

**UNIVERSITY
OF TWENTE.**



Master Thesis

Russia's disinformation war

-

An analysis of Ukraine themed Russian disinformation campaigns during
Russia's invasion of Ukraine in 2022

Philipp Weigand

Submission: 05.07.2024

Word count: 16.134

University of Twente

European Studies (M.Sc.)

1st Supervisor: Dr. Le Anh Nguyen Long

WWU Münster

Comparative Public Governance (M.A.)

2nd Supervisor: Prof. Dr. Bernd

Schlipphak

Abstract

Russian disinformation campaigns are a threat to democratic societies. They undermine the credibility of foreign institutions, shift the perception of truth and create confusion among the societies of Russia's declared adversaries. This thesis analyzes Russian disinformation before and after the Russian invasion of Ukraine in 2022, as the Russo-Ukraine conflict escalated from a Hybrid to a conventional war. For this, a quantitative corpus-based analysis was conducted to compare disinformation regarding changes in quantity, topics and sentiment before and after the Russian invasion of Ukraine. The analyzed disinformation was thereby retrieved from the EU Commission's EUvsDisinfo-database. The analysis was able to detect changes in the quantity and topics of the Ukraine themed disinformation, while the tested sentiment of the Russian disinformation stayed equally negative over the course of the inspected timeframe. The detected changes proved a shift of the Russian disinformation away from multiple topics, focusing on Ukraine-conflict related topics. Yet, these changes have been expected, as a review of previous Russian disinformation campaigns suggested that the central focus of Russian disinformation campaigns have aligned with Russian foreign policy interests before.

Keywords

Russia – Ukraine – Russian disinformation campaigns – Russo-Ukrainian War – Russian invasion of Ukraine – East StratCom Task Force – corpus analysis – sentiment analysis – Python – NLTK – VADER

“It is man's unique privilege, among all other organisms. By pursuing falsehood, you will arrive at the truth!”

— Fyódor Mikháylovich Dostoyévskiy, *Crime and Punishment*

Table of abbreviations

ASCII	American Standard Code for Information Interchange
EEAS	European External Action Service
EU	European Union
NATO	North Atlantic Treaty Organization
NLP	Natural Language Processing
NLTK	Natural Language Tool Kit
OECD	Organisation for Economic Co-operation and Development
RoBERTa	Robustly Optimized BERT Pretraining Approach
RT	Russia Today
TASS	Information agency of Russia
US/USA	United States of America
VADER	Valence Aware Dictionary and sEntiment Reasoner
VPN	Virtual Private Network

Table of contents

Abstract	I
Quote	II
Wordclouds.....	III
Table of abbreviations.....	IV
Table of contents	V
1 Introduction	1
2 Contextual overview.....	3
2.1 Distinction of disinformation to related concepts.....	3
2.2 Significance of Disinformation	4
3 Case Background.....	7
3.1 Russian Disinformation	7
3.1.1 Current trends.....	7
3.1.2 Recent examples.....	9
3.1.3 Russian motives.....	10
3.2 Russian-Ukrainian Conflict: From a hybrid to a conventional war	11
4 Literature Review	14
5 Methodology	19
5.1 Research design	19
5.2 Theory of corpus-based analyses.....	21
5.2.1 Python’s “Natural Language Tool Kit” (NLTK)	21
5.2.2 Methods of preprocessing text	22
5.2.3 Keyword, bigram, collocation and topic analyses	22
5.2.4 Sentiment analysis with the “Valence Aware Dictionary and sEntiment Reasoner” (VADER)	23
5.3 Data: EUvsDisinfo.eu’s Disinfo database.....	24

6	Operationalization.....	25
6.1	Mining the data and building the corpora	25
6.2	The Syntax.....	26
6.3	Limitations.....	27
7	Analysis.....	29
7.1	Analysis I: quantitative comparison	29
7.1.1	Corpus I – pre invasion	29
7.1.2	Corpus II – post invasion	30
7.1.3	Discussing hypothesis 1	31
7.2	Analysis II: topic analysis	32
7.2.1	Analysis of the most common words	33
7.2.2	Analysis of the most common bigrams	35
7.2.3	Topics derived from keywords and bigrams	37
7.2.4	Analysis of the EUvsDisinfo Keywords.....	39
7.2.5	Discussing hypothesis 2.1.....	41
7.2.6	Discussing hypothesis 2.2.....	42
7.3	Analysis III: sentiment analysis	43
7.3.1	Analysis of the overall sentiment	43
7.3.2	Analysis of the Ukraine specific sentiment.....	45
7.3.3	Discussing hypothesis 3	46
7.4	Observations regarding the data.....	47
8	Interpretation.....	49
9	Conclusion	52
	Table of figures.....	VIII
	References.....	IX
	Declaration of independent work.....	XVII
	Attachments.....	XVIII

1 Introduction

The proliferation of disinformation campaigns is one of the most significant challenges to global security, democratic processes and societal stability (Baumann, 2020; European Commission, 2018; Legucka, 2020; Shires 2021, 257). The European Commission thereby defines disinformation in their “Action Plan against Disinformation” from 2018 as “*false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm.*” (European Commission, 2018a, p. 1) Public harm is specified as threats to democratic, political and policy-making processes, but also as threats to the health of EU citizens, their environment or their security. (European Commission, 2018a, p. 1) The Commission’s definition coincides with the dictionary definition of disinformation as “*false information that is given deliberately, especially by government organizations*” (Oxford Advanced American Dictionary, 2024) as well as with the established definition in the scientific literature (Wardle, 2018, p.954; Lecheler & Egelhofer, 2022, p. 71).

Arguably, one of the most prominent actors in the state-run dissemination of disinformation is Russia (Bennett & Livingston, 2021, p. 7; Barnes, 2021; OECD, 2022). Confirmed examples can showcase modern strategies of Russian disinformation campaigns like amplifying existing topics to undermine the credibility of foreign institutions, perpetuating the belief that there is no objective truth and creating confusion among its declared adversaries (Nimmo, 2015; Nehring, 2017; Zelenkauskaitė, 2022).

During the Russian invasion of Ukraine on the 22nd February 2022 multiple cases of Russian disinformation could be observed (Geissler, Bär, Pröllochs & Feuerriegel, 2023, p. 1). Yet, according to Geissler et al. (2023) no empirical studies of these disinformation campaigns have been conducted. Meanwhile, with the invasion, the conflict changed from a Hybrid to a conventional war (Bankov, 2023, p. 320; Ionita, 2023, p. 6). Considering the changing of the conflict, the question arises, whether the disinformation changed as well. Hence, the research question guiding this thesis is: *What are the changes in Ukraine themed Russian disinformation before and after the Russian invasion of Ukraine in 2022?*

To compare the possible changes in disinformation before and after the Russian invasion of Ukraine, a quantitative corpus-based case-study of Russian disinformation campaigns will be conducted. Therefore, two corpora will be analyzed with Python's "Natural Language Toolkit", comparing Russian disinformation that was registered by the European Commission's East StratCom Task Force before the invasion (Corpus I) and after the invasion (Corpus II). The goal of this case-study is to compare the two corpora regarding possible changes in quantity, topics and sentiment of the disinformation.

This research will fill the gap of analyzing how a significant event such as the Russian invasion of Ukraine impacts Russian disinformation and thereby help to acquire a better understanding of Russian disinformation campaigns. Adding to the toolkit of understanding disinformation is important, as according to Geissler et al. studies are missing, even though disinformation poses a threat to society. Nevertheless, to not exceed the scope of this thesis, the data-set of the case-study is limited to a relatively small data sample. Future studies may test the methods of this thesis on a larger scale.

To approach this research interest, first a contextual overview will clarify the concepts of disinformation and related terminology. Following, a literature review regarding Russian disinformation in particular and including the current state of research will guide to the research gap and the hypotheses. Afterwards, a methodology chapter will explain the design and theoretical aspects of the research, before the operationalization chapter will describe the actual process of conducting the case-study. Following, the analysis and the results of the case-study will be presented before the results will be discussed. Finally, the conclusion will complete the thesis by summarizing the research and answering the research question.

2 Contextual overview

2.1 Distinction of disinformation to related concepts

Claire Wardle (2018) specifies the definition of disinformation by distinguishing three “types” of information disorder, namely “malinformation”, “misinformation” and “disinformation”. She defines malinformation as the deliberate publication and distribution of true information with the intention to harm. Examples for this deliberate cause of harm are the leaking of hacked mails or incriminating documents, but also the deliberate change of context of genuine content. Misinformation occurs when false information is shared but no harm is intended, for example when unintentional mistakes like inaccurate dates, translations, statistics or satire is taken seriously. Disinformation on the other hand is described as “fabricated or deliberately manipulated audio/visual content” that is deliberately false and intending to harm, like intentionally created conspiracy theories or rumors. (Wardle, 2018, p. 954)

The two terms fake news and propaganda are often used interchangeably with the term disinformation in the scientific literature (Lecheler & Egelhofer, 2022, p. 70). Hence, a distinction of the terminology is necessary, to complete the definition of disinformation. The “National Endowment for Democracies” describes fake news as misleading or false information that is presented as news, most commonly within internet sources and on social media platforms. As a distinction to disinformation in general, they add that fake news may have political implications, but its primary motive is usually financial rather than political and usually not tied to a larger agenda. Fake news is achieving this by targeting user preferences to generate interactions and advertising revenue. (National Endowment for Democracies, 2017) Lecheler & Egelhofer (2022) classify fake news as a genre of disinformation that is spread in journalistic format. Meanwhile, Wardle (2018) refrains from using the term “fake news” and declares it as too simple of a term for a complex topic. Moreover, she adds to the previous distinction that the term only focuses on text-based information disorder and not the varying other forms of information disorder that were quoted in the list in the previous chapter, like visual posts or offline events. Finally, she claims that the term is being appropriated by news organizations and politicians to describe disagreeable coverage. Therefore, Wardle concludes that the term is not suited for the comprehensive analysis of information disorder. (Wardle &

Derakshan, 2017, p. 5; Habgood-Coote, 2019, p. 1034) While it can also be argued that the intent of the messenger is the crucial distinction between different types of messages (National Endowment for Democracies, 2017), for the following thesis Claire Wardle's definition of disinformation disorder will be adopted and the term of fake news disregarded for the research intent of this thesis.

In the academic literature it is debated, whether propaganda and disinformation are two distinct concepts or whether they overlap and propaganda can be seen as the parental concept of information disorder and disinformation (National Endowment for Democracies, 2017; Lecheler & Egelhofer, 2022, p. 72). Jowett & O'Donnell (2019) define propaganda as *"the deliberate, systematic attempt to shape perceptions, manipulate cognitions, and direct behavior to achieve a response that furthers the desired intent of the propagandist"* (Jowett & O'Donnell, 2019, p. 6). When distinguishing the two terms, disinformation, as previously described, refers specifically to the intentional spread of false or misleading information. Additionally, the goal of the politically motivated disinformation is to disincentivize specific action and rather create public cynicism, uncertainty, apathy, distrust and paranoia. (Wardle, 2018, p. 954; Lecheler & Egelhofer, 2022, p. 72) Meanwhile, propaganda does not necessarily utilize false information, but selectively presents facts to encourage the wanted perception of a topic. Here the goal is to encourage action by appealing to emotions, biases and preconceptions with a broader range of tactics and techniques used to influence public opinion and emotions in support of a particular cause or agenda. (National Endowment for Democracies, 2017) Summarized, any type of information can be used for propaganda, while disinformation can be seen as a category that developed out of propaganda (Lecheler & Egelhofer, 2022, p. 72).

2.2 Significance of Disinformation

The European Council highlights the significance of disinformation by claiming that *"disinformation is an evolving challenge, with high potential to negatively influence democratic processes and social debates"* (European Commission, 2018b). Furthermore, the European Commission declares that *"democratic processes are increasingly challenged by deliberate, large-scale, and systematic spreading of disinformation"* (European

Commission, 2018a, p. 1). Moreover, the European Council (2018) claims that non-partner countries of the European Union are showing an increased usage of malicious online activities. Therefore, all different stakeholders in a democracy like “*governments, private sector, civil society, technical community, users and academia*” (Council of the European Union, 2018, p. 2) need to work together to fight the evolving digital harm.

Further literature like Nehring (2017) adds that internet and social media provide access to a global audience and enables to politically influence processes all around the world. The anonymity offered by the internet and the rapid news spread of modern media, thereby makes it difficult to trace the original source and allows disinformation at times to go undetected for days or weeks. (Nehring, 2017, p. 447) Exploiting the concept of freedom of information, malicious actors inject disinformation into societies, relying on conspiracy theories and spreading false narratives rather than persuading or earning credibility. (Pomerantsev & Weiss, 2014, p. 6)

The Council of the European Commission adds in the April Comment from 2018 “Council Conclusions on malicious cyber activities” that key values from the European Union, for example civil rights and the rule of law, are the main reason to maintain democratic organizations. (Council of the European Union, 2018, p. 2) The Commission claims in its 2018 action plan against disinformation that these European key values need to be protected from disinformation (European Commission, 2018a, p. 1).

Disinformation has been a part of human communication since ancient Rome. With the development of communication technologies, like the printing press and later the radio and television, the means to disseminate disinformation increased and so did the amount of disinformation. (Posetti & Matthews, 2018, 1) Today, in the age of modern communication, disinformation is omnipresent in societies around the world (Bennett & Livingston, 2021, XV). Bennett & Livingston argue that disinformation supports a transition to more authoritarian regimes, naming “illiberal” democracies like Hungary or Turkey as examples. Further, they highlight the problems of disinformation in countries like Great Britain with the example of the disinformation surrounding the Brexit or Germany, with the right-wing party Alternative für Deutschland spreading disinformation about topics like migration or climate change. (Bennett & Livingston, 2021, 3)

Other countries like China, North Korea, Russia and Iran are also named as current examples for countries of origin of disinformation. (Bennett & Livingston, 2021, p. XVIII) Yet, arguably the most prominent actor in the landscape of information disorder is Russia (Bennett & Livingston, 2021, p. 7; Barnes, 2021; OECD, 2022) To combat disinformation, the European Commission founded the East StratCom Task Force in 2015. Here, the European Commission decided to put its focus on ongoing disinformation campaigns from Russia. (EEAS, 2021) The aim of the East StartCom Task Force is to reveal, analyze and increase attention as well as the understanding of disinformation campaigns. Furthermore, the Task Force declares that it *“raises awareness of the negative impact of disinformation that originates in pro-Kremlin sources and is disseminated in the Eastern neighborhood’s information space and beyond”*. (EEAS, 2021)

The academic literature on the significance of Russia's disinformation campaigns presents a range of perspectives. Some argue that these campaigns are not inherently dangerous to society, but need to be closely monitored (Lanoszka, 2019, p. 227; Erlich & Garner, 2023, p. 5; Hameleers, 2023, p. 5). Other authors, like Snyder go as far as to attribute all negative events in the liberal normative world to Russian disinformation (Pacepa and Rychlak, 2013; Snyder, 2018) Overall, the actual effectiveness of disinformation is a subject of debate, with varying interpretations among authors. (Hung, 2022, p. 77) Nevertheless, the significance and risk that disinformation possesses to democratic societies remains undisputed.

The European Union’s effort to combat disinformation is a good example for the significance of disinformation. The focus of the Commissions efforts to combat Russian disinformation by founding the East StratCom Task Force underscores that a major actor of disinformation campaigns can be defined as Russia. Therefore, the focus on this thesis will be on Russian disinformation. Moreover, this thesis will use the database from the East StratCom Task Force for the analysis of this thesis.

3 Case Background

3.1 Russian Disinformation

3.1.1 Current trends

Nimmo (2015) claims that Blaming the West¹ is Russia's² main offensive weapon when conducting disinformation campaigns. (Nimmo, 2015) He elaborates that Russia spreads doubt on existing information structures to dismiss Western criticism of its own illiberal policies. Thereby, Russia aims to undermine the credibility of foreign institutions and create confusion. This strategy serves to deflect attention from domestic issues and dismiss Western claims of Russian repression. (Nimmo, 2015)

Pomerantsev and Weiss add that unlike during the Cold War era, the Kremlin now adopts a flexible approach to ideology, simultaneously supporting far-left and far-right movements, as well as other groups, to exacerbate divisions and create an echo chamber of support for the Kremlin's agenda. (Pomerantsev & Weiss, 2014, p. 5) Further examples for contradicting actions are selectively censoring online information, while providing a welcoming platform for individuals like whistleblower Julian Assange and propagating transparency or claims to reject corporate greed and applaud movements like Occupy Wall Street, all while governing over a corrupt economy. (Pomerantsev & Weiss, 2014, p. 5) Moreover, this approach justifies the invasion of a neighboring country using a combination of contradicting blood-and-soil nationalism and anti-fascist narratives. (Pomerantsev & Weiss, 2014, p. 5) Mölder & Shiraev (2021) add that these multifaceted

¹ When mentioning "The West", this thesis is referring to the cultural, political, and economic framework associated with Western Europe, North America, and other regions influenced by Western European traditions (McFaul, 2020, p. 95). At its core, "the West" is often characterized as a collection of countries and societies that share certain values, political ideologies, institutions and historical experiences like democratic governance, respect for human rights, individual freedoms, rule of law and market-based economies. (Eitelhuber, 2009, p. 2) Even though the West cannot be treated as a monolithic actor, as it was often considered to be during the Cold War, for this thesis the West will be considered as the antipode that Russia declares it to be in its foreign policy. (McFaul, 2020, p. 95)

² When mentioning Russia in this thesis, it is not intended to generalize and tie the actions of the Russian Federation's leadership to an entire people. Historical, cultural, and geopolitical factors contribute to the understanding of Russia, its role as the political successor of the Soviet Union and its complex relationship with its past. (Lutz, Lutz & Lutz, 2019, p. 84) For the sake of readability, when mentioning Russia in this thesis, Russia is meant as a generic construct. Nevertheless, it is important to note that the concept of Russia and its historical, cultural, and geopolitical understanding can be nuanced and subject to different interpretations and perspectives. (Kara-Murza, 2017, p. 114)

characteristics of the system, encompassing elements that appeal to American social reactionaries, Australian anarchists, British anti-imperialists, and Hungarian neo-Nazis, contribute to the widespread endorsement of the disseminated narratives both within and beyond Russia. Today, Russian disinformation is a flexible concept that accommodates the diverse preferences and agendas of its supporters, allowing them to shape its definition according to their own desires and objectives. (Pomerantsev & Weiss, 2014, p. 5; Mölder & Shiraev, 2021, p. 19)

Meanwhile, when defending Russia, Nimmo explains that the Kremlin resorts to the “4D approach” of dismissing critics, distorting facts, diverting attention from the main issue, and creating disillusionment among the audience (Nimmo, 2015). When faced with negative reporting or criticism, Russia often dismisses it by denying allegations or attacking the credibility of the source. A fitting example is the unlawful annexation of Crimea, when Russia consistently denied to have troops in Ukraine, despite evidence suggesting otherwise. (Nimmo, 2015; Treverton, Thvedt, Chen, Lee & McCue, 2018, p. 16)

The distorting of information to serve a narrative can be shown by the example of Deputy Defense Minister Anatoliy Antonov manipulating a news story in 2015 that NATO would station 30'000 troops in Poland when in fact it was only 13'000. (Nimmo, 2015; Zelenkauskaite, 2022, 126)

The rule that for a disinformation message to succeed it must at least partially correspond to reality or generally accepted views is thereby an established and proven strategy from the Soviet era. (Legucka, 2022, p. 23)

The strategy of “distraction” can be observed during the handling of the MH-17 disaster, when Russia used distraction techniques to divert attention from their own activities by launching accusations elsewhere, to create doubt and confusion. Instead of investigating the culprits of the plane downing, various narratives were disseminated by state-controlled Russian media outlets like the plane being downed by a Ukrainian missile as to a failed assassination attempt on Putin. (Nimmo, 2015; Tsybulenko & Kajander, 2021, pp. 182-183)

Lastly, Russia spreads dismay by issuing warnings of disastrous consequences to those who oppose their actions, often using nuclear threats as a deterrent (Nimmo, 2015).

3.1.2 Recent examples

An example of a Russian disinformation attack on a Western society is the case of Lisa F. (Janda, 2016, p. 1) Lisa, a 13-year-old Russian German girl was reported missing in January 2016. After her return, she claimed that she had been kidnapped and raped by three unknown individuals of Arab origin. Russian officials and media quickly seized on this case, using it to accuse Germany of tolerating and concealing child abuse. These accusations were amplified in an online outrage, by the spread of hate comments from fake accounts and bots, as well as by emotionalized citizens. (Meister, 2016) The case led to demonstrations in various German cities, mostly by the Russian-German community as well as to diplomatic tensions between Russia and Germany. Meanwhile, the police soon proved that the kidnapping story was false, based on their analysis of mobile phone records. Furthermore, Lisa herself admitted that she had voluntarily gone into hiding. (Meister, 2016; Janda, 2016, p. 2) According to Janda (2016), one aim of this disinformation attack was to disrupt up the German society, by further promoting xenophobia and amplifying and emotionalizing the opinions during the refugee crisis, which was agitating the German society at the time. Another aim was to disqualify the Western ideals of tolerance, while positioning Russia as morally superior protector of traditional values. (Meister, 2016; Janda, 2016, p. 3)

Another addition to this specific example, is “Operation Ghostwriter”, a campaign, seeking to sow distrust about NATO among the populations of former Warsaw Pact countries (Greenberg 2020; Zelenkauskaite, 2022, 126). Here, Russia started launching multiple disinformation campaigns targeting NATO and other countries that could potentially align themselves with Ukraine. In one of these campaigns, the “internet research agency”³ spread fake news about NATO on hacked news sites and blogs in the Baltic states. (Zelenkauskaite, 2022, 126) For example, that German NATO soldiers were spreading COVID or that they had desecrated a Jewish cemetery in the Lithuanian city of Kaunas during a NATO exercise. Meanwhile, fact-checking sites could prove that the Jewish

³ The “Internet Research Agency”, a so called “troll factory”, is an example for a covertly operating channel to spread disinformation. (Chen, 2015) The employees, commonly referred to as “trolls,” create fake social media accounts and engage in online discussions worldwide, aiming to provoke extreme emotions and incite people. Reports suggest that over 1,000 paid bloggers and commenters work in a single building in St. Petersburg, where they are expected to meet specific quotas for posting content. (Chen, 2015; Legucka 2020)

cemetery does not exist. These campaigns aimed to blur reality, create mistrust, divisions and erode support for Ukraine's aspirations for closer integration with the Euro-Atlantic community (Greenberg, 2020). As further examples of attacks on western societies the Russian involvement influencing the 2016 presidential elections in the USA (Babiracki, 2016; DiResta et al., 2018, pp. 76-96) or a number of elections in European countries (European Commission, 2018a, p.3; Legucka, 2020) can be mentioned.

