

Corso di Laurea Magistrale
in Digital and Public Humanities (LM-43)
ordinamento D.M. 270/04

Tesi di Laurea

**A study on journalism from a
computational linguistic
perspective**

A corpus based analysis on Italian journalism

Relatore

Prof. Fabiana Zollo

Correlatore

Prof. Gianluca Lebani

Laureando

Gabriele Dettori

Matricola 884898

INDEX

INDEX	3
ABSTRACT	4
INTRODUCTION	5
1.THEORETICAL REASONS: DIGITAL HUMANITIES AND INFORMATION OVERLOAD.....	7
1.1.The role of the digital humanist.....	7
1.2.Information society.....	12
1.3.Journalism, social network and algorithms	15
2.LINGUISTIC ANALYSIS ON ITALIAN JOURNALISM.....	20
2.1.Corpus extraction.....	21
2.2.Journals' history.....	23
2.3.Computational tools.....	24
2.3.1.Data extraction methods and tools on La Repubblica and Il Fatto Quotidiano web archives.....	24
2.3.3.Linguistic analysis methods and tools.....	25
2.3.4.Stanza NLP processing language.....	26
2.3.5.Topic modeling methods and tools.....	26
2.3.6.Gensim.....	27
2.3.7.Sentiment analysis methods and tools.....	27
2.3.8.FEEL-it.....	27
2.4.Analysis and visualizations.....	28
2.4.1.Topic modeling and sentiment analysis.....	35
2.4.2.Comparison of recurring topic over the years - La Repubblica & Il Fatto Quotidiano	37
2.4.1.Polarity and sentiments over the years - La Repubblica	45
CONCLUSION	50
BIBLIOGRAPHY	54

ABSTRACT

The change that modern technologies have massively brought to the way people communicate and inform themselves is now more evident than ever. The objective of this thesis is to check if we can use computational methods to understand more on this change and to develop a reflection on the evolution of language used in Italian journalism by examining a large corpus of journalistic articles published online by two famous Italian newspapers. Through the use of computational tools, I extracted data on style variations in Italian newspapers over the past years in order to obtain information relating to recurring topics and the feelings with which they are treated and I compared the journalistic lexicon with that of more modern media, detecting correlations useful to study the impact of new media on journalistic communication.

INTRODUCTION

More and more studies highlight the problems about today's information and the ways in which fake news spreads. The exponential speed of technological progress and the changes it brings to people's lives figure a society in a state of cognitive shock. Individuals, constantly flanked by electronic devices through which they receive thousands of information every day, passively undergo the information mostly unaware of the functioning of the computer systems that provide it to them. The fragmentation of information translates into the lack of a unitary and shared narrative, which distances individuals from taking precise political positions, and allows the development of a distrust which translates into extremism and fake news. In such a context, the figure of the digital humanist can help can help translate the complexity of our hyper-technological society into a language more understandable to the people, and the historical gaze of the humanist can allow the recognition of recurring communication and behavioral patterns in the history of human beings. The technical-scientific tools achieved by the social sciences today more than ever need to be accompanied by humanistic analyzes that attempt to bring the particular analysis back to more universal considerations. In an attempt to respond to the problem of information overload, a term used to refer to the problems encountered by processing too much information from the Internet, my thesis aims to exploit the availability of huge amounts of journalistic material online in order to extract useful information and to gain insight into the long-term behavior of certain newspapers.

The discursive part of the thesis starts with a reflection on what I believe the role of the digital humanist should be and why I believe this role is fundamental in this precise historical period. Starting from this assumption, I will then describe the transformations that have taken place in our society as a result of the advent of modern information technology, and then try to describe the effect of this change on the way we receive journalistic news today. These contextualizations, in addition to providing the theoretical reasons for the thesis, can be useful to demonstrate how analyses of social phenomena should be carried out by the digital humanist on two fronts: the essayistic and the technical-analytical. Subsequently, I will show the results of the technical analyses applied on the extracted journalistic corpses, using computational methods to reconstruct in an alternative way a history of the topics covered in recent years by newspapers, merging topic modeling

extractions and sentiment analysis. Alongside this analysis of the recurring topics, always through the use of computational linguistics methods, I will trace a history of the linguistic variations of the language used by newspapers and a comparison of this with the language used by newspapers on Twitter to try to see if the advent of social media has left obvious influences on the communication style of newspapers.

1. THEORETICAL REASONS: DIGITAL HUMANITIES AND INFORMATION OVERLOAD

1.1. The role of the digital humanist

To begin this dissertation in Digital and Public Humanities, I decided to start by talking about the role of the digital humanist because I think that, in this new subject of integration between computer science and humanities disciplines, the meaning of Digital Humanities needs to be explored in depth in order to make the audience understand what we are dealing with in this subject. The Digital Humanities are so new that there is still no understanding of what theoretical contribution digital humanists can make to more general humanistic reflection. This discipline, however new, has already changed its name despite its few years. A few years ago we referred to this field of study by talking about "humanities computing". The term changed to "digital humanities" thanks to John Unsworth, Susan Schreibman and Ray Siemens who, as editors of the anthology "A Companion to Digital Humanities"¹, tried to prevent the field from being viewed as "mere digitization." Indeed, I am deeply convinced that the digital humanist must distance himself or herself from the initial idea of the "cultural digitizer" who exploits computer tools simply to enhance his or her analytical capabilities or simply to use digital media to spread cultural works. The role of the humanist, in my opinion, should remain unchanged: to reflect on the human condition. The digital humanist, should harness the power of information technology to understand the change technology brought to our society. Precisely because technology, never before in history, is radically altering every moment of the lives of the masses, the humanist who reflects on the human condition must be familiar with the digital so as not to run the risk of being left writing about the "remnants" of humanities productions that are not being affected by technology. Although today, in common thinking, areas such as philosophy and literature are considered to be extremely distant from computer science, on a theoretical level this distance is less great than is generally believed.

The philosophy of the 20th century with its insights into language and its analytical developments created fertile ground for the development of information technology and I

¹ Unsworth, Schreibman and Siemens, 2004.

think that without the philosophical complexity of Western thought which in the 19th century lead to the modern philosophy of language, computer science would have been slower to arrive. Alan Turing, considered one of the fathers of information technology, went to lesson Ludvig Wittgenstein's Lectures on the Foundations of Mathematics and probably this allowed him to obtain the most suitable theoretical foundation to work on his research that led to the creation of the Turing machine². Wittgenstein and other philosophers from the school of analytic philosophy such as Russell and Frege studied the logical principles of language in depth and allowed to realize the logical limitations of spoken language and to try to understand the language suitable for formal logic, which is the kind of language we use today to talk to computers³.

Although the world of humanities research seems so distant from that of scientific research we have to remember that research paradigms already changed and are still changing. Before the development of modern science the distinction between humanistic knowledge and scientific knowledge did not exist or in any case was not so relevant. Today's technological advancement and modern artificial intelligences, according to many, are introducing a cognitive paradigm shift. As Paolo Benanti explains in his lecture "Algoetica" held at FestivalFilosofia⁴, today's most cutting-edge scientific instruments do not base their decisions on causal linkages, but on correlations between data. If today we try to ask a satellite navigator using "Tree decision making" algorithms how to get to a place, he will point us to the road with the least traffic, without asking why that road is the least traffic. The development of artificial intelligences based on neural networks allow even more precision in computing and processing answers, but by basing their computation on correlation between data, when they make mistakes it is even more difficult to understand their errors. Are we willing today to abandon the search for causes in exchange for a calming control of the reality around us? In the belief that empirical analysis of reality is the main resource for performing scientific analysis but that reflection on the human condition cannot be exhausted by empirical analysis and requires a philosophical gaze, I am convinced that the digital humanities perspective, aware of the history of human thought and its limitations and up-to-date on the most cutting-edge scientific research methodologies, can enable the perpetuation of a search for correlations by leading it back to reflections that take into account the limitations of the paradigms of modernity, and can reduce the distance between the scientific

² Cfr. Murphy, 2021.

³ Gillies, 2002.

⁴ Cfr. Benanti, 2022.

and humanistic worlds, allowing philosophers to understand progress without being sidelined by it.

The goal that Digital Humanities should set for itself today is to add the value of the humanistic perspective to the analysis of contemporary society, taking advantage of the methods that computational sciences offer. For this reason, I wanted to write a thesis that could represent an initial experiment of analysis that seeks a balance between humanistic reflection and technical-scientific analysis. My thesis aims to describe the recent evolution of Italian journalistic language and topics through computational methods in order to elaborate reflections on the issue of information in today's society. The need to simplify the meaning of a multitude of data into a representation that is simple to understand has always been the role of the philosopher who absorbs information from new scientific discoveries to try to attribute meaning to human existence and to reality through the use of common language. With less ambitious goals, the Information Visualization field does something similar: representing data in a meaningful, visual way that users can interpret and easily comprehend. For this reason I chose this discipline for my thesis: today's great availability of data and analysis places a great responsibility in the hands of those who deal with presenting us the facts and news. I think that for there to be true democracy, the technological tools of data extraction, processing and presentation must be used in both directions: journalists to convince us of their ideas, readers to understand the reasons behind these ideas. My project does not claim to draw enlightening conclusions on journalistic communication but it is a starting point for developing a critical analysis method on how information is presented to us. Furthermore, I am convinced that the humanist has lost the authority that was one day recognized to him and now he's seen by most of the people exclusively as a creative. The complexity of programming and its links with the field of logic are the means that the digital humanist must know how to use to regain the lost intellectual authority and enhance his analytical skills on the human condition.

From a purely cognitive point of view, programming and computer knowledge more generally can produce an improvement in every man's reasoning abilities. The coldness of machine language and the need for spatially ordered writing can bring that attention to measurement and numerical values that is often lacking in the career of the humanistic intellectual. There are studies that have shown cognitive improvements in those who learn to programming. In 1991, Liao and Bright, two academics expert in education, made a study to asses the value of computers in grade-school education. The results of their analysis indicate that computer programming has slightly positive effects on student cognitive outcomes:

«Students are able to acquire some cognitive skills such as reasoning skills, logical thinking skills and planning skills, and general solving skills through computer programming activities.»

- *Liao and Bright*⁵

However, the researchers also admitted that this effect was moderate and the study did not compare whether other disciplines produced such an improvement in learning. Surely, however, programming and computational tools allow the student to concentrate on the reasoning and logic of his own actions, allowing the iteration of all repetitive operations. Beyond the cognitive level, I am profoundly convinced that understanding the principles of computer science is fundamental to making an analysis of contemporary. If the cognitive benefit of knowing how to program is that of being able to develop logical procedures useful for processing information in an orderly and fast way and for planning operations oriented towards a goal, the philosophical benefit can be precisely that of understanding how companies, corporations or even states today can perform certain operations and to what end their operations are oriented.

Exploring this perspective is Douglas Mark Rushkoff, an American media theorist, writer, columnist, lecturer, graphic novel author and documentary filmmaker. As Rushkoff said in 2010, “to the person who understands the code, the world appears as a series of decisions made by planners and designers about how the rest of the world should live”⁶. Being able to see the human choices behind marketing operations, works of art, political campaigns, is crucial to understand contemporary society. The perspective highlighted by Rushkoff is somewhat reminiscent of John Carpenter's science fiction movie "They live", where the protagonist gets a pair of glasses which, when worn, show the "ideology" behind the things he sees: looking at a billboard, he simply reads the words “spend money”, looking at another that exploits the body of a young attractive girl, he reads the words “reproduce yourself”⁷. In my opinion, the glasses of “They Live” movie today can be achieved by the ability to merge a critical look at society typical of the humanities with a technical look at design intentions typical of engineering and computer science subjects. In today's society, where all information is easily replicable and communicable, the humanist must be able to recognize the decisions behind the large corporations and large multinationals which constantly increase their ability to carry out cultural operations and influence masses. In my opinion,

⁵ Liao and Bright, 1991.

⁶ Rushkoff, 2010.

⁷ Carpenter, 1988.

the digital humanist should become the example of the man who, aware of the power of machines, knows the machine so as not to be controlled by it, and who tries to suggest to other men how to do it, like Plato's caveman who, once he realizes he lives in a cave, tries to convince others as well. Leaving the power of machines in the hands of a few is a danger to democracy. For this reason I completely agree with Rushkoff when he states:

«We teach kids how to use software to write, but not how to write software. This means they have access to the capabilities given to them by others, but not the power to determine the value-creating capabilities of these technologies themselves. [...]. The more humans become involved in [new technologies'] design, the more humanely inspired these tools will end up behaving. We are developing technologies and networks that have the potential to reshape our economy, our ecology and our society more profoundly and intentionally than ever before in our collective history.»⁸

- *Rushkoff*

If the humanist is unable to understand the tools that are changing our world, he will no longer be able to reflect on the human condition. Thanks to computational tools, the humanist has access to text sources that he would never have dreamed of consulting before, and he has the power to develop tools that analyze texts for him. To demonstrate how much IT tools can help the development of critical thinking, it is necessary to start from a linguistic analysis. Having machines processing linguistic data from newspaper articles allows us to read the content of the articles in a colder and more distant way and can lead us to new and unexpected conclusions. I think that nothing better than a linguistic analysis of journalism can demonstrate on the one hand the power of computational tools and the enormous amount of data usable by the humanist, on the other hand it can bring out a discussion on how society is changing in relation to new technological developments.

