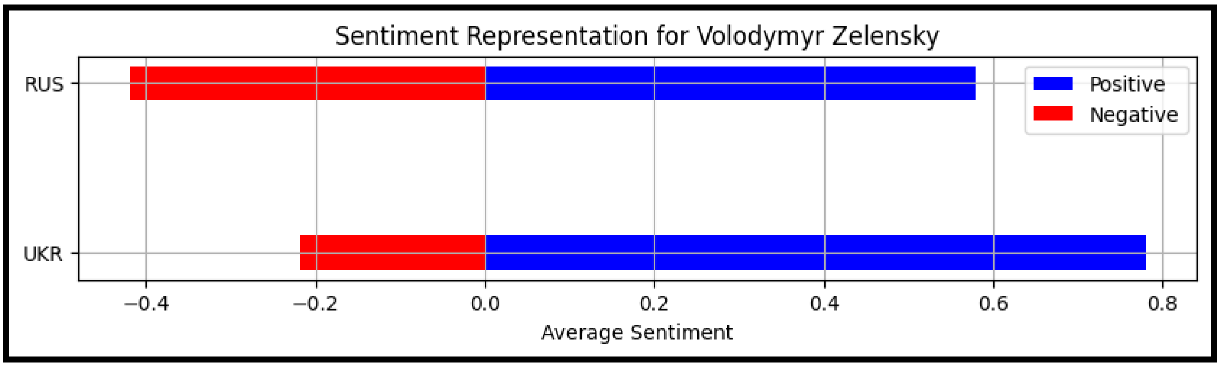


Figure 4: Sentiment Framing of the entity “Volodymyr Zelensky”.

Figure 4 presents an interesting sentiment framing of Volodymyr Zelensky. The sentiment from Ukrainian sources towards their president is very positive, with minimal negative sentiment. This suggests that Ukrainian media actively strive to present Volodymyr Zelensky in a positive light. This approach contrasts sharply with the Russian media's portrayal of Vladimir Putin, which, as previously observed, is more neutral.

The sentiment from Russian sources towards Volodymyr Zelensky is again relatively neutral, despite being the president of the nation with which they are at war. One might expect Russian media to adopt a similarly negative stance towards Zelensky as Ukrainian media do towards Putin. However, the relatively neutral sentiment may indicate a more dismissive attitude from Russian media. This likely stems from the prevailing Russian narrative that portrays Zelensky as a mere puppet, which leads Russian media to use more reserved and less overtly negative language when describing him.

This contrast highlights how media in both countries frame the president of Ukraine differently. Ukraine seeks to bolster the image of their leader positively, reinforcing national unity and resilience, while Russia aims to downplay Zelensky’s significance by maintaining a neutral tone.

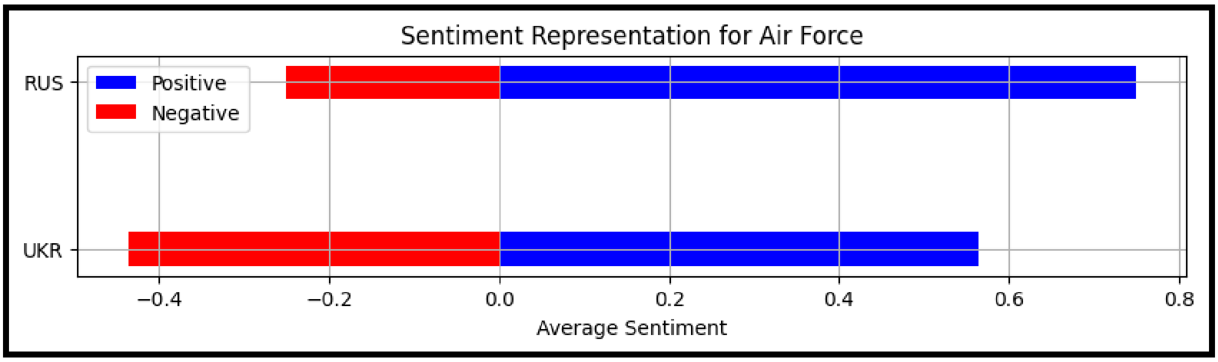


Figure 5: Sentiment Framing of the entity “Air Force”.

Figure 5 concerns the entity "Air Force," which appears in both news corpora and likely refers to the air forces of both countries as well as their adversary's air forces.

The sentiment from Russian sources towards "Air Force" is overwhelmingly positive. This likely reflects Russia's frequent reporting on their air force's missile bombings in Ukraine and its overall battlefield performance, which they claim is highly successful in combatting Ukraine's armed forces. The positive sentiment underscores a focus on the effectiveness and achievements of their own air force.[20]

In contrast, the Ukrainian sentiment towards "Air Force" is largely neutral with a mix of positive and negative sentiments. This neutrality likely results from the Ukrainian media's coverage of both their own air force and that of their adversary. The negative sentiment can be attributed to reports of Russia's aggressive bombing campaign, which regularly inflicts losses on both Ukrainian civilian and military targets. Additionally, there are likely reports on the need to refurbish Ukraine's aged and degraded air fleet. [21] On the other hand, the positive sentiment arises from reports of successful Ukrainian air strikes on Russian military installations and interceptions of Russian drones.

This shows the different narratives each country promotes regarding their air force's role and effectiveness in the conflict. Russian media emphasize the strength and success of their air force, while Ukrainian media provide a more balanced view, reflecting both the challenges they face and their own achievements.

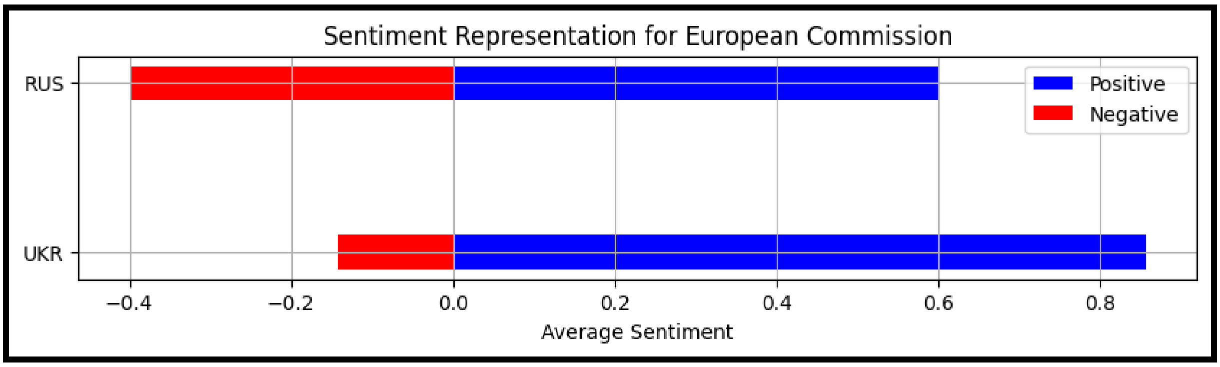


Figure 6: Sentiment Framing of the entity “European Commission”.

Figure 6 shows the sentiment framing for the European Commission, the main executive body of the European Union. As indicated by the sentiment graph, Ukraine and Russia hold markedly different attitudes towards the European Commission.

The Ukrainian sentiment towards the European Commission is very positive. This is likely due to favorable reporting on the European Commission's decisions to sanction and bar Russia from various economic sectors and overall trade with Europe. Additionally, the European Union's decision to grant Ukraine candidate status for membership probably contributed to the positive sentiment.[22][23]

Conversely, the Russian sentiment is more neutral but includes significantly more negative scoring compared to Ukraine. This likely reflects Russian media's criticism of the European Commission's sanctions against Russia. However, the presence of positive sentiment could indicate a milder treatment of the European Commission's decisions, possibly to downplay the impact of the sanctions. Additionally, the positive sentiment might also stem from Russian media gloating about the perceived limited effect of the European Union's sanctions on Russia's economy, which only experienced a minor reduction in growth.[24]

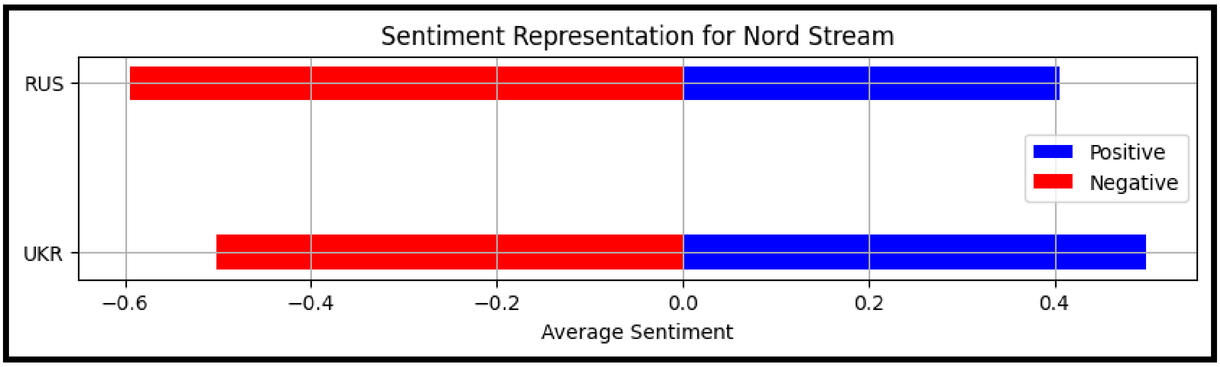


Figure 7: Sentiment Framing of the entity “Nord Stream”.

Figure 7 illustrates the difference in reporting on the Nord Stream pipelines between Russian and Ukrainian sources. These pipelines, joint projects between Germany and Russia, were designed to facilitate the supply of natural gas from Russia to Germany. Both pipelines were sabotaged in September 2022 and are currently inoperative.

