EDICIONES COMPLUTENSE

Estudios sobre el **Mensaje Periodístico**
ISSN-e: 1988-2696

ARTÍCULOS

La construcción del mensaje perfecto en un entorno digital. El caso de Vinicius Junior y el racismo

José Cabeza-San-Deogracias
Universidad Rey Juan Carlos

María Antonia Paz-Rebollo
Universidad Complutense de Madrid

Raúl Casado-Linares
Universidad de Ciencias Aplicadas de Ámsterdam (Países Bajos)

https://dx.doi.org/10.5209/emp.98077 Recibido: 23 de septiembre de 2024 / Aceptado: 2 de diciembre de 2024

ES **Resumen**. Esta investigación estudia las características de los comentarios que logran más impacto en un entorno digital. Se emplea una metodología cualitativa que analiza los 50 mensajes con más *likes* del foro de una noticia, publicada en *Marca.com*, en la que el jugador del Real Madrid Vinicius Junior denuncia los episodios racistas sufridos en los estadios de fútbol españoles. Se tienen en cuenta las emociones expresadas por los autores, así como las estrategias retóricas y argumentales utilizadas en los comentarios. Se concluye que los mensajes escritos con empatía, aunque estén en contra de la opinión mayoritaria, pueden tener cierto impacto si formulan un *argumento dual* que reconoce las opiniones contrarias como valiosas en algún punto, incluso sin validarlas. En el análisis de la eficacia de los mensajes se observa que lo distinto se premia, como aportar un dato inesperado de forma inteligente (ironía) o presentar una opinión contraria a lo que se espera por pertenecer a un determinado grupo (afiliación): en este último caso el comentario resulta más auténtico y, por lo tanto, es más impactante. Se demuestra también que la percepción de los mensajes es compleja y que algunos elementos o rasgos que habitualmente se consideran eficaces pueden no serlo. Así sucede con los comentarios breves, que no siempre tienen trascendencia y pueden aparecer como inconsistentes si no incluyen una argumentación. Tampoco la diversidad o la acumulación de argumentos es positiva en cualquier circunstancia: compensa más usar los argumentos por separado para no distraer del objetivo persuasivo.
Palabras clave: Discurso de odio, Vinicius, Entorno digital, Racismo, Contranarrativas.

ENG The construction of the perfect message in a digital environment: The case of Vinicius Junior and racism

Abstract. This research examines the characteristics of comments that achieve the greatest impact in a digital environment. A qualitative methodology is used to analyse the 50 most liked messages in the forum discussion prompted by a news article published on the Spanish sports newspaper Marca.com. The article focuses on Real Madrid player Vinicius Junior's denunciation of racist incidents in Spanish football stadiums. This study considers the emotions expressed by the authors, as well as the rhetorical and argumentative strategies employed in the comments. The findings suggest that messages written with empathy, even when opposing the opinion of the majority, can have a significant impact if they present a *dual argument* that acknowledges opposing views as valuable in some respects, even without validating them. The analysis reveals that originality is rewarded, such as introducing an unexpected fact in an intelligent manner (e.g., irony) or expressing an opinion contrary to expectations based on group belonging (affiliation). In the latter case, the comment appears more authentic and, therefore, more impactful. The study also demonstrates that the perception of messages is complex, and some elements traditionally considered effective may not always be so. For example, brief comments do not always carry weight and can come across as inconsistent if they lack sufficient argumentation. Similarly, the diversity or accumulation of arguments is not universally positive; it is often more effective to present arguments separately to avoid distracting from the persuasive objective.
Keywords: Hate speech, Vinicius, Digital environment, Racism, Counter-narratives.

Cómo citar: Cabeza-San-Deogracias, J., Paz-Rebollo, M. A. y Casado-Linares, R. (2025). La construcción del mensaje perfecto en un entorno digital. El caso de Vinicius Junior y el racismo. *Estudios sobre el Mensaje Periodístico, 31*(1), 77-88. https://dx.doi.org/10.5209/emp.98077

1. Introducción

Las expresiones de odio denigran y ofenden a una persona por pertenecer a un grupo (también se pueden dirigir a todo el grupo) definido por su raza, religión, género, sexo, edad, entre otros aspectos (Paz *et al.*, 2020). Los investigadores coinciden en señalar a Internet como la herramienta más poderosa que contribuye a su expansión (Nave y Lane, 2023): se comparten anónimamente, alcanzan una rápida y amplia difusión y constituyen un grave problema social al negar los valores de tolerancia fundamentales en una sociedad democrática.

Las expresiones de odio están presentes mayoritariamente en las redes sociales, pero también en los mensajes de los usuarios publicados en la prensa digital. En estas webs las noticias políticas no sólo son las que más comentarios provocan (Boberg *et al.*, 2018), sino también las que generan mayor incivilidad (Svenja *et al.*, 2018). No obstante, existe igualmente lenguaje ofensivo en las noticias blandas, como las relacionadas con los deportes (Bonaut *et al.*, 2023; Cabeza *et al.*, 2024). En estas informaciones, el enfoque emocional y algunas tendencias de la escritura periodística, como el sensacionalismo, pueden tener un efecto explosivo de odio en una audiencia ya de por sí emocionalmente variable.

Los estudios sobre las formas de combatir los discursos de odio en Internet se han centrado sobre todo en la moderación y la vigilancia de estas expresiones (Fortuna y Nunes, 2018). Pero los resultados no son plenamente satisfactorios, puesto que, aunque la detección automática es capaz de identificar y eliminar algunos de estos discursos (Wang *et al*., 2024), los usuarios utilizan técnicas cada vez más sofisticadas de escritura (uso de guiones, signos, números, prefijos) que dificultan la detección, además del uso de la ironía, el sarcasmo o las metáforas imposibles de interpretar por ninguna IA. Los algoritmos se completan en ocasiones con la moderación humana incapaz de atender el volumen de mensajes difundidos diariamente. Tampoco son plenamente eficaces las denuncias de los usuarios, por el coste social que implica (Hansen *et al.*, 2024). Por otra parte, la eliminación sistemática de comentarios entra en colisión con el respeto a la libertad de expresión de la ciudadanía (Llansó *et al.*, 2020).

Una opción, cada vez más reclamada, consiste en la creación de narrativas que contraargumenten los comentarios ofensivos y presenten perspectivas alternativas o socaven la supuesta autoridad de los que emiten esas expresiones. Se entiende por contranarrativa cualquier forma de expresión que pretende influir en quienes simpatizan con el discurso de odio o participan en su construcción y difusión (Baider, 2023). En opinión de Chung *et al.* (2021), esta posibilidad resulta interesante porque «preserva el derecho a la libertad de expresión, se rectifican los estereotipos y se fomenta el intercambio de diversos puntos de vista». Además, la intervención de los usuarios contra los comentarios poco cívicos puede impactar en la calidad de los comentarios posteriores (Friess *et al.*, 2021).

Esta investigación estudia las características retóricas y formales de los comentarios que aparecen de manera espontánea en entornos digitales y que logran una mayor atención. Así tienen más posibilidades de construir, cambiar o rebatir tendencias de opinión en cualquier foro. Se trata de un estudio empírico *microscópico* (Golder y Macy, 2014) que analiza 50 comentarios sobre la noticia «Los 10 mensajes contra el racismo de Vinicius en su rueda de prensa más reivindicativa» (Rodríguez, 2024). A partir de 2023, los episodios racistas contra Vinicius Junior, jugador del Real Madrid, aparecen con virulencia y exposición mediática en España (colgar un muñeco con su camiseta en un puente o insultos en los estadios que incluso llevaron a suspender durante unos minutos un partido). En paralelo también aumentan las críticas sobre su comportamiento en el terreno de juego (celebraciones, gestos, actitudes). En mayo de 2023, el caso de Vinicius trascendió al ámbito internacional y se convirtió en símbolo de la lucha contra el racismo en el deporte con declaraciones de apoyo de personalidades deportivas y políticas de diferentes países, incluso del alto comisionado sobre Derechos Humanos de la ONU, Volker Türk, que exigió llevar a cabo acciones para «evitar y contrarrestar» el racismo que emerge en el mundo deportivo (Naciones Unidas, 2023). El jugador siguió denunciando comportamientos racistas y la situación llegó al límite durante la rueda de prensa anterior al partido amistoso España-Brasil. Vinicius lloró al explicar y defenderse de los episodios racistas que sufría en España (Rodríguez, 2024).

