



Approach to the concept and identification of hate in digital media in Spain. A proposal for its classification

Prof. Dr. Max Römer – Universidad Camilo José Cela

Prof. Dr. Alberto De Lucas Vicente – Universidad Internacional de la Rioja

Prof. Dr. Dámaso Izquierdo – Universidad de Navarra

Project PTDC/COM-JOR/3866/2020, funded by:



Project PID2020-114584GB-I00, funded by:



Abstract

Defining a social phenomenon requires a specific design of the methodology that will be applied. The study of hate in digital media and social networks in Spain is a way to propose an instrument to the research community for the understanding of the intensities of hate in written discourses.

The purpose of this dissertation is to expose the way in which researchers obtained the intensities of hate in Twitter and Media. The methodology used was a forum of discussion among researchers with two main objectives: 1) determine the levels of hate messages in Spain; 2) define how hate grows with the interaction in social networks.

This classification of hate provides five different intensity levels, labeled from zero to five, being the lowest number that originated from insult and affront, and the highest level the one that represents a physical threat or intimidation to a person or social group.

Each level identifies hate in digital media in Spain, and classifies the uses of hate in different matters of Spanish life: social, political, sexual identification, ideological, etc. The results will provide an understanding of the social Spanish context when hate occurs.

During the creation of the methodology for this study, the researchers debated the procedure to determine the subtleties of language, semantics, and pragmatic situations about several types of hate in media or social networks like Twitter.

On the other hand, the application of the levels of hate by researcher classifiers showed that the instrument of classification works in a simple and friendly way, and allows to know the panorama of hate in Spain.

Key words: Hate studies, Spain, Social Networks, Media, Classification, Hate intensifies

Previous considerations (1/2)

1. Delimiting “Hate Speech”:

- Not all hate or verbal violence is "Hate Speech"
- We do not understand by hate speech only what is legally established as such (speech condemnable by law)
- The antechamber of hate: when there is no hate, but it could provoke it
- Journalistic malpractice: sensationalism, clickbait...



Previous considerations (2/2)

2. Graduating the level of hate: a methodological problem

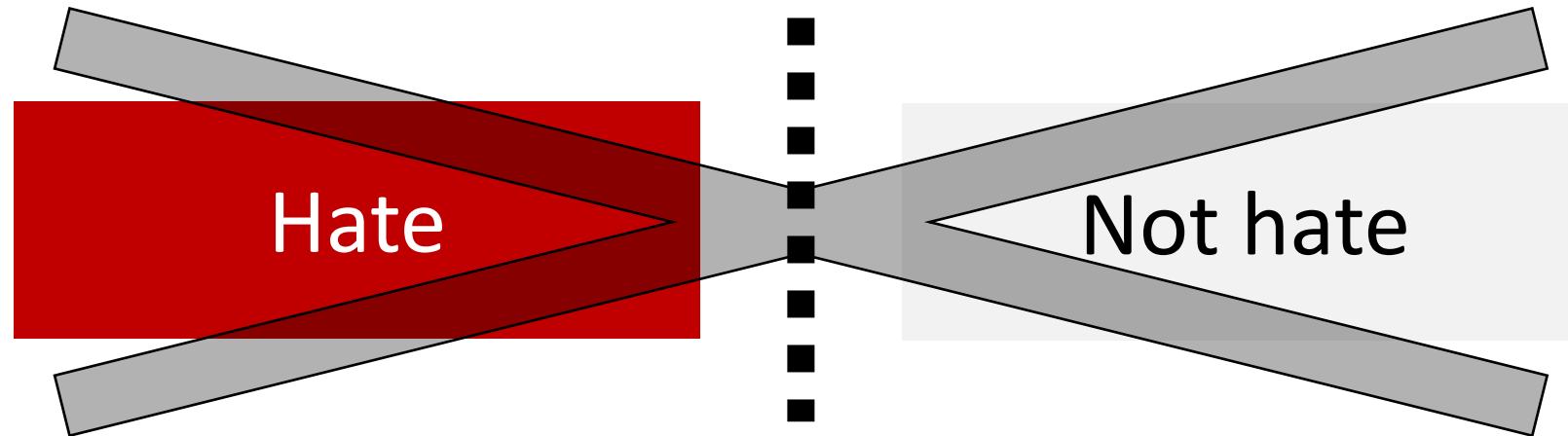
- Levels that are not permeable or continuous (ineffective)
- Equivalence problem >Level = >Intensity
- Faders and Faders (Do not change level)
- Humor problem (irony, for example): does it cancel or modify hate or its intensity? Is it humor even if it's not funny?
- Combination of levels in a single message: predominance of the highest

