ICONO 14

ICONO 14, Revista de comunicación y tecnologías emergentes, 20(2), 2022. ISSN: 1697-8293
https://doi.org/10.7195/ri14.v20i2.1858

Monográficos

Hate speech on Twitter during the Ceuta migration crisis in May 2021

El discurso de odio en Twitter durante la crisis migratoria de Ceuta en mayo de 2021

Discurso de ódio no Twitter durante a crise migratória de Ceuta, em Maio de 2021

Aránzazu Román-San-Miguel[1]*
Francisco J. Olivares-García[1]**
Salud María Jiménez-Zafra[2]***

[1] Department of Journalism II, Faculty of Communication, Universidad de Sevilla, Spain
[2] Department of Computer Science, Universidad de Jaén, Spain

* Associate Professor (Department of Journalism II). Faculty of Communication, Universidad de Sevilla, Spain
** Associate Professor (Department of Journalism II). Faculty of Communication, Universidad de Sevilla, Spain
*** Professor (Department of Computer Science). Universidad de Jaén, Spain

Received: 23/02/2022; Revised: 19/03/2022; Accepted: 29/06/2022; Published: 28/07/2022

Translation to English: ALTALINGUA, S.L.

To cite this article: Román-San-Miguel, Aránzazu; Olivares-García, Francisco J.; & Jiménez-Zafra, Salud María. (2022). Hate speech on Twitter during the Ceuta migration crisis in May 2021 *ICONO 14. Scientific Journal of Communication and Emerging Technologies, 20*(2). https://doi.org/10.7195/ri14.v20i2.1858

Abstract

This paper analyses hate speech within messages posted on Twitter from 17-25 May 2021 during the crisis caused by thousands of immigrants entering on the Tarajal border in Ceuta. The aim of the research was to group messages that included hate speech into themes. To do this, a mixed methodology was used, and resulted in six themes being identified. Of these, four focused on political issues, accounting for 80% of the messages, and only 20% referred to racism and immigration. Up to five campaigns of unknown origin have also been identified. This paper concludes that hate speech focuses more on political issues than on the problem of immigration itself, its causes, consequences and possible solutions.

Keywords: Hate speech; Twitter; Ceuta; social media; online campaigns; immigration

Resumen

Este trabajo analiza el discurso de odio en los mensajes publicados en Twitter desde el 17 al 25 de mayo de 2021 durante la crisis producida por la entrada de miles de inmigrantes en la frontera del Tarajal en Ceuta. El objetivo de la investigación es realizar una clasificación temática de los mensajes que incluyen discurso de odio. Para ello, se ha empleado una metodología mixta y como resultado se han podido diferenciar seis temas, de los cuales cuatro se centran en temas políticos, suponiendo el 80% de los mensajes, y solo el 20% de ellos se refieren a racismo e inmigración. Además, se han detectado hasta cinco campañas de origen desconocido. Este trabajo concluye que los discursos de odio se centran más en temas políticos que en la propia problemática de la inmigración, sus causas, sus consecuencias y las posibles soluciones.

Palabras clave: Discurso de odio; Twitter; Ceuta; redes sociales; campañas online; inmigración

Resumo

Este documento analisa o discurso do ódio nas mensagens publicadas no Twitter de 17 a 25 de Maio de 2021 durante a crise causada pela entrada de milhares de imigrantes na fronteira de Tarajal em Ceuta. O objectivo da investigação é levar a cabo uma classificação temática das mensagens que incluem o discurso do ódio. Para este fim, foi utilizada uma metodologia mista, e como resultado, foram identificados seis temas, quatro dos quais se concentram em questões políticas, representando 80% das mensagens, e apenas 20% das quais se referem ao racismo e à imigração. Além disso, foram detectadas até cinco campanhas de origem desconhecida. Este estudo conclui que o discurso do ódio se concentra mais em questões políticas do que no problema da imigração em si, nas suas causas, consequências e possíveis soluções.

Palavras-chave: discurso do ódio; Twitter; Ceuta; redes sociais; campanhas em linha; imigração

1. Introduction

Social media have allowed the use of freedom of expression to degenerate into abuse, creating the need for limits. So much so, that in recent years the media debate has focused on whether freedom of expression is at risk when certain speech - especially hate speech - is being fought in the courts. As Burgos-García (2019) argues, "If people behave anonymously and fail to adhere to critical, constructive thinking, the limits on freedom of expression are exceeded to the detriment of the rights of others" (p. 138). However, a distinction should be made between different forms of expression that may be rude, and hate crimes themselves, which lead the legal system to place limits on freedom of expression (p. 149).

The birth of social networks can be traced back to the beginning of the 21st century, although they did not become popular until years later, when the number of users and the functionalities of these platforms increased. Only a decade later, researchers such as Falxa (2014) began to study hate speech on social media, claiming that "the fight against hate speech on social media comes up against numerous obstacles" (p. 104). Although steps in this direction were already being taken in Europe, such as the decision taken on 24 January 2013 by the Judge of First Instance of Paris's Tribunal de Grande Instance, requiring US-based Twitter "to set up an alert mechanism allowing French users to report any abuse or hate messages for possible removal, and to provide the French authorities with details of any users who have posted illegal tweets" (p. 104). On 31 May 2016, the European Union required Facebook, Google, Microsoft and Twitter to agree to a code of conduct that would require them to review and remove within 24 hours any illegal hate speech posted on their services (Bisht et al., 2020, p. 244).

The scientific literature approaches hate speech from different areas of study - especially law (Howard, 2019; Rollnert-Liern, 2019; Bautista-Ortuño, 2017; Falxa, 2014). However, it is significant that studies are carried out as part of communication sciences - especially when there are those who consider social networks to be media (Hütt-Herrera, 2012; Pantoja-Chaves, 2011). This is a point of view that can be disputed. However, what is not disputed is that messages posted on social media reach millions of people every day and it is important to know whether, as channels of communication, they spread hate crimes.

On this, a recent study by Arcila-Calderón et al. (2022, p. 32) found that, "in European regions where there is a higher proportion of immigrants, they receive greater public support" and, therefore, "regions in which support was greater saw a lower level of hate speech on Twitter". However, as this paper shows, there is still hate crime on social networks such as Twitter which we should work to eradicate.

