ICONO14

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

Monográficos

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

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

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

Aránzazu Román-San-Miguel¹*© Francisco J. Olivares-García¹**© Salud María Jiménez-Zafra²***©

Recibido: 23/02/2022; Revisado: 19/03/2022; Aceptado: 29/06/2022; Publicado: 28/07/2022

Para citar este artículo: Román-San-Miguel, Aránzazu; Olivares-García, Francisco J.; & Jiménez-Zafra, Salud María. (2022). El discurso de odio en Twitter durante la crisis migratoria de Ceuta en mayo de 2021. *Revista ICONO 14. Revista Científica De Comunicación Y Tecnologías Emergentes*, 20(2). https://doi.org/10.7195/ri14.v20i2.1858

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



¹ Departamento de Periodismo II, Facultad de Comunicación, Universidad de Sevilla, España

² Departamento de Ciencias de la Computación, Universidad de Jaén, España

^{*} Profesora Contratada Doctora (Departamento de Periodismo II) Facultad de Comunicación, Universidad de Sevilla, España

^{**} Profesor Asociado (Departamento de Periodismo II). Facultad de Comunicación de la Universidad de Sevilla, España

^{***} Profesora (Departamento de Ciencias de la Computación). Universidad de Jaén, España

Conflictos de intereses: Los autores declaran no tener conflicto de intereses.

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

Abstract

This paper analyses the hate speech in messages published on Twitter from May 17 to 25, 2021 during the crisis caused by the entry of thousands of immigrants at the Tarajal border in Ceuta. The aim of the research is to perform a thematic classification of messages with hate speech. For this, a mixed methodology has been used and six subjets related to hate speech on Twitter have been identified, of which four are focused on political issues, accounting for 80% of the messages, and only 20% of them refer to racism and immigration. In addition, up to five campaigns of unknown origin have been detected. This work concludes that hate speeches focus 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

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. Introducción

Las redes sociales han permitido hacer uso de la libertad de expresión que ha podido degenerar en abuso, de forma que se ha creado la necesidad de establecer unos límites. Tanto es así que en los últimos años el debate mediático se está centrando en si la libertad de expresión está en riesgo al combatirse en los tribunales ciertos discursos, sobre todo los de odio. Como sostiene Burgos-García (2019), "si bajo el anonimato se realizan conductas que no obedecen a un pensamiento crítico, constructivo, se está sobrepasando los límites

a la libertad de expresión en detrimento de los derechos del otro" (p. 138). No obstante, habrá que distinguir entre diferentes formas de expresión que pueden resultar groseras, de los propios delitos de odio, que son los que llevan a la jurisprudencia a poner límites a la libertad de expresión (p. 149).

El nacimiento de las redes sociales se sitúa en los inicios del siglo XXI, aunque su popularidad no llegó hasta años más tarde, cuando aumentaron los usuarios y las funcionalidades de estas plataformas. Solo una década después, investigadores como Falxa (2014) empezaron a estudiar los discursos de odio en las redes sociales, asegurando que "la lucha contra los discursos de odio en las redes sociales choca contra numerosos obstáculos" (p. 104). Aunque en Europa ya se estaban dando pasos en este sentido, como la "decisión del 24 de enero de 2013", del juez de primera instancia del Tribunal de Gran Instancia de París, que obligaba a Twitter, con sede en Estados-Unidos, "a habilitar un mecanismo de alerta que permita a los usuarios franceses señalar todo abuso o mensaje de odio para su eventual eliminación, y a transmitir a las autoridades francesas los datos de los usuarios autores de los tuits ilícitos" (p. 104). El 31 de mayo de 2016, la Unión Europea impuso a las plataformas: Facebook, Google, Microsoft y Twitter, acordar un código de conducta que les obligara a revisar y eliminar el discurso de odio ilegal que se hubiera publicado en sus servicios en un plazo de 24 horas (Bisht et al., 2020, p. 244).

La literatura científica se acerca a los discursos de odio desde diferentes ámbitos, sobre todo desde el derecho (Howard, 2019; Rollnert-Liern, 2019; Bautista-Ortuño, 2017; Faixa, 2014). Sin embargo, resulta relevante que se hagan estudios desde las ciencias de la comunicación, máxime cuando hay quienes consideran a las redes sociales como medios (Hütt-Herrera, 2012; Pantoja-Chaves, 2011). Un punto de vista que puede ser discutido, pero lo que no se discute es que los mensajes que se difunden a través de las redes sociales llegan a millones de personas cada día y, al menos como canales de comunicación, es interesante saber si, a través de ellas, se difunden delitos de odio.