The sentiment from Russian sources towards "Nord Stream" is predominantly negative. This negative inclination likely stems from Russia's reporting on the complications arising from the operation of the pipelines due to EU sanctions. The sabotage of the pipelines further exacerbated the negative sentiment, as these were costly projects that generated significant revenue for Russia.[25]

Conversely, Ukrainian sentiment towards the pipelines is more neutral. The negative sentiment likely reflects disapproval of the continued operation of the pipelines during the war. The positive sentiment may result from reporting on the sabotage of the pipelines more favorably, viewing it as a setback to one of Russia's primary economic levers over Germany.

Overall, the sentiment framings presented above provide a clearer understanding of how both sides report and present various topics and events. These analyses reveal the distinct narratives and attitudes each country promotes through their media.

From the sentiment analyses, we observe that Ukrainian media generally adopts a more positive tone when discussing their own entities and allies while portraying Russian entities in a highly negative light. This approach likely aims to bolster national morale and secure international support by emphasizing the righteousness of their cause and the malign nature of their adversaries.

Conversely, Russian media tends to maintain a more neutral or mildly positive tone towards their own entities, even when discussing controversial figures or events. This neutral stance may be designed to project an image of stability and rationality, both domestically and internationally. Additionally, Russian media often downplays or criticizes Western entities and their actions, aligning with their narrative of Western antagonism and interference.

Through these sentiment framings, we can discern the broader strategic goals of both countries' media. Ukraine's media aims to unify its population and galvanize international support by highlighting the threats posed by Russia and the support received from allies like the European Union. On the other hand, Russian media seeks to portray a narrative of resilience and justified action, while delegitimizing Ukrainian leadership and downplaying the impact of international sanctions.

These contrasting media strategies not only shape domestic perceptions but also influence international viewpoints, highlighting the role of media in the ongoing information warfare between the two nations. Acknowledging these sentiment framings is crucial for comprehending the broader geopolitical context and the psychological underpinnings of the conflict.

# Chapter 5

## Temporal Shifts in Sentiment



### 5.1 Shift in sentiment for the total data

In this subchapter, I will present two line plot graphs that show the shift in total negative and positive sentiment from both Russia's and Ukraine's collective entity mentions. The plots have a y-axis that corresponds to the sentiment and an x-axis that corresponds to the date. The data follows a format of YYYYMM, covering only the year and months.

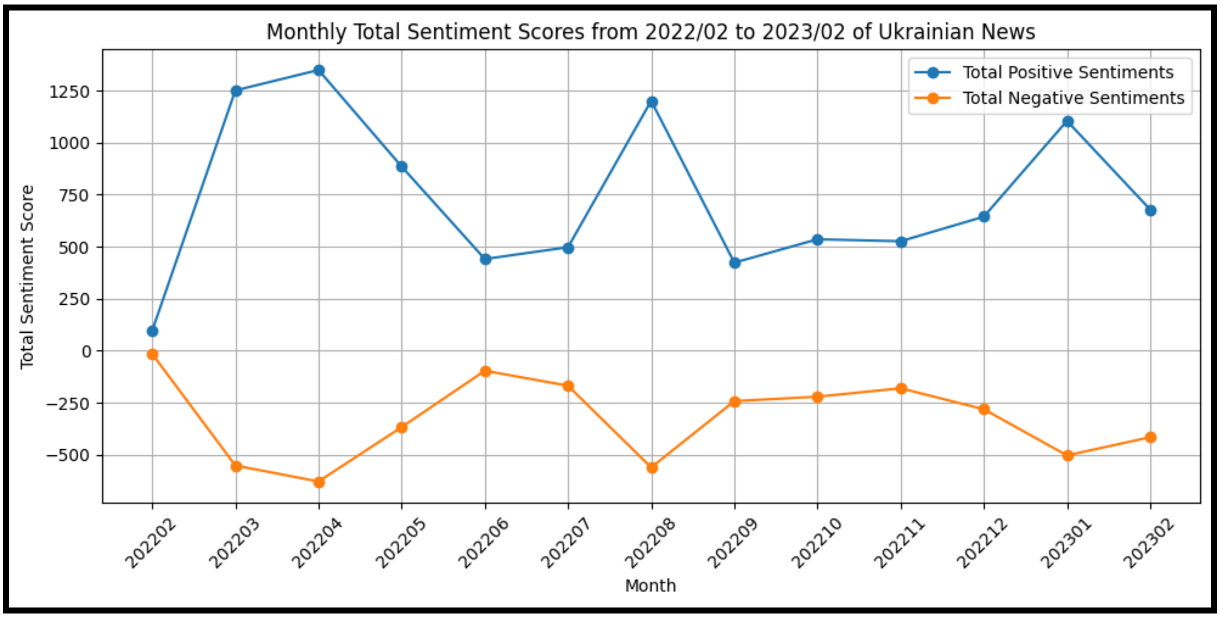


Figure 1: Sentiment shift over the whole year of Ukrainian news data

Figure 1 provides a timeline of the total sentiment in Ukrainian news during the first year of the war. The line plot illustrates how sentiment fluctuates, growing, shrinking, or remaining steady over the months. Both negative and positive scores follow a similar pattern, expanding and contracting in unison. There is an abundance of positive scoring compared to negative, which, as mentioned earlier in section 4.2, may be due to TextBlob's scoring method and the code's implementation.

As the war began on February 24, 2022, sentiment surged wildly in the following month. In March, we see significant growth in both positive and negative scores, with both peaking in April. This likely corresponds to Ukraine's coverage of Russia's large-scale attack on its sovereign territory and the subsequent battles. During this stage of the war, the frontlines were highly fluid, with reports of Russian troops advancing to many locations within Ukraine.[26]

The positive scores for March and April can be attributed to Ukrainian media reporting on the substantial international support received, especially from Europe and the United States. Additionally, Ukraine's positive coverage of some significant Russian military setbacks, particularly in the north of Ukraine, contributed to the high positive sentiment.[27]

The negative scores for March and April likely result from reports on Ukraine's significant territorial losses during these months and major military setbacks, such as losing the Zaporizhzhia Nuclear Power Plant (ZNPP), the largest in Europe, and the capture and extended siege of Mariupol. Other contributing factors include reports of war crimes perpetrated by Russia in the north of Kyiv, discovered during the Russian army's retreat from that area.[28]

In the subsequent months, we observe a positive spike in sentiment scores during August, likely due to Ukraine's ascension to EU candidate status at the end of July and the commitment of further military support from the United States.[30]

The other months show a surprisingly steady pace in sentiment scores. This is notable because September and October were among Ukraine's most successful months on the now-established frontlines. Ukraine almost fully recaptured the Kharkiv region and parts of the Donetsk region, while in November, Russian forces retreated from the right bank of the Kherson region, allowing Ukraine to fully recapture the city of Kherson.[29][31]

This steady sentiment suggests that Ukrainian media maintained a more reserved stance on their frontline successes, avoiding over-advertising their military actions. This restraint may have contributed to their success, as the 2022 Kharkiv counteroffensive largely took Russian forces by surprise.

The spike observed in January 2023 is likely due to Germany and the United States announcing their delivery of Abrams and Leopard 2 tanks to Ukraine, which significantly boosted positive reporting in Ukrainian media during that month.[32]

With Figure 1 we can see a comprehensive view of how Ukrainian media sentiment shifted throughout the first year of the war, reflecting the nation's experiences and strategic communications.

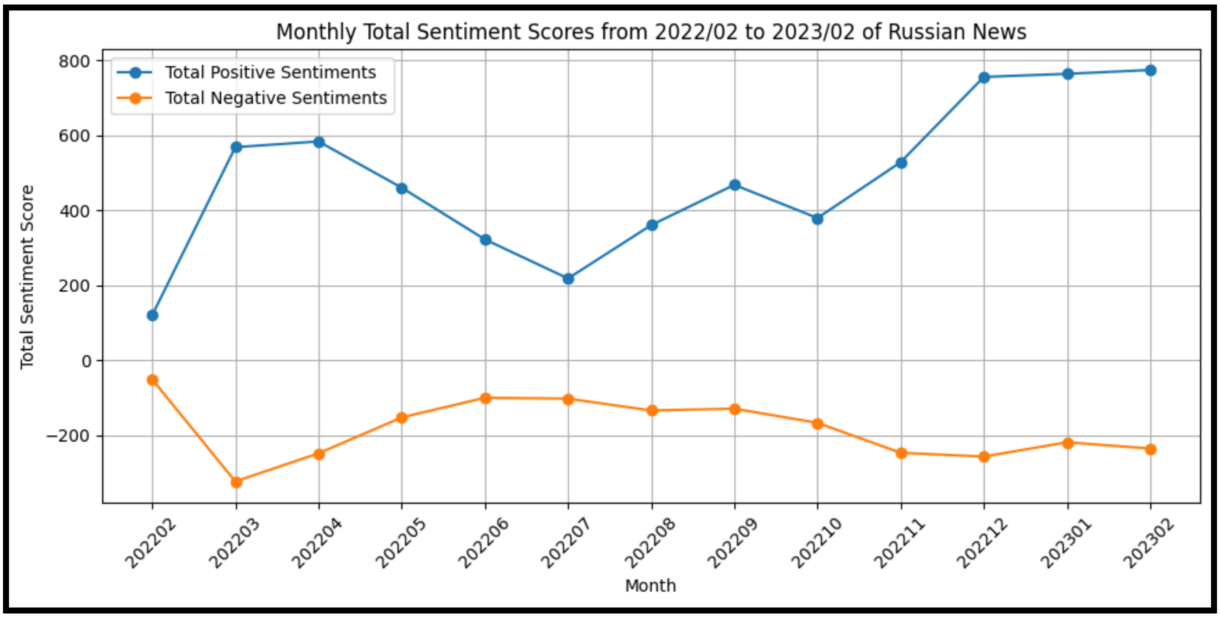


Figure 2: Sentiment shift over the whole year of Russian news data

Figure 2 illustrates the sentiment shift in Russian news throughout the first year of the war. The line plot shows distinct patterns for positive and negative sentiment scores. Initially, during March and April, there is growth in both positive and negative sentiment, though the positive sentiment grows more significantly.

