

**Dr. Marc BLASCO-DUATIS**

Universitat de Girona. España. marc.blasco@udg.edu

**Dr. Germà COENDERS**

Universitat de Girona. España. germa.coenders@udg.edu

## **Análisis de sentimiento de la agenda de los partidos políticos españoles en Twitter durante la Moción de Censura de 2018. Un enfoque de datos composicionales**

### ***Sentiment analysis of the agenda of Spanish political parties on Twitter during the 2018 motion of no confidence. A compositional data approach***

**Fechas** | Recepción: 13/03/2020 - Revisión: 02/06/2020 - En edición: 08/06/2020 - Publicación final: 01/07/2020

#### **Resumen**

Este artículo presenta el análisis de sentimiento de la discusión de partidos políticos en Twitter, en el contexto de la Moción de Censura al gobierno español de 2018. En particular, extrajimos y analizamos 2824 tweets de las cuentas oficiales de las 13 formaciones políticas representadas en el Congreso de los Diputados. En su desarrollo metodológico aplicamos el análisis composicional de datos y su visualización a través del biplot (una herramienta de visualización que permite contrastar la importancia relativa de los elementos en estudio). A diferencia de los enfoques tradicionales, nuestro estudio enfatiza la importancia relativa de los temas dentro de la agenda, a la vez que incorpora un tercer componente, el análisis de sentimiento. La investigación concluye sobre la fiabilidad del método para representar composicionalmente la agenda y los agenda-setters, así como el análisis de sentimiento, constatando que los temas que se asocian de forma más notable con determinados partidos, también lo hacen con su proyección sobre los sentimientos. El análisis arroja luz sobre la representación de los sentimientos en los agenda-setters (*agenda de atributos*), especialmente en el campo de la comunicación política.

#### **Palabras clave**

Análisis de sentimiento; agenda-setter; agenda de atributos; datos composicionales (CoDa); Twitter; biplot

#### **Abstract**

*In this article we provide an analysis of the sentiment of the political parties' discussion on Twitter, about the 2018 motion of no confidence to the Spanish government. In particular, we extracted and analyzed 2824 tweets from the official accounts of the 13 political parties represented in the Congress of Deputies. In its methodological development we apply a compositional data analysis and its visualization through the biplot (a visualization tool that allows to contrast the relative importance of the elements under study). Unlike traditional approaches, our study emphasizes the relative importance of the issues within the agenda, while incorporating a third component, the analysis of sentiment. The main findings of this research concern the reliability of the method to represent compositionally the agenda and the agenda-setters, as well as the sentiment analysis, confirming that the issues that are most notably associated with certain parties, also are so with their projection on sentiment. In sum, this analysis sheds light on the representation of sentiment in the agenda-setters (attribute agenda), especially in the field of political communication.*

#### **Keywords**

*Sentimental analysis; agenda-setter; attributes agenda; compositional data (CoDa); Twitter, biplot*

## 1. Introducción

La pasada década fue testimonio de la consolidación de las redes sociales como un importante canal para que los dirigentes políticos vehiculen sus discursos, haciéndolos más accesibles a sus posibles votantes (Golbeck, Grimes y Rogers, 2010; Conway et al., 2015; Graham, Jackson y Broersma, 2016). La profesionalización del discurso político en las redes no solamente ha supuesto una forma más de comunicación política competente para influir en la esfera social, sino que a su vez ha demostrado su capacidad para influir en la agenda política y mediática. En este ecosistema mediático-social, Twitter se ha consolidado como uno de los principales espacios políticos para difundir información a una audiencia creciente (Conway et al., 2015), presentándose como una plataforma de referencia para recopilar información, conectarse mejor con los votantes potenciales y seguir la actividad de las fuentes (Parmelee, 2013). Asimismo, Twitter se reivindica como la herramienta estandarizada para establecer conexiones entre partidos, candidatos, periodistas y audiencias, especialmente en los procesos de campañas electorales (Jungherr, 2016), siendo este un importante canal para medir las reacciones de la opinión pública por parte de la acción política. Por ello, la inherente interactividad de la plataforma ha fomentado considerablemente la exposición de los ciudadanos a la discusión política y a la confrontación ideológica (Guo, Rohde y Wu, 2018).

Twitter se ha consagrado como una de las plataformas globales más importantes en términos de su inmediatez en la difusión de información entre la sociedad (Yousuf Al-Aama, 2015). Los acontecimientos sociopolíticos más destacados de los últimos años como la revolución egipcia en 2011, la elección de Trump, el referéndum por la paz en Colombia, el Brexit, o más recientemente movimientos como el #MeToo o el #FridaysForFuture, constituyen una prueba vigente de esto. Las nuevas dinámicas en las relaciones entre los medios y las audiencias brindan un mayor pluralismo a la esfera mediática, llevando al modelo tradicional de la comunicación política a una transformación sin precedentes en la interacción entre sus principales agentes. Todo ello forma parte de un proceso de reformulación sistémico de las dinámicas de poder e influencia, donde el mensaje no se traslada de forma unidireccional de los medios de comunicación tradicionales al público, sino que se entiende como un constructo que se alimenta y se desarrolla por la interacción horizontal y convergente de los distintos actores y plataformas (Scolari, Lugo y Masanet, 2019). Tanto es así que en la última década, desde la academia, la investigación en el campo de las agendas política, mediática y pública se ha centrado especialmente en el estudio de los medios digitales como nuevos espacios para influir en la construcción de los contenidos que en ellas se tratan (Meraz, 2009).

