



UNIVERSIDAD DE GRANADA

Medios de Comunicación y Polarización; Un Análisis del Feminismo en España mediante Procesamiento del Lenguaje Natural

Autor:

Ignacio-Jesús Serrano-Contreras

Directores:

Óscar García Luengo

Javier García Marín

Tesis Doctoral

Programa de Doctorado en Ciencias Sociales

Agosto, 2021

Editor: Universidad de Granada. Tesis Doctorales

Autor: Ignacio Jesús Serrano Contreras

ISBN: 978-84-1117-085-7

URI: <http://hdl.handle.net/10481/71409>

On paper we're ideal. But life isn't on paper.

Whatever Works de Woody Allen

*Viajar es útil, hace trabajar la imaginación. [...]
Hombres, animales, ciudades y cosas, todo es inventado. [...]
Nada más que una historia ficticia.
[...] Basta con cerrar los ojos.
Ocurre al otro lado de la vida.*

Viaje al fin de la noche de Louis-Ferdinand Céline

Medios de Comunicación y Polarización; Un Análisis del Feminismo en España mediante Procesamiento del Lenguaje Natural

Mass Media and Polarization; An Analysis of Feminism in Spain via Natural Language Processing

Resumen

La cuestión fundamental que este trabajo plantea es por qué la sociedad se divide. Son muchos autores los que, desde diferentes ámbitos, trazan patrones acerca de este elemento. Nuestra particular reflexión detenta la misión de analizar y conocer cómo este fenómeno se conforma en las ciencias sociales. De este modo, una de las misiones de este trabajo es comprender la polarización política en su variabilidad de supuestos y escenarios.

Estas disensiones entre ciudadanos cobran significancia debido a enmarañadas disquisiciones morales, éticas y políticas que dan lugar a corrientes de pensamiento como ecologismo, socialismo, liberalismo, nacionalismo o feminismo. Será por medio de esta última teoría desde la que busquemos cotejar cómo la opinión pública, la esfera política o los medios de comunicación de masas se dividen, aproximan, dispersan y reconfiguran posiciones y conceptos.

Así, este trabajo emerge desde la multidisciplinaridad y el saber acumulativo. Lo que manifestamos no es solo la necesidad de reportar un entendimiento de fenómenos como la polarización o el feminismo, sino que pretendemos seguir abriendo una estela para que, independientemente de la partícula de análisis, la sociedad pueda entender y seguir entendiendo, pueda preguntar y seguir preguntando, cómo se entrelaza nuestro mundo. Por consiguiente, la presente tesis tiene como objetivos: (i) entender qué es la polarización política; (ii), comprender qué papel juega el movimiento y la doctrina feminista presente en el ámbito público; (iii), analizar cómo estos fenómenos sociales emergen dentro de los medios de comunicación de masas; (iv), contribuir a la expansión de nuevos métodos de aprendizaje de máquina dentro de la ciencia política. Para la consecución de nuestros tres primeros objetivos se ha dado forma a un marco en donde se ponen de manifiesto las diferentes líneas teóricas y terminológicas expuestas sobre las anteriores cuestiones. Además, se busca con ello poder encontrar unos puntos de conexión con los que dar forma a nuestros planteamientos. Asimismo, y en relación con el último de los objetivos, este busca aportar nuevas líneas metodológicas para las ciencias sociales, empleando estas como modelos de análisis de nuestras temáticas. A este respecto, los enfoques utilizados surgen de la linde de distintas disciplinas como la computación, la psicología y las ciencias sociales, conformándose en lo conocido como Natural Language Processing.

Con todo ello, los resultados sugieren que el feminismo no es concebido como un fenómeno polar en los medios de masas españoles analizados. Tampoco resulta claramente polarizante al ser analizado en redes sociales, en este caso en plataformas como YouTube. A su vez, sí hemos detectado que, una parte del feminismo, el que crece a raíz de la segunda ola de este y da forma al género, es el que obtiene la mayor presencia dentro de los periódicos españoles. También, se pone de relieve que el feminismo mediático español está inserto en una serie de tendencias que se ven influenciadas por sucesos coetáneos, como el influjo del poder político o factores de índole social.

Abstract

The fundamental question that this work raises is why society is divided. There are many authors who, from different fields, trace patterns about this element. Our reflection has the mission of analyzing and knowing how this phenomenon is conjectured in the social sciences. In this way, this study is aimed at understanding the development of political polarization by bringing a new face to this complex prism.

These dissensions among citizens become significant due to entangled moral, ethical and political disquisitions that give rise to currents of thought such as ecologism, socialism, liberalism, nationalism or feminism. It will be through this last theory that we will seek to compare how public opinion, the political sphere or the mass media divide, approximate, disperse and reconfigure positions and concepts.

Thus, this work arises from mutable multidisciplinary and cumulative knowledge. What we manifest is not only the need to report an understanding of phenomena such as polarization or feminism, but we intend to continue opening a trail so that, independent of the particle of analysis, society can understand and continue to understand, can ask and continue to ask, how our world is intertwined. Therefore, the present thesis aims to: (i) understand what political polarization is; (ii), comprehend what role the feminist movement and doctrine present in the public sphere plays; (iii), analyze how these social phenomena boil within the mass media; (iv), contribute to the expansion of new machine methods within political science. In order to achieve our first three objectives, a framework has been developed in which the different theoretical and terminological lines of thought on the above issues are highlighted. In addition, the goal is to find some points of connection with which to shape our approaches. Likewise, and in relation to the last of the objectives, it seeks to contribute new methodological lines for the social sciences, using them as models of analysis of our topics. In this regard, the approaches used arise from the border of different disciplines such as computer science, psychology, and social sciences, forming what is known as Natural Language Processing.

The results suggest that feminism is not conceived as a polar phenomenon in the Spanish mass media analyzed. Nor is it clearly polarizing when analyzed in social media, in this case on platforms such as YouTube. In turn, we have detected that a part of feminism, the one that grows as a result of the second wave of feminism and shapes gender, is the one that obtains the greatest presence in Spanish newspapers. It is also clear that Spanish media feminism is part of a series of trends that are influenced by contemporary events, such as the influence of political power or social factors.

Agradecimientos

El presente texto, aún debiéndose a la formalidad de reseñar solo un padre, tiene infinitos ancestros. A todos, y sin excepción, se les dedica y agradece.

En primer lugar, dar las gracias a mi exégesis, mi oxímoron, mi hermana. Siempre le digo que de ser yo crepúsculo, ella no podría ser aurora, porque una linda con la otra, y ella y yo, por suerte o desgracia, somos dos líneas paralelas que nunca convergen, pero tampoco se separan. Por ti, por siempre, esto es tuyo.

A mis padres, porque tuvieron el capricho y las ganas de tener y aguantar a alguien como yo. No creo que debiera de encontrarse mayor elogio de agradecimiento.

A mi tío, por tener la capacidad de estar.

A Óscar y Javi, a Javi y Óscar. Por enseñarme un oficio, por mostrarme un camino, por, simplemente, esa manía de ser como son. Por dejarme ser discípulo y por poder decirlos Maestros. Mis estimados amigos: ¡Góngora vive, la lucha sigue!

Por último, a todos y cada uno de los amigos que, de algún modo u otro, estuvieron y me permitieron ser quien soy. Y en especial, aunque sin orden de importancia a: Paco, Puchi, Javi Tardío, David, Eloy, Loren, Javi Golbano, Rafa y Cristina, Víctor, Miguel Rueda, Alba, Alberto Heras, Rafa C., Tote, Manolillo, Apolo, Cespedosa, Juan Ramón, Elena Maiki, Juan Antonio, F.M. Delgado, Ana Corazón...

Breve exposición de motivos

Como se podrá apreciar, los análisis que aquí van a definirse y conjugarse han sido elaborados desde preguntas genéricas hasta el entendimiento persistente de cuestiones y dudas, por medio de objetivos e hipótesis útiles que sirven para dar carta de naturaleza científica a este proyecto. Mientras que, por un lado, tenemos la misión de construir un marco conceptual para entender nuevos fenómenos y métodos, por otro, y gracias a la anterior aspiración, encontraríamos, en la disección del feminismo en España, una pieza con la que poner en práctica todos los conocimientos adquiridos.

A su vez, me gustaría exponer que el trabajo aquí expuesto es el reflejo de todo lo aprendido durante mi etapa doctoral. Parte de ella ha sido fruto de los comentarios que mis directores, muy acertadamente, han convenido en precisarme. Desde la elección de la modalidad de tesis, como el apremio por conocer e implementar nuevas metodologías, hasta el estudio del procesamiento del lenguaje natural. Por todo ello, el presente proyecto no trata solo de ser una unificación de trabajos, sino que gracias a ellos hemos podido elaborar las sustancias de un estudio genuino y deudor de la experiencia alcanzada. Por lo que se pretende una puesta en común de todo este bagaje académico, fijando para ello distintos hilos conductores por los que transitar.

Listado de publicaciones originales

Serrano-Contreras, I-J, García-Marín, J, y Luengo, Ó. G. (2021). Coberturas mediáticas, polarización y reformas educativas en España. *Revista de ciencia política (Santiago)*, Epub 27 de julio de 2021. <https://dx.doi.org/10.4067/S0718-090X2021005000109>

Revista de ciencia política (Santiago) - Factor de impacto (2020): 0.93 - Q1 Political Science and International Relations - H-Index 18

Resumen

Las leyes de educación comportan, para todos los partidos políticos, una de las correas transmisoras de su proyecto ideológico. De este modo, cada uno de los diferentes gobiernos españoles, como ocurre en otros países, han llevado a cabo modificaciones legislativas con el fin de implementar políticas de un signo más cercano a sus intereses partidistas. A su vez, el contexto mediático, siguiendo los postulados de los académicos en comunicación política, ha tendido a apoyar u oponerse a dichas reformas, generando una escena polarizada. La presente investigación realiza un análisis del panorama español, durante los últimos cuarenta años, sobre los impactos que las reformas educativas han suscitado en el ámbito mediático. Se emplean para ello nuevas metodologías de aprendizaje de máquina (machine learning), complementadas con el desarrollo de una serie de medidores diseñados para captar fenómenos sociales complejos dentro de escenarios de variabilidad semántica. Los resultados han evidenciado una clara relación entre partidos políticos y medios, así como el posicionamiento que estos últimos, en consonancia con su línea editorial, mostraron con respecto a las reformas educativas que se iban sucediendo a lo largo del periodo de estudio.

Serrano-Contreras I-J. (2021). NLP on YouTube: A Look on Feminism. En M. Musial-Karg, O. G., Luengo (Eds), *Digitalization of Democratic Processes in Europe. Studies in Digital Politics and Governance*. Springer. https://doi.org/10.1007/978-3-030-71815-2_10

Springer - SPI - Orden 4 - ICEE - 670.000

Abstract

Feminism is a complex phenomenon that has evolved in recent times. We can see it through different optics such as social media or traditional media. The aim of this project is to analyze, descriptively, how it has grown on a platform like YouTube. We are using new computer techniques and trying to understand how the concept evolves. Hence, in this paper, we will provide a language analysis that examines deliberative frameworks and online roles. Our study focuses on adding new methods that offer insights into the Digital Agora. The main outcomes show that the deliberation on YouTube, with regard to Spanish Feminism, has received an important and enormous increase especially since 2016, revealing two possible souls within that debate.

Serrano-Contreras, I.-J., García-Marín, J., y Luengo, Ó. G. (2020). Measuring online political dialogue: does polarization trigger more deliberation? *Media and Communication*, 8(4), 63-72. <https://doi.org/10.17645/mac.v8i4.3149>

Media and Communication - Factor de impacto (2020): 0.8 - Q1 Communication - H-Index 19

Abstract

In recent years, we have witnessed an increasing consolidation of different realms where citizens can deliberate and discuss a variety of topics of general interest, including politics. The comments on news posts in online media are a good example. The first theoretical contributions called attention to the potential of those spaces to build a healthy (civic and participatory) public sphere, going much deeper in the process of political dialogue and deliberation (Fung, Gilman, & Shkabatur, 2013; Lilleker & Jackson, 2008; O'Reilly, 2005; Stromer-Galley & Wichowski, 2011). Polarization has been configured as a constant feature of the quality of the mentioned dialogues, particularly in Mediterranean countries (polarized pluralists' cases). One of the research challenges at the moment has to do with the scrutiny of polarization within the political deliberation provoked by news stories. The goal of this article is the analysis of political dialogue from the perspective of the polarization in the increasingly popular network YouTube, which is presenting very particular characteristics. Using a sample of almost 400,000 posted comments about diverse topics (climate change, the Catalanian crisis, and Political parties' electoral ads) we propose an automated method in order to measure polarization. Our hypothesis is that the

number of comments (quantitative variable) is positively related to their polarization (qualitative variable). We will also include in the examination information about the ideological editorial line of newspapers, the type of topic under discussion, the amount of traceable dialogue, etc. We propose an index to (1) measure the polarization of each comment and use it to show how this value has behaved over time; and (2) verify the hypothesis using the average polarization of comments for each video.

Lista de abreviaturas

AP – Alianza Popular

API – Application Programming Interface

BOW – Bag Of Words

BTM – Biterm Topic Model – Short Text

CSV – Comma-Separated Values

DFM – Document-Feature Matrix

EE.UU. – Estados Unidos

H – Hipótesis

HTML – HyperText Markup Language

KDT- Knowledge Discovery in Textual

LDA – Latent Dirichlet Allocation

LSA – Latent Semantic Analysis

NLP – Natural Language Processing

OMS – Organización Mundial de la Salud

ONU – Organización de las Naciones Unidas

PP – Partido Popular

PSOE – Partido Socialista Obrero Español

RAE – Real Academia Española (de la Lengua)

STM – Structural Topic Models

TERF – Trans-Exclusionary Radical Feminist

TIC – Tecnologías de la Información y la Comunicación

UCD – Unión de Centro Democrática

URL – Uniform Resource Locator

VIOGEN – Violencia de Género

Figuras

Figura 1. La polarización (pg. 25)

Figura 2. Modelo de Estaban y Ray (1994) (pg. 25)

Figura 3. Bramson et al. (2017); tipo 5: *Comunidad fracturada* (pg. 26)

Figura 4. Poole y Rosenthal (pg. 26)

Figura 5. Polarización afectiva (pg. 28)

Figura 6. Representación espacial (Ej. Covid) (pg. 39)

Figura 7. Similaridad vectorial por coseno (pg. 48)

Figura 8. Frecuencias retóricas (pg. 52)

Figura 9. Referencia a “igualdad” (pg. 52)

Figura 10. Nomenclaturas de violencias (pg. 54)

Figura 11. Agregados “Mujer” y “Hombre” (pg. 54)

Figura 12. Agregados por periódico (pg. 55)

Figura 13. Incrustación de los principales tokens (pg. 61)

Figura 14. Fecha de publicación de los vídeos más vistos en YouTube sobre “Feminismo” (pg. 63)

Figura 15. LOESS (log10) de los comentarios presentes en los vídeos seleccionados (pg. 63)

Figura 16. Polarización mediante índice (pg. 64)

Figura 17. (LOESS y media móvil) Polarización mediante índice tomando el año (pg. 65)

Figura 18. Titulares con Feminismo (pg. 66)

Figura 19. Frecuencias de usos y (LOESS) mediante DFM – *El Mundo* (pg. 67)

Figura 20. Frecuencias de usos y (LOESS) mediante DFM – *El País* (pg. 67)

Tablas

Tabla 1. Top 30 términos más prominentes (2009-2014) (pg. 56)

Tabla 2. Top 30 términos más prominentes (2015-2019) (pg. 56)

Tabla 3. Top 30 términos más relevantes por topic ($\lambda = 1$) (2009-2014) (pg. 57)

Tabla 4. Top 30 términos más relevantes por topic ($\lambda = 0$) (2009-2014) (pg. 57)

Tabla 5. Top 30 términos más relevantes por topic ($\lambda = 1$) (2015-2019) (pg. 58)

Tabla 6. Top 30 términos más relevantes por topic ($\lambda = 0$) (2015-2019) (pg. 58)

Tabla 7. Estimación 'theta' (2009-2014) (pg. 59)

Tabla 8. Estimación 'theta' (2015-2019) (pg. 59)

Tabla 9. Selección de Tokens por BTM (2009-2014) (pg. 59)

Tabla 10. Tokens presentes por BTM (2015-2019) (pg. 60)

Tabla 11. Probabilidad de Tokens seleccionados por género BTM (2009-2014) (pg. 60)

Tabla 12. Probabilidad de Tokens seleccionados por género BTM (2015-2019) (pg. 61)

Tabla 13. Selección de analogías de Coseno (θ) por incrustación (pg. 62)

Tabla 14. Similaridad de de Coseno (θ) mediante DFM (pg. 66)

Tabla de contenidos

Resumen	5
Abstract.....	6
Agradecimientos	8
Breve exposición de motivos.....	9
Listado de publicaciones originales	10
Lista de abreviaturas.....	12
Figuras.....	13
Tablas	14
Tabla de contenidos.....	15
1. Introducción	16
Objetivos de la Tesis	18
2. Marco teórico.....	19
2.1. La comunicación (política).....	19
2.1.1. Medios de comunicación de masas	20
2.2. La polarización	24
2.2.1. Breve tipología; de la postura ideológica al sentimiento de pertenencia	27
2.2.2. Sistemas mediáticos y lineamientos ideológicos; recapitulando el caso español	29
2.3. Las olas del feminismo	31
2.3.1. Sobre la cuestión de género	34
2.4. NLP; Descifrando el (enigma del) lenguaje	37
2.4.1. Cerrando el círculo, abriendo el espacio; <i>Text as data</i> y NLP como respuesta	39
3. Metodología y datos.....	42
3.1. Extracción de datos	42
3.2. Minería de texto.....	43
3.3. Indexar la polarización.....	45
3.4. Codificar, incrustar, tematizar, clasificar y generar	46
3.5. Descripción de las muestras	49
3.5.1. T.1. From Catholicism to Gender Equality: Political Rhetoric in Spain from 1980 to 2020..	49
3.5.2. T.2. NLP on YouTube: A look on Feminism.....	49
3.5.3. T.3. A machine learning analysis of gender images in Spanish Press.....	50
3.5.4. Datos supletorios (T.4. y T.5.)	50
4. Resultados	51
Conclusiones	69
Conclusions	72
Bibliografía.....	76
Anexos.....	97
Publicaciones originales.....	97
Coberturas mediáticas, polarización y reformas educativas en España	97
NLP on YouTube: A Look on Feminism.....	115
Measuring Online Political Dialogue: Does Polarization Trigger More Deliberation?.....	128

1. Introducción

El ser humano, en palabras de Aristóteles (trad. en 2018), es un animal social. Juntos, entre las partes, como hacen otros animales, somos capaces de construir desde un primitivo núcleo -como el parentesco- sociedades tecnológicamente complejas como los actuales estados nación (Haidt, 2019). La cuestión que ahora aquí subyace es cómo el pegamento social emana o, una vez adherido, se disipa.

Todo esto parece virar hacia una prospección de la retórica de la confrontación (Scott y Smith, 1969) -o confrontación retórica- donde se congregan tipologías como crispación, radicalidad o polarización. Quizás, la polarización política es la que ha cobrado más presencia en el debate público (Hernández et al., 2021). Su difusa concreción, más allá de la expresión que revela la distancia entre grupos -polos- (Esteban y Ray, 1994; Converse, 2006), invita a una exploración sosegada en la que discernir los modos y formas con los que encajar este puzle. Para ello, es conveniente poder aportar un calibre que sirva para, dependiendo de nuestra partícula de análisis, estipular si las brechas de nuestras poblaciones se encuentran en un estado o, por el contrario, en un proceso de dispersión (DiMaggio et al., 1996).

En la actualidad, hablar de polarización lleva incardinado múltiples satélites que enmarañan la cuestión; medios de comunicación, redes sociales, populismos, además de la incorporación al debate de vocablos que desplazarían el foco hacia la elaboración de nuevas lógicas mediáticas (Altheide, 2016); *fake news* (Quandt et al., 2021), *clickbait* (Munger et al., 2020) o recientes conjeturas sobre el influjo de la era de masas, tales como las cámaras de eco (Cinelli et al., 2021; Scruggs, 1998) o los filtro burbuja (Pariser, 2017). Toda esta entretrejida escena asienta nuevas cuestiones acerca de, por ejemplo, cómo de débiles son los lazos y relevantes los vínculos (Granovetter, 1973) engendrados en las sociedades contemporáneas.

A pesar de la cada vez más vasta literatura sobre la temática, el término continúa aún hoy sin un encaje claro y cerrado sobre el que de forma sistémica construir una exégesis. En virtud de este pretexto aspiramos a dar forma a unas aportaciones que puedan ser extrapolables a regiones y complejidades sociales múltiples, más allá de los constantes estudios que desde los 70 se vienen alumbrando sobre el contexto de los EE.UU (Abramowitz y Saunders, 2008; McCright et al., 2014). Por tanto, pretendemos incorporar y desarrollar en el debate, cuestiones acerca de cómo poder graduar cuándo una sociedad ha sido o está inmersa en un proceso de polarización.

A día de hoy, a causa de la expansión y adopción de nuevos paradigmas como el *Big Data* (Boyd y Crawford, 2011), podemos tener acceso a infinidad de datos estructurados¹ que hasta hace poco eran impensables (Maté Jiménez, 2014). En

¹ Los datos estructurados son aquellos que albergan información dividida y esquematizada en variables, con filas y columnas tales como fecha o edad. Mientras que los datos desestructurados, aún pudiendo contener esta arquitectura, son aquellos otros que, por su naturaleza y recogida, incluyen principalmente texto libre de, entre

consecuencia, y gracias a las enormes dotes que esta tecnificación aporta, se estipula que podemos dar cabida a investigaciones más nítidas y empíricas sobre una amplia multitud de campos. De ese modo, se baraja la idea de que, debido a estas contribuciones, y los que más tarde se expondrán acerca del entendimiento del lenguaje humano por medio de máquina, seremos capaces de situar el distanciamiento semántico, así como el enmarcado ideológico de los clústeres para percibir posibles fisuras sociales. Por lo tanto, nuestro aporte hace hincapié en que el empleo de distintos significados para referirse a un mismo significante da origen a posturas adversas a la hora de encarar una cuestión. Por este motivo, se busca aclarar cómo la dispersión semántica genera posturas y constructos que son utilizados como arma arrojadiza y que pueden ser tomados en consideración para estudiar la polarización.

Para poder enjuiciar tales distanciamientos, se decide poner bajo la lupa fenómenos que pudieran producir picos o brechas entre grupos. De esta manera, el feminismo, debido a su amplio prisma sistémico y la creación paulatina de un argot propio (Cobo, 2012; Cobo y Ranea, 2020), nos suministra las provisiones pertinentes para poder chequear cómo los cimientos de las variadas ópticas se asientan en el plano español. Para llevar a cabo tal diagnóstico de la cuestión, y su correspondiente reflexión, se analizarán distintos ámbitos de la esfera pública y de la sociedad de masas: redes sociales, prensa escrita, discursos políticos... Siendo todos ellos sondeados mediante nuevas técnicas computacionales.

Por consiguiente, partiremos de una cuestión genérica:

¿El feminismo en España se ha erigido como un fenómeno polarizante?

Esta idea reporta una posición multifactorial y taxonómica que se inicia con el estudio de la acción humana como modo de examinar al conjunto. Por lo que tanto la parte micro, como la macro, estarán influenciadas por la variabilidad temporal y estocástica. Por lo cual, las preguntas suscitadas son múltiples: ¿Qué es la polarización? ¿Cómo se mide? ¿Es el distanciamiento semántico un predictor de la polarización? ¿Es el feminismo un fenómeno polarizante? ¿Feminismo, ¿agregador o disgregador? ¿Cómo se ve afectado el feminismo por los nuevos fenómenos de masas? ¿Qué papel juegan los medios de comunicación de masas? Quizás, y aquí arranca este supuesto, las respuestas a estos y otros interrogantes podrían ser resueltos desde el examen de los procesos de expresión y comunicación congregados en el ágora digital. Por lo que, en busca de complejidades más precisas, exponemos:

H.1. Las bases del feminismo en España se configuran en las últimas dos décadas, aunque su mayor explosión cobra fuerza a partir del último lustro.

H.2. El tratamiento mediático del feminismo en España está enmarcado bajo los conceptos de las tesis de género, lo que a su vez no supone una clara polarización entre medios de comunicación.

Objetivos de la Tesis

Tal y como se señala en el título y la introducción, esta investigación tiene un conjunto de núcleos bien definidos: medios de comunicación, polarización, feminismo y procesamiento del lenguaje natural. Así, nuestra finalidad se basa en la combinación de estos para, de forma sumarial, exponer todos los conocimientos alcanzados durante este periplo y dar forma a un trabajo genuino. Este partirá de lo expuesto previamente; el porqué las sociedades se dividen y polarizan. El tema que hemos asumido para afrontar esta dispersión de posturas ha sido el feminismo y el lugar en donde poder encontrar tal fragmentación han sido los medios de comunicación. De este modo, además del conocimiento teórico de las anteriores líneas de investigación, también asumimos el uso del procesamiento del lenguaje natural como un modo con el que poder estudiar y aportar, de un modo más amplio, todo este abanico de estadios.

Por lo tanto, y de forma esquematizada, señalamos que los cuatro objetivos principales de este trabajo son: primero, entender qué es la polarización política; segundo, comprender qué papel juega el movimiento feminista en el ámbito público; tercero, analizar cómo estos fenómenos sociales bullen dentro de los medios de comunicación de masas; cuarto, contribuir a la expansión del aprendizaje de máquina dentro de la ciencia política

Entre los objetivos particulares que posee este proyecto, destacamos:

- Plantear un nuevo paradigma que sea capaz de calibrar la dispersión como eje para contrastar la polarización.
- Examinar y dar cobertura a nuevos enfoques metodológicos que, irradiados del campo de la computación y la lingüística (a saber, Natural Language Processing), provean la presente y futuras aportaciones científicas.
- Aplicar los conocimientos adquiridos al análisis del feminismo en España.
- Conocer cómo es el feminismo que marca la escena pública española.

2. Marco teórico

A lo largo de las siguientes páginas vamos a dar forma al armazón teórico que sustentará nuestros estudios y que, a cambio, nos dotará de un esquema común. Por lo cual, partiremos de una explicación pormenorizada acerca de qué son o qué entendemos por medios de comunicación y su relación con la esfera política, pasando después a desglosar determinadas aproximaciones sobre qué es la polarización, y continuando con unas breves acepciones históricas y doctrinales sobre el feminismo. Este último punto se aborda desde una perspectiva genérica, sin entrar en juicios de valor o adscribirse a ninguna corriente en concreto y con el fin de que el lector pueda interpretar un campo tan mutable. En última instancia, y con la aspiración de facilitar una contribución sobre la materia, también se proporciona un breve repaso sobre las ramificaciones computacionales empleadas en nuestro proyecto. En resumidas cuentas, proveemos de un entorno para que la utilización de determinado argot no lleve al equívoco, y tanto investigador, como los posibles interesados en esta aportación, hablen el mismo lenguaje.

2.1. La comunicación (política)

La comunicación política es un amplio surtido de teorías, conceptualizaciones y nociones que emanan en su origen de un eje tan grueso como el de las ciencias sociales. De ahí nace su precepto actual, derivado de la fusión de disciplinas que van desde la ciencia de la información, pasando por el marketing o, de forma cada vez más acuciante, la psicología o las neurociencias (Luengo, 2016). Algunos autores, tratando de canalizar esta abundancia, y con el fin de congregiar conceptos comunes, enuncian que la comunicación política se entiende como el nexo limítrofe entre partidos políticos, ciudadanos y medios de comunicación (Norris, 1997; López-García, 2017). Este cruce de caminos aguarda en su seno el estudio de una extensa panoplia de recursos lingüísticos, lógicas mediáticas, metodologías y prospecciones científicas que tratan de dilucidar las vías por las que la comunicación se efectúa (McNair, 2003). Con todo ello, obtenemos a una disciplina ocupada de la investigación y reflexión que la difusión de la información, a través de los medios masivos, puede producir en los individuos y sus relaciones interpersonales. Además, y de manera crucial, del análisis e implementación que tanto investigadores, como medios y esfera política, hacen de la variabilidad de técnicas y métodos útiles para tratar de repercutir en las acciones del conjunto.

A pesar de esta abrupta aproximación al concepto de comunicación política, lo cierto es que la construcción de su término se viene elaborando desde el final de la II Guerra Mundial, considerando a la capacidad y al poder ejercido por la comunicación como la piedra angular que vertebra todas estas intersecciones. Siendo todas estas las múltiples perspectivas que se aúnan acerca de los diversos modos de encarar la cosa pública (Canel, 1999a). Algo que también se cerciora en factores agregadores que los medios masivos aportan a la cosa, o causa, política; como sería la capacidad conformadora que los medios tienen acerca de lo

entendido por sociedad o cultura (Viswanath et al., 2007), y que da lugar a estudios de comunicación política desde una visión comparada (Canel y Voltmer, 2014).

En tal sentido, lo que se pone de relieve es cómo la coligación de fuerzas entre procesos de comunicación y política abren el espectro para modelar nuevos fundamentos en donde los medios de comunicación favorecen un nuevo conjunto de sinergias (Curran, 2002; Luengo, 2005). Estas propician una hibridación de nuevas praxis entre lo habido y lo contado, entre lo formal y lo mediático (Aguirre Romero, 2009), dando paso a una espectacularización más pensada para captar una audiencia que en solventar las fallas de la cosa pública. Surge así una fusión entre el papel que juegan los medios para transformar lo público, y la astucia con la que la esfera política asume esas reglas con el mismo fin. Asistimos a una confluencia entre dos mundos que colisionan y producen nuevos nichos de estudio o fenómenos como los de mediatización (Mazzoleni y Schulz, 1999; Mazzoleni, 2008; Strömbäck, 2008; Esser y Strömbäck, 2014) o americanización (Negrine y Papathanassopoulos, 1996; Schulz, 1998).

Por tanto, y tal y como ya preconizaba Luengo (2016), lo que tenemos entre manos es un área de estudio divergente y en constante cambio, apoyada en la multidisciplinariedad como eje para la edificación de nuevas sendas de tránsito teórico. Habida cuenta de ello, no paran de ser incorporadas nuevas nociones y técnicas (véase, del filtro burbuja (Pariser, 2017) al deep fake (Kietzmann et al., 2020)) que deben de ser estudiadas desde ámbitos cada vez más amplios, yendo desde la psicología a las neurociencias y pasando por la pujanza que la ciencia de la computación aporta.

2.1.1. Medios de comunicación de masas

Si echáramos la vista atrás podríamos comprobar cómo el desarrollo de los métodos de comunicación ha ido incorporando una múltiple heterogeneidad de disciplinas a su desarrollo. Una gran prueba de ello sería lo denominado como TICs. Técnicas que, desde la invención de la imprenta (Cazorla Pérez, 2008), han ido aportando las piedras para que la difusión en masa tenga lugar.

Al hilo del anterior planteamiento, y con la intención de sellar un término que determine nuestra posición sobre qué entendemos por medios de comunicación de masas, definimos a estos como canales por los que, sirviéndose de las tecnologías para una reproducción masiva, podrían alcanzar múltiples escenarios, espectadores, oyentes o lectores. Esta asunción propia bebe directamente de las pinceladas que un dilatado compendio de autores ha dilucidado (a saber, McLuhan, 1959; Spitulnik, 1993; Turow, 2011 o Campbell et al., 2017 entre otros destacados académicos). No obstante, conviene precisar que los estudios de comunicación se han dirigido en su mayoría, y con especial atención desde la segunda mitad del siglo XX, a la revisión concreta de unos determinados: la prensa, el cine, la radio o la TV (López-García, 2005).

Todos estos análisis han incentivado reflexiones que han servido para que la comunicación genere unas lógicas mediáticas (Altheide, 2016) que en función del medio (McLuhan, 1965) produzcan distintas arquitecturas a través de las cuales mediatizar el orden social (Altheide, 2013). Lo que no cabe duda es que el desarrollo de las tecnologías de la información favorece un terreno comunicacional dispuesto para insertar subgéneros y categorías dependiendo del modo en que se asuma la acción comunicativa. En cambio, en la actualidad, la cuestión entraña otros ejes; como la retroalimentación y la facilitación que una invención como Internet ha sido capaz de aportar a la expansión de lo que entenderíamos por comunicarse de forma masiva.

La anterior disquisición apunta a cómo los flujos de comunicación de masas clásicos, emisor, receptor y mensaje², se han visto alterados de forma ambivalente en el mundo digital. O lo que es lo mismo, que tanto emisor como receptor pueden intercambiarse los roles. La virtud de esta nueva tesitura se debe a las dotes de Internet. A diferencia de los, ya hoy, viejos medios (Schultz, 2000; Manovich, 2013; Scrivens y Conway, 2020) –o medios tradicionales–, el actual estado de los medios derivados de la red supone unos réditos poco vistos hasta ahora. Esto ha de agradecerse a los protocolos y arquitecturas (Berners-Lee, 2000) conferidos para la causa, capaces de tejer los nodos necesarios para constituir las edificaciones de lo que podría denominarse aldea global (McLuhan, 1969). Todo ello incorpora un factor clave, la paulatina disminución del valor de los costes de producción de estas redes ha propiciado que cualquier usuario pueda tener la capacidad de reproducir masivamente su contenido.

La génesis de esta sociedad mediada, y que perfila lo que hoy día conocemos como plataformas sociales, redes o medios sociales (Carr y Hayes, 2015), aflora de un razonamiento basado en la idea de comunicación por canales deslocalizados que implosionaron hace solo unas décadas (Leiner et al., 2017). Hablamos de la evolución (tómese Stevenson (2018) como bitácora de las tecnologías que han protagonizado este soberano cambio en la forma de relacionarnos y acceder a la información) que supone el pasar de lo que se denominó web 1.0, a su primer gran salto cualitativo y exponencial, la web 2.0 (O'Reilly y Batelle, 2009) –paso previo de las 3.0 y 4.0 (Latorre, 2018)–, y que ha permitido la proliferación de fenómenos de masas que no han necesitado de una difusión centralizada.

Esta breve armonización terminológica señala cómo el desarrollo de la comunicación induce a una alteración del revestimiento del mensaje y también de la interpretación de su significado. Ya que al igual que ocurrió con la primacía de la prontitud que la radio impuso sobre el papel, la preponderancia de la imagen (Druckman, 2003) hizo lo propio con la onda modulada y la media. Por lo tanto, esta nueva calibración digital –a la par convergencia– del siglo XXI apunta a una

² Aunque siguiendo, por ejemplo, las funciones expuestas por Jakobson (1960), también deberíamos contar con ambientes que están influidos por el contexto, el canal y el código, alterando de este modo la recepción, así como la emisión del propio mensaje.

nueva reconfiguración del tablero mediático, confeccionando formas y modos para que el mensaje pueda alcanzar a una masa cada vez más ingente³.

2.1.1.1. Efectos mediáticos

A este respecto, debemos prestar atención a las investigaciones que la academia ha realizado de las sociedades modernas de la segunda mitad del siglo XX, las cuales ya se encontrarían arbitradas por el impacto de los medios masivos (Bennett y Segerberg, 2014). De esta forma, esa omnipresencia que éstos sostienen (Aupers, 2020), ha supuesto un conjunto de aportes y retos tanto para las democracias liberales como para una parte de los académicos (Bennett y Entman, 2001; Gurevitch et al., 2009; Brants y Voltmer, 2011); desde la mencionada función socializadora o movilizadora (Petersen, 2020), hasta una capacidad agregadora con la que conducir a nuevas tónicas de cambio estimuladas por la fuerza de Internet (Luengo, 2014).

Esta mediatización cobra un mayor interés desde que, entre las décadas de los 40 y 60, múltiples investigadores comenzaran a acotar los estudios de un modo más empírico, dejando atrás posturas en donde los aspectos informales se alzaban como evidencia científica. Dentro de aquella nueva terna destacan los clásicos aportes de Lazarsfeld et al. (1948) o Berelson et al. (1954), así como la divergente postura de Klapper (1960). Por lo que vemos, y al igual que ocurre actualmente, existen distintos enclaves y procedimientos para diagnosticar los efectos que los medios concentran. No obstante, la evolución paulatina del papel que los medios masivos juegan sí ha conseguido aportar y concordar unas reflexiones, dejes y estructuras alejadas de posturas maniqueístas. Además, esto ha servido para aglutinar características comunes, útiles para entender determinados fenómenos sociales y mediáticos, y sin pasar por alto la extensa matriz multifactorial que incide en las decisiones humanas (McQuail, 2000).

A este parecer, el paradigma que se destaca por encima de todos es el consumo de un mismo producto por una gran cantidad de personas e irradiado por una o varias fuentes. Este hecho, unido a la construcción del relato, la imprimación de una realidad social o la exposición selectiva (Frey, 1986) en busca de la confirmación del propio yo (Nickerson, 1998; Groseclose y Milyo, 2005; Gentzkow y Shapiro, 2006), da lugar a un abanico de escenarios marcados por el individuo y los contextos.

A fin de cuentas, esta posición del consumidor de medios se erige bajo otros supuestos psicológicos como el de no querer enfrentarnos a opiniones que desafíen nuestras conjeturas sociales, morales, políticas... (Sunstein, 2007). Por lo que, mientras que, por un lado, nos encontramos con que los medios actúan con arreglo a su forma de entender y expresar el mundo (McQuail, 1979), por otro, tenemos a los ciudadanos haciendo lo propio bajo su autonomía. Al menos eso es

³ A este respecto surgen nuevos alumbramientos que divagan acerca de lo que provoca el atomizar por completo la difusión de la información (véase Turov (2011) para una sutil aproximación al planteamiento de diversificación mediática).

lo que también se destaca de propuestas teóricas como los modelos de probabilidades (Petty et al., 2009) o aquellos trabajos que arguyen una causalidad multifactorial marcada por la posible susceptibilidad de los afectados (Valkenburg y Peter, 2013).

Bajo esta perspectiva surge no una tesis fija acerca del papel que podrían efectuar los medios en el conjunto de los receptores, sino una amalgama de evaluaciones focalizadas en las diferentes técnicas que los medios adoptan para cimentar sus miradas de la realidad. En este particular, señalamos lo que tuvo en denominarse como “teorías de la negación” (García-Marín, 2007), una sucesión de criterios de investigación en los que se prepondera el examen de las estrategias con que los medios plasman sus posiciones a la hora de difundir las noticias (McQuail, 2010). Bajo este paraguas crecen, en una primera instancia, las primitivas teorías de fijación de la agenda, *agenda-setting* (véase McCombs y Show (1972, 1993) como padres del término). Estas se establecieron como eje de un paquete de teorías más exhaustivas –denominadas de segundo nivel– que desde finales de los 90 lideran el estudio de medios (Weaver, 2007). Las teorías de *priming* (Iyengar y Kinder, 1987) y, con una especial significancia para este proyecto, las de *framing* (véase a Goffman (1974) y Entman (1993) como impulsores teóricos en los estudios de comunicación) o una herramienta fagocitada (Chyi y McCombs, 2004; Schildkraut y Muschert, 2014) por estas, la *media salience* (Kiousis, 2004), aportan los sustentos doctrinales con los cercar los modos en que los medios entienden y proyectan la realidad.

Además, y tal y como se glosó en el apartado terminológico de medios de comunicación, las redes sociales son en la actualidad las que acaparan un mayor número de investigaciones en el ámbito político comunicativo. Aupados por determinados fenómenos políticos de los últimos lustros (véase, sobre el particular, el trabajo de Kramer et al. (2014), así como otro posterior de Matz et al. (2017)), estos nuevos estudios, al igual que sus antecesores, también agregan nuevas nomenclaturas para tratar de entender complejos acontecimientos. En tal sentido, e influenciados de nuevo por la recapitulación de las lógicas mediáticas, brotan significantes como *fake news* (Quandt et al., 2021) o *clickbait* (Zannettou et al., 2018), así como otros nuevos surgidos del diseño algorítmico, como, por ejemplo; exponentes *offline* refundados⁴; el de cámara de eco (Scruggs, 1998; Cinelli et al., 2021) o planteamientos más novedosos como el de filtro burbuja (Pariser, 2017). Sin olvidar factores interesantes que poco a poco resquebrajan premisas propias de los viejos medios; como sería el de la individualización del consumo⁵, la pérdida en determinadas instancias de un *gatekeeper* de entrada (Bro

⁴ La idea de refundación o recapitulación *offline* es deudora en gran medida del trabajo llevado a cabo por Guess et al. (2018), quienes imputan de nuevo a la exposición selectiva, la motivación, los contextos, los amigos o familiares, un papel más acuciante que el que pudieran tener las nuevas lógicas algorítmicas, o como ellos tildan de forma brillante e irónica en el título de su trabajo: *Avoiding the echo chamber about echo chambers* [Evitar la cámara de eco de las cámaras de eco].

⁵ Spiller (2019) contribuye con un ejemplo muy evocador. Sugiere que, a diferencia de lo que ocurre en la actualidad, el consumo de información hace solo unas décadas estaba prefijado por el noticiario televisivo que, entre familias o compañeros de piso, se hacía en su mayoría a las horas de comer. Aun pudiendo saltar chispas de aquella exposición de motivos o reflexiones al respecto, lo cierto es que entraban en juego múltiples factores

y Wallberg, 2014) o las suspicacias concatenadas en la expresión digital - *affordances* (Hayes et al., 2016; Rojas Mesa y Leal Urueña, 2017)-. Circunstancias todas ellas que aún deben de ser puestas de forma más sistemática y congruente bajo la lupa, ya que en el marco de recientes estudios estas parecen seguir un camino más próximo a ejercicios para la elaboración del relato y la mercadotecnia -en especial el *clickbait* o las *deep fakes*⁶- que para un influjo realmente masivo y consistente capaz de alterar comportamientos de forma sustancial (tómense Coppock et al., 2020; Munger et al., 2020; De Francisci, 2021; Yu et al., 2021 como algunos de los recientes aportes desmitificadores acerca de las cuestiones expuestas).

2.2. La polarización

Hoy en día, los estudios y concreción de la polarización política son deudores de las singularidades americanas (téngase a Iyengar (2021) como una de las últimas pruebas de ello), por otra parte, como suele ocurrir en los estudios de esta área. Tal que así, y empleando como punto de partida la brecha política americana⁷, su bagaje sirve para tratar de entender de qué hablamos cuando hablamos de polarización. Por ello, y con arreglo a la cuestión de por qué las sociedades se distancian, los investigadores sociales han cercado su interés en algunas esencias clave. Eso sí, con una marcada atención desde mediados de los 70 (Sartori, 1976).

Es evidente que estamos ante un término pujante (Hernández et al., 2021). Por ello, su adscripción y definición, en ocasiones difusa, ha de ser expuesta de la forma más clara. Aunque debamos asumir que, como la mayoría de términos, este se encuentre perturbado por una gran cantidad de contextos y singularidades. Por este motivo, nos gustaría partir de una piedra de toque que nos permita calibrar y acompañar todos estos amplios escenarios. O sea, hablar de polarización es, llanamente, hablar de la acción de polarizar. Según la propia RAE, polarizar⁸, desde su variedad de acepciones, orienta sus nociones al hecho de que existen posturas contrarias que, llevadas a su gerundio -polarizando-, estarían tomando direcciones contrapuestas pero que, tomadas desde su participio -polarizado-, ya habrían alcanzado un estado de discrepancia (Di Maggio et al., 1996). De este

que marcaban una mayor diversificación que la de la era actual. Debido, primordialmente, a tener que asumir o alternar el canal, así como repensar o generar argumentos para defender una postura. Algo que con el auge del internet está terminando por perderse, eso sí, sin tener en cuenta los posibles desvarios anonimizados que segregan las redes sociales.

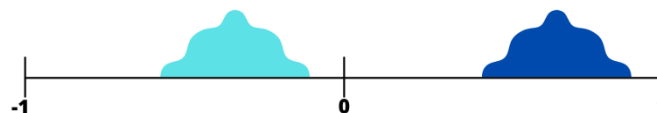
⁶ Kietzmann et al. (2020) destilan un concienzudo e ilustrativo diagnóstico para entender el término.

⁷ Jonathan Haidt (2019), en su libro *La mente de los justos*, disecciona, desde diferentes y enconadas esferas, algunas de las conjeturas morales - ¿Y las genéticas? (Alford et al., 2005)- de por qué la religión o la política dividen a las sociedades. Un ejemplo paradigmático a este respecto, y que nos permite encarar esta ardua disyuntiva, señala una anécdota -cuanto menos curiosa- acerca de la política americana. Haidt, de boca de Jim Leach -excongresista republicano-, detalla cómo, el por aquel entonces presidente de la Cámara de Representantes, Newt Gingrich, incentivó para que los republicanos electos en Washington, no se mudaran a la ciudad con sus familias, sino que viajaran hasta la capital de lunes a jueves, y ese mismo día regresaran a casa. Así se evitarían relaciones que no fueran del ámbito laboral, lo que también limitaría las prácticas que no estuvieran relacionadas con el fragor político; que las parejas de rivales pudieran entablar amistad; o que sus hijos pudieran compartir colegios y equipos deportivos. A este respecto también puede tomarse Darmofal y Strickler (2019) como revisión histórico-geográfica de la división bicolor estadounidense.

⁸ De hecho, si atendemos al origen de su raíz, denotamos que polarizar es llanamente la comprensión de sus dos léxicos latinos: *polus* (polo) e *izare* (convertir en) (Dechile, 2021)

manera, y con arreglo a lo expuesto, la polarización ha sido asumida en el campo de la ciencia política –pero también en el de otros ámbitos de las ciencias sociales y del estudio de la psicología humana–, como un proceso o un estado en el que los individuos asumen o se hallan en posturas divergentes. Algo que, inserto en una distribución espacial, podría tomar unas valencias definidas en cuanto a un eje X donde 0 retribuye al punto intermedio y sus estandarizados contrapuestos estarían en [-1:1].

Figura 1. La polarización



Fuente: Elaboración Propia

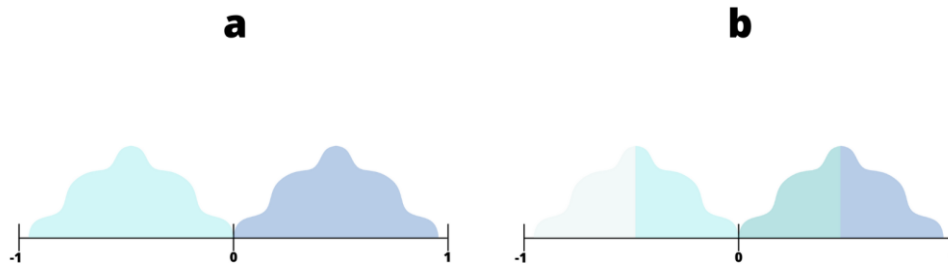
Ciertamente, y aún simbolizando de forma patente cómo se puede vislumbrar esa dispersión entre polos, la polarización puede oscilar entre distintas distribuciones; tanto espaciales como cuantitativas. A nuestro juicio, los trabajos de, por un lado, Esteban y Ray (1994) –aunque este desde una perspectiva económica–, Bramson et al. (2017) por otro, así como la exposición dicotómica de los años 80 de Poole y Rosenthal (Poole et al., 2008), aportan unos interesantes enfoques para comprender de primera mano qué se oculta tras este fenómeno. Ya que, si diseccionamos las hasta nueve tipologías de Bramson et al. (2017), podremos cerciorar cómo las posturas polares aguardan en su seno complejidades que van mucho más allá de la simple espacialidad entre los polos (+) y (-).

Figura 2. Modelo Esteban y Ray (1994)



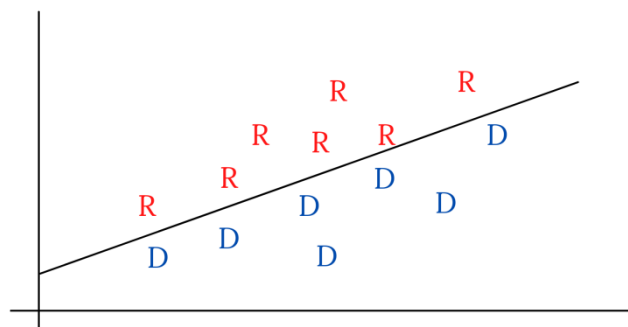
Fuente: Elaboración Propia

Figura 3. Bramson et al. (2017); Tipo 5: Comunidad fracturada



Fuente: Elaboración Propia

Figura 4. Poole y Rosenthal



Fuente: Elaboración Propia

Habiendo ya instaurado unos supuestos básicos de qué se entiende por polarización, nos toca asumir la postura de atestiguar por qué estas cuestiones han sido objeto de interés para los investigadores. En armonía con los diferentes trabajos, lo que se constituye del todo evidente es por un lado el interés de por qué las sociedades se distancian en polos, mientras que la otra preocupación de los investigadores políticos es cómo ese hecho puede minar a las democracias liberales (Boxell et al., 2020; Carothers y O'Donohue, 2020); y, por supuesto, afectar a circunstancias de interés politológicas como las fluctuaciones del voto (Abramowitz y Saunders, 1998; McCarty et al., 2009). Con todo ello, los académicos, en busca de captar estos distanciamientos, han establecido una amalgama de formas y métodos con las que medir la polarización en los múltiples escenarios donde esta se pudiera encontrar. Bien fuera en el estudio de los sistemas de partidos (Taylor y Herman, 1971; Wagner, 2021), en las divergencias acerca de las manifestaciones de corte ideológico (Lelkes, 2016), la exposición selectiva a determinados medios de comunicación (Baum y Groeling, 2008; Lodola y Kitzberger, 2017), así como interesantes aportes donde poder enjuiciar la polaridad y el distanciamiento ideológico habido en los textos (Slapin y Proksch, 2008; Goet, 2019; Serrano-Contreras et al., 2020; García-Marín, 2021). Todos ellos revelan contribuciones heterogéneas en donde lo que importa es la variable que se quiere analizar.