Some studies about social media that have been published have focused on hate crimes in regard to gender or sexual orientation - such as Arcila-Calderón et al. (2021a). The authors also present "the first prototype for automatically detecting hate speech on Twitter in Spanish, specifically were it is motivated by gender and sexual orientation".

This research is based on events in Spain on 17-18 May 2021, when thousands of people entered the country through the border with Morocco at El Tarajal (Ceuta). This prompted Spain’s Prime Minister, Pedro Sánchez, to visit the area on 18 May, and on the same day Reuters provided the media with an image of a Spanish Red Cross volunteer comforting a sub-Saharan immigrant, which generated numerous negative tweets.

The trigger for Morocco allowing thousands of people to enter Spain was the Polisario Front leader, Brahim Ghali, being admitted to a Spanish hospital on an Algerian passport, without prior notice being given to the Moroccan authorities, at a time of great tension, as in November 2020 heavy clashes had erupted between Morocco and the Polisario Front over control of Western Sahara.

All of this provoked major reaction on social networks, both in Spain and Morocco. Some users took advantage of the occasion to post xenophobic, violent, racist or hate speech messages.

Immigration has always been a focus for fomenting speech around rejection and hatred. Accordingly, we can cite studies on the hatred extolled by nationalism in the United States, with migrants being positioned as a threat (Morgenfeld, 2016); and Germany, where the Alternative for Germany (AfD) party - which calls itself anti-refugee and anti-immigration - became the third strongest party in the German parliament in 2017 (Valdez-Apolo et al., 2019, p. 365). The latter study concludes that the predominant tone on Twitter is negative, which the authors attribute to the fact that immigration is frequently related to "hostile events or circumstances of a bellicose, economic, political, demographic and/or climatic nature" (p. 379). However, it has the limitation that "the study was done by manually analysing content" (p. 385). As we will see below, this article has used a mixed method for the analysis. Another focus of hateful and violent publications is Islamophobia (Awan, 2014, Zamora et al., 2021).

1.1. Freedom of expression and social media

The fact that freedom of expression is a right does not mean that it is without a number of limitations that have been legally put in place over the years. These limitations are unrelated to the expression of ideas that are perceived as negative - as, if this were the case, pluralism, tolerance and broadmindedness in a democratic society would disappear (González-Herrera, 2018, p. 3). On this, the media are particularly protected as informers and opinion formers, even where this right has not always been exercised in an ethical manner. By way of example, González-Herrera cites the false information published in the UK during the campaign leading up to the referendum to leave the European Union and Donald Trump's 2016 presidential election. According to some studies, the US president had shown himself to be 'an inveterate liar, immune and defiant to fact-checking' (Echevarría, 2016, p. 10).

It was not the media, but social networks that spread the most fake news during the US Republican candidate's campaign, which was accentuated during the Covid-19 pandemic - fake news that was then passed on to the media that used social networks as sources of information (Fernández-Muñoz et al., 2022; Gutiérrez-Vidrio, 2020; Pérez-Curiel and Limón-Naharro, 2019). Indeed, social media and especially Twitter had to intervene to curb Donald

Trump's misuse of the platform, especially after the assault on the US Capitol on 6 January 2021 and his false messages, if not inciting hatred and violence. On 8 January, Twitter shut down Donald Trump's account and later Reddit shut down the account of Trump supporters for promoting hateful behaviour, followed by Twitch shutting down Trump's account.

On Twitter's help website there is an entire section dedicated to the company's policies where hate speech is detected in posted content (Twitter, 2021). During the course of this investigation, several accounts were found to have been deleted for posting hate speech content on a number of occasions.

1.2. Hate speech on Twitter

Hate is deemed to be reprehensible conduct in the moral, religious, social and legal spheres, as it can harm rights such as a person or group's honour or dignity (Moretón-Toquero, 2012, p. 4). As early as 2012, an attempt was made to give a name to hate-motivated behaviour on the internet, which Moretón-Toquero (2012) called "cyberhate" - even though, in this study, the internet is referred to in a very generic way, without studying the impact of social media. What it highlighted was the difficulty of legislating in such an open and international domain like the internet, but also the fear that trivialising this issue in a medium accessed by young people as could be aroused by a more influential public (p. 15).

To define which material is racist and xenophobic, we referred to the Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of racist and xenophobic acts committed through computer systems, which was agreed in Strasbourg on 28 January 2003 and published in Spain's Official Government Gazette Number 26, Section 1 on 30 January 2015:

> "Racist and xenophobic material" means any written material, images or any other representation of ideas or theories that advocates, promotes or incites hatred, discrimination or violence, against any person or group of persons, on the grounds of race, colour, descent or national or ethnic origin, as well as religion insofar as religion is used as a pretext for any of these factors. (p. 7.216)

The Hate Crime Survey Report for the period December 2020 to March 2021, by the Secretary of State for Security of Spain's Home Office (López-Gutiérrez et al., 2021), classifies hate crime as: racism/xenophobia, sexual orientation/ LGBTphobia, religion, ideology and disability/illness. All these manifestations of hate could be found among the messages downloaded from Twitter for this project.

According to a report published by Barcelona City Council, as early as 2014 there were around 10,000 tweets a day that included racist insults in English - equivalent to one in every

15,000 tweets (Cabo-Isasi and García-Juanatey 2017, p. 5). Years later, the media would echo Twitter's decision to delete more than a thousand tweets in which racist abuse was detected and it was stated that the British government was working on laws in this area (Cano, 2021).

We found some recent studies about hate speech on Twitter, such as that of Valdez-Apolo et al. (2019), who claim that the prevailing sentiment in messages about immigrants is negative, and more so than about refugees. Those authors consider that this "negative tone" may have its origin in the fact that "the migration phenomenon" in the media is "negative in nature", as it is frequently "related to hostile events or circumstances of a bellicose, financial, political, demographic and/or climatic nature" (p. 379). They also address the stigmatisation of refugees through hate messages on social media Frías and Seoane (2019), and Arcila-Calderón et al. (2021b).

This is something that can be seen in a latent form in the study at hand, as the massive entry of immigrants into a country - and encouraged by a neighbouring country's authorities' laxity - could be seen by the receiving population as a threat to its integrity as a country. For, although online hate speech might initially be thought to cause no harm due to its *virtual* nature, studies by Williams et al. (2020) have shown a link between hate speech on social media and actual crime - especially against black and Muslim communities.

Other studies have approached the refugee issue by disaggregating the study by gender, such as Amores et al. (2020), who conclude that the image of refugee women is as passive agents in the images published by the media.