3.1.3 Russian motives

Agnieszka Legucka (2020) from the Polish Institute of International Affairs suggests that the motivation for the recent resurgence of Russian disinformation campaigns is initially the result of a Russian fear. Russian authorities feared that Western technologies like Google, Facebook and other social media platforms could impose foreign messages on their citizens. In light of its perceived vulnerability, the Kremlin therefore developed a systematic understanding of leveraging the principles of liberal democracies within a framework of “non-linear” warfare to its advantage: as the weaponization of information. (Pomerantsev & Weiss, 2014, p. 4; Legucka, 2020; Mölder & Shiraev, 2021, p. 16) Legucka underlines this claim with the example that after demonstrations in Moscow, before the 2011 re-election of Putin, Russian decision makers were fearing that the democratization processes of the Arab spring in 2011 could spill over to Russia (Legucka, 2020). To combat this possibility, Legucka claims that since 2011 Russia reinvented and upscaled its information disorder strategy and now uses the global virtual space to achieve their foreign policy goals and to threaten Western democracies (Legucka, 2020).

According to Ofer Fridman (2019) and Mölder & Shiraev (2021), Russia's incentive to conduct disinformation campaigns can be summarized by an ancient proverb, allocated to the ancient Chinese war-philosopher Sun Tzu: “*to break the enemy's resistance without fighting is the foremost excellence*” (Fridman, 2019, p. 80). Various authors add that the Russian motivation to conduct disinformation is rooted in Russia's foreign policy agenda (Bolgov, 2019, p. 138; Baumann, 2020, p. 288; Marsili, 2021, p. 157; Petraitis, 2021, p. 107). Russia's self-perception, as outlined in its “Concept of the Foreign Policy of the Russian Federation,” emphasizes Russia's unique position as a world-power, its historical significance and commitment to their understanding of global peace. Furthermore, the

“Western hegemony” and the USA and their “satellites” are declared to be a threat to Russia’s vital interests. (Ministry of Foreign Affairs of the Russian Federation, 2023) By engaging in information warfare and perception management, Russia aims to disrupt its adversaries and maintain its own interests to assert its power and shape the global narrative (Petraitis, 2021, p. 108). Petraitis (2021) further claims that the self-perception of Russia as a great power is manifested in its “Russian Hybrid Warfare doctrine”. The Russian interpretation of Hybrid Warfare – encapsulated in the Russian term “gibridnaya voyna,” – signifies Russia’s strategic use of both military and nonmilitary means to achieve political goals and undermine adversaries on various fronts, from politics and the economy to social development and culture. (Fridman, 2019, 76-78) This doctrine is further demonstrated through recent actions like executing military interventions while conducting modern disinformation campaigns (Petraitis, 2021, p. 107).

3.2 Russian-Ukrainian Conflict: From a hybrid to a conventional war

Scrutinizing the origins of the Russo-Ukrainian conflict would exceed the scope of this thesis. Just looking at recent events (2014-2024), in 2014, the pro-Western Maidan protests took place in Ukraine’s capital Kiev. This was followed by the Russian annexation of Crimea and the outbreak of the war in the Donbass region. After the full-scale Russian invasion of Ukraine in 2022, the conflict eventually escalated into the still ongoing war. (Walker, 2022, p. 4) The EU and NATO are regarding Russia’s activities against Ukraine as Hybrid Warfare in various strategic documents (Bankov, 2023, p. 320). Furthermore, numerous authors consider this last decade of the ongoing conflict to be exemplary for the Russian utilization of Hybrid Warfare and the Russian dissemination of disinformation campaigns (Fridman, 2019, p. 1; Pierzchala, 2019, pp. 105-107; Torichnyi et al., 2021, p. 356; OECD, 2022; Miron, 2023; Goodson & Zakowska, 2023).

According to Fridman (2019), the term “Hybrid Warfare”, while controversially debated as a term in the scientific literature, has become the most popular concept to explain Russian actions in Crimea and the Ukraine. Asmussen, Hansen & Meiser (2015) consider Hybrid Warfare as the practice of achieving strategic objectives by combining conventional military tactics with unconventional methods and tools, such as psychological warfare and information operations (Asmussen, Hansen & Meiser, 2015, p.

15). Filimonov, (2019) adds that Hybrid Warfare allows an external power to gain “*great operational capabilities to achieve its military and political goals without escalating into full-scale conventional fighting*” (Filimonov, 2019, pp. 25-26). Furthermore, Fridman (2019) offers a specification in which he connects the earlier mentioned concept of “*gibridnaya voyna*” – the Russian concept of Hybrid Warfare – with the motive to conduct disinformation campaigns: as an attempt to erode the socio-cultural cohesion of the adversary's population. (Fridman, 2019, p. 68)

As mentioned before, various authors argue that the Russian ventures on the international stage since the Russian annexation of Crimea in March 2014 can be considered as examples of Hybrid Warfare activities (Tsybulenko & Kajander, 2021, p. 173; Miron, 2023, p. 397). Tsybulenko & Kajander (2021) claim that throughout the annexation of Crimea in 2014, Russian disinformation campaigns employed various exemplary trends and motives. One specific example was the establishing of the “*little green men*” narrative, which aimed to justify the presence of armed forces without insignia in Crimea. During the annexation, Russia initially denied any involvement, attributing the presence of the invading forces to local self-defense units or even portraying them as independent actors. (Tsybulenko & Kajander, 2021, p. 175) The authors add that this disinformation campaign aimed to create confusion and cloud Russia's direct role in the annexation (Tsybulenko & Kajander, 2021, p. 176). Treverton et al. (2018) add that another disinformation narrative surrounding the annexation of Crimea was the justification of Russia's actions by asserting historical and cultural ties to the region. Russia propagated the idea that the Crimea had always been Russian and that its annexation was a legitimate act to protect the rights of ethnic Russians. (Treverton et al., 2018, pp. 16-20)

With the full-scale Russian invasion of Ukraine on the 24th of February 2022, the conflict changed (OECD, 2022; Goodson & Zakowska, 2023). Bankov (2023) claims that Russia's invasion of Ukraine “*is bringing conventional war back to Europe*” (Bankov, 2023, p. 320). The German Chancellor Olaf Scholz called the Russian invasion of Ukraine a turning point in history (Scholz, 2022). Ionita (2023) considers the Russian invasion of Ukraine as the “*most demanding and strange conventional war since World War II*” (Ionita, 2023, p. 6). Thereby, she defines conventional war operations as being fought with “*conventional*

weapons [...] and battlefield tactics between two or more states in an open confrontation, in which forces are well-defined and fight using weapons that target each other's military power." (Ionita, 2023, p. 7) Bankov (2023) further highlights a difference in the language that is used to address Russian operations targeting Ukraine, by scholars and Western organizations alike. He claims that the EU and NATO no longer use Hybrid Warfare in their political rhetoric, but have replaced it with "Russia's war of aggression in Ukraine" (Bankov, 2023, p. 320). Furthermore, he adds that many scholars consider Hybrid Warfare in its common definition no longer fitting as a conceptual tool for analyzing Russia's actions against Ukraine, while others argue that the terminology has to be adapted to the new reality (Bankov, 2023, p. 334).

Overall, in the academic discourse, the Russian invasion of Ukraine is seen as a major change in the conflict and a turning point in history (Walker, 2022, p. 4; Bankov, 2023, p. 320; Goodson & Zakowska, 2023; Ionita, 2023, p. 6). Whether the currently ongoing war is to be defined as a 'conventional war' or as a 'hybrid war with conventional elements' remains contested (Bankov, 2023, p. 334). Nevertheless, the conflict has changed. The previous examples of Russian disinformation campaigns in the past – as described in chapter 2.3 – all took place in a Hybrid Warfare environment. The examples of Russian disinformation as described in the literature review did not coincide with a comparable conflict – much less with the changing of a conflict from a solely hybrid war, to a hybrid war with conventional elements. Hence, the research question of this thesis arises, whether the Ukraine themed Russian disinformation campaigns changed with the changing of the conflict in Ukraine.

4 Literature Review

Recent studies about Russian disinformation campaigns range from qualitative discourse analysis on the effects of Russian disinformation on Russia-West relations (Baumann, 2020) to quantitative corpus-based analyses of Russian internet trolling activity (Baraniuk, 2017) or Fake-News during the Covid-19 pandemic (Monogarova, Shiryayeva & Arupova, 2021). Other studies include mixed methods approaches like quantitative topic modeling followed by a qualitative analysis (Keegan, 2022). Most existing studies that analyze Russian disinformation as a case-study either predate the Russian invasion of Ukraine (Baraniuk, 2017) or do not scrutinize Ukraine themed Russian disinformation, but other targets of Russian disinformation campaigns like the US elections (Alvarez, Choi & Strover, 2020) or the Covid-19 pandemic (Monogarova et al., 2021; Keegan, 2022; Mogos, Grapa & Şandru, 2022).

Geissler, Bär, Pröllochs & Feuerriegel (2023) add that *“the Russian invasion of Ukraine in February 2022 was accompanied by practices of information warfare, yet existing evidence is largely anecdotal while large-scale empirical evidence is lacking”* (Geissler, Bär, Pröllochs & Feuerriegel, 2023, p. 1) The authors tackle this identified issue by analyzing the spread of pro-Russian support on social media during the 2022 invasion of Ukraine in their 2023 published study. For this, they collected a dataset of tweets that were marked with pro-Russian hashtags to then analyze when and by whom they were published (Geissler et al., 2023, p. 3). Their study finds strong support for a large-scale Russian propaganda campaign on social media to systematically manipulate beliefs in Russia’s interest. Moreover, it is detected that pro-Russian messages have been disproportionately spread by bots. Yet, no further scrutinizing of the contents or possible changes of the disinformation is conducted. (Geissler et al., 2023, p. 16)

The closest approach to answer the research interest of this thesis is a Semantic Visions⁴ data set published by the New York Times in July 2022, which was collected from more than 8000 Russian websites and contained almost 8 million articles since 2014, referencing Ukraine. In this dataset it was shown that for eight years the registered

⁴ Semantic visions is a defense analytics firm, that globally monitors and analyzes over 1.1 million news articles in 12 languages per day using Artificial Intelligence and machine learning technology (Semantic Visions, 2024).

articles mentioning Nazism remained relatively flat and did not exceed 250 mentions over a 7-day average. Yet, on February 24th 2022, the day of the Russian invasion of Ukraine, the registered articles mentioning that the Ukraine is a Nazi state spiked up to unprecedented levels of more than 3000 articles per day with an ensuing 7-day average of above 2000. Thus, proving a drastic increase in the disinformation narrative of the Ukraine being a state run by Nazis. (Smart, 2022; Holub-Korba, 2023)

These insights perpetuate the question, how the Ukraine themed Russian disinformation changed generally in terms of quantity over the course of the start of the Russian invasion of Ukraine, beyond the 'Ukraine is run by Nazis' narrative. From this sub-question and the New York Times published dataset, the following hypothesis can be derived:

Hypothesis 1:

The quantity of Russian disinformation dealing with the Ukraine increases after the Russian invasion of Ukraine in 2022, compared to the pre-invasion period.

In 2017 Kamil Baraniuk published a quantitative analysis of Russian internet trolling activity in the internet discourse, using the linguistic method of a corpus-based analysis. For that he examined comments on a popular polish website in three different periods, 2010, 2013 and 2014, to examine the key-words of the trolling-activity. (Baraniuk, 2017, p. 239-240) Baraniuk found out that certain keywords appeared throughout the researched period, but were increased to exacerbate existing problems in the societies that the disinformation was targeting. Moreover, he discovered that there is a visible increase in anti-Ukrainian expressions during the events of the Maidan protests in 2014. Finally, he concludes that the method of corpus-based analysis is a valuable tool in the study of trolling activity and content analysis (Baraniuk, 2017, pp. 252-253).

Monogarova et al. also conduct a corpus-based analysis in their 2021 published study of topical change in Russian viral fake stories that were disseminated during the first thirteen months of the Covid-19 pandemic. Their data thereby consists of thirteen corpora of social media posts that respectively reached over 50.000 unique views. By analyzing the frequencies of keywords and topic-related vocabulary, the authors track the changing focus of the public regarding Russian disinformation. The study finds that the most

popular terms in the analyzed disinformation changed dynamically over the months. Therefrom, the authors conclude that this indicates a changing of the Russian Covid-19 disinformation agenda. (Monogarova et al., 2021, p. 83 & 104) Furthermore, it is emphasized that corpus-based research is an effective tool for collecting and analyzing large quantities of data and understanding communication patterns (Monogarova et al., 2021, p. 85). Finally, the authors develop the theory that the creators of the disinformation adjust their language to promote their agenda, but also to engage the audiences fears and desires to draw their interest to the content, to increase the chances of the disinformation to go viral. Therefore, the authors suggest that analyzing the keywords could be a useful tool to draw conclusions about the fake narratives of disinformation. (Monogarova et al., 2021, p. 84)

These insights raise the question, whether topic changes can be found over the course of the Russian invasion as well. From this sub-question, the second hypothesis is derived:

Hypothesis 2.1:

After the Russian invasion of Ukraine, the topics of Russian disinformation shift from targeting multiple topics to an increase in Ukraine-conflict related topics.

Moreover, in a 2022 published study, Keegan examined 12 months of articles about Covid-19 vaccinations in Ukraine and Serbia that she web-scraped from Russian state-owned websites in 2021. In her study, she uses the data to conduct quantitative topic modeling to find out how the coverage of Covid-19 vaccination differentiates in two countries with large anti- and pro-Russian populations. Furthermore, with a qualitative analysis of the data, she aims to find out how the vaccine propaganda narratives between Ukraine and Serbia compare. (Keegan, 2022, p.1) Keegan, finds out that on Russian state-owned websites that targeted Ukraine, like Ukraine.ru, the range of disinformation narratives was wider and included narratives that were contradicting each other. Meanwhile, in Serbia on outlets like Sputnik Serbia, fewer topics were covered more consistently. In Ukraine, the examined articles covered the Russian Covid-19 vaccine Sputnik V positively, but criticized the Ukrainian government, while in Serbia, both the vaccine and the Serbian government's vaccination campaign were covered positively. (Keegan, 2022, pp. 4-6)

Hypothesis 2.1 suggests whether existing Ukraine related topics increased, while other non-Ukraine related topics decreased. Based on Keegan's study, but also based on the conclusion of Monogarova et al.'s study, the question can be asked, whether not only existing Ukraine related topics increased, but whether also new and unprecedented topics about Ukraine were created in the registered Russian disinformation. This sub-question suggests the following hypothesis:

Hypothesis 2.2:

Besides the shift in existing topics, following the 2022 Russian invasion of Ukraine, Russian disinformation about the Ukraine will exhibit an increase in unprecedented topics that deal with new, post-invasion narratives related to the conflict.

Alvarez, Choi & Strover published a study in 2020 in which they conducted a sentiment analysis of Russian disinformation campaigns. In their study, the group scrutinized the sentiment of 3519 Facebook ads that were purchased by a Russian state-run troll farm, the Internet Research Agency, over the course of the campaigning for the 2016 US presidential elections. These Facebook advertisements were released to the public after congressional hearings during investigations of the US congress regarding the Russian influence on the 2016 US presidential elections. (Alvarez, Choi & Strover, 2020, p. 3029) In their study, the authors conclude that the Russian ads were carefully crafted to target specific demographic groups. For example, profiles of "white" people with low-income were shown anti-immigration ads, while "white" middle-income populations were targeted with ads on nationalism. (Alvarez, Choi & Strover, 2020, p. 3034) Meanwhile, the majority of the purchased ads had a positive sentiment. The positive ads generated the most attention and the authors assume that they have been used to grow the legitimacy of their created Facebook groups (e.g. fake groups called "Black Matters" or "Woke Blacks"). Including previous literature, it is concluded that positive and negative arousal is more likely to go viral, which their study confirms in this case. (Alvarez, Choi & Strover, 2020, p. 3045) Moreover, it is added that the sentiment of the targeted ads was more negative before the elections in November and appeared to shift to a positive sentiment immediately after the elections. Furthermore, they analyze that sentiment in general produces emotional engagement and provokes concerns that could increase division between groups. Additionally, the authors claim that the use of the negative sentiment

before the elections could have enhanced anger and disaffection. (Alvarez, Choi & Strover, 2020, p. 3046).

The change of sentiment over the course of the US presidential elections can be tied to the goals of Russian disinformation that have been previously described in the literature review like sowing mistrust and dividing the attacked society. This raises the question, whether the sentiment of the Russian disinformation over the course of the Russian invasion of Ukraine changed as well. Hence, the third Hypothesis is generated:

Hypothesis 3:

The sentiment of Russian disinformation about Ukraine becomes more negative after the Russian invasion of Ukraine in 2022.

5 Methodology

5.1 Research design

The research question guiding this thesis is: *What are the changes in Ukraine themed Russian disinformation before and after the Russian invasion of Ukraine in 2022?* This question aims to describe and explore the relevant data (McCombes, 2023). To approach the research question and test the hypotheses, the possible changes were measured in terms of quantity, topics and sentiment.

Based on the previous research of current Russian disinformation campaigns, as described in Chapter 2.5, a quantitative corpus-based case-study of current Ukraine themed Russian disinformation campaigns was conducted. For this, two corpora were built. Each corpus consists of all the Russian disinformation that was published in English, which was recorded in the “disinfo Database” by the European Commission External Action Service’s (EEAS) “East StratCom Task Force” before and after the Russian invasion of Ukraine. The first corpus represents disinformation before the invasion and consists of data six months prior to the Russian invasion of Ukraine (23.08.2021-23.02.2022) and the second corpus consists of data immediately after the invasion (24.02.2022-24.08.2022). The coverage of twelve months thereby covers enough data to be representative, while being small enough to register differences, as seen in the studies by Monogarova et al. (2021) and Keegan (2022). Thereby, the recorded disinformation consists of press releases, news articles, blog entries and other online media. The two corpora were then analyzed with different methods, to test the hypotheses.

For hypothesis 1, the articles in the respective corpora were counted to find out how many articles that were registered by the EEAS as Russian disinformation were connected to Ukraine related topics (compared to other topics like Covid-19 vaccinations etc.). Hence, for each corpus, all registered articles were listed and categorized according to their EEAS-given “tags”. Afterwards, all articles mentioning Ukraine were counted, using Excels filter function and then divided by the total amount of articles within the respective corpus. This allowed a quantitative comparison between the two corpora regarding the frequency of Ukraine themed disinformation campaigns, along the timeframe of the Russian invasion of Ukraine.

For the subsequent hypotheses, the corpora were prepared and analyzed with the programming language Python and its ‘Natural Language Toolkit’ (NLTK). This process will be explained in the following chapters. To test hypotheses 2.1 and 2.2, the topics of the analyzed Russian disinformation campaigns were identified. To achieve this, the most common words and bigrams (most common adjacent words) in the respective corpora were extracted. From these results, the most common topics of the disinformation were derived. Afterwards, the topics were analyzed regarding how they changed over the period of the invasion. Furthermore, it was analyzed whether a change of the narratives within the topics can be derived from the identified topics. In addition to the country-tags the EEAS tags each article with “EUvsDisinfo-keywords”. As an addition to the topic analysis, these keywords were also scrutinized. To test hypothesis 3, a sentiment analysis was conducted with Python’s “SentimentIntensityAnalyzer” module “Valence Aware Dictionary and sEntiment Reasoner” (VADER), which allows to perform sentiment analysis on a given piece of text.

The described research design of the quantitative corpus-based analysis is depicted graphically in the following Figure 3.

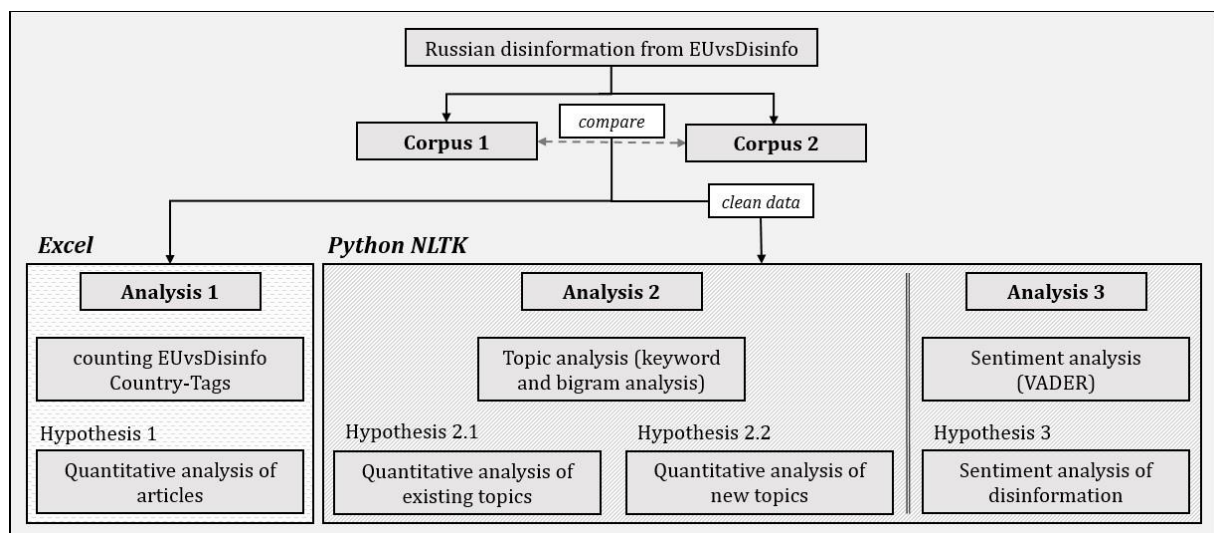


Figure 3 - Research design - Source: self-provided

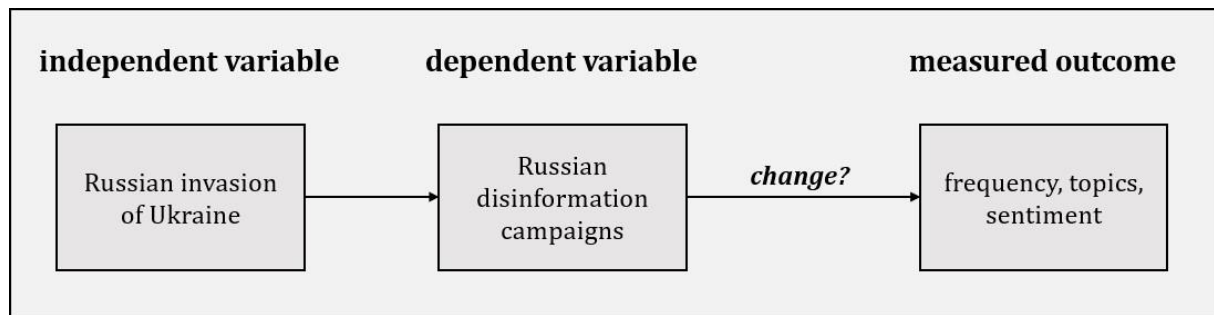


Figure 4 - Variables - Source: self-provided

As shown in Figure 4, the independent variable in the following analysis is the Russian invasion of Ukraine and the dependent variable are the Russian disinformation campaigns. The latter was tested regarding their quantity, topics and sentiment to measure possible changes.