En este sentido, un estudio reciente de Arcila-Calderón et al. (2022, p. 32) ha detectado que, "en aquellas regiones europeas en las que se registra una mayor proporción de inmigrantes existe un mayor apoyo ciudadano hacia el colectivo" y, por tanto, "aquellas regiones en las que el apoyo era mayor registraban un menor nivel de discurso de odio en Twitter". No obstante, como se aprecia en el presente trabajo, aún se observan delitos de odio en redes sociales como Twitter que se debería trabajar para erradicar.

Algunos estudios desde el ámbito de las redes sociales que se han publicado han puesto el acento en los delitos de odio motivados por el género o la orientación sexual, como es el caso de Arcila-Calderón et al. (2021a). Estos autores, además, presentan "el primer prototipo para la detección automática de discurso de odio en Twitter en español específicamente motivado por género y orientación sexual".

La presente investigación parte de lo acontecido en España los días 17 y 18 de mayo de 2021, cuando miles de personas entraron en el país a través de la frontera con Marruecos

situada en El Tarajal (Ceuta). Este acontecimiento provocó que el propio presidente del Gobierno español, Pedro Sánchez, visitara la zona el día 18 de mayo y esa misma jornada Reuters ofreció a diferentes medios de comunicación la imagen de una voluntaria de Cruz Roja Española, consolando a un inmigrante subsahariano que generó numerosos tuits denigrantes hacia ella.

El detonante de la permisividad de Marruecos para que entraran miles de personas en España fue el ingreso en un hospital español del considerado como líder del Frente Polisario, Brahim Ghali, con pasaporte de Argelia, sin previo aviso a las autoridades marroquíes, en un momento de gran tensión, ya que en noviembre de 2020 habían surgido fuertes enfrentamientos entre Marruecos y el Frente Polisario por el control del Sáhara Occidental.

Todos estos hechos provocaron una gran reacción en las redes sociales, tanto en España como en Marruecos. Algunas personas usuarias aprovecharon la ocasión para difundir mensajes xenófobos, violentos, racistas o que fomentaban un discurso de odio.

La inmigración siempre ha supuesto un foco para alentar discursos de rechazo y de odio. En este sentido se pueden citar estudios sobre el odio ensalzado por el nacionalismo en Estados Unidos, posicionando al migrante como una amenaza (Morgenfeld, 2016) o el caso de Alemania donde el partido Alternativa para Alemania (AfD), que se autodenomina anti refugiados y antiinmigración, llegó a ser el tercero más fuerte en el parlamento alemán en el año 2017 (Valdez-Apolo et al., 2019, p. 365). Este último estudio concluye que el tono que predomina en Twitter es negativo, lo cual achacan los autores a que frecuentemente la inmigración se relaciona con "acontecimientos o circunstancias hostiles de carácter bélico, económico, político, demográfico y/o climático" (p. 379). No obstante, presenta como limitación que "el estudio se ha llevado a cabo a través del análisis de contenido manual" (p. 385). Como veremos más adelante, este artículo ha utilizado un método mixto para el análisis. Otro de los focos de odio y publicaciones violentas se encuentra en la islamofobia (Awan, 2014, Zamora et al., 2021).

1.1. Libertad de expresión y redes sociales

El hecho de que la libertad de expresión sea un derecho no es óbice para que tenga también una serie de limitaciones que las leyes han tenido que ir desarrollando a lo largo de los años. Estas limitaciones no están relacionadas con la expresión de ideas percibidas como negativas, puesto que, si esto fuera así, desaparecería el pluralismo, la tolerancia y la amplitud de miras en una sociedad democrática (González-Herrera, 2018, p. 3). En este sentido, los medios de comunicación están especialmente protegidos como informadores y formadores de opinión, aunque no siempre el ejercicio de este derecho se haya hecho de una forma ética. González-Herrera pone como ejemplo las falsedades publicadas en Reino Unido en la campaña que precedió al referéndum para su salida de la Unión Europea

y las elecciones presidenciales de Donald Trump de 2016. Según algunos estudios, el presidente estadounidense se había mostrado "como un mentiroso empedernido, inmune y desafiante al *fact-checking*" (Echevarría, 2016, p. 10).

