Communication Library
Vol. 7

Branco Di Fátima
(Ed.)

HATE ON SOCIAL MEDIA SPEECH

LABCOM
COMMUNICATION
& ARTS

OdeCOM

Communication Library
Vol. 7

Branco Di Fátima
(Ed.)

HATE
ON SOCIAL MEDIA
SPEECH

LABCOM
COMMUNICATION
& ARTS

OdeCOM

Technical Specifications

Title
Hate Speech on Social Media:
A Global Approach

Editor
Branco Di Fátima

LabCom Books & EdiPUCE
www.labcom.ubi.pt
https://edipuce.edu.ec/

Collection
Communication Books (LabCom Books)
Communication Library, Vol. 7 (EdiPUCE)

Direction
Gisela Gonçalves (LabCom Books)
Marco López-Paredes (OdeCom-PUCE)

Graphic Design
Daniel Baldaia

ISBN
978-989-654-914-5 (print)
978-989-654-916-9 (pdf)
978-989-654-915-2 (epub)
978-9978-77-664-3 (EdiPUCE)

Legal Deposit
513655/23

DOI
10.25768/654-916-9

Print
Print-on-demand

University of Beira Interior
Rua Marquês D'Ávila e Bolama.
6201-001 Covilhã. Portugal
www.ubi.pt | labcom@ubi.pt

PUCE Publications Centre
Av. 12 de Octubre y Robles
17-01-2184, Quito, Ecuador
https://edipuce.edu.ec/ | publicaciones@puce.edu.ec

Covilhã, 2023
Quito, 2023

Published by LabCom Books in co-edition with OdeCom and EdiPUCE.

Contents

Preface 11

Abstracts 17

Chapter 1 — Aggravated anti-Asian hate since COVID-19 and the #StopAsianHate movement: Connection, disjointness, and challenges 35
Lizhou Fan, Huizi Yu & Anne J. Gilliland

Chapter 2 — Is it fine? Internet memes and hate speech on Telegram in relation to Russia’s war in Ukraine 75
Mykola Makhortykh & Juan-Manuel González-Aguilar

Chapter 3 — Syrian refugees in the shade of the ‘Anti-Syrians’ discourse: Exploring discriminatory discursive strategies on Twitter 95
Özlem Alikılıç, Ebru Gökaliler & İnanç Alikılıç

Chapter 4 — Disseminating and resisting online hate speech in Turkey 121
Mine Gencel Bek

Chapter 5 — Hate speech on Twitter: The LGBTIQ+ community in Spain 143
Patricia de-Casas-Moreno, Macarena Parejo-Cuéllar & Arantxa Vizcaíno-Verdú

Chapter 6 — Circulation systems, emotions, and presenteeism: three views on hate speech based on attacks on journalists in Brazil 159
Edson Capoano, Vítor de Sousa & Vinicius Prates

Chapter 7 — Clipping: Hate speech in social media against female sports journalists in Greece 185
Lida Tsene

Chapter 8 — Mapping social media hate speech regulations in Southern Africa: A regional comparative analysis 203
Allen Munoriyarwa

Chapter 9 — Ethiopian socio-political contexts for hate speech 227
Muluken Asegidew Chekol

Chapter 10 — Social media narratives and reflections on hate speech in Nigeria 255
Aondover Eric Msughter

Chapter 11 — Hate speech among security forces in Portugal 277
Tiago Lapa & Branco Di Fátima

Authors 295

Preface

A GLOBAL APPROACH TO HATE SPEECH ON SOCIAL MEDIA

Branco Di Fátima
LabCom – University of Beira Interior

Hate speech manifests itself in different social contexts, such as political debates, artistic expression, professional sports, or work environments. However, the rapid development of digital technologies, and especially of social media platforms, has created additional challenges to understanding this extreme act. Although this field of study is already over two decades old (Duffy, 2003), many questions still need to be answered.

There is no universally accepted definition of hate speech. Its characterization is a point of intellectual dispute among different worldviews, many outside the Western universe and little known. In general, hate speech is an attack on a person or group, usually targeting members of a social minority. Thus, it can be classified as sexist, racist, xenophobic, ageist, fatphobic, or homophobic, among others. Haters direct their attacks, for example, against women, Black people, immigrants, seniors, disabled people, and the LGBTQ+ community. The United Nations (n.d.) emphasizes that hate speech refers to offenses based on inherent traits, such as race, nationality, or gender.

Hate speech can also originate from and amplify religious intolerance (against Catholics or Muslims, for example), inflame tribal conflicts, or fuel prejudice against individuals within the same country (south *vs* north, capital *vs* countryside). Given the diversity of approaches, understanding the phenomenon involves the context in which it emerges. As a communicative act, the roots of hate speech are the codes and values of a particular culture (Matamoros-Fernández & Farkas, 2021).

These are only some of the challenges. Empirical studies based on Big Data show that detecting hate speech on social media is difficult (Miranda et al., 2022). Indeed, haters mobilize numerous subterfuges to obscure their intentions. For example, haters can use irony, humor, and satire to disguise a violent narrative (Schwarzenegger & Wagner, 2018). Moreover, in order to dehumanize opponents, a systematic strategy is to compare victims to repulsive animals, such as snakes, wasps, spiders, or cockroaches (Ndahinda & Mugabe, 2022).

Hate speech on social media can be verbal (posts, comments, articles, etc.) and non-verbal (emojis, stickers, photos, etc.). These multimedia attacks create and reinforce stereotypes based on toxic language. They can range from mere insults to calls for physical extermination and genocide. Sometimes they stem from emotional outbursts and go viral online, migrating from one platform to another (López-Paredes & Di Fátima, 2023). Thus, they affect both the victims and society itself by undermining democratic spaces for deliberation.

Regulating hate speech is not a simple issue. Sometimes it is driven by nationalist groups or far-right parties, going hand in hand with disinformation and conspiracy theories. Occasionally, haters use freedom of expression to justify their behaviors (Amores et al., 2021). In the name of combating hate, authoritarian states also have passed vague laws that censor the public sphere (Garbe, Selvik & Lemaire, 2023). So whose responsibility is it to regulate hate speech: the governments', social media platforms', or society's? It is a game of chess, and every move counts.

Some authors have pointed to the power of social media in shaping hate speech (Müller & Schwarz, 2021). The platforms would be open and favor violent narratives (Brown, 2018). However, how one can regulate hate speech without interfering with freedom of expression remains an open question.

First, it is urgent to map hate online and the results of its platformization, which has fostered old and new forms of abuse (Gagliardone, 2019).

Hate speech is more complex and diverse on social media. It spreads at high speed and can impact behaviors beyond the borders where it originates. Hate is ubiquitous, interactive, and multimedia. It is available 24/7, reaching a much larger audience. On social media, haters can be anonymous and find support from individuals with the same aggressive mindset. This is just a brief characterization and certainly presents many theoretical gaps that need improvement.

This book explores the nature of hate speech on social media. Readers will find chapters written by 21 authors from 18 universities or research centers. It includes researchers from 11 countries, prioritizing a diversity of approaches from the Global North and Global South – Brazil, Cyprus, Ethiopia, Germany, Nigeria, Portugal, South Africa, Spain, Switzerland, Turkey, and the USA. The analyses herein involve the realities in an even larger number of countries, given the transnational approach of some of these studies.

One can find a preview of the chapters at the beginning of the book, with abstracts organized in a separate section. It is evident that the authors study the impact of recent events on hate speech – the Covid-19 pandemic, Russia-Ukraine war, the refugee crisis – and recurrent attacks on minority groups such as women, immigrants, or the LGBTQ+ community. The authors employ classic and digital research methods, using quantitative and qualitative data gathered from platforms like Telegram, Facebook, Instagram, Twitter, and YouTube. As a result, readers will encounter taxonomic proposals, new methodological approaches, theoretical frameworks, and mapping of behavioral patterns.

While hate speech is rooted in national identity and shaped by context, it is a global phenomenon that requires transnational study to uncover its unique characteristics. For example, who are the primary targets? What forms do the messages take? How do virtual armies replicate violent narratives? What emotional drivers underlie hate speech on social media? And lastly, how can legal dilemmas surrounding regulation be resolved?

The construction of these answers is open and subject to constant dispute. Theoretical and methodological normalization needs to be improved. Currently, hate speech in digital environments challenges academia and society. This book aims to dispel some of these uncertainties.

References

Amores, J. J., Blanco-Herrero, D., Sánchez-Holgado, P. & Frías-Vázquez, M. (2021). Detectando el odio ideológico en Twitter: Desarrollo y evaluación de un detector de discurso de odio por ideología política en tuits en español. *Cuadernos.info*, 49(2021), 98-124. https://doi.org/10.7764/cdi.49.27817

Brown, A. (2018). What is so special about online (as compared to offline) hate speech? *Ethnicities*, 18(3), 297-326. https://doi.org/10.1177/1468796817709846

Duffy, M. E. (2003). Web of hate: A fantasy theme analysis of the rhetorical vision of hate groups online. *Journal of Communication Inquiry*, 27(3), 291-312. https://doi.org/10.1177/0196859903252850

Gagliardone, I. (2019). Defining online hate and its "Public Lives": What is the place for "extreme speech"? *International Journal of Communication*, 13(2019), 3068-3087.

Garbe, L., Selvik, L. M. & Lemaire, P. (2023). How African countries respond to fake news and hate speech. *Information, Communication & Society*, 26(1), 86-103. https://doi.org/10.1080/1369118X.2021.1994623

López-Paredes, M. & Di Fátima, B. (2023). Memética: la reinvención de las narrativas en el mundo digital, protestas sociales y discursos de odio. In: Márquez, O.C. & Parras, A.P. (Eds.). *Visiones contemporáneas: narrativas, escenarios y ficciones* (pp. 25-37). Madrid: Fragua.

Matamoros-Fernández, A. & Farkas, J. (2021). Racism, hate speech, and social media: A systematic review and critique. *Television & New Media*, 22(2), 205-224. https://doi.org/10.1177/1527476420982230

Miranda, S., Malini, F., Di Fátima, B. & Cruz, J. (2022). I love to hate! The racist hate speech in social media. *Proceedings of the 9th European Conference on Social Media* (pp. 137-145). Krakow: Academic Conferences International (ACI).

Müller, K. & Schwarz, C. (2021). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association*, 19(4), 2131-2167, https://doi.org/10.1093/jeea/jvaa045

Ndahinda, F. M. & Mugabe, A. S. (2022). Streaming hate: Exploring the harm of anti-banyamulenge and anti-Tutsi hate speech on Congolese social media. *Journal of Genocide Research*, 1-15. https://doi.org/10.1080/14623528.2022.2078578

Schwarzenegger, C. & Wagner, A. J. (2018). Can it be hate if it is fun? Discursive ensembles of hatred and laughter in extreme right satire on Facebook. *Studies in Communication and Media*, 7(4), 473-498. https://doi.org/10.5771/2192-4007-2018-4-473

United Nations (n.d.). *Understanding hate speech: What is hate speech?* https://shre.ink/cVjq

1. AGGRAVATED ANTI-ASIAN HATE SINCE COVID-19 AND THE #STOPASIANHATE MOVEMENT: CONNECTION, DISJOINTNESS, AND CHALLENGES

Lizhou Fan
University of Michigan, USA
lizhouf@umich.edu

Huizi Yu
University of Michigan, USA
huiziy@umich.edu

Anne J. Gilliland
University of California, USA
gilliland@gseis.ucla.edu

As the COVID-19 pandemic has unfolded, there has been a dramatic increase in incidents of anti-Asian hate, including violent hate crimes such as the 2021 Atlanta Spa Shootings. Documenting and analyzing hate and counterspeech is essential and urgent work that can both record history in the making, and provide new insights for those working to de-escalate hate and diminish social inequity. By building two social media archives of hate and counterspeech on Twitter and using them to conduct different kinds of computational discourse analyses, we identified how anti-Asian hate has increased since the beginning of the COVID-19 pandemic, and how the #StopAsianHate movement has responded to many aspects of this hate, including stereotyping, stigmatization, and use of derogatory language. However, our research suggests that it remains challenging to counter anti-Asian hate speech

and the associated movement by responding in in direct and actionable ways that could attract more public attention and result in systemic changes in how Asians and Asian Americans are regarded in US society. We also argue that the forms of analysis we describe here show strong potential for use the emerging field of computational archival science – supporting archival digital intelligence by assisting archivists and researchers to identify important themes related to emerging social issues efficiently, and connections between very large digital collections, especially those of social media archives. **Keywords:** hate speech, counter-speech, anti-Asian, Covid-19, social media, Twitter

2. IS IT FINE? INTERNET MEMES AND HATE SPEECH ON TELEGRAM IN RELATION TO RUSSIA'S WAR IN UKRAINE ↗

Mykola Makhortykh
University of Bern, Switzerland
mykola.makhortykh@unibe.ch

Juan-Manuel González-Aguilar
International University of La Rioja, Spain
jm.gonzaguilar@gmail.com

The rise of digital platforms has changed the ways hate speech is disseminated today. Internet memes, namely digital content units sharing features of content and form, are one of the new formats in which hate speech is spread across different online platforms. Distinguished by their virality and frequent use of humoristic remixing of popular culture elements, memes are increasingly used by extremist groups to normalize hate speech towards vulnerable communities. However, the relationship of Internet memes and hate speech in the context of armed conflicts, where the use of hate speech is both particularly common and worrisome, currently remains under-studied. Using a sample of memes from pro-war Russophone Telegram channels, we examine this relationship in the context of the ongoing Russia's war in Ukraine. Relying on the intertextual discourse analysis, we identify three main functions of memes: 1) spreading hate speech; 2) amplifying personal attacks; and 3) glorifying the Russian army and its officials.
Keywords: memes, war, Telegram, Russia, Ukraine, hate speech

3. SYRIAN REFUGEES IN THE SHADE OF THE ‘ANTI-SYRIANS’ DISCOURSE: EXPLORING DISCRIMINATORY DISCURSIVE STRATEGIES ON TWITTER

Özlem Alikılıç
Yaşar University, Türkiye
ozlem.alikilic@yasar.edu.tr

Ebru Gökaliler
Yaşar University, Türkiye
ebru.gokaliler@yasar.edu.tr

İnanç Alikılıç
Malatya Turgut Özal University, Türkiye
inanc.alikilic@ozal.edu.tr

Along with the increase of user-generated content in social media, immigrants are often subject to hate speech. Recently, Turkey has become an important region for migrants from Syria, and the refugee problem has become a frequently shared issue by the Turkish public on social media. This study intended to evaluate the hatred dimension of contents produced on Twitter regarding the Syrian refugees in Turkey. For two months, 245,587 tweets in total, posted under the hashtags of ‘#suriyeli’ (Syrian), ‘#mülteci’ (refugee), ‘#suriyelimülteci’ (Syrian refugee), ‘#suriyelileriistemiyoruz’ (we don’t want the Syrians), and ‘#suriyelilerdefolsun’ (Syrians piss of), were collected, and discourse strategies were applied. Findings from the tweets showed that those who have negative views about Syrian refugees use discriminatory language to glorify the ‘we’ phenomenon

while separating the refugees into ‘others’. The findings also showed that positive tweets about Syrian refugees consisted of content around religion and supporting government policies. Among the negative contents, the excesses of criticisms regarding the Turkish government and its policies are remarkable. **Keywords:** Syrian refugees, Turkey, online hate speech, discriminatory discourse, Twitter

4. DISSEMINATING AND RESISTING ONLINE HATE SPEECH IN TURKEY ↗

Mine Gencel Bek
Universität Siegen, Germany
gencel.bek@sfb1472.uni-siegen.de

The chapter aims to contribute to the book with the Turkish case. It first reviews the literature on hate speech in Turkey with a special focus on the studies supported by the Hrant Dink Foundation which was established after the killing of Hrant Dink in 2007. A case study on hate speech recently directed to popular singer Sezen Aksu follows that. It reveals how hate speech is directed at the popular singer on different axis, including womanhood, LGBTQI, non-Turkish, and non-Muslim identities in the name of religion and Islam, as well as the association with animals as a hate object. Finally, the chapter discusses the ideas and attempts against hate speech and its limitations and potentials.

Keywords: hate speech, Twitter, popular culture, Turkey, sexism

5. HATE SPEECH ON TWITTER: THE LGBTIQ+ COMMUNITY IN SPAIN

Patricia de-Casas-Moreno
University of Extremadura, Spain
pcasas@unex.es

Macarena Parejo-Cuéllar
University of Extremadura, Spain
macarenapc@unex.es

Arantxa Vizcaíno-Verdú
University of Huelva, Spain
arantxa.vizcaino@dedu.uhu.es

The Internet and specifically social media became an area of interaction where hate speech gained visibility. Several minority groups have been exposed in an explosion of hateful comments due to their gender identity. In this case, the LGBTIQ+ collective group known as Lesbian, Gay, Bisexual, Transgender, Intersex, Queer and other identities not included in the above, became a target for their sexual orientation. This study intends to compile a comprehensive theoretical framework, as well as detailed case studies in Spain to offer an overview of the current panorama of the aforementioned group. We also outline the prevailing hate speech through social media such as Twitter. We conclude that there is still much to debate in this context and that platforms should be encouraged to strengthen their anti-speech measures to prevent and avoid this kind of discourse. **Keywords:** social media, LGBTIQ+, hate speech, Twitter, toxicity, Spain

6. CIRCULATION SYSTEMS, EMOTIONS, AND PRESENTEEISM: THREE VIEWS ON HATE SPEECH BASED ON ATTACKS ON JOURNALISTS IN BRAZIL

Edson Capoano
University of Minho, Portugal
edson.capoano@ics.uminho.pt

Vítor de Sousa
University of Trás-os-Montes and Alto Douro, Portugal
vitorsousa@jutad.pt

Vinicius Prates
University Presbyterian Mackenzie, Brazil
vinicius.prates@mackenzie.br

This text starts from the hate speech promoted during the presidency of Jair Messias Bolsonaro (2019-2022) to reflect on how we got here as individuals, communicators and society and what are the characteristics of this contemporary communicational phenomenon. For this, we will present three perspectives on hate speech to understand hate speech in an interdisciplinary way. The first will be the individual and biological sphere, on the neurological triggers of anger, the emotion that sustains hate speech, a theme so dear to the social sciences that it has caused the so-called emotional turn in the field. Next, the systemic issue of the hate circuit of narratives in communication environments will be presented, how they arise, how they propagate through networked information supports, how they feed back between contents crisscrossed. Finally, we will broaden the debate to the issue of historical presentism, a phenomenon of postmodernity that makes heterogeneous discourse something threatening to homogenizing groups, without spaces for the historical nuances necessary for the understanding of complex themes, simplified by hate speech, which circulate at the speed of digital social networks.With this approach, we hope to better understand what are the motivators of hate speech, such as those reported at the beginning of this text, and perhaps understand how to stop this spiral of narrative violence that affects the current society. **Keywords:** hate speech, communication, circulation, emotions, presenteeism

7. CLIPPING: HATE SPEECH IN SOCIAL MEDIA AGAINST FEMALE SPORTS JOURNALISTS IN GREECE

Lida Tsene
Open University of Cyprus, Cyprus
ltsene@gmail.com

The web 2.0 gave us the opportunity to explore new ways of collaboration and communication. Digital platforms and social media became a fertile ground for people to interact and express their opinions unfiltered, while the non-obligation to reveal oneself directly added an extra level of freedom in the way they shared news, thoughts and observations. But unfortunately, there is also the other side of the same coin. This democratisation facilitated somehow heated discussions which frequently result in the use of insulting and offensive language. In this chapter we are discussing sexist hate speech towards female sports journalists in Greece. Our research hypothesis drives from two basic facts related to the underrepresentation of women both in media and in sports. Through content analysis and in depth interviews we attempted to explore whether women working in the sports journalism field in Greece have been targets of online abuse, with a special focus on sexism hate speech, how do they respond and the impact this might have on their professional development and mental health, the role of Internet and social media as well as possible solutions to this challenge. **Keywords:** equity, gender, hate speech, sexist hate speech, social media, sports journalism

8. MAPPING SOCIAL MEDIA HATE SPEECH REGULATIONS IN SOUTHERN AFRICA: A REGIONAL COMPARATIVE ANALYSIS ↗

Allen Munoriyarwa
University of Botswana, Botswana
munoriyarwaa@ub.ac.bw

This chapter provides a comparative content analysis of social media hate speech in seven selected Southern African countries of South Africa, Zimbabwe, Eswatini (formerly Swaziland), Lesotho, Zambia, Democratic Republic of the Congo (DRC) and Botswana. Its aim is to examine how these countries, regulate social media hate speech, and how they legally sanction it. The chapter observes that as a preventive measure of social media hate speech, regulations have failed in these countries. It notes the weaponisation of hate speech to haunt legitimate anti-regime forces in some of these countries, and further notes how social media hate speech is increasingly blurring the lines on the maintenance of social order, political authoritarianism and free speech. The chapter concludes that an overhaul of social media hate speech regulations is necessary in Southern Africa if the laws are to serve their legal purposes. **Keywords:** hate speech, social media, Southern African region, authoritarianism, weaponization

9. ETHIOPIAN SOCIO-POLITICAL CONTEXTS FOR HATE SPEECH

Muluken Asegidew Chekol
Debre Markos University, Ethiopia
fulday02@gmail.com

Continuous Ethiopian youths' protests in Ethiopia for two years, forced the EPRDF's government to reform that has brought Abiy Ahmed to the Prime Minister position on April 2, 2018. This change has resulted in so many improvements on content and structure of the media including the online platform. Mostly, media had been filled with unison messages. Nevertheless, the situation did not last long; ethnic tension has risen again; ethnically motivated conflicts have become prevalent and caused peoples' death, and displacement. Hate speech and fake news also seemingly become common both on some mainstream and online media, which ultimately forced the state to endorse a law to suppress hate speech and fake news. This chapter prepared base on empirical studies. The study employed a mixed method research approach to understand and explain the prevalence, natures, severity, and regulation of social media hate speech in Ethiopia. As a data source, using a multi-stage sampling of users' comments offered on three purposeful selected Ethiopian Ethnic-media's social media sites, namely ASRAT, OMN, and DWTV, hate speech analysis were made. In addition to the content analysis of the online comments on the Facebook pages and the YouTube channels of the three media, the study included focus group discussions, interviews, and documents analysis tools to owe relevant data. Accordingly,

the study found a substantial prevalence of social media hate speech, dominated by offensive severity, and less incitement to violence, and genocide. It is also found that the ethnic-politics based hate was overriding. Identity-driven contesting and reform incidents were the main trigger factors of social media hate speech. It is argued, the law in place to minimize hate speech, may be used by the executive body for political interests to silence critical voices. As such, the prevalent of hate speech on the online media will have severe effects on the Ethiopian community. Along with the law, political dialogue to dig out the root causes of the hate speech, and enhancing media literacy in the country could be the potential solutions to deter hate speech in Ethiopia.
Keywords: hate prevalence, hate severity, hate natures, speech regulation, Ethiopia

10. SOCIAL MEDIA NARRATIVES AND REFLECTIONS ON HATE SPEECH IN NIGERIA

Aondover Eric Msughter
Caleb University Imota, Nigeria
aondover7@gmail.com

All over the world, hate speech represents a form of threat to damage the lives of individuals and increase the sense of fear. The recent trend in journalism malpractice in the country is the dissemination of hate speech and vulgar language. Within this context, the paper examined social media narratives and reflections on hate speech in Nigeria. The theoretical postulations of Castells' Theory of Network Society, Durkheim's Social Fact and Weber's Social Action or Relations Theory, The Functional Theory of Campaign Discourse, Critical Discourse Analysis Theory and Critical Race Theory were used as theoretical framework. Based on the literature, the paper argues that while still countering hate speeches in the traditional media, the emergence of social media has broadened the battlefield in combating the hate speech saga. Social media offers an ideal platform to adapt and spread hate speech and foul language easily because of its decentralised, anonymous and interactive structure. The prevalence of hate speech on social media bordering on political and national issues, and even social interaction in Nigeria, especially on Facebook, Twitter, YouTube and LinkedIn is becoming worrisome. This is because apart from undermining the ethics of journalism profession, it is contributing in bringing disaffection among tribes, political class, and religion or even among friends in the society. The paper concluded that Nigerian public is inundated with negative social media usage such as character assassination and negative political campaigns at the expense of dissemination of issues that help them make informed choices. **Keywords:** hate speech, narratives, Nigeria, reflections, social media

11. HATE SPEECH AMONG SECURITY FORCES IN PORTUGAL

Tiago Lapa
Iscte - University Institute of Lisbon, Portugal
tiagolapasilva@gmail.com

Branco Di Fátima
LabCom - University of Beira Interior, Portugal
brancodifatima@gmail.com

The European Union and the United Nations recognize hate speech as a threat to democracy, human rights, and peace. However, there is no universal definition of what hate speech is. Its meaning has been fluid and diverse, varying across countries, governing bodies, and disciplinary lenses. There are also considerations about the distinction between offline and online hate speech, since digital platforms might allow anonymity, invisibility, the instantaneous spread of hateful content and the clustering of hate speakers with like-minded individuals (Brown, 2018) that might be instilled with a sense of empowerment and exemption. It has been argued that online hate speech can be described as toxic behavior and in cases outright unlawful, exacerbated by Internet culture and the digital underworlds. On social media, hate speech can take different forms, but has been characterized by its hurtful or potentially harmful (visual and/or textual) language. This chapter presents a brief case study on the use of closed Facebook groups by security force officers to propagate hate speech against activists and minorities in Portugal. In this context, academics and legislators have always been faced with the contraposition between hate speech and freedom of expression. Where does one begin and the other end? One may question the efficacy of hate speech regulations, especially when law enforcement officers use social media to promote hate speech as if it were acceptable in democratic societies. **Keywords:** hate speech, social media, security forces, Facebook, Portugal

Chapter 1

AGGRAVATED ANTI-ASIAN HATE SINCE COVID-19 AND THE #STOPASIANHATE MOVEMENT: CONNECTION, DISJOINTNESS, AND CHALLENGES

Lizhou Fan
University of Michigan, USA

Huizi Yu
University of Michigan, USA

Anne J. Gilliland
University of California, USA

1. Introduction

Anti-Asian hate is a growing social problem in both the US and around the world. Anti-Asian hate incidents, including hate speech and hate crimes, have seen unprecedented increases in the US since the beginning of the COVID-19 pandemic. From 2020 to 2021, anti-Asian hate crimes increased by 833% in New York City and 700% in Sacramento (Levin, 2021). Anti-Asian physical assaults in the US doubled from 8.1% of total hate incidents in early 2020 to 16.2% in late 2021 (the Stop AAPI Hate coalition, 2020, 2022). Anti-Asian hate, however, is not a new social issue but rather is deeply rooted in long and systematic racism in the US towards people of Asian origins. Among targets of anti-Asian hate, anti-Chinese hate has a particularly long and specific history, especially on the west coast, dating all the way back to the arrival of the first Chinese immigrants in the nineteenth century.

Two of the first US immigration laws, *The Page Act of 1875* (1875) and *The Chinese Exclusion Act* (1882), intentionally and explicitly prohibited Chinese laborers from entering the country. Chinese laborers were seen as endangering "the good order of certain localities" (*The Chinese Exclusion Act*, 1882), creating workplace competition, and as foreigners who could not become US citizens and therefore should be kept out of the country. In addition to this legal discrimination and the daily prejudice experienced by Chinese immigrants, there is also a history of scapegoating Chinese people and communities during epidemics (Zhou, 2021). For example, when smallpox broke out in San Francisco in 1875, city health officers blamed Chinese immigrants in Chinatown; as they did again for the prevalence of venereal disease and during an unusual outbreak of bubonic plague between 1900 and 1904 (Trauner, 1978).

With such entrenched historical discrimination and stigmatization and the emergence of the COVID-19 virus as a direct triggering event, it is perhaps not surprising that toxic racism and even violent hate crimes escalated rapidly as the ensuing pandemic spread across the US and around the globe. When the Chinese city of Wuhan was identified in 2020 as the location of the first known cases of the virus, aggressive discrimination and stigmatization began, both online and in daily life towards people of Chinese origins. Egged on by a US President who persisted in referring to COVID-19 as the "Chinese virus" and "Kung flu", anti-Asian and anti-Chinese haters again treated Asians and Asian Americans, especially those of Chinese heritage, as medical scapegoats, repeating disinformation that associated the virus with Chinese food, eating habits and hygiene (King, 2020; Q. Yang et al., 2021). On March 16, 2021, six women of Asian origin were killed in Atlanta, possibly out of racist motivations. This tragic event, known as the 2021 Atlanta Spa Shootings (Stewart, 2022), precipitated counter anti-Asian hate and a heightened sense of social urgency in the US of the imperative to address anti-Asian hate. The #StopAsianHate movement, an online-offline hybrid movement is an innovative form of social activism that combines the internationality of hashtag activism with local protests, and was one of the first responses to the aggravated anti-Asian hate. It has given a heightened presence to Asian and Asian American communities that heretofore had lower-than-average involvement in social movements and limited political influence. Social media in particular has amplified their voices and provided diversified channels for being heard.

This chapter discusses the processes and outcomes of our research that applies mixed computational and human methods to identify the dynamics of anti-Asian hate speech and counterspeech on social media and provide insights into the effectiveness of that counterspeech. After a brief review of recent research on computational techniques for analyzing hate speech and the effectiveness of counterspeech, the chapter describes the processes and methods we used to build and analyze two archives of Twitter relating to anti-Asian hate and summarizes our findings in four main areas:

1. The trending anti-Asian hate speech categories on Twitter and their changes in volume during the early stage of the COVID-19 pandemic;
2. The volume and hashtag discourses of the counter anti-Asian hate movement, #StopAsianHate;
3. The connection and disjointness between anti-Asian hate and counterspeech, as well as current challenges in tackling anti-Asian hate;
4. The implications for documenting and analyzing social media data streams, which prototype and work towards "archival digital intelligence".

2. Related Work

The prevalence of hate speech and counterspeech on social media has attracted increasing research interest over the past five years. Earlier research argued that counterspeech is a promising way to respond to and mitigate the harms caused by hate speech (Lepoutre, 2017). Mathew et al. (2018) proposed counterspeech as an effective method for tackling hate speech without harming freedom of speech. Other researchers sought to detect and classify hate and counterspeech, understand their dynamics, and provide suggestions for countering hate speech.

Finding hate and counterspeech and differentiating between them can be challenging because of the complexity of how humans use language to express themselves, especially how they use language to discuss controversial topics, engage in emotional or heated exchanges, express prejudiced notions, or counter comments they find objectionable. Today, social media are widely used for such discourse, and because their content can be captured and archived in digital form, they can yield a rich text base on which to perform research relating to different types of speech, their dynamics and their impact. With advances in computational tools and natural language processing (NLP) techniques, developing effective hate and counterspeech detection and classification systems has become possible and increasingly nuanced. For example, Mathew et al. (2020) developed a classifier based on social media user data and linguistic patterns that can detect whether a user is a hateful or a counter speaker. Garland et al. (2020) used an ensemble learning algorithm that pairs a variety of paragraph embeddings with regularized logistic regression to classify hate and counterspeech. Yu et al. (2022) found that neural networks for identifying hate and counterspeech can perform better if context such as the preceding comment in a conversation is taken into consideration. Although no direct comparisons have been undertaken of the variety of methods that are now available, context-aware and neural network-based NLP methods are widely believed to perform well in detecting and classifying hate and counterspeech.

Regarding the dynamics between hate and counterspeech, in addition to counterspeech's effects on hate speech, recent research also examines the differences and interactions between them. Mathew et al. (2020) studied the topical difference between hate and counter speakers on Twitter and concluded that hate speakers, who often use subjective and negative expressions associated with envy, hate and ugliness, attracted more popularity than the counter users, who use more words related to government, law, and leadership. Garland et al. (2022) investigated the interactions of hate and counterspeech with different degrees of organizational behavior and found that organized counterspeech may help more than unorganized counterspeech in curbing online hate discourse.

In addition to policy-level insights, NLP researchers have developed and suggested implementing automatic and large-scale counterspeech generation to tackle hate speech, which can serve as "a third voice" to inform social media users of their inappropriate language uses without harming the principles of freedom of speech (Alsagheer et al., 2022). In general domains and rich resource languages, recent work showed that it is possible to combine pre-trained and large-scale language models, using for example, GPT-2 for synthetic text production, with additional tuning methods, for example, stochastic decodings, to generate counterspeech (Tekiroğlu et al., 2020; Tekiroğlu et al., 2022). In cross-domain and multilingual settings, it is also possible to create datasets and develop models that can help create counter-hate rhetorics (Chung et al., 2019, 2021). Beyond data and model, Zhu and Bhat (2021) found that a pipeline containing a generative model, a filter model, and a retrieval-based model can improve diversity and relevance in generating counterspeech for online hate speech.

While the current research on hate and counterspeech is extensive and has demonstrated application potential, due to the complexity of the origin and development of hate speech, hate speech targeted towards specific marginalized groups has been much less subject to analysis and can also be missed if the scope of the research is too general. This scarcity of granular research focusing on specific communities that have been the target or subjects of hate and counterspeech is a critical absence. Among the few studies to date, He et al. (2021) analyzed the development and diffusion of anti-Asian hate and counterspeech since COVID-19 on Twitter and found that counterspeech discouraged users from becoming hateful. However, due to the limited scope of this study's data collection, some unpredictable but closely related events may not be covered by the preselected fixed set of keywords. For instance, the #StopAsianHate social movement is not directly connected to hate speech data collected using keywords related to COVID-19 and

Asians. Moreover, none of the previous research differentiates different types of hate speech or tries to understand the motive or origin of the hate speech.

Thus, it is imperative to come up with an analytical framework, as well as a prototype, of hate and counterspeech that can make connections between discourses by applying detailed contextual aboutness that goes beyond binary classification. Instead of depending on in-dataset chronological closeness, overlapping keywords, or entangled user networks, as many other studies do, we report here on our efforts to find a cross-dataset connection between hate and counterspeech using granular information about the mappings between different categories of hateful and countering rhetoric.

3. Data and Methods

3.1. Data

To document and analyze hate speech related to China and counterspeech in the #StopAsianHate movement, we used the Twitter search API[1] to obtain public social media discourse that contains hate and counterspeech. Using the query “china+and+coronavirus”, we collected about 3.5 million tweets and published them in the COVID-19 Hate Speech Twitter Archive (CHSTA)[2] (Fan et al., 2020). Using the query term “StopAsianHate”, we collected more than 5.5 million tweets and published them in the Counter-anti-Asian Hate Twitter Archive (CAAHTA)[3] (Fan et al., 2021). In this section, we introduce the overall volumes and discourses of hashtags of CHSTA and CAAHTA respectively.

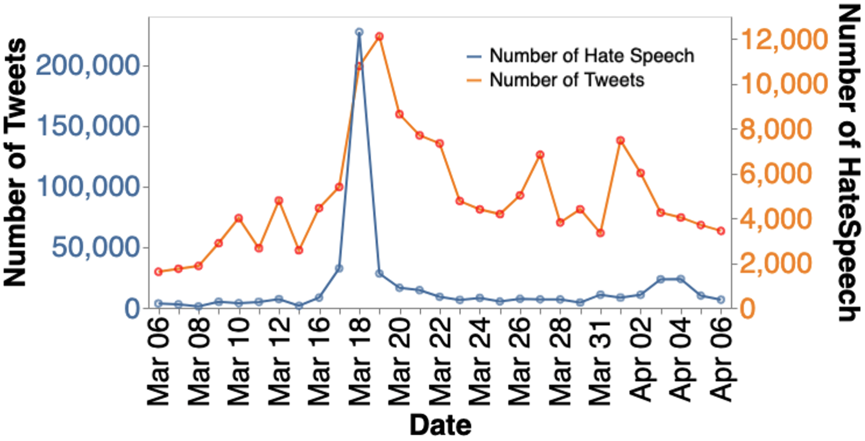
3.1.1. CHSTA: The Covid-19 hate speech Twitter Archive

The blue line in Figure 1(a) shows the volume of total tweets obtained between March 8 and April 6, 2020, using the query “china+and+coronavi-

[1] Developer Twitter: https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/api-reference/get-search-tweets
[2] GitHub: https://github.com/lizhouf/CHSTA
[3] GitHub: https://github.com/lizhouf/CAAHTA

rus”. There is an overall increasing trend of tweets across this period, with a substantial increase in the number of tweets between March 16 to March 19, 2020. We believe this occurrence is the result of burgeoning confirmed cases in the U.S. and growing media attention. Of the 3,457,402 tweets related to “china+and+coronavirus”, 25,467 are labeled as Hate Speech (Fan et al., 2020). We identify a tweet as hate speech if it contains at least one word from the Hatebase dictionary[4], which contains thousands of discriminatory words. We note that although Hatebase is a continuously updating dictionary of hate words, it might not fully capture all hate speech that is used. It may also not be completely up to date with the language being developed and used at speed in social media streaming. We use the red line in Figure 1(a) to show the trend in the volume of hate speech over time. This trend has a high association with the trend of total tweets, both peaking between March 16 and March 29, 2020.

a) Volume change of tweets



[4] Hate Base: https://hatebase.org/ (the API access retired in 2022)

b) Wordcloud of hashtags

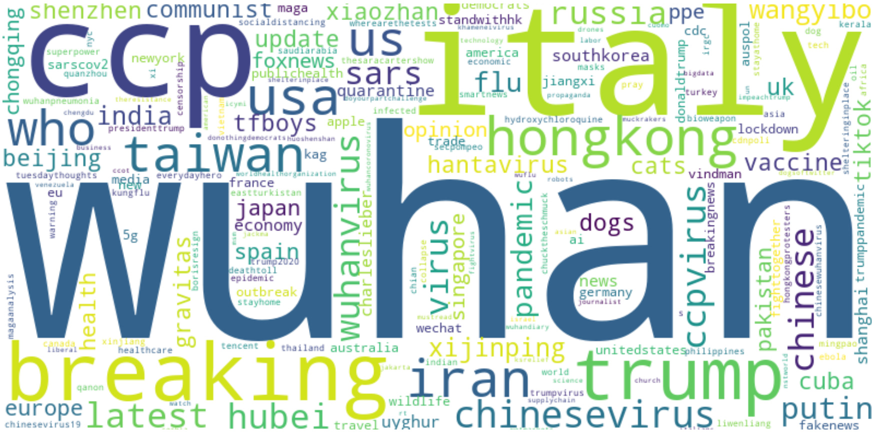


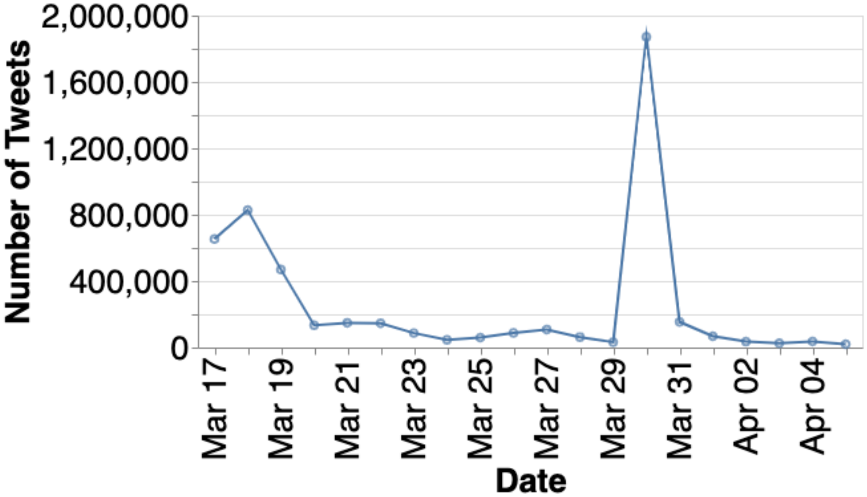
Figure 1 – Volume change and hashtag wordcloud of CHSTA.

The wordcloud in Figure 1(b) shows the discourse of tweets as reflected in their hashtags and provides a summary of user input topics. By analyzing the most frequently used hashtags, we identified several main categories of speech contained in the archive, including location, person, organization and abstract concept. The location hashtag “wuhan” is the most prevalent hashtag used in the archive, with over 24,000 occurrences. Other location hashtags such as Italy, USA, Hongkong and Hubei (of which Wuhan is the capital city) are also prominent. A large number of location hashtags appears to support the hypothesis that the trend of speech discourse is associated with demographics and geographic location. Additionally, hashtags such as “chinesevirus” and “wuhanvirus” carry discriminatory connotations and are a violation of the WHO’s convention for naming new human infectious diseases (World Health Organization, 2015). As suggested by this exploratory analysis, Twitter users frequently use discriminatory or prejudicial language against particular groups. Based on this result we conducted further analyses using lexicon-based information extraction methods.

3.1.2. CAAHTA: The counter-anti-Asian hate Twitter Archive

Figure 2(a) shows the longitudinal trend of the total number of tweets obtained using the query term “StopAsianHate”. We noticed two significant spikes: one on March 18, 2021, and another on March 30, 2021. To identify the driving force behind such sudden increases in anti-Asian hate public discourse, we searched for key events around those times. The Atlantic Spa Shooting, which took place on March 17, 2021, led to the increase in anti-Asian hate discourse (first spike) on March 18, 2021, and precipitated all subsequent social movements and events. On March 26, 2021, the popular South Korean boy band BTS advocated for “#StopAsianHate” on Twitter, which led to notably increased media attention and public discussion on Twitter in the subsequent days. By further analyzing the traffic peaks in relation to the key events, we observed that social influencers play vital roles in advocating for and promoting social events.

a) Volume change of tweets



b) Wordcloud of hashtags

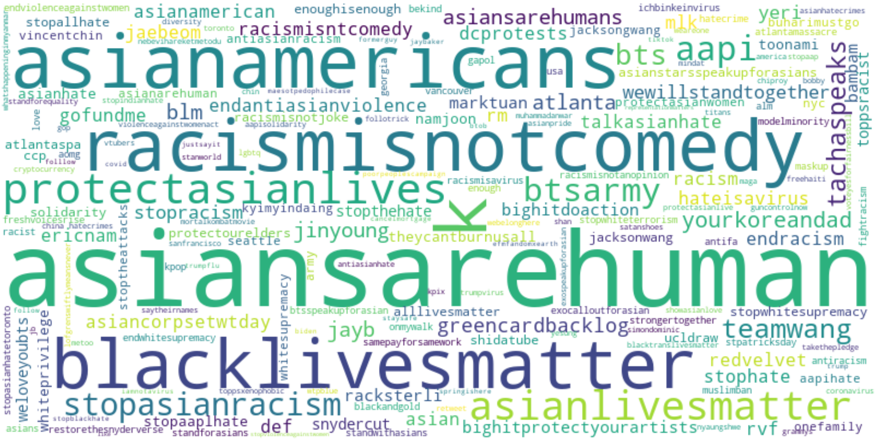


Figure 2 – Volume change and hashtag wordcloud of CAAHTA.

Similar to the first hate speech archive (CHSTA), with CAAHTA we identified a set of more than 300 frequently used hashtags that could be used as specific query words in future archival ingest activities. From the wordcloud, we identified a few emerging topics and concerns. Besides hashtags that advocate for general actions such as #stopasianracism and #endantiasianviolence, we also observed hashtags that call for specific actions such as #gofundme (advocating fundraising for survivors of Asian hate crimes) and #racismisnot before comedy (responding to specific types of anti-Asian hate speech). Additionally, we observed hashtags of related racial movements such as #blacklivesmatter and #blm. These preliminary findings suggested that there might be multiple dimensions of counter-anti-Asian hate speech discourse, which we have analyzed systematically as described in the following sections.

3.2. Methods

In this study, we used Computational Discourse Analysis (CDA) to analyze the connection between the two separate social media data archives. CDA is a mixed method that uses natural language processing (NLP) to automatically detect cohesion and local coherence, which can then be used in making summative inferences (Dascalu, 2014), while the inclusion of theoretical frameworks enhances the applicability and specificity of the results of the computational analysis. As a data-driven method, CDA is widely applied to harvest data and build archives from social media and web pages (Andreotta et al., 2019; Fan et al., 2022), as well as in conducting predictive modeling (Emmert-Streib & Dehmer, 2021).

The CDA method is useful in this study because it operationalizes theoretical frameworks computationally, and combines the strengths of both humans and algorithmic processing. As already introduced in the Data section, we need to connect the hate and counterspeech discourse identified from the two social media archives that is neither collected in the same time period nor developed through queries that used overlapping keywords. Thus, human-identified contextual aboutness is key to making connections between the two corpora, since traditional computational methods such as topic modeling or text clustering are unable to construct this high-level connection.

We therefore first clustered (through computation) and labeled (through human annotation) tweets in CHSTA into three categories, namely Stereotyping, Stigmatization, and Derogatory Language, that are potentially indicative of anti-Asian hate speech. We then analyzed the top hashtags in CAAHTA by labeling (through human annotation) counterspeech categories, including Advocating Action, Influencing Narrative Change, and Building Identity, that corresponded to the three hate speech categories already applied to CHSTA, as well as three dimensions of social movement and ten sub-dimensions, which are discussed below. Finally, we combined the above analyses, focusing on any matching between hate and counter categories.

3.2.1. Clustering and analyzing hate speech in CHSTA

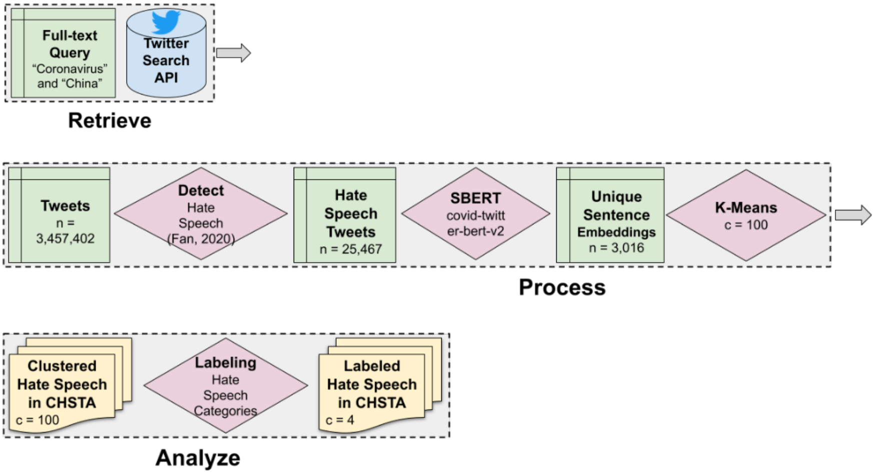

Full-text Query "Coronavirus" and "China"
Twitter Search API
Retrieve
Tweets n = 3,457,402
Detect Hate Speech (Fan, 2020)
Hate Speech Tweets n = 25,467
SBERT covid-twitt er-bert-v2
Unique Sentence Embeddings n = 3,016
K-Means c = 100
Process
Clustered Hate Speech in CHSTA c = 100
Labeling Hate Speech Categories
Labeled Hate Speech in CHSTA c = 4
Analyze

Figure 3 – The workflow of computational discourse analysis for CHSTA.

Note: The color of each frame represents the category of the sub-step – green frames are in-step text sources, yellow frames are end-of-step text results, red frames are actions of computational decision-making or human analysis references, and blue frames are API querying actions.

As Figure 3 shows, after retrieving and detecting hate speech in CHSTA, as described in the Data section above, we proceeded through the following processing steps. To cluster and analyze hate speech in CHSTA, we first used Sentence-BERT (SBERT), a transformer-based pre-trained NLP model to derive semantically meaningful sentence embeddings (Reimers & Gurevych, 2019). SBERT sentence embeddings, or sentence vectors, can support effective comparison between sentence meanings and can put together semantically similar sentences. Our implementation uses SBERT[5] with Python and the pre-trained model 'covid-twitter-bert-v2' (Müller et al., 2020), which is trained on COVID-19-related Twitter data and can map tweets in CHSTA to a 512-dimensional vector space. We then use the K-Means to cluster these sentence embeddings based on the scikit-learn

[5] Sbert: https://www.sbert.net/

package in Python[6]. We use Lloyd's K-Means clustering algorithm (Lloyd, 1982), since it is simple and efficient in processing large-scale language embeddings with high dimensions[7], and we obtained 100 clusters of potential anti-Asian hate speech.

To further analyze the clusters of hate speech, we began with the widely-accepted United Nations (UN) definition of COVID-19-related hate speech: "a broad range of disparaging expressions against certain individuals and groups that have emerged or been exacerbated as a result of the new coronavirus disease outbreak – from *scapegoating*, *stereotyping*, *stigmatization* and the use of *derogatory*, misogynistic, *racist*, *xenophobic*, Islamophobic or antisemitic language[8]." Focusing specifically on anti-Asian hate speech during COVID-19 and considering the expressiveness of concepts in short expressions in natural language, we used the categories mentioned in the UN definition to come up with an aggregated characterization framework. As Table 1 shows, our hate speech analytical framework has three categories: Stereotyping, Stigmatization, and Derogatory Language. Notably, these labels contain combined elements in the UN definition and we provide simple examples as explanations.

[6] Scikit-Learn: https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html
[7] For technical details, see Appendix II. A Algorithm 1
[8] United Nations Guidance Note on Addressing and Countering COVID-19 related Hate Speech

Category	Definition and Notes	Relation to UN Definition	Simple Example
Stereotyping	A fixed idea that many people have about a thing or a group that may often be untrue or only partly true. Often without hate words.	Corresponds to *stereotyping*, *racist*, and *xenophobic* language.	“Chinese eat bat.”
Stigmatization	An action of describing or regarding someone or something as disgraceful or with great disapproval. Often uses misinformation or disinformation for reasoning. May contain hate words.	Corresponds to *stigmatization* and *scapegoating*.	“China is f*ck*ng evil because it created coronavirus.”
Derogatory Language	Language showing a critical or disrespectful attitude without apparent reason. Often because of xenophobia and racism. Often contains hate words.	Corresponds to *derogatory*, *racist*, and *xenophobic* language.	Racial slurs.

Note: The simple examples above are provided as examples of hate speech and may cause discomfort. The authors do not agree with and strongly condemn any form of hate speech or hate crime, including the contents above. Some of the letters in the hate words have been masked with asterisks.

Table 1 – Categories of hate speech in CHSTA.

3.2.2 Categorizing hashtag activism in CAAHTA

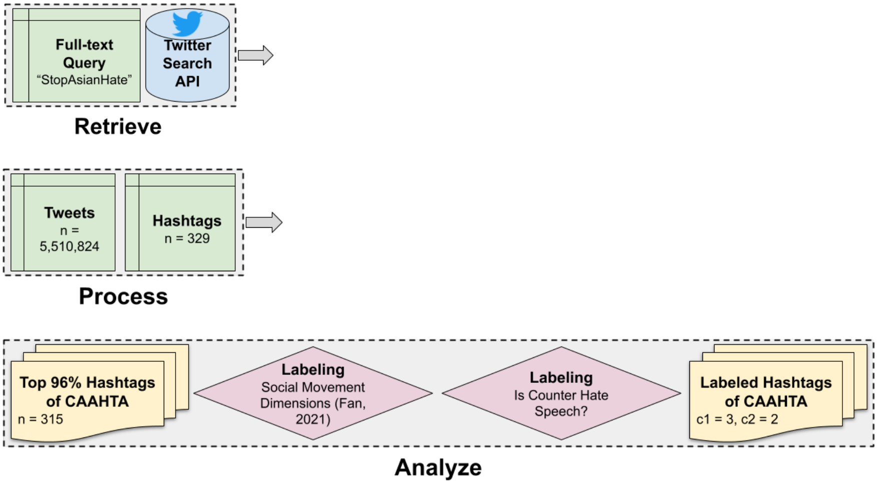


Figure 4 – The workflow of computational discourse analysis for CAAHTA.

Note: The color of each frame represents the category of the sub-step – green frames are in-step text sources, yellow frames are end-of-step text results, red frames are actions of computational decision-making or human analysis references, and blue frames are API querying actions. C1 corresponds to the number of dimensions of hashtag activism and C2 shows the binary annotation of counterspeech correspondence.

As Figure 4 shows, after retrieving tweets in CAAHTA, we extracted the hashtags for analysis. On social media, hashtags are representative keywords that are used to build up public support for affirmative social-political changes and social movements (Goswami, 2018). When a hashtag or a group of hashtags are used intensively, these short user inputs can serve similar purposes to slogans in protests, and they promote searching, liking, and forwarding the contents behind the hashtag through online social networks. When hashtags trend, narrative agency and activist messages associated with these hashtags will disseminate rapidly and can become the catalyst for online social movements (e.g., #BlackLiveMatter and #MeToo) (Xiong et al., 2019; G. Yang, 2016).

We then based our analysis of the hashtag activism in the #StopAsianHate social movement on 315 frequently used hashtags that encompassed more than 96% of all the hashtag uses. We applied Fan et al. (2021)'s adaptation of Reuning and Banazak's analytical framework for social movement phenomena (Reuning & Lee, 2019), which resulted in the identification of three dimensions, Advocating Action, Influencing Narrative Change, and Building Identity as well as 10 sub-dimensions of the hashtags' functionality in representing activism. As Table 2 indicates, the Advocating Action dimension includes *Specific Advocate*, which advocates for specific actions, and *General Advocate*, which contains the non-specific or overarching advocacy. The Influencing Narrative Change dimension includes different sub-dimensions based on identity groups and includes *AAPI Influencer*, social influencers with Asian origins, and *General Influencer* –social influencers with ethnicities other than Asian or Asian American. The Building Identity dimension includes six sub-dimensions of broader contexts related to Asian and Asian American identities and covers frequently mentioned concepts and incidents related to the unfolding social movement. We also provide examples for each dimension and its sub-dimensions in Table 2.

Dimension	Sub-dimension	Example Hashtags
Advocating Action	Specific Advocate	#racismisnotcomedy, #gofundme, #apologize_to_ bts
	General Advocate	#stopasianhate, #stopaapihate, #stopasianhatecrimes
Influencing Narrative Change	AAPI Influencer	#got7, #bts, #teamwang
	General Influencer	#tachaspeaks, #mlk, #biden
Building Identity	Related Racial	#asianamericans, #asiancorpsetwtday, #filipino
	Related Hate Inequality	#racism, #whiteprivilege, #chinesevirus
	Related Movement	#blacklivesmatter, #metoo, #guncontrol
	Related Entity	#wtpblue, #117thcongress, #tiktok
	Related Event	#greencardbacklog, #covid19, #toppsracist
	Related Place	#atlanta, #dcprotests, #atlantaspa

Table 2 – Social movement dimensions and sub-dimensions with example hashtags related to #StopAsianHate (Fan et al., 2021).

Finally, we sought to find connections between the social movement dimensions of hashtags and the hate speech categories. For each hashtag, we made binary annotations to check if it was counterspeech, either in general or specifically countering a category of hate speech. In so doing, we believe that analyzing hashtags is an appropriate way of finding hate and counterspeech connections because the use of hashtags shows both topical focus on tweets, which are the countering rhetorics we want to summarize[9], and the categories of hate speech.

[9] Semantic-based summary methods (e.g., topic modeling) are not working well on CAAHTA. The experimental results of topic modeling using Tomotopy (Minchul Lee, 2022), a topic modeling Python package, is provided in Appendix II.B.

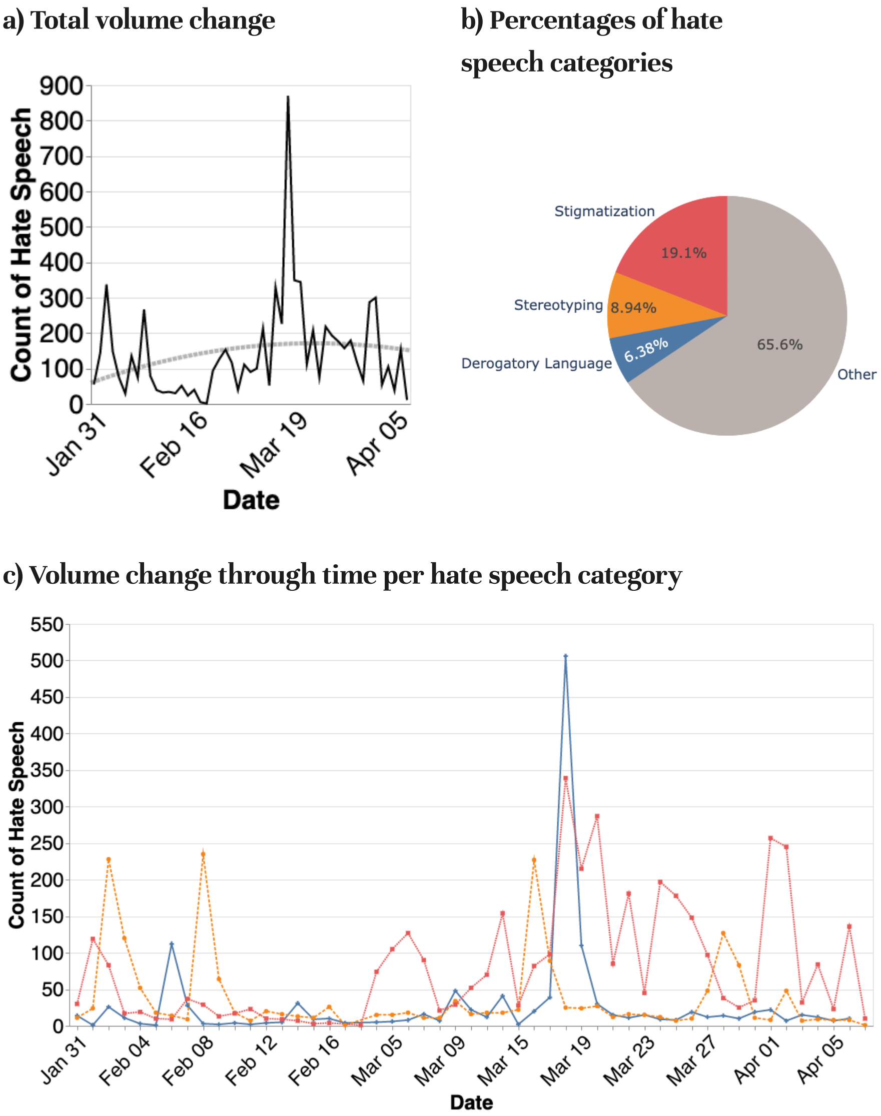
4. Results

In this section, we present what we refer to as the rising tide of anti-Asian hate speech during COVID-19 at the start of the pandemic (January to April 2020), where we analyze the hate speech volume change per category with examples. We also demonstrate the dimensions of the #StopAsianHate social movement after the 2021 Atlanta Spa Shootings and its countering connections to anti-Asian hate.