The growth in positive sentiment in March and April likely stems from Russian media framing their military actions positively, presenting the war as a special military operation aimed at assisting Russian minorities in eastern Ukraine.[2] Additionally, the early military successes, particularly in southern Ukraine, contributed to this positive reporting. However, despite the setbacks in April, such as the pullback from northern Ukraine, the positive sentiment remained high, and the negative sentiment decreased.

The reason for maintaining an overall positive score and lowering negative sentiment in April may be due to Russian media's portrayal of its army's retreat from northern Ukraine and around Kyiv, as a strategic pullback and goodwill gesture towards Ukraine. At that time, Russian and Ukrainian delegates were meeting in Istanbul to negotiate a peace deal, and the retreat was framed as a measure to facilitate these negotiations.[33]

In the following months, we observe a gradual decline in both positive and negative sentiment scores. This likely reflects the stabilization of the frontline and the slower pace of advances by the Russian army.

From June to September, there is a noticeable increase in positive sentiment. This period corresponds to Russian media reporting on the successful defense against Ukrainian counterattacks in the Kherson region during the summer. The peak in positive sentiment in September coincides with President Vladimir Putin's announcement of a partial mobilization.[34] However, in October, there is a slight drop in positive sentiment and a gradual increase in negative sentiment. This change likely results from the significant loss of territory in the Kharkiv region due to a surprise Ukrainian counteroffensive. Despite this setback, the sentiment scores did not drastically change, as Russia framed the loss as a strategic retreat after completing its objectives in the Kharkiv region.[35]

From November to February 2023, positive sentiment grows to its highest levels. This increase could be attributed to Russia's major missile bombardment campaign, which began in November and December, with Russian media highlighting successful airstrikes. Additionally, by December, the Russian front had stabilized against

Ukrainian counteroffensives, and by February 2023, the Russian army achieved partial successes in the Donetsk region.[36]

This analysis of sentiment shifts in Russian news throughout the first year of the war reveals how media narratives adapt to military and political developments, maintaining a positive portrayal of Russian actions while mitigating the impact of setbacks.

### 5.2 Shift in monthly sentiment for entity mention

In this subchapter, I will present several line plot graphs that illustrate the shift in total negative and positive sentiment for specific entity mentions in both Russia's and Ukraine's collected news data.

Each plot features a y-axis representing the sentiment scores and an x-axis representing the date, formatted as YYYYMM, covering the year and months. These graphs provide insights into how sentiment towards particular entities has evolved over the course of the conflict.

**Russian Entity Mentions:**

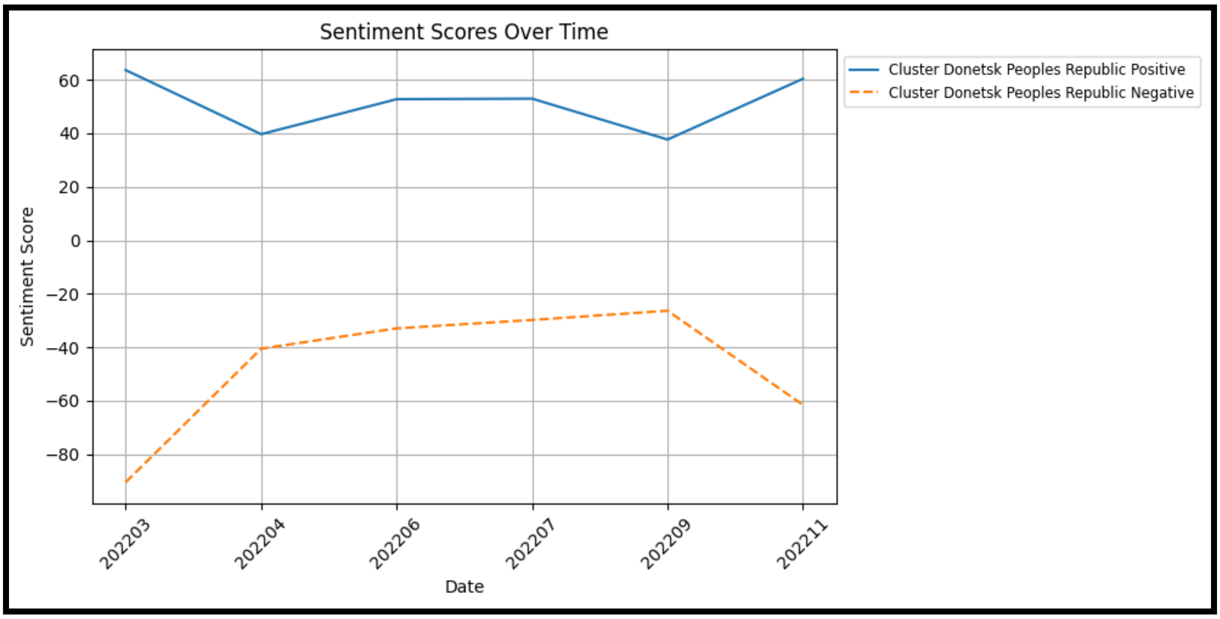


Figure 3: "Donetsk Peoples Republic" entity mention sentiment over time.

Figure 3 indicates that Russian online publications have maintained a relatively high and positive sentiment towards the Donetsk People's Republic (DPR), reflecting Russia's long-standing support for the region's ambitions to join Russia. Consequently, Russian media likely cover topics related to the DPR with an overall positive attitude.

The presence of high negative sentiment is primarily attributed to the region's role as a major frontline in both the Ukrainian civil war and the subsequent Russo-Ukrainian War. The negative sentiment likely arises from Russian reports on the ongoing conflict and the inherently negative connotations associated with war and destruction.

In March, the first occurrence of the sentiment cluster, we observe a significant negative spike that continues into April. This is likely due to the intense fighting during the siege of Mariupol, a city in the Donetsk region, which saw heavy combat as Russian forces encircled and assaulted the city. The negative sentiment during this period can also be attributed to reports of Ukrainian shelling of Donetsk City and other frontline clashes occurring simultaneously.

Figure 3 I believe, reflects the complex nature of the region's coverage, balancing positive reports of support and integration with the harsh realities of ongoing conflict.

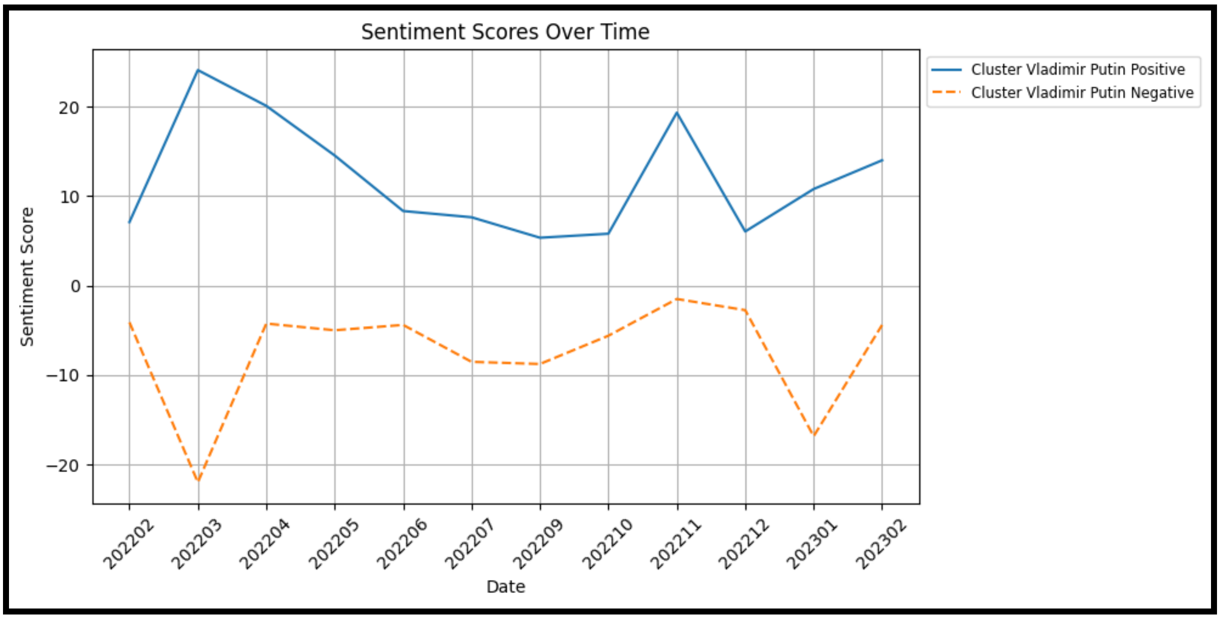


Figure 4: “Vladimir Putin entity” mention sentiment over time.

Figure 4 shows the sentiment scores for the "Vladimir Putin" entity mention, illustrating its presence in all months of the Russian news corpus. This consistent presence is expected, given that Russian media regularly report on the actions, speeches, and decisions of Vladimir Putin as the president of the Russian Federation.

In March, we observe a significant spike in overall sentiment scores. This likely corresponds to Russian media coverage and commentary on Vladimir Putin's decision to initiate what he termed a "special military operation" against Ukraine. The substantial reporting on this event would naturally result in numerous adjectives with varying sentiment scores associated with his name.

Following March, the overall sentiment gradually decreases in the subsequent months, reaching a lower point by October. The spike in November is likely attributable to Russian media coverage of Vladimir Putin's speech at the 19th Annual Meeting of the Valdai Discussion Club, which took place at the end of October. The Valdai Discussion Club is a prominent forum where Vladimir Putin annually attends and shares his perspectives on various topics, and his speeches there generate significant media attention.[37]

Figure 4 ultimately, highlights how major political events and speeches by Vladimir Putin influence the sentiment trends in Russian media over time.