No serían los medios de comunicación, sino las redes sociales las que difundieron más noticias falsas en la campaña del candidato republicano estadounidense, acentuándose durante la pandemia de la COVID-19, unas noticias falsas que luego pasaron a los medios de comunicación que usaron las redes sociales como fuentes de información (Fernández-Muñoz et al., 2022; Gutiérrez-Vidrio, 2020; Pérez-Curiel y Limón-Naharro, 2019). De hecho, las redes sociales y especialmente Twitter tuvieron que intervenir para frenar el mal uso por parte de Donald Trump de la plataforma, sobre todo tras el asalto al Capitolio de los Estados Unidos que tuvo lugar el 6 de enero de 2021 y por sus mensajes falsos, cuando no incitando al odio y la violencia. El día 8 de enero, Twitter cerraba la cuenta de Donald Trump y más tarde lo hizo Reddit con la cuenta de los seguidores de Trump por promover conductas de odio, a lo que se sumó Twitch con la cuenta de Trump.

En la página web de ayuda de Twitter hay un apartado completo dedicado a las políticas de la empresa en el caso de detectar discurso de odio en los contenidos publicados (Twitter, 2021). Durante el transcurso de esta investigación se han detectado varias cuentas que han sido borradas por haber publicado en diversas ocasiones contenido que incluía incitación al odio.

1.2. Los discursos de odio en Twitter

El odio es una conducta reprochable tanto en el ámbito moral, religioso y social como en el jurídico, puesto que puede lesionar derechos como el honor o la dignidad de una persona o un colectivo (Moretón-Toquero, 2012, p. 4). Ya en 2012 se intentaba denominar de algún modo a las conductas motivadas por el odio realizadas a través de internet a las que Moretón-Toquero (2012) llamaba "ciberodio", aunque en este estudio se habla de internet de un modo muy genérico, sin entrar a estudiar la incidencia de las redes sociales. Lo que ponía de manifiesto era la dificultad de legislar en un ámbito tan abierto e internacional como es internet, pero también el temor que podía suscitar la banalización de este asunto en un medio al que accedían los jóvenes como un público más influenciable (p. 15).

Para definir qué es material racista y xenófobo recurrimos al 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 y publicado en el número 26, sección1 del Boletín Oficial del Estado de España el 30 de enero de 2015:

Por "material racista y xenófobo" se entenderá todo material escrito, toda imagen o cualquier otra representación de ideas o teorías, que propugne, promueva o incite al odio, la discriminación o la violencia, contra cualquier persona o grupo de personas, por razón de la raza, el color, la ascendencia o el origen nacional o étnico, así como de la religión en la medida en que ésta se utilice como pretexto para cualquiera de esos factores. (p. 7.216)

El informe de la encuesta sobre delitos de odio, realizado entre diciembre de 2020 y marzo de 2021 por la Secretaría de Estado de Seguridad del Ministerio del Interior del Gobierno de España (López-Gutiérrez et al., 2021), clasifica el delito de odio en: Racismo/Xenofobia, Orientación sexual/ LGTBfobia, Religión, Ideología e Incapacidad/ enfermedad. Todas estas manifestaciones de odio se han podido encontrar entre los mensajes descargados de Twitter para este trabajo.

Según ponía de manifiesto un informe publicado por el Ayuntamiento de Barcelona, ya en 2014 había unos 10.000 tuits cada día con insultos racistas en lengua inglesa, lo que equivalía a uno de cada 15.000 tuits (Cabo-Isasi y García-Juanatey 2017, p. 5). Años más tarde, los medios se harían eco de la decisión de Twitter de eliminar más de mil tuits en los que se detectaba un abuso racista y se afirmaba que el Gobierno británico estaba trabajando en leyes en este sentido (Cano, 2021).

Encontramos algunos estudios recientes sobre los discursos de odio en Twitter, como el de Valdez-Apolo et al. (2019), quienes aseguran que el sentimiento que prevalece en los mensajes sobre inmigrantes es negativo, más que entre los refugiados. Estos autores consideran que ese "tono negativo" puede tener su origen en que "el fenómeno migratorio" en los medios presenta una "naturaleza negativa", por encontrarse con frecuencia "relacionado con acontecimientos o circunstancias hostiles de carácter bélico, económico, político, demográfico y/o climático" (p. 379). También abordan la estigmatización de los refugiados a través de mensajes de odio en las redes sociales Frías y Seoane (2019), y Arcila-Calderón et al. (2021b).

Una circunstancia que puede verse de forma latente en el objeto de estudio que nos ocupa, pues la entrada masiva de inmigrantes en un país, propiciada por la laxitud de las autoridades del país vecino, pudo ser vista por la población receptora como una amenaza a su integridad como país. Ya que, aunque en un principio podría pensarse que el discurso de odio en línea no causa daño debido a su ser virtual, los estudios de Williams et al. (2020) han demostrado una relación entre los mensajes de odio en redes sociales y los delitos reales especialmente contra comunidades negras y musulmanas.