Durante el segundo trimestre de 2018 la plataforma Twitter registraba 338 millones de perfiles a nivel mundial, y 4,9 millones en España. Sus usuarios publicaron una media de 500 millones de tweets por día en todo el mundo (Statistic Brain, 2018). Además, es significativo que el número de usuarios activos en Twitter supera el 22% de los usuarios de Internet en el mundo (Kayser y Bierwisch, 2016). Estas estadísticas reflejan el alcance global de esta red social y su impacto potencial. Además de tener una cobertura global sobre los principales asuntos del día, proporciona un espacio que permite compartir opiniones fácilmente, utilizando varios recursos de contenido, incluidos texto, imágenes, enlaces, a diferencia de muchas otras plataformas de medios sociales. Además, proporciona acceso casi en tiempo real a sus publicaciones a través de la API, impulsándola como una plataforma adecuada para contrastar estados de opinión a gran escala casi en tiempo real. En este sentido, decidimos concentrarnos en Twitter como la principal fuente de datos entre otras redes sociales para llevar a cabo el análisis propuesto en el estudio.

En este contexto, la dimensión emocional de una discusión política en las redes sociales resulta ser de particular importancia, ya que un debate emocional sobre un tema controvertido a menudo se desarrolla de manera más dinámica e impredecible que una discusión objetiva (Kušen, Strembeck, Cascavilla y Conti, 2017). Los métodos de análisis temáticos han contribuido a clasificar y comprender los intereses suscitados entre el público y en base a las distintas agendas. Sin embargo, la complejidad en los estudios sobre las dinámicas de emisión-recepción hacen que el análisis de los sentimientos generados en estas acciones discursivas sean una difícil tarea (Fan y Gordon, 2014). Es así como los estudios de caso sobre acontecimientos políticos del mundo real son de particular interés, ya que contribuyen a comprender el comportamiento humano, detectar patrones e identificar enfoques genéricos para analizar la conducta de los actores implicados en la gestión de las redes sociales en línea (ver, por ejemplo, Ahmadian Azarshahi y Paulhus, 2017; o Ferrer-Rosell et al., 2019, 2020b).

El análisis de sentimiento en Twitter se ha venido utilizando en una amplia gama de áreas relacionadas con la gobernanza y la confianza pública, que van desde la predicción del resentimiento contra las políticas gubernamentales hasta la predicción de resultados en elecciones generales (Tumasjan, Sprenger, Sandner y Welpe, 2010; Calderon et al., 2015). Este campo emergente tiene como objetivo comprender y predecir dicho comportamiento, y aunque esta área de estudio aún está evolucionando

y generando mucho entusiasmo, existe un debate candente sobre la eficacia en el uso del análisis de sentimiento de Twitter para predecir elecciones y otros eventos del mundo real.

Con respecto a la investigación sobre el uso de Twitter en la agenda política, los investigadores se han interesado inicialmente en la forma en que la plataforma influye en determinar la creación de dicha agenda temática para los principales medios de comunicación (Jungherr, 2014; Wallsten, 2014), pero ha ido mucho más allá (Calvo y Aruguete, 2020). El análisis también ha centrado el interés de los investigadores en el estudio de la agenda de segundo nivel (o agenda de atributos), donde el objetivo es descifrar el enfoque dado sobre el tema, lo cual tiene sin duda relación con el análisis de sentimiento. Un ejemplo reciente es el de Lee y Xu (2018). Otro campo de estudio es el de las relaciones dinámicas entre agendas (Neuman, Guggenheim, Jang y Bae 2014). Asimismo, mediante el método de análisis de redes sociales se ha estudiado lo que se ha venido a llamar agenda de tercer nivel, en la que se ponen de manifiesto las asociaciones entre temas percibidos por distintos públicos (Guo y Vargo, 2015).

Desde los inicios del estudio de la teoría del establecimiento de la agenda se sostiene que ciertos medios de comunicación o agentes sociales se erigen, o intentan erigirse, como mediadores de la información, en una praxis basada en intentar incluir u omitir ciertos temas en base a un escenario de jerarquización de los contenidos (McCombs y Shaw, 1972; McCombs, 2004). Ello ha suscitado que las relaciones entre los que podemos llamar agenda-*setters*, o al menos pretenden serlo, y los contenidos (temas de la agenda) se estudiaran tradicionalmente desde una base descriptiva, contabilizando frecuencias de repetición de un tema o mediante las correlaciones de Spearman, donde solamente se contempla el orden en el ranking de temas y se omite cualquier otro tipo de información.

Desgraciadamente los métodos de análisis tradicionales mencionados no centran el análisis genuinamente en la importancia relativa de los temas de la agenda, ni proporcionan herramientas de visualización. Para resolver dicha limitación, en los últimos años se ha empezado a trabajar con la aplicación del análisis de datos composicionales (CoDa) al campo de estudio de la comunicación política (ver a Blasco-Duatis et al., 2018a, 2018b, 2018c, 2019), una metodología estadística estándar que se emplea cuando el investigador está interesado en la importancia relativa de las partes de un todo (Aitchison, 1986).

En nuestra investigación, nos centramos especialmente en una herramienta conocida de visualización de datos, el biplot composicional, que nos permite ver qué temas se priorizan por cuáles partidos políticos (que actúan o intentan actuar como agenda-*setters* cuando usan sus cuentas de Twitter), qué temas varían su presencia de modo proporcional, y qué partidos tienen un comportamiento parecido con respecto a los temas, teniendo en cuenta que lo importante no reside en el volumen de información, sino en el peso relativo de cada tema para con cada partido.

