uc3m | Universidad **Carlos III** de Madrid

Master Degree in Computational Social Science
Academic Year (2022 - 2023)

Master Thesis

"Mass Media Coverage of International Conflicts: Differences by Countries and Ideologies"

Artem Urlapov

Francisco Villamil Fernández

Madrid, August the 31st 2023

Abstract

While this piece of research is certainly not the first one that has been undertaken as an attempt to rigorously analyse mass media coverage of international conflicts, there is a reason to believe that a high degree of novelty is brought in.

This is so because the theme in question is addressed following the framework of the emerging field of Computational Social Science, and –for the first time in a study of this kind– the concepts of “psychic distance” and “multipolarity” are introduced.

Furthermore, the present author not only demonstrates the validity of usage of the aforementioned ideas for an exhaustive analysis of international conflicts in post-Cold War era, but the notion of “psychic distance” is further refined by making a distinction between *positive* and *negative* psychic distance.

Even though –due to the fact that computational requirements for this body of work are exceedingly high– Russo-Ukrainian War is analysed (for which, a total of 1.440 pieces of data have been extracted for a time interval of 60 days alone), a wide-ranging structure has been put in place.

This includes both an extensive and highly systematic review of a series of historical insights, sociological considerations, computational aspects, time series analysis, sentiment analysis, and / or frequency analysis.

Thus, theoretical and computational insights provided allow for manifold lines of research, that can be applicable not only to mass media coverage of international conflicts, but even to the study of other important events, such as climate change or economic crises (as long as data is available).

Key words (ELSST): Mass Media Coverage, International Conflict, Data Collection Methodology, Computer Languages, Statistical Analysis, Social and Political Polarization, Linguistic Analysis, International Relations, Cultural Change.

Dedication, Acknowledgements, and Humble Words of Wisdom

Dedication

The present piece of research is dedicated to the memory of my late grandmother, Ekaterina Sedova (1924 – 2020), who –through her great effort and constant dedication- helped me in large part become the person I am today, and whom I miss dearly.

Acknowledgements

On an academic level, I feel greatly indebted to a series of highly useful suggestions, insights, and remarks received from (both during the course in "Master's Thesis Seminar" and, in some cases, outside):

- Prof. Francisco Villamil Fernández (above all, for his infinite patience with me, and for his kindness and consideration in attentively responding to my enquiries).

- Prof. Iñaki Úcar Marques (for having pointed out to me for how to work best with time intervals in the given context, providing me with a piece of advice on how to make the R code of the "dynamic version" even more computationally efficient, as well as for his useful feedback on the possibility of bias in sentiment analysis –given that different languages in this body of work are employed–). Additionally, Prof. Iñaki Úcar Marques was in charge of "Data Visualization" course, skills gathered in which are applied throughout the present body of work.

- Prof. Margarita Torre Fernández (to whom I am thankful for having helped me keep focus on what my academic priorities should ideally be, given that this is an enormously time-consuming and computationally expensive piece of research).

- Prof. Gonzalo Génova Fuster (my esteemed friend and former Professor of the course in "Social and Ethical Issues of Big Data & A.I."). Without any aim of diminishing the highly valuable contributions of Professors in other courses, I can maintain that –without shadow of a doubt– this course was the one that helped me actually become a Computational Social Scientist. The one in which I became fully aware of all the implications of what being a Computational Social Scientist was all about.

- Prof. Carmen Torrijos Caruda (my former Professor of the course in "Text Mining" –Neuro-Linguistic Programming–), and Prof. Jorge Cimentada (my former Professor of the course in "Data Harvesting" –in which "web scraping" plays an important role–). Knowledge acquired in these two courses has been essential to undertake this challenge.

- Isabela Zeberio Aguerrevere (my former colleague, whose remarks during the course in "Master's Thesis Seminar" on the effects of "war fatigue" were particularly useful, and that were further addressed in an interaction with Prof. Iñaki Úcar Marques afterwards).

- Yassin Abdelhady (my former colleague, who was the first one to point out, during the course in “Master’s Thesis Seminar”, the computational challenges that I was possibly going to encounter myself with, in terms of interaction with the server).

- Prof. Rafael Castro Balaguer (my dear friend and Professor at Autonomous University of Madrid, whose vision as an Economic Historian was highly valuable to assess both quality and depth of some of my insights).

- Prof. Antonio Martín Arroyo (an esteemed friend and Professor at Autonomous University of Madrid, whose perspective on several of my points as a renowned Econometrician –among other merits, he holds an MBA in Statistics from the University of Chicago, and has worked extensively in the fields of Financial Econometrics and Bayesian Statistics– was, again, a deeply constructive one).

- Prof. Ricardo Correia Vaz Antunes Pereira (my former Professor of “International Finance” at Autonomous University of Madrid, who first acquainted me with the concept of “psychic distance”, which is one of the pillars of the present body of work).

On a personal level, I owe a tremendous sense of gratitude to:

- Lyudmila Doronina (my aunt –and, as of today, my only family–; for having supported me financially: due to the one-year nature of the Master in Computational Social Science, this was a demanding programme, and I could not combine my graduate studies with a full-time job during most part of the course).

- Luis Gómez Rufián (a dear friend and member of Gabinete Económico de La Moncloa, for his relentless support in all kinds of circumstances, good or bad).

- Romualdo Emilio Mira García (a great friend and PhD Candidate in Theoretical Mathematics at Complutense University of Madrid; I am particularly grateful to him for having further pushed me to maintain academic discipline at all times, dealing with a theme as affectively demanding for me as Russo-Ukrainian War undoubtedly is).

- Again, to Prof. Fernando Villamil Fernández, Prof. Iñaki Úcar Marques, and Prof. Margarita Torre Fernández; the three of whom were incredibly supportive of me at all times, knowing the emotional challenges that I was facing in view of Putin’s aggression against Ukraine: particularly, the feelings of enormous shame and guilt that I had to deal with, given the fact that I was born in Russia –even though I have lived in Spain two thirds of my life, and that I now hold Spanish nationality–. Quite honestly, I would not have finished my graduate studies in the present academic year, had it not been for the relentless support received from these three Professors.

Humble Words of Wisdom

"Only education is capable of saving our societies from possible collapse, whether violent or gradual.“ – Jean Piaget (1896 – 1980).

Contents

1. Introduction

1.1. Personal and Academic Motivation.

Both politics and geopolitics have always occupied an important place in my sphere of personal and academic interests. This is so, not least because I was brought up by my maternal aunt and my maternal grandmother, having been the latter (Ekaterina Sedova, 1924 – 2020) a top-tier Soviet diplomat throughout most of her life. Additionally, my mother (Taísia Urlapova –née Sedova–, 1954 – 2009) worked largely as a telecommunications engineer in the Soviet, and later, Russian Air Forces.

Thus, on a personal note –one that would years later enable me to do this piece of academic research–, I remember quite vividly three key events that took place relatively early in my life, but that have stayed with me forever since:

1. My mother (the only member of my –otherwise, quite left-leaning in ideological terms– family to do so) voting for Vladimir Putin in 2000 Russian Presidential Elections, in hope of a better and brighter future, after the disastrous decade of the 1990s.
   The rest of my family and myself (when I was still a Russian citizen) have always voted for Yabloko centre-of-left socio-liberal political party, which is –as of today– the only platform that stands in stark opposition to Putin's act of aggression against Ukraine (which has led to some prominent politicians, such as Lev Schlosberg –awardee of "Boris Nemtsov Prize" or "Free Press Prize"– and one of the few individuals in Russia able to win even in a scenario of completely rigged elections, being banned from running for public office on "Foreign Agent" charges).

2. American illegal occupation of Iraq, in what would come to be known as the Second Iraq War (2003 – 2011); which was widely broadcast by the media both in Russia and in Spain (the country in which I have resided since 2004 on a permanent basis, and whose citizen of I am of now).

3. Russia's gradual descent into authoritarianism, and eventually –as of now– totalitarianism, with the aggravating factor that not only do people of Russia suffer such a terrible and unfair regime, but another country –Ukraine– is currently being the victim of an aggression, ongoing since 24th of February of 2022. While some renowned scholars, such as John Mearsheimer –for instance–, have provided a rationale for the invasion of Ukraine; there certainly can be no justification whatsoever for war crimes committed by Russia: a clear example of this is the Bucha Massacre, an unspeakable atrocity committed in a small Ukrainian village by the Russian troops in March of 2022.

This third point has had a profound psychological and emotional impact on me. While international conflicts are – sadly enough– a rather common occurrence, never could I envisage my country of birth invading a territory the like of the country of Ukraine, one with which Russia has traditionally had enormously close ties in terms of history, culture, and language.

When I first met Prof. Francisco Villamil Fernández, little needs to be said about how I was feeling emotionally regarding the events of the 2022 Russian invasion of Ukraine.

Furthermore, what was even more troubling for me to understand and process was the total and complete demonisation of everything pertaining to Russia, something that did not occur even at the heights of the American illegal invasion of Iraq (in which many more war crimes –and on a higher scale– took place: we can all remember the infamous 1996 quote of the former United States Secretary of State Madeleine Albright aired on "60 Minutes" on how 500.000 dead Iraqi children was "a price worth it", or pictures of German shepherds –otherwise, very beautiful dogs– being thrown onto naked Iraqi prisoners, or music of Nine Inch Nails –one of my favourite music bands– being used to torture prisoners).

It was always common knowledge that responsibility lied with the government, not with the citizens. However, what we witnessed in the context of Russo-Ukrainian War has been, quite simply, astounding: orchestras refusing to perform the works of Tchaikovsky, Russian and Belarusian athletes being prevented from competing even under a neutral flag, Russian political refugees being expelled from multiple countries or being denied asylum, and so on and so forth.

By no means this is a stated as an attempt to diminish the unspeakable suffering of the Ukrainian citizens.

Rather, what is meant by this is to demonise an entire nation is a dubious strategy on a rational level. Resentment can only breed even more resentment, and –what is more worrying– prolongs the conflict:

On purpose of this, CSIS –Center for Strategic & International Studies– clearly states the following (with basis on data compiled by the Uppsala Conflict Data Program since 1946): "Of wars that last over a month but less than a year, only 24 percent end in a ceasefire. When interstate wars last longer than a year, they extend to over a decade on average, resulting in sporadic clashes." [1]

All in all, as I had the chance to meet and see more in detail some of the work that Prof. Francisco Villamil Fernández was carrying out, the more convinced did I become of the possibility of being able to turn something as dark and horrendous as the Russo-Ukrainian War –with all of the aforementioned consequences– into something academically constructive, and of social relevance and interest to my fellow colleagues, Professors, and a wider audience.

A final side-note: the previously described experiences have constituted an important personal motivation to rationalise these events, but they do by no means constitute an impediment to address the conflict in full compliance with the most rigorous academic objectivity standards, as will become evident throughout the entire piece of research, in what follows.

1.2. Antecedent Research.

In chronological order, a number of studies in the field of Mass Media Coverage of International Conflicts are presented:

- Arnaud Mercier in his 2005 research paper on *War and media: Constancy and Convulsion* reaches the following conclusion: "Rather than colliding head-on with the power of journalists, Western armed forces have shown themselves adept at playing the media game, offering tokens of openness as a means of better blunting public and media vigilance. From Kosovo to Iraq the procedure is the same: claim that you are limiting censorship to that which is needed for strategic effectiveness and protection of your forces; deter journalistic enterprise without actually disallowing it; dominate the proceedings by staging the action, if need be, but without doing so in an excessive manner. Like other entities in the public eye, the military has fully adapted to the demands of the media society and grasped the need to professionalize the mechanism by which it communicates with journalists. The media have become part of war. The military strategy today includes them as one of its objectives. Military operations are accompanied by media plans, media relations are handled by professionals and the armed forces invest in internal training to make their officers aware of the need to master the media process and nurture good relations with journalists. The military has acquired the know-how to provide "products" (reports, press kits) that meet journalists' needs." [2]

- Eytan Gilboa in his 2007 research paper on *Media and International Conflict: A Multidisciplinary Approach* makes a series of valid points.[3] Concretely, he states the following: "Since the end of the Cold War, the nature and level of international conflicts have considerably changed. Until the end of the Cold War, most international conflicts occurred between and among states, but afterward they mostly occurred at the intrastate or global levels." He further outlines the fact that research in the field has been not nearly multidisciplinary enough in order to rigorously analyse the realities of an increasingly more diverse and complex post-Cold World era world.

- Richard C. Rubens in his 2009 research paper on *The Impact of News Coverage on Conflict: Toward Greater Understanding*, largely echoes Gilboa (2007): "Despite the pervasiveness of both the media and conflict, the question has received surprisingly little scholarly attention." [4]

  Furthermore, he critically questions the role of mass media in coverage of international conflicts: "Conflict theory suggests that conflict escalates destructively when one or both of the parties view the conflict or a dispute as necessarily something that is won by one party and lost by the other—that the dispute is zero-sum. News media coverage can perpetuate such an understanding of conflict or a dispute. News stories about conflict frequently follow a structural paradigm that is sometimes called "issue dualism", in which the news media reduces complex issues to two competing sides that are articulated by familiar, predictable sources and that get roughly equal weight in their coverage."

  Another highly important insight of Rubens (2009) is the following: "An important contributor to the destructive escalation of conflict, particularly sustained conflict, is the delegitimization of the other side's perspective. News coverage can foster destructive escalation by promoting the denigration of one of the disputants, such as by marginalization or demonization."