2. Marco teórico

Las opiniones y los razonamientos que aparecen en estas conversaciones *online* son importantes porque las creencias y los comportamientos que las personas perciben de los demás influyen en sus actitudes, sobre todo de índole política (Garland *et al.*, 2022). Es lo que los psicólogos denominan «efecto contagio» (Buerger, 2021). En este sentido, resulta especialmente relevante el análisis de las argumentaciones, puesto que los argumentos son más eficaces que los simples insultos o la presentación de hechos sin conclusiones explícitas a la hora de desafiar al discurso de odio. En este sentido, cabe distinguir el juicio de valor que es una afirmación que no contiene hechos ni datos, sino que ofrece una valoración personal y revela la actitud del orador ante el tema (Monakhova y Tuluzakova, 2022); el punto de vista como encuadre (*frame*) que un autor otorga a un tema; y las expresiones argumentativas que ponen de relieve opiniones o afirmaciones que defienden un punto de vista interactuando con otros usuarios (Furman *et al.*, 2022).

Los estudios actuales de la teoría de la argumentación clasifican los argumentos desde propuestas empíricas, por lo que tienen interés para el análisis de conversaciones *online*. Pero, precisamente por tratar casos reales, estos esquemas de argumentación no resultan precisos (Walton *et al*., 2008, pp. 12-13), ni coincidentes sobre su naturaleza y número, lo que dificulta el trabajo de los investigadores. Entre las propuestas esquemáticas de tipos de argumentos destacan los trabajos de Walton (1996) y el de Wagemans (2016). El primero porque ofrece una clasificación que identifica muchas de las formas más comunes de argumentación. El segundo porque elabora una Tabla Periódica de Argumentos práctica para identificar tipos de argumentos. Ambas pro-

puestas se utilizan indistintamente. Por ejemplo, Saha y Srihari (2023) experimentan con modelos computacionales para detectar el tipo de razonamiento presente en un texto, seleccionando seis aspectos de los incluidos en la clasificación de Walton («Medios para la meta», «Meta a partir de los medios», «A partir de la consecuencia», «Conocimiento de la fuente», «Autoridad de la fuente» y «Regla o principio»).

Sin embargo, Furman *et al.* (2022) aplican la propuesta de Wagemans (2016) en su investigación para detectar esquemas argumentativos presentes en tuits con odio y comprobar cómo interactúan con otros argumentos. Su investigación se centra en la capacidad de la inteligencia artificial de identificar argumentos y construir modelos de razonamiento que puedan dar respuesta a expresiones de odio. Concluyen que hay aspectos, como la identificación de conclusiones o lo que los autores denominan «pivote», en los que la IA todavía no consigue los resultados esperados. La propuesta de Chung *et al.* (2019) intenta identificar varias categorías en las respuestas a los comentarios de odio con resultados eficaces: presentar hechos, evidenciar la hipocresía y la contradicción, advertir sobre las consecuencias, utilizar el humor, el tono positivo, y contra-pregunta para provocar la reflexión. Con estas categorías se realizó un experimento: el 85 % de las veces los participantes vincularon correctamente el discurso de odio con la contra-narrativa adecuada. A partir de estos resultados se creó en X el proyecto Conan para ofrecer automáticamente una lista de contranarrativas multilingües con las que responder a comentarios islamófobos (Chung *et al.*, 2021).

Mun *et al.* (2023) inciden en la importancia de la refutación específica que aporte ejemplos o muestre las razones que conducen a la creación de un estereotipo frente al uso de conceptos más genéricos porque resultan menos convincentes. Estos autores experimentan con seis estrategias, que diseñan para contrarrestar estereotipos, basándose en la literatura sobre la psicología social, cognición social y filosofía del lenguaje. Estas estrategias son: poner en duda la veracidad de esos estereotipos; relacionar ese grupo con otros; exponer cualidades alternativas; ofrecer ejemplos que demuestren que el estereotipo no se puede aplicar a todos los miembros del grupo; explicar factores externos que condujeron a esas ideas; demostrar que un grupo no puede valorarse por una sola cualidad. Concluyen que la respuesta correcta está muy condicionada por el contexto, es decir, el contra-discurso eficaz depende del estereotipo y del grupo aludido: no hay fórmulas universales. A ello, hay que sumar los niveles de incivilidad que también inciden en las respuestas generadas por los usuarios (Yu *et al.*, 2024).

Se subrayan otras tácticas que pueden resultar positivas y convincentes. Por ejemplo, la utilización de un tono empático y la inclusión de imágenes (Wright *et al.*, 2017) o simplemente evitar un lenguaje tóxico en la conversación y tener un tono positivo (Saha *et al.*, 2022). Pero también se ha demostrado que el apelar a la comprobación de los hechos para evitar la desinformación y las noticias falsas puede promover un cambio de ideas (Porter y Wood, 2021) o, al menos, permitir que las personas intenten emitir juicios con un mejor conocimiento de la realidad.

Se apuesta, en general, por utilizar mensajes sencillos. En estos no se considera importante que los argumentos no estén respaldados por hechos. Sin embargo, se señala como relevante que no incluyan insultos ni emociones tanto negativas (ira) como positivas (entusiasmo), porque ambas alientan las conversaciones incívicas (Lasser *et al.*, 2023). De hecho, en muchos casos, las emociones positivas no reciben respuestas (Baider, 2023). En general, en los foros de los medios de comunicación, predominan las emociones hostiles, como la ira y el miedo, especialmente cuando los usuarios consideran que la información de un hecho es una provocación porque no coincide con su visión (Ihlebæk y Holter, 2021). Para Potash y Rumshisky (2017) es importante el número de participantes que argumentan en contra, así como la diversidad en la contraargumentación, entre otros aspectos como el lenguaje utilizado (el uso de pronombres personales, por ejemplo) porque, como también afirman Mun *et al.* (2023), la especificidad es una característica esencial frente a los estereotipos que son afirmaciones genéricas.

Se evidencia así la necesidad de realizar estudios sobre las estrategias necesarias para reducir el discurso de odio basados en el análisis de conversaciones *online* porque ofrecen pistas a los investigadores sobre los comportamientos reales de las personas en estos foros de discusión.

3. Objetivos y metodología

El objetivo de esta investigación es identificar los comentarios que logran más impacto en una conversación *online*. Se busca la forma y la elaboración retórica del comentario perfecto en términos de eficiencia: el que más adeptos consigue independientemente de si transmite odio o lo neutraliza. Se intenta detectar, en primer lugar, qué características tiene; y, en segundo lugar, qué elementos circunstanciales lo condicionan y aumentan o disminuyen su potencial para ser eficaz en un entorno digital determinado.

Se plantean las siguientes preguntas de investigación:

> P1. ¿Qué influencia ejercen los elementos emocionales en el impacto de una intervención, ya sea por dinámicas generales o subjetivas del autor?
> P2. ¿Cuáles son las características relevantes de los mensajes, respecto a la extensión, recursos retóricos y argumentativos y construcción formal, para la obtención de apoyo o de rechazo?

El impacto se mide en función del número de *likes* y de *dislikes* obtenidos de otros usuarios, puesto que se considera un indicador objetivo de la interacción y el apoyo o rechazo logrado (Drummond, O'Toole y McGrath, 2020). Se ha seleccionado la rueda de prensa de Vinicius Junior antes del partido amistoso España-Brasil y sus quejas sobre el racismo que había sufrido en España para este análisis. Se estudia el hilo que se generó a partir de esa rueda de prensa y la noticia «Los 10 mensajes contra el racismo de Vinicius en su rueda de prensa más reivindicativa», publicada en el diario deportivo *Marca.com*, el cuarto sitio web más visitado de España, sólo superado por las páginas de Google, Amazon y Youtube (Asocia-

ción para la Investigación de Medios de Comunicación, 2024, p. 88). Del total de mensajes publicados (1133), más de un tercio (407) fueron eliminados al no cumplir las normas de participación del foro. Aun no pudiendo acceder al contenido de estos mensajes, es probable que contuviesen insultos o faltas de respeto dirigidos a Vinicius o a otros usuarios.

