

Páginas: 222-233
Recibido: 2022-01-13
Revisado: 2022-01-19
Aceptado: 2022-03-11
Preprint: 2022-03-15
Publicación Final: 2022-05-31









www.revistascientificas.us.es/index.php/fuentes/index

DOI: <https://doi.org/10.12795/revistafuentes.2022.20240>

Hate speech: a systematic review of scientific production and educational considerations

Discursos de odio: revisión sistemática de la producción científica y consideraciones educativas

-   **Alberto Izquierdo Montero**
Universidad Nacional de Educación a Distancia (España)
-   **Noemi Laforgue-Bullido**
Universidad Nacional de Educación a Distancia (España)
-   **David Abril-Hervás**
Universitat de les Illes Balears (España)

Abstract

Hate speech is a specific form of expression targeting certain vulnerable or minority groups. It is a social phenomenon which has gained strength in recent years due to its normalisation from the public sphere and escalation on social media, particularly in contexts of crisis such as the circumstances that have arisen around COVID-19. This paper presents a systematic review of the scientific literature produced over the past two decades in English, Portuguese and Spanish as a tool to examine, from social science, the landscape of studies around one of the main threats democratic societies face today. In addition to an evidence-based idea of the state of the art, the results show a substantial increase of research into hate speech over the past five years and how focal points and disciplines have evolved from those approaches and are gaining in importance, along with the analysis of regulations and laws, computational linguistics and discourse analysis. Furthermore, the dearth of research from the field of communication and education is confirmed, which would allow us to examine not only its management and analyse its impact, but rather its understanding from critical pedagogical perspectives committed to the deepening and extension of democracy.

Resumen

Los discursos de odio constituyen una forma de expresión específica dirigida contra determinados colectivos vulnerabilizados. Se trata de un fenómeno social que ha ganado fuerza en los últimos años de la mano de su normalización desde la esfera pública y su propagación a través de las redes sociales, especialmente en contextos de crisis como la generada en torno a la COVID-19. En este trabajo se presenta una revisión sistemática de la literatura científica producida en las dos últimas décadas en inglés, portugués y español, como herramienta para acercarnos desde las ciencias sociales al panorama de estudios acerca de una de las principales amenazas a las que se enfrentan hoy en día las sociedades democráticas. Además de disponer de un estado de la cuestión basado en evidencias, los resultados muestran un incremento sustancial de la investigación sobre discursos de odio en los últimos cinco años, así como una evolución de los enfoques y disciplinas desde los que se aborda, en los que cada vez tienen más peso, junto al análisis de normas y leyes, la lingüística computacional y el análisis del discurso. Se constata, además, una escasez de investigaciones que desde el campo de la comunicación y la educación permitan abordar no solo su gestión y el análisis de su impacto, sino su comprensión desde perspectivas pedagógicas críticas comprometidas con la profundización y extensión democráticas.

Keywords / Palabras clave

Scientific activities; Speech; Emotions; Discrimination; Disinformation; Media education; Pedagogy; Democracy.
Actividad científica; Discurso; Emociones; Discriminación; Desinformación; Educación sobre medios de comunicación; Pedagogía; Democracia.

1. Introduction

Hate speech is an ever-present social phenomenon in our societies, particularly since its normalisation in the public sphere (Fernández de Castro & González-Páramo, 2019) and the rise of so-called “authoritarian populism” (Norris & Inglehart, 2019). This speech poses a threat to social harmony and has prompted different national and international institutions to create expert opinions and recommendations. One of these institutions is the European Commission Against Racism and Intolerance (2016), which has warned of hate speech’s reach, defining it as the use of one or more specific forms of expression, or its justification and backing, to promote the rejection of specific social groups which have, contextually, become more vulnerable — namely, women, and ethnic or religious minorities and migrants — in a non-exhaustive list of categories. Fish (2019) has suggested the patent link between hate speech and “action”, in as much as they involve performative messages of reality and, therefore, cause suffering to certain collectives, despite a lack of recognition in legislation.

By and large, hate speech includes a repertoire of communication strategies able to fuel the creation of predatory (Appadurai, 2007) and murderous (Malouf, 1999) identities in a present day characterised by a “neoliberal time which threatens to be an anti-democratic time” (Fassin, 2018, p. 128). This discursive strategy looks to connect with certain collective emotions that arise from successive economic and social crises (Foessa Foundation, 2019), stoking fear and a dismissal of those not considered worthy of sharing the increasingly scarce aid and guarantees of the “welfare state”. Such discourse deliberately appeals to that “deep history” we recount so as to shape our image and self- and group esteem (Hochschild, 2016), and which is linked, in turn, “not so much to a sense of inequality that must be repaired but a sense of privilege that must be conserved opposite those who wish to attack it” (Rancière, 2021, p. 1). Situations of mounting uncertainty such as the COVID-19 pandemic intensify this trend (United Nations, 2020).

Within this context, a global rise in the presence of different anti-rights movements can be observed and documented, a kind of “Reactionary International” (Ramos et al., 2021) or “International of Hate” (Tamayo, 2021), which, on a European stage, focus their efforts on attacking the rights of women, migrants and LGBTIQ+ collectives (Cabezas & Medina, 2021). This entails a framework of social agents, mostly connected to the extreme right (political parties, media, think tanks, web spaces, influencers, swathes of trolls and bots sitting in virtual private communities, etc.), which look to impose a narrative framework — and with it a political, economic and social agenda — capable of normalising that which today we understand to be “politically incorrect” (Stefanoni, 2021) through a tautening of the right to the “freedom of expression” and the progressive displacement of the Overton Window¹ (Marantz, 2021). Thus, there is the search for an antidialogical situation (Tiburi, 2019) through the expansion of a specific framework to define the “common sense” of our times (Lakoff, 2007).