Political discourse has also been related in different studies to hate messages via Twitter (Arcila-Calderón et al., 2021c; Arcila-Calderón et al., 2020; Casero-Ripollés et al., 2020; Paz, 2021).

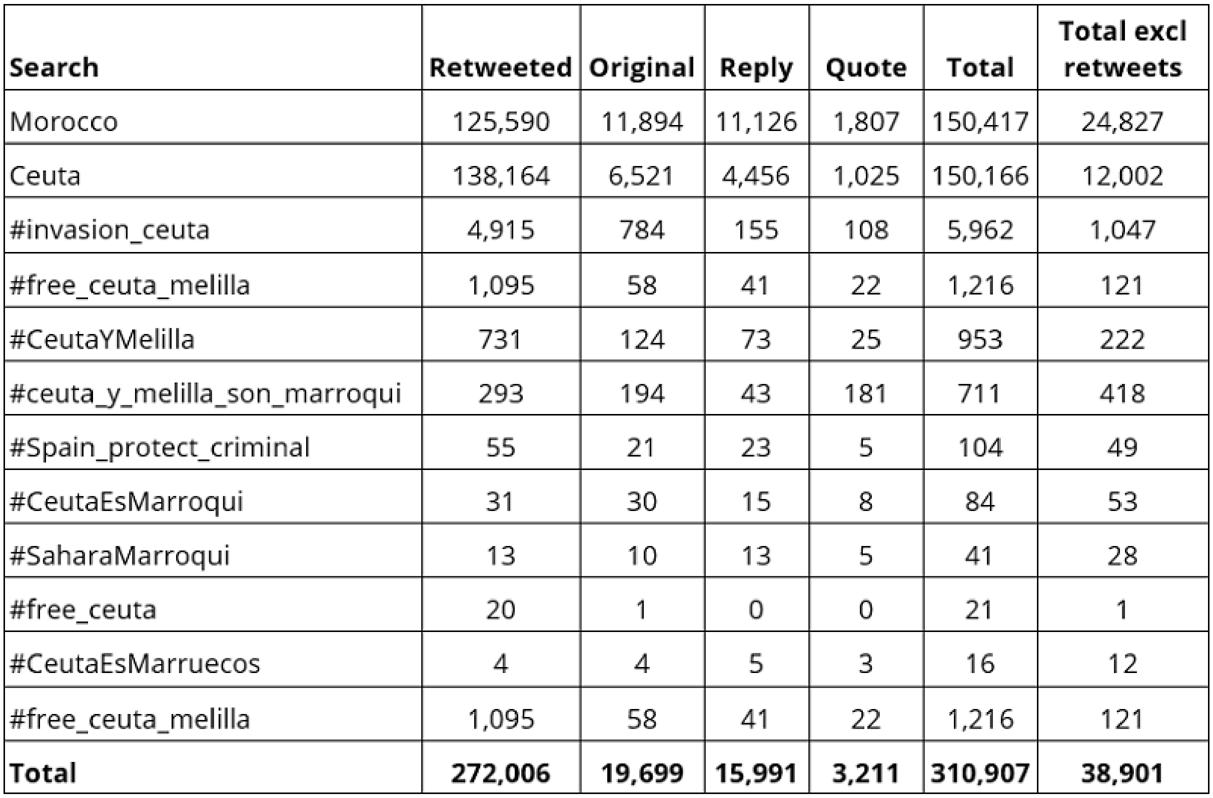
2. Material and methods

A mixed methodology was used for this project - very similar to that seen in other projects related to hate speech and Twitter. For example, in Miró-Linares (2016), in which both quantitative analysis with data mining techniques and qualitative analysis of content are used. Accordingly, techniques that automatically filter Twitter content based on the use of specific lexicons have also been used - as can be seen in Jiménez-Zafra et al. (2021).

First, all messages published in Spanish between 17-25 May 2021 containing the words Ceuta and/or Morocco were downloaded using Twitter's API V2. An additional download was also done based on a series of hashtags, as shown in Table 1. The list of tags was obtained using the Trendinalia, Account Analysis and Google Trends websites, based on the topics that most interested users and the tags used in the messages. This was then all converted into a spreadsheet.

A total of 310,907 tweets were downloaded. Many of them were retweeted posts - that is, original posts that other people simply copied and pasted as they were via their own accounts, without contributing any content. Thus, these messages needed to be eliminated, leaving 38,901 posts at the end of this process. Table 1 shows the number of messages retrieved according to keyword and tag classification:

Table I. Complete list of downloaded messages organised by keyword and tag



Search	Retweeted	Original	Reply	Quote	Total	Total excl retweets
Morocco	125,590	11,894	11,126	1,807	150,417	24,827
Ceuta	138,164	6,521	4,456	1,025	150,166	12,002
#invasion_ceuta	4,915	784	155	108	5,962	1,047
#free_ceuta_melilla	1,095	58	41	22	1,216	121
#CeutaYMelilla	731	124	73	25	953	222
#ceuta_y_melilla_son_marroqui	293	194	43	181	711	418
#Spain_protect_criminal	55	21	23	5	104	49
#CeutaEsMarroqui	31	30	15	8	84	53
#SaharaMarroqui	13	10	13	5	41	28
#free_ceuta	20	1	0	0	21	1
#CeutaEsMarruecos	4	4	5	3	16	12
#free_ceuta_melilla	1,095	58	41	22	1,216	121
Total	**272,006**	**19,699**	**15,991**	**3,211**	**310,907**	**38,901**

Source: Author's own

A second filtering was then performed to remove duplicate entries, as a tweet can include several keywords and tags at the same time. After the new filtering process, the number of publications was reduced to 36,887. In total, 18,441 original messages were identified, the rest being those that were quoted (2,932) or replied to (15,514).