El foro permaneció abierto del 25 de marzo a las 17 horas hasta el 26 de ese mismo mes a las 22 horas aproximadamente. A las 5 horas de actividad ya había 600 comentarios, 1000 a las 16 horas y se cerró con 1133 a las 30 horas. Estos comentarios se descargaron mediante un proceso de *web scraping* utilizando la herramienta *Instant Data Scraper*. Se generó una base datos con los siguientes campos: nombre del usuario, número de comentario, fecha y hora de la publicación, el texto del mensaje, orden de la valoración en el caso de los 572 más valorados (según el diario). Se anotó manualmente el número de *likes* y de *dislikes*. Uno de los autores llevó a cabo un primer análisis temático (*frames*) sobre el total de mensajes. Clasificó los comentarios en función de estos y añadió si estaban a favor del jugador o en contra. Esta clasificación permitió obtener tablas de frecuencia: número de comentarios de cada *frame* y su peso relativo a lo largo de la conversación. En total se identificaron nueve *frames* contra Vinicius: *Es un provocador*, *Se hace la víctima*, *Comparación con otros jugadores afrodescendientes*, *Que se vaya de la Liga*, *Actúa para un documental de Netflix*, *Tiene problemas psicológicos*, *Ataques a su nivel económico*, *En Brasil sí hay racismo* y *Otros* (comparaciones con *Martin Luther King o Nelson Mandela)*. Los encuadres a favor del jugador identificados son tres: *En defensa de Vinicius*, *Denuncia de odio*, *Otros* (por ejemplo, apelaciones de apoyo directas dirigidas al jugador). En *Miscelánea* se agruparon los mensajes que recogían *Disputas entre usuarios* y *Ataques al Real Madrid* (un total de 209 comentarios)[1].

El análisis textual se aplicó a los 50 mensajes más apoyados con *likes* por los usuarios, correspondientes a los tres *frames* en contra del jugador con mayor presencia en la conversación (*Es un provocador*, *Se hace la víctima*, *Comparación con otros jugadores afrodescendientes*) y a los dos centrados en la *Defensa de Vinicius* y *Denuncia de odio*. Se analizaron las emociones negativas de los participantes (Ihlebæk y Holter, 2021), expresadas a través de sus comentarios (decepción, ira, vergüenza, indignación, entre otras), con las que se creó el sentir general del foro. Igualmente se identificaron las figuras retóricas (humor, ironía, metáforas, elipsis, anáforas, entre otras), el tipo de argumentos de los comentarios, según la clasificación de la Tabla periódica de Wagemans (2016) y, en general, todas las especificidades que pudieron tener influencia en el impacto del comentario y hacerlo más movilizador. En este sentido, se tuvo en cuenta la apelación a diferentes interlocutores (el jugador, participantes en el foro o uso del impersonal), el impacto del insulto y la extensión del enunciado.

Se analizaron, en primer lugar, los cien mensajes más valorados en la conversación; a continuación, se diseccionaron los comentarios con más *likes* dentro del encuadre que se posiciona en contra de Vinicius y, finalmente, se valoraron los mensajes que intentan bloquear las opiniones en contra del jugador.

4. Resultados

Los cien mensajes más valorados de todo el foro reflejan proporcionalmente los encuadres que más insistentemente aparecen. El rechazo masivo a las declaraciones de Vinicius y, sobre todo, a su argumentación (la existencia del racismo) hace que la proporción de mensajes de apoyo al jugador del Madrid sea nula en los comentarios que más éxito tuvieron entre los usuarios (Tabla 1).

Tabla 1. Frecuencia de aparición de todos los *frames* en los 100 mensajes más valorados.

Frame	Mensajes	Frecuencia
Contra Vinicius	114	97%
Es un provocador	38	32%
Comparación con otros jugadores afrodescendientes	19	16%
Se hace la víctima	14	12%
Problemas psicológicos	11	9%
Actúa para su documental (Netflix)	10	8%
Que se vaya de la Liga	9	8%
Ataques por su nivel económico	5	4%
Otros	4	3%
En Brasil sí hay racismo	4	3%
Miscelánea	4	3%
Otros	2	2%
Disputa entre usuarios	1	1%
Ataques al Real Madrid	1	1%
TOTAL	118	100%

Fuente: elaboración propia.

1 El sumatorio total en número de mensajes es mayor que 1133, porque hay mensajes que se adscriben a más de un *frame*.

Sólo diez comentarios en todo el foro están por encima de los 500 *likes*: el primero tiene 2047 apoyos, mientras que el décimo logra 592. El momento de su publicación influye de forma decisiva en su éxito: los diez más respaldados están dentro de los primeros 24 comentarios publicados y un 50 % incluso se encuentra entre los diez que aparecieron antes. Ser de los primeros asegura un flujo importante de lectores, aunque también influye tanto o más la conexión emocional con el foro: todos critican a Vinicius. Paradójicamente, los dos únicos comentarios que están por encima de los 2000 *likes* no pertenecen a ninguno de los encuadres predominantes (Tabla 2), aunque sí recogen observaciones muy próximas al magma emocional que sostiene esos encuadres: *Actúa para su documental (Netflix) y En Brasil sí hay racismo*. Hay dos diferencias claras entre ambos mensajes: la forma de expresión y la cantidad de rechazos. El más apoyado utiliza la ironía («Va a quedar divino en el documental. ¿Salió a la primera o hubo que repetir toma?»), aprovechando un dato que los lectores ya conocían, aunque no consta en la información: Netflix estaba grabando un documental con Vinicius.

Tabla 2. Frecuencia de aparición de todos los *frames* en todo el foro.

Frame	Mensajes	Frecuencia
Contra Vinicius	536	42%
Es un provocador	120	9%
Se hace la víctima	86	7%
Comparación con otros jugadores afrodescendientes	85	7%
Que se vaya de la Liga	84	7%
Actúa para su documental (Netflix)	49	4%
Problemas psicológicos	33	3%
Ataques por su nivel económico	32	3%
En Brasil sí hay racismo	24	2%
Otros	23	2%
A favor de Vinicius	124	10%
En defensa de Vinicius	58	5%
Denuncia de odio	44	3%
Otros	22	2%
Miscelánea	209	16%
Disputa entre usuarios	83	7%
Otros	77	6%
Ataques al Real Madrid	49	4%
Eliminados	407	32%
Comentarios eliminados	407	32%
TOTAL	1.276	100%

Fuente: elaboración propia.

El segundo comentario en lograr más adeptos sigue otra estrategia; la vía directa: «Donde más racistas hay es en Brasil. Empieza por ahí». En esta ocasión se establece una comparación con el país de origen del futbolista. Las dos argumentaciones encuentran un respaldo casi idéntico y mayoritario, pero la segunda genera más oposición que la primera: un 49 % más (257 vs. 172) etiquetan como negativa la observación. Si la intervención incluye al menos una argumentación que se sienta como válida («está actuando para un documental») y no como extrema o absurda («donde más racistas hay es en Brasil») puede generar más comprensión y, por lo tanto, bloquear una primera reacción que lleva al rechazo después de su lectura.

La comparación entre el quinto y el sexto mensajes que más aprobación obtienen deja también un dato interesante sobre el peso de una argumentación, es decir, sobre cómo la simple atribución de razones a juicios directos sobre algo o alguien puede influir en la percepción de un mensaje:

> Mensaje número 5 (1240 *likes*; 140 *dislikes*, posición en el foro: #9):
> «La gente racista SOLO contigo, con Camavinga, Rüdiger, etc no lo son, Provocador. Lágrimas de cocodrilo».
> Mensaje número 6 (1252 *likes*; 192 *dislikes*, posición en el foro: #2):
> «Lágrimas de cocodrilo».