⁸ Rushkoff, 2010.

1.2. Information society

In 1985, Daniel Bell introduced the concept of an "information society." At the time, the widespread use of the Internet was not yet a reality and people primarily obtained information through newspapers and television. However, Bell saw that society was undergoing a transformation into post-industrial societies centered around information. He believed that the United States was no longer primarily an industrial society, as manufacturing and production were no longer the main focus of their economy. Bell described this transformation as an "information explosion", which he defined as a set of interrelated factors including the growth of science and technology, increasing demand for news, entertainment, and knowledge, a rapidly growing and more educated population, and improved communication technologies. He also highlighted the impact of television and computerized data banks in creating global awareness and interconnectedness. It is clear from Bell's considerations that he believed that this transformation was not just an economic shift, but also a change in population behavior. In the years since Bell's ideas were introduced, there have been rapid technological advancements that have accelerated this transformation, such as the exponential growth in the capacity to store and exchange information. In the second half of the noughties, technological power had already reached a point where it was possible to allow very good data transfer speeds over the Internet, but it took commercial phenomena capable of making the user experience pleasant and easy for the Internet to become popular and affordable. The combination of two commercial successes in particular led the world to be constantly connected to the net: the diffusion of smartphones and the birth of social networks. The smartphone permitted people to be able to connect to the internet everywhere, while social networks, which started as a simple resource used to communicate among friends, become the central vehicle for exchanging all kinds of informations. This has allowed the digital information to be more and more easily within everyone's reach in most parts of the world.

But now that we have achieved what could have seemed to be one of humanity's greatest achievements, which is having an enormous amount of information always available, it is turning out to be something we are not yet cognitively ready for and this is evident especially on young people which are the most connected to the web while their cognitive development phase is still ongoing. Today, the issue of use of social networks by young people is much debated, with more and more studies testifying to an increase in mental disorders related to

overexposure to information sources⁹. We talk about “cognitive overload” which is a disorder characterized by spending more and more time online surfing from one site to another in search of more and more up-to-date and complete information, without being able to interrupt or decrease the actual time spent connected to the network. Initially, the person experiences a sense of pleasure and excitement once he or she finds the information sought, which later gives way to a vicious cycle in which the person finds himself or herself trapped: the initial information is no longer sufficient and the search for more is perceived as a duty and a necessity¹⁰.

The definition which in my opinion is more explicative of information overload process was given by Orrin Edgar Klapp in 1986: information overload occurs when "the information available exceeds the user's ability to process it"¹¹. When faced with a large amount of information, it is often difficult or almost impossible to mentally organize all the content at hand properly. This definition has been deepened in 2003 by Speier, Valacich and Vessey that in “The Effects of Interruptions, Task Complexity and Information Presentation on Computer-Supported Decision Making Performance” stated that:

« Information overload occurs when the amount of information received by a system exceeds its relative processing capacity. Individuals have a limited information processing capacity. Consequently, when information overload occurs, it is possible that it will result in a reduction in the quality of the decision to be made.»¹²

- *Speier, Valacich and Vessey*

It may seem paradoxical, but receiving too much information can lead to a reduction in the ability to process information, and thus end in a decrease in the amount of information learned and consequently to reduce the quality of the decisions to be made.

Herbert Simon, an American economist and psychologist who studied the concept of “bounded rationality” focus on how every decision we take is affected by cognitive limitations. The author was one of the first to analyze the link between amount of information and degree of attention, stating that “information richness breeds attention poverty”. These lead individuals to make mostly satisfactory and suboptimal decisions, as

⁹ Durlofsky, 2020.

¹⁰ Rutkowski, Saunders, 2018.

¹¹ Klapp, 1986.

¹² Speier, Valacich and Vessey, 2007.

they are unable to process all the information that would be necessary to make a perfectly considered choice¹³.

Another author who was visionary in predicting the cultural, economic and social changes brought about by modern technology is Talvin Toffler. Toffler popularized the concept of “Information overload” through his best-selling book "Future Shock"¹⁴, published in 1970, in which he tried to predict the effects of the industrial revolution and the arrival of new technologies. He associate the term “information overload” with the feeling of being inundated with data and knowledge and defined the “future shock” as the anxiety brought on by “too much change in too short a period of time”¹⁵. Toffler was prescient about the impact of the emerging digital world and did not believe its impact would be restricted to specific spheres such as business, but would influence everything including our personal lives and the way we interacted with others. If Daniel Bell spoke of an “information society”, Toffler preferred to use the term “post-industrial society” describing the arrival of a society that would decouple our sense of the world and our place within it, while the old social foundations of family, church, community, nation and profession would be swept away. Prosperity would continue to grow indefinitely in this new era, but the price would be a loss of identity. "We must look for totally new ways to anchor ourselves, because all the old roots, whether religion, nation, community, family or profession, are now shaking under the impact of the hurricane of accelerating momentum," he wrote¹⁶.

If we receive too much informations, we cannot expect to understand well what we are being told and it will also be more difficult to distinguish the true ones from the false ones. Newspapers, adapted through social pages to new forms of communication, are still one of the primary sources of information about current events. If our ability to comprehend information is compromised, it is right to try to understand more about how information is communicated throughout history to be able to see how our perception of events is filtered by the narrative that journalists make, but also how information is produced today in relation to new technologies.

¹³ Simon, 1957.

¹⁴ Toffler, 1970.

¹⁵ Ibidem.

¹⁶ Ibidem.

1.3. Journalism, social network and algorithms

The advent of social networks has revolutionized the way information is done and forced major newspapers to adapt to the fast communication brought by technological innovations. Journals, looking for an alternative and less formal space with which to connect with their readers, can take advantage of interaction with readers and to have appeal to a youth audience. Social media have become the main vehicle for journalistic information. To increase data traffic and the level of interaction with readers, newspapers can ask opinions to their readers, make them vote in a survey and have a clear feedback about latest news being published. According to Richard Sambrook, director of the BBC's Global News division, the impact of social network on information has been underestimated. Traditional media are adopting social media, especially with blogs and Twitter, he admitted, but no one discusses the long-term effects¹⁷. The first long-term effect he pointed out is that journalists are no longer the owners of the news. From a custodian of information, the journalist is turning into a mere disseminator. On the one hand, citizens having a camera always in their pocket can document numerous phenomena even without it being their job, and on the other hand, newspapers are also specializing in getting their information directly from the Internet. The second long-term effect Sambrook described, which is even more alarming, relates to journalistic objectivity. He states that social networks are bringing a new type of objectivity. According to him, objectivity was once designed to provide journalism that could be trusted and today news still needs to be accurate and fair, but it is equally important for readers, listeners and viewers to see how news is produced, where the information comes from, and how it works. The emergence of news is as important as its dissemination. While it is right that by increasing the number of sources, readers need to filter out news that comes from unknown sources, this can also lead to an unfounded reliance on official sources, which benefited by the trust placed in them can afford to take sides on certain issues and lose their objectivity. In my opinion, it is not correct to talk about a new type of objectivity: we can't stop defining objectivity as being "above the sides". If the level of trust is defined by where the news comes from or how it is presented, we are not talking about objectivity but about the principle of authority, a philosophical principle that mankind has not considered valid since the Middle Ages.

But despite this disagreement on the meaning of the word "objectivity," I very much agree with Sambrook when he explains that, in order to maintain their authority, journalists must not bend too much to the most effective communication techniques to gain the readers' attention, but must be disciplined, analytical and explanatory:

¹⁷ Sambrook and Cushion, 2016

« Information is not journalism, he explained further. You get a lot of things, when you open up Twitter in the morning, but not journalism. Journalism needs discipline, analysis, explanation and context, and therefore for him it is still a profession. The value that gets added with journalism is judgment, analysis and explanation - and that makes the difference. So journalism will stay - he was optimistic about that. However, journalists must understand one rule: if you believe you are in competition with the internet, find your way out. Collaboration, openness and link culture are rules, you can't deny at the moment.»¹⁸

Richard Sambrook

If so far we have focused on analyzing the changes that the internet has brought about in the way journalistic content is used, it is right to start spending a few words also on how information is brought to users. The issue of competing 'with the internet' and its technologies is not only about the availability of information and its speed of transmission, but also about the speed of data processing. Lately, there has been a lot of talk about the risks associated with artificial intelligences and the analysis of the risks that artificial intelligence can bring to humanity is central to the current ethical-philosophical debate. In "Brief Answers to the Big Questions", Stephen Hawking focuses precisely on these risks, assuming that the main danger of artificial intelligence is not their mistakes, but their competence. If artificial intelligence, trying to pursue its goals, develops a will that conflicts with ours or with human welfare we will be in trouble¹⁹. The concept of the danger of "excessive competence" of machines is useful to us in understanding the risks of algorithm-driven information. The potential of artificial intelligence is now expanding in all the ambitions of human knowledge, but ordinary users probably know it best because they know that, when they open a social network, the first contents that they will see are the contents that the algorithm deems most interesting for them. In fact, a feature of all web sites through which we use information is to acquire data on users and offer content based on this personal data. Social networks make use of artificial intelligence algorithms designed to use the users' personal data to present information by attracting the user's attention as much as possible, and this is the core business of social networks, because they make money by selling user data or through clicks on advertisements. Since artificial intelligence makes it possible to understand how intelligent computer systems capable of simulating the capacity and behavior of human thought can be created, these algorithms don't just think about what you might be interested in 'in general', they go even deeper: they think about what you might be interested in at a specific time of day, they try to extrapolate information about your emotional state to understand when you are most likely to make a purchase, and, as

¹⁸ Oxford Social Media Convention, 2009.

¹⁹ Cfr. Stephen Hawking, 2018

explained in the film “The Social Dilemma”, they can even go so far as to understand your sentimental situation and recommend something based on it²⁰. For this reason we can say that social media’s algorithms are often designed to induce precise behavioral patterns that cause users to stay connected and continue their activity on the platform, which will then get more personal data to sell and more clicks on advertisements banners. These algorithms do not take into account the type of information presented (except in the cases of the recent pandemic where social network initiated campaigns against vaccine misinformation); they are simply interested in presenting information that captures your attention.

Tim Wu in his book “The Attention Merchants” describes the history of professions that have profited from attracting the attention of the people in order to advertise products or services. These professions were made possible by the dissemination of the press but, today, thanks to the technological means of the “information society”, the companies that do this are the richest companies in the world. The dependence of media on advertising is not a new thing: in 1833 Benjamin Day, the founder of The Sun, was the first newspaper editor who created his business on advertisement and brought to popular use an object that was previously considered exclusive to the elite²¹. It was precisely the ability to sell advertising space and thus sell readers' attention to advertisers that made it possible to make the price of the newspaper cheap and allow their circulation to the poorer classes. Having decisively demonstrated that a business could be founded on the resale of human attention, Benjamin Day became the first attention merchant. Tracing the history of attention merchants, Tim Wu comes in the last chapters to affirm that the revolution that Google brought to the world of attention merchants was precisely a renegotiation of the cost of that attention.

« The key was in renegotiating the terms under which the public was asked to tolerate ads. For as Google became a part of everyday life over the 2000s—its name becoming a synonym for search—it presented what seemed by the standards of the day a very reasonable tradeoff, which few would ever question, and which caused no degradation, to judge by earnings growth. So unintrusive was AdWords that some people didn’t even realize that Google was ad-supported: it just seemed to be there for the taking like manna from the heavens. Google had, in fact, laid bare what had originally been so miraculous about the attention merchant model—getting something truly desirable at no apparent cost. For what really seemed like nothing, the public got the best search ever designed and, in time, other goodies as well, like free email with unlimited storage, peerless maps, the world’s libraries, and even research devoted to exciting innovations like self-driving cars.»²²

- *Tim Wu*

²⁰ Orolsky, 2020.

²¹ Wu, 2016.

²² Ibidem.