4.1. The rising tide of anti-Asian hate speech during Covid-19

Anti-Asian hate speech during the start of COVID-19 is a rising tide. Figure 5(a) shows the overall trend of anti-Asian hate speech in CHSTA: the volume quickly increased and doubled to its peak around March 19, 2020, while slowly decreasing afterward. Figure 5(b) further indicates the percentages of anti-Asian hate speech among the 25,467 potential incidences of hate speech on Twitter, where 19.1% are anti-Asian stigmatization, 8.94% are anti-Asian stereotyping, and 6.38% are uses of derogatory language. Notably, 65.6% of the potential hate speech is not anti-Asian hate speech, possibly because the tweets are selected using a dictionary-based method. After the clustering and labeling, we moved some potential hate speech clusters not targeting Asian or Asian Americans to the Other category, together with those Twitter-specific irrelevant tweets.

Regarding the discourse of the three categories of anti-Asian, and especially anti-Chinese hate speech, Figure 5(c) indicates the changes in their volumes. Stigmatization had local peaks (50-100 occurrences) in February and early March 2020, while the global peak at the start of COVID-19 reached more than 500 occurrences, which is also the single-day peak among all three categories. Stereotyping had local peaks throughout the start of the COVID-19 pandemic, often with higher volumes (more than 200 occurrences) than the local peaks of stigmatization. Derogatory language followed an increasing trend, which usually had less than 50 occurrences before early March, while soaring to more than 5 times the occurrences per day on average. We provide examples for each of the three anti-Asian hate categories with explanations in Table 3.



a) Total volume change

b) Percentages of hate speech categories

c) Volume change through time per hate speech category

Figure 5 – Percentage of Hate Speech Categories in CHSTA.

Categories	Example Tweets	Explanations	Cluster
Stereotyping	“China’s Shenzhen bans the eating of cats and dogs after coronavirus | Article [AMP] | Reuters They will eat anything that moved, if it doesn’t moved they pushed it! They can start eating horse, camel, donkey, monkey, panda, bear and gold fish”	Cultural stereotyping: wrong impressions relating to eating habits in China	6
	“China distributed the coronavirus and covered it up in order to try and cripple capitalist nations who rely on them for their medical supplies. Which is why they’re angry at Taiwan and 3M for sending the United States aid. Is this info not available here or something?”	Political stereotyping: assuming the tension based on the type of political system	17
	“Pakistan government left its students in China and they will be used as Guinea pig after being infected to test Coronavirus vaccine. Just wait and watch the true nature of Pak and an eyeopener for many in India.”	Political stereotyping: wrongly assumed mistreatment against foreigners in China during COVID	40

Stigmatization	“China invented it, take credit for it! You opened an advanced virology lab in Wuhan and now you shy away from the fruits of its labor? Embrace it! It’s yours! Now, don’t you dare shame us for acknowledging the fruits of your labor, especially when it’s killing us. #coronavirus”	Stigmatization based on misinformation: conspiracy theory about the origin of COVID	4
	“They did a mock simulation in 2018... read it people and you’ll think it’s todays headlines.. (change names to china-coronavirus)”	Stigmatization based on misinformation: conspiracy theory about China’s anticipation about COVID	12
	“‘Ebola was named after Ebola River in the Congo, Nipah Virus after the village Sungai Nipah in Malaysia. Then you have Guinea Worm, MERS But guess what: 2 viruses that originated in China, corona and SARS, aren’t linked to place of origin. #ChineseVirus ”	Stigmatization based on inappropriate reasoning	93

Derogatory Language	"B*tch.... what if the coronavirus was controlled in China but when them n*gg*s saw we was cracking jokes they said f*ck it and sent allem infected n*gg*s ova here ?"	Derogatory language with hate words: disrespectful to China using racial slurs	47
	"#TaiwanCanHelp Taiwan is the real b*tch. China has dealt with coronavirus already but Europe and North American is reacting slow as sh*t.Taiwan is still l*cking the American t*st*cl*s and providing the masks. What a shame. And stop asking China for help."	Derogatory language with hate words: disrespectful to Taiwan using disrespectful analogy	82
	"Thanks China Bat soup Coronavirus China virus Covid 19 Kung flu Wuhan clan Ch*ng Ch*ng coughs Flat faced fever Shanghai shivers Kung pow killer Sweet and sour sickness Poo poo lung Wet market weakness Pay back for Italians clams of starting pizza"	Derogatory language with hate words: disrespectful to various racial groups and regions	99

Note: The example tweets above are provided as example hate speech and may cause discomfort. The authors do not agree with and strongly condemn any form of hate speech or hate crime, including the contents in the examples above. Some of the letters in the hate words are masked with asterisks.

Table 3 – Example anti-Asian hate speech with categories.

4.2. The #StopAsianHate movement and counter hate speech

As Figure 4 shows, we use a Sankey diagram to indicate the dimensions and sub-dimensions of counterspeech and their correspondences to hate speech. After removing the thematic hashtags "#stopasianhate", "#stopaapihate" and "#stopasianhatecrimes", we observe a diverse discourse across the three dimensions of the #StopAsianHate social movement, where hashtags for *Building Identity* are mostly used (43.8%), followed by hashtags for *Advocating Action* (31.3%) and *Influencing Narrative Change* (25.0%).

For the sub-dimensions, we observe that some are clearly dominant in both the building identity and influencing narrative change dimensions: AAPI influencer hashtags are used about 10 times as frequently hashtags of general influencers; *Related Movement* hashtags are also used at least twice as frequently as hashtags of other sub-dimensions. For the advocating action hashtags, *General Advocate* hashtags are used more frequently than *Specific Advocate* hashtags.

Regarding the connection between the #StopAsianHate movement in CAAHTA and the anti-Asian hate speech in CHSTA, we see that 12.9% of hashtags are responding to hate speech, either generally or specifically to a category of anti-Asian hate. As Table 4 shows, some hashtags specifically respond to the hate speech categories of derogatory language (e.g., #racismisnotcomedy), stereotyping (e.g., #toppsxenophobic), and stigmatization (e.g., #equality). More examples of hashtags with their frequency are also provided in Table 4.

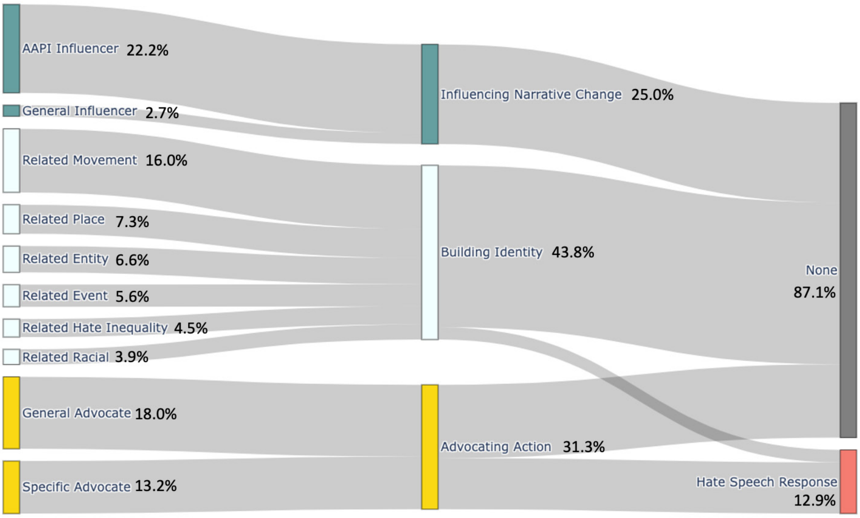


Figure 6 – Dimensions of counterspeech and correspondences to hate speech

Note: The proportions on the Sankey diagram are presented after removing the thematic hashtags "#stopasianhate", "#stopaapihate" and "#stopasianhatecrimes". The percentages of bars in each column do not add to one because of rounding issues.

Hashtag	Counter Hate Speech Category	Count
#racismisnotcomedy	Derogatory Language	24022
#racismisntcomedy	Derogatory Language	2820
#racismisnotjoke	Derogatory Language	862
#toppsxenophobic	Stereotyping	363
#webelonghere	Stereotyping	344
#equality	Stigmatization	237
#inclusion	Stigmatization	233
#justice	Stigmatization	228
#chinesevirus	Derogatory Language	218

Table 4 – Top hashtags directly responding to anti-Asian hate speech.

5. Discussion

Despite the long history of prejudice and racism experienced by Asians and Asian Americans within the US, they are often held up in popular conception as “model minorities” and are, therefore, unlikely to experience hate. Public awareness of Anti-Asian hate grew considerably after the outbreak of the COVID-19 pandemic.

In this section, we provide a detailed analysis of the (mis)match between anti-Asian hate and counterspeech in the #StopAsianHate movement through a comparative analysis of data in the CDA results. We then discuss the broader implications of this study through the lens of what we are calling “archival digital intelligence”, whereby applying computational archival science methods in collecting and managing social movement archives can benefit both documentation and analytical goals. Finally, we critically analyze the limitations of our study and outline some possible solutions.

5.1. The (Mis)match: Connection and disjointness between anti-Asian hate and counterspeech

We can make several observations about the developments of both the rising tide of anti-Asian hate and the unprecedented #StopAsianHate social movement. On the one hand, as Figure 3(a) shows, overall anti-Asian hate speech soared since the start of the COVID-19 pandemic and quickly reached its volume peak. As Figure 3(c) shows, the three categories of anti-Asian hate speech also indicate the origin and the potential development of anti-Asian hate speech: Stereotyping is associated with pre-existing impressions (e.g., eating habits and health conditions) that show up as hate speech periodically; Stigmatization becomes more intense when triggering events happen (e.g., the declaration of a global or local health emergency); Derogatory Language soars to a peak when there is strong hate speech from influencers (e.g., widespread racist speech or acts by politicians and celebrities) and remains at a high volume.

On the other hand, as we discussed in Section 4.2, the topical discourse represented by the hashtags in the #StopAsianHate movement shows that the counter anti-Asian hate speech covers all three dimensions of hashtag activism. The hashtag examples in Table 2 also show the potential popularity that hashtags can achieve if they focus on short and clear phrases for activism that lend themselves to rapid information diffusion via social media. A few examples in Table 4 also show that clear activist acts can develop into activist movements. For example, "#racismisnotcomedy" clearly criticizes the ignorance underlying making anti-Asian jokes and "#webelonghere" also challenges the "forever foreigner" misbelief to which Asian Americans have been subject throughout their history in the US.

In addition to how their respective discourses have evolved, there are clear overall connections between anti-Asian hate speech in CHSTA and counterspeech in CAAHTA. A substantial proportion of the slogans, i.e., hashtags, used in the #StopAsianHate movement are responding to and pushing back against the stereotyping, stigmatization, and derogatory language used in anti-Asian hate speech, indicating that the #StopAsianHate movement has successfully gathered momentum. However, we can also observe a mismatch between the demands and messages in the counter hate speech and hate speech categories: counterspeech in #StopAsianHate is often more general, less actionable, and less resonant than the intense, targeted language used in hate speech. This mismatch, or disjointness, suggests a challenge to the effectiveness of the #StopAsianHate movement, which can be a shared challenge of similar hashtag movements (e.g., #BlackLiveMatters and #MeToo).

5.2. Towards archival digital intelligence: Documenting and analyzing online hate and counterspeech

Documenting historical and contemporary actions and events, appraising social media to identify specific content, and making that documentation readily analyzable by subsequent researchers such as historians and policymakers all pose major challenges for archivists working with social media. Archivists began to grapple with the appraisal and description of social media content as early as 1992, with early studies suggesting that statistical approaches might be useful in exploring latent patterns within that content and user behaviors (Gilliland-Swetland & Hughes, 1992). More recently, advanced data science methods, a.k.a. computational archival science, artificial intelligence implementations such as machine learning, text classification, natural language understanding, and graph data mining have increasingly been used for large-scale digital collections management applications in archives (Colavizza et al., 2022; Fan et al., 2022; Fan et al., 2022; Fan & Presner, 2022; Franks, 2022; Yin et al., 2020). Such collections include not only bureaucratic records and other more traditional historical materials, but also new forms of social documentation including very large scale corpora of archived social media.

Compared to bureaucratic records that are generated through the activities of a single entity such as an institution or a research collaboration, digital content generated through social media is likely to be more heterogeneous in content and less predictable or understandable in terms of its provenance or intent. It is also likely to be far more immediate and voluminous, and the platform on which it originated, more opaque about its technology, practices and user base. This makes social media uniquely difficult to appraise for archival acquisition, and to describe or otherwise bring latent themes and patterns to the surface in ways that might be of relevance to subsequent researchers. Since social media archives also tend to drive and capture public responses in a very immediate way on emerging social issues, archives have been faced with a new problem of how to rapidly understand and process social media content in real-time, rather than after it has become inactive, in order to be able to provide researchers, policymakers and others with needed insights into this complex data almost at streaming speed (Fan et al., 2021). Computational archival science innovators, therefore, have been striving to identify methods that will help archivists quickly develop a sophisticated understanding, or digital intelligence, of a new data archive, that will enable them to process it and make it available for further analysis at the production-level speed.

A flexible analytical framework is key to promoting archival digital intelligence for broader applications. In our case, we mainly use the computational method (text clustering) for CHSTA and the qualitative method (hashtag close reading) for CAAHTA, while we combine the analytical results and make further inferences of the connected problems, i.e., anti-Asian hate and counterspeech movement. The flexibility here is regarding the choice of analytical methods. As Appendix II.B Figure 5 shows, topic modeling, a potential computational method, is not useful for analyzing CAAHTA, while the simple and effective qualitative analysis of hashtags works well. In this sense, we use a mix of computational methods and human analysis to examine the anti-Asian hate and counterspeech in two Twitter archives.

Our work prototypes a case where two archives are related in topics but not in data collection, and demonstrates how this can generate new archival digital intelligence that would not be possible through purely manual methods. By comparing the two archives using these flexible analytical approaches, we can optimize the combination of human and machine strengths to identify high-level theoretical connections and obtain knowledge that can be useful to both archivists (e.g., for making appraisal decisions or in the description) and researchers (e.g., to apply additional machine learning models, data pipelines, or analytical frameworks).

5.3. Limitations

There are a few technological and methodological limitations in our analysis. Because of infrastructural constraints, we conducted the data collection process for both CHSTA and CAAHTA in waves (Appendix I, Table 5 & 6). As a result, there are several data gaps in the archives. However, since the data gaps are small, we expect the data discontinuity to have minimal impact on the analysis. Additionally, the query keywords “china+and+coronavirus” and “#StopAsianHate” might not fully capture all tweets related to the topics of interest. For example, the delayed adoption of the term “COVID-19” in place of “coronavirus” might have diverted some of the information streams. In future research, or similar situations where it is impossible to collect complete data in one pass, archivists and researchers might be able to mitigate such problems by iteratively collecting data and updating the terms used in data collection based on term evolution or introduction that has been highlighted by analyses of the archive captured to date.

Additionally, we note some limitations caused by Twitter search API constraints. Our current CHSTA and CAAHTA archives are limited by the scraping rate of the Twitter API and contain approximately 1% of all Tweets containing the query keywords. Although the archives are not exhaustive, we believe the extracted samples are representative of the public discourse. Second, Twitter search API has a 7-day limit, which only allows for searches against a sampling of recent Tweets published in the past 7 days. This limitation unavoidably introduces additional data gaps in the archive, as we needed to initiate a new search manually every seven days. Lastly, deleted tweets are no longer retrievable by the search API, which might result in data loss. We found that approximately 1% of the tweets are not available, a small percentage that will not significantly impact the quality of the analysis. For future research, we will use the Twitter Academic API for exhaustive search results.

Our analysis of hate speech in CHSTA leverages both computational and human labor, which unavoidably introduces computing randomness and human judgment into the analysis. First, to validate that the scraping process has behaved as expected, we have employed a manual process of content relevance checking. We randomly sampled 100 tweets from both archives and inspected their content. Out of the 100 tweets in CHSTA, 96% were related to both “coronavirus” and “china”, and 3.3% were related to at least one of the topics. Similarly for CAAHTA, 98% were related to “#StopAsianHate”. The strong topical coherence of both archives ensures the validity of the subsequent analysis. Additionally, to mitigate the potential bias introduced by human annotators, the categorization and grouping of hate speech clusters were conducted by two researchers. We conducted intercoder reliability checks through the calculation of percent agreement and obtained a high agreement rate (83%) that confirms the internal consistency and validity of this study. Of the total 100 clusters, 83 % were placed in the same categories by two annotations, and the differences (17%) were resolved after discussion.

6. Conclusion

Anti-Asian hate is a deeply rooted social problem in the US and globally. COVID-19 is not the cause of anti-Asian hate, but it has triggered an increasing number of anti-Asian hate incidents, including hate speech. The 2021 Atlanta Spa shooting is not the main reason for the initiation of the unprecedented #StopAsianHate movement, but this violent event has unified voices of fighting against anti-Asian hate online and in local protests. Through our studies, we have analyzed relationships, including match and mismatch, between anti-Asian hate and the #StopAsianHate movement, and sought to identify the current progress in countering anti-Asian hate, as well as the detailed discourses of anti-Asian hate speech and the potential challenges in tackling them. Our studies indicate that there was a rapid increase in hateful rhetoric, including stereotyping, stigmatization, and the use of derogatory language towards China and Chinese people from the first appearance of COVID-19. The #StopAsianHate movement had a broad discourse on countering anti-Asian hate, including discourse advocating action, influencing narrative change, and building identity, but it remains challenging for this counter discourse to build sufficient momentum through Asian influencers to improve public perceptions of Asians and Asian Americans. Finally, we believe that these studies suggest potentially helpful approaches for the nascent field of computational archival science as it grapples with massive volumes of digital social media, and thereby, more effective archival digital intelligence about the contents of that social media.

Acknowledgments

This work is an extension and comparison of short conference papers by the co-authors. The co-authors thank our previous collaborator Zhanyuan Yin for his initial contribution to the work on which this research is based.

Appendix

I. Notes on data collection

Wave	Start Date	End Date	Total Number of Tweets
1	2020-01-31	2020-02-07	445,893
2	2020-02-07	2020-02-17	297,654
3	2020-03-03	2020-03-12	422,146
4	2020-03-15	2020-03-22	1,031,969
5	2020-03-22	2020-03-29	673,625
6	2020-03-31	2020-04-07	610,910

Table 5 – Metadata of tweets in CHSTA

Date	Total Number of Tweets
2021-03-19	469,305
2021-03-20	133,634
2021-03-21	148,861
2021-03-22	146,332
2021-03-23	87,021
2021-03-24	46,578
2021-03-25	59,207
2021-03-26	87,656
2021-03-27	108,444
2021-03-28	62,723
2021-03-29	32,347
2021-03-30	1,871,442
2021-03-31	154,039
2021-04-01	67,824
2021-04-02	35,139
2021-04-03	25,528
2021-04-04	36,744
2021-04-05	19,760

Table 6 – Metadata of tweets in CAAHTA.

II. Notes on the CDA method and related computing

A. K-means algorithm for clustering CHSTA

As Algorithm 1 shows, it first randomly initiates a number (a predefined hyperparameter) of centroids in the 512-dimension vector space and randomly assigns a centroid label to each of the 3,016 tweets embedding vectors. It then uses Euclidean distance as a measurement to recursively update the centroid assignments.

The updating loop stops when the centroid assignment to each of the tweets no longer changes and we keep the assigned centroid as their cluster labels. Regarding the number of centroid hyperparameters that we need to predefine, we experiment with three different numbers of clusters: 50, 100, and 200. Because 50 clusters lead to overly large clusters containing several topics and 200 clusters result in sparse clusters, we decided to cluster the corpus into 100 clusters.

While the tweet embeddings remain the same, the clustering results can be slightly different with different statuses (seeds) of randomness. Since the order and the trivial difference among the clusters of tweets do not matter for our research on discourse, we pick one of the results for demonstration in the Results section.

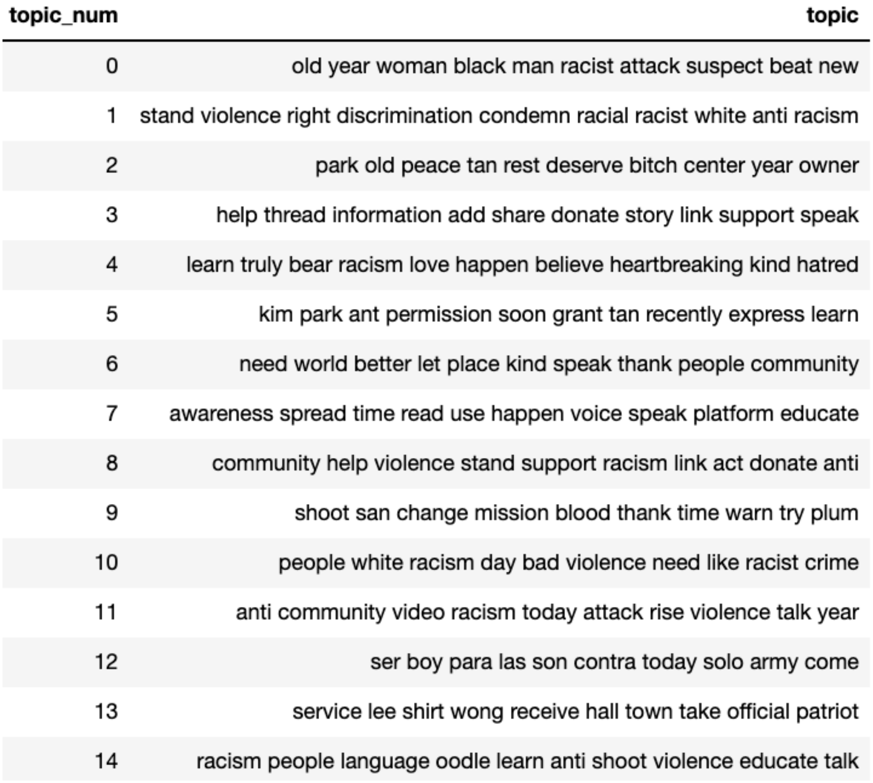
```
**Require**: Tweet embeddings vectors $\mathbf{S} = S_i$ for i = 1, 2, ..., 3,016 where $S_{ij}$ are vector space positions for dimensions j = 1, 2, ..., 512

**Ensure**: Labels $\mathbf{L} = L_i$ for i = 1, 2, ..., 3,016 for tweet embeddings

Initiate random centroids $\mathbf{C} = C_i$ for Clusters i = 1, 2, ..., 100 where $C_{ij}$ are vector space positions for dimensions j = 1, 2, ..., 512
Randomly initiate and assign labels of 1, 2, ..., 100 to $\mathbf{L'} = L'_i$ for i = 1, 2, ..., 3,016
**Repeat**
  Assign **L'** to **L**
  **for** m = 1, 2, ..., 3,016 **do**
    Initiate a vector of Euclidean Distance $\mathbf{D} = D_i$ for Clusters i = 1, 2, ..., 100
    **for** n = 1, 2, ..., 20 **do**
      Calculate Euclidean Distance $D_n = \sqrt{\sum_{k=1}^{512} (S_{mk} - C_{nk})^2}$
    **end for**
    **if** $D_p$ = argmin $||\mathbf{D}||$
      Assign $p$ to $L'_n$
  **end for**
**until** **L** = **L'**
**return** **L**
```

Algorithm 1 – Lloyd's K-Means algorithm for clustering hate speech in CHSTA

B. A topic modeling result of CAAHTA



topic_num	topic
0	old year woman black man racist attack suspect beat new
1	stand violence right discrimination condemn racial racist white anti racism
2	park old peace tan rest deserve bitch center year owner
3	help thread information add share donate story link support speak
4	learn truly bear racism love happen believe heartbreaking kind hatred
5	kim park ant permission soon grant tan recently express learn
6	need world better let place kind speak thank people community
7	awareness spread time read use happen voice speak platform educate
8	community help violence stand support racism link act donate anti
9	shoot san change mission blood thank time warn try plum
10	people white racism day bad violence need like racist crime
11	anti community video racism today attack rise violence talk year
12	ser boy para las son contra today solo army come
13	service lee shirt wong receive hall town take official patriot
14	racism people language oodle learn anti shoot violence educate talk

Figure 5 – A topic modeling result of CAAHTA

References

Alsagheer, D., Mansourifar, H. & Shi, W. (2022). *Counter Hate Speech in Social Media: A Survey*. https://doi.org/10.48550/ARXIV.2203.03584

Andreotta, M., Nugroho, R., Hurlstone, M. J., Boschetti, F., Farrell, S., Walker, I. & Paris, C. (2019). Analyzing social media data: A mixed-methods framework combining computational and qualitative text analysis. *Behavior Research Methods*, *51*(4), 1766-1781.

Chung, Y. L., Kuzmenko, E., Tekiroglu, S. S. & Guerini, M. (2019). CONAN - COunter NArratives through Nichesourcing: A Multilingual Dataset of Responses to Fight Online Hate Speech. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/p19-1271

Chung, Y.-L., Tekiroğlu, S. S. & Guerini, M. (2021). Towards Knowledge-Grounded Counter Narrative Generation for Hate Speech. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. https://doi.org/10.18653/v1/2021.findings-acl.79

Colavizza, G., Blanke, T., Jeurgens, C. & Noordegraaf, J. (2022). Archives and AI: An Overview of Current Debates and Future Perspectives. *Journal on Computing and Cultural Heritage*, *15*(1), 1-15. https://doi.org/10.1145/3479010

Dascalu, M. (2014). Computational Discourse Analysis. In *Analyzing Discourse and Text Complexity for Learning and Collaborating* (pp. 53–77). Springer.

Emmert-Streib, F. & Dehmer, M. (2021). Data-Driven Computational Social Network Science: Predictive and Inferential Models for Web-Enabled Scientific Discoveries. *Frontiers in Big Data*, *4*, 591749. https://doi.org/10.3389/fdata.2021.591749

Fan, L., Lafia, S., Bleckley, D., Moss, E., Thomer, A. & Hemphill, L. (2022). *Librarian-in-the-Loop: A Natural Language Processing Paradigm for Detecting Informal Mentions of Research Data in Academic Literature* (arXiv:2203.05112). arXiv. http://arxiv.org/abs/2203.05112

Fan, L. & Presner, T. (2022). Algorithmic Close Reading: Using Semantic Triplets to Index and Analyze Agency in Holocaust Testimonies. *Digit. Humanit. Q.*, *16*(3). http://www.digitalhumanities.org/dhq/vol/16/3/000623/000623.html

Fan, L., Yin, Z., Yu, H. & Gilliland, A. J. (2022). Using Machine Learning to Enhance Archival Processing of Social Media Archives. *Journal on Computing and Cultural Heritage*, 3547146. https://doi.org/10.1145/3547146

Fan, L., Yu, H. & Gilliland, A. J. (2021). #StopAsianHate: Archiving and Analyzing Twitter Discourse in the Wake of the 2021 Atlanta Spa Shootings. *Proceedings of the Association for Information Science and Technology*, *58*(1), 440–444. https://doi.org/10.1002/pra2.475

Fan, L., Yu, H. & Yin, Z. (2020). Stigmatization in social media: Documenting and analyzing hate speech for COVID-19 on Twitter. *Proceedings of the Association for Information Science and Technology*, *57*(1), e313. https://doi.org/10.1002/pra2.313

Franks, J. (2022). Text Classification for Records Management. *Journal on Computing and Cultural Heritage*, *15*(3), 1–19. https://doi.org/10.1145/3485846

Garland, J., Ghazi-Zahedi, K., Young, J.G., Hebert-Dufresne, L. & Galesic, M. (2020). Countering hate on social media: Large scale classification of hate and counter speech. *Proceedings of the Fourth Workshop on Online Abuse and Harms*. https://doi.org/10.18653/v1/2020.alw-1.13

Garland, J., Ghazi-Zahedi, K., Young, J.-G., Hebert-Dufresne, L. & Galesic, M. (2022). Impact and dynamics of hate and counter speech online. *EPJ Data Science*, *11*(1). https://doi.org/10.1140/epjds/s13688-021-00314-6

Gilliland-Swetland, A. J. & Hughes, C. (1992). Enhancing Archival Description for Public Computer Conferences of Historical Value: An Exploratory Study. *The American Archivist*, *55*(2), 316-330. JSTOR.

Goswami, M. P. (2018). Social media and hashtag activism. *Liberty Dignity and Change in Journalism*, 2017.

He, B., Ziems, C., Soni, S., Ramakrishnan, N., Yang, D. & Kumar, S. (2021, November 8). Racism is a virus. *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. https://doi.org/10.1145/3487351.3488324

King, M. T. (2020). Say no to bat fried rice: Changing the narrative of coronavirus and Chinese food. *Food and Foodways*, *28*(3), 237–249. https://doi.org/10.1080/07409710.2020.1794182

Lepoutre, M. (2017). Hate Speech in Public Discourse: A Pessimistic Defense of Counterspeech. *Apollo - University of Cambridge Repository*. https://doi.org/10.17863/CAM.15815

Levin, B. (2021). *Report to the nation: Anti-Asian prejudice and hate crime*.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137.

Mathew, B., Hardik Tharad, Subham Rajgaria, Prajwal Singhania, Maity, S. K., Goyal, P. & Animesh Mukherjee. (2018). Thou shalt not hate: Countering Online Hate Speech. *Unpublished*. https://doi.org/10.13140/RG.2.2.31128.85765

Mathew, B., Kumar, N., Goyal, P. & Mukherjee, A. (2020, January 5). Interaction dynamics between hate and counter users on Twitter. *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*. https://doi.org/10.1145/3371158.3371172

Minchul Lee. (2022). *bab2min/tomotopy: 0.12.3* (v0.12.3). Zenodo. https://doi.org/10.5281/ZENODO.6868418

Müller, M., Salathé, M. & Kummervold, P. E. (2020). *COVID-Twitter-BERT: A Natural Language Processing Model to Analyse COVID-19 Content on Twitter* (arXiv:2005.07503). arXiv. http://arxiv.org/abs/2005.07503

Reimers, N. & Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks* (arXiv:1908.10084). arXiv. http://arxiv.org/abs/1908.10084

Reuning, K. & Lee, A. (2019). Measuring Social Movement Phenomena: Action, Message, and Community. *Unpublished Manuscript*.

Stewart, L. Z.-L. (2022). *The yellow figment of East Asian American women: A case study of the 2021 Atlanta spa shootings*.

Tekiroglu, S. S., Bonaldi, H., Fanton, M. & Guerini, M. (2022). *Using Pre-Trained Language Models for Producing Counter Narratives Against Hate Speech: A Comparative Study*. https://doi.org/10.48550/ARXIV.2204.01440

Tekiroğlu, S. S., Chung, Y.-L. & Guerini, M. (2020). Generating Counter Narratives against Online Hate Speech: Data and Strategies. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/2020.acl-main.110

The Page Act of 1875, U.S. Congress (1875).

the Stop AAPI Hate coalition. (2020). *3-Month Report*. https://stopaapihate.org/3-month-report/

the Stop AAPI Hate coalition. (2022). *National Report*. https://stopaapihate.org/national-report-through-december-31-2021/

Trauner, J. B. (1978). The Chinese as Medical Scapegoats in San Francisco, 1870-1905. *California History*, *57*(1), 70–87. JSTOR. https://doi.org/10.2307/25157817

Chinese Exclusion Act, (1882) (testimony of U.S. Congress). https://www.loc.gov/resource/llsalvol.llsal_022/?sp=85&st=image&r=-1.492,-0.017,3.984,1.745,0

World Health Organization. (2015, May 8). *WHO issues best practices for naming new human infectious diseases*. https://www.who.int/news/item/08-05-2015-who-issues-best-practices-for-naming-new-human-infectious-diseases

Xiong, Y., Cho, M. & Boatwright, B. (2019). Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of Twitter during the #MeToo movement. *Public Relations Review*, *45*(1), 10–23. https://doi.org/10.1016/j.pubrev.2018.10.014

Yang, G. (2016). Narrative Agency in Hashtag Activism: The Case of #BlackLivesMatter. *Media and Communication*, *4*(4), 13–17. https://doi.org/10.17645/mac.v4i4.692

Yang, Q., Young, I. F., Wan, J. & Sullivan, D. (2021). Culturally Grounded Scapegoating in Response to Illness and the COVID-19 Pandemic. *Frontiers in Psychology*, *12*, 632641. https://doi.org/10.3389/fpsyg.2021.632641

Yin, Z., Fan, L., Yu, H. & Gilliland, A. J. (2020). Using a Three-step Social Media Similarity (TSMS) Mapping Method to Analyze Controversial Speech Relating to COVID-19 in Twitter Collections. *2020 IEEE International Conference on Big Data (Big Data)*, 1949–1953. https://doi.org/10.1109/BigData50022.2020.9377930

Yu, X., Blanco, E. & Hong, L. (2022). *Hate Speech and Counter Speech Detection: Conversational Context Does Matter.* https://doi.org/10.48550/ARXIV.2206.06423

Zhou, L. (2021, March 5). *The long history of anti-Asian hate in America, explained*. https://www.vox.com/identities/2020/4/21/21221007/anti-asian-racism-coronavirus-xenophobia

Zhu, W. & Bhat, S. (2021). Generate, Prune, Select: A Pipeline for Counterspeech Generation against Online Hate Speech. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. https://doi.org/10.18653/v1/2021.findings-acl.12

Chapter 2

IS IT FINE? INTERNET MEMES AND HATE SPEECH ON TELEGRAM IN RELATION TO RUSSIA'S WAR IN UKRAINE

Mykola Makhortykh
University of Bern, Switzerland

Juan-Manuel González-Aguilar
International University of La Rioja, Spain

1. Introduction

Hate speech is increasingly recognized as a major threat to human rights and the rule of law (ECRI, 2016). While there are multiple definitions of hate speech (see, for instance, Gagliardone et al., 2015; Fortuna & Nunes, 2018), most of them agree that it is constituted by content attacking specific groups distinguished by certain attributes, such as race or gender. Hate speech can serve different goals, but usually, it aims to incite and justify hate, discrimination, or violence toward the targeted group (Fortuna & Nunes, 2018). Consequently, hate speech often contributes to societal radicalization and the erosion of individual and collective empathy (Bilewicz & Soral, 2020). It can be an important factor in enabling different forms of mass violence, including genocide (Schabas, 2017).

The rise of online media has facilitated the spread of hate speech. Some of the factors that contribute to the distribution of hate speech online include the ease of producing and disseminating digital content that limits the effectiveness of its removal, the transnational nature of online platforms that prompts the need for cross-jurisdictional collaboration to legally counter hate speech, and anonymity that encourages the production of hate speech and complicates the identification of actors responsible for it (Gagliardone et al., 2015). Furthermore, the rise of new formats of digital content expands possibilities for spreading hate speech among certain types of audiences (e.g., youngsters in the case of TikTok; Weimann & Masri, 2020) by going beyond traditional forms of hate speech, which might be easier to recognize and counter.

One particular format which is increasingly employed for disseminating hate speech is Internet memes. Often defined as digital content units which share common features of content and form (Shifman, 2013), memes are a popular form of digital communication due to their highly affective potential, which is enhanced by memes' intertextuality (Wutz & Nugteren, 2018). The ability to adapt a broad range of cultural texts, ranging from popular culture products (Knobel & Lankshear, 2007) to mnemonic tropes (Makhortykh & González-Aguilar, 2020) contributes to memes' virality, namely the ability to quickly spread across online platforms and stimulate user engagement. This viral nature of memes contributed to their intense use for a broad range of purposes, from identity-building and public mobilization (Bozkuş, 2016) to challenging as well as reinforcing hegemonic historical narratives (González-Aguilar & Makhortykh, 2022) to dealing with collective tensions and anxiety (Steir-Livny, 2016). In addition to these uses, however, Internet memes are known to be adopted by extremist actors to facilitate the dissemination of hate speech, often by normalizing extreme messages with the help of humor (Lamerichs et al., 2018).

Despite the growing number of studies looking at the relationship between Internet memes and hate speech (e.g., Lamerichs et al., 2018; Bogerts & Fielitz, 2019; Woods & Ruscher, 2021; Nayak & Agrawal, 2022), only a few studies (e.g., Rodley, 2016; Gaufman, 2022) so far look at the use of memes in the context of armed conflicts. Yet, at the time of war, hate speech in its different mediatized forms is particularly concerning due to its potential to facilitate the dehumanization of the opponents and amplify ongoing hostilities (Ndahinda & Mugabe, 2022). Under these circumstances, it is important to look at the role of Internet memes in disseminating hate speech; specifically, we aim to answer the following questions: To what degree are Internet memes used to normalize hate speech in the context of armed conflicts? What features of Internet memes make them an effective means of spreading hate speech in online environments? And what are the other aims Internet memes can serve in the context of modern wars?

To address these questions, we look at Internet memes used in the context of the ongoing Russian-Ukrainian war, particularly following the large-scale Russian invasion in 2022. Specifically, we are interested in Internet memes used in Russophone Telegram channels, a medium that is integral for the mediatization of Russia's war in Ukraine (Bergengruen, 2022) and also known as "a favourable environment for the spread of hateful speech and disinformation" (ISD, 2022). Using intertextual discourse analysis, we examine different functions of memes used in the context of the war and discuss their relationship with the phenomenon of hate speech.

The rest of the chapter is organized as follows. First, we provide a short overview of existing research dealing with the relationship between Internet memes and hate speech, with a particular emphasis on the context of armed conflicts. It is followed by the examination of our approach for data collection and data analysis, supplemented with a discussion of the limitations of the chosen approach. Then, we introduce our findings concerning the three main functions of Internet memes in the sample of memes we collected: the dissemination of hate speech, the amplification of personal attacks, and the glorification of the Russian army. The chapter ends with a discussion of our findings and their implications for the research on hate speech.

2. Theoretical background

Internet memes are a digital communication phenomenon that has gained prominence in recent decades and plays an essential role in different social contexts due to memes' virality that enables their rapid diffusion across online platforms. Dawkins (1976) introduced the concept of a "meme" in the 1970s to denote gene-like cultural units (e.g., catchphrases) that virally reproduce themselves. The rise of online platforms, which contributed to Internet users increasingly moving beyond content consumption to content produsage (Bruns, 2008), signified a new stage in the memes' lifecycle. Amplified by the anonymity (Wiggings, 2016) and the connectivity (Makhortykh & González-Aguilar, 2020) enabled by online platforms, the use of memes has become an important component of online cultures across the globe. In this context, the concept of a meme was redefined by Shifman (2013, p. 177) as a "group of digital content units sharing common characteristics of content, form, and/or stance" which are created with awareness of each other and circulated by Internet users. Since then, multiple typologies of memes have been proposed based on their visual formats (e.g., memes based on still images; Shifman, 2014) or the interpreted functionality of the memes (e.g., memes reinforcing or challenging a particular narrative; Makhortykh, 2015). However, the majority of Internet memes to date still follow a conventional format of an image accompanied by a short text that offers additional cues for interpreting the meme's meaning.

The rise in popularity of Internet memes is due, among other factors, to their versatility, which allows Internet users to utilize memes for communicating different ideas and adapting them to any situation, no matter how dramatic or banal it might be. Memes have become a common format for making a societal commentary on a wide range of topics (Seiffert-Brockmann, Diehl & Dobusch, 2018), particularly as memes are polysemous and open to diverse interpretations (Paz-Rebollo, Mayagoitia-Soria & González-Aguilar, 2021). Likewise, the ability of memes to provoke strong emotional reactions (Shifman, 2014) and their capacity to remix different cultural texts (Wutz & Nugteren, 2018) makes them an indispensable element of online communication practices as well as a vital component of the "vibrant remix culture" (Xu et al., 2016, p. 106) which flourishes in online environments.

The new possibilities for producing and disseminating online content, however, contributed not only to the rise of Internet memes but also facilitated the dissemination of hate speech (Mathew et al., 2019). In some cases, memes become part of cyberhate campaigns (Castaño-Pulgarín et al., 2018), which can be defined as the use of violent, aggressive, or offensive language to target a particular group in digital environments. Given their versatility, easily comprehensible nature, and virality, memes turned out to be an effective means of conveying hate messages, promoting extremist messages for new audiences and new followers, and justifying bigotries (Bogerts & Fielitz, 2019; Lamerichs et al., 2018). Furthermore, the humorous nature of memes which contributed to them being commonly viewed as a form of entertainment (Burgess, 2008), can facilitate the use of memes for spreading hate messages and distorting the public discourse (Schwarzenegger & Wagner, 2018) as well as bypassing censorship and normalizing the extremist messages (Bogerts & Fielitz, 2019).

The frequent appropriation of Internet memes for disseminating hate speech is of particular importance at the time of crises, in particular considering the essential role played by digital media in mobilizing public support and shaping how the causes and solutions for these crises are understood by the public (see, for example, Tufekci and Wilson, 2012; Tufekci, 2013; Papacharissi, 2014). In the case of armed conflicts, in particular, memes become essential elements of what Rodley (2016) defines as “viral agitprop,” namely the instrumentalization of online creative practices as a form of digital propaganda. In addition to enabling additional possibilities for manipulating public opinion for political gains, memes become an important component of the process of representing and interpreting ongoing warfare, for instance, by giving voice to civilians caught in the middle of the conflict (Wolf, 2015) or creating an opportunity for the combatants to sustain connections to their popular culture (Silvestri, 2016). However, in many cases, the affective potential of memes is used primarily to degenerate or even dehumanize the groups perceived as enemies (e.g., Gaufman, 2015; 2022).

3. Methodology

To conduct the study, we collected data from Telegram, a popular messaging service. Our decision to focus on memes coming from Telegram is attributed to several reasons. First, the platform plays an integral role in mediating the ongoing Russian-Ukrainian war by turning into a digital battleground and serving as "an instrumental tool for both governments and a hub of information for citizens" (Bergengruen, 2022) in Ukraine and Russia. Second, due to minimal moderation, in particular when compared to mainstream Western platforms, such as Facebook or Twitter, and its strong focus on user privacy, Telegram has a reputation as a censorship-free medium that contributes to its use for public mobilization (e.g., by anti-authoritarian protest movements; Urman et al., 2021; Wijermars & Lokot, 2022), but also for spreading hate speech by extremist groups (Rogers, 2020; Urman & Katz, 2022).

Public communication on Telegram is organized through channels, where channel owners publish messages and can enable users to react to them via emoticons or text messages (in the case the owner enabled a user chat). To find memes related to the Russian-Ukrainian war, we looked at a selection of Russophone Telegram channels. Our selection was guided by two factors: first, we were interested in channels that are actively involved in the mediation of the war (i.e., regularly posting content related to the war) and attracted a large number of followers. Second, we wanted to select channels associated with different groups of pro-Russian Telegram actors, such as political bloggers, individual combatants, and war journalists.

Our selection of Telegram channels consists of four channels: 1) @boris_rozhin, the channel of the political blogger Boris Rhozin (also known as Colonelcassad) with 817 thousand subscribers (here and for the following channels the data on subscribers is given as of January 2023); 2) @Topaz_Govorit, the channel of a former pro-Russian combatant and a member of the right-wing Rusich group, Eugeny Rasskazov, with 41 thousand subscribers; 3) @SergeyKolyasnikov, the channel of the political blogger, Sergey Kolyasnikov (also known as Zergulio) with 236 thousand subscribers; and 4) @vladlentatarsky, the channel of the war journalist, Maksim Fomin, with 523 thousand subscribers. Because of the large volume of data published in the selected channels, we looked at their content for a fixed period: from 2 December 2022 to 2 January 2023.

The final sample of collected memes is comprised of 115 items retrieved from the four Telegram channels listed above. To analyze data, we relied on intertextual discourse analysis. Specifically, we were interested in what groups of memes can be identified based on their perceived functionality. Based on our examination, we identified three main groups of memes, which we will discuss in more detail in the Findings section below: 1) memes focusing on spreading hate speech in the form of messages attacking or degrading large groups of people; 2) memes amplifying personal attacks directed against individuals or small groups of people; and 3) memes focused on glorifying the Russian army and its officials.

Finally, it is important to mention some limitations of our methodology, which must be considered when examining our findings below. First, we relied on a relatively small sample of Russian Telegram channels and examined their use of memes for a relatively short period of time. In future research, it will be important to look at a broader range of actors, including anti-war Russian and Ukrainian channels. Second, for this study, we focused on the use of memes by channel owners but did not look at their use by channel audiences (e.g., the users discussing channel posts in the chats). Future research will benefit from examining how memes can be used in chat-based communication and whether their selection differs from the ones used by channel owners. Third, we focused on three of the most common functions of Internet memes in the sample, whereas some of the more niche functions were not examined in this chapter.

4. Findings

Dissemination of hate speech

The first category of memes is constituted by the content disseminating messages which attack large groups of the population, usually Ukrainians. The exact selection criteria of people who are attacked vary from the broad national/ethnic groups, such as Ukrainians presented as pigs (Figure 1a), to the supporters of specific ideology (e.g., the presumed supporters of

Stepan Bandera, a WWII Ukrainian nationalist group leader; Figure 2a) to the professional communities, such as the members of the Ukrainian army (Figure 2b).

Unlike many other forms of hate speech, messages communicated via Internet memes we examined in most cases did not call for violence directly. Instead, they often focused on the dehumanization of the targeted group, thus potentially aiming to justify or normalize the violence and discrimination. To achieve this, many memes presented Ukrainians as non-human beings, usually animals or insects. Two such memes are shown in Figure 1: Figure 1a shows an anthropomorphic pig with the traditional Ukrainian haircut (i.e., the so-called oseledets) forcefully countered by the human arm associated with a conspiratorial pro-war Russian media outlet, whereas Figure 1b combines images of pigs with an image of a Colorado potato beetle with a signature made of words "khohol" (a derogatory Russian term denoting Ukrainians) and "fascist."

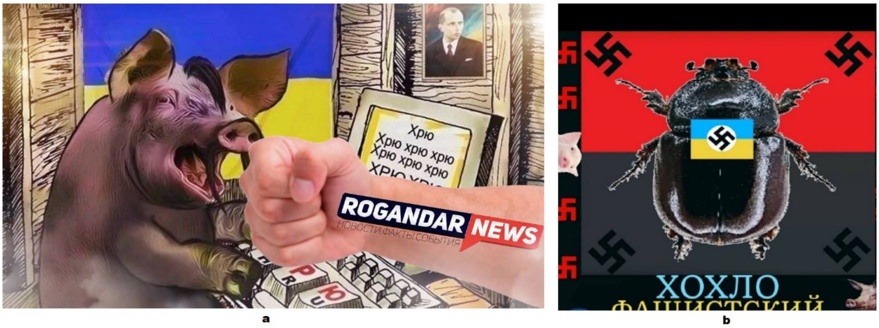


Figure 1 – Memes disseminating hate speech (animal-/insect-based dehumanization).

The memes similar to the one shown in Figure 1 illustrate two important aspects which differentiate the use of memes for spreading hate speech during the large-scale Russian invasion in 2022 from the earlier stages of the Russian-Ukrainian war. The first aspect concerns the different hate speech strategies used in 2022: during the earlier periods (e.g., in 2014, during the first hot phase of the war), the use of hate speech relied less on dehumanization and more on derogation of Russia's opponents, mainly by drawing comparisons between them and certain groups treated as marginal in the Russian society (e.g., LGBTQ; Gaufman, 2015, 2022) or by applying negative historical tropes (e.g., Nazi or Latin American juntas; Gaufman, 2015; Makhortykh, 2018).

Several memes in the sample we collected still utilize the above-mentioned derogation strategies. Figure 2b shows one meme which draws parallels between Nazi Germany's soldiers and the Ukrainian soldiers, utilizing the strong negative image of Nazis embedded in the Russian collective memory through the Great Patriotic War narrative to stigmatize Ukrainians serving in the army. A similar strategy is used in Figure 2a, which combines religious and historical tropes by showing Stepan Bandera in the infernal cauldron and wishing his supporters "to burn in Hell as your hero." However, such memes were relatively few compared to those focusing on animal-based dehumanization.

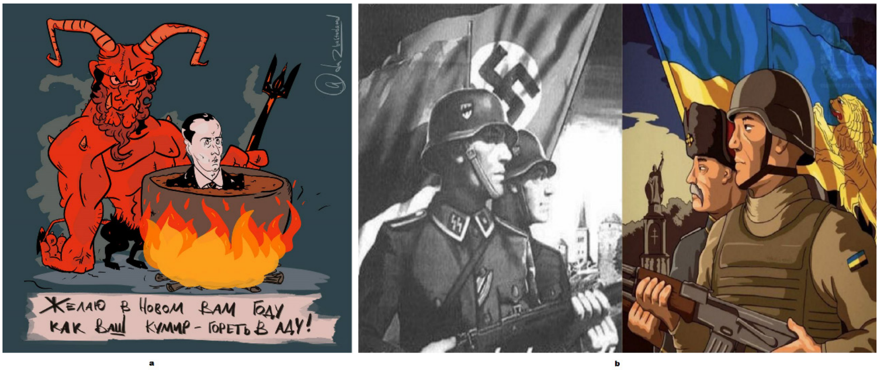


Figure 2 – Memes disseminating hate speech (derogation based on historical references).

The second aspect differentiating the use of memes for spreading hate speech in 2022 is their broader target groups. In earlier periods of the war, the targets of hate speech were often similar to the groups targeted by memes in Figure 2: namely, the members of the Ukrainian military and the alleged supporters of Ukrainian nationalistic movements. However, with the beginning of the large-scale Russian invasion in 2022, the target group of many meme-based hate speech messages expanded to all Ukrainians who are dehumanized. Such a shift is particularly troubling considering that the Russian military campaign against Ukraine might also involve actions that can be interpreted as genocide (see, for instance, Hook, 2022), and dehumanization is an important factor facilitating it (Ndahinda & Mugabe, 2022).

Amplification of personal attacks

Similar to memes used to disseminate hate speech, memes amplifying personal attacks are constituted by content items that aim to offend or derogate. However, the difference between the two groups of memes is that memes amplifying personal attacks focus on individuals (for instance, Western/Ukrainian politicians or oppositional Russian journalists) or small groups (for instance, Russian anti-regime media outlets such as Dozhd or Meduza).

The main aim of such memes usually involves establishing or amplifying a negative image of the individual or a small group of targeted individuals. Similar to some memes used to disseminate hate speech which we discussed above, this aim is usually achieved by attributing to targeted individuals/groups characteristics associated with the LGBTQ community. Such a choice aligns with the intensive anti-LGBTQ campaigns in pro-Kremlin Russian media, which became particularly common in recent years and aimed to demonize the LGBTQ community and justify legal repressions against LGBTQ representatives in Russia, reinforcing the hegemonic masculinity discourse promoted by the Kremlin (Legeido, 2022).

Two typical memes of such type are shown in Figure 3. Both build on the meme showing an old White man kicking a rooster. In the Russian criminal culture, the rooster is a symbol of homosexuality associated with prisoners who were converted by force into passive homosexuals (Yusupova, 2015). Figure 3a completely recreates the older meme but just adds the title "Dozhd" to the image of the rooster to present the Russian oppositional journalistic outlet as a group of homosexuals. Figure 3b remixes the meme in a slightly different format to replace the image of a rooster with the image of a Georgian-Russian singer, Valery Meladze, who caused an outrage among the supporters of the Kremlin after voicing a pro-Ukrainian slogan, "Glory to Ukraine," during a stage performance in Dubai. The statement attributed to the kicking man - "Farewell, gypsy Valera" - adds racist meaning to the scene by suggesting that both homosexuality and being a Roma are derogatory characteristics.

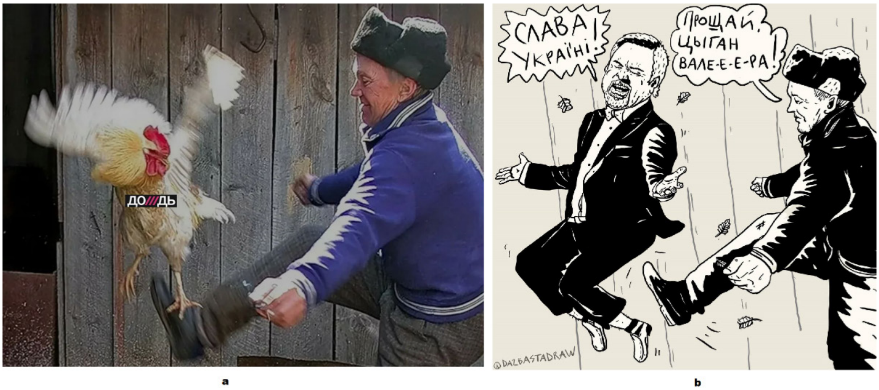


Figure 3 – Memes amplifying personal attacks (derogation based on LGBTQ references).

Another common strategy for amplifying personal attacks observed in the examined sample of memes is the addition of characteristics connecting the targeted individual or groups with the negative historical tropes, in particular Nazi Germany. Figure 4 shows two examples of such memes: Figure 4a contrasts the photo of crying Natalya Sindeyeva, a chief executive officer of oppositional journalistic media Dozhd, with the image of German soldiers from the war movie titled "The Bridge"; by doing so, it effectively equals Russian oppositional journalists with potential war criminals. A similar approach is used by the meme in Figure 4b, where an image of a Wehrmacht hat is added to the photo of a Polish official, Mateusz Morawiecki.

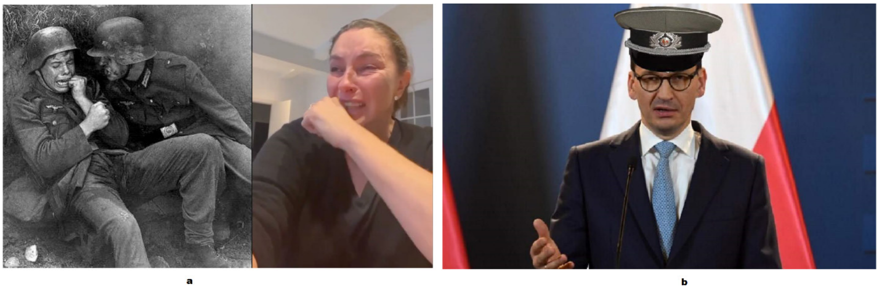


Figure 4 – Memes amplifying personal attacks (derogation based on historical references).

In addition to these two most common approaches for amplifying personal attacks, some memes also used other strategies. Some of them relied on racist tropes: for instance, by showing Ukrainian president Zelensky as an Indigenous person with the assumption that such an image is derogatory due to Indigenous people being a symbol of non-civilized groups, in particular when compared with Russian or Western people. Other memes relied on the combination of visual and verbal satire to amplify personal attacks against Russian anti-Kremlin opposition leaders, such as Alexei Navalny or Ilya Yashin, for instance, by adding counterfactual claims to their photos.

Glorification of the Russian army

In contrast to the other two groups of memes, memetic content devoted to glorifying the Russian army does not aim to offend or derogate. Instead, these memes focus on promoting a positive image of the Russian army in general and a few selected military officials, in particular general Sergey Surovikin, who in October 2022 was assigned as a commander of the Russian forces in Ukraine. Consequently, the virality of memes in this context is primarily used to amplify the message about the superiority of the Russian army or, in some cases, mobilize public support in relation to the invasion.

The two common memes belonging to this group are shown in Figure 5. Figure 5a shows a collage of two images, the one of Ukrainian president Zelensky claiming that there will be no fireworks in Kyiv on New Year's eve (on the top) and the one of Russian general Surovikin stating that the fireworks will happen (on the bottom). The meme not only promotes the image of the Russian army as a proactive force (despite losing the strategic initiative at the time when the memes were produced) but also emphasizes the masculinity of Surovikin, presented as a physically impressive man of few words. Other similar memes praised Surovikin's campaign against the civil infrastructure in Ukraine, humorously referring to him as the leading environmental activist of 2022 who decreased Ukraine's carbon output or calling him the Blackout General.

Figure 5b adopts a classic Anakin and Padme 4-part meme, with Anakin claiming that each Russian tank army has to include a mixed air division and a brigade of military aviation, with Padme stating that there is only one tank army and then starting doubting it. In this case, the purpose of the meme is to use humor to promote the idea of Russian military superiority, which is far beyond what is publicly known.

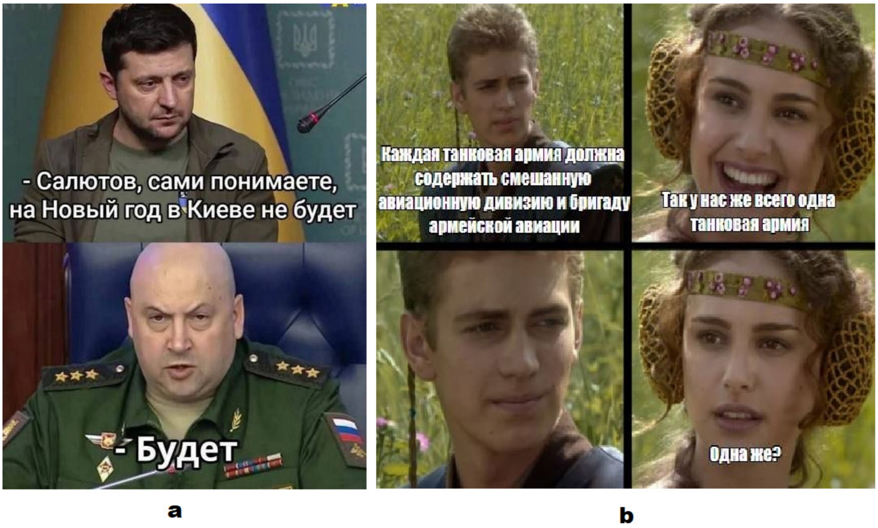

- Салютов, сами понимаете, на Новый год в Киеве не будет
- Будет
a
Каждая танковая армия должна содержать смешанную авиационную дивизию и бригаду армейской авиации
Так у нас же всего одна танковая армия
Одна же?
b

Figure 5 – Memes glorifying the Russian army (humor-based).

