Università degli Studi di Modena e Reggio Emilia

DIPARTIMENTO DI STUDI LINGUISTICI E CULTURALI

DOTTORATO IN SCIENZE UMANISTICHE - XXXIV CICLO

DIGITAL HUMANITIES AND DIGITAL COMMUNICATION

Social Media Analysis using word embedding: Exploring echo chambers and filter bubbles

Dottorando:

Leonardo Sanna

Tutor:

Prof.ssa Marina Bondi

Università di Modena e Reggio Emilia

Co-tutor

Prof. Dario Compagno

Université Paris Nanterre

Anno Accademico 2020/2021

SOMMARIO

Introduction ..... - 3 -
1 Social Media Analysis ..... - 6 -
1.1 The last decade of Social media Analysis ..... - 6 -
1.1.1 The beginning: from virtual to digital, the need of new methods ..... - 7 -
1.1.2 Social Networks, user generated content: social media analysis in 2010-2015 ..... - 10 -
1.1.3 The fake news debacle: the pessimistic turn (2015-2020) ..... - 12 -
1.2 Echo Chambers and filter bubbles ..... - 15 -
1.2.1 Echo chamber: redefining the problem ..... - 16 -
1.2.2 Filter Bubble ..... - 24 -
1.3 Research questions ..... - 27 -
2 Methods ..... - 30 -
2.1 Introduction ..... - 30 -
2.1.1 Reasons behind the methodological choices ..... - 30 -
2.2 Word Embedding ..... - 31 -
2.2.1 What is a vector ..... - 33 -
2.2.2 How does word embedding work ..... - 34 -
2.2.3 Word2vec. ..... - 37 -
2.3 Topic modelling and keywords ..... - 43 -
2.3.1 Topic modelling choice: why DHC and not LDA ..... - 44 -
2.3.2 Corpus-assisted tools: keywords ..... - 46 -
2.4 The Appraisal framework ..... - 47 -
2.4.1 Introduction to appraisal theory ..... - 48 -
2.4.2 Corpus-driven approaches to appraisal ..... - 53 -
3 Filter Bubbles ..... - 58 -
3.1 Introduction ..... - 58 -
3.2 Facebook ..... - 58 -
3.2.1 Experiment ..... - 62 -
3.2.2 Discussion ..... - 70 -
3.3 YouTube ..... - 72 -
3.3.1 YouTube Tracking Exposed ..... - 73 -
3.3.2 Approaching the filter bubble on YouTube. Does it exist? ..... - 80 -
3.3.3 Evidence of filter bubbles ..... - 81 -
3.3.4 Experiment on the American elections ..... - 86 -
3.4 Conclusions ..... - 95 -
4 Echo chambers ..... - 97 -

4.1 Introduction - 97 -
4.2 Corpus analysis: Preliminary overview - 99 -
4.3 Analyzing *hoax* in the news - 102 -
4.4 *Hoax* on Twitter - 106 -
4.5 Inside the inferential path: the embeddings - 110 -
4.6 Echo chambers and appraisal: looking for dialogic contraction - 115 -
4.7 Conclusions - 141 -
5 Conclusions - 145 -
5.1 Research questions recap - 145 -
5.2 Filter bubble - 146 -
5.3 Echo Chambers - 153 -
5.4 Methods - 160 -
References - 164 -

INTRODUCTION

In this thesis, we present a methodological proposal for a semio-linguistic approach to social media analysis. The nature of this work is inherently experimental, meaning this term in both the most strictly technical but also philosophical sense. The thesis will present in the central chapters (3-4), experiments on the study of filter bubbles and echo chambers, experiments that would have the ambition to pose as a first attempt for a structured linguistic approach of the two phenomena.

The topic of social media has become central during the past decade. After the so-called "Web 2.0" era, the world wide web has evolved to something that has never existed before on this scale. In 2022 Facebook counts 2,9 billion users and on You Tube 500 hours of new videos are uploaded every minute[1]. With this dramatic data production new professional figures have been created to analyse this huge amount of information and new powerful methods have been developed to investigate and process what is called "Big Data". These data are what Celli (2016) called "human data", data that are produced, more or less consciously, by human beings during online activities. Hence, the comprehension of digital spaces like social media is not possible without collecting and analyzing these huge quantities of data.

Nevertheless, the contributions of the humanities to the analysis of these very large quantities of data are still quite scarce. Most studies about social media have been realized by social scientists, within a framework that is also known as "computational social sciences" but linguistics, philosophy, semiotics and other humanities disciplines still miss important contributions in the study of social media. This fact is a paradox, because Big Data are complex and their interpretation cannot overlook the analysis of the socio-

[1] For Facebook active users see Statista, for YouTube statistics see Brandwatch

cultural context in which they are generated, nor a semantic analysis of their content. In other words, computational and statistical methods are not enough to fully understand the human use of digital platforms. However, it is also true that Big Data have characteristics like enormous dimensions, dynamicity, interactivity and inconsistency (Zikopoulos et al. 2012); these features are a real challenge for human sciences' methodologies and they make necessary to develop an interdisciplinary methodological effort. Even disciplines with an analytical vocation, struggle to produce literature on these issues. For instance, the semiotics of the new media has been studying signification over the world wide web since its foundation; one of the hottest topics has been interface analysis (Cosenza 2014), especially evaluating the usability of websites and their interactivity (Adami 2013; Derboven et al. 2012; Nazrul, Tétard 2014) collaborating with other disciplines like cognitive ergonomics. Nonetheless, the literature on the analysis of digital platforms is still quite limited, due to methodological constraints facing very large quantity of data.

We will try to overcome this methodological difficulty in the five chapters of this dissertation. In chapter 1, we start by presenting the latest ten years of Social Media Analysis and its approaches to the digital. Then, we introduce the concepts of echo chamber and filter bubble, which constitute the focus of this work. The research questions, presented in 1.3, will focus on how to study and measure these two phenomena, using a data-driven approach to further develop their theorization.

Chapter 2 explains in great detail methods and technology used in chapters 3-4. In particular we illustrate the functioning of word embedding, a machine learning technique that we are using to enhance the qualitative analysis of the pragmatic dimension of our corpora.

Chapter 3 and 4 present the experiments on filter bubbles and echo chambers. Filter bubbles were tested on Facebook and YouTube, while echo chambers on Twitter. In particular, section 3.2 experiments the use of word embedding as a qualitative tool, which we believe it is one of the most important proposals of this thesis.

Finally, chapter 5 presents the main theoretical advancements that we may draw from our experiments, also illustrating the main methodological innovations and limitations.

1 SOCIAL MEDIA ANALYSIS

1.1 THE LAST DECADE OF SOCIAL MEDIA ANALYSIS

The social media platforms that we are used to nowadays were born more or less a decade ago. Facebook, Twitter, YouTube and other popular social media were born in the mid '00 but they were something else.

In our opinion, it is important stress immediately the difference between the expressions *social media* and *social networks*. Although in everyday language these terms are mostly used as synonyms, they refer to two different things. The expression *Social Network* focuses, indeed, on the social groups and their connection and interaction on the digital platforms. On the other hand, *Social Media* refers to the mediatic use and effects of digital platforms. Another important difference between these two expressions is that we can still find and analyze a social network also without digital platforms, while we cannot have social media without Facebook, Twitter, You Tube etc. In this work we are interested in the mediatic aspect of digital platforms, although sometimes it is inevitably blurred with social dynamics. Therefore, we will use uniquely the term *social media*.

Nonetheless, we can see how the use of the term social media has evolved especially at the beginning of last decade, around 2011-2012, when it became clear that those digital platforms were something more than an aggregator for old friends. The first years of social media could be defined as an enthusiastic era, reenacting what had already been described by Umberto Eco (1964) well before the birth of social media and Internet, as a division between enthusiasts - also well before the social media rise (Kerckhove 1997; O'Reilly 2007) - and a minority of rather apocalyptic critics (Lovink 2003, 2012; Morozov 2011).

We might say that this era, focused mostly on the positive aspects of social media and on their positive impact on society, begins with the foundation of social media and it ends around 2015, with the so-called "fake news debacle" (Rogers, 2018). To better understand these two radically different periods we have to go back at the origins of social media. At that time, the optimistic views on the web were far more popular than the pessimistic. The definition of *web 2.0* as a place where finally the distances between the audience and medium ended is in fact renowned, overused and misused. Communication scholars talked enthusiastically about *prosumers*, defining them as the new audience: everyone on the web produces and consumes content at the same time and this was seen as an epochal revolution, with almost only positive consequences.

Indeed, by the time we are writing this thesis (2021), it is still true that on digital platforms we are, potentially, producers and consumers at the same time. However, it is certainly not true that a mediatic hierarchy does not exist any longer. Conversely, the content creators that gained success during the past decades are by far at the top of the pyramid, while most of the users will remain anonymous (or at least not famous) for their whole life. Still, social media were for sure a game changer at their foundation, as they created a digital space in which the old mediatic relations were effectively subverted. This transformation, however, created new digital-born power dynamics in a process of remediation (Bolter and Grusin 1999). Nowadays in fact we can still observe old and new mediatic dynamics on digital platforms (e.g., influencers and content creators that effectively have more mediatic power than common users).

1.1.1 The beginning: from virtual to digital, the need of new methods

During the enthusiastic web 2.0 era, there was the dichotomy of virtual and digital, an opposition between the online world and the offline reality. An important contribution

for the evolution towards the concept of digital has been given by Floridi with his famous concept of *onlife* in 2013 (Floridi 2015). For the sake of truth, the definition of *onlife* in was linked by Floridi to the idea of "Web 6.0", a sort of final stage of the web 2.0 project, which Floridi himself said was "ill-defined" although very promising.

This concept by Floridi then becomes clearly linked to the perception and narration of the self, which according to the philosopher is closely linked with another characteristic element of the digital turn, namely the infosphere (Floridi 2014).

Infosphere is a neologism coined in the seventies. It is based on 'biosphere', a term referring to that limited region on our planet that supports life. It is also a concept that is quickly evolving. Minimally, infosphere denotes the whole informational environment constituted by all informational entities, their properties, interactions, processes, and mutual relations. It is an environment comparable to, but different from, cyberspace, which is only one of its sub-regions, as it were, since the infosphere also includes offline and analogue spaces of information. Maximally, infosphere is a concept that can also be used as synonymous with reality, once we interpret the latter informationally. In this case, the suggestion is that what is real is informational and what is informational is real. It is in this equivalence that lies the source of some of the most profound transformations and challenging problems that we will experience in the near future, as far as technology is concerned. (Floridi 2014 pp- 40-41)

Rogers (2009) also wrote about the end of the virtual, with a crucial work that stated 1) the end of the dichotomy real/virtual in the Internet studies 2) the need of having a specific framework to do research on digital media, namely "digital methods".

The main difference between virtual and digital is that a virtual world is completely separate from real life, in every aspect. What happens in a virtual world has no effect on the reality. Conversely, the digital world is integrated in our everyday life, what happens online has effects offline and viceversa. The evolution from virtual to digital was crucial for the development of social media analysis because it allows to rethink social media both as an integral part of human reality and as a specific feature that requires specific methods of investigation.

Digital methods are an extensive and multidisciplinary framework to approach digital-born data (Rogers 2013). During last decade this discipline has been dealing with a variety of case studies mainly in the area of social sciences (Rogers 2010, Rieder 2012, Borra et al. 2014, Rieder et al. 2015, Helmond et al. 2017, Rogers 2017, Nielborg et al. 2019). Even though digital methods are mainly developed by social scientists, the framework is epistemologically broad enough to work, with the necessary adjustments, in every discipline within the area of the humanities. Hence, we propose to use digital methods as the main discipline to take into consideration for methodologies of social media analysis.

On the other hand, computational linguistics is necessary to have the know-how on Natural Language Processing (henceforth NLP). Computational linguistics has a quite long tradition of social media analysis, using a set of different tools all structured on a corpus-based approach (Zappavigna 2011, Bamman et al. 2014, Aditya et al. 2015, Mewari et al. 2015, Apoorva 2017, Fang 2017, Yang et al. 2018). Recently there have been some attempts of including NLP in a semiotic analysis (Chartier et al. 2019) and also some research on the interaction between semiotics and quantitative methodologies (Ulanowicz 2002, Barnham 2015, and more recently also by Compagno 2018a)

Semiotics suffers these limits in its approach to objects, even if over the past few years there have been some attempts to deal with large digital corpora (Cosenza, Colombari, Gasparri 2016) and to reflect on user generated content (Ferraro, Lorusso ed. 2016). A few studies have tried to deal with theoretical matters (Kress 2009; Maggi 2014; Mirsarraf et al. 2017), while most studies have produced applied research based on text analysis (Peverini 2014; Finocchi ed. 2015; Bonilla 2015; Madison 2016;). The applied studies that have been carried out on social media are case studies on small datasets like the works of Marrone (2017) and Peverini (2017).

1.1.2 Social Networks, user generated content: social media analysis in 2010-2015

As we have already said, in the first part of the past decade, the expression *social media* was not common yet in social media analysis. Instead, research referred to social media with a variety of terms such as *social networking sites* (SNS) or, generally speaking, *new media.* In the first part of the decade, we might say that research on social media analysis was user-centered. In fact, the topics covered between 2010 and 2015 ranges from user interaction to online activism

Looking at the literature, we can draw an outline of three main thematic areas covered in this period.

1) User interaction. Intended both as the study of user interaction with social media and the use of social media as new communication means. (Marwick & Boyd 2011; Ellison, Steinfield & Lampe, 2011; Hasinoff 2012; Ruths, 2014)
2) Politics and social movements. This topic is still at stake social media research even nowadays, although with different perspectives. The studies on these themes analyse the use of social media by politicians and the use of social media to organize political movements and actions (Harlow 2012; Strandberg, 2013).
3) Brands and social media marketing. The first studies on how to use social media for marketing-oriented goal, ranging from corporate communication to actual digital-born marketing campaign. (Nah & Saxton, 2013)

We said that these studies are mostly user centered, meaning that they deal with social media as tools and means of communication. The perspective is always oriented to analysing what users can do with social media, intending users in its broadest possible meaning, thus including companies and politicians. Humphreys (Humphreys, 2010)

provides a fascinating example by analysing the possible impact of social media (called *mobile social networks*) on the negotiation of public space. The study is particularly interesting because it is conceptually seminal for current research. Using a qualitative social-science oriented approach, Humphreys analyzes the concept of public space in relation to social media.

> *Mobile social networks can help to turn public realms into parochial realms through parochialization. Parochialization can be defined as the process of creating, sharing and exchanging information, social and locational, to contribute to a sense of commonality among a group of people in public space. Sharing information through mobile social networks can help to contribute to a sense of familiarity among users in urban public spaces. (Humphreys 2010, p. 768)*

We can consider this a sort of anticipation of the theory of echo chambers that we will discuss later in this chapter. In fact, the concept of parochial space is basically a closed social group that lives and interacts on social media. In the study by Humphreys, differently from the present literature, this phenomenon is not considered a problem, nor a potential problem. Instead, the role of the social media is explicitly viewed as positive. The conclusive example is a case in point:

> *For example, perhaps Facebook can be understood as a parochial realm, that is, a site of familiarity and comfort because of the social relations found therein. (Humphreys 2010, p. 776).*

After eleven years it seems almost ironic to consider Facebook a familiar and comfortable online space. However, in a user-centered perspective this is still true even today, as the evolution of Facebook has shown. This concept of parochial space is crucial to understand (or at least theorize) the genesis of the phenomena that are at the stake of this work.

Regarding the second topic, digital activism, it is important to highlight that the interest around these themes was alive already in 2009 in social sciences (Hintz & Milan, 2009).

Specifically, the main interest was around marginalized actors and citizen journalism (Goode, 2009). This interest was possible at its maximum around 2013, when all around the world social media played, for the first time, a significant role in protests and rebellions (cfr Arab Spring) (Valenzuela, 2013).

The evaluation of their social impact on political participation was utterly positive across the literature at that time, especially regarding the alleged empowered dynamic allowed by digital platforms. The crucial part, to better understand the successive post-2015 shift, is that also in this case the center of the research are users, while social media are tools that seem to be used to empower offline dynamics. This might be considered an early-days approach to the digital, in which the peculiarity of the digital spaces was blurred and somehow flattened by the old offline research perspective. Of course, digital activism, as the name suggest, is a digital born phenomenon. However, in these years it does not emerge as that, instead it emerges as something that was already present offline, repressed, that is now coming to power thanks to social media.

This quick and surely partial review of social media studies from 2010 to roughly 2015 is not intended to show outdated themes. Instead, it was meant to contextualize what we will introduce in the next section and in section 1.4. The shift from a user-centered perspective to a platform-centered perspective does not come out of the blue. It has its roots in social science research on social media and their users.

1.1.3 The fake news debacle: the pessimistic turn (2015-2020)

What is commonly known as the fake news debacle is actually a series of event taking place on social media between 2015 and 2016, with a diffuse political success of populist political movements all over the world.

The most famous case is certainly Donald Trump's victory at the 2016 presidential elections, which was a shock not only in the U.S. context. In particular, for the first time the social media became, in mainstream media but even in academia, something somehow dark and dangerous. Fake news, trolls, echo chambers, polarization of the debate, these all are themes that emerged around 2016 and that intensified their presence in academic literature after Donald Trump victory.

If we had to summarize three main topics in this five-year period, we might say that they were:

1) The polarization of the debate
2) Hate speech online
3) Fake news and conspiracy theories

Regarding fake news, further studies confirmed this view, adding the fact the conspiracy-oriented content seems to live longer online, besides spreading more (Del Vicario et al. 2016a). The majority of studies on these themes focused on fake news detection (Conroy et al. 2015; Kucharski 2016; Perez-Rosas et al. 2017; Shu et al. 2019), in a variety of disciplines (Albright 2017; Vargo et al. 2018; Shu et al. 2020; Gray et al 2020; Melchior & Oliveira 2021; , Di Domenico et al. 2021).

Also hate speech, despite being a linguistic issue, has been approached by multiple research fields. The main topics in this area deals with the identification of hate speech targets (Silva et al. 2016; Ben-David et al. 2016; ElSherief 2018) and the quantification of hate speech on specific topics (Schmidt et al. 2017; Soral et al. 2018; Caldèron et al. 2020). Concerning more specifically linguistic phenomena, it is worth mentioning the theory of blocking (Langton 2018), that is a strategy of counterattacking an opposite standpoint by highlighting what was left implicit in their discourse, thus exposing the hateful position.

More broadly, linguistics has been dealing with the field of aggressive language and impoliteness since 2011 (Culpeper 2011, 2017, 2021; Tagg et al. 2017; Kienpointner 2018; Oliveira et al. 2020), extending this type of approach also to social media analysis (Oz et al. 2018, Kienpointer 2019, Teneketzi 2022).

For our goals, however, the most relevant topic is surely the polarization of the debate. The most important study for this area it is the renowned paper titled "Debunking in a World of tribes" (Zollo et al. 2017), which highlighted social media as a place in which ideological isolation is fostered and amplified by the platform itself. This of course is in opposition to the web 2.0 philosophy that saw the web as a friendly space in which users could coexist with their reciprocal differences and enrich each other's perspectives. The study presents a disturbing view in which it appears clear for the first time that, on social media, users interact only with people who share their same beliefs. In their experiment, the authors monitored two Facebook pages, one pro-science and the other openly against official science, which we should call conspirationist. The result of the study showed that the two communities did not interact with each other: people who believed in science only interacted with science content, whereas conspiracy believers only interacted with their own content. Actually, the study pays attention on both polarization and fake news, since the title addresses a specifical fake-news-related problem, i.e. debunking. According to the authors of the study, debunking seems to be useless as it has no impact on the people that is meant to persuade. This work has probably been the starting point for a renewed interest in echo chambers, moving from a theoretical level to an empirical analysis. Further research made in this area, namely computational social science, has confirmed these first findings, showing also that conspiracy discourse tends to last more and spread easily online (Del Vicario et al. 2016b, Zollo et al 2017, Fernandez & Harith 2018).

Nonetheless, in linguistics and semiotics few studies have dealt with this topic. Besides the aforementioned issue of hate speech, linguistic research on social media has mostly been focusing on a variety of approaches to topics closely connected with echo chambers, such as studies on political discourse (Burgers et al 2019, Breeze 2020), toxicity in online debates (Pascual Ferrà 2021) and stance taking on Twitter (Cotfas et al. 2021). However, in all these studies, echo chambers are (more or less explicitly) in the background but not the central point of the research. On the other hand, when the echo chamber is more central (Demszky et al, 2019; Bliuc et al. 2020), the research always entails a computational social science approach, namely the study of online interaction among groups.

We believe that, to design a methodological framework, we should have in mind the linguistic dimension of the polarization dynamics. The methodological and theoretical frameworks must be interdisciplinary of course but we need to make more clear the linguistic contribution on these themes. It should be noted that interest towards social media is rising in linguistics, with important contribution also on the methodological aspects (Rüdiger and Dayter 2020) and on the investigation of specific social media affordances, such as hashtags (Zappavigna 2011, 2015; Zappavigna and Martin 2018).

In the next section we try to draw a theoretical outline for linguistic research on these themes. Specifically, we will attempt to have a linguistic theorization of echo chambers, while we will clarify some crucial aspects on the filter bubble theory, illustrating the main methodological challenges that we must face to study it.

1.2 ECHO CHAMBERS AND FILTER BUBBLES

Since the start of the last decade, filter bubbles and echo chambers have become quite popular in social media studies and in a variety of academic disciplines. The two

phenomena are sometimes blurred and confused, as the two terms as used as synonyms. Instead, the difference between the two is crucial. The main difference between echo chambers and filter bubbles is that the former is an active phenomenon, while the latter is fundamentally passive. Echo chambers are studied by analyzing how communities interact with social media content, whereas filter bubbles are studied by analyzing what the algorithms select for the users. We might say that echo chambers are built by users while filter bubbles are made by algorithms (Zimmer et al. 2019). In the following paragraph we detailed the studies on the phenomena and their relative conceptualizations.

1.2.1 Echo chamber: redefining the problem

The first definition of echo chamber is the one of Sunstein (2007), although it is not a proper definition. In his book, Sunstein introduces and widely discusses a problem with the web and the newborn social media (or their ancestors). The point that Sunstein makes is that the interaction with technologies that personalize content is at risk of putting users in what he called *information cocoons.* In this perspective, the problem is the interaction with users and technology.

According to Sunstein in fact, the Internet changed the rules of communication and group interaction. On the Internet it is in fact very easy to communicate and, at the same time, it is very easy to reach a vast amount of content that we are interested in. This can be seen as a problem for polarization:

Group polarization is unquestionably occurring on the Internet. From the evidence thus far, it seems plain that the Internet is serving, for many, as a breeding group for extremism, precisely because like-minded people are deliberating with greater ease and frequency with one another, and often without hearing contrary views. (Sunstein 2007 p. 69)

This statement by Sunstein seems very topical today, but it was written almost ten years before group polarization became an issue discussed in social media studies. The fact that Sunstein includes aspects of interaction with features of the technology is probably the reason why echo chambers and filter bubbles are often confused. However, we must keep in mind the context in which Sunstein wrote his book. At that time, algorithmic personalization was at its very beginning. In fact, the only social media platform mentioned is YouTube. In fact, in Sunstein's work there is also a sketchy bit of critique of algorithmic personalization by talking about Amazon and Netflix, but it remains a secondary theme and left in the background of the introductory chapter.

Sunstein made it noticeably clear that echo chambers are, actively, made by users. One of the examples that is brought into the discussion of the echo chamber effect is Google News, that is for sure a content aggregator and provides a personalized experience but with very different degrees of algorithmic personalization compared to social media like Twitter, Facebook or YouTube. Nowadays Google has indeed enough data on us to personalize our Google News newsfeed, but the Google News is just a selection of news, indeed; the peculiarity of algorithmic personalization is the selection of likeminded content along with likeminded context, such as comments, group suggestions, video recommendation.

Sunstein's perspective on YouTube is also still user-oriented:

YouTube is a lot of fun, and in a way it is a genuine democratizing force; but there is a risk that isolated clips, taken out of context, will lead like-minded people to end up with a distorted understanding of some issue, person, or practice. (Sunstein 2007 p. 69)

The focus is undoubtedly on users' interaction; Sunstein is worried that "out of context videos" might have a negative effect on some like-minded community. Instead, as we will explain in the next section, in a filter bubble perspective the problem is the context and the fact that this context is provided by a non-transparent algorithmic selection.

Besides our personal interpretation of Sunstein's work, in the book there are two passages that explicitly talk about "creation", in relation to echo chambers.

It is entirely reasonable to think that something of this kind finds itself replicated in the blogosphere every day. Indeed some bloggers, and many readers of blogs, try to create echo chambers. Because of self-sorting, people are often reading like-minded points of view, in a way that can breed greater confidence, more uniformity within groups, and more extremism. (Sunstein 2007 p.145)

The Internet is hardly an enemy here. It holds out far more promise than risk. Indeed it holds out great promise from the republican point of view, especially insofar as it makes it so much easier for ordinary people to learn about countless topics, and to seek out endlessly diverse opinions. But to the extent that people are using the Internet to create echo chambers, and to wall themselves off from topics and opinions that they would prefer to avoid, they are creating serious dangers. And if we believe that a system of free expression calls for unrestricted choices by individual consumers, we will not even understand the dangers as such. (Sunstein 2007 pp. 222-223)

Going back to the definition of echo chamber, as we said before, also Sunstein does not provide a clear-cut definition of the phenomena. To simplify, we might say that Sunstein's echo chambers are basically informational cocoons, using his words.

Recently, Nguyen (2020) draws an epistemological distinction between echo chambers and divergence of opinion.

Loosely, an epistemic bubble is a social epistemic structure in which some relevant voices have been excluded through omission. Epistemic bubbles can form with no ill intent, through ordinary processes of social selection and community formation. We seek to stay in touch with our friends, who also tend to have similar political views. But when we also use those same social networks as sources of news, then we impose on ourselves a narrowed and self-reinforcing epistemic filter, which leaves out contrary views and illegitimately inflates our epistemic self-confidence. An echo chamber, on the other hand, is a social epistemic structure in which other relevant voices have been actively discredited. (Nguyen 2020 p.142)

In this perspective the active dimension of echo chambers becomes clear once for all, in terms of their being user-driven distortions. On the one hand we have epistemic bubbles, that we might want to consider as normal forms of homophily, while on the other hand we have echo chambers. This view of echo chambers as dysfunctions of mass media

communication is also shared in computational social science, as we saw in section 1.1.3 (Del Vicario, et al., 2016b, Zollo et al., 2017, Di Marco et al. 2021) where echo chambers are structures that foster the polarization of debates and the spreading of misinformation (Gallacher 2009, Törnberg 2018). Even Sunstein refers to echo chambers as a problem for democracies, outlining their problematic facets.

Yet, it should be noted that the tendency to prefer information that confirms existing beliefs and to seek aggregation with likeminded individuals is definitely not new. Looking at the literature, we can go back to 1960, with selective exposure theories in social psychology (e.g., Klapper 1960). Perhaps for these reasons, some recent works have questioned the concept of echo chamber saying it is a product of academic theories rather than a social reality (Dubois & Blank 2018, Bruns 2019).

Indeed, Dubois and Blank make an important point, showing that people access information on different types of media, while echo chambers are studied only in one specific platform at time.

Focusing on a single medium may not give us useful information about how political information flows across offline media and other online media. (Dubois & Blank 2018: 730).

However, this is a problem of interpretation rather than conceptualization. Empirical studies on echo chambers contextualize their work around a specific platform to ensure reproducibility of the experiment. It is in fact true that every digital platform has its specific affordances and therefore its own definite effects on social interaction. None of the studies made on single platforms should be generalized to other platforms prior to realizing cross-platform research. So, it is true that traditional studies on echo chambers do not provide any information about the flow of information across different types of media.

Bruns (2019) also contested:

From this whole-of-system perspective, then, it appears exceptionally unlikely that ordinary social media users would find themselves entirely enclosed in connective echo chambers or communicative filter bubbles, even if they actively pursue homophilous connections with like-minded others in the context of specific interests or activities: on the mainstream social media platforms themselves, and even more so across the contemporary media ecology as a whole, the forces of context collapse in a complex and thoroughly interconnected mediasphere are simply too powerful. (Bruns 2019. P. 8)

It should be noted that Dubois and Blank did not use computational social science methods to question the existence of echo chambers: they used surveys, while Bruns' work is a review on the theoretical aspects. Conversely, empirical studies on echo chambers have the ability to capture live user interaction and to compare that with the chosen opposite community. Surveys cannot capture live interaction, but they capture an orientation. Nonetheless, a phenomenon such as echo chambers requires data-driven research, methodologies that can capture and analyse what is happening on social media. It might be that a vast majority of users, represented in Dubois and Blank's study, is effectively out of echo chambers dynamics and still echo chambers could exist for some communities. Otherwise, even the users interviewed by Dubois and Blank might be in some echo chambers during their online lives and not being aware of that.

In fact, echo chambers are given by ideological isolation within social groups; everyone might be in an echo chamber even if they try to diversify its information sources. For instance, someone who follows ten types of different newspapers of different political ideologies, might still be interacting mostly with content that confirms some of his deepest beliefs or else he might engage very easily with content that disturbs his ideological values. For example, they might be against Brexit and thus actively

commenting on news that discredit Brexit and meanwhile also actively commenting and engaging on pro-Brexit news or with pro-Brexit comments.

However, the study of Dubois and Blank surely imposes a reflection on the concept of echo chamber, since in computational social science this is always associated with exaggerated forms of isolation, for the sake of reproducibility. On the other hand, we have to disagree with the edgy conclusions of Bruns who refuses to see echo chambers as a valid metaphor for social media studies. We can discuss the nature of echo chambers on a purely conceptual level, but to disprove their existence we need empirical evidence.

Within the current paradigm of communication and information science, we observe increasing attention to echo chambers (Edwards 2013, Guo et al. 2015, Jacobson et al. 2016, Duseja and Jhamtani 2019, Calderón et al. 2019). While using different methods and having different goals, most of these studies tend to accept the current definition of echo chamber as an anomaly in the communication process.

In linguistics, the research on echo chambers is quite limited, so the understanding of their linguistic dimension is scarce. The effects of debate polarization on discourse and language use are still definitely unexplored. Discourse studies have shown a long-standing interest in the representation of ideology, but their attention has mostly focused on representation of conflicting positions in the news. A good example of this line of research is for instance the work on Critical Discourse Analysis (Van Dijk 1998; Wodak 2002; Fairclough 2010) and corpus-based discourse studies (Baker 2006).

Nevertheless, from a linguistic perspective, we should not take the dysfunctional dimension of echo chambers for granted. As we have written above, empirical studies on echo chambers sacrifice a part of social media complexity in order to be reproducible, e.g. considering only two different opposing factions. Hence, their dysfunctional aspect might

be something that is amplified by the methodological approach, as of course no one lives in completely isolated informational systems.

This approach generates perhaps a bias, since we observe polarized debates in specific digital contexts (e.g., Conover et al 2011 on Twitter) or in selected communities of interest (e.g., van Eck et al 2021 on the Climate Change blogosphere). However, this polarization might also not be digital borne, namely it might be caused by other reasons rather than social media affordances. It remains an open question whether we are observing an offline phenomenon that has moved online, and thus is now more visible, or if we are inquiring a specific social media effect on mass mediatic communication.

Moreover, from a semiotic perspective, the concept itself of debate polarization considers a binary relationship that is often overlooked in studies on echo chambers. For instance, it is reasonable to imagine that the discursive strategies of opposing factions may be specular, meaning that echo chambers are nothing more than a discursive manifestation of semantic contradictions.

If we define the echo chamber as the strong rejection of the opponent's ideologies along with the strong adherence to an inherited cultural space, we should take into account both positions and look at echo chambers as dichotomic controversies. For instance, both the opponents and the promoters of vaccination campaigns can end up in an echo chamber, i.e. the irreducible beliefs that the counterpart is completely wrong and that it should be (semiotically) annihilated. Let us take a concrete example in the following two sentences.

(1) Vaccines are crucial to fight COVID-19 and they should be mandatory

(2) Vaccines are dangerous and mandatory vaccinations are criminal

If we take the definition of Nguyen (2020), or what we can grasp out of the empirical studies on echo chambers, it is almost too easy to consider (1) the normality while (2) as a product of echo chambers, such as people who do not believe in science.

However, if we look at the implicatures (Grice 1975) both of these sentences carry a strong adherence to a belief (vaccines are safe/vaccines are not safe) and conversely a strong rejection of those who believe the opposite. This strong rejection includes a determination to fight the counterpart. From a linguistic perspective, therefore, it is hard to say that one of these sentences clearly represents an echo chamber while the other does not. It is even worse if we move out of the anti-scientific conspiracy theories:

(3) ESM is crucial for Italy development. Those who do not agree are ignorant

(4) ESM is dangerous for Italy. Those who want to take it are EU-slaves

Now, in this case, we have no explicit proposal of elimination of the counterpart and yet both sentences are completely "closed", there is no possibility to acknowledge the opposite point of view, instead the opposite side is attacked with some classic ad hominem argument. Is this likely to be part of an echo chamber or just a normal divergence of opinion? It is unclear, taking into account the sole discoursive dimension, whether we have a manifestation of an echo chamber or just disagreement. We cannot agree on everything and surely every individual in the world has some non-negotiable values in her cultural background.

We propose then to soften a bit the concept of echo chambers, considering them as dichotomic structures that imply ideological conflict. More precisely, they are indeed ideological structures (Eco 1968) that are observable when ideological conflict occurs (Rogers 2018b).

The aim of this work is to analyse the echo chambers' linguistic dimension. Buder et al. (2021) showed that we may find symptoms of polarization within the use of language. Their work, focused on the use of negatively toned language on Twitter, proving a link between negativity and polarization. They evidenced how users' individual negativity is the most influential towards polarization; we may then argue that this finding suggests that we can observe (and perhaps predict) polarization following textual patterns rather than social interaction. However, in their work Buder et al. did not attempt to clarify the linguistic dimension of echo chamber, although they provide an extensive review on the latest discussion on this topic.

Hence, following their conclusions, we will endeavor a corpus-driven approach to echo chambers in chapter 4, trying to redefine echo chambers analyzing textual insights.

1.2.2 Filter Bubble

The filter bubble is definitely more complicated to study. The most important problem is that it has to do with non-transparent algorithms, making it quite difficult to have full comprehension of the phenomenon. Moreover, social media algorithms use such a vast range of criteria that it is difficult to set up an experimental environment that could be reproducible.

The theory of the filter bubble was born with Eli Pariser's book (2011) and since then, to the best of our knowledge, there have been no attempts of empirical research on its effects. What Pariser said, is basically that we live in a world full of personalization algorithms that select information for us. This is due to the information overload in which we live daily. If Netflix, Spotify, Facebook or YouTube had no algorithm to select content for us it would almost be impossible to use them, since the amount of content that they

have is massive and constantly increasing. According to Pariser this service is not completely free of cost, as also stated by Fuchs (2010, 2013, 2018).

Personalization is based on a bargain. In exchange for the service of filtering, you hand large companies an enormous amount of data about your daily life—much of which you might not trust friends with. These companies are getting better at drawing on this data to make decisions every day. But the trust we place in them to handle it with care is not always warranted, and when decisions are made on the basis of this data that affect you negatively, they're usually not revealed. (Pariser 2011, p.14)

Differently from the echo chamber, which is a phenomenon that has been theorized in academia and then observed with a dedicate methodology, the filter bubble exists for sure. There is no need to theorize the filter bubble, contrary to what Bruns affirmed. The filter bubble exists, by definition, any time we use a personalization algorithm. The heart of the matter however is not the existence of the filter bubble but, again, whether this is a problem for us or not.

It is amazing to have technologies that can select content for us, choosing what is relevant for us. On the other hand, as Pariser says in the quote above, it is also a very delicate game of trust. On which basis can an algorithm decide what is relevant for us?

The short answer is that we do not know. Every social media algorithm is a black box that is constantly evolving, and some social media are so big that perhaps even the computer scientists working on them are not in full control of all the variables. Nonetheless, we have some general ideas on the principles on which social media algorithms work. They decide what is relevant for us looking at how we engage with content, thus proposing what is most engaging to us.

A very effective metaphor reported by Pariser (attributed to Danah Boyd), is that we are biologically evolved to seek for sugar and lipids because these elements are rare in our natural environment. Similarly, in our social environment we evolved to be attentive to things that stimulate us.

Our bodies are programmed to consume fat and sugars because they're rare in nature....In the same way, we're biologically programmed to be attentive to things that stimulate: content that is gross, violent, or sexual and that gossip which is humiliating, embarrassing, or offensive. If we're not careful, we're going to develop the psychological equivalent of obesity. We'll find ourselves consuming content that is least beneficial for ourselves or society as a whole. Just as the factory farming system that produces and delivers our food shapes what we eat, the dynamics of our media shape what information we consume. Now we're quickly shifting toward a regimen chock-full of personally relevant information. And while that can be helpful, too much of a good thing can also cause real problems. Left to their own devices, personalization filters serve up a kind of invisible autopropaganda, indoctrinating us with our own ideas, amplifying our desire for things that are familiar and leaving us oblivious to the dangers lurking in the dark territory of the unknown. (Pariser 2011,p.13)

However, in this work we are not interested in studying psychological effects of algorithmic personalization, instead we are interested in analysing whether or not there are issues at a linguistic level.

At the moment there is almost no literature on empirical studies on the filter bubble. Mostly, this is due to the fact that algorithms are not transparent and also because the empirical study of the filter bubble entails studying individual users' experiences. In particular these two aspects have an impact on methodology design, since it is particularly difficult to share a common methodology among different research groups.

In this context, the work of Tracking Exposed has been crucial. One first tentative experiment has been made by Tracking Exposed (TREX) research group [2] during the Italian elections in 2018[3]. The TREX group has as main goal to highlight (expose) user profiling and algorithmic personalization. In 2018 they made an experiment (Hargreaves et al. 2018a) during the Italian elections. To the best of our knowledge, this is the first empirical study on the Facebook filter bubble.