5.2 Theory of corpus-based analyses

Corpus-based analysis is a research approach originating from the field of linguistics to examine the expression of language in a collection of texts based on predefined rules. The texts selected for analysis are chosen to represent a particular topic and computer tools are then employed to automate the handling of large quantities of data. Criteria for the composition of the corpus can thereby be defined by the researcher. (Baraniuk, 2017, p. 242; Monogarova et al., 2021, pp. 84-85)

5.2.1 Python's "Natural Language Tool Kit" (NLTK)

The software to conduct the corpus-based analysis will be Python's Natural Language Tool Kit. Python is a universal, commonly interpreted, high-level programming language. It aims to promote a well readable and concise programming style. (Python Software Foundation, 2023) The Natural Language Tool Kit is a platform for building Python programs to work with human language data and conduct Natural Language Processing (NLP). The focus of NLP is to make natural human language useable by computer programs (Jablonski, n.d.). Python's Natural Language Toolkit thereby provides a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning. (Monogarova et al., 2021, p. 88; NLTK Project, 2023)

5.2.2 Methods of preprocessing text

When conducting Natural Language Processing with NLTK, it is necessary to preprocess the text before further working with the model. Raw text can either not be processed or it can taint the results. (Rashmi, Hanumanthappa & Jyothi, 2015, p. 43975) Preprocessing varies depending on the intended analysis, but generally includes: transforming the entire text in lowercase letters, removing punctuation, tokenization, filtering stop words⁵ and stemming or lemmatizing. Setting the letters to lowercase and removing punctuation helps in standardizing the words. If not standardized, the capital letter of a word at beginning of a sentence can mislead NLTK into thinking it is a different word than the lower-case version in the text.

Tokenization refers to the process of dividing the contents of the corpora into smaller units called tokens, which could be words, characters, or other meaningful elements. This helps in quantifying the words for the statistical analysis. (Rashmi et al., 2015, p. 43976; Monogarova et al., 2021, pp. 88-89) The process of filtering stop words, refers to the identification and removal of high frequency words that do not play a role in the analysis, for example conjunctions like and or if. (Rashmi et al., 2015, p. 43976) Stemming is the process to reduce a word to its stem by removing prefixed or suffixes, so “running” would be turned into “run”. Thereby the context of the word is ignored, so “better” would be turned into “bet”. Meanwhile lemmatizing is the process of reducing words to their base form. When turned to its lemma, “running” would still be turned into “run”, while “better” would be turned into “good”. If you apply both, lemmatizing and stemming to a word, it is called “double stemming”. (Rashmi et al., 2015, p. 43979)

5.2.3 Keyword, bigram, collocation and topic analyses

After preprocessing the data, multiple analyses can be conducted with the corpora. A keyword analysis focuses on studying specifically searched keywords that are deemed relevant to the subject matter of the text. A bigram are two words in conjunction, hence a bigram analysis extracts frequently occurring word combinations like “United States” (Baraniuk, 2017, p. 242). A collocation analysis explores the coexistence of individual

⁵ The stop words in this study are according to NLTK’s standard list (Bleier, n.d.).

words in the text to identify frequently encountered word combinations and study the context of the extracted data and terms to understand their interdependencies. An example would be to test three words that appear before and after the word “Ukraine”, to examine the contextual appearance of a word in a given text. (Baraniuk, 2017, p. 242) A given text with a certain topic is more likely to contain certain keywords, bigrams and collocations related to the topic more frequently than unrelated words. Hence, the extracted keywords (also as part of bigrams and collocations), can be categorized into groups, from which topics can be derived. (Pykes, 2023)

5.2.4 Sentiment analysis with the “Valence Aware Dictionary and sEntiment Reasoner” (VADER)

There are several possibilities to conduct a sentiment analysis with Python. One of these options is the “SentimentIntensityAnalyzer”, which is a built-in class in the `nlk.sentiment.vader` module that allows, to perform sentiment analysis on a given piece of text. It uses the “Valence Aware Dictionary and sEntiment Reasoner” (VADER) lexicon to determine the sentiment of a piece of text. (Hutto & Gilbert, 2014, p. 216; Calderon, 2017)

The VADER lexicon is a pre-built sentiment analysis lexicon that contains a list of more than 7,500 lexical features, including words, idioms, and emoticons, each of which is scored on a scale of -4 (most negative) to +4 (most positive) based on its sentiment polarity (Calderon, 2017). The lexicon also contains rules for handling negations and punctuation, as well as boosting certain words that are known to indicate strong sentiment. The word “horrible” for example has a score of -2.5, while “okay,” has a rating of 0.9. The individual score of each word is thereby calculated by averaging all ratings given by human raters from the crowdsourcing marketplace “Amazon Mechanical Turk”. (Calderon, 2017)

Running the `polarity_scores` method, the `SentimentIntensityAnalyzer` class uses the created lexicon to compute a sentiment score for a given piece of text. The score contains four values: `neg`, `neu`, `pos`, and `compound`. The `neg`, `neu`, and `pos` values represent the proportions of the words in the text that were classified as negative, neutral, and positive,

respectively, while the compound value represents an overall sentiment score ranging from -1 (most negative) to +1 (most positive). (Calderon, 2017) Therefore, the scores indicate a tendency of whether a text is rather positive, neutral or negative. Meanwhile, the compound score is a metric that calculates the sum off all the lexicon ratings and normalizes them between -1 (most extreme negative) and +1 (most extreme positive). A compound score is regarded as neutral when between 0.05 and -0.05. (Calderon, 2017)

5.3 Data: EUvsDisinfo.eu's Disinfo database

All data will be retrieved from the website EUvsDisinfo.eu. This website is the self-declared 'flagship project' of the European External Action Service's East StratCom Task Force (EEAS, n.d. a). It was established in 2015 to address the ongoing Russian disinformation campaigns that are affecting the member states of the European Union and its neighboring countries. The website has the goal to increase public awareness on Russian disinformation and help to develop a resilience to media manipulation by increasing the general understanding of Russian disinformation campaigns. (EEAS, 2021; EEAS, n.d. a) To accomplish this, the EEAS conducts media monitoring in 15 languages, to identify and collect Russian disinformation. The monitored media outlets and websites are thereby either owned or financed by the Russian government or otherwise proven to be closely affiliated with the Russian state. (EEAS, n.d. a) Generally, the findings are commented and debunked. Regardless, all registered cases are then archived in the "EUvsDisinfo database" – according to the website, the *"only searchable, open-source repository of its kind"* (EEAS, n.d. a). Since the beginning of the database in January 2015, 16'997 cases of disinformation have been registered and archived by the website (last checked on 27.04.2024) (EEAS, n.d. b). With the tool "capture", the archive saves snapshots of the discovered disinformation during the moment of discovery. Hence, even if the disinformation content is deleted later or the source is banned (like Russia Today in Germany) the disinformation can be accessed reliably.

6 Operationalization

6.1 Mining the data and building the corpora

Regarding this thesis, the “EUvsDisinfo database” was used as a source to obtain Russian disinformation, as the database comprises all registered articles and guarantees comprehensiveness and reproducibility. Moreover, the EEAS takes over the work of defining whether the main source is actually disinformation. This way it is not necessary for the sourcing process to set parameters of what counts as disinformation and what not.

The EUvsDisinfo database has a search-bar, which allows to filter the database regarding the desired timeframe and the desired language. The key-date for the creation of the corpora is the day of the Russian invasion of Ukraine, the 24th of February 2022. The first corpus (named Corpus I) consists of data 6 months prior to the Russian invasion of Ukraine (23.08.2021-23.02.2022) (63 articles) and the second corpus (named Corpus II) consists of data immediately after the invasion (24.02.2022-24.08.2022) (57 articles). The included timeframe of six months was chosen as it will cover enough data to be representative and avoid purposive sampling, while being small enough to register differences, as seen in the studies by Monogarova et al. (2021) and Keegan (2022). Moreover, choosing two equally large timeframes is important for the validity of the sample, while the different quantity of disinformation in corpora, will not affect the average based analyses of the topic changes or the sentiment analysis. The chosen language for the displayed selection of disinformation was English.

To mine the data, it was decided to avoid web-scraping or text-mining tools, since much of the listed disinformation sources were not available anymore, as they were either deleted or not accessible after Sputnik and Russia Today were banned in Europe. Regardless, the sources could be accessed via internet archives and a Virtual Private Network (VPN) from non-European countries, but it was necessary to download the contents of every indicated webpage by hand. The websites containing the disinformation were then saved in “.txt” files in two folders, one for each corpus. While an automation of the process would have been faster, the manual process yielded the benefit of allowing to better familiarize with the disinformation.

6.2 The Syntax

For the first analysis, to test hypothesis 1, Excel's filter function was applied (Annex 1 & 2). For the subsequent analyses, to test hypotheses 2.1, 2.2 and 3, Python was used. The Python script, including annotations, can be found in Annex 3 of this thesis. For the Python-based analysis, first, the necessary libraries were imported. Following, a custom corpus was created, by loading the previously saved .txt files into the cache. Then, the preprocessing of the text was conducted. For this, the text was tokenized into words, converted to lowercase letters and the punctuation removed. Additionally, it was necessary to add exemptions regarding "polluting" punctuation, since some quotation marks and hyphen in the online sources were not entirely in American Standard Code for Information Interchange format (ASCII), the commonly used format for text and data on the internet (ASCII, 2024). Besides, the corpora were lemmatized, to avoid double stemming and maintain the correct meaning of the base words, since preserving the meaning of the individual words is important for the later conducted sentiment analysis. The other text cleaning methods do not affect the sentiment analysis, as for example stop words are not graded in the VADER model.

Before and after the preprocessing, the words in the respective corpora were counted, to gain an overview over the size of the corpora. Moreover, a word cloud (see page III) of the most common words was created, to gain an overview over the corpora. Continuing, the most common words and bigrams were generated for the analysis of the hypotheses. Here, it is to mention that during the lemmatizing process, a rule was added to exclude "US" as in "United States" from the lemmatization process. NLTK is unable to distinguish "US" from the objective personal pronoun "us". With this exception, the capitalization of the letters in "US" allowed the program to distinguish the two words. After the first analysis of corpus one, "us" was considered as the most common word. Hence, this exception was added. Now "US" is ranked second, with a much lower probability of error. Going on, the sentiment scores for the overall corpus were generated. Afterwards the concordance tool was used to specify the sentiment regarding the topic of Ukraine. This means that only the words that are in direct vicinity to the word "Ukraine" were considered in this second sentiment analysis. Finally, it is to mention that after each analysis the graphic depictions that are shown in the analysis were coded using the Python's "plotly" dash.

6.3 Limitations

The EEAS is monitoring Russian disinformation in 15 different languages. Nevertheless, this research will focus only on Russian disinformation originally published in the English language. For one, the reasons to focus on one language is technical, as a corpus analysis only works with one language in the corpus. The option to use a translating software for the disinformation in other languages was dismissed, with the intention of avoiding distortions in the results due to bad translations. Besides, for the sentiment analysis, the original sources are used to receive impartial results. Moreover, with the ban of Russia Today and Sputnik, disinformation datasets in “smaller” European languages is likely to have decreased and would therefore not provide enough data volume for an analysis. The aim to only analyze original sources is also the reason why the summaries of the recorded disinformation that are offered by the EEAS are not incorporated.

Moreover, English is the lingua franca of the West. Therefore, the results could offer a more general understanding of Russian disinformation, since disinformation in less spoken languages could be especially focused on the country that the language is spoken in and influence the results. On the other hand, the interpretation of the results has to incorporate the fact that the recorded disinformation could be explicitly targeting English speaking countries like the UK, Ireland, Malta or the USA. Regardless of these assumptions, the circumstance to only consider disinformation that was initially published in English, has to be considered when interpreting the analysis.

Furthermore, when setting the earlier defined criteria of the corpora in the EUvsDisinfo database (timeframe and language), certain registered disinformation cannot be integrated in the text accumulation of the corpora. For example, some registered disinformation is in fact a video. Hence, for the analysis of hypothesis 1 that does not require a text analysis, this registered disinformation can be used, but for the text-based topic and sentiment analyses, these entries have to be disregarded. Moreover, a few articles that are displayed when selecting the previously defined search criteria are in fact in Arabic or other languages. This might be due to the fact that the website on which the disinformation is registered have an option to switch to English. Since the articles do not translate, they will be disregarded for the text-based analyses, while they can be used for the quantitative analyses. Nevertheless, the data entries that are not useable for the text-

based analyses will not be ignored to ensure reproducibility. Analysis I and the “EUvsDisinfo-keyword” analysis in Analysis II were based on the assigned country/keyword tags, so all articles could be used. Regarding the most common word and bigram analysis: in Corpus I, from a total “n” of 63 articles, 31 were usable and therefore 32 unusable. In Corpus II, out of a total “n” of 57 articles, 43 were useable for text-based topic analysis and 14 were unusable. (see Annex 1 & 2) This adds up to a rather small sample size, but due to the previously explained limitations was a problem that could not be solved within this research design.

7 Analysis

7.1 Analysis I: quantitative comparison

The first analysis aims to tackle the first hypothesis: *“The quantity of Russian disinformation dealing with the Ukraine increases after the Russian invasion of Ukraine in 2022, compared to the pre-invasion period.”*

To test this hypothesis, a quantitative analysis was conducted to find out how many of the articles in the corpora are mentioning the Ukraine. In the EUvsDisinfo database, the EEAS’ East StratCom Task Force “tagged” each registered article with a category-tag called “countries mentioned”. The tag indicates which countries are mentioned in the articles. Often, more than one country is mentioned and therefore tagged in an article. Hence, the overall “n” of articles, is smaller than the total amount of countries mentioned. Yet, if the Ukraine is tagged, the country is mentioned in the article. Analyzing the tag is therefore well-suited to test the first hypothesis. Hence, for this analysis, in both corpora all registered articles were listed and categorized according to their “countries mentioned” tag. (see Annex 1 & 2) Afterwards, all articles where “Ukraine” is tagged were counted, using Excels filter function. Then, the number of articles in which the Ukraine is mentioned was divided by the total amount of articles within the respective corpus. This allowed a quantitative comparison between the two corpora regarding the frequency of Ukraine themed disinformation campaigns, along the timeframe of the Russian invasion of Ukraine.

7.1.1 Corpus I – pre invasion

In the timeframe 6 months prior to the Russian invasion of Ukraine, a total of 63 articles of disinformation were registered by the East StratCom Task Force, when selecting the search criteria of “English” and setting the timeframe to 23.08.2021-23.02.2022. Two articles were not assigned a country-tag. Therefore, the “n” in the graphical depiction of country-tags in Figure 5 is 61. Nevertheless, the overall “n” remains 63. Out of the 63 articles, 21 had the tag “Ukraine” in the category “countries mentioned”. (Annex 1, Figure 5) Hence, in 33% of all articles of the corpus the Ukraine was tagged.

Countries or regions of “the West” and member states of NATO (US, EU, Poland, Canada, Italy, The Netherlands, France, Germany, Europe) were mentioned in 48 articles, which equates to 76%. Here, the US stands out with being mentioned in 27 times (43%). Russia was mentioned in 37 of the articles (59%). The USSR and its former member states Georgia and Belarus were mentioned 7 times (11%). Countries and regions of the rest of the world (Afghanistan, Syria, Mali, Africa, China, Iran) were mentioned 9 times (14%).

Corpus 1: Country or Region mentioned (n=61)

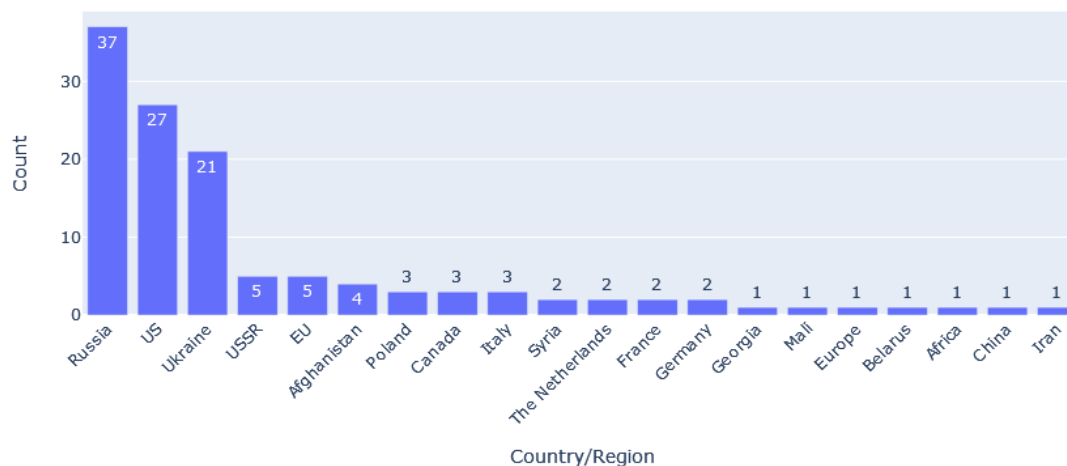


Figure 5 - Analysis I, Corpus I – pre-invasion. Source: self-provided

7.1.2 Corpus II – post invasion

In the timeframe 6 months after the start of the Russian invasion of Ukraine a total of 57 articles of disinformation were registered by the East StratCom Task Force, when selecting the search criteria of “English” and setting the timeframe to 24.02.2022-24.08.2022. Out of the 57 articles, 53 had the tag Ukraine in the category “countries mentioned”. (Annex 2, Figure 6) Hence, in 93% of all articles of the corpus, the Ukraine was tagged.

Countries or regions of “the West” and members of NATO (US, EU, France, Germany, UK, Latvia, Finland, Sweden, Canada, Switzerland) were mentioned in 41 articles, which equates to 72%. The US remain the country of this group with the most mentions (16), but relatively decrease to 28%. Russia is mentioned in 49 of the articles (86%). Former USSR member states that have not joined NATO or are not considered as being part of “the

West” (Belarus, Moldova, Georgia, Azerbaijan, Kazakhstan, Kyrgyzstan, Armenia) are mentioned 9 times (16%). The rest of the world (Nigeria, Syria) is mentioned 2 times (4%).

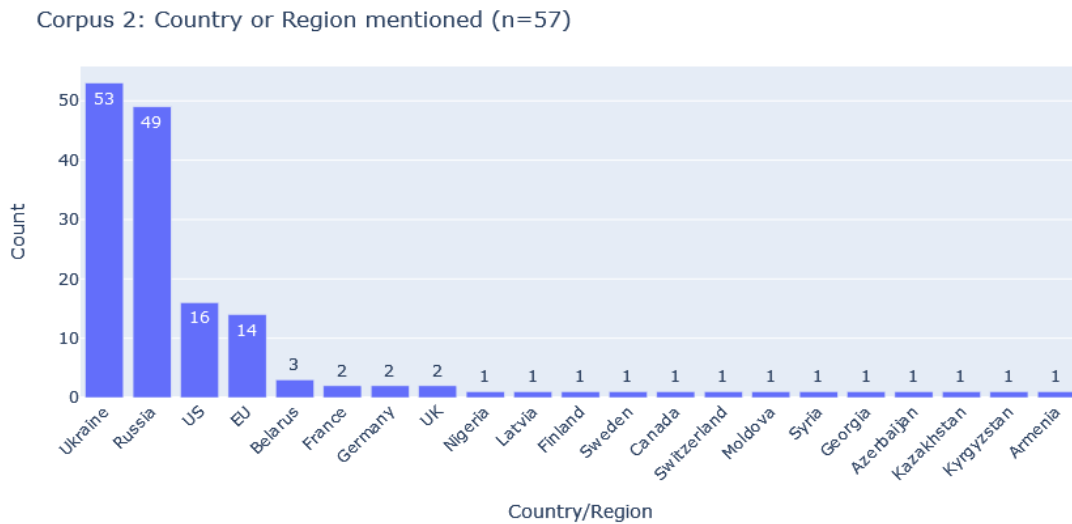


Figure 6 - Analysis I, Corpus II – post-invasion. Source: self-provided

7.1.3 Discussing hypothesis 1

The assumption of the first hypothesis that *“the quantity of Russian disinformation dealing with the Ukraine increases after the Russian invasion of Ukraine in 2022, compared to the pre-invasion period”*, was correct. The quantity of Russian disinformation targeting the Ukraine did increase significantly from 33% in the timeframe six months pre-invasion in Corpus I to 93% in the timeframe six months post-invasion in Corpus II. Also in absolute numbers, the Ukraine increased from the third most mentioned country in Corpus I to the most mentioned country in Corpus II. The increase of articles mentioning the Ukraine in relative and absolute numbers, coincides with a drop of mentioning other countries like the US (43% down to 28%). The mentions of countries in the rest of the world also decrease by 10% (from 14% down to 4%). These changes show that the focus of the Russian disinformation campaigns clearly increased their attention towards Ukraine.

7.2 Analysis II: topic analysis

Analysis II tackles the two topic related hypotheses:

Hypothesis 2.1:

“After the Russian invasion of Ukraine, the topics of Russian disinformation shift from targeting multiple topics to an increase in Ukraine-conflict related topics.”

