

www.uoc.edu/idp

Monográfico «Ciberdelincuencia y cibervictimización»

ARTÍCULO

Taxonomía de la comunicación violenta y el discurso del odio en Internet^{*}

Fernando Miró Llinares

Catedrático acreditado de Derecho penal

Universidad Miguel Hernández

Fecha de presentación: marzo de 2016

Fecha de aceptación: mayo de 2016

Fecha de publicación: junio de 2016

Resumen

Aunque el Derecho siempre se ha ocupado de las palabras ofensivas, lesivas, de odio o que incitan a la violencia, en los últimos años ha crecido la preocupación social relacionada con la capacidad del ciberespacio de incrementar los efectos nocivos de algunas expresiones o formas de comunicación violenta. El presente trabajo aborda la cuestión a partir de la observación y el análisis de más de 250.000 tuits publicados en español tras los ataques sufridos por los responsables y trabajadores del semanario satírico francés *Charlie Hebdo* en París. Tras varios cribados y la lectura de los tuits se han identificado las diferentes tipologías de comunicación con contenido violento y de odio en la red social Twitter, lo que ha permitido la construcción de una taxonomía de la comunicación violenta y el discurso del odio en Internet. Además, se analizan los resultados de prevalencia comprobando que la comunicación violenta es más bien testimonial incluso en eventos violentos como los analizados y que las formas más graves son menos representativas que otras formas de expresión de mal gusto o que afectan a la sensibilidad social.

Palabras clave

comunicación violenta, discurso del odio, ciberespacio, ciberodio, taxonomía de la comunicación violenta y del odio, *Charlie Hebdo*

Tema

Derecho penal y Criminología

* Artículo realizado en el marco del proyecto DER2014-53449-R titulado «Incitación a la violencia y discurso del odio en Internet. Alcance real del fenómeno, tipologías, factores ambientales y límites de la intervención jurídica frente al mismo», del MINECO.

Taxonomy of violent communication and the discourse of hate on the internet

Abstract

Even though the Law has always focused on words that are offensive and damaging, and which convey hate or incite to violence, during the last few years there has been a growing preoccupation with the capacity of cyberspace to augment the harmful effect of some expressions or forms of violent communication. This paper approaches the topic beginning with the observation and analysis of more than 250,000 tweets published in Spanish after the Paris attacks that victimized those responsible for the satirical French weekly *Charlie Hebdo* and its workers. After several screenings and readings of the tweets, the different typologies of communication with violent and hate content have been identified and this has allowed for the construction of a Taxonomy of Violent Communication and Hate Speech on the Internet. Also, the results of prevalence studies are analysed and they demonstrate firstly that the nature of violent communication on the Internet is testimonial, even in violent events such as those analysed; secondly that the most serious forms are less representative than other forms of expression of bad taste or that impact social sensibility, and finally that hate speech has a relevant presence even if it does not represent the majority of the expressions tweeted during the event.

Keywords

violent communication, hate speech, cyberspace, Taxonomy of Violent Communication and Hate Speech, *Charlie Hebdo*

Topic

Penal Law and Criminology

1. «Solo palabras»: acerca de la alarma social en torno a la comunicación violenta en el ciberespacio

Hoy no parece exagerado afirmar que Internet lo ha cambiado todo. La aparición y desarrollo de ese nuevo ámbito de intercomunicación ha dado lugar, en algunos casos, a fenómenos sociales que no preexistían. En otros, Internet simplemente ha modificado el evento, a veces sustancialmente, otras de modo apenas trascendente. Pero en pocas ocasiones la popularización del ciberespacio como ámbito de intercomunicación social no ha tenido efecto alguno en las relaciones, interacciones y construcciones de la socie-

dad, incluso en las que parecen estar alejadas de él y se dan en el espacio físico.

Antes de entrar a valorar cómo ha modificado Internet el fenómeno de la comunicación violenta y de odio (en adelante, CVyDO) podríamos aceptar al menos dos cosas: la primera, que Internet no ha creado tales formas de comunicación social; la segunda, que estas ya no son las mismas desde la aparición del ciberespacio. En cuanto a lo primero, la CVyDO existe desde mucho antes de la aparición de Internet. Los actos de habla que, desde una perspectiva comunicativa, pueden considerarse violentos, debieron de aparecer con las primeras relaciones sociales.¹ De hecho, al sistema jurídico penal siempre le ha interesado y preocupado la comunicación verbal anterior al hecho delictivo consumado o intentado y relacionada con él por su potencialidad para

1. Afirman los autores de *Words that wound* con ingenio, que la comunicación de odio, y lo mismo podría decirse de la comunicación violenta, pese a que haya habido un alarmante crecimiento de su incidencia, es tan antigua y tan americana como «la tarta de manzana». Lawrence III et al. (1993, pág. 1).

generar violencia física en el propio sujeto o en otros,² así como la propia capacidad de algunos actos de habla de constituir un daño o una ofensa susceptibles de ser castigados por sí mismos.³ Refiriéndonos en particular al discurso del odio, la propagación de concretas formas de expresión y comunicación, dirigidas contra grupos definidos por su raza, religión, orientación sexual, discapacidad, etnia, nacionalidad, edad, género, grupo social, afiliación política, o por otras características personales, funcionales o sociales, o simplemente contra quienes no comparten una ideología, se remonta a los orígenes de las propias sociedades. Y aunque tales expresiones de odio fueran mucho más antiguas, fue especialmente tras la Segunda Guerra Mundial, y debido a la aparición de propaganda racista en Estados Unidos y de propaganda sobre el Holocausto en muchos lugares del mundo, cuando surgieron las primeras regulaciones jurídicas del denominado discurso del odio primero a nivel nacional⁴ y luego en el plano internacional.⁵

Sin embargo, y como hemos dicho, aunque Internet no haya creado el fenómeno de la CVyDO, sí puede decirse que este ya no es el mismo al interactuar con el ciberespacio. Internet, al fin y al cabo, ha modificado las propias posibilidades y condiciones de la comunicación en sociedad, de modo que también la comunicación «desviada» se ve modificada. Las peculiaridades estructurales del ciberespacio, su carácter transnacional, su neutralidad o ausencia de censuras para el acceso de los usuarios, su universalidad y popularización y su permanente desarrollo, le definen como un nuevo ámbito de oportunidad, también delictiva, distinto al espacio físico,⁶ en el que los eventos delictivos pueden ver modificadas sus características, su significado social o sus concretas manifestaciones.

Internet ha modificado la CVyDO de forma más que significativa y lo ha hecho en dos etapas que se corresponderían con los dos grandes momentos «evolutivos» del ciberespacio. En la primera, con la aparición de la World Wide Web como

ámbito de intercambio de información en red, Internet se convierte en un extraordinario vehículo para la difusión sencilla y universal de mensajes, sean violentos o no, de modo que se incrementa la facilidad de acceder comunicativamente a miles de usuarios. Las palabras, escritas en libros que se vendían y pasaban a quienes los compraban o a quienes los hojeaban en librerías conocidas, comenzaron a volar a finales de los años ochenta sin barreras de distancia física, incrementando los potenciales receptores de cualesquiera mensajes, también los de mensajes violentos,⁷ tal como descubrieron pronto los grupos racistas.⁸ Los prospectos y folletos repartidos antes a mano, eran en Internet difundidos por webs y blogs fáciles de hacer y resultaban mucho más eficaces para transmitir ideas odiosas a millones de personas en todo el mundo.⁹ Pero hay una segunda etapa, la de la web 2.0,¹⁰ donde el desarrollo de las redes sociales convierte el ciberespacio en un ámbito de intercomunicación personal y social popular y global. En ella Internet se convierte en una herramienta comunicativa con más poder (cuanto menos a nivel potencial) de disuasión, donde la comunicación vertical del emisor del mensaje (violento o no) al receptor ya no es la única forma de comunicación ni la prioritaria, puesto que aparece otra comunicación en red, incluso una autocomunicación de masas,¹¹ en la que las personas contactan, comparten y se relacionan con un mensaje mucho más personalizado que antes, incrementándose la posibilidad de que este incida en el receptor-emisor.

Estas dos etapas también han incidido significativamente en la evolución de la alarma social relacionada con la difusión de «meras palabras» e influirán, y en realidad lo han hecho ya, con la aparición de reformas legislativas que traten de responder a tales demandas. La preocupación en la primera de las etapas, la de la web 1.0, se focalizaba en lo cuantitativo, esto es, en la potencialidad de Internet como vehículo difusor de la comunicación violenta o de odio a millones de personas.¹² Ante esa realidad, la criminalización

2. L. F. Ruiz Antón (1998, págs. 5-34).

3. Véase T. S. Vives Antón (1987); o en el Derecho penal anglosajón, J. Feinberg (1985).

4. Con más profundidad, sobre el origen y evolución de las mismas, A. Turienzo Fernández (2015).

5. F. Kubler (1998, pág. 335).

6. F. Miró Llinares (2012, págs. 145 y ss).

7. En este sentido, véase E. T. Pollock (2006). También en Y. Akdeniz (1997).

8. D. K. Citron; H. L. Norton (2011, pág. 1435).

9. Y. A. Timofeeva (2003, pág. 256).

10. D. C. Caldevilla Domínguez (2014, págs. 1 y ss).

11. M. I. Castells (2009).

12. E. T. Pollock (2006).

de conductas de difusión de mensajes de odio racial o discriminatorio fue extendiéndose más allá de los países que, por motivos históricos, ya disponían de tales leyes, y comenzó una discusión sobre los límites de la criminalización de las conductas que analizamos, que dura hasta nuestros días.

La preocupación, ahora, con la web 2.0, es otra, y tiene un sentido más cualitativo que cuantitativo: ya no se trata de que Internet facilite que la comunicación llegue a más personas, sino de cómo llega esa comunicación, particularmente de los efectos «nocivos» de la misma por el hecho de que donde se comunica es en Internet. Lo que preocupa en la actualidad es que Internet potencie sus efectos comunicativos. En particular esta preocupación se ha concretado en dos problemáticas distintas: por un lado, Internet como foro de radicalización violenta usado, particularmente aunque no solo, por grupos terroristas yihadistas para el reclutamiento de miembros o para la mera difusión de mensajes de odio o de terror;¹³ por otro, la aparición de todo un conjunto de conductas ofensivas y expresiones de comunicación violenta más allá del propio discurso del odio tradicional, particularmente en redes sociales como Facebook y Twitter;¹⁴ y que ha dado lugar a una preocupación social que llevó a la Fiscalía General del Estado a intervenir ante varios hechos y que seguramente incidió también en la reforma, de carácter extensivo, del denominado delito de incitación al odio en la reforma de la LO 1/2015 CP.¹⁵ La preocupación, pues, ya no es que se puedan difundir a miles de personas materiales violentos y de odio producidos por grupos específicos, sino que sean miles de personas las que comuniquen odio y violencia, en muy distintas formas, a través de las redes sociales, interaccionando entre sí y generando, en muchos casos, una indignación social considerable. Las redes sociales parecen poblarse de palabras ofensivas y violentas en un entorno que, quizá por el supuesto potencial anonimato al que se asocia la comunicación en él o por otros factores relacionados con la percepción de falta de lesividad de las conductas en él realizadas, parece potenciar tales formas de comunicación y/o, al menos, incrementar la visibilidad de las mismas. Esto ha conllevado una inmediata exageración sobre la dimensión de un fenómeno que, además, engloba formas de comunicación muy diferentes entre sí. Y ello ha derivado en una demanda de intervención jurídica que exige una profunda reflexión sobre los límites de la intervención

penal pero, también, quizá antes, un análisis básico acerca de las diferentes modalidades de comunicación que configuran lo que podríamos llamar «CVyDo en Internet».

El presente trabajo pretende aproximarse a la problemática de las palabras que dañan y ofenden, de aquellas que incitan a la violencia o facilitan la radicalización violenta de otros. El objetivo, sin embargo, no es aquí el de discutir sobre la criminalización de las mismas, debatiendo acerca de las razones que pueden esgrimirse a favor o en contra de su tipificación penal y de la concreta forma y los límites para llevarla a cabo. El objetivo, distinto y quizá anterior a este, es el de analizar el fenómeno mismo, identificarlo y diferenciar sus expresiones significativas para poder, posteriormente, tanto realizar valoraciones ético-normativas en aras de decidir acerca de su criminalización, como tomar otro tipo de decisiones preventivas a raíz de la incidencia real del fenómeno, para lo cual se ha llevado a cabo un estudio empírico basado en la observación del fenómeno en una red social.

2. Hacia una taxonomía de la comunicación violenta y el discurso del odio

2.1. Objetivos de la investigación y metodología para la construcción de una primera taxonomía

La presente investigación, como paso previo a otras futuras, pretende describir las diferentes formas de comunicación violenta existentes en Internet sobre las que gira la discusión social acerca de si deben ser criminalizadas o no, y conceptualizar las mismas en una taxonomía básica que permita identificar cada una de las categorías como distinta a las otras. El método inicial es la observación del fenómeno, en este caso, de la comunicación en Internet y, dentro de ella, de la comunicación violenta en Twitter. A partir de ello, se aplicarán criterios axiológicos, relacionados con la decisión criminalizadora de tales comportamientos, para conceptualizar y construir una taxonomía que clasifique las diferentes formas de comunicación violenta en Internet entre las cuales está el discurso del odio. Esto conlleva un

13. A. Stenersen (2008); M. Á. Cano Paños (2008); A. Silke (2005); R. Thompson (2011).

14. N. Djuric *et al.* (2015).

15. Véase S. Rodríguez, Ferrández (2014, págs. 165-232). También en G. Portilla Contreras (2015).

proceso analítico de tres pasos: 1) delimitación del objeto de la observación y la categorización de la CVyDO en Twitter; 2) observación de los mensajes a partir de una muestra comunicativa determinada, y 3) construcción de la taxonomía a partir de la conceptualización basada en lo observado y apoyada en criterios valorativos de interés.

