Title: A Comparison of Online News Media Framing of the 2022 Russia-Ukraine Conflict in Ukraine, Russia, the U.S. & China.

July 2022

University of Glasgow ID: 2574393Q

Dublin City University ID: 20109351

Charles University ID: 18329097

Appendix A: Codebook for Online Media Frame Analysis

Table of Contents

1. Introduction

This code book describes the content, structure and layout of the data collected for the comparison of online news media framing of the 2022 Russia-Ukraine conflict.

2. Research Objectives

2.1. Research Question

The main research question for this project is "what are the differences and similarities between the online news media's framing of the 2022 Russia-Ukraine conflict in Ukraine, Russia, the United States and China"?

2.2. Aims & Objectives

This research project aims to combine both quantitative and qualitative empirical data about online news reporting on the 2022 Russia-Ukraine conflict in order to compare how the conflict has been framed within on websites predominantly viewed by the respective audiences in Ukraine, Russia, the United States and China. The quantitative data comprises the news stories and their corresponding metadata while the qualitative data consists of academic literature, government, news reports and organizational publications on Russia-Ukraine relations prior and during the conflict. This project examines online news reports from 1st December 2021 when warnings of a potential Russian invasion of Ukraine started to emerge in international news to 30th April 2022, an end date chosen purely chosen for practical reasons to allow adequate time to meet this project's deadline.

This project seeks to:

i. Identify empirical differences and similarities in the framing of 2022 Russia-Ukraine conflict by the online news media within Ukraine, Russia, the U.S. and China through the use of Artificial-Intelligence based Sentiment Analysis and Topic Extraction software.

ii. Compare and contrast the online news media's framing in terms of the emphasis, sentiments expressed, phrasing and terminology used as well as the volume of the reporting.

iii. Interpret the identified empirical similarities and differences in the online media framing through the lens of post-soviet Russia-Ukraine relations along with the other major geostrategic actors like the U.S. and China.

3. Data Sampling

3.1. Sources

The choice to restrict the timeframe of online news articles relating to the 2022 Russia-Ukraine war from 1st December 2021 to 30th April 2022 was informed largely by practical considerations concerning this research project's submission deadline. The earliest warnings that Russia was planning to invade Ukraine started appearing on major international news websites like the BBC (2021), the Washington Post (Harris and Sonne, 2021), Al-Jazeera (Varshalomidze and Child, 2021) and the Associated Press (Madhani et al., 2021) in the first week of December 2021. While the full-scale conflict commenced after Russia invaded Ukraine on 24th February 2022, there was significant news media interest in the build-up to the invasion from early December 2021. Although the war in Ukraine was still raging at the end of April 2022, with no clear indication of when it would cease, the end point for the news articles' publication date was set to 30th April 2022 to ensure ample time to process and analyse the research data before this research project's submission deadline.

3.2. Date Range

The criteria for selecting which news websites to collect articles relating to the 2022 Russian-Ukraine conflict was based on three main criteria. Firstly, the news website had to have the majority of its audience traffic coming from one of the countries of interest. This was verified by checking website analytics

reporting companies like Alexa Analytics and SimilarWeb. Secondly, the news website had to have news stories relating to the 2022 Russia-Ukraine war which were either searchable using the relevant keywords or had a dedicated section of the website for the war as was evident in many Ukrainian news websites. Lastly, the website had to technically and legally permit the download of its content for academic research purposes, or at least not explicitly prohibit the latter. Based on these criteria, three news websites were selected each from Ukraine, Russia, the U.S. and China. The limit to three website per country was due to practical reasons associated with the quantity of data, cost of data extraction and time involved in processing and analysing the data.

4. Sampled Data

The sampled news articles about the 2022 Russia-Ukraine conflict consisted of 24,422 stories obtained from 12 news websites from Ukraine, Russia, the U.S. and China. 24.00% of the stories were published by Ukrainian news websites while 49.17% were posted on Russian News websites. 23.59% were from American news websites and only 3.23% from Chinese news websites. The stories were published between 1st December 2021 and 30th April 2022. While 38.29% of the news articles were originally published in English, 10.17%, 49.40% & 2.15% were originally published in Ukrainian, Russian and Mandarin respectively and needed to be translated into English using the Google Translate language translation software. Table 1.1 shows the distribution of the news articles per website source country.