5. Discussion

In this chapter, we examined the relationship between Internet memes and hate speech in the context of the war in Ukraine. Our examination shows that memes serve as one of the formats through which Russophone Telegram channels propagate hate speech. Using offensive humor to facilitate the dissemination of hateful messages, these memes primarily focus on dehumanizing Ukrainians by presenting them as less-than-human beings (e.g., pigs or insects). Such a strong emphasis on dehumanization contrasts with earlier periods of the Russian-Ukrainian war when offensive online content (including memes) usually focused on stigmatizing specific groups of the Ukrainian population by attributing to them characteristics associated with the LGBTQ community and negative historical tropes (for example, see Gaufman, 2015, 2022; Makhortykh, 2018).

At the same time, it is important to note that memes disseminating hate speech and focusing on dehumanization constituted only a relatively small part of the memes we collected. While the limited use of memes for propagating and normalizing hate speech can be viewed as a positive sign, it is likely attributed to the lack of need for such normalization. In contrast to Western extremist groups aiming to mask hate speech as a form of dark Internet humor (e.g., Lamerichs et al., 2018), many pro-Russian channels we examined rather openly promote hate speech by calling for murdering Ukrainians resisting Russian aggression. Under these circumstances, there is little motivation to normalize hate speech with the help of memes, and the major benefit they can bring is virality.

Regardless of the reasons for the relatively limited presence of memes promoting hate speech, our examination shows that there are other functions played by memes in the context of Russia's war in Ukraine. Many memes we examined are used to amplify personal attacks, particularly against individuals known for their anti-Kremlin stance (e.g., Western and Ukrainian politicians, Russian journalists, and anti-war public figures). Such amplification usually relies on attribution to these individuals characteristics associated with groups that are targeted and stigmatized by the Kremlin (e.g., LGBTQ community) as well as references to the negative historical tropes, in particular the ones related to Nazi Germany. Other memes, by contrast, focused on glorifying the Russian army and its officials by using humor to stress the superiority of the Russian forces and mobilize public support in relation to the Russian invasion in Ukraine.

Together, these observations prompt the importance of a multi-layered assessment of the relationship between Internet memes and hate speech. Specifically, it is important to account that the role of Internet memes in the context of armed conflicts goes beyond propagating hate speech and acknowledge that some of these non-hate speech-related uses of memes can can still promote violence. For instance, memes glorifying the Russian army do not attack any specific group directly, but indirectly they do justify Russian war crimes by turning them into a source of entertainment and, potentially, making the audience more accepting of memes dehumanizing other ethnic or national groups. This complexity is important to acknowledge both when conceptualizing the role of memes - as well as other forms of user-generated content - in the propagation of hate speech in online environments as well as looking for ways to counter the spread of hate speech across digital platforms.

References

Bergengruen, V. (2022, March 21). How Telegram Became the Digital Battlefield in the Russia-Ukraine War. *Time*. https://time.com/6158437/telegram-russia-ukraine-information-war/

Bilewicz, M. & Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology*, 41, 3-33.

Bogerts, L. & Fielitz, M. (2019). “Do you want meme war?”: Understanding the visual memes of the German far right. In M. Fielitz & N. Thurston (Eds), *Post-Digital Cultures of the Far Right: Online Actions and Offline Consequences* (pp. 137-153). Transcript.

Bozkuş, Ş. B. (2016). Pop polyvocality and Internet memes: As a reflection of socio-political discourse of Turkish youth in social media. *Global Media Journal: Turkish Edition*, 6(12), 44-74.

Bruns, A. (2008). *Blogs, Wikipedia, Second Life, and beyond: From production to produsage*. Peter Lang.

Burgess, J. (2014). 'All your chocolate rain are belonging to us?': Viral video, YouTube and the dynamics of participatory culture. N. Papastergiadis & V. Lynn Eds), *Art in the Global Present* (pp. 86-96). UTS ePRESS.

Castaño-Pulgarín, S. A., Suárez-Betancur, N., Vega, L. M. T. & López, H. M. H. (2021). Internet, social media and online hate speech. Systematic review. *Aggression and Violent Behavior*, 58, 1-7.

Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.

ECRI. (2016). *ECRI General policy recommendation No. 15. On combating hate speech*. Council of Europe.

Fortuna, P. & Nunes, S. (2018). A survey on automatic detection of hate speech in text. *ACM Computing Surveys*, 51(4), 1-30.

Gagliardone, I., Gal, D., Alves, T. & Martinez, G. (2015). *Countering Online Hate Speech*. Unesco Publishing.

Gaufman, E. (2015). World War II 2.0: Digital memory of fascism in Russia in the aftermath of Euromaidan in Ukraine. *Journal of Regional Security*, 10(1), 17-35.

Gaufman, E. (2022). Damsels in distress: Fragile masculinity in digital war. *Media, War & Conflict* (online first). https://doi.org/10.1177/17506352221130271

González-Aguilar, J. M. & Makhortykh, M. (2022). Laughing to forget or to remember? Anne Frank memes and mediatization of Holocaust memory. *Media, Culture & Society,* 44(7), 1307-1329.

Hook, K. (2022, July 28). Why Russia's War in Ukraine Is a Genocide: Not Just a Land Grab, but a Bid to Expunge a Nation. *Foreign Affairs*. https://www.foreignaffairs.com/ukraine/why-russias-war-ukraine-genocide

ISD. (2022, October 26). A false picture for many audiences: How Russian-language pro-Kremlin Telegram channels spread propaganda and disinformation about refugees from Ukraine. *ISD*. https://www.isdglobal.org/digital_dispatches/a-false-picture-for-many-audiences-how-russian-language-pro-kremlin-telegram-channels-spread-propaganda-and-disinformation-about-refugees-from-ukraine/

Knobel, M. & Lankshear, C. (2007). Online memes, affinities, and cultural production. In M. Knobel & C. Lankshear (Eds.) *A New Literacies Sampler* (pp. 199-227). Peter Lang.

Lamerichs, N., Nguyen, D., Melguizo, M. C. P., Radojevic, R. & Lange-Böhmer, A. (2018). Elite male bodies: The circulation of alt-Right memes and the framing of politicians on Social Media. *Participations*, 15(1), 180-206.

Legeido, V. (2022, October 21). 'There are no homosexuals in this country' How Putin's embrace of homophobia echoes dictators of the past. *Meduza*. https://meduza.io/en/feature/2022/10/21/there-are-no-homosexuals-in-this-country

Makhortykh, M. (2015). Everything for the lulz: Historical memes and World War II memory on Lurkomor'e. *Digital Icons: Studies in Russian, Eurasian and Central European New Media*, 13, 63-90.

Makhortykh, M. (2018). #NoKievNazi: Social Media, Historical Memory and Securitization in the Ukraine Crisis. In V. Apryshchenko & V. Strukov (Eds.) *Memory and Securitization in Contemporary Europe* (pp. 219-247). Palgrave Macmillan.

Makhortykh, M. & González-Aguilar, J. M. (2020). Memory, politics and emotions: Internet memes and protests in Venezuela and Ukraine. *Continuum*, 34(3), 342-362.

Mathew, B., Dutt, R., Goyal, P. & Mukherjee, A. (2019). Spread of hate speech in online social media. In *Proceedings of the 10th ACM Conference on Web Science* (pp. 173-182). ACM Press.

Nayak, A. & Agrawal, A. (2022). Detection of hate speech in Social media memes: A comparative Analysis. In *Proceedings of the Third International Conference on Intelligent Computing Instrumentation and Control Technologies* (pp. 1179-1185). IEEE.

Ndahinda, F. M. & Mugabe, A. S. (2022). Streaming Hate: Exploring the Harm of Anti-Banyamulenge and Anti-Tutsi Hate Speech on Congolese Social Media. *Journal of Genocide Research* (online first). doi: https://doi.org/10.1080/14623528.2022.2078578

Papacharissi, Z. (2015). *Affective Publics: Sentiment, Technology, and Politics*. Oxford University Press.

Paz-Rebollo, M. A., Mayagoitia-Soria, A. & González-Aguilar, J. M. (2021). From polarization to hate: Portrait of the Spanish political meme. *Social Media+ Society*, 7(4), 1-15.

Rodley, C. (2016). FCJ-200 When Memes Go to War: Viral Propaganda in the 2014 Gaza-Israel Conflict. *The Fibreculture Journal*, 27. https://twentyseven.fibreculturejournal.org/2016/03/18/fcj-200-when-memes-go-to-war-viral-propaganda-in-the-2014-gaza-israel-conflict/

Rogers, R. (2020). Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. *European Journal of Communication*, 35(3), 213-229.

Seiffert-Brockmann, J., Diehl, T. & Dobusch, L. (2018). Memes as games: The evolution of a digital discourse online. *New Media & Society*, 20(8), 2862-2879.

Schabas, W. A. (2017). Hate speech in Rwanda: The road to genocide. In E. Lattimer (Ed.) *Genocide and Human Rights* (pp. 231-261). Routledge.

Shifman, L. (2013). *Memes in Digital Culture*. The MIT Press.

Shifman, L. (2014). The cultural logic of photo-based meme genres. *Journal of Visual Culture*, 13(3), 340-358.

Silvestri, L. (2016). Mortars and memes: Participating in pop culture from a war zone. *Media, War & Conflict*, 9(1), 27-42.

Steir-Livny, L. (2016). Is it OK to laugh about it yet? Hitler Rants YouTube parodies in Hebrew. *The European Journal of Humour Research*, 4(4), 105-121.

Schwarzenegger, C. & Wagner, A. J. (2018). Can it be hate if it is fun? Discursive ensembles of hatred and laughter in extreme right satire on Facebook. *Studies in Communication and Media*, 7(4), 473-498

Tufekci, Z. (2013). “Not this one” social movements, the attention economy, and microcelebrity networked activism. *American Behavioral Scientist*, 57(7), 848-870.

Tufekci, Z. & Wilson, C. (2012). Social media and the decision to participate in political protest: Observations from Tahrir Square. *Journal of Communication*, 62(2), 363-379.

Urman, A. & Katz, S. (2022). What they do in the shadows: examining the far-right networks on Telegram. *Information, Communication & Society*, 25(7), 904-923.

Urman, A., Ho, J. C. T. & Katz, S. (2021). Analyzing protest mobilization on telegram: The case of 2019 anti-extradition bill movement in Hong Kong. *PLOS ONE*, 16(10), 1-21.

Weimann, G. & Masri, N. (2020). Research note: spreading hate on TikTok. *Studies in Conflict & Terrorism*, 1-14. https://doi.org/10.1080/1057610X.2020.1780027

Wijermars, M. & Lokot, T. (2022). Is Telegram a “harbinger of freedom”? The performance, practices, and perception of platforms as political actors in authoritarian states. *Post-Soviet Affairs*, *38*(1-2), 125-145.

Wiggins, B. E. (2016). Crimea River: Directionality in memes from the Russia-Ukraine conflict. *International Journal of Communication*, 10, 451-485.

Wolf, H. (2015). Paper is patient’: tweets from the ‘# AnneFrank of Palestine’. *Textual Practice*, 29(7), 1355-1374.

Woods, F. A. & Ruscher, J. B. (2021). Viral sticks, virtual stones: addressing anonymous hate speech online. *Patterns of Prejudice*, 55(3), 265-289.

Wutz, I. & Nugteren, L. (2018, April 30). Brexit and Online Political Activism. On Vox Populism, Slacktivism and Online Intertextuality. Diggit Magazine. https://www.diggitmagazine.com/papers/brexit-and-online-political-activism

Xu, W. W., Park, J. Y., Kim, J. Y. & Park, H. W. (2016). Networked cultural diffusion and creation on YouTube: An analysis of YouTube memes. *Journal of Broadcasting & Electronic Media*, 60(1), 104-122.

Yusupova, M. (2015). Masculinity, Criminality, and Russian Men. *Sextures: E-journal for Sexualities, Cultures, and Politics*, 3(3), 46-61.

Chapter 3

SYRIAN REFUGEES IN THE SHADE OF THE 'ANTI-SYRIANS' DISCOURSE: EXPLORING DISCRIMINATORY DISCURSIVE STRATEGIES ON TWITTER

Özlem Alikılıç
Yaşar University, Türkiye

Ebru Gökaliler
Yaşar University, Türkiye

İnanç Alikılıç
Malatya Turgut Özal University, Türkiye

1. Introduction

Negative judgments and stereotype mindscapes regarding the 'others' may direct the people to hatred-themed conversations and even hate crimes. And for this reason, this hate speech is a significant factor separating a specific group (others) from the remaining part of the society (us). Today, the inequality in power relationships and nationalist trends handles the formation and aggravation of hate speech and ethnic, socio-cultural, religious, and economic conflicts.

As the foreigners and immigrants fall into different ethnical groups and as minorities in the societies they step in, they are experiencing hate speeches directed by the public of the unknown country. The 'foreigners', who try to adapt to the society, later on, remain out of the concept of 'us' (Bilge, 2016). The refugee communities frequently experience facts such as alienation and hate speech at the places they migrate to. Syrian refugees are also being called ‘refugees’ because of the forced migration based on the civil war in their country, and they are being tagged as ‘foreigners’ in other societies, and they remain beyond the concept of ‘us’.

The participatory construct of social media allows everyone to reflect their opinions and express their opinions within the community (Kirschenbaum, 2004). But while this freedom of expression is gradually transforming to a worrisome state because of the unrestrained content power of social media, the freedom of expression is giving its place to hate speech and hate crime. As a result, the individuals, groups, and societies are deeming themselves superior to the ‘others’ in the sense of nationality, race, religion, or culture, and as a result, they are humiliating and despising the ‘others’ whom they consider as ones not being one of them (Mihajlova, Bacovska & Shekerdjıev, 2013). And the hatred, developing and being generated through social media, is gradually being naturalized, preventing the individual from facing his hatred and cause. Due to Twitter, immigration also has had its share of the hatred contents that can be generated easily.

The examination of interaction observed in social networks regarding Syrian refugees provides the social scientists a new research ground in terms of both subject and method. Twitter, where the providers’ opinions and interactions can be observed relatively easily, provides the opportunity for alternative and rich data to measure the perceptions and attitudes of individuals on different subjects. Twitter, where the mission is not ‘keeping in contact with friends like the other social media platforms, is mainly a media where the individuals, elite agenda identifiers and media actors (Erickson and Lilleker, 2014) share their instant news flows with each other (Verweij, 2012).

This study classifies the posts regarding Syrians in Turkey as per their hatred content based on discriminatory discursive strategies and searches whether it can include new strategies in the current typology. 245,587 tweets posted under the hashtags of ‘#syrian, #refugee, #syrian refugee, #wedontwantthesyrians, and #syrianspissoff’ were analyzed. Critical discourse analysis (CDA) was applied to the collected contents. This research tried to discover mainly under which categories the discourses directed to refugees develop. This study was determined mainly under which typology the discourses regarding the refugees gather, and met with surprising results. Especially the hatred against the Syrian refugees was observed within criticisms against Turkish government.

2. The refugees problem in the framework of online hate speech

Because of the civil war in Syria, which is one of the most important domestic and foreign policy problems of the entire world and Turkey since 2011 (BBC News, 2018), Turkey has embraced the asylum requests of millions of Syrian refugees fleeing from the war (Asaf, 2017). And this circumstance has caused the increase of hate speech, especially against the Muslim immigrants (Soral, Bilewicz & Winiewski, 2018). Millions of refugees, the negative effect of the civil war on the country's borders and public, and problems in a socio-cultural adaptation of refugees to the country have caused to the frequent observance of hatred contents in comments made against news on Syrian refugees (Memişoğlu & Ilgıt, 2017; Öztürk & Ayvaz, 2018; Yıldız, 2018).

The number of Syrians living in Turkey under temporary protection status is 3 million 635 thousand 410 as per the official declaration of the Ministry of the Interior Directorate General of Migration Management of 25 July 2020 (Directorate General of Migration Management, 2020). Along with the uncertainty of civil war in Syria, the refugees who are still coming from Syria, and the increasing population of the Syrians in the country, have now lived in the cities besides the camps.

In Turkey, the gradually increasing population of Syrian refugees, the current economic crisis, tensions in foreign policy, and political polarization are increasing the xenophobia and hate speech against the refugees day by day. Gorodzeisky and Semyonov (2019) verified this circumstance by drawing attention to subjects such as perceived economic threat, increased crime rates, perceived cultural threat in terms of the formation of negative emotions against the immigrants, and especially negative emotions attitudes against specific immigrant groups. Likewise, in research made in Turkey in 2019, it had been determined that news had been made on the subjects of the potential criminal state of Syrians because of crimes of murder, theft, abuse, and on social security concerns such as unemployment and on the threat against the demographical structure of Turkey (Media Watch On Hate Speech Report, 2019, p. 9).

In Turkey, the dehumanization of Syrian refugees and a sign of them as causes of the continuing economic problems and increase of unemployment are increasingly taking place on social media (Duran, 2019). Along with the developing technology, the online environment has become a significant platform for activism, politicization, and mobilization. Especially many social networks such as Twitter, Instagram, and Facebook have allowed each user in the world to generate content in an uncontrolled manner and to develop new discussions and actions by sharing them. On this, various research on hate speech against the refugees determines that hate speech against refugees is frequently finding a place on both traditional media and social media (Öztürk & Işık, 2020; Akşak, 2019; Soral et al., 2018; Media Watch On Hate Speech Report, 2019; Sayımer & Rabenda, 2017; Aslan, 2017; Kuş, 2016).

From the perspective of refugees and immigrants, while the individuals are becoming familiar with derogatory discourses and clichés about the limited interaction between the hosting societies and refugee groups, fear of threat and uneasiness may occur within the hosting society. And the minority groups, which are expressed as 'outgroup' and that try to enter the living space of the hosting group, are vulnerable in this process (Sulaiman-Hill et al., 2011). The Syrian refugees, who are being isolated by society, have remained outside the concept of 'us' (Yıldız, 2018).

In Turkey, it is interesting that this circumstance is standing far from the centerline of religion. The religion of both populations being Islam, has not prevented the marginalization of Syrian Muslims in a Muslim society such as Turkey. In Turkey, it is being observed that digital lynch attempts are being started in case of any incidence and that such digital lynches are turning to acts and hitting the streets at most places. For instance, when the videos regarding the Syrians dancing, chanting, and waving Syrian flags at Taksim Square of Istanbul on the new year night of 2019 had been shared on Twitter, hate contents had increased in the whole country, and it had caused conflicts and personal injuries at different cities and had caused bringing damage to the real estates of Syrians (Öztürk & Işık, 2020; Yeni Çağ, 2019; Haberler.com, 2019; EnKocaeli.com, 2019). This circumstance shows how significant online platforms and social networks, especially Twitter, have in the transition of hate speech to hatred acts.

Studies in Turkey also draw attention to the high frequency of hate speeches in online environments. Kuş (2016) found that the user comments on Facebook are bearing negative emotions against the refugees, Yıldız (2018) found that more hate speech is being generated on Twitter compared to mainstream media, and Yazıcı (2016) noted that hate speech regarding the refugees is being generated on the E-dictionary website. Aslan (2017) noted that the Syrian refugees are stigmatized with a negative discriminatory hate tone by phrases such as 'traitor,' 'potential threat,' cause of the country's economic problem' on YouTube. Öztürk and Işık (2020) had also performed discourse analysis on Twitter, and they had found that the discourses are mainly tending to nationalist and marginalizing dimensions. The findings of Çoban (2016) showed that the refugees are being conceptualized on the news with metaphors such as 'flow, flood, wave', and in the same manner that they are being shown as an economic load as being defined with the metaphors of the economy (cost, expense, invoice, expenditure).

It is drawing attention that similar findings are also matching in the world (Rettberg & Radhika, 2016; Bhatia & Jenks, 2018; Kreis, 2017; Kalav & Fırat, 2017). They observed that social media, and especially Twitter could become the center point where hate speech is being generated and spread, that Twitter is creating a suitable ground for the spread of hate speeches, and that it is also facilitating the supporter gathering activities of the extreme right, xenophobic organizations being fed by extreme nationalism.

3. Hate speech and discriminatory discursive strategy are reshaping in Twitter

Discourse analysis is frequently being preferred in the process of analysis of both conventional and digital contents. Considering the factors making up the basis of critical discourse analysis such as 'hyper intertextuality' (Reisigl & Wodak, 2001, p. 185), 'interdiscursivity' (Reisigl & Wodak, 2001, p. 37), 're contextualization' (Bernstein, 1990, p. 183-184), and repeating the same words with creative discourses' (Van Leeuwen & Wodak, 1999, p. 96; Richardson & Wodak, 2009; Wodak & Fairclough, 2010, p. 24), suitability of Twitter for discourse analysis can be indicated. Furthermore, Twitter's characteristics such as facilitation of the reconstruction of discourse, instantaneous re-contextualization, ease of reinterpretation of social practices (Van Leeuwen & Wodak, 1999, p. 96), and of course, justification by retweet may also be considered.

Twitter is an effective field of research for social scientists regarding analysis regarding content and discourse. CDA has also been used in many research. In his study, Wodak (2015) revealed five discourse strategies classified as 'referential, predication, argumentation, perspectivization, and mitigation' by the major categories of 'positive self-representation, and 'negative other representation'. Van Dijk had developed discourse analysis codes supporting the opinion of Blumer (1958) to draw attention to conflicts among groups arising from biases and in terms of disturbance, tension, and formation of incorrect thoughts against the ones not being from one's side. Van Dijk (2011) evaluates the discourse strategies as 'in-group and out-group within a semantic construct, and specifies that the in-groups are being positively represented, and out-groups are negatively represented socially. Chen and Flowerdew (2019), analyzing the comments of YouTube videos on 'Umbrella Movement' in Hong Kong, found 4 main categories among discriminatory discursive strategies as being 'negating the others, frightening tactics, accusing the victim, and illegitimacy'. The studies showed that the importance of CDA in the new media order increased more than its study of traditional media that had mainly been considered since today. Among the causes of this is, of course, the fact that hatred can circulate and act for 7/24 over the Internet and social media.

Another reason is that social media – especially Twitter, where opinions and interactions can be observed relatively easily provide the opportunity to get alternative and rich data to measure the perceptions and attitudes of individuals on different subjects. Besides such conveniences provided by Twitter, it has some disadvantages compared to conventional methods. One of these is the fact that the contents on Twitter don't just comprise texts. Mobile visuals such as emojis, photographs, GIFs are obstructing the performance of automated content-coding (Hatipoğlu et al., 2019, p. 190). Such interferences influence the result of the research (Grimmer & Gary, 2011). In recent years, social media research, which drew attention to hate speech, has been directed to Twitter because of its technical infrastructure facilities it is providing in data retrieval. Öztürk and Ayvaz (2018) actualized by using modern analysis methods, had also drawn using modern analysis methods, and drew attention by their tweets examining Syrian refugees via the Neuro-Linguistic Programming method.

4. Methodology

This study searches whether the contents generated on social media in Turkey regarding Syrian refugees bear hate speech or not to classify the posts containing hate speech based on discriminatory discursive strategies (Flowerdew et.al., 2002, pp. 329-341; Chen & Flowerdew, 2019), and to check if the current classification operates in different societies, and to discover if different strategies had developed. The research questions are: (RQ1) Are there posts on Twitter containing hate speech against Syrian refugees? (RQ2) When these posts are examined, is there a difference between the classification got and the classification made by Flowerdew et al. (2002)

based on discriminatory discursive strategies? (RQ3) What are the new categories (theme), or sub-categories determined, if they exist?

CDA is an analysis method, which analyzes text and speech (Bennett, 2015; Krzyzanowski, 2010; Reisigl & Wodak, 2001). While Reisigl and Wodak (2001) say that it comprises a three staged analysis, Krzyzanowski (2010, 2013) claims that it is 2 staged. According to his claim, this two-staged approach comprises intertextuality and interdiscursivity. In the first stage, the texts are read carefully, and headings of discourses are derived from them (Krzyzanowski, 2013, p. 116), and in the second stage, in-depth text analysis is performed to determine how the intertextuality and interdiscursivity are being ensured (readdressing the issue along with the surrounding elements–by re-contextualization method). In this research, the CDA method was used by performing the above stages to get a response to the research questions.

A purposive sampling method was used. 245,587 tweets in total posted under the hashtags of #suriyeli (Syrian), #mülteci (refugee), #suriyelimülteci (Syrian refugee), #suriyelileriistemiyoruz (we don't want the Syrians), and #suriyelilerdefolsun (Syrians piss of) were collected by the codes written in R language and by the use of R Studio program (V. 1.1.463), which was inspired by S language (Ihaka & Gentleman, 1996). CDA strategies developed by Chen and Flowerdew (2019) were used for the classification of strategies on Syrian refugees.

Three coders analyzed 200 posts randomly selected from the sample in accordance with the code list. All three coders agreed by making the required changes in the coding of the study and entered the data in R studio program by performing the coding as per the determined date ranges. Krippendorff's alpha coefficient was calculated to test the reliability among the coders. K ALPHA value was determined as α=0,762 (Krippendorff, 2011). While retrieving data from Twitter, in requests made over API, the number of requests returned by Twitter is between 18 and 50 per 15 minutes. According to the search request, Twitter is returning search results at this range, and this figure is coming out as a constraint in large-scaled searches. Again, Twitter specifies that the search API on its site is not an extensive tweet source and that all tweets are not being indexed (Twitter, 2019).

5. Findings

Totally, 245,587 tweets were examined. It shows that 37.39% of the tweets are ones with negative content (Table 1).

Theme	n =	%
Negative Tweets	91828	37.39
Positive Tweets	69913	28.47
Government Criticism	57614	23.46
Irrelevant Tweets	26232	10.68
Total	245587	100

Table 1 – Analysing Turkish tweets under anti-Syrian hashtags

It was also observed that there are irrelevant contents (10.68%) among the comments made under the hashtags subjected to research. We excluded such contents from the research. And the Turkish society has actually divided into two main opinions on Twitter regarding Syrian refugees. One section of the society is marginalizing the Syrians as refugees, and developing hate speech regarding them, and the other section is embracing the Syrian refugees over 'religious fellowship' by deeming them as 'oppressed Muslim public'. But the third opinion explored as the result of the current research, which is 'government criticisms', is new, and maybe it is the point of making this research exploratory.

Discursive Strategy	n =	%
1 – Negative Other	16916	18.42
2 – Scare Tactics	52851	57.55
3 – Blaming the Victim	7345	8.00
4 – Delegitimation	14716	16.03
Total	91828	100

Table 2 – Discursive strategies in Turkish.

CDA was performed for the negative tweets (N = 91.828 / 37.39%), forming the basis of the research. The negative tweets were coded by the taxonomy of Chen and Flowerdew (2019) under the four main categories of classification as (1) Negative other (2) Scare tactics (3) Blaming the victim (4) Delegitimation (Table 2). All the examples were translated to English from Turkish.

1. Negative Other	**n =**	**%**
1 – 1. Negative Attributions	8435	49.86
1 – 2. Labelling	3372	19.93
1 – 3. Construction of Conspiracy Theories in Group	2578	15.24
1 – 4. Dehumanisation	2531	14.96
Total	16916	100
2. Scare Tactics	**n =**	**%**
2 – 1. Predicting Threats to Interest of the in-group	15484	29.30%
2 – 2. Predicting Threats to Public Order and Political Stability	20173	38.17%
2 – 3. Use of Quasi-objective Statistics	8208	15.53%
2 – 4. Use of Quasi-theory	8986	17.00%
Total	52851	100
3. Blaming the Victim	**=**	**%**
3 – 1. Self-justification / Positive Discrimination	1704	23.20%
3 – 2. Distortion	2202	29.98%
3 – 3. Concession	3439	46.82%
Total	7345	100
4. Delegitimation	**n =**	**%**
4 – 1. Magnifying Voice the Group	10702	72.72%
4 – 2. Pointing to Illegitimate Status / Activities	4014	27.28%
Total	14716	100

Table 3 – Taxonomy of discriminatory discursive.

Posts regarding forming negative attitudes about others, stigmatizing and labeling the others, unfavorably exaggerating the characteristics of others and forming conspiracy theories against them, and dehumanizing the others form the main discriminatory discursive strategies classification. Among the negative tweets, this group is getting a share of 18.42% (Table 2). However, when the sub-typologies of the dominant strategy of are examined, the sub-typology of ‘Negative Attitudes’, which brings the negative characteristics of Syrians to the forefront, is ranking first by 49.86%. Posts exemplifying this classification are:

> (1.1) Syrian, who is fleeing without firing a gun in his country, is shooting the cats in our country and having pleasure! The police are coming and taking the rifle.
>
> (1.1) Isn’t the one on the left smoking the water pipe in the sea a Syrian?

Following the above classification, the classification of labeling (1.2), forming the second-largest class, covers 19.93%:

> (1.2) Look at his style. He just looks like Syrian.
>
> (1.2) Let’s continue with taking care of the Syrian parasites.

The sub-typology of (1.3) follows the above two classifications by 15.24%:

> (1.3) By the way, the USA and Russia split the petroleum between themselves, and our job remained as constructing the houses of 5 million parasites and feeding them by giving them electricity and water for free.

And finally, there is the sub-typology of (1.4) with a rate of 14.96% (Table 3):

> (1.4) This many Syrian hungry pigs will eat their fill in some way.

The largest group of posts in classifying main discriminatory discursive strategies are ‘Scare Tactic’ by 57.55% (Table 3). This classification comprises four sub-typologies: Anticipating the threats against the interests of small groups within the society (increasing the concerns against the group by exaggerating the statistics among in-group members, abnormalizing and criminalizing the out-group by exaggerating the threats against public order, manipulating the statistical data for the sake of the interests of the group and distorting the statistical data, and finally developing discourses by the use of quasi-theories sometimes by specifying the source, but mostly without specifying the same). The largest sub-class under (57.55%) is 'estimating the threats against public order (2.2)' by 38.17%. This typology is also the second-largest class among all negative tweets with a share of 21.97%. Such post:

> (2.2) One day, the Syrians may attain the municipality elections. Unfortunately, their population is too much. Their population is increasing faster than us!

> (2.2) One day the Turks will become the minority population in Turkey, the Syrians will enter the parliament by establishing a political party.

The second-largest sub-class under (57.55%) is 'Estimating the threats against the interests of the group' (2.1) by 29.30%. This class is also the second-largest class among all negative tweets with a share of 16.86%. Such posts are criminalized and abnormalizing the out-group by exaggerating the threat against public order (Table 3):

> (2.1) Number of asylum seekers coming from Syria is about 4 million. 47% of them are under the age of 18. About 400 Syrian babies are being born per day.

> (2.1) While hundreds of thousands of teachers are waiting to be assigned, they are assigned thousands of Syrian teachers. They gave citizenship to 110 thousand Syrians.

The above two typologies are being followed up by the sub-classification of 'Use of Quasi-objective Statistics' (2.3) by 15.53%, and finally the sub-classification of 'Use of Quasi-theory' (2.4) comes by 17% (Table 3).

> (2.3) Following the 110 thousand Syrians to whom citizenship they provided will also provide citizenship to the remaining 4 million Syrians.

> (2.4) Why do you think the NATO ships are at the Aegean Sea? Turkey is the buffer zone of Europe. It is the refugee dumpsite of Europe, no need to fool each other. If you are such a foolish community that doesn't look after the country, you bear the consequences.

The strategy of ranks last by 8%. This strategy comprises three sub-typologies: (3.1) (legitimizing the characteristics of another ill-famed group, the biased attitudes regarding the other group) by 23.20%, (3.2) (generating negative and falsified news about the others) by 29.98%, and finally (3.3) (first starting by positive content, and then telling negative contents) by 46.82% (Table 3):

> (3.1) The most radical terrorists are the Syrians. Something inclined them to war. They hate Turkey. And what did we do? We opened the borders and took them to Turkey when ISIS pressured them. Unfortunately, it's our fault. We have pity on our enemies.

> (3.2) Just now it was on the news. They will place 2 million Syrians in İzmir. They didn't want to be placed in Ankara. Good luck with it.

> (3.3) I also have many Syrian students. They are all nice, I love them, but apart from the children, should we want the water pipe smokers, thieves, also perverts to stay?

Strategy ranks third by 16.03% (Table 2). This class comprises two sub-classes: 'Magnifying voice against the group' (4.1) (72.8%), and 'Pointing to illegitimate status/activities' regarding the others (4.2) (27.28%) (Table 3):

> (4.1) The citizens at Arnavutköy rioted the abuses of Syrians.

> (4.2) They have all kinds of tricks, they were collecting money from around by acting as if collecting things from the garbage.

Government Criticisms	n =	%
1 – Government's	28621	49.68
2 – Government's policies	19704	34.25
3 – Economic	7453	12.94
4 – Injustice/unlawfulness	1836	3.19
Total	57614	100

Table 4 – Taxonomy of government criticims under Anti-Syrian tweets.

This research searched whether the classification of Chen and Flowerdew (2019) functions in Turkey or not, and findings showed that all discriminatory discursive sub-strategies are being used specifically to Syrian refugees in Turkey. But the new findings extens the current taxonomy for Turkey. The new critical discourse strategy being present in the extended taxonomy is 'Government Criticisms'. It showed that 23.46% of the posts mainly comprise criticisms against the government policies. Twitter users prefer posting their criticisms regarding the government under the negative hashtags regarding Syrians. In this new strategy, four sub-typologies were found. These are government criticisms regarding the government's domestic policies, economy, government's foreign policies, and injustice/unlawfulness (Table 4).

The problem of hundreds of thousands of people in Turkey, who cannot get retirement pay despite having retired as they cannot meet the precondition of age although having fulfilled the period of working days that is required for retirement, is called Victims of Delayed Pension Age (VDPA). In the state of emergency periods practiced in the country, the government had ended thousands of individuals, through decree-laws, who had become unemployed. In this section, the posts with the themes of the problem of VDPA, and the problem of public officers discharged by Decree-Law, and posts telling about the problems of the country by exemplifying the Syrian refugees and criticisms against the government's policy of Syria are collected.

For instance; it was observed that many tweets such as problems of #EYT (VDPA) victims aren't solved.

'The end of AKP came by this VDPA. Good luck with it. I'm calling you trolls. You have no information on this subject; you are completely gammoning; you have no information on the details and cost of the incidence'.

They presented these tweets under the hashtags of #syrian, #refugee, #syrianrefugee, #wedon'twantthe syrians, and #syrianspissoff. It is being preferred to share the posts drawing attention to the government's administration manner, negativities in the society, and social problems and remonstrations again by the hashtags regarding the Syrians, and it is being aimed at drawing the attention of the public to these subjects. The following tweets may be given as examples of government criticism.

Criticisms regarding the government's foreign policies:

> (5.1) Erdoğan is at the beck and call of Trump!
>
> (5.1) You have been a sandwich between Russia and the USA. The country has lost its composure. What are you talking about?

Government criticisms within the economy:

> (5.3) We paid premiums and taxes for about 30 years, but we are not worth the Syrian bastards.
>
> (5.3) You had spent 40 billion Dollars by collecting 4 million Syrians, 4 citizens had committed suicide because of poverty, and you're not spelling a word.

Criticisms regarding the government's domestic policies:

> (5.2) Let me ask not theoretically, but concretely. Where is the money of public affiliates transferred to Wealth Fund while making a profit, and that suddenly made a loss?

> (5.2) Debts have increased. Suicides have increased. Drug consumption has increased. Prostitution has increased. Abuse has increased. Women's murders have increased. What are you talking about?

Government criticisms within justice/law:

> (5.4) His son and daughter are imprisoned, he is looking after 3 grandchildren: I'm leukemia; I have no power to stand!

> (5.4) If you are selecting these subjects yourself as the press, I'm condemning you as a citizen of the Republic of Turkey. And if it is due to pressure, I'm feeling pity.

6. Conclusion

This study analyzes the marginalizing and discriminatory discourses on Twitter regarding the Syrian refugees living in Turkey. 245,587 tweets total were analyzed, and it should ensure a general understanding of how the refugees are being depicted in online public discourses by analyzing the selected tweets. It was examined whether there are posts on Twitter covering hate speech against the Syrian refugees. The study determined that public discourse in online media has a sharper tone and that it is being collected under negative frames (RQ1). This circumstance also shows that Twitter is a medium allowing many online users to generate marginalizing and discriminatory discourses about the Syrian refugees. Interestingly, when the comments are collected under the selected hashtags, these hashtags are also juxtaposing negative word groups (for instance, traitors, thief, coward, murderer, terrorists, etc.) gathering negative characteristics and threats regarding the refugees and concerns are detrimental to the country. Posts regarding the formation of negative attitudes regarding the Syrian refugees, stigmatizing and labeling them, unfavorably exaggerating their physical or social characteristics, had also observed the formation of conspiracy theories about them regarding the country's interests and even dehumanization of Syrian refugees.

And findings showed that 28.47% of total tweets are comprising positive content defending the Syrians. When the positive posts were examined, the number of Syrians returning to their country is continuously emphasized. They are being reflected as ‘they are going to their country, they are returning’, and it was observed that the discourses in positive tweets are being formed as addressing the most victim group (Syrian children, babies, and women) by making a ranking of ‘least victim, mid-level victim, most victim’ among the refugees and that posts activating the instincts of appropriation, helping and protecting are much over the theme of ‘our Syrian brothers.

The result from here is that it divided the society on Twitter to two different basic views regarding the Syrian refugees. One section of the society is marginalizing the Syrians as refugees and developing hate speech regarding them, and the other section is embracing the Syrian refugees over religious fellowship by deeming them an oppressed public.

When comments with negative discourses were considered, it was observed that an oppressive discourse is being formed such as ‘Let the Syrians go to Syria, Turkey belongs to Turks’ by the plucking of the rising nationalism movement in Turkey away from the centerline of religion. It was observed that the ones writing negative comments on Twitter are presenting themselves as a group, and they are referring to themselves as ‘us’, but even if which group they belong to is not clearly being specified, that they are legalizing the discrimination, they make against the ‘others’. While some tweets refer to ‘us’ ‘us’ (Turks), most of the tweets contain ambiguous expressions, and who the ‘us’ ‘us’ is not clearly being specified. It was observed that ‘others’ the Syriansto exclude those is being used to exclude those whodon’t belong to ‘us’ in the sense of group identity. This circumstance emphasizes the threat against the ideology of sovereignty and the superiority of the Turkish nation, and In addition, it shows the rise of right populism in Turkey. These findings are parallel with Wodak (2015) similar studies and Kreis (2017) performed in Europe.

It was observed that the positive comments under the selected tweets are being made on the centerline of religion, that the positive contents are being provided within the frame of religious fellowship because both nations have the same religious belief, and that it is being expressed that it is obligatory in terms of religion for the Muslims (Turks) to take Muslims (Syrians) under their wing in difficult circumstances.

A difference was determined between the classification got as the result of the research and the classification made by Flowerdew et al. based on CDA strategies (RQ2). New categories or sub-categories were obtained (RQ3). Among the new findings is that a significant rate of 23.3% of the comments, except the negative comments, reflects increasing negative emotions against the migration and refugee policies within the frame of Turkey's foreign policy. Among the posts included in this rate, posts are criticizing the Syria policy of Turkey, and criticizing the government policies (posts regarding the ones who cannot retire because of being hindered by age in early retirement practice, and posts regarding the ones who are discharged by decree-law, who are being prohibited from working at another position) and criticizing the country's problems (increasing inflation, unemployment, economic bottleneck, bankrupts, violence against women, terrorism, wars continuing at the Middle East countries, etc.) by exemplifying the Syrian refugees. They may be evaluated among the negative tweet contents. Interestingly, the users are making comments covering such government policies under the relevant hashtags, and maybe by this means they are trying to give a message to the government like 'You have solved my problem rather than dealing with the Syrians'. In the contents, there are ones linking the negative course of Turkey's economy with the Syrian refugees, and the ones perceiving the Syrian refugees as an economic threat are a lot.

On Twitter, the citizens' communication of their displeasure arising from government policies (domestic and foreign) over the Syrian refugees is indirectly being represented by the negative hashtags regarding the Syrian refugees. Turkish society expresses its discomfort regarding government policies and the country's problems as reflecting over Syrian refugees. This circumstance is causing the increase of the frequency of hate speech by different contents. The consequences of government policy, such as the employees hindered by age limits in early retirement practice, unemployment, and economic crisis, are being associated with Syrian refugees. Likewise, other research (Aslan, 2017) is also supporting this. Aslan (2017) had determined by the videos on YouTube regarding the Syrians that they are being accused of being betrayer and ungrateful, and also that they are being deemed as the source of economic problems in Turkey. And he had also specified that they are being associated with some crimes in the country.

This study has not just shown the researchers that Twitter is an extremely open platform for hatred discussions and political violence. However, it has also concluded that agendas are being formed through tweets regarding the country's problems such as economy and unemployment that are placed under the hashtags regarding the Syrians. In this context, Twitter has a significant place that determines the agenda regarding the refugees in the country and the world. Also, allows the formation of new agendas by associating the generated contents with actual subjects. Furthermore, along with Twitter's increasing usage in Turkey since 2006, it has experienced an increase in the number of Turkish users during the civil commotion of Gezi Park protests. The report published by eMarketer disclosed that the number of Twitter users in Turkey had increased to 11.3 million after the Gezi Park protests (Kural, 2013). Thus, this platform is an extremely effective platform for forming, developing, and acting out of hatred.

While this platform gives the researchers the opportunity to collect extensive data, it is causing the echo chambers–created by the individuals of similar opinions via the networks- to be similar with the findings (for instance, attitude, idea, and discourses) got by the researchers. But this also shows us that the dominant opinion and discourse on Twitter may not represent the general population. For this reason, researchers should care to abstain from generalizations, especially in qualitative research to be performed on Twitter (Mitchell & Hitlin, 2013). And thus, it is required to abstain from making status evaluations regarding general Turkish society as only being based on findings focused on the results of Twitter. For further studies, it is suggested to examine the viewpoints of the Turkish public regarding Syrian refugees in their content generation process on social media by supporting this subject with a questionnaire.

References

Akşak, E. Ö. (2019). Discursive construction of Syrian refugees in shaping international public opinion: Turkey's public diplomacy efforts. *Discourse & Communication*, 175048131989376. https://doi.org/10.1177/1750481319893769

Asaf, Y. (2017). Syrian women and the refugee crisis: Surviving the conflict, building peace, and taking new gender roles. *Social Sciences*, 6(3), 110. https://doi.org/10.3390/socsci6030110

Aslan, A. (2017). Online hate discourse: A study on hatred speech directed against Syrian Refugees on Youtube. *Journal of Media Critiques*, 221-256. https://bit.ly/377LYFK

BBC News, (2019, December 9). *Suriye'de 8. yılına giren savaş*. https://bbc.in/2THxeu1

Bennett, S. (2015). *Constructions of migrant integration in British public discourse (Unpublished doctoral dissertation)*. University of Adam Mickiewicz in Poznan, Poznan, Poland.

Bernstein, B. (1990). *The social construction of pedagogic discourse: Vol. IV. Class, codes and control*. London: Routledge.

Bhatia, A. & Jenks, C. J. (2018). Fabricating the American Dream in US media portrayals of Syrian refugees: A discourse analytical study. *Discourse & Communication,* 12(3), 221–239. https://doi.org/10.1177/1750481318757763

Bilge, R. (2016). Sosyal medyada nefret söyleminin inşası ve nefret suçlarına ilişkin yasal düzenlemeler. *Yeni Medya 1*(2), 1-14. https://bit.ly/3f7NOuA

Blumer, H. (1958). Race prejudice as a sense of group position. *Pacific Sociological Review,* 1(1), 3-7. https://bit.ly/3l8KfIc

Chen, M. & Flowerdew, J. (2019). Discriminatory discursive strategies in online comments on YouTube videos on the Hong Kong Umbrella Movement by Mainland and Hong Kong Chinese. *Discourse & Society*, 30(6), 549-572. https://doi.org/10.1177/0957926519870046

Çoban, K. H. (2016). Metaforun ayrımcı hegemonyanın inşasındaki rolü: Suriyelilerin haberleştirilmesinde metafor kullanımı. *Gaziantep University Journal Of Social Sciences*, 15(2), 253-280. https://bit.ly/3f7BfPT

Directorate General of Migration Management. (2020, November 29). *Geçici Koruma İstatistikler.* https://bit.ly/374ZJoq

Duran, H. (2019, January 7). The rise of hate speech against Syrian Refugees in Turkey. *The New Turkey*. https://bit.ly/2Vn083a

EnKocaeli.com. (2019, September 5). *Genç kızlara tecavüz eden Suriyelileri feci şekilde dövdüler.* https://bit.ly/3BUYt5C

Erdoğan, O. Y. & Isik, G. H. (2020). Discourses of exclusion on Twitter in the Turkish Context: #ülkemdesuriyeliistemiyorum. *Discourse, Context & Media*, 36(2020), 100400. https://doi.org/10.1016/j.dcm.2020.100400

Erickson, K. & Lilleker, D. G. (2014). Elite Tweets: Analyzing the Twitter communication patterns of labour party peers in the House of Lords. *Policy and Internet*, 6(1), 1-27. https://doi.org/10.1002/1944-2866.POI350

Flowerdew, J. L., David, C. S. & Tran, S. (2002). Discriminatory news discourse. Some Hong Kong data. *Discourse & Society*,13(3), 319-345. https://doi.org/10.1177/0957926502013003052

Gorodzeisky, A. & Semyonov, M. (2019). Unwelcome immigrants: Sources of opposition to different immigrant groups among Europeans. *Front Sociol l.*4, 24. https://doi.org/10.3389/fsoc.2019.00024

Grimmer, J. & Gary, K. (2011). General purpose computer-assisted clustering and conceptualization. *Proceedings of the National Academy of Sciences*, 108(7), 2643–2650. https://doi.org/10.1073/pnas.1018067108

Haberler.com. (2019, December 12). *Genç kızı taciz eden Suriyeliyi taksi durağında dövdüler*. https://bit.ly/3l9qpNe

Hatipoğlu, E., Gökçe, O. Z., Arın, İ. & Saygın, Y. (2019). Automated text analysis and international relations: The introduction and application of a novel technique for Twitter. *All Azimuth: A Journal of Foreign Policy and Peace,* 8(2), 183-204. https://dx.doi.org/10.20991/allazimuth.476852

Ihaka, R. & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299-314. https://doi.org/10.1080/10618600.1996.10474713

Kalav, A.& Certel, A. B. (2017). Amerikan sosyal medyasında göçmen karşıtlığı ve dijital nefret söylemi: Twitter özelinde bir inceleme. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *22*, 2209-2222. https://bit.ly/3BSTpi5

Kirschenbaum, A. (2004). Generic sources of disaster communities: A social network approach. *International Journal of Sociology and Social Policy*, 24(10/11), 94-129. https://doi.org/10.1108/01443330410791073

Kreis, R. (2017). #refugeesnotwelcome: Anti-refugee discourse on Twitter. *Discourse & Communication*, 11(5), 498–514. https://doi.org/10.1177%2F1750481317714121

Kuş, O. (2016). Dijital nefret söylemini anlamak: Suriyeli mülteci krizi örnek olayı bağlamında BBC World Service Facebook sayfasına gelen yorumların metin madenciliği tekniği ile analizi. *İstanbul Üniversitesi İletişim Fakültesi Dergisi*, 2, 97-121. https://doi.org/10.17064/iuifd.289373

Krippendorff, K. (2011). Computing Krippendorff's alpha-reliability. The Basics Of Communication Research. Cengage Learning. https://bit.ly/3l4FqQh

Krzyzanowski, M. (2010). *The discursive construction of European identities. A multi-level approach to discourse and identity in the transforming European Union*. Frankfurt am Main, Germany: Peter Lang.

Krzyżanowski, M. (2013). Policy, policy communication and discursive shifts: Analysing EU policy discourses on climate change. In: Cap, P, Okulska, U (eds) *Analysing New Genres in Political Communication*. Amsterdam: John Benjamins Publishing Company, 101–135. https://doi.org/10.1075/dapsac.50.05krz

Kural, S. (2013, December 9). Gezi Parkı olayları Twitter kullanıcı sayısını patlattı (mı?). https://bit.ly/37mrl8V

Media Watch On Hate Speech Report. (2019, March 3). Hırant Dink Foundation, May-August 2019. https://bit.ly/2Vkjx4J

Memişoğlu, F. & Ilgit, A. (2017). Syrian refugees in Turkey: Multifaceted challenges, diverse players and ambiguous policies. *Mediterranean Politics 22*(3), 317-338. https://doi.org/10.1080/13629395.2016.1189479

Mitchell, A. & Hitlin, P. (2013, December 10). Twitter reaction to events often at odds with overall public opinion. https://pewrsr.ch/3i9mT3b

Mihajlova, E., Bacovska, J. & Shekerdjıev, T. (2013). Freedom of expression and hate speech. OSCE, Mission to Skopje, Skopje: Polyesterday. https://bit.ly/3ibiTQ0

Öztürk, N. & Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telematics and Informatics*, 35(1), 136-147. https://doi.org/10.1016/j.tele.2017.10.006

Reisigl, M. & Wodak, R. (2001). *Discourse and Discrimination. Rhetorics of Racism and Antisemitism*. London, New York: Routledge.

Rettberg, J. W. & Radhika, G. (2016). Terrorists or cowards: Negative portrayals of male Syrian refugees in social media. *Feminist Media Studies*, 16(1), 178-181. https://doi.org/10.1080/14680777.2016.1120493

Richardson, J. E. & Wodak, R. (2009). Recontextualising fascist ideologies of the past: Right-wing discourses on employment and nativism in Austria and the UK. *Critical Discourse Studies*, 6(4), 251-267. https://doi.org/10.1080/17405900903180996

Sayımer, I. & Rabenda, D. M. (2017). Syrian Refugees as victims of fear and danger discourse in social media: A Youtube analysis. *Global Media Journal TR Edition*, 8(15) Fall 2017, 384-403. https://bit.ly/3zJL2Ub

Soral, W., Bilewicz, M. & Winiewski, M. (2017). Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior,* 44(2), 136-146. https://doi.org/10.1002/ab.21737

Sulaiman-Hill, C. M., Thompson, S. C., Afsar, R. & Hodliffe, T. L. (2011). Changing images of refugees: A comparative analysis of Australian and New Zealand print media 1998– 2008. *Journal of Immigrant & Refugee Studies*, 9(4), 345-366. https://doi.org/10.1080/15562948.2011.616794

Twitter. (2019, January) 5. *Search Tweets*. https://bit.ly/3rIduD5

Van Dijk, T. A. (2011). Discourse and ideology. In: Van Dijk TA (ed.) *Discourse Studies: A Multidisciplinary Introduction*. London: SAGE, 379-407. http://dx.doi.org/10.4135/9781446289068.n18

Van, L. T. & Wodak, R. (1999). Legitimizing immigration control: A discourse-historical analysis. *Discourse Studies*, 1(1), 83-118. https://doi.org/10.1177%2F1461445699001001005

Verweij, P. (2012). Twitter links between politicians and journalists. *Journalism Practice*, 6(5-6), 680-691. https://doi.org/10.1080/17512786.2012.667272

Wodak, R. (2015). Discrimination via discourse. In: Bonvillain N (ed.) *Routledge Handbook of Linguistic Anthropology*. London: Routledge, 366-383.

Wodak. R. & Fairclough, N. (2010). Recontextualizing European higher education policies: The cases of Austria and Romania. *Critical Discourse Studies*, 7(1), 19-40. https://doi.org/10.1080/17405900903453922

Yazıcı, T. (2016). Yeni medyanın nefret dili: Suriyeli mültecilerle ilgili ekşi sözlük örneği. *Global Media Journal*, 7(13), 115-136. https://bit.ly/3zNyT0t

Yeni Çağ. (2019, December 13). *Türk kızlarına taciz edip görüntülerini paylaşan suriyelileri böyle dövdüler*. https://bit.ly/2TH94zH

Yıldız, E. (2018). Twitter'da ve çevrimiçi bir gazetede yer alan nefret söylemlerinin karşılaştırılması: Suriyeli mülteciler örneği. *OPUS Uluslararası Toplum Araştırmaları Dergisi*, 9(16), 760-793. https://doi.org/10.26466/opus.478176

Chapter 4

DISSEMINATING AND RESISTING ONLINE HATE SPEECH IN TURKEY

Mine Gencel Bek
Universität Siegen, Germany

1. Introduction

The history of the use of the Internet as a tool for the circulation of hate speech goes back to the 1980s with the actions of neo-Nazis, and since then, it continued with a wide range of activities of different individuals and groups, including racist, xenophobic, homophobic, transphobic, religious, misogynic (Doğu, 2010). Hate speech is visible in different forms on social media platforms: By relying on their research on political extremism on Facebook in Spain, Ben-David and Matamoros-Fernandez (2016) argue that it can be overt hate speech (such as direct threats and insults against immigrants and foreigners) or covert discrimination (for example, associating discourses of immigrants and denigrating Muslims with danger and crime). Müller and Schwarz (2021), through their research on anti-refugee sentiment on Facebook, argue that there is a causal link between social media and hate crime: Right-wing social media can lead to violent crimes through the propagation mechanisms. Online speech with the potential to incite violence is often referred to as “cyberhate” (Gagliardone, 2019).

There has been a significant increase in hate speech on social media recently (Jakubowicz, 2017; Udanor & Anyanwu, 2019). Hate moves globally beyond borders as a ‘global racist subculture’ (Perry & Olsson, 2009).

However, Matamoros-Fernández and Farkas (2021), in their extensive review of the literature on racism and hate speech, argue that there is a lack of geographical diversity and that the United States is more studied than the rest of the world. The chapter which aims to contribute to the debate with the Turkish case first reviews the studies on hate speech in Turkey. A case study on hate speech directed to popular singer Sezen Aksu follows that. Finally, the chapter discusses the ideas and attempts against hate speech and its limitations and potential.

2. Hate speech in the Turkish media

The killing of Hrant Dink in 2007 has been a turning point to discuss hate speech in Turkey. Hrant Dink was the founder and the editor in chief of the Turkish Armenian community's Agos newspaper. Still, the murder has not been solved. The role of the media in showing Hrant Dink with his Armenian identity as a target has been criticized by many anti-racist intellectuals, academics, journalists, and activists. Fethiye Çetin (2010) states:

> Now we all know that there was a systematic attack on Hrant Dink a year later, that is, in 2004. We can see that he was somehow made a target in the public opinion, with an isolation policy directed at him. And throughout this entire process, the media played a very important function...
>
> ... There was this campaign run by the media, which prevented people, not only those consciously involved in this plan but also those who had no connection whatsoever with it, from taking steps to protect Hrant Dink. They could not dare it. I think the media has a very important role here. Discussing the role of the media is important in terms of understanding how not only the Hrant Dink murder but also the hate crimes in general gain a systematic nature.

Journalist Kemal Göktaş (2010: 94) lists a series of news full of hate and adds:

> Of course these articles appearing in the media were reflected with a triple effect on racist websites. Exactly one year before the assassination of Hrant Dink, a web site published an article which included the following sentences: “The snarling dog Hrant Dink, … since infidel courts are established in the homeland of Turks, and as a slap against these Zionist courts, the slaying of these two dogs (here the reference is to Orhan Pamuk) is an urgent necessity.” And no legal action was initiated with regard to these sentences.

Media studies scholars Yasemin İnceoğlu and Ceren Sözeri (2012) analyze the press before the murder of Hrant Dink with many examples full of hate speech and argue that media has neither confronted the issue of hate speech and the consequences of hate speech nor realized a self criticism. Eser Köker and Ülkü Doğanay (2010, p. 116) in the conclusion of their analysis of the media after the murder of Hrant Dink discuss how the media foregrounded the frame of ‘Turkey’ rather than focusing on the issues of racism:

> … the printed media covered the news of Hrant Dink’s murder mostly through strategies of denial, blaming, ignoring or rendering invisible. Hence, the press not only overlooked the human dimension of the murder but also pushed into the background the link between the murder and the racism targeting Armenians in Turkey. The words used by columnists who remained outside of this framework and who emphasized the rising nationalism and racism lying behind the murder are lost amid the news stories and articles that evaluate the murder as a conspiracy against Turkey, and are rendered ineffective as they fail to receive any comments.

On the one hand, we can say that these critical studies and perspectives did not lead the mainstream media to develop an overall self-criticism that led to any change. On the other hand, we have seen a solidification of counter-racist mobilization in academia, journalism, and civic circles. There even emerged initiatives against the hate crime after Hrant Dink’s murder. It is titled DurDe (Say No to Racism and Nationalism). Levent Şensever (2010) explains that in 2007 establishing the network with these emotions: "that something should be done about the anger, the desperation and the frustration we felt after the murder" (p. 290).

Hrant Dink Foundation, which was established in 2007 has become a source of many studies to struggle against hate speech in the media. Their scope is based mainly on monitoring hate speech in print media rather than online platforms. In 2019, the report Hate Speech and Discriminatory Discourse in Media detected discriminatory discourse that directly or indirectly excludes and marginalizes Kurds, associates Kurdish identity with crime, and portray Kurds as the actor of negativities while Kurds remain passive and silent in these media texts. The media targets at the same time non-Muslims which is not random, as Arus Yumul (2021) states. Besides Hrant Dink Foundation, LGBTQI associations such as KAOSGL monitor hate speech against LGBTQI, carry out awareness campaigns and develop reactionary policies against the state's homophobic discourses and policies (Yıldız, 2021; Cerit, 2010). This issue has been a theme in the work of Hrant Dink Foundation as well: In the report titled *Discriminatory Discourse in News Stories on Murders of Transgender Women*, Burak Demiryakan and Pınar Ensari (2017) problematize invisibility and indifference to transgender murders in the media. According to their analysis, those represent also use the perpetrator's statement to legitimize murders and tabloidize in sensational ways. These were strategies being used previously against violence towards women in the media (Gencel Bek, 2012).