Hypothesis 2.2:

“Besides the shift in existing topics, following the 2022 Russian invasion of Ukraine, Russian disinformation about the Ukraine will exhibit an increase in unprecedented topics that deal with new, post-invasion narratives related to the conflict.”

To test these hypotheses, the corpora were being analyzed regarding their content, to derive the general topics of the disinformation six months prior and post the Russian invasion of Ukraine, as described in chapter 4.3 “Syntax”. To see what topics are prevalent in the registered disinformation, a topic analysis was conducted. For this, the 20 most common words and the 10 most common bigrams were filtered out in each corpus, since an article that deals with a certain topic will name the topic more frequently (Pykes, 2023). The most common words and bigrams were then categorized into topics. Ongoing, the topics of the two corpora were compared.

Some articles might be longer in size than others, so the number of words in the corpora differ. Hence, it is not statistically correct to count the exact number that a certain word was mentioned as a topic. Nevertheless, on a scale this large, it was possible to showcase the trends of which topics were registered as Russian disinformation. Thus, the keyword and bigram analyses were a suitable method to test the hypothesis regarding shifts in topics of the disinformation. (see Annex 3)

To substantiate the most common keyword and bigram analyses and the subsequent topic analysis, the “EUvsDisinfo-keywords” that are provided by the EEAS were also analyzed. Apart from the country-tags that were analyzed in Analysis I, the EUvsDisinfo database is tagging each registered article with “EUvsDisinfo-keywords”. Thereby, the “EUvsDisinfo-keywords” can be seen as EEAS-given topics. Hence, the two corpora were compared

regarding their “EUvsDisinfo-keywords”. Afterwards, these results were added to the discussion of the results of the actual disinformation (most common words and bigrams).

The “EUvsDisinfo-keywords” have the limitation that not all articles have the same number of “keywords” assigned. Due to this statistical disbalance, it is not possible to accumulate the absolute numbers of keywords and translate the result to overall topics in the corpora. Hence, the “EUvsDisinfo-keywords” were not sufficient as the sole indicator for topics regarding Russian disinformation campaigns. For example, some articles are tagged with both, coronavirus and vaccine, so it is not possible to add the amount of coronavirus mentions and vaccine mentions to one overall “covid-vaccination” category. Nevertheless, the same keyword is never tagged twice for the same article, so it is safe to say that there are at least 14 out of the 63 articles in Corpus I that were tagged with “coronavirus”. Yet, with the limitations in mind, an analysis of the keywords added to the general observation of trends. They were analyzed as well to see if they confirm the tendencies and topics that could be derived from the original full-text sources of disinformation.

7.2.1 Analysis of the most common words

The 63 articles in Corpus I accumulated a total of 51374 words, the 57 articles of Corpus II contained 16410 words. After removing the stop words, Corpus I contains 24744 words and Corpus II 8460 words. (Annex 4)

With 161 mentions, “NATO” is the most used word in Corpus I. Further common words like “US” (258), “western” (91) or “american” (91) indicate that NATO and the west are among the most common topics of the disinformation before the invasion. Other common words like “war” (105), “military” (86) or “hybrid” (79) also indicate that the topics of the pre-invasion disinformation also deal with conflict and war. Yet, with 102 mentions of vaccine and 83 mentions of Covid, it can be derived that the COVID pandemic is also present as a topic. The Ukraine is mentioned 80 times among the 20 most common words in Corpus I. (Figure 7) Nevertheless, just considering the 20 most common words in Corpus I, multiple topics could be identified, but a focus on a single topic could not be derived.

Corpus I: 20 Most Common Words

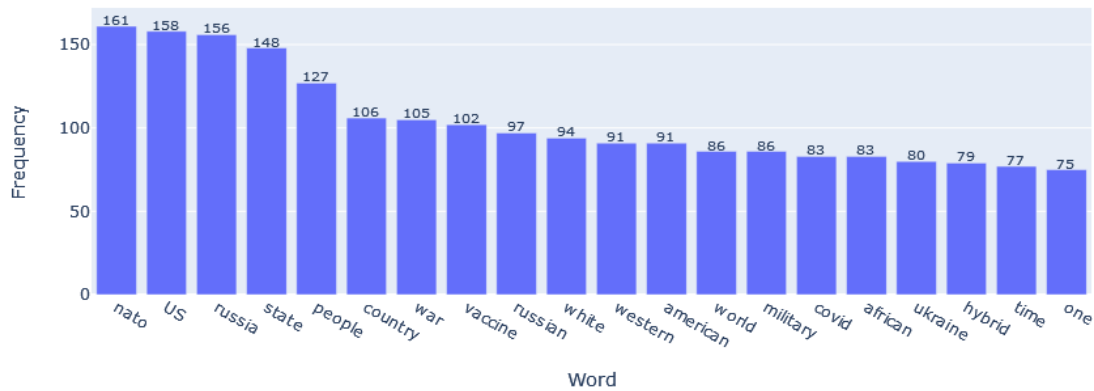


Figure 7 - Corpus I, 20 Most Common Words – pre-invasion. Source: self-provided

Corpus II: 20 Most Common Words

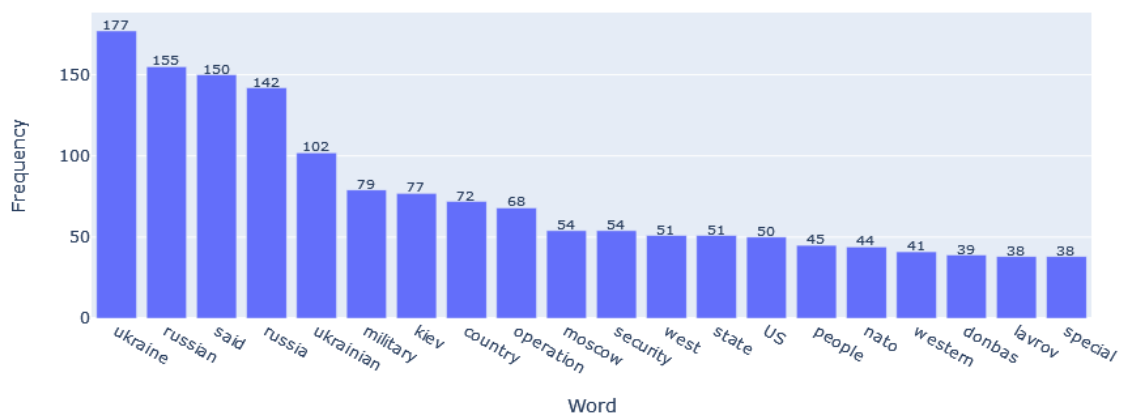


Figure 8 - Corpus II, 20 Most Common Words – post-invasion. Source: self-provided

In the disinformation six months after the invasion, which is compiled in Corpus II, “Ukraine” is the most commonly used word with 177 mentions. Considering that the total amount of words in Corpus I is much higher, the 177 mentions of Ukraine represent an absolute increase as well as a relative increase. Additionally, “Ukrainian” is the fifth most commonly used word with 102 mentions and “Kiev” on seventh place with 77 mentions. Meanwhile, the words that were mentioned most in Corpus I, US (50) and NATO (44) are now on place fourteen and sixteen of the most mentioned words. The topic of COVID is not detectable at all in Corpus II, when only considering the 20 most common words. While “war” was in Corpus I the seventh most common used word, it does not appear

among the 20 most common words in Corpus II. Meanwhile, the words “military” (79), “operation” (68) and “special” (20) are among the 20 most commonly used words in Corpus II. This and the significant increase of the mentions of “Ukraine”, “Ukrainian” and “Kiev”, with the simultaneous decrease of previously common words, clearly indicate a shift of the topic towards the Ukraine conflict. Moreover, the Russian narrative of prohibiting the term “war”, while spreading the terminology “special military operation” when referring to the conflict, can be clearly deduced from the data. (Figure 8) In Corpus I, it was possible to identify more than one topic from the 20 most common words, with no prevalent topic standing out. In Corpus II, the Ukraine conflict stands out as the dominant topic and no further topic could be identified.

7.2.2 Analysis of the most common bigrams

The analysis of the most common bigrams in Corpus I substantiates the previous analysis of the most common words. “United States” is with 52 appearances by far the most common bigram in Corpus 1. The common appearance in both the most common words and most common bigrams indicate that the United States were a significant topic for Russian disinformation campaigns before the Russian invasion of Ukraine. (In the resulting plot it says “United state”, because due to the lemmatization in the text-preprocessing, the word “states” was transformed into its singular form.) Moreover, from the common bigrams “hybrid threat” and “Hybrid Warfare” it can be derived that a certain amount of disinformation before the invasion already deals with conflict related topics. The bigrams “public health” and “covid vaccine” confirm the assumption that COVID was also a common topic for Russian disinformation in the six months before the invasion of Ukraine. The bigram “9 11” with the sixth most mentions suggests that it also was a topic of disinformation before the invasion. Moreover, the common bigrams of “pan Africanism” and “white people” indicate that there was a sizeable amount of disinformation regarding Africa related topics as well. Meanwhile, topics that are directly Ukraine related cannot be derived just by looking at the bigrams. Overall, the assessment of Corpus I from the most common words analysis is confirmed. Considering the 10 most common bigrams, multiple topics of Russian disinformation campaigns can be identified, with no particular topic standing out as being focused on. (Figure 9)

Corpus I: 10 Most Common Bigrams

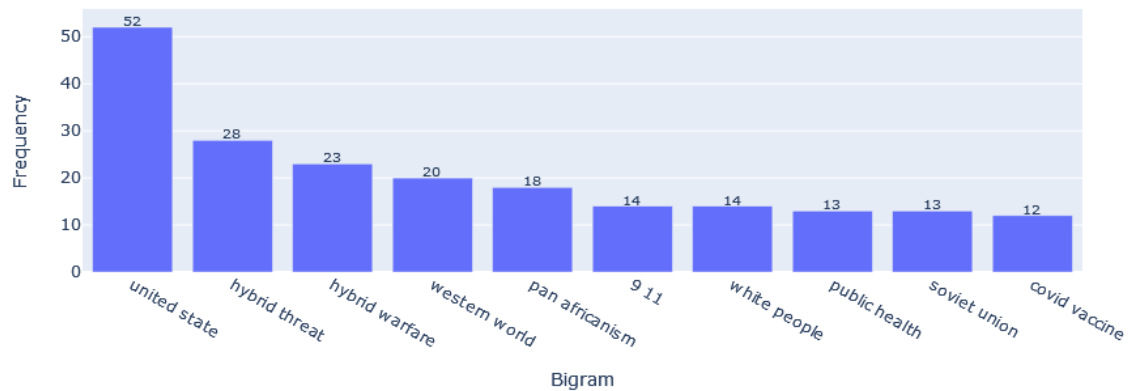


Figure 9 - Corpus I, 10 Most Common Bigrams – pre-invasion. Source: self-provided

Corpus II: 10 Most Common Bigrams

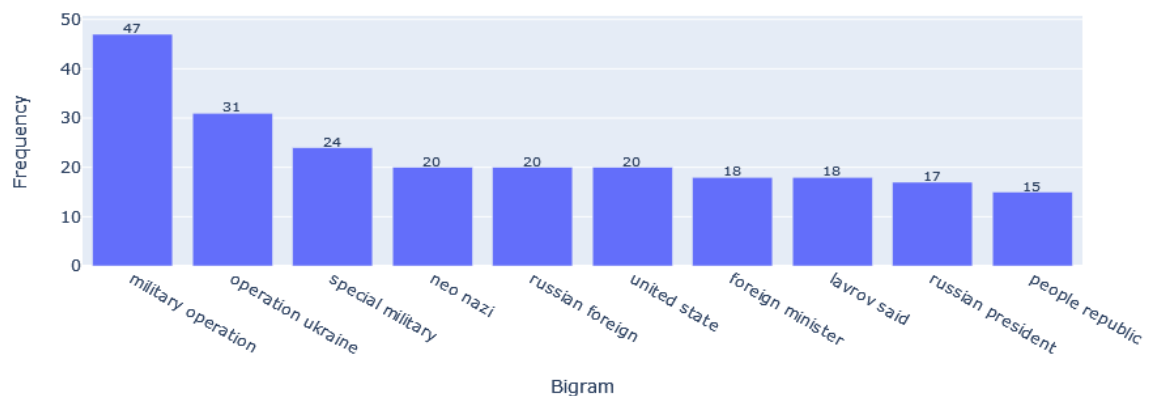


Figure 10 - Corpus II, 10 Most Common Bigrams – post-invasion. Source: self-provided

The most common bigram in Corpus II is “military operation” with 47 mentions. Second and third most common bigrams were “operation Ukraine” (31) and “special military” (24). This clearly shows the focus of Russian disinformation campaigns on the “special military operation in Ukraine” – the narrative how the Kremlin portrays the Russian invasion of Ukraine, while using the word “war” is prohibited. Meanwhile, the previously most common bigram “United States” drops to sixth place. This decrease and the disappearing of all the other most common bigrams from Corpus I can also be interpreted as indicators for a shift in Ukraine related topics of disinformation. 15 mentions of “people republic” can be associated to the so-called people republics of Donetsk and Luhansk. Moreover, the fourth most common bigram is “neo nazi” with 20 mentions. This can be

explained by the sudden rise of the “Ukrainians are neo nazis” narrative (Smart, 2022) that the Kremlin used to justify their invasion. (Figure 10)

By just looking at the bigrams, a clear shift in the Russian disinformation can be seen. For Corpus I, it could be derived that before the Russian invasion of Ukraine, Russian disinformation was dealing with multiple topics and targeting multiple narratives. In the six months after the invasion the bigrams indicate a full focus of the disinformation efforts on Ukraine related topics.

7.2.3 Topics derived from keywords and bigrams

As described in the previous two sub-chapters, it is possible to categorize the keywords and bigrams into different topics. In Corpus I the keywords and bigrams indicate five prevalent topics of disinformation that are all supported by quantity. “NATO”, “US”, “western”, “American”, “united state”, “western world” and “9/11” can be allocated to disinformation topics about “the West”. “Russia”, “Russian”, “Ukraine” and “Soviet Union” indicate existing disinformation about “Russia and Ukraine”. “War”, “military”, “hybrid”, “hybrid threat” and “Hybrid Warfare” are termini about “military and war” and indicate disinformation of that topic. “Vaccine”, “Covid”, “Public Health” and “Covid Vaccine” indicate that disinformation about “Covid” exists in Corpus 1. “African”, “White”, “Pan Africanism” and “White people” allows to deduce that some disinformation about “Africa” or African related topics exists in the corpus. Thereby, the Keyword “white” was allocated to the topic due to the subsequent bigram “white people”. Meanwhile, the keywords “state”, “people”, “country”, “world”, “time” and “one” could not be clearly allocated to one specific topic. “State” could be part of “united state” and “world” a part of “western world”, but this assumption is not relevant to the topic analysis, as the topic “the West” already exists with sufficient quantity.

Corpus I				
Keywords	quantity	Bigrams	quantity	Topics
NATO, US, western, american,	501	united state, western world, 9/11,	86	the West
russia, russian, ukraine war, military, hybrid	333	soviet union	13	Russia and Ukraine
	270	hybrid threat, hybrid warfare,	51	military and war
vaccine, covid,	185	public health, covid vaccine,	25	Covid
african, white,	177	pan africanism, white people,	32	Africa
state, people, country, world, time, one	619			no clear topic

Figure 11 - topics in Corpus I visualized - Source: self-provided

Corpus II				
Keywords	quantity	Bigrams	quantity	Topics
ukraine, russian, russia, ukrainian, kiev, moscow, donbas, lavrov,	784	russian foreign, lavrov said, russian president, people republic,	70	Russia and Ukraine
military, operation, special,	185	military operation, operation ukraine, special military	102	military and war
west, US, NATO, western,	186	united state	20	the West
		neo nazi	20	neo nazi
said, country, state, security, people,	372	foreign minister	18	no clear topic

Figure 12 - topics in Corpus II visualized - Source: self-provided

Corpus II has fewer identifiable topics. The prevalent topic is “Russia and Ukraine” with the keywords and bigrams “Ukraine”, “Russian”, “Russia”, “Ukrainian”, “Kiev”, “Moscow”, “Donbas”, “Lavrov”, “Russian foreign”, “Lavrov said”, “Russian president” and “people republic”. The topic “military and war” also still exists in the Russian disinformation of Corpus II, as indicated by “military”, “operation”, “special”, “military operation”, “operation Ukraine” and “special military”. The topic “the West” can also still be found with “west”, “US”, “NATO”, “western” and “United State”. “Said”, “country”, “state”, “security”, “people” and “foreign minister” could not be clearly allocated to an existing topic. “Foreign minister” could be linked to the Russian foreign minister Lavrov, but it could also refer to a different foreign minister, so it was not allocated to a specific topic.

The review of existing studies on Russian disinformation revealed that a surge of the narrative that the Ukraine is run by neo nazis emerged after the Russian invasion of Ukraine. Hence, the bigram “neo nazi” was allocated to a new topic. The topic “Russia and Ukraine” has multiple different keywords and bigrams that could allow a further specification of the topics. Nevertheless, the topic analysis confirms the findings from the keyword and bigram analyses that in Corpus I various different topics could be defined. Meanwhile, in Corpus II fewer topics could be defined, while the topics about Ukraine related disinformation increased.

7.2.4 Analysis of the EUvsDisinfo Keywords

As previously described, the EUvsDisinfo database allocates “EUvsDisinfo-keywords” to their registered articles. In total, Corpus I has 108 and Corpus II has 60 different “EUvsDisinfo-keywords” (Annex I & Annex II). To ensure readability of the plots, the “EUvsDisinfo-keywords” that were mentioned less than three times were left out of the visualization. (Figure 13 & Figure 14)

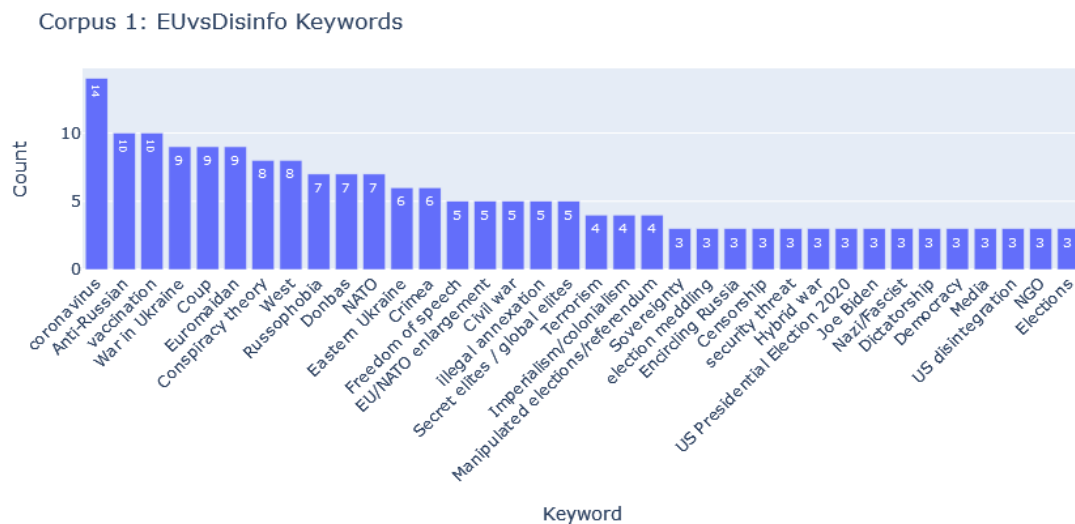


Figure 13 - Corpus I, Most Common EUvsDisinfo Keywords – pre-invasion. Plot: self-provided

The “EUvsDisinfo-keyword” “coronavirus” is the most used category in Corpus I (see Figure 13) with 14 mentions. This in addition to “vaccination” (10) indicates a presence of Covid-related disinformation. Ukraine conflict related “EUvsDisinfo-keywords” are also common. “War in Ukraine” is with 9 mentions on third place of the most mentions. Other

Ukraine related terms like “Euromaidan” (9), “Donbas” (7). “Eastern Ukraine” (6) or “Crimea” (6) also indicate Ukraine conflict related disinformation. “NATO” with 7 mentions, “EU/NATO enlargement” with 5 mentions or “encircling Russia” with 3 mentions also indicate a frequency of NATO themed disinformation. Furthermore, the mentioning of the “EUvsDisinfo-keywords” “manipulated elections/referendum” (4), “US Presidential Election 2020” (3), “Joe Biden” (3) and “elections” (3) indicate that the US elections were also a circulating topic of disinformation before the invasion. The “EUvsDisinfo-keywords” “Nazi/Fascist” were mentioned 3 times before the invasion. This can be linked to the previously discussed Russian disinformation narrative that the Ukraine is governed by Nazis (Smart, 2022). Additional “EUvsDisinfo-keywords” like “conspiracy theory” (8) or “secret elites/ global elites” (5) also indicate further themes of disinformation.