Thanks to Adword, advertising techniques became so sophisticated that they were not perceived as intrusive. In addition to being non-intrusive, online advertising is the perfect place to develop target marketing strategies. Approximately ten years ago, tech firms initiated the employment of algorithms to tailor our internet experience. By utilizing advanced technology and copious amounts of customer information, these companies started to correctly anticipate our preferences more accurately than we could ourselves. This led to the growth of e-commerce and the transformation of journalism to cater to the individualized attention economy. Targeted advertising refers to a type of advertising, including online ads, that are aimed towards a specific group of people with particular characteristics, related to the product or individual being advertised. These characteristics can be based on demographics, such as race, socioeconomic status, gender, age, education, income, or employment. Alternatively, they can be based on psychographic factors, like consumer values, personality, attitudes, opinions, lifestyles, and interests. Nevertheless, there is a concealed adverse consequence which was referred to as "the filter bubble"²³ by Eli Pariser. The expression "filter bubble" refers to a situation of intellectual seclusion that has arisen due to the use of algorithmic capable of filtering the information that is presented to users based on their previous preferences, actions and interests. Eli Pariser described it as a "personalized ecosystem of information created by algorithms"²⁴. The intellectual seclusion that has arisen due to the use of algorithmic systems leads to the creation of multiple versions of the internet that correspond to the individual tastes, opinions and interests of different users. This "information bubbles" can lead to individuals having a distorted perception of reality, where they believe that their own interests are the only ones that exist since they are not exposed to new ideas and informations. In terms of practical implications, filter bubbles have been linked to the rise of populist and nationalistic movements, including the election of Donald Trump as President of the United States. Therefore, it is important to be aware of the existence of filter bubbles and the role that algorithms play in shaping the information that is presented to us. This can help to mitigate the negative effects and promote a more diverse and inclusive online environment. In response to the problem of the filter bubble, my thesis wants to start tracing the route for a humanistic method of analysis that exploits in favor of an analysis above the parts precisely the technical potential used by large companies to sell us their advertising.

In order to better balance humanistic and technical skills, I decided to carry out my analysis on Italian journalism. Being aware of the most important events that took place in my home country, I can more easily try to attribute the data extrapolated from the analysis of

²³ Pariser, 2011.

²⁴ Ibidem.

the journalistic corpus to specific historical events. Furthermore, according to the Word Press Freedom Index, a comprehensive study conducted by "Reporter sans frontieres", Italian information has many problems. The study evaluates the state of global press freedom and categorizes it based on five indices: political, legal, economic, socio-cultural, and security. Unfortunately, Italy has seen a significant drop in press freedom and now ranks 58th, down 17 spots from 41st place just a year prior. This decline can be attributed to a number of factors, including criminal sanctions, economic dependence, and a polarized public opinion fueled by the pandemic. This has resulted in a rise in verbal and physical assaults on journalists, with 44 reported cases of intimidation in the first quarter of 2022 alone.²⁵ The country's freedom of the press is being threatened by organized crime, especially in the south, and by various violent extremist or protest groups, which have increased in number during the pandemic.

As a democratic and culturally rich country with, it is unacceptable for Italy to have such a low rank in press freedom. This highlights the importance of developing tools to analyze journalistic material, so as to gain a clearer understanding of how journalists communicate information. I have taken the initiative to create an embryonic versions of this tools, with the goal of utilizing Digital Humanities to conduct a critical analysis of journalistic communication.

²⁵ Press Freedom Index, 2022.

2.

LINGUISTIC ANALYSIS ON ITALIAN JOURNALISM

Linguistic analysis

- **Journals' tweets compared with journal articles' titles:** A comparative analysis between the tweets of the newspapers' Twitter account and the web article titles to verify if there are correlations and if we can speak of a linguistic impact on newspapers due to the language of social media.
- **Linguistic features over the years:** Extracting data on linguistic features divided by year in order to be able to produce a visualization of the evolution of these values over the years and to verify if there were any evident changes in the style of writing, richness of languages and textual complexity by journalists.

Topic extraction and sentiment analysis

- **Comparison of recurring topics and their polarity:** An analysis through artificial intelligence of the themes that have been more recurring since 2010 and an extraction of the emotion polarity with which these themes are treated, trying to reconstruct a chronology of the most important historical events of these years. This operation will be feasible both on the corpus of the Repubblica and on that of the Fatto Quotidiano and will allow us to compare the two newspapers both in terms of editorial direction and in terms of opinions expressed.
- **Polarity and sentiments over the years:** An analysis through artificial intelligence of the themes that have been more recurring since 1986 and an extraction of the emotion polarity and sentiments with which these themes are treated, trying to reconstruct a chronology of the most important historical events of these years. This operation will be possible only through the analysis of the corpus of La Repubblica, sufficiently extended in time to include 4 decades.

2.1. Corpus extraction

With modern tools of computational linguistics and thanks to artificial intelligence algorithms that allow natural language processing, topic modeling and sentiment analysis, it will be possible to analyze and extract visualizations that allow us to get an idea of the linguistic style used by these newspapers, the topics that are of most interest to the newspapers distributed over the years, the sentiments with which these topics are treated, and I will try to verify whether journalistic language has adapted to that of social networks in recent years. In order to ensure that the type of analysis could be reproducible by anyone, I chose to use material available online at a low price, and therefore decided to download the articles from the web archives of two famous Italian newspapers. In analyzing the impact that social networks and the digitization of news media has had on journalism I compared the two newspapers' most recent tweets (Twitter does not allow access to the archive to see older tweets) and the headlines of their articles. This will allow us to analyze, beyond the content, also the formal and stylistic developments that recent journalism has undergone as a result of the more general social communicative change that has occurred on a large scale with the spread of social networks.

In choosing the reference newspapers, I took into consideration all the most popular Italian newspapers but at the end I selected the two journals whose archives have a decryptable structure. This permitted me to iterate calls to archives' endpoints to automatically obtain a list of links of the articles to which I could subsequently extract the text of the articles through BeautifulSoup python package. Indeed, BeautifulSoup allows to extract the source code of the web page. Once I got pages downloaded in HTML format I had to analyze the HTML structure to understand under which tags the information I needed was enclosed, so as to be able to script the extraction processes of this data and relative cleaning. Between all the newspaper web pages consulted, the only two that I was able to decrypt were La Repubblica and Il Fatto Quotidiano. It would certainly have been more useful for analysis to be able to extract data from as many different newspapers as possible, but unfortunately technical limitations did not allow this. I therefore decided to create two very large corpora for the two newspapers, which fortunately are representative of quite distinct and often conflicting strands of thought.

About 44 000 articles were extracted for each newspaper. The articles from La Repubblica, beginning in 1985, allow us to cover almost 40 years of contemporary Italian history that on the archive of Il Fatto quotidiano cannot be there because of its youth: the Chernobyl disaster, the collapse of the Soviet Union, Gulf War, Operation "Mani Pulite" and

the end of the first republic, Silvio Berlusconi's rise to power, 9/11, Italy's entry into the euro, and the economic crisis of 2009. Both papers cover the last 10 years of history: the birth of the smartphones and social networks, the rise of populist movements in politics and the Covid-19 pandemic.

2.2. Journals' history

La Repubblica is an Italian daily newspaper headquartered in Rome and owned by GEDI Gruppo Editoriale. It is the second largest newspaper in Italy in terms of circulation and readership, with Corriere della Sera being the first. La Repubblica was founded in 1975 by Eugenio Scalfari, former director of Il Messaggero, with Caracciolo and Mondadori investing 2 billion 300 million lire. In the 1980s, CIR, owned by De Benedetti, increased its stake in the publishing company to 50%. De Benedetti aimed to become the majority shareholder of Mondadori by purchasing the shares from Arnaldo Mondadori's heirs, but this was blocked by Silvio Berlusconi. This resulted in the 'Segrate war', a legal dispute that lasted for more than two years and ended in 1991 with the separation of the book and magazine sector, which went to Berlusconi's Fininvest, and La Repubblica, L'Espresso, and local newspapers, which went to Gruppo Editoriale L'Espresso. In 1998, Società Editoriale La Repubblica merged into l'Espresso, which merged with Itedi in 2016 to become GEDI Gruppo Editoriale. In 2020, CIR sold its shares in GEDI to Exor, owned by the Agnelli family, and Giano Holding completed a public tender offer for the remaining shares, becoming the owner of 100% of the shares and delisting GEDI from the Italian Stock Exchange.

Il Fatto Quotidiano is a daily Italian newspaper that was founded in 2009 by Marco Travaglio, Peter Gomez, Marco Lillo, Bruno Tinti, and Antonio Padellaro. The newspaper has an average circulation of around 50,000 and an average readership of around 475,000. The website, ilfattoquotidiano.it, is directed by Peter Gomez and registered as an independent publication. Until February 28, 2014, the website was available under the Creative Commons Attribution - Noncommercial - No Derivative Works 2.5 Italy license, and the content remains under this license even though the terms and conditions of use were subsequently modified. The company's articles of association require that no controlling shareholder can own more than 16.67% of the share capital, and entrepreneurial shareholders cannot own more than 70% of the capital in total. This means that the consent of journalists cannot be disregarded for important resolutions, and a qualified majority of 70% plus one is required for decisions regarding the editorial line or the appointment of the editor-in-chief.

Both newspapers have waged political and legal battles against Silvio Berlusconi, who can be considered their common enemy. However, the two newspapers diverge on many issues, Il Fatto Quotidiano tends to be more tolerant with populist and anti-political opinions, while La Repubblica is considered closer to Italian and European governmental establishment.

2.3. Computational tools

2.3.1. Data extraction methods and tools on La Repubblica and Il Fatto Quotidiano web archives

The La Repubblica archive allows subscribers to perform search queries, being able to specify keywords and date. If we try to connect to the archive without typing anything in the search will show all articles from 1985 onwards, but despite the fact that the archive and the web page show that there are thousands of pages, having arrived at the 51st page of articles the website will respond with an error message "Searches can show a maximum of 50 pages of articles." To get around this problem, I made an extraction script that would randomly select commonly used words from a long list and iterate for each month from 1985 to the present a search with some of these words. In this way I was able to make sure that I took articles evenly distributed over time whose contents were not too much affected by the words that I compulsorily had to include in the search. For the Fatto Quotidiano, on the other hand, the extraction was easier since the online archive consists of the web pages divided by month and which already had the date in the url. So a simple loop that iterated month by month and extracted the links to the articles reported was enough. The complete extraction of the approximately 90,000 articles took about 3 weeks to complete. The computer systems of online newspapers obviously have systems in place to block excessive download requests from the same IP address to prevent server overload and allow all users smooth browsing. To get around these blocks, it is necessary to take advantage of Python options that allow the process to be paused, anticipated by an "if" condition that would acknowledge the error message due to too many requests. I strongly recommend that those who want to perform this kind of data mining operation make sure that the script allows monitoring of the process by sending out in print the information useful to understand if the mining is happening correctly and to be able to optimize the mining time. In my case the script that downloaded the articles sent to print the title of the article, the url of the article the first 500 characters of the article, the exact time the download ended, the index number of the article, and the percentage of completion of the process for every article.

```
Titolo: Sondaggi, Fratelli d'Italia supera la Lega: ora è il primo partito. M5s guadagna uno 0,6% nella settimana dello scontro Grillo-Conte
```

```
Giorgia Meloni supera Matteo Salvini e Fratelli d'Italia ora è il primo partito. Almeno secondo l'ultimo sondaggio di Swg per il Tg La7. Negli ultimi sette giorni Fdi guadagna due decimali e ora è al 20,7%. Nello stesso periodo di tempo, però, il Carroccio di decimali ne perde tre, fermandosi al 20,3%. Guadagna poco - uno 0,2% - anche il Pd, che è terzo col 18,8%. Si confermano quarti i 5 stelle col 16.6%, che però nella settimana dell'attacco di Beppe Grillo a Giuseppe Conte (il sondaggio è stato
```

```
data: 2021-06-28 author: [Tg La7. url: https://www.ilfattoquotidiano.it/2021/06/28/sondaggi-fratelli-ditalia-supera-la-lega-ora-e-il-primo-partito-m5s-guadagna-uno-06-nella-settimana-dello-scontro-grillo-conte/6244561/
```

```
2023-01-13 12:28:01.663396
```

```
INDEX: 39463
```

```
Process: 87.76 %
```

Fig. 1 - Data extraction process monitoring sample

2.3.2. Web scraping through BeautifulSoup

Web scraping, also referred as web harvesting or web data extraction, is a computerized process of gathering data from websites using software programs. The programs typically imitate human navigation on the World Wide Web through the use of the Hypertext Transfer Protocol (HTTP) or by accessing the website through browsers such as Internet Explorer or Mozilla Firefox. This technique is related to the indexing of websites by search engines, which mostly rely on bots for the process. However, web scraping is distinct in that it focuses specifically on converting unstructured data found on the web, commonly in HTML format, into metadata that can be stored and analyzed locally in a database. It is also similar to web automation, which involves the use of computer software to simulate human navigation on the web. To prevent web scraping, some websites use methods such as detecting and blocking bots from accessing their pages. To counter this, there are web scraping systems that utilize techniques such as DOM parsing, computer vision, and natural language processing to emulate human web navigation. These techniques make it possible to collect the contents of web pages for offline analysis, which can be used for purposes such as price comparison, weather data monitoring, detecting changes in a website, scientific research, web mashups, and web data integration. BeautifulSoup is a well-regarded Python package for parsing HTML and XML documents, even those with malformed markup, as its name suggests, "soup" refers to the structure of unclosed tags. It creates a parsing tree from parsed pages, enabling the extraction of data from HTML, making it a valuable tool for web scraping.