In the experiment they built six virgin Facebook profiles (henceforth "bots") to do some automatic data acquisition. Each profile automatically scrolled on his feed 30 posts per

[2] Official website https://tracking.exposed/

[3] Dataset available on Github at https://github.com/tracking-exposed/experiments-data

hour, collecting and parsing HTML code with the browser extension "Facebook Tracking Exposed"[4]. The extension was used to collect evidence about what was being shown to the bot.

To control for the algorithmic selection of the FB algorithm, all bots followed the same 30 different sources, covering the Italian political spectrum of that time (Hargreaves et al. 2018b). However, each bot was interacting (by liking) only with the content coming from one particular political segment. Therefore, we had a right-wing bot, a left-wing bot, a far-right bot, a far-left bot, a populist bot and finally an undecided bot that was not interacting with any content in its feed.

The dataset is hence composed of two main parts: the *sources* dataset, including the totality of posts made by the 30 pages (between January 10th and March 6th, 2018) and the *impressions* dataset, including only the posts that have been shown to the bots. In their 2018 study, the FBtrex researchers found evidence of uneven exposure to the sources of information, discovering that usually exposure was unbalanced towards the political bias of the bots. We will use this dataset in section 3.3 to start our text analysis on Facebook filter bubble. Although TREX experiments are quite new, as they experiment a novel approach, it is a good starting point to discuss issues and evidence of filter bubble effects.

1.3 RESEARCH QUESTIONS

We have reviewed the main lines on social media analysis research, also focusing on specific linguistic interests towards social media. As we shown, themes are quite varied and multidisciplinary. Nonetheless, the linguistic understanding of echo chambers and filter bubbles is still quite unexplored.

[4] More information on the extension here https://facebook.tracking.exposed/

This work is aimed to test an interdisciplinary methodological approach, experimenting new lines of exploration of these two phenomena within linguistics and semiotics. Therefore, our investigation covers both methodological and conceptual aspects of social media analysis.

For what concerns the methodological aspects we will:

1) Experiment a novel linguistic approach towards filter bubbles, trying to reproduce the effect of personalization within online discourse.
2) Experiment a corpus-driven approach towards echo chambers, trying to let them emerge directly from the text.

The research questions that we will try to answer within this experimentation will be covering various aspects of the two phenomena at stake of our investigation:

RQ A: existence

1) Filter bubble: is there evidence of filter bubbles?

As we have shown this is still debated, although social media do have for sure personalization algorithms. Specifically, we will search for experimental evidence of filter bubbles.

2) Echo chamber: is there semio-linguistic evidence of echo chambers and their alleged dysfunctionality?

We will look for echo chambers using a corpus-driven approach, namely observing if they are a visible effect on the textual dimension. In our opinion, this approach would remove possible bias due to selection of two opposed communities.

RQ B: measurability

1) Filter bubble: how can we measure the effects and impact of algorithmic personalization?

As we have said, we are interested in finding data-driven evidence of effects of algorithmic personalization. The second step would then be how to measure these effects within the linguistic dimension and to verify whether it is possible to make reproducible experiments.

2) Echo chamber: how can we measure the degree of severity of an echo chamber?

If we find linguistic evidence of echo chambers, we should be able evaluate the degree of their alleged dysfunctionality. In other words, we should be able to find corpus-driven metrics to measure the degree of polarization and, hopefully, some quantitative patterns that would guide qualitative exploration.

RQ C: conceptualization

1) Filter bubble and echo chamber: which new concepts and theorization could we have in semiotics for the filter bubble and echo chambers?

The ultimate goal of this work is to use the findings of our experiments to proceed in further semiotic theorization and conceptualization. This data-driven theory is necessary due to the complexity and volume of data to be studied. Without starting from solid empirical evidence, it becomes extremely difficult to think about advancing semiotic theoretical knowledge about social media.

2 METHODS

2.1 INTRODUCTION

In this chapter we review the methods used in this dissertation. In section 2.2, word embedding is introduced, explaining the theory and the design of this machine learning method. We also provide a basic introduction to neural networks and machine learning so that it is clear how the technology we are using works. For the goals of our work, we will use word embedding to create a semantic model of our corpora that will allow us for a qualitative exploration of the pragmatic dimensions of each word. In section 2.3 we introduce topic modelling and keywords, that we use to select the relevant word to investigate within our word embedding model. The use of topic modelling and keyword in facts enable us to a data-driven investigation of our data.

Finally, we introduce the appraisal framework in 2.4, which is a theory developed for a qualitative analysis of the dialogistic dimension. Along with this framework, in 2.5 we discuss the interaction between quantitative and qualitative methodologies, proposing an approach for an integration of the two.

2.1.1 Reasons behind the methodological choices

We will explain with great detail the functioning of word embedding in 2.2. Some general knowledge of Natural Language Processing is probably required to understand the most technical parts. We recommend the reading of Manning and Schütze (1999) and the latest edition of Jurafsky and Martin (2021) book "Speech and language processing"[5].

[5] The book has also a section dedicated to word embedding
https://web.stanford.edu/~jurafsky/slp3/6.pdf

However, it is important for us to clarify the reasons behind the choice of this technology. In a few words, we need a technology that would allow us to process a very large amount of data and at the same time would allow us to have a computational model of the semantic relationships built within our corpus. Word embedding is the technology capable of performing these two tasks. However, the aim of this work is not to test and experiment the best possible model of word embedding. We are proposing to use word embedding as a tool for qualitative exploration.

Our idea is that we can use machine learning to create a solid and data-driven semantic model to qualitatively explore our corpus at a pragmatic level. The semantic model built on top of word embedding is in fact a network of probabilities, which provides us an accurate snapshot of the pragmatic possibilities allowed by the semantic relationships within our corpus. So, we can compute the word “cat”, observe that the word “tree” is related to cat and following the inferential we may discover that in our corpus the word “cat” is primarily related to a frame of natural exploration.

Furthermore, we believe that is important to remark the added value of qualitative exploration. Word embedding knowledge of language is limited to data used to train the model, but the researcher can connect the dots that lies outside the dataset, enriching the understanding of the pragmatic dimension.

2.2 WORD EMBEDDING

Word embedding is a machine learning technique used in natural language processing to create a computational semantic model. With the term “semantic model” we refer to the ensemble of the semantic relationships that exist in a corpus. As we explained in 2.1.1, in this work we use word embedding to model the pragmatic moves suggested by the texts

in our corpora. In other word, we compute the inferential paths allowed within our corpora.

The theory on which word embeddings are based is distributional semantics. In this framework, the assumption is that similar words would occur in similar contexts, so that the semantics of each word is distributed across its occurrences. The very first theorization of this linguistic feature is actually quite old (Harris 1954), and it is known as the distributional hypothesis.

[…] if we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference of meaning correlates with difference of distribution. (Harris 1954 p.156)

In other words, we can think at the distributional hypothesis in mathematical term, saying that "the degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear" (Lenci 2008).

This intuition has been crucial for the developing of models such as word embedding. However, it is a long way from 1954 to 2013, that is the year in which word embedding became widely used and popular. In the same timespan there has been a second crucial theorization, which is what Sahlgren (2008) called the geometric metaphor of meaning. In this framework, the main idea is that we can represent words in a geometric space, a space in which similar words will be close to each other.

Just like the distributional hypothesis, this allows us to think of semantic similarity in mathematical term, calculating words relatedness (Budanitsky and Hirst 2006) quantifying the geometric distance between the two vectors representing them (Bruni et al. 2014). While referring to models such as word embedding, we often use the term of "Distributional Semantics Model" or indeed, more broadly "Vector Space Models".

2.2.1 What is a vector

The simplest possible definition is that a vector is a machine-friendly representation of a word. As human beings we can represent words in many different ways, by speaking, writing and sometimes even drawing. Very briefly, we might say that we learn language by associating sign and sounds to a meaning, so that words for us are always under graphical or phonetical form. For a machine is quite different, as every information must be under numerical form, since it must be represented in binary language. Hence, a computer thinks at words as a sequence of letters which are completely unrelated to meaning, until we have a mathematical representation of this connection. This representation is done using vectors.

Vectors are basic elements of Euclidean geometry; a vector is an object, in a geometric space, which has both a magnitude (i.e., a numeric value) and a direction. Hence, we can imagine the word "dog" having as random value [001] and pointing to some direction within a geometric space. For the sake of simplicity, we can just imagine vectors as if they were arrows, pointing towards the infinite in a specific direction, all starting from the same origin.

The similarity between two words is hence calculated looking at the angle that these two vectors create; the narrower the angle, the more similar the two words would be. In fact, we might say that the similarity between the two words is given by the direction onto which both are pointing in the space. Back to our example, the word "dog" is pointing in a direction that might be represented, in two dimensions, with coordinates 1;2; a similar word - e.g., "dachshund", a hyponym of dog - would be pointing in a very similar direction, maybe with coordinates 1:3.

However, it is more complex than that. Each vector has in fact as many dimensions as the number of words we have in our corpus, and this creates some computational

problems. Imagine having a corpus of one million words. Each of the words must have a unique computational representation in binary language, so that each word is associated to a sequence of bits. This type of computational representation is called *one-hot encoding* and it creates a digital object that is a sequence of digits with all zeros and a single one. Hence our first word would be [100000000…] our second word would be [0100000000…] and so on.

Yet, *one-hot* vectors have two big issues. First, each of the word must have as much digits as the words in our corpus and each word representation share no semantic information, meaning that is completely unrelated to the other words. This is the point where word embedding take the stage.

2.2.2 How does word embedding work

We said before that word embedding is a machine learning technique. In a nutshell, machine learning is any type of algorithm that allows a machine to learn from its experience and to use this knowledge to accomplish some given tasks. (Mitchell 1997). In the case of word embedding, the computer can learn the semantic model of a corpus by simply observing the co-occurrences. As we illustrated in the previous section, we have serious computational issues to have a machine-friendly representation of words and their semantics, therefore is not surprisingly that distributional semantics took almost 60 years to find a technological application.

The game changer was the ability to reduce the dimensions of each vector. The name "word embedding" indeed refers to the fact that the machine can represent each term as a *dense vector*, with no more that some hundreds of dimensions. Without getting into technical details, we might summarize saying that machine learning enables us to calculate semantic similarity using dense vectors (embeddings) instead of sparse vectors (one-hot vectors).

What is important for us to understand is the added value for a linguist. In a *one-hot* representation, even if I manage to have an infinite computational power, I have no information on the relationship among words. Just like in a wordlist, I only know that "dog" is equal to a vector, "cat" to another vector, "Bill Gates" to another vector and so on. Instead, the dense vector reduces the dimensionality of each word by also capturing its context, following the distributional hypothesis. Hence, with word embeddings we have a computational solution to represent a word *w* along with all its possible contexts *c* within the corpus.

Nonetheless, dense vectors are not sufficient to represent a semantic model, they are just the pieces that we need to build our semantic puzzle. Indeed, we need to train our machine, so that our embeddings would capture the semantic similarity among words.
To train our model there are two main ways:

1) Count-based model
2) Predictive models

Count-based models are perhaps the most old-school way to train a machine learning algorithm with word embeddings. In this framework, semantics is learned using a co-occurrence matrix, so that the relationship between two words is captured among co-occurrences. A renowned technique that relies on count-based dense vector representation is GloVe (Pennington et al. 2014). Of course, in a large corpus, our matrix would be huge. Hence, we use some techniques for dimensionality reduction such as PCA or SVD (Abdi 2007, Ringnér 2008), that allow us to visualize our model.

However, in this work we are using a predictive model, that is called word2vec (Mikolov et al. 2013)[6]. The main difference is that vector embeddings are not learned via

[6] Other predictive word embedding models, similar to word2vec. are *fastText* (Bojanowski et al. 2017), *Star Space* (Wu et al. 2018), *RAND-WALK* (Arora et al 2015).

a co-occurrence matrix but instead with a neural network. The reason of this choice is twofold. First, as shown in the literature, predictive models outperform count-based models (Baroni et al. 2014). Still, we might also say that performance is not crucial for our goals if we can still compute efficiently our model. However, the increased performance of predictive models also includes a crucial feature for inquiring pragmatics, that is the ability to capture semantic relations between words that rarely (perhaps never) co-occur together in our corpus. A said by Riedli and Biemann (2017) in fact, we should instead choose our embeddings according to our goals, rather than looking for mere performance.

To investigate pragmatics, it is crucial that the semantic model is built in a way that we can explore inferences in our model. It is also possibly more accurate to have a predictive model, rather than a count-based, as our semantic knowledge is not represented in a fixed and rigid scheme. Working on pragmatics, our semantic model in fact should be a representation of probability of co-occurrences, rather than actual co-occurrences. We are interested in capturing also weak relationships, namely those indirect semantic affinities that are necessarily built-in *absentia* within a corpus.

2.2.2.1 What is a neural network

Word2Vec is a machine learning technique that uses what is called a neural network, that is a computing system that is inspired to brain structures. Just like in a biological brain, in a neural network we have a series of connected artificial neurons. Each neuron is in fact nothing but a mathematical function that takes one or more input *z*. A neural network has usually at least three levels, called *layers*. The first layer, called *input layer* just receives its input. In word embedding, it is the neural level that receives our words to start the embedding process. Then we have at least one *hidden layer* that is where the model is actually learned. In this layer, the neurons apply their mathematical functions to transform the original input. In our case, this is where each word is embedded to a dense

vector. Finally, we have what is called *output layer* which is the level that is responsible for the final result. For word embedding this is where the probability of co-occurrence is calculated.

2.2.3 Word2vec.

Word2vec neural network is simple and powerful at the same time, as it uses only one hidden layer, calculating the probability of co-occurrence in the output layer. Word2vec actually comes in two different "flavors".

1. The Skip-Gram (SG): a neural network that learns to predict the context given a word. In this algorithm, words are fed to the output layer as a bag of word, thus regardless of their order.
2. The Continuous Bag of Words (CBOW): this second word2vec algorithm works in the opposite direction - starting from a context - with the goal of predicting a given word.

The input and the hidden layer are the same in both cases. The input layer is a vocabulary where each word is a sparse vector, while the hidden layer contains the weights, namely a matrix with n numerical values, usually between 50 and 500. These values are actually the column of the matrix, while we have one row for each word in our vocabulary. The number of weights is actually the number of dimensions that each of our embedding will have. Hence, a matrix with 100 columns would create dense vectors with 100 dimensions. In this layer we have thus the function that maps each sparse vector with a dense vector.

The difference between SG and CBOW, that will be covered in detail in following paragraphs, is actually in the way the output layer works. Each model is trained using a parameter that is called window. This specifies how many words should the neural

network take into account while calculating. What changes between our two algorithms is the way in which the probability is calculated. For instance, if we take the word *cat* and we put it through a Skip-Gram, with a window of 10, our neural network takes 10 random words and calculates the probability that each word has to occur in the context of *cat*. On the other hand, if we feed the word *cat* within a CBOW, the output layer will calculate the probability that *cat* has to occur in a 10-word context.

For the neural network this produces some changes in the input and in the output layer, as well explained in this illustration by Lilian Weng[7].

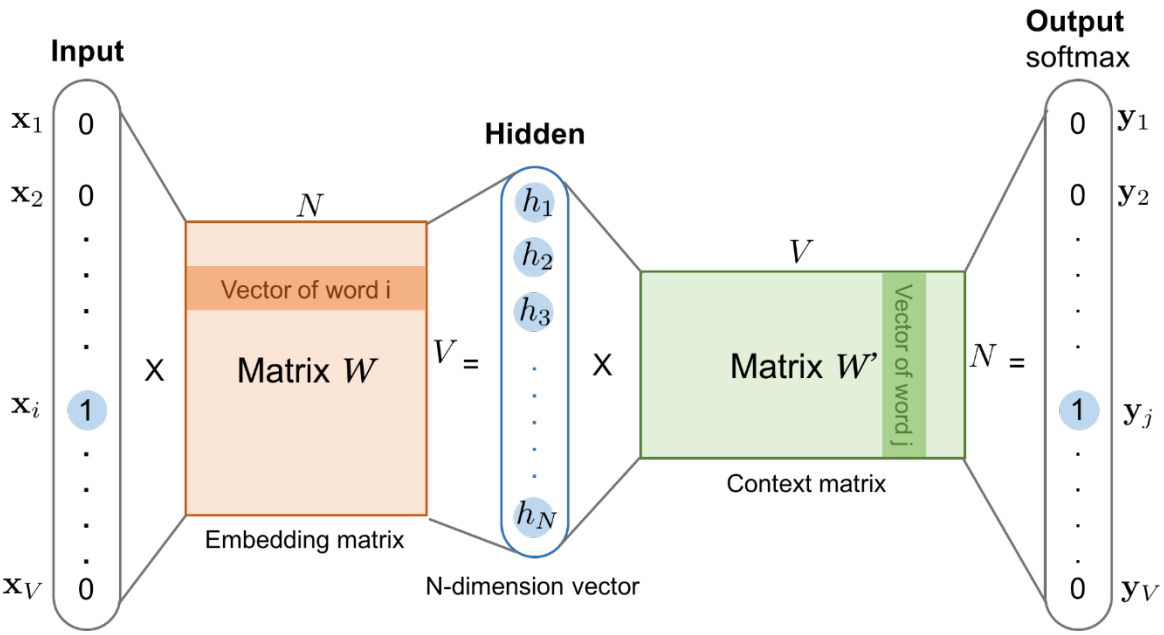


Fig. 1: the skipgram model

[7] https://lilianweng.github.io/posts/2017-10-15-word-embedding/

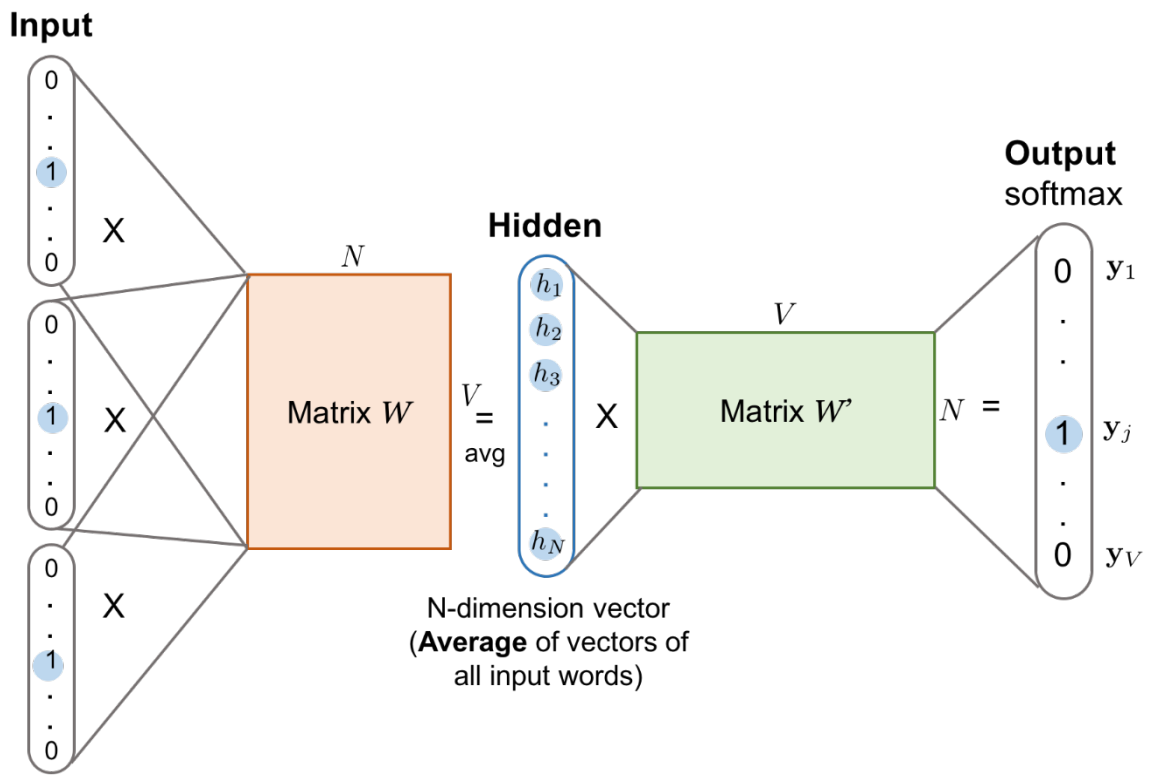


Fig.2: the CBOW model

The main change is that the CBOW creates, in the input layer, a dense vector representing the context and it then uses this vector to match the target word in the second matrix, which then leads to the output layer. We might then say that the main difference is that the CBOW learns from context embedding, as it feeds to the hidden layer an average of all input words.

Let us turn to two concrete examples to clarify this passage. If we take the sentence from Wikipedia "The cat is a domestic species of small carnivorous mammal", with target word *cat*, then it would be fed in our neural network as follow:

(1) For the SkipGram, "cat" would pass to the matrix *W* in figure NN to produce its dense vector. Then, it would pass to matrix *W'*, which is the context matrix. The context matrix, completely independent from *W*, represents the actual distributed

meaning of each word. The rows of the matrix are indeed the words in our vocabulary, while the columns are the dimensions of *W*. Supposing that our corpus is the whole Wikipedia, in this matrix we will have all the words of Wikipedia, including our sentence. Finally, in the output layer the SkipGram calculates the probability of each context to co-occur with *cat*.

(2) For the CBOW, the model would learn an *n*-dimensional vector representing a context, in our case it will be "the", "is", "a", "domestic", "species", "of", "small", "carnivorous", "mammal". All these words are averaged within a single context vector. Hence, the second matrix *W'* in the CBOW matches each context vector with the words in the vocabulary, finally computing the probability of our context vector to occur with "cat"

However, how is this probability calculated? There two ways of doing it and they are both valid for SkipGram and CBOW. The first one is called Hierachical Softmax, while the latter is called negative sampling. The softmax is an algorithm that calculates the probabilities as it were a decision tree (Quinlan 1997).

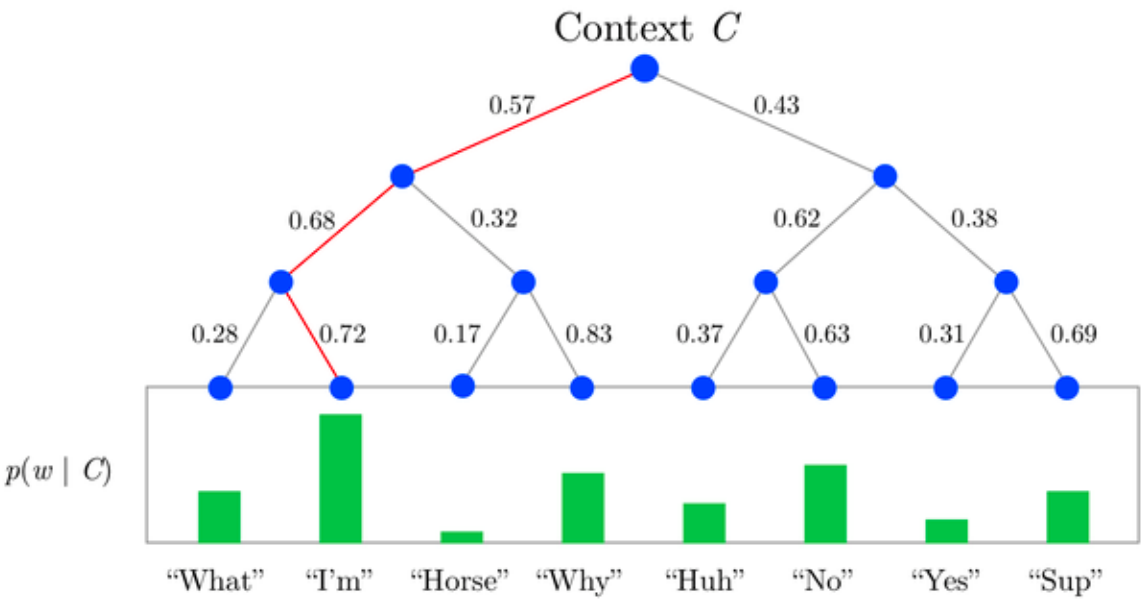


Fig. 3: Hierachical Softmax Credit to Steven Schmatz[8]

The main problem of the hierarchical softmax is that it is particularly expensive to compute for large corpora and the authors of word2vec were aware of that (Mikolov et al. 2013b). In fact, in this second publication they introduced what then became a standard for training word2vec, indeed negative sampling.

2.2.3.1 Negative sampling and downsampling: canceling the noise

As we explained in paragraph NN, the power of the neural network is that each function (neuron) is connected to the others. This means that, in the process of learning, the calculation of each neuron is adjusted to improve the learning. For instance, going back to the example of "cat", if the probability of "cat" to occur with the word "Mars" is 0, the *weights* of the neuron would be adjusted in order to retain this information in the following calculations.

This means that with a vocabulary of 1 million words we should update all the rows of our matrix and this of course takes a lot of time, as each line has *n* dimensions too. Instead, negative sampling selects a fixed number of words to be updated.

For instance, using a hierarchical softmax on a corpus that has 1 million words and 100 dimensions would mean that we should update 10 million weights. On the other hand, with negative sampling we might have 10 negative words to be updated, plus our target word. In this case, we only have 1100 weights to update, that is of course much faster and increases the quality of our embeddings (Lau and Baldwin 2016, Chamberlain et al. 2020). The words for negative sampling are selected using a unigram distribution, where more frequent words are more likely to be picked up as negative sample.

[8] https://paperswithcode.com/method/hierarchical-softmax

Moreover, in their paper Mikolov et al. (2013b) introduced also a technique called *downsampling* to account for data sparsity. As we know, words distribute in a corpus roughly following a Zipfian distribution, so that we will have few words with many occurrences and a lot of words occurring very few times. The subsampling is done again calculating the probability for a word to occur in a corpus with the following formula

$$P(w_i) = 1 - \sqrt{\frac{t}{f(w_i)}}$$

This equation calculates the probability for each term to be discarded; $f(w_i)$ is the frequency of the word w_i, while t is a chosen threshold that is usually a value that ranges from 0 to 10^{-5}, where 0 represents no downsampling.

Let's have two concrete examples to better understand this mechanism. We have a corpus of one million words, the word *cat* occurs 100 times. So, $f(w_i)$ would be 100/1000000, that is 0.0001. Let's say that we chose 10^{-6} as threshold, that would mean 0.00001/0.0001. Hence, resolving the equation would give us a probability of 0.6. It might seem a very high probability for such few occurrences of cat. However, a word occurring 1000 times in the same corpus would have $P(w_i) = 0.9$, meaning that is almost 50% more likely to be discarded. According to the authors, it is actually the threshold value that plays a key role.

We chose this subsampling formula because it aggressively subsamples words whose frequency is greater than "t" while preserving the ranking of the frequencies. Although this subsampling formula was chosen heuristically, we found it to work well in practice. It accelerates learning and even significantly improves the accuracy of the learned vectors of the rare words, as will be shown in the following section. (Mikolov et al. 2013.b)

Indeed, the subsampling might seem overly aggressive. Nonetheless, because of the Zipfian distribution, this aggressiveness would be almost all concentrated on the most frequent words in the corpus. Take for instance the word *the*, that would be very likely to represent our most frequent token in an English corpus. Imagining that it would occur around 50000, this would give us have $P(w_i) = 0.99$, very close to 1. Hence, the downsampling would affect the most frequent and less informative words, increasing computational efficiency and enhancing the quality of our embeddings. To maximize efficiency the threshold should be adjusted to corpus size. A small corpus is likely to need a greater threshold or perhaps even no downsampling.

2.3 TOPIC MODELLING AND KEYWORDS

Along with word embedding, we will be using topic modelling and keywords. This is necessary in order to explore our model following the evidence that emerges from our data.

For topic modelling we chose Reinert Descending Hierarchical Classification (henceforth DHC). For DHC we will be using the software Iramuteq (Ratinaud 2009) while we will compute keywords using WordSmith software (Scott 2020).

The aim of using these methods in combination with word embedding is to have data-driven evidence of some key aspects of our corpus:

1) Topic modelling provides us with a distribution of topics and the words that are associated with them.
2) Keywords helps us highlighting the differences between corpora

Regarding the first point, it is particularly useful for the study of algorithmic personalization as it allows us to explore the different semantic framing of each topic. For instance, lexicon associated with a political topic such as taxation might have different framing caused by algorithmic personalization.

On the other hand, keyness is crucial while dealing with polarization, echo chambers and in general all the phenomena in which we need to have a quick look at the macro difference among different parts of our data. In the case of echo chambers, computing keywords is useful to let emerge lexical markers of conflict or to highlight the differentiation among the main topics of interests.

2.3.1 Topic modelling choice: why DHC and not LDA

LDA (Blei et al. 2003) uses the Dirichlet distribution to discover, indeed, latent topics into our corpus. The unit of this model is a document, which is processed as a bag-of-words. Unlike word embedding, LDA does not calculate semantic similarity among words but the probability that each word would belong to a topic *k.* Hence, what LDA calculates is the proportion of terms that is assigned to a topic k for each document. Then it calculates the probability of t over the entire corpus, determining which are the most typical words, hence calculating the probability of each word of belonging to *k.* Finally, the model updates the probability of w over the whole model:

$$p(w|k) = p(k|d)\, p(w|k)$$

On the other hand, the Reinert method is an algorithm for hierarchical clustering. It works creating a document-token matrix that is weighted as binary (Lebart 1997), which means that DHC does not take into account frequency but the presence of the word into the document. The documents (called ICU, *initial context units*) are split into smaller textual segments called *elementary context unit* (ECU). The length of these segments is defined in the hyperparameters of the model. Finally, the top-down hierarchical classification is applied to the ECU. The final output is a set of classes that are defined by an exclusive set of ECUs and the words that are associated with those classes. The word clusters are created maximizing the inter-cluster Chi-squared distance (Lapalut 1995). A detailed explanation of the functioning of the algorithm could be found in Ratinaud and Marchand (2016),

The two methods are similar, but they are useful to achieve different goals. LDA is a particularly fast and accurate way to explore a dataset discovering the most relevant topics and the words associated with it. Moreover, it is useful when we need to explore the different use of lexicon in those topics, as each word might be assigned to multiple topics. Instead, the DHC provides a strong top-down classification in which words are assigned exclusively to one cluster. This is particularly important for our experiments because it also accounts for reproducibility of the results, as probabilistic methods are, by definition, variable.

With DHC we have an unambiguous and fully reproducible topic modelling that allows us to do a data-driven exploration of our word embedding model. The problem of reproducibility of word embeddings has been addressed by Hellrich (2019), that noted that models such as Word2Vec are never fully reproducible because of their predictive nature. In fact, Word2Vec is a stochastic model, meaning that, even with the same parameter, it would produce a slightly different output on different machines.

To address this problem Hellrich experimented a Singular Value Decomposition of a PPMI matrix with weighting-based downsampling to generate reliable word embeddings without losing performance. In this case, it is possible to have a fully reproducible word embedding model, meaning that with the same parameters we always get the same vectors. SVD works on three different matrixes, two orthogonal containing vectors and one diagonal containing values. These three matrices are then decomposed each time in the same way, so that the same input would always generate the same matrix. The solution proposed by Hellrich is particularly interesting, as it introduced a variant using a PPMI matrix populated via weighting, to enrich its embeddings.

However, this would shift our semantic model to a fully count-based vector space, which might be a very good solution for machine translation like in Hellrich work but may not be the best for computational pragmatics. As we have already said before, for our goals a predictive method would be the best solution as it is the best modeling to capture pragmatic relation. However, we should not underestimate the problem of reproducibility even when working on pragmatics. For

this reason, we decided to prefer DHC over LDA, to introduce a stable and reproducible step to guide our embedding analysis.

2.3.2 Corpus-assisted tools: keywords

In this work we use corpus-assisted discourse analysis along with word embedding. In fact, Corpus-based discourse analysis has a long tradition in the field of studies on representations of ideology, starting from Stubbs' emphasis on the contribution of corpus-based lexico-grammatical analysis to critical studies of culture and ideology (e.g. Stubbs 1996, 2001), and reaching work at the intersection of corpus linguistics and CDA (e.g. Mautner 2001, Baker et al 2008; see Nartey and Mwinlaaru for a recent overview). Corpus perspectives contribute to discourse analysis by providing attention to frequencies and to words in combination, i.e., phraseology in a wider sense. Corpus tools like frequency lists and keywords provide us with easy access to quantitative data whereas concordances allow the study of meaning in text. Sinclair's extended-units-of-meaning model (Sinclair 1996) offers four levels of analysis: collocation (words that occur regularly with the node word, the word under investigation), colligation (grammatical categories that define the immediate context of the node word), semantic preference (the tendency to co-occur with words sharing the same semantic features) and semantic prosody (the tendency of the word to occur in specific pragmatic contexts, its relation to a specific speech act and/or evaluation) (Hunston 2007).

In particular, we are using keywords, namely words that occur significantly more frequently in a corpus than in the other (Bondi and Scott 2010; Gabrielatos and Marchi 2011, 2012; Pojanapunya 2018). It should be noted however that we are using keyness in one of its possible specific implementations. In particular we are using Scott's (1997) idea of keywords as "words that occur with unusual frequency in a given text". As also explained by Gabrielatos (2018) raw frequency cannot determine keyness, instead it is

the measurement of the statistical significance of this over- or under-occurrence. In this case there are several methods to determine the significance of keywords; in Wordsmith 8, their statistical significance is calculated using the Bayesian Information Criterion (BIC) score, as suggested by Wilson (2013). The Bayesian Information Criterion is a statistical criterion for model selection which is closely related to the likelihood function. Hence, in our case, the BIC score measures the statistical significance of each word keyness, in the direct comparison with the other corpus.

Computing keywords allows us to explore polarization, highlighting differences between two groups. Just like topic modelling, it allows us to investigate our word embedding model starting from evidence that emerges from our textual data. This enhances the reproducibility of the experiments and it includes a specific corpus-driven indicator to explore textual polarization.

2.4 THE APPRAISAL FRAMEWORK

In this thesis we refer to the theory of J.R. Martin and P.R.R. White as the Appraisal framework (Martin and White 2005)[9]. Specifically, we intend to use their proposed theory to explore pragmatic relations that cannot be fully captured only relying on word embeddings. These pragmatic aspects are those related to the dialogistic dimension, which are particularly important for our case study.

As we said in chapter 1, the distinctive feature of an echo chambers is ideological isolation, the extreme polarization of two irreconcilable positions, while for filter bubbles this ideological isolation is an effect of algorithmic personalization.

[9] Their approach is part of the Systemic Functional Linguistics (SFL), which considers language as a social semiotic system. More details on this approach could be found in Halliday (1961, 1978), the founder of SFL and also Martin (2016).

We can address ideological isolation starting for the semantic framing of each word, as we do with word embeddings, however this is not enough to have a full comprehension of the phenomenon. Echo chambers are indeed an active phenomenon, meaning that from a linguistic perspective they develop within the dialogic dimension. It is not the simple opposition between two semantic frames that would define an echo chamber, instead we need also an active rejection of the counterpart. In particular, for our goals we are interested in the category of engagement, which explores the different possibilities of dialogic expansions and contractions. We might speak of echo chambers when we have a discourse that is closed to other voices. Hence, we should consider as part of an echo chamber all those heteroglossic (i.e., introducing external voices) and monoglossic (i.e. strong evaluative language) utterances that challenge the opposite view, that Martin and White call *dialogic contraction*.

In this work we will focus on heteroglossic statements in the engagement domain, since they are easier to identify in texts, e.g., by searching for specific markers. On the other hand, the monoglossic dimension is difficult to operationalize, while surely it sometimes may contribute to the polarization of the debate, especially when it is used to make a clear stance-taking.

2.4.1 Introduction to appraisal theory

In this subsection we briefly review the general theory proposed on the engagement category by Martin and White.

*As indicated, we include within the category of **engagement** those meanings which in various ways construe for the text a heteroglossic backdrop of prior utterances, alternative viewpoints and anticipated responses*. (Martin and White 2005, p.97)

In other words, the category of engagement includes every utterance that is meant to reject or consider opposite viewpoints. It is therefore a way to analyze what goes beyond

stance taking, taking into account what in semiotics is called enunciation strategies (Benveniste 1970, Benveniste 1971, Greimas and Courtes 1979, Manetti 2008).

Now, the category of engagement is quite articulated. First of all, as we said earlier, it is composed by two other macro categories which are expansion and contraction, where the expansion is the possibility to introduce other voices while the contraction is the opposite. However, inside these categories, the lexico-grammatical elements highlighted by Martin and White introduce further shades that makes the distinction between expansion and contraction somehow also blurred.

2.4.1.1 Dialogic Expansion

Inside expansion we find two other different possibilities that are called *entertain* and *attribution*.

The *entertain* is perhaps the strongest textual opening as it explicitly validates opposite viewpoints that are openly, or at list potentially, in contrast with the textual stance presented by the text.

They construe a heteroglossic backdrop for the text in which the particular point-of-view is actually or potentially in tension with dialogistic alternatives. By this, they project for the text an audience which is potentially divided over the issue at stake and hence one which may not universally share the value position being referenced. (Martin and White 2005, p.108)

The *entertain* is introduced with locutions like modal verbs and adverbs expressing other epistemic modalities such as direct states of believing, probability or other veridiction strategies. It also includes all those locutions expressing deontic modalities and all those statements that we can include under the category of evidentials. The authors also include what Goatly (Martin and White, 2005 p.110) calls expository question, a particular type of rhetorical questions.

On the other hand, *attribution* includes a vast range of discursive configurations that are used to attribute something to an external voice. This might be done in two different ways: with what is called acknowledgement, that is the mere reporting of external voices, and with distance, e.g., the introduction of an external voice with, indeed, explicit distancing from its standpoint.

2.4.1.2 Dialogic Contraction

Opposite to dialogic expansion, Martin and White introduce the category of dialogic contraction which include quite a vast range of different possibilities. *Disclaim* includes two sub-types that are *deny* and *counter*.