Es evidente que taxonomías de la comunicación violenta podrían hacerse muchas. De hecho, ya hay clasificaciones de discurso del odio atendiendo al tipo de discriminación,¹⁶ como también hay categorizaciones de la comunicación violenta en Internet que atienden al tipo de mensaje que se comunica.¹⁷ La que se propone aquí es distinta por varias razones. En primer lugar, por el punto de partida. No es objeto de interés exclusivamente el discurso del odio, el *hate speech*, la violencia comunicativa consistente en la defensa o la incitación a la discriminación, ni siquiera las ofensas o las injurias perpetradas con una intención discriminatoria, sino cualquier acto comunicativo violento en Internet tal y como luego se precisará. En segundo lugar, también será distinta la taxonomía por el diferente objetivo pretendido. Puesto que el objetivo principal es comprender el fenómeno para, en futuros trabajos, tener criterios a la hora de valorar las decisiones sobre la criminalización y para evaluar decisiones de política preventiva en relación con las mismas, es obvio que la taxonomía debía tener un carácter valorativo-jurídico de fondo. Así, la clasificación parte de una valoración ético social sobre las conductas que suele tener reflejo (si bien de forma distinta según el sistema jurídico) en las normas. En concreto toma como referencia los criterios usualmente utilizados para tomar decisiones de criminalización en relación con estas conductas.

Por último, y pese a que lo que se pretende es realizar una taxonomía de la CVyDO en Internet por medio de la observación del fenómeno, el rigor exigido a una investigación de

estas características obligó a restringir la observación a una muestra de una concreta red social, en particular Twitter. Hay otras muchas redes sociales en las que se produce comunicación violenta, particularmente en Facebook y en foros específicos de páginas web en los que el discurso del odio y la violencia comunicativa pueden ser incluso de mayor entidad¹⁸ y en las que la expresión concreta de la comunicación violenta va a ser distinta por las particulares características de cada red social.¹⁹ Pero hay razones tanto para limitar la observación a la red social Twitter, como para pensar que la taxonomía de la comunicación violenta que derive de ella puede ser válida para la expresada en otras redes sociales. En cuanto a las razones para elegir como red social en la que se va a observar la comunicación violenta, Twitter, el nuevo «ágora social» donde se comenta todo aquello que produce un sobrecogimiento o interés social,²⁰ podrían esgrimirse tanto la propia repercusión pública de lo difundido en esta red social y que incrementa la alarma social de la comunicación en ella vertida,²¹ como la mayor simplicidad metodológica que conlleva la observación en esta red social frente a otras (Facebook, foros sociales) en los que la “censura” de comunicaciones violentas es más rápida y en las que resulta más difícil acceder a una muestra grande y homogénea de comunicación. El principal motivo, en todo caso, es que en Twitter, por su propia limitación de red de *microblogging* que publica mensajes de 140 caracteres, se expresan algo así como unidades básicas de comunicación, por lo que es más difícil (aunque no imposible) expresar varios sentidos comunicativos y más sencillo identificar un sentido único. En Twitter se publican mensajes cortos que, si bien pueden contener información compleja, difícilmente dispondrán de los matices de la comunicación en otros lugares donde no exista esa limitación.²² Siendo el objetivo de este trabajo definir una taxonomía básica de la comunicación violenta entiendo adecuado centrar el análisis en este tipo de mensajes más básicos.

16. M. J. Matsuda (1993). Otra clasificación, en L. Silva *et al.* (2016).

17. Véase W. Jacks; J. R. Adler (2015, págs. 64-89). En el mismo sentido McDevitt y sus colegas contrastan esta tipología en J. McDevitt, J. Levin, S. Bennett (2002, págs. 303-317). P. Sobkowicz y A. Sobkowicz (2010, págs. 633-643).

18. Entre otros, R. Cohen-Almagor (2009). También sobre el análisis de Facebook, en esta ocasión con el antisemitismo en A. Ben-David; A. M. Fernández (2016). Otras investigaciones sobre el discurso del odio en redes sociales como la de J. M. Maweu (2013). En relación con el discurso del odio en blogs: M. Chau; J. Xu (2007). Y sobre el discurso del odio en la Web, la investigación de C. Nobata *et al.* (2016).

19. Twitter ha desarrollado sus propias características comunicativas, algunas de las cuales dificultan la automatización de los análisis por medio de métodos de computación. Véase, A. Z. Khan *et al.* (2015).

20. A. Fernández-Planells, C. Feixa Pampols, M. Figueroas-Maz (2013).

21. La literatura científica ha descrito ampliamente la relación entre Twitter y la generación de alarma social, especialmente en situaciones de emergencia. En la investigación de Oh y sus colaboradores, se describe el papel de Twitter en la difusión de rumores, sus consecuencias y control en relación con el terremoto de Haití. Véase O. Oh *et al.* (2010, pág. 231). En el mismo sentido, otras investigaciones como O. Oh *et al.* (2013).

22. C. P. Sabater *et al.* (2015).

Ello no significa que la taxonomía no pueda extrapolarse a otros entornos donde el discurso no sea tan reducido como lo es en Twitter. Lo que no podrán extrapolarse serán las tasas de prevalencia. Pero una vez definamos la taxonomía, la misma podrá aplicarse a otras redes siempre comprendiendo que en ellas lo más usual será que una misma expresión comunicativa violenta (imaginemos, un post en Facebook) pueda ser ubicada en dos o tres, y no en solo una, de las categorías de la taxonomía, puesto que un incremento del texto comunicado dará lugar en ocasiones a un mayor número de mensajes en él contenidos.

2.2. Observación del fenómeno en Twitter (I). Objeto

2.2.1. Delimitación del objeto de investigación y criterios de inclusión de la categoría (más allá del discurso del odio)
Antes de proceder a una observación del fenómeno es imprescindible delimitar con claridad qué es aquello que va a ser observado. Como se ha dicho, nuestro interés no se centra exclusivamente en el discurso del odio, sino en cualquier forma de expresión o comunicación violenta, y es el momento de explicar qué es y por qué.

Aunque el *hate speech* ha sido definido de múltiples formas, y no siempre de modo coincidente,²³ el protocolo adicional al Convenio sobre la Ciberdelincuencia ha consolidado en Europa la consideración de discurso del odio como «todo material escrito, toda imagen o cualquier representación de ideas o teorías, que propugne, promueva o incite al odio, la discriminación o a 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 esta se utilice como pretexto para cualquiera de esos factores». Aunque esta definición podría entenderse superada por otras posteriores de la ECRI, lo relevante de la misma es que el elemento central del concepto lo constituye el hecho discriminatorio que es, a su vez, el que ensambla el discurso del odio dentro del más general concepto de los *hate crimes*.²⁴

Es cierto que el centro de la «preocupación mediática» relacionada con la comunicación violenta en Twitter es el discurso del odio,²⁵ en general, y algunas formas de odio en particular, como las relacionadas con el odio racial,²⁶ el odio religioso (y en particular la islamofobia)²⁷ o el odio por razón de género o de identidad sexual.²⁸ Sin embargo, la preocupación social por la comunicación violenta en Internet no se ceñía exclusivamente a la comunicación violenta por motivo de odio. Como se ha señalado previamente, lo que preocupa en el entorno web 2.0 son los efectos lesivos que determinadas conductas ofensivas podían producir, así como la capacidad disuasoria de algunos actos de habla en relación con la potencial realización por parte de otros de actos violentos. Y es evidente que expresiones violentas en Internet como la incitación terrorista a la violencia, las injurias o las calumnias contra personas concretas, entre otras, muestran que hay formas de comunicación violenta más allá del *hate speech*.

A los efectos de este trabajo, pues, pretendo huir del restrictivo concepto del *hate speech* y buscar un marco comprensivo más amplio que aborde cualquier comunicación violenta que no sea exclusivamente la discriminatoria. La comunicación violenta engloba cualquier forma de expresión que pueda conceptuarse como violenta, independientemente de que se lleve a cabo por motivos discriminatorios. La violencia constituye, pues, el elemento definitorio de este tipo de comunicación, integrando en la misma tanto la violencia física anunciada, incitada, deseada, justificada o valorada positivamente, como la violencia moral, aquella que resulta de la causación de un daño no físico o de una ofensa a intereses morales dignos de tutela de personas concretas o de una colectividad. El *hate speech* o discurso del odio, pues, tal como ha sido definido, podría considerarse comunicación violenta, en cuanto que la incitación a la violencia lo es en sentido de violencia física, y la incitación al odio o a la discriminación constituye una forma de violencia moral. Pero hay otras formas de comunicación violenta distintas al *hate speech*, concretamente: a) todas las formas de incitación (directa o indirecta), o amenaza específica de causación de

23. Sobre el Hate Speech véase, entre otros, en M. J. Matsuda. También en C. R. Massey (1992). Otro ejemplo en S. Walker (1994). Y más recientemente, en J. Waldron (2012).

24. B. Perry (2001).

25. Así lo demuestra la literatura científica que se ocupa mucho más del *hate speech* que de otras formas de comunicación violenta en redes sociales. Así, entre otras obras de referencia podría citarse C. A. Noriega *et al.* (2012); pero también en G. Titley (2014).

26. A. Harell (2010).

27. I. Awan (2016).

28. V. Reddy (2002).

violencia física que no sean por razón discriminatoria o de grupo; b) toda ofensa o daño al honor o a la dignidad de personas concretas, y c) todo comportamiento que pueda considerarse ofensivo o vejatorio para la sociedad aunque no vaya dirigido a una persona en concreto.

Sobre esta idea de «CVydO» versó, pues, la observación cuyo proceso se explica a continuación.

2.2.2. Recogida de la muestra y resultados de los cribados

En la red social Twitter se manifiestan de manera condensada y en gran cantidad, diferentes tipos de fenómenos de odio y violencia, es por ello que se seleccionó un suceso cuya difusión en este medio fuera muy amplia, con un gran impacto y repercusión, y por tanto sobre el que se tuviera la certeza de que podría registrarse una comunicación a gran escala con un gran volumen de eventos comunicativos, pero también, que tuviese la potencialidad de generar múltiples interacciones y expresiones violentas y de odio.²⁹ En este sentido, se recurrió a la información proporcionada por los usuarios de Twitter como reacción al atentado terrorista perpetrado en París en la sede del semanario satírico francés *Charlie Hebdo*, una revista cómica que, desde su refundación en 1992, había publicado controvertidos artículos y dibujos por los que había atraído las críticas y provocado la indignación de musulmanes, judíos y cristianos, y que había sido objeto de ataques de diversa índole. Uno de ellos, el que tuvo lugar con ocasión de la reproducción solidaria en febrero de 2006 de algunas de las caricaturas satíricas que, en septiembre de 2005, publicó el periódico danés *Jyllands-Posten* y que indujeron a una oleada de disturbios y reacciones en todo el mundo,³⁰ puede ser considerado como origen de los atentados de principios del año 2015.

Como es sabido, a las once de la mañana del día 7 de enero de 2015, los hermanos Said y Chérif Kouachi, de nacionali-

dad francesa, padres argelinos, y pertenecientes al grupo terrorista Al-Qaeda,³¹ armados con fusiles automáticos Kalashnikov y actuando en respuesta a la publicación por la revista *Charlie Hebdo* de unas viñetas sobre el profeta Mahoma, abrieron fuego en la sede de este semanario sobre periodistas y dibujantes, mataron a doce personas, entre ellos el editor, y causaron heridas graves a otras cuatro. Tras su huida, en la que abatieron a un policía, Ahmed Merabet, en unas imágenes de las que se hicieron eco los medios, comenzó una espiral de terror en la que participó también Amedy Coulibaly³² y que terminó días después con el abatimiento de todos ellos por las Fuerzas de Seguridad de Francia.

La violencia con la que se perpetraron los atentados, ampliamente comentados en los medios de comunicación,³³ en los que se publicaron impactantes imágenes del asesinato del policía francés o del asalto por parte de los grupos especiales de la policía al supermercado judío, así como la raíz islamista de estos hechos en los que, como en acciones precedentes, jóvenes franceses de procedencia extranjera causaban el terror entre la población de este país, provocaron una importante repercusión pública que se trasladó a las redes sociales. Casi de inmediato, se produjeron reacciones de solidaridad y apoyo a la revista, pero también en contra del Islam y los musulmanes, y surgió en Twitter el *hashtag* #CharlieHebdo, el cual se convirtió en *trending topic*.³⁴ Rápidamente esta fue una de las diez palabras más relevantes en la página principal de Twitter con más de seis millones y medio de menciones. El miércoles 7 de enero, el día del ataque, en torno a las 21.30, se registró el punto álgido de actividad con más de 6.300 tuits por minuto con esa etiqueta. Otros términos, como #JesuisCharlie, #StopIslam, #MuslimAreNotTerrorist, #IslamIsTheProblem o #IslamNon-Coupable registraron igualmente un importante número de tuits.³⁵ Con ellos, se expresaban condolencias a los familiares de las víctimas, solidaridad y apoyo a la revista y a la libertad de expresión, también odio y rabia por lo sucedido.

29. Diferentes investigaciones han recogido muestras de Twitter y otras redes sociales, con base en comunicaciones referidas a distintas acciones terroristas y otros eventos de carácter disruptivo, habiéndose convertido en fuentes de datos muy relevantes para el estudio del comportamiento humano. Véase, por ejemplo, entre otros F. Glibetto y Y. Lee (2015). También J. An *et al.* (2016), o en P. Burnap *et al.* (2014).