A este respecto, habría que tener en cuenta –aunque sin llegar a profundizar en ellos porque alejarían las miras de este trabajo– algunos factores que incidirían en ese distanciamiento ideológico⁹. En nuestro caso, el foco se centra en el papel de los medios masivos. Rescatando de nuevo a Bennett y Segerberg (2014), los medios de masas trazan un rasgo diferenciador con respecto a sociedades contemporáneas precedentes. Por lo que ha generado investigaciones que se focalizan desde los estudios de los medios tradicionales (Bennett e Iyengar, 2008; Campante y Hojman, 2013; Prior, 2013) hasta las actuales redes sociales (Tucker et al., 2018; Kearney, 2019) o sus constructos algorítmicos (Cho et al., 2020), imponiendo una mirada de interés sobre las posibles incidencias que estos pudieran ocasionar en la fractura de la sociedad.

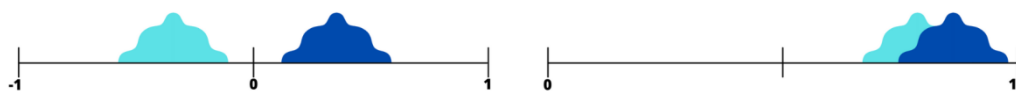
2.2.1. Breve tipología; de la postura ideológica al sentimiento de pertenencia

Una vez se ha abonado la idea de qué se entiende por polarización, es de recibo señalar una pieza clave cuando se aborda el término desde la mirada politológica: su tipología. Aún pudiendo encontrar divagaciones complejas sobre cómo exponer estos términos, quizás la forma más somera de comenzar es poder saber si este fenómeno es o no perjudicial para la salud democrática de las naciones. Esta idea es la que resplandece cuando saltamos de la noción ideológica a la afectiva; siendo ambas las que marcan las principales exposiciones sobre los tipos. Incluso, si se cambia el marco, podría efectuarse esta mirada desde una de tipo binario; polarización de las élites y polarización de las masas (McCarty, 2019). Aunque en el fondo, lo que se defiende con estos estudios y diferenciaciones (Mason, 2015), es que el distanciamiento ideológico no tiene por qué lastrar el llegar a acuerdos y, a la par, no tendría por qué minar la convivencia democrática. Simplemente sería perspectivismo ideológico a la hora de encarar el devenir de la sociedad. Por lo que habría que centrar la misión de la investigación, los términos y las formas para tratar de atajar de un modo más sustancial y empírico este dilema.

A diferencia de la distribución de la Figura 1 del epígrafe anterior, cuando cambiamos la ideología por el sentimiento de pertenencia, la representación espacial también varía.

⁹ Véanse, a modo de breve sumario, las posturas socioeconómicas conjugadas por McCarty et al. (2016); Arrigada et al. (2010), Suárez-Cao (2013) o Baker et al. (2020), así como la reflexión planteada por Sigelman y Yough (1978) acerca del difuminado de la ideología por estos supuestos.

Figura 5. Polarización afectiva¹⁰



Fuente: Elaboración Propia

Como bien apuntan Iyengar et al. (2012, 2019), no estábamos hablando de ideología –o al menos no solo (Lelkes, 2018)–, estamos hablando de afecto. Una afectividad en el sentido más irracional del término, expuesta bajo una óptica maniquea más cercana a las preferencias deportivas por un determinado equipo –e incluso por una postura científica sobre un hecho en cuestión (Rekker, 2021)–, que a la exposición sosegada del porqué una persona considera que su forma de afrontar una disyuntiva social es mejor que otra. De hecho, al encarar el problema desde el ángulo más parcial, se constata que se termina por recurrir a una falacia *ad hominem* para desprestigiar tanto al argumento como al que lo esgrime, pudiendo llegar a la aversión (Mason, 2018) e, incluso, a la deshumanización (Harel et al., 2020). Y sin olvidar otra variable a tener en cuenta, la consideración de nuestro par, independientemente de su postura y simplemente por continuar bajo la adscripción común al mismo grupo (Iyengar y Westwood, 2015; McWilliams, 2021). Por lo que se consustanciaria otro aspecto evidente; el sí estaríamos ante una divergencia o bien ante un alineamiento o radicalización (Lelkes, 2016; Almagro y Villanueva, 2021), como mostramos en la Figura 5. Por añadidura, se inserta a su vez otro ámbito clave para abordar la polarización –muy bien apuntado por Druckman y Levendusky (2019) y Knudsen (2020)–; que el lineamiento afectivo dependerá de qué tomemos como variable modificante si, por ejemplo, un partido político o los simpatizantes de este.

A la luz de estas posiciones nos encontramos que la simple disentía ideológica encierra una necesidad de pertenencia social que se ve marcada por el tribalismo, que es el que prende las discrepancias entre partes (Hawkins et al., 2018). Por lo que, expuesto de forma más clara, y a modo de ejemplo, quizás nuestra postura se encuentre, dentro de un eje X, ubicada en el centro del espectro ideológico, lo cual, a priori, no tendría por qué insertarnos en una deriva de polarización ideológica. Pero, con arreglo a los estudios y disquisiciones terminológicas afectivas, el tipo de polarización habida está construida al amparo de la postura cerril para defender que el centro del espectro es el *único*, fundamental e indispensable modo para sacar del atolladero a la sociedad, o sea, extremo centro. Por lo que estaríamos ante asunciones dogmáticas, basadas en las creencias y subrayadas por nuestra parte más irracional. Lo que podría generar, a la postre,

¹⁰ Como se aprecia en el diseño de la Figura 5, si bien la polarización ideológica (aquí representada como $\epsilon[-1:1]$) puede mostrar posiciones cercanas entre polos, la polarización afectiva señala unos lineamientos de estos muy cercanos al máximo. Por lo que la defensa de sus posturas será, como referenciamos, más cerril e irracional, basado en la pertenencia y defensa del grupo.

una mayor dificultad para encarar determinados acuerdos. Todo ello marcado mediante la lupa del individuo por, entre otros factores: un sentimiento de pertenencia a la tribu (Clark et al., 2019); las identidades nacionales (Torcal y Martini, 2013; Wojcieszak y Garret, 2018); la primacía intragrupo (Tajfel y Turner, 1979; Devine, 2015); o el papel que juega el lazo de proximidad (Granovetter, 1973) en la toma de decisiones que se emprende frente a un dilema político o social. Supuestos todos ellos influenciados por, como esgrime Pinker (2012), el preeminente papel de los factores biológicos y genéticos (Alexander, 1982), y su correspondiente determinismo biológico (Pinker, 2008). Factores a tener en cuenta por los investigadores que tratan de arrojar luz a la cuestión de la polarización social y política, derivada esta desde una parte micro y hasta conformar un todo.

2.2.2. Sistemas mediáticos y lineamientos ideológicos; recapitulando el caso español

Ante esta última idea, y al añadir el papel que juegan los medios de masas, nos encontramos ante otro de los pilares que buscan sostener nuestras posiciones. Ya que, desde nuestra asunción y la de los sustentos teóricos que aportamos, la polarización es una característica que puede ser engendrada por estos métodos de difusión masiva cuando retratan a las sociedades. De ahí que, en nuestra particular aportación, de lo que nos sirvamos es de entender si las diferentes posturas que cada medio utiliza, influyen en la masa para distanciarla.

Partiendo de la anterior exposición, lo que este subepígrafe detenta, es una exposición acerca de qué surge a la hora de sumar medios de masas, poder político, sociedad y, por consiguiente, las distribuciones polares existentes. Asistimos pues a una serie de lineamientos ideológicos en el que los medios, mediante su forma de presentar la sociedad y seleccionar lo noticioso (Baron, 2006), poseen, al menos de forma descriptiva, una capacidad para congregar determinadas perspectivas políticas. Ya que su importancia después de todo se basa en que los medios de comunicación son la principal fuente de transmisión de opiniones públicas (Stroud, 2010).

En consecuencia, los medios de masas tradicionales¹¹, a la luz de los estudios, han sido capaces de crear determinados frentes mediáticos y, por tanto, ser un modo mediante el que encarar las temáticas. Por ejemplo, siguiendo a Iyengar y Hahn (2009), estos apuntan a que la adscripción ideológica de los ciudadanos influye en la elección de un canal para informarse, siendo este elegido en función de las creencias e ideologías de los espectadores (Lazarsfeld et al., 1960). Esto, unido a Kull et al. (2003), ocasiona que el entendimiento de determinados sucesos mediáticos, como fue en su día la posible tenencia de armas de destrucción masiva

¹¹O complejas influencias de nuevo cuño, como podrían ser las de corte algorítmico configuradas en las redes sociales (Spohr, 2017).

por parte de Sadam Hussein, se viera marcado por el noticiero que se había consumido.

Alumbrados por esta mezcla entre partes y sus posibles inscripciones políticas y mediáticas, nos apoyamos en esta ocasión en una obra matriz para continuar describiendo una tendencia particular. El aporte de Hallin y Mancini (2004) ha servido, para gran parte de los investigadores, como punto de partida para refrendar, apoyar o refutar determinados supuestos. En este particular, España juega un papel muy importante en la comprensión de los sistemas de comunicación democráticos. Esta idea insertaría a nuestro país en un sistema político/mediático denominado de pluralismo polarizado. Algo que, en línea con lo determinado por Hallin y Mancini, estaría marcado por unos medios armonizados con los respectivos partidos de su misma línea editorial y en donde la primacía del papel del estado se vuelve fundamental. Una situación que en el caso español ha estado claramente marcada por la capilaridad engendrada a lo largo de 40 años de dictadura (Strömbäck y Luengo, 2008), así como, a posteriori, por una intrusión de los principales partidos políticos –PSOE y PP– en la formación de definidas configuraciones mediáticas empresariales (de Mateo, 2004).

Esta línea de actuación integrada en el sistema de medios español segrega diferentes perspectivas desde las que retratar las situaciones que acaecen en el país y, en especial, las ligadas a los agentes políticos. Quizás, los casos más paradigmáticos estarían en la prensa escrita. Aunque también se debe tener en cuenta que algunos de ellos, como ABC o *El País*, llevan más tiempo presentes en la sociedad, a diferencia de los medios televisivos creados al calor de la liberalización del sistema. De hecho, estos medios, junto con *El Mundo*, son los considerados hegemónicos de la prensa española. Principalmente por tener su sede en la capital, además de un gran arraigo y cobertura. Si bien, dentro de esta terna también entrarían otras cabeceras clásicas, con sede central en Barcelona, como serían *La Vanguardia* o *El Periódico de Cataluña*. Esta idea linda de forma manifiesta con la tesis del pluralismo polarizado. Como refrenda Canel (1999b), los tres medios – en referencia a ABC, *El País* y *El Mundo*– hacen una visión partidista y situada en determinados polos del espectro por lo que, dependiendo del periódico, estos presentan las noticias en función de sus lineamientos. Es más, siguiendo las líneas expuestas por Fletcher y Jenkins (2019) acerca de la polarización de las noticias en Europa, estos ponen de manifiesto cómo *El País* y *El Mundo* sostienen posiciones claramente alineadas con los partidos próximos a sus ideologías y defendidas de un modo afectivo. Dichos medios asumen la crítica al rival como forma de defensa de su postura ideológico-estratégica, siendo estimulada por, entre otros métodos, la del *salience media* (Baumgartner y Bonafont, 2015). Técnica que, como apuntan Helbing y Tresch (2011), va más encaminada a la preponderancia temática que a una posible alteración de la realidad acaecida.

2.3. Las olas del feminismo

El feminismo en la actualidad alberga heterogéneas nociones morales, así como difusas estructuras que no permiten construir un eje sistemático unificado. Partiendo de esta base, la meta que aquí se persigue repercute en la idea de tratar este fenómeno, movimiento y doctrina, como una recapitulación histórica del hecho en cuestión. Su posición central en el debate público, inundando múltiples esferas, lo convierten hoy en día, al menos en el mundo occidental, en un concepto *mainstream* (de Lora, 2019), en especial si repasamos la ingente literatura académica que desde hace algunos lustros se está produciendo. De ahí que sea de recibo comenzar reportando que este compartimiento estanco se encuentra mayoritariamente vertebrado en cuatro periodos, u olas, como algunos doctos en la materia defienden. Por lo que una concreción nítida se vería lastrada por la persistente mutabilidad e indefinición.

El punto de partida inicial aflora al fijar qué es el feminismo -aunque aquí entrarían juicios sobre el propio concepto de igualdad, los de hombre y mujer, y hasta los de sexo y género-. En un primer peldaño, este se basa en la búsqueda de la igualdad de derechos entre hombres y mujeres (RAE, 2021). A pesar de ello, esta escueta definición no sería capaz de captar de un modo preciso la inmensa tipología presente en la doctrina. Desde el más clásico feminismo liberal al radical, podemos encontrar una panoplia tan extensa que recogería reflexiones concretadas como postfeminismo, ecofeminismo, afrofeminismo, transfeminismo, feminismo de la igualdad, de la diferencia, marxista, abolicionista, postcolonial, lésbico, TERF, disidente...

Todas estas intersecciones pueden ser aglutinadas por medio de unas secuencias que refrendarían una evolución ubicua. De este modo, y como bien apunta León-Mejía (2010), una primera toma de contacto puede ser apuntalada gracias a lo que podrían denominarse como las primeras manifestaciones de dialéctica feminista recogidas, habidas en este caso en el seno de la religión. Por ejemplo, en la Edad Media ya se recababan posturas subversivas de mujeres que se negaban a contraer matrimonios concertados, conllevando consigo penas como el internamiento forzoso en alguna orden eclesiástica. Las condenas, según apunta León-Mejía, se convirtieron para algunas de ellas en un modo de evasión, además de una vía para reivindicarse y dar rienda suelta a sus intelectos e intereses una vez eran alejadas del yugo que las sometía.

Esta actitud contestataria abre su brecha contemporánea en algunas famosas revueltas del siglo XVIII, como la Revolución Francesa, aunque toma un cuerpo doctrinal sólido décadas más tarde a raíz de las reivindicaciones de las mujeres por conquistar el derecho a voto. Las denominadas como sufragistas (véase Walters, 2005 para un entendimiento amplio del término) son una serie de uniones femeninas, creadas a finales del siglo XIX y principios del XX, que demandaban una igualdad humanista. Esta implosión dentro del status quo fisura el orden de la época haciendo que varias de las instigadoras sean condenadas al ostracismo

(Estrada Esparza, 2017), aunque su valía y tesón aportaron para la causa las primeras piedras sobre las que construir el futuro.

Este proto-feminismo, de partida atomizado, fue poco a poco cobrando fuerza y desde distintos puntos de la esfera terrestre muchas fueron las que dieron consistencia al entramado hilvanando sus ansias de libertad. La siguiente frontera a ser derribada estaba puesta en conquistas más ambiciosas, dando forma a lo que actualmente se conoce como derechos de la mujer (Hosken, 1981). Para ello, esta es establecida como sujeto político y, por ende, sus singularidades. Se abren así los primeros debates sobre la sexualidad, el aborto, la prostitución o la tutela de los hijos (Gordon, 2002). No obstante, la irrupción de nuevos frentes, junto con la aparición de nuevas sensibilidades será lo que, poco a poco, vaya difuminando la promulgación de un concepto inequívoco en el que no será fácil determinar segmentos concretos. Al igual que ocurre en las distintas etapas de la historia, o en otras reflexiones filosóficas y políticas, el feminismo se compone de un gran abanico de enfoques, para nada monolíticos, y en donde priman posturas diversas, aunque también adversas (Schuster, 2017).

A partir del final de la Segunda Guerra Mundial brotan las primeras conjunciones de la segunda ola -marcada en cierta medida por la creación por parte de Naciones Unidas de una comisión sobre la mujer-, en la que se agregan temas sobre la igualdad de derechos, la disyuntiva sobre el matrimonio (Walters, 2005) o la institución de la familia. A su vez, la incorporación al debate de prominentes figuras o el surgimiento de otras íntimamente derivadas de la causa, además de trabajos académicos y ensayísticos, permiten perfilar sustentos teóricos sobre los que construir los puntales del movimiento. Hablamos, esencialmente, de obras como las bosquejadas por Simone de Beauvoir, que se deslindan de la tónica general y emergen con una mirada más minuciosa e incisiva sobre la disparidad y el agravio femenino.

Esta nueva cresta naciente, como señala Bartra (2020), toma conciencia de que el cuerpo femenino es la esencia anhelada por la estructura dominante. Es así como entran en juego dialécticas de corte constructivista y tesis estructuralistas como las de patriarcado (a este parecer leer las aportaciones de Hartmann (1976) y Walby (1989) para un conocimiento más completo de las tesis patriarcales). Se parte de la idea primigenia expuesta por Beauvoir en su obra, *Segundo sexo* (2019 [1949]), en la que, como reza su título, habla de la mujer como la posición opacada por el sexo dominante, el masculino (Spencer-Wood, 2017).

A partir de estas prenociencias derivadas de la acepción de cuerpo, sexo y mujer, los 60, en especial en Estados Unidos, asentarán la formulación de las pautas de las siguientes olas, debido, en cierto modo, a la creación de los primeros estudios sobre la mujer. Estos bullen en torno a la idea de Beauvoir acerca de que la mujer no nace, sino que, y haciendo una clara distinción entre lo conocido como sexo y género, la identidad se alcanza de forma escalonada.

Estamos ante un caldo de cultivo formulado por autoras como: Betty Friedan (1977 [1963]) y sus posturas acerca del concepto de la feminidad; Shulamith Firestone (1977 [1970]) que ofrece un manifiesto esencial sobre cómo encumbrar la lucha dialéctica del sexo como clase; Kate Millett (2015 [1970]) señalando a la literatura y al arte como hacedores de parte del yugo soportado por las mujeres; Germaine Greer (2012 [1970]) que describe a la mujer como castrada de su condición de ser por las convenciones sociales; Gerda Lerner y Mónica Tussell (2017 [1986]) quienes realizan una evolución histórica sobre el concepto de patriarcado o Carol Hanisch (Kelly, 2017) que pone de relieve la premisa de que lo íntimo y lo privado también es político (de Lora, 2019). Todas ellas sirven de impulso para el empleo de retóricas y posturas que tienen como tesis central la materialización de la mujer como ente político y la fundación de un colectivo unido, aunque no por ello unificado. Para ello, y como apunta Nordmann (2011), la cuestión del género comienza a consolidarse como una bisectriz desde la que afianzar nuevas corrientes de pensamiento (Schuster, 2017). Buscan revelar la dominación y subordinación vivida, atacando mitos congregados y forjados sobre la mujer. Adicionalmente, y en una línea descendente desde las disciplinas académicas creadas sobre la figura femenina, germinan nuevas raíces que apuestan no solo por conquistar el ámbito académico, sino también el espacio público.

Comprobamos cómo el espectro se va abriendo cada vez más hacia reflexiones más complejas que abordan de lleno espacios y supuestos hasta ahora poco, o difícilmente, alcanzados. Por ejemplo, la teorización sobre los estereotipos y roles concebidos para encorsetar a los individuos abre camino para profusas conjunciones que subrayan que las represiones sufridas ya no han de estar solo en la clase o el sexo, sino que el germen que anhela la verdadera revolución feminista ha de ir más allá (Firestone, 2003). Las feministas que ya salen formadas por la academia desde los 60, diseccionan y retratan la opresión como mutable y divergente dependiendo de las razas, culturas, etnias o religiosidades. De esta forma, todo este compendio florece con nuevas propuestas que traten de mostrar, y combatir, con este desamparo (Biswas, 2004). Asimismo, las teóricas de esta nueva vertiente emprenden una lucha más hondable, en la que buscan dilapidar unos esquemas exógenos aceptados por la segunda ola (Reverter, 2010).

El culmen de la crítica, que apunta a una tercera ola, y que tanta disputa supone, fue impulsado por, principalmente, dos libros de Judith Butler: *Gender Trouble* (1990) y *Bodies that Matter* (1993). En ellos se induce a la conceptualización de que la identidad es subjetiva y volátil, apartando del debate estructuras binarias y abogando por la búsqueda performativa de no alienación (Butler, 1986). Por consiguiente, se busca resquebrajar la ya entredicha postura cultural; género, pero, sobre todo, la impuesta por la ciencia biológica; sexo. Esta nueva cosmovisión feminista es sustentada por reflexiones posestructuralistas que tratan de revertir desigualdades que se dejaban de lado al asumir una superestructura agregadora que no tiene en cuenta las micro-estructuras y, por ende, las micro-opresiones multifactoriales generadas al candor del espacio íntimo.

Todo este entramado suscitó, y suscita, determinadas divergencias dentro del movimiento, aunque este pueda considerarse polifónico (Varela, 2020). Aun así, y con especial atención en los años 90 del pasado siglo, las corrientes comienzan a inundar distintos sectores sociales como la ONU, con una especial atención a la conferencia mundial de la mujer de 1995 en Pekín. En ella se adopta la elaboración de medidas que palien las diferencias de género y sustenten un programa de empoderamiento femenino (UNWOMEN, 2021). Algo que desde entonces es asumido por diferentes gobiernos como parte de sus políticas públicas. En el caso español podemos comprobarlo a partir de la primera legislatura de José Luis Rodríguez Zapatero (2004-2008), en la que se acometieron reformas con las que encarar tales compromisos internacionales; la ley de VIOGEN (2004)¹² y la de paridad (2007)¹³ son una clara prueba de ello.

Estos alentadores designios son, en parte, catapultados gracias al engranaje de nuevos mecanismos que marcan nuestra presente era. Internet, en este nuevo salto, se descubre como un catalizador crucial a la hora dar cabida a nuevos rumbos (Chamberlain, 2017). Como apuntan Charles y Wadia (2018) la tercera ola despegó gracias a las bondades del ciberespacio (Shiva y Nosrat, 2019), aunque termina por colmar su influencia en lo que empieza a conocerse como cuarta ola; donde las historias diseminadas se aglutinan -como ya propugnaban autoras de décadas previas- en una causa común.

Otro de los ejes que potencia esta nueva fase es la fuerza que aportan las nuevas generaciones. Tal y como ocurrió en el anterior tránsito entre olas, a las nuevas cohortes de edad más jóvenes adheridas al movimiento (Edelman, 2001; Heywood y Drake, 2007) se agrega el interés por dilapidar las diferencias entre autoras (Kaplan, 2003) y el papel de las redes sociales, gracias a las cuales penetrar en todas las capas de la sociedad, incluidas artistas pop y mainstream (Rivers, 2017). Estamos ante un cambio sintomático, en el que, a través del ágora digital, centenares de miles y millones de personas, se conectan para tomar las calles en un acto de músculo social que recorre numerosos rincones del globo. Asentando, de este modo, la evolución constante y persistente que se venía fraguando.

2.3.1. Sobre la cuestión de género

Debido a la importancia que juega el género y sus posturas al respecto, se emplea este subapartado a modo de expresión unificada del término dentro de las olas del feminismo. Además, no se ha de olvidar que el debate acerca de la temática da lugar a variados enfrentamientos entre distintas ramas del mundo académico¹⁴.

¹² Ley Orgánica 1/2004, de 28 de diciembre, de Medidas de Protección Integral contra la Violencia de Género.

¹³ Ley Orgánica 3/2007, de 22 de marzo, para la Igualdad Efectiva de Mujeres y Hombres.

¹⁴ Existe un extenso abanico de posturas biológicas, neurocientíficas o clínicas, que muestran sus reticencias a los actuales postulados de género cosechados por el feminismo dominante en las últimas décadas, incluso dentro de la teoría feminista como sería el feminismo disidente. En este heterogéneo conjunto se encuentran autores como Simon Baron-Cohen (2009) y sus estudios sobre la influencia del cerebro en las diferencias entre hombres y mujeres; las aportaciones a propósito la condición humana que conjuga Steven Pinker (2012); las críticas doctrinales que agrega Camille Paglia (1994; 2001) sobre la sexualidad o la violación; las posturas en cuanto a los roles de género como los enunciados por Talcott Parsons (1939); las críticas académicas que realiza Jordan

Esta disputa dialéctica y científica aglutina un amplio surtido de maneras de entender el mundo; desde el concepto de condición humana y hasta el modo en el que instrumentar el devenir de la sociedad futura. Debido a ello, son muchos autores los que construyen enconadas teorías sobre las que, quizás, obtener, o construir, una sociedad mejor. Por consiguiente, a continuación, presentamos una aproximación conceptual de lo que, en la actualidad, el feminismo hegemónico o mainstream (de Lora, 2019) –el que desde los 60 con múltiples voces marca la escena y la doctrina– entiende por género¹⁵. Siendo esta una piedra de toque, a nuestro juicio, fundamental para comprender y calibrar los distintos advenimientos que el feminismo ha congregado y que, por tanto, ha introducido en la esfera pública.

Para afrontar de forma inicial esta disquisición, lo que proponemos, como punto de partida, es un abreviado significado desde el que constituir este marco. Por ello, y según la RAE, género hace referencia a:

“Grupo al que pertenecen los seres humanos de cada sexo, entendiendo este desde un punto de vista sociocultural en lugar de exclusivamente biológico.”

Esta conjetura teórica recrea algunas ideas básicas. Primero, la de que el género son diferenciaciones que se manifiestan más allá de las concepciones biológicas. Y segundo, que estos hechos se tornan presentes debido a estructuras socioculturales que improntan en los individuos el género al que pertenecen o han de pertenecer. Prosigue en la misma línea que organizaciones internacionales como la OMS, que señala que es una característica relacional influenciada por un constructo social prefijado en base a comportamientos, roles y normas (WHO, 2021). Lo que, influenciado también por la época y el ámbito geográfico, produce discrepancias y desigualdades. En consecuencia, la persona, independientemente de sus cromosomas, hormonas o sexo biológico, se siente, por influjo sistémico, contaminada por los estereotipos que a la postre repercuten en su devenir (Hyde et al., 2019).

Sobre la base de una cronología, y mediante una evaluación general de la temática en cuestión, aducimos determinados trabajos claves para asentar, desde diferentes enfoques ensayísticos, clínicos y antropológicos, la actual terminología de género empleada por las feministas. En esta serie convergen algunos puntales básicos que van desde los trabajos de campo de Margaret Mead (1928; 1935) sobre los roles sociales en tribus indígenas (Conway et al., 2013), pasando por el sostén clave de Beauvoir acerca del llegar a ser mujer y hasta alcanzar una de las primeras

Peterson (2017) sobre el postestructuralismo como hacedor de las últimas olas del feminismo; la posición materialista y divergente de Christina Hoff Sommers (1995) en lo que concierne a aspectos difusos de lo que ella denomina “feminismo de género”; las reflexiones jurídico-filosóficas del profesor Pablo de Lora (2019) en referencia al encaje normativo del constructo social o el derecho penal de autor; así como el trabajo de Susan Pinker (2009) acerca de la influencia del sexo biológico en el aprendizaje.

¹⁵ El concepto de género, como tal, comienza a tomar forma en las primeras décadas del siglo XX. De hecho, su conceptualización previa en idiomas como español o inglés estaba anclada a su locución latina *genus*, y hacía referencia a su vertiente netamente gramatical. Su viraje actual emerge, por tanto, como sustancia que explica las dimensiones sexuales más allá de las notoriamente expuestas por la biología o la neurociencia reciente.

menciones académicas de identidad en Money et al. (1955), quienes inyectan a la cuestión una variable fundamental, la idea de que el rol de género puede ser adquirido, y más en concreto, educado. Y así, desde los 60, lo versará la fecunda obra de Robert Stoller, quien afianza una división clínica entre género y sexo (Green, 2010). Algo que también se apunta en la obra de Lopata y Thorne (1978), encargados de desechar la retórica de roles de sexo, focalizando sus miras en lo que hoy es conocido y versado como género (Risman y Davis, 2013).

Esta línea de pensamiento se sitúa como piedra angular desde la que irradiar futuras hipótesis. La conjunción aflora como soporte de las principales reflexiones de la segunda ola, acerca de la posición dominante que ejerce el hombre en distintos ámbitos, como el mercado laboral (Hartmann, 1976), y que sirve, en consonancia con los preceptos del aparato estructural, para subordinar a la mujer (New, 2020). De esta forma es como entrarían en juego construcciones sociales como el patriarcado o retóricas sobre el poder simbólico (León-Mejía, 2010). La literatura feminista se respalda en la premisa de cómo, a través del cuerpo, la estructura dominante posterga a un sexo y prepondera a otro. Las teóricas arguyen, en base a las definiciones biológicas de macho y hembra, que el patriarcado promueve los estándares del cuerpo apoyándose en el de hombre, mientras relega al de la mujer por su déficit físico e incluso pasional, llegando a catalogar a esta como la posición distorsionada (Annandale y Clark, 1996). Se habla, por tanto, de un primer sexo que superpone y dispone. Por lo que los postulados de esta presunción proveen una visión acerca de cómo el factor social subyuga a la mujer para que no sea o no quiera ser.

Comprobamos cómo el género se aloja como el quid desde donde el debate feminista, a partir de la segunda ola, construye su argumentario (León-Mejía, 2006). Se preconiza como esencia matriz escapar del determinismo biológico (Aguilar, 2008). Sin embargo, y como se expuso previamente, las raíces del feminismo no se encuentran para nada fosilizadas, por lo que han ido constituyendo nuevas reformas contestatarias que buscan subvertir esta dualidad binaria sexo/género y naturaleza/cultura. La tercera ola, de lógica post-estructuralista, atacó de lleno no solo la línea de flotación biológica¹⁶, sino que también argumentó que las posturas bimodales de género ayudaban más a reforzar la opresión de la mujer que a su emancipación (Annandale y Clark, 1996). Asistimos a una división dentro del cuerpo teórico. Por un lado, están las que abogan por la erradicación de la opresión que no deja a la mujer ser y, por otro, en línea con Butler (Sabsay, 2010), las que apuestan por implosionarlo mediante la atomización y la performatividad.

A pesar de las diferencias, en ocasiones insalvables (Schuster, 2017), lo que se muestra de forma clara es que el género se convierte en uno de los cimientos más sólidos utilizados por el feminismo para sustentar su argumentario político (Lamas,

¹⁶ Como señala una parte del feminismo, la propia concepción de sexos tampoco es muy antigua, y tiene un corto bagaje donde el acuerdo científico del siglo XVIII estipuló su carácter dual que llega hasta nuestros días (Fernández-Llebrez, 2004).

2013). De este modo, y si establecemos unos criterios comunes desde los que recabar el sentir del movimiento, ya desde la segunda ola, con respecto al género, diríamos que:

Primero, el ser humano tiene un carácter roussoniano donde, este buen salvaje (González Alcantud, 1987), es regulado y moldeado por el constructo social (Pinker, 2012).

Segundo, vivimos en una sociedad patriarcal creada por el hombre en distintas esferas y culturas (Walby, 1989) mediante la fuerza física, para oprimir a la mujer.

Tercero, es a través de esa estructura, en virtud de la cual el estereotipo masculino dominante dispone y genera una realidad dual dicotómica, como la de sexo/género, con la que atentar contra la libertad de ser (Knights, 2019), fundamentalmente, de lo consabido como mujer.

2.4. NLP; Descifrando el (enigma del) lenguaje

En esta sección, y a modo de revisión del caso, se presenta una evolución teórica de la técnica que ha sido empleada y que será, mediante un abanico de modelos, expuesta de forma pragmática y ordenada en nuestra metodología. Como referenciamos, la misión, o una de las misiones que aquí se arma, es la de acercar a nuestra disciplina un nuevo ámbito de tareas que sirvan para atajar el presente continuo de una sociedad cada vez más capitulada por el envite de las TICs. Para ello, abogamos por subirnos a la ola de nuevas incursiones que la multidisciplinaridad está recabando para la academia en general, y la ciencia política en particular (Luengo, 2016)

A día de hoy, hablar de NLP (Zampolli et al., 1994) es hablar sobre una pieza indispensable de nuestra vida diaria. Este puede encontrarse desde el SPAM de nuestro correo, pasando por aparatos domóticos o hasta coches autónomos. Todos estos ejercicios, funciones y acciones están siendo empleadas como aportaciones para solventar desafíos que hoy se tornan fundamentales. Para ello, se ponen en liza metodologías algorítmicas¹⁷ que puedan hacer frente a las ingentes cantidades de datos que, minuto tras minuto, generamos con todos los artefactos digitales que albergamos a nuestro alrededor. Esto, por consiguiente, y a través de unos criterios e infraestructuras determinadas, da forma a lo que se conoce como *Big Data* (Boyd y Crawford, 2011; Halavais, 2015; Jones, 2019). Por lo que este se erige como una parte crucial para que los procesamientos de máquina

¹⁷Desde los orígenes de la algoritmia, a saber, de Euclides a Alkhô-warizmi (Trajtenbrot, 1977; Angius et al., 2021), su cometido ha sido siempre el mismo, definir un método con el que solucionar problemas por etapas. Este origen netamente analógico está caracterizado por unos criterios taxonómicos e imperativos, precisos, definidos y finitos (Joyanes, 2008, p.45; Hill, 2016), y en la línea de lo que Markov (1954) emplea como exégesis sobre la cuestión. Por consiguiente, de lo primero de lo que se trata es de atajar una problemática, para lo que se diseña un algoritmo que después será codificado en un lenguaje inteligible entre humano y máquina, donde, por último, será ejecutado (Vázquez, 2012, p.17).

sean cada vez más sofisticados y certeros, y por añadidura poder avanzar hacia nuevos estadios.

La función principal que se aloja tras el NLP sería, como su nombre recaba, un entendimiento, mediante el uso de la razón, entre humano y máquina (Eisenstein, 2018). Partiríamos del uso sistemático del computador para descifrar mensajes -lo que se denominó inteligencia artificial¹⁸ (Turing, 1950; Minsky, 1961; Vergeer, 2020) o *machine learning* (Bi et al., 2019)-, algo que, unido la conjunción de distintas disciplinas como la lingüística, la psicología, la lógica o la estadística (Vijayarani et al., 2015), nos posibilitan para alcanzar un conocimiento de lo textual -o en sus siglas en inglés KDT- (Justicia de la Torre, 2017). Con ello, seremos capaces de nivelar el terreno para llevar a cabo un procesamiento e interpretación del lenguaje humano habido en nuestras muestras de análisis, en general desestructuradas, y así extraer un significado más completo de lo que allí se aloja (Kao y Poteet, 2007).

Esta técnica parte de la base de que el lenguaje humano, además de encerrar dejes y leyes (ej. La ley Zipf (1949)), también se compone de subjetividades y propósitos particulares (Sonntag y Stede, 2014). Así, siguiendo esta evaluación del caso, comprobamos que tanto los morfemas, la sintáctica, la semántica o la retórica (Chowdhury, 2003), pondrían en curso una amplia gama de niveles en los que debajo de cada uno encontrar una intención determinada con la que tratar de expresar, guiar, concienciar, persuadir... a nuestros interlocutores, o no (Mio, 1997).

Nuestro interés por esta rama se inserta en la línea de otros autores que han refrendado esta tónica en producciones académicas dentro de las ciencias sociales y, a su vez, en la ciencia política (véase como muestra este conjunto de publicaciones: Baumer et al., 2015; Figueiredo Filho et al., 2016; Wilkerson y Casas, 2017; García-Marín et al., 2018; Moreno-Mercado y García-Marín, 2020; Webb Williams et al., 2020). Por lo tanto, y sin despegarnos de nuestra visión pragmática de este ámbito de estudio, lo que de lleno buscamos es una capacitación mediante la cual poder adquirir y transformar desde cero datos de una pieza de análisis (texto, imágenes, vídeo...) a través diferentes técnicas de computación (Collobert et al., 2011).

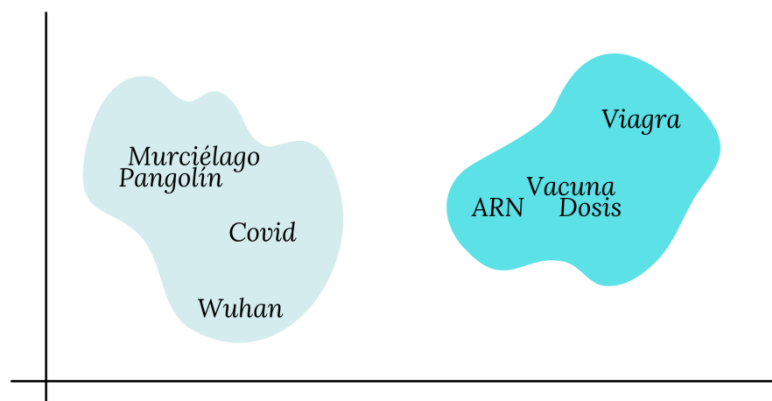
¹⁸ A pesar de la fama del término, y de que en determinadas acepciones de *inteligencia* que la RAE recoge sí que podríamos estar ante un hecho como el descrito, consideramos que en su carácter ontológico, y más en concreto en su acepción filosófica sobre el haber y el ser, la inteligencia como tal, también debería de incluir -al menos desde nuestra noción- la predisposición del pensante para saber diferenciar entre patrones, más allá de la simple conjetura binaria en la que se siguen unas reglas fijadas que no permiten lidiar con externalidades (Hui, 2020). Por tanto, test como la habitación China (Searle, 1985) o la dificultad de que un mismo modelo algorítmico pueda afrontar a la vez, y con garantías, distintas disciplinas como el ajedrez, el *go* o una partida de *Texas hold 'em*, pondrían en entredicho esta concepción. Quizás, y de ahí la reflexión, las dudas no solo estarían en el término inteligencia, sino también en el de artificial, ya que, aceptando el argot, podríamos hablar de *Inteligencia Delegada* o *Inteligencia Suprerepresentada*. De este modo, apuntamos más a quién aporta nuestra conceptualización [del intelecto] y a la forma en la que esta, gracias a la máquina, se manifiesta. Aun así, este debate es demasiado amplio como para ser abordado de una forma tan nimia, por lo que señalamos que esta nota es una aclaración brevísima sobre esta ardua polémica.

2.4.1. Cerrando el círculo, abriendo el espacio; Text as data y NLP como respuesta

Recapitulando acerca de cómo los medios de comunicación transmiten sus contenidos a las audiencias (Muñiz, 2017); mediante una interpretación filtrada de la realidad (Igartúa y Humanes, 2004; Arce y Menéndez Menéndez, 2018), convenimos en hilvanar algunas nociones acerca de este supuesto, y a través del cual tratar de un modo clínico y cuantitativo la interpretación que los medios hacen de la realidad.

Para dar forma a algunas de estas ideas partiremos de dos textos clásicos, pero que, a raíz de las cada vez más numerosas aportaciones computacionales, se constituyen como contemporáneos. Hablamos de trabajos fundacionales como son los de Riechert (1996)¹⁹ y Miller (1997), así como la puesta en común de Miller y Riechert (2001). Estos plantean y aportan nuevos modos con los que abordar la medición y tratado de los marcos textuales.

Figura 6. Representación espacial (Ej. Covid)



Fuente: Elaboración Propia

Como ilustramos en la Figura 6, el *frame mapping*, como los autores lo han bautizado, es una forma eficiente y precisa de eludir los posibles sesgos (a este respecto Valera Ordaz (2016) brinda una interesante reflexión sobre el papel de los investigadores del framing) que los estudiosos también aportan cuando llevan a cabo una investigación. En tales circunstancias, la línea que abren Miller y Riechert trata de desplazar posibles presunciones y, a su entender, elucubraciones. Pero, sobre todo, sus fundamentos sacan partida de las posturas de Entman (1991), quien subraya que la cohabitación de determinadas palabras transmite significados temáticamente consonantes.

¹⁹ A pesar de la referencia aportada, no ha sido posible acceder a la misma, ya que no se encuentra publicada, tal y como se recoge en nuestra propia bibliografía, pero también en el texto citado Miller y Riechert (2001). Aun así, y en base a lo que ellos exponen en el referenciado trabajo de 2001, el texto del propio Riechert es una de las piezas clave para entender cómo han podido alcanzar la composición actual de lo entendido por *frame mapping*.

La aportación de estos académicos abre la puerta a investigaciones que ya hoy son un hecho, y que algunas toman cuerpo en este proyecto. Aunque en los trabajos de partida ellos sustancian también algunos modelos y métodos particulares con los que llevar a la práctica estas ideas, el avance de la tecnología computacional catapultó sus conjeturas de forma exponencial. A pesar de ello, es cierto que sus nomenclaturas, sus nociones y sus formas de atajar esta cartografía de la cuestión, ya empleaban algunos de los sustentos que el actual NLP implementa en las ciencias sociales.

De este modo, de la fusión del NLP con las variadas sendas que el framing congrega en su seno (Goffman, 1974; Entman, 1993; Pfeffer, 1992; D'Angelo, 2002; de Vreese y Semetko, 2004; García-Marín, 2011; Morstatter et al., 2018; entre otros prolíficos expertos sobre la materia), obtenemos un modelo multidisciplinar con el que dar una respuesta al uso de las palabras. La idea del mapeo presenta una vetusta aportación que emerge de la lingüística más formal; el papel de los campos semánticos en línea con el *esquema* kantiano (Nerlich y Clarke, 2000). En una trayectoria similar, y tal y como señalan Winston (1995) y Lakoff (2008), las palabras ejecutan en el receptor marcos mentales que cercan el campo semasiológico. Este planteamiento sobre cómo el cerebro se activa ante un hecho como el de comunicarse, apunta en cierta medida a lo que subyace en la idea del *frame mapping* y, también, del *priming*. De conformidad con este razonamiento, nosotros también podemos encontrar en los textos núcleos relacionales, clústeres y frecuencias de palabras que abran la senda para el hallazgo de nuevos enfoques en nuestra disciplina (Conceição et al., 2017).

Gracias a este mecanismo asistimos al estallido de un nuevo modo de afrontar el estudio de medios, aunque en definitiva de corpus textuales, incorporando técnicas -no solo de mapeo- que nos permitan conocer de forma más amplia qué hay en nuestras muestras. A este respecto, y vinculándolo de nuevo con las dotes del procesamiento del lenguaje, conseguimos un abanico de métodos que nos aportan unos suministros cada vez más robustos. Uno de estos ejemplos sería el *Topic Modeling*. En armonía con lo descrito por Ramage et al. (2009) o Sagi et al. (2013), esta técnica de modelado de temas se basa en una forma de disminuir la ingente cantidad de datos generados y, a la par, reducir las variables ocultas en ellos, sin olvidar la reproducción espacial que podemos obtener. A su vez, y adelantando algunas de las funciones que se aclararán en la metodología, encontraríamos otro método para importar y fundar regiones de palabras, en este caso por medio de lo denominado como *Word Embedding*. Esta, parte en una primera instancia de la conjetura de proximidad como factor clave en el análisis de contenido (Arnborst, 2017). Los aportes emanados del *Word Embedding* nos proporcionan a los investigadores la posibilidad de acceder a datos interesantes sobre por qué en determinadas circunstancias y medios, algunas palabras estarían más próximas a otras, revelando con ello, campos semánticos, marcos y conjeturas ideológicas o mediáticas. O, dicho desde una mirada más próxima en el tiempo, por

qué las palabras *pangolín* y *murciélago* están distribuidas espacialmente al lado de Coronavirus.

Bajo esta serie de pretextos, son variados los enfoques que buscan añadir la distribución espacial de las palabras a los estudios del framing. En este particular, y siguiendo algunas de las técnicas computacionales citadas, nos encontramos que, en la actualidad, a diferencia de lo que se producía hace unas décadas (García-Marín, 2007), los frames surgen más de una instrucción *post hoc* que de una búsqueda *ad hoc* prefijada por determinados patrones. Sin duda, son dos formas más que interesantes de alcanzar tales ideas, de hecho, si recurrimos a aportes como el de Benoit et al. (2018), podríamos fusionar ambos constructos. De este modo, asistimos a una evolución donde los aspectos macro de los marcos, se ven, en cierta medida, virados por el influjo de microframes (Hansen, 2015). O, siguiendo a Kwak et al. (2020), cómo el uso de un simple antónimo (ej. legal o ilegal) fijaría el enmarcado de la noticia o texto en particular.

3. Metodología y datos

A lo largo de las siguientes páginas se emprende una descripción de los métodos y criterios utilizados en nuestras investigaciones. De este modo, es de recibo señalar que cada uno de los estudios está sustentado bajo unos parámetros singulares que marcan cada una de las fases de la metodología. Se lleva a cabo esta aclaración debido a que, dependiendo del repositorio, web o red social minada, el tipo de datos a obtener varía en función de nuestro interés, nuestra hipótesis y de los mecanismos a nuestro alcance. Así pues, en este primer peldaño apuntamos que en el epígrafe 3.5 podrán encontrarse las explicaciones pertinentes de cada investigación. En ellas se pueden localizar fechas, tipos de búsqueda (Codina, 2020), así como el lugar donde fueron recabadas las piezas de análisis.

Como se ha podido apreciar, el presente trabajo sigue una arquitectura prediseñada con el fin de ofrecer un curso legible y unificado, aferrado a una lógica de ordenación. Este inventario aúna, desde la recogida de datos, hasta la posterior clasificación mediante el empleo de algoritmos y los cruces de estos. Los resultados que esas fases ofrecen son insertados en figuras, tablas y visualizaciones con las que podremos emitir una evaluación y, consiguientemente, unas conclusiones esbozadas en torno a algunas ideas más clásicas, como puede ser el análisis de contenido²⁰.

Antes de pasar a explicar los distintos métodos implementados es razonable mencionar que, para poder desembocar en este camino, nos hemos servido de diferentes softwares con los que ir escalonando nuestras acciones. En este particular, nuestro proyecto se apoya en el empleo de lenguajes de programación, más allá de los paquetes unidireccionales expuestos en los programas informáticos. A este saber, debemos mencionar el conocimiento de lenguajes como JAVA, Javascript, Python y, en mayor medida, R. Con ellos, hemos sido capaces de perfilar tareas complejas y muy precisas que de otro modo no hubiera sido posible.

3.1. Extracción de datos

Tal y como se ha desgranado de antemano, el tipo de dato obtenido es fundamental en el momento de decantarse por unas técnicas u otras. Bien estemos ante posibles datos estructurados o desestructurados, se torna primordial saber hacia dónde queremos dirigir la búsqueda. Aun estando ante un axioma, no es menos cierto que este se considera crucial a la hora de abordar unos procesos u

²⁰ El análisis de contenido es uno de los métodos más extendidos dentro del estudio de medios de comunicación, aunque también, por supuesto, de todo aquello que pueda ser, de forma sistemática, cuantitativamente analizado en la búsqueda de patrones y símbolos. Por ende, y aun simplemente referenciándose a esta instancia previa, lo que se evidencia es que, de conformidad con la propia estructura de nuestro epígrafe metodológico, una de las piezas que vertebra este trabajo es ese estudio sistematizado de los corpus –en los que podría albergarse también imagen o vídeo–, ya sean de prensa o de otra índole, por lo que, sin ahondar de forma precisa, el planteamiento que incluye el análisis de contenido es la exégesis neta de este epígrafe. A pesar de ello, sí nos gustaría acopiar algunos autores mediante los que condensar este modo de adquirir y aportar conocimiento, así, siguiendo a Hansen y Manchin (2019), podríamos dibujar un extenso recorrido, a saber: Berelson, (1952); Altheide y Snow, (1979); Picard y Adams (1991); van Dijk, (1991); Krippendorff, (2004) [originalmente publicado en 1980]; Atanasova y Koteyko, (2016).

otros. Por este motivo, y dejando a un lado repositorios y bases de datos genéricas como *Mynews*²¹ o *Nexis Uni*²², este trabajo se ha valido también de partículas de análisis extraídas de fuentes externas, las cuales no cuentan con criterios estandarizados. Estos hechos señalados han propiciado que se haya tenido que recurrir a técnicas novedosas en nuestro campo, aunque sumamente contrastadas en disciplinas como la computación o, de una forma más utilitaria, el marketing. Esta circunstancia propicia que para ejecutar nuestra captación de datos deban llevarse a cabo, previamente, estudios de la arquitectura web o el empleo de APIs²³ que faciliten la extracción.

Esta tesis subraya tres piezas clave. Por un lado, tendríamos las URL²⁴, que son las direcciones webs en donde se aloja la información que queremos analizar. Seguidamente, y ya dentro del emplazamiento online, estaría el documento HTML, al que podemos acceder a través del código fuente. Por último, y una vez allí, ingresaríamos en la composición en forma de árbol desde la que poder obtener nuestra información (Kouzis-Loukas, 2016) de un modo automatizado (vanden Broucke y Baesens, 2018).