Under disclaim we cover those formulations by which some prior utterance or some alternative position is invoked so as to be directly rejected, replaced or held to be unsustainable. (Martin and White 2005, p.118)

2.4.1.2.1 Deny

According to Martin and White, this particular type of contraction has a particular dialogic status, since it acknowledges an external voice while rejecting it. This is a well-known problem, as negation always includes the positive negated term, while the opposite is uncommon (a positive term including the negative). From a dialogistic and even a semantic perspective, for instance, formulations such as "No-Vax" or "No Global" are dependent on what they are denying: vaccines and globalization are present within the pragmatic context needed to interpret these two expressions. Also, the denial might vary a lot, as it might be a way to align the reader to author's positions or a way to address expected values or beliefs that reader might have. The following examples clarify these two different types of denial.

(5) Labour said that Brexit will be a tragedy. I can promise that this is false.

(6) Brexit is not a problem: we have a detailed plan on how to relaunch our economy

In (5), the denial is used to align the reader to the enunciator voice, while introducing an opposite and discredited standpoint. On the other hand, in (6) the denial targets the reader, assuming that he/she believes that Brexit would cause problems for the economy.

In both cases however, denial is used to create a model reader (Eco 1979) that is aligned with the enunciator standpoint. However, because it is a negation, its effect is not completely under the control of the author.

In this context, it is useful to recall the renowned textual intentions theorized by Umberto Eco (1990)

1. Intentio Auctoris: the standpoint of the author that he may have wished to communicate regardless of the text s/he actually produced
2. Intentio Operis: the meaning that emerges from the text, with reference to the author's intention as it is reconstructed within interpretation
3. Intentio Lectoris: the interpretation of the addressee regardless of the textual clues allowing for the reconstruction of the author's intention

It is crucial to introduce also the concepts of *model reader* and *model author* (Eco 1979), as well as those of empirical reader and empirical author. The former are semiotic devices while the latter are two concepts used to indicate the actual readers and the actual authors, intended as human beings with their social and psychological traits. The two are distinguished because the text, by means of its structures, is an autonomous semiotic object. According to Eco, the model author is a textual strategy. On the other hand, the model reader is the pragmatic competence needed to interpret a certain text with reference to its producer's intention.

Model reader and model author are indeed two sides of the same coin. The model author is produced by an empirical reader that interpret the textual strategy proposed

(hence what produced the *intentio operis*) while the model reader has been produced to guide readers' interpretation. In (4) for instance the model reader assumes that the empirical reader would be worried about Brexit. However, this is somehow true also for (3), although less evident as the addressee is not directly involved.

2.4.1.2.2 Counter

A similar reasoning applies also to the second sub-type of disclaim, that is *counter*. In this category, we have all those utterances introduced by adversative elements that, just like negation, introduce a conflictual aspect within the dialogistic dynamics. We argue then we could refer to the category of *disclaim* as the symptom of dialogic on-going conflict.

Hence, we can see denial and counter as a particular form of dialogic contraction, namely markers of dialogic conflict, as they introduce the notion of a conflict with the counterpart (counter discourse).

2.4.1.2.3 Proclaim

The second macro-type of dialogic contraction is *proclaim*.

We group together under the heading of 'proclaim' those formulations -which, rather than directly rejecting or overruling a contrary position, act to limit the scope of dialogistic alternatives in the ongoing colloquy (Martin and White 2005, p.121)

This category of contraction has three related sub-types which are *concur*, *pronounce* and endorse. We will briefly sum up these last three.

1. Concur is introduced by all those formulations, mainly adverbs, that highlight points of agreement or common knowledge among the addresser and addressee (*unsurprisingly, surely)*

2. Endorsement happens when external voices are introduced within the discourse and presented as undoubtedly true and trustworthy. These utterances are often introduced by factive verbs.
3. Pronounce, finally, is an explicit emphasis on the author's point of view, hence all those formulations which restrict the dialogistic space without introducing or referring to external voices.

2.4.2 Corpus-driven approaches to appraisal

The appraisal framework is often used in corpus linguistics along with annotation (Read and Carroll 2012, O'Donnell 2014, Cavasso and Taboada 2021). An example of this approach is Fuoli (2012) which has focused on quantifying markers of appraisals in corporate reports. His work was aimed at quantifying the elements of appraisal present in the texts, comparing two different textual strategies of corporate social responsibility. Fuoli in this work was especially devoted in clarifying that the annotation process was well evaluated according to the most common principles in the discipline and that more than an annotator contributed to the process.

Later, Fuoli also wrote a specific paper (Fuoli 2018) on the annotation process for appraisal, pointing out that Martin and White present their coding choices as self-evident and unproblematic. Although we agree with Fuoli that evaluative language is complex to analyse, there are some aspects of the appraisal framework that we might accept as self-evident, i.e., engagement. Hence, we argue that dialogic expansion and contraction could be highlighted as objective textual structures that emerge from the discourse.

The refusal or the endorsement of external voices is in fact something that emerges, clearly, from the enunciation strategies of the text itself. Of course, as we saw in paragraph 2.3.2, even in the engagement frameworks there are some elements that are

somehow borderline and surely, we will have to deal with problematic and non-dichotomic utterances. We will of course address this part of analysis with great care of the detail, nonetheless, defining a reliable methodology for annotating appraisal is outside the scope of our research goals.

The real added value of using word embedding is that we can approach textual elements as they naturally emerge from text, without the need of manual coding. In our approach, the appraisal analysis will be part of the last quali-quantitative step, in which we will explore the dialogistic dimensions within the pragmatic status of our words.

The methodological intent is twofold:

1) We are interested in a particular category of appraisal, i.e., engagement. This category will certainly be present in every political discourse, as there is always at least a counterpart. In particular we are interested in quantifying the appraisal, looking for evidence of polarization, i.e., the abundance of dialogic contraction.
2) We would like to keep a local dimension on the texts, studying appraisal on small excerpts of collocations, to contextualize the theoretical framework of Martin and White: However, we argue that this second step should necessarily follow the quantification of appraisal, along with a statistical overview on the distribution of the appraisal markers. The goal is to have a data-driven fine-grained qualitative enquiry.

These two different types of approaches might be summed up as semio-linguistic driven and data-driven. We apply a semio-linguistic approach when we start from an idiographic dimension and then we move to a quantitative perspective, using for instance manual coding to quantify appraisal markers. On the other hand, we apply a data-driven approach when we start from a quantitative dimension and then we move to an in-depth

analysis, as in the case of corpus-driven discourse analysis or word embedding. However, these two approaches should not be intended as a rigid dichotomy, as in an operational context there is rarely a clear line between the two. For instance, the work of Fuoli as well might be considered data-driven as they start their analysis solely after a robust quantitative investigation, although as we said it is an analysis that is based on a qualitative paradigm that is then put through the lens of a quantitative approach.

Since in this work we are trying to have a hybrid approach we need to reflect thoroughly on what the interaction of these two methodologies entails. Qualitative and quantitative have both strengths and limitations. For instance, if we adopt a standard data-driven approach we can start our analysis on very robust premises, while qualitative methodologies allow us to explore in depth the linguistic facets of our corpus, grasping in detail the different nuances of signification. However, quantitative often also means large datasets, which brings great complexity. Hence, a multidisciplinary approach is likely to be the best solution, as the quantitative provides us with a big picture, missing some of the granularity that a linguistic inquire usually requires. Conversely, a pure idiographic approach prevents us from generalizing our conclusions, regardless of the size of our corpus.

We propose to structure the methodology following these milestones:

(1) *Data-driven first*. Given the complexity and vastness of the data, it is always necessary to highlight what naturally emerges from the data, following quantitative evidence. This is the reason why we are using word embedding along with topic modelling and keywords, to shed some light on our corpora.

(2) *From big to small*. The semiotics-driven, the linguistics-driven, generally speaking the humanities-driven, is always a key part of the exploration.

Quantitative evidence is not enough to respond to our research questions, as the granularity that we need is often hidden into the folds of our corpus. It might be immediately clear what might be a keyword or how a given word is framed semantically but understanding the reasons of the data-driven stage is really the core work of the humanities. This means that our word embedding output should be explored with a qualitative exploration, evaluating the semantic frames of the most relevant words and that we should use appraisal to investigate the dialogic dimension.

(3) *Mutual Enhancement*: Data-driven and humanities-driven have different goals and different strengths. A theoretical framework, such as Appraisal, might be introduced in a data-driven perspective in many ways. Instead, computational tools as word embedding are often used with purely operational tasks such as translation or sentiment analysis. The idea of this methodological proposal is that the combination of the two approaches creates a new and improved research protocol, where the theoretical frameworks could be experimented on large data and where the computational tools can be used to explore solid and complex theoretical frameworks.

The idea is not particularly new. For instance, distributional semantics itself could be seen as a quali-quantitative methodological framework as the distributional hypothesis was, tautologically, a simple hypothesis. What was an expression of a theoretical assumption has then evolved in cutting-edge technology that we now take almost for granted. However, the real advantage of DSM is not the particularly lucky compound of a right hypothesis and technological advancement but instead the hybrid composition that

allowed to translate a linguistic concept, namely the regularity of contexts, in a computational task and finally in a usable technology.

At this point, a reasonable criticism would be to suggest that also corpus-driven discourse analysis is, in fact, a quali-quantitative methodology that is already self-sufficient. Of course, it is true that corpus linguistics has a long-standing tradition in exploring research questions coming from the humanities using a solid quantitative framework. However, in this thesis we are proposing a methodology for social media analysis, which is something that goes beyond the pure language dimension and, as we illustrated, is a relatively unexplored topic in corpus linguistics. For instance, to approach the filter bubble we would need to use a digital methods framework, using a specific tool for data collection and following a specific methodology meant to investigate specific platforms. This goes, indeed, beyond the linguistic dimension and thus beyond the effort that a traditional corpus-driven approach would provide, because it involves different disciplines and issues that we believe should be investigated in a new methodological perspective.

3 FILTER BUBBLES

3.1 INTRODUCTION

In this section we are introducing an experimental approach to the study of filter bubbles. In section 3.2 we focus on the study of the Facebook platform, working on the data of a pre-existing experiment made in 2018 during the Italian elections. In this section we explore the effects of algorithmic personalization on the political debate, introducing the use of word embedding to simulate the model reader (Eco 1979), hence the pragmatic dimension of our texts.

On the other hand, in section 3.3 we focus on the study of YouTube, seeking data-driven evidence of YouTube filter bubble and presenting a specific case study, showing the potential effects of algorithmic personalization on the polarization of the debate. The chapter will show that Facebook and YouTube algorithms pander to the ideological preference of their users, creating a noticeable difference within the discourse of key political topics that they propose to each group of users.

This chapter is also aimed at showing how we can experiment filter bubble analysis within a digital methods framework, creating a corpus that we can explore with the use of word embedding, to understand the inferential paths proposed to the final user.

Section 3.2 is published in Sanna and Compagno (2020) while Section 3.3, other tan 3.3.4, is published in Sanna et al. (2021).

3.2 FACEBOOK

Most of the content of this section is published in: Sanna, L., & Compagno , D. (2020). Implementing Eco's Model Reader with Word Embeddings. An Experiment on Facebook Ideological Bots. In JADT 2020: 15th International Conference on Statistical Analysis of Textual Data.

Facebook is still (by 2022) the most used social media platform in the world, with almost 3 billion of monthly active users. Facebook it is also the platform that is been at the

center of mediatic attention for one of the most important scandals in the social media era, namely the Cambridge Analytica case[10]. For our goals this is a crucial passage for two reasons. On the one hand, Cambridge Analytica has made it clear that platforms can play an important role in influencing political dynamics. On the other hand, it started what is now known as post-API research (Freelon 2018, Perriam et al. 2020, Tromble 2021), namely an era in which is difficult for the researchers to collect independent data on social media platforms.

For these reasons, also on Facebook it is necessary to use a specific tool for data collection. The dataset on which we conducted our experiment was collected by the Tracking Exposed (TREX) research group[11] during the Italian elections in 2018[12], via the browser extension of Facebook Tracking Exposed (FBTREX)[13]. We already covered their experiment in section 1.2.2 (Hargreaves et al. 2018a). To sum up, they create six automated profiles that simulated an ideological preference interacting (by liking) only with the content coming from one particular political segment. The researchers found evidence of uneven exposure to the sources of information, discovering that usually exposure was unbalanced towards the political bias of the bots. Starting from the same pages, the six bots accessed six different feeds. Now, if we consider algorithmic personalization from the perspective of Umberto Eco's semiotics, the politically biased feeds of the six bots would contribute to the shaping of six different model readers.

The concept of model reader was created by Umberto Eco (1979) to represent the pragmatic competence needed to interpret a certain text with reference to its producer's intention or, better, to a given hypothesis about this intention. We could think of the model

[10] https://en.wikipedia.org/wiki/Cambridge_Analytica
[11] Official website https://tracking.exposed/.
[12] Database available on Github at https://github.com/tracking-exposed/experiments-data.
[13] https://facebook.tracking.exposed/

reader as a set of implicit instructions to formulate legitimate inferences from sentences and their combinations in texts. One of the main components of the model reader is what Eco called encyclopedic competence, accessing a cultural context shared by the text's producer and reader. For instance, to understand the sentence "Macron nomme Philippe" [Macron nominates Philippe], the reader has to know that Macron is the actual French President, that the President has the power to nominate a Prime Minister, and more importantly the reader has to guess also (by abduction) that this sentence refers to such nomination. In Eco's theory, these steps are accounted for by the competence of a model reader adequate to that sentence, including knowledge about the actual world, it's entities and their interactions. In absence of such encyclopedic interpretation, it would be impossible for the reader even to guess the correct meaning of the verb "nommer" in this context. This competence has to be acquired by a previous knowledge of other texts and verbal exchanges. For example, to fully understand the last chapter of the Star Wars movie franchise, I am required to know the previous episodes, since the plot has plenty of intertextual references pointing to them. Hence, we should intend the encyclopedic competence of the model reader as a shared cultural background, that in many cases might be expressed by intertextuality.

In order to treat the concept of model reader empirically, we propose a tentative formalization of it. We argue that it can be seen as an inferential model, produced as output from a function that takes as input a target text and a larger corpus of texts. This model should then be able to add to the target text the implicit information needed for its interpretation. Reading, as an activity, therefore is accounted for by two circular steps: the identification of some reading instructions from a text and their application to the text itself. If such formalization of the model reader could be implemented computationally, it would make it easier to treat textual production and interpretation automatically. We also

argue that it would become possible to perform experiments so to observe whether or not a model reader, and the textual interpretation deriving from its application, is affected by algorithmic personalization.

As we highlighted in chapter 1, the theory of the filter bubble has pointed out that in an era of algorithmic personalization, there is a danger of not sharing any common ground anymore, each of us living in his or her personal information bubble[14]. This is because each person has access to a richer variety of sources compared to the pre-Internet era, and so people make use of algorithms to retrieve the information that is pertinent to them. In this work, we are interested in understanding whether (and to what extent) the model reader of a set of texts is influenced by filter bubbles. In other words, how the access to information affects textual interpretation, by constraining the constitution of a largely shared common knowledge. We point out that this research goal is particularly complex since the study of algorithmic personalization has in itself some methodological issues that are still unsolved (i.e., how to control for all the variables of users' behavior that may influence the filtering algorithm, such as user activity on and outside social media platforms). Another important factor to take into consideration is that a large amount of digital data has to be collected and analyzed to study algorithmic personalization. As a first step towards accomplishing this task, we used word embedding to implement the model reader of some text fluxes.

In other words, the textual flux shown to a FB account influences the underlying instructions that guide the reading of the flux itself: for example, a person who gets his or her information mainly from far-right sources should realize certain inferences more easily than another person reading far-left sources, and vice versa. Because of this, we

[14] The hypothesis of filter bubbles does not make a universal scientific consensus. Some studies show that individuals may have access to a large and shared background (Compagno et al. 2017, Bechmann 2018).

expect to find six different inferential models in the TREX dataset, one for each bot. Our aim is to simulate these six inferential models by training six different word embedding spaces and studying their behavior.

3.2.1 Experiment

First of all, it is important to highlight that the inferential model we implement with word embedding may perform two different kinds of inferences:

1. *Necessary inferences* are all those inferences driven by common word co-occurrences in the corpus. These inferences are necessary for a basic understanding of the semantic content of the text. They include *mandatory inferences* (i.e., compound words) and *natural inferences* such as technical lexicon (names of people, institutions, laws). For instance, “step” and “out” –> “step out” and “Macron” –> “President” are examples of necessary inferences.
2. *Embedded inferences* are all those inferences made without starting from direct co-occurrences of words in the corpus. The capacity to detect these inferences is the real peculiarity of word embedding. These inferences are interesting because they provide interpretation paths that would not be easily discovered by standard semiotic analysis or corpus linguistics. For example, “Leader” + “Germany” –> “Merkel” is an example of embedded inference.

To study the behavior of the inferential models simulated by word embedding, we wanted to identify some words that were used all along the entire Italian political spectrum. Starting from these words we could then observe which inferences they generated in different model readers (associated to our six bots). We first used Iramuteq[15] to perform a hierarchical classification (Reinert 1983) on the sources' dataset: we identified the main topics in the entire debate and the words characterizing these topics, independently from the eventual political filtering operated by the FB algorithm. Figure 4 shows a correspondence analysis displaying the 11 clusters and the most specific words for each, that is, the words which are most associated with a cluster and less with all the others.

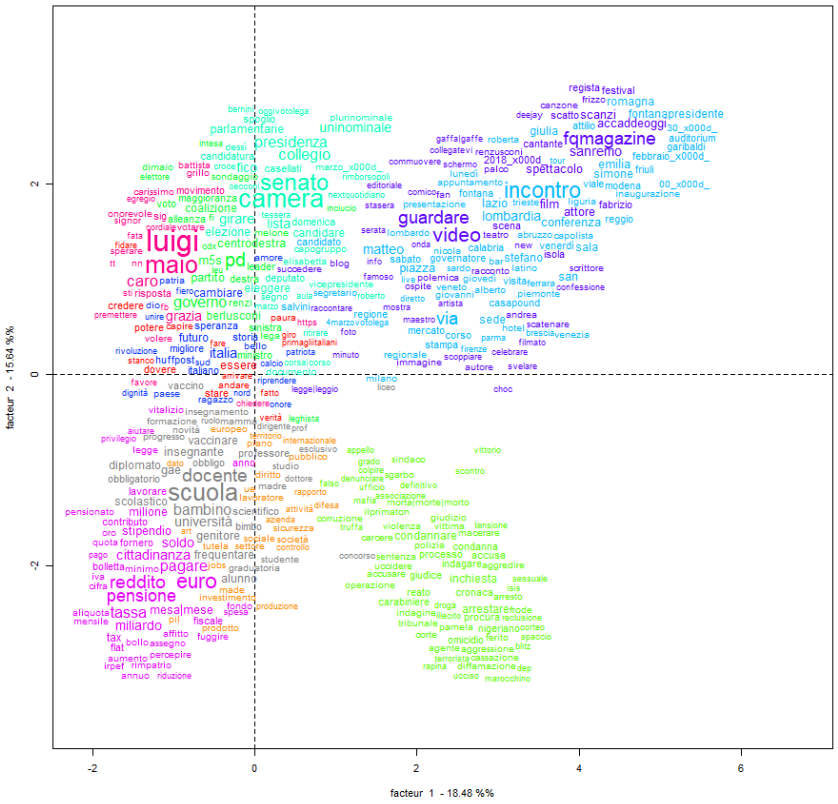


Fig. 4: Correspondence analysis of the clusters in the sources' corpus

[15] Iramuteq (0.7 alpha 2), by Pierre Ratinaud, 2020, http://iramuteq.org/.

Class 1 (14,7%): "Verb Modalities"	Class 2 (12,39%): "Occupation"	Class 3 (14,52%) "Crime News"	Class 4 (10,78%) "Taxes & Pensions"
1. essere/stare	1. lavoratore	1. arrestare	1. euro
2. credere	2. settore	2. indagare	2. reddito
3. potere	3. europe	3. inchiesta	3. pensione
4. dovere	4. investimento	4. cronaca	4. tassa
5. pensare	5. sociale	5. nigeriano	5. pagare
6. vedere	6. tutela	6. condanna	6. cittadinanza
7. capire	7. assunzione	7. procura	7. soldo
8. andare	8. diritto	8. sentenza	8. stipendio
9. paura	9. sicurezza	9. omicidio	9. "flat tax"
10. governare	10. made in Italy	10. processo	10. fornero
11. sentire	11. economia	11. reato	11. lavorare
12. fare	12. prodotto	12. giudice	12. fiscale
13. promettere	13. salute	13. carabiniere	13. bolletta
14. popolo	14. ue	14. accusa	14. disoccupazione
15. verità	15. occupazione	15. polizia	15. versare

Table 1: Main clusters' composition

After having identified these 60 frequent and supposedly neutral words, we performed word embedding on the impressions' dataset, creating one model for each bot. Then, by using the lists of words obtained with our classification above, we explored our models so to see how our bots "make inferences": which new words does each bot associate to the 60 extracted from the four clusters?

We created an embedding for each bot, building both a skip-gram model and a C-BOW model, using in both cases a window context of 10 words and 200 dimensions. We used the plain text of each post to create our models[16]. In Table 2 we summarize the size of each model.

[16] For each post in the impressions dataset we have: the date of publication, the date of impression, the number of comments, the name of the publisher, impression order (the position in the newsfeed at the time of the impression), the permaLink of the post, the URL of the post (sometimes external), the post ID, the publisher orientation, the bot political orientation, the number of visualizations. We filtered our posts by political orientation, mapping the post ID within the sources dataset in which we also have the "post message" available. The post message is the plain text of each Facebook post. For our goals we only need the post message of each impression.

	Center-Left	Far-Right	Left	Populist	Right	Undecided
Word Types	12615	13712	18326	24930	25893	18945
Word Tokens	89600	107254	180629	312954	376813	187033
Posts	3694	3417	6437	13331	16085	7597

Table 2: The size of the different sub-corpora for each embedding. Duplicated posts have been removed

Each sub-corpus has been preprocessed removing URLs, emojis and terms shorter than 2 characters; each word was also turned to lowercase during the tokenization process. Word2vec works by default with a downsampling[17].

This basically ignores the most frequent words (like articles and prepositions), that do not add any semantic information to our model (see section 2.2.3.1 for details). However, in our corpus we obtained better results without downsampling[18]. This is probably caused by the relatively small size of our dataset; word embedding algorithms are designed to work with very large datasets and the default downsampling threshold has been determined heuristically, as stated in the original paper (Mikolov et al. 2013a).

For each of the four clusters, we computed the cosine similarity of the top-20 most similar words. The results are summarized below (Figures 5-8). For visualization we selected samples of words including of:

- The two most similar adjectives
- The two most similar nouns
- The two most similar verbs

[17] According to Mikolov et al. (2013a) "each word w_i in the training set is discarded with probability computed by the formula (1), where $f(w_i)$ is the frequency of word w_i and t is a chosen threshold, typically around 10^{-5}."

[18] This is part of the Gensim implementation that recommends using a threshold in a range between 0 and 10-5. In our work we determined, heuristically, that the best value for us was 0.

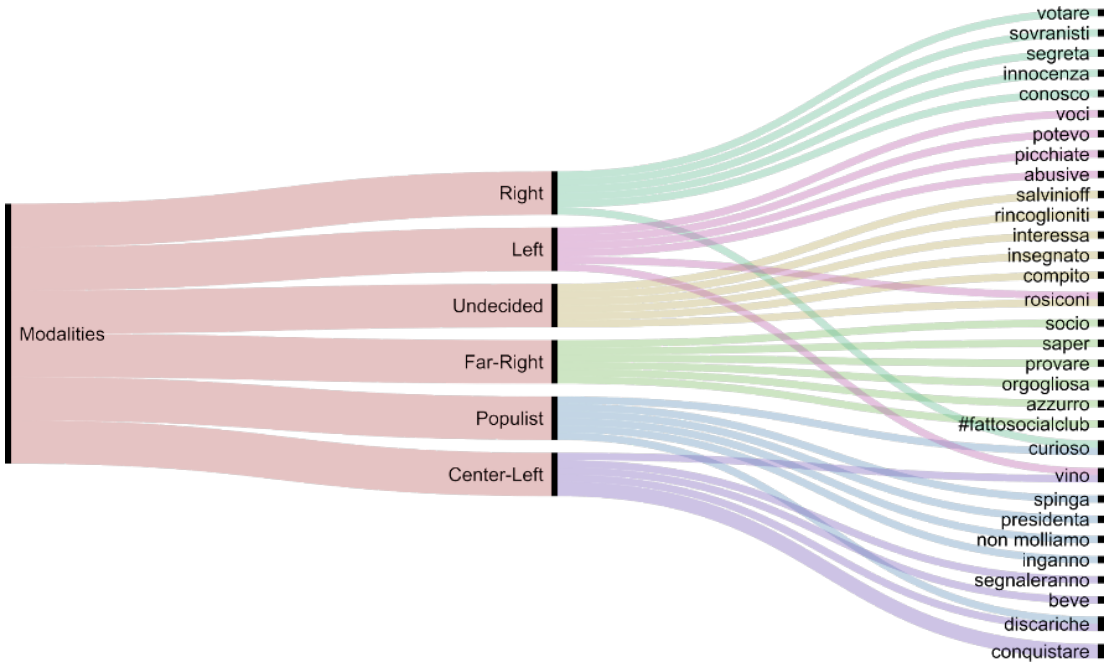


Figure 5: Inferences within the "Modalities" cluster

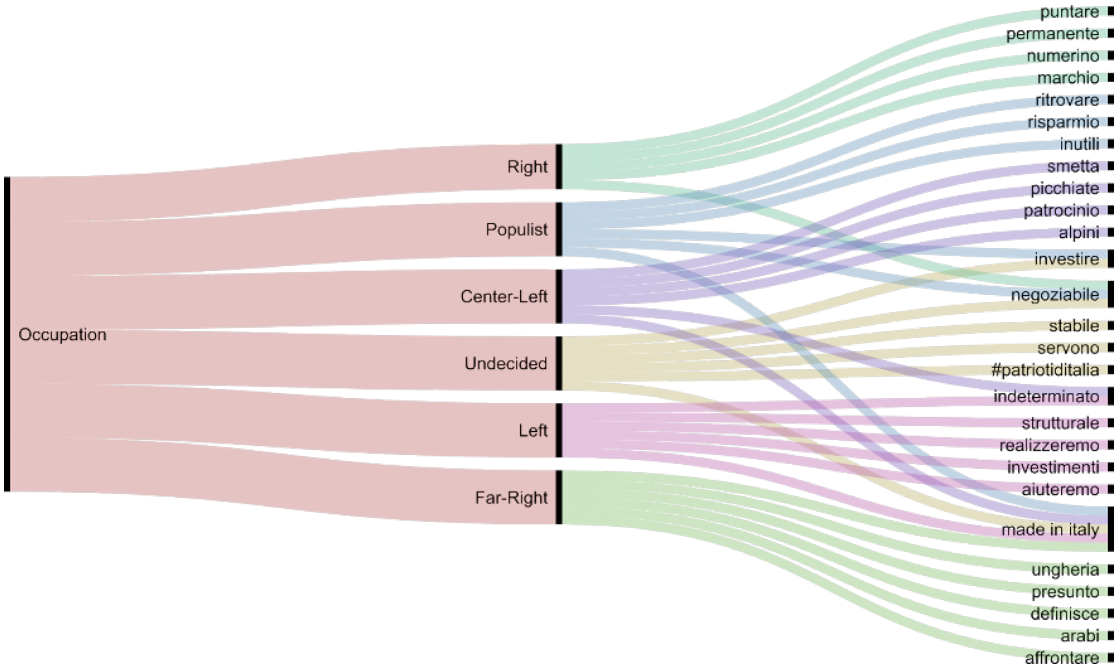


Figure 6: Inferences within the "Occupation" cluster

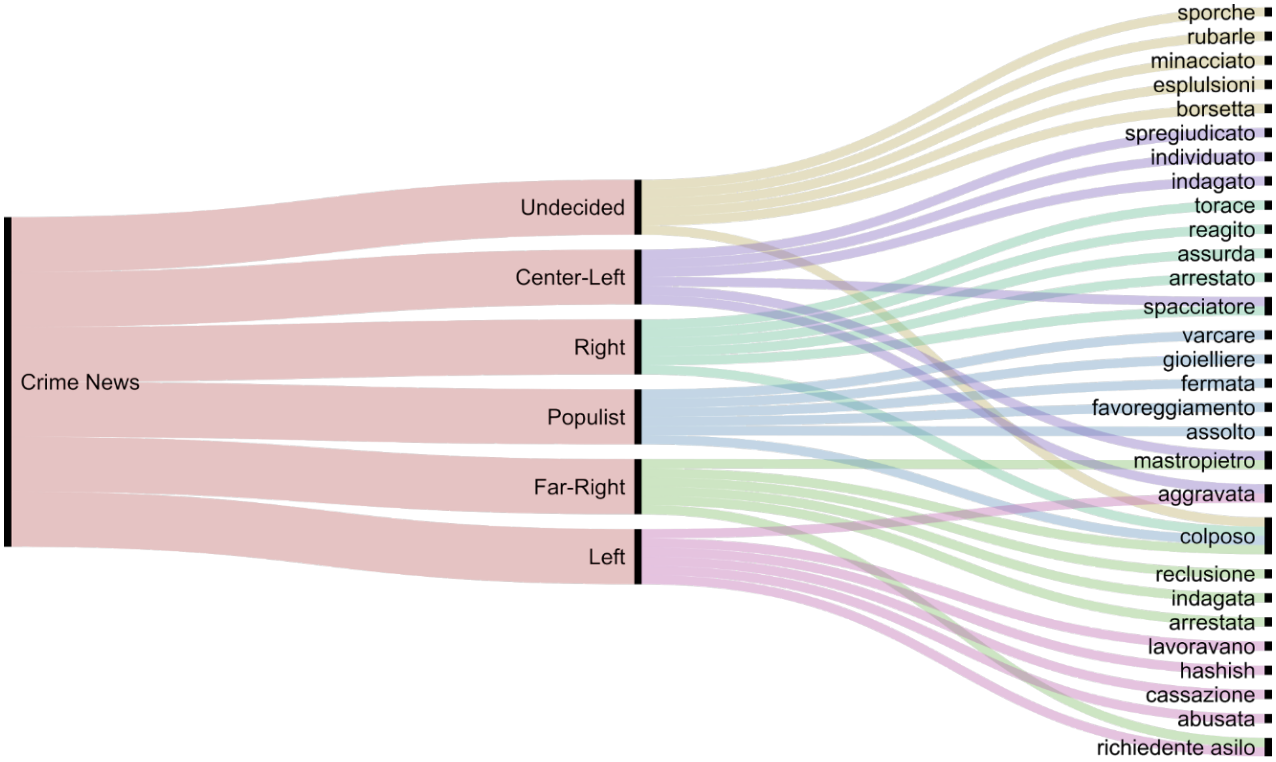


Figure 7: Inferences within the "Crime News" cluster

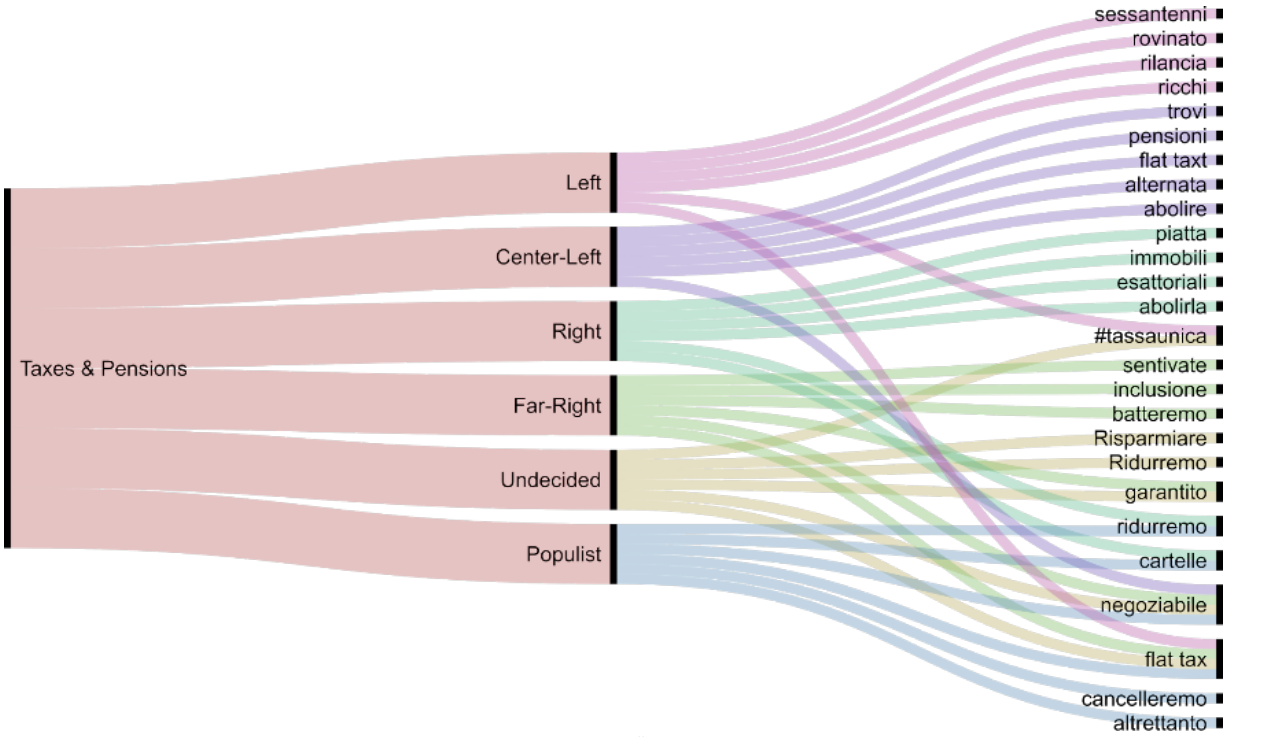


Figure 8: Inferences within the "Taxes & Pensions" cluster

This first level of inferences portrays a situation in which there is a lot of lexical diversity, suggesting that we can really see different inferential models in action. For example, each bot associated very specific words to the *Occupation* cluster, not shared by the other bots, so reframing the problematic from a unique perspective. There is an exception: all bots seem to recall the expression "Made in Italy" when making inferences in the occupation domain. However, this does not mean that the inferred words necessarily have the same connotations for all bots: in the *Crime News* cluster, the expression "richiedente asilo" (asylum seeker) is inferred both by the Left and the Far-right bots, but probably a deeper analysis would show two contrasting axiological positionings towards the expression. Still, by qualitative analysis of the inferred terms, we observed that those used by right-wing parties seem dominant in the global debate. In the two economic clusters, the expression "flat tax" and "made in Italy" are shared by many bots, and both expressions were crucial in the right-wing parties electoral discourse. As first partial result it seems, then, that if different bots can be associated to different model readers, each with its own perspective, they are also influenced by the global information as a whole.

We can try to go more into the details of the inferential models by calculating *snowball inferences* (Rogers 2017) to compare different interpretation paths. For brevity of exposition, we focus here on one single case. We saw that "flat tax" is a term appearing in many inferential models, even if it originates from the electoral discourse of Italian right. Its mention alone therefore does not say much about how different model readers may give a value to it. Hence, we tried to unpack the semantic frame (Fillmore 1976, Eco 1979) of the expression "flat tax". To prevent potential biases in our findings, we started by taking a step back and working on the more generic word "tassa" (tax). This allows us to

observe a larger frame about taxes in our corpus. Figure 9 displays the indirect inferences realized by the politically undecided bot.

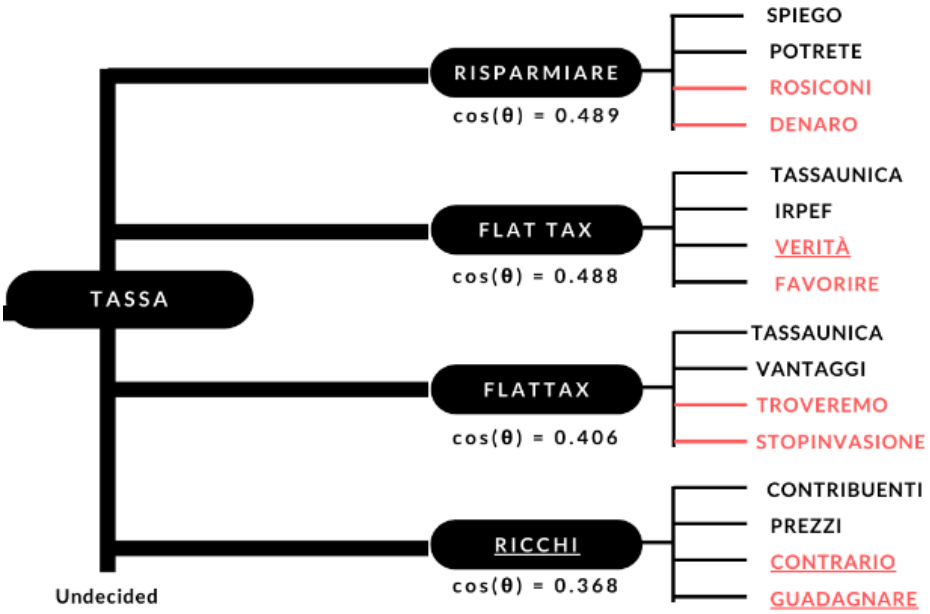


Figure 9: Examples of inferential paths made by the politically undecided bot

We analyzed qualitatively the resulting spaces of word embedding to identify the positive or negative axiological perspectives given to the main word vectors. We also manually distinguished embedded and necessary inferences by analyzing collocations. In Figures 9 and 10 we show the embedded inferences in red and the necessary ones in black. We selected four out of the ten most similar words, including at least two embedded inferences for each. This allows us to visualize and possibly better understand the ideological differences of the six bots. The underlined words evoke negative axiological frames in the word embedding model of this bot.