- Teresa Joseph in her 2014 research paper on *Mediating War and Peace: Mass Media and International Conflict* elaborates a deeply critical assessment of mass media being complicit in promoting and amplifying the established domestic foreign policy agenda.[5] She refers to the Gulf War (1991 – 1992) as the first conflict which was entirely broadcasted, albeit following a clear "framing" narrative. Furthermore, Joseph (2014) makes several important points, in what to selective reporting, the sidelining of dissent, and censorship it refers. To provide three examples out of her important contribution:

“It has been widely argued that the reason why Americans were persuaded to support the invasion of Iraq was because the United States' media coverage during the days leading up to the war portrayed protest as unpatriotic and arguments against war as irrelevant. CNN's coverage of the 1991 Gulf War and other international conflicts during the following years led to advocates of the 'CNN effect' arguing that the media play an important role in determining foreign policy action, particularly with regard to 'human intervention’. The understanding was that the media set the agenda, and when they focus on one issue, political action follows”. (Gilboa, 2007, equally mentions the “CNN effect”.)

“Aware of the importance of the mass media in justifying or legitimising conflict, participants in international conflicts increasingly attempt to manipulate news coverage and use the media as a 'force multiplier.' Where the media are controlled by the state there is a more explicit bias towards official interpretations of events and securitisation of the 'enemy'. The media are needed not only to report from the battlefield, but also to justify or legitimise war. This has its impact on the understanding of international conflicts.”

“News involves the conscious selection of events. This selection of news is often based on the interests of the home country. Conflicts and suffering within the Third World often go unreported by the inter-national media unless the West has its own interests in the region. The Iran-Iraq War resulted in over a million deaths and was of major political importance but received relatively little media coverage in the West, as both parties were out of favour with the American public.”

The four aforementioned authors coincide on a series of facts:

- Very little research has –thus far– been carried out on such an important topic as Mass Media Coverage of International Conflicts.

- Mass Media is, nowadays, largely complicit with Dwight D. Eisenhower’s notion of “Military-Industrial Complex”. Furthermore, there is a tendency for self-censorship among journalists (Mercier, 2005).

- As it has been acknowledged by Gilboa (2007), multidisciplinarity has been a clearly lacking factor.

- Mass media is often complicit in promoting domestic foreign policy, using strategies such as denigration of the enemy (a notion addressed in 1.1.) and introducing selection bias.

- Complexity appears to be a highly expendable factor, and mass media is seldom interested in exploring and explaining to the general public the roots of the conflict (Rubens, 2009).

2. Research Objective

As can be seen from antecedent research, even though a series of authors make some deeply relevant points, many of them –to the mind of the present author– are not nearly comprehensive enough.

Especially, while Gilboa (2007) confirms the lack of multidisciplinarity in addressing Mass Media Coverage of International Conflicts, he does not try to amend the *status quo* by making one proposal or another.

Quite on the contrary, the objective of the present body of work is to present a comprehensive piece of research, one in which these shortcomings would be specifically and exhaustively addressed.

Furthermore, all the aforementioned studies include theoretical considerations only; while no endeavour to provide detailed data analysis is realised.

The main objective of the present body of work is, thus, to provide a detailed layout –on both a computational and theoretical level– that can serve as a "footprint" for a broad analysis not of Mass Media Coverage of International Conflicts only, but for many other purposes as well (this point will be covered broadly in the next section), given the manifold challenges of an increasingly complex and diverse world.

3. Hypotheses

The present body of work rests primarily on two main hypotheses:

3.1. "Psychic Distance"

In what to the first hypothesis it concerns, the present author owes his gratitude to Prof. Ricardo Correia Vaz Antunes Pereira. It was during his course in "International Finance" at Autonomous University of Madrid that Artem Urlapov –during his studies of the degree in Economics– first grasped the importance of the fact that –many a time– "psychic" and not so much physical distance plays a significantly larger role in cross-border commerce.

For instance, in the case of Spain, it is a fact that Mexico is located geographically further away than, say, Hungary. Nevertheless, when a Spanish company makes the choice to expand its business, there is much more likelihood that a branch will be established in Mexico rather than in Hungary, due to the aforementioned factor of cultural proximity (shared language, similar social norms, an environment of mutual trust, etc.).

Even though the term "psychic distance" is mostly used in international finance, the truth is that it was first coined by Edward Bullough in 1912, as pertaining to a strikingly different area (art and aesthetics).[6] In fact, in his seminal paper, the author makes mention of the notion that this phenomenon was first indirectly noticed by Aristotle in his "Poetics". Eventually, as time went by since the beginning of the Twentieth Century, the concept became widely adopted in international finance and marketing, and even –more recently– in computer science.[7]

In what follows, an empirical argument will be made as to why the use of "psychic distance" is a perfectly valid approach to analyse international conflicts.

It is a widely accepted fact that the current global situation can be characterised as one in which the world faces a "polycrisis" (a term that originated in the 1970s, but that has only recently been popularised by the British economic historian Adam Tooze).[8] To quote the author directly (in his response to when did this moment originate in time):

"I think you could make a good case that really the unfolding of this current moment starts in 2008, which is simultaneously not just the financial crisis, which we all remember, but also Putin's first aggression against Georgia. It is also the breakdown of the WTO in the Doha Round; it is setting the stage for the disappointment of the Copenhagen climate talks. And then on top of everything else, there was a swine flu epidemic in 2008/9.

The key things for me are economics, politics, geopolitics, and then the natural environment blowing back at us. And those four things, they don't reduce to a single common denominator. They don't reduce to a single factor."

Indeed, this seems to be a very valid argument.

However, while a case can be made with regard to the fact that economics or climate change are independent of “psychic distance” (everyone, to a greater or lesser extent, is affected), the scenario is different for politics and –quite certainly– for international conflicts.

Focusing on more recent times (for which data is available), one can instantly think of two main conflicts where “psychic distance”, for two different reasons, has played a crucial role.

The first one is the Syrian Civil War (2011 – ongoing), and the subsequent Syrian Refugee Crisis, which put a non-trivial strain on an already weakened Welfare State by years of fiscally restrictive (austerity) economic policy in the aftermath of the 2007-2008 Global Economic-Financial Crisis.

But this is not, by far, the only factor. In fact, in this particular scenario, one can clearly think of the term *negative* “psychic distance”. This is to say, a country like Germany receiving an influx of refugees with whom the population felt that no cultural bonds were shared.

As is well-known, this led to the rise of extreme right movements such as Pegida (Patriotic Europeans Against the Islamisation of the West), and further propelled the rise of Alternative für Deutschland far-right political party.

The second one is the Russo-Ukrainian War (2014 – ongoing), with its most critical phase commencing on 24th of February of 2022.

In this case, the term *positive* “psychic distance” can certainly be used. As such, in the year 2022 alone –according to statistics provided by United Nations Refugee Agency–, Poland (a country that shares a strong “psychic distance” with Ukraine) received 958.147 refugees in a highly welcoming way, with no cultural backlash whatsoever.[9] The same applies to Germany, which received 1.002.999 Ukrainian refugees in 2022, with arms wide open.

On the other hand, for comparison, during the rise of Pegida and Alternative für Deutschland (2013 – 2015), only 177.851 Syrian refugees were received in Germany (again, according to United Nations Refugee Agency official statistics). However, this number was enough to generate enormous discontent, especially in Eastern Germany.

Another deeply significant point for the purpose of this analysis is that, at the same time that the events of the Russo-Ukrainian War were unfolding, Germany received 522.575 Syrian refugees in the year 2022. Yet, this fact went notoriously unnoticed by the mass media.

This is not to mention that only 280 Syrian refugees were received by Poland in the entire year of 2022 (it has already been mentioned that the number is of 958.147 in the case of those originating from Ukraine).

One final point that further highlights the importance of “psychic distance” is that –during the rise of Pegida or Alternative für Deutschland– there was a widespread claim that only male Syrian refugees were coming.

Statistics point out in a different way, however. As such, in the case of Germany (as per United Nations Refugee Agency), out of a total of 177.851 Syrian refugees received between 2013 and 2015, 51.78% were female while 48.22% were male. In what to Ukrainians it concerns, 63.49% were female while 36.51% were male. Though there certainly is a difference, it would be demonstrably untrue to claim that most Syrian refugees were men. Certainly, one could make the argument that mass media should have played a much more critical role in assessing and purposefully addressing these blatantly false allegations. And, in what to this first hypothesis it refers, it should be fairly obvious to the interested reader by now that only "psychic distance" can explain these variations in the general perception of the public. Unfair and unethical as it may be, it is not the same to accept those who are closer to one in cultural terms rather than to do the same for those who do not share these socially important bonds.

All in all, with these official statistics in mind, it would certainly be difficult to reject the importance of "psychic distance", both in international conflicts and in mass media coverage of these.

3.2. "Multipolarity"

The core of this hypothesis rests that in post-Cold War world, there are less differences by individual countries in terms of ideology. Rather, ideology is commonly shared across a number of globes (also known as spheres of influence, or poles).

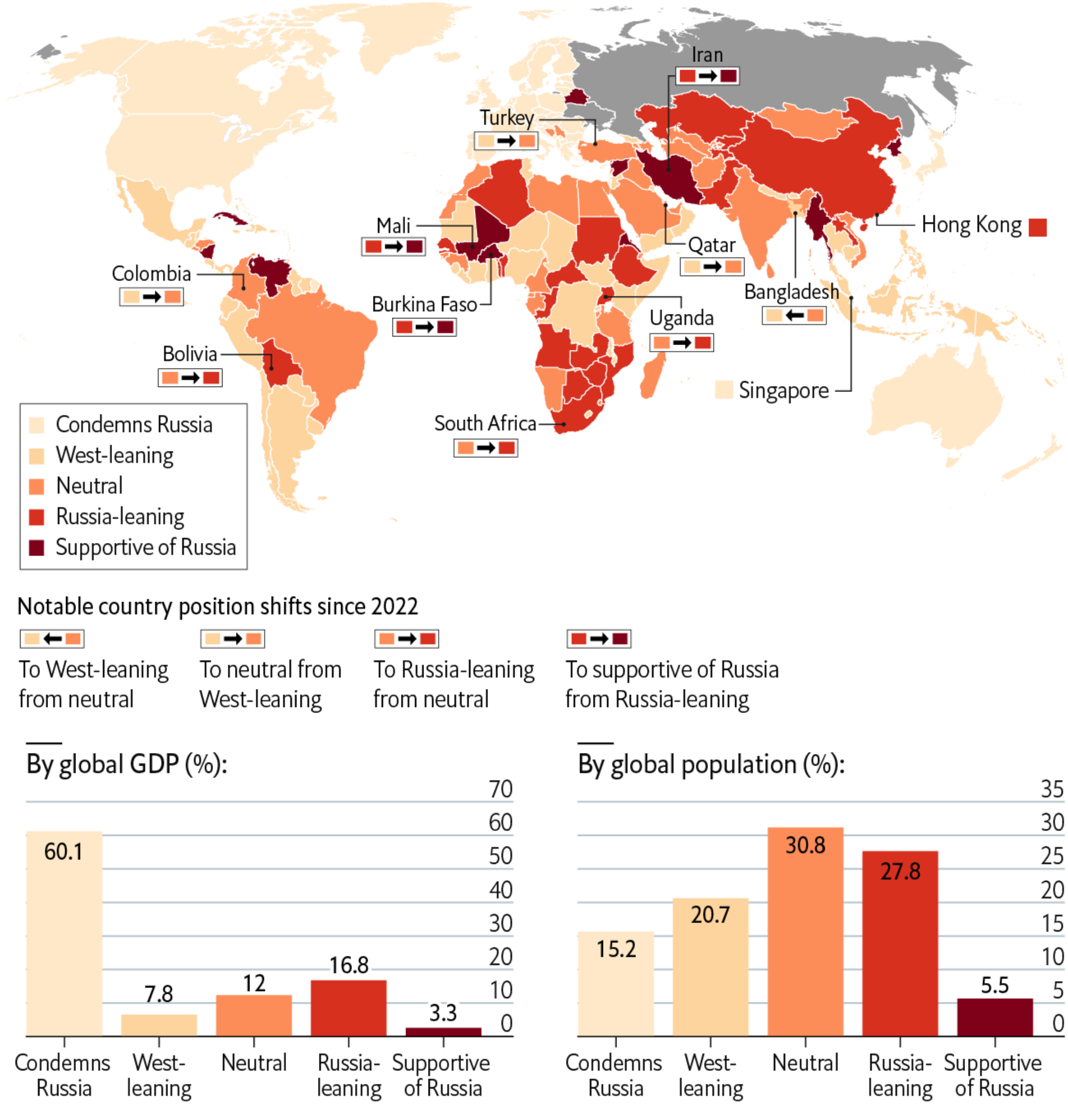
For instance, in the case of mass media coverage of international conflicts, something that has to do clearly with geopolitics, there is much less diversity (that is to say, the main argumentative line is very similar) in a newspaper coming from one same pole, whether it is from the United States of America, United Kingdom, or Spain; as will be seen shortly. On the other hand, if a mass media outlet from United Kingdom was compared with another one from Brazil, a significant difference would be found.

Among many proponents of multipolarity, a key thinker is the often controversial Russian philosopher Alexander Dugin, who proposes a number of established and potential poles, building on the work of Samuel Huntington (1996).[10] Additionally, to put it in the terms of the heterodox economist Hyman Minsky on stabilising an unstable economy –only that it is the sphere of International Relations in this case–, this idea has found acceptance (albeit in slightly different terms) particularly among the representants of the Realist School of International Relations. As such, John Mearsheimer is a leading proponent of having a bipolar world (with the United States of America and China as the two hegemons) in order to secure stability.[11]

Good as theory may be, empirical grounding is still necessary to further support this hypothesis.