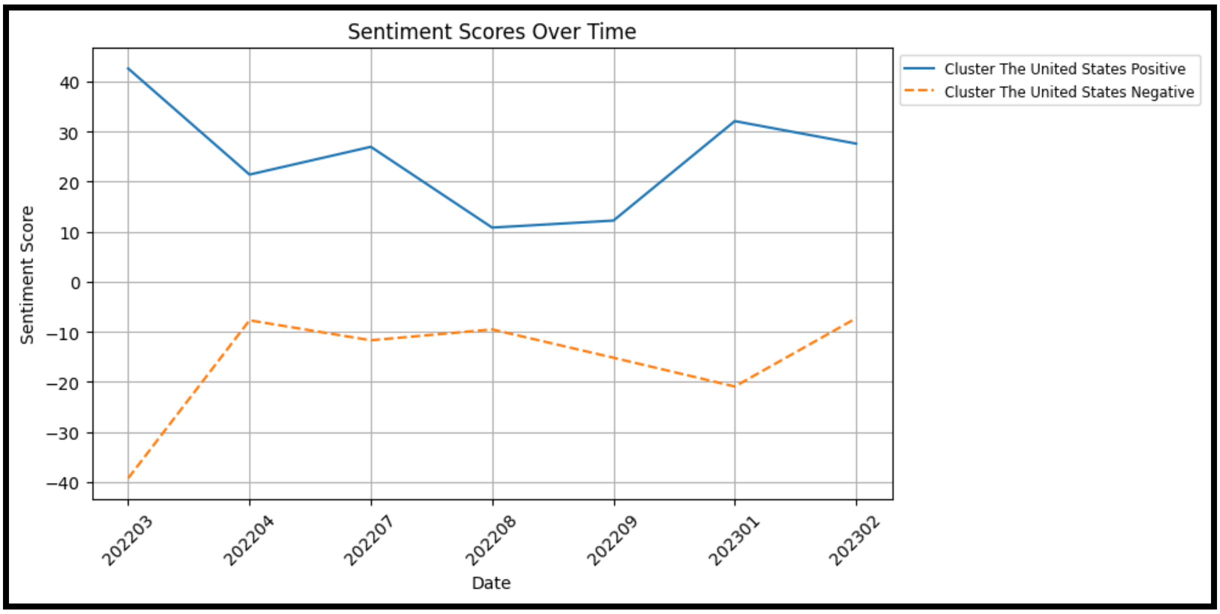


Figure 5: “United States” entity mention sentiment over time.

Figure 5 we see the sentiment scores for the "United States" entity mention, capturing how Russian media have reported on the U.S. throughout the conflict. In March, we observe a significant negative sentiment score. This is likely indicative of the United States' prominent role in supporting Ukraine both financially and militarily during February and March. Russian media likely reported these actions with a generally hostile attitude, reflecting the adversarial relationship between the two nations.

Throughout the subsequent months, both positive and negative sentiment scores generally decrease. However, there is a noticeable relative increase in both positive and negative sentiment scores in January 2023. This uptick likely corresponds to increased U.S. support for Ukraine, as additional military and financial aid was provided to bolster Ukraine's defenses and prepare for the anticipated 2023 counteroffensive.[32]

The sustained negative sentiment reflects the ongoing criticism by Russian media of U.S. actions and policies that are perceived as antagonistic to Russian interests. Meanwhile, the positive sentiment, though consistently higher than negative sentiment, likely results from Russian media's portrayal of certain U.S. actions or statements in a manner that can be spun positively within the context of their domestic narrative.

This sentiment analysis demonstrates the complex and evolving portrayal of the United States in Russian media, influenced by geopolitical developments and strategic communications. It highlights how the U.S.'s role in the conflict is framed to support Russian narratives, both critical and occasionally acknowledging the impact of U.S. actions.

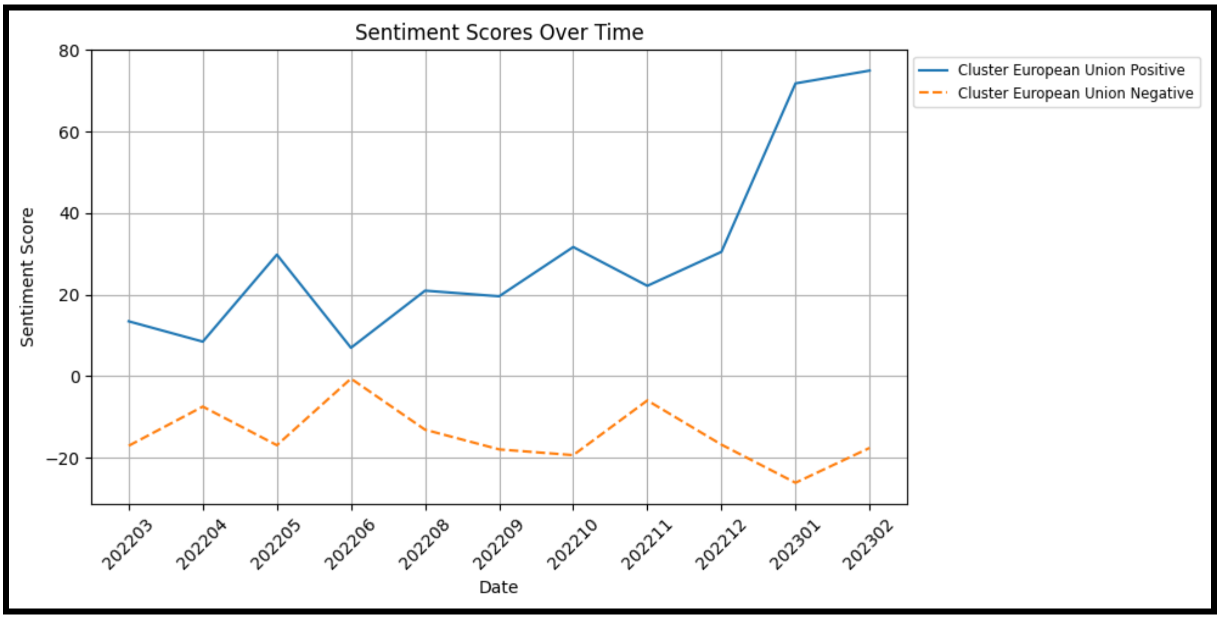


Figure 6: “European Union” entity mention sentiment over time.

Figure 6 illustrates the sentiment scores for the "European Union" entity mention over time, highlighting how the EU is portrayed in Russian media throughout the conflict.

The regular appearance of the "European Union" entity in Russia’s news data is expected, given the EU’s active and oppositional stance towards Russia's actions in Ukraine. The sentiment spikes observed in the graph likely correspond to reactions from Russian media to the various sanctions imposed by the EU over the first year of the war.

The positive sentiment scores generally outweigh the negative ones, though both exhibit fluctuations. The peaks in sentiment scores during the winter months, particularly in January and February 2023, are notable. This increase in sentiment is likely related to the EU and G7’s decision to finally implement a price cap on Russian oil, after prolonged debates among EU and G7 member states. [38] Interestingly, the sentiment during these months is predominantly positive.

This positive sentiment may indicate that Russian media approached the news of the price cap with a certain level of calm, possibly downplaying its significance and avoiding major negative hype. By framing the price cap as manageable or ineffective, Russian media could be aiming to reassure their domestic audience and mitigate any potential concerns about the economic impact of the sanctions.

The overall trend in sentiment scores suggests that Russian media maintain a complex narrative around the European Union, balancing criticism with a portrayal of resilience and stability. This approach helps to sustain a narrative that sanctions and opposition from the EU, while challenging, do not severely undermine Russia's strategic goals or economic stability.

This analysis underscores the nuanced portrayal of the European Union in Russian media, reflecting broader geopolitical strategies and the ongoing information warfare between Russia and the West.

**Ukrainian Entity Mentions:**

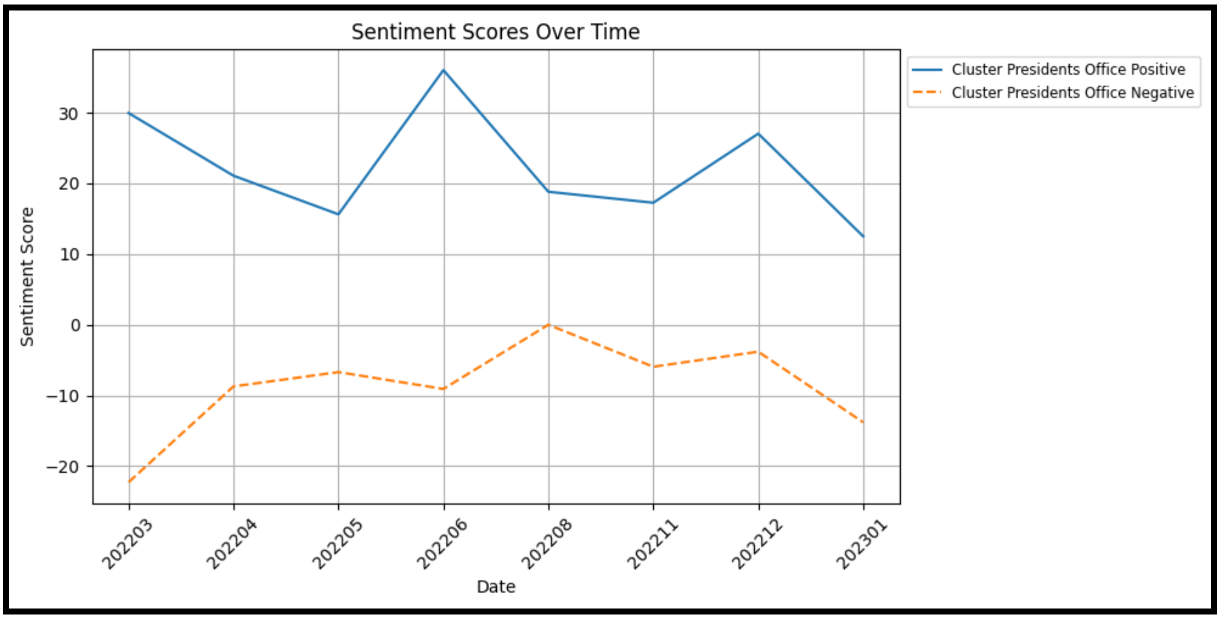


Figure 7: "Presidents Office" entity mention sentiment over time.