For the quantitative analysis, four specific lexicons were used that can be helpful in detecting hate speech. These lexicons have been satisfactorily evaluated in the study by Plaza-del-Arco et al. (2020). The number of words contained in each lexicon along with a couple of significant examples are shown below:

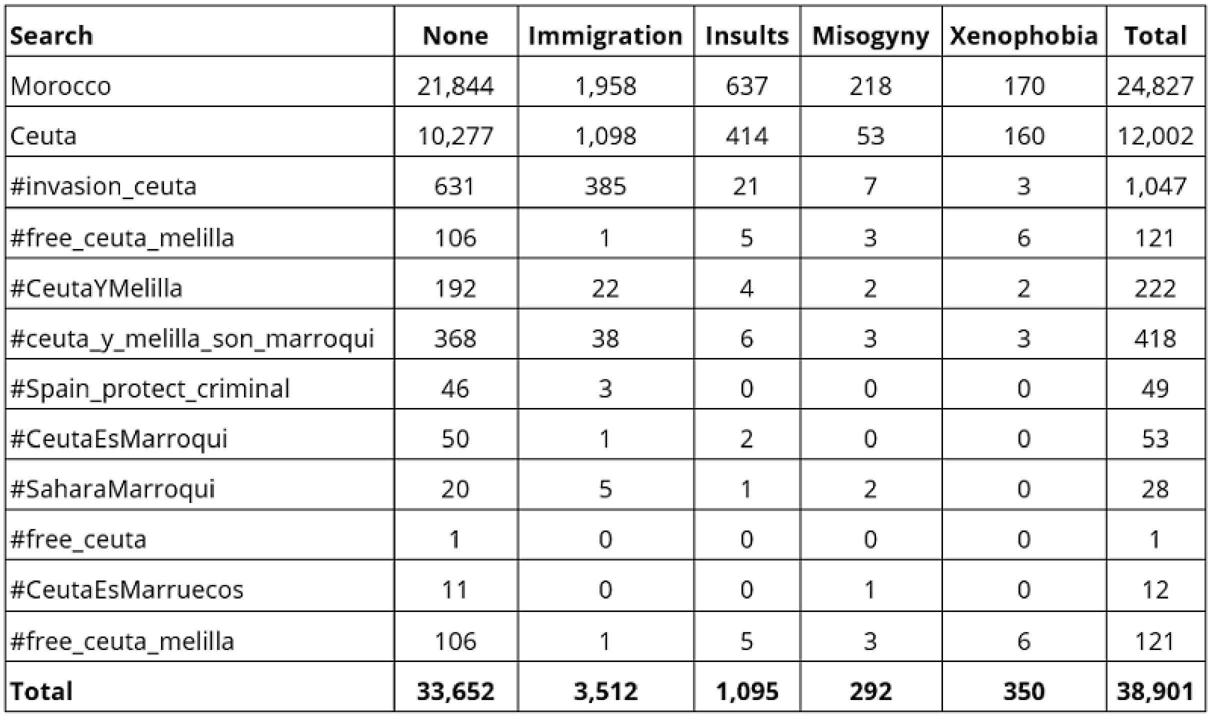
Xenophobia: 44 words (e.g. Moor/Muslim), nigger)
Immigration: 250 words (e.g. deport him, refugee)
Misogyny: 183 words (e.g. bitch, slut)
Insults: 279 (e.g. bastard, faggot)

To facilitate the analysis, all information was organised in a spreadsheet consisting of 18 fields. The first ten fields were directly related to the information extracted via the Twitter API; the other eight were created based on the content provided by the four lexicons. Each of these lexicons generated two fields in the spreadsheet: one indicating the presence (or not) of words included within them and the other containing the terms that were found:

Id: Unique identifier for the tweet.
Text: Message written in the tweet.
Topic: Field automatically generated based on use of the lexicon filter, which classifies messages according to whether they deal with immigration, xenophobia, insults or misogyny. The result of this classification can be seen in Table 2.
Type: Original, reply or quote, depending on whether the message is an original post or whether it is a comment on or quoted from another.
Sensitive content: Field provided automatically by Twitter, showing either true or false.
Retweet count: Number of times an original post has been retweeted.
Comment count: Total number of comments received for each tweet.
Quoted count: Number of times a user has quoted a message.
Likes count: Total number of users who have liked a message.

These four fields have been used to assess the popularity of tweets.
Source: The origin of the publication, depending on whether it was tweeted from a mobile or desktop application.
Total number of words on immigration: Number of words in the tweet that are included in the immigration lexicon.
Number of words on immigration: Number of words from the immigration lexicon that are used in the tweet.
Total number of insults: Number of lexicon insults used in the tweet.
Insults: Words from the lexicon related to insults and used in the tweet.
Total number of misogynistic words: Number of words from the misogyny lexicon that are used in the tweet.
Misogynistic words: Words from the misogyny lexicon that are used in the tweet.
Total number of xenophobic words: Number of xenophobic words used in the tweet.
Xenophobic words: Words from the xenophobia lexicon that are used in the tweet.

Table II. Themed classification based on the lexicon filters



Search	None	Immigration	Insults	Misogyny	Xenophobia	Total
Morocco	21,844	1,958	637	218	170	24,827
Ceuta	10,277	1,098	414	53	160	12,002
#invasion_ceuta	631	385	21	7	3	1,047
#free_ceuta_melilla	106	1	5	3	6	121
#CeutaYMelilla	192	22	4	2	2	222
#ceuta_y_melilla_son_marroqui	368	38	6	3	3	418
#Spain_protect_criminal	46	3	0	0	0	49
#CeutaEsMarroqui	50	1	2	0	0	53
#SaharaMarroqui	20	5	1	2	0	28
#free_ceuta	1	0	0	0	0	1
#CeutaEsMarruecos	11	0	0	1	0	12
#free_ceuta_melilla	106	1	5	3	6	121
Total	**33,652**	**3,512**	**1,095**	**292**	**350**	**38,901**

Source: Author's own

Following this automated organisation, the content was analysed manually by reading the messages marked by the lexicon theme filter, along with all the messages with hashtags. Thus, of the 36,887 tweets downloaded without duplicates, a total of 7000 tweets suspected of containing hate speech were manually reviewed. Of these, 4964 include one of the words in the lexicons, while 1817 relate to the sum of all messages that included one of the hashtags. The fact that a post has a word from a lexicon does not imply that the tweet is violent, just as there are messages that are not marked by lexicons that include hateful content. To find the largest number of these messages, an additional search for the words war, invasion, border, army, humiliation, death and expulsion was done. This added 219 messages to those analysed, bringing the total to 7000. These messages were manually reviewed and 1000 tweets were found that actually included hate speech. Among the rest, although words were used that could indicate hatred, they were not in the context used. The exact number was 1005, but was reduced to 1000 to simplify the calculations.