Otros estudios han abordado la temática de los refugiados disgregando el estudio por género, es el caso de Amores et al. (2020), quienes concluyen que la imagen de las mujeres refugiadas es como agentes pasivos en las imágenes publicadas por los medios.

El discurso político también se ha relacionado en diferentes estudios con los mensajes de odio a través de Twitter (Arcila-Calderón et al., 2021c; Arcila-Calderón et al., 2020; Casero-Ripollés et al., 2020; Paz, 2021).

2. Material y métodos

Para la realización de este trabajo se ha usado una metodología mixta, muy similar a la que se puede ver en otros trabajos también relacionados con el discurso de odio y Twitter, por ejemplo, en Miró-Linares (2016), en el que al mismo tiempo se emplea el análisis cuantitativo con técnicas de minería de datos, y el cualitativo para el análisis del contenido. En este sentido, se han usado también técnicas que filtran automáticamente el contenido de Twitter a partir del uso de lexicones específicos, como puede verse en Jiménez-Zafra et al. (2021).

En primer lugar, se descargaron, usando la API V2 de Twitter, todos los mensajes publicados en castellano entre el 17 y el 25 de mayo de 2021 que contenían las palabras Ceuta y/o Marruecos. Además, se realizó una descarga adicional en función de una serie de etiquetas (hashtags), que pueden verse en la Tabla 1. El listado de etiquetas se obtuvo usando las páginas web de Trendinalia, Account Analysis y Google Trends, a partir de los temas que más habían interesado a los usuarios y las etiquetas presentes en los mensajes. Todo este material se convirtió en una hoja de cálculo.

El número total de tuits descargados fue 310.907. Muchos de ellos eran publicaciones retuiteadas, es decir entradas originales que otras personas se limitan copiar y pegar tal cual, en sus propias cuentas, que no aportaban contenido, por lo que hubo que hacer un filtrado que eliminara estos mensajes, obteniendo tras este proceso 38.901 publicaciones. En la Tabla 1 se puede observar la cantidad de mensajes recuperados en función de la clasificación por palabra clave y por etiquetas:

Tabla I. Listado completo de mensajes descargados organizados por palabra clave y etiqueta

Consulta	Retuiteado	Original	Respuesta	Cita	Total	Total sin retuits
Marruecos	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

Fuente: Elaboración propia

Posteriormente, se realizó un segundo filtrado para eliminar las entradas duplicadas, ya que un tuit puede incluir varias palabras clave y etiquetas a la vez. Tras el nuevo proceso de filtrado, el número de publicaciones se redujo a 36.887. En total, se detectaron 18.441 mensajes originales, siendo el resto citas (2.932) o respuestas (15.514).

Para el análisis cuantitativo se usaron cuatro lexicones específicos que pueden ser útiles para detectar discurso de odio. Estos lexicones se han evaluado de forma satisfactoria en el estudio de Plaza-del-Arco et al. (2020). A continuación, se muestra el número de palabras que contiene cada lexicón y un par de ejemplos significativos:

Xenofobia: 44 palabras (ej. moro, negrata)

Inmigración: 250 palabras (ej. deportarlo, refugiado)

Misoginia: 183 palabras (ej. perra, zorra)

Insultos: 279 (ej. cabrón, maricón)

Para facilitar el análisis, toda la información se organizó en una hoja de cálculo compuesta por 18 campos. Los diez primeros estaban directamente relacionados con la información extraída a través de la API de Twitter, mientras que los ocho últimos se crearon a partir de los contenidos aportados por los cuatro lexicones. Cada uno de estos lexicones generó dos campos en la hoja de cálculo, uno indicando la presencia o no de palabras incluidas en ellos y otro con los términos encontrados:

Id: Identificador unívoco del tuit. **Texto:** Mensaje escrito en el tuit.

Tema: Campo generado automáticamente a partir del uso del filtro de los lexicones, que clasifica los mensajes en función de si tratan sobre inmigración, xenofobia, insultos o misoginia. En la Tabla 2 puede verse el resultado de esta clasificación.

Tipo: Original, respuesta o cita, en función de si el mensaje es una publicación original o si es un comentario o cita de otro.

Contenido sensible: Campo facilitado automáticamente por Twitter con valores de verdadero o falso.

Recuento de retuits: Número de veces que se ha retuiteado una publicación original.