In the case of Figure 9, showing data from the politically undecided bot, three out of four of the main inferred words are positive or neutral towards the term “flat tax”. If we expand further the frame of each sub-word, it becomes clearer that positive evaluations are predominant and that the ideology proposed to this bot is close to that of the Italian right. For instance, the word “verità” (truth) is associated with a semantic frame of

mistrust and falsehood. This method allows for manual comparisons among different inferential models, as it can be shown by visualising the inferences induced by the same term “tassa”(tax) in the far-right bot (Figure 10).

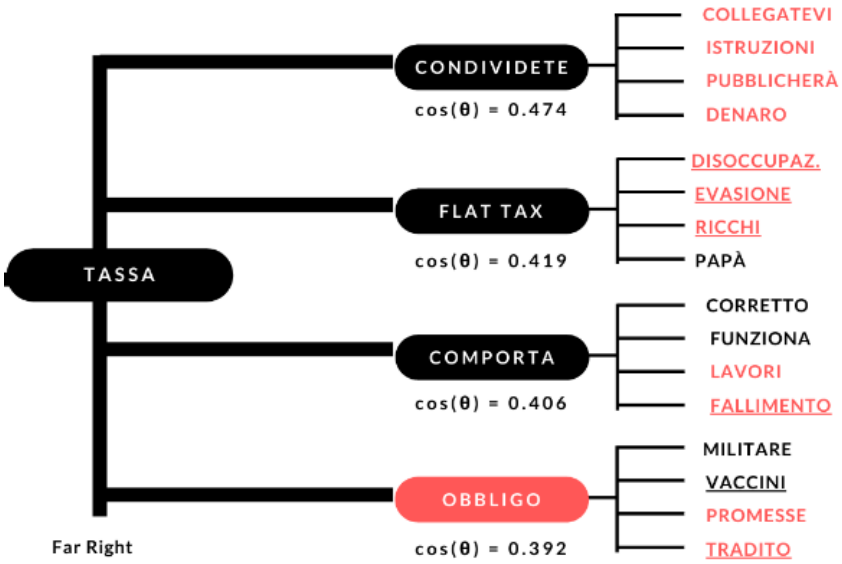


Figure 10: Examples of inferential paths made by the politically undecided bot

In Figure 22 we observe a different axiological configuration compared to Figure 8, as “flat tax” leads to embedded inferences such as “disoccupazione” (unemployment), “evasione” (tax evasion) and “ricchi” (the rich), inducing a negative evaluation of introducing a flat taxing system in Italy, as it would become an advantage only for the rich. We reiterate that the inferential models that we can extrapolate with word embedding are complex to interpret, and for now they can only be used as tools for supporting quali-quantitative analysis.

3.2.2 Discussion

With our study we distinguished how different bots infer some given words from others, depending on the information they are shown by the FB algorithm. Despite the expected differences among them, we also found evidence that, during the 2018 campaign, in Italy, the right-wing arguments were dominant across all different bots, showing that algorithmic personalization may actually propose to readers information in

contrast to their individual political perspective and oriented by more global tendencies. If confirmed, these results would be important as it would show that Facebook influences the public discourse as it performs a sort of automated agenda setting skewed towards the most popular perspectives.

It may also be that what we observed was an artefact due to our choice of using word embedding. In this case it would still be interesting to understand why right-wing information managed to affect the construction of all the other vector spaces. This may be due to two factors: repetition and corpus size. First, in the right-wing sub-corpus there are a lot of reposts and more generally repetitive content. Words that occur a lot may then be overrepresented, and the interpretation of more rare words risks being problematic. However, the word "flat tax" despite being associated with the Italian right-wing discourse, does not appear in our top-20 most similar words list of the inferences of the right-wing bot. This might be caused by the size of the right-wing sub-corpus, which was the largest; hence the probability of occurrence of the expression "flat tax" is smaller compared to the other dataset. It is anyway evident that Facebook algorithms spread right-wing content more, since "flat tax" is found in the models of all other bots.

Despite the fact that the sub-corpora are of different size, we managed to compute acceptable semantic models, capturing meaningful semantic relations for every bot. However, the quality among models is variable. An effect of corpus size may explain why the term "vino" [wine] appears in one of the models, apparently without any reason. We also observed what was suggested by Rogers (2018): the use of ideologically charged terms differentiates vocabularies across bots, making it difficult to compare them with purely computational methods.

As this is a first attempt to formalize the concept of model reader so to simulate it with word embeddings, we do not have enough data to validate the effectiveness of our

simulation. Anyway, literature is still quite uncertain about evaluation methods for word embeddings (Mikolov et al. 2013b, Gimenez et al. 2015, Lai et al. 2016, Sahlgren and Lenci 2016, Faruqi et al. 2016, Naili et al. 2017, Wang et al. 2019) and, to the best of our knowledge, there are no evaluation methods to evaluate the accuracy of a pragmatic model. In general, for our ends, we did not notice any relevant difference between the use of skip-gram or C-BOW models, nor between negative sampling and hierarchical softmax. Hierarchical softmax could allow us to control better for repetitive content, but to validate this assertion a dedicated work is needed.

3.3 YOUTUBE

Most of the content of this section is published in: Sanna, L., Romano, S., Corona, G., Agosti, C. (2021). YTTREX: Crowdsourced Analysis of YouTube's Recommender System During COVID-19 Pandemic. In: Lossio-Ventura, J.A., Valverde-Rebaza, J.C., Díaz, E., Alatrista-Salas, H. (eds) Information Management and Big Data. SIMBig 2020. Communications in Computer and Information Science, vol 1410. Springer, Cham. https://doi.org/10.1007/978-3-030-76228-5_8

Every day, we use a large number of services that use algorithms to select relevant information for different users. YouTube is no exception, as it uses an algorithm that decides what might be most important for each one of us.

However, we have a transparency issue, meaning that we do not know which criteria YouTube uses to operate its selection. YouTube provides some information about its algorithm, but just related to the general structure of the recommender system algorithms (Covington et al. 2016, Zhe et al. 2019). In other words, we do not know how these algorithms work, nor how they decide whether the information is relevant or not.

We believe that is important to remark that algorithmic personalization is not a technological issue but rather a problem of opacity; in fact, we do need algorithms, as we have to select an enormous amount of information daily in our online experiences. This problem of opacity may seem trivial when using algorithms in our spare time on

entertaining services, but the issue becomes more serious when the same opacity applies to the selection of political news or other sensitive content that impact on social behavior.

As we shown in section 1.2.2, empirical research on algorithmic personalization is still quite fragmented and we believe that this happens because of the lack of a shared methodology among the researchers. We propose then a novel approach, using a tool that collects evidence of the personalization that happens within YouTube to explore the filter bubble effect within the platform. In the following section we illustrate the functioning of this tool, that is the YouTube Tracking Exposed (YTTREX) browser extension. Finally, in sections 3.2.2, 3.2.3 and 3.2.4 we present an experimental setting in which we use YTTREX to collect our data and explore the filter bubble effects.

3.3.1 YouTube Tracking Exposed

3.3.1.1 Why a browser extension

The majority of studies in the literature do not really focus on the filter bubble, because they use methodologies to study user behavior (Abisheva et al. 2016, Airoldi et al. 2016, Song et al. 2017, Rieder et al. 2018, Bishop 2018, Arthurs et al. 2018).

On the contrary, algorithmic personalization is essentially a passive phenomenon; users are subject to personalization and therefore the only way to approach it is to study the full range of user experiences. For its part, YouTube provides information about its algorithm, which is, however, simply a description of its general structure. The site also provides an official API[19], often used by researchers, but this may differ significantly from actual user experience.

Therefore, there are two ways to approach algorithmic personalization.

1. With "bots", synthetic profiles specifically designed for research

[19] Acronym for « Application Programming Interface". On Social Media we can use API to retrieve content from the platform, if allowed.

2. By using real people to carry out a collective experience (Crowdsourcing).

The crowdsourced approach is, in our opinion, the most appropriate for obtaining evidence of the filter bubble, as it provides a general picture of the fragmentation of content recommendations.

In order to have a crowdsourced approach, we need a tool that can collect data directly from the users' screens while they're experiencing content within the platform. The only way to do that is to use a web scraper. In our experiments we are used a browser extension developed by tracking.exposed (TREX) to collect data on YouTube. The following section explains in great detail how does this tool work.

3.3.1.2 A deep dive into YTTREX

The browser extension (add-on) of Tracking Exposed[20] collects evidence from the metadata that is observable on the web page when the user lands on the homepage, watches a video, or does research on the YouTube website. The data collection is completely anonymous; the privacy is ensured by creating a cryptographic key pair that allow each user to access her/his private data. The tool collects separate contributions for each browser with the add-on installed.

The data are collected in three phases:

1. **Collection**: the add-on takes a copy of the HTML when the browser is watching a video. Four buttons appear on the top left of the screen (Fig.11), when the add-ons are installed and enabled by the popup. The color code represents the different status.

[20] https://youtube.tracking.exposed, AGPL3 code: https://github.com/tracking-exposed/yttrex/

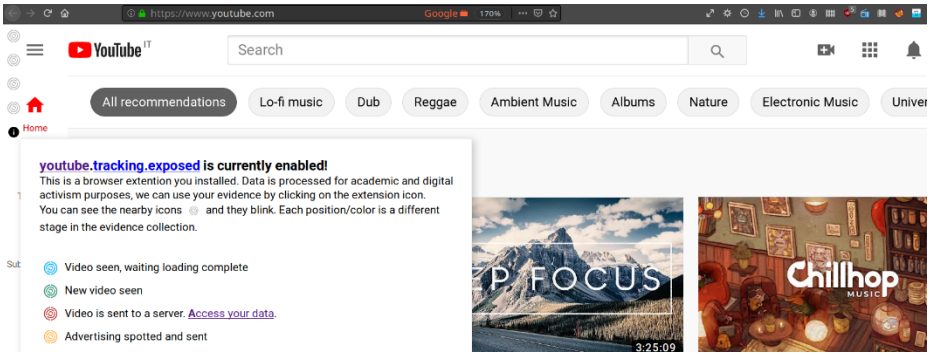


Fig. 11: Screenshot of what the browser extension shows while navigating on YouTube.

2. **Parsing**: server side, the HTML is processed and metadata are extracted. The information is then organized in a dataset. In the HTML there are many different data that might be analyzed to extract metadata. We did not yet extract all possible information, especially we avoided any unique tracker that might become personal data if collected. On the other hand, the YTTREX project still has room for improvement, and we might not have yet mapped 100% of the potentially interesting metadata for YouTube algorithm analysis.

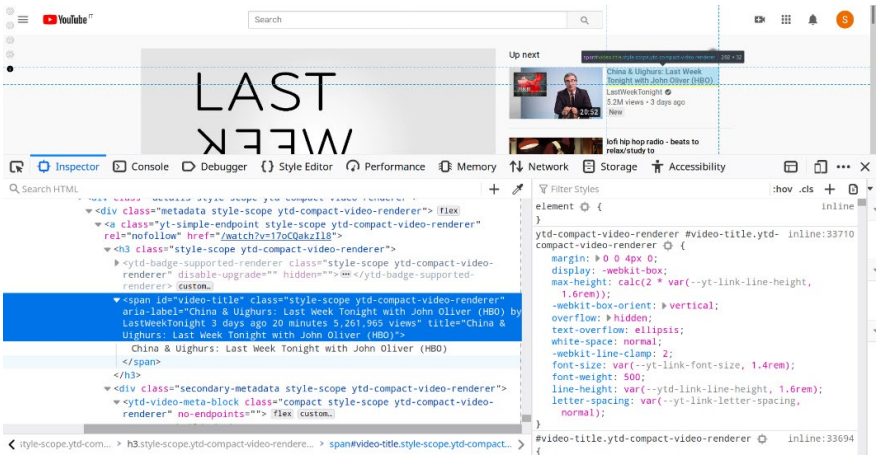


Fig. 12. HTML inspection of a recommended video on YouTube and its aria-label

We were also able to record the users' interface, detect the language, record related videos, the number of views, and duration. Inspecting the HTML of a recommended video

(Fig. 12), you might see the data field named aria-label[21]. This text field is meant for accessibility and contains a compacted, but human formatted, set of information useful for researchers. Because of the localization, YouTube produces aria-label with strings that change accordingly to the user interface Language. For example, the aria-label: "*Crise pétrolière : coup de poker sur l'essence | ARTE by ARTE 6 days ago 58 minutes 213,982 views"* is composed by the information shown in Table 3.

Title	Crise pétrolière : coup de poker sur l'essence | ARTE
UX Language dependent *stopword*	by
Publisher name	ARTE
Relative human readable publication time	6 days ago
Human readable video length	58 minutes
Number of views formatted as per UX locale standard	213,982 views

Table 3. Aria-label composition.

We might externalize this natural language conversion, managed by our aria-label parsing library[22], as an independent library, once we figure out how to maintain the list of fixed terms that scale up proportionally to the language supported by YouTube. The

[21] For reference see: https://mzl.la/33dMuRN
[22] https://github.com/tracking-exposed/yttrex/blob/master/backend/parsers/longlabel.js

sum of session information, video watched, and recommended videos, produces the data unit with the format detailed in Table 4.

Field Name	Data Type	Description
login	Boolean	True if the profile was logged on YT
id	String	Unique identifier for each installed extension
savingTime	ISODate	GMT hour when evidence get saved
clientTime	ISODate	Date on the users' browser
uxLang	ISO 639-1 code	Browser language
recommendedId	String	Unique identifier of the data unit
recommendedVideoId	String	Video unique ID used in YT URL
recommendedAuthor	String	Publisher of the recommended video
recommendedTitle	String	Title of the recommended video
recommendedPubTime	ISODate	Date of recommended video publication
recommendedRelativeS	Number	Seconds between recommended publication and

		access to watch the video
recommendedViews	Number	Views at *savingTime* for the recommended video
recommendedForYou	Boolean	True if YT explicitly says recommended for you
recommendedVerified	Boolean	True if publisher has the blue check ✓
recommendedKind	String	Live streaming or video
recommendedLength	Number	Duration of the video in seconds
recommendedDisplayL	String	Human formatted duration of video
watchedVideoId	String	From YT URL, the Video ID
watchedTitle	String	Title of the watched video
watchedAuthor	String	Publisher of the watched video
watchedChannel	String	Relative URL of YouTube channel
watchedPubTime	ISODate	Publication time of the watched video
watchedViews	Number	Amount of views at *savingTime*
watchedLike	Number	Amount of thumbs up at *savingTime*
watchedDislike	Number	Amount of thumbs down at *savingTime*
sessionId	String	Unique identifier of users' sequence
hoursOffset	Number	Amount of hours after the 25 March 2020 GMT, the beginning weTest1
experiment	String	'weTest1', the experiment of this paper
pseudonym	String	A unique pseudonym for each browser plugin
top20	Boolean	True if *recommendationOrder* < 20

isAPItoo	Boolean	True if recommended is also in YT API related
step	String	Human readable language of watched video

Table 4. Data structure

3. **Research and data-sharing**: YTTREX was created to support independent analysis and privacy-preserving sharing of the algorithmically powered circulation of videos. Every video observation has a dynamic number of related videos (if the watcher scrolls the video page down, the browser loads 80 or more related videos, but for users who do not scroll down the default is to receive and display only the first 20 related videos). Every related video becomes a single row, a data record with its own unique ID. Interconnecting these with *metadataId*, the researcher might re-group all the related videos belonging to the same evidence, as they were displayed to the watcher. Certain fields such as logged, pseudo, and *savingTime*, are the same across the same id because they depend on the collection condition. *recommendedVideos*, *recommendedAuthor*, and other recommended-fields, changes in each row according to the related video described*; recommendedId* is generated for each row and should be used as guarantee of unique field.

According to the definition provided by Sandvig (2014), the tool enables the user to potentially four of the five methods of algorithmic audit: Noninvasive User Audit, Scraping Audit, Sock Puppet Audit, if they have the know-how to use bots, and Crowdsourced or Collaborative Audit, as the experiment presented in this section.

The database collected for this paper is available on Tracking Exposed website[23] and the code is available on GitHub[24] protected by AGPL v3 license.

3.3.2 Approaching the filter bubble on YouTube. Does it exist?

We used a crowdsourced approach to test whether there was empirical evidence of filter bubbles on YouTube. We made a call for participation on our website to select the participants[25]. Every participant joined the experiment for free and voluntarily. The procedure involved the following protocol:

1) The participants had to install and enable the YTTREX extension
2) Each participant had to watch five videos about COVID-19 prevention, produced by the BBC channel, one for each of the most spoken languages in the world: Chinese, Spanish, English, Portuguese, Arabic.
3) The YTTREX extension recorded the recommendations made to each user.
4) The participants shared their data collection to compare the different recommendations.

Originally the idea of this experiment comes from our doubt that YouTube could not effectively take down conspiracy theory on COVID-19[26], differently to what is claimed. We suspect English language and recommendation might benefit from a better curation, thus by comparing the recommended videos close to equally accurate COVID-19 videos. Still, in different languages, we could neither confirm nor reject the hypothesis.

We did not provide additional information about the minimum time that had to be spent watching the videos: loading the page was enough to collect the HTML. Participants could choose to perform the test logged with their personal account or without, the tool

23 https://youtube.tracking.exposed/data/
24 https://github.com/tracking-exposed/youtube.tracking.exposed
25 https://youtube.tracking.exposed/wetest/1
26 https://www.nytimes.com/interactive/2020/03/02/technology/youtube-conspiracy-theory.html

records if the user is logged or not, without collecting any data related to the specific account.

The same day of the test, we retrieved via the official YouTube's API the related videos for the five videos included in the methodology.

Since language is an option for the API request, we performed five requests, one for each language. 50 videos were retrieved in each API request. We then stored this information using the metadata *isAPItoo* (see Table 4) for each of our evidence collected via YTTREX.

3.3.3 Evidence of filter bubbles

The distribution of the recommended videos is clearly skewed as shown in Fig. 6. We investigated the distribution of recommended videos considering the language of the starting video, the browser's language, and considering whether the user was logged or not. No matter of which variable we took into account we always obtained a skewed distribution, as shown in the example of Figg. 13-14.

Fig. 13: Frequency distribution of recommended video in our dataset.

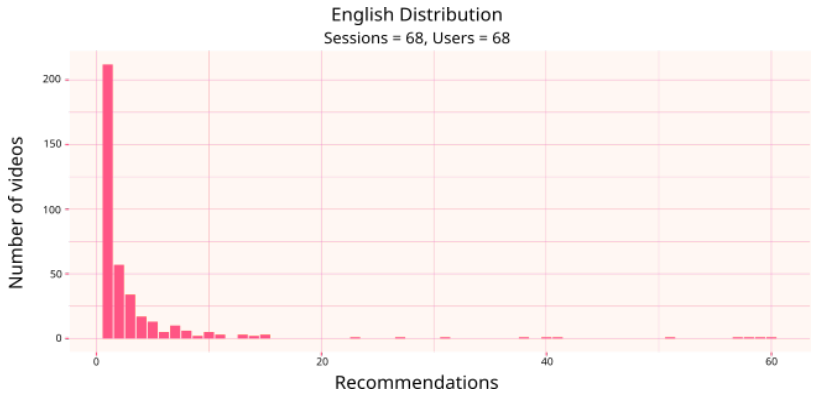


Fig. 14: Frequency distribution of recommended videos starting from the BBC video in English.

Our findings show that the vast majority of videos are recommended very few times (1-3 times), regardless of the variable considered. This distribution is significantly positively skewed according to Fisher's skewness coefficient (>2). Summing up, 57% of the recommended videos have been recommended only once and only around 17% of the videos have been recommended more than 5 times during our experiment.

These results highlight that the filter bubble is real, and that algorithmic personalization produces a high fragmentation of recommended content among YouTube users. Another relevant finding for the study of algorithmic personalization is the huge difference we found using YT API and our tool. For users logged into their Google account only 11% of the recommended videos could be retrieved using the API as showed in the following section, in Fig. 18.

Finally, we calculated Lorenz curve (Lorenz 1905) [27] over the distribution of the recommended video, confirming that the inequality in the distribution (Gini > 0.5) of the recommended videos. In a nutshell, the Lorenz curve show us that very few videos are recommended to more than one user. This is empirical evidence that might prove the existence of the filter bubble on YouTube.

[27] The Lorenz Curve is a method to measure distribution of wealth and it is used along with Gini index to measure social inequalities. However, this method finds a quite handful application also in measuring the inequality of the video recommendations' distribution.

We also calculated the Gini index for the number of videos selected for each user, since the result shown in Fig. 15 might be caused by an uneven number of videos selected for each user. However, with a Gini coefficient around 0.2, we have evidence that the algorithm is selecting an equal number of videos for each user, while distributing unevenly the recommendations for each video.

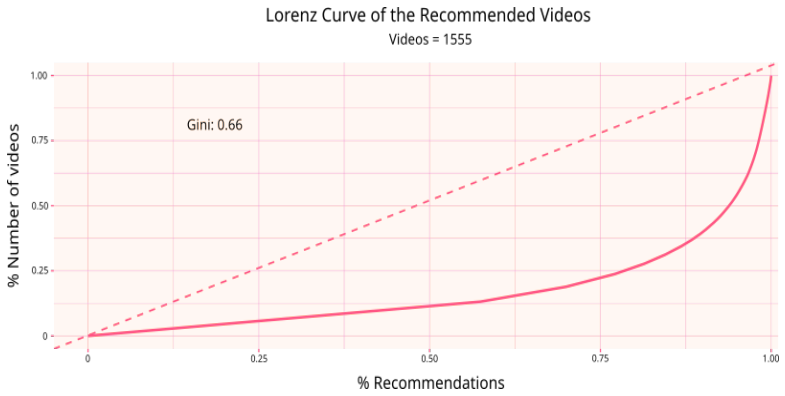


Fig. 15: Lorenz Curve and Gini Coefficient of the recommended videos

Finally, we performed a network analysis using Gephi (Bastian et al. 2009) to better understand and visualize how the recommender system creates a filter bubble around users watching the same video the same day. Thanks to the Medialab's tool *Table2net*[28] we extracted a network file from the csv file. We created a bipartite network linking two types of nodes: users' pseudonyms and suggested video's ID.

In the graphs (Fig. 16, 17, 18) we used a circular layout algorithm (Six et al. 2006) to dispose of all the users in a circle. We aimed to show all the participants in the same positions, pointing in the same direction, because they were performing the same task: in the examples they are watching the video from the English version of BBC channel "How do I know if I have coronavirus? - BBC News.". This representation allowed us to show how, even if they were all watching the same video, they were getting a different configuration of suggested videos.

28 https://medialab.github.io/table2net/

The size of the nodes is based on the degree of each node: in a range between size 15 and size 60, each user and each recommended video is big in relation to the number of links that it has. The videos in the center of the graph are bigger because they have been suggested more than the others. Because of a graphical compromise, the nodes with a degree minor than 15 have the same shape, likewise the nodes with a degree higher than 60 are all the same.

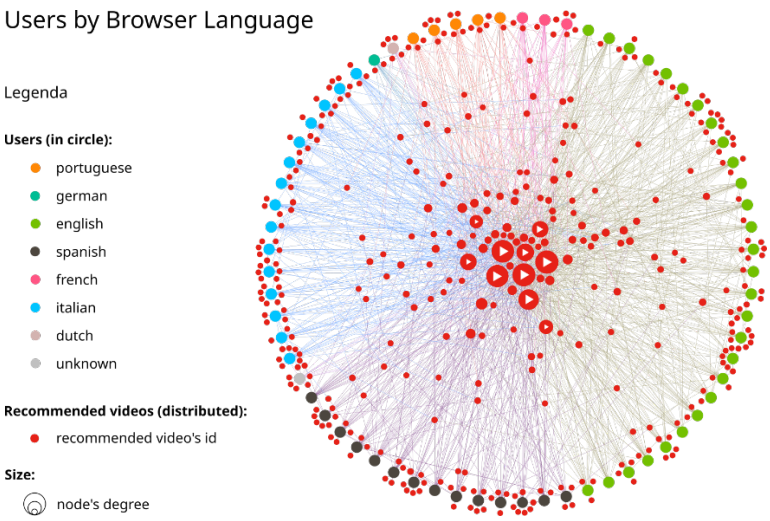


Fig. 16: Graph of the videos suggested to the participants while watching the video "How do I know if I have coronavirus? - BBC News"

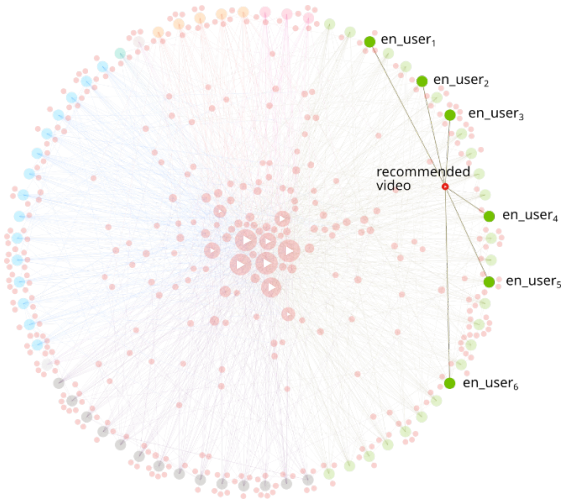


Fig.17: Zoom of Fig. 9, an example of video suggested only to users with English interface.

In Fig. 7 we highlighted how some of the videos recommended appear only to users with English browsers. This shows that the participants in the experiment received

personalized suggestions according to their characteristics, despite watching the same video. This type of analysis can demonstrate differences in the users' experiences tracing the most influential features that can generate changes in the platform experiences.

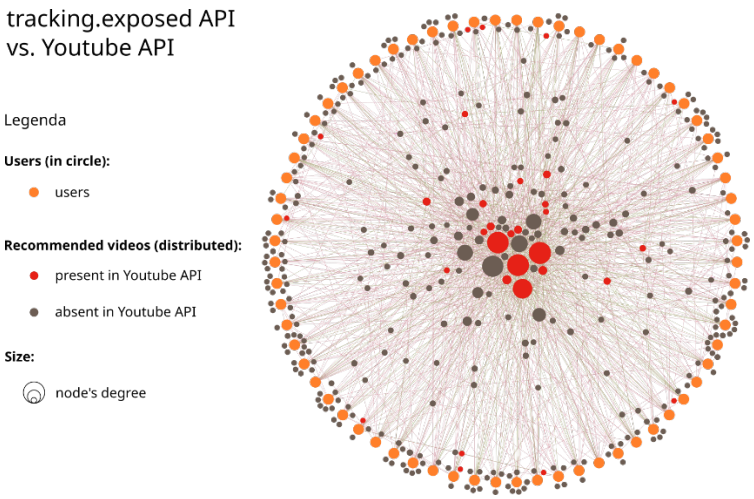


Fig. 18: Same graph of Fig. 9, here the colors highlight the differences between the videos recorded with Tracking Exposed and the ones retrieved with YouTube official API.

As we already said in the previous section, there is a huge difference between the recommended videos that we retrieved from the API and the actual recommendations (Fig. 11). The majority of the videos retrieved by the Tracking Exposed tool are not present in the database created with YouTube's API. Some of the most suggested videos (biggest nodes in the center of the graph) neither. This is relevant because it is evidence against the usability of official YT's data in academic research. The official API cannot represent the real variance of suggestions present in the actual recommended videos. Many scientific articles (Brbić et al. 2012, Ledwich and Zaitsev 2019, Marchal et al. 2020) rely on these data to explain the circulation of videos on the platform, but according to our findings we might say that API data are just a generic representation of an ideal user that is really difficult to find in reality (no one of the users in our experiments gets the same recommendations as in the API).

The official API does not represent the various levels of personalization that occur in relation to the structural users' characteristics and to their past online behaviors. Thus, we cannot use API data to make inferences about personalization, polarization and filter bubbles, because these phenomena presuppose the study of real users in real context.

3.3.4 Experiment on the American elections

The second step on YouTube filter bubble, once confirmed its existence, was to verify algorithmic personalization within the platform in the run-up to the inauguration ceremony of new US President Joe Biden in January 2021.

The data collection was divided into two phases. In the first one - the personalization phase - our goal was to simulate echo chambers on YouTube and the influence of the filter bubble. To do this, the research group simulated behavior on YouTube, watching ideologically oriented content. All the participants in the project were divided into two groups (15 people), each group representing a different orientation in American politics (conservative / republican and progressive / democratic). The classification of "progressive" or "conservative" videos and channels was based on the one done by the project transparency.tube. Users were randomly assigned to each group to simulate browsing behavior. Using a clean browser[29], each user watched six videos from channels considered progressive or conservative depending on the assigned group.

In the second phase - the filter bubble simulation phase - all users performed three searches on YouTube:

1. 1 US elections
2. Coronavirus
3. New Year (as control variable)

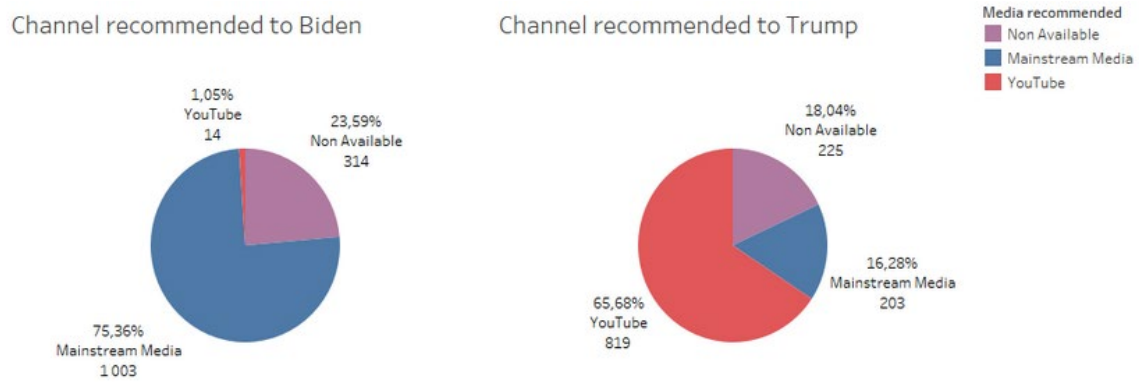
[29] Meaning a browser without cookies or google login, so that it could not be linked to the actual users' profiles.

In this section we will discuss only the US election query, as during the experiment happened the capitol hill assault. Therefore, surely the most interesting part is to look for evidence of filter bubbles and polarization within this particular topic, due to its newsworthiness and sensitivity at the time of the experiment.

3.3.4.1 Results

As can be seen in the images, the two groups received different content with regard to both the types of media (Fig. 19) and the political orientation of the content (Fig. 20)[30].

Fig. 19: Media types suggested to the Biden and Trump profiles. Biden profiles gets 75% of recommendations of



mainstream media (e.g., CNN) while Trump profiles gets 66% of native YouTube content. It should also be noted that native YouTube videos are recommended to Biden profiles just 1 out of 100.

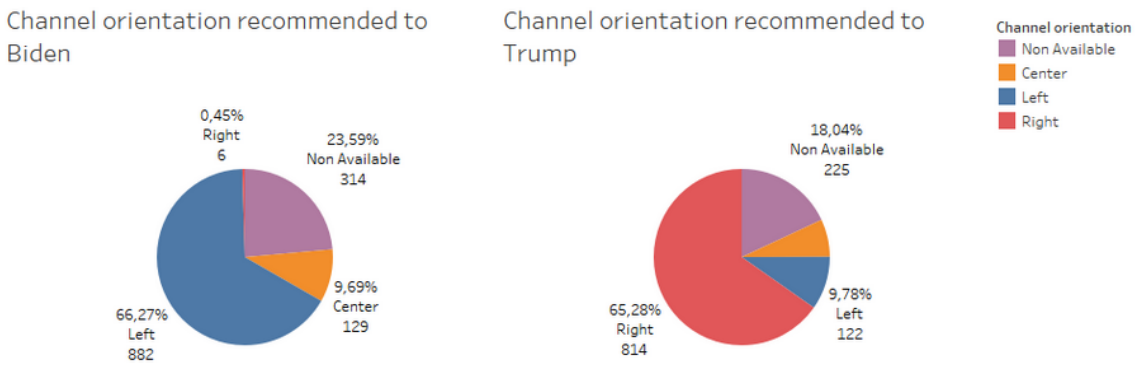


Fig. 20: Political orientation of the channels suggested to the Biden and Trump profiles. Both groups get a vast majority of content coming from their political part, while very few content of the counterpart is recommended.

[30] The full experiment report is available at https://wiki.digitalmethods.net/Dmi/WinterSchool2021FIterTube

During the experiment, we also built a corpus of comments of our videos, retrieving comments via the official API. This resulted in around 400,000 comments (Table 5), which were split between videos suggested to "progressive" users and those suggested to "conservative" users. In order to optimize the analysis, to highlight the difference between them, we have excluded the comments from the videos recommended to both of them. For the query "US election", this amounted to about 65,000 comments for "progressive" suggestions and 130,000 for "conservative" suggestions.

	Tokens	**Types**
Corpus Conservative (Trump)	3,844,007	78,854
Corpus Democratic (Biden)	2,999,920	75,882

Table 5: Corpora overview

A first analysis allowed us to highlight the linguistic differences and therefore to select the most relevant words for each corpus. In this case, we resorted to keyword analysis, using WordSmith software. There is strong evidence of statistical significance when the BIC score is over 10 (Gabrielatos 2018).

KW prog.	**Freq. Prog**	**Freq. Cons.**	**KW Cons.**	**Freq. Cons.**	**Freq. Prog.**
Vice	3.177	470	Romney	1.490	25
Farmers	984	35	Gary	1.393	14
India	1.831	439	Fox	2.464	299
Prize	599	43	Christmas	1.599	133
Indian	852	189	Pat	1.024	19

Nobel	472	25	Fauci	970	42
Police	1.460	609	Bill	2.261	504
Mars	425	27	Mitt	795	16
Native	439	48	Merry	877	38
Nuclear	559	107	Gates	1.133	129
Kashmir	341	20	Governor	1.183	151
Boys	639	164	Pelosi	1.187	170
Hong	383	39	Robertson	556	2
Districtrep	250	0	Sharpton	544	6
Seattle	360	34	Patriots	1.100	164
Kong	353	40	Omar	539	9
Puertorican	298	25	Barr	960	123
Farmer	260	16	Melania	724	54
Fbc	197	0	Swamp	1.082	195
Algorithm	321	41	Nancy	818	105
Pakistan	764	319	Graham	610	42
Indigenous	215	5	Juan	454	10
Kilo	195	1	Tucker	431	10
Kevin	270	29	Maxine	432	12
Coffee	334	61	Elijah	363	1
Oil	416	114	Gitmo	459	21
Fur	173	2	Michelle	381	7

Punjab	177	5	Lindsey	466	26
Cocaine	254	37	Blm	1.319	358

KW cons .: conservative keywords

KW prog ..: progressive keywords

Freq. Cons. : frequency in the conservative corpus

Freq. Prog: frequency in the progressive corpus

Words are ordered by BIC score.

Table 6: Keywords in the two corpora

The keywords indeed show a deep differentiation with regard to the subjects treated (see Table 6). Among the key words used by conservatives, we notice a clear majority of proper nouns, all referring to political actors or other persons relevant to the American political debate (for example, Bill Gates). On the contrary, in the progressive corpus, most of the most significant keywords are associated with topics of the political debate, with a clear predominance of international rather than national topics (for example, the protests of the peasants of Punjab).

We then explored the keywords using word embedding. We created two semantic models, one for each corpus. Next, we calculated the thirty most similar terms for each of our keywords; this allowed us to compare the semantic framework of each word for the two parts, therefore the respective visions proposed to the two parts.

The semantic frame is a typical context with which the word is associated. This context has a fixed part (Violi 2000) and a variable part. Taking the example of Trump. to understand the term, you must at least know that he was president of the United States; this part is the fixed semantic component of the framework. On the other hand, the variable part is determined by the cultural context; for example, a conservative context

would activate, along with the fixed component, other semantic traits related to a positive assessment of Trump's policy, while a progressive context would probably do the opposite.

Word embedding allows us to highlight these differences between semantic frameworks and also to explore other pragmatic relationships. Romney, a Republican senator, seems to have a bad reputation in both corpora, although in the conservative corpus he is associated with a very specific semantic field: betrayal. This is particularly clear because Romney's name is associated with the adjective "TRAITOR," which of course is an explicit negative assessment. More precisely, another term that appears in its semantic frame is "RINO", sometimes also spelled "RHINO" which is the acronym for "republican in name only", yet another indication of betrayal towards its political ideology. Another word used in conservative comments to express negative assessment of Romney is "SPINELESS". On the other hand, in the progressive corpus does not emerge an image of Romney as a traitor, despite the presence of certain negative words.

While on most other political actors, nothing particularly relevant to our subject emerges, Nancy Pelosi, former president of the American parliament, brings out a certain polarization. In fact, in the conservative corpus, Nancy Pelosi is associated with the word “MALFEASENT” (evil), while in the progressive corpus her name is closely related to the adjective “EXTRAORDINARY” (extraordinary). On the contrary, Melania Trump in the conservative corpus is defined as a chic and elegant woman, while the former first lady in the progressive corpus she is associated with other negative terms such as drugs and divorce (Fig. 21).