In recent years, we have seen studies on deciphering hate speech against refugees in the media as well. Syrian refugees are framed in the mainstream media with the discourses of threat, economic burden, and security problem (Doğanay & Çoban Keneş, 2016). Not only anger is used in hate speech. Erbaysal Filibeli and Ertuna (2021) detect sarcasm in comments on social media platforms against Syrian refugees in Turkey as an instrument for discrimination and expressing superiority. There are also studies on countering these discourses as one step beyond conducting critical discourse analysis. One of these recent studies, *Towards a New Discourse: Training Manual*, published by Hrant Dink Foundation, is a collaborative attempt (Yilmaz, 2021) between the public, media, and civil society organizations to build an alternative discourse altogether. The project also has an online media dimension. Adaptation to the handbook on hate speech produced by ARTICLE 19 to the context of Turkey in 2015 as a guideline titled *Tüm Yönleriyle Nefret Söylemi-Kılavuz* (Yilmaz, 2022) also is another crucial publication of Hrant Dink foundation.

Another series of recent studies has been on the critical analysis of the hate speech depicted on social media during COVID-19. According to an elaborative report by Dirini and Özsu (2020), hate speech disseminated on social media is even faster than the spread of the COVID-19 virus. The target groups this time were expanded to include Chinese people, elderly people older than 65, and LGBTİ+ people with mostly insults, swearing, and downgrading. The authors warn that the strong religious references, which were emphasized with the dichotomy of us versus them, have a stronger impact. Our research also inhabits the religious dimension, as we discuss below.

3. Speech on Twitter against Sezen Aksu

On 15 January 2022, we witnessed a very popular singer Sezen Aksu being targeted for insulting the sacred figures and Islam because of the words 'Say hello to the ignorant Eve and Adam' in her song Şahane Bir Şey Yaşamak (What a Wonderful thing to Live). Interestingly, the song was not new but was released five years ago, in 2017. The online lynch campaign against Aksu started with the social media posts of Erdem Özveren, a member of the İsmailağa cult, introducing himself as the head of The Islam Defense Action (Medyatava, 2022). While thousands of tweets were posted on the same day, the pro-government Internet news websites and newspapers' Twitter accounts shared the news with offensive and provocative language. Yeni Safak, known as a pro-government newspaper, shared the story with a manipulative headline: "Insult against religious values by Sezen Aksu: the lyrics of her new song drew wrath on social media" (Yeni Safak, 2022). The hashtag #SezenAksuHaddiniBil ("Sezen Aksu, know your limit") appeared in the Trending Topics. On the same day, a group from Islam Defense Action protested against Aksu in front of her house in Istanbul. A group of lawyers has filed a judicial complaint against her that she has degraded religious values' (Al-Monitor, 2022). On 17 December, the Directorate of Religious Affairs Diyanet publicly stated, "The careless attitude towards religious figures is disrespectful, to say the least" (Gazete Duvar, 2022).

On the other hand, following the hashtag campaign, the hashtag ##sezenaksuyalnızdeğildir (Sezen Aksu is not alone) was initiated by the Sezen Aksu supporters and became a trending topic in response to the preceding anti-Aksu hashtag campaign. (Euronews, 2022). RTUK Radio and Television Supreme Board warned the channels by underlining that they would face severe sanctions if they broadcast the song (Şimşek, 2022). RTÜK vice president Uslu advocated this action following the criticism that RTÜK threatens the channels as preventing the channels from possible harm.

Nationalist Movement Party leader Devlet Bahçeli, who is also an ally of Erdogan's Justice and Development party, said that "If you are a sparrow, know your sparrowship, don't be tempted to be a raven," referring to the iconic singer's nickname of "Minik Serçe" (Gazete Duvar, 2022). On the other hand, opposition politicians like Istanbul Mayor Ekrem İmamoğlu and many artists expressed their support for Sezen Aksu on Twitter.

A week after the controversial debate on social media and traditional media, President Recep Tayyip Erdogan was involved in the debate without mentioning her name in his speech after Friday prayer at Grand Çamlıca Mosque on 21 January: "We cannot allow people to badmouth the first prophet and the mother of humanity. "We will tear out the tongues of those who do so, if necessary." (Duvar English, 2022). The statement of Erdogan as the President of Turkey was regarded as a direct threat against Sezen Aksu. Erdogan's harsh statement invoked massive support from social media users as well as from political circles. As a reply to the President's statement declared with the public during the Friday prayers at İstanbul's

Grand Çamlıca Mosque, singer and songwriter Sezen Aksu shared the lyrics of her new song "Hunter" from her Facebook account on 22 January. She thanked those who announced support for her and said, "As you know, the matter is not me; the matter is the country." The lyrics of the song titled "The Hunter" (Avcı) were translated into more than 30 languages in a few days on social media.

In a TV interview on 27 January, Erdogan stated that he was not referring to the legendary singer Sezen Aksu, saying that she is an important name for Turkish music and an interpreter of emotions. However, his words were regarded as a step back due to the massive support of public opinion when he did not name whom he was referring to in his harsh statement. Still, some popular figures are being targeted, subject to investigation, and even arrested in Turkey for degrading religious values with Article 216/3 of the Turkish Penal Code, which sets out a penalty of imprisonment from six months to one year if the act disturbs the public peace. The article has been criticized for vagueness and increasing misuse by the government against the opponents in the last ten years (Al-Monitor, 2022). Also, 160,169 investigations have been started for "insulting the president" over the last eight years, based on article 299 of the Turkish Penal Code, which was demanded to be changed by The European Court of Human Rights (ECHR). These two articles were criticized for restricting freedom of speech and creating a chilling effect on journalists, artists, academics, and ordinary citizens (Al-Monitor, 2022).

For our research, we collected tweets containing the name "Sezen Aksu from 1 January 2022 to 24 January 2022, as the song was released on Youtube by the official account of Sezen Aksu on 30 December 2021. 581256 tweets were captured over 29 days by the TCAT Twitter Capture and Analysis Toolkit[i], a software suite developed by OILab, and the Digital Methods Initiative at the University of Amsterdam. Then we chose those 100 overtly hate speech content and coded them using MAXQDA. Any Turkish-speaking person can easily see the basic discourses being used in the tweets: If we translate the most used words in the tweets, including harassment and insults (excluding the words Sezen, Aksu, conjunctions, verbs, and irrelevant words such as https): Jewish, slut, monkey, atheist, Muslim, ignorant, HDP (People's Democratic Party), PKK, Armenian, insult, curse, lip, vagina, death penalty, bitch, her father, his daughter, CHP, with lips".

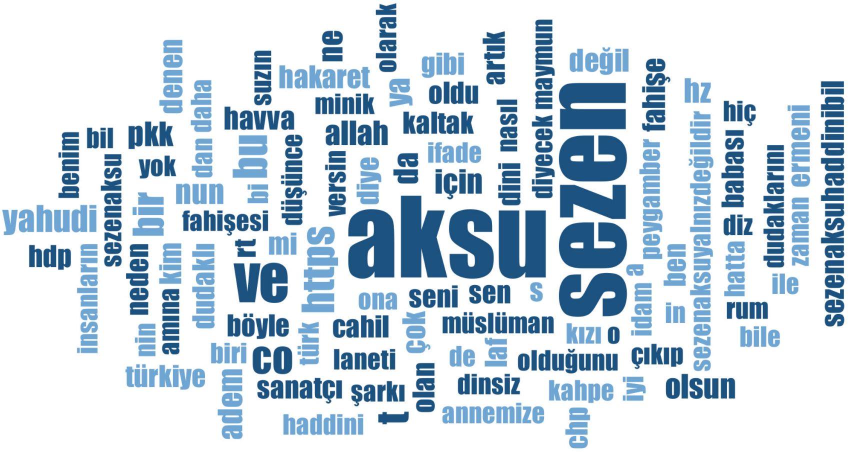


Figure 1 – Word cloud of hate tweets targeting Sezen Aksu (MAXQDA)

As previous research shows, hate speech in the new media environment can sometimes target one group while sometimes targeting many groups (Çomu & Binark, 2013). By relying on the MAXQDA analysis of 100 tweets, we grouped the tweets under the headings and the dominant themes even though there are often overlappings. The first three headings are religion based (56 times), on womanhood (40 times), and on the body and physical features (12 times) despite the overlaps. In the chapter, we did not use the address of the tweets in quoting. Instead, we used the number MAXQDA automatically assigned. We also translated the quotes directly without intervening in their grammar and other mistakes.

Religion-based hate speech tweets mostly touch upon two polarized discourses. While they mention the merit of Islam, in the same tweet, they swear at Sezen Aksu by claiming her being non-Muslim. She is claimed to be Jews or Armenian (these are used as insults) or evil. It is so striking that these both dimensions are interlinked: While the tweets using the religious frame argue that Sezen Aksu insults Islam and religious values with the songs downgrading holy prophets Adam and Eve, the tweets are full of hate, swearing with calls to mobilize, not to be silent, speak up, getting ready to punish her more.

> You prostitute, bitch, miserable Sezen Aksu. Who are you so that you can talk to Adam prophet, and our mother Eve. God damn you. (501305)

> Are those listening to "Sabetay Sevi crypto" Jews SEZEN AKSU from the same religion as her? Let's open our eyes now... Who is with who? We can not tolerate those who insult our values. ENOUGH IS ENOUGH, REALLY ENOUGH. Muslims should be awake against the children of satan. https://t.co/fselbw1n4S (515938)

In reply to the hashtag #sezenaksuyalnızdeğildir (meaning sezen aksu is not alone), another one also states she is satan:

> Of course, sezen aksu is not alone because the demon has many friends. I swear to God that we bring those and their satans together and certainly make them ready in hell with their knees bent down.

Forty *tweets targeting Sezen Aksu as a woman* very often use the word "bitch" in different ways. Her being middle-aged is emphasized by the word "with menopause ."These tweets also mostly continue with threats of "fucking" and rape. Her being immoral is emphasized in this tweet with the hashtag #SezenAksuHaddiniBil meaning Sezen Aksu know your place, know your limits:

> If Sezen Aksu were moral, she would not make such an album. God damn you, the daughter of a prostitute. The whole world f.ck you prostitute bitch #SezenAksuHaddiniBil (532267)

The threats of rape are extended to other women supporting Sezen Aksu, even men. As this tweet shows, again, the target is wifes and daughters of men:

> ... I shit down to Sezen Aksu's p.ssy, f.ck the wives and daughters of those defending her (477209)

Body and physical features are often used as insults in tweets. She is known with her beautiful voice and short height as "Serçe" (sparrow) in society. This has become another occasion to use as insults:

> Disgrace knee high... The little sparrow (Sezen Aksu) shitted and even screwed up. After menopause, some things happen to these shrewish. They then attack our religion and our sacredness. Who is going to say stop to them? Those from Mars? Enough is enough. Enough... (551764)

> Sezen Aksu is going through her menopause with an inferiority complex. So pity that sparrows have become garbage. Sezen Aksu aşağılık kompleksi içinde menopoz yaşıyor çok yazık serçeler çöp oldu çöp. (524641)

Animalization is related to insulting body features and works as an insult seven times. In another tweet again, she is targeted her title as "serçe" (sparrow) and transforms into a crow:

> How dare these Greek bitches? Know your limit; how dare you, tiny crow (Sezen Aksu).

Her age, as well as body features such as her thick lips, get insults in association to animals, such as 'lips with chimpanzee a..'.21613; "monkey lipped S. AKSU" (486858); "Kenia #Monkey Sezen Aksu, you would not talk like that about Islam if you expanded your brain rather than your lips #sezenaksuhaddinibil" (542992).

Hate speech directed at the *LGBTQI community* is seen in six tweets targeting Sezen Aksu even though she is not known for her lesbian identity. The opportunity is used to attack LGBTQI, as often practiced by Islamists in the last decade, as we have seen in the discussions on ending the Istanbul Convention membership of Turkey (Gencel Bek, 2022). It is used even in this case irrelevantly, either with Sezen Aksu's support of LGBTQI rights or by generalizing her as "artist fagots" ('sanatcı ibneleri') whose punishment is to remain in the fire of the hell (506025). Being *Armenian* is used as an insult in some tweets in a hateful way five times. Sezen Aksu is claimed to be Armenian just because she is not Turkish and Muslim, as can be seen in this tweet written with hate:

> Suzin Aksu, the main ignorance is to attack the religious values of a society... For example, you are not Turkish and Muslim. That situation then confirms that you are Armenian. Not Sezen, suzin #sezenaksu (505415)
>
> Her lyrics they did not like are related to her *using drugs* in some tweets:
>
> Did they not tell you not to use cocaine or white heroin, otherwise, you would be confused. B.TCH SEZEN Aksu. https://t.co/SgOAtzCraR (552141)

Not only LGBTQI or Armenians but all political others which were mentioned in the article as the repeated hate speech targets can also be seen in the tweet, sometimes even in the same tweet: FETO (six times), especially with the claim that her father was a FETO supporter ("F.ck the month of Sezen Aksu who is the daughter of FETOist) (60576); PKK was used ten times "PKK bitch who enlarged her lips in order to take more d.ck" (446386). In some tweets, different parties CHP (four times) and HDP (three times) used together as enemies, and Sezen Aksu is claimed to support them with insults again with the claim that a war is opened against Muslims and the Turkish Islamic world (https://t.co/BgEzre68UL, 363727). In the photos attached as memes by these tweets, on the first one, we see an implication that Sezen Aksu supports PKK, and the 'evidence' shown is her dress in red, green, and yellow, her hugging a Kurdish MP Sebahat Tuncel. In the second one, her father's presumably Jewish identity and FETÖ connection are foregrounded.

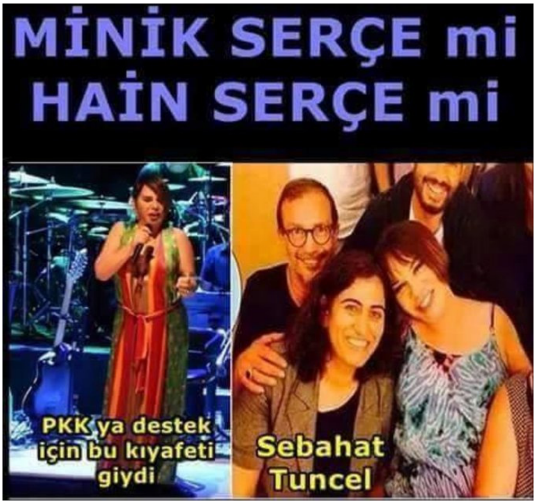


Figure 2 – 'Is she a small sparrow or traitor sparrow? She wore this dress to support PKK'.

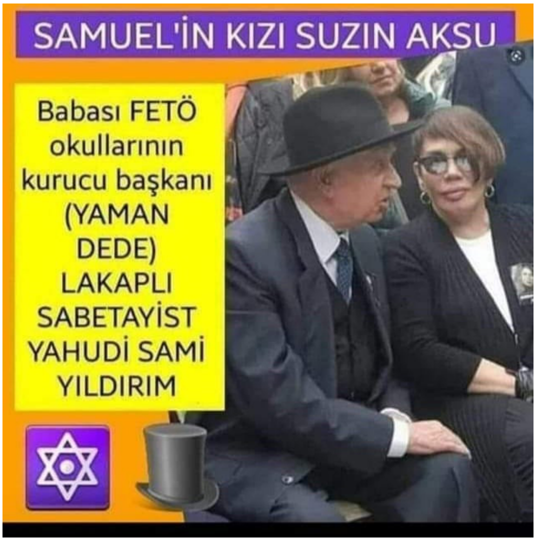

SAMUEL'İN KIZI SUZIN AKSU
Babası FETÖ okullarının kurucu başkanı (YAMAN DEDE) LAKAPLI SABETAYİST YAHUDİ SAMİ YILDIRIM

Figure 3 – 'Samuel's daughter Suzin Aksu. Her father is the founder member of FETO schools. Sabetayist Jews Sami Yıldırım so-called Yaman grandfather'.

As seen from the tweets quoted, Sezen Aksu is often asked to be punished. According to most of the tweets, that is her being raped. This is quite disturbing but put here to show how the people writing these tweets consider themselves in parallel with the state as the voice of the state in a violent and pathetic call of action:

> ... if Sezen aksu does not shut up, then the state of Turkey and political parties would rape her and enter their penis to her p.ssy and her a.shole (406665)

Besides rape, she was asked to be killed in different forms, such as stone to death, hanged, stoned, or burned. Some even declare all:

> 'The b.tch called Sezen Aksu should be hanged, killed with bullets, and burned alive.' #sezenaksuhaddinibil (496073)

Our case study revealed how hate speech is directed at the popular singer on different axis, including womanhood, LGBTQI, non-Turkish, and non-Muslim identities in the name of religion and Islam, as well as the association with animals as a hate object.

4. Discussion

In this section, we would like to summarize and discuss the counter attempts. In the Sezen Aksu case, we have already mentioned the lyrics of her new song, Hunter, which can be seen as resistance when we think that these lyrics were translated into more than 30 languages and shared on social media:

> **The Hunter (Avcı)**
> You can't make me sad
> I'm already very sad
> I see pain wherever I look at
> I see pain wherever I look at
> I'm the prey and you're the hunter
> Shoot me then...
> You can't sense me
> You can't crush my tongue
> I see pain wherever I look at

I see pain wherever I look at
Who is the traveler, who is the innkeeper
We'll see...
You can't kill me
I have my voice, my instruments, my words
When I say 'me', I am everyone

The translation and the circulation of these words certainly can be counted as resistance. Yet we should also emphasize that when social media, in accordance with populist politicians in power, target more ordinary people with not much power and use other forms of violence and subordination, such as the corrupted legal system and working conditions, the fragility of the subjects is more the case. That is especially when they are not active parts of collectivities. That deserves further attention from scholarly research. Matamoros-Fernández and Farkas (2021) already urge for more work using intersectionality in order to show the impact of capitalism (class), white supremacy (race), and heteropatriarchy (gender) on social media. Developing literature on the role of policies and affordances of social media platforms such as likes and sharing (Gerlitz & Helmond, 2013; Ben-David & Matamoros-Fernandez, 2016) can also be the sources of further study.

The struggle against hate speech is continuing around the world. Iginio Gagliardone et al. (2015) in their report published by UNESCO, list the mechanisms to counter online hate speech by reminding the need to balance freedom of expression and respect for human dignity. According to the report, the current international standards are one means of countering hate speech, such as the International Convention on the Elimination of all forms of Racial Discrimination; the International Covenant on Civil and Political Rights or ICCPR (Article 19). Besides these, the report lists these responses as complementary to legal actions. These are 1. Monitoring and analysis by civil society. 2. Individuals promoting peer-to-peer counter-speech 3. Organised action by NGOs to report cases to the authorities

4. Campaigning for actions by Internet companies hosting the particular content. 5. Media and Information Literacy.

Expanding the debate to the media is essential, especially when we consider the lack of quality public service broadcasting and media being used as the instrument of the government and social media users' hate speech campaigns to be transferred to media, as we have seen. In the report titled *Hate and Propaganda Media in Turkey: Affiliations, Models and Patterns*, Sinem Aydınlı (2020, p. 23) lists a series of policy recommendations, including mapping media ownership and financial investments, developing a shared platform and/or coalition as well as a self-regulatory body to scan, monitor and report hate speech on social media and producing a glossary and booklets to prevent hate speech. Besides these, the importance of the ombudsman system Faruk Bildirici (2021) can be added as well as the development of ethical guidelines by professional organizations such as the Turkish Journalists Association TGC.

The role of counter organizations in developing strategies against hate speech is also crucial in both developing an awareness of the general public and targeting hate groups and actions (Doğu, 2010), as we see in the case of Hrant Dink Foundation. The foundation was mentioned, especially with their work on deciphering hate speech in the media discourse, but that is evolving. Az et al. (2021) summarize the changing focus of Hrant Dink Foundation from revealing and criticizing the hate speech to discussing how to change and develop an alternative, as seen in the development of workshops to 'discuss, feel better and find an alternative since 2017' and building a new discourse, producing guidelines since 2018 as seen in the case of refugees (*Towards a New Discourse Training Manual*). Here, we should also mention the activities, such as developing an app called KarDes to discover the multicultural legacy and memory of the city of Ankara.[1]

[1] Multicultural legacy and memory of the city of Ankara: https://hrantdink.org/tr/bolis/faaliyetler/projeler/kulturel-miras/3819-kardesten-yeni-icerik-ankar

Using Youtube to disseminate counter ideas on hate speech is also a good strategy to target the younger generation. Another one is developing non-media ideas and linking even with the media: For example, making mantı (dumpling) together, as a popular food of diverse cultures and publishing on these and similar experiences, especially at Dumpling Post as open access in cooperation with IKSV İstanbul Foundation for Culture and Arts.[2]

Fethiye Çetin's (2022) explanation of the origin of the idea of organizing a dumpling festival can be our final words and be an inspiration to imagine alternatives even in oppressive contexts:

'The Dumpling Festival was Hrant Dink Foundation's response to the ban on the conference. The announcement of the festival was accompanied by a touching video that said the dumplings would be made together in workshops for children and adults, then eaten together, and that there would be discussions on 36 different kinds of mantı (dumplings) ranging from the Kayseri to the Circassian style, in a celebration of diversity.

'That's it!' I remember thinking as I jumped up from my desk. 'This may be the new way we are seeking to resist bans and pressure, to protest what is being imposed upon us, in order to overcome the human rights crisis with such actions.'

Oppressive regimes not only violate our rights but also take away our capacity to imagine new and creative forms of resistance. Many of us are focused on standing our ground, at whatever cost, rather than strengthening solidarity and challenging the oppression by producing and acting together. While it is important to stand firm, I think it is even more important to develop ways to overcome a crisis in order to move forward'.

[2] Dumpling Post: https://hrantdink.org/en/site-of-memory/announcements/3818-dumpling-post-is-published

Acknowledgements

I would like to thank my colleague Dr. Duygu Karataş for sharing the tweet corpus which was collected for the research project funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1472 "Transformations of the Popular" – 43857702.

References

Aydınlı, S. (2020). *Hate and Propaganda Media in Turkey: Affiliations, Models and Patterns*. SEENPM: Ljubljana.

Az E. I., Ensari, P. & Özkan, Ş. (2021). Hate speech in Turkey's print media: The experience of the Hrant Dink Foundation's Media Watch on Hate Speech project. E.İrem Az, A. Yılmaz. (ed.). *Media and Hate Speech. Perennial Questions, Current Debates.* Mas pub.

Ben-David, A. & Matamoros-Fernández, A. (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, 1167–1193 1932–8036/20160005

Bildirici, F. (2021). Ten years of ombudsmen in Turkey's media: The role of the reader representative in the fight against hate speech and discrimination. Altuğ Yılmaz. (ed.). *Media and Hate Speech. Perennial Questions, Current Debates*. Mas pub.

Cerit, O. (2010). LGBTT organizations: Black Pink Triangle.Ayse Cavdar, Aylin B. Yildirim (ed.) *Hate Crimes and Hate Speech*. Istanbul. International Hrant Dink Foundation, pp. 296-301

Çetin, F. (2010). Introduction. Translated by Suzan Bölme Ayse Cavdar, Aylin B. Yildirim (ed.) *Hate Crimes and Hate Speech*, Hrant Dink. Istanbul. International Hrant Dink Foundation. 4-75.

Çetin, F. (2022). Thoughts on the Dumpling Festival. *Manti Postasi*. September.

Çomu, T. & Binark, M. (2013). Yeni Medya Ortamlarında Nefret Söylemi. (ed.) Mahmut Çınar. *Medya ve Nefret Söylemi*. İstanbul: Hrant Dink Vakfı.

Demiryakan, B & Ensari, P. (2017). *Discriminatory Discourse in News Stories on Murders of Transgender Women*. Translated by Cansen Mavituna. İstanbul: Hrant Dink Vakfı.

Dirini, İ. & Özsu, G. (2020). *COVID-19 Pandemi Sürecinde Sosyal Medyada Nefret Söylemi Raporu*. Ankara: Alternatif Bilişim Derneği.

Doğanay, Ü. & Çoban Keneş, H. (2016). Yazılı Basında Suriyeli 'Mülteciler': Ayrımcı Söylemlerin Rasyonel ve Duygusal Gerekçelerinin İnşası . *Mülkiye Dergisi*, 40 (1), 143-184.

Doğu, B. (2010). Sanal Nefret Pratikleri: İnternette Nefret Söylemi ve Karşı Örgütlenmeler. *Yeni Medyada Nefret Söylemi*. Tuğrul Comu (ed.). Ankara: Kalkedon.

Erbaysal Filibeli, T. & Ertuna, C. (2021). *Sarcasm Beyond Hate Speech: Facebook Comments on Syrian Refugees in Turkey*. International Journal of Communication 15, 2236–2259 1932–8036/20210005

Gagliardone, I. (2019). Defining Online Hate and Its "Public Lives": What is the Place for "Extreme Speech"? *International Journal of Communication* 13, 3068–3087 1932–8036/20190005

Gagliardone, I., Gal, D., Alves, T. & Martinez, G. (2015). *Countering Online Hate Speech*. Paris: UNESCO 2015

Gencel Bek, M. (2012). "Ataerkillik, Piyasa ve Mesleki Değerler: Medyada Aile İçi Şiddetin Temsili ve Üretim Pratikleri". (Patriarchy, Market and Professional Values: Representation of Domestic Violence in the Media and Production Practices) (ed.) Serpil Sancar. *Birkaç Arpa Boyu...21. Yüzyıla Girerken Türkiye'de Feminist Çalışmalar*. Prof. Dr. Nermin Abadan Unat'a armağan. İstanbul: Koç University, 649-678

Gencel Bek, M. (in print). Gendered Violence, Populism and (Social) Media in Turkey. Caroline Williamson Sinalo and Nicoletta Mandolini (eds.) *Representing Gender-Based Violence: Global Perspectives*. Palgrave.

Gerlitz, C. & Helmond, A. (2013). The like economy: Social buttons and the data-intensive Web. *New Media & Society*, 15(8), 1348-1365.

Göktas, K. (2010). How did the media make Hrant Dink a target. Ayse Cavdar, Aylin B. Yildirim (ed.) *Hate Crimes and Hate Speech,* Hrant Dink. 2010. Istanbul. International Hrant Dink Foundation, 85-96

Hrant Dink Foundation (2019). *2019 report Hate Speech and Discriminatory Discourse in Media*. İstanbul: HDV.

İnceoğlu, Y. & Sözeri, C. (2012). Nefret Suçlarında Medyanın Sorumluluğu: "Ya sev ya terk et ya da..." İnceoğlu, Y. *Nefret söylemi ve-veya nefret suçları* [Hate Speech and/or Hate Crimes]. İstanbul: Ayrıntı Publishing, 23-38.

Jakubowicz, A. (2017). Alt_Right White Lite: trolling, hate speech and cyber racism on social media. *Cosmopolitan Civil Societies: an Interdisciplinary Journal*/Vol. 9, 3. http://dx.doi.org/10.5130/ccs.v9i3.5655

Köker, E. & Doğanay, Ü. (2010). They shot Turkey: Press failed to name Hrant Dink murder. Ayse Cavdar, Aylin B. Yildirim (ed.) *Hate Crimes and Hate Speech*, Hrant Dink.. Istanbul: International Hrant Dink Foundation, 97-118

Matamoros-Fernández, A. & Farkas, J. (2021). Racism, Hate Speech, and Social Media: A Systematic Review and Critique. *Television & New Media*, 22(2), 205–224. https://doi.org/10.1177/1527476420982230

Müller, K. & Schwarz, C. (2021). Fanning the Flames of Hate: Social Media and Hate Crime, *Journal of the European Economic Association*, Volume 19, Issue 4, August, Pages 2131–2167, https://doi.org/10.1093/jeea/jvaa045

Perry, B. & Olsson, P. (2009). Cyberhate: the globalization of hate. *Information & Communications Technology Law*, 18, 185-199.

Şensever, L. (2010). Dur De Initiative to Say Stop Racism and Nationalism. Ayse Cavdar, Aylin B. Yildirim (ed.) *Hate Crimes and Hate Speech*.. Istanbul: International Hrant Dink Foundation, 289-295

Tar, Y. (2021). In search of sentences with hidden subjects: LGBTI+ representation in the media. Altuğ Yılmaz. (ed.). *Media and Hate Speech. Perennial Questions, Current Debates*. Mas pub.

Udanor, C. & Anyanwu, C. C. (2019). Combating the challenges of social media hate speech in a polarized society. A Twitter ego lexalytics approach. *Data Technologies and Applications*, 53(4), 501-527.

Yılmaz A. (2022). (ed.). Tüm Yönleriyle Nefret Söylemi Kılavuz. Article 19. translated by Bülent Çınar İstanbul: HDV.

Yılmaz, A. (2021). (ed.). *Towards a New Discourse: Training Manual.* Istanbul: HDV.

Yumul, A. (2021). "The great chain of being" from religious worldview to secular nationalism. Altuğ Yılmaz. (ed.). *Media and Hate Speech. Perennial Questions, Current Debates*. Mas pub.

References (websites)

Al-Monitor (2022). https://www.al-monitor.com/originals/2022/01/turkeys-pop-queen-draws-sudden-wrath-over-five-year-old-lyrics#ixzz7mgmSjbuG

Bianet English (2022). https://bianet.org/english/freedom-of-expression/256346-singer-sezen-aksu-faces-threats-complaint-over-song-from-five-years-ago

Birgün (2022). https://www.birgun.net/haber/rtuk-ten-kanallara-sezen-aksu-tehdidi-374069

Duvar English (2022).https://www.duvarenglish.com/turkish-pop-star-sezen-aksu-targeted-over-2017-song-referencing-eve-and-adam-news-60160

Duvar English (2022). https://www.duvarenglish.com/erdogan-claims-he-wasnt-referring-to-sezen-aksu-when-threatening-to-cut-tongues-news-60237

Duvar English (2022). https://www.duvarenglish.com/erdogan-threatens-legendary-singer-sezen-aksu-it-is-our-duty-to-cut-those-tongues-news-60193

Euronews (2022). https://www.euronews.com/culture/2022/01/18/turkish-pop-star-sezen-aksu-slammed-by-islamist-groups-for-attacking-traditional-values

Gazete Duvar (2022). https://www.gazeteduvar.com.tr/diyanetten-sezen-aksu-aciklamasi-en-hafif-tabirle-saygisizliktir-haber-1549612

Medyatava (2022). https://www.medyatava.com/haber/erdem-ozveren-hedef-gosterdi-sezen-aksu-gundem-oldu_242342

Yeni Şafak (2022). https://www.yenisafak.com/hayat/sezen-aksudan-dini-degerlere-hakaret-sahane-bir-sey-yasamak-sozleri-nedir-sarkisinda-ne-dedi-3730397

Chapter 5

HATE SPEECH ON TWITTER: THE LGBTIQ+ COMMUNITY IN SPAIN

Patricia de-Casas-Moreno
University of Extremadura, Spain

Macarena Parejo-Cuéllar
University of Extremadura, Spain

Arantxa Vizcaíno-Verdú
University of Huelva, Spain

1. Introduction

The visibility of hate speech is a rising concern on the cyberspace according to the Anti-Defamation League's (2021) *Online Hate and Harassment* report. Since 2018, this report highlights the uncontrolled advance of this sort of discourses. These results stem from the relationship between the online and offline environment. This means that messages posted on social media are linked to the behaviors in which society participated until now in traditional media (Olmos et al., 2020).

According to Martínez-Valerio (2021), social media promotes hate crime, mainly among the younger audience. Nevertheless, these platforms are continuously seeking to adopt measures to prevent the spread of this kind of discourse. Due to the growth and visibility of these harmful and abusive comments, videos and reactions, these platforms face their daily monitoring and prevention (Miró, 2016). In fact, the Spanish Penal Code revision carried out in 2015 notes that hate speech sentences increase when the message is likely to be virally disseminated through mass media (Carratalá & Herrero-Jiménez, 2019).

The LGBTIQ+ collective (Lesbian, Gay, Bisexual, Transgender, Intersex, Queer and other identities not included in the above) is the most targeted community in online contexts because of their sexual orientation. This concept emerged in the 90s and included recently the symbol "+" to refer to all those individuals not listed in the previous acronyms who are gaining greater visibility in society (Vila, 2019). This movement gained popularity for its social resistance against heteronormativity, which advocates for the rights of the collective since June 26, 1977. After 50 years, the community has succeeded in legally enforcing much of its rights on a legal and social stage. Every June 28th, LGBTIQ+ Pride Day is celebrated around the world, which is a date on which this community symbolizes the claim for their rights. This event commemorates the riots that took place in 1969 at the Stonewall Inn in New York. In Spain, the recent approval of the Law for the real and affective equality of trans people and for the guarantee of the rights of LGBTIQ+ people ensure the rights and prevention of harmful attitudes against this group from the health, education and family environment. In this way, Spain is one of the ten European countries with the highest visibility and recognition of LGBTIQ+ equality (ILGA-Europe, 2021).

The following critical review provides a theoretical framework of the phenomenon in Spain, introducing some of the most media representative cases. Our purpose is to outline the current scenario of this group in Spain, mapping the prevailing hate discourse through Twitter.

2. Hate speech in the digital arena

Social movements find the opportunity to make visible their political claims, social practices and identity inclusion in the current digital context. Following Olmedo-Neri (2019), today's social movements focus on calling for the acceptance and inclusion of the individual and collective human identity in the face of the meta-storytelling created and disseminated by the Internet.

Social media enhance political movements and citizen debates, constituting the most popular venue for the mobilization of social groups. As Lagares et al. (2021) point out, these digital communities create a new mode of participation, socializing new generations who build partisan networks. Consequently, Garay (2018) states that these initiatives within the digital environment facilitate the visibility of traditionally discriminated groups, which strengthen cooperation strategies against potential social challenges.

Although the media historically was a hostile forum against the LGBTIQ+ community, in the last few decades it helped to challenge social injustices. These efforts were associated with the labor of communication departments and associations who advocate for the rights of the collective as a newsworthy topic. This led to an increased coverage of news about sexual diversity. Consequently, we have experienced a relentless informative approach to such facts. As a result, there is a new social, creative and active audience that seeks to modify the prevailing heteronormative discourse, to which the media must react with plurality and truth (Carratalá & Herrero-Jiménez, 2019).

In light of these events, we introduce two critical concepts: hate speech and homophobia. According to the United Nations Organization (UN, 2019), hate speeches constitute behaviors or offenses through pejorative or discriminatory expressions directed to another person because of his or her individual status (religion, sexual orientation, race, or any other factor related to identity). On the other hand, Gagliardone et al. (2015) point out that hate speech is related to those expressions that prompt acts of discrimination on different grounds (ethnicity, gender or sexual orientation). Additionally, Megías (2020) defines the concept of homophobia as the irrational fear and aversion to homosexuality, which is supported by prejudices traditionally manifested in the society. Following Allport (1954), pioneer of social psychology theories, human being feels preference for prejudice, creating myths and hostile attitudes towards a specific group. Multiple studies analyze the hate speech directed against the LGBTIQ+ collective known as LGTB-phobia (Wilhelm et al., 2020; Ferré-Pavia & Zaldívar, 2022). These studies suggest that this group is often the victim of discrimination, exclusion and stigmatization due to the establishment of hegemonic standards and the development of stereotypes.

In social media, the rise of hate speech is growing steadily (Domínguez-Fuente et al., 2016). Kaufman (2015) considers that a message is hateful when it matches four basic criteria: (1) the group to which the message is directed is in a vulnerable position, (2) and humiliation (3) and malignity (4) are exercised against them purposefully. Despite social awareness of human rights, we find that anonymity when disseminating these messages in online environments provokes those criteria, which often result in violent behavior. For Bustos et al. (2019), this response, outside any social morality, goes beyond anonymity by bringing into play other aspects such as the feeling of group belonging (avoidance of isolation). They also suggest that these attitudes seek positive reinforcement from a dominant group that legitimizes its position through racist or xenophobic comments. Another distinctive factor is that these offensive groups tend to construct a collective hate discourse based on the overabundance of information and the decentralization of content. In other words, a hateful message could go viral and generate an unstoppable multiplier effect that intensifies hate speech and discriminatory stereotypes (Cabo, 2017).

3. Social media and the LGBTIQ+ community

Beyond being venues for communication, social media are also a common arena for social interaction in which prosumers share information. In this context, the socialization process plays a major role, generating a network of online networking interaction (García-Calderón & Olmedo-Neri, 2019). Collective and organized movements appear to resist discrimination and repression caused by the status quo (Olmos et al., 2020).

The popularity of these platforms generated an important debate concerning the mental health of users. These apps serve as social support, avoiding the exclusion of certain groups or individual identities (Macías-Marcelino, 2022). However, depending on the context, they are used for harmful purposes that can injure the morals and self-awareness of some citizens (Ramos-Soler et al., 2018). Addressing Yudes-Gómez et al. (2018), the evolution of discrimination in this sort of platforms generally starts with mockery, cyberbullying, and other types of offenses.

Several studies show that more than 40% of LGBTIQ+ individuals experience cyberbullying because of their sexual orientation (Tinoco-Giraldo et al., 2021). They also note that these assaults are often anonymous, leading to serious consequences on the self-esteem and mental health of young people (Morell-Menguall et al., 2020). Nevertheless, with the rise of digital moderation systems, the perpetrators are easily identified and sentenced.

In this regard, we must be aware of some types of cyberthreats targeting the LGBTIQ+ community. Firstly, (a) *sextortion* demands a large amount of money in order not to spread intimate content to the public. Secondly, (b) *LGTBphobic* cyberbullying, considered the most frequent threat through adverse comments about a person's sexual orientation. Thirdly, (c) *fraping*, which is the creation of a fake profile to impersonate someone's identity. Fourthly, (d) *wokefishing*, which is a technique used by cybercriminals to create fake profiles pretending to hold an ideological position related to a collective or social movement. However, they are ruthless individuals who deceive the younger population with the aim of stealing, threatening or extorting. Finally, (e) *mobbing*. This kind of harassment also occurs in the digital environment, prompting permanent hostile attitudes towards people because of their sexual orientation (Europapress, 2022).

The LGBTIQ+ community is one of the most affected by this sort of hate crime on the Internet. On the basis of the Comisión Europea's (2021) sixth Report on the *Code of Conduct on Combating Hate Speech Online*, sexual orientation is the most common reason due to hate speech (18.2%), followed by xenophobia (18%), and anti-Gypsyism (12.5%). The European Union Agency for Fundamental Rights explains that every tenth LGBTIQ+ person in Europe has reported feeling discriminated against. In Spain, the report of the Ministry of the Interior on the *Evolution of Hate Crimes in 2021* shows a total of 1,802 cases, in which 466 cases (25.86%) are related to sexual orientation and identity (Unión Europea, 2020).

Furthermore, a study by Amores et al. (2022) emphasizes that there are users in social media who spread homophobic and hateful discourse, claiming their right to freedom of expression. They have also observed that there are comments that deny the existence of sexual diversity. Debates on this topic have been widely politicized and polarized, leading to the banalization of the human rights (Riviera-Martín et al., 2022).

4. Spanish LGTB-phobic case studies

Social media make visible the daily experiences of several traditionally marginalized groups. However, the anonymity promoted by these platforms facilitates the rise of hate speech (Martínez-Valerio, 2021), especially directed towards the LGBTIQ+ community. In this context, those who show their identity experience cases of humiliation, denigration, harassment, threats, among other types of abuse. In the following, we will explore some Spanish examples of hate speech against the LGBTIQ+ community on Twitter. We decided to anonymize the identities of the victims in order to protect and preserve their privacy.

Firstly, being the first transgender woman to be elected to a parliamentary representative position in Spain did not make many people happy. The activist we will mention as *@case1*, pointed out that social media is like the "gateway to hell". She reports that she received death threats on several occasions through Twitter. Fear and insecurity have been the costly price she paid for making visible and sharing her sexual orientation. Her testimony also illustrates the normalization of certain transphobic comments, such as the one she experienced in 2019. Under the pseudonym "This is

Spain", a user launched the following message against her position: "The first man with a cut penis and operated tits in the Parliament. You are not what you want to be, but what you decide to be". The political leaders did not hesitate and decided to denounce this person promptly to the social media, but Twitter did not consider an offense against its rules of use. Hundreds of users concerned with this case supported *@case1*, making the microblog rectify and close the bully's profile for defamation and homophobic hate crime.

The testimony of *@case1*, of course, is not unique. A survey conducted by the European Agency for Human Rights (2019) reveals that nearly 70% of LGTBIQ+ people have experienced hatred through social media in recent years. Sometimes, these behaviors go viral and cause severe moral consequences among the victims. An example of this is the story of *@case2*, a 23-year-old transgender woman. She tells that the hate speech began with the posting of a video in an assembly of several feminist collectives in which the young woman participated as a member. Within hours, dozens of unknown users interacted with her Twitter profile, attacking her privacy. Different individuals revealed where she lived, where she studied, among other aspects and details of her private life simply because she was transgender. The student describes the psychological impact of this abuse, explaining that not only her privacy was harmed, but also her identity and mental wellbeing (Público, 2022). Even people close to her were victims of this hatred just because they had a personal relationship with her.

Reputational injury to the victims as a result of these acts becomes highly concerning. "Do I take out the gun or punch?" is the viral tweet that turned the life of a young 25-year-old activist, *@case3*, into a nightmare. From this post in 2015, different messages were published in chain insulting or threatening the boy with comments like "fucking faggot" or "we're going to stick our bats up your ass". However, *@case3* reacted to the harassment, waited until he was of legal age and presented the case to the courts. Although he reported more than twenty tweets, the police only found the author of one who was sentenced to nine months in prison.

This same legal challenge was initiated by the actress and teacher *@case4*. A user posted the following message on Twitter: "...prototypical faggot with boobs. She can't stand that I am a woman and has a pathological grudge against me". The post was followed by a photo of the victim. Soon after, different people spread the tweet via WhatsApp. From that moment on, @ *case4* felt compelled to explain his sexual identity. Despite the harassment she had to endure, the judicial reports helped her recover some self-esteem because they determined that the perpetrator was belligerent and felt animosity towards trans people who have not undergone genital reassignment. In other words, the justice system considered the aggressor to be a person with mental health problems. The victim, moreover, knew her. She was a close friend with whom "she had no relationship since her transition" (López, 2022). However, these instances are not common. Most often, the aggressors are mostly profiles and people unknown to them (Público, 2022).

Anonymity in social media facilitates creating and disseminating hate speech towards the LGTBIQ+ community without thinking about the consequences (Rivera-Martín et al., 2022). However, these platforms should be faithful to their commitment to society by avoiding the spreading of this kind of discourse (Martínez-Valerio, 2021; Aguiló, 2020). Even more so, when the European Commission and the platforms Facebook, Twitter, YouTube and Microsoft assumed and made public a behavior code to prevent and stop hate speech online. For six years the report's assessments were promising. These companies and government agencies found that the monitoring of harmful messages and posts decreased. Nevertheless, today there is still a weakness in terms of the inadequate response of their mediators to deal with user complaints (Comisión Europea, 2021).

In the face of these results, the State Federation of Lesbians, Gays, Transgender, Bisexuals, Intersexuals and others, claims that hate crimes have not been eradicated. In contrast, they have only evolved, moving from the physical realm to social media and especially to Twitter. Attending to Hidalgo's words (2022), for the moment there is no change in the criminal behavior because instead of listening to "fucking faggot", we now see it written digitally. Such lack of monitoring produces serious consequences on mental health when the abusive mistreatment is permanent. In these cases, the victims are not able to endure it and even if they report it, they ultimately choose to suicide (Pinargote et al., 2022). This is what happened to *@case5*, a 17-year-old girl from Galicia, and *@case6*, another 20-year-old girl who lived in Jaén (both from Spain). Both decided to suicide because they could not stand cyberbullying as a result of their sexual orientation (Puga, 2020). The LGTBIQ+ community accused that no one did anything to prevent it, knowing the circumstances of these girls. The fact that they belonged to small localities made the social response even more passive on the part of the agencies and those responsible for the platforms. The digital context invisibilized these cases, increasing the suffering of the victims (Puga, 2020).

Something similar experienced *@case6*, a 20-year-old girl from Navas de San Juan in Jaén. In this case, her identity was supplanted on Twitter to damage her image (Público, 2022). In addition to stealing her profile, they sent her insulting and obscene messages. In this assault, they even published her phone number on dating sites. The victim reported the incident several times, but it was too late to bring the case to court. Her family's representative declared to the media that “LGTBIphobia kills”. This quote was used from that moment on to condemn this reality on other social networking sites, which are the possible weapon used by the perpetrators.

Finally, we emphasize the involvement of celebrities in such assaults. One of the most popular cases involve *@case7* from a Spanish series. The actor plays a transsexual woman in the series. Due to his role, this actor received multiple injuries that he made public – e.g. “Fucking faggot, son of a bitch, I hope you get beaten, transvestite”. He reported the case to the police, but at the same time asked his followers to report the profile, which was finally removed from the platform. In the same way that we found local cases invisibilized by macro-platforms and justice systems that resulted in tragic endings, we find that celebrities are a leading force against such offenses. This relies on the phenomenon known as *fanbullying*, whereby people target an actor or actress for performing an immoral fictitious role (Vizcaíno-Verdú et al., 2020). At this point, we find a gap of attention to vulnerable groups such as LGBTIQ+, where it seems that the algorithm and moderation does not pay attention to those cases that require the same protection and guarantee of equality for the expression of individual-collective identity.

5. Discussion and conclusions

As we discussed, nobody is exempt from hate speech through social media. Although this chapter is locally focused in Spain, it is just one example of the hate speech that the LGBTIQ+ community constantly faces.

Hate speech increased exponentially in the last few years, mainly since the rise of social media. This kind of behaviour is a particular concern nowadays because of the multiple means of message spreading, which results in harmful consequences for certain minority groups (Isasi & Juanatey, 2017).

According to Ródenas (cited in Rivera-Martín et al., 2022), hate speech is punishable, while pejorative speech is not always so, as it may be interpretable. The author also points out the evidence that most of these speeches in social media are not reported due to the complexity of the system. The reason for this is that many servers are outside the country and the proceedings are difficult and time-consuming.

In this review we underline that despite the progress achieved for the protection of the rights of this community, there are still several citizens who deny their inclusion, spreading harmful messages.

On the other hand, focusing on the platforms, we could note that Twitter has promoted communication and social activism, driving collective efforts on behalf of this community. Even though, the LGBTIQ+ community faces a hostile arena in which prevail uncontrolled speeches against their sexual orientation, resulting in physical and psychological consequences for those affected. Studies such as that of Piñeiro-Otero and Martínez-Rolán (2021) show the toxicity of Twitter for some minority groups that are unprotected by anonymity.

Dealing with hate speech has become a daily issue in the political and media environment. This reality to which the LGBTIQ+ community is continually confronted demands awareness and the reinforcement of measures to stop such hostility. Following Bonet-Martí (2021), it is imperative to take action to stop the spread of rumors and fake news, as well as to address the spread of cyber-violence and LGTBphobia.

We think that in Twitter it is essential to strengthen the control over this sort of discourse. Even though at the beginning of 2021 this social media announced a pilot project in which users of the application could report offensive and sensitive content. However, further efforts to achieve a more equanimous and tolerant online community are still needed. Recent events involving CEO Elon Musk have left the platform in the spotlight. It seems that what was considered a forum to debate and freedom of expression for all, is now branded as the "realm of surveillance and censorship". In other words, Twitter is a platform that is far from ensuring the well-being of a socially and historically discriminated community via just a few characters.

Funding agency

This work was supported by 'Alfamed' (Euro-American Research Network), under Grant R+D+I Project (2019-2021), entitled "Youtubers and Intagrammers: Media competition in emerging prosumers", with code RTI2018-093303-B-I00, financed by the Spanish Ministry of Science, Innovation and Universities and the European Regional Development Fund (ERDF), and the R+D-i Project (2020-2022), entitled "Instagrammers and youtubers for the transmedia empowerment of Andalusian citizens. The media competence of instatubers", with code P18-RT-756, financed by the Andalusian Regional Government in the 2018 call (Andalusian Research, Development and Innovation Plan, 2020) and the European Regional Development Fund (ERDF).

References

Aguiló, B. & Giró, X. (2020). *Análisis multimodal de las publicaciones de los perfiles de Instagram de La Vanguardia, El País, El Mundo y Ara durante las protestas en Barcelona contra la sentencia del "Procés"*. (Tribunal Supremo. Causa Especial núm. 3/20907/2017 Sentencia núm. 459/2019). (Màster Universitari en Mitjans, Comunicació i Cultura). https://ddd.uab.cat/record/237117?ln=ca

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.

Amores, J. J., Blanco-Herrero, D., Sánchez-Holgado, P. & Frías-Vázquez, M. (2021). Detectando el odio ideológico en Twitter. Desarrollo y evaluación de un detector de discurso de odio por ideología política en tuits en español. *Cuadernos.info*, *49*, 98-124. https://doi.org/10.7764/cdi.49.27817

Anti-Defamation League. (2021). *Online Hate and Harassment. ADL's Center for Technology and Society*. http://bit.ly/3tIblJm

Bonet-Martí, J. (2021). Los antifeminismos como contramovimiento: una revisión bibliográfica de las principales perspectivas teóricas y de los debates actuales. *Teknokultura. Revista de Cultura Digital y Movimientos Sociales*, 18(1), 61-71. https://doi.org/10.5209/tekn.71303

Carratalá, A. & Herrero-Jiménez, B. (2019). La regulación contra el discurso de odio hacia el colectivo LGTBI en los medios: análisis comparado de diez leyes autonómicas. *RAEIC, Revista de la Asociación Española de Investigación de la Comunicación*, *12*(6), 58-80. https://doi.org/10.24137/raeic.6.12.3

Comisión Europea. (2021). *6º Informe de Monitoreo sobre el Código de Conducta para la lucha contra el discurso de odio ilegal*. Publications Office European Comission. http://bit.ly/3UU9LAb

Domínguez-Fuente, J. M., García-Leiva, P. & Hombrados-Mendieta, M. I. (2016). *Transexualidad en España. Análisis de la realidad social y factores psicosociales asociados*. http://bit.ly/3UU9U6H

Europapress. (2022). *El 40% de los adolescentes LGBT han sufrido ciberacoso por su orientación sexual*. http://bit.ly/3UTBaSz

Garay, L. M. (2018). Colectivos de diversidad sexual, redes sociodigitales y ciberactivismo como escenario de visibilidad. En J. Candón-Mena (Coord.), *Actas del II Congreso Internacional Move.net sobre Movimientos Sociales y TIC* (pp. 92-108). Compolíticas. http://bit.ly/3TQ3W5u

García-Calderón, C. & Olmedo-Neri, R. A. (2019). El nuevo opio del pueblo: apuntes desde la Economía Política de la Comunicación para (des) entender la esfera digital. *Iberoamérica Social*, *12*, 84-96. http://bit.ly/3Olih8G

Gagliardone, I., Gal, D., Alves, T. & Martínez, G. (2015). *Countering online hate speech*. UNESCO. https://bit.ly/3zTqwAc

Ferré-Pavia, C. & Zaldívar, G. (2022). El feminismo trans excluyente en Twitter: un monólogo sesgado en #ContraElBorradoDeLasMujeres. *ICONO 14*, *20*(2). https://doi.org/10.7195/ri14.v20i2.1865

Hidalgo, Carlos (2022, March 6). Crecen un 30% los delitos de odio por redes: ya son ocho de cada diez casos que llegan a la Policía. *ABC*. https://bit.ly/3XqPYdf

ILGA-Europe. (2021). *Annual review of the human rights situation of lesbian, gay, bisexual, trans, and intersex people 2021*. http://bit.ly/3Oo3z0B

Isasi, A. C. & Juanatey, A. G. (2017). *El discurso del odio en las redes sociales: Un estado de la cuestión (Hate speech on social media: A state of the art).* Ajuntament de Barcelona Progress Report.

López, N. (2022, March 31). La batalla judicial de Violet Ferrer, acosada en redes sociales por ser trans: “Todo el mundo en el trabajo se enteró a través de Twitter de que no estaba operada”. *Newtral*. https://bit.ly/3TZjRhF

Macías-Marcelino, D. C. (2022). *Percepción de las y los jóvenes transexuales de la transfobia en las redes sociales.* [Trabajo Final Grado, Universidad de la Laguna]. http://bit.ly/3XjSZfq

Martínez-Valerio, L. (2021). Mensajes de odio hacia la comunidad LGTBIQ+: análisis de los perfiles de Instagram de la prensa española durante la “Semana del Orgullo”. *Revista Latina de Comunicación Social*, *80*, 363-388. https://doi.org/10.4185/RLCS-2022-1749

Megías, I., Amezaga, A., García, M. C., Kuric, S., Morado, R. & Orgaz, C. (2020). *Romper cadenas de odio, tejer redes de apoyo: los y las jóvenes ante los discursos de odio en la red.* Centro Reina Sofia sobre Adolescencia y Juventud. https://doi.org/10.5281/zenodo.4288486

Ministerio del Interior. (2021). *Informe de la encuesta sobre delitos de odio. Dirección General de Coordinación y Estudios.* http://bit.ly/3EibiJ8

Miró, F. (2016). Taxonomía de la comunicación violenta y el discurso del odio en Internet. *IDP: Revista de Internet, Derecho y Política*, *22*, 93-118. http://bit.ly/3TUeY9L

Morell-Mengual, V., Gil-Llario, M. D. & Gil-Juliá, B. (2020). Prevalencia e influencia de la violencia homofóbica sobre la sintomatología depresiva y el nivel de autoestima. *Información Psicológica*, *120*, 80-92. https://doi.org/10.14635/IPSIC.2020.120.6

Ramos-Soler, I., López-Sánchez, C. & Torrecillas-Lacave, T. (2018). Online risk perception in young people and its effects on digital behaviour. [Percepción de riesgo online en jóvenes y su efecto en el comportamiento digital]. *Comunicar, 56*, 71-79. https://doi.org/10.3916/C56-2018-07

Olmedo-Neri, R. A. (2019). Los medios en la inclusión de la diversidad sexual en la Ciudad de México. *Revista Internacional de Ciencias Sociales Interdisciplinarias*, *7*(2),187-200. https://doi.org/10.18848/2474-6029/CGP/v07i02/187-200

Olmos, A., Rubio, M., Lastres, N. & Martín, P. (2020). *Jóvenes, redes sociales virtuales y nuevas lógicas de funcionamiento del racismo: Etnografía virtual sobre representaciones y discursos de alteridad e identidad*. Centro Reina Sofia sobre Adolescencia y Juventud. https://doi.org/10.5281/zenodo.3666178

Pinargote-Vinces, G. J., Maldonado-Zuñiga, K., Pin-Menéndez, C. Y. & Pérez-Chilán, D. L. (2022). Uso de Internet por parte de los jóvenes y dependencia de los teléfonos móviles. *UNESUM-Ciencias*, *6*(3), 20-30. https://doi.org/10.47230/unesum-ciencias.v6.n3.2022.471

Piñeiro-Otero, T. & Martínez-Rolán, X. (2021). Say it to my face: analysing hate speech against women on Twitter. *Profesional de la Información*, 30(5), e300502. https://doi.org/10.3145/epi.2021.sep.02

Público (2022, January 3). Una joven de 20 años se suicida tras ser acosada en redes por su orientación sexual. *Público*. https://bit.ly/3GBVyn6

Puga, N. (2020, Octuber 2). Una joven de 17 años se quita la vida en Galicia tras sufrir acoso por su orientación sexual. *El Mundo*. https://bit.ly/3i6pgq9

Rivera-Martín, B., Martínez-de-Bartolomé-Rincón, I. & López-López, P. J. (2022). Discurso de odio hacia las personas LGTBIQ+: medios y audiencia social. *Revista Prisma Social*, *39*, 213–233. http://bit.ly/3hZEaOY

Tinoco-Giraldo, H., Torrecilla-Sánchez, E. M. & García-Peñalvo, F. J. (2021). An analysis of LGBTQIA+ university students perceptions about sexual and gender diversity. *Sustainability*, 13(21), 11786. https://doi.org/10.3390/su132111786

ONU. (2019). *La estrategia y plan de acción de las Naciones Unidas para la lucha contra el discurso de odio*. https://bit.ly/3BCAeb6

Unión Europea. (2020). *A long way to go for LGBTI equality*. http://bit.ly/3Et2b8E

Vila, L. (2019). *¿Qué significan las siglas LGBTIQ+?* http://bit.ly/3AuN5OH

Vizcaíno-Verdú, A., Contreras-Pulido, P. & Guzmán-Franco, M. D. (2020). Construcción del concepto fanbullying: Revisión crítica del acoso en redes sociales. *Pixel-Bit*, *50*, 211-230. https://doi.org/10.12795/pixelbit.2020.i57.09

Wilhelm, C., Joeckel, S. & Ziegler, I. (2020). Reporting hate comments: Investigating the effects of deviance characteristics, neutralization strategies, and users' moral orientation. *Communication Research*, *47*(6), 921-944. https://doi.org/10.1177%2F0093650219855330

Yudes-Gómez, C., Baridon-Chauvie, D. & González-Cabrera, J. (2018). Cyberbullying and problematic Internet use in Colombia, Uruguay and Spain: Cross-cultural study. [Ciberacoso y uso problemático de Internet en Colombia, Uruguay y España: Un estudio transcultural]. *Comunicar, 56*, 49-58. https://doi.org/10.3916/C56-2018-05

Chapter 6

CIRCULATION SYSTEMS, EMOTIONS, AND PRESENTEEISM: THREE VIEWS ON HATE SPEECH BASED ON ATTACKS ON JOURNALISTS IN BRAZIL

Edson Capoano
University of Minho, Portugal

Vítor de Sousa
University of Trás-os-Montes and Alto Douro, Portugal

Vinicius Prates
University Presbyterian Mackenzie, Brazil

1. Introduction

The Brazilian ideological polarisation and the communicative practices of the political sphere have generated a toxic and dangerous environment for Brazilian journalists and their institutions. Since the rise of Jair Messias Bolsonaro as President of the Republic (2019-2022), however, the scenario has worsened with the systematic production of defamation of government opponents, including journalists; attacks on communication companies and the discrediting of their news; disinformation and fake news about reported happenings and the press being portrayed as an enemy of Brazil.

Since the beginning of his presidential mandate in 01/01/2019, he has forced journalists who covered the event to wait for hours in a room with no structure – water, food or chairs – until there is clearance from the security team. In other public events, he prohibited the entry of reporters from the country's main vehicles, silenced questions during press conferences using violent expressions, and allowed presidential security to attack journalists. Such events circulated on social media, through the president's supporters' channels, implying that a new era had begun against government opponents, journalists, and journalism. Since he became president, he has adopted a posture contrary to the work of the press, such as avoiding giving interviews to journalists of the great media, adopting Twitter as a communication channel without journalistic mediation and giving statements always surrounded by supporters, who pressure reporters with hateful messages or threats of physical violence.