Recuento de comentarios: Total de comentarios recibidos por cada tuit.

Recuento de citas: Número de veces que un usuario ha citado un mensaje.

Recuento de me gusta: Total de usuarios que han marcado cada mensaje como favorito.

Estos últimos cuatro campos se han usado para valorar la popularidad de las publicaciones.

Fuente: El origen de la publicación en función de si se ha publicado desde un móvil o desde una aplicación de escritorio.

Total de palabras sobre inmigración: Número de palabras del tuit que aparecen en el lexicón de inmigración.

Palabras sobre inmigración: Palabras del lexicón de inmigración presentes en el tuit. **Total de insultos:** Número de insultos del lexicón usados en el tuit.

Insultos: Palabras del lexicón relacionadas con insultos y presentes en el tuit.

Total de palabras misóginas: Número de palabras del lexicón de misoginia usadas en el tuit.

Palabras misóginas: Palabras del lexicón de misoginia que aparecen en el tuit.

Total de palabras xenófobas: Número de palabras xenófobas utilizadas en el tuit.

Palabras xenófobas: Palabras del lexicón de xenofobia usadas en el tuit.

Tabla II. Clasificación temática según el filtro de los lexicones

Consulta	Ninguno	Inmigración	Insultos	Misoginia	Xenofobia	Total
Marruecos	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

Fuente: Elaboración propia

Una vez realizada esta organización automatizada, se procedió al análisis manual del contenido leyendo los mensajes marcados por el filtro temático de los lexicones, más todos los mensajes con hashtags. Así, de los 36.887 tuits descargados sin duplicados, se revisaron manualmente los tuits considerados sospechosos de contener discurso de odio, en total 7.000. De ellos, 4.964 incluyen alguna de las palabras de los lexicones, mientras que 1.817 se corresponden a la suma de todos los mensajes que incluían alguno de los hashtags. El hecho de que una publicación tenga una palabra de un lexicón no implica que el tuit sea violento, del mismo modo hay mensajes que no aparecen marcados por los lexicones y que incluyen contenido de odio. Para encontrar el mayor número de estos mensajes se realizó una búsqueda adicional de las palabras: guerra, invasión, frontera, ejército, humillación, muerte y expulsión; que aumentó en 219 la selección de mensajes analizados, llegando así a los 7.000 indicados. Estos mensajes fueron revisados manualmente y se encontraron 1.000 tuits que realmente incluían discurso de odio. En el resto, aunque se usaban palabras que podrían ser indicadoras de odio, no lo eran en el contexto usado. El número exacto fueron 1.005, pero se redujo a 1.000 para simplificar los cálculos.

3. Resultados y discusión

A partir del análisis del contenido de los mensajes, se ha establecido una clasificación en seis áreas temáticas. Para esta clasificación no se han tenido en cuenta los mensajes que forman parte de campañas organizadas, que se estudian en el apartado 3.2.

3.1. Clasificación temática

A pesar de que todo el corpus está en castellano, se han encontrado muchos mensajes de usuarios marroquíes publicados en castellano con la intención de insultar a usuarios españoles o hacer reclamaciones con lenguaje violento que la ciudadanía española pudiera leer. No obstante, aunque la mayoría de los mensajes de odio encontrados expresan violencia contra Marruecos, en el estudio no se han hecho distinciones entre los mensajes de odio o violentos de unos y de otros.

En este ejemplo, se observa un tuit en castellano en el que se hace referencia a Ceuta y Melilla como ciudades ocupadas y se alienta la creación de grupos terroristas:

Tabla III. Relación de temas detectados y número de mensajes

Política interna española. Insultos a diferentes cargos políticos.	167
Reclamación/afirmación de la soberanía de Ceuta y Melilla. Invasión, humillación.	123
Política interna marroquí, insultos al rey de Marruecos.	92
Defensa militar de las fronteras y mensajes que llaman a la violencia física.	86
Racismo en general. Todo tipo de insultos racistas.	76
Inmigración en general, menas, ayudas a los inmigrantes.	41

Fuente: Elaboración propia

Durante el análisis de la muestra de 1.000 mensajes seleccionados en los que se ha podido constatar la presencia de lenguaje violento o discurso de odio, se han encontrado seis temas en los que se pueden clasificar los contenidos (Tabla 3). Un mismo mensaje puede formar parte de varias categorías.