Ambos mensajes aparecen casi inmediatamente en el foro, ambos son bastante concisos y apenas hay diferencia en la aceptación (0.9 %). Los dos comparten un mismo núcleo comunicativo: subrayan que las lágrimas de Vinicius no son auténticas. Sin embargo, tienen un nivel muy diferente de rechazos. Uno intenta argumentar la falta de veracidad de la afirmación con un argumento de similitud (Wagemans, 2016): otros jugadores afrodescendientes no sufren el racismo que él denuncia. El otro autor decide no matizar, ni poner en contexto, ni ampliar la carga significativa de lo que quiere transmitir. En este último caso, casi un 37 % más de personas etique-

tan el comentario de forma negativa. Que un mensaje sea extremadamente breve puede facilitar la adhesión por su simplicidad, de hecho siete de los primeros diez mensajes de los que reciben mayor aprobación están compuestos por menos de 24 palabras. Ahora bien, la concisión también comporta ciertos riesgos. Se deja menos espacio para incluir a lectores que quizás necesitan leer algo más para ser convencidos o que simplemente les gusta tomar una postura más neutral o de cierta comprensión. Una valoración que no está respaldada por un hecho provoca más rechazo.

Por otra parte, el exceso de argumentación puede resultar negativo. El mensaje número 10 es bastante más extenso que cualquier de los 100 primeros mensajes mejor valorados: 185 palabras. Recoge cinco argumentaciones ya expuestas, directa o indirectamente, en los primeros 100 mensajes que consiguen más adeptos de forma aislada que cuando aparecen todos juntos. Se produce así un *efecto de saturación*: ideas o elementos retóricos que sí funcionan individualmente en otros mensajes, cuando se dan de forma sumada o combinada o bien embotan la percepción del que lee, y ya no les da el mismo valor, o compiten por la atención y eso provoca un menor efecto de adhesión. Cuando todos aparecen a la vez se anulan.

Igualmente se evidencia que los juegos de palabras o el uso del lenguaje de forma sofisticada pueden lograr que el argumento construido sea mejor, simplemente por su carácter rítmico o visual, y tenga así más posibilidades de quedar anclado en la memoria del lector y desee hacerlo suyo. Si estos mensajes se comparan con otros que expresan una idea similar o idéntica, pero con elaboración formal más pobre, se comprueba la diferencia. «No es racismo, es tu gilipollez» es el primer mensaje que logra romper con el dominio de los que aparecen antes (#1-24): se publica en el puesto #49 y aun así recibe una valoración alta: 458. En principio, la moderación elimina los insultos, pero, en este caso, la falta de respeto se envuelve en una estructura más sofisticada (uso de asíndeton) que hace que el juego verbal disminuya el rechazo que provoca la presencia de una palabra malsonante. También el mensaje 13 en aprobación incluye una figura retórica (anáfora) de manera eficaz: «El jugador más desequilibrante de la liga y de Europa. Desequilibrante de la cabeza».

No quiere decirse que la forma venza siempre. Hay mensajes que son atractivos gramaticalmente, pero su capacidad de atraer la atención puede ser más limitada frente a una argumentación expresada con llaneza. El mensaje 75 logra ser el 19 más aplaudido: «El problema no es que sea negro, el problema es EL, únicamente!». Pero apenas un mensaje después, el número 76, se sitúa en la posición 14: «Cuando se aplazó el juicio porque estabas de vacaciones y no te querías conectar por videollamada, entonces no llorabas, no?». Este mensaje recibe 310 *likes*, 200 el anterior. La extensión y el momento en el que aparecen son similares y los dos pertenecen a encuadres muy seguidos. La diferencia está en el fondo y en la forma: uno más ingenioso, el otro recupera un dato conocido: Vinicius decidió no declarar en verano cuando se le invitó a ello.

4.1. Dinámica de los encuadres en contra de Vinicius

Los comentarios, analizados de manera individual, pueden parecer inconexos, como se ha visto en otros foros (Paz *et al.*, 2023), pero observados en conjunto presentan una coherencia argumental interna. En concreto, los encuadres de los argumentos en contra de Vinicius pretenden desmentir las acusaciones de racismo por parte del jugador. Los ataques se centran en su actitud. De esta manera se alejan del discurso de odio, que denigra a una persona por pertenecer a un grupo vulnerabilizado, en este caso, por ser afrodescendiente.

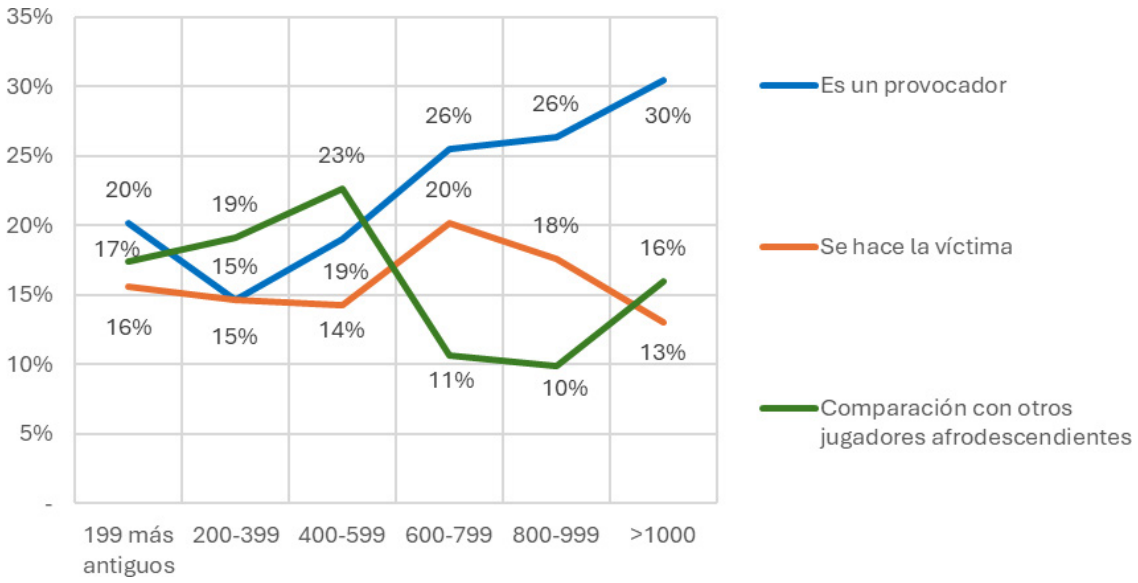
El que obtiene una mayor valoración de los participantes en el foro, en función del número de *likes* obtenidos por el 5 % de los mensajes más apoyados (6986 *likes*), es el que considera que el jugador provoca con su actitud los insultos que recibe (*Es un provocador*). Le siguen, con 4036 *likes*, los comentarios que consideran que hay otros jugadores afrodescendientes en diversos equipos y son tratados con respeto (*Comparación con otros jugadores afrodescendientes*). A distancia (717 *likes*) se sitúan los que opinan que se presenta falsamente como víctima (*Se hace la víctima*). Este es también el que obtiene un porcentaje menor de *dislikes* (un 8,2 %) frente a los otros dos que reciben en torno al 11 % de *dislikes*. Quiere decirse que, a pesar de que agrupa prácticamente el mismo número de mensajes que *Comparación con otros jugadores afrodescendientes* (85 comentarios), *Se hace la víctima* no moviliza ni a favor ni en contra, porque los comentarios de este encuadre se centran mayoritariamente en insultos y, como se ha visto, éstos no funcionan.

Ahora bien, la secuencialidad de estos *frames* evidencia que los comentarios que plantean ciertas proposiciones de hecho se agotan en el tiempo, mientras que los que emplean proposiciones de valoración se mantienen (Gráfico 1). En concreto, el encuadre *Comparación con otros jugadores afrodescendientes* decae, a partir de las cinco horas de debate (comentarios 600 a 799), porque ya se han planteado todas las posibles similitudes y, cuando estas se llevan al terreno de la analogía, no funcionan. Por ejemplo, el comentario 16 («si hay 4 obesos, y solo le gritan e insultan a uno, porque es un chulito... Háztelo mirar tu») es el 194 en valoración. En general, los usuarios prefieren concentrarse en el tema y no buscar paralelismos que puedan confundir acerca de la cuestión que se está debatiendo. Como se verá en el siguiente apartado, la analogía funciona si se relaciona con ámbitos que afecten a la mayoría de los participantes. *Es un provocador,* sin embargo, permite plantear diferentes razones y consecuencias de esa provocación. También *Se hace la víctima*, pero al centrar su interés en la variedad de insultos como se ha señalado, decae a partir del comentario 1000, mientras que *Es un provocador* se mantiene con vigor.

