Preface

ANALYSING HATE SPEECH: GENERAL FRAMEWORKS AND RESEARCH PROBLEMS

The unleashed feelings circulating in the streets are probably the worst nightmare for any authority, even for people. As long as they limit their public nature to the setting in which they occur, these outrages can be more or less conspicuous, even constitute a crime, but little else. A surprise, a complaint, a show or an intervention by the police can follow them. They may not get a place until a fifth or sixth click in a digital newspaper. In overcoming the barriers of order, the definitive thing is the volume that it acquires. This dimension is defined by the number of people involved in that mother's outing. When the masses are the protagonists of a movement (cultural, sports, political, social, among others), the only solution is to make way for them to reduce disasters because cancelling them is impossible.

Outbursts of joy are more bearable. However, even these generate problems for lovers of order and the rest (even if they do not say so for fear of hurting sensibilities). "Popular" celebrations of sporting and political success show it quite often, but nothing compares to hate explosions' destructive effects. Indeed, in all cultures, there have always been authentic "hate professionals." They are the ones who generate, encourage, lead or take advantage of these outbursts. Whether they are psychopaths to lock up or tolerable elements in necessarily unstable societies are assessments that fit this highly unequal scale. In addition, getting the diagnosis right is only possible at the end and by then, things have already gone wrong, and it is only possible to rebuild with what is left.

Our next history has witnessed the construction and consolidation of the announced global village. In this relatively new urban context, the public expression of emotions has also been reorganised, especially the manifestly negative ones. These have had to adapt to the dimensions and characteristics of a macro city that intertwines the real with the virtual, the analogical with the digital, the enormity of the dimensions with the neighbourhood's proximity, and the multiplication of possible poles of identity with depersonalisation of relationships. In this complexity, merely face-to-face has lost the sense of the only channel whose rupture can drag down the rest. On the contrary, the triggering role of the virtual is increasing and it is the potential to involve elements (human or not) of a social and material nature in the new overflows that hate feeds (Nortio et al., 2020). Without ruling out maximum limits: the liquidation of the stick figure that represents the execrable and that the masses inevitably confuse with real institutions and people. A first approach to this issue, also framed in the particularly sensitive territory, is "The Theme of Hated in Modern Communication: The Case of Palestine," one of the chapters of this book.

This dangerous journey threatens to turn into a dark Matrix, presented as Mac Luhan 's happy Global Village. It is an increasingly widespread fear. This is shown by an example that analyses one of

the present contributions ("The Southernification of the Pandemic in Italy: Images of the South, Fears of Contamination, and the First Wave of COVID-19 in Italy"). For this reason, it is intended to redirect the feared process to avoid destroying civilised life in the style of an enviable "European garden." It is witnessed little by little that a colossal effort is being made to avoid the disorder that any form of hate implies, which always translates into violence: from the one generated by clean (and above all sustainable) technologies in the first world to the that appear full of filth, dirt and misery and armed to the teeth, from failed states.

Academics (from all areas of knowledge that label the most diverse databases of the highest quality journals) have answered the call in this fight against hate in our societies, especially in the most advanced ones, perhaps because they live in them. Some have joined institutional organisations that try to get ahead of these crimes. Others have decided to analyse the indicators that offer clues to promote policies that prevent or counteract both crimes and incitement to them. Both of us carry out our activity in particular contexts. Of course, hatred is not absent in them. His presence is clear and probably inevitable. However, if it is measured in actions (fatalities, attacks, physical violence, brutal discrimination) and compared with the rest of the world, the levels of intensity and amplitude that they acquire in failed or unrecognised states are not even remotely reached. and legal respect for human rights. The contribution "Hate Speech or Hate Shot? finding patterns of the anti-Muslim Narratives in Italy" is an example of the importance of this first distinction. This asymmetry must be highlighted whenever the analysis of the social presence of hate is addressed: whether or not it is institutionalised.