Figure 7 illustrates the sentiment scores for the "President's Office" entity mention. This entity refers to President Volodymyr Zelensky and his administration, the office of the president of Ukraine. The overall positive sentiment indicates that Ukrainian media generally report on their government and president in a favorable light.

There are three major spikes in positive sentiment: March, June, and December. The positive sentiment in March is likely due to the onset of the war, with Ukrainian media reporting positively on President Zelensky's actions and decisions in response to the invasion. The media coverage during this period likely focused on his leadership and efforts to rally the nation and secure international support.

The spike in June corresponds to President Zelensky's successful efforts in convincing Ukraine's European partners to grant the country EU candidate status. Achieving candidate status for European Union membership has been a significant goal for Ukraine, and the positive sentiment reflects the media's praise for this diplomatic success.[30]

The increase in positive sentiment in December is likely due to President Zelensky's continued efforts to secure foreign aid for Ukraine, particularly military support from Germany and the United States. The media coverage would have highlighted these achievements, emphasizing the ongoing international backing for Ukraine's defense against Russian aggression.

Generally, the positive sentiment trend underscores the media's role in bolstering national morale by focusing on the president's successful initiatives and international support. This analysis provides insights into how Ukrainian media frame the actions of their leadership, reflecting the broader strategy of maintaining public confidence and unity during the conflict.

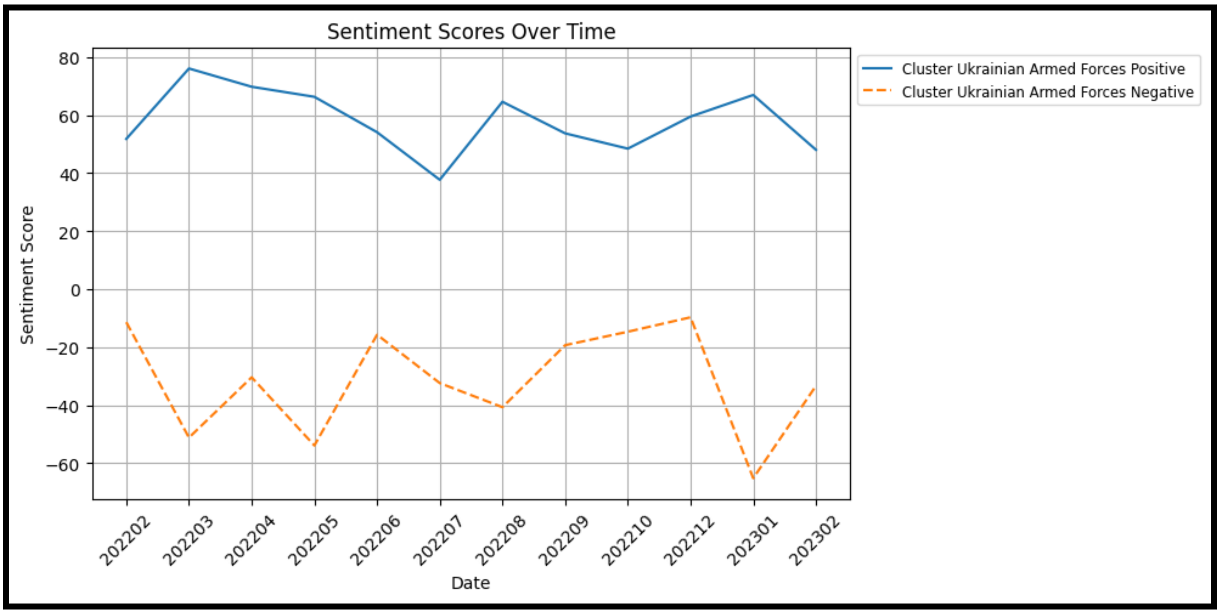


Figure 8: “Ukrainian Armed Forces” entity mention sentiment over time.

Figure 8 illustrates an overall mixed sentiment for the "Ukrainian Armed Forces" entity mention, as I mentioned at beginning of this chapter this score, could be influenced by TextBlob’s sentiment scorer, potentially misinterpreting some words without considering their context. However, there are also clear reasons for the varied sentiment regarding the Ukrainian Armed Forces.

For the most part, the sentiment is predominantly positive. This positive outlook is likely due to Ukrainian media portraying their armed forces in a favorable light to maintain national morale, especially in the early stages of the war when Russia held the initiative and won several battles. Significant spikes in positive sentiment are observed during the early phases of Ukraine’s counteroffensives in the Kherson and Kharkiv regions around August and September 2022.

On the other hand, there are also notable spikes in negative sentiment. These spikes correspond to Ukrainian media reporting on significant military setbacks. For example, during April and May, the negative sentiment spikes align with the siege of Mariupol and the subsequent siege of Azovstal, where Ukrainian forces faced substantial challenges and major losses. Another major spike in negative sentiment is observed in January 2023, when Russia achieved some combat success in the town of Bakhmut in the Donetsk region and managed to halt Ukraine’s advances from their successful Kharkiv counteroffensive.

Figure 8 highlights the dynamic and evolving portrayal of the Ukrainian Armed Forces in the media. Positive sentiment generally reflects I believe, periods of success and strategic victories, while negative sentiment corresponds to significant military challenges and losses. This dual narrative illustrates the complex reality of the conflict and the efforts of Ukrainian media to balance optimism with the realities of war.

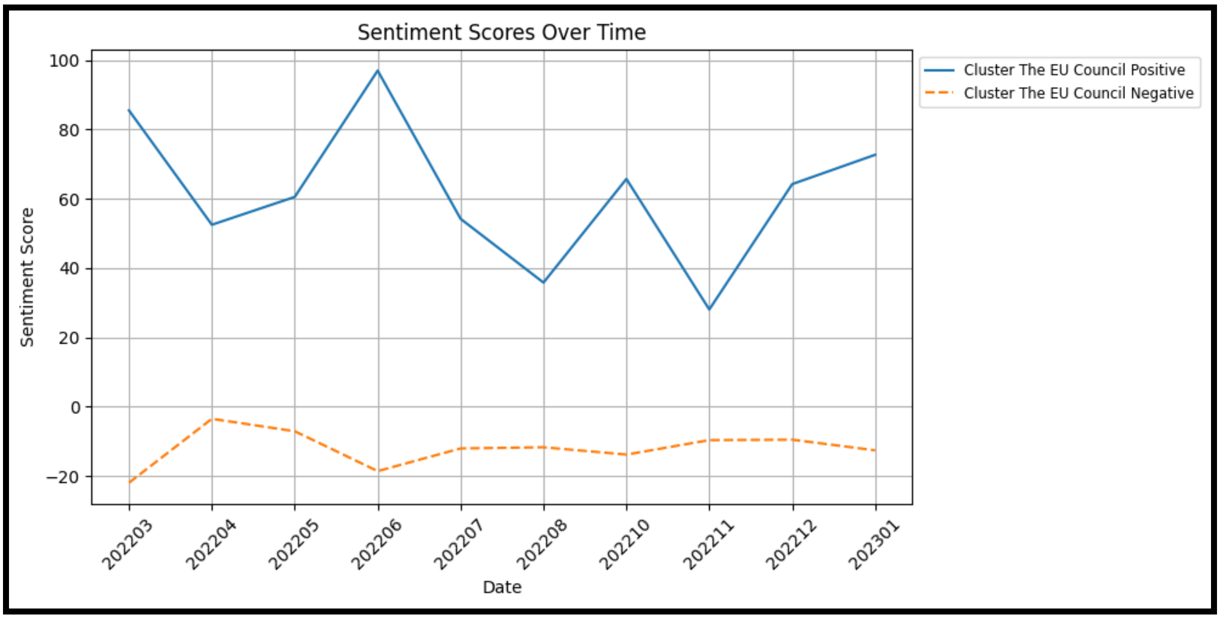


Figure 9: “EU Council” entity mention sentiment over time.

Figure 9 shows the sentiment scores for the "EU Council" entity mention, highlighting its overall positive sentiment in the Ukrainian news corpus. The "EU Council" refers to the Council of the European Union, one of the legislative bodies of the European Union. The European Union has been very supportive of Ukraine, providing significant financial and military assistance through its member states.

The sentiment analysis reveals three notable spikes in positive sentiment, occurring in June, October, and December. The spike in June is likely due to the European Union granting EU candidate status to Ukraine, a significant milestone for the country. The October spike corresponds to the adoption of further sanctions and financial assistance by the European Union. The spike in December is likely related to the adoption of the oil price cap by the EU, a move aimed at limiting Russia's revenue from oil exports.

These positive sentiments, coupled with the overall very low negative sentiment, indicate that Ukrainian media generally present the European Union in a favorable light. This aligns with the Ukrainian government's goal of eventually becoming a member of the EU. The consistent positive portrayal of the EU by Ukrainian media reflects the importance of European support for Ukraine during the conflict and Ukraine's aspirations for closer integration with Europe.