Digital humanists especially appreciate this tool as it enables them to create scripts that can automatically extract vast amounts of textual data from any website whose code structure is not encrypted.

2.3.3. Linguistic analysis methods and tools

Natural language processing (NLP) is a multidisciplinary field that combines the areas of linguistics, computer science, and artificial intelligence to study the interactions between human language and computers. The objective is to develop computer systems capable of comprehending and analyzing vast amounts of natural language data. By understanding the contextual complexities of language, these systems can extract relevant information and insights from documents, as well as classify and arrange them in a structured manner. However, the pursuit of this goal presents several challenges, such as speech recognition, language comprehension, and the ability to generate natural language. These difficulties

require the development of sophisticated algorithms and models to accurately process and analyze language data.

2.3.4. Stanza NLP processing language

Stanza is a comprehensive collection of tools for linguistic analysis across various human languages. This advanced natural language processing package, developed in Python, can transform raw text into meaningful insights, such as sentence and word lists, base forms, parts of speech, morphological features, and named entity recognition. Additionally, Stanza provides a dependency parse of the text's syntactic structure, utilizing Universal Dependencies formalism, with support for over 70 languages. Featuring highly accurate neural network components, Stanza enables the option to train and evaluate the models with custom annotated data.

2.3.5. Topic modeling methods and tools

Topic modeling is a statistical method for uncovering the latent, abstract topics present in a collection of documents. Utilized in both statistics and natural language processing, this text-mining tool provides a means for discovering the semantic structures hidden within a text corpus. The idea behind topic modeling is that a document about a specific topic will contain certain words that appear more frequently compared to words associated with other topics. For example, in a document that is primarily about dogs, one would expect words such as "dog" and "bone" to be frequently used, while in a document about cats, words such as "cat" and "meow" would be more prevalent. Topic models mathematically capture this idea by creating clusters of similar words, allowing for the examination of a set of documents and determining, based on the word statistics, what the topics might be and the proportions of each topic in each document. These models are often referred to as probabilistic topic models, as they rely on statistical algorithms to discover the latent semantic structures of a text corpus.

In the age of information overload, topic models provide a means for organizing and making sense of large amounts of unstructured text data. Originally developed as a text-mining tool, they have since been applied to a variety of data types, including genetic information, images, and networks, and have a wide range of applications in fields such as bioinformatics and computer vision.

2.3.6. Gensim

Gensim is an open-source software library that offers a variety of natural language processing (NLP) capabilities, including unsupervised topic modeling and document indexing and retrieval based on similarity. The library utilizes state-of-the-art statistical machine learning techniques to achieve these functionalities. Designed to be both fast and efficient, Gensim is implemented using Python and is optimized to handle large text collections. The library employs data streaming and incremental online algorithms, allowing it to process large amounts of data without sacrificing performance. This unique approach sets Gensim apart from many other machine learning software packages, which often focus solely on in-memory processing.

2.3.7. Sentiment analysis methods and tools

Sentiment Analysis is a multi-disciplinary field that leverages the power of Natural Language Processing, Text Analysis, Computational Linguistics, and Biometrics to systematically recognize, extract, measure, and analyze emotional states and subjective information. This technology finds extensive usage in a broad range of sectors, including customer feedback obtained through reviews, surveys, and social media platforms. It's also utilized in healthcare to gather and analyze patient feedback. The purpose of Sentiment Analysis is to assist businesses with marketing, customer service, and even clinical medicine, by providing them with a comprehensive understanding of how their customers feel about their products and services.

2.3.8. FEEL-it

Feel-it is a novel benchmark corpus of Italian Twitter posts annotated with four basic emotions: anger, fear, joy, sadness. By collapsing them, it can also be used for sentiment polarity analysis.

2.4. Analysis and visualizations

2.4.1. Journals' tweets compared with journal articles' titles

Among the linguistic analyses of Italian newspapers that I consulted, Giusy Scarfone's master's thesis jumped out at me. She analyzed the tweets of La Repubblica and La Stampa from a linguistic point of view, but without the use of computational tools. In her thesis, she highlights a series of linguistic characteristics that lead her to conclude that the language used by these newspapers on Twitter is extremely simplifying and tends towards communicative efficiency. In response to this thesis, I wanted to try to verify if this type of "web" communication had somehow translated into journalistic communication outside of Twitter. The features most determined to lead her to this conclusion were the monopropositionality of the tweets and the use of nominal sentences, as well as the use of evocative expressions such as the use of coordinating conjunctions.

« From the analysis of the corpora of the posts and tweets of 'La Repubblica' and 'La Stampa,' the tendency to extreme synthesis of information, aimed at obtaining maximum communicative efficiency, stands out. Thus, from a syntactic point of view, the tendency to a fragmented and simple style, characterized by little articulation, is not surprising: monopropositionality and nominal style are frequently used in the writing of journalistic contributions on social networks, on the one hand in the wake of that desire for extreme synthesis and speed that characterizes contemporary journalism, and on the other hand to increase the incisiveness of the news, increasing its communicative impact»²⁶.

- *Giusy Scarfone*

My comparative analysis between the headlines of newspaper articles and tweets verified that these characteristics aren't actually attributable to contemporary journalism but they belong to a communicative style that has been characteristic of newspapers for a long time. According to my analysis, this communicative style highlighted by Scarfone corresponds to the communicative style already used by newspapers when writing headlines. The need to condense information into an 'attention-grabbing' headline does not appear to be higher on social network tweets than on newspaper headlines; on the contrary, tweets, being 140 characters in length, are less often composed of simple sentences or nominal sentences compared to articles' titles. The communicative simplicity that Scarfone attributed to the use of simple and nominal sentences is therefore not attributable to the trait of a modern communicative style characterized by extreme synthesis, but by the simple need to

²⁶ Scarfone, 2017.

summarize an article in a few words. In fact, newspaper articles' titles are even shorter than tweets and for this reason more often composed of simple and nominal sentences.

Table 1: Comparison between tweets and articles' titles

	Tweets		Titles	
	La Repubblica	Il Fatto quotidiano	La Repubblica	Il Fatto quotidiano
<i>Mean number of words</i>	15	16	8	15
<i>% Nominal phrases</i>	89%	89%	99%	80%
<i>% Simple phrases</i>	90%	84%	99%	78%
<i>% Phrases with coordinating conjunctions</i>	27%	31%	12%	27%

From the values shown in Table 1, we can see that the headlines of La Repubblica's articles are significantly shorter than both the tweets of the two newspapers and the headlines of the Fatto Quotidiano. In fact, since 1986, the average length of La Repubblica's headlines has ranged from 7 to 8 words. Preferring to write short headlines, La Repubblica's journalists have to exploit simple or nominal sentences as much as possible, and in fact 99% of the headlines fall into these categories. Again due to the same need to make short headlines, we can also notice less use of conjunctions, which are often used in coordinated sentences. Since Il Fatto Quotidiano is a newspaper that was born when social media was already undergoing great development, it can be assumed that it has developed a style of headline writing more similar to that of tweets: in fact, the values for the number of words, nominal phrases, simple sentences and sentences containing conjunctions are similar to the tweets of newspapers. I agree with Giusy Scarfone's analysis on the fact that, in order to always maintain a simple style, the use of subordinate clauses is highly avoided in the titles and tweets and the few times that a mono-propositional sentence is not used, the use is preferred of simple coordinates such as "e" or "but". We can therefore say that comparing the tweets of newspapers with the headlines of news articles has not revealed any impact of the language of the Internet on journalistic styles and that the need for simplicity and communicative

efficiency is not attributable to social networks but to the need newspapers have always had to condense essential information into a headline of a few words.

2.4.2. Linguistic features over the years

Type/token ratio

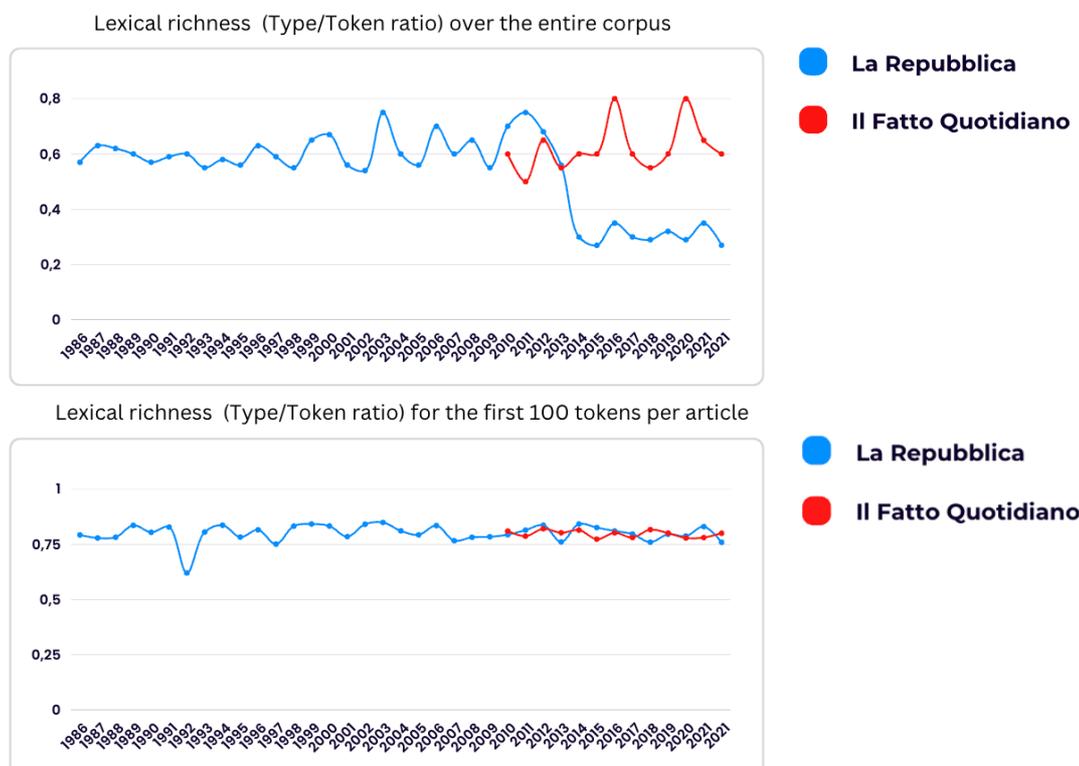


Fig. 2 - Type/token ratio

The Type/Token Ratio (TTR) or lexical richness measures the ratio between the number of lexical types (i.e. unique words) and the number of tokens, providing a metric of lexical variation. The value of lexical richness can therefore give us an idea of the linguistic skills of the authors of the texts, but also of their tendency to reflect and re-elaborate. As reported by Dominique Brunato in her text “Operationalizing Linguistic Complexity from a NLP Perspective”²⁷, lexical richness values can also be indicative of text complexity. In fact, it is expected that a text with few variations in linguistic register is less useful for informing and more complex, and in fact bureaucratic and legal or professional texts have a low level of lexical richness.

²⁷ Brunato, 2015

This value can be very important for us, on the one hand because it is somewhat indicative of the lexicon of the writers of the articles, and allows us to try to establish which journalists with a richer lexicon belong to which journalists. Secondly, the value can allow us to see if there have been variations in lexical richness over the years.

In computational linguistics, this value calculated for objects of different length is not considered reliable: it is claimed to be too dependent on the object length. To test whether this is true, I calculated this value both by taking whole articles into account and by taking into account only the first 100 tokens of each article. The articles extracted from La Repubblica before 2014 had an average length of about 500 words, a length that doubled from 2014 onwards. By calculating the type/token ratio on whole articles, it emerged that the value had halved since 2014. When instead I applied the analysis taking into consideration only the first 100 tokens, the values showed a very linear continuity over time. We can therefore confirm that this value cannot be considered reliable if the analysis is not applied to linguistic objects of the same length.

We can say that at a temporal level no evident variations in lexical richness have been recorded.