3. Distinguish types of hate by its receiver

- The infinite casuistry > the problem of “Others” as a mixed bag
- Need for distance and objectivity: "What if the hated person deserves it?"
“What if he really is a fascist/terrorist/useless...?”

Hate intensity levels

- Hate speech is a complex phenomenon of a scalar nature.

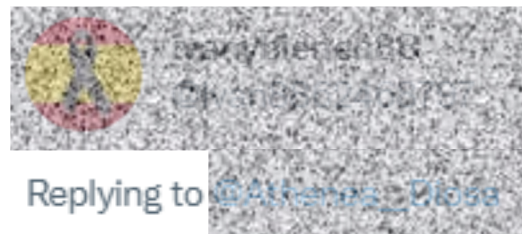


- Proposal of 5 (+1) levels of intensity



Intensity level 5

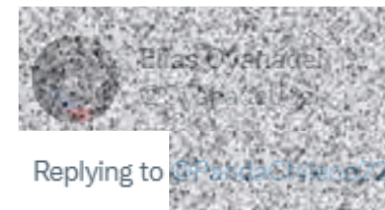
- Expressions that call to act (explicit incitements or threats) with physical violence against a social group or against something directly linked to it.
- Expressions of wishing that another die or suffer some physical harm.



Hay que matarlos a los moros de mierda

[Translate Tweet](#)

4:54 PM · Jul 6, 2019 · Twitter for Android



Venezolanos de mierda, ojalá los maten y que después no salgan los progres pasao a culo con sus discursitos baratos sobre la Xenofobia y de que "Ningún humano es ilegal"

[Translate Tweet](#)

5:35 PM · May 2, 2022 · Twitter for Android

Intensity level 4

- Implied threats.
- Expressions of positive emotion in the face of death, aggression or physical harm that another person has caused against someone or something.
- Intimidating expressions that are not physical in nature (for example, bullying).

Project PTDC/COM-JOR/3866/2020, funded by:



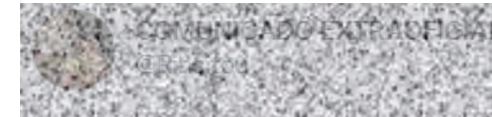
Project PID2020-114584GB-I00, funded by:



Cierra ya la puta boca facha de mierda. Ten cuidado cuando vayas de caza y no te vuelas los sesos racista de mierda.

[Translate Tweet](#)

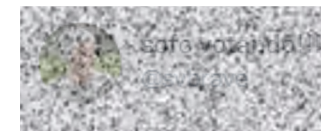
11:32 PM · Mar 26, 2020 · Twitter for Android



Pues yo me alegro de que se haya muerto el torero... Venid a pedir carcel por mi cabrones

[Translate Tweet](#)

5:43 PM · Jul 13, 2016 · Twitter for Android



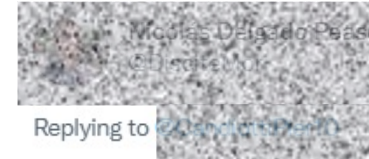
si eso fue para mí pedazo de hormonadas mononeuronales les voy a hacer la vida imposible y se las voy a hacer pasar tan como el orto, el que avisa no traiciona muak