Thus, for the case of Russo-Ukrainian War, the following can be observed:

One year after the start of the war, an increasing number of countries are siding with Russia



Source: EIU.

As can be seen, in terms of global population, we can clearly identify at least four different poles in this particular scenario:

1. Canada, Australia, United Kingdom, United States of America, and European Union.
2. China (and, to a certain extent, the Asian continent).
3. Africa (a significant part of the African continent).
4. Russia.

In conclusion, both hypotheses seem to be confirmed on both theoretical and empirical ground, with basis on a series of antecedent –yet not fully exhaustive (the objective of the present piece of work is precisely to be as comprehensive as possible)– analysis.

4. Research Questions

Two main Research Questions are fully addressed, as per hypotheses presented in the previous section:

- How large a role does “psychic distance” play in the mass media perception (needless to say, “mass media” is not an entity as such; rather, what is meant is how journalists from a number of different countries perceive and cover the events) of an international conflict?

  (How is this reflected on both the intensity and the subsequently arisen sentiments? Are these feelings more or less intense depending on the newspaper and the chosen time interval? Are the aforementioned sentiments more negative or more neutral? In order to describe a military confrontation, are similar or different words –in terms of frequency– referenced?)

- To what extent does the academic notion of “multipolarity” reflect the current geopolitical *status quo*, in accordance with mass media coverage?

  (Are different poles –or spheres of influence– clearly identifiable, as per empirical data with the basis on which the present piece of research was elaborated?)

5. Research Design

Russo-Ukrainian War (with February the 24th 2022 as starting point, the beginning of the critical phase of the confrontation) is used as a "footprint". Due to the process being extraordinarily expensive in computational terms (this point will be addressed shortly), only this conflict is examined in the present body of work.

However, it needs be said that the technique followed for the analysis would be extremely similar for a comprehensive analysis of other military confrontations. Thus, the aim is to provide, so to speak, a full guide on how to properly inspect any conflict (as long as data is available) by means of R (though it can easily be extrapolated to other programming languages: for instance, Python).

Thus, a total of six newspapers are considered. In accordance with the two main hypotheses these are:

1. The Guardian, The New York Times, and El País (all of them pertaining to countries that are members of the NATO alliance: United Kingdom, United States of America, and Spain; respectively).
2. Folha de S. Paulo, Daily Maverick, and The Hindu (which belong to Brazil, South Africa, and India; this is to say, "neutral" countries).

All data is obtained by means of Wayback Machine (also known as WebArchive), and front pages are carefully examined.[12]

Research Design takes place in four phases:

1. A static version for the front page of each of the six aforementioned newspapers is generated for the 24th of February 2022. The extracted information is the headline, the weighting (adjusted in the range [0, 1] for further clarity), and the tag (by which a headline is identified in the HTML –HyperText Markup Language– script).
   In each of the cases, a customised vector of words for the Russo-Ukrainian War is used:

   The English vector of words includes a match between the words "war", "invasion", "conflict", "offensive", "occupation", "aggression", "battle", "assault", "campaign", "attack", "peace", "sanction" on the one hand; and "Putin", "Ukraine", "Zelenskiy", "Zelensky", "Zelenskyy", "Russia", "Biden", "Blinken", "Leyen", "Borrell", "Xi", "NATO" on the other hand. (Several transcriptions for the surname of the President of Ukraine are used, due to the fact that it is often spelled differently.)

   The Spanish vector of words includes a match between the words "guerra", "invasión", "conflicto", "ofensiva", "ocupación", "agresión", "batalla", "asalto", "operación", "ataque", "paz", "sanción", "sanciones" on the one hand; and "Putin", "Ucrania", "Zelenski", "Zelensky", "Zelenskiy", "Zelenskyy", "Rusia", "Biden", "Blinken", "Leyen", "Borrell", "Xi", "OTAN" on the other hand. (It should be noted that these are not necessarily direct translations from English; the criterion has been, rather, to choose those words that are most often associated with “war” in each language.)

Finally, the Portuguese vector of words includes a match between the words "guerra", "invasão", "conflito", "ofensiva", "ocupação", "agressão", "batalha", "assalto", "campanha", "ataque", "paz", "sanções" on the one hand; and "Putin", "Ucrânia", "Zelenski", "Zelenskiy", "Zelensky", "Zelenskyy", "Rússia", "Biden", "Blinken", "Leyen", "Borrell", "Xi", "OTAN" on the other hand.

(An example of R code can be consulted in Appendix A.)

2. A dynamic version of each of the newspapers is now created by means of a “loop” in R. The extracted information out of each front page is now the date, the headline, the weighting, and the tag. For this purpose, different time intervals were initially considered. However, as will be clearly seen in the following section, 60 days forward from the 24th of February 2022 seems to be a reasonably good choice to identify all the relevant patterns.
   A brief note on why this process is so computationally expensive seems appropriate now: as such, for each iteration in the loop, two requests are made to WebArchive (one to randomly choose a “screenshot” of the front page –on occasions, there can be as many as nearly 100 “screenshots” per day–, and another one to identify the headline with the highest weighting). The main difficulty that arises here is due to the fact that for commonly accepted in HTML scripts “h1”, “h2”, and “h3” tags there is no pre-established size. For example, “h3” can be of one size in one section of the front page, and of another completely different size in another section. Thus, the only way to address this challenge is to iterate through all of the headlines to know the exact weighting of each one (as per size of the front page), and to extract the most preponderant one in terms of weighting. This is why precisely the process is so expensive on a computational level.
   Despite this significant challenge, there exists a solution for “debugging” server overload issues. Quite simply, this line needs to be introduced in Windows Command Prompt (patience is needed, but solution is usually guaranteed):
   curl --ssl https://web.archive.org/

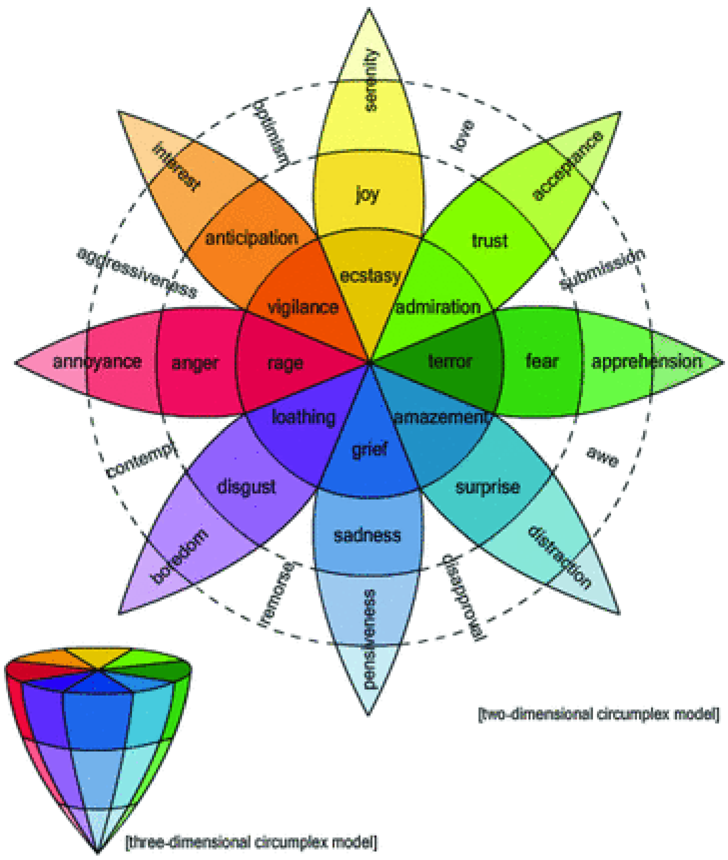
   (An example of R code can be viewed in Appendix B.
   Note: a full “loop” is provided; however, depending on WebArchive server internal demands and computational prowess of the computer, the code may not run in full. Subsequently, an update –of date and the number of iterations– may be required on behalf of the person undertaking this computationally expensive endeavour.)

3. Time series are elaborated. Again, this is a task that requires certain degree of expertise. In order to avoid having an abrupt piece of time series, Holt’s Linear Exponential Smoothing method is used to extract the trend, for clarity purposes. In addition to this, some time series are quite peculiar:

For instance, in the case of El País and The Hindu newspapers, there occasionally appear no headlines after the 29th and 36th day of conflict, respectively (due to two possible circumstances: either that the event commences to lose its importance –the so-called "war fatigue" in this case– or because there is no match between the two parts of the vector of words). Thus, LOWESS (Locally Weighted Scatterplot Smoothing –a type of Local Regression–) is used to handle missing values.
As for Folha de S. Paulo, general trend begins to present non-trivial oscillations after the 32nd day of the conflict. 5-Day Moving Average approach is consequently implemented in order to still represent these oscillations, but not quite as drastically.

(Full R code is provided in Appendix C.)

4. Finally, Sentiment and Frequency analysis are carried out.
As for sentiment analysis, the chosen method is "nrc" (as developed by Saif M. Mohammad and Peter Turney in 2011). This is mainly due to two facts.
The first one is that is available in more than forty languages (including English, Spanish, and Portuguese).
The second, and perhaps the most important fact, is that it is a method based on Plutchik's wheel of emotions:[13]



Source: (Mohammad and Turney, 2012)

Robert Plutchik, an American psychologist, identified eight key emotions: “anger”, “fear”, “sadness”, “disgust”, “surprise”, “anticipation”, “trust”, and “joy”. (As for other sentiment analysis methods, such as AFINN or Bing et al., they are significantly more limited with regard to both of the aforementioned aspects.)
Generally, positive emotions are “joy” and “trust”. “Sadness”, “anger”, “fear”, “disgust” are negative emotions. Finally, “anticipation” and “surprise” are normally considered neutral.
However, in the context of an international conflict, a choice to consider “anticipation” and “surprise” as negative emotions is made.
Additionally, “nrc” method includes two sentiments: positive and negative.

The main challenge encountered while carrying out sentiment analysis is that, without further refinement, aiming to extract the main ten words, one can perfectly obtain –for instance– “Russia” and “Russian”, or “Ukraine” and “Ukraine’s” as four out of ten main words.

Consequently, “tokenization” and “lemmatization” (both are Neuro-Linguistic Programming approaches) are required:
“Tokenization” is the process of segmenting running text into individual terms or tokens, typically words, phrases, or sentences.
“Lemmatization” refers to the reduction of inflected or derived words to their base or root form, such as converting "running", "runs", and "ran" to "run".

For the present body of work, a particularly sophisticated R library is made use of: “udpipe” (which is, fortunately, enough, available in more than 65 languages). It allows to categorise words into nouns, verbs, adjectives, and / or adverbs; thus, precision increases greatly.

Besides, many headlines include “indicators” such as “Live Russia-Ukrainian war:” before the main part of the headline. In order to avoid bias in both sentiment and frequency analysis, these indicators need to be removed.

Last but not least, a sentiment score in the range of [-1, 1] (as per “nrc” convention) is obtained for each mass media outlet.

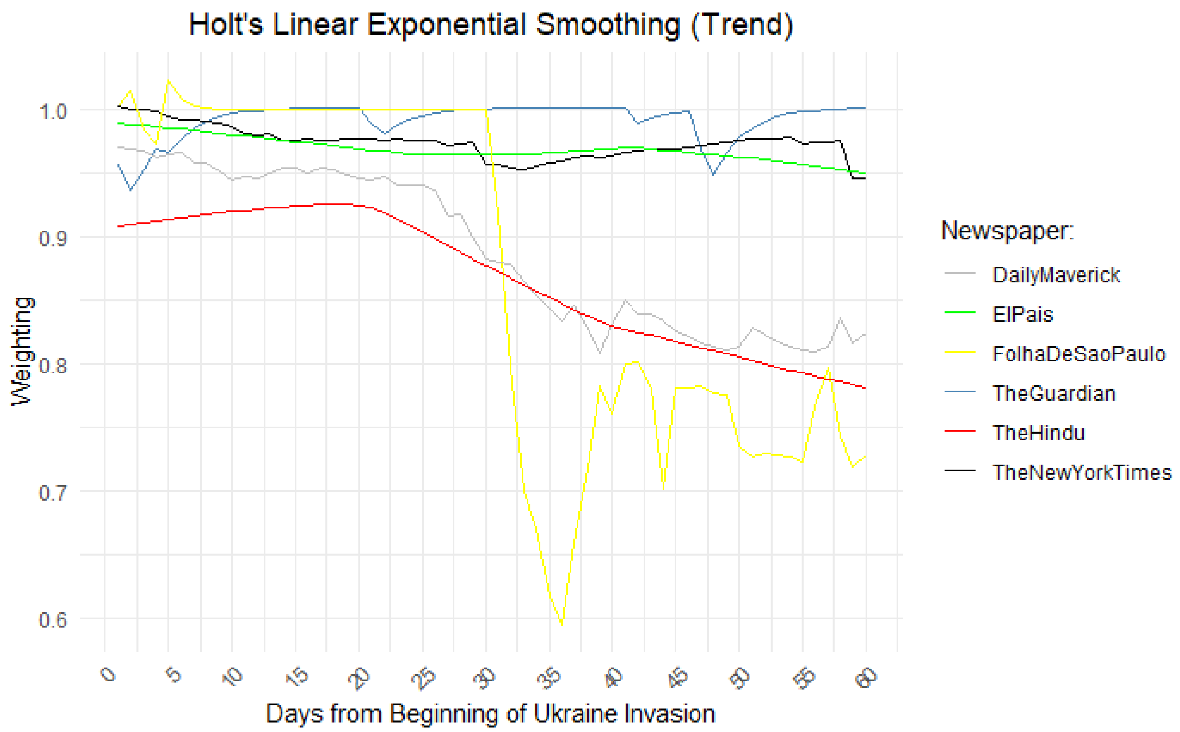
(An example of R code is provided in Appendix D.)