Algunas de las figuras retóricas utilizadas en estos *frames* son semejantes. Como se ha comentado, los mensajes a modo de eslóganes, con construcciones gramaticales breves (asíndeton, anáfora, elipsis), impactan independientemente del *frame* («Respeta aficiones y rivales» #38, el 20º

en valoración; «No hay Dios que se fume a este chico. Que pesado», #885, 55 en valoración). En otros se respalda la verdad del argumento con la afiliación del autor. Se defiende una idea que, en teoría, es contraria a lo esperado por ser hincha de un equipo rival o del propio equipo. Esta afirmación inesperada cala no por lo que se dice, sino por quién lo dice. Se pueden citar diversos ejemplos: «Soy del madrid y boy todos los fines de semana al bernabeu y este chico va a acabar mal con la afición» (104 *likes*); y «Ojalá te venda mi Real Madrid. No me representas...» (ocupa el puesto 63 en valoración). Funcionan porque parece una opinión objetiva y sincera.

Gráfico 1. Desarrollo temporal de los *frames* (distribución porcentual por *frame*).



Fuente: elaboración propia.

Los autores coindicen también en manifestar sus emociones, tanto a través de los adjetivos como de los verbos. Son emociones negativas y responden al sentir general de los *frames* en contra de Vinicius. La expresión con más éxito localizada es la que plantea un hecho positivo junto al sentir que provoca —en este caso, decepción— que permite hacer una valoración crítica: «El problema es que después de lo de Mestalla había mucha gente de tu lado, pero después de hacer lo mismo en los demás campos ya sólo te compran tu discurso los madridistas y los periodistas afines al régimen de florentino» #12. Es el cuarto en valoración por los usuarios (1672 *likes*). Pero, en general, los mensajes que expresan estas emociones cuentan sólo con un apoyo medio, es decir, se posicionan entre los 500 primeros en número de *likes*. Los comentarios en los que el autor se manifiesta indignado, así como en los que el enfado deriva en ira y en aversión, son los más apoyados, además de la decepción ya mencionada: «No aguanto más tu cruzada sin sentido» (#218, 35 en valoración, 80 *likes*); o «... es acojonante...» (#381, 36 en valoración, 71 *likes*). La vergüenza por lo que se considera un comportamiento inapropiado del jugador, posiblemente por mencionarse de manera reiterada (se menciona en 23 comentarios), alcanza menos notoriedad salvo si se acompaña de una propuesta de solución al problema, en concreto de la expulsión del jugador del equipo incluso del país, como se puede leer en el comentario 693 (puesto 88 en valoración): «Esto roza ya la vergüenza ajena. Qué hace este tío en el Madrid? Jajajajajajajajaa», frente a los que sólo expresan ese sentir: «Vergüenza da este personaje, madre mía» (# 569).

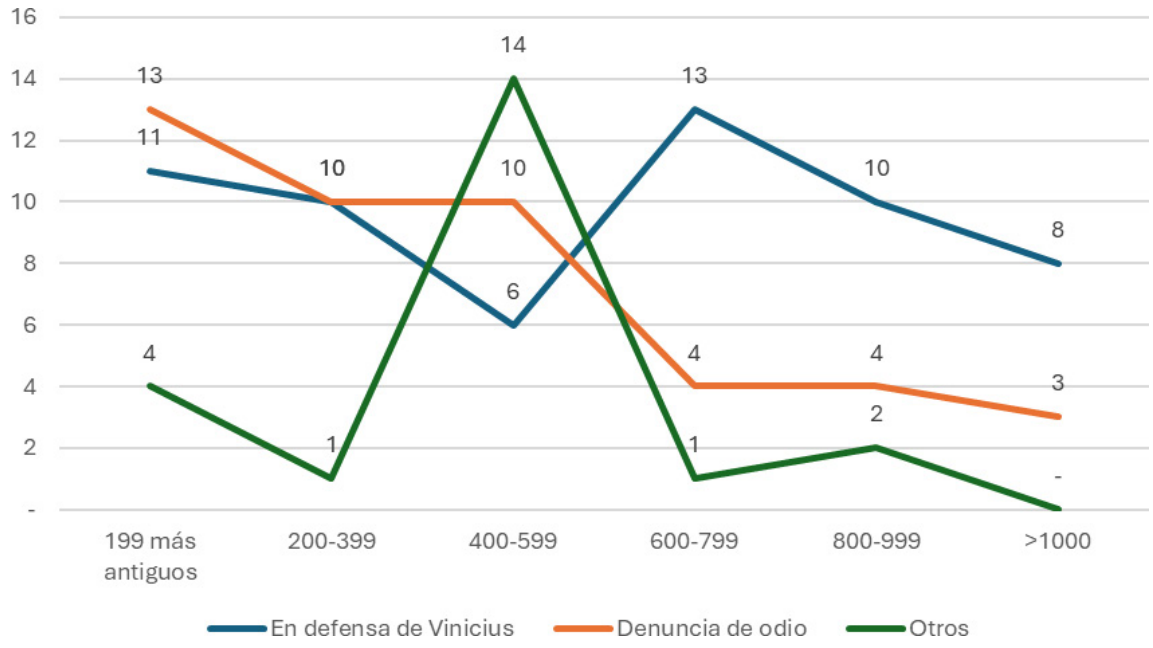
No obstante, se observa igualmente que, aunque un mensaje reitere una idea sobre la que se ha insistido en el foro, puede mover a la audiencia si apela a su consideración. Así el comentario 653 apareció después de cinco horas de conversación, pero se posicionó el 38 en la valoración de los usuarios: «El gran problema con Vinicius no es el color de su piel...», pero añade «y lo sabéis». Los participantes clican en *like* para ratificar que es así. El patrón se repite en otros ejemplos. Pero en cada encuadre funciona con eficacia un interlocutor diferente, es decir, el tema determina a quién dirigirse para buscar el apoyo del resto de usuarios: en *Se hace la víctima* predomina el impersonal («Cansino, está falta de vara de fresno», #716, 28 *likes*, 158 en valoración); y, en *Comparación con otros jugadores afrodescendientes,* funciona más el dialogar directamente con Vinicius («Que no Vini, que no...» #57, posición 15, 314 *likes*), porque en este encuadre los participantes en el foro dictan al jugador lo que debe hacer o cómo comportarse. Estas proposiciones que Wagemans (2016) denomina «de política» son apoyadas por los participantes en la conversación: de alguna forma son una consecuencia lógica de la idea principal en torno a la cual se desarrolla el *frame* («... Dedícate a jugar que lo haces maravillosamente bien y verás como se hablará de ti de otra forma», #57, puesto 15 con 314 *likes*).

4.2. El argumento dual o cómo bloquear el odio

Los comentarios que argumentan que Vinicius está siendo víctima de una campaña de odio se encuen-

tran en clara minoría respecto a los que expresan críticas a las declaraciones y al comportamiento del jugador. El clima emocional del foro provoca, primero, que cualquier comentario que apoye al jugador brasileño tenga un número significativamente menor de apoyos (el máximo son 30 *likes* frente a los miles de algunos comentarios críticos) y, segundo, que se produzca la desaparición progresiva de los comentarios positivos sobre Vinicius. El encuadre *Denuncia de odio*, que está compuesto por comentarios que no argumentan y que sólo afean a otros usuarios la forma de expresarse, se reduce a medida que avanza el tiempo de intervención en el foro: mientras que tienen un peso de un 30 % al principio, termina con apenas un 7 % en los últimos mensajes (Gráfico 2). Si se tiene en cuenta que el número de mensajes borrados es alto y frecuente, se puede deducir que hay motivos suficientes para que las denuncias a este tipo de comentarios sean relativamente constantes. Por lo tanto, si éstas no se mantienen, se debe a la ausencia de interés, no de motivación.

Gráfico 2. Desarrollo temporal de los *frames* a favor de Vinicius (número de comentarios).