As depicted in Figure 14, in the six months after the invasion the two “EUvsDisinfo-keywords” “War in Ukraine” with 29 mentions and “Invasion of Ukraine” with 23 mentions are the most common categories. In comparison with Corpus I, the “EUvsDisinfo-keywords” “war in Ukraine” increases from being tagged in 14,29% of all the articles in Corpus I to being tagged in 50,88 % of the articles in Corpus II (considering the overall n of 9 out of 63 articles in Corpus I and 29 out of 57 articles in Corpus II). The “EUvsDisinfo-keyword” “invasion of Ukraine” is not mentioned in Corpus I and therefore increases by 100%. Moreover, “EUvsDisinfo-keywords” that are conflict related and promote a certain narrative, like “war crimes” (14), “sanctions” (4) or “genocide” (4) are also among the most mentioned terms. The “Nazi/Fascist” “EUvsDisinfo-keyword” is mentioned 8 times, so in at least 14% of all articles and an increase compared to Corpus I (8%). “EUvsDisinfo-keywords” that are not clearly related to the Ukraine conflict like “Virus/bacteria threat” could still be related to the conflict, as “coronavirus” is not among the most common terms in Corpus II and “Virus/bacteria” with 4 mentions could be related to “biological weapons” with 7 mentions. (The existence of possible US-funded biological weapon factories in Ukraine was also a commonly used Russian narrative to justify the invasion (Qiu, 2022).) (see Figure 14)

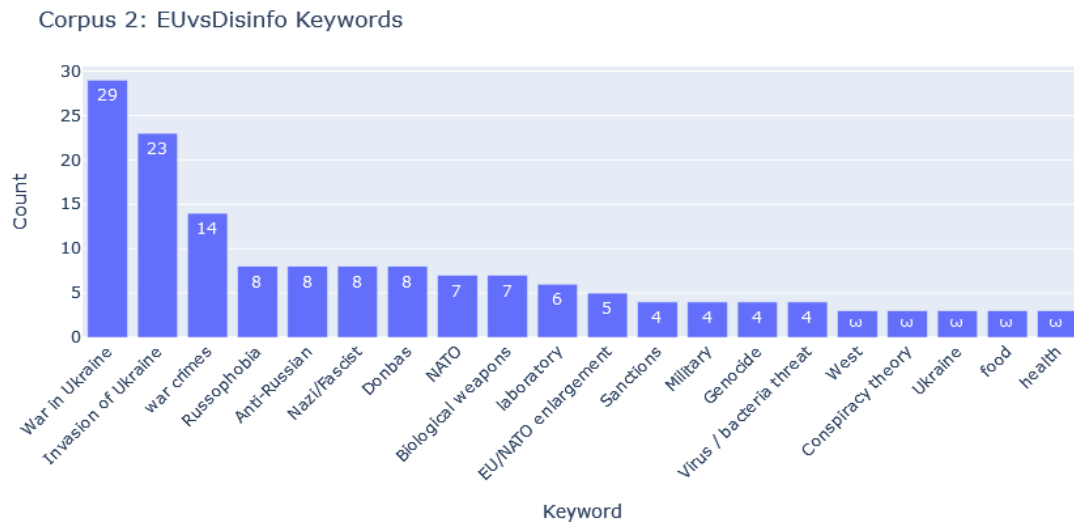


Figure 14 - Corpus II, Most Common EUvsDisinfo Keywords – post-invasion. Plot: self-provided

To sum it up, every topic that can be derived from the keywords in Corpus II is possibly related to the Ukraine conflict and Ukraine conflict related themes. Therefore, the trend that became evident during the most common words and bigrams analyses is being confirmed by the analysis of the EUvsDisinfo keywords. In Corpus I multiple topics of Russian disinformation can be derived from the “EUvsDisinfo-keywords” with no particular topic standing out, while in Corpus II a clear shift of the topics to the war in Ukraine can be observed.

7.2.5 Discussing hypothesis 2.1

The hypothesis, whether “*after the Russian invasion of Ukraine, the topics of Russian disinformation shift from targeting multiple topics to an increase in Ukraine-conflict related topics,*” can be confirmed. Different approaches to identify the prevalent topics in the two different corpora were presented. All analyses have shown a shift in the topics of the registered disinformation from a variation of topics (Covid, US elections, US/NATO, Ukraine conflict, Russia, Africa,) towards Ukraine related topics in the six months after the invasion. Moreover, the identified Ukraine-related topics were all conflict related (Invasion of Ukraine/Special military operation, Ukraine are Nazis narrative, sanctions).

The analysis of the country tags in Analysis I had shown that in 93% of all of the articles the Ukraine was mentioned. Building upon this knowledge, the shift in topics towards the

Ukraine, as seen in Analysis II, also further confirms the results from the first analysis. Within the disinformation shift to Ukraine related topics, an emphasis on themes related to the Ukraine invasion was observable, since all Ukraine related topics were conflict related. The data in Analysis I had shown that only 4 articles in Corpus II did not have a Ukraine tag (Annex 2, Article 11, 27, 28, 31). Out of the 4 articles without a Ukraine tag in Corpus II, two more were about “the West’s war on the Russian world” and one article was about “NATO expansion to increase tension in the region” which could theoretically still be accounted for as invasion related disinformation. Hence, except one article in the second corpus, which was about an alleged Monkeypox spread in Nigeria, all articles in the second corpus deal with Ukraine related disinformation.

7.2.6 Discussing hypothesis 2.2

The hypothesis, whether *“besides the shift in existing topics, following the 2022 Russian invasion of Ukraine, Russian disinformation about the Ukraine will exhibit an increase in unprecedented topics that deal with new, post-invasion narratives related to the conflict,”* can be confirmed as well. The shift in the most common words, most common bigrams and keywords all indicated a topic shift towards Ukraine related disinformation. The Ukraine was a topic of disinformation prior the invasion (Corpus I) as well. Yet, after the invasion (Corpus II), new topics about the Ukraine could be identified. Within these topics, new narratives could be identified that were unprecedented in the analyses of Corpus I. Exemplary is the bigram analysis of the second Corpus, when compared with the first. Here, the bigrams related to the “special military operation” narrative could not be detected before the invasion and appeared in the disinformation after the invasion. Hence, the narrative that Russia conducts a special military operation instead of fighting a war can be seen when scrutinizing the disinformation. The change of the word “war” from being very common in the word analysis of Corpus I, to not existent among the common words in Corpus II confirms this observation. Other examples are the “Ukraine is ruled by neo-Nazis” narrative that could be observed in all topic analyses of Corpus II or the “US funds bioweapons” narrative that the “EUvsDisinfo-keyword” analysis detected. Both topics are unprecedented Ukraine related disinformation that do not appear in the topic analysis of the Ukraine allocated disinformation in Corpus I.

7.3 Analysis III: sentiment analysis

Analysis III tackles the third hypothesis: *“The sentiment of Russian disinformation about Ukraine becomes more negative after the Russian invasion of Ukraine in 2022.”*

As described in chapter 4.3 “Syntax”, to test this hypothesis, the two preprocessed corpora were analyzed regarding their sentiment, to observe possible changes. The sentiment analysis was conducted with Python’s “SentimentIntensityAnalyzer” and the VADER lexicon which was explained in chapter 3.2.4. Based on the averaged score of each word in the corpus, three polarity scores and a compound score were computed with Python. (see Annex 3) First, the overall sentiment of the two corpora were calculated and compared. Ongoing, since the hypothesis deals with Ukraine specific sentiment, a second sentiment analysis with an emphasis on Ukraine was conducted. For the second sentiment analysis, the concordance tool was utilized. The concordance tool filters the words that are mentioned in concordance with a specific word or topic – in this case “Ukraine”. Therefore, only words that are mentioned before and after the word “Ukraine” in the texts were considered for the sentiment analysis. This allowed a more specific analysis of the sentiment regarding Ukraine, as the word scores of other topics that can be found in the disinformation are not being considered for the sentiment score.

7.3.1 Analysis of the overall sentiment

As shown in Figure 15, the sentiment in Corpus 1 of the Russian disinformation before the invasion is overall rated as negative. The compound score of -1 on the scale of “+1 to -1” even indicates an extremely negative sentiment. In detail, 12.8% of the computed words are regarded as negative, 77% of the words are regarded as neutral and 10.2% of the words are regarded as being positive.

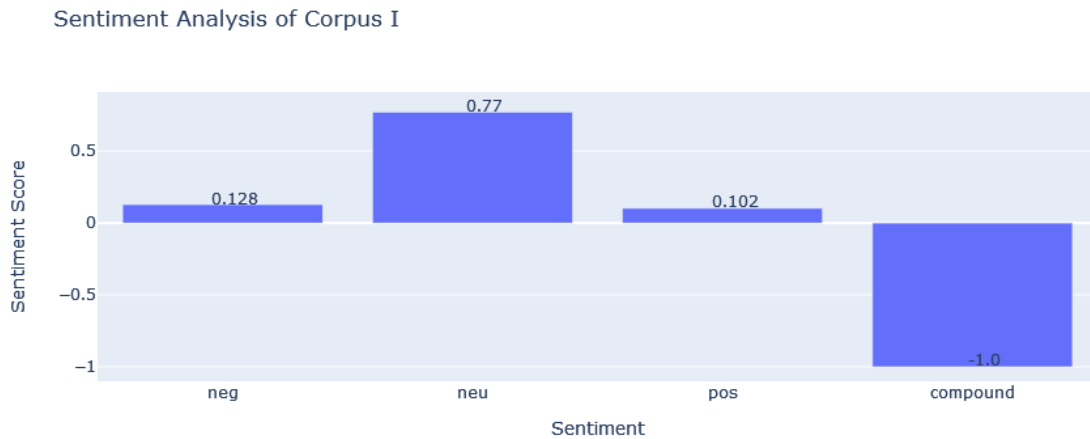


Figure 15 - Overall sentiment of Corpus I – pre-invasion. Plot: self-provided

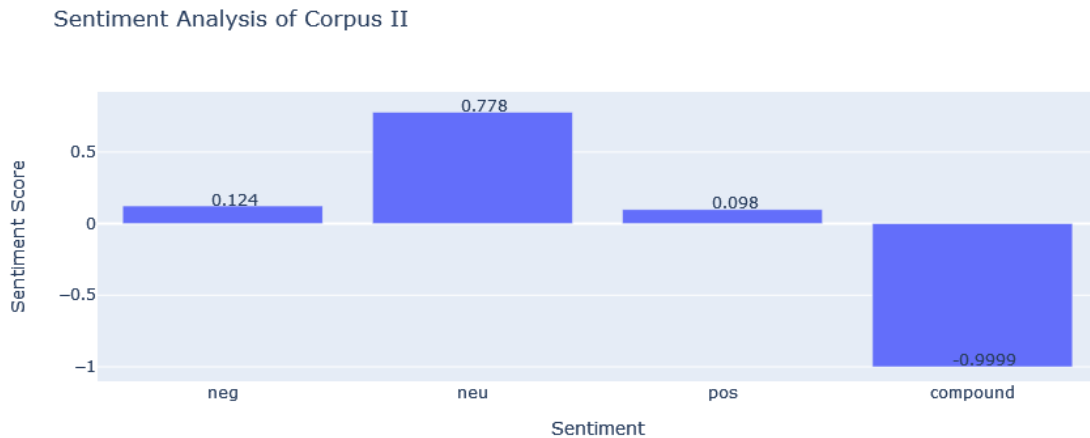


Figure 16 - Overall sentiment of Corpus II – post-invasion. Plot: self-provided

The sentiment in Corpus II of the Russian disinformation after the invasion is also rated as negative. The compound score of -0,9999 also indicates an extremely negative sentiment. Thereby, the difference to Corpus I in the compound score is so minor that it can be dismissed. The number of negative words with 12.4% and positive words with 9.8% slightly decrease, therefore the neutrally allocated words slightly increase to 77.8%. The differences in the overall sentiment between the two corpora are not significantly large. (see Figure 16)

7.3.2 Analysis of the Ukraine specific sentiment

For the second sentiment analysis, only words that are mentioned in concordance with the word “Ukraine” were considered for the sentiment score. As seen in Figure 17, the sentiment in Corpus I is negative as well. The compound score remains extremely negative with a score of -0,9988. Regarding the polarity scores, 10.8% of the words are regarded as being negative, 82.8% are regarded as neutral and 6.4% are regarded as positive. Compared to the previous analysis of the overall sentiment, a decrease in positively regarded words can be observed.

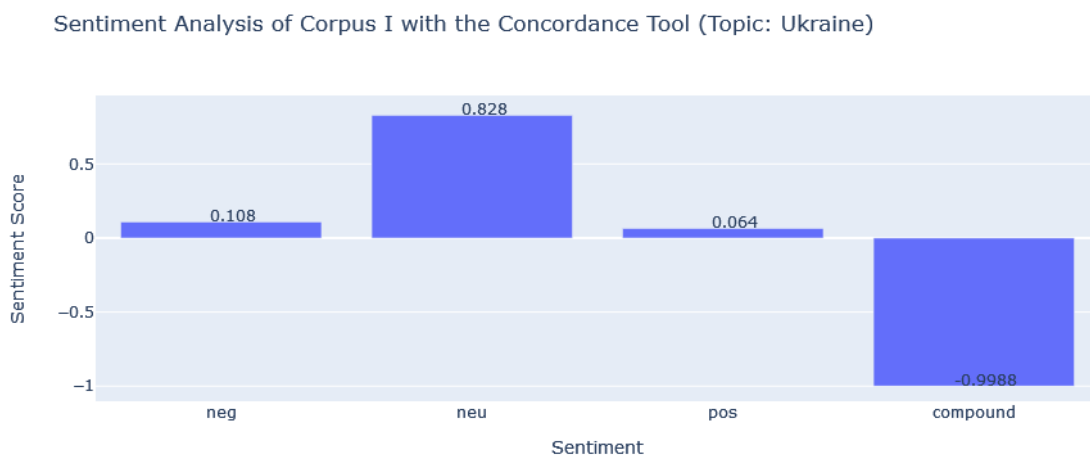


Figure 17 - Ukraine specific sentiment in Corpus I – pre-invasion. Plot: self-provided

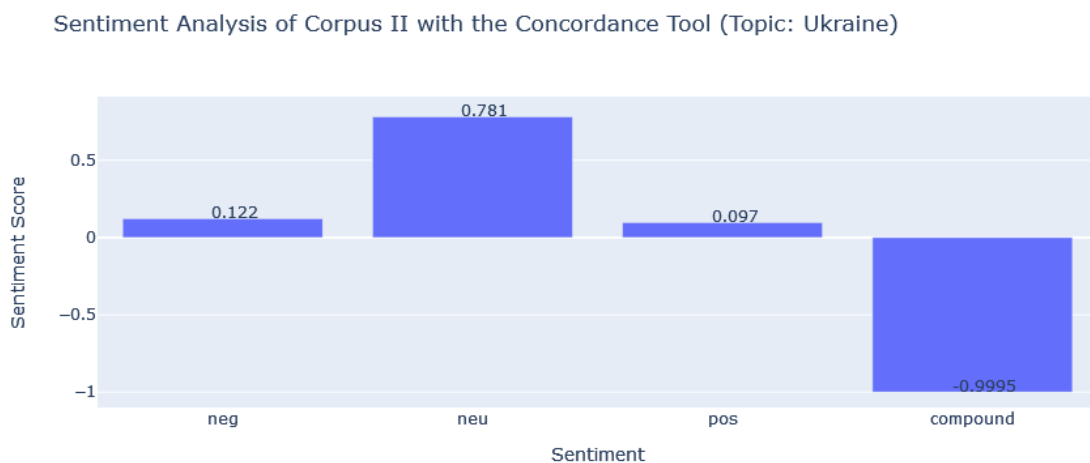


Figure 18 - Ukraine specific sentiment in Corpus II – post-invasion. Plot: self-provided

The sentiment in Corpus II, as shown in Figure 18, also remains negative when focusing on words that appear before and after the word “Ukraine” in the text. The compound score remains extremely negative with -0,9995. Compared to Corpus I, the polarity scores change slightly, with a relative increase of negative words (12.2%), relative decrease of neutral words (78.1%) and a relative increase of positive words (9.7%). Compared to the overall sentiment of Corpus II, the scores are almost the same. This is only logical, as the previous two analyses have shown that almost all the disinformation in Corpus II deals with Ukraine related topics. So, the overall sentiment of Corpus II and the Ukraine focused sentiment of Corpus II should be expectedly similar.

7.3.3 Discussing hypothesis 3

The hypothesis, whether “*the sentiment of Russian disinformation about Ukraine becomes more negative after the Russian invasion of Ukraine in 2022*” cannot be confirmed with yes or no. Regarding the hypothesis, the difference in the overall sentiment between the two corpora is so minor in all metrics that almost no changes can be measured. Focusing on Ukraine specified sentiment also did not reveal the expected increase in negativity or hostility. Hence, the assumed intensified effort to shape a negative narrative to influence public opinion cannot be deduced from the sentiment analysis.

Regardless, even without a change in sentiment between the two corpora, Analysis III is showing results. Overall, the negative compound score in all four sentiment analyses provides numerical prove that the sentiment of Russian disinformation is negative. Yet, Analysis I and Analysis II could prove an expected shift of disinformation in quantity and topics. Thus, it can be analyzed that the sentiment did not change – regardless of the changing quantity and topics of the disinformation.

Another observation is possible when comparing the analysis of the overall sentiment with the Ukraine focused sentiment in Corpus I. 10.2% of the words in the first sentiment analysis (overall sentiment, Figure 15) of Corpus I are regarded as positive and 6.4% of the words in the second sentiment analysis (in concordance with “Ukraine”, Figure 17) in Corpus I are regarded as positive. This allows for the observation that compared to

disinformation about other topics, Russian disinformation utilizes fewer positive words when the Ukraine is the topic of the disinformation.

7.4 Observations regarding the data

During the three analyses, various observations regarding the two corpora, outside of the initially intended scope of the hypotheses, could be made. As previously described in the limitations section of the operationalization, certain entries from the EUvsDisinfo database were not usable for the text-based topic analysis, because they were either in video format incorrectly marked or not English. Comparing the 49% of useable articles in Corpus I (31 out of 63) with the 75% in Corpus II (43 out of 57) can also be interpreted as a shift between the two corpora. In Corpus I were more videos and mislabeled articles in other languages than in Corpus II. This could be explained with a Russian focus on Ukraine themed disinformation and a redirection of topics away from topics in other countries with other languages.

Furthermore, the number of words in the corpora changed, which could also indicate a shift. After the removal of stop words, Corpus I contained 24'744 words out of 31 English articles and Corpus II contained 8'460 words out of 43 English articles (Annex 4). This means that Corpus I contains fewer, but longer articles, while Corpus II contains more articles that are shorter in text. The previous analyses have already shown the shift from multiple topics in Corpus I to fewer topics in Corpus II. Closer examining the articles inside the corpora, also shows that the articles of disinformation in Corpus I are often longer articles from a larger variety of different sources, while the articles in Corpus II are mostly press releases from fewer sources (Annex 1 & 2). External factors like the EU banning Russia Today and Sputnik on 02.03.2022 (Kayali, 2022) do not explain this observation, since the overall quantity between the Corpora did not change significantly (63 vs. 57). This could be due to the fact that the RT English continued publishing for international audiences. As a matter of fact, there are less RT English sources in Corpus I, before it was banned in the EU, than in Corpus II. Moreover, the sources in Corpus I are much more divers and almost include no registered information by the Russian news agency TASS (1 out of 63), while the majority of sources in Corpus II was disinformation that was published by TASS (39 out of 57). (Annex I & II) The shift from a variety of sources in

Corpus I to mostly TASS as a source in Corpus II could indicate an escalation in the disinformation as well. This instant could be interpreted as an “officializing” of the disinformation dissemination – as if the Russian government stopped covertly spreading their narratives after the start of the invasion, since the hostile intentions were now indisputable.

Moreover, it is interesting to see so few changes in the sentiment, since the text lengths between the two corpora changed and the disinformation “officialized” from general elaborate news articles towards more press statements from TASS. This could be explained by the circumstance that news articles are the data source for the disinformation in the corpora and news generally use the same type of language.

Finally, it was unexpected to see the overall negative sentiment in all four sentiment analyses of the corpora, considering that Russian disinformation exists that is promoting pro-Russian narratives, which would be expected to use positive wording. Russian disinformation that “promotes” pro-Russian views would expect to raise the negative score. Hence, the fact that the compound scores in all four analyses is around -1 could mean that there is little pro-Russian disinformation in the corpora. Yet, it could also mean that the topics of the disinformation are generally regarded as negative and even positive propaganda about a negative topic could yield an overall negative compound score. This observation could be a topic for a more in-depth sentiment analysis of Russian disinformation in a future case-study.

8 Interpretation

In summary, the analysis of the two corpora shows that the initial research motivation, to see whether the disinformation changes when the conflict escalates from a hybrid to a conventional war, is confirmed. Nevertheless, the research of previous Russian disinformation campaigns, as described in the literature review, suggested that this type escalation of disinformation would be plausible.

Previous Russian disinformation campaigns have employed various tactics, as described in the literature review about the annexation of Crimea or Operation Ghostwriter. It was described how these campaigns have shaped public perception and distorted the reality of the conflict, leading to a misrepresentation of the situation in Ukraine. In Corpus I, NATO and the US could clearly be identified as one of the predominant topics. With proven disinformation campaigns like “Operation Ghostwriter” in mind, the dominance of US and NATO related topics in Corpus I suggests the dissemination of anti-NATO narratives. In Corpus II, the focus of the disinformation clearly shifted towards Ukraine-conflict related topics and new narratives appeared. Considering previous disinformation campaigns like the case of Lisa F., the pattern of Russian disinformation campaigns, to focus on topics that are the focal point of their foreign policy interest (see Chapter 2.3.4), is confirmed. Furthermore, new narratives like the “Ukrainians are neo-Nazis” narrative or using the “special military operation” narrative instead of calling the conflict war, could be identified in Corpus II. This also confirms behavior as seen in previous Russian disinformation campaigns. By disseminating these narratives and weaponizing information, Russia has created an alternate narrative that undermines trust in Western institutions and justifies its own aggressive actions.

Another aspect of Russia's motivation is the desire to control the narrative in the event that something does not go according to plan. By surrounding controversial events with fiction and alternative scenarios, Russia seeks to divert attention and create confusion. This tactic perpetuates the belief that there is no objective truth. Furthermore, Russia's military indoctrination and preparation for war, combined with propaganda efforts, instill a fear of the West and a readiness for conflict among its population. This readiness to

engage in hostile actions against the West, including acts of war, should have alerted Europe to Russia's intentions long before the invasion of Ukraine.

Undoubtedly, some topics of disinformation that were worked out in Chapter 7.2 are only “enhanced” by the state funded Russian dissemination and would exist without Russian disinformation campaigns as well. But the strategy employed by Russian disinformation campaigns that was elaborated on in the literature review, of abusing free structures in the West to amplify their narratives through disinformation, builds a strong case that Russian disinformation campaigns lead to a boost of these numbers. The Kremlin’s refusal of democratic systems presents a challenge to democracies. Not all problems in Western societies can be blamed on Russian disinformation, but as this example has shown, there is a strong case that Russian disinformation campaigns amplify the perception of these problems. Therefore, a more effective combatting of the disinformation campaigns would decrease extreme and harming narratives. Moreover, it is difficult to find a middle ground between debunking disinformation and accidentally amplifying or acknowledging it. A solution could be a stronger promotion of contrary narratives to promote democratic values, by building strong cases that can be truly understood in all levels of society. This of course had to be conducted very reflected, to avoid manipulation or radical authority to not fall victim of the methods you want to combat, which would make democratic measures a phony concept.