Countries	**Article Count**	**Percentage**
Ukraine	5,861	24.00%
Russia	12,009	49.17%
USA	5,762	23.59%
China	790	3.23%
Total	**24,422**	**100%**

Table 1.1

5. Data Collection Process

The data collection process involved using the OctoParse web content extraction software to download the metadata on the selected news websites. The metadata collected included the headline of the news article, publication data, source Uniform Resource Locator (URL), the publishing website's primary domain name and the language the article was originally published in. The data collection sequence was as follows:

i. Search for articles on the website related to the 2022 war in Ukraine by sequentially entering terms such as "War + Ukraine" or "Russia invasion + Ukraine" into the website's search bar and proceeding to the results page. Some news websites however, had sections dedicated to the 2022 war in Ukraine which displayed all news articles related to the war, effectively making it easier to find the relevant news articles.

ii. The search results were then filtered by publication date to ensure the news articles were between the desired time period, i.e. from 1st December 2021 – 30th April 2022.

iii. The web address of the final search results page was then parsed to the OctoParse software for the metadata extraction.

iv. The metadata of news articles extracted from non-English language websites were then translated into English using the Google Translate software. The news source languages that required translation into English were *Ukrainian*, *Russian* and *Mandarin (Chinese).*

v. All the news article headlines were run through the MeaningCloud Sentiment Analysis software to identify the polarity of the words the terms and phrases that constitute each news article title. Subsequently, the MeaningCloud Topic Extraction software identified the individual entities and concepts within each news article's title.

6. News Articles Metadata Variables

Table 1.2 shows details about the news article dataset's meta-data including the variable names, description, data type as well as the variable formats and examples of each variable within the dataset.

Variable Name	Description	Variable Data Type	Variable Format/ Example	Source of Variable Value
Article Title	The title or heading of the news article	Text	e.g. Putin's war on verge of expanding outside Ukraine as it cries 'terrorism' in Moldova	Extracted directly from news websites
Publication Date	The date the article was published	Date	Format: dd/mm/yyyy eg. 22/03/2022	Extracted directly from news websites
Article URL	The universal resource locator link to the news article	Text (Web link)	Eg. https://www.pravda.com.ua/eng/news/2022/04/5/7337312/	Extracted directly news websites
Source	The website the article was obtained from	Text (Web Link)	Eg. PRAVDA.COM.AU	Extracted directly from news websites
Country	The country the source of the news article is based in	Text	Eg. Ukraine, Russia, China or USA	Identified through research
Original Language	The original language of the news article	Text	Eg. English, Ukrainian, Russian or Chinese	Listed on news' source website
Translation Tool	The language translation tool or software used	Text	Google Translate	Google Translate
Summary/Intro	The summary or introductory text to the news article	Text		Extracted directly from news websites

Table 1.2

7. Dataset File Structure

The dataset is broken into 8 Microsoft Excel (.XLSX) document stored within one folder. The filename and content of each file is explained in Table 1.3:

#	File Name	Content Description
1	1. Ukraine - All News Combined [Sentiments Analysis].xlsx	Online news article headlines from Ukraine and associated sentiment polarity output
2	2. Ukraine - All News Combined [Topic Extraction].xlsx	Online news article headlines from Ukraine and extracted entities and concepts
3	3. Russia - All News Combined [Sentiments Analysis].xlsx	Online news article headlines from Russia and associated sentiment polarity output
4	4. Russia - All News Combined [Topic Extraction].xlsx	Online news article headlines from Russia and extracted entities and concepts
5	5. USA - All News Combined [Sentiments Analysis].xlsx	Online news article headlines from the U.S. and associated sentiment polarity output
6	6. USA - All News Combined [Topic Extraction].xlsx	Online news article headlines from the U.S. and extracted entities and concepts
7	7. China - All News Combined [Sentiments Analysis].xlsx	Online news article headlines from China and associated sentiment polarity output
8	8. China - All News Combined [Topic Extraction].xlsx	Online news article headlines from China and extracted entities and concepts

Table 1.3

Table 1.4 shows an example of the metadata of a single online news article within the dataset.