30. C. Guimelli *et al.* (2010).

31. Sobre la relación de los Kouachi con Al-Qaeda y su radicalización véase, entre otros, G. Woo (2015, pág. 33); también en Z. Wang y otros (2016); J. Khader (2015). Said, el mayor de los hermanos, había recibido entrenamiento militar en Yemen en 2011; por su parte, Chérif había cumplido una condena en 2005 por tratar de unirse al movimiento yihadista.

32. I. Wolska-Zogata (2015).

33. I. Wolska-Zogata (2015).

34. Sobre el procedimiento de determinación del *trending topic* ver entre otros K. Lee *et al.* (2011).

35. W. Magdy *et al.* (2015).

Así, con el fin de obtener una muestra que contuviese una importante cantidad de mensajes referidos a reacciones frente a los atentados, se siguieron dos criterios de selección. En primer lugar, se recurrió a la extracción de tres *hashtag* que en algún momento en los días en los que sucedieron los acontecimientos hubieran sido identificados como *trending topic* en España. El segundo criterio, trataba de equilibrar la muestra para su análisis, desde el punto de vista de las tendencias del contenido comunicativo.³⁶ En este sentido, dentro de los *hashtag trending topic*, se seleccionó una etiqueta con contenidos solidarios y de apoyo #JeSuisCharlie;³⁷ otra referida a la descripción del evento conforme iba sucediendo, #CharlieHebdo;³⁸ y, en tercer lugar, un *hashtag* que pudiera expresar actitudes negativas hacia los atacantes y su procedencia, #StoptIslam.³⁹ Con estos criterios iniciales y ciertos filtros que permitiesen acotar la muestra a mensajes construidos con el idioma español por personas, y un periodo temporal que abarcara entre el día 7 y el 12 de enero, se extrajeron en un archivo de datos en formato JSON por medio de la API de Twitter⁴⁰ liberada para tal efecto, un total de 255.674 mensajes en esta red social, incluyendo tanto los mensajes originales como los retuits,⁴¹ con 41 variables cada uno.

Sabiendo que la muestra recogida contendría un importante sesgo determinado por la misma selección del evento de estudio, pero con el objetivo fundamental de conceptualizar las modalidades de comunicación violenta en una red social, se partió del concepto de CVyDO para tratar de identificar cuántos de los mensajes incluidos en la muestra entraban dentro de tal categoría. Ello exigía, conforme a la metodología utilizada y al no confiar en algoritmos de análisis de contenido incapaces de integrar los matices de la categorización tan amplia como la realizada,⁴² la lectura de los tuits y la comparación de cada uno de ellos con el concepto de CVyDO.

Previamente, sin embargo, fue necesario llevar a cabo un esencial cribado del total de la muestra de mensajes. El mismo consistió en la eliminación de los retuits, es decir, las réplicas de los mensajes originales, que no aportan información adicional y por tanto relevante para el análisis de la comunicación. Para ello se utilizaron técnicas de filtrado consistentes en la identificación de las cadenas de caracteres iniciales que marcan a los mensajes como tales. De este modo, se obtuvieron 53.860 mensajes sobre los que se realizó un segundo cribado para identificar el discurso del odio contenido en los mismos.⁴³

Tabla 1. Muestra dividida en *hashtags* y sin retuits (n = 255.674)

Hashtags	Muestras totales	Muestras sin RT
#CharlieHebdo	194.206	41.444
#JeSuisCharlie	49.383	8.397
#StoptIslam	18.159	4.019

Dada la magnitud de la muestra se recurrió a un análisis interjueces y al uso, para garantizar la concordancia en sus valoraciones, del coeficiente Kappa que mide el índice de concordancia y, por tanto, la fiabilidad, interjueces. Se procedió a una selección de investigadores que deberían codificar el material objeto de estudio. Después de un entrenamiento con los jueces sobre una muestra de cien retuits en el que se discutían los matices de cada uno de los criterios de inclusión alternativa, se hizo una prueba piloto de 200 tuits para comprobar que había alta fiabilidad en la valoración de todos y cada uno de ellos. Se analizó el índice de concordancia de Kappa que, como es sabido, viene a establecer que cuando su resultado arroja un valor de k inferior a 0,20 la fuerza

36. Sobre el *hashtag* como vehículo de identificación social, F. Giblietto y Y. Lee (2015).

37. L. Merzeau (2015).

38. Ver el tratamiento de los *hashtag* en el caso de *Charlie Hebdo* en E. Herrera-Viedma *et al.* (2015, pág. 8).

39. Sobre la utilización de esta etiqueta G. Blanquart y D. M. Cook (2013); también en W. Magdy *et al.* (2015, pág. 12).

40. P. Barbera, P. Maintainer y Roauth (2015).

41. Para conocer el funcionamiento del retuit, véase entre otros, D. Boyd *et al.* (2010).

42. Existen distintas aplicaciones como *Qda Miner* junto a *WordStat (Provalis)* o *Alceste (Image)*, entre otras muchas que, por medio de algoritmos de clasificación, realizan agrupaciones o *clusters* sobre la base de la comparación de un corpus de texto. P. Burnap y M. L. Williams (2015). No obstante, en las mismas solo se analizan de momento la suma de palabras, y no el discurso propio cuyo sentido solo se puede obtener mediante un análisis de pragmática lingüística que aquí sí realizamos como paso inicial a la construcción de un algoritmo futuro que si lo tenga en consideración.

43. Las diferencias que se observan entre los *hashtag* totalizados en la tabla 1 (261.748 mensajes etiquetados si sumamos las tres categorías) y el totalizador de mensajes, se deben a que cada uno de los tuits analizados podían contener una, dos o las tres etiquetas, lo que, dado el gran volumen de datos, no afecta a los resultados finales obtenidos.

de concordancia es pobre, entre 0,21 y 0,40 es débil; de 0,41 a 0,60 moderada; de 0,61 a 0,80 buena, y entre 0,81 y 1,00 muy buena. El resultado para la prueba piloto de 200 tuits fue de 0,91, por lo que la concordancia interjueces se consideró más que satisfactoria.

Se crearon tres grupos de dos jueces que procedieron a la lectura de 17.953 tuits. Respecto a cada uno de ellos debían afirmar si el mensaje suponía la expresión clara de alguno de los siguientes contenidos:

1. La voluntad de realización directa, o a través de otros a los que se incite directamente, de actos de violencia física contra personas concretas o indeterminadas, así como expresiones de referencia en positivo (en forma de defensa, enaltecimiento, justificación, banalización, comprensión, alegría) a la causación de tal violencia.
2. El insulto o la ofensa grave dirigida a personas concretas y determinadas, así como la atribución a estas de la realización de hechos delictivos o ilícitos graves con conocimiento de la falsedad o con temerario desprecio hacia la verdad.
3. El desprecio o expresión de odio hacia grupos determinados, especialmente hacia aquellos que de algún modo han visto, o pueden ver, privados sus derechos y que sufren actividades intolerantes, y en particular aquellas expresiones que usen términos despectivos contra los mismos y que pidan o justifiquen la restricción de derechos contra tales grupos.
4. Aquellas expresiones especialmente desagradables y de muy mal gusto referidas a sucesos que causan grave dolor a algunas personas, en particular las que muestran odio a dichas personas o las que deshumanizan totalmente al que las realiza, incluyendo chistes y humor negro especialmente grave y en relación con eventos que, no siendo violentos (muerte natural o accidental), causan mucho dolor a víctimas indirectas y otras que afectan a sentimientos o creencias aceptadas por una colectividad.

En el caso de que el mensaje fuera una de estas formas de expresión quedaba definido como mensaje de CVyDO. En

el caso de que no fuera así por alguno de los dos jueces, quedaba descartado. De nuevo, con el fin de establecer el grado de acuerdo en la clasificación de los mensajes entre los investigadores, se llevó a cabo una prueba de contraste de concordancias con ajuste del efecto del azar en elementos cualitativos, utilizando para ello el índice de Kappa. Para la muestra seleccionada la prueba indicó que el grado de acuerdo fue satisfactorio, ya que como se muestra en la tabla 2 la fuerza de la concordancia interjueces podría clasificarse como "muy buena" según los parámetros descritos.

Tabla 2. Índice de Kappa

Grupo	P
Grupo 1	0,98
Grupo 2	0,86
Grupo 3	0,98

Concluidos por tanto los ajustes previos necesarios, aplicados los filtros, eliminados los elementos que no aportaban información relevante desde una perspectiva comunicativa y realizados los análisis interjueces, se obtuvo finalmente una submuestra de 2.274 tuits originales con contenido que podía clasificarse como discurso violento y del odio. A ellos debían sumarse, después, y para los análisis de prevalencia, los retuits. Es tras este proceso inicial y a partir de la observación de estos mensajes, desde donde se construyó la taxonomía de la CVyDO.

2.3. Taxonomía de la comunicación violenta y el discurso del odio en Twitter

Mucho antes de Austin y Searle, los sistemas penales habían aceptado que al hablar se realizan actos conforme a determinadas reglas,⁴⁴ y no es necesario discutir aquí la obviedad de que la mera expresión de gestos o el mostrar imágenes pueden ser conductas llenas de significado, también ilícito.⁴⁵ Los delitos de amenazas, el falso testimonio o las injurias y calumnias, forman parte de la tradición de muchos códigos penales, y la inducción como forma de participación delictiva y formas preparatorias del delito como la provocación o la conspiración para cometer delito son aceptadas como insti-

44. J. L. Austin (1975); J. Searle (1994).

45. Así, ya decía Ruiz Antón que «mostrar de una cierta manera un palo convencionalmente es tanto como hacer una amenaza y se puede conseguir el efecto de intimidar; hacer ciertos gestos con las manos convencionalmente constituye una acción despectiva, y a su través es posible producir un efecto vejatorio» (L. F. Ruiz Antón, 1998, pág. 34).

tuciones del sistema penal. La expansión del sistema penal ha ido incorporando más tipos penales que criminalizan la expresión de actos de habla, como los delitos de odio, los delitos contra los sentimientos religiosos o los de enaltecimiento de la violencia y de humillación a las víctimas. Además, hay muchos delitos que se pueden cometer con actos ilocucionarios, con significado por su capacidad perlocucionaria, como las estafas (y otros delitos económicos de naturaleza defraudatoria). El interés, aquí, sin embargo, no estriba en describir todos los actos de habla que pueden tener consecuencias penales y en analizar cuántos de ellos se dan en el ciberespacio, sino en crear una taxonomía de las formas de comunicación que, conforme a su significado social axiológico, pueden considerarse expresiones violentas, y de hacerlo teniendo en consideración los criterios usados tradicionalmente por los sistemas jurídico penales para criminalizar tales tipos de conductas.

Existen dos vías de argumentación usualmente utilizadas en los sistemas jurídicos penales para la criminalización de conductas que consisten, esencialmente, en actos comunicativos verbales: la relación performativa potencial del acto de habla con la perpetración, por parte de otros, de actos delictivos (a los efectos de lo que nos interesa, de conductas que pueden causar un daño físico a otros), por un lado y, por otro, la capacidad del propio acto comunicativo ilocucionario de afectar por sí mismo, e independientemente de la intervención de un tercero, a los intereses individuales o, incluso colectivos, dignos de tutela. Los argumentos de criminalización en uno y otro caso son claramente distintos, cuanto menos de inicio.

Existe un grupo de actos de habla cuya criminalización se razona y legitima por la capacidad de los mismos para crear decisiones criminales o para incidir en el riesgo de que los actos criminales terminen perpetrándose. La lógica de la

criminalización, aquí, es la de los actos preparatorios o la de los actos de participación en fase anterior a la tentativa, en cuanto que se arguye el riesgo que conlleva la conducta con respecto a la lesión del interés digno de tutela,⁴⁶ o el daño remoto que se produce al realizar aquello que puede dar lugar, posteriormente, a un daño efectivo.⁴⁷ Si en el supuesto de la inducción y con respecto a algunos actos preparatorios, y siempre que se tipifiquen de forma expresa, se acepta generalmente la legitimidad de tal forma de intervención, hay otros institutos como la apología del delito y otras figuras delictivas específicas como las distintas formas de incitación al odio y a la discriminación y las de incitación al genocidio,⁴⁸ el enaltecimiento del terrorismo o la humillación a las víctimas⁴⁹ y otros delitos relacionados con el terrorismo que entrarían dentro de la categoría de los *precursor crimes*,⁵⁰ en los que no hay consenso acerca de la legitimidad de tal criminalización o, más bien, sobre los elementos que deben incorporarse a tales formas de incitación indirecta en relación con los distintos intereses en juego para que la criminalización sea legítima. Al fin y al cabo, mientras que en algunas formas de incitación directa es más sencillo argüir la relación entre el acto ilocucionario y los potenciales efectos perlocucionarios de que se perpetre un crimen que justifican la criminalización, en los casos de incitación indirecta es más difícil encontrar el «*clear and present danger*»⁵¹ que, tradicionalmente se suponía que debía producir el acto comunicativo para poder ser criminalizado.⁵² Lo cual ha llevado a algunos autores, ayudados por la redacción de algunas leyes, a justificar la criminalización de conductas de incitación indirecta no tanto en la relación del acto de habla con el riesgo real que el mismo generaría de que se perpetren hechos delictivos de carácter violento, sino en la afectación a la paz pública que tales expresiones conllevan⁵³ o por el «clima de violencia, hostilidad, odio o discriminación» que las mismas generan.⁵⁴ El matiz, más que relevante a los efectos conceptuales de nuestra taxonomía

46. J. L. Fuentes Osorio (2006).

47. D. J. Baker (2007).

48. Véase C. Alastuey Dobón (2014); R. Alcácer Guirao (2013); M. L. Cuerda Arnau (2001); V. Gómez Martín (2012); J. M. Landa Gorostiza (2012); P. Lorenzo Copello (1996); S. Rodríguez (2014); P. T. Rodríguez Montañés (2012); G. Rollnert Liern (2008); G. M. Teruel Lozano (2015); A. Turienzo Fernández (2015).