Fuente: elaboración propia.

Los argumentos intentan desactivar dos de los principales enfoques: *Es un provocador* y *Comparación con otros jugadores afrodescendientes*. Los usuarios justifican por qué Vinicius actúa de forma agresiva en muchos partidos (presentación de hechos), aludiendo a las agresiones que ya sufría en otros equipos. También usan la analogía que se traslada al terreno personal: si en tu trabajo te trataran así, ¿cómo reaccionarías? (pregunta para provocar reflexión). O la expresión de experiencias propias para probar que el odio y el racismo están muy presentes en España. Se presenta así el odio como algo normal y no excepcional. El mensaje número 7 es el primero que apoya a Vinicius: no aparecerá otro hasta el número 29. El hecho de ser el primero que aparece defendiendo a Vinicius tiene un coste dentro de un foro emocionalmente predispuesto en contra de su contenido, como se ha señalado: es el que más rechazos provoca (45 *dislikes*).

Este mensaje, junto con los otros tres más vilipendiados en el foro de los que apoyan a Vinicius (#740, 33; #744, 32 y #854, 31), tienen una característica común: son agresivos contra los otros usuarios que no opinan igual. Ya sea de forma general, sin dirigirse a otros usuarios, «les ha escocido a todos» (#744) o «le insultan porque ha cerrado muchas bocas de un tiempo a esta parte» (#740), o bien juzgando a aquellos que no comparten su opinión: «Hacéroslo mirar lo del odio» (#7). Ninguno deja espacio a otras argumentaciones y cuestionan de forma belicosa otros puntos de vista. Sin embargo, si se analizan los mensajes en los que hay *argumentos duales*, que mezclan críticas y apoyo a Vinicius, la reacción de los usuarios es diferente. Los dos mensajes del grupo *En defensa de Vinicius*, que tienen mayor aceptación (20 *likes*), reciben un número de desaprobación muy distinto: uno cuenta con 32 rechazos, mientras que el otro se queda en poco más de la mitad: 18. Ambos están muy próximos en cuanto al momento de publicación: #744 y #752. El que registra las 32 valoraciones negativas (#744) sólo defiende a Vinicius, es breve (56 palabras) y su escritura es iracunda contra los que no opinan como él.

Por el contrario, el mensaje que suma +2 combinando aprobaciones y rechazos (20 vs. 18) tiene una extensión bastante mayor (186 palabras) y comienza de forma muy diferente, incorporando posibles errores propios y ajenos en la interpretación de las acciones de Vinicius: «Honestamente creo que este tema se ha salido de madre por parte de todos (...)». Después su argumentación es compleja y dual. Por un lado, refleja la injustica que se comete con Vinicius: otros jugadores en situaciones parecidas no sufrieron sus críticas. Por otro lado, justifica las reacciones airadas y desproporcionadas del jugador brasileño en el campo de juego: «Es un chaval de unos 20 años

que recibe todo tipo de acoso dentro y fuera del campo de parte de gente que en muchos casos le dobla la edad». Sin embargo, al mismo tiempo que defiende su figura, también se muestra crítico («Todos tenemos mucho que mejorar en nuestras actitudes, incluido él»), incluso se posiciona claramente contra él: «No es santo de mi devoción». Estas dos críticas ocupan menos espacio en el comentario y son más directas, pero se sitúan en la parte final. De hecho, el segundo mensaje con más diferencia entre los que están a favor y en los que están en contra (#175: +10) tiene una argumentación dual: incluye una crítica a Vinicius («algunas veces tiene actitudes no muy positivas»). También los dos mensajes del grupo *Denuncia de odio* que más diferencia presentan entre los *likes* y los *dislikes* (+14 y +13), y que son los únicos que están posicionados entre los 250 más valorados de esta categoría, se abren a este argumento dual, por un lado, reconociendo que Vinicius no es especial o diferente: «Vinicius es uno de los muchísimos futbolistas que sufren insultos» (#86) o subrayando sus errores: «Está claro que vinicius es víctima de racismo. Está claro que vinicius no es ejemplo de nada» (#77).

5. Discusión y conclusiones

A pesar del sistema de moderación de *Marca.com*, cuya actuación se manifiesta en los comentarios eliminados (32 %), se publican mensajes incívicos, ofensivos y con expresiones de odio. Se confirma por tanto la importancia de las contranarrativas para ofrecer otros pareceres (Howard, 2021) que puedan influir en los lectores y desafiar las narrativas negativas. Este estudio ha identificado algunas características de los mensajes publicados en una conversación *online* que pueden tenerse en cuenta a la hora de crear esos mensajes positivos.

En primer lugar, en respuesta a la P1, cabe señalar la importancia de la atmósfera emocional existente en el foro porque genera cohesión grupal. Los comentarios a favor de la tendencia mayoritaria, en este caso en contra del jugador, reciben más apoyo y ese apoyo favorece a su vez la creatividad en la forma y en el argumento de los mensajes: cuantos más puntos de vista y reflexiones sobre un mismo aspecto, más posibilidades hay de encontrar comentarios que sobresalgan por puro sentido estadístico. No sucede lo mismo con la manifestación de emociones individuales en los comentarios: su exposición no siempre proporciona los resultados esperados, en consonancia con las conclusiones de Lasser *et al.* (2023). De las emociones negativas localizadas (ira, enfado, aversión, vergüenza), la decepción obtiene mayor apoyo posiblemente porque se respalda con lo que hemos denominado un *argumento dual*: «Antes te apoyábamos» (concesión al jugador), «pero ahora has cambiado» (justificación de los ataques).

Las opiniones diferentes al sentir general que han tenido más apoyo (y menos rechazo) lo han conseguido gracias también gracias al *argumento dual:* presentar un reconocimiento a la forma de sentir y opinar de los otros para, al menos en parte, validar la idea contraria. Esta estrategia se relaciona con la conveniencia de mostrar empatía, señalada por Wright *et al.* (2017), pero aquí se concreta. En algún caso, el apoyo masivo a una de las partes provoca una rendición en la otra con el transcurso del tiempo: los comentarios de *Denuncia de Odio* pasaron de un 30 % al principio del hilo a apenas un 7 % al final, una reducción sensible. Igualmente, la secuencialidad determina también los apoyos: el arranque de la conversación favorece la presencia de más lectores. No obstante, caben excepciones, puesto que la incidencia de un mensaje también depende de otros aspectos.

En este sentido, cabe señalar que resultan más eficaces los mensajes breves, en torno a 30-40 palabras, como anticiparon Potash y Rumshisky (2017), especialmente si se construyen a modo de consigna, porque llaman la atención (P2). Pero en este estudio se comprueba que muchos de esos mensajes breves pueden pasar desapercibidos y que sólo incrementan su impacto si existe una argumentación que respalde el concepto principal que se pretende transmitir para evitar el *efecto de inconsistencia*: cuando no hay un argumento sólido o, al menos perceptible, detrás de una construcción formal seductora. Los insultos, por muy ingeniosos que sean, se quedan en la *impetuosidad del desahogo*: no calan y terminan agotándose ante la falta de nuevas propuestas. Pero también el exceso de argumentación resulta negativo porque produce, como se ha visto, el *efecto de saturación*: los usuarios no necesitan mucha insistencia para estar convencidos y les vale con una sola buena excusa argumental. La complejidad les abruma o les distrae. En otras palabras, más es menos.

La afiliación constituye otro elemento importante, puesto que otorga objetividad y veracidad a un mensaje: cuando la opinión de un autor contradice lo que se espera de él por su pertenencia a un determinado grupo (en este caso al equipo de fútbol de Vinicius o defender al jugador siendo de un equipo rival), ese comentario resulta más auténtico y, por lo tanto, genera más impacto. Otra táctica movilizadora, relativa a los autores, es solicitar directamente la opinión o el respaldo del resto de los usuarios, puesto que se genera una «respuesta a la llamada».