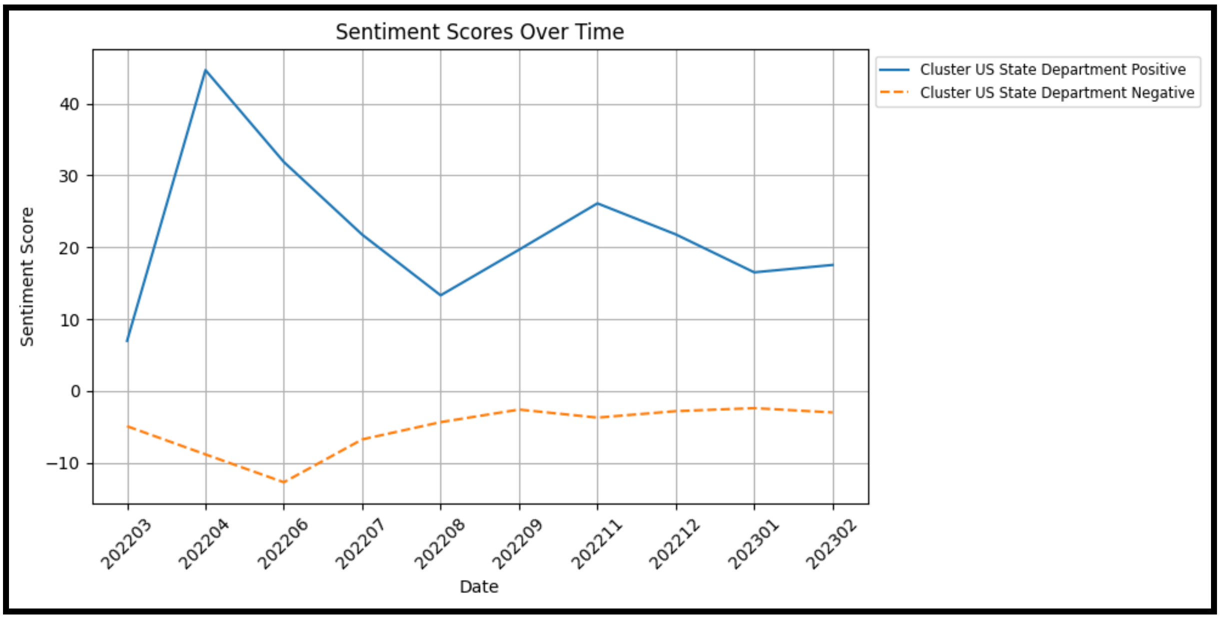


Figure 10: "US State Department" entity mention sentiment over time.

Figure 11 shows a very positive sentiment towards the entity mention "US State Department." The US State Department, equivalent to a ministry of foreign affairs, has been a key supporter of Ukraine's efforts against Russia. The United States, through the State Department, has regularly provided substantial material and financial support to Ukraine.

This ongoing support explains the consistently positive sentiment scores. The spikes in positive sentiment correspond to months when the US State Department announced new packages of aid or additional support. These sentiment scores likely reflect how Ukrainian media report positively on the support from the US, highlighting the significant assistance they receive and reinforcing the strong alliance between the two nations.

The positive sentiment illustrates the US State Department's regular presence in Ukraine’s war effort and how Ukrainian media use these announcements to boost national morale by showcasing international support.

With the analysis in sections 5.1 and 5.2, we gain a more detailed understanding of how the news media in both warring states, Ukraine and Russia, shift their stances and respond to major events over time. The sentiment trends reveal how each country’s media adapts to evolving circumstances and frames their narratives accordingly. By examining the sentiment changes, we can observe how both positive and negative developments are reported, offering insights into the media strategies used to influence public perception.

Furthermore, the comparison of sentiment towards similar events or topics between the two states highlights the stark differences in media narratives. This contrast provides a deeper view of the propaganda and information warfare tactics employed by each side. Through these analyses, we can better see how both Ukrainian and Russian media construct their stories to align with their respective national agendas and shape the views of their domestic and international audiences.

### 5.3 Shift in monthly similarity between select words

In this section, I will present three graphs where I tested selected pairs of words and mapped their changing usage between the Russian and Ukrainian models created using CADE. The difference in usage of these words is measured using cosine similarity.

The y-axis of the graphs corresponds to the cosine similarity of the words between the state-trained CADE models. The x-axis represents the date, formatted as YYYYMM.

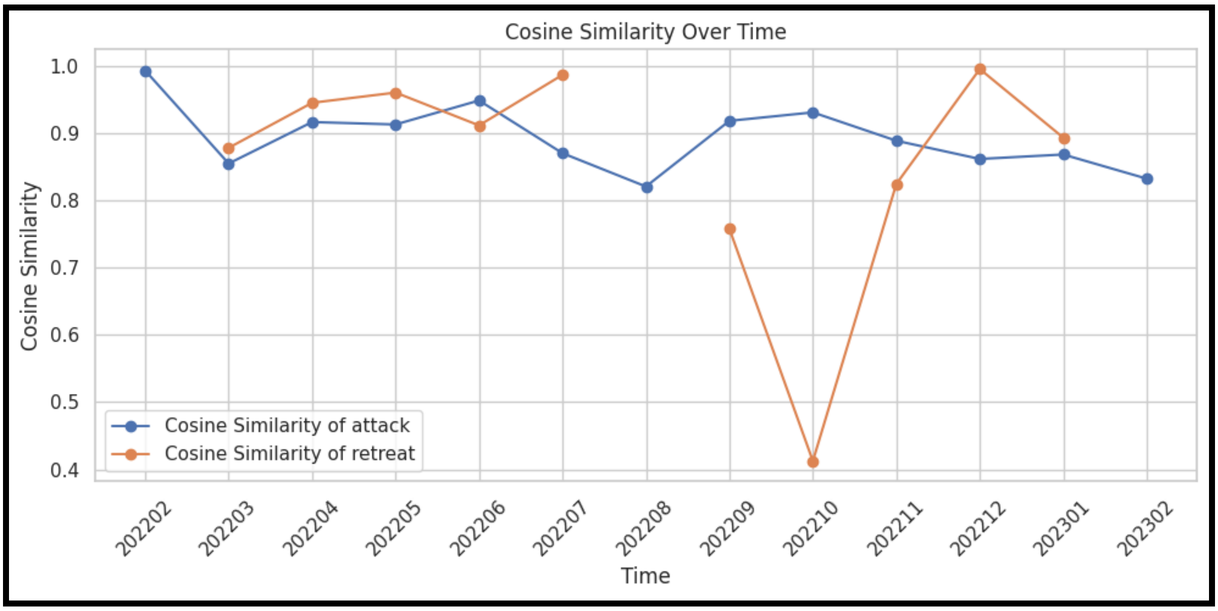


Figure 12: “attack” and “retreat” cosine similarity over time.

Figure 12 illustrates the changing differences in the usage of the words "attack" and "retreat" between the Russian and Ukrainian models.

The graph shows that the word "attack" maintains relatively similar usage between the two sides, with a slight decrease in similarity observed in August. This decrease likely coincides with Ukraine’s counterattacks in the Kherson region, which faced resistance from Russian forces.

The word "retreat," however, exhibits a more varied pattern. There is a noticeable absence of data points for "retreat" in one of the models during August, indicating that one side may not have used the word "retreat" at all during that month. A significant divergence in the similarity of "retreat" occurs in October. This is likely due to Russia's retreat from the Kharkiv region, which Russian media framed not as a military failure but as a tactical withdrawal.[39]

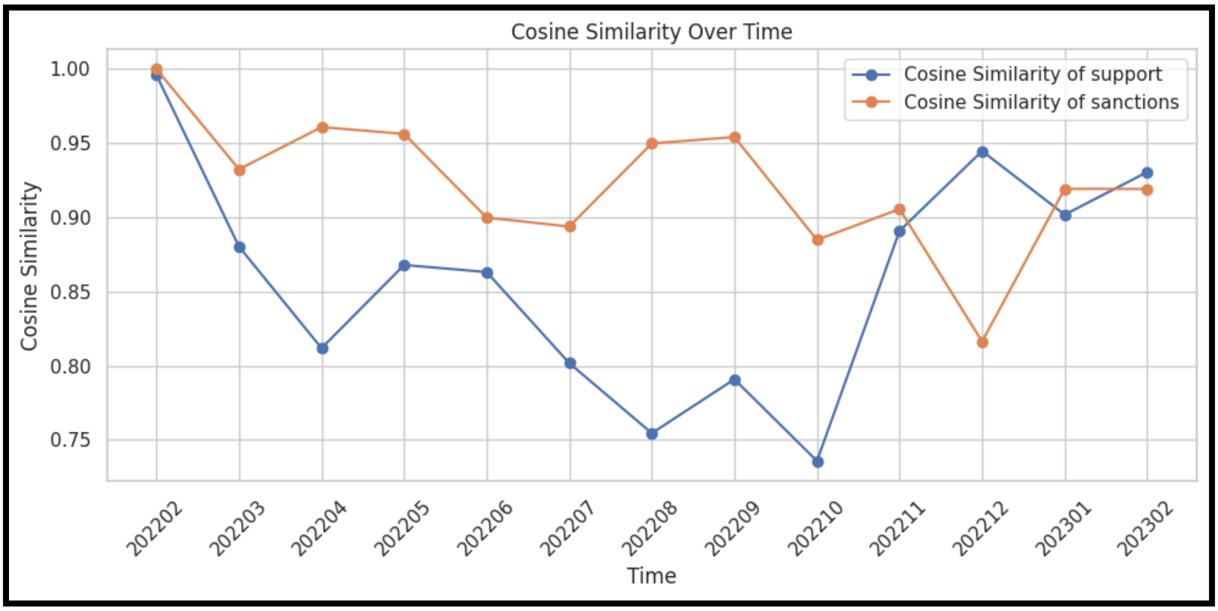


Figure 13: “support” and “sanctions” cosine similarity over time.

In Figure 13, the word "sanctions" shows a relatively consistent usage throughout the year of the war, as indicated by the high cosine similarity between the Russian and Ukrainian models. This consistency is likely because both sides were reporting on the same set of sanctions imposed against Russia. The only small deviation occurs in December, which could be due to differences in reporting on the gas and oil price cap imposed on Russia. Russia might have framed the price cap as another aggressive action by the West, while Ukraine likely viewed it as a significant and positive step against Russia, enhancing the impact of previous sanctions.