49. Véase el trabajo de A. Alonso Rimo (2010).

50. C. Walker (2015, pág. 545 y ss).

51. G. Barnum (2006).

52. G. Rollnert Liern (2014, págs. 233 y ss).

53. R. Hefendehl (2001, pág. 150).

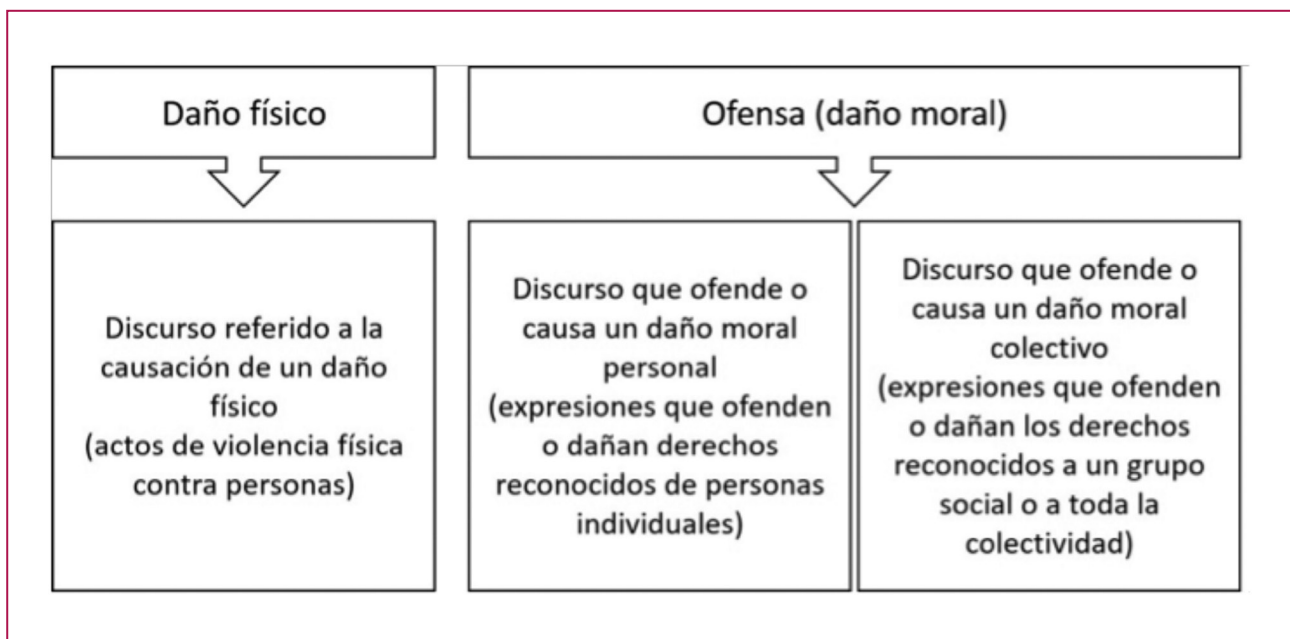
54. Así Turienzo Álvarez, quien sin embargo es capaz de relacionar ese clima con un daño remoto, al señalar que con tal criminalización se evita «una materialización de acciones delictivas como consecuencia de la cultivación de ese clima pernicioso propiciado mediante la difusión de los postulados negacionistas» (A. Turienzo Fernández, 2015).

es que, con respecto a ese tipo de «conductas con palabras», ya no se argumenta que las mismas pueden producir actos delictivos lesivos contra intereses como la vida o la salud, pero sí que al expresarlas pueden generar otro tipo de efectos perniciosos para las personas o para la sociedad, como por ejemplo un clima de hostilidad o intolerancia,⁵⁵ que, a su vez, tienen una supuesta capacidad potencial de generar efectivamente violencia física, aunque ya no sea necesario demostrar tal relación. La cuestión de interés, que no será resuelta sino solo planteada aquí, es que con ello se está dando un salto argumentativo significativo para evitar reconocer que la incriminación de tales conductas no tiene que ver con lo «lesivo» o con el daño, sino con lo «ofensivo» o con la ofensa, lo que, por sí mismo, no supone, a mi parecer, un obstáculo para la criminalización aunque sí una limitación del sentido y características de la misma.⁵⁶

Porque lo cierto es que hay otro «itinerario discursivo» para defender la legitimidad de la criminalización de actos comunicativos como los que nos interesan y que, en ocasiones, se mezcla en relación con algún delito concreto con el argumen-

to recientemente analizado. Se trata del daño u ofensa que, por sí mismo, conllevan (o causan) algunos actos de habla. Esta es la razón que se arguye, por ejemplo, en el caso de los delitos de amenazas: el acto de habla, por sí mismo, adquiere la fuerza ilocucionaria suficiente⁵⁷ como para considerarse que afecta a la libertad en el proceso de formación de la voluntad y, por tanto, lesiona un interés digno de tutela. Otro ejemplo paradigmático de esto son los denominados delitos contra el honor, particularmente las injurias y las calumnias, cuya criminalización se fundamenta en el daño (u ofensa) al honor personal que las mismas producen.⁵⁸ Y lo mismo sucede con la criminalización de delitos contra los sentimientos religiosos dentro de las cuales podrían entrar algunas expresiones hirientes que, en sí mismas, pudiera entenderse que afectan a la libertad religiosa de las personas.⁵⁹ En estos casos, independientemente de cual se diga que es el interés jurídico digno de tutela, en el discurso argumentativo de la legitimidad de la criminalización no se relaciona de modo alguno la conducta con ningún tipo de daño a intereses relacionados con la salud o la vida, sino con otros bienes personales que son violentados o dañados por el acto ilocucionario.

Gráfico 1. Bases de la taxonomía. Elaboración propia.



55. De ahí que, por ejemplo, Waldron (2012), sea capaz de hablar de daño y no de ofensa al referirse al *hate speech*.

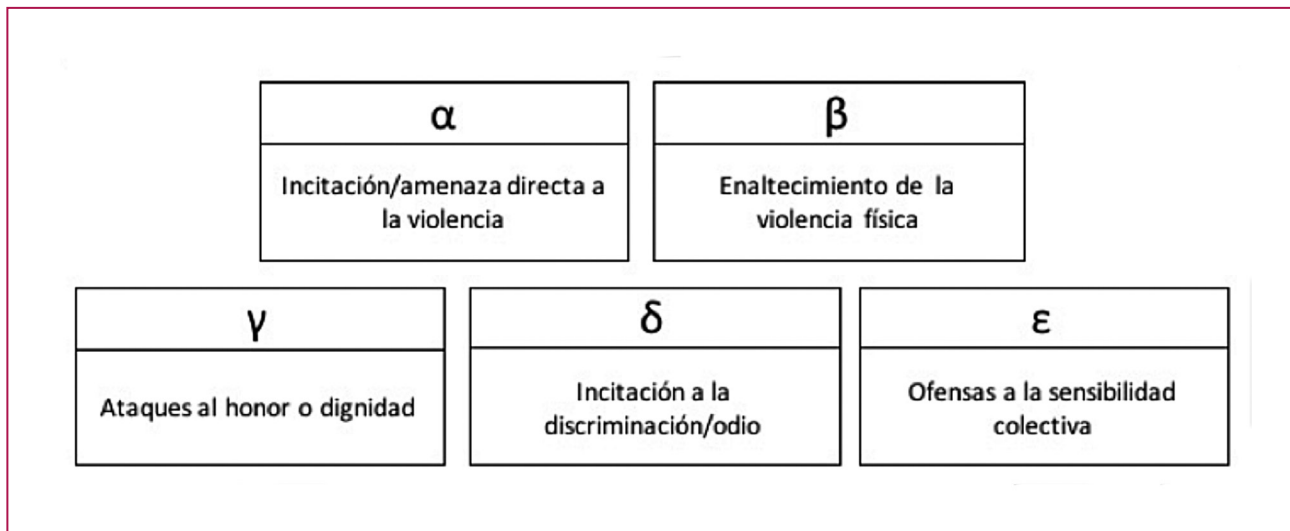
56. F. Miró Llinares (2015).

57. L. F. Ruiz Antón (1998).

58. En el sistema anglosajón, la argumentación sobre la criminalización de las expresiones injuriosas no se centra en el daño sino en la ofensa que conlleva tal forma de comunicación. Véase F. Miró Llinares (2015).

59. M. Alonso Álamo (2012).

Gráfico 2. Taxonomía de la CVyDO. Elaboración propia



La taxonomía de la CVyDO que propongo parte de la identificación de estos dos modos de argumentación de la criminalización de este tipo de actos comunicativos, y se inicia, pues, con una distinción entre aquellos actos de habla, imágenes y demás expresiones comunicativas que conllevan una capacidad de daño, en un sentido de daño físico o de daño a intereses personales como la vida o la salud, y aquellas otras expresiones cuya criminalización deriva del daño moral (no físico, o no a intereses personales como la vida o la salud) que las mismas causan. Esto podría simplificarse, sin modificar, a mi parecer, el sentido último de la clasificación, utilizando otros términos, concretamente distinguiendo entre actos de habla que causan o se refieren a la causación de un daño, y actos que no dañan (por lo menos en un sentido de daño físico) pero sí ofenden. Los primeros son los actos de habla referidos a la violencia física, los segundos son referidos más bien a la causación de violencia «moral».

Las expresiones escritas, imágenes, sonidos, vídeos, *gifs* animados y demás elementos que pueden configurar un tuit o cualquier otra forma de comunicación en cualquier red social en Internet, que nos interesan, son aquellas que, o bien son definidas como «violentas» por referirse a la causación de un daño físico a una persona, o bien son violentas por ofender y constituir violencia moral contra

alguien o contra un grupo determinado o indeterminado de personas. A partir de esta primera distinción es posible realizar algunas diferenciaciones más teniendo en cuenta por un lado el tipo de referencia (más o menos directa) del discurso al daño, por un lado, y, por otro, si la ofensa lo es por dirigirse a personas individuales o más bien por razones de grupo, todo lo cual nos permite una primera categorización en cinco modalidades básicas de la CVyDO en Internet.

Las dos primeras categorías englobarían las diferentes formas de discurso referido a la causación de un daño físico, incluyendo lo que tradicionalmente ha sido conceptualizado en Europa, particularmente a efectos de su regulación jurídica en los distintos códigos penales, como incitación a la violencia. Entrarían aquí, eso sí, solo la incitación a la realización de actos de violencia física contra alguien, quedando fuera de esta primera categoría la incitación a otro tipo de conductas, lícitas o no, como la discriminación o el odio racial.⁶⁰ Se podría argumentar que algunas formas de comunicación de odio que no se refieren a la perpetración de actos violentos causan un clima violento que puede generar, a su vez, actos de violencia, pero como se ha dicho el propósito de la taxonomía es el de diferenciar las categorías básicas comunicativas de violencia sin entrar en disquisiciones sobre la legitimidad de la intervención penal frente a cada una de ellas. La incitación a realizar actos violentos

60. A mi parecer, e independientemente de que la regulación jurídica diferencie o no entre ellas y de si ello resulta aceptable o legítimo, para una taxonomía de la CVyDO es esencial comenzar aceptando que existe una diferencia etiológica esencial entre el incitación a actos de violencia física por motivos discriminatorios, y la incitación al odio o a la propia discriminación.

o el enaltecimiento de la realización de actos violentos no son lo mismo que la incitación a la discriminación, a la humillación por razón de grupo o el enaltecimiento del racismo, independientemente de que ambas formas de comunicación se criminalicen, del mismo modo o de forma diferente, o no.⁶¹ Además, y esto es relevante, no es necesario que la incitación, directa o indirecta, a la violencia física sea por razón de grupo o por motivo discriminatorio. El motivo por el que se incita a actos violentos, a efectos de esta primera clasificación, es indiferente para que sea efectivamente un discurso referido a la realización de violencia sobre alguien. Lo que sí es cierto, también, y la doctrina penal ha puesto de manifiesto recientemente,⁶² es que el motivo por el que se hace algo puede cambiar el significado de lo realizado, por lo que sí podría distinguirse, y se diferenciará más adelante a la hora de realizar los análisis entre lo que es discurso del odio y lo que es comunicación violenta, entre aquellos discursos referidos a la causación de violencia física por motivos discriminatorios y los demás. Los primeros formarían parte, junto a algunas categorías de esta taxonomía, del Discurso del odio, los segundos más bien serían formas de Comunicación Violenta pero no de odio.

Dando un paso más, la diferenciación entre las dos categorías que engloban la incitación se puede realizar a partir de la consideración de si el discurso conlleva una petición o expresión de voluntad de la realización de tales actos de violencia física o, por el contrario, implica una valoración en positivo de la violencia física ejercida contra otros. En el primer caso estaríamos dentro de la categoría de discurso de incitación directa, o amenaza, de la realización de actos de violencia física, en el segundo, ante un discurso de enaltecimiento de la violencia.