In digital social networks, Jair Bolsonaro has a parallel structure to the government's official communication, composed of web publications by his three political sons, government members, and by users who work on social networks. The "Gabinete do Ódio" (Hatred Bureau), regularly reported in the Brazilian press, is a network of producers of counter-information about what is published in the press, composed of government officials and structure, coordinated by his son, the parliamentary Carlos Bolsonaro.

The Globo Communication group, the largest in Brazil and creator of the most watched news program on Brazilian TV, is their biggest target for attacks on digital social networks. The expression "Globolixo" (Globotrash), popularised by the president's son, Eduardo Bolsonaro, circulates on social networks and messaging apps through memes and videos (Figures 1 and 2). The president also uses his Twitter channel to attack journalists during live broadcasts (Figure 3).

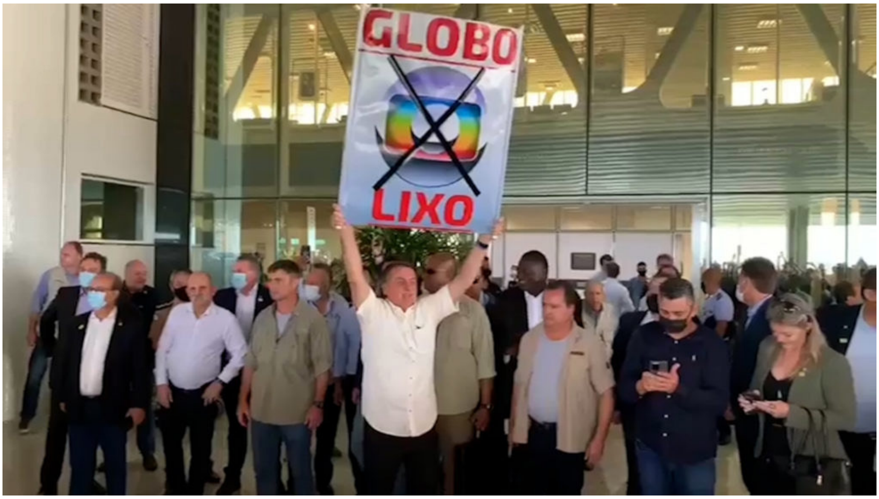


Figure 1 – President Jair Bolsonaro displays a poster with the expression “Globo trash”. **Source:** Printscreen from video by Poder 360.

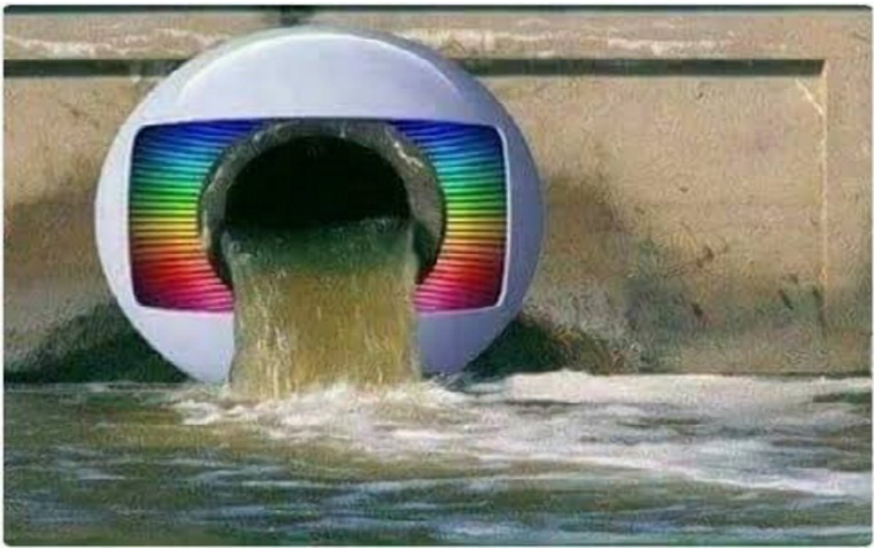
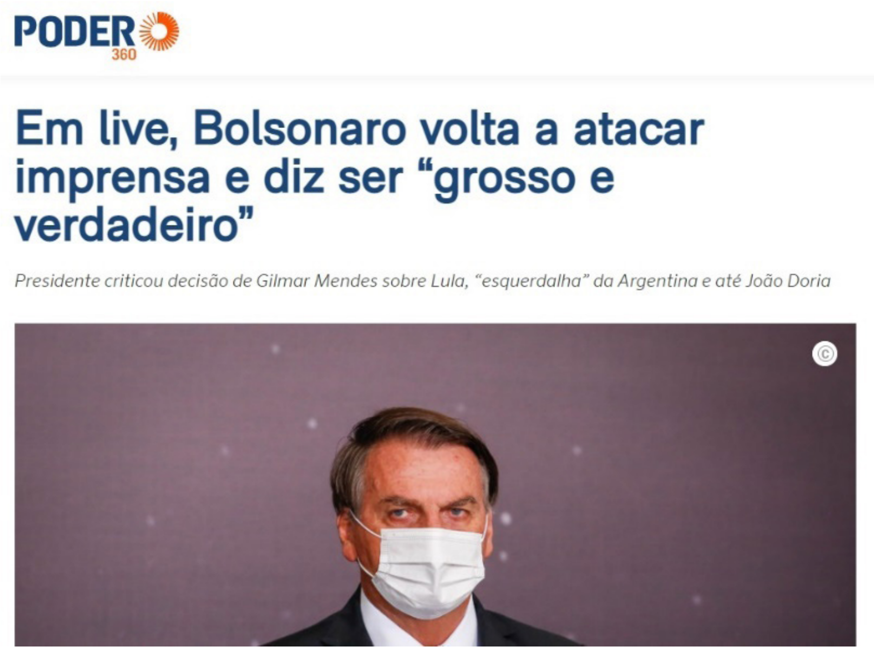


Figure 2 – A meme was found on the web after searching for the mention “Globo trash” in the Google search engine. **Source:** Globolixo.com.



PODER 360

Em live, Bolsonaro volta a atacar imprensa e diz ser "grosso e verdadeiro"

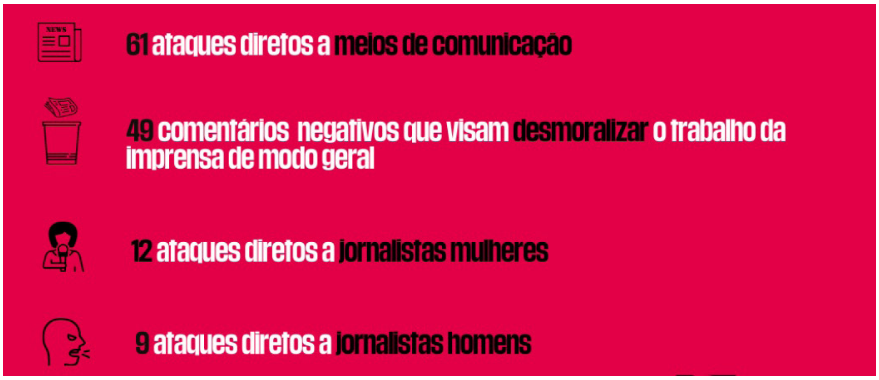
Presidente criticou decisão de Gilmar Mendes sobre Lula, "esquerdalha" da Argentina e até João Doria

Figure 3 – Headline "In live, Bolsonaro attacks the press again and says he is 'thick and true'". **Source:** Printscreen by Poder 360.

The media was also related to the poor evaluation of the president by his supporters during the Covid19 pandemics, being accused of trying to overthrow the government, after the publication of surveys in which the population evaluated the management of the pandemic crisis by the Bolsonaro government as bad/very bad. In 2020, a survey by DataFolha recorded that 79% of respondents thought the pandemic was out of control in Brazil.

A report from "Repórteres Sem Fronteiras" (Reporters Without Borders) (RSF, 2021) states that the working conditions of journalists have deteriorated considerably because of constant pressure from the president and his allies. During the Covid-19 pandemic, journalists were accused of disrupting Bolsonaro's government, reporting the number of dead and infected, encouraging the use of masks and social distancing, and finally supporting vaccination, all practices questioned by the president. Bolsonaro accused them as enemies of the population in his statements on social networks, having great repercussions among his followers.

In 2022, the Brazilian Association of Investigative Journalism – Abraji – released the first edition of a report monitoring attacks on journalists in Brazil, with data from the previous year. It states that 453 attacks against communicators and media (whose attack typologies are shown in Figure 4) were recorded. In 69% of the cases, the aggression was provoked by state agents (among them the parliamentary sons of the president and government ministers, Figure 5). The president of Brazil, Jair Bolsonaro, alone attacked the press 89 times, representing 19.64% of the total. Along with Bolsonaro's ministers, advisors and sons, this percentage rises to 55% of the total attacks, and when the attacks by supporters and protesters in events favourable to the president are included, they reach 271 – 60% of the total records (Abraji, 2022).



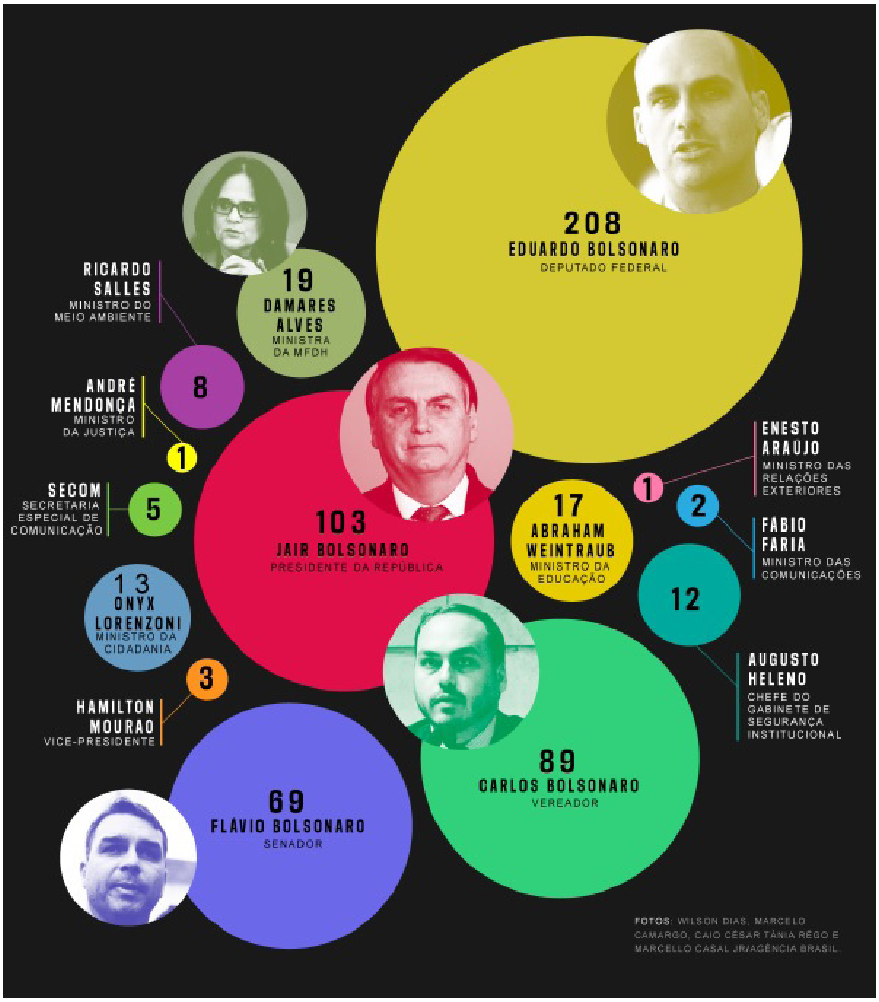
61 ataques diretos a meios de comunicação

49 comentários negativos que visam desmoralizar o trabalho da imprensa de modo geral

12 ataques diretos a jornalistas mulheres

9 ataques diretos a jornalistas homens

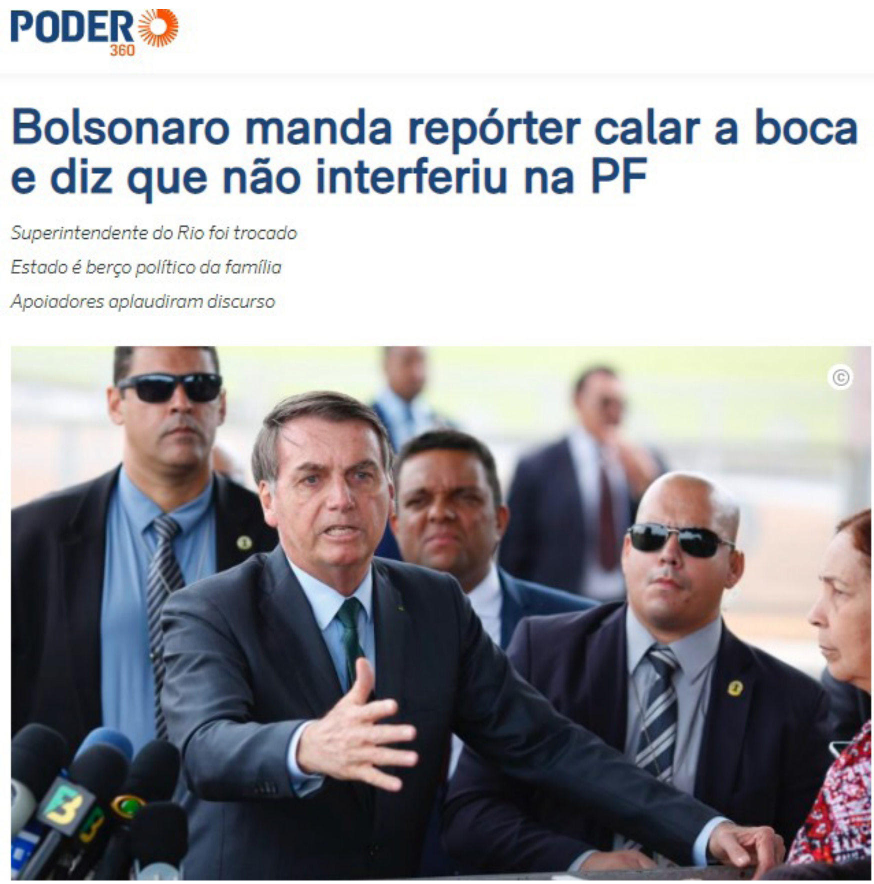
Figure 4 – Attacks on Brazilian journalists in the 4th quarter of 2021: 61 direct attacks on media; 49 negative comments aimed at demoralizing the work of the press in general; 12 direct attacks on female journalists; 9 direct attacks on male journalists. **Source:** RSF (2021).



FOTOS: WILSON DIAS, MARCELO CAMARGO, CAIO CÉSAR TÂNIA RÊGO E MARCELLO CASAL JR/AGÊNCIA BRASIL.

Figure 5 – Ranking of attacks - Bolsonaro's network for attacking the press with several actors, but is played out by his sons, who are also politicians (in yellow, green and purple bubbles). **Source:** RSF (2021).

According to the institution Reporters Without Borders (2021), attacks on journalists intensify on female professionals (Figure 6), who have been the target of the president's verbal aggression and his followers' hate speech, both in press conferences and via digital social networks. Among the various cases, on October 23, 2020, the Brazilian president threatened a journalist with aggression ("I feel like beating you up."), when asked about his wife's alleged involvement in a crime of corruption; on 21/06/2021, ordered a reporter to "shut up" when asked about the fine he received for not wearing a mask during a visit to the State of São Paulo, removing the mask again, this time to criticise the Globo Communication group and their professionals. On 04/04/2022, the president's son, Eduardo Bolsonaro, attacked a journalist who criticized his father's government, referring to the period in which the professional was tortured during the Brazilian military dictatorship.



PODER 360

Bolsonaro manda repórter calar a boca e diz que não interferiu na PF

Superintendente do Rio foi trocado

Estado é berço político da família

Apoiadores aplaudiram discurso

Figure 6 – Headline "Bolsonaro tells a reporter to shut up and says he did not interfere with the Federals". **Source:** Poder 360.

Hate speech was accentuated during the Brazilian electoral period, according to the Brazilian Association of Investigative Journalism – ABRAJI. The entity recorded a 250% increase in attacks on women journalists in September 2022, compared to the previous month, and 47.7% compared to September 2021; 63% of the cases were directly linked to the coverage of electoral events and 50% of the attacks came from politicians and government officials; 67.9% of hate speech was based on stigmatisation and 64.3% of the cases had their origin or repercussion in the online environment.

On 06/09/2022, Bolsonaro called the journalist frivolous and said that she should not question him because “her husband voted for him”, when asked about his family’s practice of buying real estate with cash, according to an investigative report by the newspaper Folha de Sao Paulo; about the author of the aforementioned investigative report – Juliana Dal Piva and her work, “A Vida Secreta de Jair” (“The Jair´s Secret Life”, Source: UOL) –, she stated on social media that the reason for the report would be “lack of sex” in her life; in the first presidential debate of 2022, he refused to answer questions from the journalist mediating the event, Vera Magalhães, saying that she “sleeps thinking of him” and that she was “the shame of the category”. Days later, on 09/13/2022, the same journalist was attacked again during a debate, this time by a supporter of President Bolsonaro who was broadcasting live saying she was “the shame of the category”.

These are some of the many cases perpetrated by the Brazilian president against journalism and journalists. However, such a communication strategy based on hate speech is applied to several other groups and individuals, such as environmental activists, artists, indigenous people, independent communicators and influencers, lesbian, gay, bisexual, transgender and queer - LGBTQIA+, former co-religionists, leaders of Parliament, representatives of the opposition to the government and judges of the Supreme Court –STF, among others.

Concurrent with the themes and targets of Bolsonarist hate speech, the circulation form of the speech varies among the president’s statements to the professional press, participation in digital debates with influencers, use of channels on digital social platforms and message circulation within digital communities, through organic sharing of supporters, driven by amounts paid to platforms or by fake profiles and massive message-shooting bots. It is a circuit, as discussed below.

This text starts from the Brazilian case to reflect on how we got here as individuals, communicators and society and what the characteristics of this contemporary cultural wound are. For this, we will present three perspectives on hate speech to understand the phenomenon in an interdisciplinary way.

The first will be the individual and biological sphere, on the neurological triggers of anger, an emotion that sustains hate speech, a topic so dear to the social sciences that it caused the so-called emotional reversal in the field. Next, the systemic issue of the hate circuit of narratives in communication environments will be presented, how they arise, how they propagate through networked information supports and how they feed back among intersecting contents. Finally, we will expand the debate to the question of historical presentism, a phenomenon of postmodernity that makes heterogeneous discourse something threatening to homogenising groups, without room for historical nuances necessary for the understanding of complex themes, simplified by hate speech, that circulate in the speed of digital social networks.

With this approach, we hope to better understand what the motivators of hate speech are, such as those reported at the beginning of this text, and perhaps to understand how to stop this spiral of narrative violence that affects the (lack of) knowledge society.

2. The emotional turn of the social sciences

With the increasing clarification of the functioning of emotions and feelings in individuals, other scientific areas besides the biological sciences began to turn to this object and confront it with their own objects of study. Sociology, for example, which has discussed the sociology of emotions for more than a century, defines the sociological politics of emotions (Barbalet, 2002; Demertzis, 2006) as a discipline that recognizes the centrality of emotion and the role of individual feelings in politics. In this context, the study of emotions within the social sciences has obvious functionality to understand hate speech.

This approach and many others come from a movement in the social sciences that, in the late 1990s and early 21st century, overcame an epistemological barrier by considering emotions not as a by-product of reason, but as constitutive of logical-scientific thinking, whose movement was called the emotional turn.

In turn, poststructuralism's critique of the reaffirmation of a binary world model, which maintains emotions as the opposition of reason, reaffirms its concern with the "death of the subject" (Terada, 2001, p. 3) in post-modernity. Clough (2008) calls for an "affective turn with the help of emotions" (2008, p. 1), a way to affect, deconstruct and auto organise the self in the face of contemporary demands:

> The growing importance of affect as a focus of analysis in a range of disciplinary and interdisciplinary discourses comes at a time when critical theory is confronted with the analytical challenges of war, trauma, torture, massacre, and the struggle against violence. If these world events can be considered symptomatic of ongoing political, economic and cultural transformations, the turn towards affection may be registering a change in the political, economic and cultural co-functioning. (Clough & Halley, 2020, p. 1)

The emotional turn does not attempt to investigate the meanings of emotions in societies - as Social Anthropology does - but to understand what emotional discourse communicates, whether verbal or non-verbal, conscious or unconscious (Athanasiou et al., 2008, p. 10). This way, the emotional turn approaches the postmodern perspective, as it deconstructs the modern subject (who denies affectivities and emotions due to its Cartesian foundation) and, instead, proposes a subject with multiple modernities (from their intimacies) (2008, p. 14). It is an affective turn to critical theory, in the sense of placing the subject's performance at the centre of the debate about the social environment that composes it and of which it is composed.

Indeed, affection and emotion are driving forces in contemporary societies. But due to a contemporary life developed according to the values of late capitalism (Jameson, 1991), resentment, as a diffuse feeling of impotence and desire for hasty reactions in the form of political and ethnic identity (Demertzis, 2006, p. 104), arises. What Betz (2002) calls resentment, "in full resemblance to anger, involves an intense feeling of frustration, illegitimate harm, the identification of a responsible agent and the desire to retaliate" (2002, p. 198), is also identified by Fukuyama (2018) as the "era of the politics of resentment" (2018, p. 25), by identifying that one of the political motivators of contemporary human beings derives from emotions. A kind of moral anger (Demertzis, 2006), a kind of "emotional opposition to unequal and unfair situations" (2006, p. 105), which implies attribution of legitimate guilt and promotes action against the offender (Rico et al., 2017).

One of the aggregators of popular resentment and its use in the form of hate speech is populism, an emotional movement (Fieschi, 2004), whose characteristic for obtaining support for policies is the manipulation of particular negative feelings (Muller, 2016). In this context, anger can materialise in physical form or through narrative, as in hate speech.

> Anger motivates a person to take action against the responsible agent, thus promoting a corrective response. More specifically, the angry citizen's reaction is confrontational, not deliberative, such that new considerations are avoided in favour of previous condemnations (...) By triggering individuals' reliance on pre-existing beliefs, anger - particularly when elicited by public issues - could thus be a factor in activating these latent and widespread attitudes towards politics. (Rico, Guinjoan & Anduiza, 2017, pp. 447, 449)

3. Emotions and anger for neuroscience

To debate hate speech in the media, it also makes sense to relate this cultural phenomenon to the biological phenomena that sustain it, human emotions and feelings.

For neurosciences, emotions are responses of the brain system, when specific regions combined result in reactions such as anger, fear, surprise or joy (Damasio, 2018, p. 158). Feelings are human catalysts for action, the motivators for an individual to react to an external stimulus. They are also responsible for the reverse route, when they monitor the success of this reaction, making the individual feel through the body whether the response to a stimulus was successful or not, according to their intentions (2018, pp. 22, 31). Thus, while emotions are actions accompanied by ideas and ways of thinking, feelings are perceptions that the body makes during emotion (2018, p. 143), thus altering rational decisions (2018, p. 163). Temperament, personality and character, as well as sociocultural background and the environment to which they relate, make up the system that will modulate the emotional reaction, by altering the weights and measures of emotions and feelings in each individual.

According to Damásio (2018), emotions are divided into universal, background and social emotions. The first ones – formerly called "low", as they refer more to the human instinct to defend against threats, such as fear, anxiety and alertness – come from the limbic system. They trigger the second type of emotions, the background ones, which are hidden in human behaviour and serve as internal motivators for action and for universal emotions such as enthusiasm and anger. Finally, social emotions are "recent" phenomena, according to the evolution of the human brain -as well as the part of the brain that generates them, the neocortex- and the social organisations that developed them. They are considered the emotions that most define the human being - therefore, also called high emotions, such as admiration, contentment and morality (2018, p. 158-161).

Emotions are linked to feelings as they are “the experience of certain aspects of an organism’s state of life” (Damásio, 2017, p. 151). For the brain processes that make up emotions extend to the rest of the human body, materialising in organs and muscles, through the activation of these by neurotransmitters, as already mentioned. Thus, it can be said that feelings make the individual feel the emotion bodily, in the same way that the body helps in the elaboration of emotion in feeling.

In this context, feelings are internal content generation systems, as they produce bodily reactions that can be externalised through emotions and, at the other end of the model, they are systems that receive external emotions and assimilate them bodily, translating the external environment to the receiver. Emotions are already the communicative tool that translates internal information into universal codes (anger, joy, sadness, fear, surprise, anger, etc.), which are also received by the receiver through emotions. They might be interpreted in a different form from the one produced by the sender, since the emotional process is individual. They are also felt within the body, being translated into the nervous system and migrating to the brain, influencing decision-making. And they do it with great influence, because “according to the evolutionary imperative, the oldest is stronger. New systems rarely subordinate older and more powerful ones. Therefore, the emotional brain (limbic system) is one of the systems that generally prevails in the fight against the cerebral cortex” (McCroskey & Beatty, 2000, p. 4).

In this picture, anger is an emotional element triggered by the most primitive circuits of the human brain. It is responsible for a physiological response linked to survival, which stimulates the fight or flee the stimulus that caused the individual to become angry. It is known that anger is linked to the limbic system, such as the amygdala, where it would be generated, and to the prefrontal cortex, which would regulate the bodily impulse of this emotion. In fact, this part of the brain, when injured, can reduce the ability to control anger, irritability and aggression.

The effects of anger occur differently from person to person, but they usually last until the supposed threat perceived by the individual no longer stimulates the aforementioned physiological phenomena. Thanks to the release of neurotransmitter hormones such as adrenaline and noradrenaline, the face of a person imbued with anger may flush, the skin may sweat, the heart rate increase and the breath become shorter.

Such emotion - also called wrath or rage, depending on the historical period in which it was recorded - is provoked by real and material elements, such as physical danger and threat, or by mental and subjective elements, such as personal frustration or subjective perception of evil. Just as its manifestation is innate in individuals, the use of anger as a tool is also culturally universal, most of the time controlled to obtain advantages over other individuals or groups. Likewise, anger is used by collective social institutions, when controlled in and applied to specific situations, as well as instrumentalized and immaterialized against alien groups to which one belongs. Just like hate speech.

Since the triggering mechanism for anger is not just biological but socio-cultural, individual upbringing and the collective environment can shape the propensity to feel and react to anger. Therefore, there is no consensus that the constitution of this emotion (as well as that of the others) is totally innate. In this context, it is possible to infer the manipulation of this primary emotion for the purpose of hate speech, since the emotion of hate is a social elaboration of anger, which is why it is also called a secondary emotion or social emotion.

Hate speech, therefore, uses the brain systems of anger, the primary emotion, to capture the attention of recipients through the biological defence/escape system that activates the attention of individuals. In addition, hate speech activates bodily and mental reactions that elaborate the feeling of hate, carried out through narratives. Hate speech would be the “training” of anger, through narrative.

4. The circulation system of hate speech

With the dissemination of network communication technologies, there was, at first, at the turn of the millennium, euphoria with the new perspectives of democratisation of speech that they could provide. However, if in fact there have been immense concrete advances in the possibilities of interaction from the diffusion of communication technologies, it is also true that a series of deleterious effects that cannot be disregarded have arisen. Among them are the proliferation of fake news that emulate journalistic language to lead to deception (Prates, 2023, in press), and also, in a related way, the intensification of hate circuits.

The proliferation of hate speech in socio-technical networks has entered an upward spiral in the last decade (Pereira, Prado & Prates, 2022), and the environment in which they are inserted is the circulation of content (Braga, 2012; Fausto Neto, 2019). This is precisely the system that breaks the traditional poles of emission and reception that characterised the mass media as described in the 20th century. The traditional "broadcaster" and "receiver" roles, as depicted by communication theories on mass society, vanish. This was a unidirectional process in which, at one end, the sender, in a reduced number, was the one who produced the senses and, at the other end, there were a large number of people subjected to this procedure, whose diversity was transformed into homogeneity, becoming a "mass".

Communication in the current stage of communicational capitalism (Prado; Prates, 2017), in turn, takes place in continuous flows, always directed forward, in a system of "circulation". In other words, there is no longer the role of broadcasting and receiving since all "receivers" are also "broadcasters". Fausto Neto (2019, online) calls this scenario of social networks "contact strategies", in which interactors seek to explore new interface conditions to maximise their interpenetration. In this way, flow communication does not form a simple, closed loop, but is continually directed forward. According to Braga (2012, p. 49) (our translation), agents that were traditionally just receivers, now put the answers back into circulation, not redirected to the sender, but inserting them in a social space in diffuse processes.

5. Antagonisms and hate circuits

This circulation system, in which the actants are diffracted, dimming the broadcaster and receiver roles, generates – from its dissemination – the possibility of establishing complex affective circuits. In this "extended circuit", the contents intertwine, at times in homologation and recognition, at times in antagonism and refusal. This is the environment of political polarisation, which closes a "hate circuit" (Pereira, Prado & Prates, 2022).

It arises from the breach of a "trust contract". What is this contract, which is gone? The "public sphere", in which the exchange of meanings creates an intersubjectivity that favours the production of "truths" (Habermas, 2014). In liberal democratic societies, this space of regulated inter-incomprehension (Maingueneau, 2005) is crystallised in institutions such as party politics, the university – but above all in the press. This is basically the "illuminist" scheme, in which the diffusion of knowledge is seen as the solution to all ills, and in which journalism has the role of sanctioning addresser, the "fourth power" capable of watching over the other powers and inserting them into axiological standards based on public interest.

However, precisely those who should guarantee the existence of the contract are depicted as deceptive. It is not necessary, it must be made clear, that we need to agree with the political understanding that the press "lies", breaking the proposed communication contract. There is no judgement in this case, but only the observation of the acting roles in the economy of symbolic exchanges. Nor is it necessary, from the establishment of the subjects of the discourse, that in fact a concrete promise has been established, to then be broken: "It is about the construction of simulacra, of these imaginary objects that the subject projects outside of himself and that, even without having any intersubjective foundation, effectively determine the intersubjective behaviour considered as such" (Greimas, 2014, p. 238).

The subject thus evoked by hate speeches feels frustrated because they are "fooled" by those who should demonstrate good intentions to participate in language exchange in the public sphere (Pereira & Prates, 2022). He then becomes frustrated because he is deprived of an asset or an advantage that he believed he could count on, but through another (Greimas, 2014, p. 235). This can be said according to the following formula: *the subject of waiting manifests a want-to-be that depends on the subject of action; this subject of waiting, therefore, assigns to the subject of the action a must-do, putting it in conjunction with an object of value* (Pereira & Prates, 2022). The discontent that follows is described thus:

> Eventually, another kind of discomfort is added to the dissatisfaction that arises after the non-attribution of the object of value, resulting from the behaviour of the subject to do, which is interpreted as not conforming to the expectation. As this behaviour, which in the eyes of the subject of fiduciary waiting is modalized by a must-do, is not carried out, the belief of the subject of state is suddenly revealed to be unjustified. The resulting disappointment is a crisis of confidence from a double point of view, not only because subject 2 frustrated the trust that had been placed in them, but also – and perhaps above all – because subject 1 can blame themselves for the misplaced trust. [...] These two forms of dysphoria, together, are caused by "frustration" and constitute, according to the dictionaries, the "lively discontent" that leads to the explosion of wrath (Greimas, 2014, p. 241).

What had been benevolence, the trust placed in the contract between subjects, gives way to malevolence, and from there there is a new rule for relationships, rising to polemics and finally antagonism. Hatred, from then on, can be diffracted into two programs: either the exacerbation that dominates the subject and manifests itself as anger; or else a better organised program of revenge. If the latter prevails, the subject of waiting will be transformed into a subject of action to inflict the evil back on the one who provoked it, and thus promote a kind – if we can say so – of homeostasis, of finding a lost balance.

In the case of polarisation, the semiotic regime of exclusion, as defined by Zilbelberg and Fontanille, is particularly relevant. For the authors, there are two possible valence regimes, the principle of exclusion and the principle of participation, which summon values two by two:

> The regime of exclusion has triage as its operator and, if the process reaches its end, it leads to the contentious confrontation of the excluder and the excluded and, for the cultures and semiotics that are directed by this regime, to the confrontation of the "pure" and the "unclean". The participation regime has mixing as its operator and produces the distensive confrontation of the equal and the unequal: in the case of equality, the quantities are interchangeable, while in the case of inequality, the quantities are opposed as "superior" and "inferior" (Zilbelberg & Fontanille, 2001, pp. 28-29, italics by the authors).

The exclusion regime has a disjunction as its operator, that is, a relationship in which one of the antagonistic poles must be chosen, and appears as an "either...or'' proposition (Zilbelberg & Fontanille, 2001, p. 27). Thus, attempts at improvement and approximation between the sides are always harmed, and the record is of mutual pejoration between the parties. A durative process of putting pejorative wording into discourse creates, as irreconcilable antagonists, the Other enemy that must be removed or eliminated, while reinforcing the identifications of belonging, of the Same (Prado & Prates, 2019).

Thus, a circuit is established in which there is, at first, the deposit of trust in the press fiduciary contract / this trust is broken, generating frustration / the frustration turns into malevolence, which can lead to a record of anger, or a lasting regime of revenge. These are the meanings of the circuits of hate, which turn as anger or as revenge to the subject of action: parties, scientific institutions, universities, and above all the press, which should have guaranteed the point of view of the subject of waiting and the cohesion of the fiduciary contract of communication. Ingrained in the circulation system of sociotechnical networks, the circuits of hate move forward, diffracting and penetrating the discursive spaces.

6. Presenteeism and the new media

The timbre of contemporary public debate is uniquely inscribed in the present, to which we are all summoned (Martins, 2011). In this regard, writer Javier Cercas warns against simplifying the present to the point of failing to understand it. In an interview with *Expresso*, he states that what is not from today is already past and what happened three weeks ago, prehistory. It is a situation that creates a totally falsified view of reality because the past is actually an active dimension of the present, without which the present is mutilated (Leiderfarb, 2020). This brings up the idea of "presentism", a concept coined by François Hartog (2003) and which is based on the idea that there is a risk that everything that belongs to history is compressed into contemporary history, as happens in contemporaneity. The "modern regime of historicity" would have been broken around 1989 with the idea of "the end of history", by Francis Fukuyama, "certainly a caesura of time" (Hartog, 2003, p. 188), leaving behind the Koselleck's ideas on the tendency of modernity to move away from experience and expectation, which are configured as "the main traits of this multiform and multivocal present: a monster present. It is at the same time everything (there is only present) and almost nothing (the tyranny of the immediate)" (Hartog, 2003, p. 259).

Political leaders take advantage of the dynamics of presenteeism, in which everything that emerges in society seems to have started today without a past history that contextualises the procedures. Which means that simultaneity was responsible for a new regime of historicity, a kind of continuous present, characterised by acceleration, and in which the present and the past are shown in a disruptive way. It is within this framework that Enzo Traverso (Observing Memories, 2018) underlines the urgency of freeing presentism from its cage – as if producing a world locked in the present with no ability to look at the future – by accommodating existing memories.

Paul Ricœur (2000) establishes a necessary link between memory and history, admitting that the historical brings forth the work of memory. It is, however, a contradictory process, as it selects and transforms previous experiences to adjust to new uses, as it practices forgetting, the only way to make room for the present.

Luciana Soutelo looks at Nora and Harrtg and, joining both perspectives, concludes that "presentism and prosthetic memory constitute (...) the explanatory keys to understand the culture of memory from the late 20th century" (Soutelo, 2015, p. 25).

For the rest, history should not be thought of in a linear way, but that it looks retroactively to the facts that are at the heart of dialectical reflection, towards the "absolute" knowledge. This is based on Hegel's (2018) idea that truth is not static, but results from the awareness of contradictory moments that overcome each other in a dialectical movement, towards "absolute" knowledge. (Jerónimo and Monteiro, 2020). As the same authors summarise, there are no signs of improvement, as the "trumpeters" remain "enthusiastically, on Fridays and Saturdays, fully aware of contributing to sharpen what they pompously say they want to transform, that is, the quality of the public debate" (Jerónimo and Monteiro, 2020, p. 10).

Which leads us, according to Pierre Bourdieu, to the idea of "doxosopher", quintessentially "the specialist in doxa, opinion and appearance, apparent scholar and scholar of appearance, perfectly prepared to give the appearances of science in a field where appearance always serves appearances" (Bourdieu, 1997, p. 27). And, nevertheless, as Rémy Rieffel (2003, p. 106) points out, the expression "mosaic culture" seems to faithfully translate the relationship between the media and culture, even if it does not make "any hasty judgement on any standardisation of thought or any waste of meaning", remaining light years away from Hommi Bhabha's idea of culture as a place of witness. The instrumentalization of social networks underlines the idea inscribed in the book by George Orwell, 1984 (2021), that whoever controls the present can create the past and, thus, fitting into the present, can control the future. This is a very dangerous engineering of thought, namely to serve as grazing for the hate speech that, day by day, has been increasing virtual exchanges.

It is not by chance that Peter Dahlgren (2014) states that social networks are platforms with a great deficit of democracy, since they work on the basis of replicating similarity and not promoting difference; which potentiates the emergence of an apparent consensus, shaped in unchallenged bubbles in the relationships between individuals mediated by these online platforms. José Pedro Zúquete (2022) compares populism to a chameleon. Perhaps that is why populist politicians direct their discourse almost exclusively to the new media, relegating traditional media to an unimportant place, altering the ecosystem that has been in place, regarding the scrutinising role of the media.

In *The expulsion of the other* (2018), Byung-Chul Han underlines the standardisation of globalisation and the blurring of the 'other', whatever it may be. He does not see positive things in the dissemination of what is the same and that reacts to the stimuli that capitalism determines in the same way. He refers that the proliferation of the same, presented as growth, makes the social body become pathological. In the chapter entitled "Listening", he predicts that, in the future, there will be a profession that will be called "listening", which will be paid to listen to the other, and listening gives back to each one what is theirs, reconciles, heals and redeems. Han states that the noisy society of weariness is deaf, and that, on the other hand, a society to come could be called a society of listeners and those who pay attention. What will go through a temporal revolution that makes a totally different time begin: rediscovering the time of the other. That will be a good time.

Moisés de Lemos Martins points out that human practices "are in direct relationship with temporality and have a local time, which is the time of experience", although they also have a contextual time, being that "between the time of experience and the contextual time there is a time for practice" (Martins, 2011, p. 64). Which means that it is not by chance that Umberto

Eco defends the idea that, even for philosophers, lies are more fascinating than the truth, a fact that justified his dedication to semiotics. For Eco, what makes signs interesting is not that they serve to tell the truth, but that they can be used to lie or talk about things we have never seen: "A language reveals its importance when it is used to refer to things that are not there. In my collection you will not find Galileo, but Ptolemy, because he was wrong" (Leiderfarb, 2015, p. 28-30). Furthermore, the philosopher, using Wittgenstein, observes that what cannot be theorised must be narrated, having no doubt that people prefer the lie to the truth.

If there are no measures aimed at reversing this *status quo*, such as increasing media regulation, the situation could deteriorate to levels that are difficult to recover. Even if indignation, when exercised by citizens, continues to contribute to resolving conflicts and problems (Innerarity, 2019). Which can, on the other hand, mean that the social networks that helped pave the way for Bolsonaro, can also remove him, if he does not live up to the expectations of those who elected him (Fernandes, 2018). Hate, on the other hand, is making its way through social networks and at the speed of the Internet.

Special thanks

Part of this work is funded by national funds through FCT – Foundation for Science and Technology, I.P., under the project UIDB/00736/2020 (base funding).

References

Abraji (2022). *Ataques contra mulheres jornalistas crescem 250% em setembro. Associação Brasileira de Jornalismo Investigativo*. https://abraji.org.br/noticias/ataques-contra-mulheres-jornalistas-crescem-250-em-setembro

Athanasiou, A., Hantzaroula, P. & Yannakopoulos, K. (2008). Towards a new epistemology: the "affective turn". *Historein*, 8(2008), 5-16. https://doi.org/10.12681/historein.33

Barbalet, J. (2002). Introduction: Why emotions are crucial. *The Sociological Review*, 50(2_suppl), 1-9.

Béland, D. (2017). What is Populism? Jan-Werner Müller. Philadelphia: University of Pennsylvania Press, 2016, pp. 136. *Canadian Journal of Political Science/Revue Canadienne de Science Politique*, 50(2), 633-634.

Betz, H. G. (1993). The new politics of resentment: radical right-wing populist parties in Western Europe. *Comparative Politics*, 413-427.

Bourdieu, P. (1997*). Les usages sociaux de la science. Pour une sociologie clinique du champ scientifique*. Versailles: Éditions Quæ.

Braga, J. L. (2012). La política de los internautas es producir circuitos. In Carlón, M. & Fausto Neto A. (eds.), *Las políticas de los internautas*. Buenos Aires: La Crujía.

Casacuberta, D. (4 de Noviembre de 2004). *Internet y la tercera izquierda*. Recuperado el, 12.

Clough, P. T. & Halley, J. (Eds.). (2020). *The affective turn: Theorizing the social*. Duke University Press.

Clough, P. T. (2008). The affective turn: Political economy, biomedia and bodies. *Theory, Culture & Society*, 25(1), 1-22. https://doi.org/10.1177/0263276407085156

Dahlgren, P. (2014). Participation and alternative democracy: social media and their contingencies. In Serra; E. Camilo, E. & Gonçalves, G. (eds.), *Political participation and Web 2.0.* Covilhã: LabCom Books, 61-85.

Damásio, A. (2018). *A estranha ordem das coisas: as origens biológicas dos sentimentos e da cultura*. Editora Companhia das Letras.

Demertzis, N. (2006). Emotions and Populism. In Clarke, S., P. Hoggett and S. Thompson (eds.). *Emotion, Politics and Society*. London: Palgrave Macmillan (103–122). *Europe. Comparative Politics*, 25(4), 413-427.

FENAJ (2021). Departamento de Saúde e Segurança. Federação Nacional dos Jornalistas. https://fenaj.org.br/

Fernandes, J. (2018, 29 de outubro). "Haddad é Lula" e Bolsonaro ganhou: as redes sociais nas eleições brasileiras. *Público Online*. https://www.publico.pt/2018/10/29/mundo/opiniao/haddad-lula-bolsonaro-ganhou-redes-sociais-eleicoes-brasileiras-1849274

Fieschi, C. (2004). Introduction. *Journal of Political Ideologies*, 9(3), 235-240.

Fukuyama, F. (2018). *Identity: The demand for dignity and the politics of resentment*. Farrar, Straus and Giroux.

Greimas, A. J. (2014). *Sobre o sentido II: ensaios semióticos* [About the meaning II: semiotic essays]. São Paulo: EDUSP.

Habermas, J. (2014). *Mudança estrutural da esfera pública*. São Paulo: UNESP.

Han, B-C. (2018). *A expulsão do outro*. Lisboa: Relógio d'Água.

Hartog, F. (2003). *Regimes d'Historicité: presentisme et experiences du temps*. Paris: Seuil.

Hegel, G. W. F. (2008). *Filosofia da História. Brasília*: Brasília: Editora Universidade de Brasília.

Jameson, F. (1991). *Postmodernism, or, the cultural logic of late capitalism*. Duke University Press.

Jerónimo, M. B.; & Monteiro, J.P. (2020). *Histórias(s) do Presente. Os mundos que o passado nos deixou*. Lisboa: Público/Edições Tinta-da-China.

Leiderfarb, L. (2015, 18 de abril). Entrevista a Umberto Eco. *Expresso (Revista)*, pp. 28-33.

Leiderfarb, L. (2020, 10 de abril). Não simplifiquemos o presente, ao ponto de deixar de o compreender (entrevista a Javier Cercas). *Expresso, E*, pp. E51-E55.

Lèvy, P. (2000). Cibercultura. Sao Paulo: Editora 34.

Maingueneau, D. (2005). Primado do interdiscurso. In Maingeneau, D. *Gênese dos discursos*. Curitiba: Criar Edições.

Martins, M. L. (2011). *Crise no castelo da cultura*. Coimbra: Grácio Editor.

McCroskey, J. C. & Beatty, M. J. (2000). The communibiological perspective: Implications for communication in instruction. *Communication Education*, 49(1), 1-6.

Neto, A. F. (2019). Política entre ações comunicativas e Circulações Disruptivas. *Rizoma*, 7(2), 10-25.

Observing Memories (2018, November). Interview to Enzo Traverso: About the complexity of the past. *Magazine of the European Observatory on Memories*, Second Issue. https://view.joomag.com/observing-memories-2/0021412001544464183

Orwell, G. (2021). *1984*. Porto: Porto Editora.

Pereira, H. P. & Prates, V. (2020). Propagação do vírus, disseminação do ódio: circulação dos afetos nas fake news sobre a covid-19. *Rizoma*, 8(1), 10-25.

Pereira, H. P.; Prado, J. L. A. & Prates, V. (2022). *Comunicação em Rede na Década do ódio: Afetos e discursos em disputa na política*. Digitaliza Conteudo.

Prado, J. L. A. & da Fonseca Bueno, V. P. (2019). O afastamento de Dilma Roussef: afetos e discursos em disputa na política. *Revista Famecos*, 26(2), e31913-e31913.

Prado, J. L. A. (2019). Sintoma e fantasia no capitalismo comunicacional. Editora Estação das Letras e Cores.

Prates, V. (2023). *O engano, a doença, a morte: como as fake news simularam técnicas canônicas do jornalismo durante a pandemia de Covid-19*. In: Eccom – Educação, Cultura e Comunicação. Lorena (SP): Fatea, 2023 (NO PRELO).

Rico, G., Guinjoan, M.; & Anduiza, E. (2017). The emotional underpinnings of populism: How anger and fear affect populist attitudes. *Swiss Political Science Review*, 23(4), 444-461.

Ricœur, P. (2000). *La memóire, l'histoire, l'oubli*. Paris: Éditions du Seuil.

Rieffel, R. (2003). *Sociologia dos media*. Porto: Porto Editora.

RSF (2021). *A dark year for freedom of the press in Brazil - 580 attacks against media in 2020*. Reporters without Borders. 21 of January of 2021. https://rsf.org/pt-br/um-ano-sombrio-para-liberdade-de-imprensa-no-brasil-580-ataques-contra-m%C3%ADdia-em-2020

RSF (2021). *Um ano sombrio para a liberdade de imprensa no Brasil*. Reporters without Borders. 21 of January of 2021. https://rsf.org/pt-br/um-ano-sombrio-para-liberdade-de-imprensa-no-brasil-580-ataques-contra-m%C3%ADdia-em-2020

Soutelo, L. C. (2015). *A memória pública do passado recente nas sociedades ibéricas. Revisionismo histórico e combates pela memória em finais do século XX*. Tese de Doutoramento, Faculdade de Letras da Universidade do Porto, Portugal.

Terada, R. (2001). Introduction: Feeling in Theory: Emotion after the "Death of the Subject". In *Feeling in Theory* (pp. 1-15). Harvard University Press.

Zilberberg, C.; &; Fontanille, J. (2001). *Tensão e significação*. São Paulo: Humanitas.

Zúquete, J. P. (2022). *Populismo - Lá fora e cá dentro*. Lisboa: Fundação Francisco Manuel dos Santos.

Chapter 7

CLIPPING: HATE SPEECH IN SOCIAL MEDIA AGAINST FEMALE SPORTS JOURNALISTS IN GREECE

Lida Tsene
Open University of Cyprus, Cyprus

***Clipping** is a foul that's called when a player makes contact with an opponent below the waist from behind.*

1. Introduction

The web 2.0 gave us the opportunity to explore new ways of collaboration and communication. Digital platforms and social media became a fertile ground for people to interact and express their opinions unfiltered, while the non obligation to reveal oneself directly added an extra level of freedom in the way they shared news, thoughts and observations. Freedom of speech became the new trend and everyone celebrated the democratisation of media. But unfortunately, there is also the other side of the same coin. This democratisation facilitated somehow heated discussions which frequently result in the use of insulting and offensive language. Danny Wallace, author of the book, "F*** You Very Much: The surprising truth about why people are so rude" (2018) believes that not only anonymity, but mostly the lack of eye contact in social media, reinforce digital rudeness. At the same time, according to his point of view, the loose and some times not well identified rules of those platforms, on what they perceive as bullying and maleficent content allow trolls to conquer the digital sphere.

Hate speech wave is growing globally and many countries and organisations are recognising it as a serious problem and threat for democracy, while it has been associated "to a global increase in violence toward minorities, including mass shootings, lynchings, and ethnic cleansing" (Laub, 2019).

2. What is hate speech

But what is hate speech and how can one define it? The truth is that there are more than one definitions for the term, "mainly because of the vague and subjective determinations as to whether speech is "offensive" or conveys "hate" (Strossen, 2016 in Tontodimamma et al., 2020) In common language, hate speech could be defined as the offensive discourse against individuals or groups of people based on characteristics such as race, gender, sexual orientation etc. Th United Nations Strategy and Plan of Action Hates Speech defines the term as "any kind of communication is speech, writing or behaviour that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor" (United Nations, n.d).

If we go back in time, Richard Delgado's article "Words that Wound: A Tort Action for Racial Insults, Epithets, and Name-Calling (1982) focused on racism provides with a definition that ticks the following boxes: "(l) language was addressed to him or her by the defendant that was intended to demean through reference to race;" (2) "that the plaintiff understood as intended to demean through reference to race; and" (3) "that a reasonable person would recognize as a racial insult." (Sellars, 2016). A decade later, Calvin Massey (1992), argues that "hate speech is any form of speech that produces the harms which advocates for suppression ascribe to hate speech: loss of self-esteem, economic and social subordination, physical and mental stress, silencing of the victim, and effective exclusion from the political arena.". Later on, Susan Benesch (2013) defines five characteristics of the dangerous speech, whether (1) there is a "powerful speaker with a high degree of influence;" (2) there is a receptive audience with "grievances and fear that the speaker can cultivate;" (3) a speech act "that is clearly understood as a call to violence;" (4) a social or historical context that is "propitious for violence, for any of a variety of reasons;" and (5) an "influential means of dissemination.". The complexity of the terminology expands and in other issues around hate speech such as the extent of freedom of speech today, the regulation of media and digital media, fake news and misinformation, as well as the skills a person needs to acquire in order to recognise hate speech.

3. Sexist hate speech

A distinctive category of hate speech it is sexist hate speech. Before defining the term, let us clarify what sexism means. Again, although there is not a global definition, we can argue that the following summarises in a good manner the main characteristics of a sexist behaviour, "the supposition, belief or assertion that one sex is superior to the other, often expressed in the context of traditional stereotyping of social roles on the basis of sex, with resultant discrimination practiced against members of the supposedly inferior sex" (Inter Press Service, 2010). Sexist hate speech or cyber gender harassment or cyber-sexism "aims are to humiliate and objectify women, to destroy their reputation and to make them vulnerable and fearful" (Gender Equality Unit, 2016) and "has a set of core features: (1) its victims are female [men are less targeted], (2) the harassment is aimed at particular women, and (3) the abuse involves the targeted individual's gender in sexually threatening and degrading ways" (Citron, 2009).

A recent study by the Pew Research Center (Vogels, 2021) states that women are three times more likely to become victims of online sexual harassment, with percentages increasing in younger women (under 35). Alongside sexist hate speech the phenomenon of gendertrolling arises bearing characteristics such as violent language, strong reactions to mentions of gender based inequality, credible threats beyond the online word, sex or/and gender related insults and comments (Mantilla, 2013). Notable is also the fact that sexism on social media can be benevolent (Glick & Fiske, 1996) in the format of humour, memes or even positive comments.

4. Hate speech in the digital sphere

All the aforementioned have been magnified under the prism of digital media. Social media and all the other digital platforms have changed the ways we communicate, we interact and react and of course our perception on freedom of speech. Online hate activities go back in the 80's "when a Commodore 64 desktop computer with a telephone modem connection was used to allow skinheads, Klansmen, and Neo-Nazis to communicate and download electronic bulletin boards" (Duffy, 2003). Since then, there is a growing wave of online hate speech mostly because, "is low cost, can be facilitated anonymously and pseudonymously, is easy to access, is instantaneous, can reach a larger audience, and can be spread via different formats across multiple platforms. It also raises cross-jurisdictional issues in regard to legal mechanisms for combatting it" (Netsafe, 2018).

The rise of hateful content in the web, has pushed the major online platforms themselves to develop definitions in an attempt to moderate the produced and shared content. For example, YouTube's Community Guidelines state "we don't support content that promotes or condones violence against individuals or groups based on race or ethnic origin, religion, disability, gender, age, nationality, veteran status, or sexual orientation/gender identity or whose primary purpose is inciting hatred on the basis of these core characteristics" (YouTube, n.d). In addition, Facebook identifies hate speech as "content that directly attacks people based on their race; ethnicity; national origin; religious affiliation; sexual orientation; sex, gender or gender identity; or serious disabilities or diseases" (Facebook, n.d). But although the platforms try to fight the fight against online harassment, a quick glance at the comments under various posts on the different platforms highlights how important the problem is.

5. Hate speech and Covid

Covid-19 appeared to have worsen pre-existing inequalities – especially towards minorities – and offered a fertile ground for hate speech to grow. In

May 2020, United Nations Secretary-General António Guterres raised the alarm about the "tsunami of hate and xenophobia, scapegoating and scaremongering around the world", while a report by Youth Charity Ditch and Brandwatch revealed that "online hate speech in the UK and US has risen by 20% since the start of the pandemic" (Baggs, 2021). Another survey by the International Center For Journalists and Columbia University for the Journalism and the Pandemic Project highlights that 20% of the participants to the survey claimed that their experience of online abuse, harassment, threats or attacks was much worse than usual" during the pandemic (ICFJ and Tow Center for Digital Journalism, 2020). Boredom, isolation during the lockdowns, high rates of disposable time plus the fact that people who already bully and troll might have traumas themselves that increased by the pandemic are some of the reasons we experienced a pandemic of hate lately (United Nations, n.d).

6. Sexist hate speech and female journalists

One of the most vulnerable target group, in terms of hate speech and sexist hate speech, appears to be female journalists. According to the UN's Secretary General (2017), "women who cover topics such as politics, law, economics, sport, women's rights, gender and feminism are particularly likely to become targets of online violence. While men journalists are also subject to abuse online, abuse directed against women journalists tends to be more severe". Since then, this topic has been a high priority for various international and national organisations. A recent study by UNESCO and the International Center for Journalists-ICJ (2020) shows that, 73% of women respondents said they had experienced online violence in connection with their work in the field of journalism, ,while they also said that they "had been subjected to a wide range of online violence, including threats of sexual assault and physical violence, abusive language, harassing private messages, threats to damage their professional or personal reputations, digital security attacks, misrepresentation via manipulated images, and financial threats". A 48% reported that they have been harassed "with un-wanted private messages, highlighting the fact that much online violence targeting women journalists occurs in the shadows of the Internet, away from public view". Finally, according to the same survey, "nearly half (47%) of the respondents identified reporting or commentating on gender issues (e.g., feminism, male-on-female-violence, reproductive rights including abortion, transgender issues) as a top trigger for online attacks, highlighting the function of misogyny in online violence against women journalists".

The reasons for receiving hate speech varies from misogyny and anti-feminism, to nationality, ethnicity or sexual orientation or even their appearance and the impact this online violence has on them is tremendous. The UNESCO and ICJ report (2020) reveals that "the mental health impacts of online violence against the women journalists participating in this survey were the most frequently identified (26%) consequences of being targeted. And 12% had sought medical or psychological assistance in response", while "many women journalists surveyed reacted to online attacks by making themselves less visible. 18% said they withdrew for a period of time from participating in online conversations and communities, while 11% permanently withdrew from certain online communities". The latter has a domino effect on media diversity and gender representation in the news, as women are silenced due to hate speech attacks.

A further look into sports journalism proves that the situation is quite similar, or even worse. Again, a brief glance at the comments on the social media profiles of female sports journalists are disappointing and sexist language is dominant. In their survey by Miller and Miller (1995), found out that female sports journalists experienced sexual harassment and felt invisible and silenced in their departments. Tracy Everbach (2018) on her side, claims that "Women sports journalists continue to face harassment and demeaning treatment on the job and that while most women journalists said they received mainly positive social media comments, the harassment causes them distress because of its threatening and abusive nature. In some cases, women said they suffer self-doubt because their qualifications and work product consistently are challenged".

A very characteristic example of online sexism hate speech towards female sports journalists, that later became a movement is the #MoreThanMean campaign launched by the small independent media organisation Just Not Sports back in 2016. Just Not Sports is a podcast and web community with the focus on athletes' and journalists' personal passions and interests outside of sports. The podcast was launched in November 2015 by Emmy-winning producer Gareth Hughes, sports media strategist Adam Woullard, and sports marketer Brad Burke. During the research process for their podcasts content Burke noticed "the women were being harassed in a way that was not consistent with playful banter and the harassment absolutely did not compare to what the men get." (Antunovic, 2018) as "the women were dehumanized and insulted on their looks, sexuality, and ideas" (Antunovic, 2018). This observation led to a video that became viral with more than 3.7 million views on YouTube. In the video, "several men alternate as they sit across from either DiCaro or Spain[1] and read comments from the journalists' Twitter accounts on a phone or a tablet. The tweets include abusive language, derogatory sexual comments, and death threats. The men hesitate to read, stumble upon their words, and apologize to the journalists. DiCaro and Spain sit calmly. The video concludes with a text that reads: "We wouldn't say it to their faces. So let's not type it" (Antunovic, 2018).