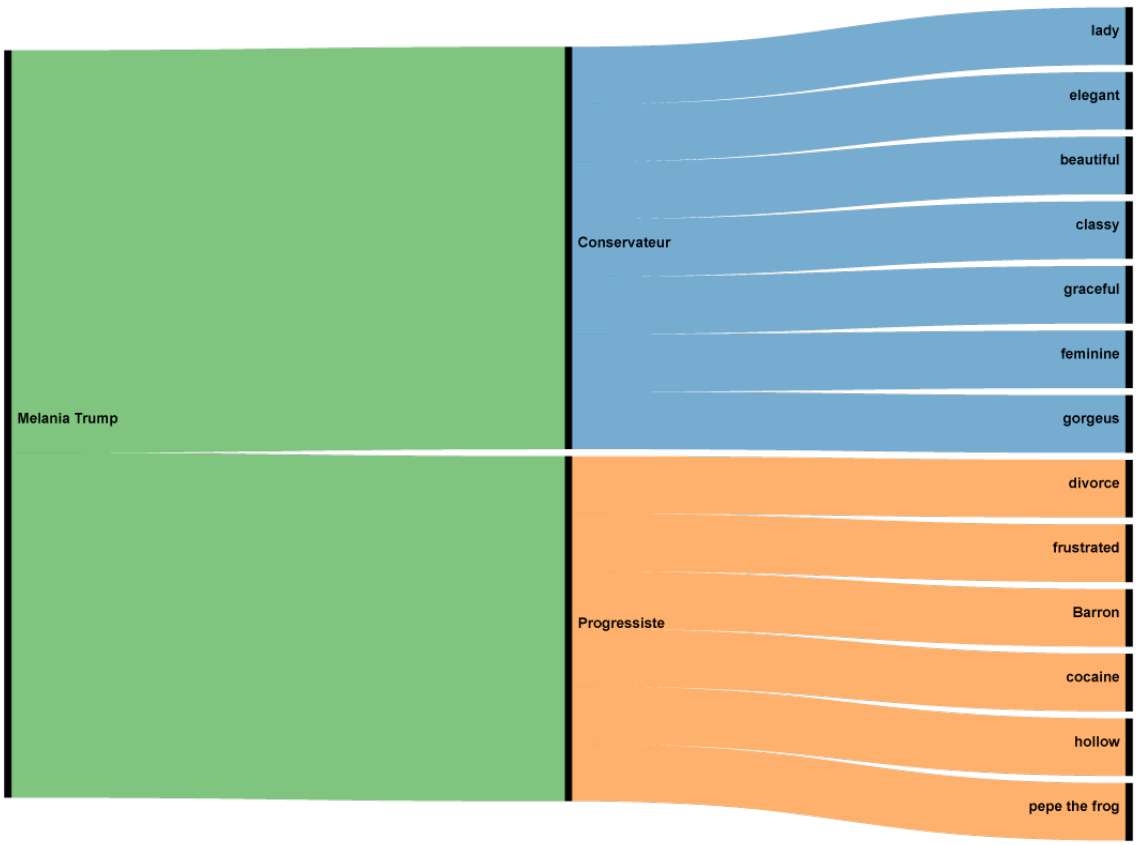


Fig. 21: The different semantic frames of Melania Trump in the two corpora

3.3.4.1.1 Polarization and themes

We delved into conservative keywords, finding varying degrees of differentiation. We had indeed highlighted other clues of polarization in "Fox", clearly referring to the media "Fox News", a prominent conservative media in the United States. This media is a good example, because in the semantic framework of the conservative corpus, Fox News has no association with the evaluative language, but it is nevertheless relevant because it evokes the lexicon probably related to trending topics, news and people. who work in the media. Conversely, in the progressive corpus, the word "FAKE" is among the most similar words associated with Fox News and also other terms such as "BIASED" and "PROPAGANDIST". However, in the conservative corpus there is a small exception with the French word "faux", in a smaller set of comments (around 130), referring to Fox News as "FAUX NEWS". These comments are actually an even more extreme part of the conservative body of

work, as they come from conversations associated with far-right conspiracy theories like Q-Anon. The phrase "FAUX NEWS" appears only twice in Biden's corpus, confirming that it is an ideological keyword of right-wing commentary.

While considering other keywords we found that the differentiation among the two corpora was quite weak and this is probably due to the fact that the keywords refer to very different topics. In fact, the majority of progressive keywords relate to international topics other than the US elections.

We again find traces of polarization in the lexicon associated with sensitive topics in the United States, as is the case with the term "police", in the progressive corpus. The fact that this term is a keyword is probably already a sign of political polarization, as it is an indicator of political preference, with police brutality being a topic often addressed by left-wing politicians. However, this is perhaps one of the most interesting terms for polarization because it shows the complexity of the phenomena. At first glance, the two semantic frames may seem similar, as they share certain terms like "COP" (cop) and definancing. However, these terms themselves have a very different framing in the two corpora. On the one hand, in the conservative corpus, "COP" has a neutral, if not quite positive, meaning, as it is used as a sort of familiar form for police officers. Definancing, on the other hand, has a negative meaning for the conservatives, because it is associated with crimes. In the progressive corpus, the word has on the contrary a negative framing, because it is associated with words such as "impostors", "criminals" and apologists of fascism. In addition, definancing is associated with police reform and the theme of police brutality which seems to be completely absent from the conservative corpus.

Another word that has shown a high degree of polarization is the keyword "patriots" in the conservative corpus. In this case, the polarization is perhaps the most extreme: if in the conservative comments the patriots correspond to the "American people", in the

progressive comments, on the other hand, they are "traitors" and "spies". This term is mainly associated with the invasion of the Capitol and there is a very strong polarization (Fig. 22).

Therefore, we can say that in the two respective filter bubbles, the algorithm effectively created the conditions of political polarization, offering each part videos with comments with a coherent context according to the positioning of the user, simulated with online behavior. Exploring the comments allowed us to gain insight into the worldview YouTube offered to both groups. In fact, comments allow us to explore a textual dimension that contextualizes the content of the video, sometimes even more polarizing than the video content itself. The added value of exploring the comments is that it allows to understand the points of intersection between algorithmic personalization and echo chambers, allowing to highlight which social contexts are proposed by the algorithm, depending on the assumed political orientation of each user group.

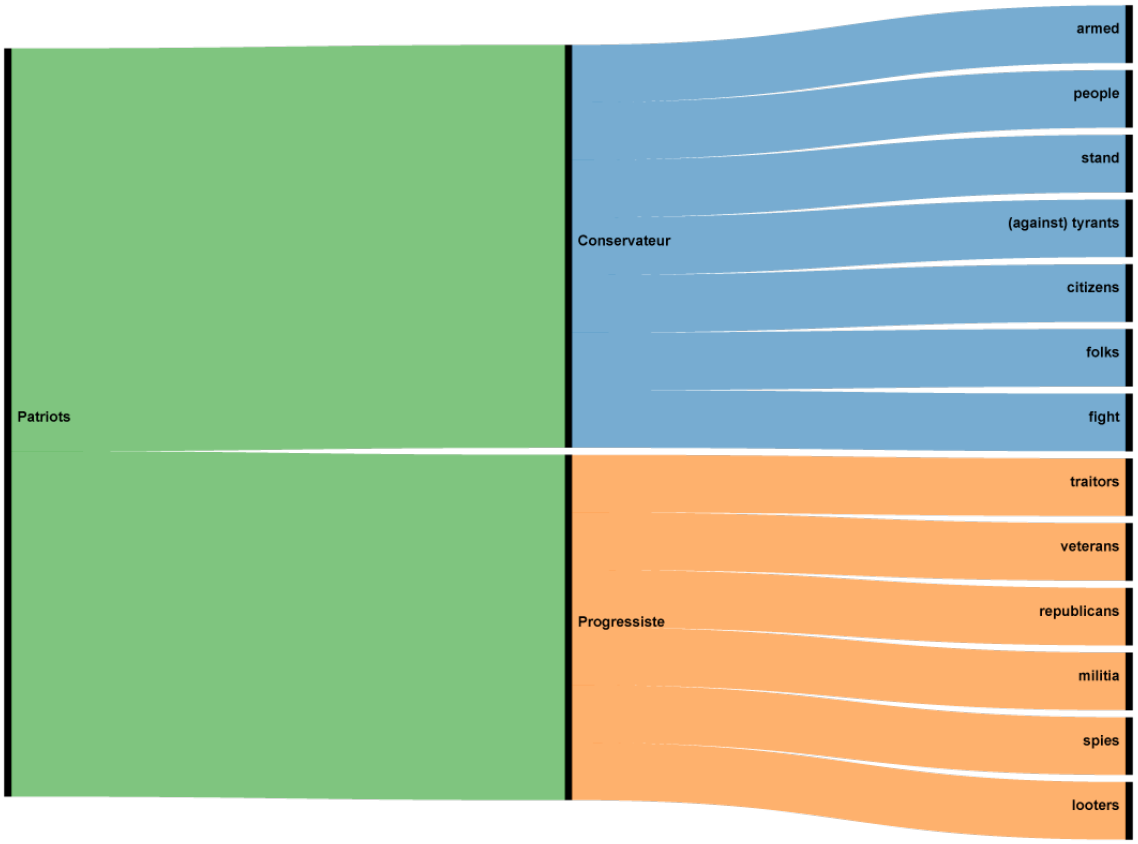


Fig. 22: Polarization on the word "Patriots"

3.4 CONCLUSIONS

The Facebook experiment showed that Eco's model reader (1979) might be formalized as an inferential model. In principle, this means that it can be simulated with word embedding to create computational inferential models, that can be built empirically and compared so to identify a possible divergence of interpretation paths.

The use of word embedding is not only useful but, in our opinion, necessary. We believe that it should be preferred to other similar methods (i.e., collocates) for two main reasons. First, we cannot explore the inferential paths of entire topics using collocates, as we did in 3.3. Second, but equally important, the semantic framing that emerges from word embedding is already cleaned from terms with high frequency and it shows a probability of co-occurrence, whereas collocates might miss semantic relationships among words that rarely co-occur. Finally, the fact that word embedding is a probabilistic model make it more accurate in reproducing the theory of the model reader; considering only semantic preference, thus actual co-occurrences, would be a limited and rather biased representation of a model reader.

We obtained a confirmation that political sources are treated unevenly by the Facebook algorithm; particularly, we collected evidence that in 2018 Italy, right-wing vocabulary had spread better than the others. We also found empirical evidence of YouTube's filter bubble existence and our experiment showed that users where exposed to conversation that might result ending in polarization dynamics. Although in this case we explored the framing of single terms, we may say that we are still exploring the inferential model proposed to each group of users, thus their model readers.

Our experiments showed that algorithmic personalization seems to impact on model readers, differentiating the most relevant lexicon of different text fluxes and proposing a remarkable ideological differentiation to different group of users.

4 ECHO CHAMBERS

4.1 INTRODUCTION

As we explained in chap. 1, we believe that echo chambers could be approached looking at them through markers of linguistic conflict. This conclusion has been reached with a study based on substantial samples of the Coronavirus corpus (Davies 2019-), available from English Corpora, along with a sample of a large corpus of tweets[31]. We decided to use a pandemic-related dataset because we believe that it is a fertile topic for polarization and that, since this topic emerged suddenly, we may be able to observe the polarization dynamics at their very beginning.

The Coronavirus Corpus was first released in May 2020 and has been regularly updated, reaching over 1000 million words in May 2021. This corpus is actually a subset of a larger corpus of news called the NOW Corpus (Davies 2016-), which has not been included in this study. The two corpora share the same data collection criteria. The NOW Corpus (Davies 2016-) grows every day by about 10,000 articles, selecting news gathered from hourly Bing News searches and from the daily scraping of more than 1,000 websites[32]. All the articles containing at least two occurrences of the word "coronavirus", "COVID" or "COVID-19" are added to the Coronavirus Corpus, along with articles having at least an occurrence of selected COVID-related vocabulary[33]. The corpus, designed to record the social, cultural, and economic impact of the coronavirus, includes online newspapers and magazines in 20 different English-speaking countries. It is extremely varied in sources, occasionally also including comments to the published articles, but

[31] https://www.english-corpora.org/corona/, https://github.com/echen102/COVID-19-TweetIDs
[32] https://www.corpusdata.org/now-sources.asp
[33] https://www.english-corpora.org/corona/help/texts.asp

clearly representative of the discourse of online news. The subset we used comprised the first seven months of the debate, for reasons of comparability.

On the other hand, the social media corpus is a repository of an ongoing collection of tweets IDs associated with the COVID-19 outbreak. The collection started on January 28th 2020[34]. The dataset is built searching a set of COVID-keywords via the Twitter's search API to gather historical Tweets from the preceding 7 days; therefore, the first tweets in the dataset date back to January 21st 2020. A detailed overview of the data collection has been published by Chen et al. (2020). As recommended by the authors, we used the software Hydrator to collect our sample of data, as Twitter's Terms and Conditions only allow the sharing of tweets' IDs[35].

We decided to compare the two corpora because of the influence of the mediatic ecosystem on public debate; previous research on a small news corpus showed that the mediatic discourse might be quite influential in shaping public opinion (Cosenza, Sanna 2021) and therefore it cannot be overlooked while studying echo chambers. Finally, the journalistic use of language is extremely different compared to social media and that would help exposing the characteristics of social media discourse.

To start our investigation, we decided to investigate the first seven months of the pandemic, focusing in particular on its first outbreak. Hence, in the Coronavirus corpus we took into consideration all the articles written between January 2020 and July 2020 (henceforth "news corpus"), while in the Twitter corpus we extracted a sample of one million tweets for each month in the same range of time.

Table 6 sums up corpus figures.

[34] A Tweet ID is a unique numeric identifier that is associated with a tweet.
[35] https://github.com/DocNow/hydrator

	N. of texts / tweets	N. of tokens	N. of types
News Corpus	650,699	442,252,000	2,086,489
Twitter Corpus	7,000,000	152,468,080	2,638,855

Table 6: The corpora used

As already stated, the methodology combined corpus-based discourse analysis and word embeddings. As a first step, we computed keywords of our social media corpus compared to our newspaper corpus using Wordsmith Tools 8 (Scott 2020). This was meant to highlight word forms that are used significantly more in social media discourse than in news discourse. We then explored the representation of the coronavirus pandemic in both corpora, looking for possible markers of ideological conflict. The main focus was on the word *hoax*, certainly identifiable as a "loaded" word which could index a strong ideological position and trigger ideological conflict.

4.2 CORPUS ANALYSIS: PRELIMINARY OVERVIEW

The exploration of keywords as elaborated by Wordsmith (Scott 2020) aimed at finding out if the social media corpus was heavily characterized by words that could be identified as markers of ideological conflict (and potentially of echo chambers). Of course, no systematic study of all the keywords could be attempted for corpora of this size in this context. An analysis of the top 100 keywords (ordered by keyness as measured by BIC score) can still provide an idea of what characterizes information on Twitter when compared to news discourse.

The data seem to suggest that the discourse of Twitter is quite understandably more focused on the virus as such, as well as on selected features of the key social actors involved and the impact of the pandemic on everyday life. The data related to these semantic areas are listed below, in Table 2, in decreasing order of keyness, with frequencies expressed per ten thousand words (pttw) contrasting Twitter and news[36].

Semantic areas	**Word forms** **(pttw frequency in the Twitter corpus vs the News corpus)**
the virus and the disease	*coronavirus* (125 vs 28), *COVID19* (22 vs <1), *corona* (19 vs <1), *COVID* (22 vs 2), *flu* (6 vs <1), *coronavirus outbreak* (2 vs >1), *epidemic* (5 vs 1), *Wuhancoronavirus* (1 vs <1)
the key social actors	*China* (47 vs 7), *Wuhan* (22 vs 2), *Realdonaldtrump* (12 vs <1), *Trump* (21 vs. 6), *Chinese* (12 vs 2), *govt* (4 vs <1), *CDC* (6 vs 1), *doctor* (5 vs 1), *china's* (2 vs. 1), *americans* (6 vs 1), *trumps* (2 vs <1), *joe biden* (2 vs <1), *CNN* (3 vs <1, *democrats* (4 vs <1), *NYTimes* (1 vs <1), *Pence* (3 vs <1), *doctors* (5 vs 2), *Fauci* (3 vs 1)
the everyday impact of the pandemic	*breaking* (9 vs <1), *lockdown* (21 vs 7), *mask* (11 vs 2), *wear* (8 vs 2), *stay* (11 vs 4), *save* (4 vs <1), *cure* (3 vs <1), *stayhome* (1 vs <1), *stop* (6 vs 2), *protesting* (2 vs <1), *die* (3 vs <1)

Table 7: Keywords and semantic areas

[36] It should be noted that we kept retweets in our dataset. The reason why we decided to keep them is because they are a typical Twitter affordance, they are used to communicate and to take a stance within the online discourse. Each retweet should therefore be considered as independent and counted in our occurrence.

This preliminary overview suggests Twitter has a marked orientation towards focusing on social actors. Using Van Leeuwen's (2008) categorization, we can say that social actors are represented mostly in specific terms, as identifiable individuals (*Trump(s), Pence, Fauci*) or institutions and media (*government, CDC, CNN, NYTimes*) in the American context, against a background of indeterminate identities categorized in terms of their nationality, position or profession (*China/Wuhan/Chinese, americans, democrats, doctors*).

The other elements in the top 100 keywords were mainly related to grammar words (e.g., *don't*, *can't*), specific abbreviations or metadiscursive organizational elements (e.g.*T/RT* for *Tweet* and *Retweet,* or *thread*), other popular tweeters or figures (*Spectatorindex, Narendramodi*) and pragmatic markers of attitude (*please, fucking*).

A word that inevitably attracted our attention was *hoax* (number 27 in ranking, with 54,028 occurrences of the token in the Twitter corpus, 3.5 pttw vs 2,492 occurrences in the news, <1 pttw). Besides the significant difference in the overall frequency, there is also an interesting differentiation in the diachronic distribution of the word. This points at a wider peak of interest in the press (distributed around February, March and April, as well as later, in July), with a more intense peak in March for Twitter.

Month	**Words**	**Hits**	**Pttw**	**Dispersion**
January	8,278,513	29	<1	0.621
February	16,337,840	139	<1	0.397

March	96,258,096	736	<1	0.852
April	119,999,000	609	<1	0.918
May	96,748,312	275	<1	0.894
June	93,851,352	225	<1	0.780
July	83,051,184	479	<1	0.919
Overall	514,524,288	2,492	<1	0.929

Table 8: Diachronic distribution of the word "hoax" in the News Corpus

Month	**Words**	**Hits**	**Pttw**	**Dispersion**
January	20,690,930	2,940	1.4	0.407
February	20,464,160	1,158	0.6	0.483
March	21,860,520	35,073	16	0.869
April	24,210,576	2,660	1.1	0.929
May	21,654,360	3,026	1.4	0.793
June	22,128,794	1,928	0.9	0.936
July	22,163,356	7,243	3.3	0.735
Overall	153,172,704	54,028	3.5	0.951

Table 9: Diachronic distribution of the word "hoax" in the Twitter Corpus

Concordance analysis can help distinguish patterns of use of the word in the two corpora, representative of online news and Twitter discourse. Differences of use may be more illuminating than just quantitative differences.

4.3 ANALYZING *HOAX* IN THE NEWS

We analyzed the word at the level of collocation, semantic preference and semantic prosody in the two corpora. When looking at collocation with full lexical items, the top collocates can be easily classified into semantic groups. In the News Corpus the top 100

collocates of *hoax*, ordered using t-score (Oakes 1998, Hunston 2002 show a dominance of the following semantic areas: potential referents of the term, potential sources and victims of the hoax, elements of negative evaluation and expressions of attribution/projection. The data including collocates related to these semantic areas are provided in Table 5. Words are ordered by raw frequencies as for our goals, namely semantic preference analysis, collocation strength was not crucial. Table 5 includes also some terms, marked with an asterisk (*) whose collocational strength was lower, hence not included in the top 100. However, semantic preference is not a simple relation of co-occurrence with the node word, instead it is a relation between the word and a semantic field (Stubbs 2001). For this reason, it is crucial to provide a full picture of the semantic areas involved in these relations.

Semantic areas	**Word forms (and absolute frequency)**
Potential referents	The virus: *coronavirus* (285), *virus* (245), *pandemic* (138) and *covid-19* (124), *flu* (15)*, *corona* (14)*, *disease* (12)* Other referents: *Russia* (105), *impeachment* (91), *climate* (38), *change* (38), *Russian* (12)*, *Chinese* (10)*
Social actors involved	*Trump* (128). *democrats* (102), *democratic* (84), *they* (78), *media* (48), *news* (42), *people* (41), *president* (41), *government* (26), *democrat* (18), *liberal* (13)*, *party* (12)*, *Mueller* (11)*
Nominal and verbal elements of projection (reported	*said* (117), *called* (115), *think* (63), *word* (55), *saying* (50), *threat* (49), *call* (46), *calls* (45), *thought* (40), *believe* (36), *claiming* (27), *investigation* (19), *claimed* (18), *criticism* (18), *statements* (18), *referring* (18), *says* (15)*, *claims* (15)*, *told* (15)*, *claim* (14)*,

speech or thought)	*concerns* (12)*, *theories* (12)*, *believed* (12)*, *know* (12)*, *says* (10)*, *statement* (10)*
Qualifications	*Just* (82), *another* (60), *political* (56), *greatest* (17), *viral* (16), *cruel* (13)*
Negative collocates	Negative connotations: *threat* (49), *fake* (44). *cooked* (*up*) (31), *conspiracy* (22), *scam* (15), *fabricated* (14), *victim* (29), *perpetrated* (20) Negatives: *nothing* (27), *never* (21)

Table 10: Hoax in the News: collocation and semantic preference.

The most frequent lexical elements unsurprisingly refer to the virus as hoax, but there are also other noticeable potential referents of the term. The social actors involved (as victims or perpetrators of a hoax) are potentially identifiable as the American president, its political opponents and the media. Many elements of premodification recall the recurrence of the word in the news and its key role in describing a number of different issues in the ongoing political debate from the impeachment to the coronavirus issue. Its frequent collocation with negative forms (*nothing, never*) and with expressions of inscribed negative evaluation seems to confirm the negative evaluative meaning of the word, inevitably evoking the whole ideological debate.

What is most noticeable, however, is that there is a substantial set of collocates that refer to processes of verbal or mental projection, i.e., reporting speech or thought. This is of course in line with the main reporting function of the news, but the frequency and range of terms suggests looking further into the question. When looking at the pragmatics of the word (and its *semantic prosody*, both in terms of illocutionary force and evaluative

meanings), what becomes most prominent is the need to distinguish its use in the speaker's discourse from forms of reported speech or thought.

Identifying the attributed nature of the claim often requires looking up a wide context; occasionally this means looking at the whole paragraph, as in Example 4 below, where the reporting element is provided by the nominal *claims* in the final sentence of the paragraph:

> (4) Bill Gates created the coronavirus. China secretly developed it in a lab as a biological weapon. A cure exists and the government controls it but won't release it to the public. The virus is no more dangerous than the seasonal flu. Coronavirus is a "fake news" *hoax* manufactured by the news. You can use hand dryers to kill the virus, vitamin C, or lemon juice. The country is going to be quarantined under martial law, and the government will shut down all grocery stores so that no one can buy food. All of these *claims* are examples of conspiracies associated with coronavirus that have been perpetrated by social media. (Counterpunch, March 27)

An analysis of 200 random concordances of *hoax* provides an approximate quantification of how often the word is directly used by the speaker (thus potentially accepting different degrees of responsibility for its use, whether in an accusation or a denial of the accusation) and how often it is attributed to other sources. The expectation, of course, is that the majority of the occurrences will be attributed to someone else in the news, rather than used in claims of falsity inevitably implying mistrust in the source reported; the actual proportion, however, is quite striking in the analysis of the 200 concordances: 87% of the occurrences (174/200) are in reported discourse, whereas only 13% (25/200) are actually used by the voice of the speaker.

The quantitative data confirm that the word is a key element of the reported debate, where the term is overwhelmingly used in affirmative statements (except for 1 reported question and 14 reported denials of the term) and mostly in reporting claims of

falsity. The focus is thus more on the context in which the term was used and on its accusatory function, than on reporting how the term was rebutted in the debate. The emphasis lies very much in reconstructing the context of the debate and interpreting the word in terms of the different components of its meaning. This involves both its semantics - i.e., the (non-)factual nature of what it refers to, and the presence or absence of reasons to support the claim (Example 5) - and its pragmatics - i.e., its accusatory value and the patterns of agreement/ disagreement that are constructed around the word (Example 6):

> (5) The victim thought the COVID-19 virus was a hoax, despite its killing more than 135,000 people in the United States so far.
>
> (6) Many Mexicans have developed a stigma around the virus and believe that it is a hoax or not as bad as it seems. It has also impacted the country's healthcare workers who are facing widespread abuse from people who believe they are helping spread the virus.

It should further be noticed that 11 of the 25 occurrences actually used by the voice of the speaker in the news corpus are in readers' comments and not in the journalist's voice and one is a case where the word is mentioned, rather than used. In the other occurrences there is a tendency to use *hoax* as a premodifier - *hoax plate/ call* (2)/ *messages* (3)/ *conspiracy-mongering*/ *stories*/ *texts*/ *text* (2) /*email* - or as a labelling noun referring to previous reports - *the hoax* (2)/ *hoax*. Here the speaker takes only partial responsibility for the accusation: the word is not used to make a claim of falsity but rather to refer anaphorically to such a claim or to the word. This leaves virtually no space to direct claims about the false nature of a position, let alone accusatory language.

4.4 *HOAX* ON TWITTER

The Twitter corpus provides a partially different picture of the word. The most obvious feature is that texts are more repetitive, due to retweets, and there is also a reduced lexical variety, probably also influenced by Twitter affordances such as characters' limit.

Table 6 below illustrates data with reference to the same semantic areas we identified in the news corpus.

Semantic areas	**Word forms (and absolute frequency)**
Potential referents	The virus: *coronavirus* (24,535), *virus* (2,037), *covid-19* (1,615), *corona* (1,471), *covid* (1,438) Hashtags: *#covid19* (983), *#coronavirusupdate* (531)
Social actors involved	*Trump* (4,930), *democrats* (3,985), *@realdonaldtrump* (3,384), *Carolina* (2,531), *administration* (2,369), *politicizing* (1552), *dems* (1,092), *media* (783), *people* (597) Hashtags: #*liberalllogic* (507)
Nominal and verbal elements of projection	*called* (4,807), *calls* (1,694), *call* (1226), *word* (1,193), *says* (1,152), *used* (893), *using* (711), *refer* (562), *think* (507), *believe* (424)
Qualifications	*just* (2,387), *another* (2,193) and *biggest* (754)
Negative collocates	Negative connotations: *failed* (2,630), *trojan horse* (454), *fake* (328)*, *dishonestly* (156)*, *blame* (144)*, *accuse* (140)*, *threat* (140)*. Role of misinformation: *pushing* (2,553), *spreading* (1,184), *hype* (503), *promoted* (131)*

Table 11: Hoax on Twitter: collocation and semantic preference.

The most common collocate is referred directly to the virus, with the word *coronavirus* being by far the most frequent collocation in the corpus, six times more common than the second collocate. The data also confirm that the word *hoax* is primarily associated - in

terms of strength of the collocation - with the social actors involved, mainly politicians, and is often qualified by a limited set of elements of premodification pointing at the recurrence of the word in the political debate. An interesting feature is the frequent occurrence of *politicizing* (1552), pointing at the action that is attributed to most of the social actors, who are "politicizing" the question of the virus. Collocates referring to reported speech or states of beliefs, on the other hand, are substantially less, when compared to the News Corpus; the difference in the distribution is also accompanied by less lexical diversity and by the fact that most of these verbs refer specifically to President Donald Trump.

It is worth highlighting the presence of a small set of hashtags that proved to be among the most frequent collocates: *#covid19* (983), *#coronavirusupdate* (531), #*liberalllogic* (507). These hashtags are attributable to the areas reported above but, as hashtags, they highlight the key role of the virus and of politics itself in defining the trending topics and in activating the contextual assumptions which may guide users' derivation of explicitly and implicitly communicated meaning (Scott 2015).

Finally, while lexical elements of negative evaluation stand out as both frequent and varied, a distinctive feature of the Twitter Corpus is the presence of collocations that refer to the propagation of misinformation: *pushing* (2,553), *spreading* (1,184), *hype* (503), *promoted* (131), thus showing the relevance of the issue on Twitter.

When considering the pragmatic status of *hoax*, the hypothesis is that the use of *hoax* would be more likely to be directly used in Twitter, and most probably in the form of claims of falsity, whether to deny a fact/claim or accuse the source of the hoax. The analysis of a random sample of 200 concordances actually shows that the picture is more varied in use (though not in forms perhaps). The formal repetitiveness that can be noticed is probably due to a very high percentage of the word in retweets (170/200) but

also as hashtag (7/200 in this particular set of concordances). The pragmatic status of *hoax,* on the other hand, is varied, with little less than half of the occurrences (97/200, i.e., 48.5%) used directly by the speaker rather than attributed. There is also a fair balance between affirmative claims (claims of falsity, 58/97) and negative claims (denials, 39/97) in the discourse of the speaker.

In terms of the pragmatic functions of *hoax*, speakers use the word mostly in nominal constructions, but often specifying its source in forms that bring out the full accusatory potential of the lexical item (Examples 7 and 8). Rebuttals often take the form of a simple denial, with no supporting argument (Example 9), leaving an interesting scope for ironical forms (Example 10).

(7) The Democrats new \"blame Trump\" hoax investigation about coronavirus

(8) Democrats are pushing a new hoax by politicizing the coronavirus situation

(9) The coronavirus is not a hoax

(10) the Coronavirus, COVID19, is a hoax in the same way that the earth is flat.

The controversy about the Coronavirus being or not being a hoax proves to be quite prominent in the quantitative data of the corpus, as shown in particular by the top clusters around the word (with the very high frequency of *is a hoax/is not a hoax*). Concordance analysis also shows that a large part of the debate is focused on how the word *hoax* is used in the political debate, as shown in Example 11:

(11) Trump says he used the word \"hoax\" NOT about the coronavirus itself but about what Democrats were saying about his administration

This “metalinguistic” focus of the debate centers more on the pragmatics of the word, its summarizing different positions and their conflict. When the word is used directly by the

speaker, a focus on its identity and accusatory value also seems to be dominant (Example 13), even if there can be attention to its truth value and to producing forms of support for the claim (Example 14):

> (13) Urging all patriots to stay focused on: General Michael Flynn Case Spygate ObamaGate Clinton Foundation COVID Hoax The Fake News and Cabal want our attention elsewhere. Together we are powerful and they know it.
>
> (14) IT IS A HOAX PATRIOTS! More people in the US have DIED from common influenza

The overall impression of concordance analysis is that, on Twitter, the word *hoax* is not only discussed as an important keyword of the political debate (as happens in the news), but also appropriated by many participants. The political connections of the word become more prominent and so does its accusatory pragmatic function, which becomes a clear index of Trumpism, as well as opposition to Democrats (or China, the media etc.). The word thus expresses not only the speakers' position on the specific issue, but also more general traits of their identity. The co-text clearly illustrates marked patterns of agreement and disagreement, while the space for providing arguments and discussing the general political context is limited.

4.5 INSIDE THE INFERENTIAL PATH: THE EMBEDDINGS

We finally created two word-embedding models to further explore the semantics and pragmatics of *hoax*[37]. In the news corpus we selected all the articles where the word

[37] We used the word2vec Python implementation of Gensim. https://radimrehurek.com/gensim/models/word2vec.html

occurred, while in the Twitter Corpus we used all the tweets to compute our model[38]. Figure 23 illustrates the semantic space of the word in the two corpora, with the top-15 most similar words. The thickness of the bands represents the number of occurrences of each term. On Twitter, unsurprisingly, there is a prevalence of usernames, indicating a direct addressing to the counterpart.

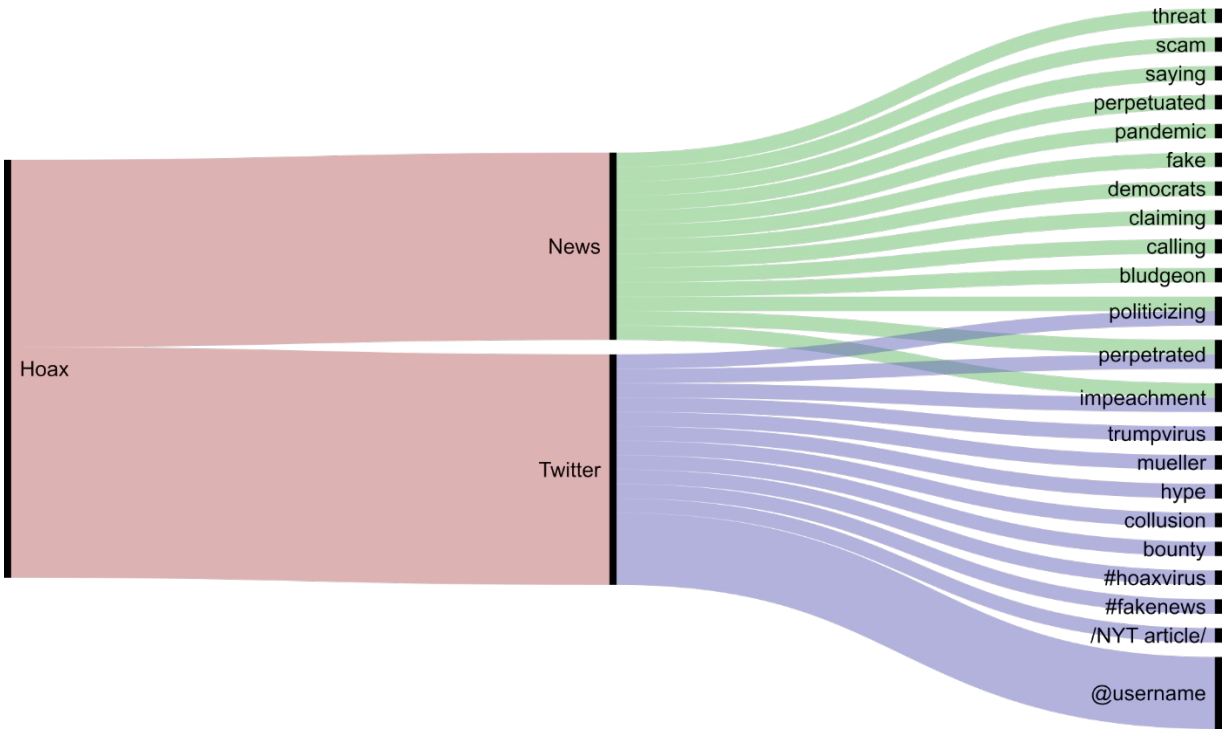


Figure 23: Hoax semantic space within the two corpora, top-15 most similar words

At first glance, our two corpora show a rather similar semantic space. In particular, both corpora see the word in semantic relationship with the words *politicizing*, *impeachment*

[38] We made this choice because of the significant differences in text sizes. In fact, although the number of types is quite comparable, word embeddings are affected by texts' length in the machine learning process. The average tweets' length in the Twitter Corpus was around 20 words, while in the News Corpus the average number of words per text was around 670. On the one hand we had a problem of computational efficiency (News Corpus was too large) while on the other hand we had a risk of computing poor embeddings (because of the characteristics of Twitter Corpus). We then came to this compromise to have a comparable result. We also tested the stability of our News Corpus model computing 2 different models, one using only the sample of texts with *hoax* occurring and another adding random samples of articles without *hoax*, finding no significant differences. The sample selected for the embeddings of the News Corpus has 1872 articles.

and *perpetrated*. On the other hand, the data show differences that are predictable both because of the composition of our corpus and because of what we have seen in our concordance analysis. Clearly, in the news corpus *hoax* occurs mostly in reported speech, as shown by some of the verbs associated with it (*calling*, *claiming*, *saying*), while in the Twitter corpus it is mostly used to address or retweet other users, as indicated by the prevalence of usernames. As shown in Figure 1, it is also associated with two debated topics. The word *Mueller* refers in particular to the *Russiagate*, where Trump was accused to have favored Russian interference during the 2016 presidential campaign. Finally, the link with a *New York Times* article[39] brings us in the debate of COVID-19 negationists. In fact, the word is associated with a specific set of retweets that share the NYT article to attack the counterpart, who believe the pandemic is a hoax.

Focusing on the words that are common to both corpora (*impeachment*, *politicizing*, *perpetrating*), it is possible to notice that in our News Corpus these terms are present in forms of reported speech (with reference to Trump and other media or political commentators). For instance, the word *impeachment* is associated to *hoax* because it is often reported that Trump called it a hoax, the word *perpetrated* is associated to Trump's political attacks to democrats and media ("hoax perpetrated by dems/media"); finally, the word *politicizing* is the first link to COVID-19 pandemic, as it is reported that Trump was accusing the democratic party of politicizing the virus.

On the other hand, in our Twitter Corpus, these terms are mostly used directly by users to support Trump, especially attacking his political opponents. The similarity with the word *impeachment* is due to a set of 130 retweets of the following sentence: "This is the man that diverted attention from the first appearance of COVID-19, while pushing the

[39] https://www.nytimes.com/2020/03/01/world/coronavirus-news.html

impeachment hoax", originally shared by CIA officer Kevin Shipptogether with a picture of Democratic politician Adam Schiff [40]. Similarly, the two verbs *perpetrated* and *politicizing* are used to accuse the Democratic party and its members of spreading misinformation, where *perpetrated* is used especially with reference to the impeachment case, while *politicizing* is used to accuse Dems of exploiting the coronavirus situation for their own political interest. Particularly interesting in these tweets is the use of the adjective *new* referred to *hoax*, as it recalls the frequency with which the word *hoax* occurred in the political debate of the period, often in connection with Trump's interpretation of Democratic positions.

To follow our inferential path, combining the different levels – invariable and optional – required to interpret a semantic frame, we then explored the embeddings of the word *collusion*, that is the most similar word in the semantic space of *hoax*. This represents the strongest semantic association; hence it is the most probable move within the inferential path[41]. The results show that the term is another marker of ideological conflict, as it evokes the opposition between *russiagate* and *obamagate*.

[40] From his Twitter profile: "Kevin Shipp: Former CIA Counter Terrorism, Counterintelligence and Staff Investigator. Author, From the Company of Shadows - CIA operations/use of secrecy". No other reliable sources have been found. The author of the tweet also owns a personal blog which seems devoted to conspiracy theories linked to alleged previous activity as CIA officer (https://kevinshipp.com/)

[41] Word similarity is evaluated calculating distances between vectors, hence calculating cosine similarity. Cosine similarity has a range of values between -1 and +1, with -1 indicating the maximum distance and +1 the maximum proximity. In our case, the word *collusion* has a cosine similarity of 0.43 with *hoax*.