3. Findings and discussion

Based on the analysis of the content of the messages, they were classified into six themed areas. This classification does not take account of messages that form part of organised campaigns, which are discussed in section 3.2.

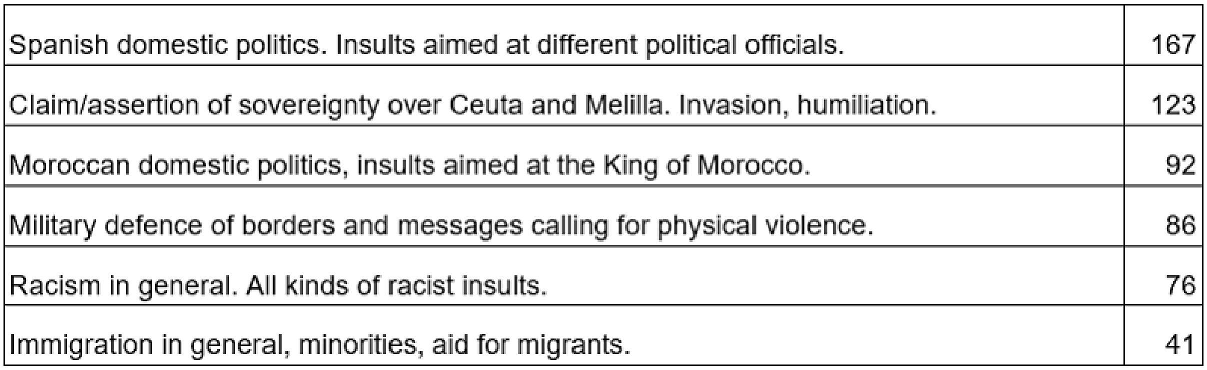
3.1. Classification by theme

Although the entire collection of tweets is in Spanish, many messages from Moroccan users were found that were posted in Spanish with the intention of insulting Spanish users or making claims with violent language that Spanish citizens could read. However, whilst most of the hate messages found expressed violence against Morocco, the study did not distinguish between hateful or violent messages from each other.

This example shows a tweet in Spanish referring to Ceuta and Melilla as occupied cities and encouraging the creation of terrorist groups:

> The inhabitants of the occupied cities and neighbouring cities should destabilise the Spaniards there and create an organised group like the Basque Eta. #########_######_#### #ceuta_y_melilla_son_marroqui

Table III. List of identified topics and number of messages



Spanish domestic politics. Insults aimed at different political officials.	167
Claim/assertion of sovereignty over Ceuta and Melilla. Invasion, humiliation.	123
Moroccan domestic politics, insults aimed at the King of Morocco.	92
Military defence of borders and messages calling for physical violence.	86
Racism in general. All kinds of racist insults.	76
Immigration in general, minorities, aid for migrants.	41

Source: Author's own

While analysing the sample of 1000 selected messages containing violent language or hate speech, six themes were found into which the content could be grouped (Table 3). One message might form part of several categories.

Spanish domestic politics. Insults aimed at different political officials: It is by far the most repeated topic, with 167 publications. There are messages from both Spaniards and Moroccans. The arrival of VOX leader Santiago Abascal in Ceuta on 24 May marks a turning point in the crisis, which has gone from being a migration crisis to a national political crisis in which the right-wing opposition parties are facing off against the left-wing parties in government: PSOE and Unidas Podemos.

A large part of the violent content is related to VOX's events organised in Ceuta, Prime Minister Pedro Sánchez's visit to Brussels and the attitude of Spanish government ministers.

It is worth noting the large number of insults in these messages, towards both right- and left-wing leaders.

Claim/assertion of sovereignty over Ceuta and Melilla. Invasion/occupation, humiliation: With 123 tweets, this topic is the most directly related to the events: The arrival of a large number of people from Morocco to Spain through the El Tarajal border in Ceuta. The messages can be divided into two types: those claiming and asserting Spanish or Moroccan sovereignty over the cities of Ceuta and Melilla, and those complaining of a humiliating invasion/occupation. Although they do not call for a violent military response, but often blame the situation on government policies. There are messages in this section from both Spaniards and Moroccans.

Moroccan domestic politics, insults aimed at the King of Morocco: In this case, there are no messages written by Moroccans against their own government, as we saw in the Spanish case, where messages were seen against the Spanish government by both Spaniards and Moroccans.

Ninety-two messages were classified under this theme, which proved to contain some of the most violent content and hate speech: Insults, xenophobia, LGBTphobia and a wide range of accusations against the King of Morocco.

Military defence of borders and messages calling for physical violence: Another of the study's most violent sections calls for a military presence and proposes a violent war scenario against Morocco or Spain. There are 86 messages in this category, which are divided between posts by Moroccans and Spaniards. Most of the messages with violent expressions, posted by Moroccan users talk about occupied territories, stolen lands and liberation, with reference to the Spanish cities of Ceuta and Melilla.

A very specific migration crisis, which has nothing to do with occupations or liberations, ends up becoming a declaration of war against the neighbouring country on both sides on social media. While the government treats migrants as refugees, on social media they are viewed as liberators or invaders depending on the nationality of the author of the messages.

Racism in general. All kinds of racist insults: There was such a high quantity of messages with racist or xenophobic content that a category was especially created. With a total of 76 posts, this category does not have the most posts, but is notable because insults is almost all that these posts contain.

Immigration in general, minorities, aid for migrants: Only 41 violent or hate speech messages appear in this category. It is one of the most frequently repeated topics in the collection of 36,887 messages analysed. However, it is not the one in which most insults are hurled.

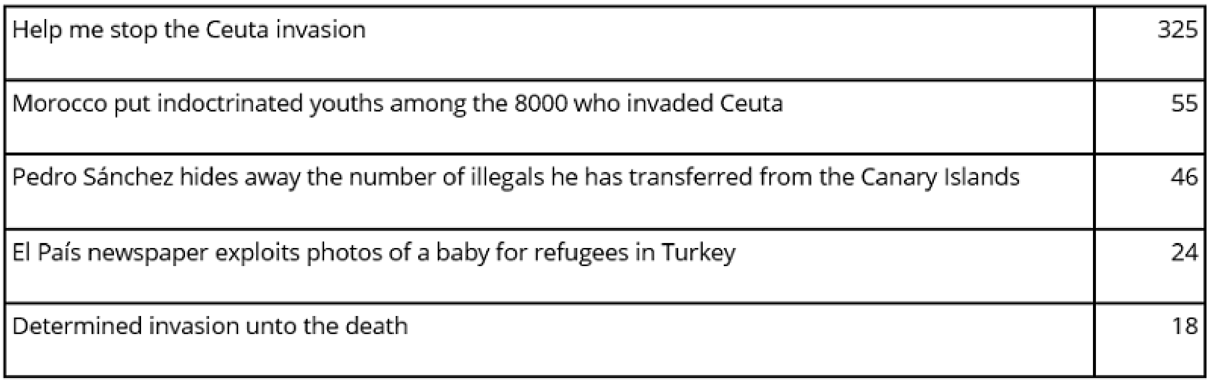
Most of the tweets found in this category are violent messages about immigration or include lies simply to discredit and criminalise Moroccan refugees or unaccompanied minors who have arrived in Spain.