7. Hate speech in Greece

As the rise of hate speech is a global phenomenon, Greece couldn't be an exception. Unfortunately, even children face online hate speech as a recent survey revealed. According to the survey, 34% of the children participated, have encountered online hate speech, while 6% claims to have experienced it as well, with girls ranking higher (SaferInternet4Kids, 2022). The refugee crisis the past years revealed a big amount of racist and xenophobic speech online towards refugees and migrants. Moreover, Greece has ranked last in the European Union on the Gender Equality Index for 2020 and 2021 was marked with the #metoo incidents making gender hate crimes more visible

[1] Julie DiCaro and Sarah Spain are Chicago-based sports journalists.

to a larger part of the local society. And although there are very few data regarding online hate speech towards Greek female journalists the problem is present. In the book, "#journaliststoo: Women Journalists Speak Out" by Irene Khan (2021), Anthi Pazianou, a Greek journalist shares her own experiences. Pazianou claims that, "in Greek, the word "journalist" is of masculine gender. Even though many other professions, as women grew stronger, became female-gendered too, "journalism" remains male", stressing out the fact that although there is a significant progress towards equity and equality in female representation and participation in Greek media, there is still a long way to be walked. Her descriptions around sexist online speech are vivid, showcasing that being a journalist in Greece and covering debatable or more "male" labeled topics could be an extreme sport. "There was a Facebook post concerning a colleague of mine with the title "Hang her"", she reports. And she continues, "On 8 September, two colleagues of mine and I covered a story about African asylum seekers training for the local football championship. After the interviews, we took a photo with the footballers and posted it on Facebook for a limited audience. On the same day, members of extreme right groups, from the village where the young girl was attacked, posted my photo with the footballers publicly on Facebook, stating I was having sexual relationships with "n*****s". They used other sexist, racist and offensive expressions, leading eventually to intervention by the Council of Europe and a number of Greek politicians" (Khan, 2021).

8. Research hypothesis and methodology

Our research hypothesis drives from two basic facts related to the under-representation of women both in media and in sports. According to the 2015 Global Media Monitoring Project, "in 2015, women make up only 24% of the people heard, read about or seen in newspaper, television and radio news, exactly as they did in 2010". And although today, "women are well represented early in the career pipeline in media and entertainment, they are a minority at the highest levels, with women accounting for only 27 percent of C-suite positions" (Beard et al., 2020).

At the same time, in sports journalism, the statistics are even worse. The recent (2021) Sports Media Racial and Gender Report Card published by The Institute for Diversity and Ethics in Sport reveals that “while women saw slight improvements in 2021, the overall record of the sports media for having women in prominent positions remains terrible”.

Moreover, there is a growing wave of online or offline violence and hate speech and women seem to be more often targets of online violence, while women working in the sector of journalism have faced hate speech at least once during their professional career. According to the survey conducted by the International Federation of Journalists (IFJ) in 2017 found that 44% out of the nearly 400 women participants had experienced online abuse. In addition, the UNESCO-ICFJ survey in 2020 highlighted that 73% (n=625) of women journalists have experience online violence.

Within this framework, we attempted to explore whether women working in the sports journalism field in Greece have been targets of online abuse, with a special focus on sexism hate speech, how do they respond and the impact this might have on their professional development and mental health, the role of Internet and social media as well as possible solutions to this challenge.

In order to explore deeper our research questions, we applied a qualitative methodology, conducting 9 in depth semi-structured interviews with women working in the fields of sports journalism and sports social media. We reached most of our interviewees from our networks of contacts within the sector, as result of the researcher own professional practices. Interviews lasted between 30 and 45 minutes and took place during September 2022.

We also managed to have a diversity in terms of age, seniority in work experience, and role within the sports journalism/media sector. Five of them are sports journalists, working in various media – sports portals, TV, radio – two of them work as social media managers for big football clubs in Greece, one of them is the marketing director of one of the most prominent sports new website and one of them used to work as a sports journalist and now is press officer in a big company related to football. The interviewees' age varied from 26 to 45 years old. Although, our sample is not big we managed to get on board most of the most representative women sports journalists working at the moment in Greece and that allowed us to unfold views and perspectives of the particular community towards hate speech and most specifically sexist hate speech.

In addition, we applied a qualitative content analysis to the comments on the Instagram profiles of three of the most popular Greek female sports journalists for the month September 2022. We analysed 1677 comments from all profiles.

9. Key findings

If we would like to summarise the key findings of the interviews those could be the following:

- The majority of the participants to the interviews claimed that they always used to be around sports and that the decision to follow a professional career in sports journalism was a deliberate one.
- Although their social and inner circle had some doubts whether they should follow such a career, they didn't stop them for doing it.
- Most of them admitted that they faced challenges during their professional career because of their sex. Although they might had the chance to find a relevant job, part of the participants agree that they don't have always the same opportunities to their male colleagues. "Of course I can remember times where I had an idea, I pitched it, but a male colleague took the job", stated one of the participants to the survey.
- Another major issue raised by the participants is the fact that although, nowadays there are more women working in the field, most of them are accompanying a male colleague. They all agree that they would like to see more women anchoring sports TV or radio shows. "I would be happy and satisfied when I will see more shows with a woman leading them,

and not with just a role in the panel", said characteristically one of the interviewees.

- Almost all of the participants said that they have faced a kind of sexual harassment during their professional career and some of them are still facing it.

- They all use social media – Instagram mostly and then Facebook and Twitter – as a professional tool. And although the content they share is most of the times related to their work, however, sometimes they do post moments from their personal lives.

- They all agreed that the Internet and social media facilitated the spread of hate speech and sexist hate speech, due to the anonymity they offer to the users and that they have been targets of sexist hate speech more than once during their career.

- Although, most of the comments are positive, they do recognise there is a sexist approach in some of them. "I cannot say that the comments I see are negative, but a big percentage are related to how do I look and not to what I am doing", a participant stated. An interesting fact is that part of our sample claims that it is stressful to think how they will cover up their appearance in order to avoid sexist comments. As an interviewee said, "it is not crime to be beautiful and we don't have to be ashamed or feared because of it".

- All of them have received negative comments about their work, always once again related to their sex. On top of that, they feel mad as they always feel that they should run the extra mile to convince everyone – employers, colleagues and fans – that they know what they are doing. As a participant to the survey claimed, "It is rather frustrating the fact that if a male colleague makes a mistake, nothing will happen, but if I make it, I will receive a thousand of negative and disparaging remarks".

- It was rather interesting the fact that most of them concluded that in the beginning of their career negative or sexist comments had an impact on them, but now they feel empowered and confident about what they do and how they do it. "I can say that the negative comments really make me want to achieve more", a participant said.

- Moreover, in the question, if they have ever thought if changing job or career orientation due to sexism online or/and offline, all of them agreed that this had never crossed their minds.
- The majority agreed that they don't pay attention to negative or sexist comments or even that they don't read them at all.
- Part of our sample admitted that there is still a sentimental impact on them if they come across negative or sexist comments and/or language, but all of them agreed that there is no severe impact on their mental health.
- They all believe there is solidarity between them and they try to discuss and raise such issues of inequalities and sexual harassment online and offline.
- And of course, all of them strongly believe that the field is in a better shape in terms of gender equality than some years ago, but there are a lot still to be done.
- In the question how we can fight online sexist hate speech, they all stressed out the role of proper education and a more concrete regulation around it.

In terms of the comments, the most interesting findings are:

- From the 1677 comments, more than half use terms such as "hot babe" or "sexy girl".
- Also, the 3/4 of the comments are related to the appearance of the journalists and not to their work. For example, in one of the profiles, there seems to be a pattern, commenting on the legs of the journalist, whether in the other, they do comment a lot on the high heels she wears.
- In addition, there are several comments using the fire emoji, again in relation to their appearance.

· Moreover, there are few comments using inappropriate language in a sexist context.

10. Discussion

It is rather sad the fact that "every 30 seconds a woman journalist is harassed online. This harassment takes the form of name-calling, sexist comments, serious accusations of physical harm such as death and rape threats, devaluing their work, threats against their partners and children, and posting of personal details online (doxing)", (GEN VIC 2019). Social media and digital platforms somehow promised a more balanced space for everyone to express themselves. The opportunity to share your opinion online encouraged underrepresented voices to be heard, women among them. But did we actually achieve an equal representation? Snježana Milivojević (2016) claims "The news industry is packed with sexist practices, glass ceilings and other forms of gender exclusion, which in turn reproduces the same reality. On digital media platforms this gender injustice persists both in terms of representation and production. Media frames and practices structure and provide patriarchal continuity and there seems to be no digital "new world" out there in terms of gender relations".

From our research we can conclude that although there seems to be a progress on the equal representation of women and men in sports journalism, still the language used around women is definitely sexist encouraging stereotypes. And though, most of the comments are positive, or use a positive language, there is a latent sexism contributing "to the strengthening power relations, gender stereotypes and sexist behaviours" (Marwick, 2013). According to a participant to the survey, "I receive a lots of comments related to my appearance, and although I cannot say that they are negative, I feel insulted, as I would love to see more comments regarding my knowledge on football, my commentary, the stories I curate etc. And I know that my male colleagues are never getting those kind of comments".

Moreover, and on top of that, the majority of the content we encounter online, either in the format of social media comments, or online articles is related to the physical appearance of the journalists and very few to their work. And while this is common for women, if someone searches respectively on Google for "Greek male sports journalists" she will end up with indexes with names, best sports articles etc. This seems not be a different case from what is happening globally. According to Chen et al. (2018), "Women have to deal with the sexual comments that males never have to deal with", explained an American online journalist. "You're viewed more often as a sexual objects ... I've been told I need to get laid... They're rare, but they're so much worse than what my male colleagues have to deal with.". And they continue, "An anchorwoman in Taiwan explained: "Most of my followers on Facebook are male. They don't really care about the news I share. They follow me because they want to see beautiful girls.".

Another pattern revealed from our research is that women sports journalists have to work far harder from their male colleagues to prove themselves. One of our interviewees highlights, "I get super mad when someone asks me or even congratulates me on knowing what an offside means. Of course I know! It's my job!". And once again this aligns with the global trends. Back in 2014, journalist Johanna Franden asked Laurent Black, then coach of the French football club Paris Saint Germain, why he switched from a 4-4-2 formation to 4-3-3 during a game, only to receive a patronising reply. "Women talking football tactics, it's so beautiful. I think it's fantastic. You know what 4-3-3 means, don't you?".

A very interesting finding is the fact that the majority of our sample has developed mechanisms of self-protection towards online abuse. But is that enough? As Posetti, Harrison and Waisbord (2020) suggest, "we need to be very cautious about suggesting that women journalists need to build resilience or "grow a thicker skin" in order to survive this work-related threat to their safety". And they continue, "It is vitally important for news organizations to have gender-sensitive policies, guidelines, training, and leadership responses. Together, these measures must ensure awareness of the problem, build the capacity to deal with it, and trigger action to protect women journalists in the course of their work". Keeping always in mind that if women journalists are silenced or becoming less visible, diversity in media and in our global societies are being injured, we should all take action towards preventing online sexist hate speech and provide them with a safe space to do their job and contribute to the democracy.

References

Antunovic, D. (2019). "We wouldn't say it to their faces": online harassment, women sports journalists, and feminism. *Feminist Media Studies*, 19(3), 428-442. http://dx.doi.org/10.1080/14680777.2018.1446454

Baggs, M. (2021). Online hate speech rose 20% during pandemic: 'We've normalised it'. *BBC*. https://www.bbc.com/news/newsbeat-59292509

Beard, L., Dunn, J., Huang, J. & Krivkovich, A. (2020). Shattering the Glass Screen. *McKinsey*. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/shattering-the-glass-screen.

Benesch, S. (2013). Proposed Guidelines for Dangerous Speech. *Dangerous Speech Project*. http://dangerousspeech.org/guidelines/

Brown, A. (2017). What is so special about online (as compared to offline) hate speech? *Ethnicities*, 18(3), 297-326. https://doi.org/10.1177/1468796817709846

Citron, D. C. (2009). Law's Expressive Value in Combatting Cyber Gender Harassment. Michigan Law Review, Vol. 108:373. http://digitalcommons.law.umaryland.edu/cgi/viewcontent.cgi?article=1687&context=fac_pubs

Chen, G. M., Pain, P., Chen, V.Y., Mekelburg, M., Springer, N. & Troger, F. (2018). *Women journalists and online harassment*. Center for Media Engagement. https://mediaengagement.org/research/women-journalists

Delgado, R. (1982). *Words That Wound: A Tort Action for Racial Insults, Epithets, and Name-Calling*, 17 Harv. C. R.-C. L. L. Rev. 133.

Duffy, M. E. (2003). Web of hate: A fantasy theme analysis of the rhetorical vision of hate groups online. *Journal of Communication Inquiry*, 27(3), 291-312. https://doi.org/10.1177/0196859903252850

Everbach, T. (2018). I Realized It Was About Them … Not Me": Women Sports Journalists and Harassment. In: Vickery, J. & Everbach, T. (eds). *Mediating Misogyny*. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-319-72917-6_7

Gender Equality Unit (2016). *Background Note on Sexist Hate Speech*. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168059ad42

Gen Vic (2019). *The horrendous online abuse a female sports journalist received highlights dangers of media that must change*. Gen Vic. https://www.genvic.org.au/media-releases/the-horrendous-online-abuse-a-female-sports-journalist-received-highlights-dangers-of-media-that-must-change/

Glick, P. & Fiske, S. T. (1996). The ambivalent sexism inventory: differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, 70(3), 491-512.

GMPP (2015). *The Global Media Monitoring Project 2015*. https://whomakesthenews.org/wp-content/uploads/who-makes-the-news/Imported/reports_2015/highlights/highlights_en.pdf

ICFJ & Tow Center for Digital Journalism (2020). *Journalism and the Pandemic: A Global Snapshot of Impacts*. https://www.icfj.org/sites/default/files/2020-10/Journalism%20and%20the%20Pandemic%20Project%20Report%201%202020_FINAL.pdf

International Federation of Journalists (2017). *One in two women journalists suffer gender-based violence at work*. https://www.ifj.org/media-centre/reports/detail/ifj-survey-one-in-two-women-journalists-suffer-gender-based-violence-at-work/category/press-releases.html

Inter Press Service (2010). *IPS Gender and Development Glossary (3rd Edition): A Tool for Journalists and Writers*. Bangkok: Inter Press Service.

Khan, I. (2021). *#JournalistsToo - Women Journalists Speak Out*. https://www.ohchr.org/sites/default/files/2021-11/JournalistsToo-en.pdf

Laub, Z. (2019). Hate Speech on Social Media: Global Comparisons. *Council on Foreign Relations*. https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons

Mantilla, K. (2013). Gendertrolling: Misogyny adapts to new media. *Feminist Studies*, 39(2), 563-570.

Marwick, A. E. (2013). Gender, sexuality, and social media. In: Hunsinger, J. & Senft, T. M. (Eds.). *The social media handbook*. Routledge.

Massey, C. R. (1992). *Hate Speech, Cultural Diversity, and the Foundational Paradigms of Free Expression*. 40 UCLA L. Rev. 103.

Miller, P. & Miller, R. (1995). The invisible woman: Female sports journalists in the workplace. *Journalism and Mass Communication Quarterly*, 72(4), 883-889.

Milivojević, S. (2016). *More platforms, less freedom: How new media reproduce old patriarchal structures in New Challenges to Freedom of Expression: Countering Online Abuse of Female Journalists*. OSCE Representative on Freedom of the Media.

Netsafe. (2018). *Online hate speech: A survey on personal experiences and exposure among adult New Zealanders*. https://www.netsafe.org.nz/wp-content/uploads/2019/11/onlinehatespeechsurvey-2018.pdf

Piñeiro-Otero, T. & Martínez-Rolán, X. (2021). Say it to my face: Analysing hate speech against women on Twitter. *Profesional de la información*, 30(5), e300502. https://doi.org/10.3145/epi.2021.sep.02

Julie Posetti, J., Harrison, J. & Waisbord, S. (2020). *Online Attacks on Women Journalists Leading to 'Real World' Violence, New Research Shows*. International Center for Journalists (ICFJ). https://www.icfj.org/news/online-attacks-women-journalists-leading-real-world-violence-new-research-shows

SaferInternet4Kids (2020). *Αποτελέσματα εθνικής έρευνας 2021-2022 για διαδικτυακές συνήθειες σε 5000 μαθητές*. https://saferInternet4kids.gr/ereynes/ereuna21-22/

Sellars, A. F. (2016). *Defining Hate Speech*. Berkman Klein Center Research Publication No. 2016-20 Paper No. 16-48. Boston University School of Law, Public Law Research, Boston University School of Law, Public Law Research. https://doi.org/10.2139/ssrn.2882244

Strossen, N. (2016). Freedom of speech and equality: Do we have to choose? *Journal of Law and Policy*, *25*(1), 185–225.

The Institute for Diversity and Ethics in Sport (2021). *The 2021 Sports Media Racial and Gender Report Card: Associated Press Sports Editors*. https://www.tidesport.org/_files/ugd/138a69_e1e67c118b784f4caba00a4536699300.pdf

Tontodimamma, A., Nissi, E., Sarra, A. & Fontanella, L. (2021). Thirty years of research into hate speech: topics of interest and their evolution. *Scientometrics*, 126, 157-179. https://doi.org/10.1007/s11192-020-03737-6

United Nations General Assembly (2017). *The safety of journalists and the issue of impunity. Report of the Secretary General*. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N17/245/44/PDF/N1724544.pdf?OpenElement

UNESCO and the International Center for Journalists (2020). *Online Violence Against Women Journalists. A Global Snapshot of Incidence and Impact*. https://www.icfj.org/our-work/icfj-unesco-global-study-online-violence-against-women-journalists

United Nations (n.d.). *Understanding hate speech*. https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech

United Nations (n.d.). *Impact and prevention*. https://www.un.org/en/hate-speech/impact-and-prevention/a-pandemic-of-hate

Vogels, E. A. (2021). *The state of online harassment*. Pew Research Center. https://www.pewresearch.org/Internet/2021/01/13/the-state-of-online-harassment/

Wallace, D. (2018). *F*** You Very Much: The surprising truth about why people are so rude*. Ebury Press.

Chapter 8

MAPPING SOCIAL MEDIA HATE SPEECH REGULATIONS IN SOUTHERN AFRICA: A REGIONAL COMPARATIVE ANALYSIS

Allen Munoriyarwa
University of Botswana, Botswana

1. Introduction

Countries in Southern Africa have, recently, faced an avalanche of disturbing hate speech incidences and dozens of horrific hate-driven violence many of them occurring due to the spread of misinformation and disinformation on social media platforms. In Kenya and Rwanda, for instance, offline hate speech has often had a relationship with online hate speech, spurred by social media platforms (Scheffler, 2015). In South Africa, the country's human rights commission has consistently (since 2016) noted the dangers posed by social media hate speech to the racial integration of the country imperative after half a decade of the brutality of apartheid (South Africa Human Rights Commission, 2016; Munoriyarwa, 2021).

A number of critical incidences of social media hate speech have dangerously hogged the limelight since, as social media platforms expand across the region. These incidences shine a light on the growing influence of social media as a conduit through which hate speech is spread. New genres of information, largely driven by social media platforms, like Twitter and Facebook now threaten to stoke further hatred and divide communities in the Southern African region. Violent and hate-laden rants laced with fearmongering and often racist rhetoric threaten both free speech and the participation of others on social media platforms (Marais & Pretorius, 2015; Munoriyarwa, 2021). Hate speech on social media has not only targeted the racially different other but has also disproportionally targeted groups like women and the LGBQTI+ communities.

This chapter examines the regulations that govern hate speech on social media in seven selected countries of Southern Africa. Drawing on textual/content analysis of the regulations, the chapter seeks to examine how these countries legislate social media hate speech, exploring areas of legal convergences and divergencies in the respective legislations of these countries. Through content analysis, this chapter makes several observations. In South Africa, it observes the inadequacies of hate speech legislation as a weapon to combat social media hate speech, as demonstrated by rising cases of social media hate speech. In Zimbabwe, Lesotho and Eswatini, the chapter notes an increasing weaponisation of these legislations for political advantage by ruling regimes. Furthermore, in Zimbabwe, Lesotho and Swaziland, the chapter argues that the laws have become too broad arguably because they are routinely and rampantly abused for political advantage by the ruling elites. Ultimately, the chapter concludes, the social media hate speech regulations in these countries stifle free speech, and in countries where democratic spaces are shrinking, the region risks creeping further into authoritarianism.

This chapter makes several contributions to scholarship. Firstly, there is little known scholarship on hate speech on social media in the Southern African region. Much of the extant work is around civil society and statutory body reports. In countries like Zimbabwe, Lesotho and Eswatini, even such reports are, hitherto, none-existent. South Africa, however, has witnessed emerging research on hate speech (Munoriyarwa, 2021) and attempts to map the prevalence of hate speech and methods of analysing it (Gagliardone, 2014). Beyond these, there is nothing up to this point, known nor systematic on research around hate speech. Therefore, the chapter contributes to this relatively virgin area. Secondly, there is no known research that provides a legal analysis of hate speech regulations in the Southern African region.

One documented analysis has focused on Kenya and Rwanda (see Scheffler, 2015). Therefore, a legal content analysis of social media hate speech regulations remains a lacuna to be filled in academic research.

I utilise a content analysis approach in order to dissect the social media hate speech regulations in the region and providing a comparative analysis. Content analysis is an empirical method commonly used in the social sciences (Mayring, 2021; Krippendorf, 2018). Qualitative content analysis involves the subjective interpretation of the content of textual data (Mayring, 2021). Content analysis works through a classification process of identifying themes that emerge in the data. It is rarely used on legal issues (Linos & Carlson, 2019). But, in the past decade, legal scholars have begun to adopt qualitative content analysis in order to understand the structure of legal statutes, regulations and government policy proposals (Bax, 2014). When used on legal statutes, content analysis helps draw out among other issues, what statutes mean, what they cover, their omissions, their targets and the people most likely to be affected by a specific statute. I use qualitative content analysis to examine, critic, and compare social media legislation in the Southern African region.

This chapter is organised as follows. In the next section, I scope the social media and hate speech context in Southern Africa. I will do this by fleshing out how hate speech has morphed out in the region. I do this by providing examples of notorious cases of social media and hate speech in the region. This section is followed by a review of literature on hate and social media. Because there is scant literature on this subject in the region, I will review literature from around other contexts where the intersection of social media and hate speech has been researched. After this section, I provide my comparative findings carefully drawn from the regulation of the selected countries. I then provide a discussion and a conclusion to the chapter.

2. Social media and hate speech: The Southern African context

In this section, I sketch the context of social media hate speech in the Southern African region. While my focus is broad, I focus more on the countries sampled for this research. Social media hate speech has become a serious crisis in the countries of the Southern African region. In Zimbabwe, for instance, most of the hate language has been directed at gay/lesbian and other LGBQTI+ people. But in the Zimbabwean case, such hate language has a historical genesis (Muparamoto, 2021). The country's founding executive president Robert Mugabe stirred hate rhetoric against the LGBQTI+ community. In one infamous address, he called the community, 'worse than dogs and pigs… that can easily identify their opposite partners' (Maenzanise, 2018). A narrative that has normalised hate language against this group has been inculcated and vigorously reinforced in Robert Mugabe's rhetoric to the effect that gay rights are not human rights.

Mugabe's hate language had also bee targeted at his political opponents. For example, he referred to his long-time political nemesis the late Morgan Tsvangirai as, "a frog…a little ant that should be crushed…a running dog of the Western imperialists…" (Munoriyarwa, 2021, p. 84). This offline hate language against perceived opponents and other groups had found reinforcement on Twitter especially by Mugabe's avid supporters, and ruling party supporters who call themselves "Varakashi' on Twitter. The group is known for its vile and diabolic attacks in opponents of ZANU PF on Twitter, and currently, they get encouragement from the ruling party itself. There are reports[1] that this army of hate speech mongers have been receiving US$ 10 a day as payment for defending the ruling party and its president on Twitter. The overall effect, however, is that Zimbabwe's Twittersphere has been weaponised, with hate language and an assortment of other hate-spewing rhetoric.

[1] Such reports of payment of this 'Twitter army' of hate speech mongers can be followed here: https://ne-np.facebook.com/hopewelljournalist/photos/a.1761564017398962/2952094438345908/?type=3.

In other countries like South Africa, hate speech and hate language have been replete on social media. The South Africa Human Rights Commission (SAHRC, 2016) noted with concern how Twitter was increasingly becoming a war zone, replete with hate speech. There are already cases of hate speech on social media that have been successfully prosecuted. For example, Penny Sparrow[2] was convicted of hate language in 2019, after referring to black Africans as monkeys. Another white Afrikaner, Vicky Momberg was also convicted[3] of hate speech after insulting members of the South Africa Police Service (SAPS) by using the 'K' word.

In Lesotho, hate speech incidences are increasing on social media. In 2022, a Twitter account has drawn public condemnation because it referred to the hotly contested 1998 elections that divided the small nation into an ethic quagmire. There has been, in 2022, a case of a celebrity subjected to hate insults because of her HIV status. This has also happened in Swaziland. Cases of hate speech and hate language have been rising on social media in the country. The civil revolt against the monarchy, that rocked the country in the middle of 2021 further polarised the country. Hate language has been directed at both the opponents and supporters of the monarchy in equal measure.

3.Findings

In the next sections, I provide the findings of my chapter. Firstly, I provide a regional synopsis of social hate speech regulations through a content/textual analysis of the regulations from the different countries in the region. This contextual synopsis is important for my comparative analysis section which comes after this section.

[2] The Penny Sparrow case can be followed here: https://www.theguardian.com/world/2016/jun/10/white-south-african-estate-agent-fined-racist-facebook-post-penny-sparrow-hate-speech.

[3] The case can be followed here: https://www.sabcnews.com/sabcnews/convicted-racist-vicki-momberg-back-court-sentencing/

3.1. Hate speech regulation: A regional synopsis of Southern African countries

This section begins with a presentation of how hate speech regulations look like in the countries sampled. Thereafter, I draw on content analysis to flesh out the similarities and differences of these laws and regulations. South Africa stands out as one country with a standalone hate speech law – the Promotion of Equality and Prevention of Unfair Discrimination Act of 2000 (also known as the PEPUDA or the Equality Act number 4 of 2000). The country's brutal history of racial xenophobia during the apartheid period has carefully guided its responses to hate speech especially the sprouting social media and other digital media platforms. South Africa's social media hate regulations are guided by the desire to ensure national healing and social cohesion. For a country, that struggles with a violent history of racial segregation, xenophobia and homophobia (Breen et al., 2016), clear anti-hate laws were, therefore, necessary. Of particular importance is section 9(3) and 9(4) of the Act which prohibits discrimination against any persons based on grounds of race and other grounds.

In September 2022, the South Africa National Assembly introduced the Prevention and Combating of Hate Crime and the Hate Speech Bill. Under the new proposed law, there are two types of specific offences; hate crime and hate speech. Hate crime, under the proposed law, is when a person commits any recognised offence motivated by prejudice or intolerance, while hate speech is when a person publishes or shares statements that intend to be harmful or incite harm (BusinessTech, 2022). The bill is explicit about hate speech on social media platforms. For instance, it proposes to sanction WhatsApp group administrators who do not delete hate content shared by members on their group. The new bill seeks to combat the weaknesses of PEPUDA. Under PEPEUDA, an aggrieved party had to approach the Equality Court within their personal capacity. The new proposed law nullifies that and allows the state to institute criminal offences against hate speech on social media. It is important to note that South Africa already has restrictions under the *crimen injuria* statute - the wilful injury to a person's dignity as a result of obscene or racially offensive language – that is common law. Furthermore, the country's constitution, "...already lists forms of speech such as propaganda for war, incitement of imminent violence or advocacy of hatred based on religion, gender or race..." (BusinessTech, 2022, n.p).

In other countries of the region, a notable feature is the lack of standalone social media hate regulation. While anti-hate regulations exist, they are spread across several other extant statutes and acts. In Namibia, Article 23 of the country's constitution prohibits hate speech and its propagation. It also renders it criminally punishable - an offence. But, perhaps as a sign of its inadequacy, there is a loosely phrased statute in article 23 which states that the Courts, and parliament can render hate speech punishable, "...for the purpose of expressing the revulsion of the Namibian people at such practices...". (Constitution of the Republic of Namibia, 2019). There is no specific reference, however, to the types of hate speech legally sanctionable. In fact, the word hate speech is not specifically referred to in the constitution.

In Zambia, the penal code section 70 and the public order Act section 13 prohibits, "offensive conduct likely to breach peace..." Section 11 of the constitution state thus, "Any person who in any public place or at any public meeting uses threatening, abusive or insulting words with intent to provoke a breach of the peace or whereby a breach of the peace is likely to be occasioned, shall be guilty of an offence...". There are several weaknesses with this law, however. Firstly, it is silent on an explicit mention of hate speech, in a country where social media has been weaponized for hate speech. Secondly, it arguably carries the most speculative of legal terms. For instance, "...whereby a breach of peace is likely to be occasioned..." is broad and speculative. Thirdly, it does not specifically protect individuals who might be subject to hate speech, referring to 'public meetings'. What about harm done outside public meetings? More importantly, the absence of specific punishable hate offences is in itself evidence of the inadequacy of the hate law clauses of the country.

In March 2021, Zambia passed the Cyber Security and Cyber Crimes Bill into law. The law has, since its passage, generated significant public rancour[4] with critics pointing out that while it is the only law close to arresting social media-driven hate speech, it is still porous in the sense that it can be politicised, and is open to very wide and problematic interpretations. These critics have pointed out that the law could be used to suppress free speech, and to shut down the Internet, a common feature in countries of Southern Africa, like Zambia and Zimbabwe. There are also problematic provisions in the law. Of particular concern is the definition of hate speech in section 2 of the act. The clause defines hate speech as including any forms of communication that, "involves hostility or segregation" towards groups of people on the basis of "economic status". (Constitution of the Republic of Zambia, 2021). Admittedly, hate speech is notorious to define. But such loose clauses have obvious dangers especially in an emerging democracy like Zambia. The most obvious dangers include lack of clarity on what is allowed and what is prohibited speech. For instance, 'segregation based on economic status' is hopelessly difficult to define as part of hate speech. What happens when, for instance, a Facebook post, or Twitter post criticises those who hold excessive wealth and economic privilege, most likely the political elites? This can easily be interpreted as 'segregation based on economic status". What then, arguably, this might mean is that people might shy away from speaking against corruption and primitive accumulation of wealth. This is likely to protect the ruling elites and other holders of capital, some of whom might have earned the wealth through illegal means (Mwananyanda, 2021).

In Democratic Republic of the Congo (DRC), Article 77 of the country's constitution – also referred to as the Freedom of Press Law criminalises incitement. Section 51 of this law clearly:

[4] The debates about the dangers posed by the new law can be followed here: https://www.dailymaverick.co.za/article/2021-04-05-free-speech-zambias-new-internet-law-fails-basic-human-rights-scrutiny/.

> criminalizes incitement to violence, discrimination, and hate against a person or group of persons on the basis of their appearance, ethnicity, nationality, race, and religion. The law appropriately uses the standards articulated in Article 20(2). The High Council for Audio-visual and Communication also has the power to suspend media outlets for incitement to violence and hate speech on the basis of tribe, ethnicity, race, and religion (Constitution of the DRC, 2009).

This is what the DRC have, that is close to hate speech law. There are a number of weaknesses notable in this clause. Firstly, it is tailored at journalists, assuming that they may be the only community of practice capable of disseminating hate speech. This is more worrisome in a country notorious for its ethnic strife, ethnic factionalism and hatred (Kongolo and Zamberia, 2016; Kabamba, 2010). And, like the Zambian law, there is no explicit mention of social media-drive hate speech in the clauses of the laws. In Botswana, the laws are equally wide as in other countries, except South Africa. The constitution of Botswana protects citizens against discrimination based on:

> race, tribe, place of origin, political opinions, colour, creed or sex whereby persons of one such description are subjected to disabilities or restrictions to which persons of another such description are not made subject or are accorded privileges or advantages which are not accorded to persons… (Constitution of Botswana, 2021).

In Zimbabwe, social media hate speech is regulated much more broadly under the general rules governing fundamental human rights and freedoms. Section 3 of the constitution articulates the values and principles of good governance and equality of all human beings, and the rule of law. The constitution of the country explicitly outlaws any, "…pronouncements that threaten, intimidate, and incite violence against persons of a different political affiliation or persuasion…" (Section 3, 67, constitution of Zimbabwe).

The Zimbabwean constitution is rather silent about hate speech on social media but has specific clauses that have been used to prosecute hate speech on social media. For example, on 30 May 2018, four people were arrested for posting and spreading hate speech on Twitter.[5] In so doing the state relied on several statutory clauses that outlaw hate speech. Section 61(5) (b) is one such clause. The section guarantees freedom of speech, but outlaws hate speech and incitement to violence. It is supported by various clauses of the Criminal Code in the Zimbabwe constitution. For example, Section 37 (ii) (c) of the Code criminalises engendering, promoting and exposing people or a person to hatred on any media platform. In the same criminal code, section 42(2) criminalises insulting or otherwise grossly provocative statements on social media or any other media.

Eswatini, like most of the countries in the region, other than South Africa, has no standalone clauses on hate speech. Rather, it relies on a collection of clauses spread across different statutes in the country's constitution. There is an overt reliance on Chapter 3 of the Eswatini constitution which protects and promotes fundamental rights and freedoms. Amongst the generic fundamental rights protected, section 18(1) of this law makes it clear that, "...the dignity of every person is inviolable..." On many instances, the state had relied on this clause to charge activists for hate speech on social media. For example, a teacher, Majahembuso Dlamini, at Nhlangano School in Eswatini, was arrested on 2 May 2020 for comments made on social media.[6]

Eswatini has had a checkered history with hate speech. It had relied, initially on a 1938 law called the Sedition and Subversive Activities Act to combat hate speech. It was a hopelessly anachronistic law and in 2016, the country's high court annulled the law. The Kingdom had also relied on the Suppression of Terrorism Act. But a full bench of the country's high court ruled that certain sections of the law were illegal. One of these was section 46 of the act. This section had been used to prosecute hate speech in a very partisan manner- that is, against democracy activists. It was also phrased and framed in a catch-all manner seemingly to target activists who

[5] The arrest stories can be read here: https://www.techzim.co.zw/2018/07/zrp-arrests-4-people-for--postinghate-speech-and-falsehoods-on-social-media/.

[6] The story of the arrest can be followed here: https://m.facebook.com/groups/142383985790674/permalink/3337175642978143/

voiced their disaffection with the monarchy. For example, the section had a clause which criminalises 'seditious intentions' and 'raising discontent and disaffection". The court annulled the clause because it was abused and was vague to be interpretable. Social media activists had been charged under this clause. It is important to point out that about four years after the courts annulled the laws, the state is still charging people using those laws. *The Mail and Guardian* reported thus:

> A striking illustration of how these laws are used to intimidate, the state continues to charge people under them, even though they have been annulled by the courts. Four years after the courts quashed the charges against … [Thulani Maseko, Maxwell Dlamini and Mlungisi Makhanya…[Mario] Masuku, Eswatini's most persecuted pro-democracy campaigner… [their charges] still stand…and [they] still comply with [their] bail conditions…[7]

Criticism of the Eswatini laws have been that they are reactionary (Dlamini, 2019) and anti-democratic (Limb, 2022). Several developments have taken place in Eswatini in the past two decades. The population of the country, while small (the country has about one and a half million people), has become youthful. The youth bulge has set in motion a series of political events that have had a huge impact on the country. For instance, from 20 June 2021, until mid-July, the youth violently protested against the absolute monarchy governing their country. They also made use of social media platforms, especially Twitter and Facebook. Critics (see for example, Freedom House, 2021; MISA, 2021), have pointed out that the country's reaction to hate speech on social media is driven by a desperation to protect the monarchy of King Mswati III from increasing online bellicosity. The monarchy face two major dilemmas. Firstly, it has become wieldy, inefficient and corrupt, yet it is absolute. This means, the failures of the state cannot be transferred, nor shared with anyone except it. Secondly, it does not connect

[7] The report by the Mail and Guardian can be followed here: https://mg.co.za/africa/2020-08-27-how--eswatini-silences-opposition-activists/

with the current generation of youths who have no nostalgic attachment nor generational relations with the monarchy.

The youth actually see the Swazi monarchy as a stumbling block to their aspirations and material dreams, and hence, see it as irrelevant and in need of, not reform, but removal. Consequently, the monarchy has taken flake on social media platforms. The charges of hate speech against activists and members of the opposition are meant to stop the anti-monarchical tide which is rising on social media. In the next section, I examine the similarities and differences in the regulation of hate speech on social media in these countries whose legal frameworks were examined here.

3.2. Divergencies and convergencies in social media hate speech regulations

There are several divergencies and convergencies in the regulation of hate speech amongst countries in the Southern African region. An obvious point of divergence is the absence of social media hate speech specific regulations in some of the countries in the region. From the section above, the evidence is that most countries of the region, for instance, Zimbabwe, Lesotho, Eswatini and Botswana, have chosen a rather broader approach to hate speech. A broader approach to the regulation of hate speech is devoid of social media specific laws on hate speech. There are laws on hate speech in these countries, but the laws do not specifically point to social media-driven hate speech. The most clear and present danger of such a generalist approach is that it backgrounds social media hate speech, which is actually the biggest menace to our society more than any form of hate speech (Wetzstein, Bartenberger & Leitner, 2013). The need, therefore, for a systematic approach to social media hate speech has never been so obvious. But countries like South Africa and Zambia have developed in their legal jurisprudence, a comprehensive legal architecture on hate speech on social media. They have social media hate speech regulations in the mould of the PEPEUDA and Cyber Security and Cyber Crimes Law, respectively. South Africa has shown greater resolve in eliminating hate speech by proposing, in 2021, the Prevention and Combating of Hate Crime and the Hate Speech Bill.

South Africa's approach to social media hate speech, as shown, draws on its brutal history with apartheid over almost half a century. The apartheid system, reliant on racial exclusion and prejudice has meant that post-apartheid regimes take more active measures to limit certain forms of speech as intolerable. In this regard, South Africa's approach to hate speech can be equated to that of Germany, which, after a checkered history with Nazism, has also taken a zero tolerance, and comprehensive approach to hate speech online (see Timofeeva, 2003, for this argument). The racial outbursts of people like Penny Sparrow and Vicky Momberg, which I have cited above point to the destabilising potential of social media-driven hate speech in post-apartheid South Africa. Countries in the Southern African region that have taken a broader approach to hate speech, which does not specifically zero-in on social media-driven hate speech, have done so at a greater cost of continued destabilisation within their communities. To illustrate this point, I want to give a more clear, relevant and contemporary example of the DRC.

The DRC has, between 2016 and 2022, faced serious social media-driven hate speech (United Nations [UN], 2021; Ntanyoma & Shacklock, 2021). This has been propagated via mainstream social media platforms like Twitter, Facebook and also on YouTube channels. The hate speech has targeted ethnic groups like the Tutsi, the Kinyarwanda and other minority ethnic groups. In 2021, the UN brought social media hate speech in the DRC to international attention[8], noting that rising social media-driven hate speech in the DRC has the potential to destabilise the country, whose peace is already very fragile. In response to the rising threat of social media-driven hate speech, the DRC government, in 2021, tabled the Racism, Xenophobia and Tribalism Bill, which seeks to combat social media hate speech in the country. This bill, if passed into law, will represent a comprehensive statute

[8] A UN public statement was issued, warning that the country was likely descend into further chaos if social media hate speech was not nipped in the bud, especially in the Eastern DRC. The statement can be followed here: https://peacekeeping.un.org/en/un-concerned-about-increase-violence-and-hate-speech-eastern-drc

targeting social media hate speech in the country. Currently, the DRC shares similarities with Eswatini, Botswana and Zimbabwe in that its hate speech regulations make broad references to the protection of human dignity and are devoid of specifications like the indignity wrought by social media hate language, the militancy of such spaces, and how legislation can successfully restrain social media spaces from being purveyors of hate speech. This, therefore, requires a holistic approach in the region that, arguably criminalises hate language on these spaces.

One other divergent difference in legislation within the region is on what aspects of social media hate speech are legally punishable. In the DRC, Zambia, Zimbabwe, Lesotho and Botswana cases, the laws generally prohibit hate speech in its broader sense, But South Africa's legislation which is comprehensive as discussed earlier, takes a rather very narrow approach which seeks to balance freedom of expression and protection of dignity. In the Qwelane v. SAHRC[9], the South African Supreme Court of Appeal held that in as far as hate speech hurts society, any act that prohibits speech which does not directly advocate for hatred and incite harm goes too far and is an infringement on the right to freedom of expression. This judgement, therefore, demonstrates the country's attempt to strike a balance between the protection of dignity and the importance of freedom of expression, an attempt which is not paralleled in other countries of the region. The Court argued that hate speech alone is not enough a justifiable limitation on free speech. It has to advocate for harm on a community or on individuals. There are no obvious parallels in the rest of the Southern African region, where such a balance is attempted. The South African court ruling is an affirmation that annoying and critical speech on social media cannot be punished for its own sake, it is in a liberal context, protected speech.

My argument is that in semi authoritarian regimes of Zimbabwe, Eswatini, for example, there is no such clear distinction of what constitutes hate speech and what is free speech because of the involvement of political

[9] The summary of the judgment can be followed here: https://globalfreedomofexpression.columbia.edu/cases/qwelane-v-south-african-human-rights-commission/.

players with political intentions in social media regulations. Social media regulations are often subjected to abuse and hence, no obvious attempt to balance free speech and human dignity. This, I argue, amounts to the weaponisation of social media to achieve political ends.

To support my argument, I want to use two relevant examples form Eswatini and Zimbabwe. On 15 October 2021, the government of Eswatini shut down all social media platforms like Twitter and Facebook, after sustained pro-democracy protests rocked the country. The Eswatini government was apparently taking notes from their counterparts in Zimbabwe who, for three days from 19 January 2019, shutdown these platforms following sustained protests by citizens. Both governments accused protesters of inflaming tensions through the spread of hate language on these platforms. The Zimbabwean government accused protesters of preaching the language of intolerance and violent dissent on social media platforms. The Eswatini government accused protesters of inciting hatred on social media. This is a clear demonstration of how vague social media regulations can play into the hands of beleaguered ruling political elites. Vague and broad laws can be used to suppress dissent in the name of combatting hate speech. In the case of the Eswatini government, it issued a statement saying that the government had, after shutting down social media for fear of rampant hate language, opened an email account where protesters could channel their grievances.

Thus, the Zimbabwean and Eswatini cases discussed above prove the argument that in semi-authoritarian regimes, legal regulations of hate speech, especially on social media are deliberately vague in order to allow for technological interventions like Internet shutdowns, when the ruling elites face power-threatening protests. This argument has also been made by Garbe, Selvik and Lemaire (2021). They argue that the character of the regime determines how they are likely to intervene to stop the spread of hate speech and fake news.

When the regime is semi-dictatorial, it is likely to resort to measure like stopping the spread of hate content through blocking access to those platforms. As I have demonstrated in this argument, for Zimbabwe and Eswatini particularly, technological interventions are meant to stop opposition actors from accessing information on social media platforms in those specific moments of protests. So, the target is no longer hate speech on social media, but legitimate protests mobilisation on those platforms. In these countries, there are institutional constraints designed not to interfere with such practices. For example, the Zimbabwe Human Rights Commission, which should push back against political interreference, like Internet shutdown, is a political stooge and a lapdog of the ruling regime (Ruhanya, 2020), heavily compromised with pro-regime sympathisers (Munoriyarwa, 2022). There is a stark difference with South Africa, which adopts a legal content regulation approach. In a liberal democratic context like South Africa, institutions like the SAHRC are independent from the ruling regime and would likely push back against technological approaches to regulation of hate speech through practices like Internet shutdowns. Social media shutdowns in these countries were, hence, justified along the need to shut hate speech.

I want to extend this argument by further noting that in semi-authoritarian regimes like Zimbabwe and Eswatini where the judiciaries are captured by the ruling party elites (see Ncube, 2019; Noyes, 2020; Fombad, 2021), several prospects on hate speech on social media exist. First, there is an unequal application of the laws governing hate speech. We can illustrate this by the case of Robert Mugabe, the former president of Zimbabwe. Mugabe's violent gay bashing rhetoric (Youde, 2017; de Saugy, 2022) is well documented, but, because of a captured judiciary, and captured state institutions, went unpunished or at the very least unrebuked. For instance, Mugabe called LGBQTI+ people as "worse than dogs and pigs because they don't even recognise their partners...dogs know the male and female ones... just like pigs...' This gay bashing rhetoric by Mugabe was not only unpunished and rebuked but reciprocated by his legion of admirers who exported this prejudice to social media platforms like Twitter and Facebook.

This signifies the dangers of judicial capture to social media hate speech in such countries where the judiciary are an appendage of the executive. The elasticity of the law is never fully recognised in order to provide constitutional protection to all groups that are victims of social media hate speech. Thus, these captured judiciaries often take a narrow interpretation of the law whenever they are called upon to adjudicate on social media hate speech, preferring to protect political players rather than at-risk communities. For example, Zimbabwean Courts have consistently[10] refused to offer constitutional protection to LGBQTI+ communities on many occasions because same sex relationships are deemed illegal in the county. This is what I mean by the lack of elasticity in the interpretation of hate speech laws. Captured judiciaries always adopt a narrow interpretation, choosing not to exercise their gift of intuition in ways that protect communities targeted by hate speech on social media.

4. Discussion

In this chapter, I undertook a comparative content analysis of social media hate speech in seven selected Southern African countries of South Africa, Zimbabwe, Eswatini, the DRC, Zambia, Namibia and Lesotho. Its aim was to examine how these countries, regulate social media hate speech, and how they legally sanction it. The chapter makes a number of interlinked findings. Firstly, the regulation of social media hate speech is variegated across the region. As demonstrated in the sections above, some countries, like South Africa and Zambia, have moved steps ahead in regulating social media hate speech. They have done this by promulgating social media hate speech specific laws in their constitutions. Other countries have adopted a 'broader approach', preferring to cover social media hate speech under existing laws of freedom of speech, expression and human dignity laws. In some countries, like Zimbabwe, there are hate speech regulations, but these are not specifically tailored at social media platforms. Yet, in other

[10] Many of the LGBQTI+ persecution cases have been met with dead silence in courts. Some of the persecution tales can be followed here: https://www.theguardian.com/world/2010/may/28/zimbabwe-gay-rights-workers-released-torture

countries, especially those that have hate speech regulations spread across different statutes of the constitutions, for example, Lesotho, hate speech is regulated under the country's freedom of information bill. It is also regulated under the country's media laws. This is also true of the DRC, where laws against hate speech are not social media specific per se, and are thinly spread out under different clauses of the constitutions.

There are obvious dangers with this approach. The most common danger is that it is difficult for people to grasp the full consequences of their actions if the laws governing hate speech are spread throughout the constitution. The second danger is that regimes always adopt this strategy to avoid exterior pressures to reform laws. When laws are thinly spread in different acts, any call for reforms of the laws would necessitate a dismantling of the whole constitution or a big chunk of it. This, regimes often argue, would not be possible. In this chapter, I have also demonstrated several issues about hate speech on digital platforms. Firstly, semi-dictatorial regimes prefer regulating hate speech through technological regulations. As noted in Eswatini, Zimbabwe and Zambia under Edgar Lungu, these countries shut down the Internet as a response to hate speech. Therefore, social media hate speech laws are weaponised to stop the opposition from organising and mobilising on these platforms. As I argued, the character of the regime (liberal-democratic, semi-dictatorial etc) determines the kind of response to hate speech. In liberal democratic contexts like South Africa, preference lies in content moderation as the ruling regimes do not have power to utilise technological interventions as a menu of manipulation. I have also, using examples, noted that in semi-dictatorial contexts, loopholes in hate speech regulations allow for the political exploitation of laws. A notable example being the 'gay bashing' practices of the late president of Zimbabwe, Robert Mugabe. I want to conclude by making a few recommendations about hate speech regulations in the region.

What this chapter points to is that the problem of hate speech is growing in the Southern African region. Therefore, this chapter makes a call for the strengthening of existing social media hate speech regulations in the region, and the promulgation of such, in countries where they do not exist. There are several pointers in this regard. The DRC, in 2021, gazetted the Racism, Xenophobia and Tribalism Bill which seeks to address a gap in social media hate speech regulation in the country. While the word ‘tribe’ itself has increasingly come under fire from anthropologists, and can be replace in this case, the country has followed South Africa in regulating hate speech laws that stand alone. South Africa has, in 2022, moved on with strengthening its responses to social media hate speech through the Prevention and Combating of Hate Crime and the Hate Speech Bill. Zambia in March 2021 passed the Cyber Security and Cyber Crimes Bill into law.

But because social media hate speech travels, there is need for the Southern African region to adopt novel approaches to combatting social media hate speech in the region. The problem of social media hate speech is a menacing growth. Strengthening country-specific legislation may not be sufficient and equal to the task. The point I make here is a call for a regional convention on social media hate speech. Regional legal agreements can go a long way to block hate speech content across the Southern African region. This would help combat social media hate speech in the region. Such a convention would place an obligation on all countries in the region to shutdown hate spewing social media platforms, or by extension, hate speech on any other platforms, in their countries, that might be aimed at another member state. A good example is Europe’s Protocol on the Convention on Cybercrime (see Timofeeva, 2003). This law criminalises cross-border hate speech targeted at some people. In addition to such a cross-border legal convention, which might take long, Southern African countries should start holding social media companies to account. There are no known incidences of such practices where these companies like Twitter and Facebook are forced to remove content. If other countries of the world are doing it, governments in the region can surely do it as well and combat hate speech.

5. Conclusion

Research on social media hate speech is still in its infancy in the region. This chapter provides a starting point by analysing existing regulations that govern social media hate speech in several countries of the region. The chapter notes numerous loopholes in most legislations. However, the region is made up of several countries which this research did not focus on. For example, Angola and Mozambique, former Lusophone countries, have not been subjected to this research because of the author's language limitations. Future research would extend this research by focusing on how Lusophone countries approach social media hate speech. This chapter might also have benefitted from interviews with legal experts who could have provided an informed legal analysis of the region's approach to social media hate speech. Future research would benefit from this approach. The strength of the chapter lies in the fact that it provides the legal synopsis, dissects existing hate speech regulations and provides practical examples of the inadequacy of these laws in the region. It, hence, not only adds knowledge to hate speech regulation in a region where the practice is rampant but provides a robust foundation on which future research might built on.

Acknowledgements

I would like to thank Tinotenda Lisa Bangajena (M.A candidate, University of Johannesburg, Department of Communication and Media), for helping with the data gathering between September and October 2022.

References

Bax, M. (2018). Qualitative content analysis for the study of legal instruments: Unravelling the concept of collective reparations in legal and quasi-legal institutions. *Methoden van onderziek in het strafrecht, de criminologie en de victimologie*, 21-34.

Breen, D., Lynch, I., Nel, J. & Matthews, I. (2016). Hate crime in transitional societies: The case of South Africa. *The globalization of hate: Internationalizing hate crime*, 126-141. Brill Nijhoff.

Business Tech. (2022) *New Hate speech Law coming for South Africa*. https://businesstech.co.za/news/government/623759/new-hate-speech-laws-coming-for-south-africa-including-legal-trouble-for-whatsapp-messages/

Constitution of the Republic of Botswana (2021). *The Constitution*. Accessible at: https://www.constituteproject.org/constitution/Botswana_2016.pdf?lang=en

Constitution of the Republic of DRC (2021). *The Constitution*. Accessible at: https://en.wikipedia.org/wiki/Constitution_of_the_Democratic_Republic_of_the_Congo

Constitution of the Republic of Eswatini (2021). *The Constitution*. Accessible at: https://www.constituteproject.org/constitution/Swaziland_2005.pdf?lang=en

Constitution of the Republic of Namibia (2019). *The Constitution*. Accessible at: https://www.ilo.org/dyn/natlex/docs/MONOGRAPH/9565/104615/F1065283089/NAM9565%202.pdfhttps://www.ilo.org/dyn/natlex/docs/MONOGRAPH/9565/104615/F1065283089/NAM9565%202.pdf. Accessed on 2 November 2022.

Constitution of the Republic of South Africa (2021). *The Constitution*. Accessible at: https://www.gov.za/documents/constitution/constitution-republic-south-africa-1996-1

Constitution of the Republic of Zambia (2021). *The Constitution*. Accessible at: https://www.ilo.org/dyn/natlex/docs/ELECTRONIC/26620/90492/F735047973/ZMB26620.pdf

Constitution of the Republic of Zimbabwe (2021). *The Constitution*. Accessible at: https://constitutions.unwomen.org/en/countries/africa/zimbabwe

de Saugy, Y. F. (2022). "We Are Not Gays": Regime Preservation and the Politicization of Identity in Mugabe's Zimbabwe. *African Studies Review, 65*(3), 591-614.

Dlamini, H. P. (2019). The Beginning of the Great Constitutional Debate: Agreeing to Disagree. In *A Constitutional History of the Kingdom of Eswatini (Swaziland), 1960–1982* (pp. 65-127). Palgrave Macmillan, Cham.

Fombad, C. M. (2021). The Struggle to Defend the Independence of the Judiciary in Africa. In *Challenged Justice: In Pursuit of Judicial Independence* (pp. 223-248).

Freedom House (2021). *The `world Democracy under siege*. https://freedomhouse.org/report/freedom-world/2021/democracy-under-siege.

Gagliardone, I. (2014). *Mapping and analysing hate speech online. Available at SSRN 2601792*.

Garbe, L., Selvik, L. M. & Lemaire, P. (2021). How African countries respond to fake news and hate speech. *Information, Communication & Society*, 1-18.

Kabamba, P. (2010). 'Heart of Darkness' Current images of the DRC and their theoretical underpinning. *Anthropological theory*, *10*(3), 265-301.

Kongolo, M. & Zamberia, A. M. (2016). State fragility and capacity building in Sub-Saharan Africa: The case of the Democratic Republic of Congo. In *State Fragility and State Building in Africa* (pp. 183-207). Springer, Cham.

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. London: Sage publications.

Limb, P. (2022). The constitution, the people and the reinvention of a royal autocracy: A Constitutional History of the Kingdom of Eswatini (Swaziland), 1960–1982. *Journal of Southern African Studies*, 48(2022), 215-219.

Linos, K. & Carlson, M. (2017). Qualitative methods for law review writing. *U. Chi. L. Rev.*, *84*, 213.

Maenzanise, J. (2018). Zimbabwe after Mugabe: Few Signs of Progress. *The Gay & Lesbian Review Worldwide*, *25*(3), 28-29.

Marais, M. E. & Pretorius, J. L. (2015). A contextual analysis of the hate speech provisions of the Equality Act. *Potchefstroom Electronic Law Journal/Potchefstroomse Elektroniese Regsblad*, *18*(4), 902-942.

Mayring, P. (2021). *Qualitative content analysis: a step-by-step guide*. London: Sage.

MISA (2021). *State of the media In Southern Africa*. https://misa.org/blog/state-of-the-media-2021-report-now-available/

Munoriyarwa, A. (2021). There ain't no rainbow in the 'rainbow nation': A discourse analysis of racial conflicts on twitter hashtags in post-apartheid South Africa. In *Hate Speech and Polarization in Participatory Society* (pp. 67-82). Routledge.

Munoriyarwa, A. (2022). The militarization of digital surveillance in post-coup Zimbabwe: 'Just don't tell them what we do.' *Security Dialogue*, *53*(5), 456-474. https://doi.org/10.1177/09670106221118796

Muparamoto, N. (2021). LGBT individuals and the struggle against Robert Mugabe's extirpation in Zimbabwe. *Africa Review*, *13*(3), S1-S16.

Mwananyanda, M. (2021). *Free speech? Zambia's new Internet law fails basic human rights scrutiny*. Daily Maverick. https://www.dailymaverick.co.za/article/2021-04-05-free-speech-zambias-new-Internet-law-fails-basic-human-rights-scrutiny/

Ncube, S. (2019). Law and the struggle for political power in Zimbabwe. *Journal of Southern African Studies*, 45(3), 617-619.

Noyes, A. H. (2020). *A New Zimbabwe Assessing Continuity and Change After Mugabe*. Rand Arroyo center Santa Monica ca Santa Monica United States.

Ntanyoma, R. D. & Shacklock, T (2021). *Hate Speech and Genocide in Minembwe, D.R. Congo*. USA: Genocide Watch

Ruhanya, P. (2020). The militarisation of state institutions in Zimbabwe, 2002–2017. In *The History and Political Transition of Zimbabwe* (pp. 181-204). Palgrave Macmillan, Cham.

SAHRC (2016, 6th July). *Increased complaints on racism means more people know their rights*. Available at: https://cutt.ly/9koNUFv

Scheffler, A. (2015). *The inherent danger of hate speech legislation: A case study from Rwanda and Kenya on the failure of a preventative measure*. Namibia: Friedrich-Ebert-Stiftung (FES).

Timofeeva, Y. A. (2003). Hate Speech Online: Restricted or Protected Comparison of Regulations in the United States and Germany. *Journal of Transnational Law & Policy*, 12(2), 253-286.

Wetzstein, I., Bartenberger, M. & Leitner, P. (2013). Hate speech on the rise: Phenomena, reflections and social media-driven concepts against cyber hate. *Web based communities and social media 2013*, 100.

Youde, J., 2017. Patriotic history and anti-LGBT rhetoric in Zimbabwean politics. *Canadian Journal of African Studies/Revue canadienne des études africaines*, *51*(1), 61-79.

Chapter 9

ETHIOPIAN SOCIO-POLITICAL CONTEXTS FOR HATE SPEECH

Muluken Asegidew Chekol
Debre Markos University, Ethiopia

1. Introduction

Socio-political issues that trigger hate speech could be various and complex to understand completely. Studies show that hate speech could be aggravated and intensified by corresponding factors such as economic suffering, migration pressures, the antagonism between groups for political power after the collapse of oppressive central regimes, and the simplicity of expressing hatred on media platforms (Minority Rights Group International, 2014). Congruent with this, Tirrell (2017) asserts that hate speech does not stand alone, but it goes with economic, migratory, and political chaos. More specifically socio-political contexts that special scourge minorities can cause some people to suffer deeply, while others remain ignorant and unsympathetic to them (Minority Rights Group International, 2014).

The social media practices also cannot be defined as phenomena that take place exclusively online; because, the online and offline world feeds each other (Althoff, Jindal & Leskovec, 2017). So that this study assumes the current Ethiopian socio-political context affects the level of social media hate speech prevalence, circulation, severity, and multiply the natures in the country. Thus, it is important to explain first the major issues happening in the country that this study assumes trigger hate speech, and determine the prevalence, nature, and severity of hate speech on any platform including social media. For this reason, a brief discussion is presented on Ethiopian politics manifestations, ethnic federalism concerns, party press relations, and the social media sphere focusing on the ongoing reform.