The worrying rise of hate speech in recent years (United Nations, 2019) is bound to and exacerbated by what some experts on post-truth and fake news have defined as “a state of widespread lies” (Aparici & García-Marín, 2019, p. 24). Such a backdrop raises a set of questions which are worth heeding: initially, we wondered whether, as a result, there will have been growing interest in this area from the scientific-academic field, as we will see in the systematic review laid out here. We maintain that to examine this hypothetically exponential increase in studies that address such a subject could stress the importance of observing how research on hate speech has been made up until now, to then continue working along lines which can complement each other and can be effectively translated into the wholesale understanding and weakening of such mechanisms of production and dissemination, and the social effects (Emcke, 2017).

Another reason which leads us to engage in a systematic review of scientific literature on hate speech is the hitherto limited presence of this kind of approach around the phenomenon we are concerned with. As we will see, among the systemised documents only two systematic reviews in this respect have been identified prior to that presented here. The first one exclusively centres on racist and xenophobic speech on social media (Bustos Martínez et al., 2019), gathering documents in Spanish and English across a seven-year time frame with the aim of offering information on online hate speech. Therefore, it includes articles, reports and documents from institutions which devote their time to reporting and eradicating xenophobic discrimination. The second (Paz et al., 2020) focuses on the compilation and analysis of legal studies and the field of

¹ The Overton Window is a political theory which sets out a range (window) to place the imaginable, normal, sensible, or acceptable for most people in a given context, highlighting how this range can be displaced via communicative strategies towards that regarded as politically incorrect, extremist or unthinkable at a particular time, thereby normalising discourse and practices which would not previously have been socially acceptable

communication. These analyse 1,112 scientific articles written in English and Spanish between 1975 and 2020, tracing the Web of Science (WoS) through different descriptors related to hate speech recommended by the Council of Europe in the handbook drawn up by Keen and Georgescu (2020). The findings show that research on hate speech is nothing new, although it was not until 2003 that it grew significantly.

The present systematic review looks to complement and broaden the evidence provided by the aforementioned, while also looking at publications in three languages — English (Eng.), Portuguese (Por.) and Spanish (Spa.) — with a greater number of available articles in this subject area and expanding the focal point to scientific-academic articles which are not necessarily indexed in the WoS. As far as we are concerned, we also add the category Approach to the proposed classification to match the type of approach from the Disciplines studied. Consequently, the main objective of the present work is to conduct a systematic review of the scientific-academic approach of hate speech over the last twenty years (2001–2020), taking as a chronological reference the start of policies framed inside the “War on Terror” in the wake of the 9/11 attacks as one of the main landmarks articulating current narratives (Velasco & Rodríguez-Alarcón, 2020). Accordingly, another three specific objectives have been pursued: 1) to identify those scientific publications which tackle the issue of hate speech from a specific criteria of inclusion-exclusion; 2) to draw up a categorisation proposal that can be used to classify the said studies and can prove useful in identifying the state of this issue and as the basis/guidance for future research around it; 3) provide knowledge about the temporary evolution of studies revolving around hate speech, indicating some of the shortcomings identified within this sphere, as well as possible lines of continuity, with a strong emphasis on the area of educocommunication, i.e. focusing on a media and information literary praxis as a question of democracy (Carlsson, 2019).

2. Methodology

The work we present in this article encompasses many of the main methodologies formulated by Grant & Booth (2009) in their “SALSA framework” to carry out systematic reviews of scientific literature, which for the sphere of social sciences this author defines as “systemised reviews”, differing from other empirical studies in the field of experimental sciences. Consequently, the procedure has been developed from a constant self-questioning around the suitability of each phase comprising it: search, selection, analysis and synthesis (Ibid.) has entailed a reflective exercise, set out below, from transparency, whilst also being an essential component of the systematic review (Codina, 2018), making it — and us — available for other researchers.

2.1. Search procedure and inclusion-exclusion criteria

We started the search process by making a comparison between the institutional search tools of Universidad Nacional de Educación a Distancia (UNED) and Universitat de les Illes Balears (UIB), universities with which we are associated as a research team. We opted for UNED’s Linceo+ search tool as it yielded a higher number of results when we carried out the search: 360 vs 246². The characteristics of such a search are specified in the following criteria: 1) search terms and languages: “hate speech” (Eng.), “discurso de odio” (Spa.), “discurso de ódio” (Por.), in singular and plural, present in the title. In a previous search with no language filters, we noted that English, Portuguese and Spanish, in that order, were by far the predominant languages among the references generated; 2) the material type: “all records” and “peer-reviewed publications; 3) time frame: 2001 to 2020 (twenty complete years).

In applying the advanced search, and also considering its conciseness (which, for instance, determined us being limited to the three higher-profile languages), it generated 360 publications, to which we applied the following exclusion criteria by means of a thorough review of scientific articles: a language that differs from those selected; the impossibility of accessing the complete text; book review; book chapter; repeated appearance (eliminating copies); no approach to the topic of hate speech, despite it being included in the title.

