ICONO14

Universidad Rey Juan Carlos

Sesgo mediático y discurso electoral: una aproximación con Procesamiento del Lenguaje Natural

Media bias and electoral discourse: a Natural Language Processing approach

Preconceitos dos media e discurso eleitoral: uma abordagem de processamento da linguagem natural

Mar Castillo-Campos[1*]

David Becerra-Alonso[2**]

David Varona-Aramburu[3***]

[1] Universidad Loyola Andalucía, España
[2] Universidad Loyola Andalucía, España
[3] Universidad Complutense Madrid, España

[*] Ayudante de investigación en el departamento de Comunicación y Educación. Universidad Loyola Andalucía, España. Email: mcastillo@uloyola.es
[**] Profesor titular en el departamento de Métodos Cuantitativos. Universidad Loyola Andalucía, España. Email: dbecerra@uloyola.es
[***] Profesor Ayudante Doctor en el departamento de Periodismo y Nuevos Medios. Universidad Complutense Madrid, España. Email: davarona@ucm.es

Esta obra está bajo una Licencia Creative Commons Atribución 4.0 Internacional

Recibido:16/02/2024; **Revisado:** 21/02/2024; **Aceptado:** 18/01/2025; **Publicado:** 08/07/2025

Para citar este artículo: Castillo-Campos, Mar; Becerra-Alonso, David; & Varona-Aramburu, David. (2025). Sesgo mediático y discurso electoral: una aproximación con Procesamiento del Lenguaje Natural. *ICONO 14. Revista Científica de Comunicación y Tecnologías Emergentes*, *23*(1): e2154. https://doi.org/10.7195/ri14.v23i1.2154

Resumen

Los medios de comunicación deben ser un canal fiable de información, especialmente en el contexto de sobrecarga informativa y de la presencia de los perfiles y canales privados de las figuras públicas. En una sociedad cada vez más polarizada y desconfiada hacia los medios, este estudio pretende analizar la cobertura mediática de una campaña electoral en España, la primera tras la pandemia de COVID-19. Utilizando TF, TF*IDF y word2vec para la cuantificación y vectorización del texto, junto con UMAP y t-SNE para el análisis de conglomerados, examinamos cómo se utilizan determinados términos en los medios de comunicación y sus asociaciones semánticas. Nuestros resultados revelan una tendencia de los medios a vincular determinados candidatos o partidos con los extremos políticos, la violencia y la negatividad, a menudo eclipsando el contenido del discurso político sustantivo. La cobertura se centra mayoritariamente en los grandes partidos y las facciones polarizadoras. Los actos de campaña reciben más atención que las propuestas políticas, que a menudo pasan desapercibidas. Estos datos coinciden con los de estudios cualitativos anteriores, lo que respalda la eficacia de este enfoque cuantitativo al ampliar el tamaño de la muestra, reducir el tiempo de análisis y revelar patrones sutiles que no se aprecian fácilmente con las metodologías tradicionales. Este estudio contribuye a una comprensión más profunda de la dinámica de los medios de comunicación durante los períodos electorales y subraya el valor de los métodos cuantitativos en el análisis de los medios de comunicación.

Palabras clave
Cobertura electoral; Análisis del discurso; Procesamiento del lenguaje natural; Comunicación computacional; Política.

Abstract

This study employs quantitative and artificial intelligence methods to scrutinize media coverage during an election campaign. Employing TF, TF*IDF, and word2vec for text quantification and vectorization, alongside UAMP and t-SNE for cluster analysis, we examine how certain terms are utilized across media outlets and their semantic associations. Our findings reveal a tendency for media to link certain candidates or parties with political extremes, violence, and negativity, often overshadowing substantive political discourse. Notably, coverage predominantly focuses on major parties and polarizing factions. Campaign events receive more attention than policy proposals, which are often neglected. These insights align with prior qualitative studies, demonstrating the efficacy of our quantitative approach in expanding sample size, reducing analysis time, and revealing nuanced patterns not readily apparent through traditional methodologies. This study contributes to a deeper understanding of media dynamics during election cycles and underscores the value of quantitative methods in media analysis.

Keywords
Elections coverage; Discourse analysis; Natural language processing; NLP; Computational communication; Politics.

Resumo

Este estudo emprega métodos quantitativos e de inteligência artificial para examinar a cobertura da mídia durante uma campanha eleitoral. Utilizando TF, TF*IDF e word2vec para quantificação e vetorização de textos, juntamente com UAMP e t-SNE para análise de clusters, analisamos como certos termos são utilizados em diferentes meios de comunicação e suas associações semânticas. Nossos resultados

revelam uma tendência dos meios de comunicação a associar certos candidatos ou partidos a extremos políticos, violência e negatividade, frequentemente ofuscando o discurso político substantivo. Notavelmente, a cobertura se concentra predominantemente nos principais partidos e nas facções polarizadoras. Os eventos da campanha recebem mais atenção do que as propostas políticas, que muitas vezes são negligenciadas. Esses achados estão alinhados com estudos qualitativos anteriores, demonstrando a eficácia de nossa abordagem quantitativa em expandir o tamanho da amostra, reduzir o tempo de análise e revelar padrões sutis que não são facilmente perceptíveis por métodos tradicionais. Este estudo contribui para uma compreensão mais profunda das dinâmicas midiáticas durante os ciclos eleitorais e ressalta o valor dos métodos quantitativos na análise da mídia.

Palavras-chave
Cobertura eleitoral; Análise do discurso; Processamento de linguagem natural; Comunicação computacional; Política.

1. Introducción

Los periodistas tienen la labor de buscar y contar lo que los ciudadanos directamente no pueden experienciar. El periodismo tiene la posibilidad y la responsabilidad de encontrar los datos, interpretarlos, proveerlos de contexto y decidir contarlos o no.

En el ámbito político, la labor periodística ha sido clave para el acercamiento de los partidos y los candidatos a los ciudadanos y para «una mayor visualización y transparencia del proceso de toma de decisiones políticas» (Holgado, 2017). Para los ciudadanos, es la principal forma de conocer a los candidatos y sus propuestas sin que los políticos las presenten directamente. Para los políticos, sirve de plataforma para ganar visibilidad, relevancia y pedir el voto. De hecho, solo nombrar a determinados partidos otorga una ventaja de posicionamiento en el imaginario público (Haselmayer et al., 2017). El papel de los medios de comunicación es mediar y conectar a ambos actores —ciudadanos y políticos— facilitando el entendimiento mutuo. Idealmente, esto permite a los partidos proponer medidas que interesen a los ciudadanos y, a estos, tomar decisiones de voto informadas.

Sin embargo, ese acercamiento no siempre está basado en las propuestas políticas de los partidos, sino, a menudo, en las acciones que estos protagonizan con objeto de ser visibilizados e identificados. Esto contribuye al malestar social, ya que los políticos en campaña electoral tienden a «descalificar al contrario», más «que explicar un programa de medidas o de acciones» (Fernández-Fernández, 2009). Hace décadas, Arredondo-Ramírez (2000) describía las elecciones como una «lucha intensa por ocupar mayor espacio público; por ser identificado, recordado, señalado y, en última instancia, elegido». La presencia y las campañas de los candidatos en las redes sociales también son de interés, ya que a menudo son replicadas por los medios de comunicación (Bright et al., 2020; Green y Gerber, 2019). Así pues, como sostienen Thurber y Nelson (2018), «las campañas electorales importan». Para Holgado (2017):

> La presencia en los medios (…) ha constituido la más importante de las formas que tienen los partidos políticos de dar a conocer su mensaje y solicitar el voto durante la campaña electoral, frente a otras más clásicas como la tradicional pegada de carteles o la celebración de mítines. De hecho, estas últimas se siguen llevando a cabo, pero con una finalidad mediática, es decir, están concebidas para ser televisadas, o recogidas en imágenes o titulares de prensa.

Las campañas electorales reciben gran cobertura mediática y son una etapa de gran competencia entre los periódicos (Arredondo-Ramírez, 2000; Denton et al., 2019). Según Castells (2009), las personas y acontecimientos políticos se presentan con narraciones poco propias de la actividad informativa, optando por un lenguaje más característico del infoentretenimiento y similar a historias de intriga, violencia o sexo. Aunque no está regulado explícitamente en el periodismo español, el uso de un lenguaje violento y el fomento de la crispación contradicen las prácticas éticas (Fernández-Fernández, 2009). Este estudio no profundiza en el impacto del lenguaje en los individuos, pues ya otros autores han probado su influencia (Van Duyn y Collier, 2019; Veres, 2006, entre otros). Además, los medios de comunicación suelen amplificar el debate entre los actores políticos, compitiendo por la atención de los titulares (Barandiaran et al., 2020). Este debate, a menudo hostil, también beneficia a los políticos, ya que «los candidatos que utilizan un tono más negativo y los que hacen un mayor uso de apelaciones emocionales reciben una mayor cobertura mediática» (Maier y Nai, 2020). Sabedores de este

efecto, los partidos participan en diversos foros públicos e influyen en las narrativas mediáticas (López-García, 2017). Otro efecto de la *mediatización,* un proceso recíproco entre la política y los medios (Hjarvard, 2004; Deacon y Stanyer, 2014) es que los partidos aprovechan la popularidad de sus líderes más carismáticos para ganar exposición mediática (Casero-Ripollés et al., 2016).

Del mismo modo que los políticos adoptan un tono más negativo o de confrontación, la prensa se centra en las elecciones en términos de victorias o derrotas, enfatiza los conflictos o los elementos emocionales (Nelson y Thurber, 2018; Maier y Nai, 2020). Todo ello repercute en la percepción que los individuos tienen de los políticos en vísperas de las elecciones (Antoniades, 2020; Casero-Ripollés et al., 2016) y configura las actitudes de la sociedad hacia los procesos electorales (Goidel et al., 2021).

No solo en este sentido, sino también en la otra dirección: los políticos influyen en la percepción que la población tiene de los medios de comunicación. Las críticas a los medios por parte de figuras políticas promueven la percepción de sesgo entre el público (Fawzi, 2019; Smith, 2010), lo que a su vez contribuye al desencanto con el Gobierno (Hutchens et al., 2016). No se trata de un efecto marginal: según una encuesta de Reuters (2023), los políticos son los críticos más frecuentes con los medios de comunicación, pues aglutinan el 42% de las críticas a medios y periodistas.

Se observa que son esferas interconectadas: la información y el lenguaje negativo influyen en la polarización, la polarización política socava aún más la confianza pública en los medios de comunicación y fomenta el discurso del odio (Reporteros sin Fronteras, 2020). Un estudio sobre la confianza del público norteamericano en los medios de comunicación (Jones, 2004) concluyó que la falta de información sobre las diferencias políticas y los programas electorales repercute negativamente en la percepción que la sociedad tiene de la prensa. Pero no solo eso, ya que la retórica vacía también hace que los ciudadanos vean el discurso político como frustrante, divisivo y algo a evitar (Duggan y Smith, 2016). En España, contexto en el que se celebran estas elecciones, la confianza de los ciudadanos en los medios de comunicación es menor en comparación con Europa, en un descenso continuado (Eurobarómetro, 2021; Asociación de la Prensa de Madrid, 2020).

1.1 Partidos políticos en España

En España, a pesar de los más de 4.900 partidos políticos registrados en 2022, solo una docena recibe cobertura mediática, lo que pone de manifiesto la persistente oligarquía dentro del llamado sistema multipartidista. La Ley Orgánica 5/1985 determina el espacio publicitario de los partidos en los medios públicos en función de sus resultados electorales anteriores, lo que contribuye a la visibilidad reducida de los partidos nuevos o más pequeños. Aunque los medios privados están sujetos a ciertas normas, gozan de gran autonomía a la hora de decidir a qué partidos dan voz. Este estudio subraya la importancia de los medios de comunicación a la hora de proporcionar una información política completa y equilibrada, y de emplear y reflejar la autonomía de sus decisiones en el marco de las democracias occidentales.

2. Estado del arte

Nos preceden investigaciones acerca de la cobertura mediática en períodos electorales, si bien no se han desarrollado con metodologías cuantitativas e incorporado inteligencia artificial al análisis, o no se plantean con un enfoque comparativo entre diversos medios y distintos candidatos (Conroy et al., 2015; Córdoba-Cabús, 2018; De Vreese et al., 2006; Ergün y Karsten, 2021; Gallardo-Paúls y Enguix-Oliver, 2015; Holtz-Bacha et al., 2014; Mancera-Rueda y Villar-Hernández, 2020; Mazaira-Castro et al., 2019; Paniagua-Rojano et al., 2020; Sánchez-Gutiérrez y Nogales-Bocio, 2018; Smith, 1997; Terkildsen y Damore, 1999, Vizoso y López-García, 2020).

Someter los datos a redes neuronales y aprendizaje automático, como plantea esta investigación, es reciente en el campo del periodismo. La mayoría de los estudios previos han implementado metodologías estrictamente cualitativas (Berrocal-Gonzalo et al., 2017; Fenoll y Rodríguez-Ballesteros, 2017; Mancera-Rueda y Villar-Hernández, 2020; Miguel-Sáez-de-Urabain et al., 2017; Sánchez-Gutiérrez y Nogales-Bocio, 2018; Van der Pas y Aaldering, 2020). Aquellos que han apostado por la vectorización de los textos lo han hecho en su mayoría con

el fin de agrupar y/o clasificar las noticias (Orden-Cruz et al., 2019; Riedel et al., 2017; Zhao y Chang, 2020), y no para un análisis de relaciones entre términos y asociaciones, como aquí se presenta. Algunos autores han utilizado técnicas similares para aprendizaje supervisado (Berven et al., 2020; Edell, 2018; Nicholls y Culpepper, 2021; Zhou et al., 2019).

Hay numerosos estudios que examinan la cobertura electoral en España. Berrocal-Gonzalo et al. (2017) utilizan *support vector machines* para analizar el conflicto en los debates electorales. Fenoll y Rodríguez-Ballesteros (2017) correlacionan la intensidad mediática con la posición de los candidatos en las elecciones generales de 2015. Mancera-Rueda y Villar-Hernández (2020) se centran en la comparación mediática del partido Vox. Sánchez-Gutiérrez y Nogales Bocio (2018) analizan Podemos desde una perspectiva crítica del discurso. Miguel-Sáez-de-Urabain et al. (2017) lo hacen con la cobertura de El País de las elecciones estadounidenses de 2016.

Por otro lado, el estudio *Anatomía del bulo electoral: la desinformación política durante la campaña del 28-A en España* (Paniagua-Rojano et al., 2020) analiza el contenido y la difusión de los bulos electorales identificados por los verificadores de noticias, al igual que Mazaira-Castro et al. (2019) y Vizoso y López-García (2020). Otros investigadores (López-Meri, 2017; Campos-Domínguez y Calvo, 2017; Chaves-Montero et al., 2017; Ballesteros-Herencia, 2020) se centran en la cobertura electoral en redes sociales que, aunque son plataformas en indudable crecimiento, siguen sin ser la principal fuente de información en periodos de campaña electoral (suponen el 67 % frente al 93 % de las elecciones de los ciudadanos, que prefieren informarse a través de prensa, televisión o radio, en sus versiones digitales o no). Los últimos estudios encontrados sobre confianza comparada indican que los medios de comunicación tradicionales, incluida su versión digital, siguen siendo los más fiables para la población (Kalogeropoulos et al., 2019), y fueron especialmente consultados durante la pandemia (Li y Sun, 2021), contexto en el que se celebran estas elecciones. Hemos encontrado artículos sobre las publicaciones realizadas por los políticos en sus propios perfiles de las redes sociales durante la campaña (por ejemplo, Diez-Gracia et al., 2023) que, aunque potencialmente complementarios, quedan fuera del ámbito de este estudio, centrado en el análisis de los medios de comunicación.

3. Propuesta

Este estudio examina la cobertura mediática de la primera campaña electoral española tras la COVID-19, y se centra en el tratamiento equitativo de partidos y candidatos y no en la clasificación de las noticias. Se derivan los siguientes objetivos:

- Analizar las diferencias de visibilidad entre partidos políticos, candidatos y temas en la cobertura mediática.
- Examinar cómo los medios de comunicación varían en su cobertura de los candidatos a la presidencia, particularmente en relación con sus programas electorales o su implicación en polémicas.

4. Muestra

Para el diseño de este estudio se ha optado por utilizar los *Really Simple Syndication* (RSS) de los medios de comunicación en abierto más leídos en España, según los informes de Reuters Institute y la Universidad de Oxford (Newman et al., 2020) y la Asociación para la Investigación de Medios de Comunicación (AIMC, 2020). Los medios que permiten la extracción mediante RSS de las noticias en tiempo real son *ABC, eldiario.es* y *El País*, los tres calificados de riesgo mínimo de desinformación (Magallón et al., 2021). Los medios *20 minutos* y *El Mundo* pertenecerían a la muestra deseada, pero no disponen de RSS abierto y no permiten la extracción automática de los textos. La muestra para este estudio son todas las noticias publicadas por estos tres medios que hacen referencia a las elecciones a la Comunidad de Madrid durante las dos semanas de campaña electoral, del 18 de abril al 2 de mayo de 2021.

5. Metodología

Como metodología cuantitativa se elige Procesamiento de Lenguaje Natural (PLN) apoyado en aprendizaje automático. Se utilizan tres métodos diferentes: TF, TF*IDF

y vectorización y cálculo mediante redes neuronales de distancias euclidianas y similitud coseno entre términos.

Term Frequency o «frecuencia de un término» es un conteo de palabras en el texto, tanto absolutas como relativas respecto al número total de palabras en el mismo documento. A partir de ella se calcula el número potencial de veces que un lector podría haber estado expuesto a diferentes términos, ideas o candidatos, conocida la posible correlación entre la intensidad de la cobertura de cada candidato y su posición en los sondeos preelectorales (Fenoll y Rodríguez-Ballesteros, 2017). Es habitual como métrica de partida en algunos estudios (Aljarah et al., 2020; Høyer and Nossen, 2015; Terachi et al., 2006; Yamamoto y Church, 2001) de análisis de medios.

TF*IDF es una métrica que permite localizar el medio que más importancia otorga a un término concreto. Se formula como un factor entre la frecuencia relativa de una palabra o conjunto de palabras en un texto y la comparación de toda una selección de textos que contienen ese mismo término. Es habitual el cálculo de TF*IDF para localizar *keywords* para un posterior análisis cualitativo (Köffer et al., 2018; Koloski et al., 2021; Xu et al., 2018) o para encontrar diferencias relativas a términos empleados por distintas fuentes (Ghosh et al., 2020). En esta investigación, se emplea TF*IDF para detectar qué medios han otorgado más importancia a determinados términos, ya sea a los candidatos, a las medidas políticas o a acontecimientos concretos que han sucedido durante la campaña.

En la fase final de la investigación, se utiliza el modelo word2vec (Mikolov et al., 2013) para transformar cada palabra en un vector multidimensional, permitiendo medir la probabilidad de que un término determinado aparezca junto a otro. Esta probabilidad se determina por la distancia entre términos dentro de una muestra y evalúa, para cada medio de comunicación, el grado de asociación entre dos conceptos. word2vec genera representaciones vectoriales de modo que las palabras similares o relacionadas se posicionan cerca unas de otras. Por ejemplo, partidos políticos como *PSOE, PP* y *Unidas Podemos (UP)* se ubican en la misma zona del espacio multidimensional, mientras que términos como *coronavirus, pandemia* y *COVID* se sitúan más alejados. Smirnova et al. (2021) utilizaron word2vec para identificar áreas temáticas de información y extraer palabras clave. Esta metodología también se ha aplicado en otros estudios de medios, incluyendo comentarios de usuarios en portales digitales y en experimentos de aprendizaje supervisado

(Budiman et al., 2019; Köffer et al., 2018). Aunque su aplicación académica es más frecuente en el análisis de redes sociales —caracterizadas por textos más cortos y menos elaborados— o en artículos científicos (Mustafa et al., 2021; Terachi et al., 2006), no es comúnmente empleada para el análisis de medios y con frecuencia no produce resultados concluyentes (Al-Omari et al., 2019). En esta investigación, la selección de términos se construye a partir de (a) las palabras más frecuentes para cada medio de comunicación (obtenidas mediante TF), (b) las palabras con el valor TF*IDF más alto para cada medio de comunicación, (c) las palabras que hacen referencia a propuestas políticas basadas en los programas electorales y (d) las palabras clave de los eslóganes de cada candidato para su propia campaña electoral.

Para visualizar de manera efectiva los datos de alta dimensionalidad generados a partir del análisis de texto, nos apoyamos en técnicas de reducción dimensional como t-SNE y UMAP. Estos métodos nos ayudan a transformar los vectores complejos y multidimensionales en un espacio bidimensional donde los patrones y las relaciones son más fácilmente distinguibles.

t-SNE (Van der Maaten y Hinton, 2008) es una técnica no lineal de reducción de dimensionalidad que minimiza la divergencia Kullback-Leibler entre las probabilidades conjuntas de los datos de alta dimensionalidad y el *embedding* de baja dimensionalidad. Preserva la estructura local y crea agrupaciones que se pueden interpretar visualmente.

UMAP (McInnes et al., 2018) construye una representación gráfica de alta dimensión de los datos y optimiza un gráfico de baja dimensión para que sea estructuralmente lo más similar posible. Este método es ventajoso para mantener la estructura local y global de los datos. La visualización realizada con t-SNE y UMAP es útil para identificar clusters, patrones y anomalías dentro de los datos. Todos estos experimentos se realizaron en Visual Studio Code utilizando Python.

6. Resultados

De los 23 partidos concurrentes a las elecciones de la Comunidad de Madrid, 18 se mencionan al menos una vez. La mayoría, sin embargo, no se nombran más que en una ocasión, pues prácticamente toda la información se centra en los seis partidos principales. ABC es el medio analizado que más partidos nombra (18), muy

diferenciado de eldiario.es (8) y El País (7). Ninguno de estos medios menciona durante la campaña electoral a las agrupaciones concurrentes Verdemunicipalistas, Unidos-Coalición Verde, Partido Demócrata Social Jubilados (PDSJE), Unión Europea de Pensionistas o Por un Mundo más Justo (PUM+J).

El partido político más nombrado es Vox (1.129 menciones en tres medios de comunicación en dos semanas), seguido de PP (960) y Unidas Podemos (730). Estos datos absolutos corresponden al número de veces que los lectores habrían recibido la mención de una agrupación política frente a otra que, siendo desconocida, no será una posibilidad de voto.

6.1 Análisis de titulares

Inicialmente, analizamos los titulares, que son clave por varias razones (Kalogeropoulos et al., 2018):

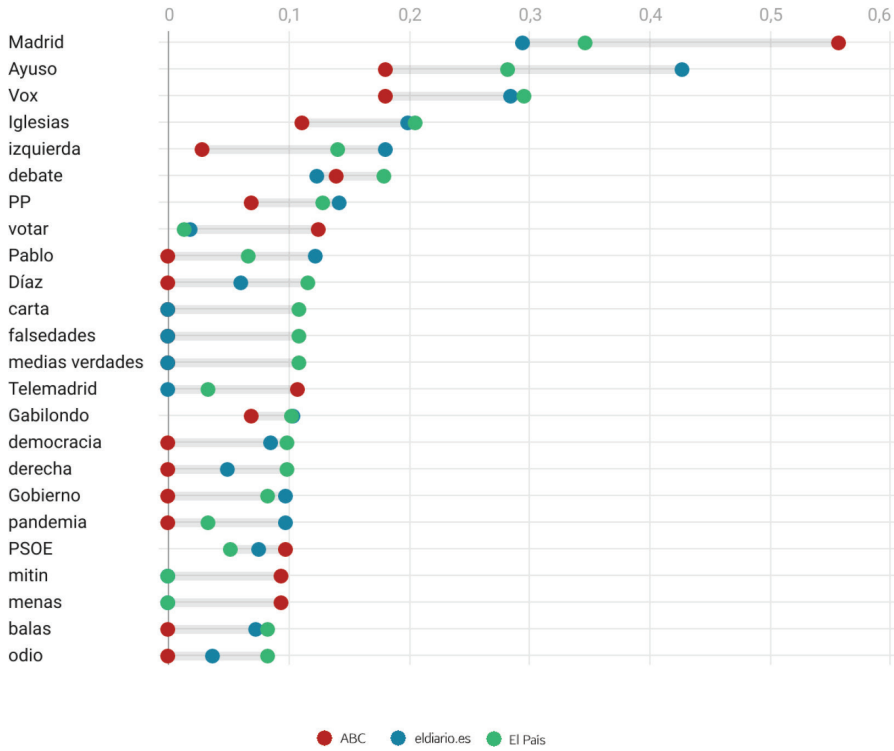
- El rápido consumo de información lleva a muchos lectores a solo leer los titulares. En España, es el 57,8 % (Castillo, 2017). Los medios tienden a elaborar titulares llamativos, a veces exagerados (Rieis et al., 2015). Y lo que es más importante, para Ecker et al. (2014), los titulares no son solo la primera impresión de los artículos de noticias, sino que pueden influir en la forma en que los usuarios perciben el artículo completo.
- Los titulares suelen ser el único texto que acompaña a las publicaciones en las redes sociales, donde la exposición de los contenidos se controla mediante algoritmos (Kalogeropoulos et al., 2018). Durante las elecciones, los medios de comunicación son especialmente activos en plataformas como Twitter (Paniagua-Rojano et al., 2020), lo que amplifica la importancia del contenido de los titulares.

Por ello, merece la pena destacar algunos datos que pueden ser de interés. Para las tres cabeceras analizadas, las palabras más repetidas son *Ayuso* (candidata del PP) y *Vox* (partido político), seguidas a distancia por *Iglesias* (candidato de UP) y *Gabilondo* o *PSOE* (candidato y partido). La representación de Ciudadanos y Más Madrid es incluso residual. Entre los términos más empleados destacan aquellos que hacen referencia a sucesos durante la campaña (*debate, balas, mitin*), y muy

pocos relacionados con propuestas de gobierno (*pandemia, educación, sanidad*). Es notable el empleo en titulares de términos relativos al conflicto: *odio, falsedades, amenazas*...

En un análisis de densidad comparada [Figura 1], esto es, evaluando la relevancia que tiene una palabra en un medio respecto a los demás medios, el índice TF*IDF para eldiario.es es 0.4269 para la palabra *Ayuso*, lo que implica que el término es más relevante en los titulares publicados por eldiario.es que en los publicados por otros periódicos. Para El País el índice máximo de TF*IDF se observa para el término *Vox* (0.2954).

Figura 1. *Palabras más relevantes en titulares. Resultados de TF*IDF de las palabras empleadas en titulares para cada medio de comunicación*



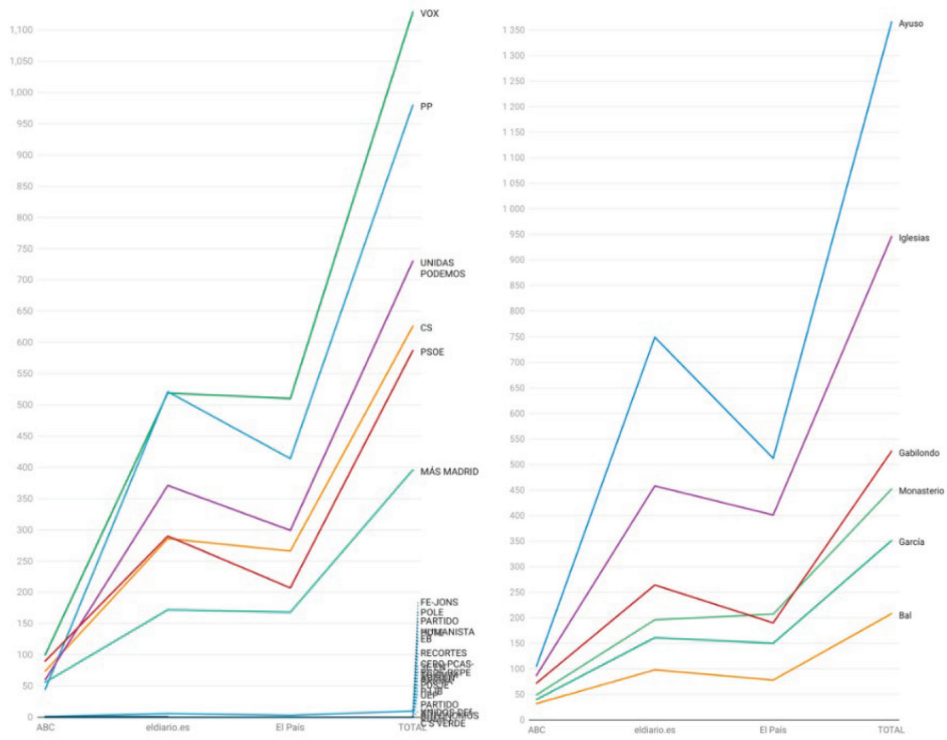
Fuente: *elaboración propia.*

6.2 Análisis de cuerpos de noticias

En este estudio, se observa que Isabel Díaz Ayuso y Pablo Iglesias personifican a sus respectivas agrupaciones políticas, PP y UP, más de lo que se menciona a los partidos. Por el contrario, Vox es mencionado con más frecuencia que su candidata, Rocío Monasterio. Pese a ser cuarto en la Asamblea de la Comunidad (quinto en la configuración de la Asamblea en 2015, antes de estas elecciones), Vox ocupa el primer lugar en menciones [Figura 2].

La aplicación de técnicas de TF*IDF ha permitido detectar algunos nombres que solo aparecen en determinados medios. Es habitual que los periodistas introduzcan una cita de Santiago Abascal (presidente del partido Vox) para reforzar o contrastar

Figura 2. *Comparación entre el número absoluto de menciones de cada partido (izquierda) y de cada candidato (derecha)*



Fuente: *elaboración propia.*

una mención de la candidata Rocío Monasterio. eldiario.es y El País, además de mencionar a Ayuso, también se refieren a Pablo Casado (presidente del partido) y a José María Aznar (expresidente del partido y expresidente del Gobierno). ABC no lo menciona en ninguna ocasión, lo que resulta en un índice TF*IDF cero [Figura 3]. Además, cuando eldiario.es o El País se refieren a Edmundo Bal (candidato por Ciudadanos), suelen mencionar a Inés Arrimadas, su homóloga en el Congreso. Por el contrario, ABC menciona con más frecuencia a Albert Rivera (expresidente de Ciudadanos, ya retirado de la política) que a Arrimadas.

Al emplear word2vec, se observa una cierta polarización: los términos con connotaciones negativas y violentas se asocian a los partidos más alejados ideológicamente, es decir, Vox y UP. Coincidentemente, estos dos partidos son los que más representación mediática reciben.

Vox y su candidato principal están fuertemente vinculados a términos como *afear, negar, dudar* y *muerte*. UP y Pablo Iglesias se vinculan estrechamente con *amenaza, duda, abandono* y *muerte*. Isabel Díaz Ayuso, la candidata más nombrada, tiene una alta probabilidad de aparecer junto a los términos *presidenta, reelección* y *victoria*. Para cada partido o candidato, ninguna de las palabras más fuertemente vinculadas se relaciona directamente con propuestas electorales.

A partir del cálculo de la distancia entre los candidatos y las propuestas, existe una mayor relación entre todos los políticos y los términos *fiscalidad, turismo, empleo* y *transparencia*, y una relación más débil con *educación*. La excepción se

Figura 3. *Densidad de menciones por cada candidato. TF*IDF aplicado a los cuerpos de las noticias para ABC, eldiario.es y El País*

TF*IDF bodies	Ayuso	Casado	Aznar	Bal	Arrimadas	Rivera
ABC	0,112	0,003	**0,000**	0,030	**0,000**	0,003
ELD	0,230	0,059	**0,010**	0,022	**0,022**	0,013
ELP	0,238	0,024	**0,001**	0,019	**0,007**	0,004

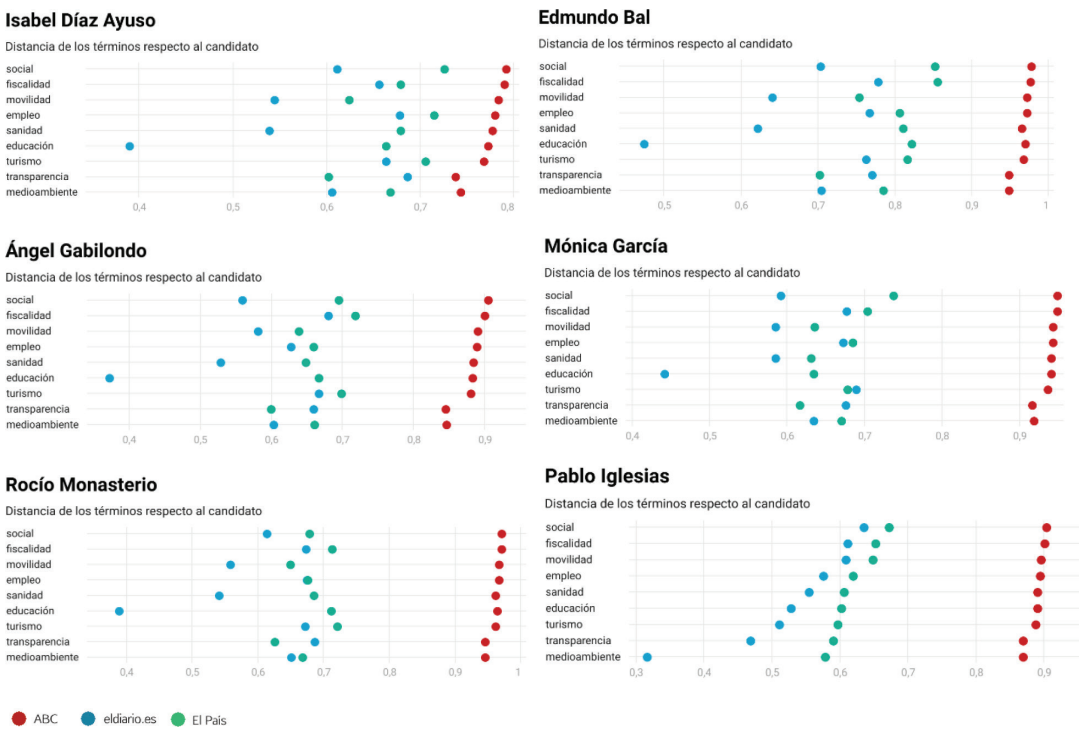
Fuente: *elaboración propia.*

da con Iglesias (UP), que tiene una mayor similitud con las propuestas sociales, y mínima con *medioambiente* [Figura 4].

Como se puede medir con TF*IDF, en la primera semana de campaña electoral (del 18 al 25 de abril), Ayuso es ya la palabra más relevante para los tres medios analizados, con un mayor índice para eldiario.es (0.265) sobre los demás (0.235 en El País y 0.148 en ABC). Le sigue *Vox*, más relevante en El País, e *Iglesias*, también en El País. ABC tiene mayor puntuación con *Gabilondo*.

En la segunda semana, del 26 de abril al 2 de mayo, *Vox* es más relevante (con mayor índice en El País) que *Ayuso*, seguido de *PP* e *Iglesias*. ABC hace un mayor uso de los términos *elecciones, electoral, voto,* y *votación*. El medio eldiario.es mantiene como más relevantes las agrupaciones de derecha: *Ayuso, PP* y *Vox*, seguidos de *UP, Iglesias* y *Díaz*. El País sigue un patrón prácticamente idéntico al de eldiario.es, aunque con *Vox* a la cabeza.

Figura 4. *Distancia entre los términos de propuestas políticas y cada candidato*



Fuente: *elaboración propia.*

Estos resultados, que no desvelan demasiada disparidad, sugieren que la cobertura de los términos y conceptos son similares. Aunque ciertos términos aparecen exclusivamente en algunos medios y no en otros (por ejemplo, «*ayusada*» o «*ayusismo*», términos peyorativos referidos a Ayuso, aparecen en eldiario.es y El País, pero no en ABC), no se utilizan con suficiente frecuencia como para ser significativos.

Isabel Díaz Ayuso, Pablo Iglesias y Vox (y en menor medida, Rocío Monasterio) son las figuras centrales durante toda la campaña. Las menciones a Mónica García y su partido, Más Madrid, son insignificantes en comparación con las de los otros candidatos. Además, a medida que aumenta la frecuencia de palabras con connotaciones violentas o negativas (por ejemplo, *amenaza, violencia, fascismo, odio, desastre*), las menciones a términos relacionados con propuestas políticas (*educación, impuestos, inmigrantes*...) disminuyen significativamente.

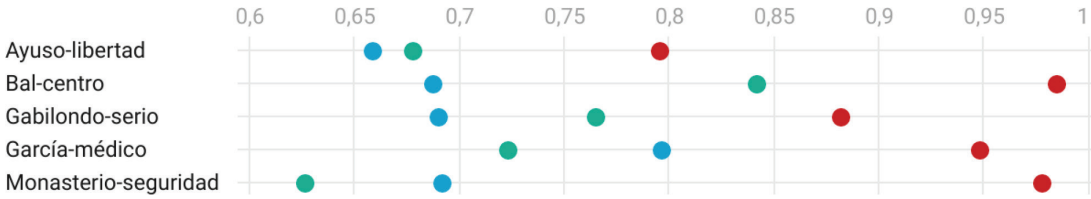
Se ha analizado la relación entre los candidatos y los términos elegidos por sus respectivos partidos políticos para la publicidad electoral. Durante la campaña, el PP asoció la candidatura de Ayuso con *libertad*; el Partido Socialista (PSOE) vinculó a Gabilondo con un *gobierno serio*; Más Madrid enfatizó la figura de García como *médica*; Ciudadanos situó a Bal como el *centro político*; y Vox asoció a Monasterio con la *seguridad* [Figura 5]. UP no empleó términos clave en su campaña electoral, excepto *mayoría*, que también se interpreta como *democracia*.

Figura 5. *Carteles de la campaña electoral. De izquierda a derecha y de arriba abajo: PP, PSOE, Más Madrid, Vox, Ciudadanos y UP*

Fuente: *Recuperado de los programas electorales y sitios web de los partidos.*

El diario ABC es el que más relaciona a los candidatos con sus términos elegidos en campaña, lo que podría ser fruto, entre otros motivos, de reproducir con más fidelidad los mensajes difundidos por los propios partidos políticos, en lugar de interpretarlos o dotarlos de contexto propio. Edmundo Bal, candidato por Ciudadanos, es frecuentemente relacionado con el *centro* [político]. En las noticias de eldiario.es, la mayor relación se observa entre Mónica García y *médico*, mientras que hay una distancia significativa para Monasterio y *seguridad*. El País vincula más fuertemente *Bal-centro* y *Gabilondo-serio* y, por el contrario, apenas relaciona *Monasterio* con *seguridad* [Figura 6].

Figura 6. *Difusión de los términos de campaña. Comparación de los términos empleados por los partidos y su reproducción en medios*



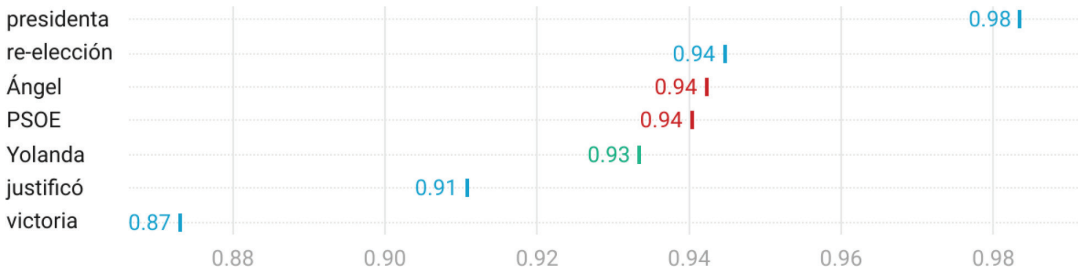
Fuente: *elaboración propia.*

Con word2vec se han obtenido las palabras de menor distancia euclídea para cada candidato y se han sometido a t-SNE (*t-Distributed Stochastic Neighbor Embedding*), una técnica de reducción no lineal que proyecta los datos en un plano de dos dimensiones y obtiene patrones de grupos en base a la similitud de los datos. En ABC, entre los diez términos más cercanos a *Ayuso* (esto es, los de mayor probabilidad de aparecer junto al apellido de la candidata), encontramos *presidenta* (0,9836) y *reelección* (0,9447), pero también *Ángel [Gabilondo]* (0,9423) y *PSOE* (0,9404). En eldiario.es destacan *reelección* (0,8831) y *victoria* (0,8731). En El País, sin embargo, se observa una estrecha distancia entre *Ayuso* y la vicepresidenta del Gobierno *Yolanda* [Díaz] (0,9335), y el verbo *justificó* (0,9108) [Figura 7].

Para Vox, la diversidad en la cobertura es notable: en ABC, los términos más cercanos a este partido son *afear* (0,998977), *anunciante* (0,998976) y tam-

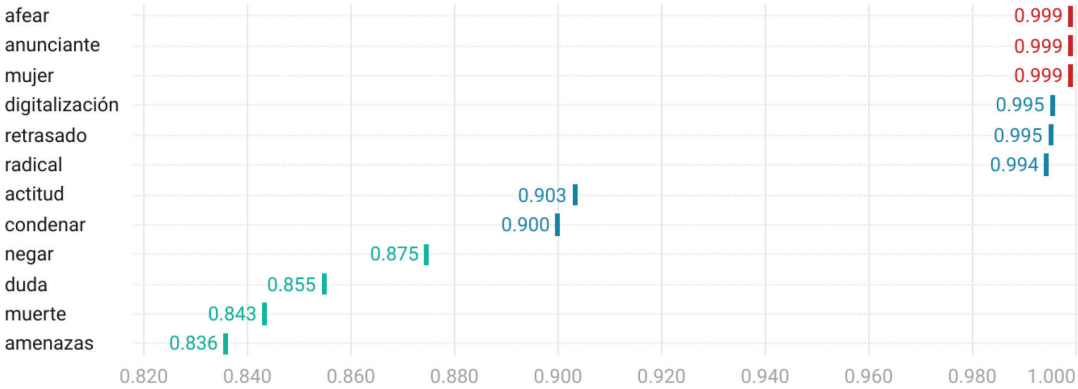
bién *mujer* (0,998958). En eldiario.es lo son *digitalización* (0,995412), *retrasado* (0,995026) y *radical* (0,994319). Con Rocío Monasterio, la candidata por Vox, se relacionan *actitud* (0,903149) y *condenar* (0,89991). El País relaciona a la cabeza de lista de este partido con *negar* (0,8746), *duda* (0,8547), *muerte* (0,8433) y *amenazas* (0,8358) entre las diez palabras más cercanas [Figura 8].

Figura 7. *Términos cercanos a Isabel Díaz Ayuso medidos con word2vec. Existe una alta probabilidad de que «Ayuso» aparezca junto a «presidenta» y «reelección» en ABC y junto a «reelección» y «victoria» en eldiario.es*



Fuente: *elaboración propia.*

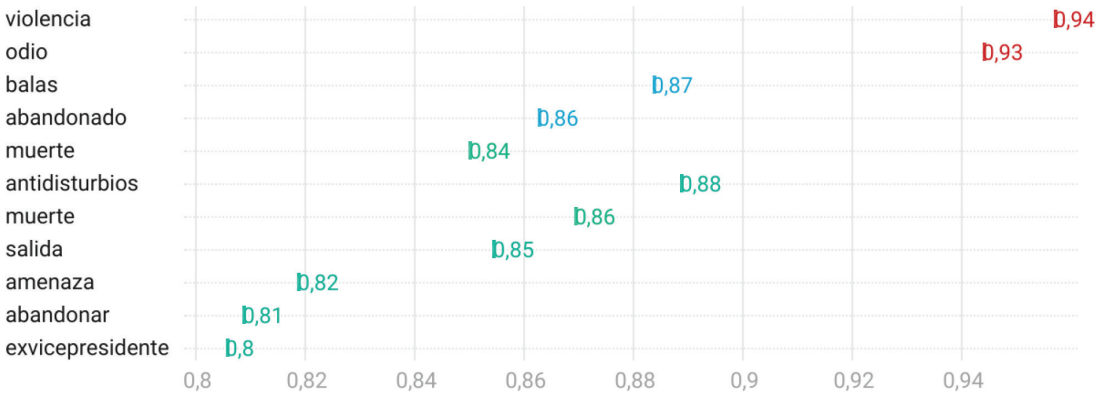
Figura 8. *Términos cercanos a Vox medidos con word2vec. Entre las palabras más cercanas a Vox se encuentran términos con connotaciones negativas como «afear», «retrasado», «radical», «condena», «negar», «duda», «muerte» y «amenazas»*



Fuente: *elaboración propia.*

Relativo a Unidas Podemos y a su candidato Pablo Iglesias, los términos más cercanos calculados por distancia euclídea tienen una connotación negativa en eldiario.es y El País. Para el primero de estos medios, hay una fuerte vinculación con *balas* (0,8734) y *abandonado* (0,855). Para El País, con *antidisturbios* (0,878), *muerte* (0,86095) y *salida* (0,84783). En ABC, aunque no son parte de los diez términos más cercanos, *violencia* y *odio* presentan índices muy elevados (0,938 y 0,9265) [Figura 9].

Figura 9. *Términos cercanos a UP y su candidato Pablo Iglesias medidos con word-2vec. Están fuertemente vinculados a «violencia», «odio» y «muerte», entre otros*



Fuente: *elaboración propia.*

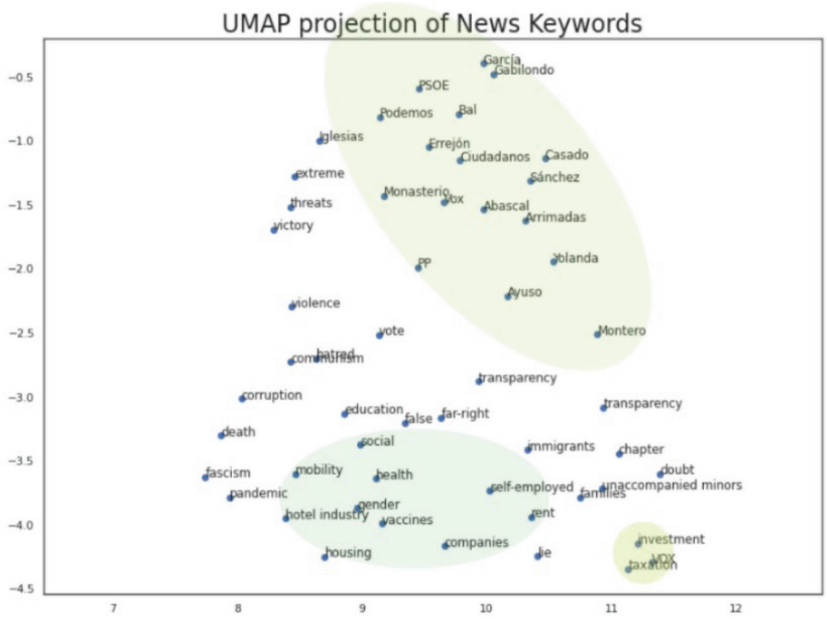
Para otros términos, se observa que en ABC es menos probable que Ayuso se vincule a *corrupción* (0,7659) o a *pandemia* (0,8118), al contrario que respecto a Bal (*violencia*: 0,9801; *corrupción*: 0,9662; *pandemia*: 0,9596). En eldiario.es se da una vinculación significativa entre *violencia* y *dictadura* (0,9692), y *violencia* y *ultraderecha* (0,9661). Para El País se halla una baja probabilidad de que *García* y *corrupción* aparezcan juntas (0,6497), o *García* y *violencia* (0,5757), siendo más alta la probabilidad con otros candidatos y, concretamente, con Bal: *corrupción*: 0,8303; *violencia*: 0,7707.

Se ha empleado también el reductor dimensional UMAP (*Uniform Manifold Approximation and Projection*), que permite plasmar los vectores generados en un plano de dos dimensiones, asumiendo criterios diferentes a t-SNE. De esta manera se pueden observar ciertas agrupaciones entre los términos, diferenciadas para cada medio. Se percibe que los términos se distribuyen conforme a tres temáticas:

- Los partidos y sus candidatos, que conforman entre sí grupos relativamente estables. Corresponde a la respuesta de *who* (quién) en la pirámide periodística.
- Las palabras relativas a la ideología (*derecha, izquierda, fascismo, odio*...) coincidirían con el *how* (cómo) en la estructura de noticias.
- Por último, aquellas palabras relativas a las propuestas políticas, que se distribuyen a lo largo del plano: *fiscalidad, empresas, salud*... corresponden al *what* (qué) de la noticia.

Se ha observado que el foco informativo se centra en el sujeto (quién) y la ideología (cómo), en lugar de priorizar el *qué* de la noticia. La distribución obtenida con UMAP agrupa candidatos y partidos cercanos a palabras relativas a la ideología y de connotación agresiva, y alejados de las propuestas electorales. En El País, *Iglesias* aparece más alejado de los otros candidatos y particularmente cerca de *extremo* y *amenaza* [Figura 10].

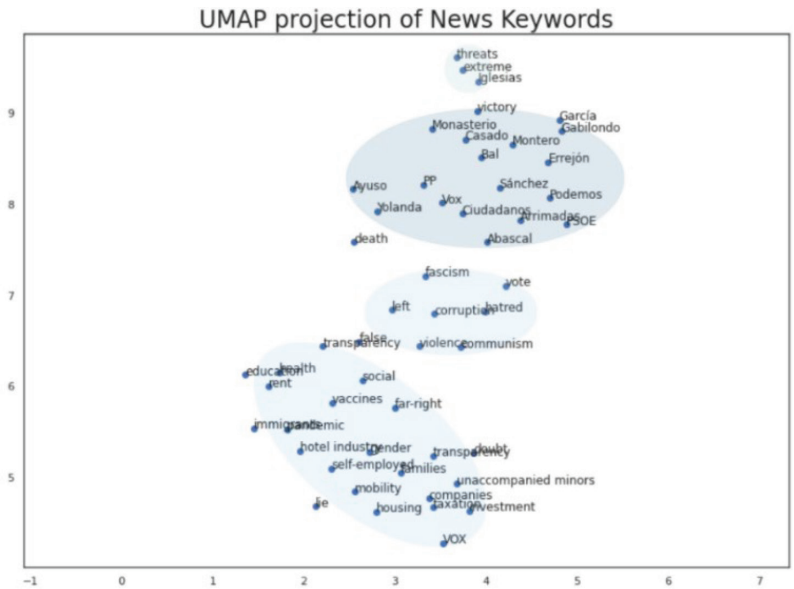
Figura 10. *Distribución de términos para El País obtenida con UMAP. Se agrupan los candidatos y los partidos. «Iglesias» y «Monasterio» se acercan a «extremo» y «amenazas»*



Fuente: *elaboración propia.*

Algunos términos referidos a propuestas, como *pandemia, transportes* o *educación*, se colocan más próximos a palabras de connotación ideológica (*falso, ultraderecha* o *fascismo*) [Figuras 10, 11].

Figura 11. *Distribución de términos para eldiario.es. Partidos y candidatos aparecen agrupados y cercanos a términos ideológicos y agresivos, y más alejados de las propuestas de gobierno. «Iglesias» es próximo a «amenazas» y «extremo», y «Monasterio» a «victoria»*



Fuente: *elaboración propia.*

7. Discusión y conclusiones

Según los resultados obtenidos, no puede afirmarse la hipótesis de que los medios de comunicación digitales presenten información desigual relativa a la campaña electoral unos respecto a otros. Si bien hay diferencias entre cómo se representan los candidatos en cada uno de los medios, hay importantes puntos comunes entre ellos:

- La información de la campaña electoral se centra exclusivamente en los seis partidos mayoritarios.
- Es llamativa la gran cobertura que se ha hecho del bloque de la derecha en estas elecciones. El Partido Popular y Vox (y/o sus candidatos), las agrupaciones consideradas conservadoras, han sido con gran diferencia las que mayor cobertura mediática han obtenido en todos los medios. Le sigue el partido más a la izquierda, Unidas Podemos, y quedan relegadas las agrupaciones consideradas de centro, polarizándose la visibilidad de los extremos del espectro ideológico.
- Además, el trato que se ha hecho de los partidos más nombrados puede considerarse negativo e incluso violento: en esta campaña electoral se recogen, en los tres diarios estudiados, 225 menciones de la palabra *odio*, 191 de *violencia*, 214 de *muerte,* y solo 153 veces la palabra *propuesta*.
- Se descuida la comunicación de los programas electorales —incluso en el contexto de pandemia que encuadra estos comicios—, en favor de los sucesos, polémicas e hitos producidos durante la campaña y del discurso exclusivamente ideológico.

Estas conclusiones refuerzan y actualizan estudios previos en los que se recalca un desigual tratamiento de los partidos minoritarios en la prensa española: según Gallardo-Paúls y Enguix-Oliver (2015), las agrupaciones minoritarias reciben menos menciones, a menudo con denominación parafrástica; su voz se emplea habitualmente para hablar de los grandes partidos y sus representantes, no para mostrar ideas de las propias formaciones. Los estudios de Levendusky (2013) y Prior (2005), citados por Haselmayer et al. (2017), demuestran que una cobertura mediática desequilibrada tiende a acentuar las desigualdades en la distribución del conocimiento político.

El empleo de términos violentos y de confrontación está tan extendido en los medios de comunicación y en el discurso político que los lectores se ven constantemente expuestos a estos estímulos negativos, los cuales incrementan o provocan tensión social y conducen a niveles peligrosos de desprecio y agresividad (Fernández-Fernández, 2009). Como se ha mencionado anteriormente, otros estudios han demostrado que la influencia del lenguaje en el individuo y en los medios puede tener efectos relevantes en la percepción de la política por parte de la ciudadanía (Casero-Ripollés et al., 2016; Gavin, 2018). Además, los periodistas suelen

utilizar palabras típicas del entretenimiento en lugar de palabras informativas en las noticias y titulares.

En este estudio se ha observado que la información sobre las candidatas suele incluir citas o ideas de sus colegas masculinos del partido —Ayuso es una excepción—, mientras que los candidatos masculinos no siempre se acompañan de sus homólogas femeninas. Las citas de varones se utilizan frecuentemente, incluso cuando no tienen una relevancia directa. Van der Pas y Aaldering (2020) ya habían detectado anteriormente un sesgo de género en las noticias de los medios políticos, donde las mujeres políticas recibían 17 puntos porcentuales menos de atención mediática. Esto no es nuevo (por ejemplo, Conroy et al., 2015; Smith, 1997), pero también puede observarse en este estudio.

Como se presentó en el estado del arte, la cobertura excesiva de los actos de campaña y la falta de información sobre las diferencias políticas y los programas disminuyen la confianza de la población en la prensa (Jones, 2004). Durante la campaña electoral del 4M, incluso en el contexto de una pandemia que aumentó la demanda de propuestas políticas, los medios de comunicación no ahondaron en las medidas electorales y las diferencias entre los partidos.

En diálogo con otras investigaciones similares, el trabajo de Mancera-Rueda y Villar-Hernández (2020), centrado en el análisis mediático del partido Vox para las elecciones generales de abril de 2019, recoge hasta 413 menciones al partido o a su entonces candidato Santiago Abascal en los titulares de 8 medios. En este caso, se recogen hasta 93 menciones de Vox o Monasterio en titulares, en 3 medios y para unas elecciones autonómicas. Si ampliamos el alcance a los cuerpos de las noticias, se contabilizan 1.589 menciones, algo que no habían medido estudios anteriores. En las elecciones de 2019, Mancera-Rueda y Villar-Hernández ya informaron de que la cobertura de Vox en los comicios se basó principalmente en actividades de la campaña y en su candidato, priorizados ante las propuestas de gobierno. Además, dichas investigaciones identificaron numerosas expresiones negativas relacionadas con Vox (Mancera-Rueda y Villar-Hernández, 2020), también detectadas en este estudio. Aunque se desconoce si esta es la intención del partido político, como ocurría con Podemos en el estudio de Casero-Ripollés et al. (2016), las intervenciones polémicas de su candidato le han permitido una mayor cobertura mediática y exposición.

Sánchez-Gutiérrez (2016) señaló, tras un análisis del discurso, que existía un tratamiento mediático negativo de manera sistemática hacia Podemos y la personalización de acciones negativas en Iglesias. Aunque del presente estudio no se pueden extraer conclusiones de tal contundencia, se percibe que Iglesias personifica a UP en un gran número de noticias, con muchas más menciones al candidato que al grupo político.

En un estudio posterior (Sánchez-Gutiérrez y Nogales-Bocio, 2018) se indicó, tras un análisis crítico del discurso, que subyace la idea del «fracaso» en el discurso mediático sobre Podemos y Pablo Iglesias. Como se observa en este experimento, existe una probabilidad alta o medio-alta de que junto a los términos referidos a este partido aparezcan también *duda*, *abandono*, *violencia*, *odio* y *muerte*. En cualquier caso, estas elecciones fueron mediáticas y agresivas, una tendencia que otros autores mencionados ya habían detectado, y que tiene efectos negativos en la población (frustración, fatiga, desconfianza y decepción con el Gobierno). Además, el peso de la información partidista a menudo recae en un candidato y su popularidad, como señalan Casero-Ripollés et al. (2016).

La principal fortaleza de este estudio radica en la metodología empleada en el campo del periodismo y en el estudio del discurso. El enfoque exploratorio —y no clasificatorio— adoptado evita la simplificación excesiva y revela nuevas líneas y objetivos de investigación para explorar en profundidad. A diferencia de otros estudios, aquí se realiza una comparación para cada medio y para cada candidato o partido, lo que permite contrastar los términos y las narrativas utilizados por algunos de los medios más importantes del país.

Sin embargo, hay claras limitaciones. El tamaño de la muestra de medios que ofrecen noticias en abierto es limitado, lo que dificulta comparaciones más amplias. La desambiguación manual de términos presenta otra limitación, que apunta a la comparación con herramientas automatizadas como BERT o ELMo. En futuros experimentos, consideraremos el uso de estas herramientas para optimizar nuestro trabajo.

En cualquier caso, este experimento revela algunas líneas de investigación muy interesantes, no solo en términos de PLN (desambiguación de términos, métricas de estabilidad, etc.), sino también en el campo de la cobertura mediática política. Aunque la evaluación de la polarización no era el objetivo principal, los resultados

sugieren su relevancia, en consonancia con las preocupaciones sobre su impacto en la democracia. En el contexto tecnológico en el que nos encontramos, es una oportunidad abordar el sesgo y la polarización mediáticas con metodologías automatizadas, que faciliten la comparación de textos y fuentes para lograr una información más diversa. Para futuras líneas de investigación, se propone el uso de nuevas tecnologías de PNL, especialmente herramientas generativas, para detectar el sesgo en los medios de comunicación.

Contribución de autores

Mar Castillo-Campos: Metodología, Investigación, Curación de datos, Redacción – Borrador original. **David Becerra-Alonso:** Conceptualización, Metodología, Software, Validación, Supervisión. **David Varona-Aramburu:** Conceptualización, Validación, Redacción – Revisión y edición, Supervisión. Los autores han leído y están de acuerdo con la versión publicada del manuscrito.

Conflicto de intereses

Todos los autores declaran no tener conflicto de intereses.

Referencias

AIMC (2020). *Marco general de los medios en España*. Asociación para la Investigación de Medios de Comunicación. https://bit.ly/4jeNJ6h

Aljarah, Ibrahim; Habib, Maria; Hijazi, Neveen; Faris, Hossam; Qaddoura, Raneem; Hammo, Bassam; Abushariah, Mohammad; & Alfawareh, Mohammad. (2020). Intelligent detection of hate speech in Arabic social network: A Machine Learning approach. *Journal of Information Science 47*(4). https://doi.org/10.1177/0165551520917651

Al-Omari, Hani; Abdullah, Malak; Altiti, Ola; & Shaikh, Samira. (2019). JUSTDeep at NLP4IF 2019 Task 1: Propaganda detection using ensemble deep learning models. *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, 113-118. https://doi.org/10.18653/v1/D19-5016

Antoniades, Nicos. (2020). Political marketing communications in today's era: Putting people at the center. *Society 57*(6), 646-656. https://doi.org/10.1007/s12115-020-00556-6

Arredondo-Ramírez, Pablo. (2000). Los medios de comunicación y la contienda electoral. *Renglones 46*(12), 68-72. http://hdl.handle.net/11117/512

Asociación de la Prensa de Madrid (2020). *Informe anual de la profesión periodística*. https://acortar.link/rsxJDc

Ballesteros-Herencia, Carlos A. (2020). Los marcos del compromiso: framing y engagement digital en la campaña electoral de España de 2015. *Observatorio (OBS*) 14*(3). https://doi.org/10.15847/obsOBS14320201507

Barandiaran, Xabier; Unceta, Alfonso; & Peña, Simon. (2020). Comunicación política en tiempos de nueva cultura política. *Revista ICONO 14. Revista científica de Comunicación y Tecnologías emergentes 18*(1), 256-282. https://doi.org/10.7195/ri14.v18i1.1382

Berrocal-Gonzalo, Salomé; Martín-Jiménez, Virginia; & Gil-Torres, Alicia. (2017). Líderes políticos en YouTube: información y politainment en las elecciones generales de 2016 en España. *El Profesional de la Información 26*(5), 937-946. https://doi.org/10.3145/epi.2017.sep.15

Berven, Arne; Christensen, Ole A; Moldeklev, Sindre; Opdahl, Andreas L.; & Villanger, Kjetil J. (2020). A knowledge-graph platform for newsrooms. *Computers in Industry,* 123. https://doi.org/10.1016/j.compind.2020.103321

Bright, Jonathan; Hale, Scott; Ganesh, Bharath; Bulovsky, Andrew; Margetts, Helen; & Howard, Phil (2020). Does campaigning on social media make a difference? Evidence from candidate use of Twitter during the 2015 and 2017 UK elections. *Communication Research 47*(7), 988-1009. https://doi.org/10.1177/0093650219872394

Budiman, Irwan; Nugrahadi, Dodon T; Faisal, Mohammad R.; & Rusli, Muhammad. (2019). *A study on effect of generated features from Word2Vec vectors for text classification*. International Conference on Wetland and Multidisciplinary Research 2019, Utsonomiya, Japan

Campos-Domínguez, Eva; & Calvo, Dafne. (2017). Electoral campaign on the Internet: Planning, impact and viralization on Twitter during the Spanish general election, 2015. *Comunicación y Sociedad* (29), 93-116. https://doi.org/10.32870/cys.v0i29.6423

Casero-Ripollés, Andreu; Feenstra, Ramón A.; & Tormey, Simon. (2016). Old and new media logics in an electoral campaign: The case of Podemos and the two-way street mediatization of politics. *The International Journal of Press/Politics 21*(3), 378-397. https://doi.org/10.1177/1940161216645340

Castells, Manuel. (2009). *Comunicación y poder*. Alianza Editorial.

Castillo, Toni. (2017, December 14). El 56,8 % de los lectores españoles de prensa se informa a través de redes sociales, aunque sólo lee titulares y alguna noticia. *Genbeta*. https://bit.ly/3H4CWi0

Chaves-Montero, Alfonso; Gadea-Aiello, Walter F.; & Aguaded-Gómez, José I. (2017). La comunicación política en las redes sociales durante la campaña electoral de 2015 en España: uso, efectividad y alcance. *Perspectivas de la Comunicación 10*(1), 55-83.

Conroy, Meredith; Oliver, Sarah; Breckenridge-Jackson, Ian; & Heldman, Caroline. (2015). From Ferraro to Palin: Sexism in coverage of vicepresidential candidates

in old and new media. *Politics, Groups, and Identities 3*(4), 573-591. https://doi.org/10.1080/21565503.2015.1050412

Córdoba-Cabús, Alba. (2018). Análisis del periodismo de datos en la campaña electoral del 20D a través de las ediciones digitales de diarios generalistas. *Estudios sobre el Mensaje Periodístico 24*(1), 137. https://doi.org/10.5209/ESMP.59942

Deacon, David; & Stanyer, James. (2014). Mediatization: key concept or conceptual bandwagon? *Media, Culture & Society 36*(7), 1032-1044. https://doi.org/10.1177/0163443714542218

Denton, Robert E; Trent, Judith S.; & Friedenberg, Robert V. (2019). *Political Campaign Communication: Principles and Practices*. Rowman & Littlefield.

De Vreese, Claes H; Banducci, Susan A; Semetko, Holli A.; & Boomgaarden, Hajo G. (2006). The news coverage of the 2004 European Parliamentary election campaign in 25 countries. *European Union Politics 7*(4), 477-504. https://doi.org/10.1177/146511650606944

Diez-Gracia, Alba; Sánchez-García, Pilar; & Martín-Román, Javier. (2023). Polarización y discurso emocional de la agenda política en redes sociales: desintermediación y engagement en campaña electoral. *Revista ICONO 14. Revista científica de Comunicación y Tecnologías Emergentes 21*(1). https://doi.org/10.7195/ri14.v21i1.1922

Duggan, Maeve; & Smith, Aaron. (2016, October 25). The political environment on social media. *Pew Research Center*. https://pewrsr.ch/4muEuSG

Ecker, Ullrich K. H; Lewandowsky, Stephan; Chang, Ee P.; & Pillai, Rekha. (2014). The effects of subtle misinformation in news headlines. *Journal of experimental psychology: Applied 20*(4), 323–335. https://doi.org/10.1037/xap0000028

Edell, Aaron. (2018, January 11). I trained fake news detection AI with >95% accuracy, and almost went crazy. *Towards Data Science*. https://bit.ly/43p8eaC

Ergün, Erkan; & Karsten, Niels. (2021). Media logic in the coverage of election promises: Comparative evidence from the Netherlands and the US. *Acta Política 56*(1), 1-25. https://doi.org/10.1057/s41269-019-00141-8

Eurobarómetro (2021). *Opinión pública en la Unión Europea: informe nacional. España*. Parlamento Europeo. https://bit.ly/4mkah8J

Fawzi, Nayla. (2019). Untrustworthy news and the media as "enemy of the people?" How a populist worldview shapes recipients' attitudes toward the media. *The International Journal of Press/Politics 24*(2), 146-164. https://doi.org/10.1177/1940161218811981

Fenoll, Vicente; & Rodríguez-Ballesteros, Paula. (2017). Análisis automatizado de encuadres mediáticos. Cobertura en prensa del debate 7D: el debate decisivo. *El Profesional de la Información 26*(4), 630-640. https://doi.org/10.3145/epi.2017.jul.07

Fernández-Fernández, Maximiliano. (2009). Lenguaje violento en los medios de comunicación españoles. In Marie-Claude Chaput & Manuelle Peloille (Coord.), *Sucesos, guerras, atentados: La escritura de la violencia y sus representaciones* (pp. 193-208). PILAR (Presse, Imprimés, Lecture dans l'Aire Romane).

Gallardo-Paúls, Beatriz; & Enguix-Oliver, Salvador. (2015). Opciones discursivas en la cobertura electoral: los temas de la campaña europea de 2014. In Eulalia Hernández Sánchez & María Isabel López Martínez (Coord.), *Sodalicia Dona* (pp. 231-252). Universidad de Murcia.

Gavin, Neil T. (2018). Media definitely do matter: Brexit, immigration, climate change and beyond. *The British Journal of Politics & International Relations 20*(4), 827-845. https://doi.org/10.1177/1369148118799260

Ghosh, Shreenita; Su, Min-Hin; Abhishek, Aman; Suk, Jiyoun; Tong, Chau; Kamath, Kruthika; Hills, Ornella; Correa, Teresa; Garlough, Christine; Borah, Porismita; & Shah, Dhavan. (2020). Covering #MeToo across the news spectrum: Political

accusation and public events as drivers of press attention. *International Journal of Press/Politics 27*(1). https://doi.org/10.1177/1940161220968081

Goidel, Kirby; Davis, Nicholas T.; & Goidel, Spencer. (2021). Changes in perceptions of media bias. *Research & Politics 8*(1). https://doi.org/10.1177/2053168020987441

Green, Donald P.; & Gerber, Alan S. (2019). *Get out the vote: How to increase voter turnout*. Brookings Institution Press.

Haselmayer, Martin; Wagner, Markus; & Meyer, Thomas M. (2017). Partisan bias in message selection: Media gatekeeping of party press releases. *Political Communication 34*(3), 367-384. https://doi.org/10.1080/10584609.2016.1265619

Hjarvard, Stig. (2004). From bricks to bytes: The mediatization of a global toy industry. In Ib Bondebjerg & Peter Golding (Eds.), *European Culture and the Media* (pp. 43-63). Intellect. https://doi.org/10.2307/j.ctv36xw5qn

Holgado, María. (2017). Publicidad e información sobre elecciones en los medios de comunicación durante la campaña electoral. *Teoría y Realidad Constitucional 40,* 457-485. https://doi.org/10.5944/trc.40.2017.20914

Holtz-Bacha, Christina; Langer, Ana I.; & Merkle, Susanne. (2014). The personalization of politics in comparative perspective: Campaign coverage in Germany and the United Kingdom. *European Journal of Communication 29*(2), 153-170. https://doi.org/10.1177/0267323113516727

Høyer, Svennik; & Nossen, Hedda A. (2015). Revisions of the news paradigm: Changes in stylistic features between 1950 and 2008 in the journalism of Norway's largest newspaper. *Journalism 16*(4), 536-552. https://doi.org/10.1177/1464884914524518

Hutchens, Myiah J; Hmielowski, Jay D; Pinkleton, Bruce E.; & Beam, Michael A. (2016). A spiral of skepticism? The relationship between citizens' involvement

with campaign information to their skepticism and political knowledge. *Journalism & Mass Communication Quarterly 93*(4), 1073-1090. https://doi.org/10.1177/1077699016654439

Jones, David A. (2004). Why Americans don't trust the media: A preliminary analysis. *The Harvard International Journal of Press/Politics 9*(2), 60-75. https://doi.org/10.1177/1081180X04263461

Kalogeropoulos, Antonis; Fletcher, Richard; & Nielsen, Rasmus K. (2018). News brand attribution in distributed environments: Do people know where they get their news? *New Media & Society 21*(3), 583–601. https://doi.org/10.1177/1461444818801313

Kalogeropoulos, Antonis; Suiter, Jane; Udris, Linards; & Eisenegger, Mark. (2019). News media trust and news consumption: Factors related to trust in news in 35 countries. *International Journal of Communication 13*(22), 3672-3693. https://rb.gy/vxm36o

Köffer, Sebastian; Riehle, Dennis M; Höhenberger, Steffen; & Becker, Jörg. (2018). *Discussing the value of automatic hate speech detection in online debates*. Multikonferenz Wirtschaftsinformatik: Data Driven X-Turning Data in Value 2018, Leuphana, Germany.