En la primera macrocategoría de discurso violento y/o de odio situaríamos todas aquellas formas de comunicación violenta en las que se anuncia o se pide la realización directa de actos violentos que dañen a otros, sea por el motivo (relevante o irrelevante) que sea. Dentro de ella cabrían, pues, aquellos actos comunicativos que conllevan el anuncio de la realización de una acción violenta que causará un daño

físico (la amenaza directa), la comunicación consistente en hacer nacer en otro la voluntad de realizar un daño de tales características (la inducción a la violencia directa) o bien la incitación a un grupo indeterminado de personas a la realización de tal forma de violencia (la provocación de violencia física). Es cierto que existe una diferencia importante entre los actos de incitación (consumada en un hecho delictivo, inducción, o no) y los de amenaza, en cuanto que en los primeros se comunica a otros la voluntad de que realicen actos violentos sobre un tercero y en el segundo se expresa que se va a realizar directamente la acción violenta de forma más o menos inmediata. De hecho las amenazas también cabrían en la categoría de la violencia moral, puesto que las mismas suponen ya la causación de un daño al verse afectado el proceso de formación de la voluntad. Sin embargo la naturaleza del discurso de las amenazas y de la incitación directa a la violencia es similar a los efectos que aquí nos interesan: en ambos casos el discurso se centra en la realización directa de actos de violencia física, y es eso lo que diferencia la amenaza de otras formas de violencia moral, aunque la razón de criminalización sea finalmente muy parecida. Por otra parte para que estemos ante una de las formas de comunicación de esta primera categoría es necesario que se exprese con claridad tal voluntad de cometer o de que se cometan actos violentos, sin que sea necesario, en cambio, que se especifique el destinatario de los mismos más que en su caracterización general.

La segunda de las macrocategorías de la taxonomía de la CVyDO engloba todas aquellas formas de discurso en las que la incitación a la violencia es indirecta, al realizarse una valoración en positivo de la perpetración de actos violentos físicos perpetrados en el pasado o a perpetrar en el futuro sobre una persona determinada o indeterminada. Aquí entrarían todas aquellas formas de expresión en las que se justifica, se defiende o se enaltece la realización de acciones violentas, aunque no se propone directamente su causación. Frente a las formas, pues, en las que se provoca la realización de actos violentos, estas comunicaciones ensalzan los actos violentos sin expresar una voluntad clara de que alguien los realice en el futuro. Dentro de ella, además, se incluirían tanto las

61. Dice Alcácer Guirao que «el odio es un mero sentimiento, sin vinculación directa o indirecta –a diferencia de la violencia o la discriminación– con la realización de conductas lesivas contra intereses de las personas; la incitación al odio no es más que el intento de generar opiniones (de rechazo, de hostilidad) a través de la opinión» (R. Alcácer Guirao, 2012, pág. 17). Añadiría yo que tampoco es lo mismo la violencia que la discriminación, aunque sea cierto que ambas resultan lesivas contra intereses de las personas. A los efectos de la taxonomía, y sin entrar en consideraciones jurídicas, entiendo que es necesario diferenciar no solo entre la incitación a la violencia física y la incitación al odio, sino también entre la primera y la incitación a la discriminación.

62. J. Milton Peralta (2012); J. A. D. Díaz López (2013).

expresiones de valoración positiva de la delincuencia, donde entrarían por ejemplo todas las formas de expresión que conllevan un enaltecimiento del terrorismo, como aquellas otras en las que hay una defensa, justificación o enaltecimiento de la respuesta violenta, esto es, de la violencia que se desarrolla o se puede desarrollar dentro de un posible marco legal, pasado o presente, como puede ser la expresión de alegría si se mata a un terrorista en una acción policial.

Las otras tres categorías de la taxonomía engloban actos comunicativos que, por sí mismos, causan una violencia moral, esto es, y frente a la física, la que no conlleva la producción de un daño físico (ni tiene la posibilidad potencial de hacerlo) pero puede afectar a valores o bienes relevantes. Tal violencia moral puede ejercerse bien contra una persona en concreto, ofendiendo a su honor o afectando a su dignidad individual, o bien contra una colectividad, ofendiendo a un grupo determinado o indeterminado y afectando a valores o intereses de la colectividad.

La primera de las categorías de comunicación violenta que causa un daño moral, y tercera de las categorías de la taxonomía, englobaría todas las expresiones injuriosas y calumniosas que afectan al honor de alguien así como otros atentados contra la dignidad personal individual.⁶³ Dentro de ellas, además, podría diferenciarse una categoría específica para la violencia moral ejercida contra víctimas de un delito, lo cual conlleva un plus de ofensividad que puede reconocerse en algunos ordenamientos jurídicos.

La cuarta y la quinta de las categorías de la CVyDO engloban todas las expresiones que, sin constituir discurso referido a la realización de actos violentos, sí ofenden pero no a una persona individual, al referirse directamente a ella o atentar contra su honor o dignidad, sino a una colectividad de personas, al suponer la expresión un atentado contra valores comunitarios. Aunque con su expresión siempre habrá personas individuales que se sentirán ofendidas, la razón de la ofensa no la constituye el ataque que el comentario supone a ellas en particular, sino el que se realiza sobre valores comunitarios que, en algunos casos, pueden incluso tener un refrendo jurídico.⁶⁴ Puede diferenciarse,

en todo caso, entre aquellas expresiones que afectan a la mores colectiva, a la sensibilidad de una mayoría o de un grupo particular, y aquel otro discurso que afecta a un valor colectivamente asumido como esencial como es la igualdad entre las personas y, derivado de él, la tolerancia hacia el que no piensa, actúa o siente como otro y que puede llegar a afectar incluso a la dignidad de la persona.

En el primer caso, y por tanto ya en la cuarta macrocategoría de la taxonomía de la CVyDO, estamos propiamente ante el discurso del odio, en el que la violencia moral la constituye el atentado al valor colectivo de la igualdad entre las personas. Es cierto que tal violencia moral puede constituir, en determinados casos, un atentado a la dignidad personal cuando alguien es tratado de forma degradante por razón discriminatoria, pero en este caso estaríamos ante la categoría número tres, al igual que cuando, por razón discriminatoria se injuria o calumnia a alguien o cuando, ya dentro de las categorías uno y dos, se incita a la realización de actos violentos, directa o indirectamente, por razón discriminatoria. En la categoría número cuatro se incluyen aquellas expresiones o comunicaciones que resultan violentas por ofender a una colectividad al incitar a la discriminación o al odio o al humillar a un grupo, por medio de expresiones discriminatorias por razón de «raza», color, linaje, origen nacional o étnico, edad, discapacidad, lengua, religión o pensamiento, sexo, género, identidad de género, orientación sexual y de otra característica personal o de estado.⁶⁵ Dentro de esta macrocategoría se incluyen, pues, tanto los mensajes en los que hay una promoción, provocación o incitación, de forma directa o indirecta, a la discriminación y denigración de un grupo de personas o de una persona por razón de grupo, como, por otro lado, la promoción o incitación al odio y la estigmatización, humillación, el insulto o la utilización de estereotipos negativos contra un determinado grupo de personas o contra alguien por razón de grupo. También entrarían dentro de esta categoría aquellos actos comunicativos en los que se desea el mal a alguien por razón de pertenencia a un determinado grupo o en las que se niega la realidad de una situación lesiva o injusta que, indiscutiblemente, está sufriendo o que ha sufrido un determinado grupo.

63. Conviene aclarar de nuevo que la taxonomía no incluye categorías cuyo ámbito se corresponda exactamente con el de instituciones penales o con tipos específicos del Código Penal que plasmen el criterio de criminalización de que se trate.

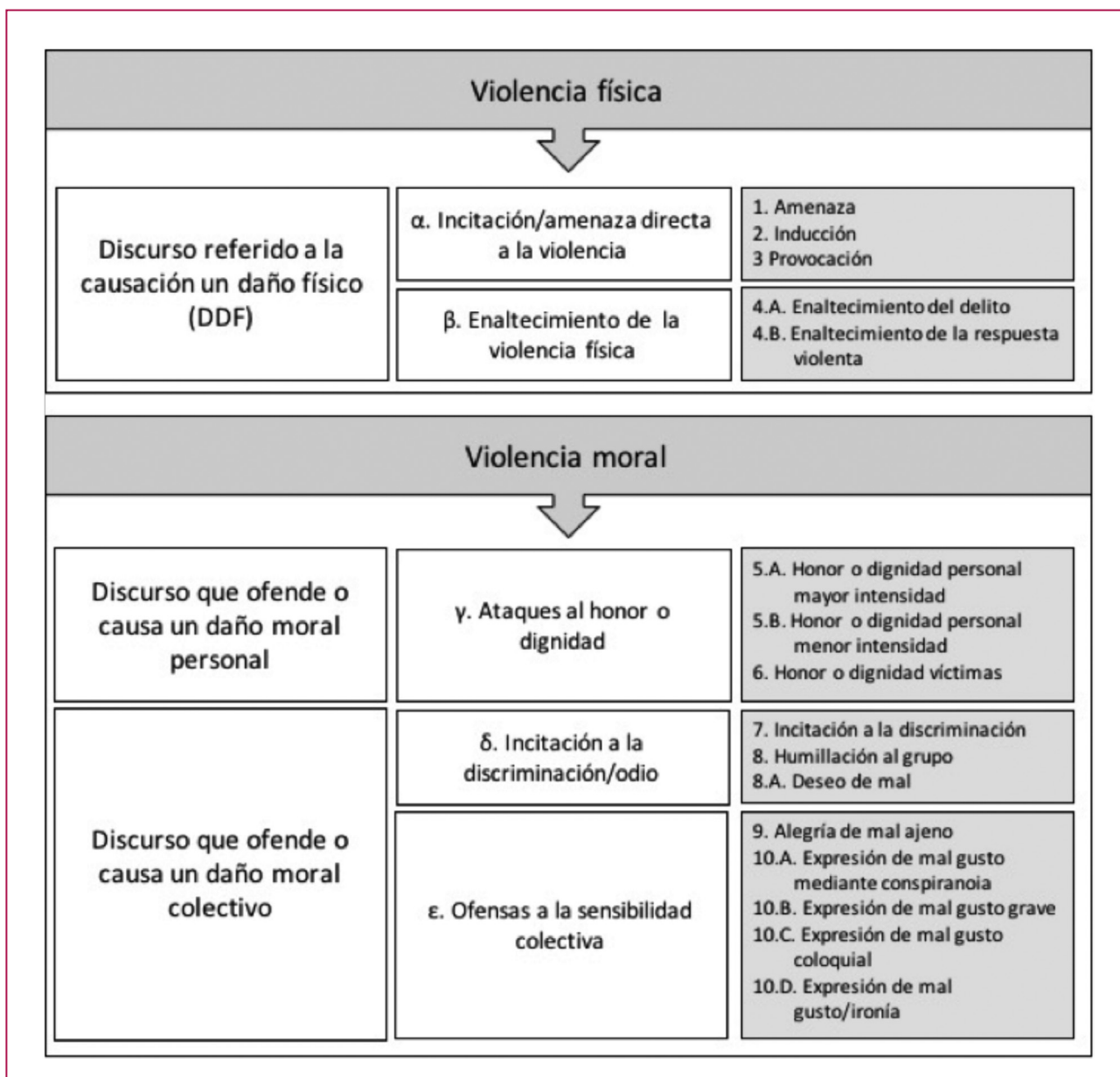
64. S. Rodríguez Ferrández (*op. cit.*, págs. 12 y ss.)

65. Conforme a las razones de discriminación enunciadas por la European Commission against Racism and Intolerance (ECRI) Council of Europe-2016 en su Ecri General Policy Recommendation n.º 15 On Combating Hate Speech Adopted On 8 December 2015.

Por último, y dentro de la violencia moral que afecta a la colectividad, tendríamos una macrocategoría que contendría otras expresiones comunicativas que resultan violentas por ofender a una colectividad, por herir la sensibilidad social, de una gran mayoría, utilizando expresiones hirientes y desagradables, carácter sobre el que existe un amplio consenso. Esta categoría engloba las manifestaciones de defensa de la violencia pasada y las expresiones de mal gusto, que afectan gravemente a la sensibilidad colectiva,

por la temática o por ser especialmente degradantes, o expresiones coloquiales de mal gusto socialmente reconocidas como tales, así como expresiones de mal gusto en las que se utiliza la ironía de un modo tal que se hiere la sensibilidad de otros. Sin entrar en consideraciones sobre si tales conductas pueden ser criminalizadas o no, lo cierto es que el legislador tiene en cuenta algunos sentimientos colectivos a la hora de criminalizar determinadas conductas, aunque usualmente exige una infracción de derechos para hacerlo.⁶⁶

Gráfico3. Taxonomía ampliada de la CVDO. Elaboración propia



66. F. Miró Llinares (2015, págs. 18 y ss.).

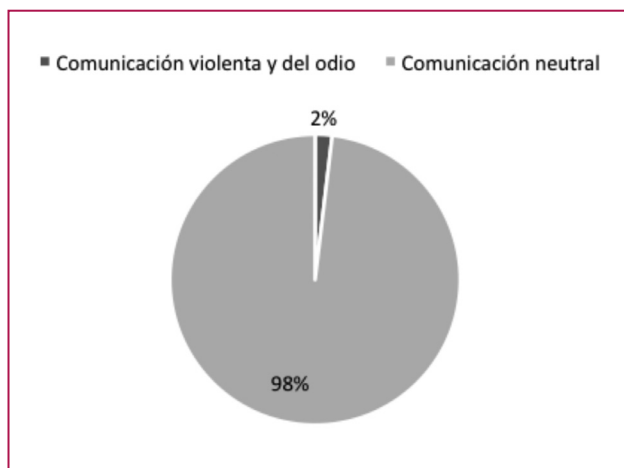
2.4. Observación del fenómeno en Twitter (II). Resultados de prevalencia de categorías

Establecidos, como se ha indicado en apartados precedentes, los criterios para la clasificación de los mensajes con contenido violento y de odio, corresponde ahora describir la muestra sobre la base de dicha categorización, analizando las proporciones de las diferentes variables objeto de estudio, en relación con las categorías a las que han sido asignados los mensajes, las etiquetas que contenían y los usuarios que los han suscrito. De este modo, tal como se observa en la Tabla 3, si consideramos el conjunto de la muestra incluidos los retuits un total de 4.838 mensajes contienen CVyDO, es decir, únicamente el 1,9% de todas las comunicaciones.