La contribución novedosa de este artículo radica en plantear una visualización de datos composicional que permita representar la agenda de atributos, incluyendo los temas propuestos por los partidos políticos y los sentimientos vinculados a los mensajes publicados en Twitter. En un primer lugar empleamos la herramienta 'Tweet Binder' para capturar los tweets publicados a través de las cuentas oficiales de los partidos políticos en estudio. Seguidamente aplicamos una metodología basada en el análisis de contenido para discernir por un lado los temas que configuran la agenda y, en una segunda fase, los sentimientos derivados de las construcciones semánticas y contextuales de dichos tweets. Ello sirve para en una tercera fase aplicar un nuevo tipo de biplot composicional que permite tratar más de una composición simultáneamente (Ferrer-Rosell et al., 2020c; Filzmoser, Hron y Templ, 2018; Kynčlová, Filzmoser y Hron, 2016), donde el foco está en las proporciones entre los distintos temas de la agenda, los sentimientos, las importancias relativas de ambos para los agenda-*setters* y la relación entre temas y sentimientos.

En suma, en este artículo no solamente proporcionamos un estudio composicional de la agenda y los agenda-*setters* en Twitter (tema que ya fue desarrollado en el trabajo seminal de Blasco-Duatis et al., 2019), sino que extendemos el análisis composicional con un segundo nivel, el análisis de sentimiento, que aporta un valor hasta el momento, y desde nuestro modesto saber, no explorado de forma composicional en el campo de la comunicación política. En concreto compilamos y analizamos 2824 tweets de las 13 formaciones políticas representadas en el Congreso de los Diputados español en 2018. Reunimos un conjunto de datos de carácter temático, de transmisión e información sobre las emociones de los agenda-*setters*, que se expresan en el contenido de los mensajes. El resto de este trabajo se estructura en una sección de presentación de la metodología, que se bifurca en dos subsecciones sobre un primer proceso de codificación-categorización, y un segundo de visualización composicional de los datos. Posteriormente se presentan los resultados sobre el análisis del sentimiento de la agenda en Twitter por parte de los partidos políticos en el marco de la Moción de Censura al gobierno de 2018. Finalmente se detalla una discusión de nuestros principales hallazgos que se eleva a unas conclusiones.

## 2. Metodología

La propuesta metodológica que presentamos constituye una contribución novedosa en el campo del análisis de sentimiento de la comunicación política, ya que aplica el análisis y visualización de datos composicionales a un estudio que, a la par de representar el mapeo de la agenda, incorpora un tercer elemento, el análisis de sentimiento, a dicho análisis (agenda de atributos). El método se desarrolla en una primera fase a través del análisis de contenido sobre el conjunto de temas tratados durante las Moción de Censura al gobierno de 2018 por parte de los principales partidos políticos españoles en Twitter (agenda de primer nivel), y la codificación de los sentimientos derivados de las construcciones semánticas de cada tweet (agenda de segundo nivel). En un segundo estadio, y dado que el volumen no es comparable porque el número de tweets y sentimientos analizados es diferente en cada partido, nos interesan los tamaños relativos. Lo que verdaderamente importa en este supuesto es el porcentaje de aparición de un tema de la agenda o sentimiento (categoría) en cada partido político (o agenda-setter), así como las relaciones entre ellos. En la literatura estadística esta situación composicional se utiliza para analizar datos que conllevan información relativa, donde el foco está en las proporciones entre los distintos vectores-categorías, o temas de la agenda y sentimientos, y sus agenda-setters, o partidos políticos (Aitchison, 1986).

### 2.1. Selección y codificación de los tweets

Existe una afirmación común en el campo del análisis de sentimiento, que los resultados exitosos dependen en gran medida del desarrollo específico que se ha empleado para la recopilación del conjunto de datos en estudio. En este sentido, la mayoría de los enfoques metodológicos en este terreno emplean sistemas basados en técnicas supervisadas de aprendizaje automático y estadístico. Si bien han demostrado ser bastante exitosos en el pasado –y no solo en el campo del análisis de sentimiento, sino en la mayoría de las aplicaciones de extracción de texto y recuperación de información– su desventaja obvia en términos de funcionalidad es su aplicabilidad limitada a dominios de datos distintos de aquellos para los que fueron diseñados (Pang y Lee, 2004). Asimismo, la codificación de sentimientos requiere de un sistema que prevea la medición sobre la orientación semántica de los textos, y no solamente sobre el término, limitando mucho el uso de los modelos tradicionales basados en la minería de textos, que suelen fundamentarse en el uso de diccionarios desarrollados manualmente a los que se asocia un cambiador de valencia.

Por ello, la presente investigación considera que la codificación manual de cada tweet no solamente asegura una clasificación asertiva sobre la temática (tema de la agenda), sino que a su vez permite delimitar con mayor precisión el sentimiento derivado del contexto semántico. Si bien es cierto que en algunos casos un contenido se podría asociar con ciertas palabras clave fáciles de detectar con este método (p.ej., el concepto machismo, claramente asociado a políticas de género), en otros casos no es posible (p. ej., el concepto diálogo, que según el contexto podría referirse a la necesidad de dialogar sobre el proceso independentista catalán, o en contra de dialogar con determinados líderes o partidos implicados con dicho proceso).