Another asymmetry that should be highlighted is that most academic production on hate speech has its origin and setting in the United States (Tontodimamma et al., 2021). In addition, the countries that follow are Great Britain, Australia and Canada. The production of others is located at a great distance (Paz et al., 2020). In the face of this overwhelming data, one can legitimately wonder if hate speech is something typical (or almost) of the Anglo-Saxon world. One might almost ask as a research question whether the rest of the Western world is less sensitive to this problem or has simply learned to live with it. Another possibility is that it has not worried those academic communities far from the privileged Anglo-Saxon campuses of the very first world. Another possible question would be whether hate speech could be described as simple dysfunctions of affluent societies. Whether academic emulation (a cynical perspective) or awareness of the problem, the academic production of hate speech has become more and more general (Di Fátima, 2023). Even the disproportion pointed out here should be tempered by the general predominance of "high-end" academic journals from the Anglo-Saxon "windows."

However, hate speech shows that something is not working well in Society. That is why they are of interest to social science analysts. These ways of saying, writing, conversing, and representing are relatively easy to detect individually. Nevertheless, it is more complicated to take charge of their leading role in specific discriminatory acts and establish cause-effect relationships, or more in accordance with the possibilities of academic studies on Society, to establish significant correlations of interest. Some of these advances and difficulties can be seen in two of the contributions in this volume led by two institutional entities that have been working for years on the detection and relationship of hate speech and crimes of the same type: "European Initiatives for the Support and Counseling of Victims of Hate Crimes" and "European Initiatives for the Support and Counseling of Victims of Hate Crimes."

There is also a prior difficulty for researchers: measuring, quantifying and counting these "discourses", although it might initially seem elementary. They are not as far as hatred is concerned: from the outset, one hates according to a plus and a minus. The biggest problem is not establishing a scale of the intensity of hate or the expressions containing it (which is not an elementary task either). The

difficulty lies in the enormity of the mass of texts that researchers face. These can indeed be reduced to specific cases, but the applicability focuses mainly on what they can contribute as a methodology, or as has been said in a metaphorical and understandable tone: what can be said about an ocean after analysing one of its drops of water. The work "The Semiotics of Xenophobia and Misogyny on Digital Media: A Case Study in Spain."

Going back to that necessary relativity in hate and its expressions, which mark different intensities, the Anti-Defamation League already proposed five general levels presented as a pyramid, distributing actions and speeches from lesser to more intense. At the base are negative stereotypes, followed by insults and expressions of discrimination. At the top are threats and genocide. Other research groups have adopted other solutions for the complete analysis of discourses: six classificatory concepts that can be reduced to three to facilitate the work of the designers of the corresponding algorithm (De Lucas Vicente, et al., 2022). Some attempts to organise and generalise these efforts are mentioned in "Creating an Online Network, Monitoring Team and Apps to Counter Hate Speech, and Hate Crime Tactics in Europe."

The presence (and detection) of hate in the expressions disseminated by the media does not necessarily imply that it is "hate speech." The prominent role of hate could be said not to ensure that there is "hate speech" in the strict sense, in which academics who study these expressions usually handle it. Hate must refer to a vulnerable group. Vulnerability is a broad and elastic concept. More so in satisfied societies such as Western ones, in which the condition, the label, of victim constitutes a distinction with a positive tone in the discourse (not necessarily, nor usually in social reality). The powerful, at least those who are far from contempt for who they are, can be envied, insulted, or wish evil. However, these expressions directed against those who move in a habitual and relatively safe world do not properly constitute "hate speech." Perhaps they are out of envy, fury, or a reaction to defeat... they will be in bad taste, insulting, uncivil and may even constitute crimes defined in the penal code. However, they are not sociologically, psychologically, or communicatively speaking "hate speech."

The digital age confronts social science researchers with the treatment of massive data if they want reliable studies for their colleagues. The first barrier to be overcome is establishing a fluent and clear conversation between social science academics, mathematicians, and computer scientists. It is not always overcome. A practical obstacle is the execution times of the processes. When the technicians affirm that they are achievable, the social analysts imagine them to be little less than instantaneous. In the meantime, months go by, and the data lose relevance, and the conclusions are of no interest for their application or their publication in journals with guaranteed diffusion among specialists in the area. This difficulty of understanding between specialists from one area or another appears overcome in "Using HurtLex and Best-Worst Scaling to Develop ERIS: A Lexicon for Offensive Language Detection," where linguists and computer scientists have managed to work together with good results. It is more frequent to find mathematicians, statisticians and computer scientists in teams that are integrated, in turn, into more extensive and diversified groups, with experts in social sciences ("Approximation of Hate Detection Processes in Spanish and Other Non-Anglo-Saxon Languages.")