Tabla 3. Prevalencia de CVyDO (CVDO) en muestra total (n = 255.674)

	f (tuits)	%
Comunicación violenta y discurso del odio	4.838	1,9
Comunicación neutral	250.836	98,1

Gráfico 4. CVDO frente a comunicación neutra.



Al analizar la distribución de los mensajes en los cinco grupos principales de la taxonomía, los resultados muestran que la categoría con mayor representación es la de ataques contra la sensibilidad colectiva, concretamente un 67,8% de las comunicaciones mostraban características que permitían clasificarlas en este grupo. La segunda agrupación de mensajes es la de incitación a la discriminación, un 26%, mostrando el resto de grupos una baja representación con porcentajes del 0,1% en el de incitación directa a la violencia,

3,7% en el de la incitación indirecta y 2,4% las afectaciones al honor o la dignidad.

Tabla 4. Prevalencia de las categorías principales

Grupo	f	%
α Incitación violencia directa	5	0,1
β Enaltecimiento de la violencia física	178	3,7
γ Ataques al honor/dignidad	115	2,4
δ Incitación a la discriminación/odio	1.258	26,0
ε Ofensas a la sensibilidad colectiva	3.282	67,8

En relación con las distintas categorías secundarias establecidas para el estudio, se observa (Tabla 6) que la mayoría de los tuits son expresiones de mal gusto con un 44,6%, humillación a un grupo con un 19,2% y expresiones coloquiales de mal gusto con un 18,2%. Por su parte, categorías como amenazas, inducción o provocación, todas ellas pertenecientes al grupo de incitación directa a la violencia, y atentados contra el honor y la dignidad de las víctimas, apenas registran mensajes. En concreto tan solo había cinco mensajes que pudieran constituir una incitación directa a la violencia y 178 que podrían considerarse una incitación indirecta, aunque la gran mayoría consistían en mensajes de enaltecimiento de la respuesta violenta contra los terroristas.

Tabla 5. Prevalencia de las categorías de CVDO.

Discurso referido a la causación de	Grupo	Categoría	f	%
Violencia física	α. Incitación violencia directa	Amenaza	0	0,0
		Inducción	0	0,0
		Proposición	5	0,1
	β. Enaltecimiento violencia indirecta	Enaltecimiento del delito	58	1,2
		Enaltecimiento de la respuesta violenta	120	2,5
	γ. Ataques al honor o dignidad	Honor /dignidad personal mayor intensidad	38	0,8
		Honor/dignidad personal	77	1,6
		Menor intensidad	0	0,0
		Honor/dignidad víctimas	0	0,0
	Violencia moral	δ. Incitación a la discriminación/odio	Incitación a la discriminación	277
Humillación grupo			927	19,2
ε. Ofensas a la sensibilidad colectiva		Deseo de mal	54	1,1
		Expresión mal gusto mediante <u>conspiranoia</u>	5	0,1
		Expresión mal gusto grave	2.159	44,6
		Expresión mal gusto coloquial	882	18,2
		Expresión mal gusto mediante ironía	43	0,9
Alegría de mal ajeno	193	4,0		

Una vez distribuidos los mensajes en las categorías de la taxonomía se decidió realizar otra clasificación entre mensajes de discurso del odio, por un lado, y mensajes que constituyeran comunicación violenta sin motivo discriminatorio, por otro. Se procedió a la lectura de todos los tuits de las tres primeras categorías α , β y γ que podían recoger mensajes clasificados como comunicación violenta o como discurso del odio indistintamente, tal como puede apreciarse en la tabla de clasificación 7, analizando respecto a cada tuit si la comunicación violenta tenía un motivo o sentido discriminatorio o no. En cambio se obvió del análisis las categorías δ , que siempre sería discurso del odio, y ϵ , que por su propia definición únicamente contendrían mensajes de comunicación violenta.

Tabla 6. Categorías principales CVDO y definición como CV, DO, o CV/DO

Grupo	Comunicación violenta	Discurso del odio
α Incitación violencia directa	+	+
β Enaltecimiento de la violencia física	+	+
γ Ataques al honor/dignidad	+	+
δ Incitación a la discriminación/odio	-	+
ϵ Ofensas a la sensibilidad colectiva	+	-

Tabla 7. Prevalencia de las categorías principales discurso del odio

Grupo	Discurso del odio	
	f	%
α Incitación violencia directa	5	0,4
β Enaltecimiento de la violencia física	98	7,2
γ Ataques al honor/dignidad	0	0
δ Incitación a la discriminación/odio	1.258	92,4

Así, con el objetivo de establecer las proporciones de tuits relacionadas, por una parte con el odio y por otra, con la comunicación violenta, se realizaron análisis de clasificación tomando en consideración en qué supuestos la comunicación violenta tenía un motivo discriminatorio (DO) y en cuáles no (CV). Resultó que el discurso del odio se agrupa en su mayoría en la categoría de incitación a la discriminación, concretamente un 92,4% de los mensajes eran compatibles

con esta categoría, mientras que la comunicación violenta se concentraba en los mensajes que afectaban a la sensibilidad colectiva, un 94,4% de ellos (Tablas 8 y 9).

Tabla 8. Prevalencia de las categorías principales comunicación violenta

Grupo	Comunicación violenta	
	f	%
α Incitación violencia directa	0	0
β Enaltecimiento de la violencia física	80	2,3
γ Ataques al honor/dignidad	115	3,3
ϵ Ofensas a la sensibilidad colectiva	3.282	94,4

Tabla 9. Prevalencia de las categorías principales desagregadas

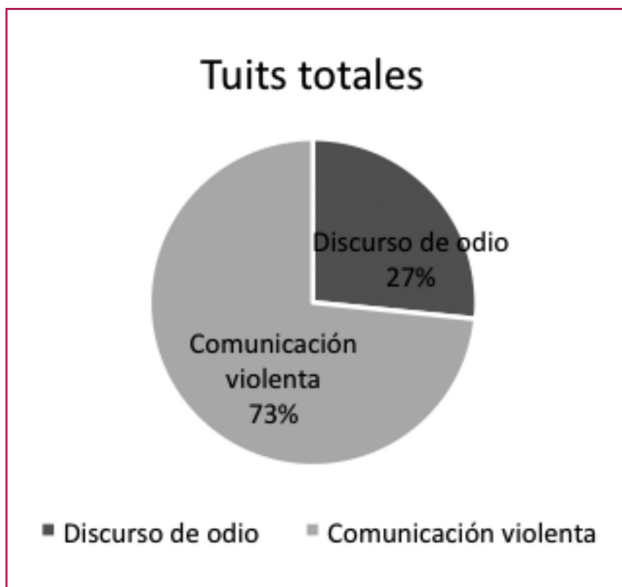
	Comunicación violenta		Discurso del odio	
	f	%	f	%
Amenazas	0	0	0	0
Inducción directa	0	0	0	0
Proposición/provocación directa	0	0	5	0,4
Apología, defensa o enaltecimiento de la violencia	80	2,3	98	7,2
Honor/dignidad personal	133	3,2	0	0
Honor/dignidad víctimas	2	0,1	0	0
Incitación a la discriminación	0	0	277	20,4
Humillación o menosprecio a grupo	0	0	927	68,1
Deseo de mal por razón de grupo	0	0	54	4
Expresión de mal gusto	3.282	94,3	0	0

Tras establecer las diferencias entre los distintos grupos principales, al estudiar las categorías secundarias en las que se desagregan la CVyDO, observamos que efectivamente en la primera casi un 94,3% son expresiones de mal gusto, seguidas de mensajes que encierran ataques contra el honor y dignidad personal, de la apología, defensa o enaltecimiento de la violencia, la defensa de la violencia pasada y deseo de mal de alguien, con una pequeñísima representación, el honor y dignidad a víctimas. Por su parte, el discurso del odio se muestra más repartido, siendo la categoría humillación

y menosprecio a grupo con un 68,1% la más representada, seguida de la incitación directa a la discriminación, la apología o defensa de la violencia y las amenazas, inducción y proposición directa (tabla 10).

Después procedimos a la suma de todos los tuits que constituirían comunicación violenta y de todos los que constituirían discurso del odio, para conocer la prevalencia de cada categoría, observando que la primera, la comunicación violenta, representa algo más del 70% del total, frente a cerca del 30% de mensajes del segundo (gráfico 3).

Gráfico 5. Proporción de DO frente a CV.



Sin embargo, la distribución en esos mismos dos grupos es diferente cuando no se contabilizan los retuits. En ese caso, la cantidad de CVyDO muestra un patrón de distribución casi uniforme, siendo la primera de un 51,1% y el segundo de un 48,9%. Es decir, se clasificaron prácticamente los mismos mensajes originales de comunicación violenta que de discurso del odio, si bien, los primeros están más representados en la muestra cuando tomamos en consideración los retuits.

Con el fin de establecer si existen diferencias entre las proporciones de CVyDO de tuits y retuits, se realizó la prueba de chi-cuadrado, resultando que efectivamente existían diferencias estadísticamente significativas ($\chi^2 = 822,17, p = 0,000$). Ello indica que son las repeticiones de los mensajes, los retuits, los que aumentan la cantidad de comunicación violenta frente al discurso del odio, puesto que estos últimos prácticamente no se repiten.

Gráficos 6 y 7. Distribución de CVDO tuits originales y retuits

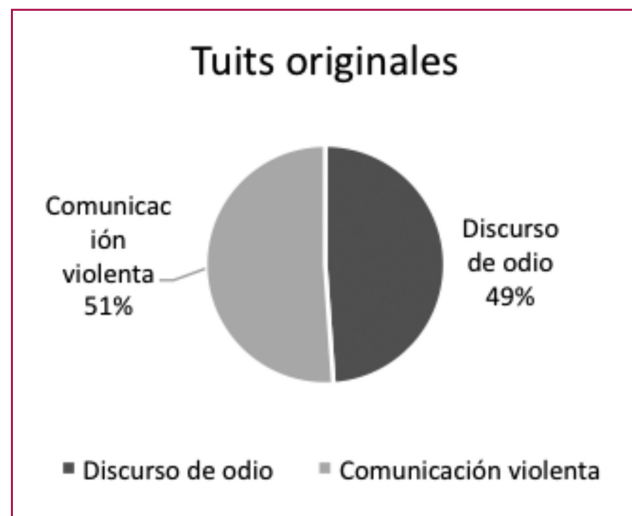


Tabla 10. Tuits originales y retuits CVDO

	f tuit	%	f retuit	%	f total	%
Comunicación violenta	1.162	51,1	2.264	88,3	3.426	70,8
Discurso del odio	1.112	48,9	300	11,7	1.412	29,2
Total	2.274		2.564		4.838	

Con el fin de establecer el número de tuits con contenido violento o de odio por usuario, se establecieron dos grupos, el primero con aquellos usuarios que habían publicado uno o dos tuits y el segundo, con los que publicaron más de dos

mensajes, obteniendo como resultado que tanto los usuarios que habían publicado uno o dos mensajes, como los que habían superado los tres, en las categorías comunicación violenta o discurso del odio, mostraban una proporción de 98,9% frente al 1,1% en el primero y 98,7% frente al 1,3% en el segundo. Sin embargo, ambos grupos se distribuyen casi proporcionalmente en el caso de la publicación de mensajes con contenido violento y de odio. De esta forma, son la mayoría de los usuarios, $f = 1.535$ usuarios, los que hacen la mayor parte de mensajes de CVyDO, $f = 1.685$, mientras que solo son $f = 112$ usuarios los que intervienen repetidas veces produciendo 588 tuits.

Tabla 15. Discurso del odio y comunicación violenta por usuario

	Comunicación violenta		Discurso del odio		Ambos	
	f	%	f	%	f	%
1-2 tuits	705	98,9	752	98,7	78	45,3
>2 tuits	8	1,1	10	1,3	94	54,7

Conclusiones y perspectivas futuras

No disponemos de datos científicos que nos permitan afirmar que Internet haya aumentado la CVyDO, pero no es difícil encontrar razones para pensar que eso debe haber sido así. Bastaría incluso con una: la comunicación, también la comunicación violenta, tiene una relación directa con la interacción social, con el contacto entre personas con necesidades comunicativas, y lo que define a la web 2.0 es precisamente que ha supuesto la aparición de un nuevo espacio de intercomunicación social, esto es, un nuevo lugar en el que las personas se comunican entre sí y en el que, también, se expresa odio y violencia. Si a ese incremento que podemos dar por seguro sumamos, además, la constatación de que el ciberespacio está siendo usado por grupos violentos para difundir mensajes de odio y de violencia, puede comprenderse la preocupación social por la capacidad de Internet para dispersar, generar y potenciar los efectos de mensajes radicales y violentos. Sin embargo, no sabemos aún ni cuánta comunicación violenta hay, ni cuál es esta, ni cómo identificarla, ni qué efectos produce, por lo que la prioridad debe ser la de comprender, cuantificar y analizar el fenómeno.

En este trabajo se ha presentado una taxonomía de la CVyDO en Internet realizada a partir de la observación del

fenómeno en Twitter con la intención de crear un marco comprensivo de estas formas de comunicación basado en los criterios utilizados tradicionalmente por los legisladores para la criminalización de actos comunicativos de habla o similares. La observación se centró en la comunicación en Twitter posterior a los atentados de *Charlie Hebdo* en París en el año 2015, evento desencadenante de una importante reacción pública en las redes sociales.