Partiendo de esta concepción, aplicamos el análisis de contenido (Berelson, 1952; Krippendorff, 2004; Wimmer y Dominick, 2006) adaptado al entorno de Twitter. Recopilamos una base de datos con la secuencia del conjunto de tweets publicados en las cuentas oficiales de Twitter de los partidos con representación parlamentaria en el momento de la Moción, en adelante denominados: PP (Partido Popular); PSOE (Partido Socialista Obrero Español); UP (Unidas Podemos y las confluencias territoriales así como la integración de IU); C's (Ciudadanos); EAJ-PNV (Euzko Alderdi Jeltzalea - Partido Nacionalista Vasco); ERC (Esquerra Republicana de Catalunya); PDeCAT (Partit Demòcrata Català); Compromís (Compromís); Bildu (Euskal Herria Bildu); CC (Coalición Canaria); NC (Nueva Canaria); FA (Foro Asturias); y UPN (Unión del Pueblo Navarro). El número total de tweets recopilados durante los 6 días de estudio (del 29 de mayo al 3 de junio de 2018), y que incluyó desde los dos días previos a la primera sesión de debate parlamentario (1 de junio) hasta los dos días posteriores al segundo debate y aprobación de la Moción (3 de junio), fueron: PP, 247 tweets; PSOE, 463 tweets; UP, 180 tweets; C's, 491 tweets; EAJ-PNV, 61 tweets; ERC, 493 tweets; PDeCat, 226 tweets; Compromís, 184 tweets; Bildu, 216 tweets; CC, 130 tweets; NC, 91 tweets; FA, 19 tweets; y UPN, 23 tweets. Estos dos últimos partidos regionales (FA y UPN) se descartaron del estudio composicional por su falta de datos y representatividad en su comparativa con la muestra del conjunto de partidos en estudio.

Siguiendo a Blasco-Duatis et al. (2019), se compiló una lista de ochenta temas relacionados con la decodificación del conjunto de tweets analizados. Posteriormente, y por similitud conceptual, los temas se agruparon en 12 categorías, de las cuales solo se seleccionaron las 9 con suficiente entidad representativa, en adelante denominadas top9: (1) proceso catalán (aplicación del artículo 155, referéndum 1-O, separatismo, independencia, TV3, Bildu y Otegui sobre el proceso catalán, paralelismo

con el proceso vasco, formación del gobierno de la Generalitat, aniversario de la Presidenta Carme Forcadell, nombramiento de los Consellers de la Generalitat); (2) política territorial (política autonómica, cuestiones estrictamente territoriales, competencias autonómicas, plurinacionalidad, Día de Canarias, la caída del gobierno municipal de Podem en Badalona); (3) economía-empleo (Presupuestos Generales de España, fondos autonómicos, pensiones, impuestos, plusvalía municipal, PNV contra UPN en las inversiones europeas bloqueadas, crisis económica global); (4) corrupción (comisión de investigación sobre la financiación irregular del Partido Popular, caso Bárcenas, cargos de confianza, caso Dolores de Cospedal, caso Tribunal de cuentas de RTVE, caso alcalde de Orihuela, caso Daniel De Alfonso, comisión sobre la corrupción del Partido Popular en el Congreso, manipulación informativa sobre corrupción); (5) género-igualdad (violencia de género, caso Manada, igualdad, conciliación familiar, casos sobre abusos de la iglesia); (6) política europea (tratado sobre el acero, debates al parlamento europeo, relaciones de los países de la UE con España, política europea asociada al caso catalán); (7) educación-cultura (pacto sobre la educación vasca, comisión de educación al Congreso, nuevos topónimos valencianos, caso Instituto Sant Andreu de la Barca, universidades, investigación, becas universitarias); (8) políticas sociales-sanidad (Menas, gestación subrogada, dependencia, medio ambiente, sostenibilidad, deporte adaptado, libertad de expresión, emergencia social, votación para la derogación de la Ley Mordaza, sanidad, narcotráfico en el Campo de Gibraltar); (9) regeneración (cambio de gobierno, bipartidismo, salud democrática, salud institucional). Cabe apuntar que en el proceso de codificación de los tweets se estimó oportuno categorizar a cada uno con tantos temas (categorías) como se identificaran en su texto. Es así como hay diversos tweets que se han categorizado con dos o más categorías (p.ej., un tweet sobre las implicaciones del proceso catalán en algún país de la UE se ha codificado con los temas (6) política europea y (1) proceso catalán).

A su vez, y en relación con el proceso de codificación de sentimientos, la mayoría de los trabajos publicados en el campo se han centrado en el enfoque dual de "aprobación o rechazo", es decir, obtener una calificación positiva o negativa (Turney, 2002). Un paso adicional a este modelo implica un intento de determinar no solo desde una clasificación binaria los indicadores textuales, sino estableciendo una escala de sentimientos intermedios que se asocian a determinadas construcciones semánticas y de contexto. En este sentido, y siguiendo el trabajo pionero de Moreno-Ortiz y Pérez Hernández (2013) sobre el análisis de sentimiento de mensajes en español en Twitter, decidimos establecer las siguientes categorías: MP (positivo, a favor); P (más positivo que negativo, más a favor que en contra); NEU (neutro, ni positivo ni negativo, de carácter informativo), N (más negativo que positivo, más en contra que a favor); MN (negativo, en contra); NING (no se asocia sentimiento, normalmente derivado de tweets vinculados con agenda de actos del partido o sus representantes). Así, y a diferencia del proceso de codificación-categorización descrito anteriormente sobre los temas de la agenda, atribuimos un único sentimiento a cada tweet. A nivel composicional, la aplicación de este método toma especial importancia en su fase de representación, ya que no solamente permite el estudio de los temas sobre sus emisores (*agenda-setters*), sino que transforma el análisis en un estudio de tres componentes y añade la composición de sentimientos a las interacciones propias de la agenda.