² At the time of writing the article, available databases on the UNED search engine had risen to 121, and the said information can be consulted at the following link: <https://uned.libguides.com/az.php>

The application of this exclusion criteria reduced $n=360$ to $n=266^3$ as a final number of scientific articles making up the systematic review we analysed in the present work, and which is presented in Figure 1 with respect to two of the initial inclusion coordinates (date and language) present in the meta-data of the selected articles:

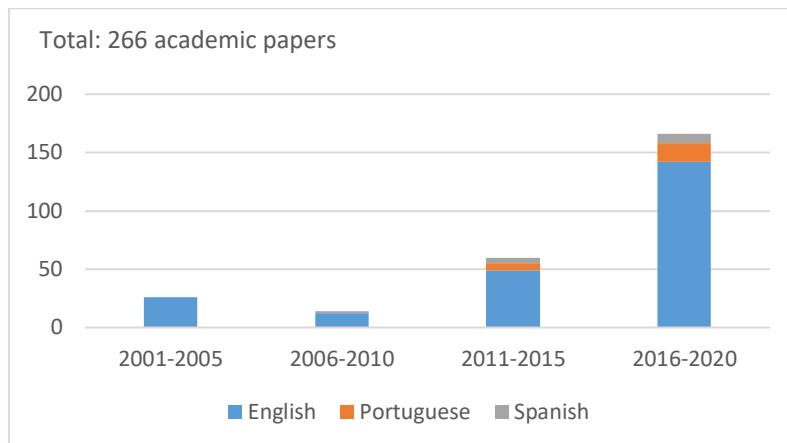


Figure 1. Scientific articles according to date and language of publication

2.2. Proposal for the classification fo scientific articles

The proposal of classification (Table 1) was generated through a repeated circular inter-rater process of agreement – individual review and group discussion around the adequacy of subcategories. The initial categories stem from screening questions to respond to the scientific approach taken in this subject area in the last twenty years, placing emphasis on: 1) the support upon which studied hate speech is manifested; 2) the focus of this hate speech; that is, the collectives at which it is aimed; 3) the scientific approach or methodology used to analyse the information and draw conclusions; and 4) the discipline or disciplines from which this subject is addressed. Furthermore, these four categories were divided into 33 subcategories to classify the different articles. The subcategories emerged during the reading and analysis of the different texts (Monge, 2015), which explains how they do not fully reflect the possible array of subjects and aspects linked to hate speech, since they were limited to those observed through the analysis of systematized scientific articles. Despite these being assembled so not to overlap — they are mutually exclusive — the same article can be categorised into subcategories from the same family and, for instance, can reference two different focal points of hate speech.

Table 1

Categories and subcategories to classify studies on hate speech and its frequency in this study

Category name	Subcategories	Frequency (n)
Category 1. Support	Legal Framework	80
	Social Media	67
	Literature	44
	Speech by Public and/or influential Figures	22
	Perception of the Population	19
	Websites	18
	Media	16
	Education Centres	8
	Others	5
Category 2. Focus	Undefined	165
	Migrants and Ethnic Minority Groups	63
	Racism with a Religious Element	31

³ Link to Figshare, where the 266 corresponding bibliographical references to scientific articles included in this systematic review can be consulted: <https://figshare.com/s/90751a81551440fcc465>

	Gender, Identifications and Sexual Orientations	28
	Political Antagonism	12
Category 3. Approach	Normative Analysis	89
	Discourse Analysis	53
	Document Review	45
	Detection, Algorithms and Machine Learning	33
	Case study	20
	Questionnaire	13
	Others	9
	Ethnography	9
	Systematic Review	2
Category 4. Discipline(s)	Law	106
	Social Sciences and Politics	77
	Computational linguistics	30
	Philosophy	12
	Communication	11
	Psychology	10
	Semiotics	7
	Education Sciences	5
	Interdisciplinary Approach	4
	Others	4

The first classification category proposed is the *Support*, understood as the place — physical or symbolic — where hate speech is positioned, analysed, and addressed in the articles that are part of the systematic review. This category is also made up of various subcategories: *Education Centres*, for speech which occurs in education institutions; *Speech by Public or Influential figures*, for speech expressed by people with a far-reaching voice, as, for instance, in the abundant literature on hate speech in US election campaigns; *Literature*, for messages starting from articles, scientific theories or entertainment literature; *Legal Framework*, for the analysis of laws, legal cases or sentences; *Media*, for speech that occurs in mass media (TV, Newspaper, Radio...); *Perception of the Population* to talk about hate speech, or the opinion of it, by the population in general terms; *Social Media*, for those messages located on online social media accounts — Twitter, Facebook, Instagram, Reddit... — *Websites*, for hate speech expressed on websites or comments in the news and on videos; and *Others*, for those spaces that cannot be classified in the aforementioned categories and which have a lingering presence.

The second category is the *Focus*; namely, the collective or collectives at which hate speech is aimed. Among the different subcategories are: *Political Antagonism* to name hate speech that attacks people or collectives with minority political ideals — more in terms of power than number; *Gender, Identifications and Sexual Orientations*, to group together misogynous LGBT-phobic speech; *Migrants and Ethnic Minority Groups* for messages discriminating against people positioned on the less privileged side of these categories; *Racism with a Religious Element* in reference to speech which discriminates against people on the basis of creed, and with a direct link to racism (Islamophobia, anti-Semitism...); and *Undefined*, for those articles which do not clearly state the hate speech they are referencing. Although these subcategories have surfaced from the review, we are also aware that they do not encompass the full range of existing hate speech or that their classification could be different.

As a third category, we set up the *Approach*; that is, the scientific method used to approach this reality: *Analysis of Speech*; *Normative Analysis*; *Questionnaire*; *Detection, Algorithms and Machine Learning*; *Case Study*; *Ethnography*; *Documentary Review*; *Systematic Review*; and *Others*.

Finally, the articles have been categorised based on the *Discipline(s)* from which they address the subject of study. The disciplines which make up these subcategories are: *Social Sciences and Politics*; *Education Sciences*; *Communication*; *Law*; *Philosophy*; and *Interdisciplinary Approach* for those studies which tackle hate speech from more than one theoretical discipline; *Computational Linguistics*; *Psychology*; *Semiotics*; and *Others*.

3. Analysis and results

As already noted, scientific output about hate speech has increased exponentially in recent years, with the papers published in the last five years representing 62.36% of the total sample, coinciding with that highlighted in other reviews (Paz et al., 2020), which point to an exponential growth of research around the issue from 2014 onwards. In the present section, we will carry out an in-depth analysis of how the study of hate speech is characterised based on the *Support, Focus, Approach and Discipline*.