2. Ethiopian politics in the reform government

Ethiopia has passed through socio-political difficulties, specifically in the last century. In 1974, the Solomon dynasty ends with a military coup, and was replaced by a unitary socialist revolutionary's government; then in 1991 the Military Derg socialist over through by the ethnic federalist Ethiopian People Revolutionary Democratic Front (EPRDF). The EPRDF government was a front formed from four ethnic-based regional parties – Tigray Liberation Front (TPLF), Amhara National Democratic Movement (ANDM), Oromo People Democratic Organization (OPDO), and South Peoples and Nations Nationalities Organization (SPNNO). The front ruled Ethiopia for 27 years (1991 - 2018) under the leading role of TPLF as a godfather to others.

During EPRDF, the country's government structure, political economy, social and cultural questions, and solutions have been orientated by ethnic, religious, and other identities difference discourses (Ayalew, 2016; Bekalu, 2017). The government claimed to be a developmental democratic state (DDS). Nevertheless, both democratic and human rights violations have been rampant (Human Rights Watch, 2018; Amnesty International, 2018), and economically even if the government claims to score double-digit growth, about "83.5 percent of the population (87,643 people) are multi-dimensionally poor while an additional 8.9 percent are classified as vulnerable to multidimensional poverty" (UNDP, 2019, p. 6). Under the name of development, the government has confiscated farmers' land in the interest of investors, in the government term "developmental investors". Moreover, in 2015 by utilizing an integrated Master Plan between the capital city, Addis Ababa, and sounding cities in Oromia regional state, the federal government has established a land bank ready for the lease market.

Cognizant to the long-lasting taped political sphere, suppression of freedom of speech, jailing of critical voices, foreign-based anti-government armed struggles, and nationwide unemployment, the integrated Master Plan becomes an immediate cause for the mass youths' protest in Ethiopia. The peak of civil unrest in 2016-2017, which included extensive violent protests in the most populous regions of Oromia and Amhara, lead to the resignation of Prime Minister HaileMariam Dessalegn in February 2018. Following this, on April 2, 2018, Abiy Ahmed Ali (Ph.D.) has become a Prime Minister of Ethiopia and the time afterward referred to as "reform". Through the course of this reform many good and bad phenomenon has been happening.

The reform government released ten thousand of political prisoners, welcome foreign-based armed fighting groups, amended the anti-terrorism law, reconnecting the mobile data and broadband Internet services that were cut off since 2016, and unblocked more than 246 websites, blogs, and news sites that have been inaccessible for over a decade, allow the foreign-based struggle media OMN and ESAT to open studies in Ethiopia, revise the media law and ease licensing and allow mainstream media to expand, and end the ethnic oriented ruling party formation and form a new party named "Prosperity" excluding the TPLF. Despite the ruling party change, many opposition parties are still under ethnic grouping rather than political-economy ideology. On the other hand, nearly three million people are displaced and hundreds are killed because of ethnic tensions (Internal Displacement Monitoring Center, 2019); religious leaders and institutions (churches and Mosques) are burnt. In-group solidarities are becoming strong and out-group ties have become weak. On top of this, still, unemployment is high, and living cost is alarming.

With all the dire happenings, following the political reform including the peaceful transition of power, and social and political inclusiveness, Ethiopian has shown improvement in the Fragile States Index (FSI) comparing to the previous trends; regarding this, the fragile States index report prepared by Fund For Peace (2019, p. 25) stated:

> The reform includes boosting political inclusiveness, [...] freeing thousands of political prisoners, [...] inviting opposition parties into dialogue, [...] increasing civic space and accountability for human rights abuses [...] lifted restrictions on websites and media, and appointed a former jailed dissident as head of the national electoral board. These reforms have been reflected in the significant improvement in Ethiopia's FSI indicator scores for State Legitimacy, Human Rights and Rule of Law, and Factionalized Elites. The Human Flight and Brain Drain indicators also dramatically improved, [...] diaspora returning home amid the political change, including exiled opposition figures.

Despite these measures and improvements, the country is still under the Fragile States Indexes alarming stage. Although the Fund For Peace 2019 fragile states indexes score confirmed excellent progress, it also pointed that the end of 2018 and the beginning of 2019 were the dangerous period of growing violence along social group-based lines. The FFP's fragile states indexes report recommended that there is a need for sustained focus on narrowing the divide among providers of services among regions, focusing on employment for the youths, and reforming governance structures to moderate the division of people along group-based lines are critical to building resistance (Fund For Peace, 2019).

The "reformist" government seems to concern about the current socio-political situation of the country, especially the rebounding of ethnic tensions and conflicts, which is also the cause for the prevalence of hate speech and fake news in the country. It can be arguable, but as a solution, Abiy's government passed the "*Hate speech and Disinformation Suppression Proclamation No. 1185/2012*" bill last year aiming at suppressing ethnic tensions ignited by offensive and dangerous speeches. As stated in its preamble, the bill aims at curving the expansion of hate speeches and solving their impacts legally to maintain citizens' wellbeing and peace. While the state initiated and emplaced the law, the government faced different critics. Starting from the drafting of the bill, critics state that the government is good at diagnosing the problems and bad at proposing solutions. They asserted that the law would gage freedom of speech[1], and present it as vague and confusing law[2]; critics also recommend the government to work on media and information literacy, which is one of the solutions to mitigate the expansion of hate speech (Eneyew, 2020; Tewdrose, 2020).

Unfortunately, on top of the above-mentioned existing predicaments, the COVID-19 pandemic adds an extra crisis to Ethiopian politics. Due to the pandemic, on 31 March 2020, the National Election Board of Ethiopia (NEBE) announced that the election preparations were forced to be suspended; the election would be delayed for some period. Within a month, the parliament approved the board claim. Then, on 10 April 2020, the parliament declared a five-month state of emergency aiming to fight the pandemic. Political opponents argued that this causes constitutional crises, because, the parliament's five-year term ends in early October 2020 without holding a new election. Some opposition political parties' quest for an interim government arrangement when the parliament term end; and some others still accept the reform government to pursue its full power until post-COVID-19 election time.

The reform government comes up with its possible solutions to the suspended election – dissolving the parliament, declaring a state of emergency, amending the constitution, and seeking constitutional interpretation. From these four the government prefers and justifies "constitutional interpretation" as constitutional, less time and budget consuming and manageable during the COVID 19 pandemic. Political opponents of the ruling party oppose this idea and argue that the ruling party, Prosperity, will not have a constitutional mandate to govern after early September 2020. Oppositions also heard voicing on the politicization of the COVID- 19 and abusing the state of emergency for political repression[3]. The conflicting argument be-

[1] Human Rights Watch: https://www.hrw.org/news/2019/12/19/ethiopia-bill-threatens-free-expression

[2] Media@LSE: https://blogs.lse.ac.uk/medialse/2019/06/04/the-problems-with-ethiopias-proposed-hate-speech-and-misinformation-law/

[3] Africa Portal: https://www.africaportal.org/features/managing-the-politics-of-ethiopias-covid-19-crisis/

tween the government and some political opponents, from TPLF and OLF, raised the country's political tension level over the pandemic concern for weeks.

In the middle of the political storm caused by the interplay between COVID-19 pandemic, suspended election, and the proceeding constitutional crisis, Amnesty International (AI) comes with a humanitarian crisis report entitled *Beyond Law Enforcement: Human Rights Violations by Ethiopian Security Forces in Amhara and Oromia*[4]. Ethiopians have divided by the methods and outcomes of the report based on political interests and the AI misrepresentation of the country's current context and warring parties. The study's focus areas, Amhara and Oromia regional states blame the report as imbalanced, partially baseless, and ignorant to human rights. Some, even non-government entities, highly criticize the report for its politically motivated content and investigation methods including sources. For example, the Amhara Association in America, the opposition political parties Amhara National Movement and political activists describe the AI report as incredible. Ethiopian Human Rights Commission General Director also considers that some critics of the AI report are accepted because the Director believes the report lacks context (Interview[5] on OMN, 1 May 2020).

On the other hand, the Tigray regional state ruling party, Tigray Liberation Front (TPLF), Oromo Liberation Front (OLF), Oromo Federal Congress (OFCo), and some other pleased by the AI report[6] and accuse the government of more unreported human rights violations. Amnesty International also responds to critics on its report, saying that AI's focus is on government security human rights violations and it is so confident in the report. Critics and the International Crisis Group[7] also interlinked COVID -19 restrictions, suspending election, and human rights violations into the current multifac-

[4] Amnesty International: https://www.amnesty.org/en/documents/afr25/2358/2020/en/
[5] YouTube: https://m.youtube.com/watch?v=nTvfanHCOSY
[6] Addis Standard: http://addisstandard.com/news-olf-ofc-urge-govt-to-take-report-by-amnesty-seriously-heed-recommendations-to-reverse-troubling-human-rights-abuse-record/
[7] Crisis Group: https://www.crisisgroup.org/africa/horn-africa/ethiopia/managing-politics-ethiopias-covid-19-crisis

eted socio-economic and political crisis and they warn the government that these will further deepen the crisis, divide the society, and potentially disintegrate the country.

3. The faults of Ethiopian ethnic federalism

Here the intention is not to argue against multicity and decentralization rather it is about instrumentalizing ethnic identity in the country's politics and administration. Many confirm that the federal government structure, in principle, offers the opportunity to decentralize and share power among regional states, easy to administer geographically big countries, and gives relative identity-based rights in multi-identity countries (Bélair, 2016; Bekalu, 2017). Ethiopia fulfills these criteria – geographically big and consists of multi-lingual and cultural societies; hence, federalism could be a preferable government system. But, some scholars do not see Ethiopian federalism as a healthy government system for some reasons; they state that territorial exclusive right for ethnic groups is emphasized more and the system also offers major ethnic groups' dominance over the minority in terms of political and resources utilization rights (Dana, 2017; Tesfaye, 2017). For this reason, the federal system misses its initial aim to empower minorities. For instance, Tesfaye (2017, p. 232) asserts, "The constitution thus misses an opportunity to respond to ethnic concerns without freezing ethnicity as an exclusive political identity."

The Ethiopian federal system signifies political power, resources rights, and ethnic identity as the same, and hence the core resources become a dividend of major ethnic groups (Abbinik, 2011; Adeto, 2019). In turn, this creates native (people whose ethnic line is originated in the region they live) and non-native (people whose ethnic line is out of the regional state where they are living) community division in their own country, and the power and the resources exclusively belong to the "natives" (The 1995 Oromia Regional State Constitution Article 8, The 1995 Revised Benishangule-Gumz Regional State Constitution Article 2, and The Revised Harari People Regional State Constitution Preamble). Consequently, scholars, for example, Legesse (2015), Bélair (2016), Bekalu (2017), Adeto (2019), and others present this ethnic federal arrangement as the main cause of ethnic conflicts in Ethiopia and they urge the government to review it sooner than the problem moves up to the worst situation.

Some of the manifestations, the above scholars mentioned as the bi-product of ethnic federalism range from competition to inter-ethnic conflicts, tension, and crisis that might push the country to civil wars to disintegrating. In this regard, Ethiopian ethnic groups are competing and conflicting along administrative bounders of regional states, results in the death and eviction of citizens (Legesse, 2015; Bekalu, 2017; Adeto, 2019). Even ethnic conflict crisis management has been taken for easy and ignored by the government with the justification of historical imbalances to be corrected as mentioned in the EFDR constitution preamble. The historical imbalance arguments rely on the Imperial period land administration and political structure, which end almost 47 years ago in 1974 Military coup Derg's lead revolution.

Further, arguments also forwarded on Ethiopia ethnic federalism that the Ethiopian government "only 'decentralized' the problems by defining the sources of conflict as local, not national" (Bekalu, 2017, p. 51); and because of the ethnic federalism, the country's stable ethnic composition which has been built for a long time has been eroded since the 1990s (Yohannes et al. 2005; Abbink, 2006). For instance, the 1991 EFDR constitution preamble aspires to correct "historical injustices" and its Article 39 (3) offers ethnic groups the rights up-to secession, which implies advocating subtle revenge against historical suppressors "*Nefetegna*"[8] and for disintegration of the country respectively. For this reason, some accuse even the federal constitution of promoting ethnic conflicts and division (Abbink, 2006; Bekalu, 2017). Besides, this federalism even does recognize neither ethno-linguistic identity nor Ethiopia people, but created unitary regional states based

[8] An Amharic word refers Ethiopians who hold arm and have the quality of shooting on target. However, under the contemporary Ethiopian context, extreme ethnic politicians use the term to portray Amhara people with a twisted historical reason that the majority of Emperor Menelik soldiers who reunified Ethiopia were Amhara descents.

on those identities (Mehari, 2010; Adeto, 2019). Adeto (2019) presented it as “territorialising identity” and “ethnically identifying with territory”.

The coming of Prime Minister Abiy Ahmed to power changes little on the perceptions of ethnic politics and changes nothing on the ethnic federalism structure. Following the reform, “While the 2019 FSI score shows improvements under both Security Apparatus and Group Grievance indicators, the end of 2018 and beginning of 2019 have indicated risks of rising violence along group-based lines that will need to be carefully monitored” (Fund For Peace-FFP, 2019, p. 26). After the FFP report, at the end of 2019 and the beginning of 2020 many more ethnic-oriented conflicts happened in the country. Based on the Ethiopian government reports, for instance, subsequent to activist Jawar Mohammed’s post on Facebook on 23 October 2019 that he was under a security threat, more than 97 people were killed in October 2019; and further, soon after an Oromo Singer Hachalu Hundessa assassination, more than 153 civilians were killed, 239 injured, and hotels, banks, commercial centers, residents, and even health centers and schools were destroyed by mob in July 2020 in Oromia region.

Consequently, since hate speech is identity sensitive, ethnic conflicts can aggravate hate speech prevalence and vice versa among social identities such as ethnicity, religious and political affiliations (Neshkovska & Trajkova, 2017); and that is why Ethiopian ethnic federalism can be considered as a potential background cause of hate speech. When identity becomes a source of political power and a means access to resources and advocated by elites, and supported by government bureaucratic machinery (Ezeibe & Ikeanyibe, 2017), it has the capacity to boost up hate speech to the worst scenario. To end division, create sustainable political settlements, national and local dialogue is vital and such dialogue requires a free media that is free from factional agents or governmental influence (Abbink, 2006; Bekalu, 2017; Adeto, 2019). The worst is the media itself is divided in Ethiopia, mainly along ethnic lines.

4. The current state of mainstream media in Ethiopia

The Ethiopian media has almost more than a century old which functions and passes in quite different government systems. It started as the Kingdom image builder, utilized as the revolutionary socialist government propaganda tool, and serves as the developmental state agenda advocator (Shimelis, 2000; Salih, Eshete & Assefa, 2018). None of the Ethiopian governments allowed the media independence and freedom to criticize governments' policies. There were exceptions during the transition times, for example, Ethiopian media enjoyed relative freedom for two years when Emperor Haileselasie's Solomonic dynasty government overthrew by the Derg military force in 1974, and when the Derg regime oust by the EPRDF in 1991 (Shimelis, 2000; Dodolla, 2019).

During the EPRDF, the Ethiopian mainstream media is politically controlled and used to reporting on so-called developmental issues lean-to echo government positions. Reports stressed that Ethiopia remains a closed society where its media is used to be restricted by the already expelled "*Charities and Societies*" and "*Anti-Terrorism Proclamation*" laws and external funding for media and human rights advocacy was controlled (International Media Support, 2018). Similarly, FOJO Media Institute (2019) report noted that the "public service broadcasters", the national radio and leading TV news station are owned and controlled by the government, and to control the media further the government was asking assistance with media policy from the West aiming to maintain the political monopoly.

Furthermore, "the [EPRDF] government surprised the satellite media owners, which are considered to be free from government control, by requesting that they reregister their licenses with EBA" (Wazema Radio, 2018, p. 5). For this purpose, as Wazema Radio revealed that, in collaboration with Ethiopian Information Network Security Agency (INSA), and Metals and Engineering Corporation (METEC), the EPRDF was working to construct a receiver that used to offer it a total media-controlling means, and opponents were criticizing the government action "as repackaging developmental propaganda through freshly branded shows on channels controlled by political loyalists" (Wazema Radio, 2018, p. 11). The EPRDF move to control everything fueled protests. The International Media Support (2018) report shows that following protests in (2016, 2017, 2018) people had been asking "whether the ruling Ethiopian People's Revolutionary Democratic Front (EPRDF), which has long survived on political suppression of the media, civil society, and opposition political parties, would be prepared to open up the political space" (p. 8).

Like the other political transition times, after the April 2018 peaceful power transitions within the same political party, from P.M HaileMariam Dessalegn to P.M Abiy Ahmed the Ethiopian media show open and new media join the market (Freedom House, 2018; Dodolla, 2019). One of the critics of the EPRDF government, Committee Professional Journalists (CPJ) in its 2019 report, states that Ethiopia, used to be one of the most censored countries in the world and the worst jailers of journalists, shows a significant political reform. Because of the reform, journalists and bloggers were freed from jails, bring back exiled journalists, and avoid jamming foreign-based media (Freedom House, 2018; World Press Freedom Index, 2019).

Subsequent to the reform, human rights abuse and mal administrations started being reported even on government media (International Media Support, 2018); although government-owned, and affiliated media are bouncing back to their agenda selling function to the current reform government. For instance, Ethiopian Television, Fana Broadcasting, Walta Media, Amhara Mass media Agency and Oromia Broadcast Network are promoting the reformist government leader, P.M. Abiy's political-economy world view '*Medemer*' (which literary mean synergy), and criticizing the EPRDF's developmental state views what they used to stand for before the reform. After showing short time support to the reform actions, the private media have been keeping its previous role of highlighting the government's fallers.

Despite some changes in the reform, still, the media is under the EPRDF's government orientation. Even, before the reform, some appreciate the government for its legal guarantee to the media mentioning the 1995 constitution Article (29), Freedom of the Mass Media and Access to Information Proclamation No. 590/2008, and for endorsing international laws (Media Progress Europe, 2018). On the other hand, the government has been criticized for imposing repressive jurisdiction on the media such as the Anti-Terrorism Proclamation of 2009 (expelled in the reform), the Computer Crime Proclamation of 2016 (Badwaza & Temin, 2018), and Hate Speech and Disinformation Suppression law ratified on February 14, 2020 (Yohannis, 2020).

One of the EPRDF's government major legacies has been establishing ethnic sensitive regional states and institutions including media. This leads to the establishment of government-owned regional ethnic-based media houses, basically, intended to serve the regions and ethnicities they belong to. Following the government's footsteps, private media such as Oromia Media Network (OMN), Amhara Satellite Radio and Television (ASRAT) Media, and Dimtsi-Weyane (DW TV) used to get a license and have been broadcasting their programs in Ethiopia following the 2018 political reform. Study shows that such identity-based media sometimes play the role of "intensifying the politics of identity and at other times transcending such politics" (BBC Media Action, 2013, p. 4). Dissimilar to this, Ethiopian Satellite Television (ESAT), which presents itself as Pro-Ethiopianism nationalist media, works to counter ethnonationalism movements. Following the reform, ESAT, the media that used to be jammed by the EPRDF government, also opens its studio in Addis Ababa.

Study shows that "there are real dangers if media is routinely co-opted by factional actors, and that fragmentation of the media can lead to the reinforcement of factional identities" (BBC Media Action, 2013, p. 25). Currently, in Ethiopia, some media abuse freedom to the extent of practicing hate speech. For instance, Oromo Media Network (OMN) broadcasted a program that divided the Ethiopian Orthodox Church[9] unity and abusing its

[9] YouTube: https://www.youtube.com/watch?v=gdp6QNLwAyI

name. Further, on March 8, 2020, when Oromo Federal Congress celebrate Women day, a young Oromo girl[10], among the participant, on live transmission, speak out passionately calling for Oromo men to divorce from their Amhara decent wives and should marry only Oromo girls. This might have happened because the Amhara are *Neftegna/ Abyssinians/ Showa* whom Oromo politicians portray as their colonizer and enemy as some Oromo scholars, for example, Mohammed Hassen (2002), and Asafa Jalata (2003) argued so. These are clear manifestations of abusing media freedom in Ethiopia's current context.

To add more, afterward the assassination of an Oromo ethnic Singer Hachalu Hundessa, in the next morning OMN, DWTV, and Tigray TV feed coverage one other in blaming *Neftegna* (Amhara) for the assassination and provoking Oromo Youth to go out and revenge, and consequently, according to government official's reports hundreds were killed in Oromia regions based on their ethnic identity. Critics and pressures mount against the Ethiopian government to act on the media, and after a day from the assassination on 30 June 30, 2020, OMN's Addis Ababa studio was locked by the government, and after five days on June 5, DWTV and Tigray TV downcast from a satellite.

In a country divided by ethnic politics and as a result attains the level of alarming fragile state (Fund For Peace, 2019), the media is also fractured and marginal social groups and vernacular language media have become more widespread. In this vein, "Both commercial and political drivers have fueled the growth of media catering for specific communities – linguistic, ethnic, religious and political" (BBC Media Action, 2013, p. 7). Ethiopian ethnonational media expansion can be an indication. Further, online tubes are growing too hence political activists and other elites are easily accessing them to address their ethnic group. It is worrisome because, fragmented media in a cracked state has the role of "reinforcing and intensifying sepa-

[10] YouTube: https://www.youtube.com/watch?v=lnMaGJdlYpU

rate identities rather than encouraging the development of shared identity" (BBC Media Action, 2013, p. 6).

Further, following the political reform, many Ethiopian media become more ethnic center than national unison voice. Regional media are competing for ethnic loyalty; and they call for in-group solidarity by representing the out-groups as an enemy, and the media in turn is portrayed as trigger instruments of violence and divergence (Negera, 2019; Addisu, 2019). Such media practices are apparent in Oromia, Amhara, and Tigray regional states in both government and private media houses. These aggravate the ethnic tensions, fears, and conflicts in the country, and hate mature over time (Negera, 2019; Addisu, 2019). According to scholars in the field, this is the consequence of the interplay among media, ethnicity, and politics (Nyamnjoh, 2010; Matsaganis & Vikki, 2014; Benequista, 2016).

Ethnic or religious or political and or other identity-based media are also accused of insisting groups seeking to leverage identity politics to destabilize from the notion of dividing citizens as "We" and "They". In the Ethiopian context the escalating of, for example, ethnic-based media is imposing a threat to society by rising conflicts and winding antagonism even among the media themselves aiming division along the peoples' identity lines (Melesew, 2019; Addisu, 2019). Following the political reform and relatively limited federal government control over the regional ethnic-based media, they are inclined to side with the ethnic groups they stand for more than any time before (Wazema Radio Report, 2018; Skjerdal & Mulatu, 2021).

When the media concentrated on political activism, their practices affect the online discourses since the offline world influence the online world (Althoff, Jindal & Leskovec, 2017). Accordingly, the more ethnicization of the current Ethiopian mainstream media, and cases of hate speech incidents in their programs, give a hint of what is happening in the Ethiopian social media sphere. In comparison, because of the absence of monitoring mechanisms and organizational editorial, the social media condition can be worse than the mainstream media, which have a firm gate-keeping system.

The current general circumstances of Ethiopia online, mainly social media (Facebook and YouTube), are presented below in brief.

5. Social media sphere in Ethiopia

Social media is a recent phenomenon, a small start-up, and the fastest scale-up invention on earth. This time, a single social media site is accessed by billions of people around the world. For instance, Facebook has 2.22 billion and YouTube[11] has more than 2 billion monthly active users worldwide from more than 4.585 billion people who can access the Internet until 2020 (Internet World Status, 2020). Developing countries are following behind the developed on engaging and utilizing social media. Still, there are a significant number of social media users. For instance, in Africa, there are 526.7 million Internet users and 213 million Facebook users (Internet World Status, 2020). Even in many of the fragile states Mobile phone use has mushroomed, and grow nearly six-fold in five years (BBC Media Action, 2013, p. 7). Hence, the online, mainly social media becomes a vital agency in politics, economy, and social affairs around the world.

Ethiopian social media, users' engagements, and government hand over it, is not a different case from other African countries if not to be the worst. According to Ethio-telecom 2019, and 2020 annual reports [[12],[13]] Data and Internet users reaches 22.3 million in June 2019 and 23.5 million in December 2020 respectively, out of estimated to be 114 million populations. In the perception of 'most' Ethiopian users, social media sometimes equates with the general Internet services, and again Facebook by itself equates social media although they are different in scope and characteristic. Mirani (2015) indicates that in the developing world, people, actually use Facebook and the Internet interchangeably in some places. *Gagliardone et. al. (2015, p. 9) notes that* for many users "Facebook is the Internet" *in Ethiopia too and de-*

[11] Omnicore Agency: https://www.omnicoreagency.com/youtube-statistics/
[12] Ethio Telecom: https://www.ethiotelecom.et/2018-19-efy-p-reporte/
[13] Ethio Telecom: https://www.ethiotelecom.et/ethio-telecom-2013-efy-2020-21-first-half-business--performance-summary-report/

bates on social media "such as Facebook, have attracted significant attention and offer a broad spectrum of the most important themes characterizing Ethiopia's political sphere." These points affirm the place of social media, particularly Facebook in developing countries such as Ethiopia.

Among social media sites, Facebook and YouTube are popular in Ethiopia. From the total 23.8 Ethiopian Internet users, Facebook had more than 4.5 million users (Yohannse, 2019); and it reaches 6.07 million[14] in 2020. YouTube also has an important share of audiences in Ethiopia. Only the top ten YouTube sites have more than 3.44 million subscribers[15], and from this, the Ethiopian mainstream media YouTube channels' audiences share is also very significant, for instance, 12 mainstream media television channels account for 3.843 million subscribers and 686 million total uploaded video views until October 2019.

From scratch, the Ethiopian governments developed the trend of restricting, blocking critical sites, and even total shutdown of Internet services. For instance, the EPRDF government had shut down the Internet six times from 2017 to 2018 (Yohannse, 2019). Like journalists, bloggers and social media activists had been the target of government security, many of them threw to jail accused of violating the anti-terrorism law (the law voided in the reform). Critical voices such as Zone 9 boogers[16], Seyoum Teshome[17], and Getachew Shiferaw[18] were the victims of engaging in online media. Getachew Shiferaw authors a book entitles "Asekaki-Demotsoch" (horrifying voices) from his eyewitness in the prison. The book tells the story of torture and torture to death prisoners for opposing the government including on social media.

[14] Internet World Stats: https://www.internetworldstats.com/stats1.htm
[15] Social Bakers: https://www.socialbakers.com/statistics/youtube/channels/ethiopia/
[16] Human Rights Watch: https://www.hrw.org/news/2015/04/23/ethiopia-free-zone-9-bloggers-journalists
[17] Al Jazeera: https://www.aljazeera.com/news/2016/10/oromo-protests-ethiopia-arrests-blogger--seyoum-teshome-161005071925586.html
[18] Committee to Protect Journalists: https://cpj.org/2017/05/ethiopian-high-court-convicts-editor-of-inciting-s.php

In Ethiopia, Telecom service is a government monopoly and it is considered as a "cash cow" and the hand of the government to control the flow of information. Using this monopoly opportunity, like that of closing the newspapers, and jamming the foreign-based broadcast media (Wazema Radio, 2018), the EPRDF government were locking websites, disconnect mobile data, and slow online information access. For instance, as Eneyew (2019) notices that the EPRDF government was blocking social media to control informally organized youths' sparking protests such as *Fano* in Amhara, *Qerro* in Oromia, and *Zerma* in Gurage. Congruent to this, International Media Support (2018) report confirmed that Ethiopian authorities responded to the wave of youths' protests by switching off mobile Internet access aiming to limit the scope of the movement against the EPRDF government.

Following the reform, the Ethiopian government reconnected the mobile data and broadband Internet services that were cut off since 2016 and unblocked 246 websites, and blogs that were inaccessible for over a decade (Badwaza & Temin, 2018). Along with an end to the blocking of the websites and restoration of access to better Internet services journalists, activists, and bloggers who used to work in exile back home and expand the accessing online information from local sources (World Press Freedom Index, 2019). The political open-up has been seen on both the social media and mainstream media discussion among opposing views.

Despite improvements still, it is recommended to lift Internet access, standardize laws that prosecute freedom of expression, refrain from surveillance, censorship, and Internet shutdowns (Badwaza & Temin, 2018). Despite the recommendation, during the ongoing political reform, Internet services were shut down twice in June 2019 alone following the students' national exam, and the killing of political and military leaders in Amhara regional state and Addis Ababa. Subsequent to the killing of Oromo Singer, the government shut down the Internet from the end of June to the first week of July 2020. Then, nobody, neither Ethio-telecom nor the security forces, explains the Internet shut down to the taxpayer and the people. Both before and after the reform, when the Internet is blocked there is no pre-warning and post justification although the reform brings positive changes including opening up the political space.

Beyond the positive changes, the reform government Internet and social media relative freedom bring the dysfunctions that even hurt the reform itself and the wellbeing of the society. Especially along with the freeing of social media, information explosion, disinformation, fake news, and hate speech becomes the new reality in Ethiopia (Yohannse, 2018). Similarly, Dibaba (2019) asserts that the dissemination of hate speech online is put in danger the human and democratic rights, the long-standing social solidarity, and eventually led to political and socio-physiological disorder by weakening the state apparatus and the country sovereignty.

Unfortunately, the expansion of social media and as Kiruga (2019) asserts that the hate speech and fake news dissemination on social media already started to affect the lives of millions of people, universities students' education, business activities due to the closure of roads in the country, the movement of citizens hindered, and millions were evicted while hundreds have been killed. Consequently, FOJO Media Institute, (2019, p. 1) stated, "Social media is the dark horse. [It] has, in a few years, turned the political landscape in Ethiopia on its head not only as a means of mobilizing people but also as a means of spreading rumors, hate speech, and disinformation." The Ethiopian government seems worried about the impact of social media hate speech and disinformation and for that reason, it proposes a bill to regulate it.

Conversely, the expansion of smartphone offers citizen journalism the opportunity to flourish as an alternative to the mainstream media (BBC Media Action, 2013). This would be taken as an opportunity to societies like Ethiopia where the mainstream media access is limited, incapable to address the multifocal living challenges. However, because of its easy and quick facility to disseminate hate speech and fake news, the open social media impacts the lives of millions by disturbing schools and universities, destructing business activities due to closure of roads by protesters, delaying the movement of citizens, evicting millions of people and killing hundreds (Seid, Abinew & Biemann, 2020). Social media in Ethiopia has become a source of intensifying ethnic, religious, and political-based antagonism that may aggravate hate speech in Ethiopia's new reform socio-political contexts where this study builds up its foundation.

6. Social media hate speech under Ethiopia interregnum

I had conducted a study on Ethiopian social media hate speech during the country's political reform. The study was held in a country under a political transition in tandem with tense ethnic politics that resulted in thousands of dead and millions of internally displaced people. Ethiopia has been in a state of interregnum waved by political turmoil including a full-fledged war in the Tigray region, insurgency civilian attacks in Oromia *Wollega*, Benishangule-Gumz *Metekel*, South Ethiopia *Gurafereda*, North *Shewa-Ataye*, and others. Eventually, such offline animosity has been duplicated to the online sphere, predominantly on Facebook and YouTube. Consequently, the study favored underscoring identity-based, specifically ethnicity, political, and religious driven hate speech under the intersection of the mainstream media and the social media; it explained young Ethiopians social media users' perceptions, exposures, and reactions towards social media hate speech; and it evaluated social media hate speech regulation under the Ethiopian laws. In other words, the study appraised the prevalence, perceptions, and regulation of social media hate speech in Ethiopia.

The study employed a mixed-method approach in tandem with a pragmatism worldview as both underscore what works best to the research problem rather than adhering to neither positivist nor interpretivists' doctrines. Using multi-stage sampling users' comments offered on three purposeful selected Ethnic media's social media outlets (ASRAT, OMN, and DWTV) were collected and analyzed under the quantitative phase. Adhering to the mixed-method explanatory sequential design, young Ethiopian users' focus group discussions, and legal scholars, lawyers, and experts' interviews together with hate speech laws analyzed under the qualitative phase; then, both the quantitative and qualitative findings were integrated and supported by theories, previous studies, and literature in the discussion under the consideration of the political reform contexts. Below is the main findings executive summery.

In line with this, regarding the prevalence of social media hate speech, a significant size of hate speech was found from users' comments offered on the ethnic-based media's social outlets content. Using binary analysis (hate vs. none hate), the size of hate comments was determined, and out of 8,525 users' comments, 2,834 (33.24%) hate comments were found. More than a third of the users' comments were identified as hate speech that fulfills the definition and conception of hate speech stated by various literatures. OMN's social outlets dominated by contributing the highest number, 1,310 (46.22%), of the total hate comments among the selected ethnic-based media's social outlets.

Regarding hate severity of social media hate speech, the size of offensive users' comments on the selected media's social outlets was the highest. Accordingly, out of the total 2,834 hate comments, 2,720 were offensive and insulting. The remaining hate comments were under the level of incitement to violence (89) and incitement to genocide (25). As of the size of the prevalence, the highest number of offensives, incitement to violence, and incitement to genocide comments were found from users' comments offered on OMN's social outlets (Facebook pages and YouTube channels) per content.

The natures of social media hate speeches were recognized as ethnic-based, politically motivated, and religious-based hate speech. The findings revealed that ethnic-politics hate speech suffocated the Ethiopian social media sphere. About 904 politically motivated, and 896 ethnic-based hate comments were found. The ethnic-politics together covered 1,800 (63.51%) hate comments out of the total, while contrary to the expectation only 71 hate comments were religious-based. The Tigray People Liberation Front (TPLF) was the most targeted political party followed by the Prosperity Party (PP) by the social media consumers of the three ethnic media houses. Among the Ethnic groups, the Oromo and Amhara ethnic groups were the most targeted ethnic group in Ethiopia during the last two and a half political reform periods.

The country's political history contesting narratives, ethnic religions, the quest for land adhere to the ethnic oriented federal structure was identified as the main trigger factors of social media hate speech in Ethiopia. Besides, various horrific incidents of the political reform including the assassination of high political and public figures have severed as trigger factors. The power struggle PP and TLF and the animosity between the Oromo, Amhara, and Tigray in tandem with identity-driven mass killings of civilians intensified the online hate speech circulation. These were the issues mentioned in the users' hateful comments frequently and thus extracted and thematized as main trigger factors of social media hate speech under the country's existing socio-political contexts with great emphasis on the ongoing political reform.

Related to users' perceptions, exposures, and reactions towards social media hate speech, the main findings from young Ethiopians' focus group discussions indicated that hate speech is understood as falsehood narratives publicly expressed against others. Hate speech has the aim of demoralizing the group members. Corresponding with literature many of the young Ethiopians believed that hate speech is targeting others' social identities such as ethnicity, religion, and political view to humiliate, insult, and lower them while boosting, and showing the strength of the self-social group. The young people blamed social media in general and Facebook in particular, as the main platform of hate speech. Consistent with the findings of the quantitative data, young Ethiopians underscored ethnicity, conflicting historical narratives, and divisive political rhetoric as major motivating factors of hate speech. They also argued that the political reform exaggerated more social media hate speech.

Many participants of the group discussion exhibited hate speech on their social media sites, as well as hear from their friends; most of the young people ignored the hate content without any reaction. Ethnic-based hate speech appears to be the most rampant on social media, and thus they supposed that if there would be a political social media hate speech could be limited. Stating the impacts of hate speech on individual and society level, users suggested a serious punitive regulation in addition to other preventive means.

The issue of social media hate speech regulation is one of the pillars of this dissertation, which is discussed based on professionals' evaluations, critiques, and arguments. Accordingly, the insights of legal scholars, lawyers, and EHRC and EMA legal experts informed the necessity, and the pros and cons of regulating social media hate speech in Ethiopia. How to regulate hate speech without affecting freedom of speech, as these two are competing rights, are central in the scholarly debates of speech regulation. Such debate has been held between the proponents (regulationists) and the opponents (minimalist) of hate speech regulation. The regulationists' approach was reflected in this study as all the interviewees agreed on regulating hate speech disseminated by any public platforms including social media. Although, many of them expressed their concern against the government's track record of using such laws for political purposes recalling what was happened in the expelled anti-terrorism law.

Related to the Ethiopian Hate Speech and Disinformation Prevention and Suppression law, legal scholars and lawyers stated that the law has the intention to manage the new technology-driven communication channel - social media and serves as a tool to govern the political fights between the pro-reform and anti-reform political actors. They believed that law contributes by securing society's wellbeing, and social coherence, as well as promotional and educational role regarding social media usage. On the other hand, the law falls short to qualify hate speech objectively, and rigorous penalties. Besides, arguably, it uses 5000+ followers as a cut point to criminalize social media users, and miss to prescribe liabilities against SNS.

Regarding institutional capacities, although the Ethiopian Human Rights Commission (EHRC) and Ethiopian Mass Media Authority (EMA) mandated under this law to work on public awareness on hate speech and disinformation respectively, the legal scholars, lawyers, and experts questioned their capacities to execute the mentioned duties, hence restructuring and capacity building is necessary to them. Some even recommended a new independent entity that specifically addresses hate speech issues. Consequently, they recommended reconsidering international experiences, clarifying operational definitions, qualifying hate speech objectively, and updating EMA and EHRC as per their duties under the law.

Cognizant of the multiple effects of hate speech, legal scholars suggested non-punitive preventive means, such as, not limited to, educating the people, awareness creation, and establishing strong and independent institutions that address hate speech concerns. Media literacy is considered vital to making people critical of media content, political actors, and extremists' messages. Furthermore, reshaping the country's ethnic-oriented political economy, including the federal and regional states' constitutions is suggested to limit the impacts of hate speech.

References

Abbink, J. (2011). Ethnic-based federalism and ethnicity in Ethiopia: Reassessing the experiment after 20 years. *Journal of Eastern African Studies, 5*(4), 596-618.

Abbink, J. (2006) Ethnicity and conflict generation in Ethiopia: Some problems and prospects of Ethno-Regional Federalism. *Journal of Contemporary African Studies*, *24*(3), 389-413. https://doi.org/10.1080/02589000600976729

Addisu, R. (2019). *The coverage of internal conflict between Oromia and Somali regional states on Ethiopian media: Comparative study of OBN and SRTV.* (Unpublished MA thesis). Addis Ababa University.

Adeto, Y. A. (2019). Preventing violent extremism in the Horn: The case of ethnic extremism in Ethiopia. *Policy paper*, European Institute of Peace. Accessed on February 16, 2021. https://www.eip.org/wp-content/uploads/2020/06/Ethnic-extremism-in-Ethiopia-policy-paper-July-2019.pdf

Althoff, T., Jindal, P. & Leskovec, J. (2017). *Online actions with offline impact: How online social networks influence online and offline user behavior.* Accessed on October 14, 2019, from https://cs.stanford.edu/~althoff/althoff_online_offline_appendix.pdf

Amnesty International. (2018). Amnesty International Report (2017/18): The state of the world's human rights. Accessed on March 16, 2020, from https://www.amnesty.org.nz/state-worlds-human-rights-annual-report-2017-

Ayalew, T. (2016). *The struggle of mobility: Organizing high-risk migration from the Horn of Africa.* In an open democracy, beyond trafficking and slavery. Accessed on 24th, November 2019 from www.opendemocracy.net

Badwaza, Y. & Temin, J. (September 2018). *Reform in Ethiopia: Turning promise into progress. Freedom House policy brief.* Accessed on December 22, 2019. https://freedomhouse.org/sites/default/files/policybrief_reform_in_ethiopia_0.pdf

BBC Media Action. (October 2013). Fragile states: the role of media and communication. (Editor: James Deane). *Policy briefing.* www.bbcmediaaction.org

Bekalu, A. (2017). Ethnic federalism and conflicts in Ethiopia. *African Journal on Conflict Resolution, 17*(2), 41-66.

Bélair, J. (2016). Ethnic federalism and conflicts in Ethiopia. *Canadian Journal of African Studies / Revue Canadienne Des Études Africaines, 50*(2), 295–301. https://doi.org/10.1080/00083968.2015.1124580

Benequista, M. P. (2016). *The moral dilemmas of journalism in Kenya's politics of belonging.* (Unpublished Ph.D. Dissertation), London School of Economics.

Benshangule-Gumz National Regional State. (1995). Revised Regional Constitution

CPJ. (2019). *Annual Report*. https://cpj.org/about/CPJ.Annual.Report.2019

Dana, M. (2017). Why the media's role in issues of race and ethnicity should be in the spotlight. Journal *of Social Issues*, 71, 1-16.

Dibaba, S. (2019). Hate speech and freedom of expression in Ethiopia. *The Ethiopian Herald, May 8/2019*. https://www.press.et/english/?p=5520#

Dodolla, N. A. (2019). *Ethiopian media industry. August*. https://doi.org/10.13140/RG.2.2.33768.42247

Eneyew, Y. (December 2019). *በኢትዮጵያ በኢንተርኔት ሃሳብን በነጻ የመግለጽ መብት እና ገደቡ;* [Internet freedom of speech in Ethiopia] Accessed in March 01, 2020. https://www.abyssinialaw.com/blog-posts/item/1842-2018-12-21-17-31-53

Eneyew, Y. (February 2020). *The new Hate Speech & Disinformation Suppression Proclamation No. 1185/2020: What it is and what it means to freedom of expression in Ethiopia*. Accessed on March 01, 2020. https://threadreaderapp.com/thread/1228320295548837889.html#

Ezeibe, C. & Ikeanyibe, O. (2017). Ethnic politics, hate speech, and access to political power in Nigeria. *Africa Today*, 63(1), 65-83. https://doi.org/10.2979/africatoday.63.4.04

FOJO Media Institute (April 2019). *Social media and journalism in Ethiopia: Setting the scene for reform*. The report accessed online on June 23, 2020, from

Fund for Peace. (2019). *Fragile states index annual report 2019*. www.fragilestatesindex.org

Gagliardone, I. et al. (2015a). Mechachal. A preliminary assessment of online debates in Ethiopia (Report One), 1-39.

Global report on Internal Displacement (May 2019). *GIRD*. Accessed in August 15, 2020 from https://www.internal-displacement.org/sites/default/files/publications/documents/2019-IDMC-GRID.pdf

Harrir National Regional State. (1995). *Regional Constitution*.

Hassen, M. (2002). Conquest, tyranny, and Ethnocide against the Oromo: A historical assessment of human rights conditions in Ethiopia, ca. 1880s-2002. *Northeast African Studies,* New Series, 9(3) 15-49. https://www.jstor.org/stable/41931279 Accessed: 29-04-2020 06:36 UTC.

Human Rights Watch. (January 2018). *Country summary: Ethiopia*. Accessed online March 16.

International Media Support. (October 2018). Ethiopia in transition: Hope aimed challenges. *Media and freedom of expression assessment*. Accessed on June 23, 2020, from https://www.mediasupport.org/wp-content/uploads/2018/11/Ethiopia.final_.spread-1.pdf

Internet World Status. (2020). *Internet world status: Usage and population statistics*. Retrieved on March 24, 2020, from https://www.internetworldstats.com/stats1.htm

Jalata, A. (2003). Comparing the African American and Oromo Movements in the global context. *Social Justice*, 30(1), 67-111. from https://www.jstor.org/stable/29768167 Accessed: 29-04-2020 06:36 UTC.

Kiruga, M. (2019). Ethiopia struggles with online hate ahead of telecoms opening. *The African Report*. https://www.theafricareport.com/19569/ethiopia-struggles-with-online-hate-aheadof-telecoms-opening/

Legesse, T. (2015). Ethnic federalism and conflict in Ethiopia: What lessons can other jurisdictions draw? *Africa Journal of International and Comparative Law*, 23(3), 462-475.

Matsaganis, M. D. & Vikki S. K. (2014). How ethnic media producers constitute their communities of practice: An ecological approach." *Journalism*, 15(7), 926-944.

Media Progress Europe. (2018). *Overview of the Ethiopian media landscape*. https://www.mediaprogress.net/downloads/Overview%20of%20the%20Ethiopian%20Media%20Landscape%.pdf in February 2, 2019

Mehari, T. (2010). Federalism and conflicts in Ethiopia. *Africa Insight*, 39(4), p. 132.

Melesew, D. (2019, May 2). Experts urge mainstream media to work in curbing hate speech and fake news. *Youth Times.* Retrieved on 12th May 2019 from https://www.youthtimes.info/news-page/experts-urge-mainstream-media-to-work-in-curbing-hate-speech-and-fake-news-5.html

Minority Rights Group International. (2014). *State of the World's minorities and indigenous Peoples 2014.*

Mirani, L. (2015). Millions of Facebook users have no idea they're using the Internet. *Quartz Web Site*, retrieved on December 12, 2019, from http://qz.com/333313/milliions-of-facebook-users-have-no-idea-theyre-using-the-internet/

Negera, T. (2019). The paradoxical Ethiopian media. *The Reporter.* Retrieved on 15th December 2019 from https://www.thereporterethiopia.com

Neshkovska, S. & Trajkova, Z. (2017). The essential of hate speech. *Ijet*, *14*(1), 71–80. https://doi.org/10.20544/teacher.14.10

Nyamnjoh, F. (2005). *Africa's media, democracy and the politics of belonging.* London: Zed Books.

Oromia National Regional State. (1995). *Regional Constitution.*

Salih, M., Eshete, A. & Assefa, S. (2018). *Reflections on expanding Ethiopia's democratic space aspirations, opportunities, choices.* Accessed on December 22, 2019, from http://fes-ethiopia.org/wp-content/uploads/2018/12/Printing-Version.pdf

Seid, M. Y., Abinew, A. A. & Biemann, C. (2020). *Analysis of the Ethiopic Twitter dataset for abusive speech in Amharic.* Accessed online on June 10, 2020, from https://arxiv.org/ftp/arxiv/papers/1912/1912.04419.pdf

Shimelis, B. (2000). *Survey of the private press in Ethiopia: 1991–1999.* FSS Monograph Series I. forum for social studies, Addis Ababa.

Skjerdal, J. & Alemayehu, M. (2020). The ethnicification of the Ethiopian media. A research report, Fojo Media Institute and International Media Support, Addis Ababa.

Tesfaye, Y. (2017). The original sin of Ethiopian federalism. *Ethnopolitics*, 16(3), 232-245. https://doi.org/10.1080/17449057.2016.1254410

Tirrell, L. (2017). Toxic speech: Toward an epidemiology of discursive harm. *Philosophical Topics*, 45(2), 139-161. https://doi.org/10.5840/philtopics201745217

UNDP. (2019). *Human development report 2019 inequalities in human development in the 21st century Congo (Democratic Republic of the Ethiopia)*, 1-10.

Wazema Radio Briefing Paper (2018). *Mapping the Ethiopian media*. http://wazemaradio.com/wp-content/uploads/2018/05/Mappingthe-Ethiopian-Media_Wazema-Briefing-2018.pdf

Workneh, T. W. (2020). Ethiopia's hate speech predicament: Seeking antidotes beyond a legislative response. *African Journalism Studies*, *40*(3), 123-139.

World Press Freedom Index (April 18, 2019). *Ethiopia rank: Winds of freedom for the media*. Accessed on March 28, 2020. https://rsf.org/en/ethiopia?nl=ok

Yohannes, G. M., Hadgu, K. & Zerihun, A. (2005). *Addressing pastoralist conflict in Ethiopia. Africa Peace Forum, Ethiopian Pastoralist Research and Development Association*. Inter Africa Group.

Chapter 10

SOCIAL MEDIA NARRATIVES AND REFLECTIONS ON HATE SPEECH IN NIGERIA

Aondover Eric Msughter
Caleb University Imota, Nigeria

1. Introduction

Hate speech is among the most significant communication issues that preoccupy the agenda of relevant governmental agencies and media analysts in contemporary Nigeria. It is an unfortunate phenomenon that manifests in the public sphere, and is fast threatening the fragile democracy which the country is struggling to consolidate. Today, most media houses in the country have condescended so low, and have given into the temptation of carrying messages that contain hate and dangerous speech, against the codes of conduct that guide journalism practice in the country. Sensationalism and the drive to sell media content are among the major reasons that entice media organizations to deviate from prescribed codes, and engage in unethical practices such as dissemination of hate speech and fake news. This has, indeed, increased the task of monitoring and regulatory agencies in the country such as the Nigerian Broadcasting Commission (NBC) and the Nigerian Press Council (NPC), among others. Perhaps, the advent of the new media in the Nigerian politico-media landscape has accelerated the occurrence and manifestation of hate speech in the country.

This is due to the fact that the new media is free-for-all terrain, a loose journalism endeavour with little or no regulation and monitoring mechanisms. This unlimited freedom that it offers makes it possible for every person, armed with the appropriate technology to disseminate uncensored content to the public (Suntai & Targema, 2017). The popularity, degree of acceptance and access to the new media is unparalleled by any other medium, and with the recent convergence of the other mainstream media of communication on the new media, it has assumed the status of the melting point of media platforms, thereby, giving a strong voice to users.

Hate speech on social media platform, therefore, is a phenomenon that has far-reaching effect in a society where a vast number of citizens are active users of its various platforms. This has increased the enormity of the question regarding social media and hate speech. Within this contextual context, this paper examines the manifestation of hate speech on social media, with the aim to interrogate the implication of the practice on democratic consolidation in Nigeria.

2. The context

Since the advent of the social media, several individuals across the globe including Nigerians have joined the bandwagon. Many Nigerians are now employing the new media technologies in their daily activities. A greater number is now using the social media platforms to not only stay in touch with their friends, loved ones and colleagues, but to publish or air their views on major developments in the society. This trend has affected every segment of the Nigerian society such that many first seek information from the social media for every new or emerging issue (news).

The new media have expanded the media landscape. Several Nigerians now have access to media as both content producers as well as content consumers. With this development, some Nigerian citizens who have penchant for the abuse of any new technology have begun to write and post whatever they want without let or restraint. Hence, social media saw a lot of people posting different and diverse political messages in the social media (Aghadiegwu & Ogbonna, 2015).

Social media have given coverage to some of these issues that showcase some obvious forms of hate speech. The propagation of political hate speech on social media could result in infringement on certain rights of the victims such as the right to be voted for, which determines political attainment. Also, it has the potential to infringe on racial, gender, and religious equality among political actors. Once an individual or group has been portrayed in a manner that promotes discrimination, such a victim may be avoided by members of the public and this could lower the victim's chances of political patronage in Nigeria.

Popoola (2019) aligns with the description that it is a globally-endorsed paradigm that the press as an important institution in the democratic process, plays a key role during elections. As the Fourth Estate of the Realm, the press provides the platform for narratives and discourses in the service of elections, political negotiations, and other features of the contestations among politicians and other civil organisations involved in election administration. However, problems associated with election reporting and media role in political contestations and machinations, particularly on the African continent, have been a recurrent clog in the wheel of politics in Africa. For instance, in Nigeria, from the 1950s up to the early 1980s, spiraling into the Fourth Republic that started in 1999 and beyond, several election problems that were rooted in perceived mishandling of the electoral process by the media, have occurred in the country. The 1965 parliamentary and 1983 general elections were faced by conflicts with accompanying widespread violence, which resulted in military interventions.

It took nearly two decades for the country to return to the path of democracy after the 1983 crisis. In subsequent elections and the attendant crisis in the country, social media have been identified to have played a somewhat negative role as disseminators and conveyors of reports, images, and analyses about the election activities; while media professionals, reporters on the political beat have also been observed to be unacquainted with certain fundamentals of election reporting (Popoola, 2019). This appears a very crucial issue, given the role that the social media are expected to play in fostering democracy. Unfortunately, the seeming unsatisfactory conduct of social media platforms in election reporting is apparent, and somewhat paucity of capacity-building resource materials on election reporting in Nigeria.

The importance of social media in a modern democracy has been a subject of discussion among participants in the political space of every democratic society, especially Nigeria as a growing democracy. Oriola (2019) is of the view that the complex and diverse nature of the 21st-century global society further accentuates audience reliance on the mass media for information about the all-important political sector of every democratic society, which dictates the space in other sectors of the society. The importance of social media could be understood firstly in their pervasiveness, which makes people learn almost everything about the world through social media, especially events and issues beyond their immediate contacts; social media's ability to disseminate information and engender citizens' political involvement in modern democracy and the ability to influence the populace on political ideologies.

This perhaps explains the position of Rasaq et al. (2017) that in Nigeria, particularly, the effects of political activities, which show hate speech have become more vivid in the successive democratic dispensation than the previous years. The deeds of politicians have only amplified the situation negatively and keeping the citizens more divided now than ever signals a great source of anxiety to Nigerians at home and in the diaspora. As noted by Pate and Oso (2017) Nigeria is a multicultural nation of diverse people, multiple identities, and colourful outlooks. It has a population of 200 million people, 400+ ethnic groups, two major religions, and dozens of political parties, 36 federating states, and additional complex platforms of diversities. The Nigerian multicultural setting is characterised by diversity, heterogeneity, and pluralism in the cultures, orientations, and attitudes of the people. In other words, it connotes diversity as a fact of life on the grounds of sex, cultural practice, ethnic origin, religious affiliation, ideological stance, political leaning, level of social development, place of habitation, and so on.

Ogbuoshi et al. (2019) maintain that today, the Nigerian polity is so heated from all political divides; there has been a resort to hate speech. There are no arguments as to how politicians have resorted to divisive comments, insinuations, and innuendoes. Not only has this hate speech pitched the North against the South but individual hatred has attained an all-time height in Nigeria. With the uncommon level of hate speech in Nigeria witnessed on social media, there is justification for venturing into an academic examination of all issues connected thereto, particularly when one views communication (hate speech) as having the capacities and infrastructures of threatening the country's collective existence. Ogbuoshi et al. (2019) establish that since there are little or no sufficient empirical studies that address the issue, such an academic gap will find a closure to explore the often less academically addressed subject of hate speech social media to the strengthening of theory, methodology, and the general knowledge based on the subject.

3. Interrogating hate speech social media and the challenges of nation building

Scholars like Ogbuoshi et al. (2019) observed that hate speech is so pervasive in Nigeria that many citizens are susceptible to it. The opposite is that people who usually complain of being insulted by other ethnic groups often use even more hateful words in describing such groups. Thus, they remark that the widening of the social distance among the different ethnicities make up the country and an exacerbation of the crisis in the country's nation-building. Several observations could be made about the interplay between ethnicity, hate speech and the crisis in the country's nation-building project. One, hate speech employs discriminatory epithets to insult and stigmatize others based on their race, ethnicity, gender, sexual orientation or other forms of group membership. This could be the reason why Ogbuoshi et al. (2019) maintain that it is any speech, gesture, conduct, writing or display which could incite people to violence or prejudicial action.

Researches by Adibe (2018) and Ogbuoshi et al. (2019) indicate that there are individuals and groups in Nigeria who openly relish the freedom to rain insults and profile others by appropriating to themselves the role of ethnic and religious champions. The implication is that hate speech is often the way to discrimination, harassment and violence as well as a precursor to serious harmful criminal acts. This perhaps explains the reason why Adibe (2018) aligns with the description that it is doubtful if there will be hate-motivated violent attacks on any group without hate speech and the hatred it purveys. Secondary, there is nothing wrong in the people celebrating pride in their ethnic and cultural identities as it is not always a manifestation of ethnicity when someone proclaims, "I am a proud *Hausa*, *Igbo* or *Yoruba*".

Therefore, most ethnic groups across the world feel that their way of life, food, dress, habits, beliefs, values and so forth are superior to those of other groups. There is nothing wrong with this. The boundary between this love for one's ethnic identity (ethnocentrism) and ethnicity (which is conflictual) could however be thin. As noted by Adibe (2018) when people's love for ethnic identify results in seeing other groups as competitors or the reasons why others are not getting what people believe they deserve to get, then there is often recourse to hate speech to vent others frustrations on the out-group. At a point, the love for one's ethnic identify has become conflictual in form and thus; crossed the boundary to ethnicity. It is important to underline that although ethnicity is rooted in the struggle for the scare societal values political positions, jobs, contracts, scholarship, etc by the various ethnic factions of the Nigeria elite, it has over time acquired an objective character such that it now exist independent of the original causative factors. Similarly, there are group of 'ethnic watcher' whose only vocation appears to be working the arithmetic of which ethnic group gets what, when and how in the proverbial sharing of the 'national cake' (Nwokoro, 2019). As such, the process of nation-building requires conscious efforts from all and sundry, irrespective of political, social or ethnic affinity.

4. Theoretical underpinning of hate speech and violence in Nigeria

Hate speech is any speech, conduct, gesture, writing, or display which could incite people to violence or execute a prejudicial action. Essentially such speeches rob others of their dignity and sense of order (Mrabure, 2015). United Nations Committee on the Elimination of Racial Discrimination (2016, p. 35) in its interpretation of the law, noted that hate speech offences include:

> (a) all dissemination of ideas based on racial or ethnic superiority or hatred by whatever means; (b) incitement to hatred, contempt, or discrimination against members of a group on the ground of their race, colour, descent, national or ethnic origin; (c) threats or incitement to violence against persons or group on the grounds in (b) above and (d) expression of insults, ridicule or slander of persons or groups or justification of hatred, contempt or discrimination on the in (b) above when it amounts to incitement to hatred or discrimination and (e) participation in organisation and activities which will promote and incite radical discrimination and violence.

As a term, hate speech can be said to have put down its roots in Nigeria since the 2015 election campaign (Ezeibe, 2015). To be specific, the term seemed to have gained public attention in Nigeria after a documentary aired on African Independent Television (AIT) about an All Progressive Congress (APC) Presidential election. Due to perceived hostility and partisanship from the transmitting station, the documentary was later described as a 'hate broadcast.' Hate speech seems mostly associated with power elites. Some utterances from political leaders, wealthy persons, religious heads and others who can be regarded as role mentors have been regarded as hate speeches (Okunna, 2018). Some examples of such instances made by leaders in Nigeria and reported in the press are as follows:

> 'Buhari would likely die in office if elected, recall that Murtala Muhammed, Sani Abacha and Umaru Yar'Adua, all former heads of state from the Northwest like, Buhari, had died in office' – The Governor of Ekiti State, Peter Ayodele Fayose, January 19, 2015, ThisDay and other national dailies.