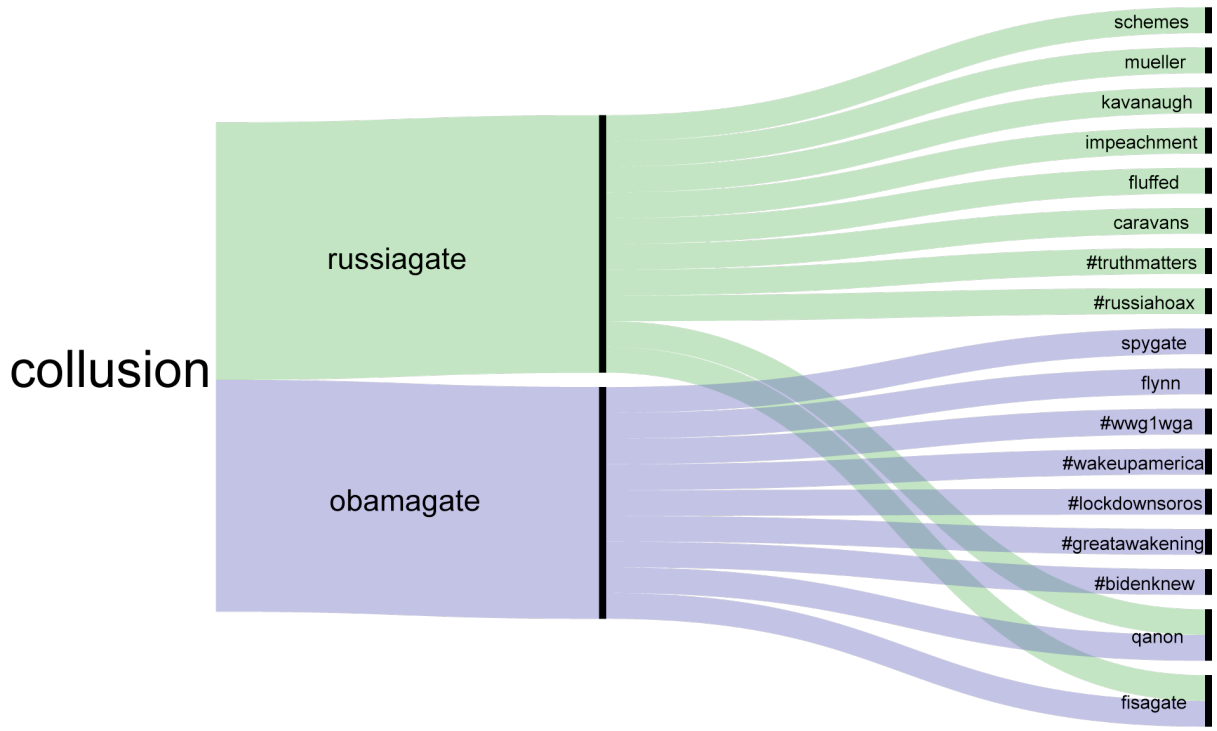


Fig. 24: Part of the inferential path of the word *collusion* in the Twitter Corpus

Going deeper into our inferential path of these two words, as shown in Figure 24, we found examples of the increasing ideological isolation that is typical of echo chambers. As shown by our embeddings, the two words are ideological keywords coming from the alt-right area and, more specifically, from QAnon members.

We might see this stage of the pragmatic process as a deep dive into the echo chamber, due to the peculiar linguistic status of hashtags (Zappavigna 2011, 2015). Due to their flexibility as a semiotic resource, Twitter hashtags are used both as a form of social tagging, ascribing the text to a particular trending topic and to indicate stance taking. In this sense, the set of Qanon hashtags shown in Figure 2 are the very deep core of the Trumpist echo chamber, as they both mark the belonging to a conspiracy theory and their irreducible conflict with the counterpart.

4.6 ECHO CHAMBERS AND APPRAISAL: LOOKING FOR DIALOGIC CONTRACTION

At this point we decided to further investigate the linguistic dimension of echo chambers using the Appraisal framework. Our intention was to explore the proportion between dialogic contraction and dialogic expansion, first, by quantifying the two types and, second, by looking for evidence of substantial qualitative differences that may emerge from the data. The goal of such exploration was moved by the idea that we could exploit dialogic markers to explore argumentation, using appraisal indicators as pivotal points to explore the argumentative function.

We used the annotation made by Fuoli (2018) to have a list of tokens that could be considered markers of appraisal. Our decision was indeed to use a given list of markers, instead of preparing a tagging for our corpus, thus combining corpus-based and corpus-driven approaches. We believe that our case study might be the right context to do it, since we are interested in highlighting linguistic patterns that may be markers of echo chambers also in the dialogic dimension. As we saw in the previous sections of this chapter, echo chambers are ideological structures that are used to express ideological positioning, thus the most appropriate type of appraisal to investigate is "Engagement", which would highlight stance-taking within the dialogic dimension.

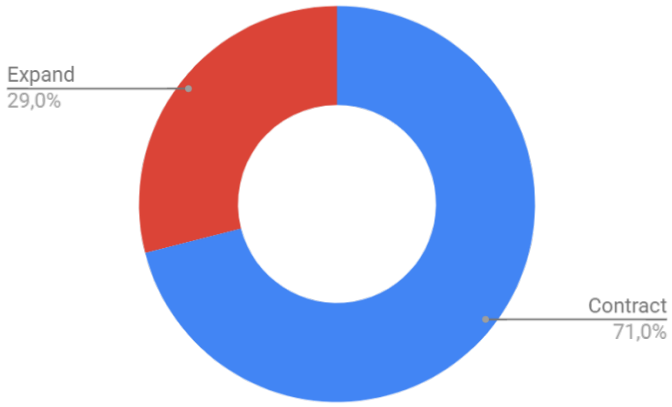
To minimize possible biases due to the nature of the original tagging process, we kept in our table only words that were the less corpus-specific possible, also avoiding multi-words expression. This allowed us to have a list of common appraisal markers to compare and quantify the dialogic dimension within our two corpora used for the echo chambers' analysis. Table 12 shows the full word list of "Engagement" tokens and the related appraisal subtype.

token	type	subtype_I	subtype_II
but	ENGAGEMENT	Contract	Disclaim
expect	ENGAGEMENT	Expand	Entertain
without	ENGAGEMENT	Contract	Disclaim
Despite	ENGAGEMENT	Contract	Disclaim
obviously	ENGAGEMENT	Contract	Proclaim
understands	ENGAGEMENT	Contract	Proclaim
While	ENGAGEMENT	Contract	Disclaim
believe	ENGAGEMENT	Contract	Proclaim
No	ENGAGEMENT	Contract	Disclaim
Indeed	ENGAGEMENT	Contract	Proclaim
expected	ENGAGEMENT	Expand	Entertain
not	ENGAGEMENT	Contract	Disclaim
confident	ENGAGEMENT	Contract	Proclaim
cannot	ENGAGEMENT	Contract	Disclaim
Naturally	ENGAGEMENT	Contract	Proclaim
recognize	ENGAGEMENT	Contract	Proclaim
demonstrated	ENGAGEMENT	Contract	Proclaim
know	ENGAGEMENT	Contract	Proclaim
unthinkable	ENGAGEMENT	Contract	Disclaim
knew	ENGAGEMENT	Contract	Proclaim
found	ENGAGEMENT	Contract	Proclaim
stated	ENGAGEMENT	Expand	Attribute
convinced	ENGAGEMENT	Contract	Proclaim
never	ENGAGEMENT	Contract	Disclaim
Clearly	ENGAGEMENT	Contract	Proclaim
nothing	ENGAGEMENT	Contract	Disclaim
Yet	ENGAGEMENT	Contract	Disclaim
might	ENGAGEMENT	Expand	Entertain
shows	ENGAGEMENT	Contract	Proclaim
may	ENGAGEMENT	Expand	Entertain
evident	ENGAGEMENT	Contract	Proclaim
believe	ENGAGEMENT	Expand	Entertain
belief	ENGAGEMENT	Contract	Proclaim
inevitable	ENGAGEMENT	Contract	Disclaim

should	ENGAGEMENT	Expand	Entertain
reflecting	ENGAGEMENT	Contract	Proclaim
reflected	ENGAGEMENT	Contract	Proclaim
reflects	ENGAGEMENT	Contract	Proclaim
However	ENGAGEMENT	Contract	Disclaim
although	ENGAGEMENT	Contract	Disclaim
anticipate	ENGAGEMENT	Expand	Entertain
convinced	ENGAGEMENT	Expand	Entertain
could	ENGAGEMENT	Expand	Entertain
Indeed	ENGAGEMENT	Contract	Disclaim
see	ENGAGEMENT	Contract	Proclaim
evidence	ENGAGEMENT	Contract	Proclaim
knowing	ENGAGEMENT	Contract	Proclaim
reflect	ENGAGEMENT	Contract	Proclaim
demonstrate	ENGAGEMENT	Contract	Proclaim
none	ENGAGEMENT	Contract	Disclaim
prove	ENGAGEMENT	Contract	Proclaim
see	ENGAGEMENT	Expand	Entertain
project	ENGAGEMENT	Expand	Entertain
recognising	ENGAGEMENT	Contract	Proclaim
did	ENGAGEMENT	Contract	Disclaim
think	ENGAGEMENT	Expand	Entertain
assure	ENGAGEMENT	Contract	Proclaim
surely	ENGAGEMENT	Contract	Proclaim

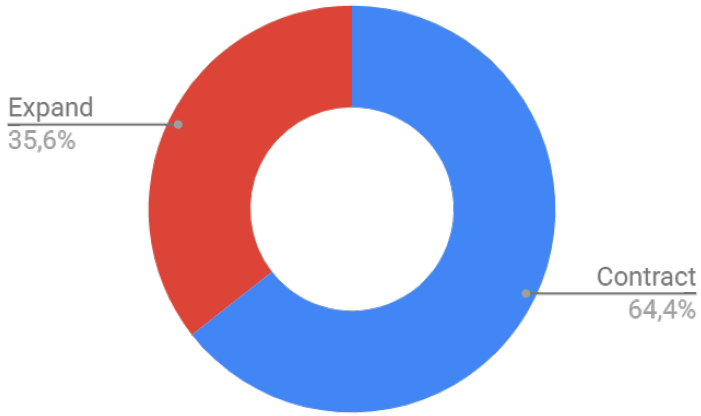
Table 12: List of all the engagement markers.

We then looked for the distribution of appraisal markers in both our corpora. In particular, we were interested in investigating the distribution of dialogic expansion and dialogic contraction within our two corpora. Figure 25 and 26 shows the distribution of dialogic expansion and dialogic contraction in the Twitter and in the News corpus, along with the normalized pttw occurrences.



Type	**Markers pttw**
Contraction	43,02
Expansion	17,55

Fig. 25: Appraisal markers distribution on Twitter



Type	Markers pttw
Contraction	51,35
Expansion	28,39

Fig. 26: Appraisal markers distribution in the News

An important limitation we should consider while quantifying the dialogic contraction and expansion is that within our tags 70% of the tokens comes from the contraction domain. It is beyond the scope of this work to investigate whether this is caused by the original CSR corpus used by Fuoli or if it is what we should expect to happen in the English language. As shown in figure 25 and 26 the distribution of markers seems to follow the distribution of our annotations, making further data cleaning operations necessary.

Hence, we started comparing our normalized pttw occurrences[42] of the markers used in both corpora, as shown in Table 13.

token	Dialogic Function	PTTW Twitter	PTTW News
although	Contract	0,08264024837	**0,7617376518**
anticipate	Expand	0,004525537411	**0,1629840001**
assure	Contract	0,02590706199	**0,1132838291**
belief	Contract	0,04381244914	**0,1455731122**
believe	Contract/Expand	2,179866107	**2,816538987**
but	Contract	**8,554774219**	8,233699339
Clearly	Contract	0,04879709904	**0,07981874587**
confident	Contract	0,05273234896	**0,367166231**
convinced	Contract/Expand	0,1170080977	**0,2264772121**
could	Expand	2,516920263	**5,268647739**
demonstrate	Contract	0,04007396171	**0,1540072176**

[42] We also tried to compare our occurrences with a third corpus, used as external reference, namely the TenTen Corpus. For reference see *enTenTen20* on Sketch Engine https://www.sketchengine.eu/ententen-english-corpus/ However, this comparison was unfruitful. We should expect, given that the dialogic dimension is quite prevalent in our corpora, a higher pttw within our corpus compared to the TenTen. Instead, we observed that most of words have often a significant lower frequency according log-likelihood, calculated as implemented by Rayson and Garside (2000).
At the present time there is indeed a variety of corpus approaches to Twitter (Pak and Paroubek 2010, McMinn et al. 2013, Refaee and Rieser 2014, Derczynski et al. 2016, Verhoeven et al. 2016, Sanguinetti et al. 2018, Rüdiger and Dayter 2020) but all of them are meant for specific case studies and cannot used a reference corpus, as they are not an accurate snapshot of English use on Twitter. Moreover, as we already explained in section 3.3, Twitter does not allow the sharing of tweets' full text (McCreadie et al. 2012) so that also one past attempt of creating an all-encompassing corpus (Petrović et al. 2010) is not available anymore. We should also consider that the TenTen corpus itself might have important limitations as it is created collecting texts from the web, intending the web as any website coming under an English-speaking domain (e.g. .uk, .us).

demonstrated	Contract	0,025382362	**0,1641598003**
Despite	Contract	0,2092897084	**0,6580863399**
did	Contract	2,015241485	**3,110534266**
evidence	Contract	0,4804940155	**0,7500248727**
evident	Contract	0,02334914954	**0,117738303**
expect	Expand	0,3129835438	**0,9886444832**
expected	Expand	0,2940287567	**1,980477194**
found	Contract	0,7778677347	**1,871602616**
However	Contract	0,2162419832	**2,246841167**
Indeed	Contract	0,1082193729	**0,325877554**
inevitable	Contract	0,06781747366	**0,179354757**
knew	Contract	0,4180547168	**0,4377142444**
know	Contract	**4,17785808**	3,050319727
knowing	Contract	0,1598367344	**0,2692356394**
may	Expand	1,905185662	**4,689362626**
might	Expand	0,9831566056	**2,086864503**
Naturally	Contract	0,005443762393	**0,02537015095**
never	Contract	1,493296171	**1,560467788**
No	Contract	**2,241911881**	1,570462089
none	Contract	0,153999447	**0,3494387815**
not	Contract	**12,69360774**	10,08146939
nothing	Contract	0,9602665686	**0,7969438239**
obviously	Contract	0,1405540097	**0,3250183153**
project	Expand	0,2046985835	**0,6301158615**
prove	Contract	0,1239603726	**0,3010726916**
recognising	Contract	0,005706112388	**0,04018071145**
recognize	Contract	0,06230812377	**0,2125032787**
reflect	Contract	0,06427574873	**0,4896303465**
reflected	Contract	0,0100348873	**0,1543463907**
reflecting	Contract	0,01679039967	**0,1124245905**
reflects	Contract	0,02853056194	**0,1706267015**
see	Contract/Expand	6,084552255	**7,162477502**
should	Expand	3,930921148	**4,396475313**
shows	Contract	0,7193636858	**1,131436376**
stated	Expand	0,09372453565	**0,6816249559**

surely	Contract	0,1067764479	**0,1333402675**
think	Expand	**3,121440239**	2,403154762
understands	Contract	0,02931761192	**0,1299937592**
unthinkable	Contract	0,004263187416	**0,02817398225**
While	Contract	0,560379589	**2,393070014**
without	Contract	1,154208802	**2,847177627**
Yet	Contract	0,225948933	**0,3609932798**

Table 13: Normalized occurrence of appraisal markers on Twitter and in the News. Bold indicates higher frequency

In table 13 we observe, not surprisingly, that the engagement level is often higher in the News Corpus, while on the other hand we observe a lack of argumentation markers in the Twitter Corpus. This emerges by comparing the normalized occurrences of the modal verbs ("could, "may") and other elements typical of the negotiation of the dialogistic dimension, namely "however" and "although". Certainly, this could be a structural feature of the Twitter environment, where arguing is more complex also because of the character limit. In fact, we note that on Twitter prevail four words compared to the News Corpus, namely two negations ("No" and "Not") and two verbs ("know" and "think") that are clearly linked to the expression of epistemic modalities.

We decided to take a closer look at these terms on Twitter, as they might be further markers of ideological conflict within the platform. We focus on Twitter because it is on social media that echo chambers happen, therefore when we can observe linguistic expressions aimed at aggregate likeminded people.

In particular, looking at a small excerpt of 200 collocations for each of these terms, we notice that "think" is used to introduce a personal standpoint (i.e. "I think") in 30% of the cases, while in the rest of the cases is either uses to introduce external voices (both in

contraction and expansion) and to formulate rhetorical questions to negate the counterpart voices ("Do you think you know COVID so well?"). Particularly interesting is the fact that the presence of the verb "think" seems to exacerbate ideological confrontation (see examples 14-15).

(14) You know, I don't think it's too complicated to wear a mask

(15) You don't think the new world order would exploit something like covid-19 to subjugate us all into slavery, do you?

This behavior is actually very typical of echo chambers, and it entails the mutual non-acknowledgment of the counterpart, which is in fact systematically denied .

Instead, the verb "know" seems marginal in the dialogistic dimension as in the vast majority of the cases (>80%) is it used to self-express the lack of knowledge ("I did not know that this happened") or just to share information ("Don't know who needs to hear this but…"). However, this particular formulation is used to achieve something that goes beyond the mere informative function, as also indicated by the amount of retweets sharing this formulation. In fact, know is very rarely used with epistemic intent, on the contrary it is often used to express a standpoint

(16) I don't know if I'm keen to trust data in the middle of a pandemic where adequate testing hasn't been accomplished and attributable death totals are questionably

(17) I don't know if I want my hair to come from China. I'm scared

In the sparse cases in which know it is used as a marker of ideological conflict it is used to discredit external voices ("they don't know if the COVID cases will double").

A similar discourse applies to the words “No” and “Not”, which, being two very common negations, are often used without any dialogic function[43]. The case of “No” is rather illustrative, as it is a very versatile determiner and therefore present in many different contexts. In section 4.6.1.1 we address its analysis in detail. On the other hand, "Not" is an adverb used for negation, thus making it easier for us to understand how much it is used as a dialogic contraction in our corpus by selecting the most significant verbs that occur with “Not”.

Because of this feature we can in fact be more precise in our exploration and directly select the most significant verbs that occur with "Not" in our corpus. We selected the 5 most significant ones according to the t-score, i.e. the verb "to be" in the present tense ("this is not"), the verb "to do" in the present tense ("does not"), the verb "should", the verb "to have" in the present tense ("to have") and finally the verb "to be" in the past tense ("were not"). What emerges is that when "not" is used to negate a verb in the third person singular it is largely part of dialogic contraction processes (>90% of cases out of 200 random collocations for both), while in the other cases the tendency is the opposite, in less than one case out of ten we can identify an episode of dialogic contraction.

In the cases of "does not" and "is not" the negation is used in most cases to deny epistemic validity to voices outside the dialogue or the direct counterpart we are addressing (e.g. "5G does not spread coronavirus"), sometimes also with a direct reference to veridical aspects (e.g. "information circulating on social media a confirmed case of COVID 2019 Corona Virus is not true").

[43] It should be acknowledged that in the appraisal theory there is always a role of dialogic contraction for the denial. However, in this work we are only considering the dialogic contraction when relevant as argumentative function towards the theme of the pandemic.

4.6.1.1 An attempt of exploring the dialogic dimension of echo chambers

Since echo chambers should be considered as dichotomic structures we should investigate appraisal tokens that are shared by the two corpora. As we said in section 4.1, we might think as echo chambers as expression of a specular ideological conflict. Hence, as we did with "hoax", we should look for words around which this conflict express. As shown in 13, the News Corpus is characterized by a wider range of word forms, which surely due to the nature of journalistic discourse, while on Twitter we have scarcity of argumentation markers and a prevalence of ideological positioning using negation and verbs expressing cognitive processes such as "think" and "know".

Despite the differences, the two corpora are quite linked, as we are assuming that the mediatic dimension is quite influential on the echo chamber dimension. The idea is that on social media is where echo chambers get operational, while within the news they have their genesis in the shaping of public opinion. The main issue is to find quantitative evidence of words that are relevant to explore for both corpora, as they will be probably the most used markers of ideological conflict, thus the terms around which most of the dialogic dimension of echo chambers is built.

To proceed in the exploration of our quantitative findings we also calculated the Inverse Document Frequency (IDF) score (Jones 1972, Robertson 2004) for each marker that we found in our corpus. We chose IDF to evaluate the saliency of each appraisal markers along with its normalized occurrence, with the goal of highlighting the most informative and yet most occurring words for each of our corpora.[44]

[44] We chose to use pttw and IDF instead of the more classic TF-IDF approach since we are not interested in comparing different documents of our corpus, instead, we are interested in determining which appraisal markers are occurring the most while being the most informative, namely having an uneven distribution within our documents. For this reason we keep the normalized frequency over the whole corpus rather than the relative frequency of each markers in our documents.

Furthermore, assuming a Zipfian distribution of our appraisal markers, we calculated the *elbow point* of the two distributions[45]. The *elbow* of the distribution allowed us to determine the threshold value to consider along with the IDF score of each word, to determine if the most frequent terms were also the most informative for us. In other words, we are looking for terms that have the highest possible IDF in combination with a pttw that is very close to our elbow point. The goal of this combination is to determine which words are the noisiest, namely which words are more likely to occur a lot, for reasons other than their role in the dialogistic dimension.

The results are shown in figures 27-28 and in tables 14-15.

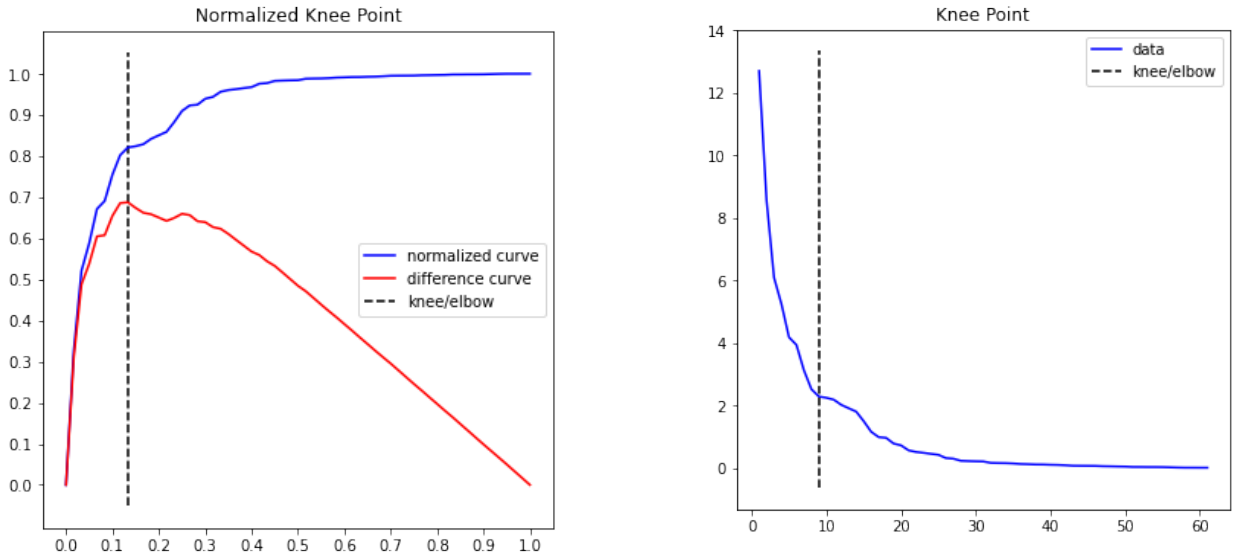


Fig. 27: Elbow point calculation for Twitter normalized frequency. The curve starts to flatten around 2,28 pttw

[45] The elbow method is usually used in clustering to determine the optimal number of clusters to calculate. Instead, we are using it to determine at which frequency the curve of word occurrences starts to flatten, thus determining which of our tokens should be considered outliers.

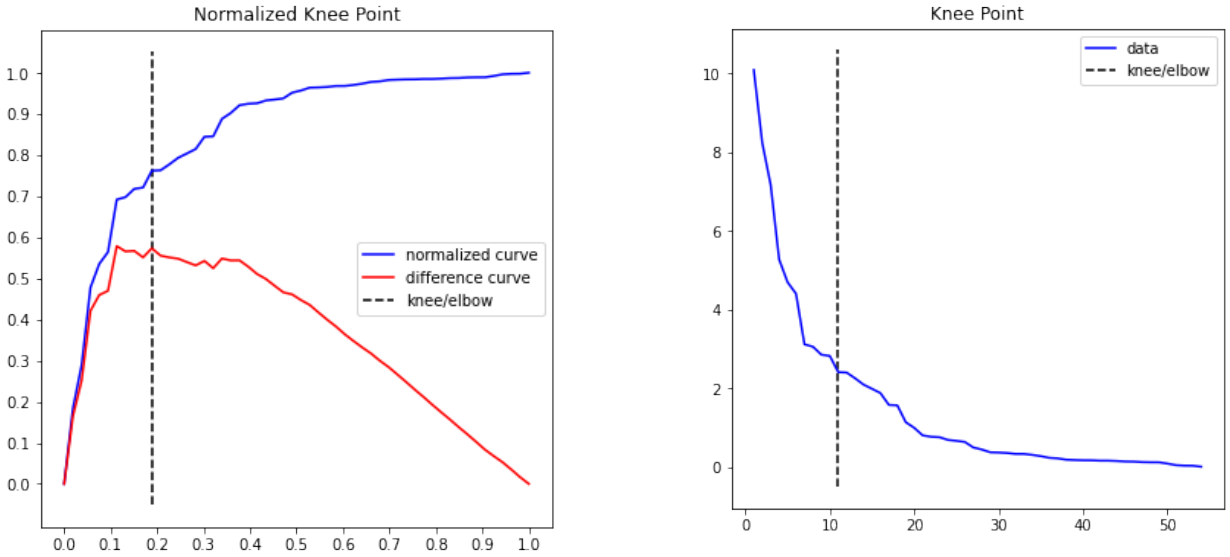


Fig. 28: Elbow point calculation for News normalized frequency. The curve starts to flatten around 4,68 pttw

token	Frequency	PTTW	IDF
No	34182	2,241911881	16,76142071
While	8544	0,560379589	16,76142071
Yet	3445	0,225948933	16,76142071
However	3297	0,2162419832	16,76142071
Despite	3191	0,2092897084	16,76142071
Indeed	1650	0,1082193729	16,76142071
Clearly	744	0,04879709904	16,76142071
Naturally	83	0,005443762393	16,76142071
anticipate	69	0,004525537411	11,99924677
unthinkable	65	0,004263187416	11,75078541
recognising	87	0,005706112388	11,45315301
reflected	153	0,0100348873	11,1930762
reflecting	256	0,01679039967	10,64112329
demonstrated	387	0,025382362	10,27831336

evident	356	0,02334914954	10,1821695
reflects	435	0,02853056194	10,17662931
demonstrate	611	0,04007396171	9,667185861
belief	668	0,04381244914	9,523642515
understands	447	0,02931761192	9,48417298
reflect	980	0,06427574873	9,237939394
convinced	1784	0,1170080977	9,211811542
inevitable	1034	0,06781747366	9,192008915
recognize	950	0,06230812377	9,142187291
confident	804	0,05273234896	9,038300615
stated	1429	0,09372453565	8,845707508
prove	1890	0,1239603726	8,595772782
assure	395	0,02590706199	8,468872193
knowing	2437	0,1598367344	8,369790739
obviously	2143	0,1405540097	8,257920843
surely	1628	0,1067764479	8,24163089
although	1260	0,08264024837	8,203469523
project	3121	0,2046985835	7,808297867
expect	4772	0,3129835438	7,673247969
expected	4483	0,2940287567	7,534223218
none	2348	0,153999447	7,080951714
cannot	7749	0,5082375275	6,962460684
knew	6374	0,4180547168	6,954003623
evidence	7326	0,4804940155	6,708091746
might	14990	0,9831566056	6,550558755
found	11860	0,7778677347	6,22871135
nothing	14641	0,9602665686	6,22228268
without	17598	1,154208802	6,075105246
shows	10968	0,7193636858	5,994768975
believe	33236	2,179866107	5,97282354
never	22768	1,493296171	5,49507376
did	30726	2,015241485	5,381820942
could	38375	2,516920263	5,278305176
think	47592	3,121440239	5,253045579
may	29048	1,905185662	5,203187324

see	92770	6,084552255	5,157403036
should	59934	3,930921148	4,905231832
know	63699	4,17785808	4,704230974
but	130433	8,554774219	3,789208352
not	193537	12,69360774	3,527084921

Table 14: Appraisal markers on the Twitter corpus ordered by IDF score.

trigger	Frequency	PTTW	IDF
Yet	15965	0,3609932798	14,38580399
While	105834	2,393070014	14,38580399
No	69454	1,570462089	14,38580399
Naturally	1122	0,02537015095	14,38580399
Indeed	14412	0,325877554	14,38580399
However	99367	2,246841167	14,38580399
Despite	29104	0,6580863399	14,38580399
Clearly	3530	0,07981874587	14,38580399
cannot	19	0,0004296193121	11,08996712
unthinkable	1246	0,02817398225	7,223406488
recognising	1777	0,04018071145	6,746642814
convinced	10016	0,2264772121	5,859254699
assure	5010	0,1132838291	5,850378026
evident	5207	0,117738303	5,781332786
reflecting	4972	0,1124245905	5,739338458
understands	5749	0,1299937592	5,726590534
belief	6438	0,1455731122	5,598125746
reflected	6826	0,1543463907	5,55467635
demonstrate	6811	0,1540072176	5,553362194
surely	5897	0,1333402675	5,508422031
demonstrated	7260	0,1641598003	5,493055216
anticipate	7208	0,1629840001	5,493055216
reflects	7546	0,1706267015	5,451085371
inevitable	7932	0,179354757	5,40098499
recognize	9398	0,2125032787	5,220774836

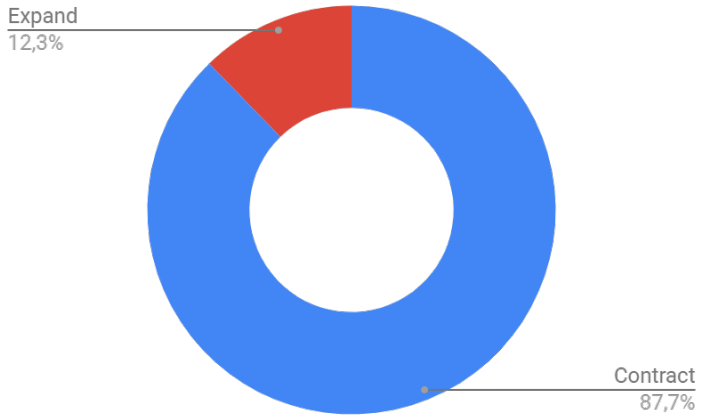
knowing	11907	0,2692356394	4,892090813
prove	13315	0,3010726916	4,882719375
confident	16238	0,367166231	4,681438328
knew	19358	0,4377142444	4,511384855
obviously	14374	0,3250183153	4,474001101
none	15454	0,3494387815	4,422068353
reflect	21654	0,4896303465	4,39875684
stated	30145	0,6816249559	4,064955693
project	27867	0,6301158615	3,949719533
evidence	33170	0,7500248727	3,919306412
nothing	35245	0,7969438239	3,79962371
expect	43723	0,9886444832	3,676175445
shows	50038	1,131436376	3,554591323
although	33688	0,7617376518	3,385105125
believe	124562	2,816538987	3,336279992
never	69012	1,560467788	3,159494495
found	82772	1,871602616	3,056128508
expected	87587	1,980477194	2,992360499
might	92292	2,086864503	2,946546589
think	106280	2,403154762	2,782571746
without	125917	2,847177627	2,582442743
know	134901	3,050319727	2,55183179
did	137564	3,110534266	2,538100576
see	316762	7,162477502	2,372873226
should	194435	4,396475313	2,189786726
could	233007	5,268647739	2,020763709
may	207388	4,689362626	1,828511247
but	364137	8,233699339	1,468422584
not	445855	10,08146939	1,368694463

Table 15: Appraisal markers in the News corpus ordered by IDF score.

As shown in our results, none of our most informative markers was also in the top occurring words. The only terms having highest possible IDF score in combination with

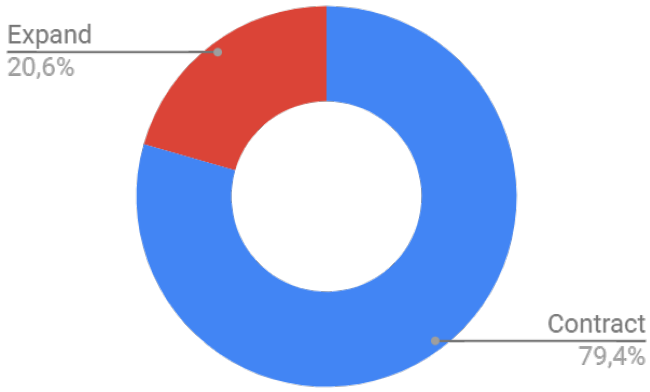
the pttw as close as possible to the elbow point are the words “No”, “While” and “However” within the News corpus and the word “No” within the Twitter corpus.

The calculation of the *elbow point* allowed us also to have a threshold value to determine which words were significantly occurring more than the others; given that the IDF score showed that the less informative words within both corpora are also the top-occurring. Hence, we removed all the words with a pttw value higher than the *elbow* and we computed the cleaned distribution for dialogic contraction and expansion within our corpora. The results are summarized in Figures 29-30. They show the proportion of dialogic contraction on Twitter (fig.29) and in the news (fig.30) after considering only words below our threshold frequency value.



Type	Markers pttw
Contraction	13,51
Expansion	1,89

Fig. 29: Cleaned distribution on Twitter



Type	Markers pttw
Contraction	25,21
Expansion	6,53

Fig. 30: Cleaned distribution in the News

To sum up, we defined echo chambers as linguistic structures of ideological isolation. Hence, considering these structures from the point of view of Appraisal theory, we should look for dialogic contraction, and in particular the ratio between contraction and expansion. According to what we analyzed so far in this chapter, contraction should be more present or more dominant in the social media than in the news. Within the Twitter corpus, nearly 90% of our markers introduce dialogic contraction while in the News Corpus, more than 20% of the markers introduce dialogic expansion. If we look at the normalized frequency, we see that in the Twitter corpus it is seven times more likely to find contraction (13 pptw) rather than expansion (2 pttw), while in the News Corpus it is only four times more common to find dialogic contraction (25 pttw vs 6 pttw).

Of course, the mere frequency of dialogic contraction is not quite informative for our goals. In fact, the absolute pttw of both dialogic types is influenced either by the two

corpora characteristics, from the actual distribution of the English words, and by the distribution of our list of markers. For this reason, it is necessary to look at the proportion between contraction and expansion, because this normalizes the first two types of biases. Indeed, having a sufficiently large couple of corpora, we should observe a comparable distribution of dialogic contraction and expansion if our search was biased by words' frequency. Instead, we observe an important difference in their distribution, although not statistically significant[46].

Regarding the use of "while" and "however" in the news corpus, we observe that when they are used to discuss the pandemic they are used as synonyms, as they share the same semantic preferences according to top collocates by t-score. This also emerges looking at the word embeddings, where we can also observe that they are close within the semantic space, indicating a high degree of semantic similarity. Given that they are both two adversative conjunctions we might conclude that they are effectively used as potential elements of *disclaim* (Martin and White 2005), the News Corpus. We should also acknowledge that "while" is often used as a temporal conjunction but very rarely to discuss COVID-19 issues. In a random sample of 200 concordances we observed this use only 1% of the times (examples 18-19).

(18) Denmark acted early , imposing a strict lockdown while paying wage subsidies that limited unemployment .

(19) Press Muslim worshippers wearing to help stop the spread of the coronavirus, offer Eid al-Adha prayer while maintaining social distance.

[46] Running a chi-squared test on our normalized occurrences resulted in a non-significant (>0.5) p value.

For our goals it sure more interesting to focus on the use of “While” as adversative, namely as synonym of “However”. The two are effectively used to contrast a presented standpoint, participating in the mechanism of *disclaim* (Martin and White 2005). In the following examples we present some cases of uses of “while” and “however, the two terms are highlighted in bold.

(20) Premier Inn claimed the world 's strongest brand title , **while** Airbnb was the most valuable leisure and tourism brand , with a value of $10.5bn (? 8.5bn) . **However** , Brand Finance said the pandemic would undoubtedly wreak havoc on hotels in the coming year and in terms of reputational damage , those that did not manage to avoid any association with the outbreak could suffer lasting harm . **While** Hilton 's revenue will take a hit following Covid-19 , Brand Finance said it was consistently boosting its reputation during the crisis.

(21) Many students may face greater food insecurity, loss of family income , loss of family members to the coronavirus , and fear of catching the virus themselves . **While** the scale of the COVID-19 school closures is novel , the inequalities in our school systems are unfortunately anything but new . Our models cannot account for the reality that the crisis is having an unequal impact on our most underserved communities.

(22) Incentives are needed to engage manufacturers for the large-scale capacity to guarantee sufficient production of SARS-CoV-2 vaccines . In line with this , various global organizations have come forward to expedite the process , such as Gavi , CEPI , and WHO . **However** , considering the pandemic scenario of COVID-19 , much stronger initiatives are required . In April 2020 , the Bill and Melinda Gates Foundation plans to help fund factories for seven promising vaccines , even before seeing conclusive data . The foundation aims to help scale up

manufacturing during testing , rather than after the vaccines have passed the trials .