6. Results

While both the generation of a static and –particularly– a dynamic version of each newspaper are two indispensable preliminary steps (without them, there would be no data), both quantitative and qualitative conclusions are obtained from time series analysis, as well as from sentiment and frequency analysis.

To this, attention shall be turned towards now.

6.1. Time Series Analysis



As can be clearly seen, all six mass media outlets strongly capture the decisive moment of the Russian aggression against Ukraine, which took place on the 24th of February of 2022.

However, while there is a persisting trend –which translates in almost a constant throughout the sixty days covered in the present body of work– in the case of all three newspapers that belong to NATO countries (The Guardian, The New York Times, and El País), which never decreases below 0.95; the story is quite different for the other three media outlets.

As such, there is a steady decrease in the intensity of reporting as for Folha de S. Paulo, The Hindu, and Daily Maverick. The only slightly dissimilar behaviour takes place in the case of Folha de S. Paulo, which occasionally captures attention-grabbing headlines (reflecting significant events, such as –for instance–, Bucha Massacre, which was an atrocity committed by the Russian troops in a Ukrainian village named Bucha in March of 2022).

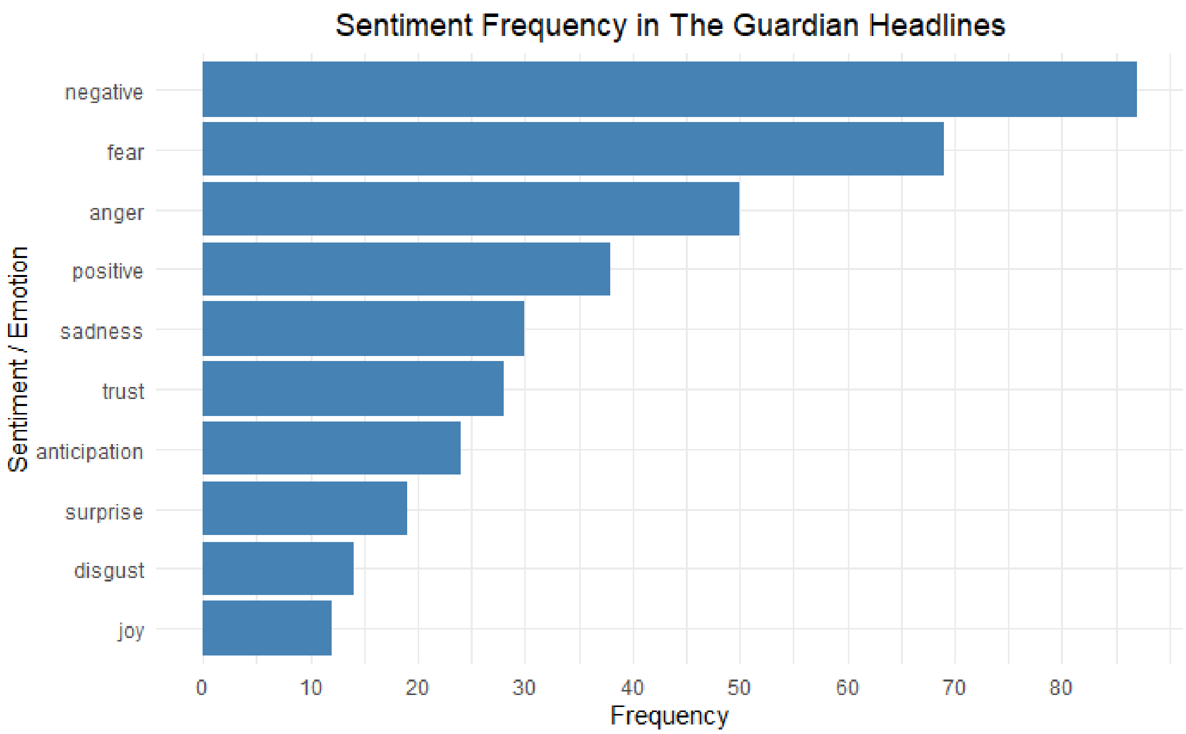
In conclusion, the obtained data clearly confirms both initial hypotheses ("psychic distance", due to the involvement of NATO countries in the conflict; and "multipolarity", consisting in the separation of the world into different spheres of influence or poles).

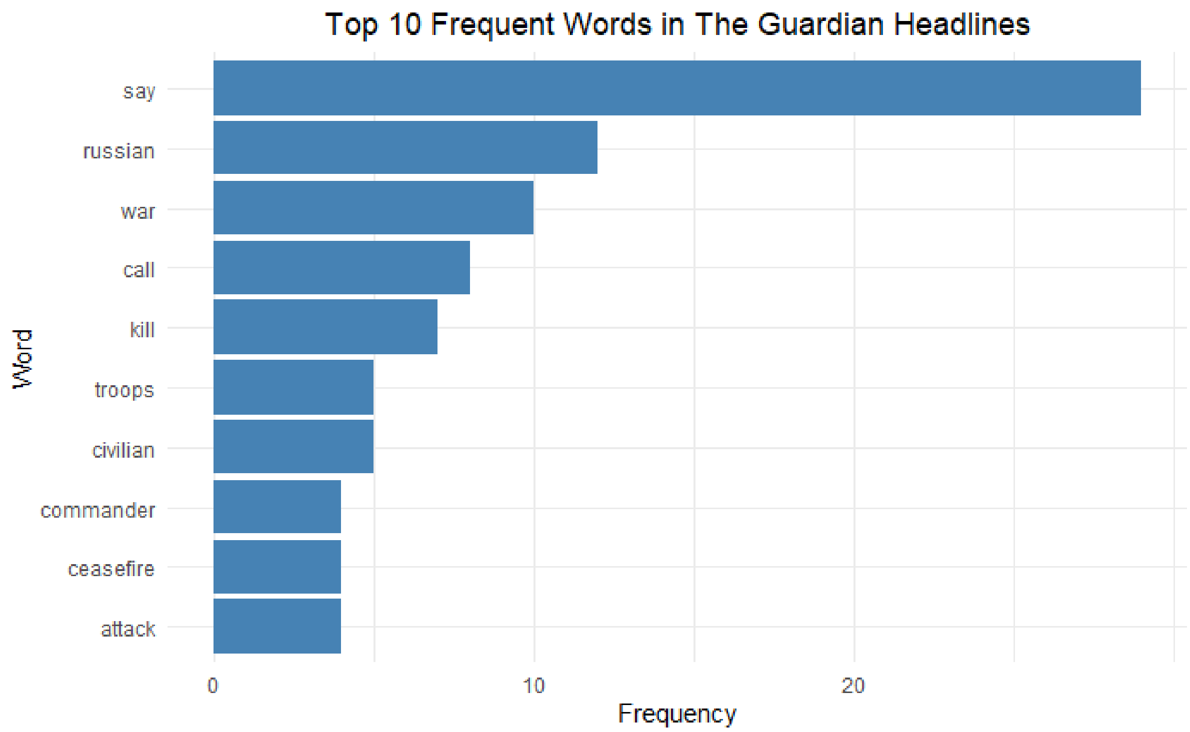
6.2. Sentiment and Frequency Analysis

In what follows, results will be presented for each newspaper. After this, further general observations will be made.

6.2.1. The Guardian

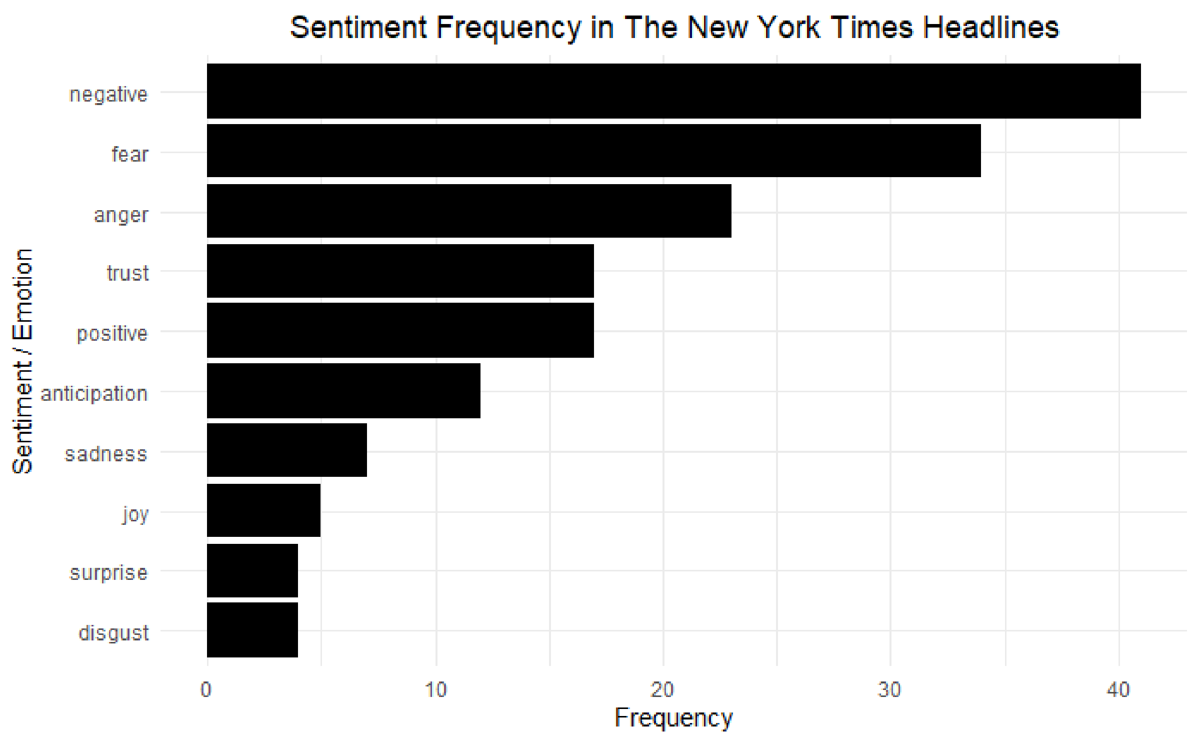
Sentiment score is of -0.67 (strongly negative, as in the range [-1, 1]).