Por tanto, y procedente del inglés, no es baladí el empleo del término raspar a la hora de hablar de obtener información de medios online. La conocida como *scraping*, es la técnica encargada de que, una vez hemos estudiado el desglose de la web, podamos generar nuestros *sitemaps*²⁵, en los que indicar unos *inputs* claros y ordenados con la intención de obtener unos *outputs* en forma de archivos, generalmente, CSV. Estos serán los que nos posibiliten materializar los siguientes tramos de análisis y el desvelo del conocimiento (Thomson, 2019) allí albergado.

3.2. Minería de texto

Una vez se han perfilado los archivos, se efectúa una fase de minado en la que conseguir estadísticos y patrones que nos serán útiles en la investigación. Esta etapa busca explotar los guarismos –entre comillas– ocultos. Gracias a una suma de procesos informáticos, podremos hacer acopio de nuestras muestras –hasta el momento sin una estructura legible– y, en consecuencia, armonizar un entendimiento máquina-humano. Por lo que una vez los ficheros se encuentran claramente parametrizados se inician los trámites de preprocesamiento.

A continuación, se sanean los textos, eliminando cualquier anomalía (ej. caracteres extraños o mal codificados) y signos de puntuación, y se configuran los siguientes niveles para la instrumentalización del lenguaje humano. Por

²¹ <https://ugr.mynews.es/hu/>

²² <https://bit.ly/3edxqYf>

²³ *Application Programming Interface (API)* son un conjunto de servicios proporcionados por un software con el fin de extraer datos tales como objetos, urls... (Meng et al. 2018)

²⁴ A pesar de que las direcciones webs y sus correspondientes hipervínculos podrían ser obtenidos en repositorios previamente dispuestos, existen procedimientos como pueden ser el *crawling*, o *araña web*, que, mediante el desarrollo de bots, secuenciarían el esquema y las centenas de miles de ventanas que un portal podría hospedar (Jarmul y Lawson, 2017)

²⁵ En palabras de Ceci y Lanotte (2021) los *sitemaps* son especificaciones claras y jerárquicas enunciadas para indexar la arquitectura de una web.

consiguiente, y ya que esencialmente nuestro trabajo se apoya en la palabra escrita, las técnicas que vamos a adoptar se enmarcarían dentro de lo denominado como minería de texto (Tandel et al., 2019). Esta busca sacar partido de los estadísticos habidos en los datos desestructurados. A tal efecto, podemos encontrarnos con unidades de análisis textual que pueden comprender desde las noticias de prensa, las publicaciones en redes sociales, los discursos políticos o hasta los programas electorales. Todos ellos congregan a su vez singularidades que deben ser estudiadas en función de nuestra prioridad. Por ello, antes de pasar a una fase posterior en la que se pueda eliminar el ruido o los *outliers* poco significantes, se ha de instar por una determinada partícula a estudiar. A este respecto, y dependiendo del tipo de prueba a la que queramos enfrentar nuestras muestras, podremos decidir subdividir el texto en párrafos, frases e incluso palabras como nodos independientes, o lo que en el argot del NLP se denomina *token*²⁶. Una decisión que, en definitiva, mejoraría los resultados y también la calidad de las clasificaciones y etiquetados (Bunescu y Mooney, 2007; Esuli y Sebastiani, 2009).

Continuando las operaciones fijadas por el preprocesamiento para una eficiente limpieza de la muestra (tómese Feinerer et al. (2008) y Talib et al. (2016) como fuentes para una mayor profundización), pasamos a definir unas líneas de actuación con las que recabar variables más nítidas de los datos crudos. A tal efecto, los pasos siguientes se sustentan en²⁷:

- Limpieza de *Stopwords*²⁸. En esta acción se depuran todas aquellas palabras que como “en” o “y” no reportarían una gran significancia a conteos o correlaciones de pares, así como a posteriores acciones de clasificación de textos o modelado de temas presentes.
- *Stemming*. Esta técnica sirve para disminuir la estructura de las palabras a través del examen de su morfología. La operación se lleva a cabo partiendo del pretexto de que la raíz de la palabra puede aportar la definición precisa que esta tiene (Lovins, 1968). Por este hecho, se pueden recortar tanto

²⁶ Si bien, en una primera instancia, apoyada por el grueso de la literatura, el uso del término *token* se emplea ampliamente como un sinónimo de palabra -como partícula aislada-, no es menos cierto que esta idea suscita un interesante debate al respecto. En línea con los planteamientos esbozados por, entre otros, Youmans (1990) o Webster y Kit (1992) nos encontramos con un punto de vista muy lícito, interesante y sustancial; el realizado por los lingüistas y estudiosos de la comunicación acerca de si una palabra puede, por sí sola, aportar realmente el significado de lo que el emisor de esta quería expresar. Por ejemplo, como los autores mencionados exponen, la grafía y léxico chino, los *idioms* del inglés o, si apuntamos al español u otros idiomas; conjuntos como *Casa Blanca*, no podrían separarse, ya que estos perderían, al ser disociados, su fundamentación.

²⁷ Todas y cada una de estas líneas de actuación se consideran de carácter contingente, por lo que, como se ha venido resaltando a lo largo de esta metodología, algunos pasos pueden ser eludidos o adoptados de un modo diferente. Como decimos, los procesos a desarrollar solo dependen del interés que tengamos o de si las técnicas son lo suficientemente consistentes como para ser aplicadas. Se hace este breve inciso porque en esta fase se puede depender de bibliotecas y diccionarios conformados de antemano y que, predominantemente, son trabajos formulados en lengua inglesa. Esta disyuntiva propicia que existan bibliotecas, por ejemplo, de *stemming*, que pueden carecer de una solidez, desde nuestro punto de vista y a día de hoy, lo suficientemente robusta como para poder captar determinadas singularidades. Aun así, sí es de recibo mencionar que, en este particular, la existencia de la propia técnica invita a que en la actualidad solo se haya de precisar de registros de palabras lo debidamente abundantes y con la suficiente capacidad para allanar la mutabilidad del lenguaje, en especial, el español.

²⁸ Son todas aquellas palabras que, dentro de los análisis de minado o aprendizaje de máquina, no agregarían unos resultados sólidos. Es más, el no cribado de estas dificultaría el entendimiento de modelos temáticos u otros de frecuencias en el que la presencia de este tipo de palabras desvirtuaría nuestras indagaciones.

género como conjugaciones, quedándose solo con el infinitivo del verbo o el genérico neutro del término en cuestión. Además, y con el fin de disminuir aún más la cantidad de palabras, existen librerías, dentro de los múltiples lenguajes de programación, que incluyen también modelos de sinonimia (Fellbaum, 1998; Wallace, 2007; Feinerer y Hornik, 2020b), con los que compactar aún más la muestra.

- *Part of Speech*. Esta tarea de etiquetado, como su nombre referencia, sirve para listar gramaticalmente cada una de las palabras. Se torna útil para emprender una pesquisa sobre la tipología del lenguaje o, por ejemplo, explorar cuáles son los verbos o adjetivos más usados en un discurso, sin la necesidad de hacer encrucijadas posteriores. Para ejecutar esta función nos apoyamos, generalmente, en librerías de NLP ya confeccionadas (a saber, (Bird et al., 2009) o (Hornik, 2019)).

A esta breve retahíla se puede incorporar otro conjunto de técnicas que podrían ofrecer un mayor surtido de ratios u otros rendimientos (para un acercamiento pormenorizado de la cuestión véase, por ejemplo, Serrano-Contreras (2020) o el magnífico libro de Silge y Robinson (2017) donde se expone un amplio arsenal de pautas a tener cuenta). Además, es muy interesante señalar que, gracias a algunas de estas librerías de programación (a saber, Benoit et al., 2018; Feinerer y Hornik, 2020a), y a diferencia de programas informáticos cerrados, podemos dar forma a nuestros propios diccionarios y con ellos ampliar las nociones del entendimiento del lenguaje -bien sea para una depuración, un filtrado o una ágil primera clasificación- que el investigador requiere en cada determinado estudio.

Por último, y una vez los corpus han sido cribados y definidos, ponemos en marcha actividades más complejas. Como señalan diferentes trabajos de minería de texto (Gentzkow et al., 2017; Kwartler, 2017; Justicia de la Torre et al., 2018; Jockers y Thalken, 2020) pasaríamos a formular estrategias que irían desde el puntaje con lexicón (Hu y Liu, 2004; Mohammad y Turney, 2013; Cruz et al., 2014; Jockers, 2017) hasta técnicas no supervisadas de clasificación como el clustering K-medias o el análisis de componentes principales (Figueiredo Filho et al., 2014; Wilkerson y Casas, 2017), ideadas para adquirir un conocimiento que, desde nuestra mirada humana, nos era opaco e ininteligible debido a la vasta cantidad de información imperante.

3.3. Indexar la polarización

Gracias a los tramos previamente fijados, y en nuestra misión por transformar el texto en datos, y estos a su vez en resultados legibles para el entendimiento humano, procedemos a dar cabida a uno de los aportes que agrega esta tesis. Partimos de los pilares fijados por, entre otros, Poole y Rosenthal (1983; 1984; 1985) con su medidor geométrico dicotómico o los calibradores expuestos por Esteban y Ray (1994). Estos, como otros tantos coetáneos y posteriores, integran representaciones espaciales multimodales de la sociedad con el fin de encontrar

los segmentos que la segregan. Esta interesante cuestión se reexamina con trabajos como los de Laver et al. (2003) o Slapin y Proksch (2008), que abren la senda para la inclusión de modelos computacionales en los que se recurre al texto como fuente con la que dar forma datificada a las disensiones.

Nuestra aportación toma cuerpo en base a la elaboración de un método con el que indexar el distanciamiento; la polarización habida en nuestras piezas muestrales. Este calibre surge, en primer lugar, de la necesidad de lo que hemos venido a denominar polarización semántica, intrínsecamente relacionada con la polarización política. En segundo lugar, la idea germina gracias a que las técnicas de análisis de sentimientos no son lo suficientemente robustas como para aportar una ponderación de la polarización. Lo que estas hacen es captar segmentos aislados que no se tienen en cuenta, por lo que, la realidad polar presente, no puede verse reflejada sin una combinación de valores, como haría un índice (Ferrando, 1987, p.34). Es más, incluso aunque los puntajes no estuvieran hechos con lexicones, sino con modelos preentrenados por el propio investigador, los ratios enmarcados, aun pudiendo ser más precisos, no contarían con los cruces pertinentes mediante los que cerciorar los lineamientos de nuestra muestra, además de no tener en cuenta la alta dimensionalidad de la naturaleza del discurso (Goet, 2019).

Avanzando sobre esta trayectoria, decidimos estimar la polarización operacionalizando esta como (Pc) que sería la distancia entre el ratio calculado para esta partícula –ya sea un artículo de prensa, comentario en una red social o un discurso político– (S) y la mediana²⁹ del agregado de los ratios –ya sean los artículos de ese medio, los comentarios de una publicación o los discursos de un partido político– (Me), en número absoluto. De esta manera, obtenemos un número que puede tomar cualquier valor entre 0 y 2, donde 0 muestra que no hay polarización y 2 muestra una polarización máxima. El ratio que establece nuestro índice tiene la capacidad de medir las posibles divergencias que, dentro de un arco temporal, pueden generarse en un medio, un partido político o entre usuarios de una red social:

$$Pc = |(S - Me)|, \text{ donde } S \in [-1, 1], p \in (0, 2)$$

Tanto la pieza de análisis, como la elección de los agregados, pueden variar en función del interés. Así se evidencia en Serrano et al. (2020; 2021), donde no solo varía lo analizado, sino que también cambia el modo en que se puntúa.

3.4. Codificar, incrustar, tematizar, clasificar y generar

Los verbos que titulan esta parte del proyecto son las acciones que se llevan a cabo para que, por ejemplo, “feminismo” pueda ser entendida por nuestra máquina.

²⁹ Ya que este índice busca una forma robusta de calibrar dónde está el centro y cómo las facciones se distancian de él, se decide emplear una distribución con mayor consistencia como es la mediana. Dado que, a diferencia de la media aritmética, la mediana se ve menos afectada por valores extremos y, por tanto, hará un mejor retrato de en dónde se halla la posición divisoria entre los valores.

Después, gracias a esta traslación, podremos encontrar grupos, estructuras y clústeres con los que reducir las dimensiones de nuestros textos.

Una de las principales técnicas del NLP para este fin sería el *Topic Modeling*. Bajo el paraguas de esta perspectiva se congregan una mixtura de modelos útiles para que miles de corpus textuales se vean reducidos a unas pocas temáticas. De este modo, lo que obtenemos es un número de k^{30} -temas- que en unas pocas palabras describen los temas presentes. Por ejemplo, si en nuestro caso tenemos textos sobre el feminismo, este algoritmo podría aducir que, en base a sus procesos, los temas podrían ser fijados en dos ámbitos concretos como la lucha de la mujer y la violencia ejercida sobre esta. Este paradigma sirve para que un investigador con miles de comentarios de una red social solo tenga que etiquetar cada una de las k que este procedimiento devuelve como salida.

En armonía con el anterior planteamiento, son múltiples los algoritmos que se han implementado para dar forma a esta premisa. Quizás, desde nuestro punto de vista, el más famoso de todos, y base de alguno de los prototipos empleados, sea el conocido como LDA –*Latent Dirichlet Allocation*- (Blei, et al., 2003). Este modelo generativo parte de un *input* de frecuencias de palabras (en inglés *Bag of Words*, –BOW-) basado en conceptos bayesianos en donde las palabras albergadas en los corpus generan probabilidades de presencia conjunta. Por ejemplo, de “feminismo” e “igualdad”, este asume que son parte del mismo tema, por lo que los inserta dentro de la misma k . Bajo el amparo de esta noción se han conjugado otras ideas en las que ampliar las capacidades de este algoritmo. Por un lado, tendríamos el STM –*Structural Topic Model*-, empleado para incorporar metadatos tales como fecha o variables categóricas, con los que poder recabar una representación escala temporal en donde examinar posibles fluctuaciones (Roberts et al., 2014; Roberts et al., 2019). Por otro lado, encontraríamos el denominado BTM –*Biterm Topic Model – Short Text*- (Yan et al., 2013). Su principal tarea es la de salvar la escasa agregación habida en los textos cortos (ej., titulares de periódico, publicaciones en redes sociales...) y que de este modo no se pierda la esencia de la muestra (Cheng et al., 2014). Finalmente, presentamos el último que hemos estudiado y utilizado. El denominado como *text2vec* (Selivanov et al., 2020) parte de la naturaleza del LDA, aunque empleando una vectorización previa que trata de salvar cuellos de botella de procesamiento de RAM.

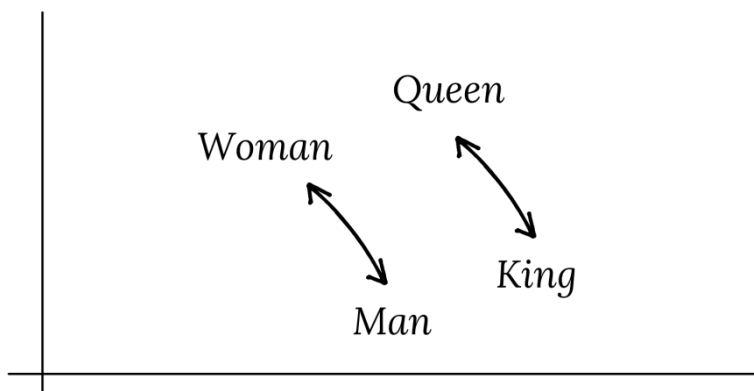
Ahondando dentro del ámbito del *Topic Modeling*, también se ha articulado durante estos años de aprendizaje, uno de tipo LSA –*Latent Semantic Analysis*- (Deerwester et al., 1990; Landauer y Dumais, 1997; Hu et al., 2007) mediante el cual reducir los posibles significados de las palabras con el fin, de nuevo, de obtener unas k presentes. Esta tipología se constituye también mediante una matriz de co-ocurrencia la cual ve reducida su dimensionalidad a través del truncamiento

³⁰ Aunque pueden estimarse el número de k gracias a la observación in situ de los *outputs*, existen técnicas con las que, de forma previa, obtener el número de temáticas más sólidas antes de proceder a la inclusión de los datos en los arquetipos de *Topic Modeling* (a saber, Griffiths y Steyvers, 2004; Cao et al., 2009; Arun et al., 2010; Deveaud et al., 2014).

ejecutado por una SVD -*Singular Value Decomposition*-. Con lo cual, una vez se ha obtenido la representación vectorial de los términos, la similitud (1, -1) entre las palabras se obtiene gracias a una distancia de coseno³¹ (Altszyler, et al., 2017).

Por último, y para hacer frente de otro modo el reto de cómo obtener la significancia terminológica -o lo que es lo mismo, por qué, por ejemplo, “feminismo” no tendría igual significado en un medio liberal que un uno conservador-, emprendemos una técnica de incrustación. La conocida como *Word Embedding* (para conocer algunos de los ejemplos más pujantes de este algoritmo, ver *Word2Vec* (Mikolov et al., 2013) o *GloVe* (Pennington et al., 2014)), emerge desde una perspectiva muy clara: *Rey* es a *Hombre* lo mismo que *Reina* es a *Mujer*, o *Rey* - *Hombre* + *Mujer* = *Reina*. Este sugestivo enfoque parte primero de una clasificación vectorial³² (x, y) con tantas palabras como tenga nuestro corpus, para, en una siguiente instancia, compactar esta información y reducir su dimensionalidad en un mapeo (Ghatak, 2019), en donde nuevamente podamos abordar la similitud mediante las distancias. Esta técnica sirve para conocer de forma más visual qué realidades alberga el texto o, en base al ejemplo, extraer sugerentes analogías. Dicho lo anterior: Si a *España* le corresponde *español*, a *Italia* le corresponde X. Por tanto, la respuesta a esa X es lo que resulta de codificar e incrustar espacialmente un término (Allen y Hospedales, 2019), generando con ello su significado dentro del corpus. Así, lo que conseguimos, es continuar la exposición que exponen Gentzkow y Shapiro, (2010); Sagi et al. (2013); An et al. (2018) o Kwak et al. (2020); una simple palabra empleada en un ámbito geográfico o temporal determinado puede entrañar interpretaciones totalmente distintas.

Figura 7. Similitud vectorial por coseno



Fuente: Elaboración Propia

³¹ Siguiendo a Huang (2008) o Han et al. (2012) la distancia de coseno se refiere a un ángulo entre dos vectores espaciados geoméricamente, por lo que esta medida, a diferencia de la euclídea, no se ve forzada por el tamaño del documento. O dicho de otro modo, podremos primar más la significancia de un corpus, que no la desviación ocasionada por la ocurrencia constante de determinadas palabras (a saber en Han et al. (2012): $\text{Similaridad} = \cos(\theta) = \frac{x \cdot y}{\|x\| \|y\|}$ siendo x e y vectores comparados, donde $\|x\|$ es la norma euclidiana del vector $x = (x_1, x_2, \dots, x_p)$. Conceptualmente, es la longitud del vector. Del mismo modo, $\|y\|$ es la norma euclidiana del vector y” (p. 78).

³² Representación binaria comprendida entre x e y.

3.5. Descripción de las muestras

A continuación, vamos a exhibir y pormenorizar todas las muestras utilizadas. Tanto las diseminadas en diferentes estudios ya publicados, como aquellas otras que se han incorporado para aportar a esta tesis una coherencia y homogeneidad.

3.5.1. T.1. From Catholicism to Gender Equality: Political Rhetoric in Spain from 1980 to 2020

El muestreo de esta pieza de análisis se emprende mediante la revisión de todos los debates del estado de la nación acaecidos en España (1983-2015). Para ello se optó por una recolección de los discursos realizados por los presidentes del gobierno y las correspondientes réplicas de los líderes de la oposición. Por lo que el número total, sin tener en cuenta otro tipo de intervenciones, es de 50, 25 por cada partido político, en este caso PSOE y PP³³.

Para proceder a su posterior explotación se implementaron los ya citados modelos de preprocesamiento y limpieza. Asimismo, y gracias a las innumerables librerías y modelos útiles para la minería textual, confeccionamos un conjunto de diccionarios que nos permitieran la búsqueda de retóricas habidas en los corpus. Mientras que, por un lado, se creó uno de palabras religiosas (véase Diócesis de Santander³⁴); por otro, se conformó uno que incumbe a este trabajo, el relacionado con lo estipulado como género y las teorizaciones que le dan consistencia (a saber, *GenFem*³⁵).

3.5.2. T.2. NLP on YouTube: A look on Feminism

Para la obtención de la muestra de este artículo se realizó, el 16 de enero de 2020, una búsqueda parametrizada dentro de la plataforma YouTube, usando como patrón de búsqueda “feminismo”. Los comentarios obtenidos se cribaron teniendo en cuenta que el título del vídeo debía incluir la palabra “feminismo”. Después, fueron ordenados por número de visualizaciones y descartados aquellos que pertenecieran al 2020 ya que, en aquel momento, no se podía obtener el arco temporal completo del propio año. Así, se obtuvieron, gracias a técnicas de *scrapeo*, un total de 690.138 comentarios. Todos estos comentarios fueron limpiados y estructurados siguiendo los pasos citados en esta metodología con el fin de realizar estadísticos computacionales más profundos.

Como resultado de estas prospecciones se llevaron a cabo una serie de minados que están recogidos en el propio artículo: los agregados temporales, la

³³Nota aclaratoria. Teniendo en cuenta que este partido nace en 1989 de, principalmente, la extinta AP y la reunificación de otros partidos y sensibilidades del espectro conservador/liberal, lo que hicimos fue tomar a la propia AP -y así se reflejaba en el turno de réplicas- como la oposición real del PSOE de aquellos años, siendo, a la postre como decimos, el origen primigenio del actual PP.

³⁴<https://www.diocesisdesantander.com/>

³⁵ “Diccionario feminista para miembros atónitos del patriarcado” por Kohan (2018) y “Breve diccionario de feminismo” por Cobo y Ranea (2020).

regresión local implementada en los comentarios y una muestra donde aplicamos nuestro índice de polarización.

3.5.3. T.3. A machine learning analysis of gender images in Spanish Press

Los datos que aquí se desgranar son producto de un trabajo académico aún sin publicar. Estos recogen un periodo (2009-2019) en el que, por medio de algunas de las cabeceras más importantes en este país (*El País* y *El Mundo* (versión web y papel), *El Confidencial* y *ElDiario.es*), se compilan –desde *My News*–, todos los titulares en los que se incluyeran tanto la palabra “mujer” como “hombre”. Lo que dio como resultado un total de 66.248 titulares.

Una vez más, y después de seguir las líneas ya mencionadas para el trabajo con texto, se aplicaron diferentes y variados modelos de machine learning. Desde unos de *Topic Modeling* (a saber, *Text2vec* o *BTM*) a ejercicios de incrustado de palabras mediante *Word Embedding*.

3.5.4. Datos supletorios (T.4. y T.5.)

Debido a que este trabajo surge de la mixtura de investigaciones, hemos considerado que sería de recibo aportar otros datos que dieran una mayor firmeza a la presente tesis, pudiendo así cobrar una entereza como trabajo propio y, a su vez, servir de análogo a los ya presentados. Con este fin decidimos suplementar los datos habidos con dos indagaciones.

T.4 Por un lado, una recopilación de agregados sobre violencias de pareja, obtenida de *Nexis Uni* de los medios *El Mundo* y *El País* (ambos versión papel). Para ello se recaban todos los titulares que, durante las últimas dos décadas (2000-2020), se han publicado acerca de la cuestión. Para que estos artículos fueran incluidos en la muestra, todos ellos debían de contener alguno de estos conjuntos: “violencia doméstica”, “violencia de género” o “violencia machista”. Si alguno formaba parte del titular de la noticia estas eran incorporadas, lo que produjo: $n = 25085$. La pretensión de esta búsqueda es la de mostrar una evolución de los conjuntos citados en prensa, teniendo en cuenta la disensión editorial de ambos medios.

T.5. Por otro lado, se hizo acopio de todos los artículos presentes en *Nexis Uni* que tenían en su titular la palabra “feminismo”, dando un total de 377 para el periodo (2000-2020). Al igual que en la búsqueda anterior, esta también se ejecuta en los citados medios (*El País* y *El Mundo*, versión papel), aunque en esta ocasión se recaba también el cuerpo de la noticia. Los artículos obtenidos fueron también introducidos en otros modelos computacionales donde buscar posibles distanciamientos.

4. Resultados

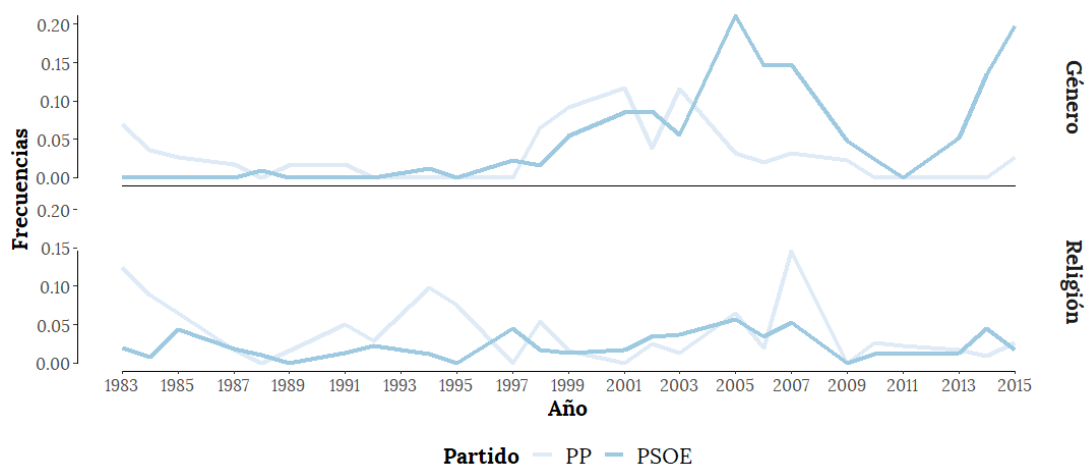
Buscando compactar de manera uniforme y estándar los precedentes trabajos académicos ya revisados, así como aportaciones conformadas para otorgar una mayor robustez a la presente tesis, surge este marco de discusiones. Consecuentemente, y en la senda fijada por los objetivos e hipótesis de este estudio, lo que adelante esbozamos será un flujo de datos que parte de cifras netamente descriptivas, hasta alcanzar entornos más complejos en donde los métodos teorizados en párrafos anteriores puedan sustentar disquisiciones más profundas.

Las muestras exploradas comprenden un periodo que se extiende a lo largo de las últimas décadas en España, aunque con especial hincapié en los años finales de nuestra época. Los ratios obtenidos beben de fuentes totalmente dispares. Mientras que por un lado tendríamos el análisis de debates políticos –en este caso los debates del estado de la nación–, por otro aportaríamos tanto el estudio de contenido de medios de comunicación clásicos, como es la prensa escrita, además de alternativas del ágora digital, como son las redes sociales.

A pesar de la amplia amalgama de guarismos y nichos de extracción tomados, brindamos como un primer apunte, que todos exhiben una cadencia generalizada, al menos de forma agregada, siendo exaltada durante los últimos lustros. Esto ya se refrenda en el chequeo de los debates del estado de la nación, y con especial consideración en la figura de frecuencias de usos retóricos (Figura 8.), donde la presencia de uno de nuestros ámbitos de estudio –el género– adquiere un mayor vigor en las últimas dos décadas. Para empezar, estos hechos se revelan sintomáticos si inspeccionamos detenidamente las Figuras 8 y 9. Por ejemplo, el gráfico 8, pone de manifiesto cómo la cuestión de género es un tema que se instaura en el discurso político desde el final de los 90. Como revelan las frecuencias de uso, este es primero enarbolado bajo prisma del Partido Popular, aunque su actual raigambre no es alcanzado hasta la primera legislatura de Jose Luis Rodríguez Zapatero (2004-2008). Él es quien, con los datos en la mano, impulsa una bandera patente acerca de estos temas, disparando la presencia de los mismos en el debate político y social³⁶.

³⁶ Tómense como prueba de ello las leyes de VIOGEN (2004) o paridad (2007).

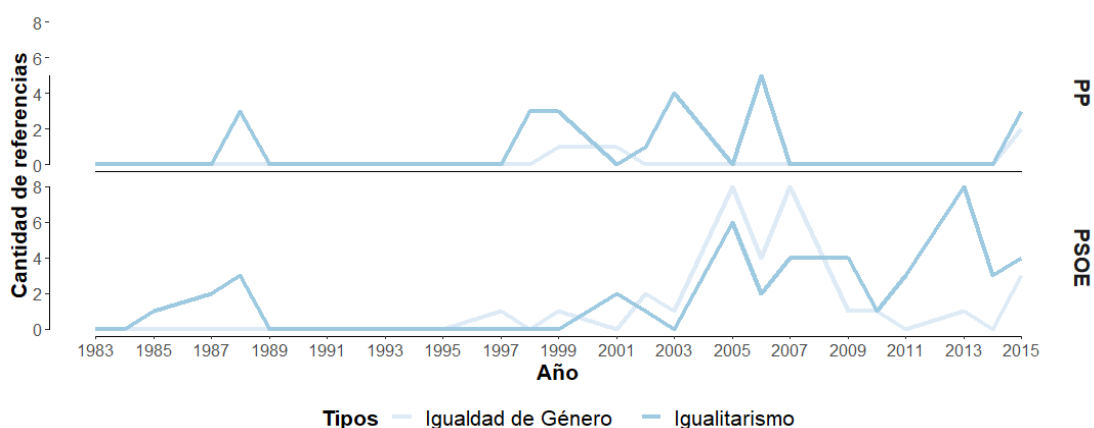
Figura 8. Frecuencias retóricas



Fuente: Elaboración Propia

Este cambio producido en aquella época a raíz de la llegada de Rodríguez Zapatero es refrendado en la Figura 9. Como los resultados de esta revisión proporcionan, el discurso sobre la igualdad de sexos/géneros no solo se queda en una mera expresión por el deseo de alcanzar unas cotas, sino que se conjuga mediante la adscripción y elaboración de un argot y un discurso propio. Siendo esto lo que enuncia la Figura 9, donde la palabra evaluada tiene dos posturas diversas de entender y afrontar la realidad social. Por un lado, la que apuesta por un igualitarismo clásico, mientras que por otro, la que aboga por una igualdad defendida desde las posturas del feminismo de género.

Figura 9. Referencias a “igualdad”



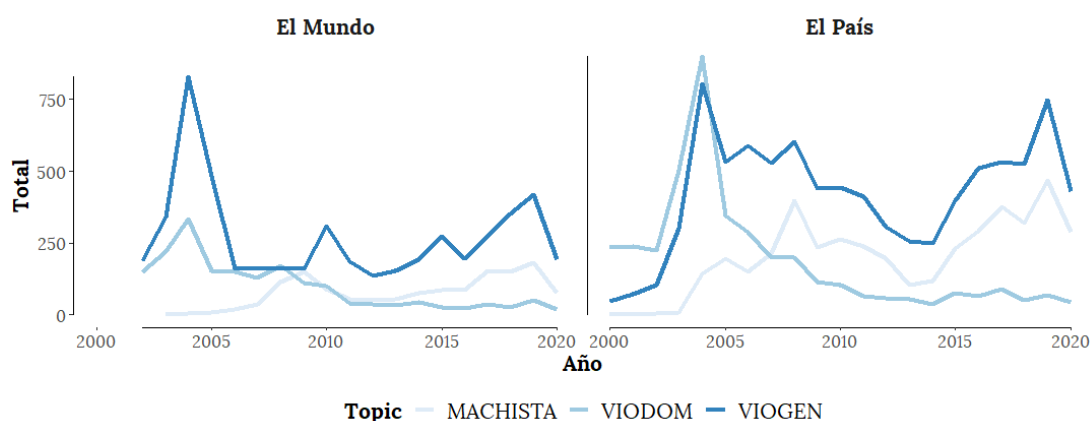
Fuente: Elaboración Propia

Al igual que ocurría con las frecuencias del uso retórico -Figura 8-, la formulación y uso de “igualdad de género” dentro de la Figura 9, cobra fuerza a partir de periodos similares. Por lo que, el empleo de la palabra “igualdad” da lugar al surgimiento de dos bloques que se encuentran distanciados semánticamente, al menos en sus usos. La Figura 9 también introduce otra idea sugerente; el término

igualdad se asume en los últimos años (2003-2015) de un modo frentista y correlativo. Mientras que el PSOE, durante las últimas dos décadas asume el término con sus dos matices, el PP solo aboga por la preferencia igualitarista. Algo sintomático de la diferencia entre alas y también de la capacidad, en este caso del PSOE, de dar sustento a nuevas acepciones dentro de su argot político. Además, se revela cómo el empleo de las dos nociones de igualdad se correlaciona negativamente una vez el país entra en la crisis económica de 2008. Al menos para el partido que detentaba el gobierno en aquella primera época de recesión, el PSOE.

A consecuencia de un ejemplo extraído de la anterior muestra y pronunciado por Mariano Rajoy en 2007: “la ley de violencia doméstica que ustedes llaman de violencia de género”, decidimos, ante la escasez de datos acumulados en los debates del estado de la nación, abonar esta disyuntiva con una breve revisión descriptiva (véase 3.5.4. apartado T.4.). Para ello, llevamos a cabo unas extracciones en dos cabeceras españolas ya clásicas; *El País* y *El Mundo*, adscritas cada una a líneas ideológicas diferentes y, a su vez, cercanas a los dos principales partidos que han ostentado la presidencia del gobierno durante este tiempo –con la excepción de UCD-. Esta exposición visual de la Figura 10 se torna interesante, ya que agrega una nueva arista a este amplio conglomerado. Si nos fijamos en las direcciones, estas prueban de forma manifiesta una predisposición por una denominación en concreto a la hora de hablar de violencias de pareja. Para ambos medios, el conjunto “violencia de género” se vuelve el dominante a partir del primer lustro del siglo XXI. Aunque podría ser sintomático que *El País* ofreciera datos antes que el otro periódico, lo más relevante, a nuestro juicio, es cómo este medio es el que, en primer lugar, se decanta por usar el ensamble terminológico “violencia doméstica” en lugar de la de “género”, algo que en cambio no ocurre en *El Mundo*. Por último, hemos de señalar que se incluye también en el análisis la partícula “violencia machista”, pero consideramos que su surgimiento no entraña significaciones específicas, y su apelación siempre es cuasi correlativa con “violencia de género”, por lo que esto se puede deber a una simple sinonimia o complemento retórico por parte de los estilos de los medios analizados, ya que no encontramos una tendencia que indique lo contrario.

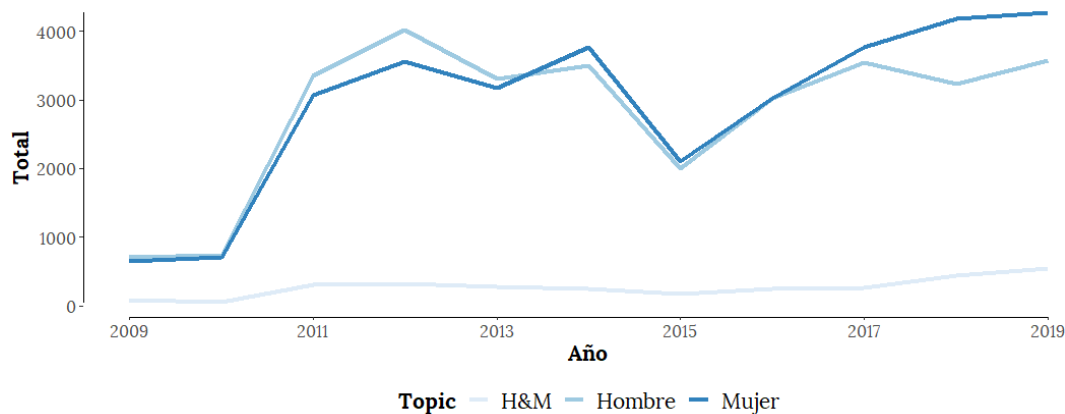
Figura 10. Nomenclaturas de violencias



Fuente: Elaboración Propia

Para suplementar los anteriores resultados, nos servimos de otro trabajo ya armonizado sobre también titulares de prensa (véase T.3, epígrafe 3.5) en un conjunto de medios emblemáticos (*El País*, *El Mundo* (ambos en versión web y papel) además de *El Confidencial* y *Eldiario.es*)³⁷. Por ejemplo, decidimos centrarnos en cómo los periódicos, a través de los titulares, representan tanto la figura de “mujer” como la de “hombre”. Algo que, a tenor de los resultados, termina focalizándose más en las relaciones entre ambos, que, en otro tipo de hechos, y de ahí que estos aportes se presenten como una continuación de la anterior figura.

Figura 11. Agregados “Mujer” y “Hombre”



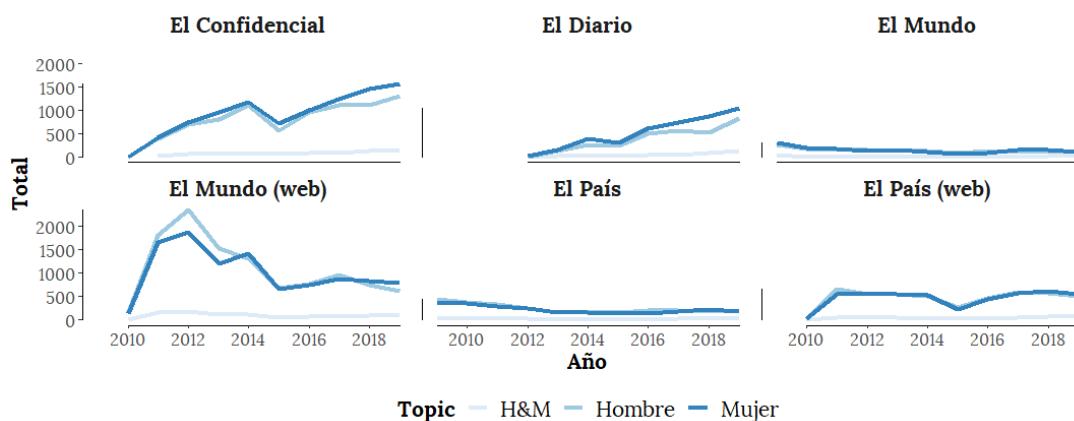
Fuente: Elaboración Propia

Aún teniendo unas temáticas mucho más genéricas, donde cualquier noción pudiera acoplarse a la búsqueda, en la Figura 11 no se aprecian, de forma manifiesta, unos cambios significativos. Principalmente por la influencia que ejerce *El Mundo* (versión web) en los primeros años –como se aprecia en la Figura 12–, pero también por la incorporación de nuevos medios a la muestra. Aun así, lo más

³⁷ Se ha de señalar que, si bien el periodo de análisis comienza en 2009, algunos de los medios no aparecen en la muestra hasta años posteriores. Estos serían el caso de *El País* y *El Mundo* en su versión web y *El Confidencial* (recogidos por My News a partir de 2010), además de señalar que *Eldiario.es* nace en 2012.

relevante de la Figura 11 sería cómo “mujer” cobra una mayor prominencia durante el último lustro. Finalmente, si nos fijamos en la cohabitación de ambos términos en el mismo titular (“H&M”), esta se considera residual y poco relevante, aunque sí significativa para entender posibles conjeturas una vez observemos más adelante las temáticas presentes.

Figura 12. Agregados por periódico



Fuente: Elaboración Propia

Sin embargo, la Figura 12 sí agrega algunas líneas interesantes. Mientras que los dos medios de versión papel se comportan totalmente similares, los nativos digitales (a saber, *El Confidencial* y *ElDiario.es*), son los que implementan unas fluctuaciones más diferentes, sobre todo si se compara con las versiones web de *El Mundo* y *El País*. Aunque *El Mundo* (web) tiene un comportamiento totalmente anómalo, influenciado quizás, por un cambio de tendencia en su línea editorial³⁸. Pese a ello, todos los medios online tienen al 2014 como fase de descenso y al 2015 como un año de ascenso paulatino.

Con el fin de proveer una mayor consistencia a nuestras contribuciones descriptivas, tomamos en consideración la utilización de diferentes modelos de *Topic Modeling*. Para ello, lo primero que hicimos fue dividir el corpus en dos subsets temporales (2009-2014 y 2015-2019)³⁹, con los que pretendemos encontrar determinadas evoluciones a lo largo del tiempo. De este modo, y después de chequeos tantos automatizados, como manuales, se consideró que el corpus tenía unas temáticas coherentes fijando $k = 3$. Así, en primera instancia, lo que llevamos

³⁸ Aunque no existe un razonamiento claro, a nuestro juicio, esto se debe a un cambio de posturas dentro del medio. La sustitución del clásico director de la cabecera, Pedro J. Ramírez, pudo provocar, al menos a tenor de los datos mostrados, un cambio de tendencia en la línea editorial del periódico. Produciendo así que la versión web fuera amoldándose a unos derroteros distintos de los planteados por el otrora director, y emparejando más su tendencia a la del otro medio hegemónico analizado, *El País*. Aun así, son factores que deberían ser abordados por otras investigaciones ya que, de acuerdo con la profundidad del estudio que debería llevarse a cabo, no tienen cabida en este trabajo.

³⁹ Nota aclaratoria. Se realizó esta subdivisión en dos lustros al considerar que el último de estos es el que podría aportar una mayor diferenciación entre ambos. Principalmente, debido a los cambios y manifestaciones acaecidas en el seno del feminismo. Además, se ha de considerar al último de estos lustros como un periodo de cambio en la escena política y social española.

a cabo fue la aplicación de los modelos orquestados mediante *tex2vec* (Selivanov et al., 2020) y su correspondiente visualización recurriendo a *LDavis* (Sievert y Shirley, 2014).⁴⁰

Tabla 1. Top 30 términos más prominentes (2009-2014)⁴¹

Tokens
muere, detenido , fallece, mujer , anos, matar, mas, cadáver, tras, prisión, hallan, grave, herido , acusado , incendio, herida , primera , cárcel, muerta , detienen, accidente, ser, mato , hallado , pareja, muerto , agredir, condenado , caer, playa

Fuente: Elaboración Propia

Nota. Se señalan en negrita las palabras que aportan una mayor consistencia a la temática

Tabla 2. Top 30 términos más prominentes (2015-2019)

Tokens
muere, detenido , tras, anos, fallece, matar, mujer , mas, prisión, cadáver, primera , acusado , incendio, hallan, grave, pareja , ser, herida , dia, muerta , agredir, herido , carcel, mundo, mata, detienen, vivienda, coche, dos, mato

Fuente: Elaboración Propia

Nota. Se señalan en negrita las palabras que aportan una mayor consistencia a la temática

En estas primeras dos tablas mostramos la probabilidad de palabras más prominentes en base al número de temáticas establecidas. De esta forma, contemplamos que “mujer” es un término más probable que “hombre” en los dos lustros, en referencia a esta subdivisión por temáticas. En ellas, en proporción, el término “mujer”, a pesar de ser un patrón de búsqueda inicial como era “hombre” y de acuerdo con las Figuras 11 y 12, cuenta para el algoritmo con una presencia más consistente que “hombre”. Esta primera revisión indica un factor claro, las temáticas principales de la muestra van a ir enfocadas en torno a la violencia (ej. “agredir”), el homicidio/deceso (ej. “matar” o “fallece”) y problemas de ámbito legal (ej. “carcel”).⁴²

⁴⁰Nota aclaratoria. Se decide no llevar a cabo trabajos individualizados por medio ya que las muestras de palabras, al tratarse de titulares, no arrojarían una robustez muy contundente. De ahí que, en este particular, se prime más por la muestra global y no por la pormenorizada por cabecera.

⁴¹ Prominencia (term w) = frecuencia (w) * [sum_t p (t | w) * log (p (t | w) /p (t))] por topic t; véase Chuang et al. (2012, p.2)

⁴²Nota aclaratoria. A lo largo de las siguientes tablas y figuras se usarán tokens, considerando a estos como palabras limpias sin sus correspondientes acentos. Esta puntualización se hace por dos motivos; el primero el aclarar que no se utilizarán los citados acentos; segundo, porque la no presencia del propio acento puede hacer que la palabra pueda coger acepciones variadas y, a la par, esto pueda agregar un cambio de tendencias en cuanto a la mayor presencia de determinados tokens (ej. “género” y “genero”, serán recogidas las dos como “genero”)

Tabla 3. Top 30 términos más relevantes por topic ($\lambda = 1$) (2009-2014)⁴³

Topic 1	Topic 2	Topic 3
hombre , mujer, detenido , anos, matar, dos, prision, pareja , detienen, muerte, tras, acusado , mato, cárcel, detenida , mata, ex , hija , policia, tres, agredir, condenado , hijo , detenidos , hijos , marido , piden, casa , condenan, robar	hombre , mujer, muere , anos, tras, fallece, cadáver , ser, hallan, grave, herido , incendio , muerta, herida , pareja , casa , accidente , hallado , mata, vivienda , muerte, muerto, coche , dos, caer , playa , resulta, piso , madrid, apunalada	mujer , hombre, mas, primera , mundo , violencia , dia, fuerte, dice, espana, juez, barcen , nuevo, da, ex, caso, vez, euros, vida, luz, pide, nueva , pp , rico, solo, policia, dos, gobierno , eeuu

Fuente: Elaboración Propia

Nota 1. Se señalan en negrita las palabras que aportan una mayor consistencia a la temática

Tabla 4. Top 30 términos más prominentes en cada topic ($\lambda = 0$) (2009-2014)

Topic 1	Topic 2	Topic 3
acusado , agredir, condenado , violar, abusar, ingresa, amenazar, sexualmente , fianza, jurado, judicial, alejamiento , juzgan, violador , disposicion, quemar, robarle, maltratar, golpear, nueve, sexuales , abusos, pistola, queria, absuelto, provisional, atropellar, supuesto, acuchillar, malos	hallan, grave, herido, incendio, herida, accidente , hallado , muerto , playa , resulta, apunalada , atropellada , apunalado , atropellado , encuentran, ahogado , desaparecido , signos, recibir, camión , heridos, rescatan, vehículo , via, chocar, tren , turismo, volcar, salirse, tractor	mundo , vez, rico, historia , segun, quiere, machista , rajoy , confianza, india, igualdad , derechos , puede, tiempo, ley española, mejor, mexico, ministro , política , genero , cada, papa, carlos, clave, garcia, rica , blanco, cancer , internacional

Fuente: Elaboración Propia

Nota 1. Se señalan en negrita las palabras que aportan una mayor consistencia a la temática

Los resultados que arrojan las Tablas 3 y 4 comienzan a ser muy reveladores. Estos parecen divisarse de forma nítida, sobre todo al comparar los tópicos que las tres temáticas pueden apuntar en direcciones distintas, aunque la primera y la segunda puedan estar coaligadas. Por un lado, el relacionado con el Tema 1, que sería las violencias de pareja y también las de ámbito sexual, de ahí que aparezcan palabras como, por ejemplo: “pareja”, “hijos”, “marido”, “alejamiento”, “maltratar”, “sexualmente”, “abusos”, “violador”. Una esfera que podría estar íntegramente enlazada con el Topic 2, pero que, de acuerdo con las palabras que están más presentes en este tema -Tabla 4- y no en otros, asumimos un corpus distinto. Esto puede verse por ejemplo en tokens que son sintomáticos (ej. “incendio”, “playa”, “ahogado”, “atropellada”, “vehículo”, “turismo”, “chocar”). A nuestro juicio, y persistimos en el aspecto probabilístico del algoritmo, este punto señalado en el Topic 2 tiende más a un tipo de incidente relacionado con violencias como los asaltos, los accidentes de coche u otros de índole de lo que vendría a denominarse en el argot comunicativo como sucesos. Por su parte, el Tema 3 es el único que señala a la “mujer” como el factor más relevante. Esto, y con arreglo a las palabras obtenidas, demuestra que el Tópico 3 ofrece como lugar común el ensalzamiento del papel de la mujer (ej. “primera”, “historia”), aunque íntegramente relacionado con factores mediáticos de aquel lapso temporal (ej. “barcen”, debido en este caso a la presencia del conjunto “la mujer de Bárcenas”).