3.1. Disciplines that address hate speech

After carrying out 266 subcodifications, a broad representation of subjects related to *Law* (39.85% of the total) has been encountered in a detailed analysis of the disciplines approaching hate speech (Figure 2). This involves studies which analyse this problem from the perspective of the legal sphere and an analysis of sentences, comparative law, and legal trends. Other disciplines with a high representation in this area are: *Social Sciences and Politics* (28.95%) and *Computational Linguistics* (11.28%), and, less frequently, we find *Philosophy* (4.51%), *Communication* (4.14%), *Psychology* (3.76%) and *Semiotics* (2.63%). The studies on hate speech approached from *Education Sciences* (1.88%), *Interdisciplinarity* (1.50%) and *Other Disciplines* that differ from those previously mentioned (1.50%) are more incidental.

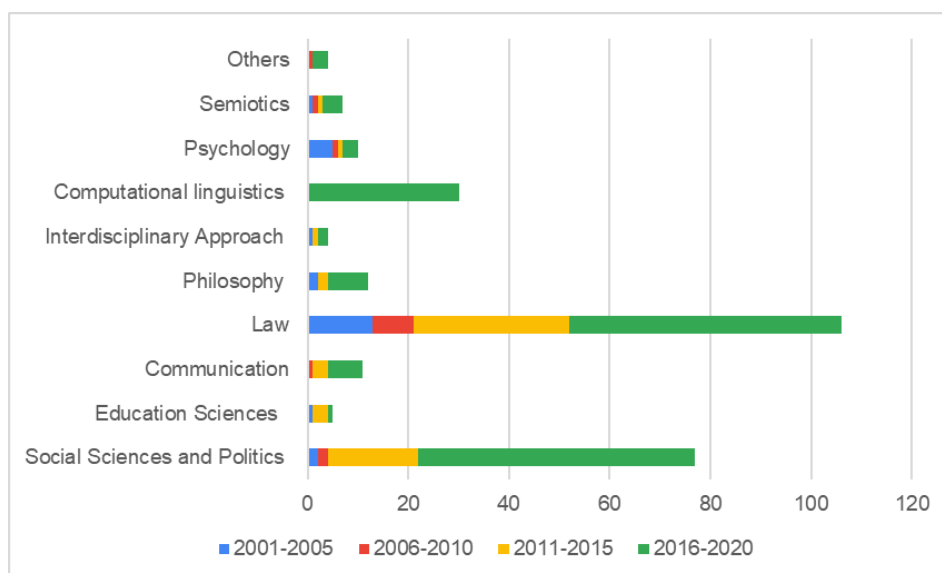


Figure 2. Disciplines addressing research on hate speech

If we focus on the temporary development of research into hate speech, *Computational Linguistics* appears as a discipline that emerges over the past five years with a representativeness of 18.07% — this category involves studies focused on creating algorithms that detect and act opposite hate speech via *Machine Learning*; a recent systematic review focusing on this issue was carried out by Poletto et al. (2021). We also found that *Psychology* had a higher representation in the first five years but was gradually generating fewer publications regarding hate speech — moving from representing 20% to 1.81% of the total in the first and last five-year period, respectively.

3.2. The methodological approach to deal with hate speech

The articles which address hate speech do so from divergent methodological focal points (Figure 3). A total of 273 codifications have been carried out in this regard, with the possibility of various methodological approaches being applied in the same article. The most frequent analysis methodology is *Normative Analysis*, carried out in 32.60% of cases. By and large, the studies approached are from the discipline of *Law*. Other methodologies frequently used are *Discourse Analysis* (19.41%), *Document Review* — articles with a more theoretical tenor — (16.48%), and Detection of Algorithms through Artificial Intelligence and *Machine Learning* (12.09%). To a

lesser degree, there were also uses of *Case Studies* (7.33%), the *Questionnaire* (4.76%), *Ethnography* (3.30%) and *Other Approaches* (3.30%). Only 0.73% of articles address this subject from the *Systematic Review*.

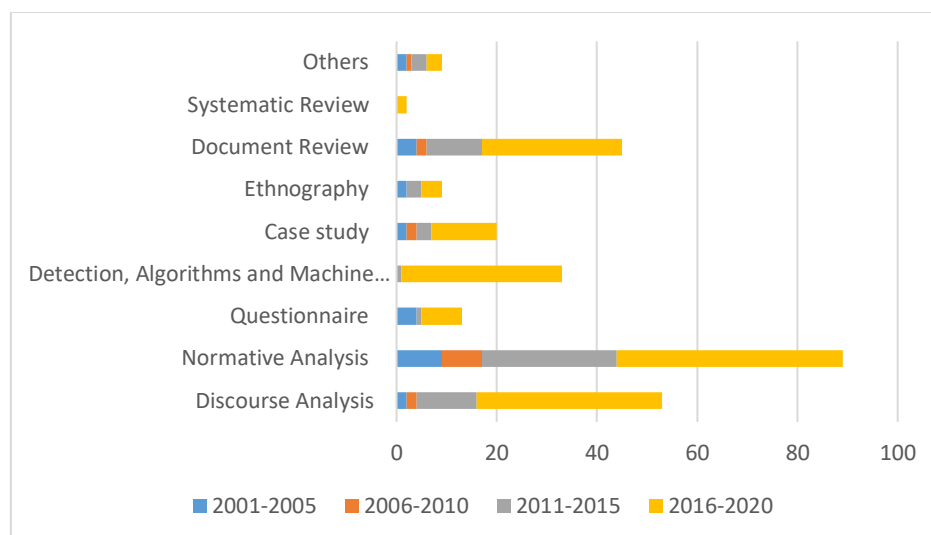


Figure 3. Approaches hate speech works are addressed from

The most frequently used approaches in *Law* are *Normative Analysis* (75,85%), *Document Review* (12,03%) and *Case Study* (3,70%). This entails studies which address legislation regarding hate speech in different territories, specific court cases or schools of thought around the dichotomy of freedom of expression/the penalisation of hate speech. The approach of *Detection, Algorithms and Machine Learning* is addressed in full from *Computational Linguistics* (96,66%). For the most part, they are articles which seek to respond to the strong presence of hate speech on the internet and, above all, on social media. *Discourse Analysis* is the approach which most in *Communication* (54,54%) and *Social and Political Sciences* (44,15%) opt for, analysing speech that materialises on social media, in different theoretical strands and in the speech expressed by public and influential figures. *Ethnography* is another of the focal points used by *Social Sciences and Politics* (7,79%), while from *Psychology* the *Questionnaire* (80%) is chosen as a technique to approach this subject area, chiefly impacting the population's perception of hate speech.

3.3. The Focus of hate speech

In relation to the Focus of hate speech (Figure 4) upon which these studies centre, 299 codifications have been carried out, with most studies concentrating on different collectives affected by this speech. The most frequent subcategory was *Undefined* (55.18%). To a large degree, these are studies which do not define the focus on which the analysis is centred and speak generically in that respect — in some cases they can cause confusion between hate speech and other types of disrespectful or antagonistic expression related to a collective or specific minority. The second most concurred subcategory is *Migrants and Racialised Minority Groups* (21.07%), and within this subcategory we find studies which reference anti-immigration speech and racist discourse aimed at ethnic minorities, for example gypsy people. Thirdly, we find *Racism with a Religious Element* (10.37%), a subcategory which includes those which refer to Islamophobic and anti-Semitic hate speech, or those referring to other minority religions in the territory in which hate speech occurs. To a lesser extent there exists speech that makes a reference to discrimination due to *Gender, Identifications and Sexual Orientations* (9.36%) or hate speech on account of *Political Antagonism* targeting minority political groups in power terms (4.01%).

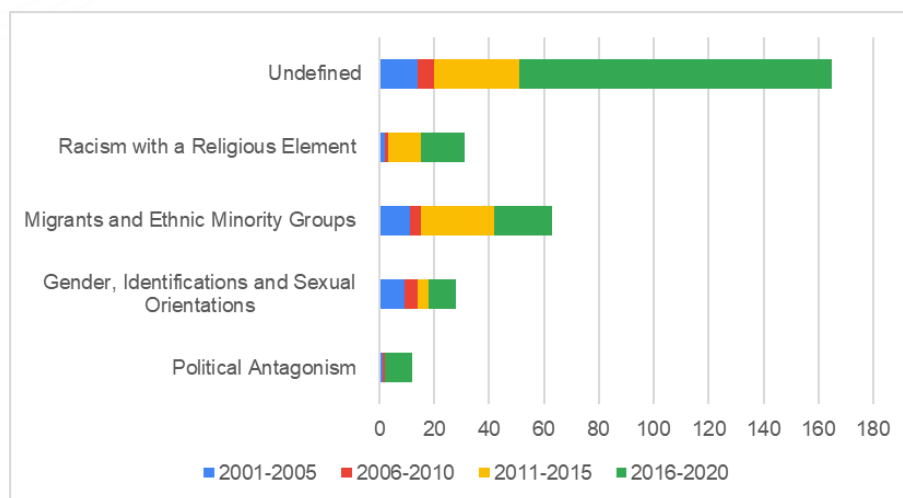


Figure 4. Focuses approached from works on hate speech

Despite this general data, other conclusions can also be drawn if we turn our attention towards the different proposed time frames. First, it is worth noting that the number of articles that do not define the types of hate speech referred to have increased exponentially, rising from 36.84% to 65.52% of the total representation in the last five years. Moreover, representation in studies encompassing *Gender, Identifications and Sexual Orientations* as the focus of hate speech has dropped, descending from 23.68% of the total in the first five years to 5.75% in the last five. On another note, the percentage of studies which address this subject area from the focus of *Migrants and Minority Racialised Groups* rose by almost 13 percentage points in the 2011–2015 period.

3.4. The Support of hate speech

Within this category (Figure 5), 279 classifications have been made — 28.67% of the total is research focused on the legal treatment of hate speech, followed by the analysis of this speech present on *Social Media* (24.01%), hate speech addressed in or from *Literature* (15.77%) and hate speech expressed by *Public or Influential Figures* (7.89%), particularly in the framework of election campaigns. The other categories have a lower presence.

The articles which address this issue from the *Legal Framework* focus their gaze on national laws or international recommendations, comparative studies between countries, case studies/sentences, security protocols or studies on the development of laws of this nature. Regarding to the *Social Networks* more commonly named in these articles, we find Facebook (19.40%) and even more so Twitter (44.77%), the latter of which is demonstrably the social media platform, which is more prone to hate speech, correlating to the reductionism of language associated with the limited number of characters in each tweet, among other aspects. The analyses focused on Literature reference, in the main, academic production — articles, theoretical trends, etc. — although to a lesser degree we find articles that focus on historical literature or literary-style texts. Those studies which address hate speech expressed by *Public or Influential Figures* refer to academic authorities, religious leaders and social media influencers.

Nevertheless, if we observe the trend from a longitudinal gaze, we can see how the chief support was the *Legal Framework* in the first three lustrums — from 2001 to 2015 — followed by *Literature* or *Speech by Public or Influential Figures*, a trend which changes in the last five years, bringing about an exponential growth of interest in the study of *Social Media* as hate speech supports, along with the mutual growth of studies which address the *Legal Framework*.