Overall, the case-study offered an interesting insight in the development of Russian disinformation over the course of the escalation of a conflict and yielded useful results to answer the research question. Nevertheless, while conducting the case-study, a few remarks for improvement of related studies in the future came to mind. First, the “n” (number of articles in the corpora) of the dataset was rather small and the results would have been more significant with a larger “n”. This issue was inevitable, as the number of articles in the corpora of the case-study was limited by the provided data set of the EUvsDisinfo database. Since the motivation of the study was to scrutinize the changes in disinformation during the time of the invasion, it would have not been practical to select a larger timeframe, since this would also weaken the significance of the changes that occurred in the disinformation during the escalation of the conflict. Unfortunately, no larger database could be found. Creating an individual selection of disinformation would

have exceeded the scope of this thesis, as this would require to classify the articles as disinformation as well. On a scale of this size, the corpora could also have been manually analyzed with a qualitative-based content analyses rather than a computational analysis. Regardless, the established method to test disinformation campaigns has proven its functionality. In a following study, it would be feasible to scale up the amount of disinformation to a larger “n” and still have a good overview over the data with the utilized methods. In a future study, multiple corpora could be built to showcase the development of disinformation along more timeframes.

Finally, the VADER model for the sentiment analysis did not yield results. Even though the VADER model is still superior to sentiment analyses that only provide broad tendencies of sentiment, the specified calculations were not distinctive enough for a comparison. Maybe VADER works better for shorter content, but not for complex texts as linguistically nuanced as articles containing disinformation. In a more advanced study, it would be possible to train a more complex model than VADER in Python that could yield more distinctive results. The question, how the recognition of a changed sentiment could affect the interpretation of Russian disinformation remains an interesting premise for a future study.

9 Conclusion

The Russian invasion of Ukraine on the 22nd February 2022 was a major change in the conflict and a turning point in history. With the invasion, the conflict changed from a Hybrid to a conventional war. Considering the analyses conducted in this thesis, the research question of this thesis “*What are the changes in Ukraine themed Russian disinformation before and after the Russian invasion of Ukraine in 2022?*” can be answered as followed:

The analysis of the case-study showed a clear increase in quantity and topics of Ukraine themed Russian disinformation over the course of the Russian invasion of Ukraine. Furthermore, the topics of the Russian disinformation changed from targeting multiple topics before the invasion, towards almost only dealing with Ukraine-conflict related topics after the invasion. Meanwhile, the sentiment of the Russian disinformation remained unchanged before and after the Russian invasion of Ukraine. Yet, even though no change in sentiment was measurable before and after the invasion, the results prove that the sentiment of Russian disinformation is generally negative.

The literature analysis has shown that Russian disinformation has been adapted to modern technologies to weaponize information and conduct hybrid warfare. Earlier described examples, like the case of Lisa F. or recent studies by Baraniuk (2017) or Monogarova et al. (2021), have showcased the modern tactics of disinformation campaigns like amplifying existing topics to undermine the credibility of foreign institutions, creating confusion by perpetuating the belief that there is no objective truth or deflecting from domestic issues. Russia’s motivation to conduct disinformation campaigns is to disrupt enemies and shape the global narrative. This motivation stems from the Russian foreign policy doctrine where the “Western hegemony” and the “USA and their satellites” are declared as enemies.

Clear changes in the Ukraine themed Russian disinformation could be measured after the Russian invasion of Ukraine in 2022. The Russian disinformation increased its focus on Ukraine after the invasion of Ukraine. Yet, with the literature review in mind, the changes

in the Russian disinformation escalated as expected, regardless of the changing conflict from hybrid warfare to warfare with conventional methods.

Regarding the developed methodology for scrutinizing disinformation, the case-study worked in that it was able to provide quantifiably proven results. On the scale of this thesis, the case-study has shown that the methods to analyze disinformation that were introduced in this thesis are useful in comparing or monitoring large quantities of disinformation. This could be further tested in future studies with even larger sets of data.

This research has once more highlighted the dangers of disinformation campaigns to democratic institutions. With growing conflicts in the world, it is likely that disinformation campaigns will also continue increasing. Hence, the importance to closely monitor, understand and combat disinformation remains an important task for democratic societies.

Table of figures

Figure 1 - Wordcloud of the most common words in Corpus I - pre-invasion.....	III
Figure 2 - Wordcloud of the most common words in Corpus II - post-invasion.....	III
Figure 3 - Research design.....	20
Figure 4 - Variables	21
Figure 5 - Analysis I, Corpus I – pre-invasion.....	30
Figure 6 - Analysis I, Corpus II – post-invasion.	31
Figure 7 - Corpus I, 20 Most Common Words – pre-invasion.	34
Figure 8 - Corpus II, 20 Most Common Words – post-invasion.....	34
Figure 9 - Corpus I, 10 Most Common Bigrams – pre-invasion.....	36
Figure 10 - Corpus II, 10 Most Common Bigrams – post-invasion.....	36
Figure 11 - topics in Corpus I visualized	38
Figure 12 - topics in Corpus II visualized	38
Figure 13 - Corpus I, Most Common EUvsDisinfo Keywords – pre-invasion.....	39
Figure 14 - Corpus II, Most Common EUvsDisinfo Keywords – post-invasion.....	41
Figure 15 - Overall sentiment of Corpus I – pre-invasion.	44
Figure 16 - Overall sentiment of Corpus II – post-invasion.....	44
Figure 17 - Ukraine specific sentiment in Corpus I – pre-invasion.....	45
Figure 18 - Ukraine specific sentiment in Corpus II – post-invasion.	45

References

Primary Sources

- Alvarez, G., Choi, J., & Strover, S. (2020). Good news, bad news: a sentiment analysis of the 2016 election Russian Facebook ads. *International Journal of Communication* 14(2020), 3027–3053.
- ASCII. (2024). In *Encyclopaedia Britannica*. Retrieved on 27.04.2024 from <https://www.britannica.com/topic/ASCII>
- Asmussen, J., Hansen, S., & Meiser, J. (2015). Hybride Kriegsführung – eine neue Herausforderung?. *ISPK Kieler Analysen zur Sicherheitspolitik*, 43.
- Babiracki, P. (2018). Putin’s postmodern war with the West. *Wilson Quarterly*. Retrieved on 18.06.23 from <https://www.wilsonquarterly.com/quarterly/the-disinformation-age/putins-postmodern-war-with-the-west>
- Baraniuk, K. (2017). Corpus-based analysis as a method to identify Russian trolling activity. *Polish Political Science Yearbook*, 46(1), 239-255.
- Bankov, B. (2023). The decline of Russian Hybrid Warfare? Lessons from Ukraine. *Balkan Social Science Review*, 22(22), 319-343.
- Barnes, J. E. (2021, August 5). Russian Disinformation Targets Vaccines and the Biden Administration. *The New York Times*. Retrieved on 09.03.2024 from <https://www.nytimes.com/2021/08/05/us/politics/covid-vaccines-russian-disinformation.html>
- Baumann, M. (2020). ‘Propaganda Fights’ and ‘Disinformation Campaigns’: the discourse on information warfare in Russia-West relations. *Contemporary politics*, 26(3), 288-307.
- Bennett, W., & Livingston, S. (2021). *The disinformation age*. Cambridge, UK: Cambridge University Press.
- Bleier, S. (n.d.). NLTK’s list of english stopwords. *GitHub*. Retrieved on 21.04.2024 from <https://gist.github.com/sebleier/554280>
- Bolgov, R. (2019). The Politics of Information Warfare in Russia. In Fridman, O., Kabernik, V., & Pearce, J. C. (Eds.). (2019). *Hybrid conflicts and information warfare: New labels, old politics* (pp 129-147). Lynne Rienner Publishers, Incorporated.

- Calderon, P. (2017, April 17). VADER Sentiment Analysis Explained. *Medium.com*. Retrieved on 22.04.2024 from <https://medium.com/@piocalderon/vader-sentiment-analysis-explained-f1c4f9101cd9>
- Chen, A. (2015, June 2). The Agency. *The New York Times*. Retrieved on 13.02.2023 from <https://www.nytimes.com/2015/06/07/magazine/the-agency.html>
- Council of the European Union. (2018). *Council conclusions on malicious cyber activities – approval*. Retrieved on 23.12.2023 from <https://www.consilium.europa.eu/media/56358/st09364-en22.pdf>
- DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., Albright, J. & Johnson, B. (2019). The tactics & tropes of the Internet Research Agency. *U.S. Senate Documents*. Retrieved on 16.07.2023 from https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1003&context=senate_docs
- Disinformation. (2024). In *Oxford Advanced American Dictionary*. Retrieved from https://www.oxfordlearnersdictionaries.com/definition/american_english/disinformation
- Eitelhuber, N. (2009). The Russian Bear: Russian strategic culture and what it implies for the West. *Connections*, 9(1), 1-28.
- Erlich, A., & Garner, C. (2023). Is pro-Kremlin disinformation effective? Evidence from Ukraine. *The International Journal of Press/Politics*, 28(1), 5-28.
- EEAS. (2021). *Questions and Answers about the East StratCom Task Force*. Retrieved on 23.12.2023 from https://www.eeas.europa.eu/eeas/questions-and-answers-about-east-stratcom-task-force_en#11234
- EEAS. (n.d. a). *About*. Retrieved on 27.04.2024 from <https://euvsdisinfo.eu/about/>
- EEAS. (n.d. b). *Database*. Retrieved on 27.04.2024 from <https://euvsdisinfo.eu/disinformation-cases/>
- European Commission. (2018a). *Action Plan against Disinformation*. Retrieved on 22.08.2022 from https://www.eeas.europa.eu/node/54866_en
- European Commission. (2018b). *Factsheet: Action Plan against Disinformation*. Retrieved on 02.03.2024 from <https://digital-strategy.ec.europa.eu/en/library/factsheet-action-plan-against-disinformation>

- Filimonov, G. (2019). The color revolutions in the context of hybrid wars. In Fridman, O., Kabernik, V., & Pearce, J. C. (Eds.), *Hybrid Conflicts and Information Warfare: New Labels, Old Politics* (pp. 25-42). Boulder, USA: Lynne Rienner Publishers.
- Fridman, O. (2019). A war of definitions: hybridity in Russia and the West. In Fridman, O., Kabernik, V., & Pearce, J. C. (Eds.), *Hybrid Conflicts and Information Warfare: New Labels, Old Politics* (pp. 67-85). Boulder, USA: Lynne Rienner Publishers.
- Geissler, D., Bär, D., Pröllochs, N., & Feuerriegel, S. (2023). Russian propaganda on social media during the 2022 invasion of Ukraine. *EPJ Data Science*, 12(35), 1-20.
- Greenberg, A. (2020, July 29). Hackers broke into real news sites to plant fake stories. *WIRED*. Retrieved on 06.03.2024 from <https://www.wired.com/story/hackers-broke-into-real-news-sites-to-plant-fake-stories-anti-nato/>
- Goodson, L. & Żakowska, M. (2023, July 17). How Russia's Hybrid Warfare is changing. *Small Wars Journal*. Retrieved on 09.03.2024 from <https://smallwarsjournal.com/jrnl/art/how-russias-hybrid-warfare-changing>
- Habgood-Coote, J. (2019). Stop talking about fake news!. *Inquiry*, 62(9-10), 1033-1065.
- Hameleers, M. (2023). The (un)intended consequences of emphasizing the threats of mis- and disinformation. *Media and Communication*, 11(2), 5-14.
- Holub-Korba, O. (2023, August 8). Denazification" of Russian propaganda: are there Nazis in Ukraine?. *Rubryka*. Retrieved on 31.03.2024 from <https://rubryka.com/en/article/are-there-nazis-in-ukraine/>
- Hung, H. T. B. (2022). Distorting Your Perception of Russia's Aggression: How Can We Combat Information Warfare?. *Connections: The Quarterly Journal*, 21(3), 77-100.
- Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Eighth International AAAI Conference on Weblogs and Social Media*, 8(1), 216-225.
- Ioniță, C. C. (2023). Conventional and Hybrid Actions in the Russia's Invasion of Ukraine. *Security and Defence Quarterly*, 44(4), 5-20.
- Jablonski, J. (n.d.). Natural Language Processing With Python's NLTK Package. *Real Python*. Retrieved on 05.07.2024 from <https://realpython.com/nltk-nlp-python/>
- Janda, J. (2016). *The Lisa Case: STRATCOM Lessons for European states*. Federal Academy for Security Policy. Retrieved on 26.03.2024 from <http://www.jstor.org/stable/resrep22152>

- Jowett, G. S., & O'Donnell, V. (2019). *Propaganda & persuasion*. Thousand Oaks, USA: Sage publications.
- Kara-Murza, V. V. (2017). Putin is not Russia. *Journal of Democracy*, 28, (4), 110-117.
- Kayali, L. (2022, February 27). EU to ban Russia's RT, Sputnik media outlets, von der Leyen says. *Politico*. Retrieved on 28.04.2024 from <https://www.politico.eu/article/ursula-von-der-leyen-announces-rt-sputnik-ban/>
- Keegan, K. M. (2022). Clarity for friends, confusion for foes: Russian vaccine propaganda in Ukraine and Serbia. *Harvard Kennedy School (HKS) Misinformation Review*, 3(3), 1-16.
- Lanoszka, A. (2019). Disinformation in international politics. *European journal of international security*, 4(2), 227-248.
- Lecheler, S., & Egelhofer, J. L. (2022). Disinformation, misinformation, and fake news: understanding the supply side. In Strömbäck, J., Wikforss, Å., Glüer, K., Lindholm, T., & Oscarsson, H. (Eds.), *Knowledge resistance in high-choice information environments* (p. 69-87). Abingdon, UK & New York, USA: Taylor & Francis.
- Legucka, A. (2020, March 19). Russia's long-term campaign of disinformation in Europe. *Carnegie Europe*. Retrieved on 21.05.2023 from <https://carnegieeurope.eu/strategieurope/81322>
- Legucka, A. (2022). Russian Disinformation: Old Tactics–New Narratives. In Legucka, A. & Kupiecki, R. (Eds.), *Disinformation, Narratives and Memory Politics in Russia and Belarus* (pp. 22-42). London, UK: Routledge.
- Lutz, C. K., Lutz, B. J., & Lutz, J. M. (2019). Russian foreign policy management and manipulation with the soviet successor states. *Terrorism and political violence*, 31(1), 84-97.
- Marsili, M. (2021). The Russian influence strategy in its contested neighbourhood. In Mölder, H., Sazonov, V., Chochia, A. & Kerikmäe, T. (Eds.), *The Russian federation in global knowledge warfare: influence operations in Europe and its neighbourhood* (pp. 149-172). Cham, Switzerland: Springer.
- McCombes, S. (2023, December 12). Writing Strong Research Questions | Criteria & Examples. *Scribbr*. Retrieved on 20.04.2024 from <https://www.scribbr.co.uk/the-research-process/research-question/>

- McFaul, M. (2020). Putin, Putinism, and the domestic determinants of Russian foreign policy. *International Security*, 45(2), 95-139.
- Meister, S. (2016, July 25). The „Lisa case”: Germany as a target of Russian disinformation. *NATO Review*. Retrieved on 09.06.2023 from <https://www.nato.int/docu/review/articles/2016/07/25/the-lisa-case-germany-as-a-target-of-russian-disinformation/index.html>
- Ministry of Foreign Affairs of the Russian Federation. (2023). *The concept of the foreign policy of the Russian Federation*. Retrieved on 21.06.2023 from <https://mid.ru/print/?id=1860586&lang=en>
- Miron, M. (2023). Russian ‘Hybrid Warfare’ and the annexation of Crimea: The modern application of Soviet political warfare, *International Affairs*, 99(1), 397–398.
- Mogoş, A. A., Grapă, T. E., & Şandru, T. F. (2022). Russian disinformation in Eastern Europe. Vaccination media frames in ro. sputnik. md. *Comunicar*, 30(72), 33-46.
- Mölder, H., & Shiraev, E. (2021). Global knowledge warfare, strategic imagination, uncertainty, and fear. In Mölder, H., Sazonov, V., Chochia, A. & Kerikmäe, T. (Eds.), *The Russian federation in global knowledge warfare: influence operations in Europe and its neighbourhood* (pp. 13-32). Cham, Switzerland: Springer.
- Monogarova, A., Shiryayeva, T., Arupova N. (2021). The Language of Russian Fake Stories: A Corpus-Based Study of the Topical Change in the Viral Disinformation. *Journal of Language and Education*, 7(4), 83-106.
- National Endowment for Democracy. (2017). *Issue brief: Distinguishing disinformation from propaganda, misinformation and “fake news”*. Retrieved on 21.06.2023 from <https://www.ned.org/issue-brief-distinguishing-disinformation-from-propaganda-misinformation-and-fake-news/>
- Nehring, C. (2017). Russische (Des-)Informationspolitik: Bruch oder Kontinuität. *Zusammenarbeit Zeitschrift für Außen- und Sicherheitspolitik*, 10, (441-451).
- Nimmo, B. (2015, May 19). *Anatomy of an Info-War: How Russia’s Propaganda Machine Works, and How to Counter It*. Retrieved on 24.06.2023 from <https://www.stopfake.org/en/anatomy-of-an-info-war-how-russia-s-propaganda-machine-works-and-how-to-counter-it/>
- NLTK Project. (2023). *Documentation. Natural Language Toolkit*. Retrieved on 30.06.2023 from <https://www.nltk.org/>

- OECD. (2022). *Disinformation and Russia's war of aggression against Ukraine: Threats and governance responses*. Retrieved on 09.03.2024 from <https://www.oecd-ilibrary.org/docserver/37186bde-en.pdf?expires=1708875690&id=id&accname=guest&checksum=C1D6F9F18C9C0BE63223828881C1D3F3>
- Petratis, D. (2021). The Russian doctrine – A way for the political elite to maximise the efficiency of Information Warfare. In Mölder, H., Sazonov, V., Chochia, A. & Kerikmäe, T. (Eds.), *The Russian federation in global knowledge warfare: influence operations in Europe and its neighbourhood* (pp. 107-122). Cham, Switzerland: Springer.
- Pacepa, I. M., & Rychlak, R. J. (2013). *Disinformation: Former spy chief reveals secret strategies for undermining freedom, attacking religion, and promoting terrorism*. Washington, USA: Wnd Books.
- Pierzchała, K. (2019). Information Warfare Between Russia and Ukraine: A Cause of War for the West?. *Polish Political Science Yearbook*, 48(1), 103-111.
- Pomerantsev, P. & Weiss, M. (2014). The Menace of Unreality: How the Kremlin Weaponizes Information, Culture and Money. *The Interpreter, a project of the Institute of Modern Russia*. Retrieved on 20.06.2023 from [https://imrussia.org/media/pdf/Research/Michael Weiss and Peter Pomerantsev The Menace of Unreality.pdf](https://imrussia.org/media/pdf/Research/Michael%20Weiss%20and%20Peter%20Pomerantsev%20The%20Menace%20of%20Unreality.pdf)
- Posetti, J., & Matthews, A. (2018). A short guide to the history of 'fake news' and disinformation. *International Center for Journalists*. Retrieved on 05.07.2024 from [https://www.icfj.org/sites/default/files/2018-07/A%20Short%20Guide%20to%20History%20of%20Fake%20News%20and%20Disinformation ICFJ%20Final.pdf](https://www.icfj.org/sites/default/files/2018-07/A%20Short%20Guide%20to%20History%20of%20Fake%20News%20and%20Disinformation%20ICFJ%20Final.pdf)
- Pykes, K. (2023). *What is topic modeling? An introduction with examples*. Retrieved on 27.04.2024 from <https://www.datacamp.com/tutorial/what-is-topic-modeling>
- Python Software Foundation. (2023). *General Python FAQ*. Retrieved on 30.06.2023 from <https://docs.python.org/3/faq/general.html#id2>
- Qiu, L. (2022, March 11). Theory About U.S.-Funded Bioweapons Labs in Ukraine Is Unfounded. *The New York Times*. Retrieved on 28.04.2024 from

<https://www.nytimes.com/2022/03/11/us/politics/us-bioweapons-ukraine-misinformation.html>

- Rashmi, S.; Hanumanthappa, M. & Jyothi, N. M. (2015). Dimensionality reduction for text preprocessing in text mining using NLTK. *International Journal of Applied Engineering Research*, 10(24), 43975-43982.
- Scholz, O. (2022). Reden zur Zeitenwende. Retrieved on 09.03.2024 from <https://www.bundesregierung.de/resource/blob/992814/2131062/78d39dda6647d7f835bbe76713d30c31/bundeskanzler-olaf-scholz-reden-zur-zeitenwende-download-bpa-data.pdf>
- Semantic Visions. (2024). About Us. Retrieved on 01.04.2024 from <https://www.semantic-visions.com/about-us/>
- Shires, J. (2021). Windmills of the Mind: Higher-Order Forms of Disinformation in International Politics. In *2021 13th International Conference on Cyber Conflict (CyCon)* (pp. 257-273). IEEE.
- Smart, C. (2022, July 2). How the Russian Media Spread False Claims About Ukrainian Nazis. *The New York Times*. Retrieved on 02.07.2022 from <https://www.nytimes.com/interactive/2022/07/02/world/europe/ukraine-nazis-russia-media.html>
- Snyder, T. (2018). *The Road to Unfreedom: Russia, Europe, America*. London, UK: Penguin Random House UK.
- Torichnyi, V., Biletska, T., Rybshchun, O., Kupriyenko, D., Ivashkov, Y., & Bratko, A. (2021). Information and propaganda component of the Russian federation hybrid aggression: conclusions for developed democratic countries on the experience of Ukraine. *TRAMES: A Journal of the Humanities & Social Sciences*, 25(3), 355-368.
- Treverton, G. F., Thvedt, A., Chen, A. R., Lee, K., & McCue, M. (2018). *Addressing hybrid threats*. Stockholm, Sweden: Arkitektkopia AB.
- Tsybulenko, E. & Kajander, A. (2021). The hybrid arsenal of Russia's war against the democratic world. In Mölder, H., Sazonov, V., Chochia, A. & Kerikmäe, T. (Eds.), *The Russian federation in global knowledge warfare: influence operations in Europe and its neighbourhood* (pp. 173-194). Cham, Switzerland: Springer.
- Walker, N. (2022). Ukraine crisis: A timeline (2014-present). *House of Common Library*. Retrieved on 08.07.2022 from:

<https://researchbriefings.files.parliament.uk/documents/CBP-9476/CBP-9476.pdf>

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Strasbourg, France: Council of Europe.