El estudio muestra que en Internet hay muy diferentes formas de comunicación violenta, pudiendo diferenciarse, a grandes rasgos, y dentro de los objetivos pretendidos por este trabajo, dos formas de categorización. La primera, que atendería esencialmente al motivo de la violencia comunicativa, distinguiría entre el discurso del odio propiamente dicho por un lado, que englobaría todas aquellas expresiones en las que la incitación a la violencia física o la violencia moral se perpetra por razón discriminatoria, y, por otro, la comunicación violenta que abarcaría todas las demás expresiones violentas en las que no hay un motivo discriminatorio. Otra forma de categorización posible clasifica las expresiones violentas tomando en consideración las razones por los que actos de habla suelen ser criminalizados. La taxonomía de la comunicación violenta propuesta parte de aquí. A partir de la constatación de que la razón de criminalización de este tipo de actos comunicativos estriba o bien en la potencial relación entre los actos comunicativos y la causación de violencia física, o bien en la propia capacidad de algunas palabras de ofender (o dañar) intereses personales o colectivos dignos de tutela, se puede diferenciar entre cinco macrocategorías de comunicación violenta: las expresiones de incitación directa a la violencia (o a la causación de daño físico), las de enaltecimiento o justificación de la perpetración de actos de violencia física, los atentados al honor o dignidad de personas particulares, los actos de incitación a la discriminación, al odio o de humillación por razón de grupo y las expresiones que ofenden a la sensibilidad colectiva. Esta taxonomía se completa con otras categorizaciones internas para, finalmente, ofrecer un marco comprensivo de cualquier forma de comunicación violenta en el ciberespacio

Puesto que el objetivo no era medir la prevalencia de la comunicación violenta en Twitter sino observarla y ofrecer una categorización de la misma, se seleccionó un evento en particular en el que fuera a haber expresiones de este tipo. Esto impide realizar afirmaciones de prevalencia de la comunicación violenta en Internet. Al seleccionarse, además, un evento relacionado con el terrorismo yihadista,

y en particular al ser uno de los hashtags que fue *trending topic* el de #stopislam, resulta razonable pensar que debe haber un sesgo de odio contra el islam que no permite asegurar cual es la prevalencia entre el odio en sentido estricto y las expresiones de “mera” comunicación violenta. Sin embargo sí se pueden realizar afirmaciones, y de ellas derivar reflexiones, de mucho interés derivadas del estudio referidas a la prevalencia de las categorías dentro de un evento comunicativo característico como el estudiado. La primera que Twitter comunica muchas más cosas que violencia y odio, incluso en relación con eventos como los seleccionados para el análisis que a priori pueden generar mucha violencia verbal. La solidaridad, el estupor, la propia descripción de la noticia, el comentario y las reflexiones sosegadas o, cuanto menos, no violentas, son la gran mayoría y representan más de un 98% de la comunicación en relación con un evento de estas características. La segunda, que incluso pese al sesgo comentado, en un evento en Twitter de las características del de *Charlie Hebdo* hay más expresiones de comunicación violenta que discurso del odio. Además, lo que predominan son expresiones de mal gusto que atentan a la sensibilidad de las personas y que ofenden a una colectividad, y son muy pocas las expresiones que constituyen una auténtica incitación a la violencia y meramente testimoniales las que, dentro de ellas, conllevan una incitación directa a la causación de daños físicos. Es significativo, no obstante, que de estas pocas expresiones que constituyen incitación a la violencia todas son por razón de odio discriminatorio, en este caso contra la religión islámica. El odio, podría decirse, es más dañino que otros motivos de discurso, en cuanto que con él se llega a la incitación directa a la violencia física. También hay, en tercer lugar, un significativo número de mensajes de enaltecimiento a la violencia, un significativo número de ellas por razón discriminatoria, aunque destacan en particular las expresiones que defienden la violencia policial contra los que perpetraron los atentados terroristas.

El evento, además, generó ataques de violencia moral directos y contra personas concretas, ofensas a su honor o ataques a su dignidad personal, particularmente a personajes públicos o a usuarios de la red social y en respuesta a sus propias opiniones. Aunque la prevalencia es pequeña en comparación con las otras categorías, eventos como el de *Charlie Hebdo* conllevan inmediatamente ataques al honor que podrían constituir delito y que se relacionan, además, con discusiones políticas o ideológicas que son propias del ámbito español y que solo indirectamente tienen que ver con el evento.

Eventos como los de *Charlie Hebdo* generan odio en Twitter. Este se manifiesta de muy diferentes modos, aunque predominan los mensajes de humillación o de estigmatización por razón de religión (en este caso), y en particular la utilización de estereotipos negativos contra los musulmanes. También hay, sin embargo, numerosos mensajes de incitación a la discriminación, en las que de forma explícita o implícita se demanda a los poderes públicos o a la sociedad la realización de acciones que limiten los derechos de los musulmanes. Es cierto que no son este colectivo los únicos que resultan discriminados. Aunque su presencia es más bien testimonial, también hay mensajes de odio religioso contra los que profesan el cristianismo y otros mensajes de odio por razón ideológica.

El trabajo, por otro lado, abre unas perspectivas de investigación muy interesantes. Algunos de los análisis realizados anticipan que hay algunos patrones significativamente distintos entre el discurso del odio y la comunicación violenta, y probablemente esta sea una de las líneas de análisis que deben desarrollarse en el futuro. Además la lectura de los mensajes muestra que hay patrones de contenido claramente identificables dentro de cada una de las categorías, lo que podría ayudar en un futuro a la construcción de herramientas que permitan clasificar, de un modo muy preciso, aquellos mensajes que presenten contenidos de discurso del odio o violento.

Referencias bibliográficas

- AKDENIZ, Y. (1997). «To link or not to link: Problems with World Wide Web links on the Internet». [artículo en línea]. *International Review of Law, Computers & Technology*. Vol. 11, n.º 2. <<http://dx.doi.org/10.1080/13600869755721>>
- ALASTUEY DOBÓN, C. (2014). «La reforma de los delitos de provocación al odio y justificación del genocidio en el Proyecto de Ley de 2013: consideraciones críticas». *La Ley*. N.º 8245.
- ALCÁCER GUIRAO, R. (2012). «Discurso del odio y discurso político: en defensa de la libertad de los intolerantes». *Revista electrónica de ciencia penal y criminología*. N.º 14.

- ALCÁCER GUIRAO, R. (2013). «Libertad de expresión, negación del Holocausto y defensa de la democracia. Incongruencias valorativas en la jurisprudencia del TEDH». *Revista Española de Derecho Constitucional*. N.º 97.
- ALONSO ÁLAMO, M. (2012). «Sentimientos y Derecho penal». *Cuadernos de Política Criminal*. N.º 106.
- ALONSO RIMO, A. (2010). «Apología, enaltecimiento del terrorismo y principios penales». *Revista de Derecho Penal y Criminología*. N.º 4.
- AN, J.; KWAK, H.; MEJOVA, Y. [et al.] (2016). «Are you Charlie or Ahmed? Cultural pluralism in Charlie Hebdo response en Twitter». International AAAI Conference on Web and Social Media (ICWSM).
- AUSTIN, J. L. (1975). «How to do things with words». [artículo en línea]. *Harvard University Press*. Vol. 120. <<http://dx.doi.org/10.1093/acprof:oso/9780198245537.001.0001>>
- AWAN, I. (2016). *Islamophobia in Cyberspace: Hate Crimes Go Viral*. London: Routledge.
- BAKER, D. J. (2007). «The Moral Limits of Criminalizing Remote Harms Review». *New Criminal Law Review: An International and Interdisciplinary Journal*. Vol. 10, n.º 3.
- BARNUM, D. G. (2006). «The Clear and Present Danger Test in Anglo-American and European Law». *San Diego International Law Journal*. Vol. 7.
- BEN-DAVID, A.; FERNÁNDEZ, A. M. (2016). «Hate Speech and Covert Discrimination on Social Media: Monitoring the Facebook Pages of Extreme-Right Political Parties in Spain». *International Journal of Communication*. Vol. 10, n.º 27.
- BLANQUART, G.; COOK, D. M. (2013). «Twitter influence and cumulative perceptions of extremist support: A case study of Geert Wilders». Edith Cowan University Research Online.
- BOYD, D.; GOLDBERGER, S.; LOTAN, G. (2010). «Tweet, tweet, retweet: Conversational Aspects of Retweeting on Twitter». [artículo en línea]. En: *System Sciences (HICSS), 43rd Hawaii International Conference on IEEE*. <<http://dx.doi.org/10.1109/hicss.2010.412>>
- BRUNS, A.; BURGESS, J. E. (2011). «The use of Twitter hashtags in the formation of ad hoc publics». En: *Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011*.
- BURNAP, P.; WILLIAMS, M. L. (2015). «Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making». [artículo en línea]. *Policy & Internet*. Vol. 7, n.º 2. <<http://dx.doi.org/10.1002/poi3.85>>
- CALDEVILLA DOMÍNGUEZ, D. C. (2014). «Impacto de las TIC y el 2.0: consecuencias para el sector de la comunicación». *Revista de Comunicación de la SEECI*. Vol. 35.
- CANO PAÑOS, M. A. (2008). «Internet y terrorismo islamista: aspectos criminológicos y legales». *Eguzkilore*. N.º 22.
- CANO PAÑOS, M. A. (2015). «La reforma penal de los delitos de terrorismo en el año 2015: cinco cuestiones fundamentales». *Revista General de Derecho Penal*. N.º 23.
- CASTELLS, M. I. (2009). *Comunicación y Poder* [trad. de María Hernández Díaz]. Madrid: Alianza Editorial.
- CHAU, M.; XU, J. (2007). «Mining communities and their relationships in blogs: A study of online hate groups». [artículo en línea]. *International Journal of Human-Computer Studies*. Vol. 65, n.º 1. <<http://dx.doi.org/10.1016/j.ijhcs.2006.08.009>>
- CITRON, D. K.; NORTON, H. L. (2011). «Intermediaries and hate speech: Fostering digital citizenship for our information age». *Boston University Law Review*. Vol. 91.
- COHEN-ALMAGOR, R. (2009). «Holocaust denial is a form of hate speech». *Amsterdam Law Forum*.

- CORRAL, A.; FERNÁNDEZ, C. (2015). «Las agendas de lo araboislámico en Twitter y *El País*». En F. SABÉS TURMO y J. J. VERÓN LASSA. (coords). *Innovación y cambio en la comunicación postindustrial*. Zaragoza: Asociación de Periodistas de Aragón.
- CUERDA ARNAU, M. L. (2001). «El denominado delito de apología del genocidio. Consideraciones constitucionales». En: G. QUINTERO OLIVARES y F. MORALES PRATS (Coords.). *El Nuevo Derecho Penal Español. Estudios penales en memoria del Profesor José Manuel Valle Muñiz*. Navarra: Aranzadi.
- DEL ROSAL BLASCO, B. (1986). *La provocación para cometer delito en el Derecho español: (exégesis del párrafo tercero del artículo 4 del Código penal)*. Instituto Universitario de Criminología.
- DÍAZ LÓPEZ, J. A. D. (2013). *El odio discriminatorio como agravante penal: Sentido y alcance del artículo 22.4ª CP*. Navarra: Cizur Menor.
- DJURIC, N.; ZHOU, J.; MORRIS, R.; GRBOVIC, M.; RADOSAVLJEVIC, V.; BHAMIDIPATI, N. (2015). «Hate speech detection with comment embeddings». [artículo en línea]. Proceedings of the 24th International Conference on World Wide Web Companion. <<http://dx.doi.org/10.1145/2740908.2742760>>
- EFRON, M. (2010). «Hashtag retrieval in a microblogging environment». [artículo en línea]. En: *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. ACM. <<http://dx.doi.org/10.1145/1835449.1835616>>
- FEINBERG, J. (1985). *Offense to Others*. Oxford: Oxford University Press.
- FERNÁNDEZ-PLANELLS, A.; FEIXA PAMPOLS, C.; FIGUEROAS-MAZ, M (2013). «15-M En España: diferencias y similitudes en las prácticas comunicativas con los movimientos previos». [artículo en línea]. *Última Década*. Vol. 21, n.º 39, págs. 115-138. <<http://dx.doi.org/10.4067/S0718-22362013000200006>>
- FUENTES OSORIO, J. L. (2003). *La anticipación de la tutela penal y los actos preparatorios del delito*. Jaén: Universidad de Jaén.
- FUENTES OSORIO, J. L. (2006). «Formas de anticipación de la tutela penal». *Revista electrónica de ciencia penal y criminología*. N.º 8.
- GIBLIETTO, F.; LEE, Y. (2015). «To Be or Not to Be Charlie: Twitter hashtags as a discourse and counter-discourse in the aftermath of the 2015 Charlie Hebdo shooting in France». *4th Workshop on Making Sense of Microposts (# Microposts2014)*.
- GÓMEZ MARTÍN, V. (2012). «Discurso del odio y principio del hecho». En: S. MIR PUIG, M. CORCOY BIDASOLO(dir.). *Protección penal de la libertad de expresión e información*. Valencia: Tirant lo Blanch.
- GÓMEZ RIVERO, C. (1995). *La inducción a cometer delito*. Valencia:Tirant lo Blanch.
- GUIMELLI, C.; MONACO, G. L.; DESCHAMPS, J. C. (2010). «The lawsuit against «Charlie Hebdo» and its effects on the social representations of the Muslim Community». *Revue internationale de psychologie sociale*. Vol. 23, n.º 4.
- HARELL, A. (2010). «Political Tolerance, Racist Speech, and the Influence of Social Networks». [artículo en línea]. *Social Science Quarterly*. Vol. 91, n.º 3. <<http://dx.doi.org/10.1111/j.1540-6237.2010.00716.x>>
- HEFENDEHL, R. (2001). «¿Debe ocuparse el Derecho Penal de riesgos futuros? Bienes jurídicos colectivos y delitos de peligro abstracto». *Revista Anales de Derecho*. N.º 19.
- HERRERA-VIEDMA, E.; BERNABÉ-MORENO, J.; PORCEL, C.; MARTÍNEZ, M. A. (2015). «Solidaridad en la redes sociales: cuando el usuario abandona su zona de confort-el caso de *Charlie Hebdo*». [artículo en línea]. *Icono14*. Vol. 13, n.º 2. <<http://dx.doi.org/10.7195/ri14.v13i2.888>>
- JACKS, W.; ADLER, J. R. (2015). «A proposed typology of online hate crime». *Psychology Unbound: Open Access Journal of Forensic Psychology*.