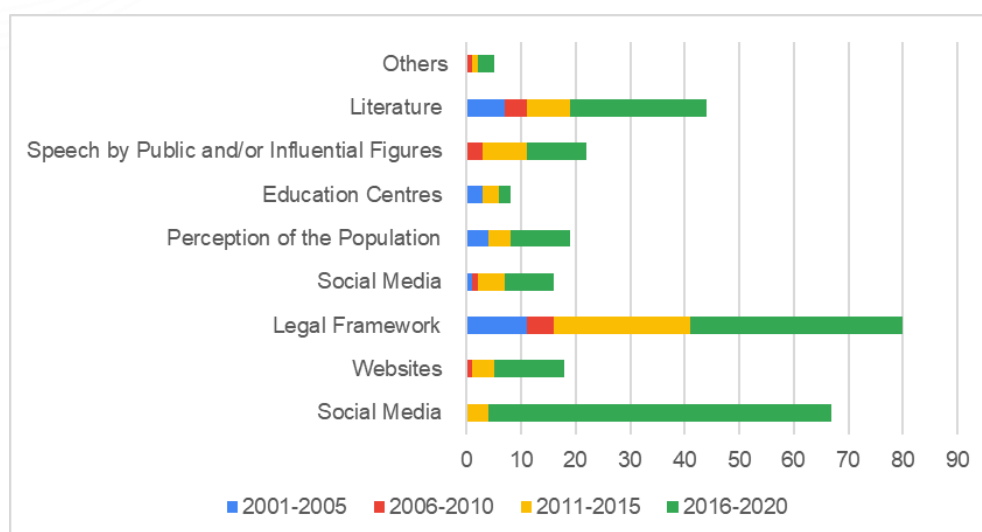


Figure 5. Supports where the hate speech analysed in publications is found

4. Discussion

Through this study we have sought to contribute to widening knowledge around the scientific-academic work being carried out to understand hate speech, particularly in terms of a “socially alive issue” (Izquierdo Grau, 2019) which is relevant for education and democratic functioning. The results of this systematic review demonstrate a notable increase in the scientific-academic interest in hate speech, specially over the last five years, and they also coincide with that observed by Paz et al. (2020), who cite 2014 as the year when publications around this issue really took off. In recent years, in which the use of the concept of “hate speech” has also spread to other spheres (the media, NGOs, observatories, social movements...). It also coincides with the emergence of nationalist and xenophobic political parties in numerous European parliaments or even in governments, within what could be understood as the “fourth wave” in the rise of the extreme right (Acha, 2021). These parties and organisations have exploited attacks on European soil and the humanitarian plight of refugees, particularly in the second half of the last decade and, from a position of power and privilege, they take on hate speech as a daily political tool (Mouffe, 2018). In fact, the public policies deployed by Bolsonaro in Brazil could explain the position of Portuguese as the second highest language represented among the publications found in this systemisation of articles; since sixteen of the 22 articles in this language (72.7%) were published between 2017 and 2020, and from a concern focused chiefly on regulatory frameworks and on democratic guarantees to regulate the hate speech that stems from power.

This democratic unrest, addressed primarily from a Law perspective, has its strongest presence in the scientific publications analysed, in which there is a deepened tension between the “freedom of expression” and the dignity and protection of vulnerable collectives (Bleich, 2014; Fish, 2019). Tension in which we encounter different stances that could be reconciled, perhaps, via midway concepts such as the “responsibility of expression” as a proposal to rethink the limits of individual freedom in democratic contexts (Todorov, 2014), particularly when this is at odds with human dignity. This legal concern has crossed over in recent years with what has been happening in virtual environments and on social media, which emerge as the main stages through where hate speech flows on account of a series of circumstances that nurture it (Jakubowicz, 2017; Cohen, 2018; Carlson, 2021; Marantz, 2021). In part, this responds to types of messages dovetailing with the tendency of the media and social media to dramatise information by prioritising sensationalist, simplified and controversial issues (Ellinas, 2010).

In this regard, in the past five years the *Detection of Algorithms and Machine Learning* emerges as a way to answer this discursive explosion on online social media (Oriola and Kotze, 2020), and, consequently, it comes as no surprise that *Computational Linguistics* is one discipline with a greater standing in studies on hate speech, in reference to the most recent years and behind *Law* and *Social and Political Sciences*. This observation raises the question of whether the same algorithmization that has been useful for managing hate speech as an in-demand commodity through fake news and other content of disinformation can also be a valid

tool to combat it, especially when there is a lack of articles and research to prevent hate speech, both in the sphere of communication and informative deontology and in educommunication, despite the recommendations oriented towards investigating and promoting media and information literacy (Muratova et al., 2019).

Regarding the *Focus* areas targeted by hate speech, we find that the majority are not centred on a social-target group in particular, but instead uphold a rationale through different examples or by reflecting on hate speech in general. This will not be problematic if there is an awareness that hate speech can only be considered as such if, among other requirements (Kaufman, 2015), it is aimed at real or imagined collectives in a vulnerable position within a specific context; thus, we can avoid conceptual corruption which could remove the predominantly protective intent of such a concept (Waldron, 2012).

We have seen that both anti-immigration speech and speech aimed at racialised groups, with or without a religious element, concur with reports and studies which place migrants under the spotlight of anti-rights and hate speech (Fernández de Castro & González-Páramo, 2019). We are surprised that, despite anti-feminist and LGTB-phobic speech also being at the heart of present-day hate speech, as we saw previously (Cabezas and Medina, 2021), its study appears to have waned. Hypothetically, this could be explained by the backing of other more specific terms, instead of “hate speech”, from gender and LGBTIQ+ studies. Another explanation for this phenomenon could be the more limited acknowledgment of this type of communicative actions as hate speech. The recognition of sexist and homophobic violence and the implementation of legal and social measures to eradicate it are still embroiled in a process of development with a long road ahead. It is unsurprising, then, that when we think of hate speech it is that which is socially recognised and reported that comes to mind, leaving to one side those struggles with less support. Therefore, we can foresee that the broadening of speech recognised as hate speech will be preceded by an increased space in the public agenda taken up by these struggles. It is worth clarifying, equally, as we highlighted in the methodology section, that the fact of having favoured the emergence of subcategories from the selected articles has meant that different existing focal points (ableism, aporophobia, etc.) have not been reflected specifically in this systemisation, a situation which does not imply a lack of social relevance.