## 2.2. Análisis y visualización de los datos

Siguiendo a Marine-Roig y Ferrer-Rosell (2018), Ferrer-Rosell et al. (2019, 2020a, 2020b, 2020c), y Ferrer-Rosell y Marine-Roig (2020), el análisis de los contenidos y los estilos de comunicación centra su interés en la importancia relativa de dichos contenidos y estilos, es decir sus proporciones, lo cual convierte dicho análisis en un análisis composicional (Aitchison, 1986; Van den Boogaart y Tolosana-Delgado, 2013; Egozcue y Pawlowsky-Glahn, 2016; Filzmoser et al., 2018; Greenacre, 2018; Pawlowsky-Glahn, Egozcue y Tolosana-Delgado, 2015).

El *análisis de datos composicionales* se define como el análisis de las partes de un todo cuando interesa el tamaño relativo de sus  $D$  elementos, *partes* o *componentes*. Una composición se expresa por un vector de partes positivas:

$$\mathbf{x} = (x_1 \ x_2 \ \dots \ x_D) \quad \text{con } x_j > 0, \ j = 1, 2, \dots, D. \quad (1)$$

En nuestro caso, la parte  $x_j$  representa el número de veces que el contenido o tema  $j$ -ésimo aparece en los tweets de un determinado partido. El tratamiento composicional presupone que dos partidos que se distinguen por el hecho de que uno emita el doble de todos los temas que el otro, constituyen dos partidos idénticos a efectos de las preguntas del investigador. Dichas preguntas así consideran un hecho trivial y

de poco interés que algunos partidos sean más activos que otros en la red, globalmente hablando, y por ello emiten más de todos los temas.

Para el cálculo de una medida de tendencia central teniendo en cuenta dicha importancia relativa y no absoluta de los temas, se usa la media geométrica, y se re-escala a un total arbitrario fácilmente comprensible, por ejemplo 1 (proporciones) ó 100 (porcentajes).

Para el cálculo de una medida de asociación, en este tipo de datos se prescinde de las correlaciones que se substituyen por el concepto de *proporcionalidad*. Si dos temas  $j$  y  $k$  se mueven de forma proporcional, su cociente se mantiene constante y el logaritmo de su cociente tiene varianza cero.

$$\text{var}\left(\ln\frac{x_j}{x_k}\right)=0. \quad (2)$$

La matriz de correlaciones se substituye así por la llamada *matriz de variación*, que contiene dichas varianzas de logaritmos de cocientes calculadas para todas las posibles parejas de temas  $j=2,3,\dots,D$ ;  $k=1,2,\dots,j-1$ . Igual que la matriz de correlaciones es simétrica. El valor cero, como se ha comentado, implica asociación perfecta directa entre dos temas. No hay, en cambio, una cota superior que indique asociación perfecta inversa. Valores altos de la varianza indican que los dos temas no varían proporcionalmente.

Como se ha constatado en (2), los logaritmos de cocientes, llamados log-ratios juegan un papel fundamental en el análisis de datos composicionales. La *transformación por log-ratios centrada (clr)* indica la importancia relativa de cada tema, con respecto a los demás temas presentes en la composición. Una manera de entender esta transformación es comparando cada parte con la media geométrica del resto (Filzmoser et al., 2018):

$$\text{clr}(\mathbf{x}) = \frac{D-1}{D} \left( \ln\left(\frac{x_1}{\sqrt[D]{x_2x_3\dots x_D}}\right) \quad \ln\left(\frac{x_2}{\sqrt[D]{x_1x_3\dots x_D}}\right) \quad \dots \quad \ln\left(\frac{x_D}{\sqrt[D]{x_1x_2\dots x_{D-1}}}\right) \right). \quad (3)$$

Cuando los datos contienen ceros (en nuestro caso, temas poco recurrentes que no son tratados en ningún mensaje por un determinado partido) las log-ratios no se pueden calcular, con lo que se hace necesario algún tratamiento de los ceros previo al análisis. En el análisis de contenido ello corresponde con los llamados *ceros de recuento*, que se suele tratar por métodos de imputación Bayesiana, como se recomienda en Martín-Fernández et al. (2015).

Igual como ocurre con el análisis estadístico clásico, el análisis de datos composicionales requiere herramientas gráficas de visualización de datos. Una herramienta muy útil es el *biplot composicional* (Aitchison y Greenacre, 2002; Egozcue y Pawlowsky-Glahn, 2016), y más concretamente el *biplot composicional de covarianza*. Se trata del mismo biplot estándar resultante del *análisis en componentes principales* basado en la matriz de covarianzas de los datos transformados con log-ratios centradas. El biplot representa conjuntamente la importancia relativa de los  $D$  contenidos o temas sometidos a la transformación *clr* y los *agenda-setters* (en este caso los partidos políticos), sobre las dos primeras dimensiones del análisis en componentes principales. Los temas se suelen representar como vectores con origen en el centro de coordenadas, y los *agenda-setters* como puntos. El biplot constituye la mejor aproximación posible de los datos originales en dos dimensiones. La calidad de esta aproximación viene indicada por el porcentaje de varianza explicada por dichas dos primeras dimensiones.