(23) The poor response to the pandemic by the US seemingly gave China an opening to gain strategic ground . On balance , **however** , China 's pandemic diplomacy in the first half of 2020 has clearly failed .

(24) **While** the COVID-19 virus has claimed more than 100,000 lives , another , more subtle aspect of the pandemic has taken its toll as well . Fear of the virus is keeping many seriously ill or injured people from seeking treatment at local emergency rooms, resulting in numerous deaths which , according to area health experts , could have been prevented with medical

(25) If and when there is significant community transmission in SA, I think people living in informal settlements will be at the greatest risk of infection because they will find the preventative measures social distancing , hand washing more difficult. The People 's Health Movement (PHMSA) has been calling on government to fix the social determinants of health that include decent housing , access to clean water and sanitation and good nutrition. **While** the Covid-19 lockdown regulations state that people should stay at home and practice social distancing , Ngqola says where she lives , this is a " mission impossible " . She shares an outside toilet and tap with dozens of other shack dwellers and this puts her at high risk . " I often listen to the radio and hear them say we must practice social distancing . How do you do that in an environment like this ? This corona is going to claim many lives in informal settlements .

The use of adversatives is quite complex and variegate and it is quite hard to reduce to its quantitative evidence. As we observe in our examples, sometimes “while” and

"however" are used to effectively contrast a standpoint, more or less explicitly suggesting the reader a possible positioning (see examples 21, 22 and 23), while in other cases (examples 24 and 25) they present different point of views, enriching the complexity and moving the positioning stage quite outside the text, leaving more space to reader's interpretation.

We also explored the word embeddings of "while" and "hower", both as single words and as unite vector[47]. We used the same subsample that we illustrated in 4.5, namely a subsample of articles having the word "hoax" in it. Despite this sample is clearly one of the most focused on misinformation of our corpus, the role of "however" and "while" does not appear to be decisive in vehiculating doubts or presenting alternative views that might be considered misinformation, nor it seems involved in systematically contracting the dialogic space to exacerbate the ideological positioning. The unite vector of "however" and "while" shows indeed a quite weak relationship with the veridiction area. In the top-100 most similar word only the word "evidence" led us to possible pragmatic paths that end up in the semantic frames closer to "hoax" and other verdictive structures.

Compared to what we showed in 3.4, surely the News corpus appears quite different from Twitter, as the textual dimension seems effectively a dialogic space rather than a positioning space. The strongest use of adversative in some cases might influence for sure the shaping of public opinion and be an active part in the creation of echo chambers, i.e., it can be used as source for epistemic validation within social media, contributing to the polarization of the online debate. Nonetheless, this type of evaluation goes beyond the scopes of this dissertation.

[47] As we did in 3.21., we can unite the semantic space of two or more vectors and explore them as they were a single semantic unit.

Our goal is in fact to highlight, if any, possible linguistic pattern that may emerge in a quantitative analysis and be able to provide us the right clues to guide our qualitative analysis of echo chambers. Hence, to explore possible relationships and similarities among the two corpora, we must explore markers that are relevant for both corpora.

The combination of elbow point, IDF score and pttw highlighted a single marker, "No", that is suitable to compare our two datasets. It is not particularly surprising that this shared word is a negation such as "No", since it is one of the simplest ways of silencing external standpoints. We started by comparing the top ten clusters in both corpora. We chose clusters as a starting point so that we could immediately highlight the most relevant contexts in which "No" is used, searching for clusters containing verbal elements that are most likely involved in dialogic contraction processes.

N	**Cluster**	**Freq.**
1	THERE IS NO	17.498
2	THERE ARE NO	4.394
3	NO HEALTH CARE	4.257
4	OUT WITH NO	4.179
5	NO TUITION REFUND	4.151
6	WITH NO TUITION	4.148
7	UNEMPLOYEMENT NO HEALTH	4.143
8	NO SOCIAL DISTANCING	3.893
9	NO ONE IS	3.386
10	NO AVAILABLE VACCINE	2.554

Table 16: Top 10 cluster of "No" on Twitter.

N	Cluster	Freq.
1	THERE IS NO	49.950
2	THERE S NO	20.773
3	THERE ARE NO	15.623
4	THERE WAS NO	15.421
5	WILL BE NO	6.890
6	THERE WILL BE	6.562
7	WE HAVE NO	6.559
8	THERE WERE NO	6.387
9	IS NO LONGER	6.109
10	NO MORE THAN	5.795

Table 17: Top 10 cluster of "No" in the News.

Although existential clauses might be used to contract the enunciation space within the discourse, with the *denial* mechanism (Martin and White 2005), we should not be surprised by their abundance in a context such as the pandemic, since most of the discourses are centered on the account of the existence of the virus. However, although this might of course be used in a context of dialogic contraction, to support a discourse on the existence of the virus or on the danger of the virus, the word "No" might be used far from the predicate, thus being less likely to be a good indicator for dialogic contraction.

Taking into account these two macro-typologies of "No" uses, we took a sample of 200 occurrences of the top-occurring clusters in both corpora, namely "There is", to evaluate

how many times it was actually used to introduce dialogic contraction. Again, in the Twitter corpus we found a more marked conflict, with a greater use of “No” as contractor of dialogic space (48 occurrences, 25% vs 30 occurrences, 15%, in the News).

However, in both cases 70% of our top-occurring cluster are not attributable to case of dialogic contraction, suggesting that our approach to appraisal might not provide enough elements for a qualitative investigation of the appraisal typologies. Indeed, as suggested by Hunston (2004), we are dealing with a subject that is rather difficult to grasp in a quantitative evaluation, since the restriction of dialogic space can also occur by means of allusions that are expressed with a variegate sequence of words on the discursive level, thus not occurring with regularity, and therefore difficult to detect by studying the collocations of lexical elements. Moreover, many lexical items on which we are basing our observations on appraisal can certainly occur in contexts where their function has nothing to do with the dialogic aspect. This ambiguity is evident on negations, such as "no" and "not," which certainly play very transversal grammatical roles, but it appears equally evident on verbs, as we have seen in the case of "know". The case of the verb "think", on the other hand, seems to lead us to identify instances of evaluative language use more easily in dialogic contexts.

Nonetheless, it is interesting to select some examples from our top occurring cluster, i.e. “There is no”, to illustrate the variety of elements that can be denied. See for example how the expression is used in denying the existence of the virus (26), in summing up a position or reporting it (27) and in rejecting the validity of a policy (28). All these examples come from the Twitter Corpus.

(26) COVID 19 = Exosomes naturally found in all cells. Cells excrete in times of stress or illness There is no virus. Only flu etc.

(27) #CureCancer_By_TrueWorship There is no such disease which cannot be cured by the devotion of Sant Rampal Ji whether it is corona virus or cancer. All diseases can be removed, but true devotion

(28) There's no point to let unlimited #coronavirus infected patients to enter Hong Kong when we are already running out of medical resources

It appears that forms introducing a proposition (rather than just the existence of an entity or process), like Examples 27 and 28 are particularly apt at taking distance from this proposition and acting therefore as forms of denial from the point of view of engagement, but also as disagreement from the point of view of argumentative dialogue. In the specific case of the COVID infodemic, forms of denial of existence were also central to the debate, when directly related to the existence of the virus, for example.

On the other hand, the News Corpus shows also some variety, thought being quite different from the Twitter Corpus.

(29) 1920 spanish flu killed 50-100 million people worldwide. It was like the end of the world for everyone then. we can be sure the people then tried everything they could. We don't know what this covid 19 will end up doing as there is no sign it is abating.

(30) There is no specific treatment for COVID-19 . Although vaccines can be developed to treat viruses , owing to the novel nature of this infection , no vaccine has currently been developed and the process to develop one may take 12 to 18 months18 .

(31) In spite of the concern on possibility of community transmission of COVID-19 ; there is no evidence to prove that deaths recorded recently were as a result of disease .

(32) While some people may be wary of using public pools when they do open , the Centers for Disease Control and Prevention states that " there is no evidence that the virus that causes COVID-19 can be spread to people through the water in pools , hot tubs , spas , or water play areas . "

It is important to notice that the role of the single markers in determining a contraction of the dialogic space relies on the lexico-grammatical patterns involved, beyond the simple presence of a word. In the case of the News Corpus this is particularly evident in examples 29-30, where the contraction of the dialogic space is projected toward the future, vehiculating the uncertainty typical of the debut of the pandemic. Another quite common use of "No" in the News is that showed in examples 31-32, with explicit veridical intent, negating the existence of something, namely how is the virus spreading and affecting population.

It should also be noted that sometimes the use of negation is determined by need of describing a situation rather than restricting the dialogic space around a specific viewpoint. This might occur when "No" is used in its predicative form, for instance when reporting an information that is not used within a dialogic exchange. This happens both in the News Corpus (33) and in the Twitter Corpus (34).

(33) There were no new cases today in London

(34) "#COVID19 has NO cure yet, but would be effectively prevented if all would wear a mask"

This latest use of "No" is also quite debatable. We might say that in example 34 the standpoint is introduced by "but", another marker of dialogic contraction; nonetheless, we should look at a larger context which is, unfortunately, not available on Twitter

anymore. In fact, "No" would be clearly playing a key role in dialogic contraction if this sentence was part of a broader conversation with someone suggesting the existence of cures for COVID-19 at the beginning of 2020. In that particular case this negation would create a denial situation, expelling from the dialogue voices stating that there is actually already a cure for the coronavirus.

Nonetheless, if that is true, we should be able to spot more potential referents for "No" other than terms used to describe the pandemic scenario or modal verbs used to determine the epistemic value of the denied voice. However, in both our corpora 90% of the top-100 collocates, ordered with t-score, are either attributable to the pandemic scenario or to verbal elements expressing modalities. There are, indeed, a few rare exceptions, i.e., experts narrowing the perimeter of possibilities on some characteristics of the virus (see examples 35, from the Twitter Corpus and 36 from the News Corpus).

(35) "Fauci: No scientific evidence the coronavirus was made in a Chinese lab"

(36) "The decisions of all levels of the government are putting Americans at risk and will speed the spread of the coronavirus . No matter what politicians say [...]"

4.7 CONCLUSIONS

As we hope to have shown in our analysis, we can detect echo chambers starting from linguistic markers of ideological conflict. For the sake of brevity, we focused on the word *hoax,* as it emerged as a keyword from the comparison between our two corpora. We found evidence that in the News Corpus the ideological conflict is mostly reported; stance is hardly taken by journalists and, when they do take a stance, *hoax* is never used as an ideological keyword explicitly manifesting the position of the journalist. On Twitter, on the other hand, the word *hoax* is not only discussed as an important keyword of the

political debate, but it is also appropriated by many participants in its accusatory pragmatic function. This becomes indexical of their position and their political identity, in contexts that highlight the conflict between positions, often without providing explicit arguments.

The word is thus not only a marker of ideological conflict, but also an ideological keyword of the republican area, a *cultural keyword* (Stubbs 2001) that evokes the whole context of the debate and derives its force from the discourse it belongs to, depending as it does, on a set of unstated premises. It is also a word that recalls the basic principles of Conspiracy Theories, as it refers to a process of intentional and concealed deception, hinting at a set of alternative explanations and pointing at the notion of misinformation and fake news, thus contributing to the construction of conflictual social identities.

Regarding the analysis of appraisal, it emerged that there was an interesting (though not significant) difference in the proportion of dialogic expansion and contraction on Twitter compared to the News. This difference, along with the findings of our qualitative exploration, seems to highlight a remarkable difference between Twitter and the News, where Twitter is a space used (almost exclusively) for ideological positioning, and the journalistic discourse has a tendency of representing a more varied and complex dialogic dimension.

This extreme tendency of ideological positioning seems to be expressed mostly using the *denial*, which is a rather typical tendency of echo chambers, where polarization is expressed precisely in this dual nature of ideological reinforcement through the denial of the other. We may therefore conclude that our work confirmed that the space of ideological positioning on Twitter might be seen as a portion of the dialogic dimension of echo chambers, even though it is surely not comprehensive.

In fact, we found some difficulties in finding data-driven explorable patterns starting from our corpus. The corpora taken from social media are difficult to segment, since it is complex to define sub-genres or any hierarchical relationships that would then allow the creation of sub-corpora made with robust criteria. Hence, in this case it was certainly helpful to define a list of terms of interest to search in our corpora, quantifying their use. This also allowed us to define in what proportions we could identify the dimensions of dialogic expansion and contraction within our two corpora.

However, a subsequent qualitative analysis of the most informative terms, such as" While", "However" and "No", seems to indicate that the use of *disclaim* is quite varied and complex, while the *denial* might participate in other functions rather than argumentation and stance taking toward the main topic. These findings show that it is quite difficult to investigate the appraisal dimension starting from its quantification. Individual markers of dialogic contraction, even when are clearly elements of *denial*, are potential markers of dialogic contraction but that are difficult to use as pivotal elements for an automatic measure of dialogic contraction. The elements involved in the appraisal mechanism are often very versatile lexico-grammatical components, which might participate in many other functions other than argumentation. In a nutshell, it emerged quite clearly that appraisal is often expressed with complex references to textual (and sometimes extra-textual) elements that escape a quantitative evaluation.

It would therefore be useful, for future research in social media, to implement a tagging methodology on a sub-corpus, to use it together with a predefined list of appraisal markers as we did in our case. The main limitation of the approach we used is that it does not allow us to focus on the various nuances of use regarding the dialogic dimension. Since the appraisal is difficult to quantify it should therefore be useful to adopt a more

qualitative approach, which would be able to highlight the complex nature of the appraisal dimension.

5 CONCLUSIONS

5.1 RESEARCH QUESTIONS RECAP

As we introduced in section 1.3, we had three main research questions on filter bubbles and echo chambers:

a) Do they exist?

b) How can we measure their effects?

c) Can we produce a new conceptualization?

Regarding their existence, we found clear evidence of existence of filter bubbles in chapter 3. The filter bubble effect is real, meaning that not only content is distributed unevenly online, but that the hyper-personalization might affect the information diet of users. On the other hand, we also found evidence of echo chambers' existence, also from a linguistic perspective, highlighting in chapter 4 that on Twitter users tend to use the platform as a space of ideological positioning.

We had quite satisfactory results developing a methodology to study and measure the filter bubbles. In 3.2.3, we showed how to collect evidence of uneven content distribution within YouTube, while in 3.2.4 and 3.3.1 we showed how to have linguistic insights of the algorithmic personalization. All the steps of this methodologies are fully reproducible, and they allow further research on algorithmic personalization. Echo chambers instead are more difficult to quantify and measure. We did find evidence in 4.4 and 4.5 that we can explore this phenomenon starting from textual patterns; nonetheless, section 4.6 showed that measuring the echo chamber is difficult as it relies on linguistic nuances that are quite difficult to quantify. Still, we may have an idea of the "thickness" of an echo chamber by looking at the proportions of the dialogic contraction compared to a reference corpus.

Finally, in this chapter, we try to use our experimental findings to seek producing a new conceptualization of the two phenomena we studied. In 5.2 we trace a hypothesis that we call "the algorithmic model reader", namely a possible semiotic theorization of filter bubbles. In 5.3 we try to reason on the semiotic aspects of echo chambers that we explored, having in mind that the methodology needs some improvements. Finally, in 5.4 we present the details on the main methodological limitations that we found during our studies

5.2 FILTER BUBBLE

Our methodological approach (Figure 31) to filter bubbles allowed us to replicate our experiments on different platforms and also to highlight the effects of algorithmic personalization on social media discourse. In fact, our experiments showed that filter bubbles often tend to pander to ideological extremism, showing users content that could potentially get them into an echo chamber. This appeared especially true and troubling in the YouTube experiment, where users were exposed to online conversations that could very well be defined within a linguistically echo chamber dynamic, a dichotomy between two mutually rejecting counterparts.

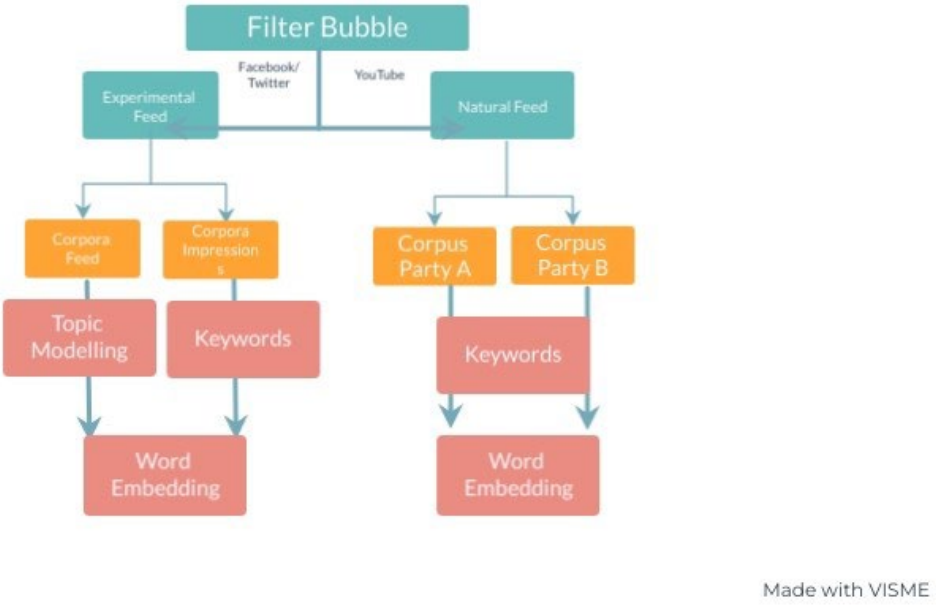


Figure 31: The methodological approach for a linguistic analysis of filter bubbles.

It is important to highlight that what emerged on echo chamber is quite worrying if combined with what emerged in our filter bubble exploration. Although research on the intersection between echo chambers and algorithmic personalization is quite recent, our findings showed that, potentially, the filter bubbles might favor the formation of echo chambers.

As we have shown throughout this dissertation, we can imagine filter bubbles as in fact a model reader, namely we can reconstruct the textual intention that is produced by the personalization process. In the filter bubble case however, this model reader is peculiar, as it involves the intent of an *algorithmic model reader*. We may say that the algorithmic model reader is the *algorithmic intention*, which is made by the ensemble of information that the algorithm has about the users, which is constantly updating due to the feedback that users given by interacting with content. This algorithmic model reader coexists with the model reader created by the aggregation of texts, and it determines its genesis. In the ecoian conception, the model reader is a semiotic device that is at the same time

embedded in the text but at the same time produced by the empirical author to reach a specific reader and guide him accordingly in the interpretive path.

In the case of filter bubbles, it is the algorithm that must imagine its potential reader and makes inferences about him based on his online behavior. After that, the algorithm selects the contents based on the previously made inferences, selecting a personalized feed for each user. The real strangeness of this model reader is twofold: 1) the algorithmic model reader does not have the semiotic goal of guiding the reader in the interpretive process, but rather the goal of engaging the reader as much as possible with the content, 2) the presence of the algorithmic model reader is invisible to the eyes of the empirical reader.

Regarding point (1), it must be said that even when it comes to communication between human beings one of the implicit goals of the text is to be sufficiently engaging and interesting to be read by the empirical reader. However, this is never the ultimate goal of communication, it is rather a mean to get a message to its destination. Social media algorithms, on the other hand, collect a set of information about users' browsing behavior, combine it with the knowledge they have about similar users, and finally they derive an output, which is the personalized feed. This personalized feed has the sole objective of pleasing the user, of rendering a service to the empirical reader. The condition of happiness of the algorithmic model reader is apparently resolved by the simple fruition of the text and it has nothing to do with the interpretation of the produced text. However, it certainly creates a text, or rather a complex network of texts and intertextual references, which reaches the user as if it were a natural selection and not mediated by his own behavior. Hence, we could say that the algorithmic model reader exists prior to the existence of the actual model reader, which is the whole of selected texts.

Coming then to point (2), the problematic of the algorithmic model reader is the union of its appearing with its invisibility. The social media feed is in fact liquid by definition, in the most extreme sense in which it was intended by Bauman (2000); it is an ephemeral text, without visible limits and without the marques (Lorusso and Violi 2004) of those who produced it, namely who made the content selection. This makes impossible to the empirical reader to become aware of the subject who produced the textual flow in which the user is immersed, that is, it makes impossible to foresee the Model Author of that text. Moreover, the personalization output (i.e., the whole feed of contents) is never fully available to the user, meaning that also the actual model reader is only partially available to the user.

At this stage, the algorithmic model reader becomes extremely influential and decisive in favoring the creation of echo chambers. It is not personalization per se that is problematic, as it is part of the human nature to affiliate with groups of individuals we believe to be similar. From a purely technological point of view, as we have already said, it is indeed desirable that we have algorithms capable of selecting the most relevant information for us.

The problem arises, however, when personalization brings with it human bias, fostering the polarization of the debate. Whether or not echo chambers are an exclusive phenomenon of digital spaces may never be possible to conclude, but what is certain is that by selecting digital content, we potentially select those that are the founding nucleus of echo chambers.

We have often remarked that echo chambers are a phenomenon actively created by people, while on the contrary filter bubbles are something suffered passively. This is true even if we think of these two phenomena in a purely linguistic perspective. The generation of an extremely polarized algorithmic model reader creates in fact a very fertile digital

space for the birth of an echo chambers. The model reader is an active device, a real guide that pushes and constricts the various pragmatic spaces, reducing the field of possible interpretations and guiding the reader in its inferential path. If within this inferential path, for the sake of the algorithmic model reader, the user finds a model reader that reduces the space of the dialogic dimension or a model reader whose condition of happiness is the participation in an ideological conflict, the users might end up very easily within an echo chamber.

Certainly, it is true that the user is able to arrive at what Eco called "aberrant decoding", that is, the use of a text outside the possible interpretations that the same text imposes on its reader. However, this is the exception to the rule and not the norm. What is likely to happen within echo chambers is that aberrant decoding is used instrumentally to silence or overturn dissonant voices with respect to the ideological core of the echo chamber, that is, everything that the echo chamber wants to reject outside itself.

In this case the dynamics of dialogic contraction may have the function of nourishing the centripetal force of the echo chamber, compressing the pragmatic space of possible interpretations to the point that it is possible to do only two things: blindly believe in the ideological dictates of the echo chamber in question. or escape from it. In this case, therefore, dialogic contractions are a real tool at the service of the echo chamber itself and its proliferation within the digital space. The systematic reduction or rejection of different standpoints, which potentially introduce other nuclei towards which inferences can gravitate, is necessary for the survival of the echo chamber itself. In a nutshell, it is really difficult to escape from an echo chamber, and it might become even impossible if we are led into that echo chamber by a non-transparent algorithm, since we would miss the semiotic and cultural tools necessary to dissect the ideologic intention of the algorithmic model reader.

The results of the experiments on algorithmic personalization illustrated in this thesis showed the need to deepen research in the field. The worrying tendency of algorithms to expose users to information that could produce a spiral of confirmation bias and ideological radicalization is not to be underestimated. At the time of writing this thesis (2022), there is almost no academic or political attention to investigating the effects of filter bubbles on the general public. The selection of personalized information and content, especially when it comes to particularly sensitive topics, such as the pandemic or an assault on parliament, should be as limited as possible or at least the algorithmic criteria that generated that precise personalization should be made explicit.

Returning to what was previously theorized, it is therefore necessary that the social media algorithms explain the algorithmic model reader developed to produce the personalized feeds, making the criteria that led to the generation of that specific customization transparent.

Indeed, we could say that the real semiotic criticality lies in the asymmetric user-algorithmic relationship, where there is an algorithmic model reader but there is no algorithmic model author, i.e., there is no user idea regarding the textual intentions of the algorithm. In this way, the real risk is that the information selected specifically for a specific user will be taken by that user as a faithful representation of ontological reality and not as a review of contents specifically intended and designed for him. It is precisely on this aspect that the fundamental difference between algorithmic personalization and simple selective exposure arises. What is often criticized about the filter bubble theory is that in fact, like the echo chambers, they would be nothing more than a digital transposition of the well-known phenomena of selective exposure, a sort of ideological cherry picking that makes us feel comfortable and quiet inside in our consolidated space. Let's take for example a devoted reader of a left-wing newspaper and a devoted reader of

a right-wing newspaper. In addition to reading their respective newspapers, it is very likely that both have political likes and dislikes that result in similar and equally differentiated reading of other newspapers, attendance of online and offline groups consistent with their political values and so on.

Although we can argue about the degree of awareness of this selective exposure, we can undoubtedly say that it is a deliberate exposure, as choices are actively made by two individuals. Furthermore, when they read their newspaper, they will be convinced and very adherent to the vision of the world that is proposed to them, but at the same time they would know they are reading a newspaper, which is a text that has clear limits and clear semiotic traits (Violi, Lorusso 2004); moreover, every newspapers carries a sufficient amount of semiotic affordances that allow the reader to have the ability to reconstruct the model author. This does not happen instead in the case of the filter bubble as 1) the exposure to the contents remains a substantially passive action and over which the user has no degree of control, 2) the selection of the contents takes place in a non-transparent way that disallow the reconstruction of the algorithmic intention and 3) the algorithmic intention has constantly available the empirical reader behavior, which cause a constant evolution of the algorithmic model reader. Hence, the degree of pervasiveness that the algorithmic model reader has been certainly much more important than the model reader of a text to which one is selectively exposed, especially because of the asymmetric relationship.

In any case, it is unreasonable to think that suddenly black box algorithms will disappear and the philosophy of approaching digital will change. It is therefore necessary to work on algorithmic explicability, as brilliantly addressed by Watson and Floridi (2021). We suggest that the humanities, united by an interdisciplinary effort, while remaining focused on each other's expertise, should work in this direction. Moreover, in

our opinion it would be necessary to adopt solutions that can empower users by giving back them control of their data; this would allow to a widespread culture about algorithmic operations, making this topic central to the public debate and mitigating the negative effects that we have described. A possible solution in this sense is that of serendipity (Reviglio 2017, Reviglio and Agosti 2020), i.e., the design of algorithms that allow users to discover content that does not belong, not even by affinity, to the sphere of their preferences.

Future research should therefore focus, more than on highlighting the problems of platforms, on which we now have over a five year period of multidisciplinary literature, on possible tools and solutions available to users and companies to be able to evaluate the algorithmic choices deployed.

5.3 ECHO CHAMBERS

In the end, we had satisfactory results for the filter bubble analysis, while the study of echo chambers needs to be integrated with qualitative analysis and deepened on different case studies and platforms. In fact, on the one hand the idea of looking for possible traces of ideological conflict within the most frequent words seemed well-founded, on the other hand it was more complex to find quantitative insights on the dialogistic part, a part that is necessary to have a full linguistic understanding of echo chamber phenomenon. The distribution of appraisal markers indeed, with all its limitations, might confirm the line proposed in the literature on echo chambers, namely that they should be not considered part of normal communication processes.

In fact, the presence of ideological conflict united to a quite extreme tendence to dialogic contraction is indeed problematic. To better understand this, it is useful to use Lotman concept of semiosphere (1984). The semiosphere is a concept developed in

analogy with the concept of biosphere, to indicate the semiotic complex that constitutes our cultural space. According to Lotman, the semiosphere has precise roles and rules that regulate the processes of semiosis and communication with the outside world.

In fact, in Lotman theory there is indeed dialogue between semiospheres

The dialogic (in the wider sense) exchange of texts is not a facultative phenomenon of the semiotic process. The isolated utopia of Robinson Crusoe, a product of 18th century thought, conflicts with the contemporary understanding of consciousness as the exchange of communication: from the exchange between hemispheres of the great brain of man to the exchange between cultures. Meaning without communication is not possible. In this way, we might say, that dialogue precedes language and gives birth to it.
(Lotman 1984, p.218)

These communications with the outside are crucial, as nothing can enter the semiosphere without first being properly translated. In other words, what is alien to the semiosphere must pass through the mediation of someone or something, be it a social actor or a linguistic act. Lotman himself gives the example of our sensory apparatuses, which translate the stimuli we receive from our surroundings into something intelligible to us.

The boundary between what is inside and what is outside the semiosphere is what divides, from the point of view of the semiosphere itself, chaos from order, the acceptable from the unacceptable. Indeed, if we consider an echo chamber as a binary structure, i.e., one that exists solely as a set of two irreducible opposing parts, we might consider the dividing line between these two poles as a fracture line that arises within a semiosphere and potentially creates new ones.

This intuition already exists in Lotmanian theory. Although Lotman focuses on literary examples, in his work he in fact introduces the concepts of asymmetry and dialogue, and in a nutshell suggests that cultures undergo cyclical transformative processes, leading

what was not the norm to become a new cultural canon. Echo chambers could therefore be considered part of these transformative processes and perhaps for a full understanding we would need to study the typology and the topology of this cultural relations, together with the study of their discursive manifestation.

In our case, we might consider the dynamic of our echo chamber as a *reflection*, as suggested by Leone (2018).

> *The third type of symmetry is observed in a semiosphere when it undergoes an operation of reflection. Such operation implies that the semiosphere contains not only a center, but also an axis, an imaginary line created by the symmetry of fields of semiotic forces created by contrasting but parallel agencies around the center. The division between Guelphs and Ghibellines, that is, the factions supporting the Pope and the Holy Roman Emperor, respectively, in the Italian city-states of Central and Northern Italy, was largely one configuring a semiosphere characterized by symmetry under reflection; the axis dividing the two orientations, indeed, would not separate different political systems, but different choices in attributing the same power, with essentially the same modalities, to either the Pope or the Emperor.*
>
> (Leone 2018, p. 177)

This process of cultural reflection also lends itself well to the description of echo chambers, irreducible dichotomies that form around a veridical dispute, to whom to attribute truth. The claim of the two disputing factions is the same, a reflection indeed, since both claim to have a model that allows to accurately describe the reality that surrounds us.

However, it must be emphasized that echo chambers are extremely local ideological grammars, i.e., originated around phenomena that are indeed part of reticulated and complex socio-semiotic systems, but at the same time are self-standing and autonomous. Taking our case study, the echo chamber with respect to COVID-19, it is possible to recognize with precision its boundaries, which are in fact those of belief/non-belief in the pandemic scenario and in the consequent sociosemiotic positioning on whether to have

confidence in the resulting social rules. These echo chambers are part of a complex and varied ideological fabric, perhaps made up of conspiracy theories, distrust in science and institutions. This intertextual network of semiotic dialogues, however, is not univocally and easily identifiable, because we can only observe its discursive manifestation, which is nothing but the facade, the product of all the complex social and semiotic dialogues that take place in the background. Thus, returning to the question we posed at the beginning of our study of echo chambers, the debate on their exclusively dysfunctional nature remains open. Besides our methodological limitation, we observed and studied only one echo chamber on a specific topic, but, as we said, the ideological conflicts of echo chambers are observable only as extremely local phenomena, i.e., limited to a specific topic. This means that our considerations might be limited to disputes around the concept of scientific truth.

Echo chambers then, from a semio-linguistic point of view, appear to us as a simultaneous contraction and distancing of semiotic spaces, where the boundaries of the semiosphere are sealed to any otherness. Everything that is other, with respect to that specific ideological space, is rejected and dismissed; nonetheless, the object of the dispute is the same, i.e. the protection of the center of that specific semiosphere, where there is the ideological core.

One of the reasons why in fact it is easier to study echo chambers on scientific issues is the fact that the hierarchical relationship between the two parties is absolutely clear, scientific norms are the center of the semiosphere while the dialogue takes place with a counterpart that rejects this center.

What really distinguishes an echo chamber and a normal divergence of opinion is probably the nature of the ideological conflict. As we have seen in the study of the filter bubble, when we observe the discourse we might find two types of polarization. A weak

polarization, which is characterized by a semantic relation of contradiction, and a strong polarization, characterized by the relation of contrariety. The weakness of the first type of polarization is inherent in the very fact that it is based on the negation of the opposite term, thus placing itself immediately in a subordinate position (Lakoff 2004), while in the relation of contrariety there is a real opposition of semantic frames and consequent underlying axiologies. Simply put, a major difference between a disagreement and an echo chamber is the robustness with which conflicting axiological planes are contrasted. If I simply deny the existence of the coronavirus, I do not directly enter an echo chamber. I can say I am in an echo chamber if I deny the existence of the virus motivating it with conspiracy beliefs and the total rejection of the scientific method. Indeed, I am also in an echo chamber if I refuse any debate on scientific issues on the virus, accusing the counterpart of being anti-scientific.

What the echo chambers, even of different topics, might have in common is the firm will to preserve and fix the belief, i.e., the deepest ideological core around which the conflictual relationship with the counterpart is built. No one is in fact willing to constantly question what he considers to be a distinctive feature of his identity, both for the cognitive and psychological fatigue that would be necessary, and because that questioning requires a repositioning within the socio-semiotic spheres that has an elevated cost. This cost corresponds to the truth value that is conferred to the credence, and it is indeed the cost of its falsification (Compagno 2018b).

To better understand the concept of the cost of falsification we need to take a step back and introduce the concept of what falsification is and what revolves around it. The concept of cost of falsification has nothing to do with empirical verifiability, but rather with the sociosemiotic cost that falsification entails. Already in Russell (1940), the concept of true and verifiable are in fact two distinct things.

"*True" and "false", we decided, are predicates, primarily of beliefs, and derivatively of sentences. I suggest that "true" is a wider concept than "verifiable", and, in fact, cannot be defined in terms of verifiability*. (Russell 1940 p.227)

Following Davidson (2001), we could say that, from a purely linguistic point of view, the truth value of any utterance is given by the relationships it has with the two speakers and with the sharing of the same dimension, which can be time but also a cultural space. It is precisely in this cultural dimension that a pragmatist theory of truth, which is the framework proposed by Compagno, finds its place and foundation. This theory is rooted in Peirce's semiotic theories (Peirce 1867, 1877), that stated that it was possible to achieve truth even in a semiotic dimension.

Two different reasoners might infer the same conclusion from the same premises; and yet their proceeding might be governed by habits which would be formulated in different, or even conflicting, leading principles. Only that man's reasoning would be good whose leading principle was true for all possible cases. (Peirce 1902 CP 2.589)

Compagno instead takes up Umberto Eco's reworking (Eco 2000, 2014) of Peircean theories, reversing the perspective and then stating that the cost of falsification corresponds to the truth value of an utterance.

1) *A statement is true relatively, not to a fact but to an effort of falsification (at a first glance, all statements are true, extreme effort falsifies any statement).*

2) The human effort needed to falsify a statement depends on a complex cost (cognitive, social, technological, economical and other conditions determine how hard it is to falsify a statement in a certain community at a certain time).

[…]

(Compagno 2018b p.286)

In the echo chamber dynamic, this cost of falsification is biased by the total absence of the dialogistic dimension, meaning that it creates a semiosphere in which it is impossible to withdraw to have any change of opinion because the cost that renunciation, is extremely socially high, because of the particular semiotic configuration we have described, so that the falsification of the echo chamber's nuclear belief would correspond with the implosion of that semiosphere itself. In this sense, the cost of falsification can be understood as the modification of what Wittgeinstein (1969) would have called grammatical utterances, i.e., the propositions whose falsification consequently causes a significant modification of the worldview.

Translating this to an echo chamber, it would result in an extreme search for confirmation of one's beliefs and a total absence of dissonant voices, thus creating an isolated and seemingly non-reducible ideological space. However, our experiment did not shed light on the dysfunctional dimension of echo chambers. In fact, we did find proof of ideological isolation following the pragmatic contexts and the inferential path of hoax, which is coherent with the social science definition of echo chambers. We also observed, trying to analyse the dialogic dimension, the tendency of using Twitter as a space of ideological positioning rather than as an informative space. None of these findings however prove, from a linguistic perspective, that echo chambers are a dysfunctional phenomenon. For sure, they might favor polarization, but it is still unclear whether they amplify a natural tendency of human society or if they are a pivotal cause of pollution of the public debate.

Nevertheless, we can still draw the conclusion that we should be able to see echo chambers as observable patterns in lit data. Further research on these themes should focus on how to distinguish echo chambers from normal disagreement and how to explain their genesis. The difference between our corpora in fact did not clarify the origins of echo

chambers. Yet, it seems to suggest that, while acknowledging its informative function, many users look at Twitter as a site for positioning themselves and others, with greater emphasis on identity construction than on developing an argument. In other words, if there is an echo chamber on Twitter, it is not because of its affordances, but rather because its users respond to a context that is characterized by a polarization of conflict. It is not the medium itself, but rather what Tagg et al. (2017) have called "context design", i.e., users' "perceptions of what the technology is for, how it functions for this purpose, who they are communicating with, and the appropriate norms for doing so." (Tagg et al 2017: 5)

Surely one possible line of research might be the investigation of patterns of dialogic expansion and contraction. In fact, although with methodological limitations, our results showed an interesting difference in the distribution of markers of dialogic contraction on Twitter, compared to the News corpus. Yet, it might be a characteristics of social media language or a peculiar quality of echo chambers; our investigation did not provide clear evidence on this aspect.

5.4 METHODS

What we tried to achieve in this thesis was to develop an interdisciplinary and, in principle, reproducible methodology for the semiotic analysis of filter bubbles and echo chambers, thought with some different research questions. Regarding the methodological side, we might say that our experiments showed promising results on the filter bubble analysis, while more work is needed for a fine tuning of the echo chamber investigation. According to our experiments, a data-driven first approach is feasible when studying algorithmic personalization, while in the study of echo chambers it seems difficult to find

solid quantitative evidence, as the argumentative functions are often made by complex and nuanced linguistic utterances.

Of course, as already discussed in the methodological section, there are limitations that need to be considered. From a technological point of view, the use of word embedding makes the analysis more powerful but also more difficult to reproduce, since they are a probabilistic model. Furthermore, the search for markers related to the dialogistic sphere may not be completely reliable as it is based on a classification of appraisal made on a corpus of CSR communication and because it is difficult to appreciate qualitative details following a full bottom-up approach. Further research is needed to deepen the methodological aspect of applying a corpus-based analysis of appraisal on a social media corpus. In particular, it was quite difficult to find relevant words to analyse following a data-driven perspective, meaning that there were no clear evidence of significant words having argumentative functions. This certainly caused a lack of qualitative exploration of appraisal markers and therefore we recommend future research to focus on the role of dialogic contraction in echo chambers, also using different methods that may not have a data-driven-first approach.