Finally, the progressive update of the present systematic review and its expansion into other languages, as well as search terms related to hate speech, will be a necessary task to undertake in order to keep on increasing knowledge of scientific approaches to this phenomenon.

5. Conclusions

As a conclusion, including lines of continuity, it would be interesting to delve deeper into works that address hate speech from a practical perspective via research and intervention proposals from educommunication. Reviewing this type of project — still a minority, as the systemisation has shown — could shed light on the strategies that are more useful for working in different social settings, both in-person and virtual. The current algorithmization of the response to hate speech raises doubts over whether its automated identification will cover the broad and subtle repertoire of words, symbols, images, etc. used to socially damage specific groups. Moreover, there must be an enquiry into the role algorithms can play in democratic societies, and the limits of their use given that visible speech is just the tip of the iceberg of a broader and more latent series of problems, in which “the paradox entailed in the global rise of the extreme right — and other actors of hate, we can add —⁴ and the danger of its ideology to democracy can only be stopped with more democratic stances” (Acha, 2021, p. 134). Therefore, investigations could delve into the relationships that can materialise between algorithms and educommunication.

On the other hand, it would also be interesting to qualitatively explore systematised articles to discover the hate speech, regarding that which we have referred to as *Focus*, considered by researchers working in these areas and the reasons behind them prioritising some over others. Furthermore, we are interested in observing how the US debate concerning the First Amendment, which protects freedom of speech, influences arguments which academics from different latitudes — not exclusively in a US context — use when it comes to justifying (or not) the regulation of hate speech. This involves a debate in which, we believe, societies as a whole should participate, and not just figures specialised in legal-judicial matters, nor exclusively from logics situated in the Global North (de Sousa Santos, 2010), given that “questions regarding the legal regulation of hate speech also raise concerns about how the term will be defined and by whom” (Carlson, 2021, p. 13), particularly taking into consideration that this type of speech “is a way to remind members of other groups of your own group’s position of dominance” (Ibid.: p. 28–29). Without doubt, this opens the opportunity to carry out a “public pedagogy”

⁴ The em-dash clarification is ours.

(Giorgy & Kiffer, 2020) to nurture the development of a profound pedagogical commitment to addressing hate speech through education (UNESCO, 2021), based on “learning to live together and learning together to live” (Garcés, 2020, p. 29) as the backbone of democracies.

Support

This systematic review is framed inside the working context linked to the research project “Teenagers against hate speech. A participatory research to identify scenarios, agents and strategies to face them” (INTER Research Group on Intercultural Education), funded by the 2019 Research Grants of the “Reina Sofía” Center on Adolescence and Youth.

References

- Acha, B. (2021). *Analizar el auge de la ultraderecha. Surgimiento, ideología y ascenso de los nuevos partidos de ultraderecha*. Gedisa.
- Aparici, R. & García-Marín, D. (2019). Historia de la mentira: más allá de Derrida. In R. Aparici & D. García-Marín. *La posverdad. Una cartografía de los medios, las redes y la política* (pp. 17-24). Gedisa.
- Appadurai, A. (2007). *El rechazo de las minorías. Ensayo sobre la geografía de la furia*. Tusquets.
- Bleich, E. (2014). Freedom of Expression versus Racist Hate Speech: Explaining Differences Between High Court Regulations in the USA and Europe. *Journal of Ethnic and Migration Studies*, 40(2), 283-300. <https://doi.org/10.1080/1369183X.2013.851476>
- Bustos Martínez, L., Santiago Ortega, P. P., Martínez Miró, M. A. & Rengifo Hidalgo, M. S. (2019). Discursos de odio: una epidemia que se propaga en la red. Estado de la cuestión sobre el racismo y la xenofobia en las redes sociales. *Revista Mediaciones Sociales*, (18), 25-42. <http://dx.doi.org/10.5209/MESO.64527>
- Cabezas, A. & Medina, P. (2021). *Mapeo de actores y repertorios de odio. El género y la migración en el epicentro de las políticas anti-derechos en España y la Unión Europea*. Asociación de Investigación y Especialización sobre Temas Iberoamericanos. <https://bit.ly/3mX1Un1>
- Carlson, C. R. (2021). *Hate Speech*. Massachusetts Institute of Technology (MIT) Press.
- Carlsson, U. (Ed.) (2019). *Understanding Media and Information Literacy (MIL) in the Digital Age. A question of Democracy*. UNESCO Chair on Freedom of Expression, Media Development and Global Policy at University of Gothenburg. <https://bit.ly/3zQ8iQT>
- Codina, L. (2018). *Revisiones bibliográficas sistematizadas. Procedimientos generales y Framework para las Ciencias Humanas y Sociales*. Universitat Pompeu Fabra. <https://bit.ly/3gY0iWm>
- Cohen, R. (2018). Taking North American White Supremacist Groups Seriously: The Scope and Challenge of Hate Speech on the Internet. *International Journal for Crime, Justice and Social Democracy* 7(2), 38-57. <https://doi.org/10.5204/ijcjsd.v7i2.517>
- Keen, E. & Georgescu, M. (2020). *Bookmarks. A manual for combating hate speech online through human rights education. [Revised edition]*. Council of Europe / No Hate Speech Movement. <https://bit.ly/3ByqoXP>
- Ellinas, A. (2010). *The Media and the Far Right in Western Europe: Playing the Nationalist Card*. Cambridge University Press.
- Emcke, C. (2017). *Contra el odio*. Taurus.
- European Commission against Racism and Intolerance (2016). *ECRI General Policy Recommendation nº 15 on Combating Hate Speech*. European Council. <https://bit.ly/3DIDrYD>
- Fassin, E. (2018). *Populismo de izquierdas y neoliberalismo*. Herder.
- Fernández de Castro, P. & González-Páramo, A. (2019). *La franquicia anti-inmigración. Cómo se expande el populismo xenófobo en Europa*. Fundación porCausa. <https://bit.ly/2YwVUrd>
- Fish, S. (2019). *The First. How to think about Hate Speech, Campus Speech, Religious Speech, Fake News, Post-truth and Donald Trump*. Atria.
- Fundación Foessa (2019). *VIII Informe sobre exclusión y desarrollo social en España*. Fundación Foessa. <https://www.foessa.es/viii-informe/>
- Garcés, M. (2020). *Escuela de aprendices*. Galaxia Gutenberg.
- Giorgi, G. y Kiffer, A. (2020). *Las vueltas del odio. Gestos, escrituras, políticas*. Eterna Cadencia Editora.
- Grant, M.J. & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, 26(2), 91-108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Hochschild, A. (2018). *Extraños en su propia tierra. Réquiem por la derecha estadounidense*. Capitán Swing.
- Izquierdo Grau, A. (2019). *Contrarelatos de l'odi a l'ensenyament i l'aprenentatge de les ciències socials. Una recerca interpretativa i crítica a l'educació secundària* [Doctoral thesis]. Universitat Autònoma de Barcelona. <https://bit.ly/3jEDywj>
- Jakubowicz, A. (2017). Alt-Right white lite: Trolling, hate speech and cyber racism on social media. *Cosmopolitan Civil Societies: An Interdisciplinary Journal*, 9(3), 41-60. <http://dx.doi.org/10.5130/ccs.v9i3.5655>
- Kaufman, G. (2015). *Odium Dicta - Freedom of Expression and Protection of Discriminated Groups on the Internet*. CONAPRED. <https://bit.ly/3t6ziJj>
- Lakkof, G. (2007). *No pienses en un elefante*. Editorial Complutense.
- Maalouf, A. (2010). *Identidades asesinas*. Alianza Editorial.