The word "support," on the other hand, shows a larger variation in usage, particularly from August to October. This difference likely corresponds to Ukraine’s counteroffensives during that period. Ukrainian media probably emphasized the international support that enabled their military successes and helped them seize the initiative in the war. In contrast, Russian media may have focused on the need for further support for their frontline troops due to the setbacks they experienced during these counteroffensives.

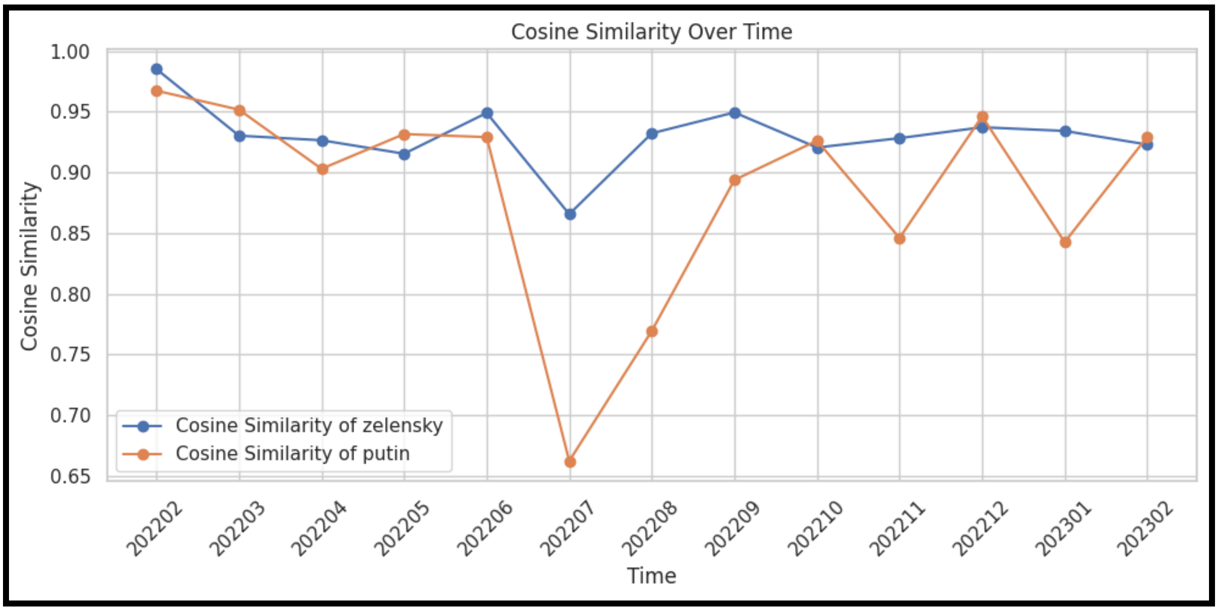


Figure 14: “zelensky” and “putin” cosine similarity over time.

Figure 14 illustrates the changes in the usage of the words "putin" (referring to the president of Russia, Vladimir Putin) and "zelensky" (referring to the president of Ukraine, Volodymyr Zelensky) over time. During the initial months of the war, both words exhibited very close and similar usage patterns. This similarity likely stems from their roles as presidents, indicating that both sides reported on their leaders in a consistent manner.

However, a significant shift occurs in July, where the usage of the word "putin" diverges noticeably. This change likely corresponds to the UN-brokered Black Sea grain corridor agreement, which was signed by both Ukraine and Russia to allow Ukrainian grain exports from the port of Odesa. The agreement may have prompted Ukrainian media to alter their reporting on President Putin, likely reacting with surprise that the Russian president agreed to such an initiative.[40]

With these experiments (Figure 12-14), we can further observe the nuanced differences in how Russian and Ukrainian media treat similar words and topics. The variations in word usage reveal distinct narrative strategies employed by each side to frame their perspectives and influence their audiences. For instance, the relatively consistent use of the word "attack" across both media outlets underscores a commonality in reporting military actions, while the varied usage of "retreat" highlights differences in how each side handles narratives of withdrawal and setbacks.

Similarly, the word "sanctions" maintains a high degree of similarity, reflecting both countries' coverage of the same sanctions imposed on Russia. However, the subtle deviation in December indicates differing interpretations of the gas and oil price cap, with Russia framing it as Western aggression and Ukraine viewing it as a positive measure. The word "support" shows more significant differences, particularly during Ukraine's counteroffensives, where Ukrainian media focused on international backing and Russian media emphasized the need for reinforcement.

The analysis of "zelensky" and "putin" also demonstrates initial similarities in reporting on their respective leaders, which shift significantly due to geopolitical events, such as the Black Sea grain corridor agreement. This shift illustrates how specific events can alter media portrayal and reflect changing narratives.

While the differences in word usage are evident, the similarities should not be dismissed. They reveal that both sides often address the same events and topics, albeit with varying degrees of emphasis and interpretation. These subtle differences provide deeper insights into the information strategies of each state, showcasing the complex interplay of media narratives in the context of the ongoing conflict.

# Chapter 6

## Conclusion and Discussion



### 6.1 Overview of similarities and differences of narratives

The collection of experiments conducted and discussed in Chapter 5 reveals that Ukraine and Russia adopt two distinct approaches to the war. Russia generally presents a more reserved attitude towards the conflict in Ukraine, reflecting its narrative that the war is a "special military operation" rather than an act of aggression against its neighbor. This framing is crucial for Russia to justify the conflict to both its domestic audience and the international community.

In contrast, Ukraine's use of words and sentiment reflects a more intense and severe treatment of the war. This approach underscores Ukraine’s narrative that it is engaged in an existential struggle against an aggressive neighbor. The heightened emotional tone in Ukrainian media serves to galvanize national and international support, emphasizing the dire stakes of the conflict.

A key similarity between the two nations' narratives is in the portrayal of their armed forces. Both Ukraine and Russia exhibit a reluctance to report on military setbacks. This is likely done for morale and propagandistic reasons, as acknowledging significant losses could lead to public disillusionment and potentially destabilize their governments. By downplaying defeats and emphasizing successes, both countries aim to maintain public support and morale.

The differences in how each country frames the conflict are stark. Russian media tend to use language that minimizes the perceived impact of the war, aligning with the government's portrayal of the operation as limited and controlled. Ukrainian media, on the other hand, highlight the aggression and brutality of the Russian forces, depicting the conflict as a life-and-death struggle for national survival.

The similarities and differences in these narratives reveal a broader strategy by both nations to weave and build a narrative that preserves their respective governments' prestige and standing among their populations and international partners. Russia's restrained narrative seeks to maintain a facade of control and legitimacy, while Ukraine's urgent and dramatic narrative aims to rally both domestic and international support against what it portrays as an existential threat.

These narrative strategies are not just about shaping public perception; they are integral to the broader geopolitical and psychological warfare being waged alongside the physical conflict. With these narrative frameworks, we gain deeper insights into the motivations and objectives driving the information campaigns of both Ukraine and Russia.

### 6.2 Possible impact on the perception of war on each side

The perception impact of these narratives is significant and multifaceted. Both Ukraine and Russia have taken steps to ban opposition media and silence dissenting voices, aiming to maintain a singular, state-approved version of events. This strategy indicates both nations' desire to control the narrative and enforce their perspective of the truth. By doing so, they distort the perception of their citizenry, creating an environment where the public can be more easily influenced by state decisions. For example, in Russia, the majority of the population holds a supportive attitude toward the war, largely due to the state's controlled narrative.

This control over media narratives extends beyond national borders, affecting international perceptions as well. Ukraine, with the backing of the US and Europe, has successfully established its narrative of the war as the dominant perspective in Europe and much of the Western world. This narrative portrays Ukraine as a nation defending itself against unprovoked aggression, thereby garnering widespread support and sympathy.

In contrast, the rest of the world, particularly the global south and most of Asia (with the exception of some US allies), often aligns with or at least gives credence to the Russian narrative. This is largely due to longstanding negative attitudes towards the West, driven by historical, political, and economic factors. As a result, these regions may view Russia's actions more sympathetically or critically assess Western motivations and actions in the conflict.

The consequence of these divergent narratives is a fragmented global perception of the war. In the Western world, the narrative is clear and supportive of Ukraine. In other parts of the world, the narrative is more complex, with some countries balancing both perspectives or favoring Russia's viewpoint.

This polarization of narratives underscores the challenges of maintaining truth and objectivity in wartime. Official narratives, shaped by state-controlled media, aim to either galvanize the population towards a specific goal or maintain compliance and support for state policies. The result is a landscape where truth becomes a tool of influence rather than an objective standard.

The preservation of truth and objectivity is thus largely compromised in the fog of war. Instead, what emerges are competing narratives designed to serve specific political and strategic interests. These narratives not only shape public opinion and international relations but also have lasting impacts on the collective memory and historical record of the conflict. Understanding these dynamics is crucial for comprehending the broader implications of information warfare and media manipulation in times of conflict.

### 6.3 Limitations and future work

**Limitations:**

One major limitation of my work on this thesis was my initial inexperience with NLP concepts and tools. However, over time, I gradually learned and improved my techniques and knowledge base.

Another limitation was the noisiness of the data. Despite my best efforts to minimize noise, the inherent nature of working with real-world data meant that some noise persisted. Nevertheless, I believe the data was sufficiently cleaned for the purposes of this analysis.

Technical limitations also posed challenges, particularly the large main memory and computational demands required to process the extensive dataset I gathered. This issue was partially mitigated by employing more efficient algorithms capable of handling large data volumes.

The contradictory presentation of events by Ukrainian and Russian media also posed a limitation in interpreting the experimental data. To address this, I adopted a predominantly objective interpretation, striving to present a balanced view despite the differing narratives.

**Future Work:**

For future work, I aim to further improve the de-noising process of my data and enhance the preservation of numbers that are part of words, such as "leopard2," "m1abrams," and "t72b3."