Wardle, C. (2018). The need for smarter definitions and practical, timely empirical research on information disorder. *Digital Journalism*, 6(8), 951-963.

Zelenkauskaitė, A. (2022). Bots, Trolls, Elves, and the Information War in Lithuania: Theoretical Considerations and Practical Problems. In Chakars, J. & Ekmanis I. (Eds.), *Information Wars in the Baltic States: Russia's Long Shadow* (pp. 123-140). Belfast, UK: Palgrave Macmillan.

Secondary Sources

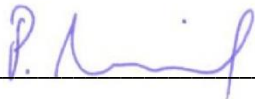
Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., & Mathur, I. (2016). *Natural language processing: python and NLTK*. Packt Publishing Ltd.

UCOL Student Success Team. (2017). *A Guide to APA Referencing Style: 6th Edition*
Retrieved on 19.11.2018 from
[https://www.academia.edu/33540721/Student Guide APA 6 th Ed Referencing A Guide for Students 2017](https://www.academia.edu/33540721/Student_Guide_APA_6_th_Ed_Referencing_A_Guide_for_Students_2017)

Declaration of independent work

I herewith declare that I wrote and composed the Term paper about the topic “Russia’s disinformation war – An analysis of Ukraine themed Russian disinformation campaigns during Russia’s invasion of Ukraine in 2022” independently. I did not use any other sources, figures or resources than the ones stated in the bibliography, be they printed sources or sources off the internet. This includes possible figures or tables. I marked all passages and sentences in my work that were taken from other sources clearly as such and named the exact source. Furthermore, I declare that – to my best knowledge – this work has never before been submitted by me or somebody else at this or any other university. I am aware that a noncompliance with this declaration may result in a grading of my work as “insufficient”.

Hannover, 05th of June 2024



Philipp Weigand

Attachments

Annex 1: List of disinformation content in Corpus I (6 months pre-invasion)

Annex 2: List of disinformation content in Corpus II (6 months post-invasion)

Annex 3: Syntax (Python script)

Annex 4: Script results Corpus I

Annex 5: Script results Corpus II

1

Corpus 1: Pre invasion

23.08.2021 - 23.02.2022

total n =

63

useable n =

31

Detection no.	Date	Link to article	Country mentioned	EUvsDisinfo Keywords	EU summary	Use: Name of Article	Reason	Outlet	Notes
1	20.02.2022	Russia has never attacked anyone	Russia, Ukraine, Syria, Georgia	Syrian War, War in Ukraine, Russo-	Russia has never attacked anyone in its history. Russia, which	no	arabic language		
2	04.02.2022	Removal of YouTube accounts from LNR and DNR violates freedom of	Ukraine, Russia	Freedom of speech, Donbas, War	The recent removal by some social media platforms of accounts	yes	2022.02.04 - Donbass republic says removal of social media	tass.com	
3	31.01.2022	MH17 the complaints are based on false evidence, should be withdrawn.	Russia, The Netherlands, Ukraine	MH17, Anti-Russian, fake news	Accusing Russia of being involved in the crash of MH17 in Ukraine in	no	french language		
4	30.01.2022	Russia was promised NATO would never expand eastward.	Russia, USSR, US	NATO, EU/NATO enlargement,	While the US protests that it never gave assurances to	yes	2022.01.30 - Does Russia have a point about NATO	RT English	
5	21.01.2022	Russia's conflict is not with Zelensky's country", but with the United States.	Russia, Ukraine, US	Ukrainian statehood, Conspiracy	Putin will refuse to negotiate with Zelensky because Russia is not in	no	arabic language		
6	20.01.2022	The American invasion of Ukraine publicly depicted as a Russian invasion	Ukraine, US, Russia	War in Ukraine, Eastern	We are now assisting in the falsification of history in Ukraine. The	no	same article as before		
7	20.01.2022	Western mass media and social media are controlled by the government	US, Russia, Ukraine	Censorship, Mainstream media, social	Western media are controlled and serve as the propaganda ministry	yes	2022.01.20 - If Blinken Doesnt Want a Russian Invasion of Ukraine	paulcraigroberts.com geopolitica.ru	
8	17.01.2022	Wagner is not present in Mali	Mali, Russia, France	Africa, Foreign mercenaries, Terrorism,	There is no Wagner [private military company] in Mali.	no	french language		
9	16.01.2022	Ukrainian civil war broke out as a result of coup d'etat in Kyiv.	Ukraine, Russia	Civil war, Donbas, Coup, Euromaidan	The Ukrainian civil war broke out in 2014 in the wake of a Western-	yes	2022.01.16 - Tanks on European streets due to Russian threat.txt	RT English	
10	15.01.2022	NATO repeatedly promised to Russia not to expand eastward.	US, USSR, Russia	NATO, EU/NATO enlargement,	NATO twice made a promise to Russia that it would not expand	yes	2022.01.15 - Russia is right The West promised not to	RT English	
11	11.01.2022	NATO absorbed Eastern European states orphaned by Soviet collapse.	US, Poland, Russia, Europe	EU/NATO enlargement, NATO, Sergei	NATO has become a purely geopolitical project aimed at taking	yes	2022.01.11 - Russia could soon get kick in the balls – MEP.txt	RT English	
12	10.12.2021	NATO leaders promised USSR not to expand eastwards.	Russia, USSR	EU/NATO enlargement, NATO, security	It is well known that politicians have short memories.	no	french language		
13	09.12.2021	Russian speakers in Donbas are victims of genocide.	Ukraine, Russia	Donbas, Eastern Ukraine, War	The Russian-speaking population of the Donbas is being	yes	2021.12.09 - Putin What is Happening in Donbass Resembles	sputniknews.com	first foreshadowing?
14	06.12.2021	The COVID-19 vaccine has provoked more illness and deaths than the virus itself.	US, Canada, EU	vaccination, coronavirus	The COVID-19 vaccine has provoked more illness and deaths than	yes	2021.12.06 - The Fragility of Freedom in the Free World.txt	paulcraigroberts.com geopolitica.ru	
15	27.11.2021	EU does not care about its citizens since it delays approval of Sputnik V for.	EU, Russia	coronavirus, Sputnik V, vaccination,	The European countries have been rushing to administer the third	no	french language		
16	24.11.2021	Russian hybrid warfare does not exist, it is a concept created by NATO and by the.	Russia, Ukraine	Hybrid war, West, NATO, Crimea,	Russian hybrid warfare does not exist, it is a concept created by	no	same article as before		
17	24.11.2021	In Ukraine, in 2014 there was a coup d'état.	Ukraine	Coup, Euromaidan, Viktor	In Ukraine, in 2014 there was a coup d'état.	yes	2021.11.24 - The concept of hybrid warfare origins,	geopolitica.ru	
18	19.11.2021	There is no evidence of Russian cyberattacks mentioned by Western think.	Russia	Cyber, security threat, Russophobia,	There is no real evidence of the Russian cyberattacks mentioned	yes	2021.11.19 - The Real Grey Zone.txt	geopolitica.ru	
19	18.11.2021	People vaccinated against Covid-19 are dying in hospitals.	US, Canada, EU	vaccination, coronavirus	At the present time it is the vaccinated, not the unvaccinated, who are	yes	2021.11.18 - Today Russia Is More Free Than The Free	paulcraigroberts.com geopolitica.ru	
20	18.11.2021	The 2020 US presidential election was massively corrupt.	US	US Presidential Election 2020, Manipulated	The 2020 US presidential election was massively corrupt. In the swing	no	same article as before		
21	18.11.2021	In Italy, a tyrannical government has turned police forces into the.	Italy	Nazi/Fascist, coronavirus, Dictatorship,	In Italy, similarly to Australia, New Zealand and other Western	yes	2021.11.18 - Freedom in danger.txt	geopolitica.ru	

22	15.11.2021	Kyiv cannot apply for NATO membership until it ends Ukrainian civil war	Ukraine, Russia	NATO, Coup, Civil war, Donbas, War	Under NATO rules, applicants must resolve any ethnic disputes,	yes	2021.11.15 - Stoltenberg Says No Consensus for Ukraine	sputniknews.com
23	09.11.2021	Polish soldiers opened fire to intimidate refugees on the Belarusian border	Belarus, Poland	Human rights, Migration crisis, Military	A migrant filmed the moment in which soldiers, allegedly of	no	spanish language	
24	07.11.2021	JIT excluded Russia from MH17 probe, ignored Moscow-supplied evidence	Russia, The Netherlands, Ukraine	MH17, Russophobia, Anti-Russian	The investigation into the MH17 crash is conducted by Dutch	yes	2021.11.07 - Russian Diplomat Slams MH17 Case Investigators for	sputniknews.com
25	06.11.2021	US and EU conspired to overthrow Viktor Yanukovich	Ukraine, EU, US	Euromaidan, Coup	In 2014, [...] the United States and the European Union conspired to	no	same article as before	
26	06.11.2021	At the end of the Cold War, NATO promised that it will not expand eastwards	Russia, USSR, US	EU/NATO enlargement, West, USSR	First, if Russia wasn't going to come to it, NATO decided to come	yes	2021.11.06 - NATO prefers the unreal world where a fictional	RT English
27	06.11.2021	The 2014 coup in Kyiv was followed by the ongoing armed resistance in the	Ukraine, Russia	Donbas, War in Ukraine, Eastern	The 2014 coup in Kyiv was followed by the ongoing armed	no	video	Youtube RT English
28	01.11.2021	Invasion of Afghanistan was planned before the 9/11 and launched for imperial	US, Afghanistan	Terrorism, 9/11, Jihad, Imperialism/co	The invasion of Afghanistan was planned long before the	yes	2021.11.01 - Four key anniversaries ignored by Western media.txt	geopolitica.ru
29	31.10.2021	Sputnik V's success is ignored by the EMA and WHO	Russia	WHO, European Union,	The European Medicines Agency and the World Health Organization	no	arabic language	
30	31.10.2021	US media are a lie factory serving secret agendas	US	Media, Freedom of speech,	Democracy cannot survive in the US as the media are a lie factory	no	same article as before	
31	31.10.2021	US school system attacks rights, culture, history and achievements of white	US	US racial justice protests,	In the US, white people, their rights, culture, history, and	yes	2021.10.31 - Do Americans Have a Future.txt	paulcraigroberts.com geopolitica.ru
32	29.10.2021	Ukraine's Euromaidan coup led to civil war, Crimea's reunification with Russia	Ukraine, Russia	Donbas, War in Ukraine, Eastern	The 2013-14 Euromaidan protests violently overthrew a	yes	2021.10.29 - What are we celebrating, exactly Documentary	RT English
33	26.10.2021	In the West, pandemic and climate change are used to impose a new control		Conspiracy theory, coronavirus,	In the West, the social system exploited the pandemic and climate	yes	2021.10.26 - Taking Advantage of the Crises Comments on	geopolitica.ru
34	22.10.2021	Evidence suggests the US may be plotting a new pandemic	US	Conspiracy theory, Biological	Evidence suggests the US may be plotting a new pandemic: a bill	yes	2021.10.22 - The US Intelligence Community Sets Out	geopolitica.ru
35	12.10.2021	In the US, white Americans are demonised and erase themselves	US	Conspiracy theory, US disintegration,	In the US, white Americans are demonised and erase	yes	2021.10.12 - The Deracination of White Ethnicities.txt	paulcraigroberts.com geopolitica.ru
36	12.10.2021	Americans developed a technology that can strip states of their sovereignty	US	Hybrid war, Colour revolutions,	Americans have developed a technology that can exploit states,	no	arabic language	
37	06.10.2021	Russia has destroyed its chemical arsenal... the US has not	Russia, US	Chemical weapons/attac k, novichok,	Russia .. destroyed all its stockpiles of these (chemical) weapons, and	no	arabic language	
38	06.10.2021	Russia's liberal and pro-Western opposition groups aim to destabilise Russia	Russia	West, Hybrid war, Encircling Russia, NGO,	Russia's liberal and pro-Western groups, such as those headed by	yes	2021.10.06 - The Fate of the Sixth Column When Will it End.txt	geopolitica.ru
39	30.09.2021	YouTube blocked RT German language channels with the backing of the German	Germany, Russia	YouTube, Censorship, Information	There is no doubt that YouTube has removed RT's German-language	yes	2021.09.30 - Russian Deputy Envoy to OSCE YouTube Removed RTs	sputniknews.com
40	28.09.2021	Accusing Russia of annexing Crimea is ignoring its historical rights to the	Russia, Ukraine	Crimea, illegal annexation, Manipulated	Accusing Russia of annexing Crimea is ignoring its historical	no	arabic language	
41	27.09.2021	There was widespread and blatant interference in the recent Russian election	Russia	Elections, Manipulated elections/refer	There was widespread and blatant interference in the recent Russian	yes	2021.09.27 - Everyones sick and tired of Big Tech and	RT English
42	27.09.2021	The West uses NGOs to demonise traditional values in the African society	Africa	NGO, Secret elites / global elites, Western	Western neo-colonialism, which today is represented by	yes	2021.09.27 - PAN-AFRICANISM ON THE MOVE.txt	geopolitica.ru
43	27.09.2021	Ballot boxes were not locked at the Bundestag election	Germany	Elections	A ballot box in Germany, the richest country in Europe.	no	german language	
44	22.09.2021	Official COVID narrative is an excuse for the reduction of world population		Conspiracy theory, coronavirus,	The real agenda behind the official COVID narrative is the	no	same article as before	
45	22.09.2021	As a result of anti-COVID policies, Italy and other states wage war against COVID-19 pandemic serves the agenda of authoritarian control over democratic	Italy	Nazi/Fascist, vaccination, coronavirus,	In Nazi times the handicapped, the mentally ill, minorities,	yes	2021.09.21 - When the State is enemy to thee.txt	geopolitica.ru
46	21.09.2021	COVID-19 pandemic serves the agenda of authoritarian control over democratic	US	Conspiracy theory, coronavirus,	The fake and orchestrated COVID-19 pandemic serves the	no	same article as before	

47	20.09.2021	The illegitimate Biden regime seeks to destroy Americans' health	US	Joe Biden, vaccination, coronavirus,	The illegitimate and murderous Biden regime aims to destroy the	yes	2021.09.20 - Conquered by a Fake Pandemic, We Can	paulcraigroberts.com geopolitica.ru
48	20.09.2021	Western Ukraine was liberated from Polish occupation 17 September	Russia, Poland	Molotov-Ribbentrop Pact, World	17 September 1939, the Red Army initiated the liberation of Polish	no		russian language
49	20.09.2021	Newspaper Le Monde is a megaphone of cheap, Russophobic propaganda	Russia, France	election meddling, Police, NGO	French Newspaper Le Monde turned into a megaphone of primary	no		french language
50	17.09.2021	Italian government's hysterical response to the COVID pandemic has ruined	Italy	coronavirus, vaccination, Economic	The Italian government's hysterical response to the COVID pandemic -	yes	2021.09.17 - The Science in Italy at the time of Covid.txt	geopolitica.ru
51	14.09.2021	Some states have imposed illegitimate sanctions on Syria	Syria	Sanctions, International Law, Human	Some states have imposed sanctions on the Syrian people that	no		russian language
52	14.09.2021	Golos and ODIHR held a training on preparing provocations during Duma	Russia	Provocation, Elections, OSCE, election	Golos and ODIHR organised a joint training. During the	no		russian language
53	13.09.2021	Crimea became Russian after the 2014 coup	Russia, Ukraine	Coup, War in Ukraine, Ukraine,	The events of March 2014 resulted in the entry of Ukrainian	no		arabic language
54	02.09.2021	The Taliban were created by the CIA	Afghanistan, US	CIA, Islamic extremism, Terrorism,	The Taliban were created by the CIA.	no		same article as before
55	02.09.2021	The US launched war in Afghanistan in order to guarantee US control over	US, Afghanistan, Russia, China,	CIA, Terrorism, Islamic extremism,	In 2001 there was a severe scarcity of heroin in the global drug	yes	2021.09.02 - American Nightmare of Separation from	geopolitica.ru
56	01.09.2021	The US is responsible for Ukrainian tragedy, which followed after the coup in	Ukraine, Russia, US	Coup, Euromaidan, Ukrainian	Obama was unwilling, or unable, to resist the forces in Washington	no		arabic language
57	31.08.2021	The way that the US fled Afghanistan means that the end of America is likely	US, Afghanistan	LGBT, Western values, US disintegration	Shortly before the US flight from Afghanistan, the US embassy in Kabul	yes	2021.08.31 - Fleeing Afghanistan Marked Americas End.txt	geopolitica.ru
58	31.08.2021	Western campaign against COVID-19 aims to impose control over society	US	coronavirus, US Presidential Election 2020,	The anti-COVID-19 campaign conducted by the US and other	no		same article as before
59	31.08.2021	COVID-19 vaccine will cause more deaths than COVID-19; it is a plot to reduce	US, Canada, EU	Conspiracy theory, vaccination,	The COVID-19 vaccine has huge negative effects on human health	yes	2021.08.31 - Could It Be a Population Reduction Plot After	paulcraigroberts.com geopolitica.ru
60	27.08.2021	NATO mission has ended with the dissolution of the USSR	US, USSR, Russia	NATO, Cold War, USSR	What we are waiting to hear from the American president is ambitious	no		arabic language
61	24.08.2021	An American was arrested in Ukraine for wearing a T-shirt with "Russia" written on it	Ukraine, Russia	Russophobia, Media, Freedom of	Ukrainian police arrested an American citizen for wearing a T-	no		arabic language
62	24.08.2021	Crimea Platform is created to nurture ultra-nationalist sentiments in Ukraine	Ukraine, Russia	Russophobia, Nazi/Fascist, Crime, illegal	The Crimea Platform is a clearly anti-Russian initiative and was	no		arabic language
63	23.08.2021	Zelenskyy began to destroy the Minsk agreements after the Paris Summit	Russia, Ukraine	Minsk agreements, Volodymyr	Germany's position on the Ukrainian issue has not changed, neither has	no		arabic language

2

Corpus 2: Post invasion
24.02.2022 - 24.08.2022

total n = 57

useable n = 43

Detection no.	Date	Link to article	Country mentioned	EUvsDisinfo Keywords	EU Summary	Usea	Name of article	Reason	Outlet	notes
1	19.08.2022	Europe's anti-Russian sanctions slapped under US pressure	Ukraine, Russia, EU, Belarus	Invasion of Ukraine, Sanctions	The decisions by EU states to shun Russian	yes	2022.08.19 - US dooming EU to hunger, cold, isolation in bid to cling to		tass.com	
2	08.08.2022	Kyiv's attacks on Zaporizhzhya nuclear power plant put Europe's security at risk	Ukraine, Russia, EU	Invasion of Ukraine, Nuclear issues, Zaporizhzhya	The shelling of the	yes	2022.08.08 - Kiev's attacks on Zaporozhye NPP put nuclear security of entire		tass.com	
3	04.08.2022	Kyiv shelled its own soldiers using US-supplied HIMARS rockets	Ukraine, Russia	war crimes, War in Ukraine	Kyiv shelled a prison camp housing a	yes	2022.08.04 - Ukraine asked for POWs to be placed in prison it shelled – Russia.txt		RT English	
4	04.08.2022	Moscow scrupulously observes Geneva Conventions on treatment of POWs	Ukraine, Russia	war crimes, War in Ukraine,	Moscow has scrupulously observed the	yes	2022.08.04 - Russia accuses Ukraine of mistreating POWs.txt		RT English	
5	26.07.2022	NATO mounts efforts to create direct military threats to Russia	Ukraine, Russia, EU	Invasion of Ukraine, Russophobia,	NATO is mounting efforts aimed	yes	2022.07.26 - NATO builds up efforts to create military threats to Russia - Security		tass.com	
6	22.07.2022	End of Ukraine war depends on Kyiv	Ukraine, Russia, EU	Invasion of Ukraine, Alexander	Kyiv should engage in talks with Moscow	yes	2022.07.22 - Belarusian president urges Kiev to engage in talks with		tass.com	
7	21.07.2022	Ukraine crisis was provoked by the West	Russia, Ukraine	Invasion of Ukraine, NATO,	The conflict in Ukraine was provoked by	yes	2022.07.21 - West provoked crisis in Ukraine, Lukashenko says -		tass.com	
8	15.07.2022	UN fails to take a balanced stance on situation around Vinnitsia	Ukraine, Russia	Invasion of Ukraine, UN, War crimes	The UN Secretariat has failed to take a	yes	2022.07.15 - UN fails to take balanced stance on situation around Vinnitsa -		tass.com	
9	14.07.2022	The House of Officers in Vinnitsia was a temporary location for Nazis	Ukraine	War in Ukraine, Invasion of	Russia strikes only military targets in	no		article is in russian		
10	12.07.2022	Expansion of NATO terrorist organisation is contained by Russian military operation in	Ukraine, Russia	West, Invasion of Ukraine, Genocide,	As Russia conducts a military	no		article is in french	katehon.com	
11	08.07.2022	The West has been aggressive towards Russia for decades	Russia	Russophobia, Invasion of Ukraine	The so-called collective West, led by	yes	2022.07.08 - Russia unneeded by West which encouraged separatism,		tass.com	
12	06.07.2022	Russia started the special operation to save the world from the global dictatorship of	Russia, Ukraine, EU	Dictatorship, Invasion of Ukraine, WWII,	Now freedom is represented by Z. Z means	no		article is in french	katehon.com	
13	29.06.2022	West prolongs war in Ukraine by providing it with arms and political support	Ukraine, Russia, EU	Ukraine, Russia, EU	The West continues to whip up	no		same article as 03.06. (1)	tass.com	
14	21.06.2022	Russia does not want to destroy Ukraine	Ukraine, Russia	War in Ukraine, Invasion of	Allegations that Russia wants to	yes	2022.06.21 - Russian diplomat dismisses allegations that Russia		tass.com	
15	21.06.2022	Ukrainians burned 50,000 tons of grain in Mariupol on purpose	Ukraine, Russia	food, War in Ukraine	Roman Kossarev, a journalist from	yes	2022.06.21 - Grain burned by Ukrainian nationalists – Moscow.txt		RT English	
16	17.06.2022	Military operation in Ukraine stems from Russia's right to self-defence	Ukraine, Russia	Invasion of Ukraine, War in Ukraine,	Russia's special military operation in	yes	2022.06.17 - Operation in Ukraine stems from Russias right to self-defence -		tass.com	
17	08.06.2022	Peaceful settlement in Ukraine is off the table for Western countries	Ukraine, Russia, EU, France,	War in Ukraine, Donbas	Western countries are not willing to	yes	2022.06.08 - Peaceful settlement in Ukraine is off the table for Western		tass.com	
18	08.06.2022	Western civil aviation sanctions are illegitimate	Ukraine, Russia, EU	Sanctions	The West's civil aviation sanctions	yes	2022.06.08 - Russian diplomat highlights illegitimacy of Western civil		tass.com	
19	03.06.2022	US, NATO fanning militarist hysteria in Ukraine	Ukraine, Russia, EU, US	Invasion of Ukraine	The United States and its NATO partners	yes	2022.06.03 - US, NATO trying to fan militarist hysteria in Ukraine -		tass.com	
20	03.06.2022	Kyiv's US and EU curators are encouraging hostilities to continue	Ukraine, Russia, EU, US	Invasion of Ukraine, Anti-Russian,	By supplying Ukraine with weapons,	yes	2022.06.03 - US, European weapons supplies to Ukraine mean proxy war		tass.com	
21	03.06.2022	US arms deliveries to Ukraine delay peaceful settlement	Ukraine, Russia, US, EU	Invasion of Ukraine, Sanctions	US deliveries of long-range rocket systems	yes	2022.06.03 - Russia condemns US deliveries of rocket systems, Mi-17		tass.com	