- Marantz, A. (2021). *Antisocial. La extrema derecha y la "libertad de expresión" en internet*. Capitán Swing.
- Monge, V. (2015). La codificación en el método de investigación de la *grounded theory* o teoría fundamentada. *Innovaciones educativas*, (22), 77-84. <https://doi.org/10.22458/ie.v17i22.1100>
- Mouffe, C. (2018). *Por un populismo de izquierda*. Siglo Veintiuno.
- Muratova, N., Grizzle, A. & Mirzakhmedova, D. (2019). *Media and information literacy in journalism. A handbook for journalists and journalism educators*. UNESCO. Retrieved from <https://bit.ly/3mVSvc0>
- Norris, P. & Inglehart, R. (2019). *Cultural Backlash: Trump, Brexit, and Authoritarian Populism*. Cambridge University Press.
- Oriola, O. & Kotze, E. (2020). Evaluating Machine Learning Techniques for Detecting Offensive and Hate Speech in South African Tweets. *IEEE Access*, 8, 21496-1509. <https://doi.org/10.1109/access.2020.2968173>
- Paz, M. A., Montero-Díaz, J., & Moreno-Delgado, A. (2020). Hate speech: A systematized review. *Sage Open*, 10(4), 1-12. <https://doi.org/10.1177/2158244020973022>
- Poletto, F., Basile, V., Sanguinetti, M. et al. (2021). Resources and benchmark corpora for hate speech detection: a systematic review. *Lang Resources & Evaluation*, 55, 477-523. <https://doi.org/10.1007/s10579-020-09502-8>
- Ramos, M. et al. (2021). *De los neocon a los neonazis. La derecha radical en el Estado español*. Fundación Rosa Luxemburgo. <https://bit.ly/3gXve9r>
- Rancière, J. (2021, January 22). Reflexiones sobre el final de la presidencia de Donald Trump. *Ctxt*. <https://bit.ly/3DJ0eU6>
- Stefanoni, P. (2021). *¿La rebeldía se volvió de derecha?* Siglo Veintiuno.
- Sousa Santos (de), B. (2010). *Descolonizar el saber, reinventar el poder*. Trilce.
- Tamayo, J.J. (2021). *La Internacional del odio*. Icaria.
- Tiburi, M. (2019). *¿Cómo conversar con un fascista? Reflexiones sobre el autoritarismo de la vida cotidiana*. Akal.
- Todorov, T. (2014). *El miedo a los bárbaros*. Galaxia Gutemberg.
- UNESCO (2021). *Addressing Hate Speech through education. World Conference of Ministers of Education, October 26, 2021*. UNESCO Education for World Citizenship and Peace Section. <https://bit.ly/3zGts4B>
- United Nations (2019). *United Nations Strategy and Plan of Action on Hate Speech*. United Nations. <https://bit.ly/3qVIPPu>
- United Nations (2020). *United Nations Guidance Note on Addressing and Countering COVID-19 related Hate Speech*. United Nations. <https://bit.ly/3BAZ1wA>
- Velasco, V. & Rodríguez Alarcón, L. (2020). *Nuevas narrativas migratorias para reemplazar los discursos del odio*. PorCausa. Investigación, periodismo y migraciones. <https://bit.ly/38z7TGp>
- Waldron, J. (2012). *The harm in Hate Speech*. Harvard Univer