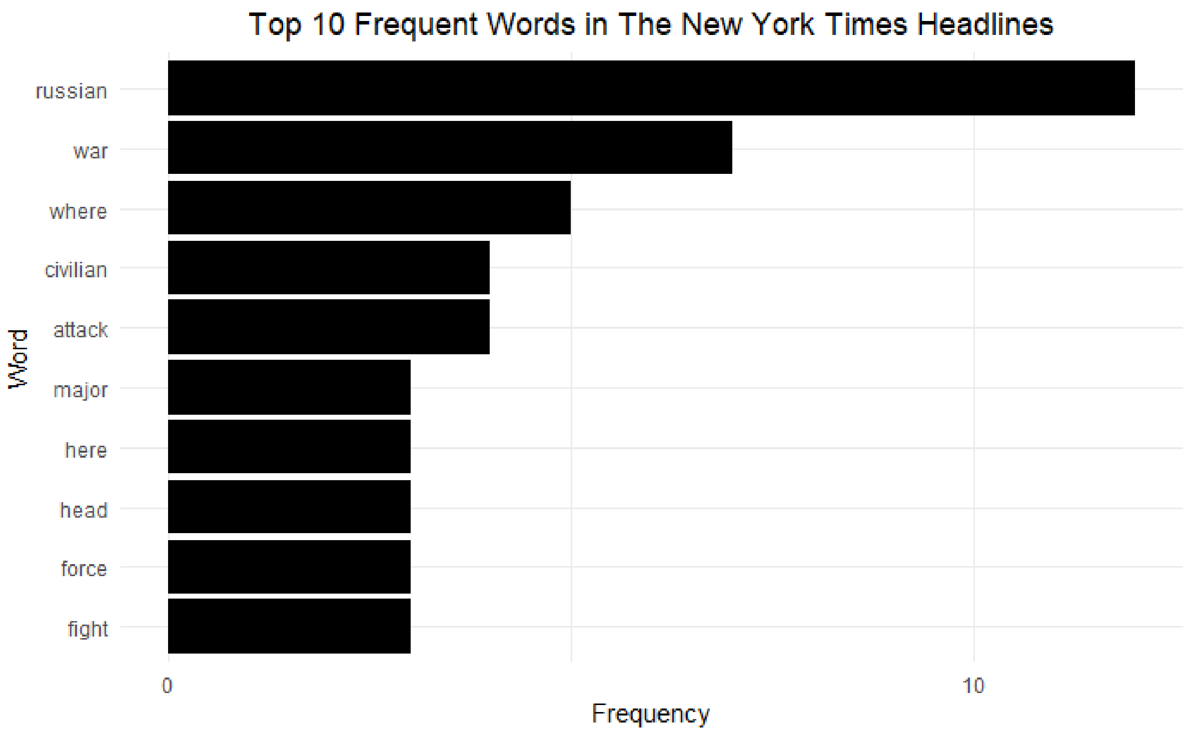




6.2.2. The New York Times

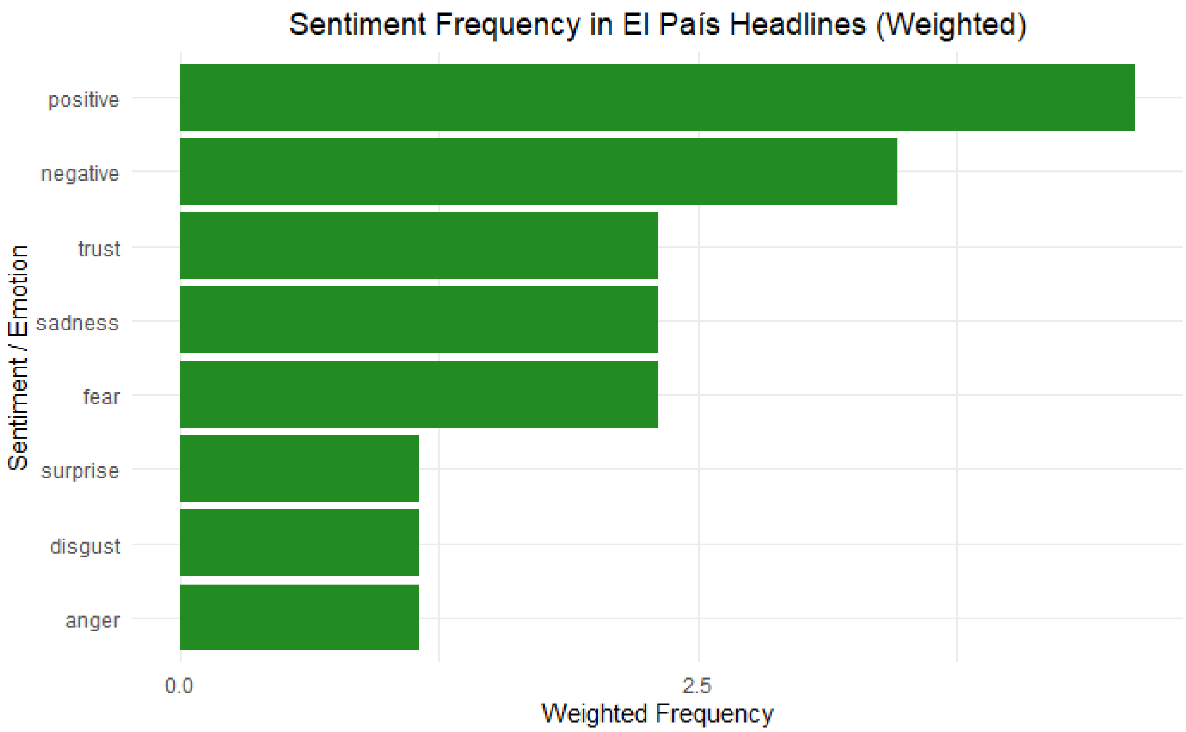
Sentiment score is of -0.58.

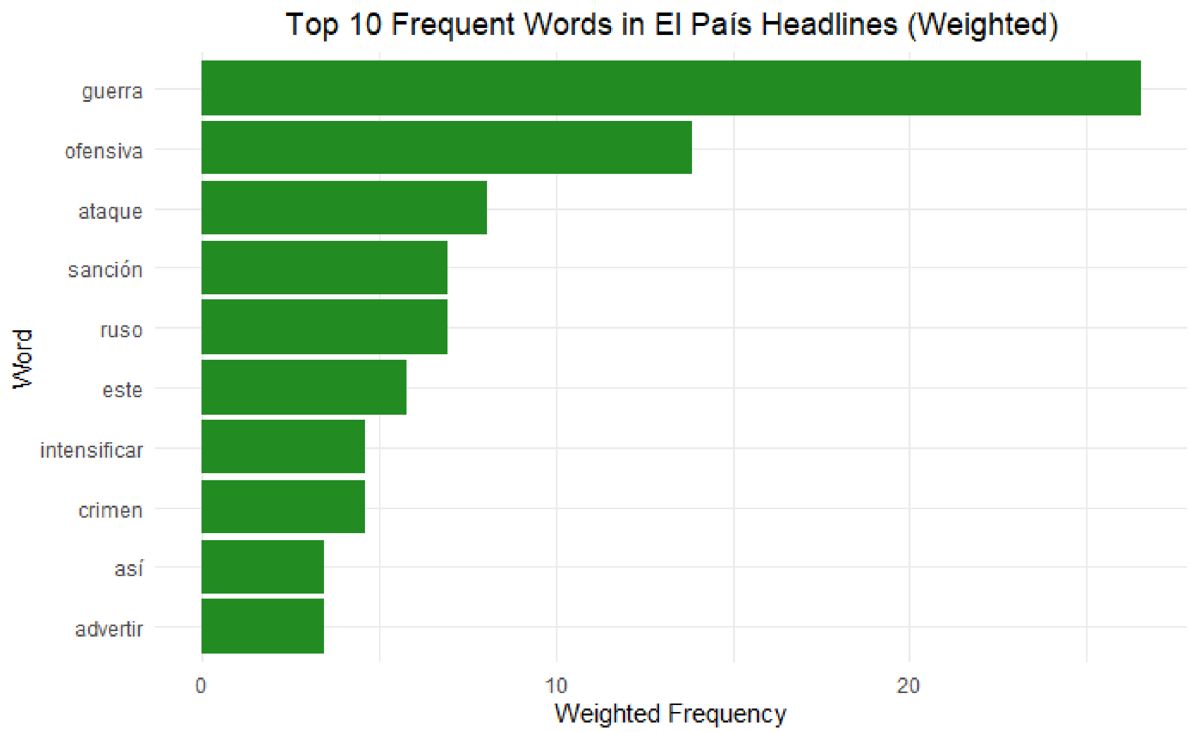




6.2.3. El País

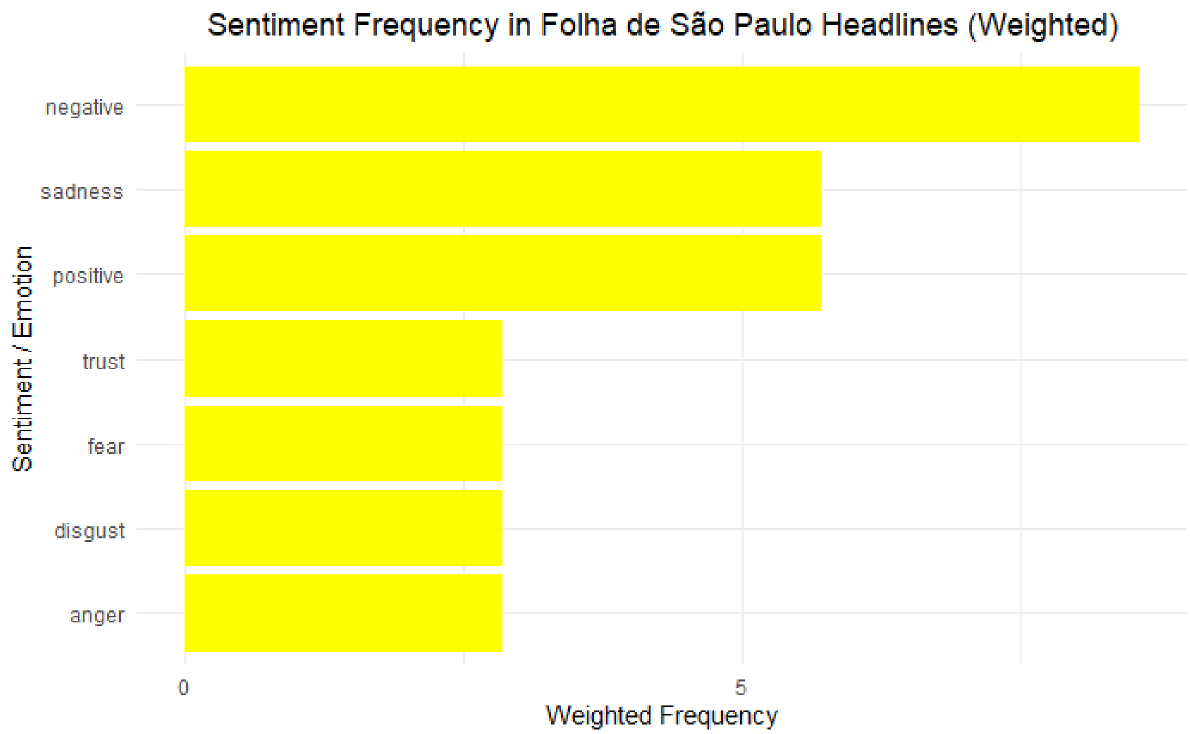
Sentiment score is of -0.56.

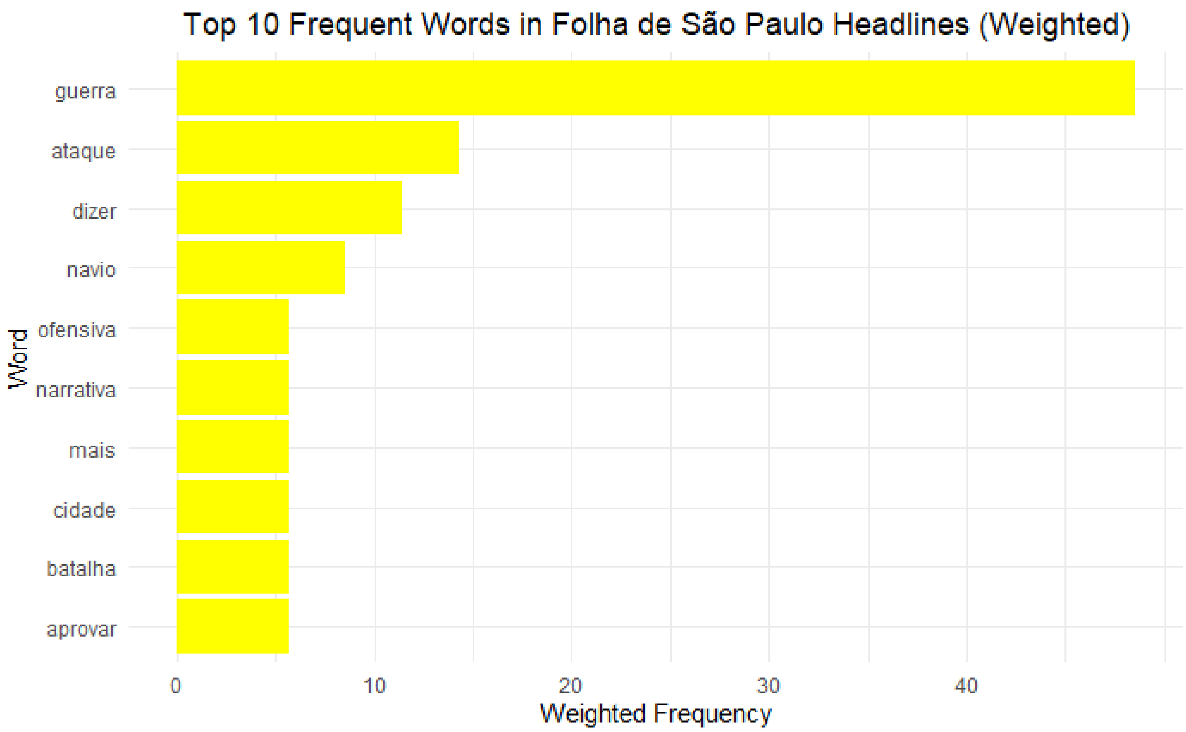




6.2.4. Folha de S. Paulo

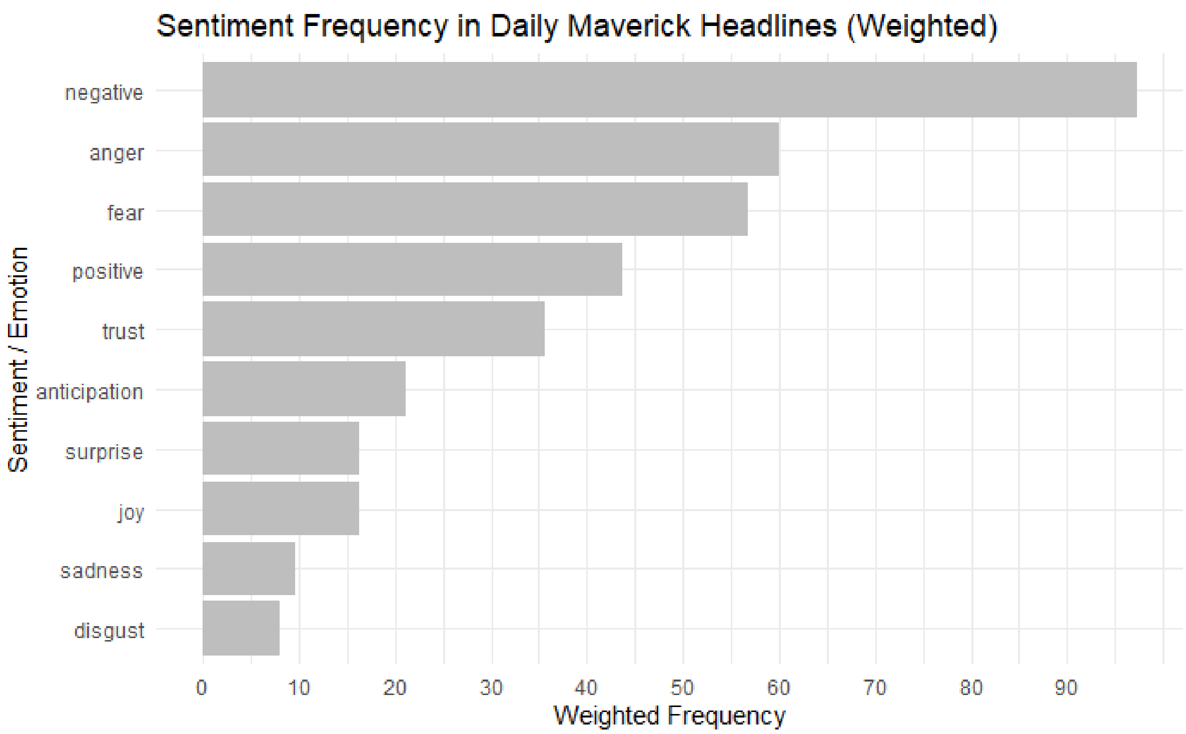
Sentiment score is of -0.67.