#	Title	Publication Date	Article URL	Source	Country	Original Language	Translation Tool	Summary/Intro	Sentiments
1079	General Staff: Russian troops have lost 815 tanks and 150 helicopters since the beginning of full-scale aggression	20/04/2022	https://www.pravda.com.ua/eng/news/2022/04/20/7340862/	PRAVDA.COM.AU	UKRAINE	ENGLISH	NONE	-	N

Table 1.4

8. Sentiment Analysis

Sentiment analysis, also referred to as opinion mining, is a field of study for analysing opinions expressed in a text using a lexicon-based algorithmic model (Alam and Yao, 2018). The sentiment analysis computations on the news article headlines analysed in this project was conducted by using the MeaningCloud sentiment analysis software. This sentiment analysis software detected the polarity contained in each news article's heading (MeaningCloud, n.d.). The polarity output of the software for a given text, in this case a news article headline, is classified into six results: positive (P) and negative (N), very positive (P+) and very negative (N+), neutral (NEU) and none (NONE). The polarity of the text is determined through what MeaningCloud describes as 'Advanced Natural Language Processing' that evaluates all aspects of the text including the morphology, syntax and semantics (MeaningCloud, 2015). Table 1.5 shows examples of news articles each type of sentiment polarity.

Online News Article Headline	Sentiment Polarity
Putin's war on verge of expanding outside Ukraine as it cries 'terrorism' in Moldova	N
Ukraine war photos: Horrific findings after Russian retreat from Kyiv outskirts	N+
US pledges to help Ukraine shore up border with Russia, Belarus: report	P
Examples of the courage of the Russian military in the Donbass	P+
Sen. Hawley calls on Biden to suspend support for Ukraine NATO membership	NEU
Biden says sending US troops to Ukraine 'not on the table'	NONE

Table 1.5

For the purposes of this study, some of the sentiment polarity output were combined to facilitate analytical expediency. Specifically, Very Positive (P+) was combined with the Positive (P), while Very Negative (N+) was combined with Negative (N). Finally, None (NONE) was combined with Neutral (NEU).

9. Topic Extraction

Topic extraction is the process of using software to automatically identifying the underlying elements and concepts contained in a document or text corpus (Jiang et al., 2018). This is done by using Artificial Intelligence and Machine Learning based linguistic and thematic models to identify the composite topics within the given text. The identified individual topics are further classified into either entities or concepts based on the model's pre-configuration and the nature of each identified topic. For example, the result of a topic extraction output for the news article headline "Britain will provide additional military aid to Ukraine - Johnson" would be the entities "Ukraine", "Britain" and ""Johnson" as well as the concept "military aid". Such an automated identification process for the news headline topics help to minimize the arbitrary tendencies that often plague many media frame identification research projects as previously highlighted by manual and hermeneutic approaches to Content Analysis.

9.1. The Topic Extraction Output

The data fields of the Topic Extraction output on each row of online news headline metadata include text (news article headline), form (entity or concept), the rank, type and a sense-id. Table 1.6 provides the definition of each of these fields.

#	Field of Topic Extraction Output	Definition
1	Text	The headline of the online news article
2	Form	Displays the name by which the topic extracted is identified. It's not configurable, so it always appears in the results.
3	Rank	Contains the order in which the topics have been detected. It's specific for each type of topic, that is, the first entity detected will be ranked 1, and the first concept will be ranked 1 too, and so on.
4	Type	Shows the type associated to the topic according to our ontology.
5	Sense ID	The unique id of the topic allocated by the MeaningCloud software

Table 1.6

Table 1.7 shows an example of the Topic Extraction output on a given online news article headline.