Noun/verb ratio

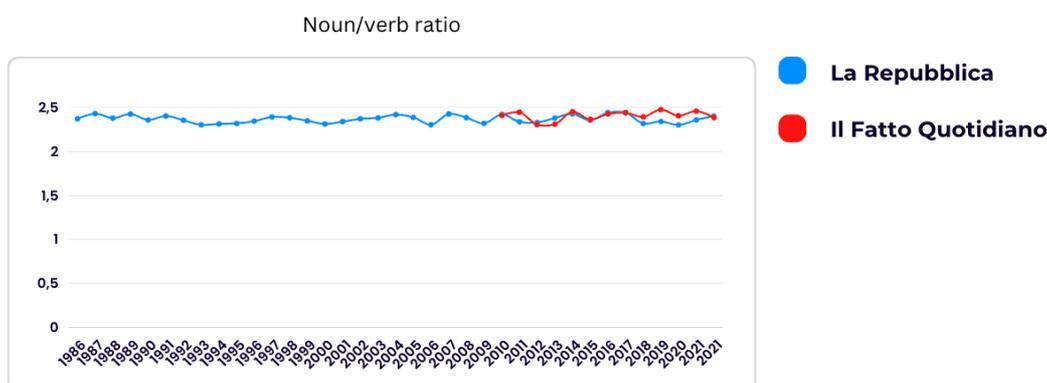


Fig. 3 - Noun/verb ratio

An interesting value for our analysis is the noun/verb ratio. In fact, in written texts it is usual for more nouns to be used while in oral texts more verbs are used. However, this is not constant and the genre of the written production may influence this value. Fictional prose,

for example, compared to informational prose, turns out to be closer to speech values. Journalistic informational texts, however, turn out to be closer to speech than academic prose or official documents. The noun-verb ratio, which is the proportion of nouns divided by the sum of the proportion of nouns and the proportion of verbs, indicates a noun bias when its value is greater than 5 and a verb bias when its value is less than 5.

I wanted to extract this value to see if the simplification of language also spills over into an increase in colloquial style and thus verbs, but the result was negative: for both journals the noun/verb ratio values have not changed over the years.

Lexical density

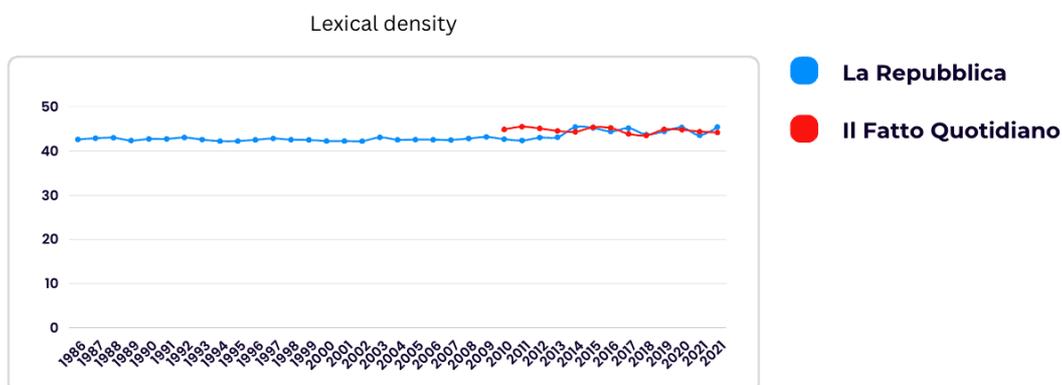


Fig. 4 - Lexical density

If type/token ratio gives an indication of lexical variety, lexical density works as a measure to quantify the informational load provided by a particular text, as it expresses the ratio between content and functional vocabulary. Lexical density is calculated as the percentage of content words in the whole text. If there are 100 words in a text and 60 are content words, then the lexical density is 60%. If there are 124 words in a text and 44 are content words the lexical density is approx. 27%. As content word we have considered verbs, adjectives, adverbs and nouns.

Also for this value, no significant changes have been recorded in recent years.

Sentence length

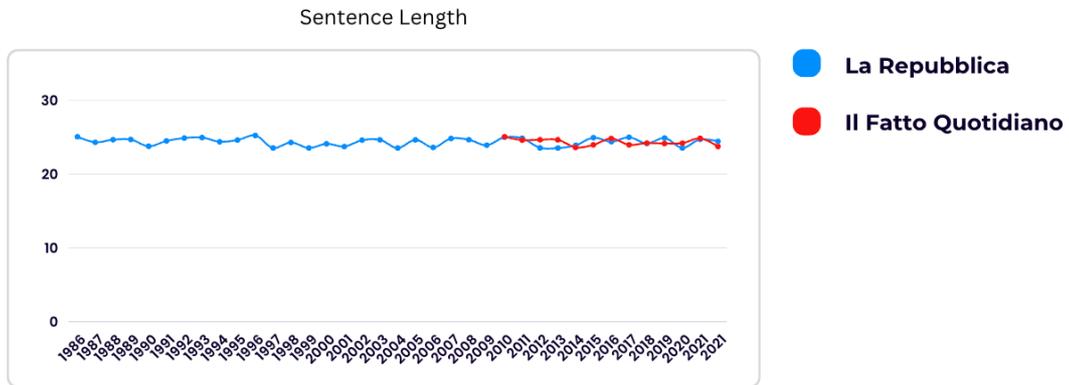


Fig. 5 - Average sentence length

It is interesting to note, as shown in the graph, that the average length of the sentences also remained extremely constant: around 25 words for sentence.

From a lexical analysis point of view, computational tools that do analysis on large amounts of data have not been shown to reveal obvious stylistic variations for any of the features taken into analysis. We can think that the formal aspect of journalistic communication has not undergone much variation in relation to the language of new media. Instead, in the next chapter we will go on to use the tools to analyze the content of articles, looking for correlations between the use of target terms and the historical events that characterized the period of their use. The target terms were selected by artificial intelligence algorithms (from Gensim) that group terms by topic. Also through artificial intelligence (from FEEL-it), I also extrapolated the level of emotional polarity associated with the words together with which the term is used.

2.4.1. Topic modeling and sentiment analysis

The idea of combining the extraction of recurring topics with sentiment analysis came in response to the need to synthesise the content of newspapers in order to have a more general prospective on the topics dealt with and the tone in which they are dealt with. I think that the reconstruction of historical events from the way they have been reported is a very useful resource in contemporary analysis. Having a large body of journalism and the tools to analyse allows us to trace recurring patterns in the journalistic narrative on the one hand, and to reconstruct historical events of the past by taking into account the popularity and frequency with which they were covered in the press on the other. Topic modeling technique identifies naturally occurring clusters of items, known as topics, even when the desired outcomes are unclear. The topic modeling methods I used is Latent Dirichlet Allocation. In LDA, documents are composed of different words, and each topic is comprised of various words as well. The objective of LDA is to determine which topics a document pertains to, based on the words it contains. I decided to apply the topic extraction algorithms on the newspaper corpus divided by year because if I had applied this model on the entire corpus I would probably have extracted only the most trivial topics, without any words that could refer to a precise historical event. Applying the year-by-year extraction instead revealed the topics that were central in the year analysed. Once I extract these topics that Gensim returns to me via a list of clustered words, I decided to apply sentiment analysis to the parts of the text where those words were used, so that I could extract an average of the emotions and tones with which these topics were treated. The script I wrote searches the articles for these target words, and analyzes the sentiments derived from the sentences that precede, contain, and succeed the word, with a 200-character range around the target word. The package I used, FEEL-it, one of the few python packages for sentiment analysis in Italian, tries to determine from the words used the sentiment with which they are used, specifically it selects between joy, anger, sadness and fear. By averaging these values it also extracts a more generic value of positive or negative polarity, which always tends more towards negativity than positivity.

In our reconstruction, we will only be able to compare the results of the analysis between the two newspapers from 2010 onwards, while for the years from 1986 to 2010 we will only be able to analyse the data from La Repubblica. It is fair to state that sentiment analysis, as well as the extraction of recurring topics using LDA models, are not tools designed for the analysis of journalistic corpuses and that my analysis is an embryonic attempt to direct these analyzes towards socio-political analysis objectives.

Recurring topics

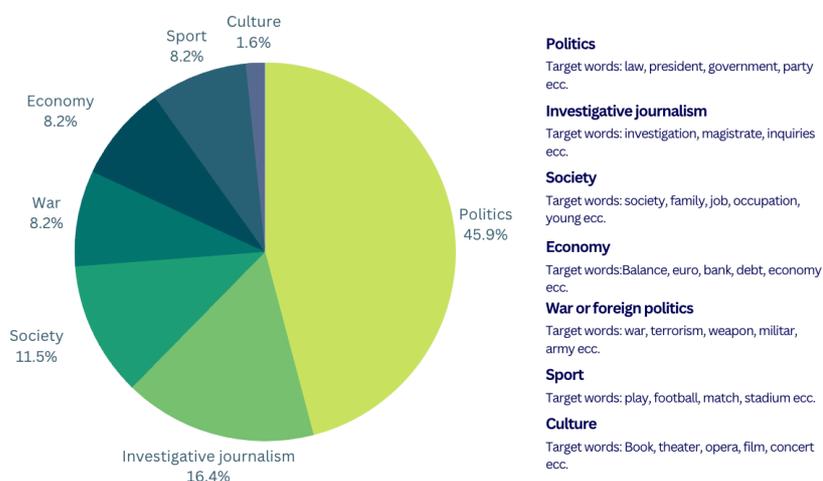


Fig. 6 - Il Fatto Quotidiano - Recurring Topics

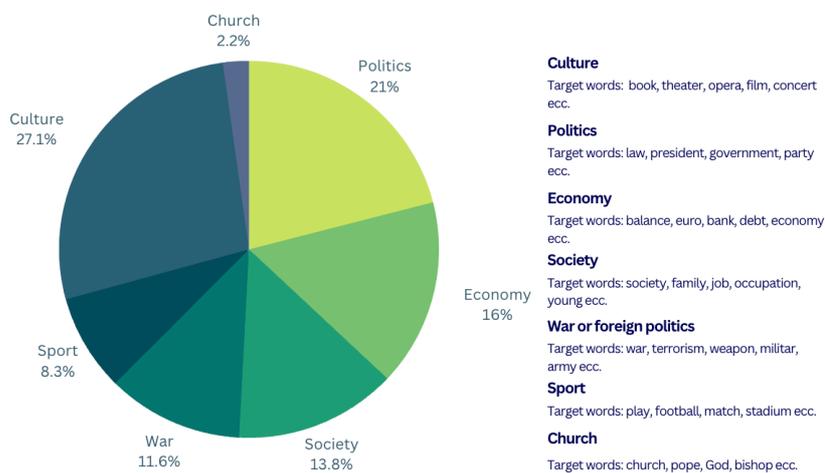


Fig. 7 - La Repubblica - Recurring Topics

2.4.2. Comparison of recurring topic over the years - La Repubblica & Il Fatto Quotidiano

The most recurring themes common to both newspapers are politics and football. Both newspapers are mainly known for their political and economic insights, especially the Fatto Quotidiano which leaves less room for extra-political content. However, La Repubblica demonstrates a recurring treatment of articles concerning art, music, culture and entertainment, but we can also note a greater interest in news and football. Il Fatto Quotidiano is more concentrated on the political topic, and for this reason the extraction of topics reveals more specific themes, while for La Repubblica the usual topics always occupy one of the available places. It is also necessary to specify that the corpus of the Fatto Quotidiano, as previously mentioned, is concentrated on the last 10 years of history and it is therefore also for this reason that it is more precise.

It is evident from these graphs that there is a much more policy-focused approach for Il Fatto Quotidiano and instead a more generalist approach for La Repubblica. The second recurring topic for Il Fatto Quotidiano is investigative journalism, which constitutes 16.4% of the recurring topics. The two newspapers turn out to be very similar in terms of the space devoted to sports. The cultural topic in the Republic appears from this data to be the main one, this is because the newspaper constantly devotes some of its pages to promoting cultural activities, and although politics is probably more centered for the newspaper, the topic of culture has earned a place among the recurring topics every year and for this reason it appears more frequently. It should also be added that the Society topic and the Politics topic, like sometimes the Economy topic are part of topics that often intertwine, and for this reason the Culture topic results more frequently when compared to the other three topics taken individually. Although La Repubblica already declared itself "a secular newspaper" at the time of its founding, it has repeatedly found religious topics to be among the main topics covered. In Italy, the subject of the church is a recurring topic, on the one hand because the country has deeply Christian roots, and on the other because the institution of the church is linked to politics and scandals.

In this section I will compare the polarity with which popular issues are treated in both newspapers, so as to try to understand similarities and differences of the two styles of journalistic communication. We will also try to understand whether it is possible to trace these data back to specific historical events that may have determined the calculation of polarity values.