En el *biplot de covarianza* el principal elemento de interpretación son las *distancias entre los extremos de los vectores-tema*. Dicha distancia es aproximadamente proporcional a la desviación típica del logaritmo del cociente entre los dos temas implicados, en otras palabras, la raíz cuadrada de (2). Extremos de los vectores próximos entre sí indican temas que guardan una proporcionalidad aproximada, en el sentido, por ejemplo, de parejas de temas tales que si se dobla la proporción de uno en unos determinados partidos políticos, también se dobla la proporción del otro. Extremos alejados indican lo contrario, en el sentido, por ejemplo, de parejas de temas tales que si aumenta la proporción de uno en unos determinados partidos, disminuye la proporción del otro.

El biplot permite también visualizar la importancia aproximada de cada tema para cada partido, en términos relativos. Si *proyectamos los partidos ortogonalmente* (los dejamos caer formando un ángulo de 90 grados sobre la dirección definida por cada vector-tema), el sentido indicado por el vector-tema señala hacia los partidos con un mayor contenido relativo de dicho tema, identificándose así qué temas contribuyen a distinguir cada partido de los demás.

Más recientemente se han desarrollado los biplots con más de una composición (Ferrer-Rosell et al., 2020c; Filzmoser et al., 2018; Kynčlová et al., 2016) que permiten complementar la interpretación tradicional que se ha explicado dentro de una misma composición, con la relación entre composiciones. El presente artículo constituye el primer análisis de contenido composicional en usar esta herramienta.

Dadas las log-ratios centradas de una segunda composición  $\mathbf{y}$  (en nuestro caso el número de veces que cada una de los  $S$  sentimientos se expresa en los tweets de cada partido):

$$clr(\mathbf{y}) = \frac{S-1}{S} \left( \ln \left( \frac{y_1}{\sqrt[S]{y_2 y_3 \dots y_S}} \right) \ln \left( \frac{y_2}{\sqrt[S]{y_1 y_3 \dots y_S}} \right) \dots \ln \left( \frac{y_S}{\sqrt[S]{y_1 y_2 \dots y_{S-1}}} \right) \right) \quad (4)$$

el biplot de dos composiciones se construye a partir de un análisis en componentes principales basado en la matriz de covarianzas entre las  $D+S$  log-ratios centradas. En este biplot, las distancias entre los extremos de los vectores de una misma composición, se interpretan del mismo modo que se ha descrito. Los partidos se pueden proyectar ortogonalmente sobre cualquier vector de cualquier composición. Adicionalmente, los cosenos de los ángulos entre dos vectores de distintas composiciones son aproximadamente iguales a las correlaciones de sus log-ratios centradas respectivas, y por lo tanto entre sus importancias relativas. Recordemos que a dos flechas paralelas del mismo sentido les corresponde un ángulo de cero grados, cuyo coseno es uno e indica la máxima correlación positiva; que a dos flechas paralelas del sentido contrario les corresponde un ángulo de 180 grados, cuyo coseno es menos uno e indica la máxima correlación negativa, y que a dos flechas perpendiculares les corresponde un ángulo de 90 grados, cuyo coseno es cero e indica nula correlación. Ángulos agudos indican pues correlación positiva y ángulos obtusos negativa.

Es aconsejable que  $D$  y  $S$  sean del mismo orden de magnitud, de otro modo una de las dos composiciones puede dominar el análisis lo que resultaría en una pobre calidad de la representación de la otra.

El reemplazo de los ceros se hizo con el paquete R "zCompositions" (Palarea-Albaladejo y Martín-Fernández, 2015) con el commando `multRep` con la opción por defecto que es el método llamado Bayesiano multiplicativo geométrico (GBM) de Martín-Fernández et al. (2015). El porcentaje de ceros, tal como se recomienda, era reducido, 6,1% en la composición de contenidos y 3,0% en la composición de sentimientos. Los otros análisis se hicieron con el programa CoDaPack (Thió-Henestrosa y Martín-Fernández, 2005).

Las ventajas de usar el análisis de datos composicionales, los biplots y la transformación por log-ratios centrada ya han sido tratadas tanto con carácter general (Aitchison, 1986; Van den Boogaart y Tolosana-Delgado, 2013; Filzmoser et al., 2018; Greenacre, 2018; Pawlowsky-Glahn et al., 2015) como en el caso particular del estudio de la comunicación política (Blasco-Duatis et al. 2018a; 2018b; 2018c; 2019).

**Tabla 1: Ejemplo de las diferencias entre los enfoques tradicionales de frecuencias relativas (arriba) y rankings (centro), con el enfoque de datos composicionales con log-ratios (abajo).**

	MP	P	NEUT	N	MN	NING
Datos originales sobre 100						
Bildu	46.3	10.7	7.0	8.4	22.0	5.6
PSOE	60.7	10.4	7.1	3.0	17.5	1.3
UP	54.2	9.5	6.7	2.8	26.3	0.6
Rankings						
Bildu	6.0	4.0	2.0	3.0	5.0	1.0
PSOE	6.0	4.0	3.0	2.0	5.0	1.0
UP	6.0	4.0	3.0	2.0	5.0	1.0
Log-ratios centradas						
Bildu	1.3	-0.1	-0.6	-0.4	0.6	-0.8
PSOE	2.0	0.2	-0.1	-1.0	0.8	-1.8
UP	2.0	0.3	-0.1	-0.9	1.3	-2.6

Fuente: Elaboración propia.