3.2. Campaigns

After analysing the extracted data, it was noted that a large proportion of the violent publications are part of campaigns in which users, in an organised or spontaneous manner, publish the same message on a massive scale with the clear intention of influencing public opinion. Five such campaigns have been identified and are discussed below. They are so important that they account for 53.2% of the content of the sample.

A summary table of the number of publications collected in each of the campaigns is shown below:

Table IV. Publications for each campaign



Help me stop the Ceuta invasion	325
Morocco put indoctrinated youths among the 8000 who invaded Ceuta	55
Pedro Sánchez hides away the number of illegals he has transferred from the Canary Islands	46
El País newspaper exploits photos of a baby for refugees in Turkey	24
Determined invasion unto the death	18

Source: Author's own

Help me stop the Ceuta invasion. Tell Marlaska to authorise the immediate expulsion of illegals in Ceuta NOW #Ceutainvasion https://t.co/U4smcLJQtB

The link leads to a petition on the hazteoir.org website, created on 18 May 2021, the aim of which was to get 50,000 signatures so that the Home Secretary would authorise the immediate deportation of all migrants described as illegal in the petition. The petition refers to the arrival of 10,000 invaders in the city of Ceuta and at all times maintains a violent tone against both the government and the refugees in Ceuta.

Although the number of signatures reached only 25,000, the campaign was a success on Twitter, as the petition was retweeted 325 times in the 1000 messages in the sample. This gives us an idea of how widespread it was.

Morocco put indoctrinated youths among the 8000 who invaded Ceuta. With 55 messages containing the same text, it ranks second in the list of campaigns identified. The article was published on 24 May in *El Español* by journalist Sonia Moreno.

This news can also be found in other media, but without the nuance of it having been a voluntary act by Morocco, as the information in *El Español* seems to indicate.

Pedro Sánchez hides away the number of illegals he has transferred from the Canary Islands. Although this issue has nothing to do with the news about the arrival of people at the Moroccan border in Ceuta, its publication in *OK Diario* on 24 May caused a controversy - especially when the country's president was personally accused of concealing information. The very wording of the news item in which it describes people as illegal is itself biased writing whose objective goes beyond information. The full text of the messages collected on Twitter does not provide much more than the title.

El País newspaper exploits photos of a baby for refugees in Turkey. Most of the messages published in this clear disinformation campaign are published in Arabic; only a small part - the 24 picked up in the sample - are in Spanish.

The photo to which it refers shows the rescue of a baby at sea by a member of Spain's diving unit, GEAS. This photo was published in hundreds of media around the world, while a group of Moroccan users were spreading the word that the photo is not about Ceuta, but is in fact part of a rescue in Turkey several months ago. The photo went viral in Spain and generated thousands of interactions across all social networks. Although almost all Spain's media echoed this information, the smear campaign focuses on *El País* newspaper. The link leads to a Twitter conversation with the photo and overlaid with Arabic text that reads:

Spain's *El País* attributes the image of a baby off the Turkish coast to the events in Ceuta. Spanish newspaper *El País* publishes an old photo of earlier events in Turkey and attributes them to events in the border area of the occupied cities of Ceuta and Melilla.

The original campaign in Arabic aims to discredit the Spanish media in order to preserve its public image with Moroccan citizens. The campaign in Spanish, made for those who know the news, seeks to promote violence by insulting the Spanish media. In all cases the content ends up creating a violent climate within the conversation, with lies and insults.

This hoax was debunked by verification platform Newtral, which publishes on its website the ways in which it searched for sources and verified the authenticity of the photograph (Maqueda, 2021).

Determined invasion unto the death is the last of the campaigns identified. In this case, with 18 messages published in Spanish. Again, these are publications that originate in Morocco and encourage a violent response to the conflict among Moroccans. The campaign always mentions Spanish media and political parties at least once - mainly VOX and its president Santiago Abascal.

The text in the messages usually includes the hashtag #Ceuta and Melilla and are in Moroccan.

4. Conclusions

By applying data mining techniques, a collection of 310,907 tweets was put together which were then filtered as explained in the methodology, leaving 1000 messages to be analysed. This enabled the research project's main objective to be achieved - namely, to classify, by theme, messages that include hate speech and analyse some of the main trends that were identified during the research.

Interestingly, the topic that appears most often in tweets containing hate speech is Spanish domestic politics, with insults against different political officials. This suggests that a political campaign has been introduced under the cover of freedom of expression and the ease with which information can be disseminated through social media.

The second most recurrent issue is that of the sovereignty of Ceuta and Melilla. To a certain extent, this also falls within the political sphere, as it is the defence of two Spanish autonomous cities on African soil that have had an impact over the years on Spain's international relations with Morocco.

Moroccan domestic politics and insults to King Mohammed VI also feature, surprisingly, in this analysis of hate speech. This is a very thorny issue for Moroccans living in their country of origin, which is why we understand that it would be included in a study of tweets in Spanish and very unlikely to be included in a sample in Arabic.

Military defence of borders and messages calling for physical violence may also have to do with political and ideological issues - an issue that is closely related to sovereignty over Ceuta and Melilla.

Of course, there are also racist insults and a strong sign of racism generally within the study, as groups that follow this ideology take every opportunity to post their messages and rabble-rousing speech.

It is of particular note in regard to the subject studied - immigration, support for immigrants and unaccompanied minors who enter Spain illegally - that this is the theme that recurs least among the messages analysed. This may lead to the conclusion that the event being studied was not seen as a major migration problem but, rather, as a political problem.