> 'Wetin him dey find again? Him dey drag with him pikin mate, old man wey no get brain, him brain don die pata pata – what else is he (Buhari) after, contesting with people young enough to be his children. The old man who lacks gumption; he is completely brain dead.' – Former First Lady, Patience Jonathan, at a PDP political party rally in Kogi State, Reported by The Express News, 4th March, 2014.

> 'God willing, by 2015 something will happen. They either conduct a free and fair election or they go a very disgraceful way. If what happened in 2011 should again happen in 2015, by the grace of God, the dog and the baboon would all be soaked in blood.' – Presidential Candidate of Congress for Progressive Change, General Muhammadu Buhari, Reported on social media.

The emphasis highlighted in italic merely indicates the depth of negative passion resentment for an individual which extended to his group and the levity with which death as a mode of dismissal is taken. Yet such hate speech needs media for widespread circulation to gain prominence. Without social media, hate speech could fail to come alive (Okunna, 2018).

Based on some theoretical postulations, Lazarsfeld and Katz's Two-Step Flow Theory was first introduced by Lazarsfeld et al., in 1944, to study the process of decision-making during a Presidential election campaign. The paper found empirical support for the direct influence of social media messages on voting intentions. Armed with this data, Katz and Lazarsfeld developed the two-step flow theory of mass communication. This theory asserts that information from the media moves in two distinct stages. First, individuals (opinion leaders) who pay close attention to the mass media and its messages receive the information. Opinion leaders pass on their interpretations in addition to the actual media content. The term 'personal influence' was coined to refer to the process of intervening between the media's direct message and the audience's ultimate reaction to that message. Opinion leaders are quite influential in getting people to change their attitudes and behaviours and are quite similar to those they influence. The two-step flow theory has improved the understanding of how the mass media influence decision-making.

The theory refined the ability to predict the influence of social media messages on audience behaviour, and it helped explain why certain media campaigns may have failed to alter audience attitudes or behaviour. The two-step flow theory gave way to the multi-step flow theory of mass communication. Although the empirical methods behind the two-step flow of communication were not perfect, the theory did provide a very believable explanation for information flow. The opinion leaders do not replace media, but rather guide discussions of media, which at times lead to issues of hate speeches. Lazarsfeld et al., in Hassan (2020), discovered that most voters got their information about the candidates from other people who read about the campaign in the newspapers, not directly from the media. They concluded that word-of-mouth transmission of information plays an important role in the communication process and that mass media have only a limited influence on most individuals.

Castells' Theory of Network Society examines the concept of the network to a high level of abstraction, utilizing it as a concept that depicts macro-level tendencies associated with the social organization in informational capitalism. He expressed the role of networks in social theory as follows dominant functions and processes in the information age are increasingly organized around networks. Networks constitute the hate speech morphology in societies, and the diffusion of networking logic substantially modifies the operation and outcomes in processes of production, experience, power, and culture. Understanding the societal context of such networks entails returning to the political economy of the social transformation of capitalist society. An analytical concept network is abstract and thus unable to frame the interpretation of real-life networks, whereas theoretical concept network is an excellent crystallization of the social morphology of informational capitalism.

As an upshot of the latter, the concept of network society has a certain intellectual appeal, even if it looks almost as if the formal description of the concept of the network was needed only to legitimate its use as a metaphor. Concerning the hardcore of the metaphor, the study comes to the true message of Castellsian political economy (where politicians metaphorically used negative words to refer to other opposition), and the network in its paradigmatic form is about the nodes and connections of powerful financial and economic institutions, which allow the flows of values in pursuit of the newspapers' accumulation of capital. This implies that 'network' in Castells' social theory is not an analytical concept but rather a powerful metaphor that served to capture the new social morphology of the capitalist system. In this context, the morphological manifestation of hate speech on social media discourse of information society gain its momentum; it went out of intellectual fashion as well as political agenda and gave its place to the visions of the creative and or smart society.

Although the critics, who looked at the theories of the information society suspiciously as ideological constructs, created for political decisions, rather than instruments for understanding the social reality. Therefore, Castells believes that McLuhan's dictum "the medium is the message" could be adequately applied in the way hate speeches flourish on social media platforms. In this perspective, there is a network (social media users), which often creates a powerful metaphor that aptly portrayed hate speech as a social morphology of information capitalism (Doda, 2015).

Durkheim's Social Fact and Weber's Social Action or Relations Theory emphasizes the importance of social collectivity and its determination over individual consciousness, pointing out concepts like *sui generis* of social facts, function, causality, generality, etc. Weber's on the other hand based his argument on concepts such as meaning, social action, interpretation, methodological individualism, etc. The study depicts both theorists to understand the social order or social reality of hate speech at the theoretical level and the approach to this social reality focused attention on individualistic autonomy in terms of ideas and desires vis-à-vis social regularity.

Weber approached the problem of social regulation through the question of how this regularity becomes possible out of the chaos of individualistic ambiguity. In this manner, he searched for the underlying rules and principles in this order. According to Weber, social continuity or social order is constructed at the individualistic consciousness level through how social actors assign meaning to their actions. Weber in Doda (2015) proposes that the reason behind regular actions is the meaning that individuals attribute to their actions. Like the action towards hate speech, the acting individual attaches a subjective meaning to his behaviour, which can be overt or covert, omission or acquiescence that is concerned with 'meaning-attributed-action' within society.

For Weber, people give meaning not only to their behaviour but also to behaviour of other people in their reciprocal relationships, because the action of each takes account of that of others. Weber understands social regularity as the harmony of individualistic social actions and meanings individuals attribute to the actions of other people. Therefore, individuals' attribution of meaning to action and social relationships gives social life its regularity; otherwise, social action would be impossible. In Weberian analysis, these regularities in social and individualistic levels merge in social action. Unlike Weber, Durkheim, when considering social order, essentially evaluates it as a whole, not as a set of individualistic actions or unique particularities.

Durkheim proposes that to understand how society thinks of itself and of its environment one must consider the nature of the society and not that of the individuals. According to Durkheim, social continuity arises from the domination of social regulations over the ambiguity of the individualistic infinite and indeterminate psychological needs and desires.

As "hate speech is a reality *sui generis*" in the case of the Durkheimian approach. For Durkheim, because individualistic needs are infinite, society imposes limits on human desires. In this manner, Durkheim's idea of social action refers to "*sui generis* of social facts," namely, the determination of "external conditions," which implies not a probability but a certainty. On the other hand, in the Weberian sense, social action has to do with 'not a certainty but probability'. For example, when Weber explains types of action orientation, he defines 'usage' as an orientation toward social action that occurs regularly, it will be called 'usage' (*Brauch*) insofar as the probability of its existence within a group is based on nothing but actual practice. As one can see, for Weber, ideas (hate speech) can assume a role in social change. On the other hand, Durkheim demonstrates that individualistic ideas and thoughts (hate speech) can never affect the path of the existing social order.

The Functional Theory of Campaign Discourse, according to Doda (2015) explains the functional theory of campaign discourse, which renders a helpful scheme to classify and synthesize political advertising and how they appear on social media. He adds that elections are intrinsically competitive, political actors deploy campaign messages, which include advertising to present a more preferable image of them. They use political ads to acclaim themselves, positive statements about their credentials as the better candidate; attack an opponent's credentials, or defend with reputations against an opponents' attack through media platforms. Candidates use two functions for themes of policy: policy themes can discuss actions or ideas related to governmental action related to past deeds, plans, and general goals. Character: character themes can discuss the candidate's perceived qualities related to personal qualities, leadership abilities, and values or principles. It would be interesting to see the extent of the use of self-acclaim, attacks, ethnic insult, blotch campaigns, religious divisiveness, or issue-based topics relating to economic, social, cultural, and political policies.

For instance, in the 2015 and 2019 elections, there were cases of sponsorship of hate advertorials by the then Ekiti State governor, Ayodele Fayose who, on January 19, 2015, ran adverts on the front pages of national dailies such as The *Daily Sun*, *The Guardian* and *The Punch* titled: "Nigeria Be Warned". In the advert, satirical reference was made to Buhari, the presidential candidate of the APC, that given his age and speculated illness and frail nature, he might die in office should he win, according to *Sahara Reporters* of January 19, 2015. As corroborated by Odera (2015) elections are intrinsically competitive, political actors deploy campaign messages, which include advertising to present a more preferable image that comes in form of hate advertorial, which serves the research goal in understanding how politicians synthesize political advertising using hate speech.

Critical Discourse Analysis Theory provides a reality that can be represented either truthfully or falsely in language. The theory assumes that it is possible to represent reality in an unmediated, neutral form; critique is then based on whether the ideal is attained or not. Neutral representations are opposed to ideological representations, which are deemed to 'distort reality. Ideology is, accordingly, conceptualised in negative terms, as the opposite of 'truth.' Critical Discourse Analysis Theory describes and analyze how the structure and content of the text encode ideas and the relation among the idea itself that is present in the text, systematically. Here, it connotes how hate speech and language, dialects, and acceptable statements are used in a particular medium across different audiences. The theory looks at discourse among people who share the same speech conventions. It also refers to the linguistics of language use as a way of understanding interactions in a social context, specifically, the analysis of occurring connected speech or written discourse like APC presidential candidate is a fundamentalist – Clarke.

Furthermore, Fairclough in Omidiora *et al.* (2019) argued that social practice has various orientations economic, political, cultural, ideological, and discourse may be implicated in all of these without any of them being reducible to discourse. The author further stated that in this line, discursive practice is constitutive in both conventional and creative ways: it contributes to reproducing society (social identities, social relationships, systems of knowledge and belief) as it is, yet also contributes to transforming society. In this context, the theory is apt in this discussion as it provides a reality that can be represented either truthfully or falsely in language.

Critical Race Theory (CRT) also provides a compelling structure by which social media concepts and hate speech can be analyzed and understood. According to Odera (2015) CRT indicates that media use phrases sponsored by politicians that refer to other opposition groups from descriptions that are not merely rhetorical but pedestals on which hate speech flourishes. Theoretically, Critical Race Theory underscores that violent political rhetoric is capable of producing the same psychological dynamics as violent entertainment.

Through Critical Race Theory, framing words on the assumption that a subtle change in the wording of the description of a situation might affect how the audience interprets the situation. This portends that media coverage can help influence how people think about candidates, events, and other issues. As a result, framing refers to the impact of news coverage on the weight assigned to specific issues in making political judgments. This means that social media may draw more attention to some aspects of political life like the elections and the aftermath at the expense of others. The interpretation of Critical Race Theory is that in choosing and displaying news, editors, newsroom staff and broadcasters play an important part in shaping political reality. Consequently, readers learn not only about given issues but how much importance to attach to that issue from the amount of information in a news story and its position.

In response to that Critical Race Theory is used to support a legal-structural response to hate speech on social media. It aims to transform the relationship between race, law, and power. CRT recognizes that the vested interests of the economic-political elite shape racial and ethnic stratification. Thus, the nexus between the theory and the paper is that the theory provides a compelling structure by which media concepts and hate speech can be analyzed and understood. This indicates that media use phrases sponsored by politicians that refer to other opposition groups from descriptions that are not merely rhetorical but pedestals on which hate speech emanate. Arguably, some literature challenged the dominant ideology of Critical Race Theory based on race and racism, the social construction of race storytelling and counter storytelling as well as the notion of white supremacy. Despite this debate among scholars, this paper considers the theory apt since social media used phrases sponsored by politicians.

5. Understanding ethical and legal position of hate speech

There are differences in socio-cultural contexts of various societies of the world and that hate speech is both a legal and sociological construct perhaps account for no generally accepted definition of the construct. However, attempts have been made to conceptualise hate speech, though most of the scholarly definitions focus on racial and religious hate. Just like how Oriola (2019) describes hated in the context of human interaction as extreme dislike of persons or groups, on the ground of their racial, ethnic, religious or gender orientation or affiliation. Such extreme dislike may be overt or covert. When it is expressed in speech form or any other non-verbal mode of message display, it becomes covert to the extent of using communication to express such kind of dislike. It is the expression of such extreme dislike which has discriminatory or denigrating consequences that constitutes hate speech.

From the legal perspective, the US legal (2016) describes hate speech as a communication that carries no other meaning than the expression of hatred or incitement to hated against some group of persons defined in terms of race, ethnicity, national origin, gender, religion or sexual orientation, especially in circumstances in which the communication provokes violence. This description rightly points generally to communication of information, which is characterized by not only articulated vocal sound speech but also to other communication written, oral and display that are intended to carry meaning to members of the public about certain groups. It also considers the potential provocation of violence as the result of such communication. However, the definition does not consider expression of hatred to individuals and it limits the potential consequences of hate speech to violence. It should be noted that hate speech affects individuals as it affects groups in the society, especially in the competitive field of politics that is characterised by struggle for supremacy and power. An individual political candidate could be targeted in an expression of hate just as a political group or party.

A bill on 'Prevention of Hate Crime and Hate Speech in South Africa, the summary of which was published in Government Gazette No. 41534 of 29 March 2018, describes an offence of hate speech. Section 1(a) of the Bill describes an offence of hate speech as an intentional communication, publication, propagation or advocacy of any message to:

> One or more persons in a manner that could reasonably by construed to demonstrate a clear intention to (i) be harmful or to incite harm or (ii) promote or propagate hatred, based on one or more of the following grounds: age, albinism, birth, colour, culture, disability, ethnic or social origin, gender or migrant or refugee status, language, nationality, includes intersex or sexual orientation (pp. 4-5).

The Bill also provides that intentional distribution or display of materials capable of being communicated or electronic communication of messages known to constitute hate speech, as provided in the paragraph above through electronic communication system to which members of the public have access and which is directed at a specific individuals who can be victims of such messages, is guilty of an offence of hate speech. However, the Bill provides exceptions to the ingredients that constitute hate speech offence.

Article 19 of the Universal Declaration of Human Right (UDHR) provides and protects the right to freedom of expression and this is given a legal force through Article 19 of the International Covenant on Civil and Political Rights (ICCPR). However, the United Nations (UN) Human Rights Committee (HRC) observes that freedom of expression is not absolute and sets limits to expressions, which may be considered offensive and discriminatory (Oriola, 2019). Therefore, under Article 19(3) of the ICCPR, a state may limit the right to free speech provide that the limitation is provided by law, in pursuance of a legitimate aim and necessary in a democratic society. Hate speech is one of such limitations, which satisfy the three-test condition for restricting free speech. Article 20(2) of the ICCPR describes hate speech as any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence and provides for the prohibition of such expression.

Ogbuoshi et al. (2019) compared hate speech with free speech doctrine of J.S Mill, which is enshrined in the constitutions of nations. They noted that hate speech is not free speech. Hylton conceived hate speech as negative while free speech is a landmark achievement of democracy. Hence, most developed democracies added a clause on freedom of speech against the use of hate speech. For example, Article 10(2) of the European Convention on Human Rights (ECHR) provides that "the exercise of freedom of expression may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law, the interest of national security for the protection of the reputation or right of others." Impressively, most doctrines that established freedom of speech and expression in Nigeria added a clause to guard against hate speech, promote human dignity, societal cohesion and peace. For instance, section 39(1) of the 1999 Constitution as amended in 2011 provides that "every person shall be entitled to freedom of expression" Similarly, section 45 provides that nothing in section 39 shall invalidate any law that is reasonably justifiable in a democratic society in the interest of public order, public morality and to protect the rights and freedom of other persons.

Sections 95 and 96 of the 2010 Electoral Act prohibited the use of any language in campaigns that will hurt tribal, religious or sectional feelings. Law of libel and slanders also protect the citizens against hateful utterances. Other legal frameworks that abhor the use of derogatory language in Nigeria are the Political Party Code of Conduct (2013) and the Abuja Accord (2015). Despite these legal frameworks, there has been notable growth in hate speech on social media. As noted by Ogbuoshi et al. (2019), there are, however, existing laws that cater for abuse of freedom or harassment of individuals and groups as pointed out by civil society and mass media groups. The law setting up the National Broadcasting Commission (NBC), Advertising Practitioners Council of Nigeria (APCON), Nigerian Press Council (NPC) and the Nigerian Film and Video Censors Board (NFVCB) filtered out offensive materials and pornography, among others.

Despite the Nigeria Electoral Act of 2010, which spells out detailed provisions specifically barring politically inspired hateful speech, still cases of offensive images of major aspirants, to create a vivid picture of a bad person flourish and have been described by Nigerians as 'one step too far'. Specifically, Section 95 of the Act provides that no political campaign or slogan shall be tainted with abusive language directly or indirectly likely to injure religious, ethnic, tribal or sectional feelings. Similarly, abusive, uncontrolled, slanderous or base language or insinuations or innuendoes designed or likely to provoke violent reaction or emotions shall not be employed or used in political campaigns. Section 102 of the Act further provides: "Any candidate, person or association who engages in campaigning or broadcasting based on religious, tribal or sectional reason to promote or oppose a particular political party or the election of a particular candidate, is guilty of an offence under this Act and on conviction shall be liable to a maximum fine of ₦1,000,000 or imprisonment for twelve months or both." Similarly, paragraph 10(c) of the Guidelines for Political Rallies issued by Independent National Electoral Commission (INEC) of Nigeria also prohibits the use of hate speech and discriminatory rhetoric.

6. Conclusion

This paper examines social media narratives and reflections on hate speech in Nigeria. Based on the literature, the paper argues that while still countering hate speeches in the traditional media, the emergence of new media has broadened the battlefield in combating the hate speech saga. Social media offers an ideal platform to adapt and spread hate speech and foul language easily because of its decentralised, anonymous and interactive structure. The prevalence of hate speech on social media bordering on political and national issues, and even social interaction in Nigeria, especially on Facebook, Twitter, YouTube and LinkedIn is becoming worrisome. This is because apart from undermining the ethics of journalism profession, it is contributing in bringing disaffection among tribes, political class, and religion or even among friends in the society. The Nigerian public is inundated with negative social media usage such as character assassination and negative political campaigns at the expense of dissemination of issues that help them make informed choices.

References

Adibe, J. (2018). *Ethnicity, hate speech and nation-building*. http://www.gamji.com/adibe/adibe19.htm, 2019

Aghadiegwu, U. C. & Ogbonna, U. A. (2015). The rise of hate and peace journalism in the Nigerian democratization process: the place of the new media. *Communication Panorama: African and Global Perspectives,* 1(1), 1-16.

Doda, M. A. (2015). *Introduction to Sociology*. Lecture note. In collaboration with the Ethiopia Public Health Training Initiative, The Carter Center, the Ethiopia Ministry of Health, and the Ethiopia Ministry of Education.

Ezeibe, C. C. (2015). *Hate speech and electoral violence in Nigeria*. Online material September 5th, 2015.

Hassan, B. S. & Owolabi, T. O. (2020). Nigerian press coverage of hate speeches and negative campaigns in the 2015 presidential elections. *Journal of Communication and Media Research*, 10(1), 52-62.

Mrabure, K. O. (2015). *Counteracting hate speech and the right to freedom of expression in selected jurisdictions*. https://www.ajol.info/index.php/naujilj/issue/view/13988

Nwokoro, C. I. (eds). (2019). Perspectives on the use of radio in countering hate speech and violence in Benue and Delta States. *Fake news and hate speech: narratives of political instability*. (6thed.). Canada University Press, Concord Ontario, Canada.

Odera, E. I. (2015). *Radio and hate speech: A comparative study of Kenya 2007 and the 1994 Rwanda genocide*. (Master's Thesis, The University of Nairobi). http://erepository.uonbi.ac.ke/bitstream/hand

Ogbuoshi, L. I., Oyeleke, A. S. & Folorunsho, O.M. (eds). (2019). Opinion leaders' perspectives on hate speech and fake news reporting and Nigeria's political stability. *Fake news and hate speech: narratives of political instability*. (6thed.). Canada University Press, Concord Ontario, Canada.

Okunna, S. (2018). *Assessment of the use of different forms of tobacco products among Nigerian adults: Implications for tobacco control policy*. https://doi.org/10.18332/tpc/87126

Omidiora, O., Ajiboye, E. & Abioye, T. (eds). (2019). *Beyond fun, media entertainment, politics and development in Nigeria. The marriage of the popular and the political: A critical analysis of Nigerian hip hop music in electoral campaign discourse*. Malthouse Press Limited.

Oriola, O. M. (eds). (2019). Mainstream media reporting of hate speech and press freedom in Nigeria politics. *Fake news and hate speech: narratives of political instability*. (6thed.). Canada University Press, Concord Ontario, Canada.

Pate, U. A. & Oso, L. (eds). (2017). *Multiculturalism, diversity, reporting conflict in Nigeria*. (1sted.). Ibadan: Evan Brothers Nigeria Publishing Limited.

Popoola, T. (2019). *Parrot journalism. A professional guide in investigative journalism*. Lagos. Diamond Publications Limited.

Rasaq, A., Udende, P., Ibrahim, A. & Oba, L. (2017). Media, politics, and hate speech: a critical discourse analysis. *e-Academia Journal*, 6(1), 240-252.

Sahara Reporters (2015, January 19). *Governor Fayose Places Death-wish Advert on Buhari in National Newspapers*, p. 17.

Suntai, D. I. & Targema, T.S. (2017). New media and democracy in Nigeria: an appraisal of the opportunities and threats in the terrain. *Brazilian Journal of African Studies, 2*(4), 198-209.

United Nations Committee on the Elimination of Racial Discrimination (2016). *General recommendation on combating racist hate speech*, CERD/C/GC/35.

Chapter 11

HATE SPEECH AMONG SECURITY FORCES IN PORTUGAL

Tiago Lapa
Iscte - University Institute of Lisbon, Portugal

Branco Di Fátima
LabCom - University of Beira Interior, Portugal

1. Introduction

Several researchers have been trying to conceptualize and delineate hate speech related phenomena, according to its variety and extent, but have also considered whether new media and the Internet have altered the character of hate speech (Brown, 2018). In the discussion whether the medium is the message, there have been concerns about the specificities of online hate speech. As stated by Brown, part of the interest in this enterprise is the identification of the characteristic difficulties in tackling hate speech online, but also to understand the attractiveness of the digital realm for gatherings of hate speakers.

In a broader sense, some definitions consider the motivations, target audience, and language used by hate speakers to attack others (Gagliardone, 2019). However, Weber (2014) argues that there is no universally accepted definition of hate speech. This state of affairs makes its meaning fluid and diverse, varying across countries, governing bodies, and disciplinary lenses.

We also frame the analyses of online discriminatory discourses with ongoing debates concerning the balancing between the democratic protection of freedom of expression and effective ways of tackling hate speech. Furthermore, we present considerations about the policies to combat hate speech at the national and European level, and its contradictions. Despite some regulatory movements, European governments have typically delegated regulation to digital platforms and Internet service providers. And it can be contended that measures of damage control against hate speech cannot be separated from regulatory policies towards digital platforms as a whole.

Finally, the present chapter also discusses the use of closed Facebook groups by Portuguese security forces officers to propagate hate speech, unraveled by a consortium of journalists, and its possible implications. Despite Facebook's public commitment to tackle hate speech, it was ineffective in enforcing its policies against such discursive forms, thus giving room to the traditional press to serve as watchdogs of the hate speech propagated among members of the security forces.

2. European Union against hate speech, but what is it?

In an attempt to somewhat harmonize national legislation within the European Union (EU), the Committee of Ministers of the Council of Europe, with Recommendation 97(20) on hate speech, defined and conceptualized it, clarifying that it should be understood as one that includes all forms of expression that disseminates, encourages, promotes or justifies racial hatred, xenophobia, sexism, antisemitism or other forms of hatred that are based on intolerance. The Recommendation also equates hate speech with "intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin" (Weber, 2014, p. 3).

By the end of 2021, the European Commission approved a Communication which proposes the extension of the current list of hate crimes and hate speech. But, although the majority of Member States of the EU have passed laws prohibiting expressions corresponding to hateful discourse, there are national variations concerning the identification and extension of hate crimes and hate speech (PRISM Project, 2015). According to the PRISM Project report (2015, p. 49), European countries tend to specify "certain bias categories in their legislation, which help to identify segments of society that may be particularly targeted in acts of discrimination, hate crime and hate speech". But, while in the Netherlands, hate crimes are solely defined as "offenses with a discriminatory background", in Lithuania, for instance, there is an all-inclusive approach to face discrimination with its Public Security Development Programme for 2015-2025.

International law prohibits incitement to discrimination, hostility and violence, rather than explicitly prohibiting hate speech (UN, 2019) and in many contexts outside the West, the meaning of hate speech is still contested. Those who propagate hate speech can take advantage of this lack of definition to always claim that "this is not what this is about".

Thus, the current state of affairs stresses Marwick and Miller's (2014) remark that defining hate speech is a challenge. However, it can be described as speech that aims to spread hatred towards a specific "minority", usually a disadvantaged one. Therefore, hate speech includes comments that are deliberately directed against a specific person or group, and encompasses a variety of situations: 1) "incitement to racial hatred", i.e., targeted hatred against people or a group because of their racial belonging; 2) "incitation to hatred on religious grounds", which can be equated with the incitement to hatred which is based on the distinction between believers and non-believers; and 3) "incitement to hatred based on intolerance", which is characterized by the manifestation of violent nationalism and ethnocentrism (Weber, 2014, p. 4).

In general, the UN (2019, p. 2) defines hate speech as "any kind of communication in speech, writing or behavior, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are". The basis for these attacks is "religion, ethnicity, nationality, race, color, descent, gender or another identity factor". This behavior can be consolidated and generate intolerance, which in certain contexts can cause division and humiliation. As the UN claims, hate speech has an influence on different areas: human rights protection; prevention of atrocity crime; preventing and countering terrorism and the underlying spread of violent extremism and counter-terrorism; preventing and addressing gender-based violence; enhancing protection of civilians; refugee protection; the fight against all forms of racism and discrimination; protection of minorities; sustaining peace; and engaging women, children and youth (UN, 2019, p. 2).

On the other hand, Silva et al. (2011) argue that hate speech is based on two fundamental principles: discrimination and externality. It is characteristic of a segregationist and relational manifestation, supported by the establishment of symbolic power and violence (Bourdieu, 1989) and a hierarchical dichotomy between the "superior" emitter (that is, the aggressor) and the "inferior" reached target (that is, the victim). Thus, hate speech is revealed by places of speech and, in a relational perspective, by others besides the speaker. The concept of place of speech used by activists of feminist, black or LGBT movements is also useful here. It confronts the knowledge produced by the hegemonic epistemologies. Therefore, places of speech do not merely reflect individuals' speech acts. They stem from unevenly positioned worldviews. This unevenly confrontation between worldviews appears in countless debates in academia and society, and is often present in discussions on social media where discourses are guided and fought (Pereira, 2018).

Hate speech can be seen as the enactment of symbolic power and violence, where, stemming from a place of speech, discourse is used to attack or socially disqualify others, often inciting violence and hatred towards a perceived group based on their physical appearance, religion, ethnicity, sexual orientation, gender identity, or other characteristics (Fortuna & Nunes, 2018). It can be used with based on various linguistic forms, and can be made subtly or by using humor, or even explicitly, based on violence (Lamerichs et al., 2018). The reinforcement of stereotypes and essentialist notions can also serve as a symbolic tool for the aggressor to validate their discriminatory and negative attitudes against specific social groups.

In addition, there have been discussions about the novelty of the characteristics and challenges of online hate speech and crimes online. Digital platforms might allow anonymity, invisibility, the instantaneous spread of hateful content and the clustering of hate speakers with like-minded individuals (Brown, 2018) that might be instilled with a sense of empowerment and exemption. Miranda et al. (2022) state that hate speech is certainly a toxic behavior exacerbated by Internet culture and the digital underworlds. Gitari et al. (2015) define hate speech on social media as language that is characterized by its hurtful or potentially harmful lexicon that can spread with unprecedented speed and reachability. It is motivated by aggressive prejudice and is directed at individuals or groups based on their inherent or perceived characteristics. For the authors, this discourse has the clear goal of being harmful, inciting hate, or propitiating hatred. This type of hate speech can be done in different digital spaces, such as news ads, comments box, online forums, and social media.

Warner and Hirschberg (2012) explain that extremists often alter their online discourse through purposely misspellings or word choices, such as using “Zionists instead of Jews”. Klein (2012) refers to this practice as a “theory of information laundering”, a set of techniques used by hate groups to legitimize their ideas through a “borrowed network of associations”. This “network” helps in spreading hatred not just through words, symbols, and images but also through hyperlinks, downloads, so-called news, threats, conspiracy theories, and even pop culture.

Concerning the extent of the problem, Kaakinen et al. (2018) point out that, while hate content production is rare overall, it gains high visibility online. The authors also indicate that the dynamics of hate speech are related to social capital in two key ways that operate in different directions. On the one hand, high social capital in offline social networks was associated with a lower probability of production of hate content. On the other hand, individuals with high social capital in online social networks were more likely to be producers of such content. This shows that, despite descriptions of social capital as a positive resource drawn from social networks and communities (Putnam, 1993; Portes, 1998), it can take a darker side online when used by certain individuals to propagate hateful content.

3. Freedom of expression and hate speech

There has been a public debate if the proper protection of freedom of expression demands the legal safeguard of so-called hate speech or not, and "whether freedom of speech should be granted priority over other political values" (Howard, 2019, p. 94). In Portugal, Article 37 of the Portuguese Constitution guarantees the freedom of expression for all citizens, allowing them to freely express their thoughts in words, images, or any other means. It also gives them the right to access and share information without apparent restrictions. Those who flout this right shall be subject to the general principles of criminal law or the unlawful of mere social ordination, "and their assessment respectively of the jurisdiction of the judicial courts or an independent administrative entity, in accordance with the law. All persons, natural or legal, are guaranteed, on a level and equal basis, the right of reply and rectification, as well as the right to compensation for the damage suffered." (Portuguese Constitution, Article 37).

Likewise, the European Union also recognizes the right to freedom of expression and information in Article 11, which must be followed by all member countries of the community. European citizens have the fundamental right to freedom of expression, including the freedom to hold opinions and exchange information without interference from public authorities and without geographical borders in Europe (European Union Agency for Fundamental Rights, 2022).

However, Gascón (2012) states that the Internet presents new challenges to tackle the spread of hate speech, a concern usually associated with freedom of expression – a privilege frequently employed by proponents of hate to justify acts of violence, particularly against minority groups. Thus, proper definitions are required, since there is no agreed definition at international level of hate speech, along with the delineation of boundaries. The world's democracies promptly define limitations to freedom of expression. In developed democracies like the United Kingdom, Denmark, France, Germany, Sweden, and so on, we can find legislation that criminalizes offenses to incite racial or religious hatred (Waldron, 2012; Brown, 2018; Pohjonen, 2018).

In this sense, the European Commission against Racism and Intolerance (ECRI, 2015), in its general policy, nº. 15, states that freedom of expression and opinion shouldn't be regarded as an unrestricted right. It should not be exercised in a manner incompatible with other rights, as they are important for a democratic and pluralistic society (ECRI, recommendation nº. 15, 2015, p. 5). This means that freedom of expression is at odds with hate speech, since hate speech discriminates against others, as well as denies recognizing their rights equally (Gagliardone, 2019).

As a matter of principle, hate speech should be fought because it is important to help prevent "armed conflict, atrocity crimes and terrorism, end violence against women and other serious violations of human rights, and promote peaceful, inclusive and just societies" (UN, 2019, p. 1). Combating hate speech does not equate to limiting or restricting freedom of expression – despite being a persistent issue, this remains a relevant question (Mihajlova, Bacovska & Shekerdjıev, 2013). Instead, it aims to prevent hate speech from escalating into more dangerous forms, such as incitement to discrimination, hostility, and violence, which are prohibited by international law (UN, 2019).

4. Policies to combat hate speech

The European Union defines illegal hate speech in European law as public incitement to violence or hatred which is based on certain perceived particularities, such as race, religion, ancestry, and national or ethnic origin. It is a discourse that is at odds with other fundamental rights and values, besides free speech, in which democratic societies are supported. It is argued that it harms not only the victims of this discourse, but also society at large. In addition, hate speech, instead of being a proper product of free speech, is seen by European institutions as an obstacle to diversity and the pluralism of ideas, due to its tendency to present a hierarchical reasoning and monopolizing worldviews and its negative effects on public debate and democracy (External Action Service of the European Union, 2022).

The European Commission against Racism and Intolerance (ECRI, 2015) admits that the duty provided for in international law to criminalize some forms of hate speech, applied to all, was designed to protect individuals from vulnerable groups. In such cases, hate speech should be monitored and penalized, especially on social media. Indeed, the European Union has sought arrangements with social media platforms to tackle the dissemination of hate speech. In 2016, for example, the European Commission signed with Facebook, Microsoft, Twitter and YouTube a Code of Conduct for combating illegal hate speech online. Two years later, in 2018, Instagram, Snapchat and Dailymotion signed this non-binding code of conduct. Respectively, Jeuxvideo (2019), TikTok (2020) and LinkedIn (2021) also adhered to the code. However, we can question the effectiveness of such arrangements in combating hate speech.

As Mansell (2011, p. 6) puts it, supporters of an open Internet, not subject to regulation, have succeeded in convincing policymakers that direct interference under conventional telecommunication or broadcasting regulatory mechanisms is unnecessary and would subdue inventive online activity (Benkler, 2000). Thus, the general regulatory rule on the Internet has been self-regulation and minimum intervention (Ben-David & Matamoros-Fernández, 2016). Governments usually delegate the control of the content to technological corporations and Internet service providers. But the procedures of providers are more flexible than regulations targeted at traditional media imposed by different countries. The authors explain that Facebook uses specific blocking techniques depending on the law of each country. For example, Nazi material is prohibited in Germany but allowed in the United

States. Thus, some hate content may have restrictions on its social media circulation in one country but continue to be freely shared in another.

The United Nations (UN) Strategy and Plan on Hate Speech, established in May 2019, acknowledges the growing trend of xenophobia, racism, and intolerance globally, including the increase of anti-Semitism, hatred towards Muslims, and persecution of Christians. The UN explains that social media and other forms of communication have been used as vehicles of intolerance, and neo-Nazi and white supremacist movements are increasing. Thus, public discourse is being used for political purposes with speech acts that recreate and dehumanize minorities, migrants, refugees, women, or anyone else considered "other". The UN (2019) also highlights that these events are not isolated, since hatred supported by the use of social media is turning into a dominant discourse - both in democracies considered liberal and in authoritarian systems, weakening the values of humanism.

In this scenario, member states of the European Union have adopted measures to combat hate speech online. In June 2017, the German parliament passed laws against social media to combat the spread of hate bear discard fake news, disseminated by users of these pages. This passed law is known as the Facebook Act and ensures that social media such as Facebook, Twitter and YouTube must delete content that explicitly is against German law within 24 hours of a report, and within seven days for material deemed offensive. Social media platforms may have to pay fines of up to €50 million if they do not comply with these rules (The Verge, 2017).

In March 2018, then-French Prime Minister Edouard Philippe announced his plan to intensify efforts to combat the daily proliferation of hate on the Internet. One of the initiatives was to enhance the accountability of Internet service providers through the implementation of new European regulations. Just like Germany, France plans to punish social media platforms that do not comply with this new law by €50 million. The government also wants to allow to use online aliases to identify the perpetrators of racist comments and publications (UOL, 2018).

At the Portuguese level, in July 2020 the Minister of State and the Presidency, Mariana Vieira da Silva, announced that the Government will monitor hate speech online and this should result in a monthly barometer of monitoring and identification of pages, with this type of speech (Público, 2020). However, although it has already elapsed two years after its announcement, this project has not yet advanced.

And, even if Nick Clegg (2020), Facebook's vice president of Global Affairs and Communication, said that "Facebook does not profit from hate", in practice, its algorithm analyzes users' links and clicks to suggest content that aligns with their interests (Pariser, 2011). In other words, this algorithm can reinforce hate speakers' discriminatory attitudes by recommending similar content from the platform. Gerlitz and Helmond (2013) also point out that digital platforms monetize from the interactions of their users, through the marketing of data, and organize the communities of users and knowledge, helping in the creation of environments where users behave in a certain way. Likewise, Facebook and its hate speech policies are driven by the motivation to monetize interactions.

Therefore, it can be argued that we cannot separate the discussion of the regulation of online hate speech from the debate of the regulation of digital platforms and of the Internet as a whole. As Silverstone (2007, p. 26) penned, "mediated connection and interconnection define the dominant infrastructure for the conduct of social, political and economic life across the globe". As the dynamics concerning online hate speech show, social media is no more a neutral configuration of technologies than previous media. As Mansell (2011, p. 1) puts it, if there are forces that are changing the Internet in ways that are not equitable or desirable from a progressive democratic perspective, then there should be ways for opposing them in the interests of positive engaged citizenry.

Ultimately, the regulation of online hate speech encompasses a contradiction between the public character of hate speech related consequences and the possibilities of public intervention in the private digital spaces where hateful discourse takes place and spreads. Although not thinking specifically about hate speech, Mansell (2011) calls for the examination of the contradictions between the means of private appropriation of digital spaces and public resistance. For him, in the interest of fostering democratic values and an engaged citizenry, public powers need to ask “what kind of information society do we want?” (Mansell, 2011, p. 16). If pro-active policies and regulatory interventions are left behind, then we can expect the erosion of online environments as inclusive communicative spaces and the maintenance of digital platforms as fertile grounds for the dissemination of hate speech.

5. “Who watches the watchmen?” Hate speech among security forces in Portugal

Recently, a consortium of journalists of newspapers *Público*, *Expresso*, *Setenta e Quatro*, and *Visão* magazine uncovered an exemplary case of hate speech in the Portuguese context. The investigation exposed Facebook pages frequented by security forces professionals – 296 from the Public Security Police (PSP) and 295 from the National Republican Guard (GNR) – containing abundant hate speech. These police officers engaged in offensive remarks that called for violence and sexual assault against women, as well as discriminatory speech based on race, national origin, gender, sexual orientation, and more.

This highlights how hate speech can undermine democratic institutions and principles from within, since security force agents used Facebook to spread hateful messages as if they were acceptable forms of expression. According to the consortium, Facebook’s closed groups like *GNR - Só Camaradas* [GNR - Just Comrades], *Forças de Segurança* [Security Forces], and *Polícias - Profissão de Risco* [Police - Profession of Risk], were frequently visited by certain security agents that acted as hate speakers, endorsing discourses and worldviews affiliated with nationalist extreme right-wing parties. This case can also be further analyzed to understand if certain police officers used their offline and online social capital among their peers to propagate hate speech.

The General Inspector of Internal Affairs (IGAI), Anabela Ferreira, said, in response to the newspaper *Setenta e Quatro*, that the security forces are "prevented from making statements that meet democratic legality, in whatever forum". Ferreira also assured that security forces are attentive primarily to the interaction in social media and that they do not want agents of authority in service in social network sites, who have to behave in a compatible way with the rule of law, defending values that are contrary to this rule of law (Teles & Coelho, 2022). However, the concern of high-level officials alone is not enough to tackle hate speech within the security forces without clear measures and laws in place. These individuals may always claim the right to free speech, even if hate speech is at odds with other public and democratic values.

The president of the Observatory of Security, Organized Crime and Terrorism (OSCOT), Bacelar Gouveia, confronted with the investigation of the consortium, said: "there are people in the security forces who do not have profile for these functions, due to their radicalized thinking" and asks "more discretion in the admissions of new agents" (Soares, 2022). The Attorney General's Office said an investigation has been opened because of the discriminatory statements of certain police members on Facebook (Público, 2022).

The hate content published on Facebook by security forces agents in the two-year period covered by the investigation is a serious issue. It might erode the public credibility of the security forces as a whole, foster distrust towards legal institutions, promote the normalization of discriminatory discourses because of the positions of authority of police officers, and endanger the minorities who were targeted by hateful comments. The fact that IGAI claims to be aware of the online behavior of members of the security forces indicates a failure in the enforcement of laws that the state imposes on its employees (Teles & Coelho, 2022). Finally, Facebook has been ineffective in monitoring and applying its policies against hate speech on its platform. Despite claiming to combat hate speech, the company has been unable to detect or take action against the hateful comments made by members of the Portuguese security forces in various groups.

6. Conclusion

The spread of hate speech on social media often involves the use of language to attack individuals based on their national, ethnic origin, race, sexual orientation, or gender. This speech can circulate quickly on digital platforms and reach many people. According to the United Nations (2019), there is no international definition of hate speech, and what is meant by "hateful" is "controversial and contested". Online hate speakers also call into question the limits of free speech, as they may use freedom of expression as a moral justification for their actions.

The member states of the European Union should take the necessary measures to ensure that perpetrators of criminal offenses are punished in accordance with the legislation in force. However, this is not always the case. In Portugal, for instance, members of the security forces have been known to utilize Facebook groups to spread discriminatory speech against minorities, and there has been limited action taken to address it. One may raise concerns about the effectiveness of hate speech regulation and the conditions under which it is implemented when law enforcement officials engage in it as if it were acceptable speech.

For two years, that social media platform was ineffective in monitoring and addressing hate speech produced in various groups and pages by Portuguese security forces agents. Thus, it raises questions about the extent to which the platform complies with the *Code of Conduct for combating illegal hate speech online*, which it signed with the European Union in 2016. Hate speech is an emerging issue in different countries, whether developed or developing. Creating a more stable definition of hate speech seems an urgent challenge, as well as strengthening the legal mechanisms to combat it in an alliance between governments, platforms, and social media users. Nevertheless, it can be contended that the regulation of online hate speech cannot be separated from the debate of the regulation of the "platform society" in its entirety. Despite the public demonstrations from digital platforms of goodwill and willfulness to work with authorities in tackling hate speech, in practice, their algorithms can strengthen the exposure to discriminatory discourses by recommending related content to users, foster the encounter and clustering of hate speakers in groups and, at the same time, they profit from all online interactions, regardless of the content.

The case unraveled by the Portuguese consortium, with further research, might also reveal the dark side of the social capital of certain influential police officers, and their capability to mobilize social networks to propagate hate speech among their peers, in spite of positive notions of social capital. This perverse link between online hate speech and social capital is more coherent with Bourdieu's (1986) neutral approach to the concept, seen as a reciprocal source of validation and acknowledgment and as a resource in the power struggles between social groups.

Notwithstanding efforts and pressures of governing bodies on social media platforms at the European and national levels, online hate speakers continue to use various expedients to spread prejudiced and intolerant content. For instance, hate speech can spread and be cloaked in the form of disinformation or misleading information. The investigation of the consortium of Portuguese journalists also demonstrates that the traditional press can embrace the role of watchdogs, either through fact-checking or through the public exposure of online hate speech. The current communicational and informational environment certainly poses serious challenges to the traditional press, but it can also constitute an opportunity to reaffirm its public relevance, especially when the regulation of online platforms is lacking.

References

Ben-David, A. & Matamoros-Fernandez, A. (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(2016), 1167-1193.

Benkler, Y. (2000). From consumers to users: Shifting the deeper structures of regulation toward sustainable commons and user access. *Federal Communications Law Journal*, 52(3), 561-579.

Bourdieu, P. (1986). *The forms of capital*. In Richardson, J. G. (Ed.), The handbook of theory and research for sociology of education (pp. 241-258). New York: Greenwood.

Bourdieu, P. (1989). *O poder simbólico*. Lisboa: Difel.

Brown, A. (2018). What is so special about online (as compared to offline) hate speech? Ethnicities, 18(3), 297-326. https://doi.org/10.1177/1468796817709846

Clegg, N. (2020). *Facebook does not benefit from hate*. Meta (1 July). https://shre.ink/cySD

European Commission against Racism and Intolerance – ECRI (2015). *ECRI general policy recommendation Nº 15: On combating hate speech*. https://shre.ink/cySe

External Action Service of the European Union (2022). *Hate speech esthetised society and foments conflict*. https://shre.ink/cySb

Fortuna, P. & Nunes, S. (2018). A survey on automatic detection of hate speech in Text. *ACM Computing Surveys (CSUR)*, 51, 1-30.

European Union Agency for Fundamental Rights (2022). *EU charter of fundamental rights*. https://shre.ink/cySr

Gagliardone, I. (2019). Defining online hate and its "Public Lives": What is the place for "extreme speech"? *International Journal of Communication*, 13(2019), 3068-3087.

Gascón, A. (2012). Evolución jurisprudencial de la protección ante el discurso del odio en España en la última década. *Cuadernos Electrónicos de Filosofía del Derecho*, 26, 310-340.

Gerlitz, C. & Helmond, A. (2013). The like economy: Social buttons and the data-intensive web. *New Media & Society*, 15(8), 1348-1365. https://doi.org/10.1177/1461444812472322

Gitari, N. D., Zuping, Z., Damien, H. & Long, J. (2015). A lexicon-based approach for hate speech detection. *International Journal of Multimedia and Ubiquitous Engineering*, 10(4), 215-230. http://dx.doi.org/10.14257/ijmue.2015.10.4.21

Klein, J. (2012). *The bully society: School shootings and the crisis of bullying in America's schools*. New York: New York University Press.

Lamerichs, N., Nguyen, D., Melguizo, M. C. P., Radojevic, R. & Lange-Böhmer, A. (2018). Elite male bodies: The circulation of alt-right memes and the framing of politicians on social media. *Participations*, 15(1), 180-206.

Mansell, W. (2011). Control of perception should be operationalized as a fundamental property of the nervous system. *TopiCS*, 3(2), 257-261.

Marwick, A. E. & Miller, R. W. (2014). *Online harassment, defamation, and hateful speech: A primer of the legal landscape*. Report (10 June). Fordham Center on Law and Information Policy. https://shre.ink/cyXG

Mihajlova, E., Bacovska, J. & Shekerdjıev, T. (2013). *Freedom of expression and hate speech*. Skopje: Polyesterday.

Miranda, S., Malini, F., Di Fátima, B. & Cruz, J. (2022). I love to hate! The racist hate speech in social media. *Proceedings of the 9th European Conference on Social Media* (pp. 137-145). Krakow: Academic Conferences International (ACI).

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. New York: Penguin.

Pereira, A. O. (2018). O que é lugar de fala? *Leitura: Teoria e Prática*, 36(72), 153-156.

Pohjonen, M. (2018). *Horizons of hate: A comparative approach to social media hate speech*. VOX-Pol Network of Excellence (NoE).

Portes, A. (1998). Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology*, 24, 1-24.

PRISM Project (2015). *Hate crime and hate speech in Europe: Comprehensive analysis of international law principles, EU-wide study and national assessments*. Fundamental Rights and Citizenship Programme of the European Union.

Público (2020). *Governo vai monitorizar discurso de ódio na Internet*. Governo (1 July). https://shre.ink/cySs

Público (2022). *MP abre inquérito a mensagens de ódio de agentes da PSP e da GNR nas redes sociais*. Forças de Segurança (17 November). https://shre.ink/cyXY

Putnam, R. D. (1993). What makes democracy work? *National Civic Review*, 82(2), 101-107.

Silva, R. L., Nichel, A., Martins, A. C. L. & Borchardt, C. K. (2011). Hate speech in social networks: Brazilian case law. *Rev. direito GV*, 7(2), 445-468. https://doi.org/10.1590/S1808-24322011000200004

Silverstone, R. (2007). *Media and morality: On the rise of the mediapolis*. London: Polity Press.

Soares, R. (2022). *Discurso de ódio nas polícias não pode ser ignorado, alerta OSCOT*. RTP (17 November). https://shre.ink/cyX4

Teles, F. & Coelho, P. (2022). *Polícias sem lei: o ódio de 591 agentes de autoridade*. Setenta e Quatro (16 November). https://shre.ink/cyXq

The Verge (2017). *Germany passes controversial law to fine Facebook over hate speech*. Tech (30 June). https://shre.ink/cyXJ

United Nations – UN (2019). *United Nations strategy and plan of action on hate speech*. https://shre.ink/cyXw

UOL (2018). *França anuncia medidas contra discurso de ódio nas redes sociais*. Opera Mundi (19 March). https://shre.ink/cyXm

Waldron, J. (2012). *The harm in the hate speech*. Cambridge: Harvard University Press.

Warner, W. & Hirschberg, J. (2012). Detecting hate speech on the World Wide Web. *Proceedings of the Second Workshop on Language in Social Media* (pp. 19-26). Montréal: Association for Computational Linguistics.

Weber, A. (2014). *Manual on hate speech*. Strasbourg: Council of Europe.

Authors

Allen Munoriyarwa

Senior Lecturer in the Department of Media Studies at the University of Botswana. His research interests are in journalism, news production practices, platforms and social media. He has also researched widely on data journalism, big data and digital surveillance. His research employs different qualitative and quantitative methodologies. He has published more than 25 journal articles and book chapters.

Anne Gilliland

Professor and Director of the Center for Information as Evidence in the School of Education & Information Studies and a faculty affiliate in Digital Humanities and the Promise Institute for Human Rights, University of California, Los Angeles (UCLA). Her research focuses on archival studies, digital recordkeeping and computational approaches to acquiring and analyzing social media and other high volume social and cultural materials, especially those relating to marginalized, displaced and diasporic groups.

Aondover Eric Msughter

Academic staff in the Department of Mass Communication, Caleb University, Imota, Lagos, Nigeria with specialization in Journalism and Media Studies. Msughter has published papers in several national and international scholarly journals and attended and participated in several conferences and workshops on communication, media, and journalism. He is a member of the Association of Communication Scholars

& Professionals of Nigeria (ACSPN), African Council for Communication Education (ACCE), and Fellow of Social Science Research Council (FSSRC), USA.

Arantxa Vizcaíno-Verdú

FPU Predoctoral Researcher and PhD in Communication at UHU. Holds a Master's Degree in Communication and Audiovisual Education from UHU, a Degree in Advertising and Public Relations from the University of Alicante, and an Advanced Technician in Plastic Arts and Design-Illustration from Massana School. Regional Head of the TikTok Cultures Research Network, Associate Researcher of the Influencer Ethnography Research Lab, member of the Comunicar Group, the Agora Group of technological-educational research (HUM-648), and the Alfamed Youth Network. Her studies focus on the analysis of transmedia narratives, fandom and popular culture in social media.

Branco Di Fátima

Journalist and non-fiction writer who holds a PhD in Communication Sciences from ISCTE-IUL. The author of the book *Dias de Tormenta* (Geração Editorial, São Paulo, 2019) and a co-organizer of the collections *Internet - Comunicação em Rede* (iGov, Lisbon, 2013) and *Outros Olhares* (Leiditathi, Belo Horizonte, 2008). He has been involved in several research projects funded by the Portuguese Foundation for Science and Technology (FCT) and the European Commission. Research interests include the application of Data Science for the extraction, processing, and visualization of Big Data. He is currently a researcher at LabCom – Communication & Arts at the University of Beira Interior, in Portugal.

Ebru Gökaliler

Professor of Advertising at Yaşar University, Turkiye. She teaches many courses on advertising and brand communication. She continues to publish national and international articles and books chapters. Her research interests include brand communication, advertising and consumer behavior.

Edson Capoano

PhD in Communication and Culture at the Latin American Integration Program, by University of São Paulo, Brazil. Master in Communication and Semiotics and Bachelor of Journalism from the Pontifical Catholic University of São Paulo. Specializations in Ibero-American Journalism, and in Environmental Journalism. Postdoctoral internships and visiting scholar at the University of Castilla-La Mancha, University of Navarra and University of California San Diego. Current FCT Researcher at CECS for the profile in Journalism, Participation, and Digital Media.

Huizi Yu

Master's student in Biostatistics at the University of Michigan. She is interested in utilizing statistics and machine learning techniques to evaluate health outcomes, particularly with respect to health policy and informatics. Her current research on the older LGBTQ+ communities' health information seeking behaviors in COVID-19 aims to understand the unique disparities and resilience of LGBTQ+ older adults during health crises.

İnanç Alikılıç

Assistant Professor at Malatya Turgut Özal University, Turkiye. His research interests include communication research, research methodology, social media analysis. He has administrative positions at his university including faculty vice dean and career development center director.

Juan-Manuel González-Aguilar

Lecturer and a researcher at the International University of La Rioja (UNIR). His main lines of research are political memes, political communication in social media and hate speech in social media. Some of his recent publications on these subjects appeared in journals such as Social Media & Society, Continuum, Media Culture & Society, and Estudios Sobre el Mensaje Periodístico.

Lida Tsene

She holds a degree in Communication, Media and Culture and a PhD on Social Media and Social Responsibility. She has been teaching communication since 2010 and working in the field for more than ten years. She is Head of PR, Art and Educational Programs of Comicdom Con Athens, founder of the Athens Comics Library, managing co-director of the Digital Comics Museum, organiser and curator of several events, exhibitions, conferences and workshops. Currently she is a researcher and teaching associate at the MA Program Communication and New Journalism of Open University of Cyprus.

Lizhou Fan

Ph.D. student in Information at the University of Michigan. He has broad interests and research experience in data science, computational social sciences, and archival science. His current research focuses on analyzing online toxicity through the lens of health and crisis informatics. He also studies data and information management that supports such research on emerging social issues.

Macarena Parejo-Cuéllar

Associate Professor at the Faculty of Documentation and Communication Sciences at the University of Extremadura (Spain), in the Journalism Area. Principal Investigator of the I+D+i project "Estrategias de traslación mediática para información pública sobre calidad del aire en Extremadura" (Comunicaire), with reference IB20081. Member of the EduTransforma-T Research Group, focused on Transformative education for a global and digital society, at the University of Extremadura (SEJ054). Research focuses on science communication, science literacy, and university media.

Mine Gencel Bek

Researcher and co-Principal Investigator for the DFG funded research group *Transformationen des Populären* (2021-2024) at the University of Siegen at the project titled *Fabricating "the people" – Negotiating Claims of Representation in Social Media in Post-Gezi Turkey*. She also teaches at both undergraduate and postgraduate programmes of the Media Studies department at the University of Siegen. The courses currently being taught are *Media Populism*; *Emerging Media and Digital Journalism*; *Digital Media and Activism and Introduction to Qualitative Methods for Media Analysis*.

Muluken Asegidew Chekol

Assistant Professor of Media and Communication at Debre Markos University, a radio Journalist, and Media and Communication Trainer. He was the Founder of Mango Magazine, Dilla University Academic publication. He served as head of the department of journalism and communication, and Director of Public and International Affairs Relations at Dilla University. He authored an academic book in 2018 in Amharic entitled *Journalism and Communication*. He had conducted his PhD project on social media hate speech in Ethiopia.

Mykola Makhortykh

He is an Alfred Landecker lecturer at the Institute of Communication and Media Studies at the University of Bern. In his research, Mykola focuses on politics – and history-centered information behavior in online environments and how it is affected by the information retrieval systems, such as search engines and recommender systems. He recently published in journals such as Telecommunications Policy, Memory Studies, New Media & Society, and Journal of Information Science.

Özlem Alikılıç

Professor of Public Relations at Yaşar University, Turkiye, and she has taught PR, CSR, Social Media and Crisis Communication. She has contributed to many books and articles and has appeared in most major international publications.

Patricia de-Casas-Moreno

PhD in Communication from the University of Huelva (Spain). Assistant Professor in the area of Audiovisual Communication and Advertising at the University of Extremadura (Spain). Member of the Ágora Research Group at the University of Huelva, which is part of the Andalusian Research Plan (Hum-648) and the EduTransforma-T Research Group at the University of Extremadura (SEJ054), which focuses on transformative education for a global and digital society. Additionally, she is also member of the Euro-American Inter-University Research Network Alfamed. The focus of her research is on media literacy, media, and narratives.

Tiago Lapa

Assistant professor and researcher of communication at Iscte - University Institute of Lisbon. He also coordinates and teaches courses in the fields of digital sociology and online methods of inquiry. Participates in international scientific networks such as the World Internet Project and the European Media Coach Initiative, related to Internet studies, the digital divide and new media literacy. He also belongs to the advisory council of the Safe Internet Center of the Portuguese Foundation for Science and Technology (FCT).

Vinicius Prates

Journalist, PhD in Communication and Semiotics, full-time professor at the Center for Comm1unication and Letters at University Presbyterian Mackenzie (CCL-UPM) in São Paulo, Brazil, and adjunct coordinator of the Media Research Group 1 Dia, 7 Dias in the Graduate Program in Communication and Semiotics at the Pontifical University of São Paulo (PPGCOS-PUC-SP). He is the author of A *map of ideology in the Anthropocene* (2020), organizer of *Symptom and fantasy in communicational capitalism* (2017), and *Network communication in the decade of hate* (2022).

Vítor de Sousa

He holds a PhD in Communication Sciences, at University of Minho (Braga, Portugal), with the thesis "Da portugalidade à lusofonia". Among his research interests are issues around National Identity, Memory, Cultural Studies, Media Education and Theories of Journalism. He is a teacher at Universidade de Trás-os-Montes e Alto Douro (UTAD, Vila Real, Portugal), and a researcher at Communication and Society Research Center (Universidade do Minho, Braga, Portugal). He is a member of Portuguese Association of Communication Sciences, where he coordinates the Intercultural Communication WG (2022-2024). He was a journalist (1986-1997) and press officer (1997-2005).

fct Fundação para a Ciência e a Tecnologia

COMPETE 2020

QREN QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL

UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional

LABCOM COMUNICAÇÃO & ARTES

UNIVERSIDADE BEIRA INTERIOR

Grupo de Editoriales Universitarias AUSJAL

OdeCOM OBSERVATORIO DE COMUNICACIÓN

IHS

Centro de Publicaciones
PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

Pontificia Universidad Católica del Ecuador

Hate speech manifests itself in different social contexts, such as political debates, artistic expression, professional sports, or work environments. However, the rapid development of digital technologies, and especially of social media platforms, has created additional challenges to understanding this extreme act. This book explores the nature of hate speech on social media. Readers will find chapters written by authors from 11 countries, prioritizing a diversity of approaches from the Global North and Global South. Hate speech in digital environments challenges academia and society. Virtual armies replicate violent narratives. This book aims to dispel some of these uncertainties.

SPI Expanded

This book series is indexed by
Scholarly Publishers Indicators

LabCom Books
www.labcom.ubi.pt

EdiPUCE
https://edipuce.edu.ec/

ISBN: 978-9978-77-664-3
9789978776643