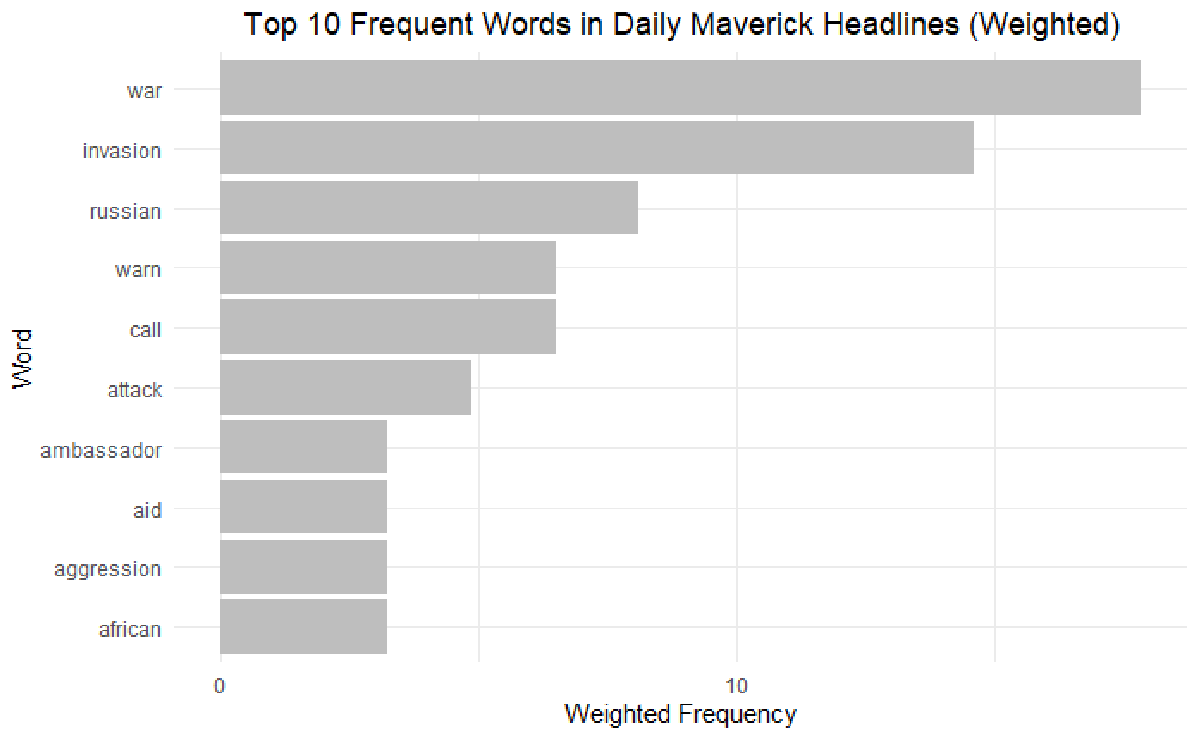




6.2.5. Daily Maverick

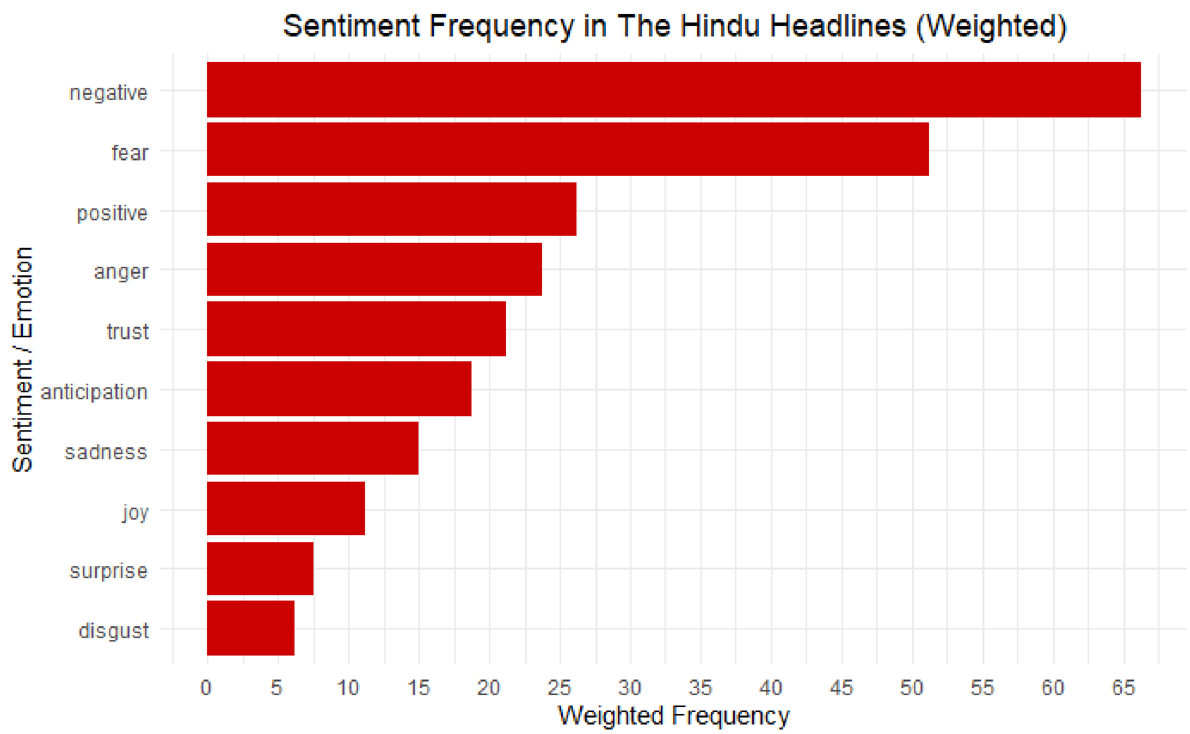
Sentiment score is of -0.54.

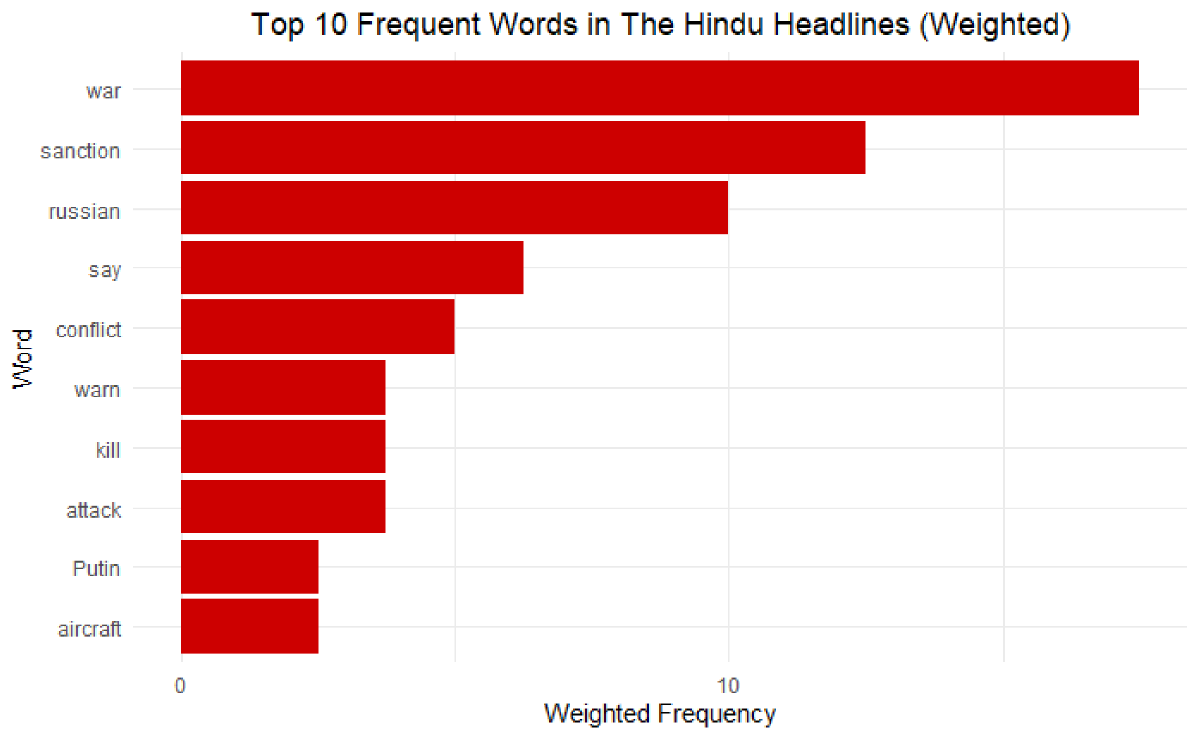




6.2.6. The Hindu

Sentiment score is of -0.58.





All in all, it can be viewed that, despite the remarkable differences identified in time series analysis, results are –nevertheless– very similar in terms of sentiment score.

In the case of frequency analysis, no matter how much or how quickly the intensity of the "signal" decays, all values are in the range of [-0.67, -0.54], while standard deviation is of only 0.05.

This leads to the natural conclusion that, despite linguistic and cross-border differences, Russo-Ukrainian war has been covered in a remarkably similar style across the six mass media outlets that have been analysed.

Equally, negative sentiment and emotions of fear and anger are the most preponderant ones.

In fact, the only outlier seems to be the Spanish newspaper El País, which presents headlines such as the following (this point will be covered more broadly in the next section):

"Rusia echará mano de las reservas en yuanes y oro para tratar de burlar las sanciones de Occidente." (01-03-2022).

("Russia will tap into its yuan and gold reserves in an attempt to evade Western sanctions.")

Clearly, "echará una mano" is perceived as something positive, even though it is objectively not in this context.

"Rusia estrecha el cerco en torno a Kiev e intensifica los ataques." (12-03-2022).

("Russia tightens the siege around Kiev and intensifies the attacks.")

Again, “estrecha” and “intensifica” may well be identified as a positive sentiment.

“Putin reaparece ante decenas de miles de rusos para defender su Guerra.” (18-03-2022).

("Putin reappears before tens of thousands of Russians to defend his war.")

Words “reaparece” and “defender” are linked to a positive sentiment.

All of this leads to a natural bias in the case of El País.

This finding, which proceeds out of handling several languages simultaneously, is indeed corroborated by Mohammad (2020) in his research paper on *Practical and Ethical Considerations in the Effective use of Emotion and Sentiment Lexicons*. As such, the author states the following: “there can be cultural differences in the emotion associations of a concept”.[14]

Finally, it should be noted that the frequency exhibited on “x-axis” scale varies.

This is due to the fact that some samples are of unequal size (even though a “weight factor” has been applied in order to compensate for this slight disadvantage).

7. Discussion

Overall, there is a reason to believe that each and every point has been well addressed in the previous two sections.

But, as always, there is room for improvement and further analysis. As such:

1. As it has previously been pointed out on a number of occasions, this is an extraordinarily expensive on a computational level piece of work. Consequently, for further research in this and / or similar areas, it would be highly advisable to make use of cutting-edge equipment in terms of hardware (ideally, a "workstation" PC or laptop), as well as considering the option of using a "cloud" platform.

2. While a manual sentiment analysis technique such as "nrc" certainly yields some quite good results, a more refined –Machine Learning (or, even, Deep Learning)– sentiment analysis approach could potentially be implemented. While being more computationally expensive and time-consuming, it presents the advantage of capturing a wider lexicon and identifying patterns in a more accurate manner – thus avoiding cases such as the ones identified for El País newspaper. For a further illustration of this point, it is enough to consider the following graph:

Lexicon	Number of words	Number of positive words	Number of negative words	Resolution	Calculation method to obtain score per year	Classification Method
AFINN	2,477	878	1,598	11	score individual words and sum	Manual
Bing	6,789	2,006	4,783	2	(number of positive words - number of negative words)/ total words	Manual
Loughran	3,917	354	2,355	2	(number of positive words - number of negative words)/ total words	Manual
NRC	5,555	2,312	3,324	2	(number of positive words - number of negative words)/ total words	Amazon Mechanical Turk
SenticNet	23,626	11,774	11,852	continuous	score individual words and sum	Machine Learning
Sentiword	20,093	8,898	11,029	continuous	score individual words and sum	Machine Learning
Syuzhet	10,748	3,587	7,161	16	score individual words and sum	Manual
SOCAL	5,971	2,438	3,530	continuous	score individual words and sum	Amazon Mechanical Turk

Source: Sonkin (2021)

As can be seen, the two main Machine Learning techniques for sentiment analysis –SenticNet and Sentiword– allow for a lexicon of 23.626 words and 20.093 words (as opposed to "nrc", which includes only 5.555 words).