Respecto a las estrategias argumentales, se observa que las definidas por Mun *et al.* (2024) para combatir los estereotipos negativos, también se utilizan para crearlos: por ejemplo, poner en duda la veracidad del estereotipo (*Es un provocador*); exponer cualidades alternativas (*Se hace la víctima*, *Problemas psicológicos*...), explicar factores externos que conducen a esas ideas (*Actúa para su documental. Netflix*), entre otras tácticas. Aunque los datos también se pueden utilizar peor o mejor: si la comparación o el dato no se percibe como de interés inmediato y se aleja de lo que centra el interés, pierde conexión y eficacia. Aquí, el foro se centra en aspectos personales del jugador, no tanto en su raza, para evitar las acusaciones de racismo.

Se muestra que los mensajes que combinan proposiciones de hecho y de valoración logran un mayor respaldo y una mayor permanencia temporal del encuadre al que se adscriben en la conversación, que las que optan sólo por uno u otro. Ahora bien, los argumentos que recurren a exageraciones o afirmaciones absurdas o fuera del tema tratado no parecen recabar apoyos (*likes*), salvo si se utiliza la ironía porque sorprende al mostrar un conocimiento específico del hecho en sí y de lo que implica. En la

conversación analizada, las ironías tienen un punto argumentativo específico (Partington, 2007), pero también un juicio.

En definitiva, los elementos más estables para lograr un mayor impacto que esta investigación aporta son la tendencia emocional de la mayoría de los comentarios, que hace la diferencia en el número de apoyos que recaba un comentario respecto a otros comentarios que tengan similares características, pero vayan en contra del flujo emocional; la brevedad; no sólo en el número de palabras, sino también en el número de argumentos; el manejo graduado de la emoción, porque el simple desahogo sirve menos que si se acompaña de una justificación; el conocimiento profundo de los hechos laterales y de la verdad, aquello que se puede traer a la argumentación y que ha sucedido; la ironía, porque permite una exposición elegante y diferente de lo anterior, y cierta empatía, que hace que las opiniones de los otros sean vistas como valiosas, aunque no sean admitidas.

El estudio de este caso permite llevar a cabo un análisis en profundidad de cada uno de los comentarios publicados en este foro, pero también cuenta con limitaciones al generalizarse unos datos muy específicos, y más porque el foro analizado se refiere a un personaje público en un ámbito concreto (deportivo). Por tanto, las características específicas aportadas deberán cotejarse en nuevas investigaciones para su validación.

6. Financiación y apoyos

Estudio apoyado por los proyectos de investigación Cartografía de los Discursos de Odio en España desde la Comunicación (ref. PID2019-105613GB-C31) y Desafiar las narrativas online de odio político y misoginia (ref. PID2023-147506OB-I00), financiados por el Ministerio de Ciencia e Innovación (España).

7. Contribución de autores

Conceptualización	Ideas; formulación o evolución de los objetivos y metas generales de la investigación.	Autores 1, 2 y 3
Curación de datos	Actividades de gestión para anotar (producir metadatos), depurar datos y mantener los datos de la investigación (incluido el código de software, cuando sea necesario para interpretar los propios datos) para su uso inicial y su posterior reutilización.	Autor 3
Análisis formal	Aplicación de técnicas estadísticas, matemáticas, computacionales u otras técnicas formales para analizar o sintetizar datos de estudio.	Autores 2 y 3
Adquisición de fondos	Adquisición del apoyo financiero para el proyecto que conduce a esta publicación.	Autor 2
Investigación	Realización de una investigación y proceso de investigación, realizando específicamente los experimentos, o la recolección de datos/evidencia.	Autores 1 y 2
Metodología	Desarrollo o diseño de la metodología; creación de modelos.	Autores 2 y 3
Administración del proyecto	Responsabilidad de gestión y coordinación de la planificación y ejecución de la actividad de investigación.	Autor 1
Recursos	Suministro de materiales de estudio, reactivos, materiales, pacientes, muestras de laboratorio, animales, instrumentación, recursos informáticos u otras herramientas de análisis.	Autor 2
Software	Programación, desarrollo de software; diseño de programas informáticos; implementación del código informático y de los algoritmos de apoyo; prueba de los componentes de código existentes.	Autor 3
Supervisión	Responsabilidad de supervisión y liderazgo en la planificación y ejecución de actividades de investigación, incluyendo la tutoría externa al equipo central.	Autores 1 y 2
Validación	Verificación, ya sea como parte de la actividad o por separado, de la replicabilidad/reproducción general de los resultados/experimentos y otros productos de la investigación.	Autores 1 y 2
Visualización	Preparación, creación y/o presentación del trabajo publicado, específicamente la visualización/presentación de datos.	Autores 1, 2 y 3
Redacción / Borrador original	Preparación, creación y/o presentación del trabajo publicado, específicamente la redacción del borrador inicial (incluyendo la traducción sustantiva).	Autores 1 y 2
Redacción / Revisión y edición	Preparación, creación y/o presentación del trabajo publicado por los miembros del grupo de investigación original, específicamente revisión crítica, comentario o revisión, incluidas las etapas previas o posteriores a la publicación.	Autores 1 y 2

8. Referencias bibliográficas

AIMC (Asociación para la Investigación de Medios de Comunicación). (2024, Marzo). *26° Navegantes en la red.* https://www.aimc.es/a1mc-c0nt3nt/uploads/2024/03/Navegantes2023.pdf

Baider, F. (2023) Accountability Issues, Online Covert Hate Speech, and the Efficacy of Counter-Speech. *Politics and Governance*, *11*(2), 249-260. https://doi.org/10.17645/pag.v11i2.6465

Boberg**,** S., Schatto-Eckrodt, T., Frischlich, L. y Quandt, T. (2018). The Moral Gatekeeper? Moderation and Deletion of User-Generated Content in a Leading News Forum. *Media and Communication*, *6*(4), 58-69. https://doi.org/10.17645/mac.v6i4.1493

Bonaut, J., Vicent, M. y Paz, M.A. (2023). Sports journalist and readers. Journalism and user incivility (2023). *Journalism Practice*, *18*(2), 356–373. https://doi.org/10.1080/17512786.2023.2222730

Buerger, C. (2021). Counterspeech: A literature review. http://dx.doi.org/10.2139/ssrn.4066882

Cabeza, J., Casado, R. y Gómez, M. (2024). El Efecto Búmeran y los Discursos de Odio en los comentarios en prensa: Lionel Messi y el Independentismo catalán (2019-2021). *ICONO 14. Revista Científica de Comunicación y Tecnologías Emergentes*, *22*(1), e2057. https://doi.org/10.7195/ri14.v22i1.2057

Chung, Y. L., Kuzmenko, E., Tekiroglu, S. S. y Guerini, M. (2019). CONAN--COunter NArratives through nichesourcing: a multilingual dataset of responses to fight online hate speech. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2819-2829. https://aclanthology.org/P19-1271/

Chung, Y. L., Tekiroğlu, S. S., Tonelli, S. y Guerini, M. (2021). Empowering NGOs in countering online hate messages. *Online Social Networks and Media*, *24*, 1-19, 100150. https://doi.org/10.48550/arXiv.2107.02472

Drummond, C., O'Toole, T. y McGrath, H. (2020). Digital engagement strategies and tactics in social media marketing. *European Journal of Marketing*, *54*(6), 1247-1280. https://doi.org/10.1108/EJM-02-2019-0183

Fortuna, P. y Nunes, S. (2018). A Survey on Automatic Detection of Hate Speech in Text. *ACM Computing Surveys*, 51(4), 1-30. https://doi.org/10.1145/3232676

Friess, D., Ziegele, M. y Heinbach, D. (2021). Collective Civic Moderation for Deliberation? Exploring the Links between Citizens' Organized Engagement in Comment Sections and the Deliberative Quality of Online Discussions. *Political Communication*, *38*(5), 624-646. https://doi.org/10.1080/10584609.2020.1830322

Furman, D. A., Torres, P., Rodriguez, J. A., Martinez, L., Alemany, L. A., Letzen, D. y Martinez, M. V. (2022). Parsimonious Argument Annotations for Hate Speech Counter-narratives. https://arxiv.org/abs/2208.01099

Garland, J., Ghazi, K., Young, J. G., Hébert, L. y Galesic, M. (2022). Impact and dynamics of hate and counter speech online. *EPJ Data Science*, *11*(3). https://doi.org/10.1140/epjds/s13688-021-00314-6