Política interna española. Insultos a diferentes cargos políticos: Es con diferencia el tema que más se repite, con 167 publicaciones. Aparecen mensajes tanto de españoles, como de marroquíes. La llegada de Santiago Abascal, líder de VOX, a Ceuta el día 24 de mayo supone un punto de inflexión en la crisis, que pasa de ser una crisis migratoria a una crisis política nacional en la que se enfrentan los partidos de la derecha, en la oposición, con los partidos de izquierda en el Gobierno, PSOE y Unidas Podemos.

Una gran parte de los contenidos violentos están relacionados con los actos de VOX organizados en Ceuta, la visita del presidente del Gobierno, Pedro Sánchez, a Bruselas y la actitud de los ministros del Gobierno de España.

Hay que destacar la gran cantidad de insultos que se recogen en estos mensajes, tanto hacia los líderes de la derecha, como de la izquierda.

Reclamación/afirmación de la soberanía de Ceuta y Melilla. Invasión/ocupación, humillación: Este tema, con 123 tuits, es el más directamente relacionado con los hechos: La llegada de un gran número de personas desde Marruecos a España por la frontera de El Tarajal en Ceuta. Los mensajes se pueden dividir en dos tipos, los que reclaman y afirman la soberanía española o marroquí sobre las ciudades de Ceuta y Melilla, y los que se quejan de una invasión/ocupación humillante, aunque no piden una respuesta militar violenta, sino que suelen culpar de la situación a las políticas de los Gobiernos. Encontramos en este apartado tanto mensajes de españoles como de marroquíes.

Política interna marroquí, insultos al rey de Marruecos: En este caso, no se encuentran mensajes escritos por marroquíes contra su propio Gobierno, como sí veíamos en el caso español, donde los mensajes contra el Gobierno de España aparecían firmados por españoles y marroquíes.

Se han clasificado 92 mensajes en esta temática, que ha resultado ser una en las que más contenido violento y discurso de odio se ha mostrado: Insultos, xenofobia, LGTBfobia y todo tipo de acusaciones hacia el rey de Marruecos.

Defensa militar de las fronteras y mensajes que llaman a la violencia física: Otro de los apartados más violentos del estudio donde se reclama la presencia militar y se plantea un escenario bélico y violento contra Marruecos o España. Hay 86 mensajes en esta clasificación, divididos entre publicaciones de marroquíes y españoles. La mayor parte de los mensajes, con expresiones violentas, publicadas por usuarios marroquíes hablan de territorios ocupados, tierras robadas y de liberación, con referencia a las ciudades españolas de Ceuta y Melilla.

Una crisis migratoria muy puntual, que nada tiene que ver con ocupaciones o liberaciones, termina derivando en las redes sociales en una declaración de guerra contra el país vecino por ambas partes. Mientras el Gobierno trata a los migrantes como refugiados, en las redes se les considera liberadores o invasores según la nacionalidad del autor de los mensajes.

Racismo en general. Todo tipo de insultos racistas: Era tal la cantidad de mensajes con contenido racista o xenófobo, que se ha planteado hacer una categoría solo para recoger estos mensajes. Con un total de 76 publicaciones, no es la categoría con

más publicaciones, pero destaca porque el insulto es casi el único contenido de estas publicaciones.

Inmigración en general, menas, ayudas a los inmigrantes: Solo aparecen 41 mensajes violentos o con discurso de odio en esta categoría. Es uno de los temas que más se repiten en el corpus de 36.887 mensajes analizados, pero no es en el que se vierten más insultos.

La mayor parte de los tuits que encontramos en esta categoría son mensajes violentos sobre la inmigración o aquellos que incluyen mentiras solo para desprestigiar y criminalizar a refugiados marroquíes o a menores no acompañados que han llegado a España.

3.2. Campañas

Tras el análisis de los datos extraídos, se ha observado que una gran parte de las publicaciones violentas constituyen campañas en las que los usuarios, de forma organizada o espontánea, publican un mismo mensaje masivamente con la clara intención de influir en la opinión pública. Se han encontrado cinco de estas campañas que se analizan a continuación. La importancia es tal, que suponen el 53,2% del contenido de la muestra.

A continuación, se muestra una tabla resumen del número de publicaciones recogidas en cada una de las campañas:

Tabla IV. Publicaciones de cada campaña

Ayúdame a frenar la invasión de Ceuta	325
Marruecos coló entre los 8.000 que invadieron Ceuta a jóvenes adoctrinados	55
Pedro Sánchez oculta a Transparencia cuántos ilegales ha trasladado desde Canarias	46
El diario El País explota las fotos de un bebé para los refugiados en Turquía	24
Invasión fuerte hasta la muerte	18

Fuente: Elaboración propia

Ayúdame a frenar la invasión de Ceuta. Dile a Marlaska que autorice la expulsión inmediata de los ilegales en Ceuta YA #invasionCeuta https://t.co/U4smcLJQtB

El enlace lleva a una petición en la web de hazteoir.org, creada el 18 de mayo de 2021, cuya finalidad era lograr 50.000 firmas para que el ministro del Interior autorizara la expulsión en caliente de todos los inmigrantes calificados como ilegales en la petición.

El texto de la petición habla de la llegada de 10.000 invasores a la ciudad de Ceuta y mantiene en todo momento un tono violento contra el Gobierno y contra las personas refugiadas en Ceuta.

Aunque el número de firmas solo llegó a 25.000, en Twitter la campaña fue un éxito ya que, en los 1.000 mensajes de la muestra, esta petición aparece repetida 325 veces, lo que nos da una idea de la gran difusión que llegó a tener.

Marruecos coló entre los 8.000 que invadieron Ceuta a jóvenes adoctrinados. Con 55 mensajes con el mismo texto ocupa la segunda posición en la lista de campañas detectadas. Se trata de un artículo publicado el 24 de mayo en El Español por la periodista Sonia Moreno.

La noticia puede encontrarse también en otros medios, pero sin el matiz de que haya sido un acto voluntario de Marruecos, como parece que indica la información de El Español.

Pedro Sánchez oculta a Transparencia cuántos ilegales ha trasladado desde Canarias, Ceuta y Melilla. Aunque este tema no tiene nada que ver con la noticia de la llegada de personas a la frontera marroquí en Ceuta, su publicación en OK Diario el 24 de mayo genera polémica, especialmente cuando se acusaba personalmente al presidente del Gobierno de ocultar información. La propia redacción de la noticia en la que califica de ilegales a las personas ya supone una redacción tendenciosa cuyo objetivo va más allá de la información. El texto completo de los mensajes recogidos en Twitter no aporta mucho más que el título.

El diario El País explota las fotos de un bebé para los refugiados en Turquía. La mayor parte de los mensajes publicados en esta clara campaña de desinformación están publicados en árabe, solo una pequeña parte, los 24 que se han detectado en la muestra, aparecen en castellano.

La foto a la que hace referencia es una en la que se ve el rescate de un bebé en el mar por parte de un miembro del GEAS. Esta foto se publicó en cientos de medios en todo el mundo, mientras que un grupo de usuarios marroquíes estuvieron difundiendo que la fotografía no corresponde a Ceuta, sino que es parte de un rescate en Turquía de varios meses atrás. En España la foto se hizo viral y generó miles de interacciones en todas las redes sociales. Aunque casi todos los medios españoles se hicieron eco de esta información, la campaña de desprestigio se centra en el diario El País. El enlace lleva a una conversación de Twitter con la foto y superpuesto un texto en árabe que dice:

El español El País atribuye la imagen de un bebé frente a las costas de Turquía a los sucesos de Ceuta. El periódico español El País publica una foto antigua de acontecimientos anteriores en Turquía y los atribuye a los hechos vividos en la zona fronteriza de las ciudades ocupadas de Ceuta y Melilla.

La campaña original en árabe pretende desacreditar a los medios españoles para salvar la imagen pública ante la ciudadanía marroquí. La campaña en castellano, realizada para que la vean quienes conocen la noticia, busca fomentar la violencia insultando a los medios españoles. En todos los casos el contenido termina generando un clima violento en la conversación con mentiras e insultos.

Este bulo fue desmentido por la plataforma de verificación Newtral que publica en su web cómo ha buscado las fuentes y verificado la autenticidad de la fotografía (Maqueda, 2021).

Invasión fuerte hasta la muerte es la última de las campañas detectadas. En este caso con 18 mensajes publicados en castellano. Se trata de nuevo de publicaciones con origen en Marruecos en las que se fomenta una respuesta violenta en el conflicto por parte de los marroquíes. La campaña incluye siempre una o varias menciones hacia medios y partidos políticos españoles, principalmente VOX y su presidente Santiago Abascal.

El texto que aparece en los mensajes suele incluir la etiqueta #Ceuta y Melilla son marroquíes.

4. Conclusiones

Aplicando técnicas de minería de datos se ha generado un corpus de 310.907 tuits que, mediante el proceso de filtrado explicado en la metodología, ha reducido el análisis al contenido de 1.000 mensajes, lo que ha facilitado la obtención del objetivo principal de esta investigación, hacer una clasificación temática de los mensajes que incluyen discurso de odio, así como analizar algunas de las tendencias más destacadas aparecidas durante la investigación.

Curiosamente, el tema que más aparece en los tuits que contienen discurso de odio es la política interna española, con insultos a diferentes cargos políticos. Un hecho que hace vislumbrar que se ha colado una campaña política amparándose en la libertad de expresión y la facilidad de difusión de información a través de las redes sociales.

El segundo tema más recurrente es el de la soberanía de Ceuta y Melilla que, en cierto modo, también se circunscribe dentro del ámbito político, pues se trata de la defensa de dos ciudades autónomas españolas en suelo africano que han tenido repercusión a lo largo de los años en las relaciones internacionales de España con Marruecos.

La política interna marroquí y los insultos al rey Mohamed VI también aparecen, por sorpresa, en este análisis sobre discurso de odio. Se trata de un tema muy espinoso para los marroquíes que viven en su país de origen, por ese motivo entendemos que se cuela en un estudio de tuits en español y sería bastante improbable que lo hiciera en una muestra en árabe.

La defensa militar de las fronteras y los mensajes que llaman a la violencia física también pueden tener que ver con cuestiones políticas e ideológicas y es un tema que está íntimamente relacionado con la soberanía sobre Ceuta y Melilla.

Por supuesto, también se hallan en el estudio insultos racistas y una fuerte significación del racismo en general, puesto que los grupos con esta ideología aprovechan cualquier oportunidad para lanzar sus mensajes y soflamas.

Resulta especialmente destacable que, tratándose de un tema, el estudiado, que tiene que ver con la inmigración, las ayudas a las personas inmigrantes y los menores no acompañados que entran en España de forma irregular, sea el tema menos recurrente en los mensajes analizados. Este hecho puede llevar a concluir que el acontecimiento estudiado no fue visto como un gran problema migratorio, sino como un problema político.

De ahí que el 53,2% del contenido de la muestra tenga que ver con campañas perfectamente orquestadas en el seno de las redes sociales, que llevan de nuevo a establecer una relación temática con la política: frenar la invasión de Ceuta; la entrada de jóvenes adoctrinados por Marruecos; una supuesta ocultación del presidente del Gobierno de España de las personas irregulares (en los tuits se usa el término ilegales) trasladados desde Canarias, hecho que nada tiene que ver con el acontecimiento que se estaba viviendo esos días en la frontera con Marruecos; la difusión del bulo de que la fotografía del guardia civil rescatando un bebé era falsa y, quizá la más dura por su contenido: invasión hasta la muerte. Unos mensajes, estos últimos difundidos por marroquíes en español, alentando a la sublevación violenta de éstos contra España, mencionando frecuentemente a VOX y a su líder, Santiago Abascal.

Por tanto, los mensajes de odio se han vertido en este caso introduciendo en las conversaciones a través de Twitter temas más políticos que la propia problemática de la inmigración, sus causas, sus consecuencias y las posibles soluciones a un drama que deja cada año cientos de muertos en el mar y tullidos en las vallas que separan, en este caso, España de Marruecos.

Contribución de autores

Aránzazu Román-San-Miguel: Conceptualización, Análisis formal, Investigación, Metodología, Recursos, Validación, Visualización, Software, Redacción-borrador original y Redacción-revisión y edición. **Francisco J. Olivares-García:** Conceptualización, Análisis formal, Investigación, Metodología, Recursos, Supervisión, Software y Redacción-revisión y edición. **Salud María Jiménez-Zafra:** Investigación, Metodología, Recursos, Visualización, Software, Redacción-borrador original y Redacción-revisión y edición. Todos los autores han leído y aceptado la versión publicada del manuscrito.

Agradecimientos

Esta publicación es parte del proyecto LIVING-LANG (RTI2018-094653-B-C21) financiado por MCIN/AEI/10.13039/501100011033 y por FEDER Una manera de hacer Europa, del

proyecto PID2020-116118GA-I00 financiado por MICINN/AEI/10.13039/501100011033, del proyecto PID2020-119478GB-I00 financiado por MICINN/AEI/10.13039/501100011033, y de los proyectos Big Hug (P20_00956, PAIDI 2020) y WeLee (1380939, FEDER Andalucía 2014-2020) financiados por la Junta de Andalucía. Salud María Jiménez-Zafra ha sido financiada por el Fondo Social Europeo y la Administración de la Junta de Andalucía (DOC_01073).

Referencias

- 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 farright 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
- 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

- 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
- 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/3nm.bbmys
- 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