Koloski, Boshko; Pollak, Senja; Škrlj, Blaž; & Martinc, Matej. (2021). Extending neural keyword extraction with TF-IDF tagset matching. arXiv preprint arXiv:2102.00472. https://doi.org/10.48550/arXiv.2102.00472

Levendusky, Matthew S. (2013). *How partisan media polarize America*. University of Chicago Press.

Li, Zeming; & Sun, Xinying. (2021). Analysis of the impact of media trust on the public's motivation to receive future vaccinations for COVID-19 based on protection motivation theory. *Vaccines 9*(12), 1401. https://doi.org/10.3390/vaccines9121401

López-García, Guillermo. (2017). Comunicación política y discursos sobre el poder. *El Profesional de la Información 26*(4), 573-578. https://doi.org/10.3145/epi.2017.jul.01

López-Meri, Amparo. (2017). Citizen contribution to the electoral debate and media coverage on Twitter. *Prisma Social* (18) 1-33.

Magallón, Raúl; Nachawati, Leila; & Seoane, Francisco. (2021). *Evaluación del riesgo de desinformación: el mercado de noticias online en España*. Madrid: GDI Global Disinformation Index. https://hdl.handle.net/10016/33587

Maier, Jürgen; & Nai, Alessandro. (2020). Roaring candidates in the spotlight: campaign negativity, emotions, and media coverage in 107 national elections. *The International Journal of Press/Politics 25*(4), 576-606. https://doi.org/10.1177/1940161220919093

Mancera-Rueda, Ana; & Villar-Hernández, Paz. (2020). Análisis de las estrategias de encuadre discursivo en la cobertura electoral sobre Vox en los titulares de la prensa española. *Doxa Comunicación 31*, 315-340. https://doi.org/10.31921/doxacom.n31a16

Mazaira-Castro, Andrés; Rúas-Araújo, José; & Puentes-Rivera, Iván. (2019). Fact-checking en los debates electorales televisados de las elecciones generales de 2015 y 2016. *Revista Latina de Comunicación Social 74,* 748-766. https://doi.org/10.4185/RLCS-2019-1355

McInnes, Leland; Healy, John; & Melville, James. (2018). UMAP: Uniform Manifold Approximation and Projection for dimension reduction. arXiv preprint arXiv:1802.03426. https://doi.org/10.48550/arXiv.1802.03426

Miguel-Sáez-de-Urabain, Ainara; Fernández de Arroyabe-Olaortua, Ainhoa; & Lazkano-Arrillaga, Iñaki. (2017). La espectacularización de la información política. El caso de El País en las elecciones estadounidenses de 2016. *Revista Latina de Comunicación Social 72,* 1131-1147. https://doi.org/10.4185/RLCS-2017-1211

Mikolov, Tomas; Chen, Kai; Corrado, Greg; & Dean, Jeffrey. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. https://doi.org/10.48550/arXiv.1301.3781

Mustafa, Ghulam; Usman, Muhammad; Afzal, Muhammad T; Shahid, Abdul; & Koubaa, Anis (2021). A comprehensive evaluation of metadata-based features to classify research paper's topics. *IEEE Access 9,* 133500-133509. https://doi.org/10.1109/ACCESS.2021.3115148

Nelson, Candice J.; & Thurber, James A. (Eds.) (2018). *Campaigns and elections American style: The changing landscape of political campaigns*. Routledge. https://doi.org/10.4324/9780429468278

Newman, Nic; Fletcher, Richard; Schulz, Anne; Andı, Simge; & Nielsen, Rasmus K. (2020). *Reuters Institute Digital News Report 2020*. Reuters Institute for the Study of Journalism. https://acortar.link/TvW7gs

Nicholls, Tom; & Culpepper, Pepper D. (2021). Computational identification of media frames: Strengths, weaknesses, and opportunities. *Political Communication 38*(1-2), 159-181. https://doi.org/10.1080/10584609.2020.1812777

Orden-Cruz, Carmen; Gómez-Martínez, Raúl; & Paule-Vianez, Jessica. (2019). Sentimiento de los medios de comunicación españoles en formato digital sobre el Ibex35. *Revista Internacional de Investigación en Comunicación aDResearch ESIC 19*(19), 56-67.

Paniagua-Rojano, Francisco; Seoane-Pérez, Francisco; & Magallón-Rosa, Raúl. (2020). Anatomy of the electoral hoax: Political disinformation in Spain's 2019 general election campaign. *Revista CIDOB d'Afers Internacionals 124,* 123-146. doi.org/10.24241/rcai.2020.124.1.123

Prior, Markus. (2005). News vs. entertainment: How increasing media choice widens gaps in political knowledge and turnout. *American Journal of Political Science 49*(3), 577-592. https://doi.org/10.1111/j.1540-5907.2005.00143.x

Reporteros Sin Fronteras (2020). *Spain: Growing polarisation, lack of transparency*. https://rsf.org/en/spain

Reuters Institute for the Study of Journalism (2023). *Digital News Report 2023*. https://acortar.link/JBHtWS

Riedel, Benjamin; Augenstein, Isabelle; Spithourakis, Georgios P.; & Riedel, Sebastian. (2017). A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. arXiv preprint arXiv:1707.03264 https://doi.org/10.48550/arXiv.1707.03264

Rieis, Julio; De-Souza, Fabrício; Vaz De-Melo, Pedro; Prates, Raquel; Kwak, Haewoon; & An, Jisun. (2015). Breaking the news: First impressions matter on online news. *Proceedings of the International AAAI Conference on Web and Social Media, 9*(1), 357-366. https://doi.org/10.1609/icwsm.v9i1.14619

Sánchez-Gutiérrez, Bianca. (2016). *La representación mediática de los partidos políticos emergentes: el caso de Podemos y Ciudadanos en Atresmedia* [Doctoral dissertation, Universidad de Sevilla].

Sánchez-Gutiérrez, Bianca; & Nogales-Bocio, Antonia I. (2018). La cobertura mediática de Podemos en la prensa nativa digital neoliberal española: una aproximación al caso de OkDiario, El Español y El Independiente. *Estándares e indicadores para la calidad informativa en los medios digitales,* 125-146.

Smith, Glen R. (2010). Politicians and the news media: How elite attacks influence perceptions of media bias. *International Journal of Press/Politics 15*(3), 319-343. https://doi.org/10.1177/1940161210367430

Smith, Kevin B. (1997). When all's fair: Signs of parity in media coverage of female candidates. *Political Communication 14*(1), 71-82. https://doi.org/10.1080/105846097199542

Smirnova, O. V; Denisova, G. V; Alevizaki, O. R; Ilyichenko, D. S.; & Antipova, A.S. (2021). Space-related scientific and technical information in current media discourse: Research results. *Scientific and Technical Information Processing 48*(2), 133-138. https://doi.org/10.3103/S0147688221020118

Terachi, Masahiro; Saga, Ryosuke; & Tsuji, Hiroshi. (2006). *Trends recognition in journal papers by text mining*. 2006 IEEE International Conference on Systems, Man and Cybernetics 2006 (Vol 6, pp. 4784-4789). https://doi.org/10.1109/ICSMC.2006.385062

Terkildsen, Nayda; & Damore, David F. (1999). The dynamics of racialized media coverage in congressional elections. *The Journal of Politics 61*(3), 680-699. https://doi.org/10.2307/2647823

Thurber, James A.; & Nelson, Candice J. (2018). Elections in a polarized America: Understanding the dynamics and the transformation of American political campaigns. In Candice J. Nelson & James A. Thurber (Eds.) *Campaigns and Elections American Style* (pp. 1-10). Routledge.

Van der Maaten, Laurens; & Hinton, Geoffrey. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research 9*(11).

Van der Pas, Daphne J.; & Aaldering, Loes. (2020). Gender differences in political media coverage: A meta-analysis. *Journal of Communication 70*(1), 114-143. https://doi.org/10.1093/joc/jqz046

Van Duyn, Emily; & Collier, Jessica. (2019). Priming and fake news: The effects of elite discourse on evaluations of news media. *Mass Communication and Society 22*(1), 29-48. https://doi.org/10.1080/15205436.2018.1511807

Veres, Luis. (2006). *La retórica del terror: sobre lenguaje, terrorismo y medios de comunicación*. Ediciones de la Torre.

Vizoso, Ángel; & López-García, Xosé. (2020). Newtral y Comprobado: experiencias de fact-checking durante la campaña electoral de las Elecciones Generales en España. In Iván Puentes-Rivera, Ana Belén Fernández-Souto & Montse Vázquez-Gestal (Eds.), *Debate sobre los Debates Electorales y Nuevas Formas de Comunicación Política* (pp. 77-98). Cuadernos Artesanos de Comunicación.

Xu, Kuai; Wang, Feng; Wang, Haiyan; & Yang, Bo. (2018). A first step towards combating fake news over online social media. In *International Conference on Wireless Algorithms, Systems, and Applications,* 521-531. Springer, Cham. https://doi.org/10.1007/978-3-319-94268-1_43

Yamamoto, Mikio; & Church, Kenneth W. (2001). Using suffix arrays to compute term frequency and document frequency for all substrings in a corpus. *Computational Linguistics 27*(1), 1-30. https://doi.org/10.1162/089120101300346787

Zhao, Jieyu; & Chang, Kai-Wei. (2020). LOGAN: Local Group Bias Detection by Clustering. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing,* 1968-1977. https://doi.org/10.18653/v1/2020.emnlp-main.155

Zhou, Pei; Shi, Weijia; Zhao, Jieyu; Huang, Kuan-Hao; Chen, Muhao; & Chang, Kai-Wei. (2019). Analyzing and mitigating gender bias in languages with grammatical gender and bilingual word embeddings. *Annual Meeting of the Association for Computational Linguistics.* ACL 2019 Montréal, QC, Canada.