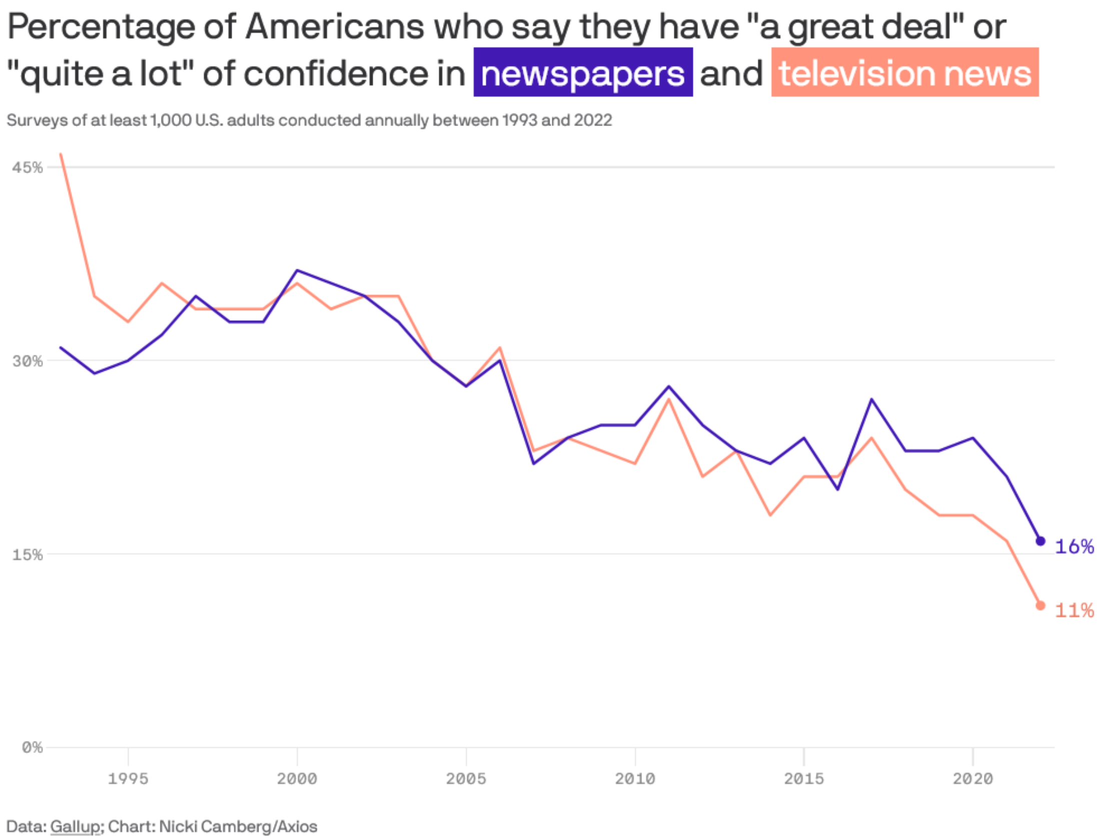
While 5.555 words suit the requirements of the present body of work well enough, the option of implementing a Machine Learning approach should not be neglected for more linguistically demanding academic studies. (This, not to mention again the fact that a Machine Learning technique is potentially better when it comes to identifying relevant patterns.)

3. It needs to be said that, while both "internal validity" and "external validity" causal inference considerations do generally hold, the researcher may be –on occasions– faced with some caveats. In the present body of work, one such inconvenience has been the fact that Folha de Sao Paulo presented only twenty-one (out of a total of sixty) non-repeated headlines. Even though that a weight factor was applied, this is still not a sufficiently statistically robust sample ($n<30$).

4. An attempt to analyse a German newspaper (Süddeutsche Zeitung) –instead of El País– was carried out. However, the present author had to desist shortly due to two reasons. The first one was that Germany is a country the economy of which has long been heavily reliant on Russian natural resources – consequently, many headlines include words such as "energy", "oil", "gas", "Nord-Stream", etc. The second one has to do with "psychic distance" – as a result, much more varied updates on the events on the Russo-Ukrainian War are provided (issues such as anti-war protests in Russia, critique against German government for not having done enough after the Crimea Crisis, assimilation of Russian political refugees in Germany, etc. were all broadly echoed). Needless to say, to broaden the German vector of words to include such diverse coverage would have violated the principle of "internal validity".

5. Had there been more time –and, particularly, if the present author could have had a more high-end technical equipment at his disposal–, it would have been very interesting to evaluate –in the mode of a "robustness check"– another recent military conflict in which "psychic distance" would not have been a factor. A good example could have been the Yemen Civil War (2014 – ongoing). In this scenario, there has been a very limited influx of refugees to Europe, and certainly there is very little cultural proximity to some of the previously identified –in the case of Russo-Ukrainian War– geopolitical poles.

6. Finally, the interested reader must be aware of the fact that trust in media in many countries (both television and newspapers) stands at a historical low:



A reasonable assumption could be made that one of the reasons behind this phenomenon is an excessively negative mass media coverage of the events, which has traditionally been a strategy to grab the reader's attention. Therefore, an interesting point for further analysis would be to choose a series of newspapers and inspect the evolution of their respective sentiment scores (as per headlines) throughout a number of years. Needless to say, this would be an enormously time- and resource- consuming process that ought to merit a separate work of its own (though computational techniques outlined in the present body of work can be perfectly applied).

7. Finally, it has been mentioned that we live in an era of "polycrisis". Another interesting task would be to inspect –by following the same computational patterns presented in this document– the impact of, say, climate change or a pandemic.

8. Conclusion

The present author holds a firm conviction in having fully demonstrated his remarkably good skills as a Computational Social Scientist, particularly in terms of multidisciplinarity, a notion that is indispensable for the emerging field of Computational Social Science.

The exhibited body of work cannot by any means considered as a mere piece of Data Science.

(This statement is under no circumstances made to diminish the much relevant field of Data Science; rather, the communicative intention here is to transmit that there is a series of objective differences between the two fields –that of Computational Social Science and that of Data Science–.)

Quite on the contrary, knowledge of computer programming, advanced statistics, and sociology –the three pillars of Computational Social Science– has been shown at all times.

In addition, many insights from other spheres of knowledge (international relations, economics and finance, history, politics and geopolitics, etc.) have been drawn on multiple occasions throughout the process of elaboration of this Master Thesis.

And, perhaps most importantly, the present piece of work has been initially conceived –and subsequently elaborated– as an exercise in open-mindedness as a key element for a good academic researcher (many challenges and some initially unexpected empirical results have been addressed and properly documented), with the aim to push further both personal and academic frontiers. This mission would have certainly not been able to be accomplished in full, had it not been for a number of individuals who are cited in acknowledgments, and to whom I feel both personally and academically indebted.

Furthermore, being aware of some of the limitations of Computational Social Science –a normal *status quo* for any emerging field–, four main considerations for an enhanced vision of Computational Social Science have been taken into account, as proposed by Petter Törnberg and Justus Uitermark in 2021: critical realism, methodological pluralism, interpretation, and explanation.[15]

Finally, while this body of work constitutes a comprehensive framework on how to proceed with the analysis of mass media coverage of international conflicts, the –hereby– presented structure allows any potential researcher to expand the lines of enquiry in a significant number of different ways –all of which have been previously discussed–.

After all, doing research means nothing if knowledge is not shared with the others.

For this reason and in the spirit of complete academic openness and transparency, R code (which can be easily adapted to other programming languages –for instance, Python–) is made available both in the present body of work (which can be consulted in the respective Appendixes), as well as uploaded onto a GitHub repository in full, along with Excel archives:

https://github.com/ArtemUrlapov/CSSMasterThesis/

References

[1] Jensen, B. (2022). *How Does It End? What Past Wars Tell Us about How to Save Ukraine.* Commentary for Center for Strategic & International Studies.

[2] Mercier, A. (2005). *War and media: Constancy and convulsion.* International Review of the Red Cross, 87(860), 649-659.

[3] Eytan G. (2007). *Media and International Conflict: A Multidisciplinary Approach.* Journal of Dispute Resolution.

[4] Reuben, R. C. (2009). *The Impact of News Coverage on Conflict: Toward Greater Understanding*. 93 Marq. L. Rev. 45.

[5] Joseph, T. (2014). *Mediating War and Peace: Mass Media and International Conflict.* India Quarterly, 70(3), 225–240.

[6] Bullough, E. (1912). *Psychical Distance as a factor in Art and an Aesthetic Principle*. British Journal of Psychology: 5, pp. 87-118.

[7] Gairola, D., and Chong, S. Y. (2012) *An Economics-Inspired Noise Model in Spatial Games with Reputation.* Advances in Intelligent Modelling and Simulation: Artificial Intelligence-based Models and Techniques in Scalable Computing, Studies in Computational Intelligence, Springer Berlin/Heidelberg, vol. 422, pp. 271-293.

[8] Whiting, K., and Park, H. (2023). *This is why 'polycrisis' is a useful way of looking at the world right now.* World Economic Forum (Blog Entry).

[9] The UN Refugee Agency. *Refugee Data Finder.* Available at: https://www.unhcr.org/refugee-statistics/download/ .

[10] Huntington, S. P. (1996). *The Clash of Civilizations and the Remaking of World Order.* Simon & Schuster.

[11] Mearsheimer, J. (2001). *The Tragedy of Great Power Politics.* W.W. Norton.

[12] Wayback Machine – Internet Archive. Available at: https://web.archive.org/ .

[13] Plutchik, R. (1982). *A psychoevolutionary theory of emotions*. Social Science Information, 21(4–5), 529–553.

[14] Mohammad, S.M. (2020). Practical and Ethical Considerations in the Effective use of Emotion and Sentiment Lexicons. arXiv:2011.03492 .

[15] Törnberg, P. and Uitermark, J. (2021). *For a Heterodox Computational Social Science.* Big Data & Society.

Appendix A.

~~~~
```{r}

library(rvest)
library(stringr)
library(V8)
library(dplyr)

EP <- "https://web.archive.org/web/20220224094037/elpais.com/"
EPWebsite <- read_html(EP)

SpanishVectorOfWords <- c("guerra", "invasión", "conflicto", "ofensiva",
                          "ocupación", "agresión", "batalla", "asalto",
                          "operación", "ataque", "paz", "sanción",
"sanciones",
                          "Putin", "Ucrania", "Zelenski", "Zelensky",
"Zelenskiy",
                          "Zelenskyy", "Rusia", "Biden", "Blinken",
"Leyen", "Borrell",
                          "Xi", "OTAN")

EPHeadlinesExtraction <- function(tag) {
  EPWebsite %>% html_nodes(xpath = paste0("//", tag)) %>%
    html_text(trim = TRUE)
}

EPHeadlinesH1 <- EPHeadlinesExtraction("h1")
EPHeadlinesH2 <- EPHeadlinesExtraction("h2")
EPHeadlinesH3 <- EPHeadlinesExtraction("h3")

EPHeadlinesCombined <- c(EPHeadlinesH1,
                         EPHeadlinesH2,
                         EPHeadlinesH3)

N <- length(EPHeadlinesCombined)
EPTagWeightings <- (N:1)/N

EPMatchingConditions <- vapply(EPHeadlinesCombined, function(x) {
    any(str_detect(x, regex(paste(SpanishVectorOfWords[1:12],
collapse="|"),
                            ignore_case = TRUE))) &&
    any(str_detect(x,
regex(paste(SpanishVectorOfWords[13:length(SpanishVectorOfWords)],
collapse="|"),
                            ignore_case = TRUE)))
}, logical(1))

EPTags <- c(rep("h1", length(EPHeadlinesH1)),
            rep("h2", length(EPHeadlinesH2)),
            rep("h3", length(EPHeadlinesH3)))

EPStaticDataFrame <- data.frame(EPHeadline = NA,
                                EPWeighting = 0,
                                EPTag = NA)

if (any(EPMatchingConditions)) {
~~~~

```
  EPStaticDataFrame$EPHeadline <-
EPHeadlinesCombined[EPMatchingConditions][1]
  EPStaticDataFrame$EPWeighting <-
EPTagWeightings[EPMatchingConditions][1]
  EPStaticDataFrame$EPTag <- EPTags[EPMatchingConditions][1]
}

print(EPStaticDataFrame)
```

EPHeadline	EPWeighting	EPTag
El ataque en mapas: las tropas rusas cruzan la frontera de Ucrania	0,98013245	h2

Appendix B.