Nonetheless we might conclude that we found first evidence of how we can look for echo chambers in a corpus-driven approach, letting them emerge from the discourse rather than observing social media interactions. However, this first approach to echo chambers has been limited to Twitter. Although we used quite large corpora, our data do not allow us to generalize our findings to other platforms and it would be surely interesting for further works to focus on other similar debate-centered social media (e.g., Reddit).

On the contrary, we were able to test the methodology proposed to investigate filter bubbles from a cross-platform perspective, working on Facebook and YouTube.

Therefore, if we could consider the considerations made so far, regarding the algorithmic influence on the polarization of the debate, more robust, we must underline that the weak point of the two case studies is the amount of data collected.

Although in fact they were both of sufficient quantity to be able to have satisfactory textual insights, data collection is difficult and less easily repeatable. In fact, despite the standard procedure, it is not possible to know for sure whether we are able to reproduce the same results. Furthermore, the data collection in this case is carried out by recording what happens on users' screens, making this step much more expensive, both in terms of time and technological effort. Another important methodological limitation is that we focused on a text-driven approach which did not include any media analysis nor a deep qualitative analysis of user experience. Both aspects are quite important in enhancing the understanding of filter bubbles' impact and they should be included in further experiments on algorithmic personalization, in order to update and expand the research protocol.

Finally, at the moment the tools used in this work operate outside what would be allowed by the platforms. Now, any scraping operation is prohibited by the terms of service, even those aimed at simply returning the data to the users' screens (Beraldo et al. 2021). These operational difficulties consequently make the collection of very large corpora quite complex.

For both phenomena there is also a final and crucial aspect to consider. As recently shown (Cinelli et al. 2021), when addressing the study of social media it must be considered that those who produce content, meaning comments by content, are only a small minority of users. The exact proportion is not known, but the rule of thumb applied is that about 99% of users are lurkers, i.e., they are passive users that do not create content or comments. This obviously cannot be overlooked when we go to do research on textual

data which also has social implications, such as those on echo chambers. Taking the traditionally adopted proportion as valid, it would mean that in our social media corpora we are analyzing a portion of online speech that is made up of 1% of the users of that platform.

This obviously does not mean that it is somehow less influential or impactful for lurker users. Indeed, as shown in the study of Cinelli et al. (2021) very few linguistic markers are sufficient to drift online discourse to hate speech or aggressive language in general. This is indicative of the pragmatic and semantic power brought about by the dialogistic dimension of the echo chambers. Further research on methodological aspects on similar topics should also attempt to use different methods for generating word embeddings, such as transformer-based models such as BERT (Devlin et al. 2018), ELMO (Peters et al. 2018) and GPT-3 (Brown et al. 2020). Both these algorithms might be used in Natural Language Generation tasks (Topal et al. 2021, Rothman 2021) to better understand the influence of filter bubbles and echo chambers on the semantic and pragmatic dimension.

REFERENCES

Abdi, H. (2007). "Singular value decomposition (SVD) and generalized singular value decomposition." *Encyclopedia of measurement and statistics*: 907-912.

Abisheva, A., Garcia D., and Schweitzer F. (2016). "When the filter bubble bursts: col-lective evaluation dynamics in online communities." In *Proceedings of the 8th ACM Conference on Web Science*, pp. 307-308.

Adami, E. (2015). What's in a click? A social semiotic framework for the multimodal analysis of website interactivity. *Visual Communication*, 14(2), 133–153. https://doi.org/10.1177/1470357214565583

Aditya, S., Baral, C., Vo, N.H., Lee, J., Ye, J., Naung, Z., Lumpkin, B., Hastings, J., Scherl, R.B., Sweet, D.M. and Inclezan, D. (2015, January). Recognizing Social Constructs from Textual Conversation. *In HLT-NAACL* (pp. 1293-1298).

Airoldi, M., Beraldo, D., & Gandini, A. (2016). Follow the algorithm: An exploratory investigation of music on YouTube. *Poetics*, 57, 1-13.

Albright, J. (2017). Welcome to the era of fake news. *Media and Communication*, 5(2), 87-89.

Apoorva, T., & Pradeep, N. (2017) Aspect Based Sentiment Analysis with Text Compression. *International Journal of Computer Sciences and Engineering* Volume-5, Issue-8 E-ISSN: 2347-2693

Arora S., Y. Li, Y. Liang, T Ma,., and A, Risteski. "Rand-walk: A latent variable model approach to word embeddings". (2015) arXiv preprint, arXiv:1502.03520

Arthurs, J., Drakopoulou, S., & Gandini, A. (2018). Researching youtube. *Convergence*, 24(1), 3-15.

Baker, P., Gabrielatos, C., Khosravinik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & society*, *19*(3), 273-306. Baker, Paul. 2006. Using Corpora in Discourse Analysis. London, Continuum.

Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014) Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18(2), 135-160

Baroni, M., Dinu, G., & Kruszewski, G. (2014, June). Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 238-247).

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. In *Third international AAAI conference on weblogs and social media*.

Bauman, Z. (2000) *Liquid modernity.* Cambridge, UK : Polity Press ; Malden, MA : Blackwell, 2000.

Bechmann, A. (2018). Filter Bubbles: From Academic Debate to Robust Empirical Analysis. International Communication Association Conference, May 2018, Prague.

Ben-David, A., & Fernández, A. M. (2016). Hate speech and covert discrimination on social media: Monitoring the Facebook pages of extreme-right political parties in Spain. *International Journal of Communication*, *10*, 27.

Benveniste, É. (1970): "L'appareil formel de l'énonciation," in *Problèmes de linguistique générale*, vol. 2, 79–88 (Paris: Gallimard).

Benveniste, É. (1971): "Subjectivity in Language," in *Problems in General Linguistics*, translated by Mary E. Meek, 223–30 (Coral Gables: Miami University Press).

Beraldo, D., Milan, S., de Vos, J., Agosti, C., Sotic, B. N., Vliegenthart, R., ... & Votta, F. Mar. 11, 2021). Political advertising exposed: tracking Facebook ads in the 2021 Dutch elections. Internet Policy Review (URL: https://policyreview.info/articles/news/politicaladvertising-exposed-tracking-facebook-ads-2021-dutch-elections/1543.

Bliuc, A. M., Smith, L. G., & Moynihan, T. (2020). "You wouldn't celebrate September 11": Testing online polarisation between opposing ideological camps on YouTube. *Group Processes & Intergroup Relations*, 23(6), 827-844.

Bishop, S. (2018): Anxiety, panic and self-optimization: Inequalities and the YouTube algorithm. *Convergence*, 24(1), 69–84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135-146.

Bolter, J. D., & Grusin, R. (1999). *Remediation: Understanding new media*. MIT Press.

Bondi, M. and Scott, M. eds., (2010) *Keyness in texts* (Vol. 41). John Benjamins Publishing.

Bonilla, Y., & Rosa, J. (2015) #Ferguson: Digital protest, hashtag ethnography, and the racial politics of social media in the United States. *American Ethnologist*, 42(1), 4-17.

Borra, Erik, et al. 2014 Contropedia-the analysis and visualization of controversies in Wikipedia articles. OpenSym.

Brbić, M., Rožić, E., & Žarko, I. P.: Recommendation of YouTube Videos. (2012). In *Proceedings of the 35th International Convention MIPRO* (pp. 1775-1779). IEEE.

Breeze, R. (2020). Exploring populist styles of political discourse in Twitter. *World Englishes*, *39*(4), 550-567.

Bruni, E., Tran, N. K., & Baroni, M. (2014). Multimodal distributional semantics. *Journal of artificial intelligence research*, *49*, 1-47.

Bruns, A. (2019)a: Filter bubble. *Internet Policy Review*, 8(4).

Bruns, A. (2019)b. "It's not the technology, stupid: How the 'Echo Chamber' and 'Filter Bubble' metaphors have failed us." *International Association for Media and Communication Research Conference 2019*.

Budanitsky, A., & Hirst, G. (2006). Evaluating wordnet-based measures of lexical semantic relatedness. *Computational linguistics*, *32*(1), 13-47.

Buder, J., Rabl, L., Feiks, M., Badermann, M., & Zurstiege, G. (2021). Does negatively toned language use on social media lead to attitude polarization?. *Computers in Human Behavior*, 116, 106663.

Burgers, C., Fa, M. J. T., & de Graaf, A. (2019). A tale of two swamps: Transformations of a metaphorical frame in online partisan media. *Journal of Pragmatics*, *141*, 57-66.

Calderón, C. A., Blanco-Herrero, D., & Apolo, M. B. V. (2020). Rejection and hate speech in Twitter: Content analysis of Tweets about migrants and refugees in Spanish. *Revista Española de Investigaciones Sociológicas (REIS)*, *172*(172), 21-56..

Calderón, F. H., Cheng, L. K., Lin, M. J., Huang, Y. H., & Chen, Y. S. (2019, August). Content-based echo chamber detection on social media platforms. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 597-600).

Cavasso, L., & Taboada, M. (2021). A corpus analysis of online news comments using the Appraisal framework. *Journal of Corpora and Discourse Studies*, 4, 1-38.

Chamberlain, B. P., Rossi, E., Shiebler, D., Sedhain, S., & Bronstein, M. M. (2020, September). Tuning word2vec for large scale recommendation systems. In *Fourteenth ACM Conference on Recommender Systems* (pp. 732-737).

Chartier, J. F., Pulizzotto, D., Chartrand, L., & Meunier, J. G. (2019). A data-driven computational semiotics: The semantic vector space of Magritte's artworks. *Semiotica*, *2019*(230), 19-69.

Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR public health and surveillance*, *6*(2), e19273.

Cinelli, M., Pelicon, A., Mozetič, I., Quattrociocchi, W., Novak, P. K., & Zollo, F. (2021). Dynamics of online hate and misinformation. *Scientific reports*, 11(1), 1-12.

Cinelli, M., Brugnoli, E., Schmidt, A. L., Zollo, F., Quattrociocchi, W., & Scala, A. (2020). Selective exposure shapes the Facebook news diet. *PloS one*, *15*(3), e0229129.

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, *118*(9), e2023301118.

Compagno, D., Mercier, A., Mesangeau, J. & Chelghoum, K., (2017) La reconfiguration de l'offre d'information médiatique opérée par les réseaux socionumériques. *Réseaux*, 205:91-116.

Compagno, D. (Ed.). (2018a). *Quantitative semiotic analysis*. Berlin: Springer.

Compagno, D. (2018b). The Cost of Truth. Motivations of a Pragmatist Trust-Conditional Approach to News Evaluation. *Versus*, 47(2), 275-290.

Conover, Michael, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. "Political polarization on twitter." In Proceedings of the International AAAI Conference on Web and Social Media, vol. 5, no. 1.

Conroy, N. K., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. Proceedings of the association for information science and technology, 52(1), 1-4.

Cosenza, G. (2014) *Introduzione alla semiotica dei nuovi media*. GLF editori Laterza.

Cosenza, G., Colombari, J., & Gasparri, E. (2016). Come la pubblicità italiana rappresenta le donne e gli uomini. Verso una metodologia di analisi semiotica degli stereotipi. *Versus*, *45*(2), 323-362.

Cosenza, G., & Sanna, L. (2021). The Origins of the Alleged Correlation between Vaccines and Autism. A Semiotic Approach. *Social Epistemology*, 1-14.

Cota, W., Ferreira, S. C., Pastor-Satorras, R., & Starnini, M. (2019). Quantifying echo chamber effects in information spreading over political communication networks. *EPJ Data Science*, *8*(1), 1-13.

Cotfas, L. A., Delcea, C., Gherai, R., & Roxin, I. (2021). Unmasking People's Opinions behind Mask-Wearing during COVID-19 Pandemic—A Twitter Stance Analysis. *Symmetry*, 13(11), 1995.

Covington, P., Adams, J., & Sargin, E. (2016, September). Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems* (pp. 191-198).

Culpeper, J. (2011). *Impoliteness: Using language to cause offence* (Vol. 28). Cambridge University Press.

Culpeper, J., Haugh, M., & Kádár, D. Z. (Eds.). (2017). *The Palgrave handbook of linguistic (im) politeness* (pp. 11-38). London: Palgrave Macmillan.

Culpeper, J. (2021). Impoliteness and hate speech: Compare and contrast. *Journal of pragmatics*, *179*, 4-11.

Davies, M. (2016). Corpus of News on the Web (NOW). Available online at https://www.english-corpora.org/now/

Davies, M. (2019). The Coronavirus Corpus. Available online at https://www.english-corpora.org/corona/

Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). Echo chambers: Emotional contagion and group polarization on facebook. *Scientific reports*, *6*(1), 1-12.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Science*s, 113(3), 554-559.

Demszky, D., Garg, N., Voigt, R., Zou, J., Gentzkow, M., Shapiro, J., & Jurafsky, D. (2019). Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. arXiv preprint arXiv:1904.01596.

Derboven, J., De Roeck, D., & Verstraete, M. (2012) Semiotic analysis of multi-touch interface design: The MuTable case study. *International Journal of Human-Computer Studies*, 70(10), 714-728.

Di Domenico, G., Sit, J., Ishizaka, A., & Nunan, D. (2021). Fake news, social media and marketing: A systematic review. *Journal of Business Research*, 124, 329-341.

Di Marco, N., Cinelli, M., & Quattrociocchi, W. (2021). Infodemics on Youtube: Reliability of Content and Echo Chambers on COVID-19. arXiv preprint arXiv:2106.08684.

Dubois, Elizabeth, and Grant Blank. 2018. "The echo chamber is overstated: the moderating effect of political interest and diverse media." *Information, Communication & Society* 21, no. 5:729-745.

Duseja, N., & Jhamtani, H. (2019, June). A sociolinguistic study of online echo chambers on twitter. In *Proceedings of the third workshop on natural language processing and computational social science* (pp. 78-83).

Eco, U. (1964). Apocalittici e integrati: la cultura italiana e le comunicazioni di massa. *Milano: Bompiani*, *23*.

Eco, U. (1968). *La struttura assente [The absent structure]*. Milan, Bompiani.

Eco, U. (1979) *The Role of the Reader*, Bloomington : Indiana UP.

Eco, U. (1990). *I limiti dell'interpretazione* (reissue 2016). *Milano, La Nave di Teseo*

Eco, U. (2000). *Kant and the platypus: Essays on language and cognition*. HMH.

Eco, U. (2014). *From the Tree to the Labyrinth.* Harvard University Press.

Edwards, A. (2013). (How) do participants in online discussion forums create 'echo chambers'?: The inclusion and exclusion of dissenting voices in an online forum about climate change. *Journal of Argumentation in Context*, *2*(1), 127-150.

Ellison, N. B., Steinfield, C., & Lampe, C. (2011). Connection strategies: Social capital implications of Facebook-enabled communication practices*. New Media & Society*, 13(6), 873–892. doi:doi:10.1177/1461444810385389

ElSherief, M., Nilizadeh, S., Nguyen, D., Vigna, G., & Belding, E. (2018, June). Peer to peer hate: Hate speech instigators and their targets. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 12, No. 1).

Fairclough, N. (2010). *Critical Discourse Analysis: The critical study of language*. London: Routledge.

Faruqui, M. et al. (2016) Problems with evaluation of word embeddings using word similarity tasks." arXiv preprint arXiv:1605.02276

Fernandez, M., Harith A:, Online Misinformation: Challenges and Future Directions. In Companion Proceedings of the Web Conference 2018, WWW '18, *International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva*, CHE, 595–602 (2018).

Ferraro, G., Lorusso, A.M. (Ed.) (2016) *Nuove forme d'interazione: dal web al mobile*, Libellula Edizioni

Fillmore, C. J. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, (1976) 280(1):20-32.

Finkelstein, E. Gabrilovich, Y. Matias, E. Rivlin, Z. Solan,, Wolfman, G., & Ruppin, E. "Placing search in context: The concept revisited". In: Proceedings of the 10th international conference on World Wide Web (pp. 406-414) (2001).

Finocchi G. (Ed.) (2015) *Carte Semiotiche. Strategia dell'ironia del web*. Volo Publisher, Lucca.

Floridi, L. (2009). Web 2.0 vs. the semantic web: A philosophical assessment. *Episteme*, 6(1), 25-37.

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. OUP Oxford.

Floridi, L. (2015) The onlife manifesto: Being human in a hyperconnected era. Springer, *Nature,*.

Fuchs, C. (2010) "Labor in Informational Capitalism and on the Internet." *The Information Society* 26, no. 3: 179-196.

Fuchs, C. (2013) *Social media and capitalism.* Nordicom.

Fuchs, C. (2018) *Digital demagogue: Authoritarian capitalism in the age of Trump and Twitter*. Pluto Press, 2018.

Fuoli, M. (2012). Assessing social responsibility: A quantitative analysis of Appraisal in BP's and IKEA's social reports. *Discourse & Communication*, *6*(1), 55-81.

Fuoli, M. (2018). A stepwise method for annotating APPRAISAL. *Functions of Language*, *25*(2), 229-258.

Gabrielatos, C., & Marchi, A. (2011, November). Keyness: Matching metrics to definitions. In *Theoretical-methodological challenges in corpus approaches to discourse studies and some ways of addressing them*.

Gabrielatos, C., & Marchi, A. (2012). Keyness: Appropriate metrics and practical issues. In *Corpus-assisted Discourse Studies International Conference*.

Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In *Corpus approaches to discourse* (pp. 225-258). Routledge..

Gallacher, J. D., & Heerdink, M. W. (2019). Measuring the effect of Russian Internet research agency information operations in online conversations. *Defence Strategic Communications*, *6*. Goode, L. ". (2009). Social news, citizen journalism and democracy. " *New media & society*, 11(8), 1287-1305.

Gray, J., Bounegru, L., & Venturini, T. (2020). 'Fake news' as infrastructural uncanny. *New media & society*, *22*(2), 317-341.

Greimas, A. J., & Courtés, J. (1970*). Sémiotique: dictionnaire raisonné de la théorie du langage.* Paris: Hachette.

Greimas, A. J., Collins, F., & Perron, P. (1989) The veridiction contract. *New Literary History*, 20(3), 651-660.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41-58). Brill.

Guo, F., Blundell, C., Wallach, H., & Heller, K. (2015, February). The bayesian echo chamber: Modeling social influence via linguistic accommodation. In *Artificial Intelligence and Statistics* (pp. 315-323). PMLR.

Halliday, M.A.K. "Categories of the Theory of Grammar." (1961) *WORD*, 17(2), pp. 241–292

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Hodder Education.

Hargreaves, E., Agosti, C., Menasché, D., Neglia, G., Reiffers-Masson, A., & Altman, E. (2018, August). Biases in the facebook news feed: a case study on the italian elections. In 2018 *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (ASONAM) (pp. 806-812). IEEE.

Hargreaves, E., Agosti C., Menasché, D., Neglia, G., Reiffers-Masson, A., Altman, E. (2018b) Supplementary Material for Biases In The Facebook News Feed: A Case Study On The Italian Elections. International Symposium on Foundations of Open Source Intelligence and Security Informatics, August 2018, Barcelona.

Harlow, S. (2012). Social media and social movements: Facebook and an online Guatemalan justice movement that moved offline*. New Media & Society*., 14(2), 225-243. doi:doi:10.1177/1461444811410408

Harris, Z.S. (1954) Distributional structure. *WORD*, 10:2-3, 146-162.

Hasinoff, A. A. (2012). Sexting as media production: Rethinking social media and sexuality. *New Media & Society*, 15(4), 449-465. doi: doi:10.1177/1461444812459171

Hellrich J. (2019) *Word embeddings: reliability & semantic change* (Vol. 347). IOS Press

Helmond, A., Nieborg, D. B., & van der Vlist, F. N. (2017, July). The political economy of social data: a historical analysis of platform-industry partnerships. In *Proceedings of the 8th international conference on social media & society* (pp. 1-5).

Hill, F., Reichart, R., & Korhonen, A. (2015). Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics*, *41*(4), 665-695.

Hintz, A., & Milan, S. (2009). At the margins of Internet governance: grassroots tech groups and communication policy. *International Journal of Media & Cultural Politics*, *5*(1-2), 23-38.

Humphreys, L. (2010). Mobile social networks and urban public space. *New Media & Society.*, 12(5), 763-778. doi:https://doi.org/10.1177%2F1461444809349578

Hunston, S. (2022). *Corpora in applied linguistics*. Cambridge University Press.

Hunston, S. (2007). Semantic prosody revisited. *International journal of corpus linguistics*, *12*(2), 249-268.

Hunston, S. (2010). *Corpus approaches to evaluation: Phraseology and evaluative language*. Routledge.

Jacobson, S., Myung, E., & Johnson, S. L. (2016). Open media or echo chamber: The use of links in audience discussions on the Facebook pages of partisan news organizations. *Information, Communication & Society*, *19*(7), 875-891.

Jurafsky, D., & Martin, J. H. (2021 - draft). *Speech and language processing: An introduction to natural language processing,* 3d edition, https://web.stanford.edu/~jurafsky/slp3/ .

Kienpointner, M. (2018). Impoliteness online: Hate speech in online interactions. *Internet Pragmatics*, *1*(2), 329-351.

Klapper, J. T. (1960*). The effects of mass communication*. New York Free Press

Kress, G. (2009). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kucharski, A. (2016). Study epidemiology of fake news. *Nature*, *540*(7634), 525-525.

Lai, S., Liu, K., Xu, L., Zhao, J. (2016) How to generate a good word embedding. *IEEE Intelligent Systems*, 31.6:5-14.

Langton, R. (2018). Blocking as counter-speech. *New work on speech acts*, *144*, 156.

Lapalut, S. (1995). *Text clustering to support knowledge acquisition from documents* (Doctoral dissertation, INRIA).

Lau, J. H., & Baldwin, T. (2016). An empirical evaluation of doc2vec with practical insights into document embedding generation. arXiv preprint arXiv:1607.05368.

Le, Q., & Mikolov, T. (2014) Distributed representations of sentences and documents. *In Proceedings of the 31st International Conference on Machine Learning (ICML-14)* (pp. 1188-1196).

Lebart, L., Salem, A., & Berry, L. (1997). *Exploring textual data* (Vol. 4). Springer Science & Business Media.

Ledwich, M., & Zaitsev, A. (2019). Algorithmic extremism: Examining YouTube's rabbit hole of radicalization. *arXiv preprint arXiv:1912.11211*.

Lenci, A. (2008) Distributional semantics in linguistic and cognitive research*. Italian journal of linguistics*, 20(1):1-31.

Leone, M. (2018). Symmetries in the semiosphere: a typology. URL: https://iris.unito.it/retrieve/handle/2318/1667169/410720/Massimo%20Leone%202018%20-%20Symmetries%20in%20the%20Semiosphere.pdf

Li, B., Drozd, A., Guo, Y., Liu, T., Matsuoka, S., & Du, X. (2019) Scaling Word2Vec on Big Corpus. *Data Science and Engineering*, 1-19.

Lorusso, A. M. (2018). *Postverità: Fra reality tv, social media e storytelling*. Gius. Laterza & Figli Spa.

Lorusso, A.M., and Violi, P. (2004*) Semiotica del testo giornalistico*. Gius. Laterza & Figli Spa,.

Lotman, Y. M. (2005). *On the semiosphere. Σημειωτκή-Sign Systems Studies*, 33(1), 205-229. Original Title O semiosfere, in Trudy po znakowm sistemamn n. 17, Tartu, 1984

Madison, N. (2016) Digital Bisexuality: The Semiotics Of Online Sexual Identity. *AoIR Selected Papers of Internet Research*, 5.

Maggi, R. (2014) Toward a Semiotics of Digital Places in Resmini, A. (Ed.) *Reframing information architecture*. Springer.

Manetti, G. (2008). *L'enunciazione: dalla svolta comunicativa ai nuovi media*. Mondadori università..

Manning, C., & Schutze, H. (1999). *Foundations of statistical natural language processing*. MIT press..

Marchal, N., Au, H., & Howard, P. N. (2020). Coronavirus news and information on YouTube. *Health*, *1*(1), 0-3.

Marrone, G. (2017). Social media e comunione fàtica: verso una tipologia delle pratiche in rete. *Versus*, *46*(2), 249-272.

Martin, J. R. (2016). Meaning matters: A short history of systemic functional linguistics. *Word*, *62*(1), 35-58.

Martin, J. R., & White, P. R. (2005). *The language of evaluation* (Vol. 2). London: Palgrave Macmillan.

Marwick, A., & Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media and Society*, 13(1), 114-133. doi:https://doi.org/10.1177%2F1461444810365313

Mautner, G. (2001). "Checks and balances: How corpus linguistics can contribute to CDA." In R. Wodak and M. Meyer (eds.), *Methods of Critical Discourse Analysis*. London: Sage, pp. 122-143.

Melchior, C., & Oliveira, M. (2022). Health-related fake news on social media platforms: A systematic literature review. new media & society, 24(6), 1500-1522.

Mewari, R., Singh, A., & Srivastava, A. (2015). Opinion mining techniques on social media data. *International Journal of Computer Applications*, *118*(6).

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013b) Efficient estimation of word representations in vector space. arXiv preprint, arXiv:1301.3781.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013a) Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, Dicembre 2013, Leake Tahoe, 3111-3119.

Mirsarraf, M., Shairi, H., & Ahmadpanah, A. (2017) Social semiotic aspects of Instagram social network. In *Innovations in Intelligent Systems and Applications (INISTA)*, IEEE International Conference (pp. 460-465).

Mitchell T. (1997) *Machine Learning,* Mcgraw-hill science. *Engineering/Math*, *1*, 27.

Nah, S., & Saxton, G. D. (2013). Modeling the adoption and use of social media by nonprofit organizations. *New Media & Society*., 15(2), 294-313. doi: doi:10.1177/1461444812452411

Naili, M., Habacha Chaibi, A. and Hajjami Ben Ghezala, H. (2017) Comparative study of word embedding methods in topic segmentation. *Procedia computer science*, 112 : 340-349.

Nartey, M., & Mwinlaaru, I. N. (2019). Towards a decade of synergising corpus linguistics and critical discourse analysis: a meta-analysis. *Corpora*, *14*(2), 203-235.

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, *17*(2), 141-161.

O'Donnell, M. (2014). Exploring identity through Appraisal Analysis: A corpus annotation methodology. *Linguistics & the Human Sciences*, 9(1).

Oakes, Michael P. (1998). *Statistics for Corpus Linguistics.* Edinburgh: Edinburgh University Press.

Oliveira, A. L. A. M., & Carneiro, M. M. (2020). A pragmatic view of hashtags: the case of impoliteness and offensive verbal behavior in the Brazilian Twitter. *Acta Scientiarum. Language and Culture*, *42*(1).

Oz, M., Zheng, P., & Chen, G. M. (2018). Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes. *New media & society*, *20*(9), 3400-3419.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. penguin UK.

Partington, A. (2003). *The linguistics of political argument: The spin-doctor and the wolf-pack at the White House*. Routledge.

Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).

Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2017). Automatic detection of fake news. arXiv preprint arXiv:1708.07104.

Pascual-Ferrá, P., Alperstein, N., Barnett, D. J., & Rimal, R. N. (2021). Toxicity and verbal aggression on social media: Polarized discourse on wearing face masks during the COVID-19 pandemic*. Big Data & Society*, 8(1), 20539517211023533.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (1802). Deep contextualized word representations. arXiv 2018. *arXiv preprint arXiv:1802.05365*, *12*.

Peverini, P. (2014) Reputazione e influenza nei social media. Una prospettiva sociosemiotica. In Pezzini, I. Spaziante L. (Ed.) *Corpi Mediali, Semiotica e Contemporaneità,* 65-83*.*

Peverini, P. (2017). Daily life in the instagram age. a socio-semiotic perspective. *Versus*, *46*(2), 285-302.

Pierce, C.S., (1867) On the Natural Classification of Arguments. In *P. E. Project, Writings of Charles S. Peirce: A Chronological Edition* - Volume 2.

Pierce, C.S., (1877) The Fixation of Beliefs. *Popular Science*.

Pierce, C.S., (1934) Collected Papers (CP) of Charles Sanders Peirce, vols. 1–6.

Pilehvar, M., & Collier, N. (2017, April). Inducing embeddings for rare and unseen words by leveraging lexical resources*. Association for Computational Linguistics*.

Pojanapunya, P., & Todd, R. W. (2018). Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*, 14(1), 133-167.

Quinlan, J. R. (1996). Learning decision tree classifiers. *ACM Computing Surveys (CSUR)*, *28*(1), 71-72.

Ratinaud, P. (2009). IRaMuTeQ: Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. Téléchargeable à l'adresse: https://www. iramuteq.org

Ratinaud, P., & Marchand, P. (2016). Quelques méthodes pour l'étude des relations entre classifications lexicales de corpus hétérogènes: application aux débats à l'assemblée nationale et aux sites Web de partis politiques. *Statistical Analysis of Textual Data*, 193-202.

Read, J., & Carroll, J. (2012). Annotating expressions of appraisal in English. *Language resources and evaluation*, 46(3), 421-447.

Rehurek, Radim, and Petr Sojka.(2010) "Software framework for topic modelling with large corpora." In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*.

Reinert, A. (1983). Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte. *Cahiers de l'Analyse des Données*, *8*(2), 187-198.

Reviglio, U. (2017, November). Serendipity by design? How to turn from diversity exposure to diversity experience to face filter bubbles in social media. In *International Conference on Internet Science* (pp. 281-300). Springer, Cham.

Reviglio, U., & Agosti, C. (2020). Thinking outside the Black-box: The case for "algorithmic sovereignty" in social media. *Social Media+ Society*, 6(2), 2056305120915613.

Rieder, B. (2012). The refraction chamber: Twitter as sphere and network. *First Monday*.

Rieder, B., Abdulla, R., Poell, T., Woltering, R., & Zack, L. (2015). Data critique and analytical opportunities for very large Facebook Pages: Lessons learned from exploring "We are all Khaled Said". *Big Data & Society*, *2*(2), 2053951715614980.

Rieder, B., Matamoros-Fernández, A., & Coromina, Ò. (2018). From ranking algorithms to 'ranking cultures' Investigating the modulation of visibility in YouTube search results. *Convergence*, *24*(1), 50-68.

Riedl, M., & Biemann, C. (2017). There's no 'count or predict'but task-based selection for distributional models. In *IWCS 2017—12th International Conference on Computational Semantics—Short papers*.

Ringnér, M. (2008). What is principal component analysis?. *Nature biotechnology*, *26*(3), 303-304..

Robertson, R. E., Jiang, S., Joseph, K., Friedland, L., Lazer, D., & Wilson, C. (2018). Auditing partisan audience bias within google search. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1-22.

Rogers, R. (2009). *The end of the virtual: Digital methods* (Vol. 339). Amsterdam University Press.

Rogers, R. (2010). Internet research: The question of method—A keynote address from the YouTube and the 2008 election cycle in the United States Conference. *Journal of Information Technology & Politics*, *7*(2-3), 241-260..

Rogers, R. (2013) *Digital methods*. MIT press

Rogers, R. (2017) Foundations of digital methods: Query design. In *The Datafied Society Studying Culture through Data 75-94* - Amsterdam University Press.

Rogers, R. (2018a). Doing Web history with the Internet Archive: screencast documentaries. In *Internet Histories* (pp. 160-172). Routledge.

Rogers, R. (2018b). Digital traces in context| Otherwise engaged: Social media from vanity metrics to critical analytics. *International Journal of Communication*, *12*, 23.

Rüdiger, S., & Dayter, D. (Eds.). (2020). *Corpus approaches to social media* (Vol. 98). John Benjamins Publishing Company.

Sahlgren, M., & Lenci, A. (2016). The effects of data size and frequency range on distributional semantic models. *arXiv preprint arXiv:1609.08293*.

Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Disability Studies*, *20*, 33-53.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry*, *22*, 4349-4357.

Scott, K. (2015). The pragmatics of hashtags: Inference and conversational style on Twitter. *Journal of Pragmatics*, *81*, 8-20..

Scott, M. (1997). PC analysis of key words—and key words. *System*, *25*(2), 233-245.

Scott, M. (2020). WordSmith Tools (Version 8.0). Stroud: Lexical Analysis Software.

Sergienya, I., & Schütze, H. (2015, September). Learning better embeddings for rare words using distributional representations. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 280-285).

Shu, K., & Liu, H. (2019). Detecting fake news on social media. *Synthesis lectures on data mining and knowledge discovery*, *11*(3), 1-129..

Silva, L., Mondal, M., Correa, D., Benevenuto, F., & Weber, I. (2016, March). Analyzing the targets of hate in online social media. In *Tenth international AAAI conference on web and social media*.

Sinclair, J. (1996). The search for units of meaning, *Textus*, 9, 1. pp. 75–106.

Six, J. M., & Tollis, I. G.: A framework and algorithms for circular drawings of graphs. *Journal of Discrete Algorithms*, 4(1), 25-50. (2006).

Song, M. Y. J., & Gruzd, A. (2017, July). Examining sentiments and popularity of pro-and anti-vaccination videos on YouTube. In *Proceedings of the 8th international conference on social media & society* (pp. 1-8).

Soral, W., Bilewicz, M., & Winiewski, M. (2018). Exposure to hate speech increases prejudice through desensitization. *Aggressive behavior*, *44*(2), 136-146.

Stelter, B. (2020). *Hoax: Donald Trump, Fox News, and the dangerous distortion of truth*. Simon and Schuster..

Strandberg, K. (2013). A social media revolution or just a case of history repeating itself? The use of social media in the 2011 Finnish parliamentary elections*. New Media & Society*., 15(8), 1329-1347. doi: doi:10.1177/1461444812470612

Stubbs, M. (1996). *Text and corpus analysis: Computer-assisted studies of language and culture* (p. 158). Oxford: Blackwell.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics* (pp. 1-267). Oxford: Blackwell publishers.

Sunstein, C. (2007). *Republic.com. 2.0* Princeton, NJ: Princeton University Press.

Tagg, C., Seargent, P., & Brown, A. A. (2017). Taking offence on social media. London: Palgrave Macmillan, 10, 978-3.

Taylor, C., & Marchi, A. (2018). Corpus approaches to discourse. *A critical review*.

Teneketzi, K. (2022). Impoliteness across social media platforms: A comparative study of conflict on YouTube and Reddit. *Journal of Language Aggression and Conflict*, *10*(1), 38-63.

Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS one*, *13*(9), e0203958.

Valenzuela, S. (2013). Unpacking the use of social media for protest behavior: The roles of information, opinion expression, and activism. *American behavioral scientist*, *57*(7), 920-942.

Van Dijk, T. A. (1998). *Ideology: A multidisciplinary approach*. London: Sage.

Van Eck, C. W., Mulder, B. C., & van der Linden, S. (2021). Echo chamber effects in the climate change blogosphere. *Environmental Communication*, *15*(2), 145-152..

Van Leeuwen, T. (2008). *Discourse and practice: New tools for critical discourse analysis*. Oxford university press.

Vargo, C. J., Guo, L., & Amazeen, M. A. (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. New media & society, 20(5), 2028-2049.

Violi, P. (1997) *Significato ed esperienza*. Studi Bompiani-

Violi, P. (2000). Prototypicality, typicality and context. *Meaning and Cognition: A multidisciplinary approach*, 103-122.

Wang, B., Wang, A., Chen, F., Wang, Y., & Kuo, C. C. J. (2019). Evaluating word embedding models: methods and experimental results. *APSIPA transactions on signal and information processing*, *8*.

Watson, D. S., & Floridi, L. (2021). The explanation game: a formal framework for interpretable machine learning. In *Ethics, Governance, and Policies in Artificial Intelligence* (pp. 185-219). Springer, Cham.

Wilson, A. (2013). Embracing Bayes factors for key item analysis in corpus linguistics. In Bieswanger, M. and Koll-Stobbe, A. (eds), *New approaches to the study of linguistic variability*. Language Competence and Language Awareness in Europe, vol. 4, Peter Lang, Frankfurt, (2013) pp. 3-11

Wittgenstein, L., Anscombe, G. E. M., von Wright, G. H., Paul, D., & Anscombe, G. E. M. (1969). *On certainty* (Vol. 174). Oxford: Blackwell.

Wodak, R. (2002). Aspects of critical discourse analysis. *Zeitschrift für angewandte Linguistik*, *36*(10), 5-31.

Wu, L. Y., Fisch, A., Chopra, S., Adams, K., Bordes, A., & Weston, J. (2018). Starspace: Embed all the things!. In *Thirty-Second AAAI Conference on Artificial Intelligence.*

Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *New Media & Society*, 13(5). pp. 788-806. DOI: 10.1177/1461444810385097

Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web* (Vol. 6). A&C Black.

Zappavigna, M., & Martin, J. R. (2018). # Communing affiliation: Social tagging as a resource for aligning around values in social media. *Discourse, context & media*, *22*, 4-12.

Zappavigna, M. (2015). Searchable talk: The linguistic functions of hashtags. *Social Semiotics*, *25*(3), 274-291.

Zhe, Z., Lichan, H., Li, W., Jilin, et al.: Recommending what video to watch next: a multitask ranking system. In *Proceedings of the 13th ACM Conference on Recommender Systems (RecSys 19)*. Association for Computing Machinery, New York, NY, USA, 43{51. https://doi.org/10.1145/3298689.3346997 (2019)

Zikopoulos, P., Eaton, C., deRoos, D., Detusch, T., & Lapis, G. (2012). Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data (IBM.). New York: McGraw.

Zimmer, F., Scheibe, K., Stock, M., & Stock, W. G. (2019). Fake news in social media: Bad algorithms or biased users?. Journal of Information Science Theory and Practice, 7(2), 40-53.

Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekhtman, L., ... & Quattrociocchi, W. (2017). Debunking in a world of tribes. *PloS one*, *12*(7), e0181821.