- JENKINS, B. M. (2011). *Stray dogs and virtual armies: Radicalization and recruitment to jihadist terrorism in the United States since 9/11*. Santa Monica, CA: The Rand Corporation.
- KARMON, E. (2015). Islamic State and al-Qaeda Competing for Hearts & Minds. *Perspectives on Terrorism*, vol. 9, n.º 2.
- KHADER, J. (2015). «Repeating Fundamentalism and the Politics of the Commons: The Charlie Hebdo Tragedy and the Contradictions of Global Capitalism». *Islamophobia Studies Journal*. Vol. 3, n.º 1.
- KHAN, A. Z.; ATIQUE, M.; THAKARE, V. M. (2015). «Combining lexicon-based and learning-based methods for Twitter sentiment analysis». *International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCE)*. Vol. 89.
- KUBLER, F. (1998). «How Much Freedom for Racist Speech: Transnational Aspects of a Conflict of Human Rights». *Hofstra L. Rev.* Vol. 27.
- LANDA GOROSTIZA, J. M. (2012). «Incitación al odio: evolución jurisprudencial (1995-2011) del art. 510 CP y propuesta de lege lata». *UNED. Revista de Derecho Penal y Criminología*. 3ª Época, núm. 7.
- LAURENZO COPELLO, P. (1996). «Marco de protección jurídico-penal del derecho a no ser discriminado. Racismo y xenofobia». *Cuadernos de Derecho Judicial*. N.º 1.
- LAWRENCE III, C. R.; Matsuda, M. J.; Delgado, R.; WILLIAMS CRENSHAW, K. (1993). «Introduction». [artículo en línea]. En: M. J. MATSUDA, C. R. LAWRENCE III, R. DELGADO y K. WILLIAMS CRENSHAW (ed.). *Words that wound. Critical Race Theory, assaultive speech and the First Amendment*. Oxford: Westview Press. <http://dx.doi.org/10.1007/978-1-4684-0444-9_1>
- LEE, K.; PALSETIA, D.; NARAYANAN, R.; PATWAY, M. M. A.; AGRAWAL, A.; CHOUDHARY, A. (2011). «Twitter trending topic classification». [artículo en línea]. En: *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on IEEE*. <<http://dx.doi.org/10.1109/ICDMW.2011.171>>
- MAGDY, W.; DARWISH, K.; ABOKHOADIR, N. (2015). *Quantifying Public Response towards Islam on Twitter after Paris Attacks*. arXiv preprint arXiv:1512.04570.
- MASSEY, C. R. (1992). «Hate Speech, Cultural Diversity, and the Foundational Paradigms of Free Expression». *UCLA L. Rev.*, 40, 103.
- MATSUDA, M. J. (1989). «Public response to racist speech: Considering the victim's story». *Michigan Law Review*, 87(8).
- MATSUDA, M. J. (1993). *Words that wound: Critical race theory, assaultive speech, and the first amendment*. Boulder, Colorado: Westview Press.
- MAWEU, J. M. (2013). «The Ethnic Hate Speech was Networked: What Social Media Online Political Discussions Reveal about the 2013 General Elections in Kenya/El discurso étnico del odio se difundió por las redes. Análisis de las discusiones políticas en las redes sociales». *Index. comunicación: Revista científica en el ámbito de la Comunicación Aplicada*. Vol. 3, n.º 2.
- MCDEVITT, J.; LEVIN, J.; BENNETT, S. (2002). «Hate crime offenders: An expanded typology». *Journal of Social Issues*. Vol. 58, n.º 2.
- MERZEAU, L. (2015). #jesuisCharlie, ou le médium identité. A parâître dans Médium N43. halshs-01121510.
- MILTON PERALTA, J. (2012). *Motivos reprochables: una investigación acerca de la relevancia de las motivaciones individuales para el derecho penal liberal*. Madrid: Marcial Pons.
- MIRÓ LLINARES, F. (2012). *El cibercrimen. Fenomenología y criminología de la delincuencia en el ciberespacio*. Madrid: Marcial Pons.
- MIRÓ LLINARES, F. (2015). «La criminalización de conductas «ofensivas»: A propósito del debate anglosajón sobre los «límites morales del derecho penal». *Revista electrónica de ciencia penal y criminología*. N.º 17.

- NOBATA, C.; TETREAU, J.; THOMAS, A.; MEHDAD, Y.; CHANG, Y. (2016). «Abusive Language Detection in Online User Content». En: *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee.
- NORIEGA, C. A.; IRIBARREN, F. J.; LENIHAN, R.; YOUNG, A.; RAMIREZ, P. H. (2012). *Social Networks for Hate Speech: Commercial Talk Radio and New Media*. N.º 2.
- OH, O.; AGRAWAL, M.; RAO, H. R. (2013). «Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets During Social Crises». *Mis Quarterly*. Vol. 37, n.º 2.
- OH, O.; KWON, K. H.; RAO, H. R. (2010). «An Exploration of Social Media in Extreme Events: Rumor Theory and Twitter during the Haiti Earthquake 2010». *ICIS*.
- OLMEDO CARDENETE, M. (2000). *La inducción como forma de participación accesoria*. Editoriales de Derecho Reunidas. EDESA.
- PERRY, B. (2001). *In the name of hate: Understanding hate crimes*. New York: Psychology Press.
- POLLOCK, E. T. (2006). *Understanding and Contextualising Racial Hatred on the Internet: A Study of Newsgroups and Websites*. Nottingham: Nottingham Trent University.
- PORTILLA CONTRERAS, G. (2015). «La represión penal del «discurso del odio»». En: G. QUITERO OLIVARES (dir.). *Comentario a la reforma penal de 2015*. Navarra: Thomson Reuters-Aranzadi.
- REDDY, V. (2002). «Perverts and sodomites: Homophobia as hate speech in Africa». *Southern African Linguistics and Applied Language Studies*. Vol. 20, n.º 3.
- ROCKSTUHL, T.; ANG, S.; NG, K. Y.; LIEVENS, F.; VAN DYNE, L. (2015). «Putting judging situations into situational judgment tests: Evidence from intercultural multimedia SJTs». *Journal of Applied Psychology*. Vol. 100, n.º 2.
- RODRÍGUEZ, S. (2014). «El ámbito de aplicación del actual art. 510 CP en retrospectiva y en prospectiva tras la reforma penal de 2015». *Revista de derecho penal y criminología*. N.º 12.
- RODRÍGUEZ MONTAÑÉS, P. T. (2012). *Libertad de expresión, discurso extremo y delito*. Valencia: Tirant lo Blanch.
- ROLLNERT LIERN, G. (2008). «Revisionismo histórico y racismo en la jurisprudencia constitucional: los límites de la libertad de expresión (a propósito de la STC 235/2007)». *UNED-Revista de Derecho político*. N.º 73.
- ROLLNERT LIERN, G. (2014). «Incitación al terrorismo y libertad de expresión: el marco internacional de una relación problemática». *Revista de Derecho Político*. N.º 91.
- RUDNER, M. (2016). «'Electronic Jihad': The Internet as Al Qaeda's Catalyst for Global Terror». *Studies in Conflict & Terrorism*.
- RUIZ ANTÓN, L. F. (1998). «La acción como elemento del delito y la teoría de los actos del habla: Cometer delitos con palabras». *Anuario de derecho penal y ciencias penales*. Vol. 51, n.º 1.
- SABATER, C. P.; FLETA, M. B. M. (2015). «ESP vocabulary and social networking: The case of Twitter». *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)*. N.º 29.
- SEARLE, J. (1994). *Ensayo de Filosofía del Lenguaje*. Barcelona: Planeta de Agostini.
- SILKE, A. (2005). «Holy Warriors. Exploring the psychological Processes of Jihadi Radicalization». *European Journal of Criminology*. Vol. 5, n.º 1.
- SILVA, L.; MONDAL, M.; CORREA, D.; BENEVENUTO, F.; WEBER, I. (2016). «Analyzing the Targets of Hate in Online Social Media». *arXiv preprint arXiv:1603.07709*.
- SOBKOWICZ, P.; SOBKOWICZ, A. (2010). «Dynamics of hate based Internet user networks». *The European Physical Journal B*. Vol. 73, n.º 4.

- STENERSEN, A. (2008). «The Internet: A virtual training camp?». *Terrorism and Political Violence*. Vol. 20, n.º 2.
- TERUEL LOZANO, G. M. (2015). «La libertad de expresión frente a los delitos de negacionismo y de provocación al odio y a la violencia: sombras sin luces en la reforma del código penal». *Indret*. N.º 4.
- THOMPSON, R. (2011). «Radicalization and the use of social media». *Journal of Strategic Security*. Vol. 4, n.º 4.
- TIMOFEEVA, Y. A. (2003). «Hate speech online: restricted or protected? Comparison of regulations in the United States and Germany». *JTLP*. Vol. 12, n.º 2.
- TITLEY, G. (2014). «No apologies for cross-posting: European trans-media space and the digital circuitries of racism». *Crossings: Journal of Migration & Culture*. Vol. 5, n.º 1.
- TSUR, O.; RAPPOPORT, A. (2012). «What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities». En: *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM.
- TURIENZO FERNÁNDEZ, A. (2015). «El delito de negación del holocausto». *Indret*.
- VIVES ANTÓN, T. S. (1987). «Libertad de expresión y derecho al honor». *Revista de la Facultad de Derecho de la Universidad de Granada*. Vol. 2, n.º 13.
- WALDRON, J. (2012). *The harm in hate speech*. Harvard University Press.
- WALKER, C. (2015). «The United Kingdom terrorist attacks on July 7, 2005, and the evolution of anti-terrorism policies, laws, and practices». En: *Zeitschrift für Internationale Strafrechtsdogmatik*. N.º 11, pág. 545 y ss.
- WALKER, S. (1994). «Hate speech: The history of an American controversy. U of Nebraska Press». En: R. J. BOECKMANN, C. TURPIN-PETROSINO (2002). «Understanding the harm of hate crime». *Journal of Social Issues*. Vol. 58, n.º 2.
- WANG, A. H. (2010). «Detecting spam bots in online social networking sites: a machine learning approach». En: *Data and Applications Security and Privacy XXIV*. Springer Berlin Heidelberg.
- WANG, Z.; YIN, Y.; AN, B. (2016). Computing Optimal Monitoring Strategy for Detecting Terrorist Plots. 30th AAAI Conference on Artificial Intelligence (AAAI).
- WEIMANN, G. (2016). «The Emerging Role of Social Media in the Recruitment of Foreign Fighters». *Foreign Fighters under International Law and Beyond*.
- WILLIAMS, M. L.; SLOAN, L.; RANA, O.; HOUSLEY, W.; EDWARDS, A.; VOSS, A. (2014). «Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack». *Social Network Analysis and Mining*. Vol. 4, n.º 1.
- WOLSKA-ZOGATA, I. (2015). «The Story of Charlie Hebdo: An Analysis of European and American Newspapers». *Mediterranean Journal of Social Sciences*. Vol. 6, n.º 2.
- WOO, G. (2015). «Understanding the Principles of Terrorism Risk Modeling from Charlie Hebdo Attack in Paris». *Defence Against Terrorism Review-DATR*.

Cita recomendada

MIRÓ LLINARES, Fernando (2016). «Taxonomía de la comunicación violenta y el discurso del odio en Internet título». En: Josep Maria TAMARIT SUMALLA (coord). «Ciberdelincuencia y cibervictimización». *IDP. Revista de Internet, Derecho y Política*. N.º. 22, págs. 93-118. UOC [Fecha de consulta: dd/mm/aa] <<http://journals.uoc.edu/index.php/idp/article/view/n22-miro/n22-miro-pdf-es>> <<http://dx.doi.org/10.7238/idp.v0i22.2975>>



Los textos publicados en esta revista están –si no se indica lo contrario– bajo una licencia Reconocimiento-Sin obras derivadas 3.0 España de Creative Commons. Puede copiarlos, distribuirlos y comunicarlos públicamente siempre que cite su autor y la revista y la institución que los publica (*IDP. Revista de Internet, Derecho y Política*; UOC); no haga con ellos obras derivadas. La licencia completa se puede consultar en: <http://creativecommons.org/licenses/by-nd/3.0/es/deed.es>.

Sobre el autor

Fernando Miró Llinares
fmiro@umh.es

Catedrático acreditado de Derecho penal
Universidad Miguel Hernández

<<http://derechopenal.umh.es/directorio/fernando-miro-llinares/>>

Universidad Miguel Hernández
Departamento de Ciencia Jurídica
Avda. de la Universidad, s/n.
Edif. Torrevalillo
ELCHE - 03201