Hence, 53.2% of the content of the sample is related to well-orchestrated campaigns on social media, and this once again leads to a thematic relationship being established with politics: stopping the invasion of Ceuta; the entry of young people indoctrinated by Morocco; the Spanish President's alleged concealing of illegal immigrants (the term 'illegal' is used in the tweets) being transferred from the Canary Islands - which has nothing to do with the event that was taking place on the border with Morocco; the circulation of the hoax that the photograph of the civil guard rescuing a baby was false; and finally, and perhaps the harshest in terms of content: invasion unto the death. The latter messages, disseminated by Moroccans in Spanish, encourage violent uprising against Spain, and frequently mention VOX and its leader, Santiago Abascal.

Hate messages have therefore flowed by introducing more political issues into Twitter's conversations than the problem of immigration itself, its causes, its consequences, and the possible solutions to a drama that leaves hundreds of people dead at sea every year and crippled on the fences that separate Spain from Morocco.

Authors' contribution

Aránzazu Román-San-Miguel: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Software, Writing original draft and Writing - review & editing. **Francisco J. Olivares-García**: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Supervision, Software and Writing - review & editing**. Salud María Jiménez-Zafra**: Investigation, Methodology, Resources, Visualization, Software, Writing original draft and Writing - review & editing.

All authors have read and accepted the published version of the manuscript. Conflicts of interest: the authors confirm that they have no conflict of interest.

Acknowledgments

This publication is part of the LIVING-LANG project (RTI2018-094653-B-C21), funded by MCIN/AEI/10.13039/501100011033 and ERDF A Way to Make Europe as part of the project PID2020-116118GA-I00, funded by MICINN/AEI/10.13039/501100011033, of the project PID2020-119478GB-I00, funded by MICINN/AEI/10.13039/501100011033, and the projects Big Hug (P20_00956, PAIDI 2020) and WeLee (1380939, ERDF Andalucía 2014-2020), funded by the Regional Assembly of Andalusia. Salud María Jiménez-Zafra has been funded by the European Social Fund and the Regional Assembly of Andalusia (DOC_01073).

References

Amores, Javier J., Arcila-Calderón, Carlos, & González-de-Garay, Beatriz (2020). The gendered representation of refugees using visual frames in the main western european media. *Gender Issues*, 37(4), 291-314. https://doi.org/10.1007/s12147-020-09248-1

Arcila-Calderón, Carlos, Amores, Javier J., Sánchez-Holgado, Patricia, & Blanco-Herrero, David (2021a). Using shallow and deep learning to automatically detect hate motivated by gender and sexual orientation on twitter in spanish. *Multimodal Technologies and Interaction*, 5(10) https://doi.org/10.3390/mti5100063

Arcila-Calderón, Carlos, Blanco-Herrero, David, Frías-Vázquez, Maximiliano, & Seoane-Pérez, Francisco (2021b). Refugees welcome? online hate speech and sentiments in twitter in spain during the reception of the boat aquarius. *Sustainability (Switzerland)*, 13(5), 1-17. https://doi.org/10.3390/su13052728

Arcila-Calderón, Carlos, Blanco-Herrero, David, Matsiola, María, Oller-Alonso, Martín, Saridou, Theodora, Splendore, Sergio, & Veglis, Andreas (2021c). Framing migration in southern european media: Perceptions of spanish, italian, and greek specialized journalists. *Journalism Practice*, https://doi.org/10.1080/17512786.2021.2014347

Arcila-Calderón, Carlos, de la Vega, Gonzalo, & Blanco Herrero, David (2020). Topic modeling and characterization of hate speech against immigrants on twitter around the emergence of a far-right party in Spain. *Social Sciences*, 9(11), 1-19. https://doi.org/10.3390/socsci9110188

Arcila-Calderón, Carlos, Sánchez-Holgado, Patricia, Quintana-Moreno, Cristina, Amores, Javier J., & Blanco-Herrero, David (2022). Hate speech and social acceptance of migrants in europe: Analysis of tweets with geolocation. [Discurso de odio y aceptación social hacia migrantes en Europa: Análisis de tuits con geolocalización] *Comunicar*, 30(71), 1-13. https://doi.org/10.3916/C71-2022-02

Awan, Imran (2014). Islamophobia and Twitter: A typology of online hate against Muslims on social media. *Policy & Internet*, 6(2), 133-150. https://doi.org/ggb54f

Bautista-Ortuño, Rebeca (2017). ¿Eres un ciberhater? Predictores de la comunicación violenta y el discurso del odio en Internet. *International e-Journal of Criminal Science*, 11, 1-28. https://bit.ly/3kiYLfL

Bisht, Akanksha, Singh, Annapurna, Bhadauria, H. S., Virmani, Jitendra, & Kriti (2020). Detection of Hate Speech and Offensive Language in Twitter Data Using LSTM Model en Jain, Shruti, & Paul, Sudip (coord.), *Recent Trends in Image and Signal Processing in Computer Vision* (pp. 243-264). Springer. https://doi.org/hft4

Burgos-García, Olga (2019). Los límites a la libertad de expresión: el discurso del odio en Internet en Marín-Conejo, Sergio (ed.). *El mundo a través de las palabras. Lenguaje, género y comunicación* (pp. 138-150). Dykinson, S.L. https://doi.org/gwbs

Cabo-Isasi, Alex, & García-Juanatey, Ana (2017). *El discurso de odio en las redes sociales: un estado de la cuestión*. Ayuntamiento de Barcelona. https://bit.ly/3Bt92vb

Cano, Manuela (12 de julio de 2021). Investigan ataques racistas a jugadores ingleses tras la final de la Eurocopa. *France 24*. https://bit.ly/2WrV3Hq

Casero-Ripollés, Andreu, Micó-Sanz, Joseph-Lluís., & Díez-Bosch, Míriam (2020). Digital public sphere and geography: The influence of physical location on twitter's political conversation. *Media and Communication*, 8(4), 96-106. https://doi.org/10.17645/mac.v8i4.3145

Echevarría, Borja (2016). Más 'fact-checking' contra la posverdad. *Cuadernos de periodistas: Revista de la Asociación de la Prensa de Madrid*, 33, 9-16. https://bit.ly/3gyBgwL

BOE (30 de enero de 2015). Instrumento de Ratificación del Protocolo adicional al Convenio sobre la Ciberdelincuencia relativo a la penalización de actos de índole racista y xenófoba cometidos por medio de sistemas informáticos, hecho en Estrasburgo el 28 de enero de 2003. BOE, n. 26, sec. I, 7214-7224. https://bit.ly/3mByWsQ

Falxa, Joana (2014). Redes sociales y discursos de odio: un enfoque europeo en Pérez Álvarez, Fernando (coord.). *Moderno discurso penal y nuevas tecnologías* (pp. 89-106). Ediciones Universidad de Salamanca.