To begin this comparison, I would start with the only topic that has obtained a positive polarity value: culture.

Culture polarity

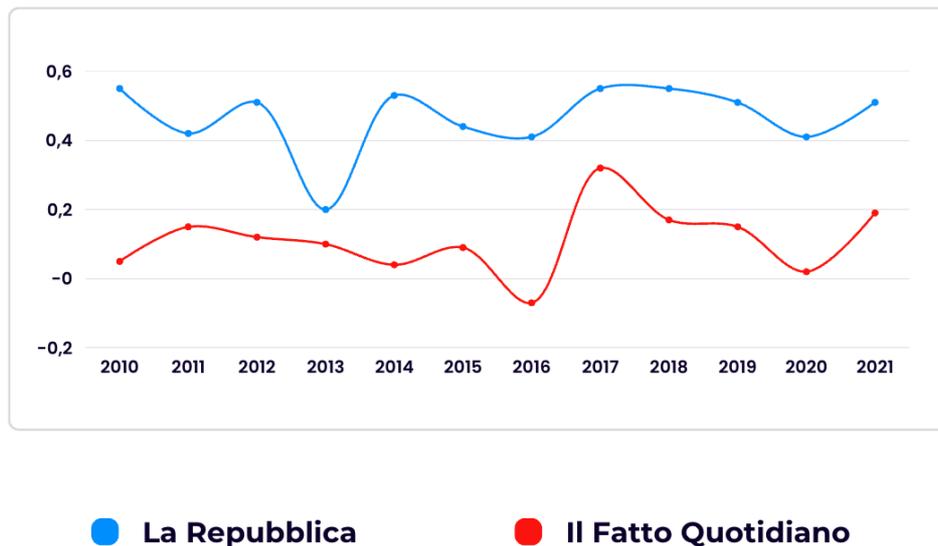


Fig. 8 - Culture polarity comparison

As shown in the graphs in Fig. 6 and 7 on the newspapers' favorite topics, we know that the cultural topic is much more frequent in La Repubblica newspaper than in the Fatto Quotidiano. Except for a single negative spike recorded in the graph of La Repubblica, we can say that the two line-charts seem to exhibit correlations and that, although the tone of La Repubblica is decidedly more positive in the narrative, the two newspapers devote themselves to the cultural topic in a similar way. Since the coverage of issues related to books, concerts, films, and exhibitions is often linked to advertising dynamics and since

there are many and varied artistic works presented each year, it is impossible to attempt to attribute precise causes to this graph.

The second topic that is treated more positively is sports.

Sport polarity

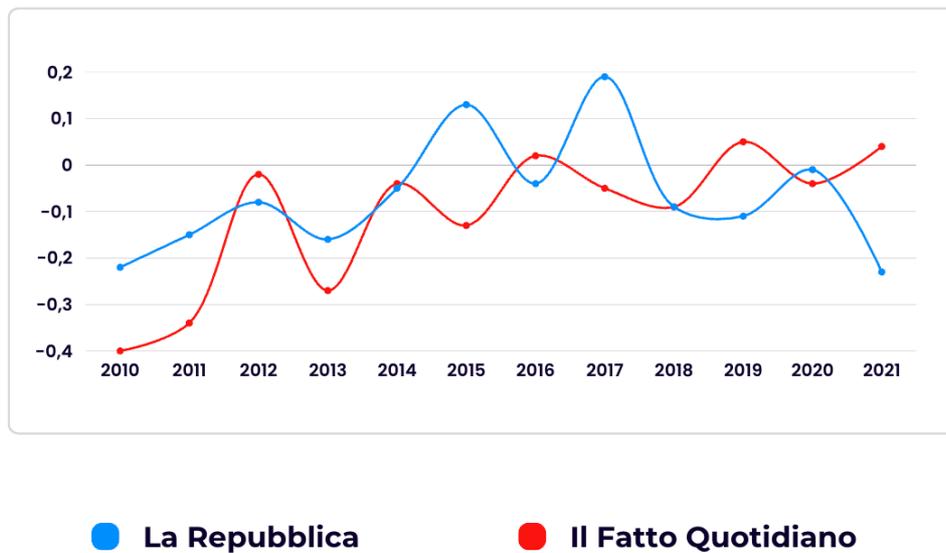


Fig. 9 - Sport polarity comparison

Again, given the many sport games that are played each year, it is difficult to attribute the spikes in this graph to precise events. Several sportive clubs play in Italy each week, and because of this, the only way to attribute the graph's trend to a precise cause might be by referring to national team matches. In fact, in 2010 the value peaked in negativity in both newspapers, and this could be attributable to Italy's defeat in the first round from the World Cup, just at the World Cup following the one Italy won in 2006. To test this hypothesis, I looked for examples in the text to see how the topic had been covered in that year and indeed found obviously unfortunate narratives of that outcome.

Just as in Berlin we wrote a great page of Italian soccer, so in Johannesburg we wrote an unfortunate one. And the responsibility does not lie with Lippi. If we in Berlin have a fault, it is that we were unable to convey the meaning of the Azzurri and the World

Cup to those who were not there in Germany." did you hear back from Lippi? "just to know how he was. I hope such a great coach does not end his career this way."

Would you like to cancel the last world cup, you who said you had a dream as ct? "I don't want to cancel anything at all. I've played three World Cups, it's like living a dream. And the joy that the national team has given me is unparalleled. I thank God for everything I've done: even bad things are good."

Despite this hypothesis we do not have enough evidence to attribute the negative spike in the graph to this specific event.

The graph probably most attributable to particular historical events is the one related to war and terrorism.

War and terrorism polarity

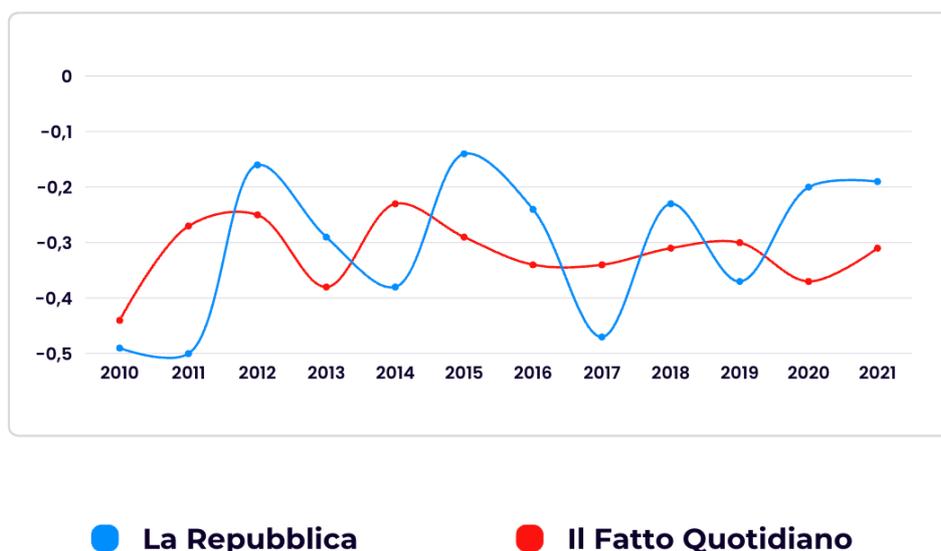


Fig. 10 - War polarity comparison

As we can see from the graph in fig. 10, there are many peaks for this topic. The subject is by its nature linked to dramatic tones and the polarity levels are therefore always very negative. As far as La Repubblica is concerned, the peaks of negativity are 2011 and 2017. 2011 is certainly characterized by the Libyan conflict which led to the killing of Gaddafi by American intelligence, while 2017 was certainly characterized by the war in Crimea. A

smaller extension of the peaks of negativity can be found in Fatto Quotidiano and is probably attributable to a more constant treatment of war issues and international politics.

The next two topics analyzed, society and politics, are very interesting to compare between the two newspapers because they can testify through data their political alignment. Initially I had assumed that the analysis related to political topics could give results more representative of the political positions of journalists, but since politics is a factional environment, the polarity values extracted at the occurrences of words such as "party" or "politics" can be contrasting and particular peaks of negativity can be compensated by positive opinions on the opposite political faction, for this reason I believe the graph in fig. 11 relating to the topic of society is more suitable for speculation on the political position because the emphasis on social emergencies, work, young people and families can be traced back not only to the description of the state of well-being of the country but also to a criticism of the policies implemented.

Society polarity

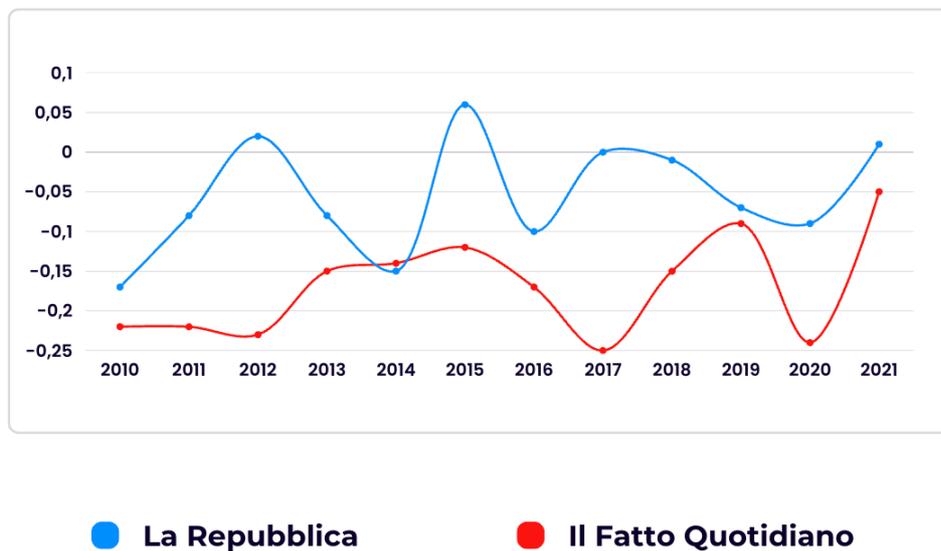


Fig. 11 - Society polarity comparison

Politics polarity

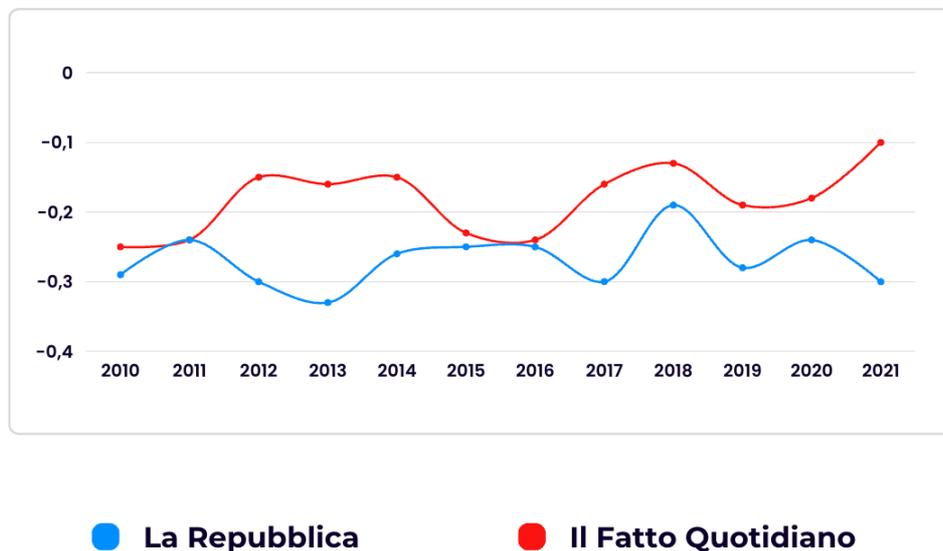


Fig. 12 - Politics polarity comparison

The first observation is that the graph on politics is the only one in which we can find a higher average of polarity values in Il Fatto Quotidiano than in La Repubblica. This is surprising since Il Fatto Quotidiano, with its investigations into political corruption and its closeness to movements born of anti-political currents, I would have imagined it more critical of the political argument. As mentioned earlier, however, this graph is not too reliable as opinions of appreciation and political criticism can be balanced with the variation of the parties in question and also because the values undergo too slight variations to be attributable to precise events. If, on the other hand, we pause to analyze the tones with which social issues are treated, some considerations can be made: la Repubblica has used much more positive tones than usual for the years 2012, 2015, 2021. The year 2012 represents the year of the Monti government, which began in November 2011 and ended in April 2013. This government was strongly supported by La Repubblica which, together with the international press, strongly pressed for the resignation of Silvio Berlusconi and the installation of a caretaker government aimed at restoring the nation's economic problems. We will recall how the end of the Berlusconi government, historical enemy of both newspapers taken into consideration, was interpreted by La Repubblica as the beginning of a new political era and the country's economic recovery. Here is a quote from 2012 from La Repubblica relating to

the theme of young people and work and a more specific consideration on the establishment of Mario Monti.