Text	Form	Topic Category	Rank	Type	Sense ID
Zelensky to Russian soldiers: You are better to survive in Russia than to die in Ukraine	Ukraine	entity	1	Top>Location>GeoPoliticalEntity>Country	7a2664fc70
	Zelensky	entity	2	Top	__14938209558940374195
	Russia	entity	3	Top>Location>GeoPoliticalEntity>Country	d2aceb9067
	soldier	concept	1	Top>OtherEntity>Vocation	1f59fae897
	soldier	concept	2	Top>Person	8aa565ccc9
	Russian	concept	3	Top	b41b92e033

Table 1.7

10. Data Analysis

The analysis of the collected web-based news metadata for this study was conducted in four main phases, each examining an indicator of frame alignment processes as stated by Snow et al (1986). The selected indicators were emphasis on specific issues and topics, sentiments expressed, volume of media coverage as well as the phrasing & terminology used. Together, these indicators form a robust framework for constructing the different aspects of the media framing of the conflict within the sampled news article dataset.

Analysing the emphasis of the online news reporting on the war, largely through frequency analysis of the extracted entities and concepts from the news article headings, was key to identifying and understanding the various aspects of the media framing that received the most focus from the news publishers. Assessing the sentiments expressed within the news headlines using the process of Sentiment Analysis enabled this study helped to establish the viewpoint of the news websites on the specific stories published and thus providing another aspect of the media framing process.

Examining the volume of online media coverage was critical for gauging the degree of frame amplification which Snow et al (1986) describe as the clarification and invigoration of an interpretive frame that bears on a particular issue, problem, or set of events". Finally, by analysing the phrasing and terminology of the news headlines relating to the 2022 Russia-Ukraine conflict identified which aspects of pre-existing notions and concepts the online news media drew upon to construct the framing of their coverage of the war, deliberately or inadvertently. The research findings contained in this study are therefore based on an aggregated analysis of the aforementioned indicators of the frame alignment process.

11. References

Alam, S. and Yao, N., (2018). 'The Impact of Preprocessing Steps on the Accuracy of Machine Learning Algorithms in Sentiment Analysis'. Computational and Mathematical Organization Theory, 25(3), pp.319-335.

BBC, (2021). 'Russia Ukraine: Biden Warns Russia Against Ukraine 'Red Lines''. [online] BBC News. Available at: <https://www.bbc.com/news/world-us-canada-59528864> [Accessed 9 February 2022].

Harris, S. and Sonne, P., (2021). 'Russia Planning Massive Military Offensive Against Ukraine Involving 175,000 Troops, U.S. Intelligence Warns'. [online] The Washington Post. Available at: <https://www.washingtonpost.com/national-security/russia-ukraine-invasion/2021/12/03/98a3760e-546b-11ec-8769-2f4ecdf7a2ad_story.html> [Accessed 12 March 2022].

Madhani, A., Knickmeyer, E., Merchant, N. and Jaffe, A., (2021). 'U.S. Intelligence Finds Russia Planning Ukraine Offensive'. [online] The Associated Press. Available at: <https://apnews.com/article/joe-biden-europe-russia-ukraine-vladimir-putin-9ef2a191af25e2920a0a1c75ccbe4c13> [Accessed 6 March 2022].

MeaningCloud, (n.d). 'Text Analytics'. [online] MeaningCloud. Available at: <https://www.meaningcloud.com/solutions/text-analytics> [Accessed 16 February 2022].

Snow, D., Rochford, E., Worden, S. and Benford, R., (1986). 'Frame Alignment Processes, Micromobilization, and Movement Participation'. American Sociological Review, 51(4), p.464-481.

Varshalomidze, T. and Child, D., (2021). 'US Warns Russia Over Ukraine As Fears of War Rise'. [online] Al Jazeera. Available at: <https://www.aljazeera.com/news/2021/12/2/us-russia-talks-as-ukraine-tensions-mount-live-updates> [Accessed 11 March 2022].