⁴³ Relevancia (term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t) / p(w)$; véase Sievert y Shirley (2014, p.4)

Tabla 5. Top 30 términos más relevantes por topic ($\lambda= 1$) (2015-2019)

Topic 1	Topic 2	Topic 3
hombre , mujer, detenido , anos, matar, prision, pareja , dos, tras, policia, acusado , mata, detienen, cárcel, mato, muerte, agredir, detenida , detenidos, hijo , agresión, hija, tres, expareja , condenado , detiene, asesinato, sexual, barcelona, Madrid	mujer , hombre, mas, primera , ser, día, mundo , violencia, espana, anos, así, vez, solo, pide, historia , gobierno, vida, trump , nueva, dice, caso, nuevo, fuerte, papel , primer, ee, uu, denuncia, gran	hombre , mujer, muere, anos, tras, fallece , ser, cadáver, muerte, incendio, hallan, grave, muerta , casa, herida , coche, vivienda, herido , madrid, playa, investigan, accidente, caer, pareja , violencia, muerto, hallado , piso, policia

Fuente: Elaboración Propia

Nota 1. Se señalan en negrita las palabras que aportan una mayor consistencia a la temática

Tabla 6. Top 30 términos más prominentes en cada topic ($\lambda= 0$) (2015-2019)

Topic 1	Topic 2	Topic 3
acusado , violar, menores, sexualmente , asesinar, abusar, apunalar, fianza, amenazar, provisional, alejamiento , jurado, ingresa, golpear, presunta , suicidarse, quemar, quemar, agredir, maltratar, secuestrar, disposicion, apunalo, juzgan, delante, estafar, amenaza, confiesa, retener, arrestado , robarle	primera , mundo , vez, solo, historia , pp, papel, ee, uu, puede, sanchez , igualdad , rico, derechos , sera, presidente, mejor, quiere, psoe , instituto , nombre, nunca, internacional, mayor, lucha, premio , gana, ana, clave, tiempo	incendio, hallan, grave, herida , accidente, caer, hallado , hallada , desaparecido , desaparecida , atropellada , atropellado , apunlada , encuentran, ahogado , recibir, signos, resulta, heridos, camion , mueren, heridas, cayo, muertos, volcar, violenta, piscina , zona, caerle, chocar

Fuente: Elaboración Propia

Nota 1. Se señalan en negrita las palabras que aportan una mayor consistencia a la temática

Si ahora contemplamos la distribución del siguiente quinquenio (2015-2019), tablas 5 y 6, vemos como es muy similar al anterior. De nuevo los topics se distribuyen entre temáticas muy parecidas⁴⁴. De nuevo, el Topic 1 retrata las violencias entre parejas y también, como antes, las sexuales. Algo que, de acuerdo a la muestra, podría estar asociado, bien por la naturaleza propia de este tipo de agresiones, bien por la construcción del relato -el titular en este caso- por parte de los medios. Ocurre igual con el Topic 3, que sería el 2 de las tablas 3 y 4. Estamos ante un suceso externo al ámbito conyugal o familiar, si bien pudieran encontrarse palabras comunes con el Topic 1, lo cierto es que de nuevo la probabilidad muestra hechos más relacionados con las reyertas, los asaltos, accidentes, etc. Para concluir, nos quedaría mencionar el Tópico 2, en el cual sucede exactamente lo mismo que su conjunto paralelo del primer lustro. En este intervalo parece incluso más enfocado a definir los satélites que engloban al ámbito de la mujer. El cambio de lustro deja a un lado lo que implicaba el Caso Bárcenas en este topic, pero agrega un factor también temporal, Donald Trump. Algo que ocurre también con la entrada “psoe” y “sanchez”, y la ausencia en este lustro de “rajoy”. Este último elemento podría estar fuertemente coaligado a la

⁴⁴Nota aclaratoria. Aunque el número del topic cambie (ej. Topic 3 (2009-2014) por Topic 2 (2015-2019)), esto no implica que varíe su importancia -o sea, el número no implica mayor presencia-. De hecho, como se ejemplifica en las tablas 7 y 8, la probabilidad de los temas no se rige en función del número que el algoritmo le aplique al topic. Por lo que un topic 2 puede tener una mayor presencia que el 1 o el 5 que el 2.

desaparición del Caso Bárcenas. Así, también subrayamos la aparición de conjeturas ligadas a la izquierda, como es la inclusión dentro del Topic 2 del token “instituto”, en clara referencia a Instituto de la Mujer, otro ámbito muy arraigado al PSOE desde su constitución en 1983 por el gobierno de Felipe González.

Con el propósito de poder ampliar el análisis de esta muestra, decidimos llevar a cabo otro de los paradigmas de máquina expuestos. En esta ocasión hacemos uso del denominado BTM. Con él, al igual que hicimos con el modelo anterior, fijamos el número de $k = 3$.

Tabla 7. Estimación ‘theta’⁴⁵ (2009-2014)

Topic 1	Topic 2	Topic 3
0.8383	0.0784	0.0832

Fuente: Elaboración Propia

Tabla 8. Estimación ‘theta’ (2015-2019)

Topic 1	Topic 2	Topic 3
0.7428	0.1523	0.1047

Fuente: Elaboración Propia

Estas dos sucintas tablas de los subsets revelan varias cosas; una, el orden de los tópicos no implica la probabilidad de la presencia; dos, que como es evidente, hay temas más imperantes que otros.

Tabla 9. Selección de Tokens por BTM (2009-2014)

Topic 1	Topic 2	Topic 3
mujer, hombre , anos, detenido , tras, muere, dos, pareja , matar, muerte, ser, mata, detienen, prision, policia, ex, casa, fallece, tres, acusado, detenida , mato, carcel, mas, hijo, hija, cadáver, muerta, marido , hallan	mujer , mas, primera , mundo, juez, barcen as, vez, rico, pide, dice, fuerte, pp , espana, caso, ruz , pujol, psoe , dia gobierno , nuevo, historia , ex, según, rajoy, aborto , euros, gonzalez, imputa, millones, frente	hombre , tras, muere, anos, fallece, grave, herido , ser, cadáver, caer, herida , resulta, hallan, accidente, incendio, coche, atropellado , playa, vivienda, atropellada , arma, ahogado , muerto, camion , recibir, blanca, hallado , salirse, turismo, tractor

Fuente: Elaboración Propia

Nota 1. Se señalan en negrita las palabras que aportan una mayor consistencia a la temática

⁴⁵Véase Wijffels et al. (2020, p.2) como explicación al respecto, a saber: “La estimación del modelo temático se realiza con el algoritmo de muestreo de Gibbs. Donde se proporcionan estimaciones para $P(w | k) = \phi$ y $P(z) = \theta$.”

Tabla 10. Tokens presentes por BTM (2015-2019)

Topic 1	Topic 2	Topic 3
mujer, hombre , anos, detenido , tras, muere , ser, dos, mas, policia, muerte, pareja , matar, prision, violencia , casa, Madrid, fallece, cárcel, detienen, acusado , hijo, mata, tres, mato, cadáver, detenida , primera, hija , pide	hombre , anos, tras, detenido , mujer , muere, pareja , dos, matar, prision, fallece, grave, policia, acusado , ser, muerte, herido , herida , cadáver, incendio , agredir, casa , tres, coche , madrid, cárcel, vivienda , hijo , hallan, mata	mujer , primera , mas, día , violencia , mundo , vez, ser, gobierno , española, pide, pp, historia , papel , dice, nueva, frente, sanchez , genero , será, solo, asi, caso, derechos , ana, igualdad , premio , psoe , rico, instituto

Fuente: Elaboración Propia

Nota 1. Se señalan en negrita las palabras que aportan una mayor consistencia a la temática

De acuerdo con las Tablas 9 y 10, y lo testado en las tablas previas con *text2vec*, este con BTM refrenda lo previamente expuesto, pero inserta unas dinámicas más lúcidas. Por un lado, y gracias a la estimación ‘*theta*’, podemos comprobar cómo la probabilidad de los tópicos ha fluctuado a lo largo del tiempo. Sin lugar a dudas, la nota más llamativa es que, como ya se había predispuesto en las tablas 1 y 2, la significación de los nodos que rodean a “mujer” es más sólida que la que flanquean a los de “hombre”, de ahí que las temáticas lo vayan difuminando del espectro a favor de “mujer” (véase la evolución de las temáticas de las tablas 9 y 10). Otro tono a destacar sería el de cómo el Tópico 3 del segundo lustro, el referido a las temáticas de la mujer, se fortalece pasando de una estimación de 0,07 (2009-2014) a una de 0,10 (2015-2019). De nuevo, y al igual que ocurría con los otros dos tópicos; violencia de pareja y el de sucesos, se destacan, en mayor medida, por situar a “hombre” como una partícula activa, siendo esta la que ejerce la violencia y la que recibe mayores castigos que “mujer”. Esta idea se concibe a tenor de los géneros de palabras, como por ejemplo: “acusado”, “detenido” o “condenado”. Lo que indica, con arreglo a la Figura 11 de agregados, donde se ve claramente que “hombre” y “mujer” apenas coinciden en un titular, es que “mujer” y la conceptualización de esta (ej. “asesinada”, “violada”, “herida”), sería un agente más pasivo de la violencia que “hombre”.

Tabla 11. Probabilidad de Tokens seleccionados por género BTM (2009-2014)

Token Topic 1	Probabilidad	Token Topic 3	Probabilidad
detenido	0.0124	herido	0.0213
acusado	0.0036	herida	0.0128
detenida	0.0036	atropellado	0.0080
muerta	0.0029	atropellada	0.0062
marido	0.0029	ahogado	0.0056
condenado	0.0021	muerto	0.0054
asesinada	0.0017	hallado	0.0048
herida	0.0016	apunalado	0.0044
muerto	0.0016	desaparecido	0.0032
hallada	0.0012	arrollado	0.0025

Fuente: Elaboración Propia

Tabla 12. Probabilidad de Tokens seleccionados por género BTM (2015-2019)

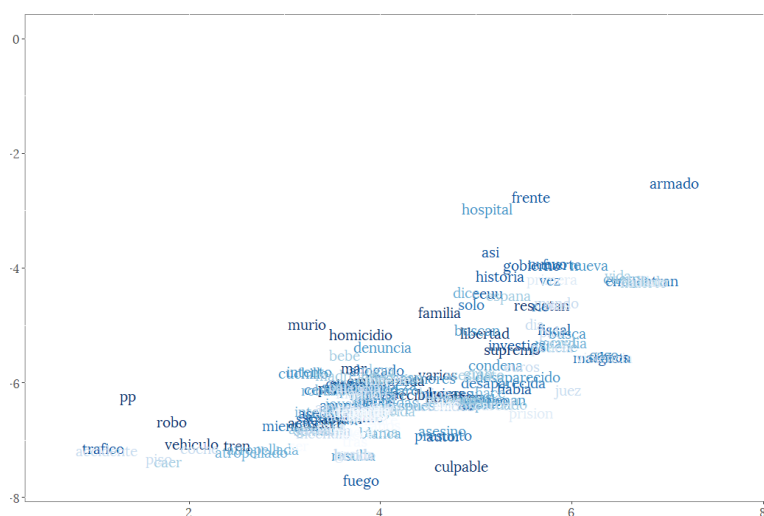
Token	Topic 1	Probabilidad	Token	Topic 2	Probabilidad
detenido		0.0097	detenido		0.0250
acusado		0.0023	acusado		0.0073
detenida		0.0019	herido		0.0060
marido		0.0017	herida		0.0057
asesinada		0.0017	muerta		0.0036
muerta		0.0016	detenida		0.0034
herida		0.0012	condenado		0.0029
condenado		0.0010	atropellado		0.0021
muerto		0.0009	hallado		0.0021
herido		0.0009	atropellada		0.0017

Fuente: Elaboración Propia

Como venimos precisando, la probabilidad de la muestra apunta a una referencia clara, el género masculino, en este caso representado por “hombre”, lleva incardinado en mayor medida la variable de ejecutor de la violencia -visto tanto en el tópicos de violencia de pareja, como en el de sucesos-. A este respecto, el conglomerado del género femenino, representado en este supuesto por “mujer”, llevaría consigo la carga de receptor en mayor proporción de la violencia. Al menos en lo que respecta a violencia de pareja, temática a su vez con mayor presencia dentro de nuestra investigación; con una probabilidad como muestra ‘theta’ de entorno a casi 4/5 de los titulares de ambos lustros.

Para ampliar estos resultados decidimos implementar un modelo de incrustado de palabras. Se ha de reseñar que este tipo de modelos, y más en este caso concreto debido a la partícula estudiada, sufren por el influjo impuesto por los sesgos de género (Ethayarajh et al., 2019), de ahí que, nuestra misión, dentro del NLP, sea la interpretación, filtrado y recomposición de posibles estereotipos que el esquema algorítmico devuelva.

Figura 13. Incrustación de los principales tokens



Fuente: Elaboración Propia

Como vemos en la Figura 13, el algoritmo de *embedding* retrata de una forma muy visual algunas de las ideas que estamos desgranando. Así, comprobamos cómo hay ciertas palabras que son coetáneas y que en función de su uso llevan correlacionadas otras, lo que hace que se creen clústeres en función de esta correlación. Este sería el caso de, por ejemplo, los tokens “vehículo”, “coche”, “tren”, “atropellada” y “atropellado”, que todos comparten una cercanía de coseno que invita a esa relación semasiológica que hemos expuesto. Esto también puede verse en relaciones como “asesino” o “presunto”, íntegramente ligadas. Por medio de esta conjetura podemos reportar un mayor entendimiento al porqué determinadas palabras aparecían presentes junto con otras en los *topics modeling*.

Para finalizar los aportes de incrustado se desarrolla un modelo de analogías mediante coseno. Lo que llevamos a cabo es una selección de varios vectores, como se ejemplifica en el caso de “España” -> “Español” : “Italia” -> “X”, con los que poder obtener algunos de los supuestos probabilísticos que los titulares albergan. Dicho de otro modo, para producir la Tabla 13 lo que hacemos previamente es seleccionar dos vectores (x, y), como decimos, “España” y “Español”, que serán anexionados, de este modo, al seleccionar un tercero “Italia”, lo que buscamos es que el algoritmo nos devuelva la respuesta del cuarto vector en función de los previamente fijados.

Tabla 13. Selección de analogías de Coseno (θ) por incrustación

Input	Output
“hombre” -> “asesina” : “mujer” ->	“intención”
“hombre” -> “desaparece” : “mujer” ->	“protagonista”
“mujer” -> “feminismo” : “hombre” ->	“bueno”
“hombre” -> “feminismo” : “mujer” ->	“comunidad”
“hombre” -> “muere” : “mujer” ->	“calcinada”
“mujer” -> “muere” : “hombre” ->	“fallece”

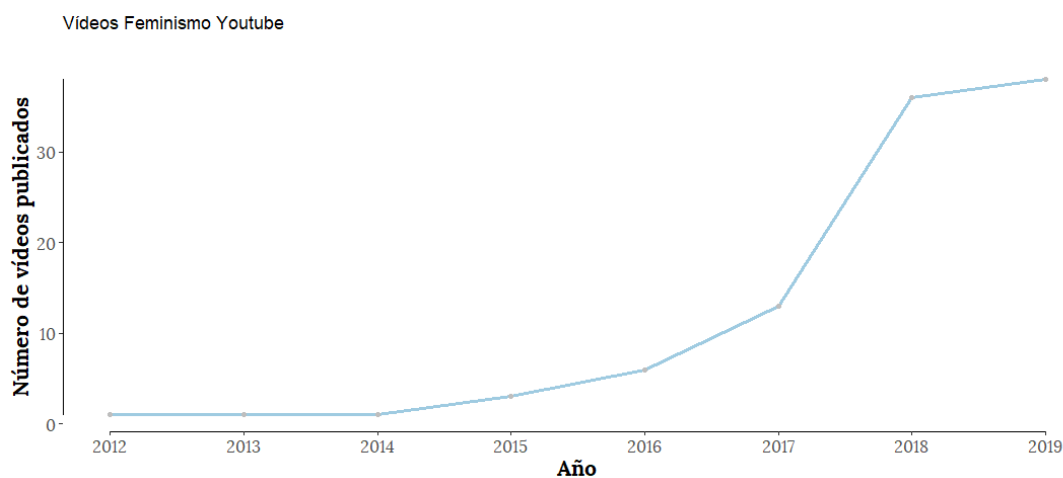
Fuente: Elaboración Propia

Versando lo expuesto en la Tabla 13 vemos casos sintomáticos. Un paradigma de esto sería: “hombre” -> “desaparece”: “mujer” -> “protagonista”. Nuestra cuestión, al plantear esa entrada, era comprobar qué incrustación era la más cercana a “mujer” una vez que [un] “hombre” “desaparece”. El algoritmo, sin embargo, muestra la siguiente idea: la desaparición de “hombre” beneficia a “mujer” ganando peso su figura y convirtiéndose en “protagonista”. O sea, la selección de los anteriores vectores señala que hay una relación espacial entre la pérdida de peso mediático del “hombre” y en beneficio de “mujer”. Otro punto interesante sería el que aporta la analogía: “hombre” -> “asesina”: “mujer” -> “intención”. Este planteamiento también deja abierta la puerta a interpretaciones variadas, pero a tenor de los diferentes análisis expuestos, consideramos que es el resultado de: “hombre” lleva a cabo esa agresión que termina en asesinato, mientras que “mujer” con arreglo a las revisiones ejecutadas, lo intenta. A pesar de ello, esta interpretación de los resultados de la Tabla 13 está sujeta a supuestos y revisiones amplias que deberían contar con un mayor número de datos y cruces

para poder ser más congruentes, así como al correspondiente papel que el investigador ha de aportar.

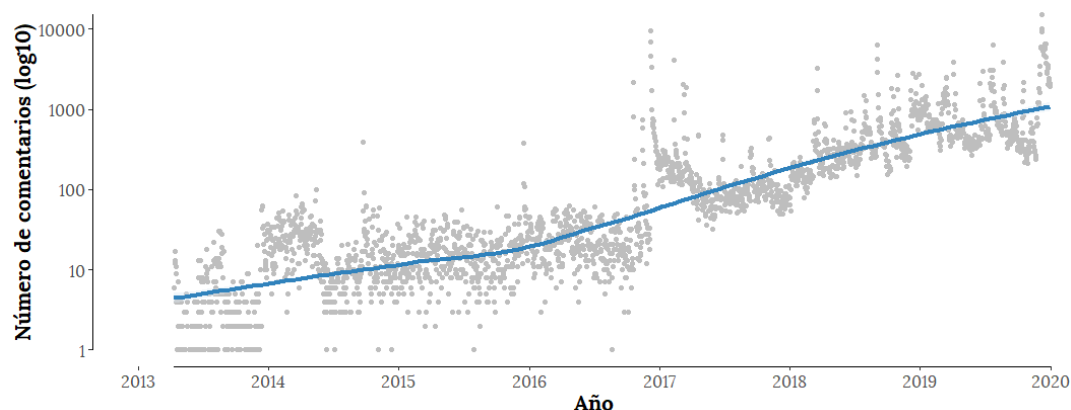
Para continuar dando forma a este trabajo nos gustaría apoyarnos en los datos recabados dentro del proyecto de Serrano-Contreras (2021) sobre YouTube. Aunque la misión aquí es llevar a la práctica los resultados congregados en torno al uso de nuestro índice de polarización (Serrano-Contreras et al., 2020) (a saber, figuras 16 y 17), también nos gustaría empezar recabando algunos valores descriptivos de esta muestra (a saber, figuras 14 y 15).

Figura 14. Fecha de publicación de los vídeos más vistos en YouTube sobre “Feminismo”



Fuente: Elaboración Propia

Figura 15. LOESS⁴⁶ (log10) de los comentarios presentes en los vídeos seleccionados



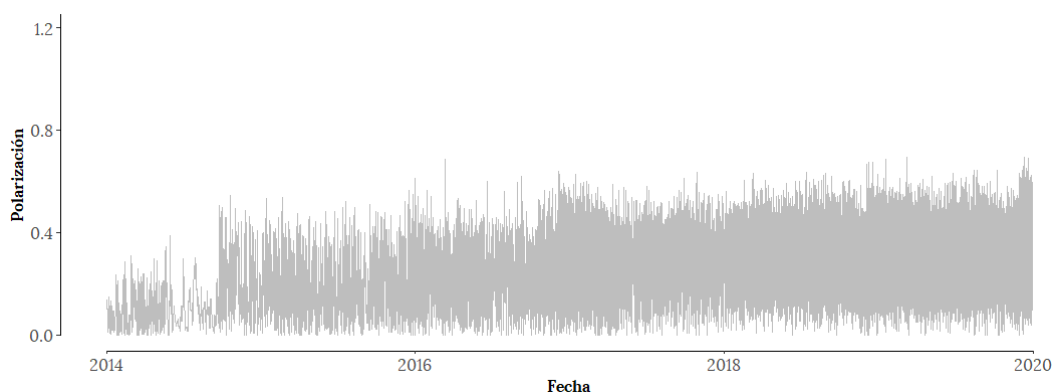
Fuente: Elaboración Propia

⁴⁶ Locally Weighted Regression and Smoothing Scatterplots.

Con arreglo a las posiciones que venimos arguyendo, los anteriores paquetes de agregados (figuras 14 y 15) manifiestan también una clara tendencia alineada con precedentes figuras y tablas. Aunque se ha de señalar que este conjunto se focaliza en un periodo de evaluación más corto, donde los rumbos alcistas se abren más abruptamente en el último lustro. Por medio de esta coyuntura, los datos de YouTube se descubren muy reveladores. De los 100 vídeos más vistos en la historia, teniendo en cuenta los parámetros de búsqueda (véase T.2 epígrafe 3.5), la mayor parte de ellos se publican a partir de 2016, Figura 14. Ocurrendo casi de forma similar con los ratios del número de comentarios, Figura 15. Por lo que, no solo la mayor cantidad de vídeos se subieron a la plataforma a partir del último quinquenio, sino que la discusión acerca de ellos también se exacerbó en consonancia⁴⁷. Lo que también proporcionan los datos es una revolución sobre cómo interactuar, participar y relacionarse en la plataforma. Como se señala en Serrano-Contreras (2021) (incluido en los anexos de este trabajo), los cambios cosechados en el plano de YouTube son bastante apreciables. Por una parte, estaría la predominancia que coge la lengua española dentro de la discusión en los comentarios de la red social. Mientras que en los primeros años de la muestra el idioma más usado era el inglés, a pesar de ser vídeos en español, a partir del gran salto cuantitativo en el número de vídeos publicados, los comentarios se comienzan a realizar en español.

A continuación, pasamos a implementar nuestro índice (Serrano-Contreras et al., 2020). Para ello primero se realizó un puntaje con Cruz et al. (2014). Esa calificación fue la que se incluyó en el índice a modo de testear la posible polarización existente (Figura 16).

Figura 16⁴⁸. Polarización mediante índice



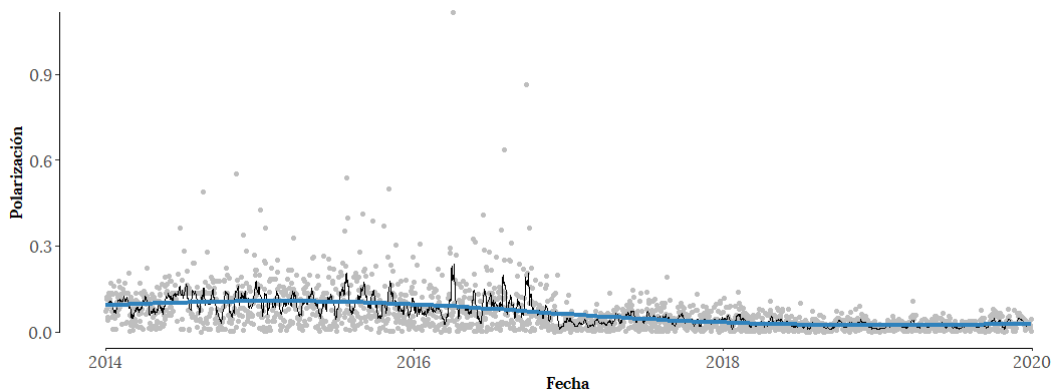
Fuente: Elaboración Propia

⁴⁷Queremos decir con esto que, a pesar de factores algorítmicos externos (véanse las modificaciones implementadas por Google en Davidson et al. (2010) y Covington et al. (2016)), si los vídeos de periodos anteriores -recordemos que YouTube nace en 2005- hubieran sido lo suficientemente famosos, su número de visualizaciones y número de comentarios, se habría disparado en años previos y no que, como muestran las Figuras 14 y 15, el rumbo cambia en fechas de carácter más simbólicas para la esfera analógica.

⁴⁸Se ha de resaltar que, como se indicó en Serrano-Contreras (2021), hay un tramo en el que el idioma dominante es el inglés, por lo que se decidió prescindir de esos comentarios.

Gracias a la dispersión resultante comprobamos que no existe, de forma evidente, polarización. A pesar de que haya subidas, ninguna se dispara en el tiempo y con la rotundidad lo suficientemente firme –más allá de (1)-, como para poder hablar de que los comentarios de cada vídeo se estaban distanciando con respecto a la mediana del mismo.

Figura 17. (LOESS y media móvil) Polarización mediante índice tomando el año

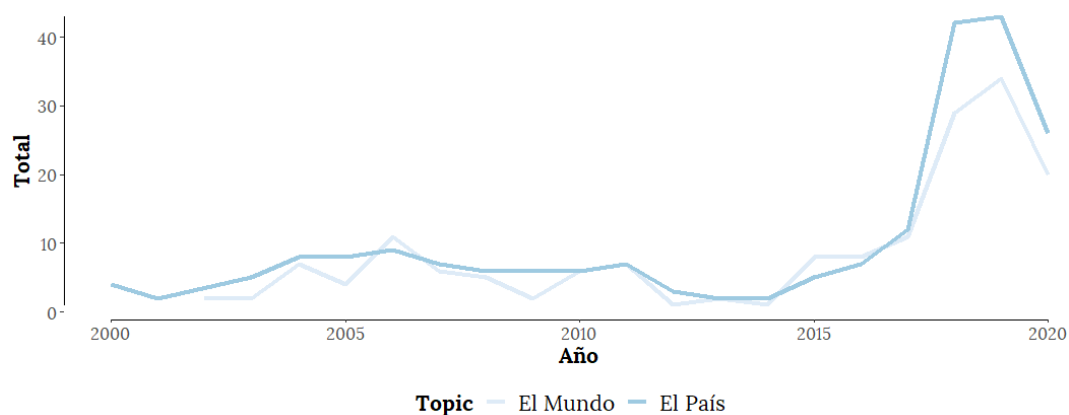


Fuente: Elaboración Propia

Por último, y a modo de prueba, en lugar de obtener la mediana por vídeo, lo que se hizo, fue simular el conjunto como un total único, ya que la temática era la misma. Este hecho propició un factor interesante; a mayor número de comentarios, menor es la polarización, lo cual generó una correlación negativa entre ambas. No deja de ser sintomática esta idea, ya que puede interpretarse de dos modos distintos. Por un lado, estaría la conjetura de que la temática fue poco a poco adquiriendo una tónica nada polarizante, y de ahí ese resultado. Del otro, y de conformidad de los datos y al no haber fijado una dirección de puntaje, el índice, a través de la mediana, lo que hace es estandarizar el resultado del distanciamiento entre los frentes -en caso de haberlos-. Aun así, este hecho invita a que se deba de seguir ahondando en esta y otras formas de medición, sobre todo en el rating o enmarcado de cada una de las partículas analizadas.

Para finalizar esta travesía por los resultados recopilados, hacemos reseña de los extraídos de una de las muestras supletorias, con la que pretendemos dar un mayor empaque y amplitud a nuestro proyecto. En este caso hablamos de la línea T.5 del epígrafe 3.5.

Figura 18. Titulares con Feminismo



Fuente: Elaboración Propia

En primer lugar, y como advierte la Figura 18, nos damos cuenta que estamos ante un fenómeno que, de forma clara y evidente, es novedoso para la línea editorial de los medios -en su versión papel-. Donde la mayoría de los artículos han sido publicados a partir de los años 2014-2015. Lo que refrenda la tónica que se viene mostrando en todos los estudios implementados. Otros detalles serían cómo *El País* (n = 209) con respecto a *El Mundo* (n = 166) muestra un mayor interés en la temática, además de ser el primero de los dos que pone el foco sobre el ámbito en cuestión.

Tabla 14. Similaridad de Coseno (θ) mediante DFM⁴⁹

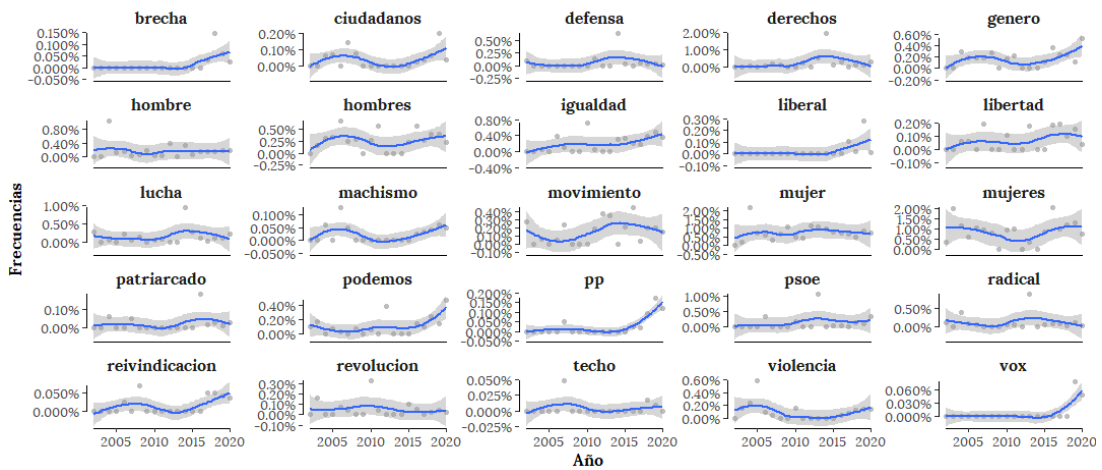
Medio por década	Similaridad (θ)
<i>El País</i> - <i>El Mundo</i> (2000 - 2009)	0.8024724
<i>El País</i> - <i>El Mundo</i> (2010 - 2020)	0.9506128

Fuente: Elaboración Propia

Para continuar con la comparación de medios construimos la Tabla 14. En ella vemos cómo, en el primer periodo, las similitudes no son tan grandes como sí ocurre en el segundo. Como apuntamos, estas discrepancias pueden estar debidas a la mayor cantidad de palabras en uso y no a que realmente hay una diferencia de hasta casi 0.15 puntos entre un lapso temporal y otro, aunque esto tampoco puede ser descartable. De hecho, para aportar un mayor sostén a tales diferencias, decidimos aplicar un modelo de frecuencias con el que comprobar las desviaciones.

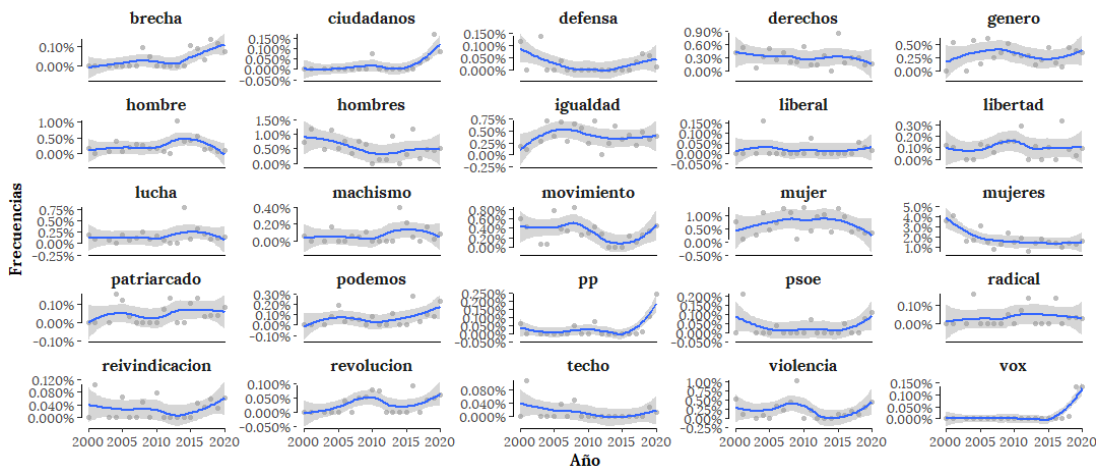
⁴⁹ DFM es una matriz de características del corpus donde recabar las propiedades de los documentos insertados dentro. En él podemos obtener el número de tokens por archivo, así como otra serie de especificidades obtenidas directamente de los metadatos y variables conjugadas. Para un conocimiento más pormenorizado véase Benoit et al. (2021). En él se encontrarán todas las singularidades habidas sobre la librería aplicada para el conformado de esta matriz.

Figura 19. Frecuencia de usos y (LOESS) mediante DFM – El Mundo



Fuente: Elaboración Propia

Figura 20. Frecuencia de usos y (LOESS) mediante DFM – El País



Fuente: Elaboración Propia

Lo que muestran las figuras 19 y 20, son los resultados obtenidos del estudio de determinadas bolsas de palabras seleccionadas previamente. Esta selección ha sido implementada con el fin de chequear si las posibles facciones del feminismo o las ideologías de uno u otro espectro han tenido mayor presencia en un medio u otro. Si bien no existen grandes diferencias que tracen una gran fisura entre los periódicos, sí es cierto que hay notas que revelan los dejes propios de cada línea editorial. Por un lado, abordando las semejanzas, los nombres de los partidos políticos incrustados (PSOE, PP, Vox, Podemos y Ciudadanos) en la bolsa de palabras, son representados con similares tendencias en ambos medios, a excepción del PSOE, que tendría una mayor presencia en los artículos publicados por *El País*. Sin embargo, en las palabras “ciudadanos” y “podemos”, consensuando que pueden presentarse de forma distinta a la referencia del nombre de los partidos, sí parece que la tendencia que ambas tienen está relacionada con la aparición de los partidos en la esfera política. Algo que también le ocurre, aunque

de forma más imperativa, a “pp” y “vox”. Por otro lado, si nos fijamos en tokens como “genero”, comprobamos que ha tenido una tendencia cambiante tanto para *El País* como para *El Mundo*, con fluctuaciones similares a lo que ya mostró la Figura 8 de los debates del estado de la nación, es decir, subidas durante la primera legislatura de Zapatero, con un descenso durante la crisis económica y alcanzando de nuevo una subida con el final de esta. Sin embargo, al fijarnos en “brecha”, esta no tiene una evolución oscilante, sino que, para los dos medios analizados, comienza su escalada en los años de mayor número de artículos sobre el tema (2014-2015) y continúa en los años siguientes. Otros comportamientos que unen a los medios son cómo las frecuencias de “hombre” u “hombres”, están infrarrepresentadas si son comparadas con “mujer” o “mujeres”, aunque en un primer lapso temporal (2000-2010), para *El País*, “hombres” sí tuviera una concepción más imperante. Otro factor, en este caso de distanciamiento entre periódicos, serían dos palabras clave “igualdad” y “liberal”⁵⁰. Al tiempo que *El Mundo* se decanta, en especial en los últimos años, por un mayor uso de “liberal” con respecto a *El País*, este último hace lo propio con “igualdad”, aunque no con tendencias similares. Para *El País*, “igualdad” es un término ya inserto en su idea de feminismo desde el comienzo de la muestra. Sin embargo, para *El Mundo*, el término “igualdad” va ganando peso a lo largo del tiempo.

⁵⁰ Preferimos decantarnos por exponer “liberal” como contraposición de “igualdad”, ya que “libertad” tendría un alcance más amplio y “liberal”, en cambio, está inserto dentro de facciones del feminismo que suponen una discrepancia.

Conclusiones

A lo largo de estas páginas hemos congregado una exposición sobre determinados factores que atañen a la sociedad del siglo XXI. Todos ellos, y como se recoge en el título de este proyecto, nos han sido útiles para tratar de entender qué acaece en nuestras relaciones humanas. A su vez, y como hemos expuesto, la amplitud de ámbitos ha servido para, de un modo u otro, presentar unas cuestiones, objetivos e hipótesis que, a lo largo de esta investigación, y de forma conclusa en este apartado, se han ido alcanzando, respondiendo, verificando y refutando.

En primer lugar, comenzaremos la exposición del tratamiento de las hipótesis como forma de cotejar nuestras indagaciones. Por tanto, nos gustaría partir del planteamiento expuesto en nuestra H.1., donde apuntábamos a las olas dominantes que han configurado el feminismo en España durante las últimas dos décadas. Es más que evidente que a la luz que arrojan los datos y las revisiones académicas, el feminismo en nuestro país adquiere su vertiente actual en los últimos años. De hecho, aunque el origen de toda esta deriva parte principalmente de la primera legislatura de Zapatero (muestra de ello también son las leyes de VIOGEN (2004) y paridad (2007)), como señalamos en las figuras 8 y 9, no es menos cierto que el auge más consistente, tanto para la prensa, como para los análisis de YouTube, comienza a ser patente durante el periodo 2014-2015, algo que puede verse refrendado, por ejemplo, en la Figura 18 acerca de los titulares que contienen la palabra feminismo. A su vez, esta tendencia también puede ser contrastable en las manifestaciones a raíz del caso de La Manada o en las primeras huelgas feministas. Una deriva relacionada con casos globales como el #Metoo (2017) dentro de las plataformas sociales, o el denominado como Time's Up (2018)⁵¹. Estos procesos, hacedores de lo considerado como cuarta ola, han propiciado también nuevas revisiones dentro del feminismo español. En especial consideración se tienen en cuenta las disensiones acaecidas durante el último año debidas al "Anteproyecto de Ley para la igualdad real y efectiva de las personas trans y para la garantía de los derechos de las personas LGTBI". Una controversia que, a tenor de las proclamas enunciadas tanto en el seno político, como en el de los movimientos populares, está dando lugar a un tipo de polarización en línea con lo refrendado por la Figura 3; Comunidad Fracturada. Pese a ello, estos debates son demasiados amplios y, a día de hoy, están insertos en un tipo de polarización ideológica que debe de ser estudiada más a fondo. En especial, si nos atenemos a las reflexiones y posiciones que tanto partidos políticos, como medios de comunicación tienen al respecto de la cuestión. Pese a ello, cambios como los aportados por este anteproyecto de ley dibujan un nuevo panorama que deberá de ser abordado en futuros trabajos pero que, ya en la actualidad, nos advierte de que emerge un enfoque divergente al que se venía construyendo desde hace dos décadas. O, en paráfrasis con las

⁵¹ Tanto el hastag #Metoo como el Time's Up fueron una serie de movimientos producidos en los EE.UU, y extendidos por el globo, como forma de crítica al abuso y acoso sexual sufrido por las mujeres. Tomó una mayor repercusión cuando un gran número de artistas de Hollywood, en su mayoría mujeres, alzaron la voz en señal de protesta por las prácticas abusivas que algunos altos ejecutivos cinematográficos ejercían

teorizaciones feministas, estaríamos en presencia de las disentías que, acerca del género, tienen la segunda y la tercera ola.

En otro término, ahora nos centraremos en la H.2: El tratamiento mediático del feminismo en España está enmarcado bajo los conceptos de las tesis de género, lo que a su vez no supone una clara polarización entre medios de comunicación. A este respecto, consideramos que nuestra hipótesis puede ser probada, al menos en el conjunto de medios que han sido analizados. Esto se evidencia tanto en el aspecto de agregados comparados de *El Mundo* y *El País* -Figura 10- acerca del uso del tipo de la nomenclatura de violencias de pareja, así como en la disección que se hace de los titulares de “mujer” y “hombre” en el conjunto de medios seleccionados.

Continuando con las afirmaciones acerca de que el feminismo hegemónico se configura en las últimas décadas, así como el papel imperante de las tesis de género en la prensa española, nos gustaría aportar otra consustancial conclusión. Siguiendo la exposición del framing compuesta por D'Angelo (2002), donde incluye tres perspectivas (a saber, cognitiva, crítica y constructorista), consideramos que la adscripción a la teoría de género de los medios se insertaría dentro de esta última. Lo que hacemos es basar nuestra idea de acuerdo con el artículo 13 de la ley de VIOGEN (2004) en relación con los medios de masas:

(i) “Las Administraciones Públicas velarán por el cumplimiento estricto de la legislación en lo relativo a la protección y salvaguarda de los derechos fundamentales, con especial atención a la erradicación de conductas favorecedoras de situaciones de desigualdad de las mujeres en todos los medios de comunicación social, de acuerdo con la legislación vigente.”

(ii) “La Administración pública promoverá acuerdos de autorregulación que, contando con mecanismos de control preventivo y de resolución extrajudicial de controversias eficaces, contribuyan al cumplimiento de la legislación publicitaria”.

Y sin dejar a un lado las condiciones que también se promueven en el artículo 14 de la misma ley, donde se recoge el tratamiento que ha de darse a las violencias ejercidas por los hombres a sus parejas mujeres. Estos planteamientos darían forma a los planteamientos de D'Angelo (2002), poniendo de manifiesto la no existencia de brechas o grandes diferenciaciones entre los medios. En especial, esto puede ilustrarse si tomamos los agregados de los tipos de nomenclaturas de violencias de parejas utilizadas por *El País* y *El Mundo*, Figura 10. A este respecto, destacar también que a pesar de las declaraciones extraídas de los discursos del debate del estado de la nación (ej. Mariano Rajoy en el debate del estado de la nación de 2007: “la ley de violencia doméstica que ustedes llaman de violencia de género”) y de la postura poco relevante que el PP ha ejercido en torno a la temática-a diferencia del PSOE-, ya que en ningún momento ha modificado ninguna de las leyes citadas, además de adherirse al pacto contra la violencia de género de 2017. Por lo que, a fin de cuentas, tanto partidos políticos como medios

de comunicación afines, no estarían retratando la realidad sobre la temática de un modo muy distinto.

Estas conclusiones extraídas de la reflexión de los datos obtenidos sustentan otra interesante idea; la de que España, en las cuestiones feministas, no reporta una polarización evidente. Los medios analizados no plasman evidencias sólidas de distanciamiento. De hecho, en el estudio elaborado sobre las diferencias de tratamiento del feminismo en *El País* y *El Mundo* –Tabla 14–, los dos periódicos han experimentado con el paso del tiempo unas posturas de acercamiento. Aunque sí es cierto que existen tintes que arrojan matices a la cuestión, como es el uso de las palabras “igualdad” y “libertad”. La famosa dicotomía simboliza muy bien las posturas de, por un lado, y con referencia al primer término, las alas socialdemócratas; mientras por otro, la segunda, recaba las líneas maestras de liberales y neoconservadores. Pero a pesar de esas evidencias, en cierta medida entendibles, no se destaca una gran diferencia entre los medios adscritos a una u otra. Algo que se puede cerciorar también en los partidos afines a las líneas editoriales de los periódicos citados. Si bien el sistema mediático español se inserta en lo conocido como pluralismo polarizado, los propios Hallin y Mancini (2004) hicieron algunas aclaraciones con respecto a la ideología de los principales partidos. Destacaban, en Hallin y Mancini (2008, p. 120), que el estilo del modelo de pluralismo polarizado español se basaba más en la apariencia que en un reflejo real de verdaderas distancias ideológicas. Una exposición de motivos que pone de relieve, a pesar de los múltiples trabajos referidos a la constatación de este pluralismo polarizado en España (véase Serrano-Contreras et al. (2021) como muestra propia de ello), que los medios de comunicación, en la temática sobre el feminismo y en especial sobre la cuestión de género, tienden a difuminar sus posturas, situándose más próximos que alejados. Por lo que, a priori, si tomáramos las técnicas de salience media principalmente, o de framing, y de acuerdo con nuestras muestras, no podríamos dictaminar que el feminismo en España genera polos mediáticos, al menos no unos lo suficientemente sólidos. Es más, podría considerarse al feminismo como un punto de encuentro de la mayor parte de medios españoles –al menos de los testeados–, tanto a la vista de los resultados agregados, como en el uso de nomenclaturas similares.

En otro ámbito de cuestiones, nos gustaría exponer, de forma particular, una de las cuestiones planteadas al inicio de esta investigación; el hecho de que el distanciamiento semántico es un predictor de la polarización. Esta idea puede encontrarse tanto en trabajos previos (véase Serrano-Contreras et al., 2020), como en las aportaciones de Miller y Riechert (2001), Morstatter et al. (2018) o Goet (2019). Asumido este anterior planteamiento, y si nos ceñimos al uso que de nuestro índice hemos hecho en los comentarios de YouTube, no podríamos indicar que con él hayamos sido capaces de encontrar una polarización evidente. Ni siquiera una que pueda ser sólida. Consideramos que esto es debido al tipo de puntaje utilizado, al no enmarcado y a la no selección de la dirección que queríamos encontrar en los corpus, ya que utilizamos calificaciones de palabras estandarizadas. Sin olvidar que en este tipo de muestras podemos hallar contextos

lingüísticos, simplicidades y figuras retóricas (ej., ironía, metáfora), con los que no es fácil apreciar esa postura divergente entre usuarios. Aun así, sí que nos gustaría aventurar que, como hacen algunos de los trabajos referidos a lo largo del texto, si decidiéramos usar modelos preentrenados con los que alimentar un algoritmo de clasificación, obtendríamos unas bases más sólidas y ponderadas por nosotros mismos, prestas para ser insertadas en nuestro índice. Pese a todo ello, es evidente que nuestro índice sí es capaz de captar determinados distanciamientos, como se mostró en Serrano-Contreras et al. (2020).

Finalmente, y con visos futuros a continuar por la senda pergeñada en esta tesis, consideramos que el abanico aquí abierto es muy amplio. Desde continuar el estudio de la polarización, los medios de comunicación y el feminismo, hasta seguir ampliando las increíbles dotes que el NLP pueda aportar a las ciencias sociales. Creemos, sin necesidad de aventurar nada nuevo, que el presente está aquí y que el futuro ahora se dirige hacia perspectivas complejas más allá del texto; como son el análisis de imágenes o vídeos con los que ofrecer miradas más amplias de la realidad. Por lo que la misión que ahora nos aguarda es la de continuar atentos a estos planteamientos, tratando de seguir ampliando nuestro conocimiento poco a poco y reportando tanto para la ciencia, como para la sociedad, lo mejor que estos puedan ofrecernos.

Conclusions

Along these pages we have compiled an exposition on certain factors that concern the society of the 21st century. All of them, as reflected in the title of this project, have been useful to us in trying to understand what is happening in our human relationships. In turn, and as we have explained, the wide range of areas has served, somehow or other, to present some questions, objectives, and hypotheses that, throughout this research, and in a conclusive way in this section, have been achieved, answered, verified and refuted.

In the first place, we will begin the exposition of the treatment of the hypotheses as to check our inquiries. Therefore, we would like to start from the approach set out in our H.1., where we pointed to the dominant waves that have shaped feminism in Spain during the last two decades. In the light of the data and academic reviews, it is more than evident that feminism in our country has acquired its current expression in recent years. In fact, although the beginning of all this drift starts mainly during Zapatero's first legislature (proof of this are also the VIOGEN (2004) and parity (2007) laws), as we point out in Figures 8 and 9, it is no less true that the most consistent boom, both for the press and for YouTube analysis, begins to be evident during the 2014-2015 period, something that can be seen endorsed, for example, in Figure 18 about the headlines containing the word 'feminism'. In turn, this trend can also be contrasted in the demonstrations following the case of *La Manada* or in the first feminist strikes. A drift related to global cases such as #MeToo (2017) within social platforms, or the so-called Time's

up (2018)⁵². These processes, which have led to what is considered the fourth wave, have also brought about new revisions within Spanish feminism. Special consideration is given to the dissensions that have occurred during the last year due to the "Draft Bill for the real and effective equality of trans people and for the guarantee of the rights of LGTBI people". A controversy which, according to the claims made by both political and popular movements, is giving rise to a type of polarization in line with Figure 3; *Community Fracturing*. Nevertheless, these debates are too broad and, as of today, are embedded in a type of ideological polarization that needs to be studied further. In particular, if we look at the reflections and positions that both political parties and the media have on the issue. In spite of this, changes such as those brought about by this draft bill outline a new scenario that will have to be addressed in future works but which, already today, warns us that a divergent approach is emerging from the one that has been built up for the last two decades. Or, to paraphrase feminist approaches, we would be in the presence of the second and third waves of gender dissent.

In another term, we will now focus on H.2: The media treatment of feminism in Spain is framed under the concepts of gender thesis, which in turn does not imply a clear polarization between media. In this regard, we consider that our hypothesis can be tested, at least in the set of media that have been analyzed. This becomes evident as much in the aspect of comparative aggregates of *El Mundo* and *El País* - Figure 10- about the use of the type of the nomenclature of intimate partner violence, as well as in the dissection made of the headlines of "women" and "men" in the set of selected media.

Following on from the assertions that hegemonic feminism has been configured in recent decades, as well as the prevailing role of gender thesis in the Spanish press, we would like to provide another consubstantial conclusion. According to the exposition of framing composed by D'Angelo (2002), where he includes three perspectives (namely, cognitive, critical and constructionist), we consider that the ascription to the gender theory of the media would be inserted within the latter. Our idea emerges in accordance with article 13 of the VIOGEN law (2004) in relation to mass media:

(i) "The Public Administrations shall ensure strict compliance with legislation regarding the protection and safeguarding of fundamental rights, with special attention to the eradication of conduct that favors situations of inequality of women in all social communication media, in accordance with the legislation in force".

(ii) "The public administration shall promote self-regulation agreements which, with effective mechanisms for preventive control and out-of-court settlement of disputes, contribute to compliance with advertising legislation".

⁵²Both #MeToo and Time's Up were a series of movements produced in the U.S., and spread around the globe, as a form of criticism of the sexual abuse and harassment suffered by women. It took on greater impact when a large number of Hollywood artists and international celebrities, mostly women, raised their voices in protest against the abusive practices exercised by some top show business executives.