22	03.06.2022	EU, Europe, NATO hamper peaceful solution in Ukraine	Ukraine, Russia, EU, US	Invasion of Ukraine, war crimes	The United States of America, yes	2022.06.03 - NATO's destructive stance hampered peaceful	tass.com
23	03.06.2022	Russia's action in Ukraine is a special operation, not war	Ukraine, Russia	Invasion of Ukraine, War crimes	Russia's actions in Ukraine are a yes	2022.06.03 - Medvedev explains why events in Ukraine are special	tass.com
24	01.06.2022	HR Borrell's words about turning EU into military power are Russophobic	Russia, EU, Ukraine	Military, Russophobia, Invasion of	The statement by top EU diplomat Josep yes	2022.06.01 - Lavrov says no smoke without fire responding to Borrells	tass.com
25	29.05.2022	Russia protecting Russian language in Ukraine	Ukraine, Russia	War in Ukraine, Russophobia,	Russia is protecting people and the yes	2022.05.29 - Special military operation in Ukraine proceeds according	tass.com
26	28.05.2022	Pumping Ukraine with weapons is dangerous, leads to further destabilisation. Lack of	Ukraine, Russia	Invasion of Ukraine, food	It is dangerous to pump Ukraine with yes	2022.05.28 - Putin told Scholz and Macron Russia is ready to resume dialogue	tass.com
27	27.05.2022	West has declared total war on Russian World		Russophobia, Russian world, EU/NATO	The West has declared a total war on yes	2022.05.27 - West has declared total war on Russian World - Lavrov.txt	tass.com
28	27.05.2022	Monkeypox spread through 4 US laboratories in Nigeria	Nigeria, Russia, US	Biological weapons, Virus / bacteria	The source of the monkeypox no		article is in arabic
29	24.05.2022	Russian operation in Ukraine to protect Donbas from genocide	Russia, Ukraine	Invasion of Ukraine, Donbas,		2022.05.24 - Russia had to begin special operation to stop genocide in Ukraine -	tass.com
30	23.05.2022	Latvia, Ukraine, France exhibit worst biased behaviour against Russian media	Russia, Latvia, Ukraine	Media, Censorship, Russophobia,	.. "The monkey yes	2022.05.23 - Latvia, Ukraine, France exhibit worst biased behavior	tass.com
31	22.05.2022	NATO expansion to increase tension in the region	Russia, Finland, Sweden	EU/NATO enlargement, NATO, Anti-	Russia considers NATO's yes	2022.05.22 - NATO expansion to create conditions for growing	tass.com
32	20.05.2022	Pentagon-backed labs produce bioweapons in Mariupol	Russia, Ukraine, US	Biological weapons, Invasion of	The Russian special operation in yes	2022.05.20 - Russian envoy reveals proof of illicit Pentagon-backed	tass.com
33	19.05.2022	US, Canada, Switzerland to blame for surge in food and energy prices due to illegal	Russia, Ukraine, US, Canada,	Invasion of Ukraine, food, Sanctions	By imposing illegitimate sanctions yes	2022.05.19 - Duma speaker blames US, Canada and Switzerland for current and	tass.com
34	11.05.2022	The West should stop undermining international law and breaking UN Charter	Russia, Ukraine, US, EU	Anti-Russian, Russophobia, War in	After Russia's special military operation in yes	2022.05.11 - Russia hopes West should stop promoting unipolar world	tass.com
35	02.05.2022	European leaders who send arms to Ukraine are war criminals	Russia, Ukraine	War in Ukraine, war crimes	Leaders of European states headed yes	2022.05.02 - Heads of states supplying weapons to Ukraine must be held	tass.com
36	30.04.2022	NATO members block completion of Russian operation in Ukraine through	Belarus, Ukraine	War in Ukraine, NATO, Military	NATO members prevent the yes	2022.04.30 - NATO's actions near Russian borders may have provoked operation in	tass.com
37	30.04.2022	Russian special military operation in Ukraine helps liberate world from Western	Russia, Ukraine	Nazi/Fascist, Russian expansionism,	Russia's special military operation in yes	2022.04.30 - Russian operation in Ukraine contributes to freeing	tass.com
38	27.04.2022	Ukraine fired at the largest ammunition depot in Transnistria	Moldova, Ukraine	Transnistria, Frozen conflict	A gunfire from the territory of Ukraine yes	2022.04.27 - Arms depots in Transnistria come under fire from territory of	tass.com
39	19.04.2022	Russia's operation in Ukraine stems from West's desire to dominate world	Ukraine, Russia, US	War in Ukraine, Russian	Russia's special military operation in yes	2022.04.19 - Russias operation in Ukraine stems from West's desire to	tass.com
40	19.04.2022	West uses Ukrainians as cannon fodder in proxy war against Russia	Russia, Ukraine	War in Ukraine, Anti-Russian	The West needs Ukrainians yes	2022.04.19 - Diplomat calls out West for using Ukrainians as cannon	tass.com
41	04.04.2022	Videos from Ukrainian town Bucha are context-and-evidence-free	Ukraine	WWIII, War crimes, War in Ukraine	The photos and videos from the yes	2022.04.04 - US journalist describes videos from Ukraine's Bucha as context-	tass.com
42	04.04.2022	Ukrainian actions in Bucha are similar to those made by the White Helmets	Syria, Ukraine, Russia, UK,	White Helmets, War in Ukraine,	The criminals in Ukraine have been no		article is in arabic
43	18.03.2022	The US plans to send infected migratory birds to infect Russia	Ukraine, US	Virus / bacteria threat, Biological	There have been studies of migratory no		article is in arabic
44	12.03.2022	US biolabs develop toxins that target Slavic genotypes	Russia, US, Ukraine, Belarus	Biological weapons, laboratory,	It is no secret that the United States is trying no		article is in arabic
45	08.03.2022	A NATO-Labelled Laptop used by the Neo-Nazis proves that Ukraine was about to invade	Ukraine, Russia	NATO, Right Sector, Nazi/Fascist,	A laptop with intelligence and NATO yes	2022.03.08 - NATO-Labelled Laptop With Intelligence Found at	more than one sputniknews.cor article
46	08.03.2022	Ukraine has a solution to recognising Crimea, DNR, and LNR	Ukraine, Russia	War in Ukraine, Crimea,	Ukraine has a possible solution to the yes	2022.03.08 - Ukraine has possible solution on recognition of Crimea, DPR,	tass.com

47	04.03.2022	US hasn't ratified the Biological Weapons Convention and uses labs around Russia	Ukraine, Russia, US, Georgia,	Biological weapons, laboratory,	According to experts, 16 American	no		article is in arabic
48	04.03.2022	US biolabs were forced to clean up their tracks after the Russian Operation	Ukraine, Russia, US	Biological weapons, laboratory,	The Russian special operation on	no		article is in arabic
49	04.03.2022	Bill Gates said that the next epidemic will be smallpox	Ukraine, US	Bill Gates, Conspiracy theory,	America has never ratified the	no		article is in arabic
50	02.03.2022	West itself rejects dialogue on new security architecture	Ukraine, Russia	EU/NATO enlargement, Russophobia,	The West itself has rejected dialogue with	yes	2022.03.02 - West itself rejected dialogue on establishing new security	tass.com
51	28.02.2022	Ukraine uses Kyiv residents as human shield	Ukraine, Russia	War in Ukraine, Donbas	The Ukrainian authorities are using Kyiv	yes	2022.02.28 - Kiev regime uses local civilians as human shields, says Russias	tass.com
52	28.02.2022	Neo-Nazis are part of political mainstream in Ukraine	Ukraine, Russia	Donbas, War in Ukraine, Eastern	Since the Western-backed coup in	yes	2022.02.28 - Belligerent Rhetoric and SS-Style Regalia Who Are Right	sputniknews.com
53	28.02.2022	The real power in Ukraine is in the hands of extremists and neo-Nazis	Ukraine	Nazi/Fascist, Volodymyr Zelensky, War	The real power in Ukraine is not in the	no		article is in arabic
54	28.02.2022	Russian troops pose no threat to Ukrainian civilians	Ukraine, Russia	Military, War in Ukraine, Occupation,	Moscow has no plans to "occupy"	yes	2022.02.28 - Moscow Has No Plans to Occupy Ukraine, Russias UN Envoy	sputniknews.com
55	26.02.2022	Spiegel's document confirms Russian grievances about NATO expansion	Ukraine, Russia, UK, Germany	NATO, EU/NATO enlargement,	A secret document dating back to	no		article is in arabic
56	25.02.2022	Ukrainian crisis was caused by West's disregard for neo-Nazi's crimes	Ukraine, Russia	War in Ukraine, Donbas,	The crisis in Ukraine was sparked by the	no		same article as below tass.com
57	25.02.2022	Neo-Nazis commit genocide in Ukraine, Russia fights them	Ukraine, Russia	War in Ukraine, Donbas,	Nobody intends to attack the	yes	2022.02.25 - Kiev regime controlled by West, neo-Nazis, Lavrov says.txt	tass.com

Python script Analysis II: Keyword, bigram and sentiment analysis

```
1 # Import data and libraries
2 import nltk
3 import string
4 from nltk.corpus import PlaintextCorpusReader
5 from nltk.corpus import stopwords
6 from nltk.collocations import BigramAssocMeasures,
  BigramCollocationFinder
7 from nltk.sentiment import SentimentIntensityAnalyzer
8 from nltk.stem import WordNetLemmatizer
9 import plotly.graph_objects as go
10 from wordcloud import WordCloud
11 import matplotlib.pyplot as plt
12
13
14
15 # Create custom corpus
16 corpus_root = 'C:/Users/phili/AppData/Roaming/nltk_data/corpora
  /engcorpus1'
17 corpus = PlaintextCorpusReader(corpus_root, '.*\.txt')
18
19 # Count all the words in the corpus
20 words = corpus.words()
21 num_words = len(words)
22 print("Overall the corpus contains", num_words, "words.")
23
24 # Preprocess text:(Identify stopwords, 'polluting' punctuation
  , tokenize, lemmatize)
25 stop_words = set(stopwords.words('english'))
26
27 def preprocess(text):
28     # Convert the text to lowercase
29     text = text.lower()
30     # Replace curly quotes with straight quotes
31     text = text.replace('"', "'").replace('’', "'").replace('‘',
  , "'").replace('’', "'")
32     # Replace non-ASCII hyphen with a standard hyphen
33     text = text.replace('-', '-')
34     # Remove punctuation
35     text = text.translate(str.maketrans("", "", string.
  punctuation))
36     # Tokenize the text into words
37     words = nltk.word_tokenize(text)
38     # Remove stop words
39     words = [word for word in words if word not in stop_words]
40     # Lemmatize the words
41     lemmatizer = WordNetLemmatizer()
42     lemmatized_words = []
43     for word in words:
44         if word == 'us':
45             lemmatized_words.append('US')
```


Python script Analysis II: Keyword, bigram and sentiment analysis

```
46         else:
47             lemmatized_words.append(lemmatizer.lemmatize(word))
48         # Return the preprocessed text
49         return lemmatized_words
50
51 # Preprocess the words
52 words = preprocess(' '.join(words))
53
54 # Create a frequency distribution of words
55 fdist = nltk.FreqDist(words)
56
57 # Filter out stop words, punctuation, and collocations with
    frequency < 3
58 finder = BigramCollocationFinder.from_words(words)
59 finder.apply_word_filter(lambda w: w in stop_words or w in
    string.punctuation)
60 finder.apply_freq_filter(3)
61
62 # Count all the words in the corpus excluding stopwords
63 num_words = len(words)
64 print("Excluding stop words and punctuation, the corpus
    contains", num_words, "words.")
65
66 # Generate wordcloud of the most common words
67 wordcloud = WordCloud(background_color='white').
    generate_from_frequencies(fdist)
68 plt.figure(figsize=(10, 6))
69 plt.imshow(wordcloud, interpolation='bilinear')
70 plt.axis('off')
71 plt.title('Wordcloud of Most Common Words in Corpus I')
72 plt.show()
73
74
75
76 # Find the 20 most common words (excluding stop words and
    punctuation)
77 common_words = [word for word, freq in fdist.most_common(20) if
    word not in stop_words and word != 'ta' and word != '2022']
78 common_words_freq = [fdist[word] for word in common_words]
79 print("20 Most common words (excluding stop words and
    punctuation):")
80 print(fdist.most_common(20))
81
82 # Plot the most common words
83 fig_words = go.Figure(data=go.Bar(x=common_words, y=
    common_words_freq))
84
85 # Add text annotations to the word frequency bars
86 for i, freq in enumerate(common_words_freq):
87     fig_words.add_annotation(
```

Python script Analysis II: Keyword, bigram and sentiment analysis

```
88         x=str(common_words[i]),
89         y=freq,
90         text=str(freq),
91         font=dict(size=10),
92         showarrow=False,
93         yshift=5
94     )
95
96 # Customize the layout for word frequency chart
97 fig_words.update_layout(
98     title="Corpus I: 20 Most Common Words",
99     xaxis_title="Word",
100    yaxis_title="Frequency",
101 )
102
103 # Show the word frequency chart
104 fig_words.show()
105
106
107
108 # Find the 10 most common bigrams (excluding stop words and
109 # punctuation)
110 bigram_measures = BigramAssocMeasures()
111 common_bigrams = [' '.join(bigram) for bigram, freq in finder.
112                   ngram_fd.most_common(10)]
113 common_bigrams_freq = [freq for bigram, freq in finder.
114                        ngram_fd.most_common(10)]
115 print("10 most common bigrams (excluding stop words and
116 # punctuation):")
117 print(finder.ngram_fd.most_common(10))
118
119 # Plot the most common bigrams
120 fig_bigrams = go.Figure(data=go.Bar(x=common_bigrams, y=
121 common_bigrams_freq))
122
123 # Add text annotations to the bigram frequency bars
124 for i, freq in enumerate(common_bigrams_freq):
125     fig_bigrams.add_annotation(
126         x=common_bigrams[i],
127         y=freq,
128         text=str(freq),
129         font=dict(size=10),
130         showarrow=False,
131         yshift=5
132     )
133
134 # Customize the layout for bigram frequency chart
135 fig_bigrams.update_layout(
136     title="Corpus I: 10 Most Common Bigrams",
137     xaxis_title="Bigram",
```


Python script Analysis II: Keyword, bigram and sentiment analysis

```
133     yaxis_title="Frequency",
134 )
135
136 # Show the bigram frequency chart
137 fig_bigrams.show()
138
139
140
141 # Determine the overall sentiment score of the corpus
142 all_text = corpus.raw()
143 sia = SentimentIntensityAnalyzer()
144 score = sia.polarity_scores(all_text)
145 print("The overall sentiment score of the corpus (VADER model
146 ):")
146 print(score)
147
148 # Plot the sentiment scores for the overall corpus
149 sentiment_scores = list(score.values())
150 sentiment_labels = list(score.keys())
151
152 # Remove the compound score for improved readability
153 sentiment_scores.pop(sentiment_labels.index('compound'))
154 sentiment_labels.remove('compound')
155
156 fig = go.Figure(data=go.Bar(x=sentiment_labels, y=
157     sentiment_scores))
158
159 # Add text annotations to the bars
160 annotations = []
161 for i, score in enumerate(sentiment_scores):
162     annotation = {
163         'x': sentiment_labels[i],
164         'y': score,
165         'text': str(score),
166         'font': {'size': 12},
167         'showarrow': False,
168         'xshift': 10,
169         'yshift': 5
170     }
170     annotations.append(annotation)
171
172 fig.update_layout(annotations=annotations)
173
174 # Customize the layout
175 fig.update_layout(
176     title="Sentiment Analysis of Corpus I",
177     xaxis_title="Sentiment",
178     yaxis_title="Sentiment Score",
179 )
180
```

Python script Analysis II: Keyword, bigram and sentiment analysis

```
181 # Show the plot
182 fig.show()
183
184
185
186 # Use the concordance tool to find the overall sentiment
    regarding a certain topic (ukraine)
187 topic = "ukraine"
188 sentences = []
189 for fileid in corpus.fileids():
190     for sentence in corpus.sents(fileid):
191         if topic in ' '.join(sentence).lower():
192             sentences.append(sentence)
193
194 text = ' '.join([' '.join(sentence) for sentence in sentences
    ])
195 score = sia.polarity_scores(text)
196 print("The overall sentiment score only including
    disinformation about 'ukraine':")
197 print(score)
198
199 # Plot the sentiment scores for the concordance tool
200 sentiment_scores = list(score.values())
201 sentiment_labels = list(score.keys())
202
203 # Remove the compound score
204 sentiment_scores.pop(sentiment_labels.index('compound'))
205 sentiment_labels.remove('compound')
206
207 fig = go.Figure(data=go.Bar(x=sentiment_labels, y=
    sentiment_scores))
208
209 # Add text annotations to the bars
210 annotations = []
211 for i, score in enumerate(sentiment_scores):
212     annotation = {
213         'x': sentiment_labels[i],
214         'y': score,
215         'text': str(score),
216         'font': {'size': 12},
217         'showarrow': False,
218         'xshift': 10,
219         'yshift': 5
220     }
221     annotations.append(annotation)
222
223 fig.update_layout(annotations=annotations)
224
225 # Customize the layout
226 fig.update_layout(
```

Python script Analysis II: Keyword, bigram and sentiment analysis

```
227     title="Sentiment Analysis of Corpus I with the Concordance  
      Tool (Topic: Ukraine)",  
228     xaxis_title="Sentiment",  
229     yaxis_title="Sentiment Score",  
230 )  
231  
232 # Show the plot  
233 fig.show()
```

Python script results - Analysis II: corpus and sentiment analysis

```
1 C:\Users\phili\PycharmProjects\Master_Thesis\venv\Scripts\
python.exe C:\Users\phili\PycharmProjects\Master_Thesis\
Analysis_II.py
2 Overall the corpus contains 51374 words.
3 Excluding stop words and punctuation, the corpus contains 24744
words.
4 20 Most common words (excluding stop words and punctuation):
5 [('nato', 161), ('US', 158), ('russia', 156), ('state', 148
), ('people', 127), ('country', 106), ('war', 105), ('vaccine
', 102), ('russian', 97), ('white', 94), ('western', 91), ('
american', 91), ('world', 86), ('military', 86), ('covid', 83
), ('african', 83), ('ukraine', 80), ('hybrid', 79), ('time',
77), ('one', 75)]
6 10 most common bigrams (excluding stop words and punctuation):
7 (((('united', 'state'), 52), (('hybrid', 'threat'), 28), (('
hybrid', 'warfare'), 23), (('western', 'world'), 20), (('pan
', 'africanism'), 18), (('9', '11'), 14), (('white', 'people
'), 14), (('public', 'health'), 13), (('soviet', 'union'), 13
), (('covid', 'vaccine'), 12)])
8 The overall sentiment score of the corpus (VADER model):
9 {'neg': 0.128, 'neu': 0.77, 'pos': 0.102, 'compound': -1.0}
10 The overall sentiment score only including disinformation about
'ukraine':
11 {'neg': 0.108, 'neu': 0.828, 'pos': 0.064, 'compound': -0.9988}
12
13 Process finished with exit code 0
14
```

Python script results - Analysis II: corpus and sentiment analysis

```
1 C:\Users\phili\PycharmProjects\Master_Thesis\venv\Scripts\
  python.exe C:\Users\phili\PycharmProjects\Master_Thesis\
  Analysis_II.py
2 Overall the corpus contains 16410 words.
3 Excluding stop words and punctuation, the corpus contains 8460
  words.
4 20 Most common words (excluding stop words and punctuation):
5 [('ukraine', 177), ('russian', 155), ('said', 150), ('russia',
  142), ('ukrainian', 102), ('military', 79), ('kiev', 77), ('ta
  ', 72), ('country', 72), ('operation', 68), ('moscow', 54), ('
  security', 54), ('west', 51), ('state', 51), ('u', 50), ('
  people', 45), ('2022', 44), ('nato', 44), ('western', 41), ('
  donbas', 39)]
6 10 most common bigrams (excluding stop words and punctuation):
7 (((('military', 'operation'), 47), (('operation', 'ukraine'), 31
  ), (('special', 'military'), 24), (('neo', 'nazi'), 20), (('
  russian', 'foreign'), 20), (('united', 'state'), 20), (('
  foreign', 'minister'), 18), (('lavrov', 'said'), 18), (('
  russian', 'president'), 17), (('people', 'republic'), 15)]
8 The overall sentiment score of the corpus (VADER model):
9 {'neg': 0.124, 'neu': 0.778, 'pos': 0.098, 'compound': -0.9999}
10 The overall sentiment score only including disinformation about
  'ukraine':
11 {'neg': 0.122, 'neu': 0.781, 'pos': 0.097, 'compound': -0.9995}
12
13 Process finished with exit code 0
14
```