```
```{r}

library(rvest)
library(stringr)
library(httr)
library(jsonlite)
library(lubridate)
library(dplyr)

StartDate <- as.Date("2022-02-24")

DailyElPaisHeadline <- function(domain, date) {

  Sys.sleep(sample(seq(22.5, 35, by = 0.1), 1))

  RandomPauseCap <- sample(seq(25, 55, by = 0.1), 1)

  WebArchiveURL <- "https://web.archive.org/cdx/search/cdx?url="

  OutputURL <- RETRY("GET", WebArchiveURL, timeout(sample(seq(120, 180, by
= 0.1), 1)),
                query = list(url = domain, timestamp = format(date,
"%Y%m%d"),
                             output = "json"), times = 90, pause_cap =
RandomPauseCap)

  OutputList <- fromJSON(content(OutputURL, "text", encoding = "UTF-8"))

  ElPaisHeadlines <- as.data.frame(OutputList[-1, ], stringsAsFactors =
FALSE)

  colnames(ElPaisHeadlines) <- OutputList[1, ]

  DesiredDate <- format(date, "%Y%m%d")

  FilteredElPaisHeadlines <- ElPaisHeadlines[grepl(DesiredDate,
ElPaisHeadlines$timestamp,
                                                   fixed = TRUE), ]

  RandomElPaisHeadline <-
FilteredElPaisHeadlines[sample(1:nrow(FilteredElPaisHeadlines),

1), ]

  Timestamp <- RandomElPaisHeadline$timestamp

  OriginalURL <- RandomElPaisHeadline$urlkey

  RegularElPaisURL <- gsub("com,elpais)/", "elpais.com/", OriginalURL,
fixed = TRUE)

  RetrievedElPaisURL <- paste0("https://web.archive.org/web/", Timestamp,
"/", RegularElPaisURL)

  cat("Retrieved from El Pais:", RetrievedElPaisURL, "\n")
```
```

```
  return(RetrievedElPaisURL)
}

SpanishVectorOfWords <- c("guerra", "invasión", "conflicto", "ofensiva",
                          "ocupación", "agresión", "batalla", "asalto",
                          "operación", "ataque", "paz", "sanción",
"sanciones",
                          "Putin", "Ucrania", "Zelenski", "Zelensky",
"Zelenskiy",
                          "Zelenskyy", "Rusia", "Biden", "Blinken",
"Leyen", "Borrell",
                          "Xi", "OTAN")

ElPaisDynamicDataFrame <- data.frame(ElPaisDate = character(0),
                                     ElPaisHeadline = character(0),
                                     ElPaisWeighting = numeric(0),
                                     ElPaisTag = character(0))

for (i in 0:59) {

  Sys.sleep(sample(seq(35, 60, by = 0.1), 1))

  CurrentDate <- StartDate + days(i)
  ElPaisTargetURL <- DailyElPaisHeadline("www.elpais.com/", CurrentDate)

  ElPaisWebsite <- read_html(ElPaisTargetURL)

  ElPaisMaxWeighting <- 0
  ElPaisMaxWeightingRow <- NULL

  for (ElPaisTag in c("h1", "h2", "h3")) {
    ElPaisHeadlines <- ElPaisWebsite %>%
      html_nodes(ElPaisTag) %>%
      html_text(trim = TRUE)

    for (ElPaisHeadline in ElPaisHeadlines) {
      SpanishVectorOfWords1 <- any(str_detect(ElPaisHeadline,
regex(paste(SpanishVectorOfWords[1:13], collapse = "|"), ignore_case =
TRUE)))
      SpanishVectorOfWords2 <- any(str_detect(ElPaisHeadline,
regex(paste(SpanishVectorOfWords[14:length(SpanishVectorOfWords)],
                                        collapse = "|"), ignore_case =
TRUE)))

      if (SpanishVectorOfWords1 && SpanishVectorOfWords2) {
        ElPaisRawWeighting <-
          length(ElPaisHeadlines) - match(ElPaisHeadline, ElPaisHeadlines)
+ 1
        ElPaisNormalisedWeighting <- ElPaisRawWeighting /
length(ElPaisHeadlines)

        if (ElPaisNormalisedWeighting > ElPaisMaxWeighting) {
          ElPaisMaxWeighting <- ElPaisNormalisedWeighting
          ElPaisMaxWeightingRow <- data.frame(ElPaisDate =
as.character(CurrentDate),
                                              ElPaisHeadline =
ElPaisHeadline,
                                              ElPaisWeighting
=ElPaisNormalisedWeighting,
                                              ElPaisTag = ElPaisTag)
```

```
        }
      }
    }
  }

  if (!is.null(ElPaisMaxWeightingRow)) {
    ElPaisDynamicDataFrame <- rbind(ElPaisDynamicDataFrame,
ElPaisMaxWeightingRow)
  } else {
    ElPaisNARow <- data.frame(ElPaisDate = as.character(CurrentDate),
                              ElPaisHeadline = NA,
                              ElPaisWeighting = 0,
                              ElPaisTag = NA)
    ElPaisDynamicDataFrame <- rbind(ElPaisDynamicDataFrame, ElPaisNARow)
  }
}

head(ElPaisDynamicDataFrame)

```

ElPaisDate	ElPaisHeadline	ElPaisWeighting	ElPaisTag
2022-02-24	Biden anuncia más sanciones y acusa a Putin de querer “restablecer la Unión Soviética”	0,974193548	h2
2022-02-25	El ataque en mapas: las tropas rusas cruzan la frontera de Ucrania	0,984375	h2
2022-02-26	Ucrania ofrece una resistencia feroz a la ofensiva rusa	1	h2
2022-02-27	Las capas de la ofensiva informativa de Putin	0,909090909	h2
2022-02-28	Rusia y Ucrania negocian mientras el Kremlin intensifica la ofensiva	1	h2

Appendix C.

```
```{r}

library(forecast)
library(ggplot2)
library(zoo)

FolhaDeSaoPauloWindowSize <- 5
FolhaDeSaoPauloUkraine60Days$FolhaDeSaoPauloWeighting <-
zoo::rollmean(FolhaDeSaoPauloUkraine60Days$FolhaDeSaoPauloWeighting,
              k = FolhaDeSaoPauloWindowSize, align = "center", fill = NA)

set.seed(123)
FolhaDeSaoPauloRandomRows <- sample(1:nrow(FolhaDeSaoPauloUkraine60Days),
4)
FolhaDeSaoPauloUkraine60Days$FolhaDeSaoPauloWeighting[FolhaDeSaoPauloRandom
Rows] <- NA
FolhaDeSaoPauloUkraine60Days$FolhaDeSaoPauloWeighting <-
zoo::na.approx(FolhaDeSaoPauloUkraine60Days$FolhaDeSaoPauloWeighting,
rule=2)

TheHinduLowess <- lowess(1:length(TheHinduUkraine60Days$TheHinduWeighting),
TheHinduUkraine60Days$TheHinduWeighting)
TheHinduUkraine60Days$TheHinduWeighting <- TheHinduLowess$y

ElPaisLowess <- lowess(1:length(ElPaisUkraine60Days$ElPaisWeighting),
ElPaisUkraine60Days$ElPaisWeighting)
ElPaisUkraine60Days$ElPaisWeighting <- ElPaisLowess$y

TheGuardianUkraineHolt <-
holt(TheGuardianUkraine60Days$TheGuardianWeighting)
TheNewYorkTimesUkraineHolt <-
holt(TheNewYorkTimesUkraine60Days$TheNewYorkTimesWeighting)
ElPaisUkraineHolt <- holt(ElPaisUkraine60Days$ElPaisWeighting)
FolhaDeSaoPauloUkraineHolt <-
holt(FolhaDeSaoPauloUkraine60Days$FolhaDeSaoPauloWeighting)
DailyMaverickUkraineHolt <-
holt(DailyMaverickUkraine60Days$DailyMaverickWeighting)
TheHinduUkraineHolt <- holt(TheHinduUkraine60Days$TheHinduWeighting)

TheGuardianUkraine60DaysHolt <-
data.frame(Date=TheGuardianUkraine60Days$TheGuardianDate,
           Holt=fitted(TheGuardianUkraineHolt))
TheNewYorkTimesUkraine60DaysHolt <-
  data.frame(Date=TheNewYorkTimesUkraine60Days$TheNewYorkTimesDate,
             Holt=fitted(TheNewYorkTimesUkraineHolt))
ElPaisUkraine60DaysHolt <-
  data.frame(Date=ElPaisUkraine60Days$ElPaisDate,
             Holt=fitted(ElPaisUkraineHolt))
FolhaDeSaoPauloUkraine60DaysHolt <-
  data.frame(Date=FolhaDeSaoPauloUkraine60Days$FolhaDeSaoPauloDate,
             Holt=fitted(FolhaDeSaoPauloUkraineHolt))
DailyMaverickUkraine60DaysHolt <-
  data.frame(Date=DailyMaverickUkraine60Days$DailyMaverickDate,
             Holt=fitted(DailyMaverickUkraineHolt))
TheHinduUkraine60DaysHolt <-
  data.frame(Date=TheHinduUkraine60Days$TheHinduDate,
```
```

```
             Holt=fitted(TheHinduUkraineHolt))

TheGuardianUkraine60DaysHolt$DateSequence <-
  seq(1, nrow(TheGuardianUkraine60DaysHolt))
TheNewYorkTimesUkraine60DaysHolt$DateSequence <-
  seq(1, nrow(TheNewYorkTimesUkraine60DaysHolt))
ElPaisUkraine60DaysHolt$DateSequence <-
  seq(1, nrow(ElPaisUkraine60DaysHolt))
FolhaDeSaoPauloUkraine60DaysHolt$DateSequence <-
  seq(1, nrow(FolhaDeSaoPauloUkraine60DaysHolt))
DailyMaverickUkraine60DaysHolt$DateSequence <-
  seq(1, nrow(DailyMaverickUkraine60DaysHolt))
TheHinduUkraine60DaysHolt$DateSequence <-
  seq(1, nrow(TheHinduUkraine60DaysHolt))

ggplot() +
  geom_line(data = TheGuardianUkraine60DaysHolt, aes(x=DateSequence,
                              y=Holt, colour="TheGuardian"), group=1) +
  geom_line(data = TheNewYorkTimesUkraine60DaysHolt, aes(x=DateSequence,
                              y=Holt, colour="TheNewYorkTimes"), group=1) +
  geom_line(data = ElPaisUkraine60DaysHolt, aes(x=DateSequence,
                              y=Holt, colour="ElPais"), group=1) +
  geom_line(data = FolhaDeSaoPauloUkraine60DaysHolt, aes(x=DateSequence,
                              y=Holt, colour="FolhaDeSaoPaulo"), group=1) +
  geom_line(data = DailyMaverickUkraine60DaysHolt, aes(x=DateSequence,
                              y=Holt, colour="DailyMaverick"), group=1) +
  geom_line(data = TheHinduUkraine60DaysHolt, aes(x=DateSequence,
                              y=Holt, colour="TheHindu"), group=1) +
  scale_color_manual(values = c(TheGuardian="steelblue",
TheNewYorkTimes="black",
                               ElPais="green", FolhaDeSaoPaulo="yellow",
                               DailyMaverick="gray", TheHindu="red")) +
  labs(colour = "Newspaper:", title = "Holt's Linear Exponential Smoothing
(Trend)",
       y = "Weighting") +
  theme_minimal() +
  scale_x_continuous(breaks = seq(0, 60, 5),
                     name = "Days from Beginning of Ukraine Invasion") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Appendix D.

```
```{r}

library(tidyverse)
library(readxl)
library(tidytext)
library(udpipe)

ElPaisModel <- udpipe_download_model(language = "spanish",
                                    model_dir = tempdir(), overwrite =
FALSE)
ElPaisModel <- udpipe_load_model(ElPaisModel$file_model)


ElPaisClean <- read_excel("ElPaisUkraineClean52Days.xlsx")


ElPaisWeightFactor <- 60 / 52


ElPaisTokenizedData <- udpipe_annotate(ElPaisModel, x =
ElPaisClean$ElPaisHeadline)
ElPaisTokenizedData <- as.data.frame(ElPaisTokenizedData)


ElPaisRelevantTokens <- ElPaisTokenizedData %>%
  filter(upos %in% c("NOUN", "VERB", "ADJ", "ADV"))


ElPaisCleanSentiment <- ElPaisRelevantTokens %>%
  inner_join(get_sentiments("nrc"), by = c("lemma" = "word")) %>%
  count(sentiment) %>%
  mutate(ElPaisWeighted = n * ElPaisWeightFactor) %>%
  arrange(-ElPaisWeighted)


ElPaisPositive <-
sum(ElPaisCleanSentiment$ElPaisWeighted[ElPaisCleanSentiment$sentiment %in%
                                          c("joy", "trust")])
ElPaisNegative <-
sum(ElPaisCleanSentiment$ElPaisWeighted[ElPaisCleanSentiment$sentiment %in%
      c("anger", "disgust", "fear", "sadness", "surprise",
"anticipation")])
ElPaisSentimentScore <- (ElPaisPositive - ElPaisNegative) / (ElPaisPositive
+ ElPaisNegative)

print(ElPaisSentimentScore)


ggplot(ElPaisCleanSentiment, aes(x = reorder(sentiment, ElPaisWeighted),
                              y = ElPaisWeighted)) +
  geom_bar(stat = "identity", fill = "green") +
  coord_flip() +
  scale_y_continuous(breaks = seq(0,
max(ElPaisCleanSentiment$ElPaisWeighted),
                                by = 10)) +
```
```

```
  labs(title = "Sentiment Frequency in El País Headlines (Weighted)",
       x = "Sentiment",
       y = "Weighted Frequency") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))


ElPaisCleanWordFrequency <- ElPaisRelevantTokens %>%
  count(lemma) %>%
  mutate(ElPaisWeighted = n * ElPaisWeightFactor) %>%
  arrange(-ElPaisWeighted)


ggplot(head(ElPaisCleanWordFrequency, 10), aes(x = reorder(lemma,
ElPaisWeighted),
                                                y = ElPaisWeighted)) +
  geom_bar(stat = "identity", fill = "green") +
  coord_flip() +
  scale_y_continuous(breaks =
                       seq(0, max(head(ElPaisCleanWordFrequency,
1)$ElPaisWeighted),
                           by = 10)) +
  labs(title = "Top 10 Frequent Words in El País Headlines (Weighted)",
       x = "Word",
       y = "Weighted Frequency") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```