Fernández-Muñoz, Cristóbal, Rubio-Moraga, Ángel Luis, & Álvarez-Rivas, David (2022). The Multiplier Effect on the Dissemination of False Speeches on Social Networks: Experiment during the Silly Season in Spain. In: Lahby, Mohamed, Pathan, Al-Sakib Khan, Maleh, Yassine, & Yafooz, Wael Mohamed Shaher. (eds) Combating Fake News with Computational Intelligence Techniques. Studies in Computational Intelligence, vol 1001. Springer, Cham. https://doi.org/10.1007/978-3-030-90087-8_12

Frías Vázquez, Maximiliano, & Seoane Pérez, Francisco (2019). Hate Speech in Spain Against Aquarius Refugees 2018 in Twitter. Paper presented at the ACM International Conference Proceeding Series, 906-910. https://doi.org/10.1145/3362789.3362849

González-Herrera, Daniel (2018). Libertad de expresión y discurso de odio en Europa: protegiendo a las minorías en tiempos de posverdad en Rodríguez-García, Nicolás, Carrizo-González-Castell, Adán, & Leturia-Infante, Francisco J. (coord.) *Justicia Penal Pública y Medios de Comunicación* (pp. 549-573). Tirant lo Blanch.

Gutiérrez-Vidrio, Silvia (2020). El discurso político en la era digital. Donald Trump y su uso de Twitter. *Estudios del discurso*, 6(1), 56-81. https://bit.ly/3lByClm

Howard, Jeffrey W. (2019). Free Speech and Hate Speech. *Annual Review of Political Science*, 22, 93-109. https://doi.org/ghrr2v

Hütt-Herrera, Harold (2012). Las redes sociales: Una nueva herramienta de difusión. *Reflexiones*, 91(2), 121-128. https://bit.ly/3EyaONC

López-Gutiérrez, Javier, Fernández-Villazala, Tomás, Máñez-Cortinas, Carlos, San-Abelardo-Anta, María Yamir, Gómez-Esteban, Jesús, Sánchez-Jiménez, Francisco, Herrera-Sánchez, David, Martínez-Moreno, Francisco, Rubio-García, Marcos, Gil-Pérez, Victoria, Santiago-Orozco, Ana María, & Gómez-Martín, Miguel Ángel (2021). *Informe de la encuesta sobre delitos de odio.* Secretaría General Técnica del Ministerio del Interior, NIPO: 126 21 071 6. https://bit.ly/3nMbMyS

Jiménez-Zafra, Salud María, Sáez-Castillo, Antonio José, Conde-Sánchez, Antonio, & Martín-Valdivia, María Teresa (2021). How do sentiments affect virality on Twitter? *Royal Society Open Science*, 8(4), 1-11. https://doi.org/gwdf

Maqueda, Adrián (19 de mayo de 2021). Ni es en Turquía ni es antigua: la imagen del guardia civil en Ceuta rescatando a un bebé es real. *Newtral*. https://bit.ly/2XoGqoG

Miró-Linares, Fernando (2016). Taxonomía de la comunicación violenta y el discurso del odio en Internet. *Revista d'Internet, Dret i Política*, 22, 93-118. https://doi.org/g4b6

Moretón-Toquero, María Aránzazu (2012). El ciberodio, la nueva cara del mensaje de odio: entre la cibercriminalidad y la libertad de expresión. *Revista jurídica de Castilla y León*, 27, 1-18. https://bit.ly/3DwlLQb

Morgenfeld, Leandro (2016). Estados Unidos: Trump y la reacción xenófoba contra la inmigración hispana. *Conflicto Social*, 9(16), 15-33. https://bit.ly/3pXix3B

Pantoja-Chaves, Aantonio (2011). Los nuevos medios de comunicación social: las redes sociales. *Tejuelo*, 12, 218-226. https://bit.ly/2ZAa6QD

Paz, María Antonia, Mayagoitia-Soria, Ana, & González-Aguilar, Juan Manuel (2021). From polarization to hate: Portrait of the spanish political meme. *Social Media and Society*, 7(4) https://doi.org/10.1177/20563051211062920

Pérez-Curiel, Concha, & Limón-Naharro, Pilar (2019). Political influencers. A Study of Donald Trump's personal brand on Twitter and its impact on the media and users. *Communication & Society*, 32(1), 57-75. https://doi.org/gwdg

Plaza-del-Arco, Flor Miriam, Molina-González, M. Dolores, Ureña-López, L. Alfonso, & Martín-Valdivia, M. Teresa (2020). Detecting Misogyny and Xenophobia in Spanish Tweets Using Language Technologies. *ACM Transactions on Internet Technology (TOIT)*, 20(2), 1-19. https://doi.org/gwhv

Rollnert-Liern, Göran (2019). El discurso del odio: una lectura crítica de la regulación internacional. *Revista Española de Derecho Constitucional,* 115, 81-109. https://doi.org/gwdh

Twitter (2021). Política relativa a las conductas de incitación al odio. *Twitter*. https://bit.ly/3w03uY3

Valdez-Apolo, María Belén, Arcila-Calderón, Carlos, & Jiménez-Amores, Javier (2019). El discurso del odio hacia migrantes y refugiados a través del tono y los marcos de los mensajes en Twitter. *RAEIC, Revista de la Asociación Española de Investigación de la Comunicación*, 6(12), 361-384. https://doi.org/dxj2

Williams, Matthew L., Burnap, Pete, Javed, Amir, Liu, H.an,& Ozalp, Sefa (2020). Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime. *The British Journal of Criminology*, 60(1), 93-117. https://doi.org/c9qh

Zamora Medina, Rocío, Garrido Clemente, Pilar & Sánchez Martínez, Jorge (2021). Analysis of hate speech involving islamophobia on twitter and its social repercussion in the case of the campaign "Remove the labels from the veil". *Anàlisi: Quaderns de Comunicació i Cultura,* 65, 1-19. https://doi.org/10.5565/REV/ANALISI.3383