Resumimos aquí algunas de las ideas más importantes en torno a un pequeño ejemplo. Comparado con los enfoques tradicionales de usar los datos originales referidos a un total común, y transformar los datos por rankings, el enfoque presentado aquí es el que tiene en cuenta las diferencias en las importancias relativas de los temas y los sentimientos. La tabla 1 muestra un fragmento de la composición de sentimientos para tres partidos. La diferencia más alta en valor absoluto la encontramos entre Bildu y el

PSOE en lo que respecta al sentimiento MP. Aún así, la diferencia en términos relativos entre 60,7 y 46,3 es solo del 31%. La diferencia más alta en términos relativos la encontramos entre Bildu y UP con respecto al sentimiento NING. La diferencia en términos relativos entre 5,6 y 0,6 es del 833%.

Sobre los datos originales, la distancia euclídea Bildu-PSOE es 16,6 y la distancia Bildu-UP 11,8, teniendo en cuenta las mayores diferencias en valor absoluto. Sobre rankings, las distancias Bildu-PSOE y Bildu-UP son ambas 1.4 porque aunque el PSOE y UP no son idénticos en sus valores, sí lo son en sus rankings. Sobre log-ratios centradas, la distancia Bildu-PSOE es 1,5 y la distancia Bildu-UP es 2,2, teniendo en cuenta las mayores diferencias en términos relativos, que constituye el objetivo del análisis.

Otro aspecto a tener en cuenta son las correlaciones espurias debido al hecho que las filas de la Tabla 1 siempre tienen suma fija. En términos relativos una parte solo puede aumentar si al menos alguna otra parte se reduce, lo cual lleva a correlaciones mayoritariamente negativas y no interpretables. De las 15 correlaciones posibles entre pares de sentimientos, 7 son negativas sobre los datos originales, 11 son negativas sobre los rankings y 12 son negativas sobre las log-ratios centradas. Por este motivo, el análisis composicional prescinde de las correlaciones entre las partes de una misma composición y las substituye por las varianzas del logaritmo del cociente entre pares de partes (2), en que no padecen este problema. Las correlaciones sí son válidas entre partes de composiciones diferentes y son las únicas que utilizamos en este artículo.

### 3. Resultados

Según las medias geométricas expresadas sobre un total igual a 100, los centros de las composiciones muestran que los temas más recurrentes son, en este orden: política territorial (21,2%), economía-empleo (17,6%), corrupción (16,1%), proceso catalán (15,9%), regeneración (9,3%), políticas sociales-sanidad (6,3%), educación-cultura (5,6%), género-igualdad (4,2%) y política europea (3,9%); y los sentimientos más frecuentemente expresados son, en este orden: MP (40,1%), P (19,2%), MN (16,4%), N (11,9%), NEUT (9,8%), y NING (2,1%).

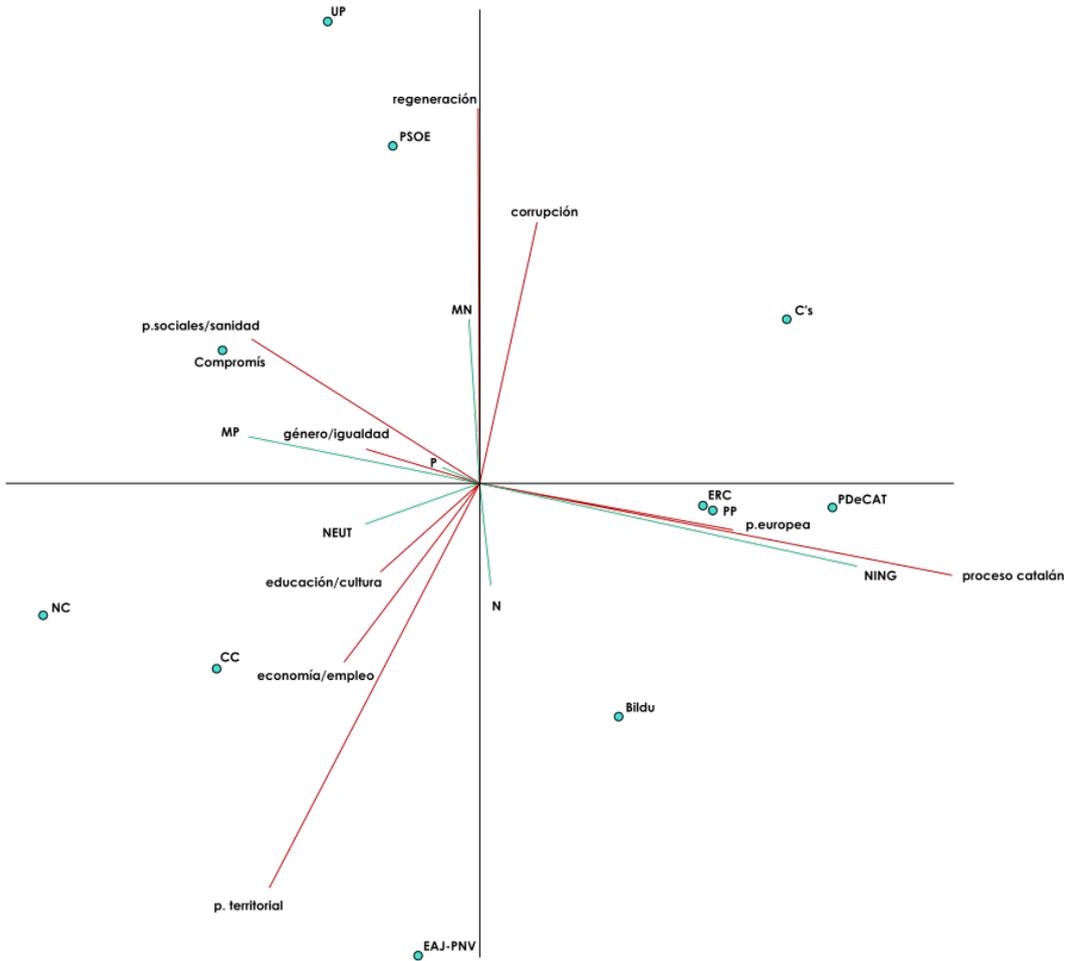
La varianza explicada por las dos primeras dimensiones del análisis en componentes principales es de 55%. Una tercera dimensión permitiría alcanzar el 71% a costa de renunciar a la visualización en un gráfico de dos dimensiones. Como en cualquier análisis en componentes principales, la calidad de la representación de cada vector depende de su porcentaje individual de varianza explicada sobre las dos primeras dimensiones, del cual el mencionado 55% es solo un promedio (Dauinis-i-Estadella, Thió-i-Fernández-de-Henestrosa y Mateu-i-Figueras, 2011). En este caso, los contenidos peor representados son género-igualdad (24,6%) y educación-cultura (21,4%). Los sentimientos peor representados son P (3,4%) y N (16,7%). Sus extremos y direcciones en el gráfico deben interpretarse con suma cautela. Por el contrario, están representados de manera especialmente fiable el proceso catalán (72,6%), la política territorial (77,5%), la corrupción (60,8%), las políticas sociales y de sanidad (66,9%), la regeneración (67,6%), los sentimientos MP (56,3%) y la ausencia de sentimientos-NING (74,7%).