"I believe that if we put ourselves together in the effort and effort to build new bridges, to change mentality and create more cohesion both in work and in society, tensions at all levels will not coagulate," said Bagnasco.

*- Bishop opinion on Mario Monti's new
government*

Few weeks we hear about Italy and Europe with new accents (a first taste, but brief, we had in the last Prodi government). The citizens approve for now, they know a kind of relief. Do they also feel comforted, in their grim rejection of politics? Maybe, but there is something ominous in this dual vision: on the one hand the politicians, on the other a prime minister who has all the qualities of a statesman, who internalizes democratic representation to the maximum, and yet is perceived as technical, extraneous to games national. Being impolitic in a broken democracy has its virtues: impolitic is whoever does not possess the cleverness of a professional politician.

*- Journalist opinion on Mario Monti's new
government*

Il Fatto Quotidiano, on the other hand, has always shown itself to be very critical of the Monti government and this is demonstrated by a further lowering of the polarity recorded in 2012.

For La Repubblica, the year 2015 may correspond to a positive judgment on the work of the Renzi government, while in 2021 both newspapers showed themselves to be very lenient with the new Draghi government, judged by almost all of the Italian press as the most competent person who could govern Italy after the pandemic.

In general, we can see an increase in positive tones from the newspapers in the years when a new government takes office, while we notice negative tones in relation to the periods preceding the fall of a government. In fact, the Italian electoral system does not allow much political stability and the vast majority of governments are unable to reach the end of their mandates due to internal conflicts within the majority parties. For this reason it is very easy to perceive the fall of a government the months before this happens, precisely because of conflicting statements between the members of the same government. From this point of view, newspapers could be attributed the tendency to "get on the winner's

bandwagon", communicating positive emotions when a government takes office and criticizing its actions when its fall is foreseen.

As for the economic topic we can see that the levels of polarity are similar. Both newspapers have sections for economic analysis, but the issue is not so central to either of them.

2.4.1. Polarity and sentiments over the years - La Repubblica

In this section I wanted to exploit the temporal extension of the La Repubblica archive to carry out an analysis that included the last 30 years of history. I added the one on the occurrences of the target terms to the already seen polarity graph so that you can get an idea of the popularity of the topic for each year. Furthermore, making the most of FEEL-it's potential, I also displayed the histogram which shows the percentage of the type of emotion recorded at the occurrences of the target word. The feelings recorded by FEEL-it are joy, anger, fear and sadness. Having a wider temporal analysis range, we could take into consideration the most evident polarity change peaks and try to associate the values with specific historical events.

Let's start with the graph that has the most positive values: the one relating to the cultural topic.

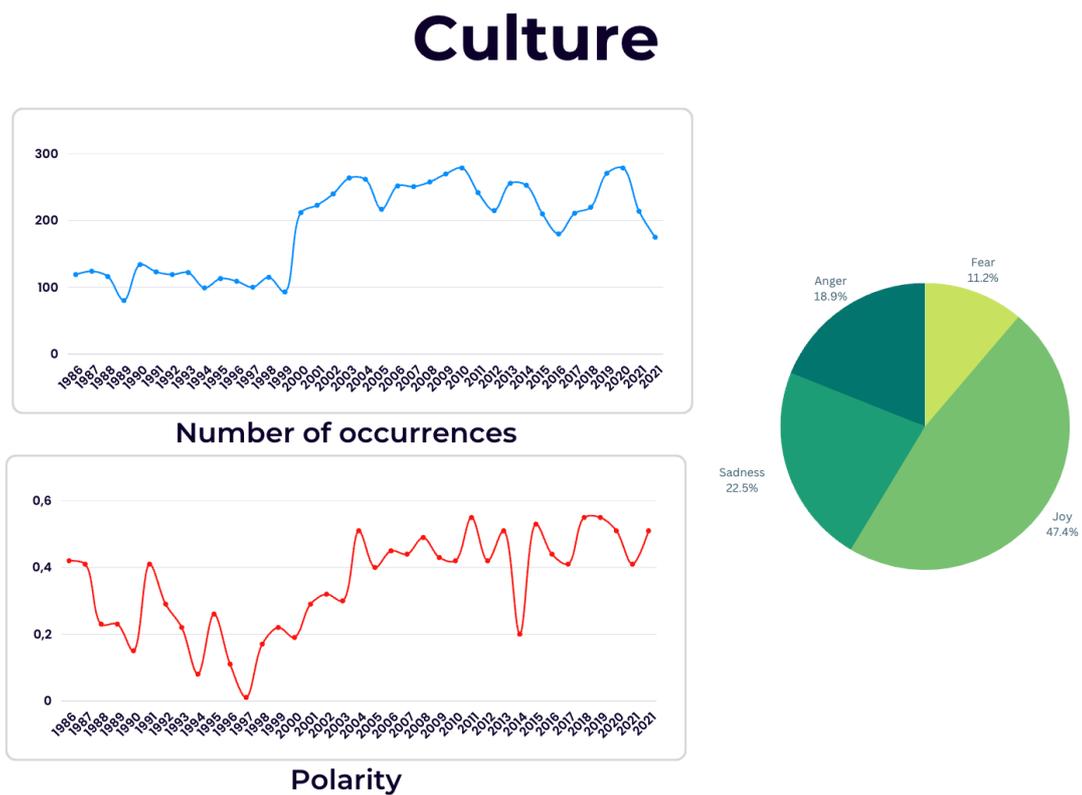


Fig. 13 - Culture - Occurrences, polarity and sentiments.

As also mentioned for the previous comparative analyses, it is difficult to attribute the values of the polarity graph to any particular event of a cultural type considering the number of books, concerts and films mentioned in the newspapers each year and the same is for sport, considering the number of different matches played every year. We can certainly say that the culture graph shows that the emotion most associated with the themes is joy.

Among the topics selected, certainly the one relating to war and terrorism is the easiest to associate with precise historical events.

War and terrorism

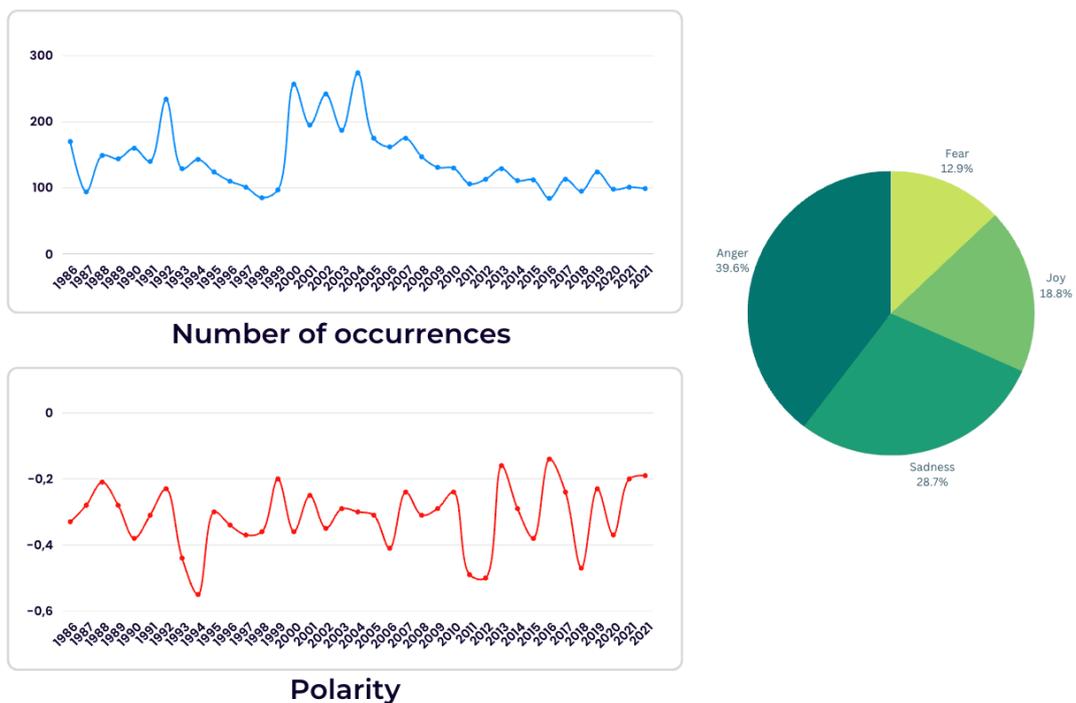


Fig. 14 - War and terrorism - Occurrences, polarity and sentiments.

From the number of occurrences in the graph we can see how the end of the 80s, with a peak in 1992, are characterized by a high presence of the war theme. This is because there was a high tension in the geopolitical equilibrium characterized by the disintegration of the Soviet Union in 1992 and the formalization of the end of the Cold War. Furthermore, the Gulf War, which took place between 1991 and 1992, also increased the popularity of the theme. The concentration on the war theme has been waning during the 90s up to re-surge in

2001, the year of the attack on the twin towers and the consequent start of the war in Iraq. Obviously the emotions most associated with the theme are negative, mainly anger and fear.

The other two graphs which are interesting to analyze are the ones about politics and society topics.

Politics

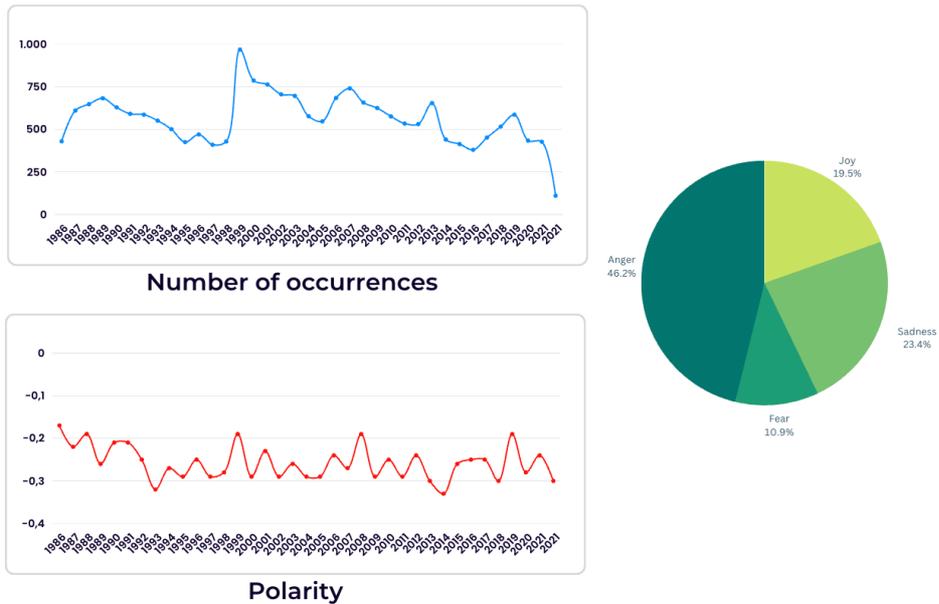


Fig. 15 - Politics - Occurrences, polarity and sentiments.

Society

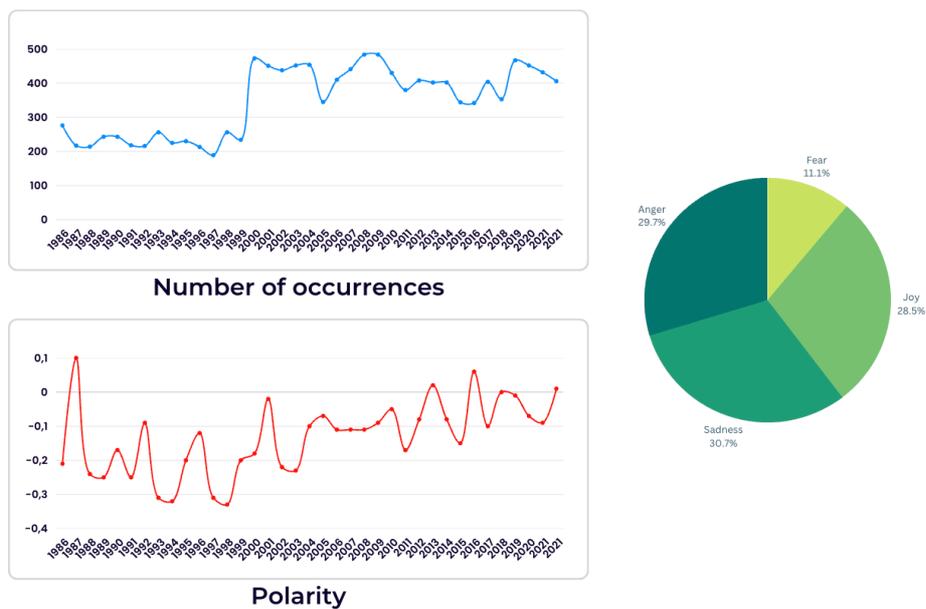


Fig. 16 - Society- Occurrences, polarity and sentiments.