There are usually fewer problems in defining precisely what data has indicator value. This aspect is the responsibility of social scholars. It is fundamental. For example, establishing categories in the intensities of hate is key in the matter that concerns us. Each must allow a news reader to classify each in the corresponding "drawer." Those who define these intensities must establish distinguishable categories and avoid classifications based on gradations of the same dimension. The limits between the intensities of hate cannot be indicated by the passage from one figure to another on a scale, and they must refer to facts or, better, to different acts: it is not the same to call for action as to insult or suggest. They are

not degrees on the same scale but very clear and differentiated actions. In each case, there may also be mitigating or intensifying factors, easy to mark grammatically. It will also be possible to assess whether or not there is irony, humour, or some other quality considered significant (Paz-Rebollo et al., 2021). In addition, the perception of hate speech has a very prominent subjective component and studies are needed to understand it and the aspects that condition it (Salminen et al., 2018; Udanor & Anyanwu, 2019).

The intensities, their definition and treatment constitute a starting point that adds value to the most frequent analyses limited to indicating the existence or not of hate in certain expressions in the media and social networks. But knowing what that number responds to is at least as important. In other words: does the abundance of expressions tending to hate indicate in each singularity a human being who expresses his opinion, or are we facing "mechanisms" (boots) that sow judgments with a single author (personal or institutional) on the networks? Even when it comes to people, won't they be hired to do that job? It does not just matter for giving value to the raw digital data. The existence or not of these procedures offers clues about other issues. The text "Spreading Organised Hate Content" addresses this interesting issue.

The intensity indicators are not exhausted in the studies on expressions that facilitate the diffusion of hate feelings. It also matters a lot how to define what is hated: women, immigrants, those of another religion, other races or ethnic groups, and those of another sexual orientation. The mere establishment of this "labelling," which looks like a simple catalogue, apparently so simple, has its difficulties. For example, those who have to program to differentiate themselves must know what to do when faced with an expression that manifests contempt for multiple reasons: a black immigrant woman and Muslim religion from Africa. This offers no difficulty in reading, but it is not easy to specify a search carried out by an algorithm, let alone prepare it for that location. If each possible thematic combination of hate constitutes a new category, the possibilities and the volume of information that must be carried out "manually" to feed the future algorithm skyrocket. Mathematicians and social analysts have to speak.

This book does not lack some contributions of interest on "specific types" of hate that are translated into expressions that both digital media and specific social networks disseminate, on what their characteristics are. Those linked to ethnic issues ("The Expression of Hate in Portuguese Digital Media: Ethnic and Racial Discrimination") or immigration ("Mapping Stigmatising Hoaxes Towards Immigrants on Twitter and Digital Media: Case Study in Spain, Greece, and Italy") and refugees ("Online Hate Speech and The Representations of Refugees in #VatanimdaMülteci (#RefugeeinMyCountry)"). Sometimes, they are only distinguishable by the focus of the study and by the methodological principles that inspire them.

Mention has already been made of some of the frontiers that must be addressed when addressing studies of "hate speech." First, what is hate speech (or not), and the relative weight that hate and speech have in them. The old principle that "a nail painted on the wall can only hang an object also painted on the same wall" helps to understand the limits of our studies on hate speech. They are, by definition, focused on the analysis of expressions, not on the analysis of facts, although they frequently denote or prepare them for even violent actions.

Although they do not mark a border but an approach, studies that propose solutions to the evil of disseminating these expressions of hate are beginning to increase. Sometimes the contributions come from the field of narratives ("New Narratives to Defuse Hate Speech"). Others propose procedures that resort to training in the understanding and assimilation of communication itself in societies that could be described as digital ("Analysis of Radicalisation Prevention Policies From the Perspective of Educommunication in Mediterranean Countries"). Not only is it a legitimate perspective, but it is also an ethical responsibility that involves all of us academics who make it our profession to analyse the environment and context in which we live. If the formula already enunciated (that of the painted nail) and the principle