[Translate Tweet](#)

4:08 AM · Dec 17, 2020 · Twitter for Android

Intensity level 3

- Verbal violence: this is not violence, or a threat linked to actions that take place in the real world.
- Abuse
- Difficulties in identifying insults:
 - Reappropriation of the insult to criticize it.
 - General insults vs. insults specialized in the denigration of certain social groups.
 - Identification of an insulting intent in words that, by default, are not necessarily insults:
 - *"Another lie from this shitty Nazi useless idiot @jairbolsonaro I hope he goes away in peace!"*



No todos los sudacas somos hispanoamericanos y no todos los hispanoamericanos somos sudacas.

[Translate Tweet](#)

Intensity level 2

- Attribution of clearly negative actions to a social group or one of its members in order to spread a negative image.
- It is not about knowing if the attributed negative action is real or not, but if the author of the message intends to discredit that social group.



Intensity level 1

- Attribution of actions that, by default, are not negative to a social group or one of its members in order to spread a negative image.



Intensity level 0

- A label is used to refer to a social group that has derogatory connotations, but the author of the message does not use it in a derogatory sense: “ultra-Catholic”, “far left”.
- When presenting a negative action carried out by one or more people, words are used that characterize such people as members of a social group that is regularly the target of hate speech. The author of the text does so without the express intention of stigmatizing that social group:

“a group of Moroccans rapes a young girl of X years”

Types of hate

Classification according to the social group to which the hate is directed

| TYPE | SUBTYPE |
|-----------------------------|---|
| XENOPHOBIA | |
| RELIGIOUS HATRED | Antisemitism Islamophobia Anti-Christianity Anti-Catholicism |
| RACISM | Antiindigenism Negrophobia Antigypsyism Sinophobia Anti-Arabism |
| MISOGYNY | |
| HATE FOR SEXUAL ORIENTATION | Homophobia Transphobia |
| IDEOLOGICAL | Against right-wing ideologies Against leftist ideologies Antifeminism Against independence |
| OTHERS | |

Application problems: from theory to practice

1. Teach “the machine”:
 - The context problem
 - The problem of perceiving humor and irony
2. Messages that combine various types of hate and different levels, but the highest level of hate does not correspond to the most obvious type
3. Quotations and polyphony, especially when it is not explicit
4. Self-referential messages

Application problems: from theory to practice

5. Difficulty distinguishing some types of hate:

- When the recipient is a politician: personal hatred (excluded), ideological (because of the political party or ideology) or class/occupation (for being a politician, especially in a pandemic)
- When the recipient is a newspaper or media: hate for malpractice or attribution of it (especially in a pandemic) or hate for representing an ideology (ideological type)

← Tweet



Mendeisa
@MendeisaF



En respuesta a [@elmundoes](#)

Lo de esta "señora" ya es de juzgado de guardia. No se puede ser más miserable, jugando con la vida de miles de madrileños para arañar unos votos. Qué asco, pero qué asco.

1:07 p. m. · 28 ene. 2021 · Twitter for Android



Application problems: from theory to practice

5. Difficulty distinguishing some types of hate:

- When the receiver is a journalist or someone who comments on a piece of news or responds to a tweet: personal hatred (because of their attitude, actions or character) or ideological hatred
- Unclear types of hate:
- To independence (ideology? xenophobia?)
- To the ores/refugees/immigrants (xenophobia? racism? aporophobia? religious?)



Approach to the concept and identification of hate in digital media in Spain. A proposal for its classification

Prof. Dr. Max Römer – Universidad Camilo José Cela

Prof. Dr. Alberto De Lucas Vicente – Universidad Internacional de la Rioja

Prof. Dr. Dámaso Izquierdo – Universidad de Navarra

Project PTDC/COM-JOR/3866/2020, funded by:



Project PID2020-114584GB-I00, funded by:

