

Rechazo y discurso de odio en Twitter: análisis de contenido de los tuits sobre migrantes y refugiados en español

Rejection and Hate Speech in Twitter: Content Analysis of Tweets about Migrants and Refugees in Spanish

Carlos Arcila Calderón, David Blanco-Herrero y María Belén Valdez Apolo

Palabras clave

Análisis de sentimientos

- *Big data*
- Discurso de odio
- Inmigración
- Refugiados
- Twitter

Key words

Sentiment Analysis

- Big Data
- Hate Speech
- Immigration
- Refugees
- Twitter

Resumen

Se analiza el rechazo verbal al extranjero como potencial detector de discurso de odio a través de dos análisis de contenido de tuits en español recogidos con la API de Twitter: el primero, manual, a 1.469 tuits; el segundo, automático, analiza otros 337.116 tuits utilizando técnicas de *big data*. El rechazo fue predominante en el primer análisis y minoritario en el segundo, mostrando la fluctuación que experimentan estas expresiones en función del contexto mediático. En ambos casos el rechazo hacia los migrantes fue significativamente mayor que hacia los refugiados, como se había observado ya en contextos internacionales. El trabajo también generó un corpus de entrenamiento sobre rechazo al extranjero y observó los aspectos negativos asociados más frecuentemente al rechazo, así como la relación entre este y la condición informativa u opinativa del tuit.

Abstract

We use Twitter to study the verbal rejection towards migrants and refugees as a potential hate speech predictor with two content analysis of tweets in Spanish collected with Twitter's API: the first analysis, manual, with 1,469 tweets; the second, automatic, uses big data techniques to study 337,116 new tweets. In the first one rejection was predominant over acceptance and neutrality. Rejection was smaller in the second one, showing how fluctuant these expressions are depending the media context. In both cases rejection toward migrants was significantly bigger than over refugees, as it had already been observed in international contexts. This work also created a training corpus about immigrant rejection, valid for future studies, and observed the negative aspects most frequently associated to rejection of immigrants, as well as the relationship existing between this and the fact of tweets being information or opinion.

Cómo citar

Arcila Calderón, Carlos; Blanco-Herrero, David y Valdez Apolo, María Belén (2020). «Rechazo y discurso de odio en Twitter: análisis de contenido de los tuits sobre migrantes y refugiados en español». *Revista Española de Investigaciones Sociológicas*, 172: 21-40. (<http://dx.doi.org/10.5477/cis/reis.172.21>)

La versión en inglés de este artículo puede consultarse en <http://reis.cis.es>

Carlos Arcila Calderón: Universidad de Salamanca | carcila@usal.es

David Blanco-Herrero: Universidad de Salamanca | david.blanco.herrero@usal.es

María Belén Valdez Apolo: Universidad del Azuay (Ecuador) | mariabelenvaldezapolo@gmail.com

INTRODUCCIÓN¹

En un contexto marcado por grandes migraciones globales, en los últimos años se ha visto cómo diversas opciones políticas abiertamente antinmigración y de corte xenófobo se han instalado en países de todo el mundo. Simultáneamente, vivimos un auge del odio hacia la otredad en plataformas digitales (Muller y Schwarz, 2018), lo que contribuye a un aumento del discurso de odio en redes sociales (Bartlett *et al.*, 2014) y, con ello, al potencial aumento de los ataques a extranjeros.

En esta línea, lo más frecuente ha sido medir las actitudes hacia los inmigrantes utilizando herramientas como la encuesta, método que puede resultar problemático por el sesgo de *deseabilidad social*, pues una persona difícilmente se autoevaluará y clasificará sus expresiones como racistas o xenófobas (Cea D'Ancona, 2009). No es objetivo de esta investigación cuestionar la fiabilidad de la encuesta como método para medir el rechazo al extranjero, sino brindar información complementaria que permita contemplar el fenómeno en todas sus dimensiones, destacando que el análisis de la opinión pública expresada en redes sociales permite visibilizar opiniones dominantes.

Partiendo de la idea de que el rechazo es una categoría construida socialmente (Berger y Luckman, 1966), y teniendo en cuenta la capacidad de las redes sociales de evidenciar en la esfera pública las actitudes de rechazo de los individuos hacia el exogrupo, este artículo busca determinar la presencia de expresiones de rechazo verbal hacia migrantes y refugiados en las re-

des sociales como potencial base de otros tipos de rechazo de mayor magnitud. Asimismo, se pretende evidenciar si los tuits en español asociados a migrantes tienen una connotación más negativa que los asociados a refugiados. Para ello, se llevará a cabo un análisis de contenidos manual y otro automatizado. A su vez, con una vocación exploratoria y complementaria, se busca conocer cuáles son las principales problemáticas que se asocian al rechazo a estos colectivos, además de descubrir si el tipo de mensaje (informativo/opinión) está relacionado con la expresión de rechazo.

Con estos objetivos, el presente trabajo amplía el conocimiento existente sobre el rechazo hacia las personas migrantes y refugiadas en los medios sociales. De manera concreta, este artículo analiza el contenido de las redes sociales para estudiar la presencia de expresiones de rechazo verbal hacia inmigrantes en el entorno hispanohablante, complementando y actualizando trabajos pasados, que han analizado la aproximación a la crisis de los refugiados en distintos países (Gualda y Rebollo, 2016) o que han medido aspectos más concretos del rechazo al inmigrante, como los estereotipos relacionados con el género (Gallego *et al.*, 2017) o el discurso de odio (Ben-David y Matamoros-Fernández, 2016), pero que no han abordado la presencia de rechazo verbal como elemento central.

Al mismo tiempo, la relevancia del discurso público en las redes sociales convierte a este tipo de análisis en una herramienta importante tanto para captar el pulso de la sociedad en torno a ciertos temas como para actuar como herramientas predictoras de futuros comportamientos (Kalyanam *et al.*, 2016). Por esto, la implementación del análisis del rechazo al extranjero en redes sociales o de trabajos que tomen los medios sociales como fuente de datos tiene un gran potencial, especialmente en temas sensibles y guiados por la deseabilidad social como este. Así, el prin-

¹ Los autores agradecen el apoyo y los recursos prestados por los proyectos Preventing Hate Against Refugees and Migrants (PHARM), financiado por la Unión Europea en el marco del programa Rights, Equality and Citizenship (REC-RRAC-RACI-AG-2019 (GA N.º 875217), y al proyecto Desarrollo y evaluación de un detector del discurso de odio en línea en español (STOP-HATE), financiado por la Fundación General de la Universidad de Salamanca en el marco del Plan TCUE 2018-2020 (PC-TCUE18-20_016).

cial aporte metodológico de este estudio es la creación de un corpus con ejemplos de muestras de aceptación/neutralidad y rechazo de extranjeros que puede servir para entrenar modelos de descarga y análisis automatizado en español.

CONTEXTUALIZACIÓN DEL ESTUDIO

Nuevas formas de medición del rechazo al migrante o refugiado

A pesar de tratarse de la herramienta más habitual, investigadores como Cea D'Ancona (2009) y Díez Nicolás (2009) han profundizado en la validez de la encuesta como metodología para medir las actitudes hacia migrantes y refugiados por el ya citado sesgo de *deseabilidad social*. Al mismo tiempo, el estudio basado en las redes sociales va ganando fuerza, ya que es aquí donde discurre una gran parte del discurso público de las sociedades. Como observaron Schäfer y Schadauer (2019), las noticias falsas y la desinformación que se propagan en la red están a menudo detrás de contenidos que fomentan el rechazo hacia migrantes y refugiados. Y dado que el fenómeno de las *fake news* es inseparable de las redes sociales (Bakir y McStay, 2018), el estudio de estas es de especial interés. Por ello, Twitter es una de las plataformas que más popularidad ha ganado en la investigación académica. Centrados en las actitudes hacia migrantes y refugiados, Chaudhry (2015) ha demostrado la capacidad de rastrear el racismo *online* usando esta plataforma. Trabajos más recientes han intentado establecer correlaciones entre el discurso de odio en redes sociales y acontecimientos violentos, como Muller y Schwarz (2018), que investigan el vínculo entre las redes sociales y los crímenes de odio usando datos de Facebook y Twitter. En general, la utilización de herramientas digitales como Twitter para descargar y procesar grandes volúmenes de datos y analizar

las actitudes hacia refugiados y migrantes va ganando peso, como demuestra el estudio con 862.999 tuits de Gallego *et al.* (2017), que incorpora una perspectiva de género al estudio de representaciones de los refugiados. Rebollo y Gualda (2017) realizaron también un estudio similar con una muestra de 151.294 tuits en español y, aunque con objetivos distintos, este método fue seguido también por Gualda *et al.* (2015). En el ámbito internacional, el presente trabajo sigue los pasos de investigaciones como la modelización del discurso de odio *online* en Twitter realizada por Burnap y Williams (2015).

En esta investigación hemos seguido estos estudios y hemos optado por Twitter, ya que, aunque no es una plataforma representativa de todos los ciudadanos, la fácil viralización de contenidos, su popularidad y la rapidez de la comunicación resultan de gran interés para su monitorización y análisis. Además, esta plataforma ofrece un registro abierto de sentimientos y opiniones acerca de asuntos de todo tipo, lo que incluye discursos de odio u otras muestras de rechazo que son expresados libremente y sin las barreras que muchas veces están presentes en espacios *offline*.

Discurso de odio hacia migrantes y refugiados

La discusión en torno al prejuicio y el rechazo hacia el exogrupo es extensa en las ciencias sociales (Brewer, 1999; Peherson *et al.*, 2011). En el campo que nos interesa, Bourhis y Dayan (2004) indican que una identidad nacional fuerte se relaciona con actitudes negativas hacia los inmigrantes, algo que también defienden Verkuyten y Brug (2004). No obstante, todavía existen importantes limitaciones en el estudio de la intolerancia hacia el diferente (Billig, 2002). Concretamente, Brown (2000) señala que, desde la perspectiva de la identidad social, el rechazo al otro puede abarcar desde

el rechazo verbal hasta el genocidio. En la misma línea, el Cuestionario de Discriminación Étnica Percibida (PEDQ) de Contrada *et al.* (2001) identifica el rechazo verbal como la forma más básica de discriminación. El presente trabajo busca abordar el problema desde su base, investigando la expresión verbal de las formas más generales de rechazo hacia migrantes y refugiados, sirviendo de punto de partida a otras investigaciones que se adentren en el estudio de otras formas más específicas de rechazo.

Desde el punto de vista teórico, existe una intensa conexión entre el uso del lenguaje y la transmisión del prejuicio hacia el otro, el denominado exogrupo (Maass *et al.*, 1989). Sin embargo, y contrario al paradigma de la hipótesis de la teoría del sesgo lingüístico intergrupar (Whitley y Kite, 2016; Gorham, 2006), en la transmisión del prejuicio a través de los medios sociales la evidencia empírica más reciente muestra que las descripciones negativas (como el rechazo) hacia el exogrupo dejan de ser vagas o abstractas para convertirse en específicas, observables y medibles cuando están amparadas en discursos oficiales (Crandall *et al.*, 2018) o en redes con capacidad de anonimato (Fox *et al.*, 2015).

Al mismo tiempo, se ha observado que el rechazo a inmigrantes, promoviendo su expulsión o prohibiendo su entrada, por figuras de elevado alcance o líderes de opinión también repercute en el potencial aumento del discurso de odio (Gualda y Rebollo, 2016). Estas descripciones negativas y muestras de rechazo son, precisamente, el principal soporte de las narrativas de odio hacia públicos altamente estereotipados y vulnerables.

El discurso del odio implica la promoción de mensajes que alienten el rechazo, el menosprecio, la humillación, el acoso, el descrédito y la estigmatización de individuos o colectivos sociales basados en atributos que van desde la nacionalidad hasta

la orientación sexual. La Comisión Europea contra el Racismo y la Intolerancia (ECRI), mediante su Recomendación General n.º 15 (2016), especifica que este discurso puede venir motivado por razones de raza, color, ascendencia, origen nacional o étnico, edad, discapacidad, lengua, religión o creencias, sexo, género, identidad de género, orientación sexual y otras características o condiciones personales. El Consejo de Europa, en su Recomendación n.º 97 (1997), añade que debe tratarse de una expresión que «propague, incite, promueva o justifique el odio racial, la xenofobia, el antisemitismo y otras formas de odio basadas en la intolerancia».

En un contexto en el que los medios digitales y las redes sociales permiten que la creación y difusión de estos contenidos sea mayor y más rápida, la relevancia del discurso de odio deriva, sobre todo, de su papel como desencadenante de crímenes de odio (Muller y Schwarz, 2018). La investigación de Muller y Schwarz sugiere que existe una relación significativa entre el discurso de odio *online* y los ataques reales y que «la exposición a la prominencia de contenido antirrefugiados y de extrema derecha es un poderoso predictor de ataques violentos hacia refugiados» (2018: 24). Con esta base, autores como Kreis (2017) han analizado el discurso de odio en Europa hacia migrantes y refugiados en Twitter, algo que también hizo Chaudhry (2015) en Canadá.

Con esto, consideramos que conviene determinar hasta qué punto las expresiones de rechazo hacia el colectivo vulnerable de los extranjeros son predominantes en los medios sociales en español, así como analizar las problemáticas y aspectos negativos asociados a ellos y el tipo de mensajes en los que ese rechazo se manifiesta, para, desde esa base, poder articular estrategias más efectivas que frenen tanto los casos más directos de discurso de odio como el rechazo en general. Así surgen las siguientes preguntas de investigación:

PI1. ¿Cuál es la opinión dominante en términos de *aceptación* o *rechazo* hacia migrantes y refugiados en mensajes en español en Twitter?

PI2. ¿Cuáles son los aspectos negativos con los que se asocian las expresiones de *rechazo* hacia migrantes y refugiados en mensajes en español en Twitter?

PI3. ¿En qué tipo de mensajes (informativos o de opinión) es más frecuente encontrar *rechazo* a migrantes y refugiados?

Diferencias entre migrantes y refugiados

Aunque en el lenguaje cotidiano mucha gente puede utilizar los términos de manera indistinta y en los grandes desplazamientos nos encontramos generalmente con personas que cumplen con ambos perfiles, conviene diferenciar entre «migrante(s)» y «refugiado(s)»². La Convención de Ginebra sobre el Estatuto de los Refugiados de 1951 indica que «refugiado» es la persona que

debido a fundados temores de ser perseguida por motivos de raza, religión, nacionalidad, pertenencia a un determinado grupo social u opiniones políticas, se encuentre fuera del país de su nacionalidad y no pueda o, a causa de dichos temores, no quiera acogerse a la protección de su país; o que careciendo de nacionalidad y hallándose, a consecuencia de tales acontecimientos fuera del país donde antes tuviera su residencia habitual, no pueda o, a causa de dichos temores, no quiera regresar a él.

² En este estudio se incluye el término «inmigrante(s)» junto al término «migrante(s)», pues en ambos casos se trata de personas que no poseen la condición de refugiado y que, por lo tanto, son percibidos por el público general como migrantes voluntarios.

Más allá de la inclusión de este término en el grupo de los «migrantes» a ojos de la Hipótesis 1, a lo largo de estas páginas se utilizarán los términos «inmigrante» y «extranjero» para hacer referencia a migrantes y refugiados de forma conjunta, pues ambos colectivos, más allá de su estatus legal, son inmigrantes y extranjeros en el país de acogida.

Los migrantes, por su parte, eligen trasladarse no a causa de una amenaza directa de persecución o muerte, sino en busca de una mejora en la calidad de sus vidas, principalmente por motivos sociales o económicos —que, no obstante, pueden resultar igualmente acuciantes, aunque no tengan la condición de refugiados—.

La importancia de esta diferencia radica, por tanto, en la protección internacional urgente y el asilo que demandan los refugiados, que, según acuerdos internacionales como la Convención de Ginebra, deben ser facilitados por los organismos nacionales de un país, ayudado por organizaciones supranacionales como ACNUR. A pesar de que conceder el estatus de refugiado a una persona puede obedecer a legislaciones o criterios que no siempre se corresponden con la realidad de cada individuo, la diferencia entre ambos grupos también se refleja en la aproximación hacia el fenómeno migratorio por parte de las sociedades receptoras, que tienden a mostrar un mayor apoyo a quienes perciben que han migrado involuntariamente —como los refugiados— que quienes lo han hecho voluntariamente —el caso de los migrantes— (Verkuyten, 2014). O'Rourke y Sinnott (2006) y Murray y Marx (2013) también apoyan esta distinción, pues, en términos generales, las personas tienden a ser menos hostiles hacia los refugiados que hacia los migrantes, sean «legales» o «ilegales» (Murray y Marx, 2013).

Con esto, es muy probable que la teoría del sesgo lingüístico intergrupual permita explicar parte de las diferencias en las formas en que se transmite el prejuicio cuando se comparan públicos con cargas de empatía distinta producidas por la cobertura mediática positiva (Park, 2012). Es decir, al ser la empatía mayor —como en el caso de los refugiados debido al tratamiento mediático de victimización—, es probable que la expresión del rechazo sea más vaga o abstracta, mientras que en los casos de menor empatía —como en el caso de los migran-

tes cuyo tratamiento mediático es de carga negativa para los países— el rechazo será más evidente y manifiesto.

Esto ha sido probado en países como Estados Unidos o Países Bajos, observando que la sociedad de acogida tiende a considerar que los refugiados no tienen alternativa —y, por lo tanto, se les considera víctimas inocentes— y son menos rechazados y más apoyados que los migrantes, pues se entiende que estos se desplazan voluntariamente (Verkuyten *et al.*, 2018; O'Rourke y Sinnott, 2006). Estas investigaciones nos permiten asumir que también encontraremos esta actitud en el contexto hispanohablante:

H1. El rechazo en Twitter es más frecuente hacia migrantes que hacia refugiados.

MÉTODO

Este estudio cuantitativo tiene alcance descriptivo y correlacional, y se basa en el análisis de contenido y en la clasificación automatizada de textos basada en el aprendizaje automático supervisado (*supervised machine learning*). Esta técnica se aplicó a mensajes de Twitter, de manera que cada tuit conformaba una unidad de análisis. Por su parte, la clasificación automatizada de textos es una técnica de *big data* que usa algoritmos de clasificación para generar modelos predictivos basados en un conjunto de ejemplos previamente etiquetados con diferentes técnicas, incluido el propio análisis de contenido. El trabajo tiene dos etapas: la primera, de análisis de contenido manual, sirvió para responder a las tres preguntas de investigación y a la hipótesis, además de para elaborar el modelo de clasificación de textos utilizado en la segunda etapa de análisis automatizado a gran escala. En esta segunda fase se da respuesta, con un volumen de datos mucho mayor, a la PI1 y a

la H1, que constituyen los dos elementos principales del estudio, permitiendo complementar y comparar los resultados de ambas fases. Dado su carácter exploratorio y complementario, las Preguntas de Investigación 2 y 3 son respondidas únicamente en la primera etapa.

Análisis manual de contenidos

Muestra y procedimiento

La primera descarga de los tuits se realizó en el entorno integrado Pycharm, a través de la herramienta Autocop (Arcila-Calderón *et al.*, 2017), conectada a la Application Programming Interface (API) de Twitter, que permite descargar tuits en tiempo real (*streaming*) o del historial (*rest*). En este caso se utilizó el API *streaming*, que descarga todos los tuits publicados en la red en cualquier lugar del mundo en el idioma seleccionado³ y que contengan una palabra clave determinada durante todo el tiempo que la herramienta esté activa. Durante los meses de abril y mayo de 2018 se descargaron de manera aleatoria 4.000 tuits en español con el único requisito de que incluyeran alguna de las palabras clave: «refugiado», «refugiados», «migrante», «migrantes», «inmigrante» e «inmigrantes».

Inicialmente se filtraron los 4.000 tuits, eliminando aquellos que utilizaran las palabras clave en otro contexto que el de la migración de personas, los tuits repetidos, los que no tuviesen sentido lógico, aquellos cuyo signi-

³ La herramienta detecta el idioma declarado en el JSON —el lenguaje de marcado del tuit— y descarga aquellos contenidos que cumplan este requisito y que incorporen las palabras clave introducidas tanto en el cuerpo del tuit como en los elementos que incorpora (imágenes, enlaces, etc.), por lo que es posible que un usuario que tenga su cuenta configurada en español introduzca uno de los términos seleccionados en un mensaje redactado en otro idioma o dialecto. Estos mensajes, que también fueron descargados por la herramienta, fueron eliminados en la posterior fase de limpieza.

ficado dependiera de un hipervínculo o una imagen, los redactados en otros idiomas y aquellos que solo contuvieran emoticonos o menciones a otros usuarios, obteniendo una muestra final de 1.469 tuits.

Medidas

Los tuits fueron clasificados manualmente por dos codificadores entrenados siguiendo las siguientes medidas:

a) *Sentido o presencia de expresiones de rechazo*: esta clasificación, la principal del estudio, exige la comprensión del sentido del tuit para conocer la actitud hacia los migrantes o refugiados, especialmente el *rechazo*, variable principal del estudio. Se codificó en las categorías *rechazo*, *aceptación*, o, si no se asumía ninguna postura, *neutral*, de manera que la *aceptación* y la *neutralidad* implicaban ausencia de rechazo. Aun cuando un tuit es informativo, este puede despertar aceptación o rechazo en función de su contenido si, por ejemplo, comparte declaraciones en uno u otro sentido. Se codificaron como *neutral* aquellos tuits en los que no se puede detectar si la opinión o la información promueve o expresa la aceptación del migrante o refugiado, o el rechazo. Conviene destacar que hay tuits que expresan solidaridad o compasión, sin embargo, si no se asume una actitud de defensa, acogida o se exigen derechos y acciones, no fueron considerados como *aceptación*, si no como *neutral*, puesto que la compasión no implica necesariamente la aceptación del inmigrante, sino su victimización, ya que la compasión no impide por sí misma que se considere, al mismo tiempo, al inmigrante como una carga para el Estado. En *aceptación*, por lo tanto, se incluyeron los tuits que manifestaron acogida, bienvenida, integración o defensa. Se identifica como *rechazo* los tuits que

reflejen la no aceptación de migrantes o refugiados, o la asociación de estos con aspectos negativos, como vincularlos con la delincuencia, con una carga económica, con una invasión o avalancha, con el empobrecimiento, etc. También es rechazo cuando se utiliza el término refugiado, migrante o inmigrante de manera despectiva o como un insulto. Siguiendo la línea del Cuestionario de Discriminación Étnica Percibida (Contrada *et al.*, 2001), se trata de contenidos que expresen rechazo a través de comentarios ofensivos contra una persona o grupo, o a través de la utilización de nombres peyorativos o descalificativos, pero sin exigir la presencia de discurso de odio ni de lenguaje ofensivo (Davidson *et al.*, 2017).

b) *Asociaciones negativas que justifican el rechazo*. Esta categoría indaga en el o los motivos asociados con el rechazo a los migrantes o refugiados. Se midió la presencia o ausencia de seis indicadores de rechazo, que se construyeron *ad hoc* para este estudio gracias a los ítems utilizados por Díez Nicolás (2009), Wike *et al.* (2016) y Cea D'Ancona (2009) en sus estudios, y que fue contrastada con las ideas que Gualda y Rebollo (2016) y Rebollo y Gualda (2017) observaron que las poblaciones autóctonas asocian con los inmigrantes. Estos indicadores constituyen, por lo tanto, una combinación y síntesis de los anteriores trabajos, agrupando en seis categorías las posibles expresiones de rechazo a los inmigrantes. El objetivo es descubrir, de estas seis grandes asociaciones, cuáles son más habituales y, por lo tanto, en qué aspectos se debe incidir para reducir ese rechazo. Se intentó primar el indicador que fuera predominante en cada tuit; sin embargo, ante la imposibilidad de seleccionar uno solo, en algunos textos se seleccionó más de un argumento, lo que generó que el porcentaje acu-

mulado de indicadores fuera superior al 100% de tuits que expresan rechazo. Los indicadores utilizados son los siguientes:

- *Carga económica* (1). Si indica que los extranjeros suponen un esfuerzo económico para el Estado o sus ciudadanos, quitando beneficios sociales o puestos de trabajo que corresponderían a los nacionales de ese país; si expresa desacuerdo con que se concedan ayudas a extranjeros, y también si considera que los extranjeros están en posición de ventaja en comparación con los nacionales a la hora de recibir apoyo estatal.
 - *Amenaza a la seguridad* (2). Si se considera que los inmigrantes son violentos, responsables de la inseguridad o que representan un peligro de cualquier clase, especialmente terrorismo.
 - *Amenaza a la identidad* (3). Se muestra que los inmigrantes amenazan la cultura del país imponiendo las de sus países de origen y se teme que la inmigración acabe provocando que el país de destino pierda su identidad, o cuando se muestra contrariedad por la «imposición» de las prácticas religiosas o creencias de los inmigrantes. También cuando se habla de multiculturalismo de una manera negativa.
 - *Amenaza de invasión* (4). Se detecta por la presencia de palabras como «manada», «ola», «miles», «millones», «invasión», «avalancha», haciendo referencia a la gran cantidad de migrantes y refugiados. Se considera que hay «muchos» o «demasiados» migrantes, y que estos deberían ser expulsados o que se deberían fortalecer las fronteras, pero no por causas concretas, sino por el miedo a ser invadido.
 - *Rechazo manifiesto* (5). Cuando las palabras expresan rechazo u hostilidad explícita, sin especificar motivos y/o se utiliza las palabras «inmigrante» o «refugiado» como un insulto, de manera despectiva o de forma negativa, con expresiones como «maldito refugiado», «inmigrante tenía que ser», etc. Esta categoría, en la que los inmigrantes son rechazados por su condición de tales, es la que más semejanza guarda con las expresiones que incluyen discurso de odio y es, por lo tanto, la más peligrosa.
 - *Prejuicio social* (6). Cuando se destaca que la presencia de los inmigrantes daña el «ambiente» de la ciudad o país, se menciona el estatus de pobreza o de clase social del refugiado, su educación, aptitudes, etc. En línea con los planteamientos de Cortina (2017), no se desprecia al inmigrante por su condición de tal, sino por su clase social.
- c) *Tipo*: categoría formal que distingue entre mensajes informativos o de opinión. Como la anterior, son variables exploratorias que buscan ampliar el análisis principal; esta categoría tiene especial relevancia por la necesidad de distinguir entre hechos y opiniones en la ola de posverdad actual, según la que se concede mayor peso a las opiniones que a los hechos en la construcción de la realidad (Oxford Dictionaries, 2016). Si se reconoce que es de tipo noticioso, objetivo, difunde invitaciones o convocatorias, presenta datos o estadísticas, se codificó como *Informativo*. Si es de tipo personal, con un carácter subjetivo, con más adjetivación, se codificó como *Opinión*. Se consideró también opinión una expresión de una cuenta de una organización u colectivo si en esta se expresa con una postura personal o subjetiva, o

cuando se hacen valoraciones sobre algún tipo de información.

Para garantizar la fiabilidad de las medidas se realizó una prueba de intercodificador con una muestra aleatoria de 150 mensajes (~10% de la muestra total). Se utilizaron los estadísticos Kappa de Cohen y Alpha de Krippendorff (medidas de 0 a 1, donde 1 refleja el máximo acuerdo). Como se aprecia en la tabla 1,

los valores de ambas pruebas son cercanos o superiores a 0,7, lo que demuestra una fiabilidad adecuada. Solo la presencia de prejuicios por la clase social como justificación del rechazo al migrante o refugiado obtuvo valores algo menores pero, al tratarse de una variable menos clara de apreciar y por acercarse al 0,6 que Neuendorf (2002) demanda en las investigaciones exploratorias, se mantuvo en el estudio.

TABLA 1. *Fiabilidad de las medidas*

Variable	Kappa de Cohen	Alpha de Krippendorff
Sentido o presencia de rechazo	0,778	0,777
Argumento: carga económica	0,754	0,753
Argumento: amenaza a la seguridad	0,784	0,784
Argumento: amenaza de invasión	0,680	0,681
Argumento: amenaza a la identidad	0,688	0,688
Argumento: rechazo manifiesto	0,687	0,687
Argumento: prejuicio social	0,580	0,580
Tipo	0,766	0,766
Media	0,715	0,715

Fuente: Elaboración propia.

Análisis computacional a gran escala

En una segunda etapa, se utilizaron métodos computacionales para ampliar la muestra original y escalar el estudio inicial con un enfoque de *big data*, que permitiera responder a la PI1 y a la H1 con mayor propiedad. Para ello, se utilizaron los 1.469 mensajes clasificados en la primera etapa como ejemplos para generar un modelo predictivo con técnicas de aprendizaje automático supervisado que permitieran estimar la probabilidad de cada nuevo tuit de pertenecer a la clase *rechazo* (el 45%, según la codificación manual explicada en los resultados) o *aceptación/neutral* (55%). Esta división se realizó para centrar el análisis en el *rechazo*, como opo-

sición a la neutralidad o a la aceptación, pues es esa categoría la que nos interesa medir, ya que podría dar lugar a un potencial discurso de odio e, incluso, la que estaría detrás de sentimientos xenófobos o racistas.

Con esta intención, se utilizaron las librerías NLTK y SciKit-Learn de Python para generar modelos de clasificación binaria con la presencia de expresiones de rechazo como categoría de referencia, utilizando seis algoritmos habitualmente aplicados a la clasificación de textos —Naive Bayes original, Naive Bayes para modelos multimodales, Naive Bayes para modelos multivariados Bernoulli, Regresión logística, Regresión logística con gradiente descendente estocástico y Máquinas de vectores

soporte con estimador SVC—. También se aplicaron técnicas de procesamiento de lenguaje natural (*natural language processing*, NLP) para extraer las características del conjunto de mensajes etiquetados.

Con estas herramientas, se procedió en primer lugar a limpiar los caracteres extraños, como emoticonos, símbolos no lingüísticos o grafías de otros alfabetos, y se convirtió todo el texto a letras minúsculas. Luego se entrenó un modelo en castellano para el etiquetado de partes de discurso (*parts of speech*, POS) basado en el corpus de ejemplo *es-cast31b* del módulo *cess-esp* contenido en la librería NLTK, que permitió seleccionar solo las palabras de los tuits que fueran adjetivos, verbos o sustantivos. Las 5.000 palabras más repetidas de estos tipos se «tokenizaron» y se convirtieron en características cuantitativas (vectores) de los ejemplos para poder generar los modelos predictivos.

Los 1.469 mensajes se dividieron aleatoriamente en dos grupos: 70% para el corpus de entrenamiento y 30% para el corpus de prueba. Se generaron clasificadores optimizados para cada uno de los seis algoritmos mencionados y se implementaron sobre el corpus de entrenamiento con el fin de generar seis modelos de clasificación. Con esto, se generó un clasificador basado

en el voto de cada uno de los seis modelos generados con un indicador de confianza basado en el grado de acuerdo de los modelos para cada predicción. Así, el clasificador elige la categoría —*aceptación/neutral o rechazo*— que la mayoría de los modelos haya predicho —si hay empate, lo hace aleatoriamente—, añadiendo un indicador de confianza basado en la proporción de dicho acuerdo (número de votos para la clase mayoritaria/Número de votos posibles), lo que permitió establecer un umbral de confianza superior a 0,8 (80%) para cada predicción.

Cada uno de los seis clasificadores, además del basado en la votación de los otros modelos, fue evaluado utilizando el corpus de prueba para comparar las etiquetas originales con las clasificaciones producidas por los modelos creados. Se utilizaron las métricas de evaluación clásicas en aprendizaje automático supervisado (Kelleher *et al.*, 2015): la exactitud (*accuracy*), la precisión (*precision*), el recuerdo (*recall*) y la media armónica (*F-score*). En la tabla 2 se puede apreciar cómo todos los valores estuvieron significativamente por encima de la línea base de 55%; en especial, el clasificador basado en la votación de los modelos obtuvo una exactitud del 76,19%, lo que se traduce en un adecuado poder predictivo de los modelos.

TABLA 2. Métricas de evaluación de los modelos

Algoritmos	Accuracy	Precision	Recall	F-score
Naive Bayes original	75,36%	78,10%	76,92%	77,50%
Naive Bayes para modelos multimodales	74,64%	79,88%	72,23%	75,86%
Naive Bayes para modelos multivariados Bernoulli	72,15%	68,80%	90,62%	78,22%
Regresión logística	74,53%	73,56%	84,05%	78,46%
Regresión lineal con gradiente descendente estocástico	71,84%	73,18%	77,30%	75,18%
Máquinas de vectores soporte (estimador SVC)	75,36%	76,11%	80,68%	78,32%
Clasificador basado en la votación de los modelos	76,19%	73,18%	77,30%	75,18%

Fuente: Elaboración propia.

Los resultados de cada modelo fueron almacenados en formato *pickle* con el fin de escalar la investigación inicial y así analizar una muestra de gran tamaño con métodos computacionales. Específicamente, como en la descarga para el análisis manual, se utilizó la API *streaming* de Twitter y, enlazándola con los modelos de clasificación de textos entrenados a partir de los mensajes etiquetados manualmente, se recogieron automáticamente todos los tuits en español producidos entre el 19 y el 29 de julio de 2019 que incluyeran las mismas palabras clave de la recolección inicial («refugiado», «refugiados», «migrante», «migrantes», «inmigrante» e «inmigrantes»). Durante este periodo se descargaron y clasificaron en tiempo real un total de 337.116 mensajes.

Finalmente, utilizando el clasificador basado en la votación de los seis modelos generados a partir de los mensajes etiquetados manualmente, y estableciendo

un nivel de confianza del 0,8 para cada predicción, se procedió al análisis. De los 337.116 mensajes recogidos, 187.305 fueron clasificados como *rechazo* o como *aceptación/neutral* con el mínimo de confianza establecido por el estudio.

RESULTADOS

Análisis manual

Respondiendo de manera preliminar a la PI1 sobre la presencia de rechazo en Twitter en español hacia el colectivo de migrantes y refugiados (tabla 3), encontramos que, de los 1.469 tuits codificados, en el porcentaje más alto de mensajes (45%) se encontraron expresiones de *rechazo* en cualquiera de sus dimensiones. En un 16,7% de los tuits se expresó *aceptación* hacia el colectivo, mientras que un 38,3% de los tuits fueron mensajes *neutros*.

TABLA 3. Clasificación manual del sentido de los tuits

	Frecuencia	Porcentaje	Porcentaje válido
Sentido	Aceptación	245	16,7
	Rechazo	660	44,9
	Neutral	561	38,2
	Total	1.466	99,8
Perdidos	3	0,2	
Total	1.469	100,0	

Fuente: Elaboración propia.

Con respecto a la segunda pregunta de investigación (PI2) sobre las problemáticas o aspectos negativos asociados al rechazo hacia los migrantes o refugiados (tabla 4), se obtuvieron los siguientes resultados: en el 35,6% de los tuits en los

que se mostraba rechazo, este se manifestaba de forma explícita de manera hostil; en el 30,2% se debía a que los migrantes o refugiados eran asociados con una amenaza para la seguridad; en el 26,1% el rechazo se debía a que estos colecti-

vos eran percibidos como una carga económica; la amenaza de invasión se mencionó en un 17% de los tuits; un 7,4% de los textos mostraban un prejuicio social; y el 4,8% de quienes mostraron rechazo lo hicieron por sentir amenazada su identidad.

Cabe destacar, de nuevo, la posibilidad de que algunos tuits incluyeran más de un

argumento o asociación para rechazar a migrantes y refugiados, motivo por el que el total asciende a 121,1%; lo más frecuente es que se mencionaran dos (57,1%) o tres (22,4%) aspectos negativos en el mismo texto. En el 14% de los tuits de rechazo no fue posible identificar de forma clara la existencia de ninguna de las problemáticas predeterminadas.

TABLA 4. Codificación de las asociaciones que justifican el rechazo

	Frecuencia	Porcentaje	Porcentaje válido	
	Carga económica	172	26,1	26,2
	Amenaza a la seguridad	199	30,2	30,2
	Amenaza de invasión	112	17,0	17,0
Justificación	Amenaza a la identidad	32	4,8	4,8
	Rechazo manifiesto	235	35,6	35,7
	Prejuicio social	49	7,4	7,4
	Total	799	121,1	121,3

Fuente: Elaboración propia.

Con respecto al *tipo* de mensaje, los resultados mostraron que en el 75,7% de los tuits analizados se mostraba una *opinión* o posición acerca del tema, frente a un 24,3% de tuits que fueron expresados de manera *informativa* o noticiosa. Respondiendo a la PI3, encontramos una clara asociación entre el *tipo* de tuit y el *sentido* del mismo [$\chi^2(2, 1464) = 208,972, p < 0,001$]: así, los tuits *informativos* tienen mayor probabilidad de ser clasificados como tuits *neutros*; mientras que los tuits de *rechazo* tienden a pertenecer a publicaciones de *opinión*. Los residuos tipificados señalan que existe una mayor probabilidad de que el mensaje sea de rechazo cuando se trata de un tuit informativo (13,6 > 3,29) y que la probabilidad

de que sean neutros (13,1 > 3,29) es significativamente mayor cuando se trata de tuits informativos, algo coherente con el predominio de los mensajes provenientes de medios de comunicación que, en su mayoría, muestran un sentido neutral. Con esto, la asociación entre el tipo de tuit y el sentido es significativa y débil: $|\Phi| = 0,378, p < 0,001$. De hecho, si estudiamos únicamente los 1.108 tuits considerados de *opinión*, asumiendo que una gran parte de los tuits informativos proceden de medios de comunicación y que el discurso de odio se nutre más de opiniones que de datos —enmarcado en la ola de posverdad contemporánea—, encontramos que 610 textos, esto es, un 55%, son de rechazo.

TABLA 5. Sentido de aceptación o rechazo según el tipo de tuit

			Tipo de tuit	
			Opinión	Informativo
Sentido	Aceptación	Recuento	178	66
		% del total	16,1	18,5
	Rechazo	Recuento	610	49
		% del total	55,0	13,8
	Neutral	Recuento	320	241
		% del total	28,9	67,7

Fuente: Elaboración propia.

En cuanto a la H1, se pudo observar una mayor presencia de rechazo cuando se trataba de migrantes que de refugiados, mientras la aceptación o la ausencia de un sentimiento explícito fueron mayores hacia el colectivo de refugiados (tabla 6). Observamos que 474 mensajes (un 32,3% de la muestra) hacían alusión a refugiados —de manera singular o plural—, mientras 994 (el 67,7%) se referían a migrantes —tanto en singular como en plural—. Las pruebas estadísticas mostraron que las diferencias entre el sentido del tuit y el tipo de inmigrante al que hace referencia fueron significativas [$\chi^2(2, 1465) = 145,815, p < 0,001$], siendo más probable que los tuits de rechazo estuvieran dirigidos a migrantes que a refugiados. Así, los residuos tipificados señalan que existe una mayor probabilidad de que el mensaje incluya expresiones de rechazo cuando se menciona a los migrantes ($12,1 > 3,29$) y que la probabilidad de

que sean neutros ($5,2 > 3,29$) o de aceptación ($8,4 > 3,29$) es significativamente mayor cuando se trata de refugiados. Por lo tanto, la asociación que existe entre la condición del extranjero y el volumen de rechazo es significativa y débil: $|\Phi| = 0,315, p < 0,001$.

Podemos añadir que la percepción de los refugiados como una carga económica [$t(171,135) = -2,977, p < 0,01, d = 0,46$] y como una amenaza a la seguridad [$t(157,788) = -2,186, p < 0,05, d = -0,35$] es significativamente menor que cuando se trata de migrantes. De hecho, los migrantes son asociados con una carga económica en el 28% de los tuits de rechazo, mientras que solo un 16% de los tuits de rechazo hacia refugiados se refiere a esta condición. Por su parte, el 32% de los textos con expresiones de rechazo hacia migrantes incluye el componente de riesgo para la seguridad, por el 22% de los tuits centrados en refugiados.

TABLA 6. Sentido de aceptación o rechazo hacia los distintos tipos de inmigrantes

			Tipo de inmigrante	
			Refugiado	Migrante
Sentido	Aceptación	Recuento	114	131
		% del total	24,1	13,2
	Rechazo	Recuento	106	554
		% del total	22,4	55,9
	Neutral	Recuento	254	306
		% del total	53,6	30,9

Fuente: Elaboración propia.

Análisis computacional

En esta segunda etapa se descargaron y analizaron 337.116 tuits producidos entre el 19 y el 29 de julio de 2019 y que hacían referencia a inmigrantes/migrantes o refugiados; 187.305 de ellos fueron clasificados con un nivel de confianza superior al 80%. Esta etapa pretendía ampliar el análisis inicial y poner a prueba en otro periodo temporal la principal pregunta de investigación e hipótesis, esto es, se quería conocer con una muestra de gran tamaño cuál es el volumen de tuits que expresan rechazo hacia inmigrantes/migrantes y refugiados en Twitter en español y comprobar si dicho rechazo era mayor hacia los migrantes que hacia los refugiados.

Como vemos en la tabla 7, y respondiendo a la PI1, el porcentaje de rechazo hacia migrantes y refugiados se situó en un 9,19% de los mensajes en los que se mencionaba a estos colectivos. Se trata

de una cifra mucho menor a la encontrada a pequeña escala quince meses antes, lo que refleja las fluctuaciones sobre este tipo de mensajes en función de eventos detonadores o de climas de opinión. Adicionalmente, debemos tener en cuenta que, al modelar las expresiones de rechazo frente a las otras categorías y controlar fundamentalmente el error tipo I (falsos positivos), el sesgo del algoritmo se inclina a clasificar como *rechazo* solo aquellos mensajes de los que esté completamente seguro. Al analizar la clasificación del total de mensajes (N = 337.116), sin tomar en cuenta o no la confianza o acuerdo entre los modelos, observamos que el porcentaje de rechazo asciende al 26,68%, lo que demuestra, precisamente, que, al elegir *rechazo* como categoría de referencia, el clasificador solo incluye en dicha categoría los mensajes de los que esté más seguro, mientras que en la otra categoría entrarían todos los demás.

TABLA 7. Clasificación de los mensajes a gran escala

		Frecuencia	Porcentaje
Sentido	Aceptación / Neutral	170.084	90,81
	Rechazo	17.221	9,19
	Total	187.305	100,0
Perdidos		0	0
Total		187.305	100,0

Fuente: Elaboración propia.

Seguidamente llevamos a cabo dos pruebas estadísticas clásicas para comprobar si existía alguna asociación entre la expresión verbal de rechazo y el tipo de colectivo, respondiendo así a la H1. Para ello, añadimos de forma automática dos variables a cada mensaje, que reflejaban, por un lado, si el tuit contenía (1) o no (0) las palabras «migrante», «migrantes», «inmigrante» e «inmigrantes»;

y, por otro, si incluía (1) o no (0) las palabras «refugiado» y «refugiados». Estas variables categóricas se cruzaron por medio de una tabla de contingencia con la variable sentido del tuit, es decir, si expresaba rechazo o aceptación/neutralidad. Las pruebas estadísticas revelaron que existe una asociación significativa entre el rechazo y la mención a migrantes [$\chi^2(1, 187.305) = 9.828,634, p < 0,001$]; y

el rechazo y la mención a refugiados [$\chi^2(1, 187.305) = 3.138,518, p < 0,001$]. Al analizar la primera asociación, los residuos tipificados señalan que existe una mayor probabilidad de que el mensaje sea de rechazo cuando se menciona a los migrantes ($140,8 > 3,29$); mientras que de la siguiente asociación se desprende que dicha probabilidad es menor cuando se menciona a refugiados ($-56,0 < -3,29$). En ambos casos podríamos hablar de una asociación significativa pero débil: $|\Phi| = 0,325, p < 0,001$ y $|\Phi| = -0,129, p < 0,001$, respectivamente. Ambas pruebas respaldan nuestra hipótesis de investigación con datos a gran escala y durante un periodo temporal diferente al primer estudio.

DISCUSIÓN Y CONCLUSIONES

La investigación ha mostrado una presencia notable, aunque fluctuante, de tuits que muestran rechazo hacia migrantes y refugiados. Conviene anotar que el porcentaje de rechazo encontrado en los mensajes estudiados no indica que en 2018 hubiera un 45% de personas de habla hispana racistas o xenóforas ni que en 2019 ese volumen hubiera caído al 9,19%, pero sí sugiere dos conclusiones clave: por un lado, la enorme variación que se puede observar en las expresiones de rechazo o aceptación de inmigrantes en las redes sociales en función de los últimos fenómenos mediáticos; y, por otro lado, la presencia de rechazo en redes sociales que en algunos casos puede estar basado en sentimientos y/o actitudes racistas o xenóforas y que en ocasiones es manifestada a través de discursos de odio, especialmente cuando se ampara en discursos oficiales (Crandall *et al.*, 2018) o en redes anónimas (Fox *et al.*, 2015), por lo que conviene seguir analizando esta materia.

Se ha apreciado, tanto en el estudio manual de 2018 como en el automatizado de

2019, que el rechazo mostrado hacia los migrantes es significativamente superior que hacia los refugiados, que son aceptados o retratados de una forma neutra de forma significativamente más frecuente. Esto coincide con lo apuntado por O'Rourke y Sinnott (2006), por Verkuyten *et al.* (2018) o por Verkuyten (2014), que observaron cómo se percibe que los refugiados no tienen opción y son menos rechazados y más apoyados que los migrantes, pues se considera que se trasladan voluntariamente o que no huyen de algo tan amenazante como una guerra; de esta forma se apoyan también las teorías sobre el sesgo lingüístico intergrupalo, por las que la distinta carga de empatía recibida por determinados públicos —mayor empatía en la cobertura mediática de refugiados que de migrantes—, implica menor rechazo (Park, 2012). No obstante, muchos de los estudios que se llevan a cabo actualmente para medir las actitudes hacia la inmigración no distinguen entre estos dos grupos, por lo que las diferencias encontradas en este estudio sugieren la importancia de continuar analizando esta diferencia y a cada grupo y sus características particulares por separado.

Al mismo tiempo, y con vocación exploratoria, se observó que lo habitual es que quienes rechazan a extranjeros lo hagan en la mayoría de dimensiones —como demuestra el hecho de que en varios tuits coincidieran más de una—. De estas dimensiones, las más frecuentes fueron la hostilidad manifiesta hacia el colectivo, la amenaza que se percibe que estas personas suponen para la seguridad y la consideración de estos colectivos como una carga económica. Se entiende que, al ser estas las asociaciones más visibles y las que más se expresan en redes sociales, las actuaciones más útiles, si se quiere contrarrestar este rechazo, deberían ir orientadas en esta dirección. Dado que esta parte de la investigación tiene un carácter preliminar, también debemos hacer hincapié en la necesidad de seguir ampliando el conoci-

miento sobre las causas que hay detrás del rechazo al extranjero para poder enfrentarlo de la manera más adecuada, tanto en los estadios iniciales que se analizan en este texto como en las formas de rechazo más dañinas que se manifiestan a través de discursos o crímenes de odio.

Por último, es importante destacar que nuestros resultados pueden servir para explorar nuevos mecanismos de detección del discurso de odio analizando el rechazo verbal en la red, especialmente en el entorno hispanoparlante. Este trabajo apunta hacia Twitter y, en consecuencia, hacia otros medios sociales, como fuentes de información valiosas para el análisis de la opinión pública y de las actitudes ciudadanas, en los casos en los que el estudio a través de encuestas resulta en gran medida limitado. En esta línea, una de las mayores aportaciones de este trabajo es la generación de un corpus de ejemplos de aceptación y rechazo hacia migrantes y refugiados⁴ con el que se pueda entrenar modelos con técnicas de aprendizaje automático supervisado, para permitir la detección automática y a gran escala de dichos discursos.

LIMITACIONES E INVESTIGACIÓN FUTURA

Este estudio, aunque utiliza información extraída de Twitter, todavía utiliza una muestra limitada. Tanto la finitud del número de tuits como su extracción en dos contextos temporales concretos —entre abril y mayo de 2018 y en julio de 2019— impiden una extrapolación absoluta, pues, como se ha podido observar, las muestras de rechazo y de aceptación fluctúan notablemente en función de fenómenos puntuales que influyen en la opinión pública. No

⁴ Dicho corpus está disponible con acceso abierto en el enlace: <https://github.com/carlosarcila/rejection>

obstante, junto a la lectura de la situación en estos dos momentos, el trabajo también incorpora un análisis preliminar de los aspectos negativos asociados con el rechazo y de la relación entre el tipo de tuit y la expresión de rechazos.

Al mismo tiempo, dado que no se geolocalizaron los contenidos, es imposible realizar una lectura más detallada por países, algo que, por otra parte, no es sencilla en los análisis que utilizan Twitter como fuente, dado que muy poca gente hace pública su localización —según Gaffney y Puschman (2014) tan solo un 1% del tráfico en Twitter se geoetiqueta— y gran cantidad de estos datos están protegidos por el medio. Así, los datos obtenidos son preliminares y generales, pero ofrecen una herramienta para que futuros trabajos puedan comparar el rechazo al extranjero en distintos contextos hispanohablantes. Por otro lado, también se excluyeron imágenes, hipervínculos y tuits compuestos únicamente por emoticonos, limitando el estudio al análisis textual.

La utilización de medios sociales en el análisis científico implica algunas limitaciones por las dificultades técnicas de las interfaces. Un ejemplo de estas dificultades, que también menciona Chaudhry (2015), es el API de Twitter, que a las cuentas gratuitas de desarrollador, como la utilizada en el estudio, solo le ofrece acceso al 1% de todos los tuits publicados descargados desde el *stream* o limita los accesos al *rest* a los últimos siete días. No obstante, la cantidad de datos es enormemente mayor de lo que cualquier otro modelo analógico de recolección de datos permitiría. Asimismo, dada la composición de usuarios de redes sociales, y de Twitter en concreto, es imposible generalizar las conclusiones de los estudios que utilicen estas plataformas como fuente de datos, ya que ciertos grupos sociodemográficos, especialmente los grupos de mayor edad, apenas están representados.

BIBLIOGRAFÍA

- Arcila-Calderón, Carlos; Ortega-Mohedano, Félix; Jiménez-Amores, Javier y Trullenque, So-fía (2017). «Análisis supervisado de sentimientos políticos en español: clasificación en tiempo real de tuits basada en aprendizaje automático». *El profesional de la información*, 26(5): 973-982. doi: 10.3145/epi.2017.
- Bakir, Vian y McStay, Andrew (2018). «Fake News and the Economy of Emotions». *Digital Journalism*, 6(2): 154-175. doi: 10.1080/21670811.2017.1345645
- Bartlett, Jamie; Reffin, Jeremy; Rumbale, Noelle y Williamson, Sarah (2014). *Anti-social media*. London: Demos.
- Ben-David, Anat y Matamoros-Fernández, Ariadna (2016). «Hate Speech and Covert Discrimination on Social Media: Monitoring the Facebook Pages of Extreme-right Political parties in Spain». *International Journal of Communication*, 10: 1167-1193. Disponible en: <https://ijoc.org/index.php/ijoc/article/view/3697/1585>, acceso el 17 de diciembre de 2019.
- Berger, Peter L. y Luckmann, Thomas (1966). *The Social Construction of Reality*. New York: Random House.
- Billig, Michael (2002). «Henri Tajfel's "Cognitive Aspects of Prejudice" and the Psychology of Bigotry». *British Journal of Social Psychology*, 41(2): 171-188. doi: 10.1348/014466602760060165
- Bourhis, Richard V. y Dayan, Joelle (2004). «Acculturation Orientations towards Israeli Arabs and Jewish Immigrants in Israel». *International Journal of Psychology*, 39(2): 118-131. doi: 10.1080/00207590344000358
- Brewer, Marilyn B. (1999). «The Psychology of Prejudice: Ingroup Love and Outgroup Hate?». *Journal of Social Issues*, 55(3): 429-444. doi: 10.1111/0022-4537.00126
- Brown, Rupert (2000). «Social Identity Theory: Past Achievements, Current Problems and Future Challenges». *European Journal of Social Psychology*, 30(6): 745-778. doi: 10.1002/1099-0992(200011/12)30:6<745::AID-EJSP24>3.0.CO;2-O
- Burnap, Pete y Williams, Matthew L. (2015). «Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making». *Policy & Internet*, 7(2): 223-242. doi: 10.1002/poi3.85
- Cea D'Ancona, María Ángeles (2009). «La compleja detección del racismo y la xenofobia a través de encuesta. Un paso adelante en su medición». *Revista Española de Investigaciones Sociológicas (REIS)*, 125: 13-45. Disponible en: http://reis.cis.es/REIS/PDF/REIS_125_011231144723167.pdf, acceso el 28 de agosto de 2019.
- Chaudhry, Irfan (2015). «Hashtagging Hate: Using Twitter to Track Racism Online». *First Monday*, 20(2). doi: 10.5210/fm.v20i2.5450
- Conrada, Richard J.; Gary, Melvin L.; Coups, Elliot; Egeth, Jill D.; Sewell, Andrea; Ewell, Kevin; Goyal, Tanya M. y Chasse, Valerie (2001). «Measures of Ethnicity-Related Stress: Psychometric Properties, Ethnic Group Differences, and Associations with Well-being». *Journal of Applied Social Psychology*, 31: 1775-1820. doi:10.1111/j.1559-1816.2001.tb00205.x
- Cortina, Adela (2017). *Aporofobia, el rechazo al pobre: un desafío para la democracia*. Madrid: Paidós.
- Crandall, Christian S.; Miller, Jason M. y White, Mark H. (2018). «Changing Norms Following the 2016 US Presidential Election: The Trump Effect on Prejudice». *Social Psychological and Personality Science*, 9(2): 186-192. doi: 10.1177/1948550617750735
- Davidson, Thomas; Warmsley, Dana; Macy, Michael y Weber, Ingmar (2017). «Automated Hate Speech Detection and the Problem of Offensive Language». En: *Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017)*. Disponible en: http://sdl.soc.cornell.edu/img/publication_pdf/hatespeechdetection.pdf, acceso el 17 de diciembre de 2019.
- Díez Nicolás, Juan (2009). «Construcción de un índice de Xenofobia-Racismo». *Revista del Ministerio de Trabajo e Inmigración*, 80: 21-38. Disponible en: http://www.mitramiss.gob.es/es/publica/pub_electronicas/destacadas/revista/numeros/80/est01.pdf, acceso el 20 de agosto de 2019.
- European Commission against Racism and Intolerance (2016). *ECRI General Policy Recommendation N.º 15 on Combating Hate Speech*. Strasbourg: European Council.
- Fox, Jesse, Cruz, Carlos y Lee, Ji Young (2015). «Perpetuating Online Sexism Offline: Anonymity, Interactivity, and the Effects of Sexist Hashtags on Social Media». *Computers in Human Behavior*, 52: 436-442. doi: 10.1016/j.chb.2015.06.024

- Gaffney, Devin y Puschmann, Cornelius (2014). «Data collection on Twitter». En: Bruns, A.; Weller, K.; Burgess, J.; Mahrt, M. y Puschmann, C. (eds.). *Twitter and Society*. New York: Peter Lang.
- Gallego, Mar; Gualda, Estrella y Rebollo, Carolina (2017). «Women and Refugees in Twitter: Rhetorics on Abuse, Vulnerability and Violence from a Gender Perspective». *Journal of Mediterranean Knowledge*, 2(1): 37-58. doi: 10.26409/2017JMK2.1.03
- Gorham, Bradley W. (2006). «News Media's Relationship with Stereotyping: The Linguistic Intergroup Bias in Response to Crime News». *Journal of Communication*, 56(2): 289-308. doi: 10.1111/j.1460-2466.2006.00020.x
- Gualda, Estrella y Rebollo, Carolina (2016). «The Refugee Crisis on Twitter: A Diversity of Discourses at a European Crossroads». *Journal of Spatial and Organizational Dynamics*, 4(3): 199-212. Disponible en: <https://www.jsod-cieo.net/journal/index.php/jsod/article/view/72>, acceso el 20 de agosto de 2019.
- Gualda, Estrella; Borrero, Juan Diego y Cañada, José Carpio (2015). «La "Spanish Revolution" en Twitter (2): Redes de hashtags y actores individuales y colectivos respecto a los desahucios en España». *Revista Hispana para el Análisis de Redes Sociales, REDES*, 26(1): 1-22. doi: 10.5565/rev/redes.535
- Kalyanam, Janani; Quezada, Mauricio; Poblete, Barbara y Lanckriet, Gerts (2016). «Prediction and Characterization of High-Activity Events in Social Media Triggered by Real-World News». *PLoS one*, 11(12): e0166694. doi: 10.1371/journal.pone.0166694
- Kelleher, John D.; MacNamee, Brian y D'Arcy, Aoife (2015). *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. London: MIT Press.
- Kreis, Ramona (2017). «#refugeesnotwelcome: Anti-refugee Discourse on Twitter». *Discourse & Communication*, 11(5): 498-514. doi: 10.1177/1750481317714121
- Maass, Anne; Salvi, Daniela; Arcuri, Luciano y Semin, Gün R. (1989). «Language Use in Intergroup Contexts: The Linguistic Intergroup Bias». *Journal of Personality and Social Psychology*, 57(6): 981-993. doi: 10.1037/0022-3514.57.6.981
- Muller, Karsten y Schwarz, Carlo (2018). «Fanning the Flames of Hate: Social Media and Hate Crime». *SSRN*. doi: 10.2139/ssrn.3082972
- Murray, Kate E. y Marx, David A. (2013). «Attitudes toward Unauthorized Immigrants, Authorized Immigrants, and Refugees». *Cultural Diversity and Ethnic Minority Psychology*, 19(3): 332-341. doi: 10.1037/a0030812
- Naciones Unidas (1951). *Convención sobre el Estatuto de los Refugiados*. Disponible en: https://eacnur.org/files/convencion_de_ginebra_de_1951_sobre_el_estatuto_de_los_refugiados.pdf, acceso el 20 de agosto de 2019.
- Neuendorf, Kimberly A. (2002). *The Content Analysis Guidebook*. Thousand Oaks, California: Sage.
- O'Rourke, Kevin H. y Sinnott, Richard (2006). «The Determinants of Individual Attitudes towards Immigration». *European Journal of Political Economy*, 22(4): 838-861. doi: 10.1016/j.ejpoleco.2005.10.005
- Oxford Dictionaries (2016). *Word of the Year 2016 is...* Disponible en: <https://en.oxforddictionaries.com/word-of-the-year/word-of-the-year-2016>, acceso el 26 de agosto de 2019.
- Park, Sung-Yeon (2012). «Mediated Intergroup contact: Concept Explication, Synthesis, and Application». *Mass Communication and Society*, 15(1): 136-159. doi: 10.1080/15205436.2011.558804
- Peherson, Samuel; Brown, Rupert y Zagefka, Hanna (2011). «When Does National Identification Lead to the Rejection of Immigrants? Cross-sectional and Longitudinal Evidence for the Role of Essentialist in Group Definitions». *British Journal of Social Psychology*, 48(1): 61-76. doi: 10.1348/014466608X288827
- Rebollo, Carolina y Gualda, Estrella (2017). «La situación internacional de las personas refugiadas y su imagen en Twitter. Un reto para la intervención desde el trabajo social». *Documentos de Trabajo Social*, 59: 190-207. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=6588971>, acceso el 28 de agosto de 2019.
- Schäfer, Claudia y Schadauer, Andreas (2019). «Online Fake News, Hateful Posts Against Refugees, and a Surge in Xenophobia and Hate Crimes in Austria». En: Dell'Orto, G. y Wetzstein, I. (eds.). *Refugee News, Refugee Politics: Journalism, Public Opinion and Policymaking in Europe*. Oxford: Routledge.
- Verkuyten, Maykel (2014). *Identity and Cultural Diversity: What Social Psychology Can Teach Us*. Hove: Routledge.
- Verkuyten, Maykel y Brug, Peary (2004). «Multiculturalism and Group Status: The Role of Ethnic

- Identification, Group Essentialism and Protestant Ethic». *European Journal of Social Psychology*, 34(6): 647-661. doi: 10.1002/ejsp.222
- Verkuyten, Maykel; Mepham, Kieran y Kros, Matthijs (2018). «Public Attitudes towards Support for Migrants: The Importance of Perceived Voluntary and Involuntary Migration». *Ethnic and Racial Studies*, 41(5): 901-918. doi: 10.1080/01419870.2017.1367021
- Whitley Jr., Bernard E. y Kite, Mary E. (2016). *Psychology of Prejudice and Discrimination*. New York: Routledge.
- Wike, Richard; Stokes, Bruke y Simmons, Katie (2016). *Europeans Fear Wave of Refugees Will Mean More Terrorism, Fewer Jobs*. Disponible en: <https://immigrazione.it/docs/2016/Pew-Research-Center-July-11-2016.pdf>, acceso el 28 de agosto de 2019.

RECEPCIÓN: 13/03/2019

REVISIÓN: 10/07/2019

APROBACIÓN: 11/02/2020