of proportionality between evils and their remedies are applied, it would be necessary to say, without a doubt, that one discourse is fought with another. Moreover, add, and not with anything else. The observatories governments and institutions have set up to analyse speeches as part of preventive policies against future (and probable) hate crimes are understandable and probably necessary. There are doubts about its effective preventive efficacy, but only doubts. Sometimes these observatories seem more interested in discovering new "pockets of social vulnerability" into which hate is unleashed in the media than in establishing effectively preventive proposals. Digital media also establish control systems, although they are not always fully effective (Paz-Rebollo et al., 2021). However, the development and dissemination of new integrating narratives seem to be a proposed solution that is already being worked on: the use of counter-discourses, not so much confrontation (Hangartner et al., 2021; Woodzicka et al., 2015).

Other limits refer to the most appropriate methodologies to detect, measure, catalogue, and classify these discourses. In this sense, we also know that the metadata (daily, date, section, place, among others) offered by the media associated with the news does not always accurately define variables that must be considered. Let us think about the scenarios in which certain expressions occur. The newspaper sections where they are published offer first cataloguing, but they may not always be helpful. It is clear that the sports sections, for example, tend to mark "settings" in which these expressions occur. Political ones also tend to serve this purpose. However, those of Society already offer principles of ambiguity. This is not to mention that each medium labels its sections differently, which makes solving this problem even more difficult. Another option, which is becoming increasingly influential, is discourse analysis since it can interpret the ironies, metaphors and double meanings most present in these expressions.

This set of observations and, above all, the contributions that follow constitute a first referential framework that has sought collaborations linked to the Mediterranean area and the need to advance in the automatic analysis of these discourses: although, unfortunately, hatred is not limited solely or mainly to narratives that we analyse.

Julio Montero Diaz
Proeduca, Spain

Elias Said Hung
Universidad Internacional de la Rioja, Spain

REFERENCES

De Lucas VicenteA.Römer PierettiM.IzquierdoD.Montero-DiazJ.Said-HungE. (2022). *Manual para el Etiquetado de mensajes de odio*. doi:10.6084/m9.figshare.18316313

Di Fátima, B. (Ed.). (2023). *Hate speech on social media*. University of Beira Interior.

Hangartner, D., Gennaro, G., Alasiri, S., & Donnay, K. (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proceedings of the National Academy of Sciences of the United States of America*, *50*(118), 1–3. doi:10.1073/pnas.2116310118 PMID:34873046

Nortio, E., Niska, M., Renvik, T., & Jasinskaja-Lahti, I. (2020). The nightmare of multiculturalism: Interpreting and deploying anti-immigration rhetoric in social media. *New Media & Society*. Advance online publication. doi:10.1177/1461444819899624

Paz, M. A., Montero, J., & Moreno, A. (2020). Hate Speech: A Systematised Review. *SAGE Open*, *10*, 1–21. doi:10.1177/2158244020973022

Paz-Rebollo, M. A., Cáceres-Zapatero, M. D., & Martín-Sánchez, I. (2021). Suscripción a la prensa digital como contención a los discursos de odio. *El Profesional de la Información*, *30*(6), e300613. Advance online publication. doi:10.3145/epi.2021.nov.13

Paz-Rebollo, M. A., Mayagoitia-Soria, A., & González-Aguilar, J. M. (2021). From Polarisation to Hate: Portrait of the Spanish Political Meme. *Social Media + Society*, *7*(4). Advance online publication. doi:10.1177/20563051211062920

Salminen, J., Veronesi, F., & Almerekhi, H. (2018). Online Hate Interpretation Varies by Country, But More by Individual: A Statistical Analysis Using Crowdsourced Ratings. *Fifth International Conference on Social Networks Analysis. Management and Security*, 88-94. 10.1109/SNAMS.2018.8554954

Tontodimamma, A., Nissi, E., Sarra, A., & Fontanella, L. (2021). Thirty years of research into hate speech: Topics of interest and their evolution. *Scientometrics*, *126*(1), 157–179. doi:10.100711192-020-03737-6

Woodzicka, J. A., Mallett, R. K., Hendricks, S., & Pruitt, A. V. (2015). It's just a (sexist) joke: Comparing reactions to sexist versus racist communications. *Humor: International Journal of Humor Research*, *28*(2), 289–309. doi:10.1515/humor-2015-0025