Golder, S. A. y Macy, M. W. 2014. Digital Footprints: Opportunities and Challenges for Online Social Research. *Annual Review of Sociology*, *40*(1), 129-152. https://doi.org/10.1146/annurev-soc-071913-043145

Hansen, T. M., Lindekilde, L., Karg, S. T., Bang Petersen, M. y Rasmussen, S. H. R. (2024). Combatting online hate: Crowd moderation and the public goods problem. *Communications*, *49*(3), 444-467. https://doi.org/10.1515/commun-2023-0109

Howard, J. W. (2021). Terror, hate and the demands of counter-speech. *British Journal of Political Science, 51*(3), 924-939. https://doi.org/10.1017/S000712341900053X

Ihlebæk, K. A. y Holter, C. R. (2021). Hostile emotions: An exploratory study of far-right online commenters and their emotional connection to traditional and alternative news media. *Journalism*, *22*(5), 1207-1222. https://doi.org/10.1177/1464884920985726

Lasser, J., Herderich, A., Garland, J., Aroyehun, S. T., Garcia, D. y Galesic, M. (2023). Collective moderation of hate, toxicity, and extremity in online discussions. *ArXiv*, 1-53. https://arxiv.org/abs/2303.00357

Llansó, E., Van-Hoboken, J., Leerssen, P. y Harambam, J. (Febrero, 2020). Transatlantic Working Group. Content Moderation, and Freedom of Expression. https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf

Monakhova, T. y Tuluzakova, O. (2022). Hate Speech in Ukrainian Media Discourse. *Cognitive studies/| Études Cognitives*, 22. https://doi.org/10.11649/cs.2624

Mun, J., Allaway, E., Yerukola, A., Vianna, L., Leslie, S. J. y Sap, M. (2023, December). Beyond Denouncing Hate: Strategies for Countering Implied Biases and Stereotypes in Language. En *Findings of the Association for Computational Linguistics*: EMNLP 2023 (pp. 9759-9777). Association for Computational Linguistics. https://aclanthology.org/2023.findings-emnlp.653/

Nave, E. y Lane, L. (2023). Countering online hate speech: How does human rights due diligence impact terms of service? *Computer Law & Security Review*, *51*, 105884. https://doi.org/10.1016/j.clsr.2023.105884

Partington, A. (2007). Irony and reversal of evaluation. *Journal of pragmatics, 39*(9), 1547-1569. https://doi.org/10.1016/j.pragma.2007.04.009

Paz, M. A., Montero, J. y Moreno, A. (2020). Hate Speech. A Systematized Review. *SAGE Open*, *10*(4). https://doi.org/10.1177/2158244020973022.

Paz, M. A., Mayagoitia, A. y González, J. M. (2023). ¿Permite TikTok un debate de calidad? Un estudio de caso sobre la pobreza. *Communication & Society*, *36*(4), 83-97. https://doi.org/10.15581/003.36.4.83-97.

Potash, P. y Rumshisky, A. (2017). Towards Debate Automation: a Recurrent Model for Predicting Debate Winners. En *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 2465–2475). Association for Computational Linguistics. https://aclanthology.org/D17-1261/

Porter, E. y Wood, T. J. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *Proceedings of the National Academy of Sciences, 118*(37), e2104235118. https://doi.org/10.1073/pnas.2104235118

Rodríguez, A. (25 de marzo de 2024). Los 10 mensajes contra el racismo de Vinicius en su rueda de prensa más reivindicativa. *Marca*. https://www.marca.com/futbol/futbol-internacional/2024/03/25/66019e52ca4741da388b45f6.html

Saha, P., Singh, K., Kumar, A., Mathew, B. y Mukherjee, A. (2022). CounterGeDi: A controllable approach to generate polite, detoxified and emotional counterspeech. *ArXiv.* https://arxiv.org/abs/2205.04304

Saha, S. y Srihari, R. (2023). ArgU: A controllable factual argument generator. *ArXiv*. https://arxiv.org/abs/2305.05334

Svenja, B., Tim, S.-E., Lena, F. y Thorsten, Q. (2018). The moral gatekeeper? moderation and deletion of user-generated content in a leading news forum. *Media and Communication*, *6*(4), 58-69. https://doi.org/10.17645/mac.v6i4.1493

Volker Türk urge a poner fin al racismo en el deporte. (26 de mayo de 2023). *Naciones Unidas. Centro Regional de Información*. https://unric.org/es/volker-turk-urge-a-poner-fin-al-racismo-en-el-deporte/

Wagemans, J. (2016). Constructing a periodic table of arguments. En *Argumentation, objectivity, and bias: Proceedings of the 11th international conference of the Ontario Society for the Study of Argumentation* (OSSA) (pp. 1-12). http://scholar.uwindsor.ca/ossaarchive/OSSA11/papersandcommentaries/106

Walton, D., Reed, C. y Macagno, F. (2008). *Argumentation schemes*. Cambridge University Press.

Walton, D. (1996). *Argumentation schemes for presumptive reasoning*. Mahwah: Lawrence Erlbaum.

Wang, X., Koneru, S., Venkit, P. N., Frischmann, B. y Rajtmajer, S. (2024). The Unappreciated Role of Intent in Algorithmic Moderation of Social Media Content. *ArXiv.* https://arxiv.org/abs/2405.11030

Wright, L., Ruths, D., Dillon, K. P., Saleem, H. M. y Benesch, S. (2017). Vectors for counterspeech on twitter. En *Proceedings of the first workshop on abusive language online* (pp. 57-62). Association for Computational Linguistics. https://aclanthology.org/W17-3009/

Yu, X., Blanco, E. y Hong, L. (2024). Hate Cannot Drive out Hate: Forecasting Conversation Incivility following Replies to Hate Speech. En *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 18 (pp. 1740-1752).

José Cabeza San Deogracias. Profesor Titular de Guion Audiovisual en la Universidad Rey Juan Carlos de Madrid y guionista de la primera película producida por Netflix en España: *7 años* (Roger Gual, 2016). Ha publicado un libro sobre metodología docente y tiene editados varios más sobre narrativa cinematográfica. Sus estudios sobre análisis narrativos y formatos televisivos se han publicado en revistas nacionales e internacionales como *Latina*, *Comunicar*, *European Journal of Communication...* Además, ha colaborado en múltiples proyectos de investigación de ámbito estatal sobre la historia de la televisión y ha estudiado la evolución del *reality show* y el *talk show* en España. Actualmente, sus líneas de investigación son las narrativas en los discursos de odio y el estudio de la repercusión de la inteligencia artificial en la forma artística. ORCID: https://orcid.org/0000-0003-1047-2733

María Antonia Paz Rebollo. Catedrática de Periodismo en la Universidad Complutense de Madrid desde 2006. Creadora y codirectora del grupo de investigación «Historia y Estructura de la Comunicación y del Entretenimiento». Es autora de más de 130 publicaciones en editoriales y revistas de primer nivel y responsable de diversas actividades de transferencia (exposiciones, peritaje profesional, participación en comités de expertos nacionales e internaciones). Ha sido la investigadora principal en cuatro proyectos competitivos nacionales, los dos últimos (ref. PID2019-105613GB-C31 y ref. PID2023-147506OB-I00) dedicados al análisis de los discursos de odio desde la comunicación: su presencia en la prensa digital y en las redes sociales, sus principales narrativas en el ámbito del odio político y misoginia y posibles estrategias para desafiarlas. Tiene 6 sexenios de investigación. ORCID: https://orcid.org/0000-0002-6664-0647

Raúl Casado Linares. Profesor Contratado en la *Amsterdam University of Applied Sciences*, donde imparte una variedad de cursos empresariales para su escuela de negocios. Sus investigaciones actuales se centran en el uso innovador de herramientas de inteligencia artificial en el ámbito empresarial, incluyendo los últimos avances en IA Generativa, y su impacto transformador en la educación universitaria como parte de un proyecto Erasmus+. Ha publicado en revistas españolas e internacionales como *Historia y comunicación social*, *Revista de Comunicación*, *Icono14*... ORCID: https://orcid.org/0000-0002-6121-8279