In addition, it is also necessary to take into consideration the conditions that are also promoted in article 14 of the same law, where the treatment to be given to the violence exercised by men against their female partners is set out. These approaches would give shape to D'Angelo's (2002) proposals, highlighting the non-existence of gaps or major differentiations between the media. In particular, this can be illustrated if we take the aggregates of the types of nomenclature of partner violence used by *El País* and *El Mundo*, Figure 10. In this regard, also note that despite the statements taken from the speeches of the state of the nation debate (e.g. Mariano Rajoy in the 2007 state of the nation debate: "the law of domestic violence that you call gender violence") and the little relevant position that the PP has exercised around the thematic, unlike the PSOE, at no time has it modified any of the laws mentioned, in addition to adhering to the pact against gender violence of 2017. Therefore, after all, both political parties and related media would not be portraying the reality of the issue in a very different way.

These conclusions drawn from the reflection of the data obtained support another interesting idea; that Spain, in feminist issues, does not report an evident polarization. The analyzed media do not show solid evidence of distancing. In fact, in the study on the differences in the treatment of feminism in *El País* and *El Mundo* -Table 14-, the two newspapers have experienced over time a closer approach. Although it is true that there are some nuances to the question, such as the use of the words "equality" and "freedom". The famous dichotomy symbolizes very well the positions of, on the one hand, and with reference to the first term, the social-democratic wings; while on the other hand, the second one, recalls the guidelines of liberals and neoconservatives. But in spite of these evidences, to a certain extent understandable, there is no great difference between the media ascribed to one or the other. This can also be seen in the parties aligned with the editorial lines of the above-mentioned newspapers. Despite the fact that the Spanish media system is part of what is known as polarized pluralism, Hallin and Mancini (2004) themselves made some clarifications regarding the ideology of the main parties. They pointed out, in Hallin and Mancini (2008, p.120), that the style of the Spanish polarized pluralism model was based more on appearance than on a real reflection of the true ideological distances. Despite the many studies referring to the confirmation of this polarized pluralism in Spain (see Serrano-Contreras et al. (2021) as an example of this), this statement of reasons highlights the fact that the media, in the area of feminism and especially gender issues, tend to blur their positions, being closer rather than farther apart. Therefore, a priori, if we were to take the techniques of salience media mainly, or framing, and according to our samples, we could not rule that feminism in Spain generates media poles, at least not sufficiently solid ones. Moreover, feminism could be considered as a meeting point for most of the Spanish media -at least of those tested-, both in view of the aggregate results and the use of similar nomenclatures.

In another area of issues, we would like to expose, in particular, one of the concerns raised at the beginning of this research; the fact that semantic distancing is a predictor of polarization. This idea can be found both in previous work (see

Serrano-Contreras et al. 2020), and in the contributions of Miller and Riechert (2001), Morstatter et al. (2018) or Goet (2019). Assuming this approach, and if we stick to the use we have made of our index in the YouTube comments, we could not say that we have been able to find an obvious polarization with it. Not even one that could be robust. We consider that this is due to the type of scoring used, the non-framing and the non-selection of the direction we wanted to find in the corpora, since we used standardized word ratings. Besides, in this type of samples we can find linguistic contexts, simplicities and rhetorical figures (e.g., irony, metaphor), with which it is not easy to appreciate that divergent stance among users. Even so, we would like to suggest that, as some of the works reviewed throughout the text do, if we decided to use pre-trained models to feed a classification algorithm, we would obtain more solid and weighted bases of our own, ready to be inserted in our index. Despite all this, it is clear that our index is capable of capturing certain distances, as was shown in Serrano-Contreras et al. (2020).

Finally, and with future visions to continue along the path outlined in this thesis, we consider that the range open here is very wide. From continuing the study of polarization, the media and feminism, to continuing expanding the incredible possibilities that NLP can bring to the social sciences. We believe, without the need to venture anything new, that the present is here and that the future is now directed towards complex perspectives beyond the text, such as the analysis of images or videos with which to offer broader views of reality. So, the mission that now awaits us is to remain attentive to these approaches, trying to continue to expand our knowledge little by little and reporting both for science and for society, the best that they can offer us.

Bibliografía

- Abramowitz, A. I., y Saunders, K. L. (1998). Ideological Realignment in the U.S. Electorate. *The Journal of Politics*, 60(3), 634-652. doi:10.2307/2647642
- Abramowitz, A. I., y Saunders, K. L. (2008). Is Polarization a Myth? *The Journal of Politics*, 70(2), 542-555. <https://doi.org/10.1017/S0022381608080493>
- Aguilar, G. T. (2008). El sistema sexo-género en los movimientos feministas. *Amnis*, 8. <https://doi.org/10.4000/amnis.537>
- Aguirre Romero, M. J. (2009). Medios de comunicación y Democracia: ¿una relación necesaria o una relación imposible? En Comunicación presentada durante el Congreso Internacional XVI Semana de Ética y Filosofía Política. Universidad de Murcia.
- Alexander, R. (1982). Biology and the moral paradoxes. *Journal of Social and Biological Systems*, 5(4), 389-395. [https://doi.org/10.1016/S0140-1750\(82\)92066-8](https://doi.org/10.1016/S0140-1750(82)92066-8)
- Alford, J. R., Funk, C. L., y Hibbing, J. R. (2005). Are Political Orientations Genetically Transmitted? *American Political Science Review*, 99(2), 153-167. <https://doi.org/10.1017/S0003055405051579>
- Allen, C. y Hospedales, T. (2019). Analogies Explained: Towards Understanding Word Embeddings. Proceedings of the 36th International Conference on Machine Learning, in Proceedings of Machine Learning Research, 97, 223-231.
- Almagro, M., y Villanueva, N. (2021). Polarización y tecnologías de la información: radicales vs. extremistas. *Dilemata*, (34), 51-69.
- Altheide, D. L. (2013). Media Logic, Social Control, and Fear: Media Logic, Social Control, and Fear. *Communication Theory*, 23(3), 223-238. <https://doi.org/10.1111/comt.12017>
- Altheide, D. L. (2016). Media Logic. En G. Mazzoleni (Ed.), *The International Encyclopedia of Political Communication*. Wiley Online Library. <https://doi.org/10.1002/9781118541555.wbiepc088>
- Altheide, D. L., y Snow, R. P. (1979). *Media Logic*. Sage.
- Altszyler, E., Ribeiro, S., Sigman, M., y Fernández Slezak, D. (2017). The interpretation of dream meaning: Resolving ambiguity using Latent Semantic Analysis in a small corpus of text. *Consciousness and Cognition*, 56, 178-187. <https://doi.org/10.1016/j.concog.2017.09.004>
- An, J., Kwak, H., y Ahn, Y-Y. (2018). SemAxis: A Lightweight Framework to Characterize Domain-Specific Word Semantics Beyond Sentiment. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 56(1), 2450-2461. <https://doi.org/10.18653/v1/P18-1228>
- Angius, N., Primiero, G., y Turner, R. (2021). The Philosophy of Computer Science. En E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. Recuperado el 13 de febrero de 2021 de <https://plato.stanford.edu/archives/spr2021/entries/computer-science/>
- Annandale, E. y Clark, J. (1996). What is gender? Feminist theory and the sociology of human reproduction. *Sociology of Health & Illness*, 18, 17-44. <https://doi.org/10.1111/1467-9566.ep10934409>

- Arce García, S., y Menéndez Menéndez, M. I. (2018). Aplicaciones de la estadística al framing y la minería de texto en estudios de comunicación. *Información, Cultura y Sociedad*, (39), 61-70. <https://doi.org/10.34096/ics.i39.4260>
- Aristóteles. (2018). *Política* (Trad. A. Gómez Robledo). Universidad autónoma de México. (Trabajo original publicado en S. IV a.C.).
- Armborst, A. (2017). *Thematic Proximity in Content Analysis*. SAGE Open. <https://doi.org/10.1177/2158244017707797>
- Arriagada Ilabaca, A., Navia Lucero, P., y Schuster Pineda, M. (2010). ¿Consumo luego pienso, o pienso y luego consumo?: Consumo de medios, predisposición política, percepción económica y aprobación presidencial en Chile. *Revista de Ciencia Política*, 30(3). <https://doi.org/10.4067/S0718-090X2010000300005>
- Arun, R., Suresh, V., Veni Madhavan, C. E., y Narasimha Murthy, M. N. (2010). On finding the natural number of topics with latent dirichlet allocation: some observations. En M. J. Zaki, J. X. Yu, B. Ravindran, y V. Pudi (Eds.), *Advances in knowledge discovery and data mining* (pp. 391-402). Berlin: Springer. https://doi.org/10.1007/978-3-642-13657-3_43
- Atanasova, D., y Koteyko, N. (2016). Obesity frames and counter-frames in British and German online newspapers. *Health*. DOI: 10.1177/1263459316649764.
- Aupers, S. (2020). Decoding Mass Media/Encoding Conspiracy Theory. En M. Butter y P. Knight (Eds.), *Routledge Handbook of Conspiracy Theories* (1ª ed.) (pp. 469-482). Routledge. https://doi.org/10.4324/9780429452734-4_7
- Baker, S., Baksy, A., Bloom, N., Davis, S., y Rodden, J. (2020). Elections, Political Polarization, and Economic Uncertainty (No. w27961; p. w27961). National Bureau of Economic Research. <https://doi.org/10.3386/w27961>
- Baron, D. P. (2006). Persistent media bias. *Journal of Public Economics*, 90(1-2), 1-36.
- Baron-Cohen, S. (2009). *La gran diferencia: Cómo son realmente los cerebros de hombres y mujeres*. Alfaomega.
- Bartra, E. (2020). El feminismo y sus olas. *Zona Franca. Revista de Estudios de Género*, (28), 516-549. <https://doi.org/10.35305/zf.vi28.179>
- Baum, M. A., y Groeling, T. (2008). New Media and the Polarization of American Political Discourse. *Political Communication*, 25(4), 345-365. <https://doi.org/10.1080/10584600802426965>
- Baumer, E., Elovic, E., Qin, Y., Polletta, F., y Gay, G. (2015). Testing and comparing computational approaches for identifying the language of framing in political news. En *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1472-1482).
- Baumgartner, F. R., y Chaqués Bonafont, L. (2015). All News is Bad News: Newspaper Coverage of Political Parties in Spain. *Political Communication*, 32(2), 268-291. <https://doi.org/10.1080/10584609.2014.919974>
- Beauvoir, S. (2019). *El segundo sexo* (Trad. A. Martorell Linares). Cátedra.
- Bennett, W. L., y Entman, R. M. (2001). *Mediated Politics. Communication in the Future of Democracy*. Cambridge University Press.
- Bennett, W. L., y Segerberg, A. (2014). De los medios de comunicación de masas a las redes sociales. *Telos*. 98. Recuperado el 13 de Octubre de 2020 de <https://bit.ly/34iAqh7>

- Bennett, W.L. e Iyengar, S. (2008), A New Era of Minimal Effects? The Changing Foundations of Political Communication. *Journal of Communication*, 58: 707-731. <https://doi.org/10.1111/j.1460-2466.2008.00410.x>
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A., Lowe, W., y Müller, C. (2020). 'quanteda'. Recuperado el 23 de marzo de 2021 de <https://cran.r-project.org/web/packages/quanteda/quanteda.pdf>
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., y Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30), 774. <https://doi.org/10.21105/joss.00774>
- Berelson, B. (1952). *Content Analysis in Communication Research*. Free Press.
- Berelson, B., Lazarsfeld, P. F., y McPhee, W. N. (1954). *Voting*. University of Chicago Press.
- Berners-Lee, T. (2000). *Tejiendo la red*. Siglo XXI de España.
- Bi, Q., Goodman, K. E., Kaminsky, J., y Lessler, J. (2019). What is Machine Learning? A Primer for the Epidemiologist. *American Journal of Epidemiology*, kwz189. <https://doi.org/10.1093/aje/kwz189>
- Bird, S., Klein, E., y Loper, E. (2009). *Natural language processing with Python* (1ª ed.). O'Reilly.
- Biswas, A. (2004). La tercera ola feminista: cuando la diversidad, las particularidades y la diferencia son lo que cuenta. *Casa del Tiempo*, 6(68), 65-70
- Blei, D. M., Ng, A. Y., y Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993-1022. <https://doi.org/10.5555/944919.944937>
- Boxell, L., Gentzkow, M., y Shapiro, J. (2020). Cross-Country Trends in Affective Polarization (No. w26669; p. w26669). *National Bureau of Economic Research*. <https://doi.org/10.3386/w26669>
- Boyd, D., y Crawford, K. (2011). Six Provocations for Big Data. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1926431>
- Bramson, A., Grim, P., Singer, D. J., Berger, W. J., Sack, G., Fisher, S., Flocken, C., y Holman, B. (2017). Understanding Polarization: Meanings, Measures, and Model Evaluation. *Philosophy of Science*, 84(1), 115-159. <https://doi.org/10.1086/688938>
- Brants, K., y Voltmer, K. (2011). Introduction: Mediatization and De-centralization of Political Communication. En K. Brants y K. Voltmer (Eds.), *Political Communication in Postmodern Democracy* (pp. 1-16). Palgrave Macmillan UK. https://doi.org/10.1057/9780230294783_1
- Bro, P., y Wallberg, F. (2014). Digital Gatekeeping: News media versus social media. *Digital Journalism*, 2(3), 446-454. <https://doi.org/10.1080/21670811.2014.895507>
- Bunescu, R. C., y Mooney, J. (2007). Extracting Relations from Text: From Word Sequences to Dependency Paths. En A. Kao., y R. S. Poteet (Eds.), *Natural language processing and text mining* (pp. 29-44). Springer.
- Butler, J. (1986). Sex and Gender in Simone de Beauvoir's *Second Sex*. *Yale French Studies*, (72), 35-49. doi:10.2307/2930225
- Butler, J. (1990). *Gender trouble : feminism and the subversion of identity*. Routledge.
- Butler, J. (1993). *Bodies that matter : on the discursive limits of "sex"*. Routledge.

- Campante, F. R., y Hojman, D. A. (2013). Media and polarization. *Journal of Public Economics*, 100, 79–92. <https://doi.org/10.1016/j.jpubeco.2013.02.006>
- Campbell, R., Martin, C. R., y Fabos, B. (2017). *Media & culture: Mass communication in a digital age*. Macmillan Learning.
- Canel, M. J. (1999a). *Comunicación política: técnicas y estrategias para la Sociedad de la información*. Tecnos.
- Canel, M. J. (1999b). El País, ABC y El Mundo: Tres manchetas, tres enfoques. *Zer. Revista de Estudios de Comunicación*, 4(6), 97–117.
- Canel, M. J., y Voltmer, K. (Eds.). (2014). *Comparing Political Communication across Time and Space*. Palgrave Macmillan UK. <https://doi.org/10.1057/9781137366474>
- Cao, J., Xia, T., Li, J., Zhang, Y., y Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72(7), 1775–1781. <https://doi.org/10.1016/j.neucom.2008.06.011>
- Carothers, T., y O'Donohue, A. (Eds.). (2020). *Democracies divided: The global challenge of political polarization*. Brookings Institution Press.
- Carr, C. T. y Hayes, R. A. (2015) Social Media: Defining, Developing, and Divining, *Atlantic Journal of Communication*, 23(1), 46–65. <https://doi.org/10.1080/15456870.2015.972282>
- Cazorla Pérez, J. (2008). *Manual de Introducción a la Ciencia Política*. Fundación Nucleo.
- Ceci, M., y Lanotte, P.F. (2021). Closed sequential pattern mining for sitemap generation. *World Wide Web*, 24, 175–203. <https://doi.org/10.1007/s11280-020-00839-2>
- Chamberlain, P. (2017). *The feminist fourth wave: Affective temporality*. Springer.
- Charles, N. y Wadia, K. (2018). New British feminisms, UK Feminista and young women's activism. *Feminist Theory*, 19(2), 165–181. doi:10.1177/1464700117723592.
- Cheng, X., Yan, X., Lan, Y., y Guo, J. (2014). BTM: Topic Modeling over Short Texts. *IEEE Transactions on Knowledge and Data Engineering*, 26(12), 2928–2941. <https://doi.org/10.1109/TKDE.2014.2313872>
- Cho, J., Ahmed, S., Hilbert, M., Liu, B., y Luu, J. (2020). Do Search Algorithms Endanger Democracy? An Experimental Investigation of Algorithm Effects on Political Polarization. *Journal of Broadcasting & Electronic Media*, 64(2), 150–172. <https://doi.org/10.1080/08838151.2020.1757365>
- Chowdhury, G. G. (2003). Natural language processing. *Ann. Rev. Info. Sci. Tech.*, 37, 51–89. <https://doi.org/10.1002/aris.1440370103>
- Chuang, J., Manning, C. D., y Heer, J. (2012). Termite: Visualization techniques for assessing textual topic models. *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '12*, 74. <https://doi.org/10.1145/2254556.2254572>
- Chyi, H. I., y McCombs, M. (2004). Media Salience and the Process of Framing: Coverage of the Columbine School Shootings. *Journalism & Mass Communication Quarterly*, 81(1), 22–35. <https://doi.org/10.1177/107769900408100103>
- Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrocioni, W., y Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118. <https://doi.org/10.1073/pnas.2023301118>

- Clark, C. J., Liu, B. S., Winegard, B. M., y Ditto, P. H. (2019). Tribalism Is Human Nature. *Current Directions in Psychological Science*, 28(6), 587–592. <https://doi.org/10.1177/0963721419862289>
- Cobo, R. (2012). Despatriarcalización y agenda feminista. En K. Uriona (Ed.), *Mujeres en diálogo: Avanzando hacia la despatriarcalización en Bolivia* (pp. 107-118). Editorial Soraya Luján.
- Cobo, R., y Ranea Triviño, B. (2020). *Breve diccionario de feminismo*. Catarata.
- Codina, L. (2020). *Fundamentos de la Búsqueda para estudiantes y profesionales de la comunicación audiovisual* [Diapositiva]. Repositorio UPF. <https://bit.ly/3sypTbo>
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., y Kuksa, P. (2011). Natural Language Processing (almost) from Scratch. ArXiv:1103.0398 [Cs]. <http://arxiv.org/abs/1103.0398>
- Conceição, S. C. O., Samuel, A., y Yelich Biniński, S. M. (2017). Using concept mapping as a tool for conducting research: An analysis of three approaches. *Cogent Social Sciences*, 3(1). <https://doi.org/10.1080/23311886.2017.1404753>
- Converse, P. E. (2006). The nature of belief systems in mass publics (1964). *Critical Review*, 18(1-3), 1-74. <https://doi.org/10.1080/08913810608443650>
- Conway, J. K., Bourque, S. C., y Scott, J. W. (2013). El concepto de género. En M. Lamas (Ed.), *El género: La construcción cultural de la diferencia sexual* (pp. 21-33). UNAM.
- Coppock, A., Hill, S. J., y Vavreck, L. (2020). The small effects of political advertising are small regardless of context, message, sender, or receiver: Evidence from 59 real-time randomized experiments. *Science Advances*, 6(36), eabc4046. <https://doi.org/10.1126/sciadv.abc4046>
- Covington, P., Adams, J., y Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations. *Proceedings of the 10th ACM Conference on Recommender Systems*, 191-198. <https://doi.org/10.1145/2959100.2959190>
- Cruz, F. L., Troyano, J. A., Pontes, B., y Ortega, F. J. (2014). Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Systems with Applications*, 41(13), 5984-5994. <https://doi.org/10.1016/j.eswa.2014.04.005>
- Curran, J. (2002). *Medios de comunicación y poder*. Hacer.
- Darmofal, D., y Strickler, R. (2019). *Demography, Politics, and Partisan Polarization in the United States, 1828-2016* (1st ed. 2019). Springer. <https://doi.org/10.1007/978-3-030-04001-7>
- Davidson, J., Livingston, B., Sampath, D., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., Gupta, S., He, Y., y Lambert, M. (2010). The YouTube video recommendation system. *Proceedings of the Fourth ACM Conference on Recommender Systems - RecSys '10*, 293. <https://doi.org/10.1145/1864708.1864770>
- De Francisci Morales, G., Monti, C., y Starnini, M. (2021). No echo in the chambers of political interactions on Reddit. *Scientific Reports*, 11(1), 2818. <https://doi.org/10.1038/s41598-021-81531-x>
- De Lora, P. (2019). *Lo sexual es político (y jurídico)*. Alianza Editorial
- De Mateo, R. (2004). Spain. En M. Kelly, G. Mazzoleni, y D. McQuail (Eds.), *The media in Europe: the Euromedia handbook*. Sage.

- De Vreese, C., y Semetko, H. (2004). *Political campaigning in referendums framing the referendum issue*. Routledge.
- Dechile. (2021). POLARIZAR, radicación. *Etimologías de Chile - Diccionario que explica el origen de las palabras*. Recuperado el 29 de Mayo de 2021 de <http://etimologias.dechile.net/?polarizar>
- Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. y Harshman, R. (1990), Indexing by latent semantic analysis. *J. Am. Soc. Inf. Sci.*, 41, 391-407. [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-ASI1>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9)
- Deveaud, R., SanJuan, E., y Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numerique*, 17(1), 61-84.
- Devine, C. J. (2015). Ideological Social Identity: Psychological Attachment to Ideological In-Groups as a Political Phenomenon and a Behavioral Influence. *Political Behavior*, 37(3), 509-535. <https://doi.org/10.1007/s11109-014-9280-6>
- DiMaggio, P., Evans, J., y Bryson, B. (1996). Have American's Social Attitudes Become More Polarized? *American Journal of Sociology*, 102(3), 690-755.
- Druckman, J. (2003). The Power of Television Images: The First Kennedy-Nixon Debate Revisited. *The Journal of Politics*, 65(2), 559-571. doi:10.1111/1468-2508.t01-1-00015
- Druckman, J. N. y Levendusky, M. S. (2019). What Do We Measure When We Measure Affective Polarization? *Public Opinion Quarterly*, 83(1), 114-22.
- D'Angelo, P. (2002). News framing as a multiparadigmatic research program: A response to Entman. *Journal of communication*, 52(4), 870-888.
- Edelman, M. (2001). Social movements: Changing paradigms and forms of politics. *Annual Review of Anthropology*, 30, 285-317.
- Eisenstein, J. (2018). *Natural Language Processing*. MIT press.
- Entman, R. M. (1991). Framing U.S. coverage of international news: Contrasts in narratives of the KAL and Iran air incidents. *Journal of Communication*, 41(4), 6-27
- Entman, R. M. (1993). Framing: towards clarification of a fractured paradigm. *Journal of Communication*, 43(4), 8-51.
- Esser, F., y Strömbäck, J. (2014). *Mediatization of politics: Understanding the transformation of western democracies*. <http://site.ebrary.com/id/10952864>
- Esteban, J., y Ray, D. (1994). On the Measurement of Polarization. *Econometrica*, 62(4), 819-851. doi:10.2307/2951734
- Estrada Esparza, O., N. (2017). Las sufragistas: narrativas de dolor, poder, libertad e igualdad. En D. Arauz Mercado (Ed.), *Pensamiento y sensibilidad en el discurso de género* (pp. 1-15). Texere editores.
- Esuli, A., y Sebastiani, F. (2009). Training Data Cleaning for Text Classification. En L., Azzopardi, G., Kazai, S., Robertson, S., Rüger, M., Shokouhi, D., Song, y E., Yilmaz (Eds.), *Advances in Information Retrieval Theory. ICTIR 2009. Lecture Notes in Computer Science*, vol 5766. Springer. https://doi.org/10.1007/978-3-642-04417-5_4
- Ethayarajh, K., Duvenaud, D., y Hirst, G. (2019). *Understanding Undesirable Word Embedding Associations*. ArXiv:1908.06361 [Cs]. <http://arxiv.org/abs/1908.06361>

- Feinerer, Ingo, Kurt Hornik y David Meyer. 2008. "Text Mining Infrastructure in R." *Journal of Statistical Software* 25(5): 1-54.
- Feinerer, I., y Hornik, K. (2020a). 'tm': Text Mining. Recuperado el 12 de enero de 2021 de <https://cran.r-project.org/web/packages/tm/tm.pdf>
- Feinerer, I., y Hornik, K. (2020b). 'wordnet': WordNet Interface. R package version 0.1-15. Recuperado el 22 de abril de 2021 de <https://CRAN.R-project.org/package=wordnet>
- Fellbaum, C. (Ed.). (1998). *WordNet: An Electronic Lexical Database*. The MIT Press. <https://doi.org/10.7551/mitpress/7287.001.0001>
- Fernández-Llebrez, F. (2004). ¿"Hombres de verdad"? Estereotipo masculino, relaciones entre los géneros y ciudadanía. *Foro Interno*, 4, 15 - 43.
- Ferrando, M. G. (1987). *Socioestadística: Introducción a la estadística en sociología*. Alianza Universidad Textos
- Figueiredo Filho, D. B., Paranhos, R., da Silva Junio, J. A., y Silva, D. (2016). Precisamos Falar sobre Métodos Quantitativos em Ciência Política. *Revista Latinoamericana de Metodología de la Investigación Social*. 11, 21-39.
- Figueiredo Filho, D. B., Rocha, E. C. da, Júnior, J. A. da S., Paranhos, R., Silva, M. B. da, y Duarte, B. S. F. (2014). Cluster Analysis for Political Scientists. *Applied Mathematics*, 5(15), 2408-2415. <https://doi.org/10.4236/am.2014.515232>
- Firestone, S. (1977). *La dialéctica del sexo: en defensa de la revolución feminista*. Kairós.
- Firestone, S. (2003). *The dialectic of sex: The case for feminist revolution*. Farrar, Straus and Giroux.
- Fletcher, R., & Jenkins, J. (2019). *Polarisation and the news media in Europe*. EPRS.
- Frey, D. (1986). Recent Research on Selective Exposure to Information. En *Advances in Experimental Social Psychology*, vol.19, 41-80. Elsevier. [https://doi.org/10.1016/S0065-2601\(08\)60212-9](https://doi.org/10.1016/S0065-2601(08)60212-9)
- Friedan, B. (1977). *The feminine mystique*. Dell.
- García-Marín J. (2021). YouTube and Traditional Media: Polarization in the Catalan Political Conflict. En Musiał-Karg M., Luengo Ó.G. (Eds), *Digitalization of Democratic Processes in Europe. Studies in Digital Politics and Governance*. Springer. https://doi.org/10.1007/978-3-030-71815-2_3
- García-Marín, J. (2007). *El comportamiento de la prensa durante los conflictos de Kosovo e Irak*. Universidad de Granada.
- García-Marín, J. (2011). Encuadres, conflictos y efectos de agenda. *Zer. Revista de Estudios de Comunicación*, 16(31), pp. 167-181.
- García-Marín, J., Calatrava, A., y Luengo, Ó. G. (2018). Debates electorales y conflicto. Un análisis con máquinas de soporte virtual (SVM) de la cobertura mediática de los debates en España desde 2008. *El Profesional de la Información*, 27(3), 624-632. <https://doi.org/10.3145/epi.2018.may.15>
- Gentzkow, M., Kelly, B., y Taddy, M. (2017). *Text as Data* (No. w23276; p. w23276). National Bureau of Economic Research. <https://doi.org/10.3386/w23276>
- Gentzkow, M., y Shapiro, J. M. (2006). Media Bias and Reputation. *Journal of Political Economy*, 114(2), 280-316. <https://doi.org/10.1086/499414>

- Gentzkow, M., y Shapiro, J. M. (2010). What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, 78(1), 35–71. <https://doi.org/10.3982/ECTA7195>
- Ghatak, A. (2019). *Deep Learning with R*. Springer Publishing.
- Goet, N. D. (2019). Measuring Polarization with Text Analysis: Evidence from the UK House of Commons, 1811–2015. *Political Analysis*, 27(4), 518–539. <https://doi.org/10.1017/pan.2019.2>
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harvard University Press.
- Gordon, L. (2002). *The Moral Property of Women: A History of Birth Control Politics in America*. University of Illinois Press.
- Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78(6), 1360–1380.
- Green R. (2010). Robert Stoller's Sex and Gender: 40 years on. *Archives of sexual behavior*, 39(6), 1457–1465. <https://doi.org/10.1007/s10508-010-9665-5>
- Greer, G. (2012). *The female eunuch*. Fourth Estate.
- Griffiths, T. L., y Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(1), 5228–5235. <https://doi.org/10.1073/pnas.0307752101>
- Groseclose, T., y Milyo, J. (2005). A Measure of Media Bias. *The Quarterly Journal of Economics*, 120(4), 1191–1237. <https://doi.org/10.1162/003355305775097542>
- Guess, A., Nyhan, B., Lyons, B., y Reifler, J. (2018). *Avoiding the echo chamber about echo chambers*. Knight Foundation.
- Gurevitch, M., Coleman, S., y Blumler, J. G. (2009). Political Communication—Old and New Media Relationships. *The ANNALS of the American Academy of Political and Social Science*, 625(1), 164–181. <https://doi.org/10.1177/0002716209339345>
- González Alcantud, J. A. (1987). El buen salvaje de Rousseau. Inflexión de la antropología y de la estética. *Gazeta de antropología*. <https://doi.org/10.30827/Digibug.13768>
- Haidt, J. (2019). *La mente de los justos: Por qué la política y la religión dividen a la gente sensata*. Deusto.
- Halavais, A. (2015). Bigger sociological imaginations: Framing big social data theory and methods. *Information, Communication & Society*, 18(5), 583–594. <https://doi.org/10.1080/1369118X.2015.1008543>
- Hallin, D. C., y Mancini, P. (2004). *Comparing media systems: Three models of media and politics*. Cambridge university press.
- Han, J., Kamber, M., y Pei, J. (2012). Getting to Know Your Data. En J. Han, M. Kamber, y J. Pei (Eds.), *Data Mining (Third Edition)* (pp. 39–82). Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-381479-1.00002-2>
- Hansen, A., y Machin, D. (2019). *Media and communication research methods* (2nd. ed.). Red Globe Press.
- Hansen, F. S. (2015). Framing yourself into a corner: Russia, Crimea, and the minimal action space. *European Security*, 24(1), 141–158. <https://doi.org/10.1080/09662839.2014.993974>

- Harel, T. O., Jameson, J. K., y Maoz, I. (2020). The Normalization of Hatred: Identity, Affective Polarization, and Dehumanization on Facebook in the Context of Intractable Political Conflict. *Social Media + Society*. <https://doi.org/10.1177/2056305120913983>
- Hartmann, H. (1976). Capitalism, Patriarchy, and Job Segregation by Sex. *Signs*, 1(3), 137-169.
- Hawkins, S., Yudkin, D., Juan-Torres, M., y Dixon, T. (2018). *Hidden Tribes: A Study of America's Polarized Landscape*. More in Common.
- Hayes, R. A., Carr, C. T., y Wohn, D. Y. (2016). One Click, Many Meanings: Interpreting Paralinguistic Digital Affordances in Social Media. *Journal of Broadcasting & Electronic Media*, 60(1), 171-187. <https://doi.org/10.1080/08838151.2015.1127248>
- Helbling, M., y Tresch, A. (2011). Measuring party positions and issue salience from media coverage: Discussing and cross-validating new indicators. *Electoral Studies*, 30(1), 174-183. <https://doi.org/10.1016/j.electstud.2010.12.001>
- Hernández, E., Anduiza, E., y Rico, G. (2021). Affective polarization and the salience of elections. *Electoral Studies*, 69, 102203. <https://doi.org/10.1016/j.electstud.2020.102203>
- Heywood, L., y Drake, J. (2007). 'It's all about the Benjamins'. Economic determinants of third wave feminism in the United States. En S. Gillis y G. Howie (Eds.), *Third wave feminism. A critical exploration* (pp. 115-124). Palgrave.
- Hill, R. K. (2016). What an Algorithm Is. *Philosophy & Technology*, 29(1), 35-59. <https://doi.org/10.1007/s13347-014-0184-5>
- Hornik, K. (2019) 'openNLP'. Recuperado el 12 de enero de 2021 de <https://cran.r-project.org/web/packages/openNLP/openNLP.pdf>
- Hosken, F. (1981). Toward a Definition of Women's Human Rights. *Human Rights Quarterly*, 3(2), 1-10. doi:10.2307/761853
- Hu, M., y Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining- KDD '04*, 168-177. <https://doi.org/10.1145/1014052.1014073>
- Hu, X., Cai, Z., Wiemer-Hastings, P., Graesser, A., y McNamara, D. (2007). Strengths, Limitations, and Extensions of LSA. En T., Landauer, D. S., McNamara y S. D., Walter Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 401-426). Routledge.
- Huang, A. (2008). Similarity measures for text document clustering. *Proceedings of the 6th New Zealand Computer Science Research Student Conference*.
- Hui, Y. (2020). *Sobre el límite de la inteligencia artificial*. Recuperado el 13 de abril 2021 de <https://lab.cccb.org/es/sobre-el-limite-de-la-inteligencia-artificial/>
- Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., y van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist*, 74(2), 171-193. <https://doi.org/10.1037/amp0000307>
- Igartúa, J. J., y Humanes, M. L. (2004). Imágenes de Latinoamérica en la prensa española. Una aproximación empírica desde la Teoría del Encuadre. *Comunicación y sociedad*, 12(1), 47-75.
- Iyengar, S. (2021). The polarization of American politics. En *The Routledge Handbook of Political Epistemology* (pp. 90-100). Routledge.

- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., y Westwood, S. J. (2019). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science*, 22(1), 129–146. <https://doi.org/10.1146/annurev-polisci-051117-073034>
- Iyengar, S., Sood, G., y Lelkes, Y. (2012). Affect, Not Ideology. *Public Opinion Quarterly*, 76(3), 405–431. <https://doi.org/10.1093/poq/nfs038>
- Iyengar, S., y Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of communication*, 59(1), 19–39.
- Iyengar, S., y Kinder, D. R. (1987). *New That Matters: Television and American Opinion*. University of Chicago Press.
- Iyengar, S., y Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, 59(3), 690–707.
- Jakobson, R. (1960). Closing statement: Linguistics and poetics. *Style in language*, 350(377), 570–579.
- Jarmul, K., y Lawson, R. (2017). *Python Web Scraping—Second Edition*. Packt Publishing.
- Jockers, M. L. (2017). 'syuzhet'. Recuperado el 20 de abril de 2021 de <https://r.meteo.uni.wroc.pl/web/packages/syuzhet/syuzhet.pdf>
- Jockers, M. L. y Thalken, R. (2020). *Text Analysis with R*. Springer
- Jones, M. (2019). What we talk about when we talk about (big) data. *The Journal of Strategic Information Systems*. 28(1), 3–16. <https://doi.org/10.1016/j.jsis.2018.10.005>
- Joyanes, L. (2008). *Fundamentos de programación*. McGraw-Hill.
- Joyanes, L. (2013). *Big Data: Análisis de grandes volúmenes de datos en organizaciones*. Alfaomega.
- Justicia de la Torre, C., Sánchez, D., Blanco, I., y Martín-Bautista, M. J. (2018). Text Mining: Techniques, Applications, and Challenges. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 26(04), 553–582. <https://doi.org/10.1142/s0218488518500265>
- Justicia de la Torre, M. C. (2017). *Nuevas técnicas de minería de textos: Aplicaciones*. Universidad de Granada. Recuperado el 13 de noviembre de 2019 de <https://digibug.ugr.es/handle/10481/46975>
- Kao, A., y Poteet, S. R. (2007). *Natural language processing and text mining*. Springer.
- Kaplan, E. A. (2003). Feminist futures: trauma, the post-9/11 world and a fourth feminism? *Journal of International Women's Studies*, 4(2), 46–59.
- Kappler, J. (1960). *The effects of mass communication*. The Free Press.
- Kearney, M. W. (2019). Analyzing change in network polarization. *New Media & Society*, 21(6), 1380–1402. <https://doi.org/10.1177/1461444818822813>
- Kelly, C. J. (2017). *The personal is political*. En *Encyclopedia Britannica*. Recuperado el 3 de enero de 2021 de <https://www.britannica.com/topic/the-personal-is-political>
- Kietzmann, J., Lee, L. W., McCarthy, I. P., y Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, 63(2), 135–146. <https://doi.org/10.1016/j.bushor.2019.11.006>

- Kiouis, S. (2004). Explicating Media Salience: A Factor Analysis of New York Times Issue Coverage During the 2000 U.S. Presidential Election. *Journal of Communication*, 54(1), 71-87. <https://doi.org/10.1111/j.1460-2466.2004.tb02614.x>
- Knights, D. (2019). Gender still at work: Interrogating identity in discourses and practices of masculinity. *Gender, Work & Organization*, 26(1), 18-30. <https://doi.org/10.1111/gwao.12338>
- Knill, E. (2010). Quantum computing. *Nature*, 463, 441-443. <https://doi.org/10.1038/463441a>
- Knudsen, E. (2020). Affective Polarization in Multiparty Systems? Comparing Affective Polarization Towards Voters and Parties in Norway and the United States. *Scandinavian Political Studies*, 44(1), 34-44. <https://doi.org/10.1111/1467-9477.12186>
- Kohan, M. (2018). *Diccionario feminista para miembros atónitos del patriarcado (hombres y mujeres)*. En Público. Recuperado el 12 de enero de 2021 de <https://www.publico.es/sociedad/feminismo-diccionario-feminista-miembros-atonitos-patriarcado-hombres-mujeres.html>
- Kouzis-Loukas, D. (2016). *Learning Scrapy: Learn the art of efficient web scraping and crawling with Python*. Packt Publishing.
- Kramer, A. D. I., Guillory, J. E., y Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788-8790. <https://doi.org/10.1073/pnas.1320040111>
- Krippendorff, K. (2004). *Content analysis: an introduction to its methodology* (2ª ed.). Sage.
- Kull, S., Ramsay, C., y Lewis, E. (2003). Misperceptions, the media, and the Iraq war. *Political Science Quarterly*, 118(4), 569-598.
- Kwak, H., An, J., y Ahn, Y.-Y. (2020). *FrameAxis: Characterizing Framing Bias and Intensity with Word Embedding*. ArXiv:2002.08608 [Cs]. <http://arxiv.org/abs/2002.08608>
- Kwartler, T. (2017). *Text Mining in Practice with R*. Wiley.
- Lakoff, G. (2008). *No pienses en un elefante: lenguaje y debate político* (1ª ed.). Ed. Complutense.
- Lamas, M. (2013). *El género: La construcción cultural de la diferencia sexual*. UNAM.
- Landauer, T. K., y Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211-240. <https://doi.org/10.1037/0033-295X.104.2.211>
- Latorre, M. (2018). *Historias de las Web, 1.0, 2.0, 3.0, 4.0*. Recuperado el 14 de Abril de 2021 de http://umch.edu.pe/arch/hnomarino/74_Historia%20de%20la%20Web.pdf.
- Laver, M., Benoit, K., y Garry, J. (2003). Extracting Policy Positions from Political Texts Using Words as Data. *The American Political Science Review*, 97(2), 311-331.
- Lazarsfeld, P., Berelson, B. y Gaudet, H. (1948). *The people's choice*. Columbia University Press.
- Lazarsfeld, P., Berelson, B., y Gaudet, H. (1960). *The people's choice*. Columbia University Press.
- Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., Postel, J., Roberts, L. G., y Wolff, S. (2017). *Breve historia de internet*. Internet Society.

- Lelkes, Y. (2016). Mass Polarization: Manifestations and Measurements. *Public Opinion Quarterly*, 80(S1), 392–410. <https://doi.org/10.1093/poq/nfw005>
- Lelkes, Y. (2018). Affective Polarization and Ideological Sorting: A Reciprocal, Albeit Weak, Relationship. *The Forum*, 16(1), 67–79. <https://doi.org/10.1515/for-2018-0005>
- Lerner, G., y Tusell, M. (2017). *La creación del patriarcado*. Katakarak.
- León-Mejía, A. (2006). *Feminismo disidente: un acercamiento a las posiciones críticas con el feminismo establecido desde la documentación y el análisis de la producción científica*. Junta de Andalucía.
- León-Mejía, A. C. (2010). *Una Aproximación analítica al feminismo del género*. Universitat Autònoma de Barcelona. <http://hdl.handle.net/10803/32084>
- Lodola, G., y Kitzberger, P. (2017). Politización y confianza en los medios de comunicación: Argentina durante el kirchnerismo. *Revista de Ciencia Política*, 37(3), 635–658. <https://doi.org/10.4067/S0718-090X2017000300635>
- Lopata, H. Z., y Thorne, B. (1978). On the term sex roles. *Signs*, 3(3), 718–721
- Lovins, J. B. (1968). Development of a Stemming Algorithm. *Mechanical Translation and Computational Linguistics*, 11 (1-2), 22-31.
- Luengo, Ó. G. (2005). Desafectos y medios de comunicación: el estado de la cuestión de una relación difusa. *Reflexión Política*, 7(14), 8-24.
- Luengo, Ó. G. (2014). Twitter vs medios tradicionales. ¿Ha implicado Twitter un espacio ciudadano real de intercambio de información? En R. Cotarelo, y J-A, Olmeda, (Eds.), *La democracia del siglo XXI. Política, medios de comunicación, internet y redes sociales* (pp. 409-428). Madrid. Centro de Estudios Políticos y Constitucionales.
- Luengo, Ó. G. (2016). Comunicación política: de la propaganda a las neurociencias. En C. Colino, J. F. Durá, J. A. Olmeda, P. Román Marugán, y J. Rubio Lara, (Eds.), *Ciencia política: Una aventura Vital* (pp. 721-740). Tirant Lo Blanch.
- López-García, G. (2005) Modelos de comunicación en Internet: desarrollo de una tipología. En G. López (Ed.) *El ecosistema digital: modelos de comunicación, nuevos medios y público en Internet* (pp. 55-85). Serveu de Publicacions de la Universitat de Valencia.
- López-García, G. (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>
- Manovich, L. (2013). *Software takes command* (Vol. 5). A&C Black.
- Markov, A. (1954). *Theory of algorithms*. Tr. Mat. Inst. Steklov, 42, 1-14. trad. por Edwin Hewitt en American Mathematical Society Translations, Series 2, Vol. 15 (1960).
- Mason, L. (2015). “I Disrespectfully Agree”: The Differential Effects of Partisan Sorting on Social and Issue Polarization: PARTISAN SORTING AND POLARIZATION. *American Journal of Political Science*, 59(1), 128–145. <https://doi.org/10.1111/ajps.12089>
- Mason, L. (2018). Ideologues without Issues: The Polarizing Consequences of Ideological Identities. *Public Opinion Quarterly*, 82(S1), 866–887. <https://doi.org/10.1093/poq/nfy005>
- Matz, S. C., Kosinski, M., Nave, G., y Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, 114(48), 12714. <https://doi.org/10.1073/pnas.1710966114>

- Maté Jiménez, C. (2014). Big data. Un nuevo paradigma de análisis de datos. *Anales de Mecánica y Electricidad*, 91(6), 10-16.
- Mazzoleni, G. (2008). Mediatization of Politics. En W. Donsbach (Ed.), *The International Encyclopedia of Communication*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781405186407.wbiecm062>
- Mazzoleni, G., y Schulz, W. (1999). "Mediatization" of Politics: A Challenge for Democracy? *Political Communication*, 16(3), 247-261. <https://doi.org/10.1080/105846099198613>
- McCarty, N. M. (2019). *Polarization: What everyone needs to know*. Oxford University Press.
- McCarty, N. M., Poole, K. T., y Rosenthal, H. (2016). *Polarized America: The dance of ideology and unequal riches (Second edition)*. MIT Press.
- McCarty, N. M., Poole, K., y Rosenthal, H. (2009). Does Gerrymandering Cause Polarization? *American Journal of Political Science*, 53(3), 666-680.
- McCombs, M. E., y Shaw, D. L. (1972). *The agenda-setting function of the press*. *Public Opinion Quarterly*, 36, 83-176.
- McCombs, M. E., y Shaw, D. L. (1993). The evolution of agenda-setting theory: 25 years in the marketplace of ideas. *Journal of Communication*, 43(2), 66-58.
- McCright, A. M., Xiao, C., y Dunlap, R. E. (2014). Political polarization on support for government spending on environmental protection in the USA, 1974-2012. *Social Science Research*, 48, 251-260. <https://doi.org/10.1016/j.ssresearch.2014.06.008>
- McLuhan, H. (1969). *La galaxia Gutenberg : génesis del "Homo typographicus"*. Aguilar.
- McLuhan, H. M. (1959). *Myth and Mass Media*. *Daedalus*, 88(2), 339-348.
- McLuhan, H. M. (1965). *Understanding media : the extensions of man*. McGraw-Hill.
- McNair, B. (2003). *An introduction to political communication (3ª ed.)*. Routledge.
- McQuail, D. (1979). The Influence and Effects of Mass Media. En J. Curran, M. Gurevitch, y J. Woollacott, (Eds.), *Mass Communication and Society* (pp. 7-23). SAGE.
- McQuail, D. (2000). *Introducción a la teoría de la comunicación de masas (3ª ed.)*. Ediciones Paidós Ibérica.
- McQuail, D. (2010). *Mcquail's mass communication theory (6ª ed.)*. Sage Publications.
- McWilliams, E. C. (2021). Affective polarization, evidence, and evidentialism. En *The Routledge Handbook of Political Epistemology* (pp. 145-155). Routledge.
- Mead, M. (1928). *Coming of Age in Samoa*. William Morrow.
- Mead, M. (1935). *Sex and Temperament*. William Morrow.
- Meng, M., Steinhardt, S., y Schubert, A. (2018). Application Programming Interface Documentation: What Do Software Developers Want? *Journal of Technical Writing and Communication*, 48(3), 295-330. <https://doi.org/10.1177/0047281617721853>
- Mikolov, T., Chen, K., Corrado, G., y Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. ArXiv:1301.3781 [Cs]. <http://arxiv.org/abs/1301.3781>

- Miller, M. M. (1997). Frame Mapping and Analysis of News Coverage of Contentious Issues. *Social Science Computer Review*, 15(4), 367-378. <https://doi.org/10.1177/089443939701500403>
- Miller, M.M. y Riechert, B.P. (2001). The spiral of opportunity and frame resonance. Mapping the issue cycle in news and public discourse. En S. D. Reese, O. H. Gandy y A. E. Grant (Eds.), *Framing public life. Perspectives on media and our understanding of the social world* (pp. 107-121). Mahwah, nj: Lawrence Erlbaum Associates.
- Millett, K. (2015). *Sexual politics*. Virago.
- Minsky, M. (1961). Steps toward Artificial Intelligence. *Proceedings of the IRE*, 49(1), 8-30. <https://doi.org/10.1109/JRPROC.1961.287775>
- Mio, J. S. (1997). Metaphor and Politics, *Metaphor and Symbol*, 12(2), 113-133. https://doi.org/10.1207/s15327868ms1202_2
- Mohammad, S. M. Y Turney, P. D. (2013). Crowdsourcing a Word-Emotion Association Lexicon, *Computational Intelligence*, 29(3), 436-465.
- Money, J., Hampson, J. G., y Hampson, J. L. (1955). Hermaphroditism: recommendations concerning assignment of sex, change of sex and psychologic management. *Bulletin of the Johns Hopkins Hospital*, 97(4), 284-300.
- Moreno-Mercado, J. M., y García-Marín, J. (2020). Conflictos armados y cobertura mediática: Aproximación al aprendizaje de máquina supervisado. *Convergencia Revista de Ciencias Sociales*, 27, 1-26. <https://doi.org/10.29101/crcs.v27i0.12872>
- Morstatter, F., Wu, L., Yavanoglu, U., Corman, S. R., y Liu, H. (2018). Identifying Framing Bias in Online News. *ACM Transactions on Social Computing*, 1(2), 1-18. <https://doi.org/10.1145/3204948>
- Munger, K., Luca, M., Nagler, J., y Tucker, J. (2020). The (Null) Effects of Clickbait Headlines on Polarization, Trust, and Learning. *Public Opinion Quarterly*, 84(1), 49-73. <https://doi.org/10.1093/poq/nfaa008>
- Muñiz, C. (2017). El análisis de contenidos mediáticos desde la teoría del framing. Abordajes metodológicos para la identificación y análisis de los encuadres noticiosos. En M. S. Paláu Cardona (Ed.), *La observación de los medios* (pp. 31-50). ROM CONEICC.
- Negrine, R., y Papathanassopoulos, S. (1996). The "Americanization" of Political Communication: A Critique. *Harvard International Journal of Press/Politics*, 1(2), 45-62. <https://doi.org/10.1177/1081180X96001002005>
- Nerlich, B., y Clarke, D. D. (2000). Semantic fields and frames: Historical explorations of the interface between language, action, and cognition. *Journal of Pragmatics*, 32(2), 125-150. [https://doi.org/10.1016/S0378-2166\(99\)00042-9](https://doi.org/10.1016/S0378-2166(99)00042-9)
- New, C. (2020). Sex and gender. A critical realist approach. En M. Van Ingen, S. Gronhmann y L. Gunnarsson (Eds.), *Critical realism, feminism, and gender: A reader* (pp. 80-98). Routledge.
- Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 2(2), 175-220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Nordmann, A. (2011). *Alltäglicher feminismus: Geschlecht als soziale Erfahrung und reflexive Kategorie* [El feminismo cotidiano: El género como experiencia social y categoría reflexiva]. Helmer.

- Norris, P. (1997). Political Communications. En P. Dunleavy, A. Gamble, I. Holiday, y G. Peele (Eds.), *Developments in British Politics* 5 (pp. 75–88). Macmillan Education UK. https://doi.org/10.1007/978-1-349-25862-8_4
- O'Reilly, T., y Battelle, J. (2009). *Web squared: Web 2.0 five years on*. O'Reilly Media, Inc.
- Paglia, C. (1994). *Vamps & tramps: New essays* (1ª ed.). Vintage Books.
- Paglia, C. (2001). *Sexual personae: Art and decadence from Nefertiti to Emily Dickinson*. Yale Nota Bene.
- Pariser, E. (2017). *El Filtro burbuja: cómo la red decide lo que leemos y lo que pensamos*. Taurus.
- Parsons, T. (1939). The Professions and Social Structure. *Social Forces*, 17(4), 457–467. <https://doi.org/10.2307/2570695>
- Pennington, J., Socher, R., y Manning, C. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. <https://doi.org/10.3115/v1/D14-1162>
- Petersen, M. B. (2020). The evolutionary psychology of mass mobilization: How disinformation and demagogues coordinate rather than manipulate. *Current Opinion in Psychology*, 35, 71–75. <https://doi.org/10.1016/j.copsyc.2020.02.003>
- Peterson, J. B. (2017). 2017/06/28: Postmodern NeoMarxism: Diagnosis and Cure. Recuperado el 6 de Julio de 2021 de <https://www.youtube.com/watch?v=s4c-jOdPTN8>
- Petty, R. E., Briñol, P. y Priester, J. R. (2009). Mass media attitude change: Implications of the elaboration likelihood model of persuasion. En J. Bryant y M. B. Oliver (Eds.), *Media effects: Advances in theory and research* (pp. 125–163). Routledge.
- Pfeffer, J. (1992). *Managing with power: politics and influence in organizations*. Harvard Business School Press.
- Picard, R. G., y Adams, P. D. (1991). Characterizations of Acts and Perpetrators of Political Violence in Three Elite U.S. Daily Newspapers. *Political Communication*, 4(1), 1–9, DOI: 10.1080/10584609.1987.9962803
- Pinker, St. (2008). The fear of determinism. En, J. Baer, J. C. Kaufman y R. F. Baumeister (Eds.), *Are we free? Psychology and free will* (pp. 311–324). Oxford University Press.
- Pinker, St. (2012). *La Tabla Rasa*. Espasa.
- Pinker, Su. (2009). *The sexual paradox: Men, women and the real gender gap*. Simon and Schuster.
- Poole, K. T., Lewis, J. B., Lo, J., y Carroll, R. (2008). Scaling Roll Call Votes with W-NOMINATE in R. SSRN *Electronic Journal*. <https://doi.org/10.2139/ssrn.1276082>
- Poole, K. T., y Rosenthal, H. (1984). The Polarization of American Politics. *The Journal of Politics*, 46(4), 1061–1079. <https://doi.org/10.2307/2131242>
- Poole, K., y Rosenthal, H. (1983). A Spatial Model for Legislative Roll Call Analysis. GSIA *Working Paper* No. 5–83–84.
- Poole, K., y Rosenthal, H. (1985). A Spatial Model for Legislative Roll Call Analysis. *American Journal of Political Science*, 29(2), 357–384. doi:10.2307/2111172

- Prior, M. (2013). Media and Political Polarization. *Annual Review of Political Science*, 16(1), 101–127. <https://doi.org/10.1146/annurev-polisci-100711-135242>
- Quandt, T., Frischlich, L., Boberg, S., y Schatto-Eckrodt, T. (2021). Fake News. En T.P. Vos, F. Hanusch, D. Dimitrakopoulou, M. Geertsema-Sligh y A. Sehl (Eds.), *The International Encyclopedia of Journalism Studies*. <https://doi.org/10.1002/9781118841570.iejs0128>
- RAE [Real Academia Española]: Diccionario de la lengua española, 23.^a ed., [versión 23.4 en línea]. Recuperado el 27 de Marzo de 2021 de <https://dle.rae.es/feminismo>
- Ramage, D., Rosen, E., Chuang, J., Manning, C. D., y McFarland, D. A. (2009). Topic modeling for the social sciences. En *NIPS 2009 workshop on applications for topic models: text and beyond*, 5, 27–31.
- Rekker, R. (2021). The nature and origins of political polarization over science. *Public Understanding of Science*, 30(4), 352–368. <https://doi.org/10.1177/0963662521989193>
- Reverter B., S. (2010). La deriva teórica del feminismo. *Daimon Revista Internacional De Filosofía*, 153–162.
- Riechert, B.P. (1996). *Advocacy group and news media framing of public policy issues: Frame mapping the wetlands debates* [Unpublished doctoral dissertation]. University of Tennessee.
- Risman, B. J., y Davis, G. (2013). From sex roles to gender structure. *Current Sociology*, 61(5–6), 733–755. <https://doi.org/10.1177/0011392113479315>
- Rivers, N. (2017). *Postfeminism(s) and the Arrival of the Fourth Wave*. Palgrave Macmillan. doi:10.1007/978-3-319-59812-3.
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., y Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, 58(4), 1064–1082. <https://doi.org/10.1111/ajps.12103>
- Roberts, M. E., Stewart, B. M., y Tingley, D. (2019). stm: An R Package for Structural Topic Models. *Journal of Statistical Software*, 91(1), 1–40. <https://doi.org/10.18637/jss.v091.i02>
- Rojas Mesa, J. E., y Leal Urueña, L. A. (2017). Affordance: constructo para la comprensión y transformación del aprendizaje en contextos interculturales. *Tecné, Episteme y Didaxis: TED*, (42). <https://doi.org/10.17227/01203916.6963>
- Sabsay, Leticia (2010) ¿En los umbrales del género? Beauvoir, Butler y el feminismo ilustrado. *Feminismo/s*, 15, 119–135.
- Sagi, E., Diermeier, D., y Kaufmann, S. (2013). Identifying Issue Frames in Text. *PLOS ONE*, 8(7), e69185. <https://doi.org/10.1371/journal.pone.0069185>
- Sartori, G. (1976). *Parties and Party Systems: A Framework for Analysis*. Cambridge University Press.
- Scheufele, D. A., y Tewksbury, D. (2007). Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models: Models of Media Effects. *Journal of Communication*, 57(1), 9–20. <https://doi.org/10.1111/j.0021-9916.2007.00326.x>
- Schildkraut, J., y Muschert, G. W. (2014). Media Salience and the Framing of Mass Murder in Schools: A Comparison of the Columbine and Sandy Hook Massacres. *Homicide Studies*, 18(1), 23–43. <https://doi.org/10.1177/1088767913511458>

- Schultz, T. (2000). Mass media and the concept of interactivity: an exploratory study of online forums and reader email. *Media, Culture & Society*, 22(2), 205–221. <https://doi.org/10.1177/016344300022002005>
- Schulz, W. (1998). Media Change and the Political Effects of Television: Americanization of the Political Culture? *Communications*, 23(4), 527–544. <https://doi.org/10.1515/comm.1998.23.4.527>
- Schuster, J. (2017). Why the personal remained political: Comparing second and third wave perspectives on everyday feminism. *Social Movement Studies*, 16(6), 647–659. <https://doi.org/10.1080/14742837.2017.1285223>
- Scott, R. L., y Smith, D. K. (1969). The rhetoric of confrontation. *Quarterly Journal of Speech*, 55(1), 1–8. <https://doi.org/10.1080/00335636909382922>
- Scrivens, R., y Conway, M. (2020). The roles of ‘old’ and ‘new’ media tools and technologies in the facilitation of violent extremism and terrorism. En Leukfeldt, E. R., y Holt, T. J. (Eds.), *The Human Factor of Cybercrime* (pp. 286–309). Taylor & Francis.
- Scruggs, J. F. (1998). “Echo chamber” approach to advocacy. Recuperadode <https://www.industrydocuments.ucsf.edu/docs/ggxn0061>
- Searle, J. (1985). *Mentes, cerebros y ciencia*. Cátedra.
- Selivanov, D., Bickel, M., y Wang, Q. (2020). Package “text2vec”. Recuperado el 7 de enero de 2021 de <https://cran.r-project.org/web/packages/text2vec/text2vec.pdf>
- Serrano-Contreras, I.-J., García-Marín, J., y Luengo, Ó. G. (2020). Measuring Online Political Dialogue: Does Polarization Trigger More Deliberation? *Media and Communication*, 8(4), 63–72. <https://doi.org/10.17645/mac.v8i4.3149>
- Serrano-Contreras, I-J, García-Marín, J, y Luengo, Ó. G. (2021). Coberturas mediáticas, polarización y reformas educativas en España. *Revista de ciencia política (Santiago)*, Epub 27 de julio de 2021. <https://dx.doi.org/10.4067/S0718-090X2021005000109>
- Serrano-Contreras, I-J. (2020). La minería de texto como método de análisis: una mirada a la administración local en la prensa española. *Revista Iberoamericana de Gobierno Local*, 16, 1-12.
- Serrano-Contreras I-J. (2021) NLP on YouTube: A Look on Feminism. En M. Musiał-Karg, O. G., Luengo (Eds), *Digitalization of Democratic Processes in Europe. Studies in Digital Politics and Governance*. Springer. https://doi.org/10.1007/978-3-030-71815-2_10
- Shiva, N., y Nosrat, Z. K. (2019). The Fourth Wave of Feminism and the Lack of Social Realism in Cyberspace. *Journal of Cyberspace Studies*, 3(2), 129–146. DOI: 10.22059/JCSS.2019.72454
- Sievert, C., y Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63–70. <https://doi.org/10.3115/v1/W14-3110>
- Sigelman, L., y Yough, S. N. (1978). Left-Right Polarization In National Party Systems: A Cross-National Analysis. *Comparative Political Studies*, 11(3), 355–379. <https://doi.org/10.1177/001041407801100304>
- Silge, J., y Robinson, D. (2017). *Text mining with R: A tidy approach*. O’Reilly.
- Slapin, J., y Proksch, S. (2008). A Scaling Model for Estimating Time-Series Party Positions from Texts. *American Journal of Political Science*, 52(3), 705–722.

- Sommers, C. H. (1995). *Who stole feminism?: How women have betrayed women*. Simon and Schuster.
- Sonntag, J. y Stede, M. (2014) Sentiment Analysis: What's Your Opinion? En C. Biemann, y A. Mehler (Eds.), *Text Mining*. Springer International Publishing. Springer. <https://doi.org/10.1007/978-3-319-12655-5>
- Spencer-Wood, S. M. (2017). Second-Wave Feminism. En B. S. Turner (Ed.), *The Wiley-Blackwell Encyclopedia of Social Theory* (pp. 1-3). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118430873.est0525>
- Spiller, T. (2019). *The breakdown: Making sense of politics in a messed up world*. <https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=2117731>
- Spitulnik, D. (1993). Anthropology and Mass Media. *Annual Review of Anthropology*, 22(1), 293-315. <https://doi.org/10.1146/annurev.an.22.100193.001453>
- Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review*, 34(3), 150-160. <https://doi.org/10.1177/0266382117722446>
- Stevenson, M. (2018). From Hypertext to Hype and Back Again: Exploring the Roots of Social Media in Early Web Culture. En J. Burgess, A. Marwick, y T. Poell (Eds.), *The SAGE Handbook of Social Media* (pp. 69-87). SAGE.
- Stroud, N. J. (2010). Polarization and Partisan Selective Exposure. *Journal of Communication*, 60(3), 556-576. <https://doi.org/10.1111/j.1460-2466.2010.01497.x>
- Strömbäck, J. (2008). Four Phases of Mediatization: An Analysis of the Mediatization of Politics. *The International Journal of Press/Politics*, 13(3), 228-246. <https://doi.org/10.1177/1940161208319097>
- Strömbäck, J., y Luengo, Ó. G. (2008). Polarized Pluralist and Democratic Corporatist Models: A Comparison of Election News Coverage in Spain and Sweden. *International Communication Gazette*, 70(6), 547-562. <https://doi.org/10.1177/1748048508096398>
- Sunstein, C. R. (2007). *Republic 2.0*. Princeton University Press.
- Suárez-Cao, J. (2013). Estados Unidos: Crisis económica, reelección presidencial y polarización política. *Revista de Ciencia política*, 33(1), 185-205. <https://doi.org/10.4067/S0718-090X2013000100009>
- Tajfel, H., y Turner, J. C. (1979). An Integrative Theory of Intergroup Conflict. En W. G. Austin, y S. Worchel (Eds.), *The Social Psychology of Intergroup Relations* (pp. 33-47). Brooks/Cole.
- Talib, R., Kashif, M., Ayesha, S., Fatima, F.: Text Mining: Techniques, Applications and Issues.(2016). *International Journal of Advanced Computer Science and Applications*, 7(11), 1-5. <https://doi.org/10.14569/ijacsa.2016.071153>.
- Tandel, S. S., Jamadar, A., y Dudugu, S. (2019). A Survey on Text Mining Techniques. 2019 5th *International Conference on Advanced Computing & Communication Systems (ICACCS)*, 1022-1026. <https://doi.org/10.1109/ICACCS.2019.8728547>
- Taylor, M., y Herman, V. M. (1971). Party Systems and Government Stability. *American Political Science Review*, 65(1), 28-37. <https://doi.org/10.2307/1955041>

- Teinemaa, I., Dumas, M., Maggi, F. M., y Di Francescomarino, C. (2016). Predictive Business Process Monitoring with Structured and Unstructured Data. En M. La Rosa, P. Loos, y O. Pastor (Eds.), *Business Process Management* (pp. 401-417). Springer International Publishing.
- Thomson, I. (2019). *Heidegger's Aesthetics > Notes (Stanford Encyclopedia of Philosophy)*. Recuperado el 28 de 2021 de <https://plato.stanford.edu/entries/heidegger-aesthetics/notes.html>
- Torcal, M., y Martini, S. (2013). *Los efectos negativos de la polarización política: confianza social, partidismo e identidades nacionales/territoriales en España*. Anuario de la Facultad de Derecho de la Universidad Autónoma de Madrid.
- Trajtenbrot, B. A. (1977). *Los algoritmos y la resolución automática de problemas*. MIR
- Tucker, J., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., y Nyhan, B. (2018). Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3144139>
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 49, 433-460. doi:10.1093/mind/LIX.236.433
- Turow, J. (2011). *Media today: An introduction to mass communication*. Routledge.
- UNWOMEN. (2021). *Conferencias mundiales sobre la mujer*. Recuperado el 31 de marzo de 2021 de <https://www.unwomen.org/es/how-we-work/intergovernmental-support/world-conferences-on-women>
- Valera Ordaz, L. (2016). El sesgo mediocéntrico del “framing” en España: Una revisión crítica de la aplicación de la teoría del encuadre en los estudios de comunicación. *ZER - Revista de Estudios de Comunicación*, 21(41), 13-31. <https://doi.org/10.1387/zer.17259>
- Valkenburg, P. M., y Peter, J. (2013). The Differential Susceptibility to Media Effects Model: Differential Susceptibility to Media Effects Model. *Journal of Communication*, 63(2), 221-243. <https://doi.org/10.1111/jcom.12024>
- Van Dijk, T. A. (1991). *Racism and the Press*. Routledge.
- Vanden Broucke, S., y Baesens, B. (2018). *Practical Web Scraping for Data Science*. Springer.
- Varela, N. (2020). El tsunami feminista. *Nueva Sociedad*, 286, 93-106.
- Vergeer, M. (2020). Artificial Intelligence in the Dutch Press: An Analysis of Topics and Trends. *Communication Studies*, 71(3), 373-392. <https://doi.org/10.1080/10510974.2020.1733038>
- Vijayarani, S., Ilamathi, J., y Nithya, S. (2015). Preprocessing Techniques for Text Mining-An Overview Dr. *International Journal of Computer Science & Communication Networks*, 5(1), 7-16.
- Viswanath, K., Ramanadhan, S., y Kontos, E. Z. (2007). Mass Media. En S. Galea (Ed.), *Macrosocial Determinants of Population Health* (pp. 275-294). Springer New York. https://doi.org/10.1007/978-0-387-70812-6_13
- Vázquez, J. B. G. (2012). *Análisis y diseño de algoritmos*. Red Tercer Milenio.
- Wagner, M. (2021). Affective polarization in multiparty systems. *Electoral Studies*, 69, 102199. <https://doi.org/10.1016/j.electstud.2020.102199>

- Walby, S. (1989). Theorising Patriarchy. *Sociology*, 23(2), 213–234. <https://doi.org/10.1177/0038038589023002004>
- Wallace, M. (2007). *Jawbone Java WordNet API*. <https://sites.google.com/site/mfwallace/jawbone>
- Walters, M. (2005). *Feminism: A very short introduction*. Oxford Univ. Press.
- Weaver, D. H. (2007). Thoughts on Agenda Setting, Framing, and Priming. *Journal of Communication*, 57(1), 142–147. <https://doi.org/10.1111/j.1460-2466.2006.00333.x>
- Webb Williams, N., Casas, A., y Wilkerson, J. (2020). *Images as Data for Social Science Research: An Introduction to Convolutional Neural Nets for Image Classification (Elements in Quantitative and Computational Methods for the Social Sciences)*. Cambridge University Press. doi:10.1017/9781108860741
- Webster, J. J., y Kit, C. (1992). Tokenization as the initial phase in NLP. *Proceedings of the 14th conference on Computational linguistics*, 1106–1110. <https://doi.org/10.3115/992424.992434>
- WHO. (2021). *Gender and health*. Recuperado el 5 de marzo de 2021 de <https://www.who.int/westernpacific/health-topics/gender>
- Wijffels, J., BNOSAC, y Yan, X. (2020). ‘Package “BTM”’. Recuperado el 23 de febrero de 2021 de <https://cran.r-project.org/web/packages/BTM/BTM.pdf>
- Wilkerson, J., y Casas, A. (2017). Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges. *Annual Review of Political Science*, 20(1), 529–544. <https://doi.org/10.1146/annurev-polisci-052615-025542>
- Winston, E. (1995). Spatial mapping in comparative discourse frames. En K. Emmorey, y J. S. Reilly (Eds), *Language, gesture, and space* (pp. 87–114). Lawrence Erlbaum.
- Wojcieszak, M., y Garrett, R. K. (2018). Social Identity, Selective Exposure, and Affective Polarization: How Priming National Identity Shapes Attitudes Toward Immigrants Via News Selection. *Human Communication Research*, 44(3), 247–273. <https://doi.org/10.1093/hcr/hqx010>
- Yan, X., Guo, J., Lan, Y., y Cheng, X. (2013). A biterm topic model for short texts. *Proceedings of the 22nd International Conference on World Wide Web*, 1445–1456. <https://doi.org/10.1145/2488388.2488514>
- Youmans, G. (1990). Measuring Lexical Style and Competence: The Type-Token Vocabulary Curve. *Style*, 24(4), 584–599. Retrieved August 6, 2021, from <http://www.jstor.org/stable/42946163>
- Yu, X., Wojcieszak, M., Lee, S., Casas, A., Azrout, R., y Gackowski, T. (2021). The (Null) Effects of Happiness on Affective Polarization, Conspiracy Endorsement, and Deep Fake Recognition: Evidence from Five Survey Experiments in Three Countries. *Political Behavior*. <https://doi.org/10.1007/s11109-021-09701-1>
- Zampolli, A., Calzolari, N., y Palmer, M. (1994). *Current Issues in Computational Linguistics: In Honour of Don Walker*. Springer Netherlands. <https://doi.org/10.1007/978-0-585-35958-8>
- Zannettou, S., Chatzis, S., Papadamou, K., y Sirivianos, M. (2018). The good, the bad and the bait: Detecting and characterizing clickbait on youtube. En *2018 IEEE Security and Privacy Workshops (SPW)* (pp. 63–69). IEEE.

Zipf, G. K. (1949). *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley Press.

Anexos

Publicaciones originales

Serrano-Contreras, I-J, García-Marín, J, y Luengo, Ó. G. (2021). Coberturas mediáticas, polarización y reformas educativas en España. *Revista de ciencia política (Santiago)*, Epub 27 de julio de 2021. <https://dx.doi.org/10.4067/S0718-090X2021005000109>

Coberturas mediáticas, polarización y reformas educativas en España

Media coverage, polarization and educational reforms in Spain

Resumen

Las leyes de educación comportan, para todos los partidos políticos, una de las correas transmisoras de su proyecto ideológico. De este modo, cada uno de los diferentes gobiernos españoles, como ocurre en otros países, han llevado a cabo modificaciones legislativas con el fin de implementar políticas de un signo más cercano a sus intereses partidistas. A su vez, el contexto mediático, siguiendo los postulados de los académicos en comunicación política, ha tendido a apoyar u oponerse a dichas reformas, generando una escena polarizada. La presente investigación realiza un análisis del panorama español, durante los últimos cuarenta años, sobre los impactos que las reformas educativas han suscitado en el ámbito mediático. Se emplean para ello nuevas metodologías de aprendizaje de máquina (*machine learning*), complementadas con el desarrollo de una serie de medidores diseñados para captar fenómenos sociales complejos dentro de escenarios de variabilidad semántica. Los resultados han evidenciado una clara relación entre partidos políticos y medios, así como el posicionamiento que estos últimos, en consonancia con su línea editorial, mostraron con respecto a las reformas educativas que se iban sucediendo a lo largo del periodo de estudio.

Palabras clave: sistemas mediáticos, cobertura mediática, reformas educativas, polarización, *machine learning*.

Abstract

Education laws involve, for all political parties, one of the most visible flagships of their ideological project. In this regard, different governments have made legislative modifications to introduce orientations in this field closer to their partisan interests. Therefore, following existing models of political communication, in Southern European countries the coverage of this information leads to a potential dichotomized media context. The present research displays an analysis of the Spanish case during the last forty years, to verify the influence that educational reforms have provoked in the press coverage. For that purpose, we have used new machine learning methodologies, complemented by the

development of self measures, to capture complex social phenomena within semantic variability scenarios. As a result of those processes, findings have evidenced a clear relationship between the ideological proposal of political parties and the editorial line of the printed press, regarding the mentioned educational reforms that were developed under the period of study.

Keywords: Media Systems, Media Coverage, Educational Reforms, Polarization, Machine Learning.

I. Introducción

La educación es uno de los ámbitos de actuación estratégica más importantes, sensibles y potencialmente conflictivos para los poderes públicos o, en palabras de Gil et al. (1998) una inversión estratégica de futuro. En la mayoría de los casos, marca la línea ideológica de los gobiernos en el poder, siendo un espacio de política pública donde no siempre es fácil encontrar el consenso entre los principales partidos políticos y, por lo tanto, se hace extremadamente complicado el objetivo de implementar políticas de estado que puedan soportar y sobrevivir a cambios de gobierno y de ciclo económico. El sector educativo es un terreno especialmente abonado para el desarrollo y defensa de posiciones beligerantes en materia religiosa, lingüística, identitaria e ideológica en España (Gortázar y Moreno 2017).

Desde los años 80, y con especial intensidad en esa década, hemos pasado por un periodo de reformas en el ámbito educativo español que han afectado a todos los niveles de formación: a protagonistas y responsables, a los estudiantes y al profesorado, o a la situación laboral de estos últimos. Aunque las primeras reformas encontraron como elemento dinamizador la reflexión y la voluntad de cambio, consolidando así el gran paradigma de la transición política, los hechos han terminado conformando una senda marcada por la gran cantidad de cambios normativos, estructurales, curriculares, organizativos y administrativos. El caso español ha sido señalado por los expertos como un ejemplo complejo donde se han sucedido reformas que, en muchos casos, han sido abortadas, frustradas, torpedeadas o abandonadas como consecuencia de cambios políticos, incluso dentro del mismo gobierno o del mismo partido (Viñao 2006: 50). De este modo, ha terminado por imponerse, como tónica general en las distintas legislaturas, una deriva sustentada en la propuesta y aprobación de nuevas leyes educativas.

Así, tras la aprobación de la constitución de 1978, se han sucedido en España casi una decena de reformas educativas a varios niveles, impulsadas por distintos partidos políticos que ostentaban el poder, y que proponían cambios de mayor o menor calado en el marco regulatorio vigente. En 1980, bajo el gobierno de la desaparecida UCD (Unión de Centro Democrático) se aprobó la LOECE⁵³, que nunca entró en vigor. El Gobierno socialdemócrata presidido por Felipe González (PSOE, Partido Socialista Obrero Español) impulsó la Ley de Reforma Universitaria

⁵³ Ley Orgánica 5/1980, de 19 de junio, por la que se regula el Estatuto de Centros Escolares.

(LORU⁵⁴) en 1983, la LODE⁵⁵ (1985) y, posteriormente, la LOGSE⁵⁶ (1990). El Partido Popular (PP) ratificó la Ley de Universidades (LOU⁵⁷) en 2001, la Ley de Formación Profesional⁵⁸ y la LOCE⁵⁹ en 2002. De nuevo el PSOE en 2006 consigue aprobar la LOE⁶⁰, ampliada un año más tarde para derogar la LOU⁶¹. A la última gran reforma educativa vigente hasta 2020, la LOMCE⁶² de 2013, promovida por el gobierno conservador del PP, le han sucedido nuevos cambios con la tramitación en noviembre de 2020 de la denominada “Ley Celaá” (LOMLOE⁶³), que está generando las mismas discrepancias que sus predecesoras (y que, por su reciente aprobación, no se incluye en este estudio).

Todos estos modelos políticos sobre la educación y la contundencia de los cambios o las estrategias, han impulsado o limitado la fragmentación de su tramitación parlamentaria, lo cual podría interpretarse como un indicador de polarización política. A la vista de las votaciones la aritmética parlamentaria refleja que la polarización mayor fue la referida a la LOMCE (con el Partido Popular como único partido a favor, con 182 votos a favor, y 130 votos en contra de más de 5 grupos parlamentarios⁶⁴), siendo además promulgada bajo un clima de gran contestación social en las calles.

Las reformas educativas de magnitud han venido reflejando presiones sociales, culturales, económicas, ideológicas, históricas, e incluso religiosas. Se constituye así un panorama con intereses y visiones difícilmente coincidentes sobre el mundo del conocimiento, la educación y sus funciones (Monarca et al. 2016). A los procesos macro le acompañan otro tipo de dinámicas más cercanas que pueden generar resistencia a las reformas y, consecuentemente, polarización (González y Bouza 2009), como la oposición gremial, la carencia de apoyos sociales o políticos, los diagnósticos erróneos, la falta de dotación presupuestaria, los objetivos y calendarios irreales, la premura (Viñao 2006), o el posicionamiento ideológico de los medios de comunicación; su beligerancia o actitud proactiva. Son estos últimos los que se revelan como el principal escenario donde se mantienen y escenifican los asuntos públicos (Bennett et al 2004), además de un agente capaz de influir tanto en los actores políticos como en los ciudadanos (Collins y Cooper 2012).

II. Pluralismo Polarizado

⁵⁴ Ley Orgánica 11/1983, de 25 de agosto, de Reforma Universitaria.

⁵⁵ Ley Orgánica 8/1985, de 3 de julio, reguladora del Derecho a la Educación.

⁵⁶ Ley Orgánica 1/1990, de 3 de octubre, de Ordenación General del Sistema Educativo.

⁵⁷ Ley Orgánica 6/2001, de 21 de diciembre, de Universidades.

⁵⁸ Ley Orgánica 5/2002, de 19 de junio, de las Cualificaciones y de la Formación Profesional.

⁵⁹ Ley Orgánica 10/2002, de 23 de diciembre, de Calidad de la Educación.

⁶⁰ Ley Orgánica 2/2006, de 3 de mayo, de Educación.

⁶¹ Ley Orgánica 4/2007, de 12 de abril, por la que se modifica la Ley Orgánica 6/2001, de 21 de diciembre, de Universidades.

⁶² Ley Orgánica 8/2013, de 9 de diciembre, para la Mejora de la Calidad Educativa.

⁶³ Ley Orgánica 3/2020, de 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de Educación.

⁶⁴ Partido Socialista Obrero Español, Convergencia i Unió, Izquierda Unida, Unión Progreso y Democracia, Partido Nacionalista Vasco y Grupo Mixto.

Las disensiones ideológicas que presentan la promulgación de cambios educativos se ven también extrapoladas a la arena mediática. Las líneas editoriales de los medios se transforman en un factor donde la neutralidad periodística se pone en tela de juicio, con lo que dejan a un lado su papel de contrapoder (Balkin 1999). Conocedores los medios del influyente rol que desempeñan en la confección de políticas públicas (algo que confirman las teorías de agenda, por ejemplo), se alinean en una posición estratégica con la que establecen tanto a unos receptores subjetivos y contextualizados (Arriagada et al. 2010; Lodola y Kitzberger 2017), como a un encuadre específico para el mensaje (Entman 1993).

Siguiendo el modelo de Hallin y Mancini (2004), los países incluidos en el pluralismo polarizado se caracterizan, entre otras cosas, por la militancia política de sus medios de comunicación, esto es, por su paralelismo político (Humanes et al. 2013). Es de esperar entonces que, en periodos de reformas educativas, las posiciones expresadas por los principales partidos políticos; tanto por el que presenta el proyecto de ley, como por el que ejerce el papel de oposición, sean asumidas por los principales medios de comunicación a través de líneas editoriales opuestas (González et al. 2010). En este sentido no existe una producción académica muy fecunda que respalde la relación entre medios y reformas educativas. La poca literatura sobre el tema ha tendido a confirmar la idea general según la cual los medios construyen y transmiten una imagen negativa de la educación (Runte 2016), señalando a los gobiernos como responsables directos.

La polarización política ha sido un objeto de investigación por parte de múltiples académicos quienes, desde una aproximación multidisciplinar, han tratado de abordar este fenómeno social. Como ejemplo podemos destacar las diversas acepciones expuestas por Bramson et al. (2017), o el medidor geométrico-dicotómico de Poole (2008) de los años 80 del pasado siglo. Todos señalan que la polarización se genera cuando se produce una posición en favor de determinadas creencias, eludiendo su contrastación (Olsson 2013; Mason 2014).

El análisis de la disensión entre polos también está presente en los estudios de comunicación política (véase, por ejemplo, la excelente contribución de Sunstein 2002), especialmente en relación con los medios de comunicación (Prior 2013) y las perspectivas políticas (Gentzkow, 2016). Por lo general, la polarización mediática puede interpretarse como la cobertura ideológicamente parcial y divergente de los medios (DiMaggio et al. 1996), aunque la mayoría de los estudios previos no ofrecen una definición operativa explícita del concepto (exceptuando el reciente trabajo de Boxell et al. 2020) puesto que, al menos a nivel más básico, se concibe como una noción de naturaleza suficientemente clara (Fletcher y Jenkins 2019). Sin embargo, la polarización también se ha definido como un efecto mediático o incluso un efecto político (Bernhardt et al. 2008). Durante los últimos años, los académicos han encontrado múltiples pruebas de que, en determinadas esferas mediáticas, no solo en los medios de comunicación convencionales sino también en las redes sociales y el entorno general de Internet, se puede producir un aumento de, y en, la polarización política o ideológica (Tucker et al. 2018), viéndose

auspiciado por una sobreexposición, directa o indirecta, de nuestro modo y forma de consumir medios (De Vries 2020). No obstante, los vínculos potenciales que cabrían esperar entre la cobertura partidista de los medios de comunicación y la creciente fragmentación política de la ciudadanía no parecen haberse visto sustentados por una extensa evidencia empírica, aunque el potencial en este sentido es señalado de forma generalizada (Fletcher y Jenkins 2019).

La mayoría de los estudios que han abordado la polarización en la cobertura mediática, que además de no ser tan numerosos, no proponen una aproximación empírica contundente, habiéndose focalizado especialmente en los medios estadounidenses (Lowry y Shidler 1998; McCluskey y Kim 2012) o en temas que generan polarización intrínsecamente, particularmente tras determinados eventos llamativos para la opinión pública (Aruguete y Calvo 2018). En este sentido destacan, por ejemplo, los análisis sobre inmigración (Benson 2010; Balch y Balabanova 2011) o corrupción (Mancini et al. 2017). Se puede percibir un aumento de la investigación sobre polarización en las redes sociales en detrimento de los medios tradicionales (Tucker et al. 2018; Boxell et al. 2020).

A tal efecto, puede argumentarse que la política educativa es un tema que, por su propia naturaleza, induce a la polarización. En el caso concreto de España, por ejemplo, es esperable que los periódicos más cercanos a las ideologías conservadoras destaquen los aspectos negativos de los resultados de la política educativa socialdemócrata. Mientras, aquellos que defienden otros valores, como la equidad y la igualdad de acceso a la escuela, encuadran de forma desfavorable las reformas emprendidas por el partido conservador (Runte 2016: 723).

El objetivo de este artículo, entonces, es el de valorar, aplicando una metodología novedosa, el grado de polarización en los medios de comunicación escritos en relación con las reformas educativas que se han llevado a cabo en España desde la transición a nuestros días. Tras este proceso, podremos verificar el grado de paralelismo político (alineamiento ideológico) mostrado por los medios de comunicación de referencia siguiendo la clasificación de Hallin y Mancini (2004). Los medios de comunicación analizados son los diarios ABC y El País. Mientras que el primero cuenta con una línea conservadora y tradicional, que evoluciona hacia postulados más liberales en la actualidad, el segundo destaca por un posicionamiento más progresista, alineándose en torno ideales socialdemócratas (Humanes 2014). Lo que se espera es que la contraposición mediática, revelada en la cobertura (coincidiendo con los periodos de reformas educativas iniciadas por los partidos políticos ideológicamente opuestos), muestre cotas de dispersión (polarización), sirviendo éstas para engrasar y apuntalar el sistema y las instituciones (Kennamer 1994; Aruguete 2013; Martínez et al. 2014).

III. Diseño y muestra

Las hipótesis de la investigación están basadas, siguiendo los postulados teóricos anteriores, en la ideología como predictor del comportamiento mediático. Así, establecemos que:

H.1: La visibilidad (cantidad) de la cobertura mediática (saliencia) estará relacionada con la ideología del medio de comunicación.

H.2: Los subtemas con los que se aborda dicha cobertura divergen de acuerdo a las diferencias ideológicas.

H.3: La cobertura mediática tenderá a una mayor polarización cuando la línea editorial del medio sea opuesta a la ideología del partido en el gobierno.

En resumen, la cobertura entre los dos medios de comunicación analizados (ABC y El País) debería mostrar diferencias significativas y mensurables en la cantidad de artículos dedicados a las reformas educativas, pero también con respecto a los temas asociados (Miller 1997) a dichas reformas, y al tono o polarización de estas. Asimismo, cabría esperar que dichas diferencias se deban a la distancia ideológica de los diarios con respecto al partido del gobierno en ese momento.

La finalidad de esta propuesta es verificar las hipótesis planteadas abarcando todo el desarrollo democrático reciente de España, es decir, los últimos cuarenta años. Por ello, se han seleccionado los dos diarios de tirada nacional más representativos del periodo, siendo, a su vez, los únicos que abarcan toda la muestra temporal de estudio.

El corpus se conformó mediante una búsqueda por palabras clave parametrizadas mediante la horquilla temporal: 6 de diciembre de 1978 a 6 de diciembre de 2018, en la que se tomó como punto de partida el referéndum de ratificación de la Constitución española y, como cierre, su conmemoración 40 años después. Las palabras clave empleadas fueron «reforma educativa» y «ley de educación». El segundo conjunto de palabras fue añadido ya que, al iniciarse la búsqueda en primera instancia, se mostraban diferencias muestrales entre uno y otro medio. Este cambio a la hora de añadir un nuevo conjunto se podría tornar sintomático en la medida en que, los temas empleados, podrían convertirse en una variable significativa. Esto podría tenerse en cuenta para la confección y tratamiento de los artículos (se han incluido piezas de información y opinión). Las palabras clave al ser buscadas podrían encontrarse tanto en el título, subtítulo o cuerpo de la noticia. El total de esta muestra fue elaborado mediante el uso de las hemerotecas online de ABC y El País (1978-1995) y la base de datos My News⁶⁵ (1996-2018). Mientras que los artículos obtenidos de la hemeroteca de ABC fueron conseguidos y conformados de forma manual, los correspondientes a El País se obtuvieron a través de un modelo de producción propia mediante la técnica de scraping. De este modo, y después de depurar la muestra y eliminar aquellos artículos que no trataban sobre reformas educativas concernientes al ámbito español, se obtuvo un

⁶⁵My News: <https://hemeroteca.mynews.es/>

total de 4872 artículos (2477 ABC y 2395 El País, un número asombrosamente similar para ambos medios a lo largo de cuarenta años).

Una vez adquirido el corpus, el texto se procesó siguiendo los estándares en este tipo de estudios (eliminación de símbolos atípicos, eliminación de espacios en blanco, tokenización, etc., Feinerer et al. (2008), muestran las diferentes estrategias para el procesamiento de textos en R (lenguaje que usamos en esta investigación) para, posteriormente, ser testeado en dos fases (más allá de las meras frecuencias). En primer lugar, se realizaron puntajes previos por medio de un análisis de sentimiento, con el fin de calificar la muestra. El proceso implementado se denomina SentimentR (Rinker 2019), y realiza un cálculo de la ambivalencia por contexto, donde se dejan a un lado los modelos clásicos por unidades de palabras (frecuencias) y se implementa un sistema contextual en el que se tienen en cuenta los cambios de valencia. Se busca, así, no distorsionar en gran medida la semántica ya que, en los análisis donde la partícula es la palabra, no se recoge una valoración íntegra, soslayando recursos del lenguaje como las expresiones retóricas. Esto propicia que la ambigüedad no sea captada en toda su riqueza y se confeccione una calificación más literal.

No obstante, como ese tipo de análisis no aporta más información que una declaración sobre lo positivo o negativo de las palabras o frases, hemos implementado una forma de medir la polarización (véase Serrano-Contreras et al. 2020 para una explicación más pormenorizada). Para ello, se operacionaliza la polarización de un artículo (p_c) como la distancia entre el análisis de sentimiento de ese artículo (S) y la mediana del agregado de los análisis de sentimiento de todos los artículos para ese diario (Me), en número absoluto. De esta forma, se obtiene un número que puede tomar cualquier valor entre 0 y 2, donde 0 muestra que no hay polarización (aunque pueda mostrarse un sentimiento tanto positivo como negativo) y 2 muestra una polarización máxima. El ratio que establece nuestro índice tiene la capacidad de medir las posibles divergencias que, dentro de un arco temporal, pueden generarse en un medio.

$$P_c = |(S - Me)|, \quad \text{donde } S \in [-1, 1], p \in (0, 2)$$

Complementariamente a lo anterior, y debido al número de artículos extraídos, hemos juzgado necesaria la aplicación de una técnica de extracción y reducción de la información contenida en la muestra. De esta forma podemos reducir la complejidad temática contenida en casi cinco mil artículos de periódico a unos pocos temas o ideas presentes en la mayoría de ellos. A este respecto, existen multitud de técnicas existentes, desde algoritmos supervisados (véase, por ejemplo, García-Marín y Calatrava 2018) hasta los no supervisados (como los LDA (*Latent Dirichlet Allocation*) y LSI (*Latent Semantic Indexing*)). Como la muestra recoge publicaciones durante más de 40 años, los algoritmos supervisados no parecen aconsejables, ya que el entrenamiento debería estar basado en todos los temas existentes (y para todas las épocas). Por ello, hemos decidido la aplicación de una técnica no supervisada, en concreto por un algoritmo LSI. Este algoritmo es

una técnica enmarcada dentro del Procesamiento del Lenguaje Natural (NLP por sus siglas en inglés) que se basa en analizar las relaciones entre un conjunto de documentos y los términos que contienen (por lo que está basado en frecuencias) mediante la producción de conceptos relacionados con esos documentos y términos. El algoritmo utiliza los conjuntos de textos de nuestro corpus para producir como resultado una serie de temas presentes (k), a través de la distancia vectorial entre las palabras y a partir de los documentos, en nuestro caso, los artículos periodísticos (una buena explicación de su funcionamiento y utilidad puede encontrarse en Letsche y Berry (1997), lo que muestra que es una técnica bastante consolidada).

IV. Análisis y resultados

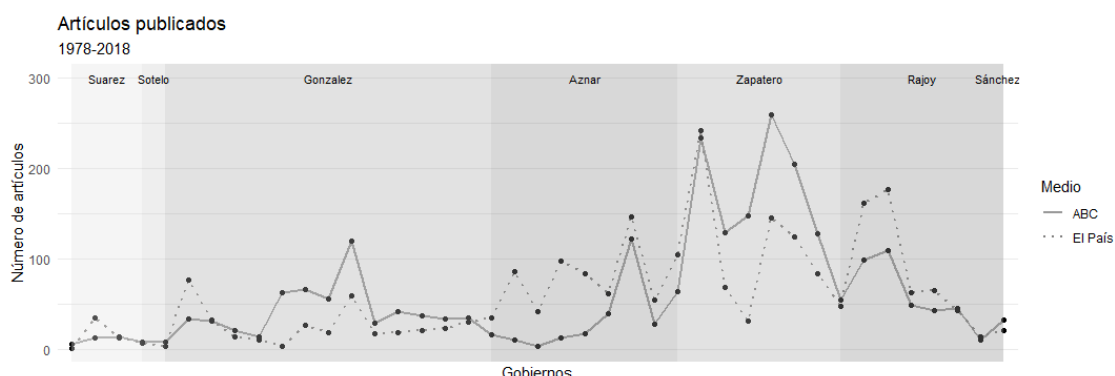
Los datos analizados se extienden a lo largo de un marco temporal de 40 años en el que se suceden siete presidentes del gobierno bajo las siglas de tres partidos políticos distintos. De un total de 4872 artículos, 3012 han sido publicados en periodos socialistas, seguidos de 1759 en gobiernos populares y 101 durante los de la desaparecida UCD⁶⁶. Desde un punto de vista más pormenorizado destacan las legislaturas presididas por el presidente Rodríguez Zapatero, quien acapara casi la mitad de la muestra, seguido de Felipe González y José María Aznar, respectivamente. Como podemos apreciar en el gráfico 1, los datos revelan una más que evidente significación entre el número de artículos publicados y el partido que se encuentra en el gobierno. Es decir que, mientras el color del gobierno se aleja de la línea ideológica del medio, este aumenta el número de artículos publicados, evidenciándose de forma directa, y casi total, en todos los periodos analizados. Por consiguiente, lo que se demuestra, en línea con lo expuesto por Budge y Farlie (1983) o Helbling y Tresch (2011), es que las discrepancias ideológicas no se manifiestan al confrontar posturas sobre un mismo tema, sino que es a través del énfasis que se da a una determinada temática lo que propicia que la atención vire y los posicionamientos tiendan a alejarse.

A su vez, cabe destacar un dato singular: cuando en el artículo se emplea alguno de los dos temas utilizados para la búsqueda, «reforma educativa» y «ley de educación», el otro no aparece, ocurriendo esta convergencia sólo en el 2,38% de las ocasiones. Algo que podría mostrar diferencias semánticas e ideológicas, aunque los datos no revelan posicionamiento ideológico alguno y se podría deber más a un simple relato de la cuestión, como refleja de forma clara el tramo temporal en torno a la ley de educación catalana (2009). En cualquier caso, parece que, efectivamente, la visibilidad del tema sí está relacionada con la diferencia ideológica entre el medio de comunicación y el partido en el gobierno, por lo que la primera hipótesis se confirma. Así se muestra en el periodo analizado, por ejemplo, en el último año -2018-, donde en los primeros meses, y con gobierno del

⁶⁶ La UCD es una plataforma política creada por Adolfo Suárez en el año 1977 y disuelta en 1983 (Martínez 2017). A pesar de su corto periodo de existencia, la formación hizo las veces de correa de transmisión entre el viejo régimen franquista y la democracia que empieza a construirse en aquellos años.

Partido Popular, El País publica más que ABC. Tendencia que se invierte en la segunda mitad del año con la llegada al poder de Pedro Sánchez.

Gráfico 1. Agregado de artículos por año

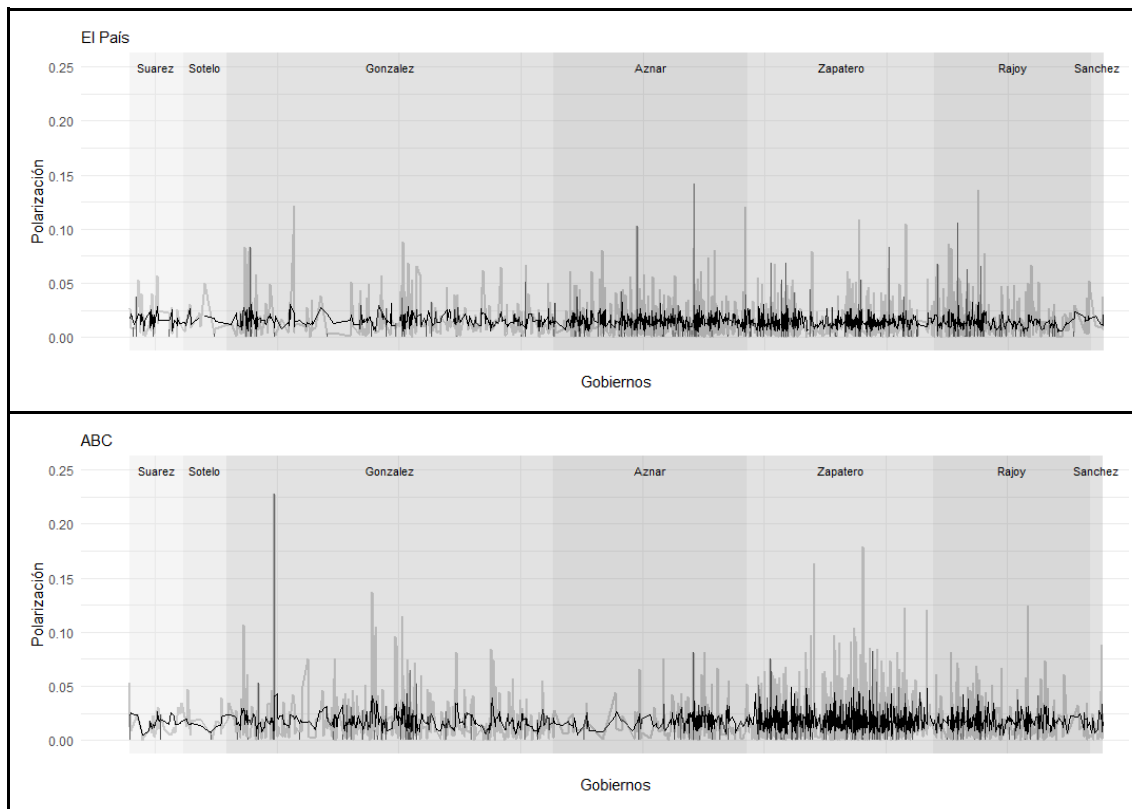


Elaboración propia. Nota sobre agregados 2018⁶⁷

Con respecto a la polarización de los medios de comunicación, el gráfico 2 muestra este parámetro de forma cronológica y contextual. Como se puede apreciar, el índice parece exponer de manera fiel cómo se produce una dispersión temporal entre medios y gobiernos. Si bien, a lo largo del periodo de análisis, la muestra se va comportando de manera diferente, demostrando que el índice al menos es capaz de mostrar variaciones consistentes con la realidad observable y, además, a nuestro parecer, de forma adecuada. A pesar de ello, no se puede asegurar que las altas cotas de polarización sean una constante dentro de los periodos de análisis, aunque sí se destilan ciertos patrones en los que los medios tienden a polarizar su cobertura en función del partido que se encuentra en el poder. Sin embargo, es importante tener presente que la polarización no indica la dirección de la misma; es decir, entendemos que la ideología es un predictor de la polarización, pero ésta puede estar motivada por el propio debate político y mediático, por lo que el apoyo a una determinada reforma educativa podría hacerse también desde posturas polarizantes, sobre todo si se producen ataques a la misma por parte de la oposición.

Gráfico 2. Polarización El País y ABC (1978-2018)

⁶⁷ Nota sobre agregados 2018. Durante el año 2018 hubo un cambio en la presidencia del Gobierno, produciendo de forma desagregada: Rajoy [ABC: 9 - El País: 11] y Sánchez [ABC: 24 - El País: 8]



Elaboración propia. La línea oscura representa la media móvil diaria

Por último, dentro de los conjuntos de análisis llevados a cabo se encuentra la aplicación de un algoritmo no supervisado, en concreto, un LSI. Este modelo de descomposición semántica mediante vectorización (Hoffmann 1999) señala la variabilidad de subtemas que esta temática provoca en los dos periódicos, tal y como reflejan las tablas 1 y 2. La tabla 1 muestra la cobertura de El País para el período de análisis. Naturalmente, los temas seleccionados tienen que ser, debido al tamaño de la muestra, forzosamente genéricos. Sin embargo, dos elementos parecen importantes: primero, que las referencias al «partido popular» han sido más o menos constantes en el período, estando presentes en tres de los temas; y, segundo, que el quinto tema parece estar dedicado a los aspectos territoriales y lingüísticos de las leyes de educación.

Tabla 1. Análisis de la cobertura de El País (1978-2018)

Tema	Términos presentes	Términos ausentes
1	educación, ley, gobierno, alumnos, reforma, centros, pp	
2	educación, centros, alumnos, enseñanza	pp, gobierno
3	ley, gobierno, pp, proyecto	alumnos, años,
4	millones, pesetas, proyecto	educación
5	alumnos, pp, gobierno, castellano, cataluña , centros, curso	educación, ley, física

Nota: Los términos se han editado para mejorar la legibilidad

En cuanto a ABC, en la tabla 2, observamos un comportamiento parecido a El País, aunque con diferencias significativas, tal y como establecía la H.2. Las semejanzas se producen en la presencia de términos como «partido popular» o el tema lingüístico. Pero, en este caso, aparece el tema religioso, ausente de los principales elementos de la cobertura de El País, lo que no es sorprendente pero sí indicativo de diferentes formas de enfocar las reformas educativas desde presupuestos ideológicos y de valores opuestos. Lo mismo puede decirse del término «padres», sólo presente en ABC. Es decir, el algoritmo ha sido capaz de mostrar diferencias consistentes en la cobertura mediante los diferentes enfoques que cada medio otorga a la realidad informativa (sobre esos enfoques se puede ver Canel 1999).

Tabla 2. Análisis de la cobertura de ABC (1978-2018)

Tema	Términos presentes	Términos ausentes
1	educación, ley, enseñanza, reforma, gobierno, centros, alumnos, educativa	
2	gobierno, pp, españa	educación, enseñanza, centro, alumnos, formación
3	ley, educacion, gobierno, pp, catalán	reforma, formación
4	años, curso, alumnos	enseñanza, ley, proyecto, derecho, libertad, religión
5	padres , centros, alumnos, enseñanza, iglesia	reforma, educacion, proyecto, formación

Nota: Los términos se han editado para mejorar la legibilidad

Nota²: Términos seleccionados en *negrita*

Pero, ¿cuál es la relación entre la polarización y los temas tratados por los medios de comunicación? A este respecto, las tablas 3 y 4 muestran la matriz de correlaciones entre las variables de análisis (nótese que los temas varían en contenido para cada medio). La primera conclusión que se puede extraer es que las correlaciones son débiles en todos los casos falsando, en principio, la última hipótesis (H.3), ya que no se aprecia un aumento o descenso significativo entre polarización y partido en el gobierno.

En concreto, hay dos temas que sobresalen en la matriz: Cataluña (tema 3) y la religión (tema 5). En los dos casos observamos comportamientos divergentes en la cobertura mediática. Es decir, los resultados del análisis siempre son inversos en todos los temas, aunque no sean demasiado significativos en algunos casos. Esto resulta interesante porque, en principio, se esperaba que el comportamiento con respecto a los temas, aunque no fuera exactamente igual (H.2), sí que estuviera tratado en ambos medios de comunicación, especialmente porque la iniciativa en la exposición de los temas debería ser gubernamental o parlamentaria. Sin embargo, los resultados parecen apoyar la idea de que los temas no dependen del

⁶⁸ Nota aclaratoria de las Tabla 1 y 2. Los términos seleccionados se incluyen con los estándares de limpieza, homogeneización y tokenización propios de este tipo de análisis. Por ello, se pueden encontrar tokens sin acento o aquellos otros que deberían de ir con la mayúscula inicial en minúscula.

nivel político, sino que estarían más relacionados con la agenda mediática. Y esto puede suceder de dos formas diferentes, desde nuestra perspectiva: por un lado, siguiendo los postulados de los teóricos de los estudios de la construcción de la agenda (el estudio más popular, clásico y descriptivo de la teoría puede ser el de Cobb et al. 1976), se podría considerar que los medios de comunicación centran el debate en los temas que consideran más interesantes o propicios (es decir, muestran los temas usando los encuadres que estimen); por otro lado, otros académicos han afirmado que los medios de comunicación se encuentran indexados al nivel político, reflejando el nivel de desacuerdo entre los grupos y, habitualmente, tomando partido por un lado u otro (Bennett 2016). La diferencia entre ambos reside en el origen de la información. Mientras que, en el primer caso, los medios de comunicación tienen la capacidad y la voluntad de afectar al debate político proponiendo temas o encuadres específicos; en el segundo, se limitan a reflejar lo que los actores institucionales exponen. Los datos que mostramos aquí podrían apoyar ambos acercamientos y sólo un detallado estudio cronológico ayudaría a dilucidar la cuestión.

Tabla 3. Correlaciones El País -Rho de Spearman, términos seleccionados-

Part_Gob	Polarización	Tema 1	Tema 2	Tema 3	Tema 4	Tema 5
PP	-0.025	.054**	-0.006	-.196**	.107**	.201**
PSOE	0.01	-.056**	0.003	.195**	-.127**	-.152**
UCD	.050	0.005	0.008	0.001	.065**	-.164**

***. La correlación es significativa en el nivel 0.01 (bilateral).*

**. La correlación es significativa en el nivel 0.05 (bilateral).*

Tabla 4. Correlaciones ABC -Rho de Spearman, términos seleccionados-

Part_Gob	Polarización	Tema 1	Tema 2	Tema 3	Tema 4	Tema 5
PP	-.055**	-.074**	.138**	-.128**	.118**	-.063**
PSOE	.051*	.077**	-.148**	.148**	-.123**	.056**
UCD	0.009	-0.015	.044	-.074**	0.025	0.018

***. La correlación es significativa en el nivel 0.01 (bilateral).*

**. La correlación es significativa en el nivel 0.05 (bilateral).*

Sea cual sea el motivo de las diferencias de cobertura entre los dos medios de comunicación (que, en todo caso, es débil), su mera existencia sí que podría confirmar que España es un caso de pluralismo polarizado, por su visible nivel de paralelismo político. Sin embargo, no sería a través de la polarización de la cobertura con respecto a los temas, sino tan sólo con respecto a la exposición de los mismos (a pesar de que el gráfico 2 sugiera que sí hay diferencias en la polarización).

V. Discusión y resultados

La meta principal de este estudio era verificar si las tesis de Hallin y Mancini (2004) se cumplen en el sistema mediático español, testando una propuesta metodológica novedosa en el campo de la comunicación política. Para ello se seleccionó un tema que tuviera un cariz polarizante dentro de la sociedad; la educación. Para focalizar esta temática decidimos centrar nuestra investigación en el empleo de las partículas semánticas «reforma educativa» y «ley de educación» a lo largo de los primeros cuarenta años de democracia en España. De este modo, tras implementar el diseño de investigación detallado, podemos afirmar que los resultados confirman en general las dos primeras hipótesis de partida, aunque no la tercera.

En primer lugar, las observaciones arrojan una relación entre las líneas ideológicas de los medios y su posicionamiento con respecto a las reformas educativas tramitadas por parte de los partidos en el poder, tal y como se aprecia de forma muy clara en las figuras de agregados por artículos. Se establece con ello una clara interrelación entre el número de artículos, que incluyen cualquiera de los topics de búsqueda, y el partido de gobierno. Consecuentemente, se evidencia que la cantidad de artículos es mayor en función de si la línea editorial del periódico se distancia de los preceptos ideológicos del partido que pretende promulgar la ley.

En segundo lugar, y con las evidencias expuestas, se procedió a completar esta investigación con un modelo de aprendizaje de máquina automatizado, LSI (*Latent Semantic Indexing*). Este proceso nos permitió ver cómo era la cobertura y en qué temáticas solían centrarse, y divergir los dos periódicos a la hora de hablar de educación. Debido a la magnitud de la muestra, hay temas que sobresalen, por comunes y sintomáticos, a la cuestión de estudio, y hay otros significantes por los que se hallan diferencias entre medios. Por ejemplo, en el caso de El País, las cuestiones «Cataluña» y «castellano» suponen una vertiente de gran interés para el medio, mientras que, en el caso de ABC, lo que se encontraría es una prenoticia por «catalán», así como por el de «iglesia», o la importancia que se suscita entre el tema educativo y «padres». Esto muestra las distancias que se generan a la hora de abordar los temas por los medios y que, en el caso analizado, abre discrepancias en cuanto a los intereses que llevan a uno y a otro a hablar de educación.

Por último, las técnicas se han ido complementando con el fin de mostrar los resultados y, a su vez, servir de campo de pruebas para futuros estudios. En este sentido, se ha buscado desarrollar nuevas líneas de actuación que sirvan para medir la polarización dentro de distintos escenarios de opinión pública, desde medios de comunicación, hasta redes sociales. El índice empleado ha mostrado un gran rendimiento (véase Serrano et al. (2020) como ejemplo), siendo capaz de captar de un modo más complejo, la frecuente simplicidad semántica que arrojaran las valencias utilizadas para el análisis de sentimiento.

A pesar de ello, todavía existen dificultades para detectar un vocabulario más discrepante, como puede ser el empleado en las redes sociales. Este hecho,

propicia que el equilibrio se acabe moderando debido a los subconjuntos semánticos que componen un texto periodístico. Es más, la polarización mostrada por la herramienta no incluye la dirección de la misma; es decir, los dos medios de comunicación polarizan en ciertos momentos su cobertura, especialmente con respecto a la línea ideológica del gobierno (lo que se aprecia en el gráfico 2), pero no lo suficiente como para detectar un patrón (de ahí la falta de significación de las tablas 3 y 4). Y, seguramente, si se analiza hacia dónde va dirigida esa polarización (el gobierno o la oposición) sí se encuentre ese último eslabón que aquí sólo se intuye, por lo que se debe virar hacia técnicas de localización más sofisticadas (a saber, Ghatak 2019; Hodnett et al. 2019; Rezaeinia et al. 2019).

Hasta el momento, ha parecido inviable un pacto entre los partidos políticos que se alternaron en el poder, PSOE y PP, como entre los bloques ideológicos que han dividido el Parlamento español en las últimas décadas. El calado de las reformas en algunos aspectos no distan mucho, pero existen dos elementos que serán clave a la hora de polarizar a los actores políticos y sociales: la iglesia católica y la privatización del sector educativo (Viñao 2006: 59). Y esto, en un sistema multinivel como el español, genera un efecto multiplicador en las pulsiones polarizantes, ya que cada reforma ha sido dificultada y desactivada hasta lo posible por las comunidades autónomas con gobiernos de ideología distinta al impulsor de la reforma a nivel central. España todavía carece de una tradición consociacional democrática, donde la educación no se considera una moneda de cambio y sí una piedra angular para el desarrollo del país. Parece que alcanzar un pacto de estado en esta materia se interpreta como asumir unos costes demasiado altos sin ningún -o pocos- beneficio, y que hacerlo supondría ignorar las preferencias de sus votantes. Pactar en política educativa equivaldría, pues, a entregarse imperdonablemente al enemigo (Gortázar y Moreno 2017).

Esta investigación abre nuevas vías y enfoques para el avance dentro de la comunicación política, y concretamente para el análisis de las influencias que el sector político ejerce en los distintos ámbitos de la esfera pública. Con este trabajo se pretende ofrecer un primer paso para el contraste evolutivo de esta temática, que se vuelve a mostrar en constante vigencia hoy día, especialmente vinculado a la transformación que ha sufrido el arco parlamentario hacia la fragmentación. Este cambio, junto con el primer ensayo de gobierno de coalición en España en 2020, muestra una nueva veta de investigación que podría alterar los patrones mostrados en el presente estudio. Por lo tanto, observando los movimientos de los grupos parlamentarios tras la formación del presente gobierno, en relación con la reforma de la ley educativa vigente, podríamos plantearnos si los paradigmas contrastados en este proyecto seguirán siendo la tónica o, si por el contrario, se producirán pactos de gran calado entre las más que complejas amalgamas ideológicas que, a su vez, se proyecten a la esfera mediática.

Finalmente, nos gustaría señalar que el empleo creciente de nuevas técnicas de investigación automatizadas es un hecho. Si bien el desarrollo y la proliferación de su uso se ha producido inicialmente en otras disciplinas más técnicas, en los

últimos años hemos observado que investigaciones propuestas desde las ciencias sociales están incluyendo estas herramientas en su diseño. Asimismo, cabe destacar que este proceso está impulsado tanto por la facilidad del acceso a datos estructurados (hasta ser prácticamente ilimitados), como la generación de procesos para estandarizar la información para el análisis.

Referencias

- Arriagada, Arturo, Patricio Navia y Martín Schuster. 2010. “¿Consumo Luego Pienso, O Pienso Y Luego Consumo?: Consumo de Medios, Predisposición Política, Percepción Económica Y Aprobación Presidencial en Chile.” *Revista de Ciencia Política* 30(3): 669-695.
- Aruguete, Natalia. 2013. “La Narración Del Espectáculo Político: Pensar la Relación Entre Sistema de Medios Y Poder Político.” *Austral Comunicación* 2(2): 205-216.
- Aruguete, Natalia y Ernesto Calvo. 2018. “Time to #Protest: Selective Exposure, Cascading Activation, and Framing in Social Media.” *Journal of Communication* 68(3): 480-502.
- Balch, Alex y Ekaterina Balabanova. 2011. “A System in Chaos? Knowledge and Sense-Making on Immigration Policy in Public Debates.” *Media, Culture & Society* 33(6): 885-904.
- Balkin, Jack M. 1999. “How Mass Media Simulate Political Transparency.” *Cultural Values* 3(4): 393-413.
- Bennett, W. Lance, Victor Pickard, David Iozzi, Carl Schroeder, Taso Lagos y C. Evans Caswe- ll. 2004. “Managing the Public Sphere: Journalistic Construction of the Great Globali- zation Debate.” *Journal of Communication* 54(3): 1437-455.
- Bennett, W. Lance. 2016. “Indexing Theory.” En *The International Encyclopedia of Political Communication*, editado por Gianpietro Mazzoleni. Hoboken: Wiley.
- Benson, Rodney. 2010. “What Makes for a Critical Press? A Case Study of French and U.S. Immigration News Coverage.” *The International Journal of Press/Politics* 15(1): 3-24.
- Bernhardt, Dan, Stefan Krasa y Mattias Polborn. 2008. “Political Polarization and the Electoral Effects of Media Bias.” *Journal of Public Economics* 92(5-6): 1092-1104.
- Boxell, Levi, Matthew Gentzkow y Jess Shapiro. 2020. “Cross-Country Trends in Affective Polarization.” Working Paper No. 26669. National Bureau of Economic Research.
- Bramson, Aaron, Patrick Grim, Daniel Singer, William Berger, Graham Sack, Steven Fisher, Carissa Flocken y Bennett Holman. 2017. “Understanding Polarization: Meanings, Measures, and Model Evaluation.” *Philosophy of Science* 84(1): 115-159.
- Budge, Ian y Dennis Farlie. 1983. “Party Competition: Selective Emphasis or Direct Confrontation? An Alternative View With Data.” En *Western European Party Systems: Continuity & Change*, editado por Hans Daalder y Peter Mair. London: SAGE Publications, 267-305.
- Canel, María. 1999. “El País, ABC y El Mundo: Tres Manchetas, Tres Enfoques de Las Noticias.” *ZER - Revista de Estudios de Comunicación* 1(6):1-11.

- Cobb, Roger, Jennie-Keith Ross y Marc Howard. 1976. "Agenda Building as a Comparative Political Process." *American Political Science Review* 70(1): 126-138.
- Collins, Todd y Christopher Cooper. 2012. "Case Saliency and Media Coverage of Supreme Court Decisions: Toward a New Measure." *Political Research Quarterly* 65(2): 396-407.
- De Vries, Gerdien. 2020. "Public Communication as a Tool to Implement Environmental Policies." *Social Issues and Policy Review* 14(1): 244-272.
- DiMaggio, Paul, John Evans y Bethany Bryson. 1996. "Have American's Social Attitudes Become More Polarized?" *American Journal of Sociology* 102(3): 690-755.
- Entman, Robert. M. 1993. "Framing: Toward Clarification of a Fractured Paradigm." *Journal of Communication* 43(4): 51-58.
- Feinerer, Ingo, Kurt Hornik y David Meyer. 2008. "Text Mining Infrastructure in R." *Journal of Statistical Software* 25(5): 1-54.
- Fletcher, Richard y Joy Jenkins. 2019. *Polarisation and the News Media in Europe*. Bruselas: EPRS.
- García-Marín, Javier y Adolfo Calatrava. 2018. "The Use of Supervised Learning Algorithms in Political Communication and Media Studies: Locating Frames in the Press." *Comunicación y Sociedad* 31(3): 175-188.
- Gentzkow, Matthew. 2016. *Polarization in 2016*. Mimeo: Stanford University.
- Ghatak, Abhijit. 2019. *Deep Learning with R*. Singapur: Springer.
- Gil, Daniel, Carlos Furió y Valentín Gavidia. 1998. "El Profesorado y la Reforma Educativa en España." *Revista Investigación en la Escuela* (36): 49-64.
- González, Juan Jesús y Fermín Bouza. 2009. *Las Razones del Voto en la España Democrática (1977-2008)*. Madrid: Catarata.
- González, Juan Jesús, Raquel Rodríguez y Antón Rodríguez. 2010. "A Case of Polarized Pluralism in a Mediterranean Country. The Media and Politics in Spain." *Global Media Journal Mediterranean Edition*, 5 (1/2)
- Gortázar, Lucas y Juan Moreno. 2017. "Costes y Consecuencias de no Alcanzar un Pacto Educativo en España." *Revista Educación, Política y Sociedad* 2(2): 9-37.
- Hallin, Daniel y Paolo Mancini. 2004. *Comparing Media Systems Three Models of Media and Politics*. Nueva York: Cambridge University Press.
- Helbling, Marc y Anke Tresch. 2011. "Measuring Party Positions and Issue Saliency From Media Coverage: Discussing and Cross-Validating New Indicators." *Electoral Studies* 30(1): 174-183.
- Hodnett, Mark, Joshua F. Wiley, Yuxi Lui y Pablo Maldonado. 2019. *Deep Learning with R for Beginners*. Birmingham: Packt.

- Hoffman, Thomas. 1999. "Probabilistic Latent Semantic Analysis." Ponencia presentada en *Proceedings of Uncertainty in Artificial Intelligence*, Estocolmo, 30 de julio a 1 de agosto.
- Humanes, María L. 2014. "Exposición Selectiva y Partidismo de las Audiencias en España. El Consumo de Información Política Durante las Campañas Electorales de 2008 y 2011." *Palabra Clave* 17(3): 773–802.
- Humanes, María L., María D. Montero, Ramón Molina y Alfredo López-Berini. 2013. "Pluralismo y Paralelismo Político en la Información Televisiva en España." *Revista Latina de Comunicación Social* 68(5): 24-16.
- Kenamer, David. 1994. *Public Opinion, the Press, and Public Policy*. London: Praeger.
- Letsche, Todd y Michael W. Berry. 1997. "Large-Scale Information Retrieval with Latent Semantic Indexing." *Information sciences* 100(1-4): 105-137.
- Lodola, Germán y Philip Kitzberger. 2017. "Politización y Confianza en los Medios de Comunicación: Argentina Durante el Kirchnerismo." *Revista de Ciencia Política* 37(3): 635- 658.
- Lowry, Dennis y Jon A. Shidler. 1998. "The Sound Bites, the Biters, and the Bitten: A Two-Campaign Test of the Anti-Incumbent Bias Hypothesis in Network TV News." *Journalism & Mass Communication Quarterly* 75(4): 719–729.
- Mancini, Paolo, Marco Mazzoni, Alessio Cornia y Rita Marchetti. 2017. "Representations of Corruption in the British, French, and Italian Press: Audience Segmentation and the Lack of Unanimously Shared Indignation." *The International Journal of Press/Politics* 22(1): 67–91.
- Martínez, Juan B. 2017. "¿Y Ahora Qué? Impacto de la Disolución de la UCD en El Mapa Político Local de las Provincias de Albacete y Murcia." En *Los Embates de la Modernidad: Debates en Torno a la Ciudadanía, El Liberalismo, El Republicanismo, la Democracia y Los Movimientos Sociales*, editado por Oriol Luján y Laura Canalías Chorrero. Barcelona: DHMC, 563-579.
- Martínez, Manuel, María L. Humanes y Enric Saperas. 2014. "La Mediatización de la Política en El Periodismo Español. Análisis Longitudinal de la Información Política en la Prensa de Referencia (1980-2010)." *Tripodos* (34): 41-59.
- Mason, Lilliana. 2014. "'I Disrespectfully Agree': The Differential Effects of Partisan Sorting on Social and Issue Polarization." *American Journal of Political Science* 59(1): 128–145.
- McCluskey, Michael y Young Kim. 2012. "Moderatism or Polarization? Representation of Advocacy Groups' Ideology in Newspapers." *Journalism & Mass Communication Quarterly* 89(4): 565–584.
- Miller, Mark. 1997. "Frame Mapping and Analysis of News Coverage of Contentious Issues." *Social Science Computer Review* 15(4): 367-378.

- Monarca, Héctor, Cecilia Simón, Soledad Rappoport y Gerardo Echeita. 2016. "Política y Cambio en Educación: El Caso de las Competencias Básicas en España." *Ensaio: Avaliação e Políticas Públicas em Educação* 24(93): 968-989.
- Olsson, Erik J. 2013. "A Bayesian Simulation Model of Group Deliberation and Polarization." En *Bayesian Argumentation Synthese Library (Studies in Epistemology, Logic, Methodology, and Philosophy of Science)* vol 362, editado por Frank Zenker. New York: Springer Publishing, 113-133.
- Poole, Keith. 2008. "Las Raíces de la Polarización de la Política Moderna en los Estados Unidos." *Revista de Ciencia Política* 28(2): 3-37.
- Prior, Markus. 2013. "Media and Political Polarization." *Annual Review of Political Science* 16(1): 101-127.
- Rezaeinia, Seyes M., Rouhollah Rahmani, Ali Ghodsi y Hadi Veisi. 2019. "Sentiment Analysis Based on Improved Pre-Trained Word Embeddings." *Expert Systems with Applications* 117: 139-147.
- Rinker, Tyler. 2019. "Trinker/Sentimentr." Recuperado el 23 de agosto de 2019 de <https://bit.ly/2VBwWBK>
- Runte, Ariadne 2016. "Pisa en la Prensa Española y su Influencia sobre las Políticas Educativas." *Opción* 32(8): 713-733.
- Serrano-Contreras, Ignacio-Jesús, Javier García-Marín y Óscar Luengo. 2020. "Measuring Online Political Dialogue: Does Polarization Trigger More Deliberation?" *Media and Communication* 8(4): 63-72.
- Sunstein, Cass. 2002. "The Law of Group Polarization." *Journal of Political Philosophy* 10(2): 175-195.
- Tucker, Joshua, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal y Brendan Nyhan. 2018. *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*. Menlo Park: Hewlett Foundation.
- Viñao, Antonio. 2006. "El Éxito O Fracaso de las Reformas Educativas: Condicionantes, Limitaciones, Posibilidades." En *La Reforma Necesaria: Entre la Política Educativa y la Práctica Escolar*, editado por José Gimeno Sacristán. Madrid: Ediciones Morata, 43-60.

Serrano-Contreras I-J. (2021) NLP on YouTube: A Look on Feminism. En M. Musiał-Karg, O. G., Luengo (Eds), *Digitalization of Democratic Processes in Europe. Studies in Digital Politics and Governance*. Springer. https://doi.org/10.1007/978-3-030-71815-2_10

NLP on YouTube: A Look on Feminism

Abstract

Feminism is a complex phenomenon that has evolved in recent times. We can see it through different optics such as social media or traditional media. The aim of this project is to analyze, descriptively, how it has grown on a platform like YouTube. We are using new computer techniques and trying to understand how the concept evolves. Hence, in this paper, we will provide a language analysis that examines deliberative frameworks and online roles. Our study focuses on adding new methods that offer insights into the Digital Agora. The main outcomes show that the deliberation on YouTube, with regard to Spanish Feminism, has received an important and enormous increase especially since 2016, revealing two possible souls within that debate.

Keywords: Feminism, YouTube, Social media, Natural Language Processing, Machine Learning

1. Introduction

Activism can be understood as struggles to bring about change, achieve rights and freedoms and lead to a new sphere in human relations. Thus, movements like “#Metoo,” International Women’s Day or “Time’s Up” are part of the new feminism wave where public presence is essential. These processes advocate taking over squares and streets by people and showing their power and using it. Events like that serve to assess the impact of these types of phenomena, but also we can look at how they can be managed by traditional media. Although mass media permitted to achieve new ways to gain weight in political control (Curran, 2005), social media was a mode to break the distance from society. It can be staged, for example, “The Arab Spring,” “15M” in Spain or “OccupyWall Street.” Certainly, the displacement of traditional media toward peer-to-peer connections allows to break the chains and archetypes of the post-industrial era (Bennett et al., 2014), reaching a new arena where the online plays a key role in the writing of our present. The same happens with Feminism where rights, which were the first steps (FirstWave of Feminism and SecondWave), marked a new path to develop new stages such as Women and Gender Studies or to use public space to transform small stories and rewrite human history, thanks to the power of the media (Haughton, 2019).

Our mission will be to take further steps in understanding social media, especially YouTube inasmuch as it is one of the most important online power aggregators, as data shows. The last point is very important in our goal, because this network, owned by Google, has some peculiarities with regard to Twitter or

Facebook. We are talking about the world's largest platform for video content, which relies on data and algorithmic models developed by the parent company. These systems utilize personal characteristics and behaviors to provide individualized content for each user. This basic idea leads to huge programming models based on deep learning, where neural networks establish complex methods with hidden layers built on the basis of the Markov Model (Graves et al., 2013). One more step within the automatic learning happens, where the “number of doors” is multiplied and the answers are improved (Davidson et al., 2010; Covington et al., 2016).

Currently, efforts are directed at learning how these invisible patterns direct our context, opinion and approach. For this reason, we must refer to the concepts that try to explain how these topics are elaborated: “filter bubble” and “echo chamber.” The definition of a filter bubble was given by Eli Pariser (2011). He talks about the indirect effects that online users ignore, such as the so-called Google wall. Meanwhile, the echo chamber has been developed by different researchers. They manifest that the disposition to reinforce our opinion is fundamental (Klapper, 1960; Sunstein, 2007; Knobloch-Westerwick, 2015; Rubin, 2002). According to this, proximity will be essential to build an environment guided by reliability scales (Granovetter, 1973; Scruggs, 1998) like an echo chamber where we can only hear our voice. There we will follow a mainstream leader or share certain information because we are ready to stand up and mobilize for some previous idea of our own (Petersen, 2020).

Our review aims to provide a process that allows us to check on the web the changes that involve all people now and, mainly, the discussion that occurs concerning feminism on YouTube. This paper is organized following different computer models that allow us to know how phenomena evolve, and what causes them, through the behavior of the users. To achieve speech recognition, it will be essential to make a transcription of the languages, where the computer can understand what humans say. Thus, we will have to find patterns within the corpus that will enable us to construct numerical matrices. This type of technique is framed in the studies of Natural Language Processing (NLP).

2. Theoretical Framework

Natural Language Processing

In recent years there has been an interesting increase in the understanding of languages through the machine. Although computer studies began decades ago, we think the most important approach was given by Alan Turing. The work done by the British mathematician in decoding Enigma marks the beginning of the road. This advance was made more interesting by the fact that we live in the Data Age which, together with the development of these techniques, generates many aspects that serve to define our current society.

For example, when we talk about BigData, we need to understand that the amount of electronic information is enormous throughout the world in the course of a single day. Algorithmic tasks operate to transform binary operations (0 and 1) into rational data, where human codes are decoded. We must translate the unstructured information into a set of data frames to be structured and reach a KnowledgeDiscovery in the Texts (KDT) (Justicia de la Torre et al., 2018), this is essential for producing quantitative data from qualitative information.

The text mining defined previously can be related to several areas depending on our interest. As Justice et al. 2018 point out, we could mention Web mining, Streaming text data mining until Opinion mining and sentiment analysis. They all point to different applications for managing human-machine understanding that are being studied as exercises within NLP. The processes seek to implement the trained knowledge to achieve new learning methods, find hidden patterns, generate new texts, vector words, dictionaries, main topics...

These contributions have provided interesting ideas for the development of a more complex understanding of human communication since the classification of hierarchically labeled text (Mao et al., 2019); the use of Maximum Entropy (Cindo et al., 2020) or with Support Vector Machine (García-Marín et al., 2018) to sort social media or news; or unsupervised techniques to find clusters in the text corpus (Filho et al., 2014).

3. Methodology

Our main assumption is that the topic of feminism has evolved:

- H1. Feminism movement has imposed a new axis into online discussion

Data

YouTube contents from the 100 most popular videos about feminism in Spanish were sampled using the keyword: “feminismo.” Our data collect was made up of a total of 690,138 comments within a period (2012–2019). A brief overview can be seen in Figs. 1 and 2. Afterward, we followed the generic steps: Pre-processing and cleaning to detect anomalies, Keyword removal, and Pattern analysis (Talib et al., 2016).

Studies Models

The main challenge is to develop some models to show how users behave, and to be able to structure works on framing or polarization in media. Therefore, we will employ different processes to achieve our mission. The first step will be to explain the extension of the R language, because we can acquire a complex structure to work with; from local regression to deep learning. For example, in the TM–Text Mining–library (Feinerer et al., 2019), we can find many functions to structure our text, with which to perform extensive and complex text mining. In our research,

the evolution of the phenomenon of feminism will be measured by year. The use of correlation, in this case Pearson, will be essential to manage our

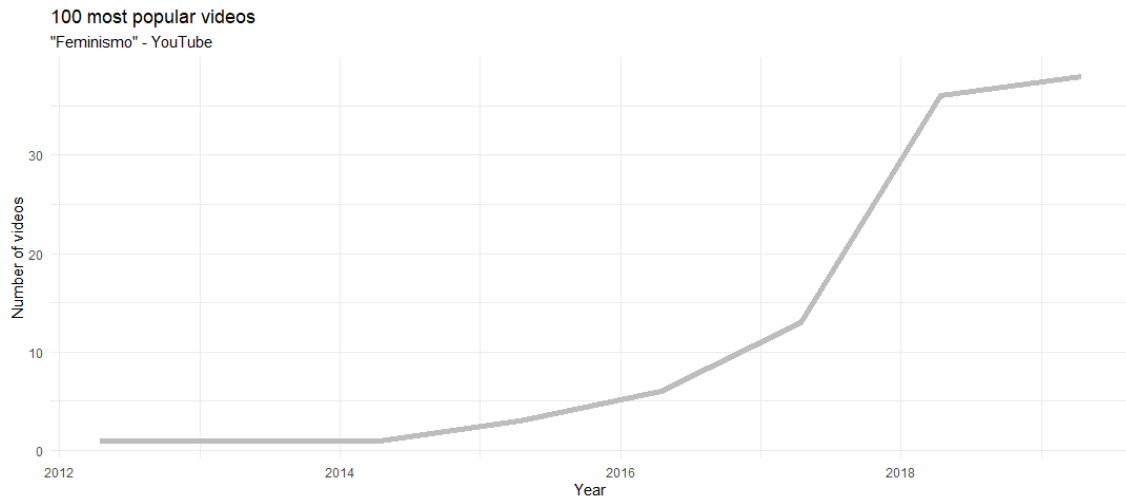


Fig. 1 Videos published by year (Source Own elaboration)

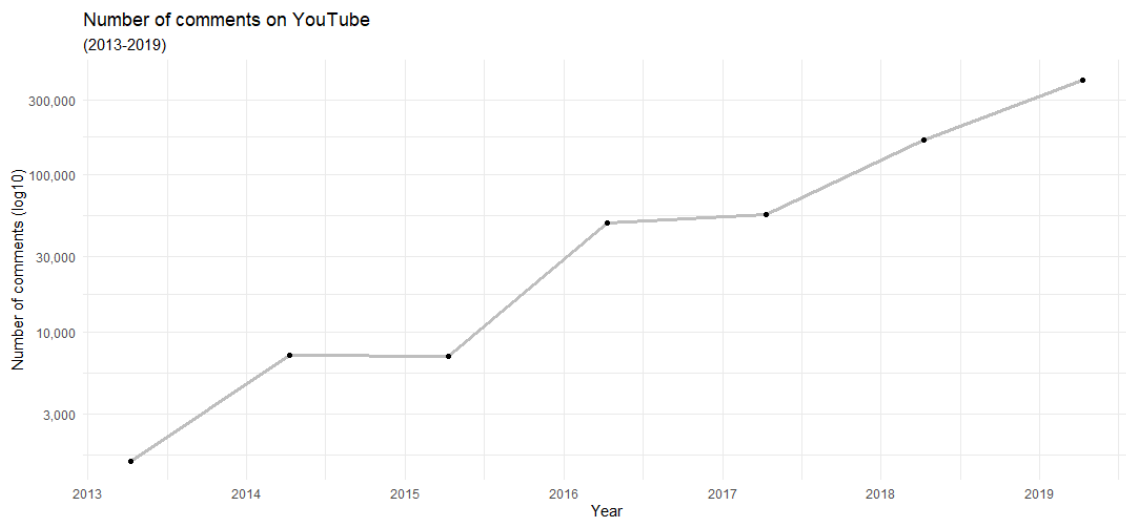


Fig. 2 Comments by year (log10) (Source Own elaboration)

understanding of the subject. Such an idea is shown within a matrix correlation where the structure will clearly draw the meanings between the years. Hence, it could offer enough information about the use of other algorithmic models, or generate a future projection in an inferential way.

The next step will be based on a semantic explanation. The theory of N-grams explore around the data set of links: two words, three words, four words... If follows the Markov chain—first about deep learning—by which through a discrete variable, the probability of the events occurs only as a function of the last event that happened. For example, when we talk about words and we want to know the pairs of “feminism,” we only need to know the previous fact:

New York (is a 2-gram)

Federico Garcia Lorca (is a 3-gram)

Once upon a time (is a 4-gram)

Now we can increase the advantage found through N-gram theory. We use probabilities to tailor specific linked words within our data set. These interactions are done by means of the Term reverse Document Frequency-Tf-idf-. With this, we can examine what the weight of the terms is and whether it yields relevant information (Spärck, 1972). According to the established steps, we must first calculate the term frequency and then check its importance as a function of the number of documents in the data set.

Once a pair theory is established we could enter another interesting model like the Phi coefficient. It's a common measure for the binary correlation between items (Silge & Robinson, 2017). Our structure, built by the WidyR package (Robinson et al., 2020), can also implement distance models (by Euclidean measure or cosine) between words or a Singular Value Decomposition where we reduce the information and we can find a low entropy with few highly relevant data.

Finally, and following the advances into machine learning, we have the option for a train process. In that case we opt for the Latent Dirichlet Allocation –LDA– through which we implement a TopicModel to find the main topics of our collection. The most important thing here will be to discover whether the terms–based on two main topics–can produce semantic distance between users–which could be defined as polarization–or whether context is very important in understanding such changes. This type of generative algorithm assumes that the presence or absence of terms causes a sampling of probabilities within the probabilities of obtaining elements that belong to certain topics (e.g., Blei et al., 2003).

4. Results

The main implication here is to show the most relevant data about “feminism” on YouTube. We use the comments from that platform to extract semantic information that could be consequential to the analysis of feminism online (Figs. 3 and 4).

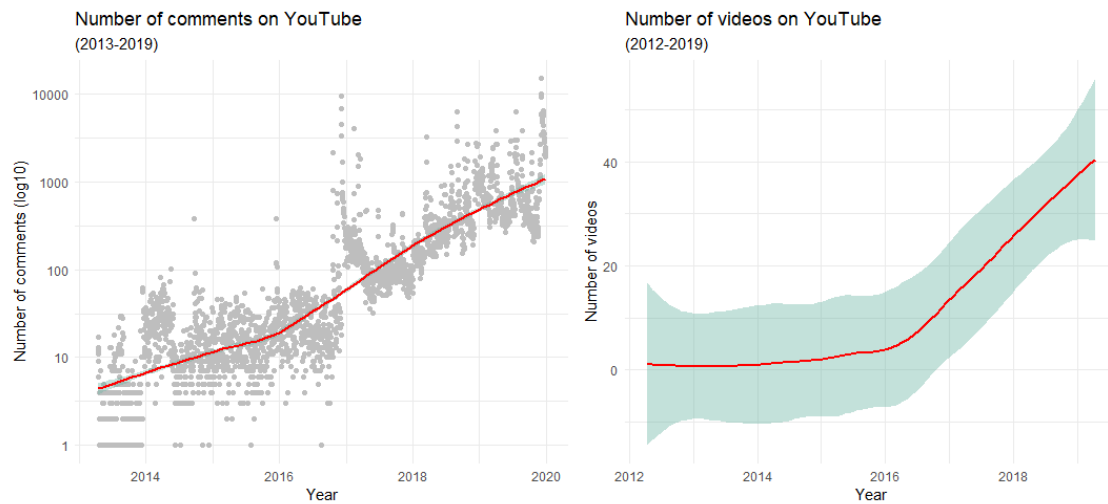


Fig. 3 Comparatives LOESS (Comments [log10] and videos) (Source Own elaboration)

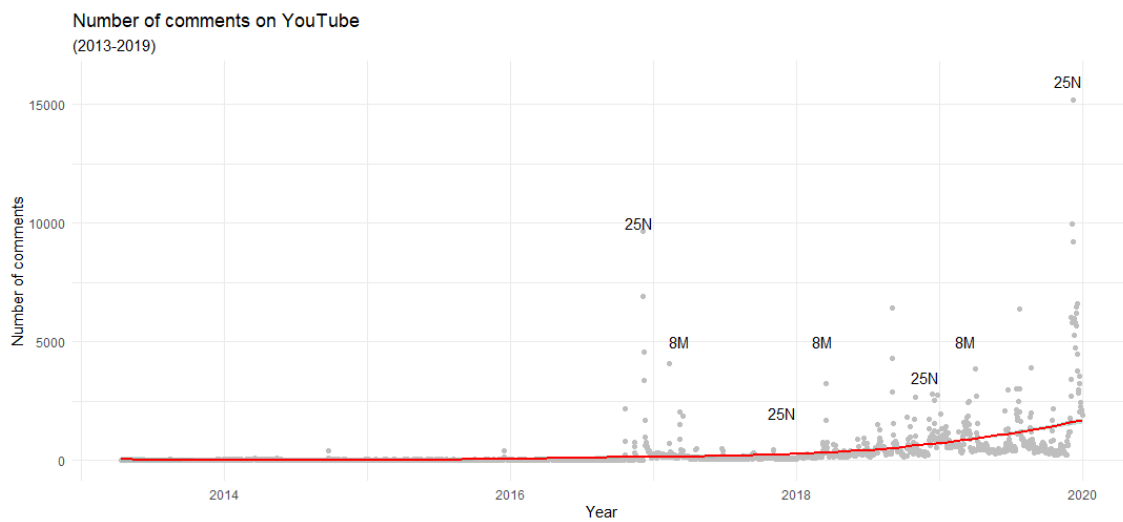


Fig. 4 LOESS and significant dates (log10) (Source Own elaboration)

Firstly, the figures have revealed similar behavior both in the video and in the comments. The parameters show an increase in the use of YouTube in addition to our main topic of interest. The changes start to occur around 2016, when the curve starts to rise. The first significant effect is the date of November 25th–International Day of Abused Women–which receives many comments. The same happens on the next 25th November, and other dates such as 8th March–International Women’s Day–but only within a general upward curve.

Now, if we start to focus on the data produced by the text mining and NLP we can find specific factors that call our attention. The latest figures show a breakdown around 2016. Figure 5 also shows us a different use of language; mainly from English to Spanish.

In a gradient between 0 and 1, depending on the absence or presence of words; the results mark a clear semantic difference. While the first years with comments

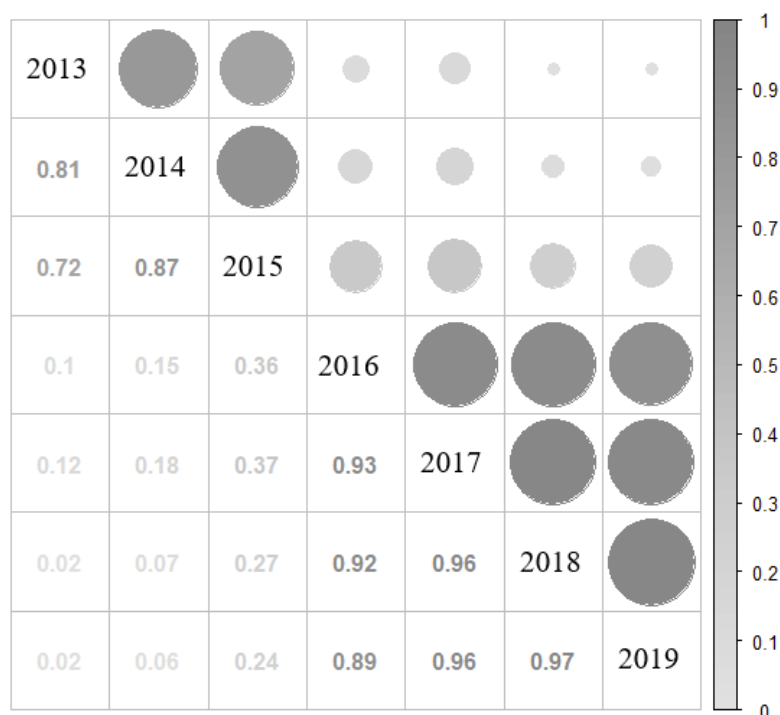


Fig. 5 Correlation matrix of language by year (Source Own elaboration)

(2013–2015) have similar behaviors, the next ones open a new period establishing 2016 as the year of change. For example, if we monitor the terms, we find that the most important reflection is based on the prevalence of English during 2013 to 2015. Later, we can see that the use of Spanish prevails from 2016 to 2019. Therefore, the correlation matrix and the 10 most important words per year in Fig. 6 are marked by the use of different languages.

Additionally, in Tables 1, 2, and 3, we show the N-grams rankings. There are three tables, one with Top 10 N-grams and two subsets parameterized with the topic “feminismo.” The complete Top 10 is not of significant interest because it does not show the relevant bigrams. The others allow us to know the descriptors used to refer to feminism, being able to find a division between two wings of feminism: Liberal

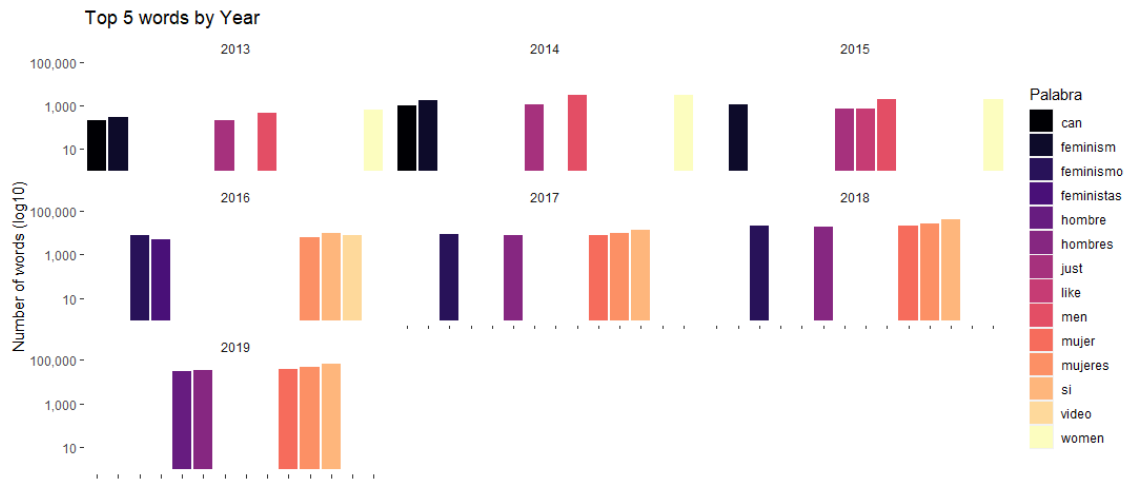


Fig. 6 Top 5 words by year (Source Own elaboration)

Table 1 N-grams (Top10)

2 n-grams	agustin laje	ser mujer	cada vez	puede ser	buen video
n	4241	3384	3260	3101	2562
2-grams	ser humano	ser feminista	brecha salarial	ver si	hace falta
n	2454	2240	2139	2067	1997

Source Own elaboration

Table 2 N-grams “feminismo” and “...” (Top 10)

2 n-grams	feminismo radical	feminismo actual	feminismo moderno	feminismo busca	feminismo si
n	1932	1866	1753	1111	681
2-grams	feminismo hegemonico	feminismo real	feminismo liberal	feminismo sino	feminismo lucha
n	384	317	260	237	235

Source Own elaboration

Table 2 N-grams “...” and “feminismo” (Top 10)

2 n-grams	verdadero feminismo	llama feminismo	palabra feminismo	unico feminismo	termino feminismo
n	658	331	263	179	157
2-grams	nuevo feminismo	falso feminismo	llamado feminismo	mismo feminismo	machismo feminismo
n	122	110	108	90	83

Source Own elaboration

and Radical. Although we could increase the N-grams (3, 4..) and even find legible sentences, the sample form and size would not be very significant. When working with data obtained from social media there are problems due to the fact that many comments have misspellings and include rare words and semantic conjunctions that are not understandable. These characteristics lead to a reduced ability to find relevant outliers, as a lot of information that is hidden is being lost. Although the

sample may be purer, the results would not be clearer due to the size of the data set.

Concluding, we exposed the results by Pairs about “feminismo,” Fig. 7, and the main topics with LDA, Fig. 8. The main conclusion is that both get similar outcomes. The average number of pairs of words or terms inside topics demonstrate the same trends. While the Fig. 4 shows qualifiers: “igualdad,” “radical,” “moderno” or “actual,” the outcome by LDA signs actions, verbs: “ser” or “hacer.”

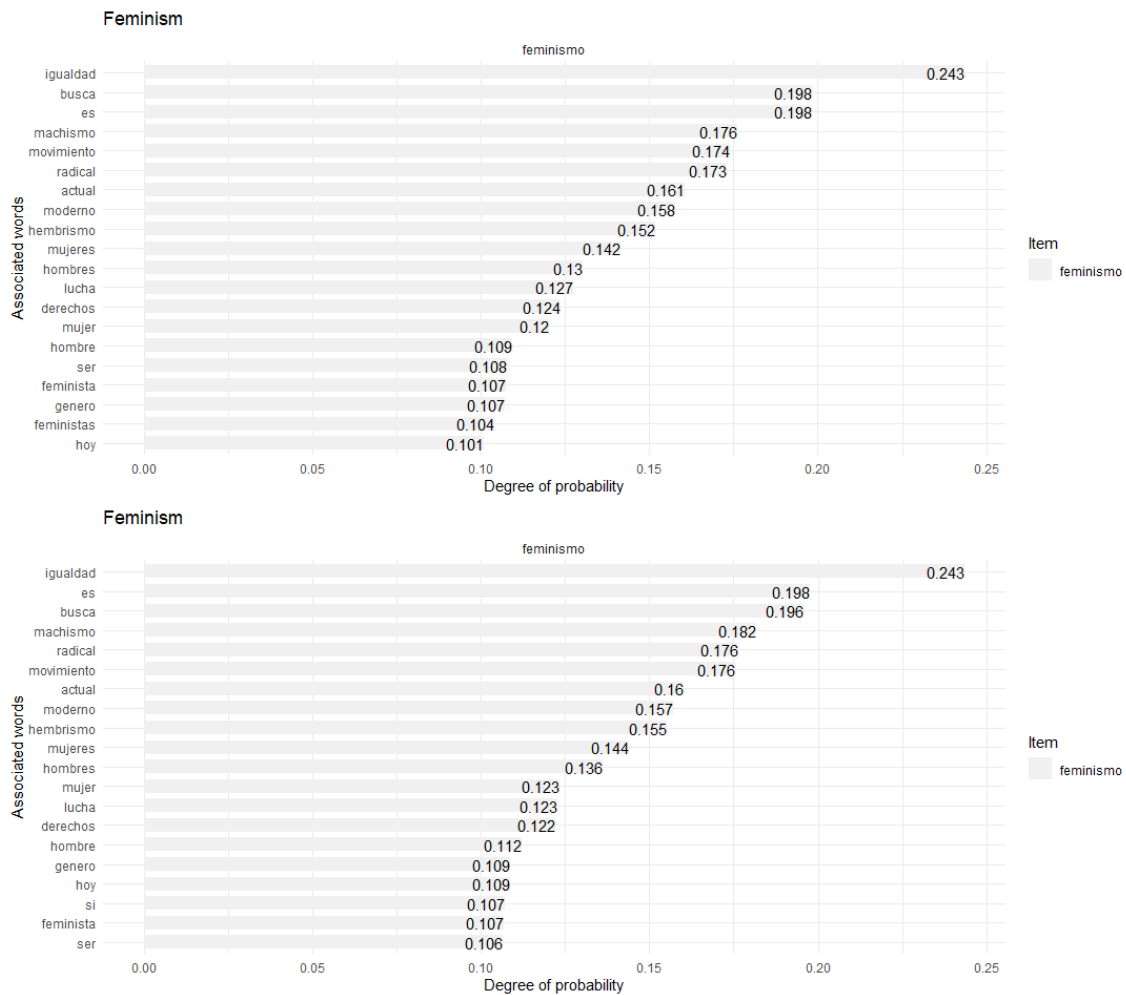


Fig. 7 Phi by (r) Pearson (Source Own elaboration)



Fig. 8 LDA-Topic modeling (1-2) (Source Own elaboration)

5. Conclusions

The mission of this project was to develop different approaches and models for analyzing feminism in Spanish on YouTube. The results reveal that the content of feminism on Google's video platform changed over time. The patterns and frames show that in recent years the subject of feminism has increased in popularity. These structures reflect how the activity on language produced a jump from English to Spanish. This point is very interesting because the first year of discussion within YouTube was marked by English, a language different from the video and the topic. This kind of evolution can be seen when looking at the correlation matrix, since it reflects exactly the expressions, words and references employed by users when they discuss online.