Additionally, I plan to experiment with combining words to treat them as single entities to observe the potential differences in results. Currently, when collecting adjectives for entity mentions, I break down the entity mentions into their base words and collect the top adjectives associated with each word. In the future, I intend to treat the entire entity mention as a single word, which may yield more representative adjectives and, consequently, more accurate sentiment analysis.

Exploring more experiments on my data, particularly in the realm of generative AI and visualization, would also be valuable. These approaches could provide deeper insights and more interesting analyses, enhancing our understanding of the narratives and sentiments in the media coverage of the conflict.

# References


[1] Nigel Walker, Conflict in Ukraine: A timeline (2014 - eve of 2022 invasion), House of Commons Library, Research Briefing, p. 4, 2023.

[2] Reuters, “Extracts from Putin's speech on Ukraine”, Reuters, https://www.reuters.com/world/europe/extracts-putins-speech-ukraine-2022-02-21/, 2022

[3] Reuters, “Pro-Russia Protesters Storm Donetsk Offices”, Reuters, https://www.nbcnews.com/storyline/ukraine-crisis/pro-russia-protesters-storm-donetsk-offices-n54046, 2014

[4] Roland Oliphant, Rozina Sabur, “Vladimir Putin admits: Russian troops 'were in Ukraine'”, The Telegraph ,https://www.telegraph.co.uk/news/worldnews/vladimir-putin/12054164/Vladimir-Putins-annual-press-conference-2015-live.html, 2015

[5] NATO, “Brussels Summit Communiqué”, https://www.nato.int/cps/en/natohq/news_185000.htm, NATO, point 69, 2021

[6] Mimi Dutta, “Word2Vec For Word Embeddings -A Beginner’s Guide”, https://www.analyticsvidhya.com/blog/2021/07/word2vec-for-word-embeddings-a-beginners-guide/, 2021

[7] Jason Brownlee, “How to Develop Word Embeddings in Python with Gensim”, https://machinelearningmastery.com/develop-word-embeddings-python-gensim/, 2020

[8] Various Authors, Documentation of spaCy’s EntityRecognizer, https://spacy.io/api/entityrecognizer/

[9] Federico Bianchi et al. Compass-aligned Distributional Embeddings for Studying Semantic Differences across Corpora, p. 15, PyPi of CADE: https://pypi.org/project/cade/, 2020

[10] Paul McLeary, Lara Seligman and Lee Hudson, “Negotiations on whether to send F-16s and Patriots to Ukraine continue but quietly”, Politico, https://www.politico.com/news/2022/09/16/f-16s-patriots-ukraine-negotiations-00057262, 2022

[11] Shaun Walker, “Azov fighters are Ukraine's greatest weapon and may be its greatest threat”, The Guardian, https://www.theguardian.com/world/2014/sep/10/azov-far-right-fighters-ukraine-neo-nazis, 2014

[12, 13] Associated Press, “Explainer: Who Were Mariupol's Last Defenders?”, Voice of America, https://www.voanews.com/a/explainer-who-were-mariupol-s-last-defenders-/6584122.html, 2022

[14] Interfax-Ukraine, “Pro-Russian activists in Donetsk prepare referendum for May 11”, Kyiv Post, https://archive.kyivpost.com/article/content/war-against-ukraine/pro-russian-activists-in-donetsk-prepare-referendum-for-may-11-343857.html, 2014

[15] RFE/RL's Ukrainian Service, “Pro-Russia Protesters Storm Government Buildings In Eastern Ukrainian Cities”, Radio Free Europe, https://www.rferl.org/a/ukraine-donetsk-government-buildings-stormed/25323219.html, 2014

[16] Arsalan Bilal, “Russia’s hybrid war against the West”, NATO Review, https://www.nato.int/docu/review/articles/2024/04/26/russias-hybrid-war-against-the-west/index.html, 2024

[17] Pavel Polityuk and Oleksandr Kozhukhar, “Some 1,000 Ukraine marines surrender in Mariupol, says Russia”, Reuters,

https://www.reuters.com/world/europe/ukraine-tells-russia-return-prisoners-if-you-want-top-ally-back-2022-04-12/, 2022

[18] Luke Harding , "Humiliation for Vladimir Putin as Ukrainians liberate key city of Lyman", The Guardian, https://www.theguardian.com/world/2022/oct/01/humiliation-for-vladimir-putin-as-ukrainians-liberate-key-city-of-lyman, 2022

[19] Al-Jazeera, "Putin announces Russian annexation of four Ukrainian regions", Al-Jazeera, https://www.aljazeera.com/news/2022/9/30/putin-announces-russian-annexation-of-four-ukrainian-regions, 2022

[20] Russian Ministry of Defense, "Russian Defence Ministry report on the progress of the special military operation in Ukraine (20 October 2022)", Russian MOD, https://eng.mil.ru/en/special_operation/briefings/briefings/more.htm?id=12442245@egNews, 2022

[21] Kristen D. Thompson, "What It Takes to Fly the F-16: Challenges for Ukraine", Council on Foreign Relations, https://www.cfr.org/expert-brief/what-it-takes-fly-f-16-challenges-ukraine#:~:text=Currently%2C%20Ukraine's%20aging%20fleet%20of,air%2Dto%2Dair%20combat., 2022

[22] European Commission, "Sanctions adopted following Russia's military aggression against Ukraine", https://finance.ec.europa.eu/eu-and-world/sanctions-restrictive-measures/sanctions-adopted-following-russias-military-aggression-against-ukraine_en, 2024

[23] Council of the European Union, "EU membership application", European Council, https://www.consilium.europa.eu/en/policies/enlargement/ukraine/, 2024

[24] Reuters, “Russian economy shrank by 1.2% in 2022, less than previously thought - stats service”, Reuters, https://www.reuters.com/markets/europe/russian-economy-shrank-by-12-2022-less-than-previously-thought-stats-service-2023-12-29/, 2023

[25] Nerijus Adomaitis and Johan Ahlander, “Nord Stream: What's known about the mystery pipeline explosions?”, Reuters, https://www.reuters.com/world/europe/qa-what-is-known-about-nord-stream-gas-pipeline-explosions-2023-09-26/, 2024

[26] Nigel Walker, Conflict in Ukraine: A timeline (2014 - eve of 2022 invasion), House of Commons Library, Research Briefing, p. 8-32, 2023.

[27] Andrew E. Kramer and Neil MacFarquhar, “Russia in Broad Retreat From Kyiv, Seeking to Regroup From Battering”, The New York Times, https://www.nytimes.com/2022/04/02/world/europe/ukraine-russia-kyiv.html, 2022

[28] Human Rights Watch, “Ukraine: Russian Forces’ Trail of Death in Bucha”, https://www.hrw.org/news/2022/04/21/ukraine-russian-forces-trail-death-bucha, 2022

[29] Marc Santora, Andrew E. Kramer, Dan Bilefsky, Ivan Nechepurenko and Anton Troianovski, “Russia Orders Retreat From Kherson, a Serious Reversal in the Ukraine War”, The New York Times, https://www.nytimes.com/2022/11/09/world/europe/ukraine-russia-kherson-retreat.html, 2022

[30] Pavel Polityuk and Vitalii Hnidyi, “EU grants Ukraine candidate status in 'historic moment'”, Reuters, https://www.reuters.com/business/aerospace-defense/ukraine-becomes-eu-membership-candidate-battle-east-enters-fearsome-climax-2022-06-23/, 2022

[31] Steve Hendrix, Robyn Dixon, Liz Sly, Serhii Korolchuk and Mary Ilyushina, "Russian troops in big retreat as Ukraine offensive advances in Kharkiv", The Washington Post, https://www.washingtonpost.com/world/2022/09/10/ukraine-kharkiv-russia-retreat-izyum/, 2022

[32] Dean Lockwood, "Abrams and Leopard 2 Tanks for Ukraine – Master Stroke or Colossal Blunder?", Defense and Security Monitor, https://dsm.forecastinternational.com/2023/01/26/abrams-and-leopard-2-tanks-for-ukraine-master-stroke-or-colossal-blunder/#:~:text=Defense%20Security%20Monitor-,Abrams%20and%20Leopard%202%20Tanks%20for,Master%20Stroke%20or%20Colossal%20Blunder%3F&text=On%20January%2025%2C%202023%2C%20the,aid%20package%20worth%20%24400%20million., 2023

[33] PBS, "Russia says it will reduce military operations near Kyiv as peace talks continue", https://www.pbs.org/newshour/world/russia-says-it-will-reduce-military-operations-near-kyiv-as-peace-talks-continue, 2022

[34] Merlyn Thomas, "Ukraine war: Putin orders partial mobilisation after facing setbacks", BBC, https://www.bbc.com/news/world-europe-62984985, 2022

[35, 39] TASS, "Forces regrouped at Balakleya, Izyum, redeployed to DPR, says Russian Defense Ministry", https://tass.com/defense/1505807, 2022

[36] The Moscow Times, "Russia Claims 'Successful' Advances Near Ukraine's Bakhmut", https://www.themoscowtimes.com/2023/02/07/russia-claims-successful-advances-near-ukraines-bakhmut-a80161, 2023

[37] Voice of America, "Putin Says West Playing 'Dangerous, Bloody Game'", https://www.voanews.com/a/putin-says-west-playing-dangerous-bloody-game-/6808544.html, 2022

[38] RTE, “EU agrees to impose price cap on Russian oil”, https://www.rte.ie/news/2022/1202/1339701-eu-russian-crude-cap/, 2022

[40] OCHA, “Joint Coordination Centre opens in Istanbul to facilitate safe export of commercial foodstuffs and fertilizers from Ukrainian ports”, https://reliefweb.int/report/turkiye/joint-coordination-centre-opens-istanbul-facilitate-safe-export-commercial-foodstuffs-and-fertilizers-ukrainian-ports, 2022