The graph of the occurrences of the topic politics shows us an interesting fact. Never as in 2021 had there been so little talk of political terms in La Repubblica. The pandemic has brought new terms and new worries into public discourse and national politics has shown, especially in the initial phase of the pandemic, close cohesion and a limited spirit of political competition. From the histogram of feelings we can see how the political topic is by far the most filled with angry emotions. As we have said, Italy is a country with high levels of corruption which have led to the development of strong anti-political sentiments which are often channeled by the newspapers demonstrating a strong tendency towards criticism and indignation. It is also interesting to note which years have reached the negative peak of polarity in the use of terms concerning politics: 1993 and 2014. The historic event that can be connected to the first year is certainly the “Mani Pulite” investigation which led to the famous scene of the coin toss against Bettino Craxi and his subsequent exile to Hammamet. The investigation conducted by the P.M. Antonio Di Pietro demonstrated the existence of a system of illicit party financing which led to the arrest of several political representatives and the dissolution of the two major majority parties of the time: the Democrazia Cristiana and the Partito Socialista. The investigation was at the center of the media spotlight for months and the indignation was such that the end of the “First Republic” was proclaimed in the media. The high negativity in the political discussion may have also influenced the theme of society, which, as shown in the figure 15, also recorded one of its highest peaks of negativity in that year. The year 2014, on the other hand, was characterized by the disintegration of the Enrico Letta government by Matteo Renzi, a party mate of the ex-premier who, with the support of a current within the Democratic Party, dissolved the caretaker government led by a member of his own party to recreate another with a similar parliamentary majority but a different executive. Matteo Renzi's attitude has been described as a betrayal of a friend by the media, who have created catchphrase by repeating the phrase "Enrico, be happy!" pronounced by Matteo Renzi to Enrico Letta shortly before bringing down his government. The phrase became one of the first Italian political “memes”.

Although limited to two newspapers, the analysis made it possible to extrapolate many data attributable to particular historical events, and the possibility of going beyond the simple popularity of certain topics by also analyzing the emotions with which they are treated made it possible to relate the treatment of such events also to the type of political position shared by the newspaper. The analysis has technical limitations: having only one computer available forced me not to be able to extract the sentiment for all the words relating to a topic that indicated the extraction of Gensim (the operation requires time if made by a single computer), the only access to the newspaper archives free does not allow me to make different comparisons between the newspapers and it would take some team-work to look for

more specific articles that reflect the meaning of the graphs. For this reason it would be unfair to draw absolute conclusions on the communicative intentions of the two newspapers taken into consideration. However, the method of analysis, if applied to more newspapers, with more technological resources and with the collaborative work of other humanists, could prove to be extremely effective in describing the communicative intentions of newspapers.

CONCLUSION

As already anticipated, on the stylistic level of the language, the research did not produce results that show significant variations. The linguistic features taken into consideration have not changed in recent years. Although certain technological innovations can rapidly change society, we can say that the formal use of language is resistant to certain social changes. Languages change over time: an example of this is the gradual process that led the populations of the Italian peninsula, after the fall of the Western Roman Empire, primarily to the development of dialects. But we are talking about a process that has lasted for centuries and technological innovations, however impacting on the habits and customs of populations, cannot influence so much on language. This may seem deleterious for the type of analysis carried out in this thesis but the constancy of the formal properties of language further supports the thesis that the use of language allows us to extrapolate useful information for the analysis of today's society. The fact that the implicit conventions (the formal aspect of the language) undergo very little variation allows us to concentrate on the content analysis without having to worry about whether the formal variations influence this second type of analysis. From the point of view of language content analysis, in fact, it is possible to extrapolate information related to precise historical events and the modalities of their journalistic treatment. History, of course, is written with words, but very often the words that are used to describe a historical period are not the words of that historical period. The possibility of synthesizing in graphs information on the frequency of use and the methods of use of certain terms and therefore on the popularity of certain topics and the emotions associated with them is, in my opinion, an essential resource for humanists today because it allows to mitigate the problems due to information overload. The too much availability of information and the computational processes that present us with information following logics that favor the business of the platform that transmits it are factors that determine the possibility of readers getting misconceptions about certain historical events, perceiving fear about events exalted by the media, and ignoring events kept in the background of public discourse. The work in this thesis is directed toward the goal of developing a gauge of the importance placed by newspapers on certain events, a tool that can be used to further the topic of press freedom, propaganda and persuasion. Whether it is necessary for a body such as "Reporters sans frontier" to draw up a ranking of countries with more or less freedom of

information based on instances of censorship, threats and persecution, a computational analysis that follows the method described in the thesis but performed on a larger and more diverse corpus could allow for a deeper understanding of the ways in which information is transmitted and would allow readers to have a clearer idea of the positions that defend certain newspapers and to be more aware of the economic or political interests that may lie behind the attempt to persuade readers of the existence or non-existence of a certain danger or enemy. In the history of Italy in the last century, there are several dark pages related to information: the P2, a Masonic lodge to which some of the most important political, military and Italian news personalities belonged, had attempted to manipulate Italian public opinion for several years probably even going so far as to help organize attacks that would contribute to the "Strategy of Tension". The strategy is based on a preordained series of terrorist acts aimed at spreading a state of insecurity and fear among the population, such as to justify, request or hope for authoritarian political changes; it can also be implemented in the form of a military tactic which consists of committing bomb attacks and attributing their authorship to others. The fact that one of the members of the P2 lodge, Silvio Berlusconi, owner of a large part of the Italian media, is still today one of the protagonists of the Italian government executive can allow us to understand why Italians have a generalized distrust towards institutions and media, distrust that creates the perfect environment for the development of anti-political and conspiratorial positions, and therefore for the development of fake news. However, the problem of fake news, anti-political sentiment, distrust of the media and more generally of propaganda is not limited to Italy, as witnessed today by the Ukrainian-Russian war. Putin has abolished the word "war", whoever pronounces or writes it, in schools, in real or virtual squares, is punishable by prison; so that the Russian pacifists, to circumvent the repression, demonstrate with unwritten placards and banners. The attempt by the Russian government is to justify the war of aggression against Ukraine with expressions such as "peacekeeping", "special or military operation" or regime change. For the EU, NATO and the US, instead, "peace is war". They say they want peace, but war is being prepared, indeed it is being done vicariously through Ukraine, with the sending of mercenaries and weapons, with economic and financial sanctions, and finally with the propaganda abuse of history²⁸. An analysis that starts from the words used to represent an overview of the emotions that a certain political faction wants to associate with a certain topic is crucial in a political environment in which even the main actors are so aware of the power of words as to limit their use.

²⁸ Martelli, 2022.

From my personal experience, when I talk about Digital Humanities to someone who has never heard it before, the reaction is full of amazement. And this is due to the fact that in the common imagination (and also in reality) information technology is mainly associated with the dynamics of capitalist profit, or with entertainment, but never with philosophy and critical thinking. Industrial society has allowed scientific progress and the development of modern technologies, but due to its capitalist nature it has directed them towards profit and not towards the improvement of the human condition. I do not in any way want this thesis to be associated with a post-modernist trend of criticism of capitalism because the political debate is not the center of the thesis but I would like to emphasize on the fact that social criticism is one of the foundations of humanistic reflection and that a field such as Digital Humanities must first of all devote itself to social analysis and the dangers in which freedom of critical thought runs today, and only then worry about digitizing museums, archives and exhibitions.

Highlighting the theoretical relationships between the philosophy of language and computer science and starting to develop social analysis tools that allow for the enrichment of philosophical reflection is, in my opinion, the starting point for allowing the humanistic disciplines to acquire the lost authority and mitigate the overwhelming power of multinationals which, although they are starting to hire humanists to demonstrate the company's ethical commitment to better sell their image, will increasingly use IT resources to enrich themselves to the detriment of the mental health of billions of increasingly users dependent from their devices and unaware of the increasingly sophistication of the technologies with which information is presented to them.

Propaganda is the price to pay for a democratic society, like explains Edward Bernays in his book "Molding public consent":

Every one is a propagandist for some platform, and it is the freedom with which all may employ the methods of propaganda that makes for safety and stability in a democratic country. No man or group of men in this country has ever had a monopolistic control of methods or media. There are no patents by which one can control the dissemination of opinion. There is no coercion by which any man can prevent any group of free adult people in America from opposing ideas disseminated by others.²⁹

- *Edward Bernays*

²⁹ Bernays, 1935.

If certain mechanisms were already evident in the eyes of Bernays in 1935, the fragmentation and polarization characterizing today's information makes even more evident the mechanisms of economic and political interest that underlie the decisions of the media to expose this or that news, to expose it with these or those tones. For this reason it is necessary to better understand the mechanisms of today's propaganda, and it is necessary that ordinary people have the tools to interpret the intentions of those who tell them what happens in the world every day.

The responsibility for developing these tools is a pride and a burden in the hands of the Digital Humanities.

BIBLIOGRAPHY

- Unsworth, J., Schreibman, S. 2004. A companion to digital humanities. *Blackwell Publishing*.
- Rushkoff, D. 2010. Program or be Programmed: Ten Commands for a Digital Age Paperback. *Soft Skull Press*.
- Durlofsky, P. 2020. Logged In and Stressed Out. *Rowman & Littlefield Publishers*.
- Rutkowski, S. 2018. Emotional and Cognitive Overload: The Dark Side of Information Technology. *Routledge*.
- Kalpp, O. E. 1986. Overload and Boredom: Essays on the Quality of Life in the Information Society. *Greenwood Press*.
- Simon, H. 1957. Models of bounded rationality. *The MIT Press*.
- Toffler, T. 1970. Future Shock. *Bantam*.
- Sambrook, R. 2018. The Future of 24 Hour News. *Peter Lang*.
- Hawking, S. 2020. Brief Answers to the Big Questions. *Hachette Collections*,
- Parser, E. 2018. The Filter Bubble: What The Internet Is Hiding From You. *Penguin*.
- Platone, Repubblica. Edited by M. Vegetti. 2007. *BUR*.
- Bernays, E. 1935. Molding Public Opinion. *IG Publishing*.
- Wu, T. 2016. The Attention Merchant. *Alfred A. Knopf*.
- Holiday, R. 2012. Trust me, I'm lying. *Portfolio*.
- Le Bon, G. 1895. Psychologie des foules. *Alcan*.
- Cialdini, R. 2007. The psychology of persuasion. *Collins Business Essentials*.

Articles

- Scarfone, G. 2017. Giornalismo e social media: un'analisi linguistica. *Unimi*.
- Brunato, G. 2017. A study on linguistic complexity from a computational linguistics perspective. *Università di Siena*.
- Murphy, P. A. 2017. When Alan Turing and Ludwig Wittgenstein Discussed the Liar Paradox. *Cantor's paradise*.
- Gillies, D. 2007. Logicism and the Development of Computer Science. *King's College London*.
- Liao, Y., Bright, G. 1991. Effects of computer programming on cognitive outcomes. *Sagepub*.
- Duff, A. S. 2016. Daniel Bell's theory of information society. *Sagepub*.
- Martelli, M. 2022. La propaganda di guerra e l'abuso della storia. *Micromega*.
- Reporters sans frontieres. 2022. Press Freedom Index. *Reporters sans frontieres*.
- Michel, O. 2017. How social media has changed journalism. *Irish Tech News*.
- Wagner, K. 2017. Three major ways social media is changing journalism. *Micromega*.
- Gottfried, J., Mitchell, A., Jorkvitz, M., Liedke, J. 2022. Journalist say social media helps at work, but most decry its impact on journalism. *Pew Research Center*.
- Bunz, M. 2009. How social networking is changing journalism. *The guardian*.
- Millman, A. 2022. Opinion: Journalism in the digital age — has social media changed the industry? *The Review*.
- Danisch, R. 2022. 7 ways to spot polarizing language — how to choose responsibly what to amplify online or in-person. *The Conversation*.
- Flora, M. 2022. I social contro la propaganda. Ma è la scelta giusta? *Formiche.net*.
- Martelli, M. 2022. La propaganda di guerra e l'abuso della storia. *Micromega*.

Movies

- Carpenter, J. 1988. They Live. *Universal Pictures*.
- Orolosky, J. The social dilemma. *Exposure Labs, Agent Pictures, Argent Pictures, The Space Program*.