Continuing with the idea about filter bubble and the echo chamber, we are going to need complex models that try to measure these effects. However, it is significant that the increase in deliberation is related to dates such as November 25th or March 8th, which could approach the studies on agenda setting. In order to talk about agenda setting within the new mass phenomena, we must advocate for the inclusion of processes that allow us to reach a deep and complex learning in which the media is understood as an elaborated phenomenon that could be monitored by influencers, political parties or our online behavior. Looking at the model presented by YouTube programmers (Davidson et al., 2010; Covington et al., 2016), the complexities of achieving our timeline include too many undecipherable layers for human processing. However, traditional media still has too much power to guide the audience. We cannot forget that we select the reinforcement to feed our own opinion (Katz & Lazarsfeld, 1955).

However, the models employed are leading us in the right direction. Especially the rules of associations used as N-grams or the Phi coefficient of pairs by Pearson, which serve to show that the models are capturing the essence of the claim movement with arguments such as "lucha", "brecha salarial" or "hace falta." We now have a good descriptive framework. While the N-grams did not show relevant meanings, the value association pairs by semantic structure presented some ideas about feminism. Mainly, YouTube users link feminism with: "igualdad," but also "movimiento," "derechos," being positive words close ideas about

Feminism. Yet, it is also evident that we can also find opposite words like “machismo,” “hembrismo” or “radical” apparently making a crack between two sides. Similarly, the process of topic modeling with LDA would not offer different information about where to find an exact frame or distance between topics. Notwithstanding, with these data we can show a progression and validation of the models used (see, specially, Figs. 3–5 and 7).

We argue that the feminism in Spanish on YouTube establishes two main forms: Liberal close to First Wave Feminism and Radical marked by the Second and Third Wave. Moreover, it is evident, with the above data, that the Fourth wave of feminism is growing on the Internet and is endorsed by the takeover of the streets. Although we must point out that both indicate that the main objective is equality, they point to different paths so the terms and context will be different (Johnson, 2017). Perhaps, if a distance measurement is included with Long Short-Term Memory–LSTM–to embed the words within the vectors, it could reflect a possible translation (Søgaard et al., 2019), as well as an evolution in terminology and rhetorics.

This paper focuses on programming ability, especially NLP, revealing that it can measure semantic behaviors very well. Perhaps the next step is to add in-depth text learning or focus on video and audio decoding, although it is complicated to show exactly one transcript of human communication made by a computer. Finally, it is important to note that the results indicate that feminism takes to the streets but the wave started online as well (Lindstedt, 2019). It will also be interesting to monitor on other social movements like “Fridays for Future,” Movements against Racism or LGBTQ+ Pride and see whether they have similar behavior online and offline.

References

- Bennett, W. L., Segerberg, A., Walker, S.: Organization in the Crowd: Peer Production in Large-Scale Networked Protests. *Information, Communication & Society*, 17(2), 232–260 (2014). <https://doi.org/10.1080/1369118x.2013.870379>.
- Blei, D. M., Ng, A. I., Jordan, M. I.: Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, 3, 993–1022 (3 January 2003). <https://dl.acm.org/doi/abs/10.5555/944919.944937>.
- Cindo, M., Rini, D. P., Ermatita, E.: Sentiment Analysis on Twitter by Using Maximum Entropy and Support Vector Machine Method. *SINERGI*, 24(2), 1–8 (2020). <https://doi.org/10.22441/sinergi.2020.2.002>.
- Covington, P., Adams, J., Sargin, E.: Deep Neural Networks for YouTube Recommendations. *Proceedings of the 10th ACM Conference on Recommender Systems*, 191–198 (2016). <https://doi.org/10.1145/2959100.2959190>.
- Curran, J.: *Medios de comunicación y poder en una sociedad democrática*. Hacer (2005).

- Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., Gupta, S., He, Y., Lambert, M., Livingston, B., Sampath, D.: The YouTube Video Recommendation System. *Proceedings of the Fourth ACM Conference on Recommender Systems - RecSys'10*, 293–296 (2010). <https://doi.org/10.1145/1864708.1864770>.
- Feinerer, I., Hornik, K.: *Package 'tm'* (2019). <https://bit.ly/2BtY53q>.
- Filho, D. B. F., Rocha, E. C. da, Júnior, J. A. da S., Paranhos, R., Silva, M. B. da, Duarte, B. S. F.: Cluster Analysis for Political Scientists. *Applied Mathematics*, 5(15), 2408–2415 (2014). <https://doi.org/10.4236/am.2014.515232>.
- García-Marín, J., Calatrava, A., Luengo, Ó. G.: Debates electorales y conflicto. Un análisis con máquinas de soporte virtual (SVM) de la cobertura mediática de los debates en España desde 2008. *El Profesional de La Información*, 27(3), 624–632 (2018). <https://doi.org/10.3145/epi.2018.may.15>.
- Granovetter, M.: The Strength of Weak Ties. *American Journal of Sociology*, 78(6), 1360–1380 (1973).
- Graves, A., Mohamed, A., Hinton, G.: Speech Recognition with Deep Recurrent Neural Networks. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1–5 (2013). <https://doi.org/10.1109/icassp.2013.6638947>.
- Houghton, P. Fourth-Wave Feminism and How Social Media Shapes Our Protests. *Hashtag Activism* (2019, March 6). <https://bit.ly/2Bppjbl>.
- Johnson, T. B.: Waves of Feminism and the Media. *Lewis Honors College Capstone Collection*, 32 (2017).
- Justicia de la Torre, C., Sánchez, D., Blanco, I., Martín-Bautista, M. J.: Text Mining: Techniques, Applications, and Challenges. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 26(4), 553–582 (2018). <https://doi.org/10.1142/s0218488518500265>.
- Katz, E., Lazarsfeld, P. F.: *Personal Influence: The Part Played by People in the Flow of Mass Communications*. Transaction Publishers (1955).
- Klapper, J. T. *The Effects of Mass Communication*. Free Press (1960).
- Knobloch-Westerwick, S.: *Choice and Preference in Media Use*. Routledge (2015).
- Lindstedt, N. C.: Structural Topic Modeling For Social Scientists: A Brief Case Study with Social Movement Studies Literature, 2005–2017. *Social Currents*, 6(4), 307–318 (2019). <https://doi.org/10.1177/2329496519846505>.
- Mao, Y., Tian, J., Han, J., Ren, X.: Hierarchical Text Classification with Reinforced Label Assignment. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 1–14 (2019). <https://doi.org/10.18653/v1/d19-1042>.

- McCombs, M., Shaw, D.: The Agenda-Setting Function of Mass Media. *The Public Opinion Quarterly*, 36(2), 176–187 (1972).
- Pariser, E.: *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Press (2011).
- Petersen, M. B.: The Evolutionary Psychology of Mass Mobilization: How Disinformation and Demagogues Coordinate Rather Than Manipulate. *Current Opinion in Psychology*, 35, 71–75 (2020). <https://doi.org/10.1016/j.copsyc.2020.02.003>.
- Robinson, D., Misra, K., Silge, J.: Package ‘widyr’ (2020). <https://bit.ly/3eSOoKl>.
- Rubin, A. M.: The Uses-and-Gratifications Perspective of Media Effects. In J. Bryant & D. Zillmann (Eds.), *LEA’s Communication Series. Media effects: Advances in theory and research*, 525–548. Routledge (2002).
- Scruggs, J. F.: The ‘Echo chamber’ Approach to Advocacy. *Truth Tobacco Industry Documents Collections* (1998). <https://bit.ly/3gj5ukR>.
- Silge, J., Robinson, D.: *Text Mining with R*. O’Reilly (2017).
- Søgaard, A., Vulić, I., Ruder, S., Faruqui, M., Hirst, G.: *Cross-Lingual Word Embeddings*. Macmillan Publishers (2019).
- Spärck Jones, K.: A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *Journal of Documentation*, 28(1), 11–21 (1972). <https://doi.org/10.1108/eb026526>.
- Sunstein, C. (2007). *Republic.com 2.0*. Princeton University Press (1972). <https://doi.org/10.2307/j.ctt7tbsw>.
- Talib, R., Kashif, M., Ayesha, S., Fatima, F.: Text Mining: Techniques, Applications and Issues. *International Journal of Advanced Computer Science and Applications*, 7(11), 1–5 (2016). <https://doi.org/10.14569/ijacsa.2016.071153>.

Serrano-Contreras, I.-J., García-Marín, J., y Luengo, Ó. G. (2020). Measuring online political dialogue: does polarization trigger more deliberation? *Media and Communication*, 8(4), 63-72. <https://doi.org/10.17645/mac.v8i4.3149>

Measuring Online Political Dialogue: Does Polarization Trigger More Deliberation?

Abstract

In recent years, we have witnessed an increasing consolidation of different realms where citizens can deliberate and discuss a variety of topics of general interest, including politics. The comments on news posts in online media are a good example. The first theoretical contributions called attention to the potential of those spaces to build a healthy (civic and participatory) public sphere, going much deeper in the process of political dialogue and deliberation (Fung, Gilman, & Shkabatur, 2013; Lilleker & Jackson, 2008; O'Reilly, 2005; Stromer-Galley & Wichowski, 2011). Polarization has been configured as a constant feature of the quality of the mentioned dialogues, particularly in Mediterranean countries (polarized pluralists' cases). One of the research challenges at the moment has to do with the scrutiny of polarization within the political deliberation provoked by news stories. The goal of this article is the analysis of political dialogue from the perspective of the polarization in the increasingly popular network YouTube, which is presenting very particular characteristics. Using a sample of almost 400,000 posted comments about diverse topics (climate change, the Catalanian crisis, and Political parties' electoral ads) we propose an automated method in order to measure polarization. Our hypothesis is that the number of comments (quantitative variable) is positively related to their polarization (qualitative variable). We will also include in the examination information about the ideological editorial line of newspapers, the type of topic under discussion, the amount of traceable dialogue, etc. We propose an index to (1) measure the polarization of each comment and use it to show how this value has behaved over time; and (2) verify the hypothesis using the average polarization of comments for each video.

Keywords: algorithms; polarization; public sphere, social media; text mining; YouTube

1. Introduction

Political interaction is essential for democratic societies. Through dialogue, citizens clarify their points of view, come into contact with the opinions of other parties, and shape the problems that people need to face (Stromer-Galley & Wichowski, 2011). Some scholars have pointed out that conversations about topics of general interest are a requirement for the integral understanding of democratic life and, consequently, are fundamental in order to provide the meaning of effective political participation, reinforcing the legitimacy of democratic systems (Rubio, 2000; Scheufele, 2001).

In the last decade, we have witnessed the proliferation of different online spaces for discussion, information exchange and deliberation that were pointed out as potential spheres for political dialogue. Around the end of the first decade of the 21st Century, when social media—understood as applications and websites employed to share information and build networks between people (Osborne-Gowey, 2014)—started to gain attention in detriment of other traditional 1.0 online platforms (online newspaper blogs), journalists, analysts, and scholars began to speculate about the real possibility to generate domains where regular citizens could find a space to share information, to portray social and political events, to exchange opinions, and to dialogue, far away (at least in theory) from the biased editorial alignment of international news corporations. Those platforms create a networked sphere of political discussions that is structurally independent from the traditional arena of politics or news; yet, it connects with the two through official affiliations and real-life interactions (Lindgren, 2011). Hence, we have observed the transformation of the public sphere, from a traditional model based on a strongly hierarchical and mainly one-way mass communication, to a network-based multidirectional and horizontal communication (López García, 2006).

Most of the communication research attention on political deliberation in the last years has been focused on the interactions observed in Twitter and Facebook (Conover et al., 2011; Grusell & Nord, 2012; Gruzd & Roy, 2014; Jaidka, Zhou, & Lelkes, 2019; Marichal, 2016; Oz, Zheng, & Chen, 2018). We assume that behind this fact is not only the traditional leadership of those social networks, but also the relatively easier logistic process of extracting and generating samples for analysis. These studies are focused on platforms that were designed to link people together. YouTube, instead, was built as a video log, although its functions are shifting to a more social media like logic. The inclusion of a recommendation system, likes and dislikes, combined with the management of data, enable us to consider YouTube as an online social network (Ma, Wang, Li, Liu, & Jiang, 2013). Thus, it is considered more than just a place to watch and share videos, since it is also used as a learning tool (Allgaier, 2019).

The goal of this article is the analysis of political dialogue from the perspective of the polarization in YouTube, which presents very particular characteristics. Launched in 2005, YouTube is the biggest online video platform worldwide, featuring a wide variety of user generated and corporate media content, and the second largest social network, in terms of monthly active users, after Facebook. YouTube has more than 1,900 million users that logged in every month, generating more than 1,000 million hours of content, with local versions in 91 countries. Our final purpose is to check what kind of polarization patterns might be observed from comments on YouTube. To this end, we will analyze the comments regarding the most popular videos on three topics: Election commercials, climate change and the Catalanian political conflict; each of which observed in the context of Spain. The final ambition of this proposal is to make methodological progress in the understanding of online political dialogue by the measuring of polarization with an innovative quantitative tool.

2. Political Deliberation

Citizens in democracies, particularly in the Western world and as a consequence of technological developments, are in the best position to participate in public discussions. New environments have displayed an impressive potential to improve democracy, creating more inclusive, receptive, deliberative and participatory political systems, in other words, contributing to the consolidation of an effective public sphere. This central concept has been defined by contemporary political theorists as a communicative space in which matters of common prominence are considered for discussion by those concerned and affected, in a way that fulfils a number of ambitious normative criteria (Dahlberg, 2004). Statements should consist of arguments, supported by an appropriate reasoning whose validity can then be checked by others. The process requires understanding from other participants, and a true and honest effort to come to a shared conclusion. All interested parties should be allowed to participate. Finally, the best arguments should prevail (Schäfer, 2016).

Ferree, Gamson, Gerhards, and Rucht (2002) proposed different models of public sphere in today's democracies. They distinguished four types of traditions where this concept finds arrangement, and specified the criteria that each perspective endorses and emphasises regarding who participates, in what sort of processes, how ideas should be presented, and the outcome of the relation between discourse and decision-making. Consequently, we can point out different combinations of criteria based on four theory types: Representative Liberal, Participatory Liberal, Discursive, and Constructionist (Ferree et al., 2002, p. 316). Social networks show great potential for achieving greater political inclusion and participation, for example, giving visibility to voices marginalized by the mainstream (Berry, Kim, & Spigel, 2010; Bimber, 1998).

Social networks also enable engaged citizens to approach other interlocutors to share information, contrast opinions or forge a position regarding a major public event. Therefore, networks have a very high potential to provide the perfect environment to foster dialogue around important policy issues. Unlike other traditional media, social media have facilitated a connection between people, regardless of distance, geographical location or political affiliation. The advent of online media has been considered as a second structural transformation of the public sphere.

Many scholars in the last years have contributed to the discussion on the updated notion of digital public sphere. This new notion includes some renewed important aspects. Following Schäfer, we approach the concept of communicative sphere from a broad perspective, including social media, websites, social network sites, weblogs, and micro-blogs; in all cases, sites where participation is open and freely available to everybody who is interested, where matters of common concern can be discussed, and where proceedings are visible to all (Schäfer, 2016).

However, the literature shows little agreement on the true scope of the political transformations by social media. Beyond some commonalities, the existing schools of digital public sphere theory differ in what kinds of communication they consider desirable. For almost three decades, different academic contributions (mainly theoretical) have asserted strong hopes and fears regarding the development of an online public sphere and its effects on society. Therefore, there are great doubts about the articulation of an effective, more inclusive and deliberative public sphere which is self-mediated (Chadwick, 2009; Hindman, 2009). We can say that social networks may not be definitive when generating a unified public sphere. Similarly, it is difficult to calibrate how the new social media will consolidate two-way communication channels with institutions, parties and candidates. On the one hand, no one seems to question that online media provide a relatively open, easy, and fast access to information, enable more people to make their voices heard in society, and help to produce new kinds of communication. On the other hand, we can point out other factors stated by some pessimistic scholars: Firstly, equal access to online sources of participation is a fiction, thanks to multiple digital divides. Secondly, the danger of fragmentation into small communities of like-minded people veers from a real exchange between participants (e.g., filter bubbles). Thirdly, the inevitable pressure of economic influences on a biased selection of topics to be discussed. And, lastly, the lack of social courtesy derived from an absent face-to-face communication leads to an irrational, emotional and somehow unrespectful interaction (Schäfer, 2016). In that sense, many skeptics understand that a high level of anonymity can exacerbate uninhibited communicative behaviors, moving in the direction of increased disrespectful and even aggressive political discussion (Rowe, 2014). In this sense, we can observe examples of polarization, radicalization or activism in the so-called trolls. This is what Byung-Chul Han calls 'swarm democracy' (Han, 2014), which could be defined as the private sum of reactive multitudes, which move on the basis of discharges of flattery or disqualification and which, like an earthquake, shake up the spaces provided by social networks. Rather than promoting rational and informed deliberation, social media work by amplifying and modulating an atmosphere of individual and collective feelings and actions (Arias Maldonado, 2016). Thus, the interaction between personal networks is altered, with small-scale movements that, together, will produce enormous effects (Granovetter, 1973). The zero threshold would be filled with individuals who are ready for action and prepared for conflict (Petersen, 2020).

In addition, in recent times we have been able to see how the dissemination of biased and malicious information (fake news), primarily aimed at altering perceptions of political discourses, distributing misinformation or directly manipulating, has influenced the dynamics of online political conversations, and even political outcomes. A recent report from the Pew Research Center estimates that two-thirds of the links tweeted on the most popular pages belong to automatic, non-human accounts (Wojcik, Messing, Smith, Rainie, & Hitlin, 2018).

Currently, polarization has been included as one of the main social and political phenomena of the present century, being a traditional object of study within political communication (see, for example, the excellent study by Sunstein, 2002), especially from the media (Prior, 2013) or political perspectives (Gentzkow, Shapiro, & Taddy, 2016). In that way, the work by Hallin and Mancini (2004) proposed a definition of polarization as partisan coverage by the media and has even constituted the basis of their different models of political and media systems. However, polarization has also been defined as a media or even a political effect (Bernhardt, Krasa, & Polborn, 2008). In addition, we can find other perspectives where this process can be considered as the distance between opposing political views. Some traditional studies in the case of the United States in the last decades are good examples: those focused on the increasing divergences between Democrats and Republicans (Mason, 2014). Those studies on polarization have increased in recent years, especially in the wake of the current political situation in the United States, where there seems to be a feeling of growing dislike and distrust between the supporters of both parties. In fact, the term 'affective polarization' is used to describe it (Iyengar, Lelkes, Levendusky, Malhotra, & Westwood, 2019). These contexts offer new meanings and measures that state a potential direction of this development. The case of Spain is not an exception: The political boundaries represented in the national parliament have become sharper in recent years, and two very different and antagonist blocs actively emerged. The phenomenon generated new parties within political wings (Bramson et al., 2017). This portrayal is also replicated at the regional level with another kind of polarization with particular variables and settings. In this sense, we are addressing different and complex involvements developed over time, across populations and comparative studies (Bramson et al., 2017). Thus, assumptions about the notion of polarization continue open. Its main meaning is related to the growth of the space between poles, caused, mostly, by the influence of emotions and beliefs, above evidence and reason (Goldman & O'Connor, 2019; Mason, 2014; Olsson, 2013). Such distinctions could drive to extreme positions (Fletcher & Jenkins, 2019; Gidron, Adams, & Horne, 2019).

Naturally, with the evolution of the mass media paradigm, new ideas are emerging that reflect the scope of change. One such change (Bessi et al., 2016) is the emergence of non-mediated processes, at least by journalists (Prior, 2013). Another is that the new public sphere is fundamentally virtual. Recently, scholars have found multiple evidence that social media and the general Internet environment can cause an increase in political or ideological polarization (Tucker et al., 2018). Particularly, research on online deliberation has identified a series of problems with online debate: Groups are often composed of like-minded people, so issues that generate disagreement or difficulty tend to be avoided. At the same time, a spiral of analogous reasoning occurs where deliberation does not exist (Wright, Graham, & Jackson, 2017). To reflect that, some authors coined the concept of 'echo chambers' (Bail et al., 2018). Those chambers have the effect of isolating individuals from opposing points of view, spinning the polarizing effects.

But there are authors who claim that these effects can come from selective exposure, for example on Facebook (Spohr, 2017). However, almost all these studies focus on the consumption of information on Facebook or Twitter (Conover et al., 2011), defined even as junk news (Narayanan et al., 2018).

Continuing in this vein, social network products must be approached from new perspectives due to new semantic styles and the large amount of data available. Our aim is to understand how deliberation occurs and whether it could be derived from an apparent polarization (Bramson et al., 2017).

3. Design

Our basic assumption is that an increase in political deliberation should trigger more polarization. Hence:

- H1. There is a direct relation between polarization and deliberation.

Given the volume of literature on social media polarization, the lack of operational definitions is surprising. Scholars often assume that polarization is a result of differing views on ideological or political issues, but offer few clues on how to measure this dynamic. From this perspective, we can define polarization as “the existence of two or more alternative and relatively coherent visions that contradict their most important elements. Polarized coverage tends to strongly criticize the opposing view, making the public more reluctant to consider the opposing position legitimate” (Balán, 2013, p. 477).

But the problem still persists: How do we measure polarization? We propose to use Balan’s definition centered in the existence of criticism to try to create a consistent way to measure this concept. Polarization, then, implies that people “hold overwhelmingly positive views of their own co-partisans and highly negative views of those on the other side of the political spectrum” (Gentzkow, 2016, p. 13). Therefore, we can affirm that the wider the distance between positive and negative statements, the greater polarization of the topic (or text). In fact, polarization itself can be the aforementioned distance. Nowadays, there are different tools for measuring positive and negative texts, fundamentally sentiment analysis. This type of analysis, used routinely by both industry and academia, is the basis of many tools and computer algorithms. Sentiment analysis—and opinion mining—is the field of study that analyzes people’s opinions, sentiments, evaluations, attitudes, and emotions in written language. It is one of the most active research areas in natural language processing and is also widely studied in data mining, Web mining, and text mining (Liu, 2012). Although sentiment analysis can use different techniques, there are usually two approaches: The use of a supervised learning algorithm trained with human coded data; or the use of a lexicon (a dictionary) to infer the tone of the text through an automated analysis (based on the presence of a specific keyword, a group of them, or a bag of words). The first approach works well when dealing within specific topics (García-Marín

&Calatrava, 2018) and the second is more transversal but may offer less accuracy if the lexicon is not adequately created (Boukes, van de Velde, Araujo, & Vliegenthart, 2019). Although sentiment analysis has been justly criticized, it does offer some useful information to researchers, as long as it is based on wellweighted dictionaries (Boukes et al., 2019).

There are not many sentiment lexicons for languages other than English, but the number is growing. However, many of them are just automatic translations using different resources, such as Google Translator, which can lead to important mismatches. In the case of the Spanish language, there are several dictionaries designed specifically for sentiment analysis. In this research we are going to use ML-SentiCon. We consider it an ideal resource: It has been evaluated (something not too common) and corrected (at least in 4 of 8 layers), and consequently is not a mere translation. It has 8 layers, depending on reliability, from more to less reliable. The first layer has 97.73% and the last 61.29%: We have decided to use 6 layers, accumulating 2848 words with a reliability of 86.09%. The results are significantly lower than those obtained for the English language but we understand that they are enough for the measurement of polarization in the Spanish language (Cruz, Troyano, Pontes, & Ortega, 2014). Since sentiment analysis is not an accurate measure of polarization, in this article we propose a way to measure polarization from our corpus. We propose the development of an index to quantify this phenomenon, through the weighting of values with which to operate (Ferrando, 1987, p. 34). To that end we have operationalized the polarization of a comment (P_c) from the distance between the sentiment analysis of the comments (S) and the median of the sentiment analysis aggregate of all the comments of each processed video (Me), in absolute number. In this way we obtain a number that can take any value between 0 and 2, where 0 is no polarization and 2 is the maximum polarization:

$$P_c = |(S - Me)|, \text{ where } S \in [-1, 1], p \in (0, 2)$$

Then, the median of the polarization measure has been calculated again to give each video an average polarization ($P_y = Me \vee (S - MeS) \vee p_y = Me(p_c)$) that was useful to relate this dimension e to the rest of the variables of each video. Therefore, the results of applying the formula are as follows: a measure of the polarization of each comment, used to show how polarization has behaved over time; and, second, a measure of the average comment's polarization of each video, used to observe the specific connections in order to demonstrate our hypothesis. The database is composed of YouTube video interactions. We decided to use YouTube because it stands out from other social media in that the video is the primary piece, leaving the text on a secondary level. However, the interactions that open up within it are multiple. In this way, we will have the possibility of analyzing very broad patterns of behavior (number of views, likes or dislikes, replies,...), as well as the potential relationships between them. Data extraction took place on October 25, 2019. The collection was based on three topics and two search patterns, producing a total of 600 videos, disaggregated as shown in Table 1. The

first pattern was the use of the following keywords in the YouTube search engine: “cambio climático” (climate change) and “independencia de cataluña” (independence of Catalonia), from which the 200 most viewed videos per topic were selected. The second search pattern was selected from the YouTube channels of the five Spanish parties with the most parliamentary representation after the April 2019 elections: PSOE, PP, Ciudadanos, Podemos, and Vox. The 40 most viewed pieces posted by political groups within their respective channels were selected, resulting in a total of 200 videos. Out of the total 600 videos, 19 of them had the comment board disabled: 12 of these were from political parties, 4 on “independencia de cataluña” and 3 on “cambio climático.” The final result of the search, and subsequent cleansing, produced a total corpus of 391,591 comments divided into:

- “Catalonian independence”: 162,789
- “Climate change” (as a hoax or scientific fact): 124,820
- Political parties’ videos (their official channels): 103,982

Of course, we decided to use these three topics because we understand that they are susceptible to producing polarizing behavior.

4. Results

We believe that the proposed measure of polarization has a great explanatory potential (Figure 1). In the three cases analyzed, the images show important details that, to a great extent, portray the actual informative dynamic. In the case of Catalonia (at the top in Figure 1), there are two clear points: September-December 2017 and September-October 2019. Both are the moments of greatest tension in the political conflict: The first represents the failed declaration of independence and, the second, the Supreme Court’s decisions on some pro-independence Catalanian politicians (with minor peaks that coincide with elections in Catalonia in recent years). The second case, on climate change, is less evident, since there are no important events in this regard, although there is a clear increase in polarization during the analyzed years. And finally, in the case of political parties, there are also some peaks that mainly correspond to electoral periods in Spain. That is, we think that, in fact, measuring polarization as the difference between positive and negative feelings has great explanatory potential.

As mentioned before, we consider the three topics as potentially polarizing. Although the proposed index shows great explanatory potential, Table 2 indicates that the hypothesis is negative. That is, the relationship between the number of comments and their polarization is not positive in all cases. Specifically, there is a positive, although weak, relationship for the cases of climate change and political parties, but not for the Catalanian political conflict.

Of course, the same behavior is reproduced with the rest of the variables: The number of visits, comments, likes or dislikes present the same results. This is not really surprising, since they are numerical variables that tend to increase as a whole (i.e., the more visits, the more likes, dislikes, and comments). We understand, then, that contextual differences, specific to each topic, affect the results.

At the moment, it seems difficult to understand how the deliberations in social media could become polarized. There may indeed be underlying factors that favor it, such as confirmation bias and content promotion algorithms. Both situations could encourage the construction of echo chambers (Bessi et al., 2016). However, in our study we have observed that there are aspects that differ. In the case of videos on climate change and those of

Table 1. Analyzed videos and comments from YouTube.

	Videos	Views	Comments	Likes	Dislikes	Views for comment	Med_Pol	SD
Catalonian independence	192	37.765.741	163.545	430.074	151.717	230	0.23	0.07
Climate change	186	68.460.219	104.175	1.019.987	67.450	657	0.19	0.10
Political parties'	159	44.239.860	124.019	672.596	123.015	357	0.20	0.08
Total	537	150.465.820	391.739	2.122.657	342.182		0.20	0.08

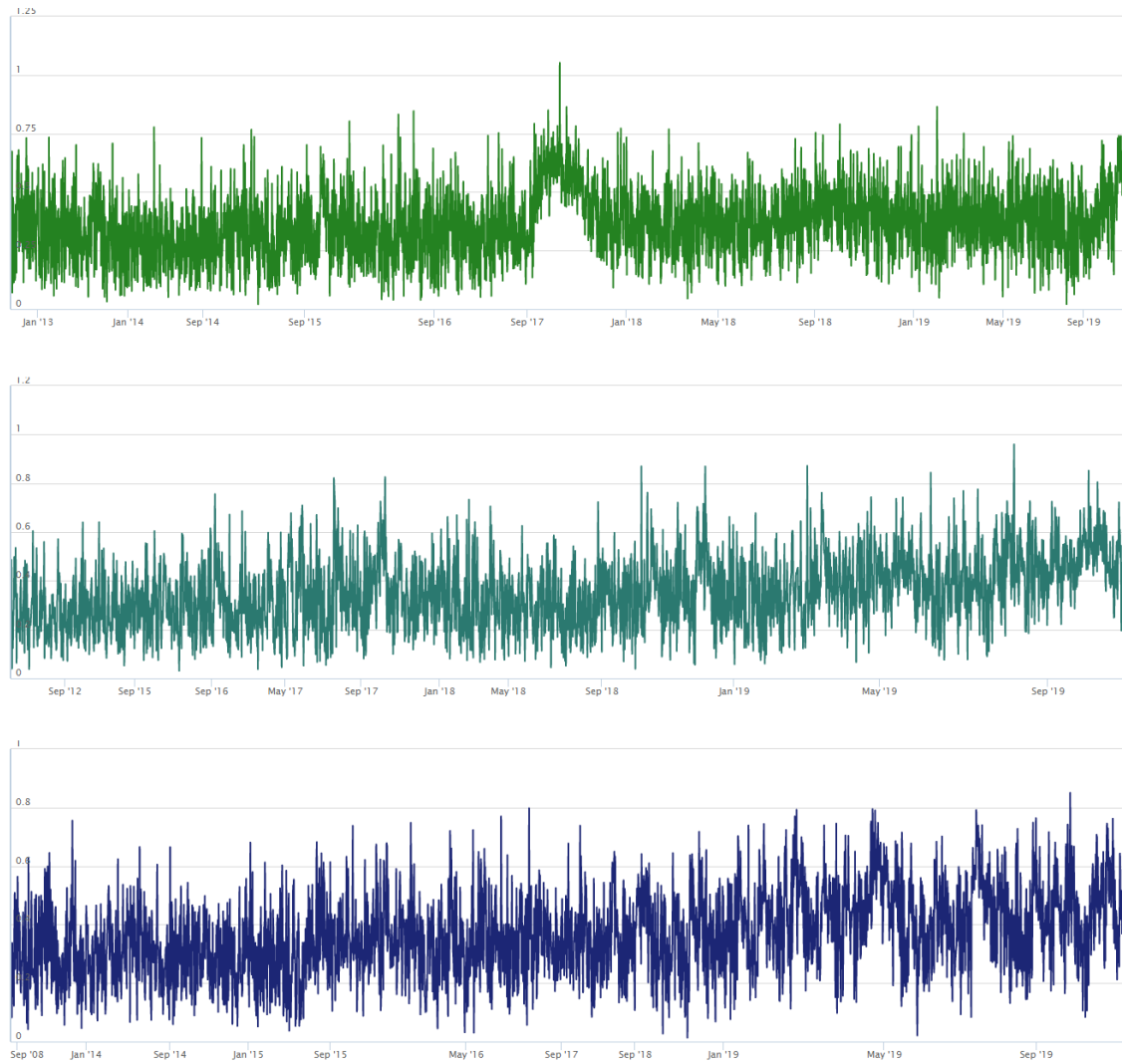


Figure 1. Polarization through time: Catalan independence, climate change and political parties (from top to bottom)

political parties, the presence of this polarization could be influenced by the presence of active trolls that radicalize the conversation. Nonetheless, there could also be external factors that we have not controlled. However, the conflict in Catalonia behaves differently. Although the index does seem to be illustrative, the correlations reject the model based on the fact that greater participation produces greater polarization. However, the behavior of deliberation in the Catalanian conflict is similar to the other issues on very specific dates, such as September and October 2017 (see Figure 1). In that year, 2017, different events took place, such as the so-called Disconnection Law or the October 1st referendum (in both cases serious and polemic attacks against the Spanish constitution).

Table 2. Spearman correlations (polarization) (R)

	All (n = 537)	Climate change (n = 186)	Catalonian independence (n=192)	Political parties (n =159)
Publish_date	0.161**	0.265**	-0.090	0.043
ViewCount	0.048	0.215**	-0.027	0.151
Comments	0.269**	0.311**	-0.005	0.400**
Likes	0.100*	0.228**	-0.076	0.220**
Dislikes	0.144**	0.248**	-0.015	0.236**

Notes: * significant at $p < 0.01$; ** significant at $p < 0.001$.

These two events attracted a great deal of media attention and, accordingly, also the interest of YouTube users. This trend seems to be manifesting itself again in the following major media event, which took place in October 2019 (ruling of the Spanish Supreme Court on the events of 2017). Undoubtedly, there are temporal trends that, unlike previous periods—such as 2014–, are more likely to relate greater participation to greater polarization, as occurred in the other two issues. These discrepancies open the door to new ideas that should be investigated, to explain how in certain scenarios polarization is not instigated by participation.

So, how do we interpret the Catalonian case? The first thing we can say is that, surprisingly, the Catalonian case is characterized by less polarization (except for some very specific dates). This fact, on the other hand, may be consistent with some findings by the academic community. Firstly, perhaps the most obvious factor is the traditional media attention. When this is greater, the behavior of polarization seems to resemble that of other issues. However, when there is not such a presence, the open discussions linked to these videos are not affected by constant polarization, leaving the polarization index without very high values (which does not mean, in any case, that it is positive or negative). Secondly, this argument can accommodate (not replace) a moderate influence driven by filter bubbles. That is, a separation in media consumption could occur according to ideological position. That consumption would be moderated, surely, by activism (the more activism, the more consumption of related information). And the effect would be similar: a low polarization, since the ideological point of view of the participants is similar.

At present, these effects act as echo chambers and filter bubbles. This conception has been intense, especially since 2009, when Google began to modify its search engine to suit the user (Pariser, 2011). While filter bubbles are generated from an unconscious action of the user, echo chambers must prevail consciously, hence one may be part of the other. The assumption that differentiates the two is the willingness and predisposition for users to manifest in the creation of an environment (Klapper, 1960; Knobloch-Westerwick, 2015; Rubin, 2002) governed by scales of reliability: constituents (unaided), major fundraisers, local media, colleagues, national media, advertising & lobbyists (Scruggs, 1998). In order to construct the bubble within the network, and to generate the alteration of

behavior, it is necessary to develop both analogical and digital reflections (Holbert & Stephenson, 2003).

In the case of the other two topics, a partisan (ideological) predisposition towards the topic under discussion is especially palpable in the videos of political parties where the predisposition towards certain channels is influenced by external factors (Barberá, Jost, Nagler, Tucker, & Bonneau, 2015). This would make focus on certain YouTube channels (not on specific videos) interesting, such as those of political parties.

If that is true, the data would show that YouTube content is permeable to the traditional media agenda (which, in turn, can be an indicator of the popularity of an issue, not its origin). In other words, there are moderators to be controlled. This dynamic could cause the debate to shift and polarize in YouTube (due to the possibilities that the platform offers). In other words, events of great media interest (the dates mentioned of 2017 and 2019 for the Catalan case or the specific dates of the political parties' election campaigns), are more favorable to a positive correlation between participation and polarization.

5. Conclusion

The first conclusion is that the main hypothesis has not been fully verified. There are no significant relationships suggesting that increased participation leads to increased polarization, at least at the aggregate level. There are indications that such a relationship may exist at specific time periods or for specific issues. This shows (especially in the Catalan issue) that there are intervening variables (the context, the issue, the time frame, etc.) that have not been controlled. These variables are not playing comparable roles in the three topics under examination. This could help to explain disparities in the significance. Despite this, new computer techniques do prove useful for a better understanding of these phenomena, especially when working with large data sets. Thus, we believe that this study sheds light on YouTubers' behavior, which is part of our purpose. However, it is necessary to develop different tools to understand the functioning of polarization in social networks. The methodology presented has proven to be reliable during the analysis in measuring the polarization of YouTube comments. According to its conception, it could be used on different samples of very different nature. In turn, YouTube has shown to be a network without as many limits as Facebook or Twitter, and less opaque.

As Sunstein (2008) points out, to the extent that there are different social media networks, and within them an almost infinite variety of compartments, regular users end up choosing their interlocutors and the most comfortable space for them. In this sense, it is argued that networks constitute 'echo chambers' in which we only hear the echo of our own predispositions. Consequently, in such spaces contact between groups with different opinions could be more difficult and

reinforcement mechanisms could become the main effects. Moreover, the fragmentation of public opinion can have a reducing effect on social cohesion.

Thus, we could expect a deeper radicalization of the opinions of Internet users who never confront their potential opponents online. Therefore, a mechanism similar to that announced by Lazarsfeld, Berelson, and Gaudet in the 1940s (1960) could emerge, based on selective perceptions, which would be reactivated in a somewhat different way. These mechanisms would point in a similar direction by reinforcing rather than challenging or opposing previous inclinations towards public issues.

As we insisted before, there is a tension between an 'optimistic' approach, which argues that social networks enable interaction between different citizens and thus foster pluralistic debates, and a critical approach, which argues that networks facilitate segregation across ideological lines and then reduce debates by avoiding the contact of contradictory opinions. Discussions about the potential and actual role of social media in public debate are often foreshortened to a mere contest between utopian and dystopian perspectives (Bruns, 2018): It is the balance between utopian and dystopian visions that unveils the true nature of the Internet as a public sphere (Papacharissi, 2002). Although, in order to conclude, we can affirm that the social media, undoubtedly, show a great capacity to respond to citizens' needs. This fact becomes much more important in the current context of disaffection and political unpredictability. The possibilities of social networks are enormous. In political terms, we can recognize beneficial effects: (a) they help to make political discourse more pluralistic; (b) they facilitate greater public involvement and allow citizens to monitor and control power and participate in decisionmaking; (c) they produce and offer more information; and (d) they provide new formats for the transmission of political content. Even so, the revolution presented by the media is complex and on countless levels. We are talking about systems that are growing all the time, changing the process and the function. This kind of evolution can be seen in the constant dynamism used in the algorithms they employ. In this way, we could find enormous improvements in the development of computing applied to the web. In that sense, we can underline the adjustment implemented by Google in its browser and platforms– YouTube is a good example of it (Covington, Adams, & Sargin, 2016), going from simple machine learning to elaborate neural networks. We assume that the contribution from academia will be to enhance models and techniques of research in order to prevent, insofar as possible, adverse potential effects.

Acknowledgments

The authors wish to thank the Editors for their excellent work and the anonymous reviewers for their valuable feedback and suggestions.

Conflict of Interests

The authors declare no conflict of interests.

References

- Allgaier, J. (2019). Science and environmental communication on YouTube: Strategically distorted communications in online videos on climate change and climate engineering. *Frontiers in Communication*, 4. <https://doi.org/10.3389/fcomm.2019.00036>
- Arias Maldonado, M. (2016). La digitalización de la conversación pública: Redes sociales, afectividad política y democracia [The digitalization of public conversation: Social media, political affectivity and democracy]. *Revista de Estudios Políticos*, 173, 27–54.
- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216–9221.
- Balán, M. (2013). Polarización y medios a 30 años de democracia. *Revista SAAP: Sociedad Argentina de Análisis Político*, 7(2), 473–481.
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10), 1531–154.
- Bernhardt, D., Krasa, S., & Polborn, M. (2008). Political polarization and the electoral effects of media bias. *Journal of Public Economics*, 92(5/6), 1092–1104.
- Berry, C., Kim, S., & L. Spigel (2010). *Electronic elsewheres: Media technology and the experience of social space*. Minneapolis, MN: University of Minnesota Press.
- Bessi, A., Zollo, F., Del Vicario, M., Puliga, M., Scala, A., Caldarelli, G., Quattrociocchi, W. (2016). Users polarization on Facebook and Youtube. *PLOS ONE*, 11(8). <https://doi.org/10.1371/journal.pone.0159641>
- Bimber, B. (1998). The Internet and political transformation: Populism, community, and accelerated pluralism. *Polity*, 31(1), 133–160.
- Boukes, M., van de Velde, B., Araujo, T., & Vliegthart, R. (2019). What's the tone? Easy doesn't do it: Analyzing performance and agreement between off-the shelf sentiment analysis tools. *Communication Methods and Measures*, 14(2), 83–104.
- Bramson, A., Grim, P., Singer, D. J., Berger, W. J., Sack, G., Fisher, S., . . . Holman, B. (2017). Understanding polarization: Meanings, measures, and model evaluation. *Philosophy of Science*, 84(1), 115–159.
- Bruns, A. (2018). *Gatewatching and news curation: Journalism, social media, and the public sphere*. New York, NY: Peter Lang.

- Chadwick, A. (2009). Web 2.0: New challenges for the study of e-democracy in an era of informational exuberance. *I/S: A Journal of Law and Policy for the Information Society*, 5(1), 9–41.
- Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on Twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 89–96). Palo Alto, CA: AAAI.
- Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. In *Rec-Sys '16: Proceedings of the 10th ACM Conference on Recommender Systems* (pp. 191–198). New York, NY: Association for Computing Machinery.
- Cruz, F. L., Troyano, J. A., Pontes, B., & Ortega, F. J. (2014). Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Systems with Applications*, 41(13), 5984–5994.
- Dahlberg, L. (2004). The Habermasian public sphere: A specification of the idealized conditions of democratic communication. *Studies in Social and Political Thought*, 10, 2–18.
- Ferrando, M. G. (1987). *Socioestadística: Introducción a la estadística en sociología* [Sociostatistics: Introduction to statistics in sociology]. Madrid: Alianza Universidad Textos.
- Ferree, M. M., Gamson, W. G., Gerhards, J., & Rucht, D. (2002). Four models of the public sphere in modern democracies. *Theory and Society*, 31(3), 289–324.
- Fletcher, R., & Jenkins, J. (2019). *Polarisation and the news media in Europe*. Brussels: European Parliamentary Research Service.
- Fung, A., Gilman, H. R., & Shkabatur, J. (2013). Six models for the Internet & politics. *International Studies Review*, 15, 30–47.
- García-Marín, J., & Calatrava, A. (2018). The use of supervised learning algorithms in political communication and media studies: Locating frames in the press. *Comunicación y Sociedad*, 31(3), 175–188.
- Gentzkow, M. (2016). *Polarization in 2016*. Toulouse: Toulouse Network for Information Technology.
- Gentzkow, M., Shapiro, J., & Taddy, M. (2016). *Measuring polarization in high-dimensional data: Method and application to congressional speech* (SIEPR Discussion Paper No. 16-028). Stanford, CA: SIEPR.
- Gidron, N., Adams, J., & Horne, W. (2019). *How ideology, economics and institutions shape affective polarization in democratic polities*. Presented at the Annual Conference of the American Political Science Association.

- Goldman, A., & O'Connor, C. (2019). Social epistemology. In E. Zalta (Ed.), *Stanford encyclopedia of philosophy*. Retrieved from <https://plato.stanford.edu/entries/epistemology-social>
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380.
- Grusell, M., & Nord, L. (2012). Three attitudes to 140 characters: The use and views of Twitter in political party communications in Sweden. *Public Communication Review*, 2(2), 48–61.
- Gruzd, A., & Roy, J. (2014). Investigating political polarization on Twitter: A Canadian perspective. *Policy & Internet*, 6, 28–45.
- Hallin, D. C., & Mancini, P. (2004). *Comparing media systems: Three models of media and politics*. Cambridge: Cambridge University Press.
- Han, B. C. (2014). *En el enjambre* [In the swarm]. Barcelona: Herder Editorial.
- Hindman, M. (2009). *The myth of digital democracy*. Princeton, NJ: Princeton University Press.
- Holbert, R. L., & Stephenson, M. T. (2003). The importance of indirect effects in media effects research: Testing for mediation in structural equation modeling. *Journal of Broadcasting & Electronic Media*, 47(4), 556–572.
- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22, 129–146.
- Jaidka, K., Zhou, A., & Lelkes, Y. (2019). Brevity is the soul of Twitter: The constraint affordance and political discussion. *Journal of Communication*, 69(4), 345–372.
- Klapper, J. T. (1960). *The effects of mass communication*. Glencoe, IL: Free Press.
- Knobloch-Westerwick, S. (2015). *Choice and preference in media use*. New York, NY: Routledge.
- Lazarsfeld, P., Berelson, B., & Gaudet, H. (1960). *The people's choice*. New York, NY: Columbia University Press.
- Lilleker, D. G., & Jackson, N. (2008). *Politicians and Web 2.0: The current bandwagon or changing the mindset?* Paper presented at the Politics: Web 2.0 International Conference, London, UK.
- Lindgren, S. (2011). YouTube gunmen? Mapping participatory media discourse on school shooting videos. *Media, Culture & Society*, 33(1), 123–136.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1–167.

- López García, G. (2006). *Comunicación en red y mutaciones de la esfera pública* [Online communication and changes in the public sphere]. *Zer*, 20, 231–249.
- Ma, X., Wang, H., Li, H., Liu, J., & Jiang, H. (2013). Exploring sharing patterns for video recommendation on YouTube-like social media. *Multimedia Systems*, 20(6), 675–691.
- Marichal, J. (2016). *Facebook democracy: The architecture of disclosure and the threat to public life*. New York, NY: Routledge.
- Mason, L. (2014). “I disrespectfully agree”: The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science*, 59(1), 128–145.
- Narayanan, V., Barash, V., Kelly, J., Kollanyi, B., Neudert, L. M., & Howard, P. N. (2018). *Polarization, partisanship and junk news consumption over social media in the US*. arXiv. <https://arxiv.org/abs/1803.01845>
- Olsson, E. J. (2013). A Bayesian simulation model of group deliberation and polarization. In F. Zenker (Ed.), *Bayesian argumentation: The practical side of probability* (pp. 113–133). Dordrecht: Springer.
- O'Reilly, T. (2005). *What Is Web 2.0: Design patterns and business models for the next generation of software*. O'Reilly. Retrieved from <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=all>
- Osborne-Gowey, J. (2014). *What is social media*. *Fisheries*, 39(2), 55–55.
- Oz, M., Zheng, P., & Chen, G. M. (2018). Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes. *New Media & Society*, 20(9), 3400–3419.
- Papacharissi, Z. (2002). The virtual sphere: The internet as a public sphere. *New Media & Society*, 4(1), 9–27.
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. New York, NY: Penguin Press
- Petersen, M. B. (2020). The evolutionary psychology of mass mobilization: How disinformation and demagogues coordinate rather than manipulate. *Current Opinion in Psychology*, 35, 71–75.
- Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, 16, 101–127.
- Rowe, I. (2014). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society*, 18(2), 121–138.
- Rubin, A. M. (2002). The uses-and-gratifications perspective of media effects. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (pp. 525–548). New York, NY: Routledge.

- Rubio, R. (2000). Internet en la participación política [Internet in political participation]. *Revista de Estudios Políticos*, 109, 285–302.
- Schäfer, M. S. (2016). Digital public sphere. In G. Mazzoleni (Ed.), *The international encyclopedia of political communication*. <https://doi.org/10.1002/9781118541555.wbiepc087>. Retrieved from <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118541555>
- Scheufele, D. A. (2001). Democracy for some? How political talk both informs and polarizes the electorate. In R. P. Hart & D. Shaw (Eds.), *Communication and U.S. elections: New agendas* (pp. 19–32). Lanham, MD: Rowman and Littlefield.
- Scruggs, J. F. (1998). “Echo chamber” approach to advocacy. Retrieved from <https://www.industrydocuments.ucsf.edu/docs/ggxn0061>
- Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review*, 34(3), 150–160.
- Stromer-Galley, J., & Wichowski, A. (2011). Political discussion online. In M. Consalvo & C. Ess (Eds.), *The handbook of Internet studies* (pp. 168–187). Malden, MA: Blackwell Publishing Ltd.
- Sunstein, C. R. (2002). The law of group polarization. *Journal of Political Philosophy*, 10(2), 175–195.
- Sunstein, C. R. (2008). *Republic.com 2.0*. Princeton, NJ: Princeton University Press.
- Tucker, J., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., Nyhan, B. (2018). *Social media, political polarization, and political disinformation: A review of the scientific literature*. Menlo Park, CA: The William and Flora Hewlett Foundation.
- Wojcik, S., Messing, S., Smith, A., Rainie, L., & Hitlin, P. (2018). *Bots in the Twittersphere*. Washington, DC: Pew Research Center. Retrieved from <https://pewrsr.ch/3cNCOye>
- Wright, S., Graham, T., & Jackson, D. (2017). *Third space and everyday online political talk: Deliberation, polarisation, avoidance*. Paper presented at the 67th Annual Conference of the International Communication Association, San Diego, CA. Retrieved from <http://eprints.whiterose.ac.uk/119308>