Según las proximidades de los extremos de los vectores-tema, las parejas de temas que tienden a moverse juntos en los distintos partidos políticos son regeneración con corrupción, política europea con proceso catalán, economía y empleo con políticas sociales y sanidad, y economía y empleo con política territorial.

Según las proximidades de los extremos de los vectores-sentimiento, los sentimientos que tienden a moverse juntos en los distintos partidos políticos son los centrales, es decir, P, NEUT y N, aunque como hemos mencionado, algunos están representados en el biplot con poca calidad. La oposición más destacada se encuentra entre NING y el conjunto de todos los demás.

Según las proyecciones ortogonales sobre los vectores representados con mayor calidad, las asociaciones más notables entre partidos y temas son Cs, ERC, PDeCAT y PP con el proceso catalán; UP y PSOE con regeneración y con corrupción; y NC, CC, Bildu y EAJ-PNV con economía-empleo y con política territorial. Las ausencias más notables de temas generalmente populares en partidos concretos son corrupción en NC, CC, Bildu y EAJ-PNV; proceso catalán en NC, UP y Compromís; política territorial en C's y UP; y economía-empleo en PSOE, UP y C's.

**Figura 1: Biplot de covarianza de los temas top9, los sentimientos y los partidos políticos en la Moción de Censura al gobierno español de 2018.**



Fuente: Elaboración propia.

Según las proyecciones ortogonales, las asociaciones más notables entre partidos y sentimientos son C's, ERC, PDeCat, Bildu y PP con NING; Compromís, UP, C's y PSOE con MN; y NC, UP, Compromís, PSOE y CC con MP. Las ausencias más notables de sentimientos en partidos son MP en C's, ERC, PDeCAT y PP; y MN en NC, CC, Bildu y EAJ-PNV.

Como hemos indicado, en este artículo presentamos por primera vez un biplot de dos composiciones en el análisis de contenido de la comunicación política. Como novedad, este enfoque abre la puerta a interpretar como correlaciones los ángulos entre partes de una composición y partes de otra, y en nuestro caso responder a la pregunta respecto de qué sentimientos tienden a expresarse en relación con qué temas por parte de los partidos en su comunicación por Twitter. Los ángulos agudos entre vectores muestran las siguientes correlaciones positivas notables entre temas y sentimientos: MP con políticas sociales-sanidad; NING con el proceso catalán y la política europea; y MN con regeneración y corrupción. Los ángulos obtusos entre vectores muestran las siguientes correlaciones negativas notables entre temas y sentimientos, MP con el proceso catalán y con la política europea; MN con política territorial; NEUT con corrupción, y NING con políticas sociales-sanidad. Los ángulos casi rectos indican ausencia de relación. Por ejemplo, expresar sentimientos MP es independiente de que el tema se refiera o no se refiera a corrupción o regeneración, expresar sentimientos MN es independiente de que el tema se refiera o no a política europea.

#### 4. Discusión

Como apuntábamos anteriormente, son varios los estudios que han analizado los datos de transmisión de Twitter como una fuente para identificar noticias de actualidad y eventos del mundo real (Thelwall, Buckley y Paltoglou, 2011; Broersma y Graham, 2012; Neuman et al., 2014; Guo y Vargo, 2015; Calvo y Arugufete, 2020; Lee y Xu, 2018), concluyendo que las tendencias en Twitter suelen ser los eventos más importantes del día y se pueden utilizar para predecir los titulares de las noticias. En nuestro estudio detectamos que los temas que centraron el debate del conjunto de partidos fueron, esencialmente, aquellos que impulsaron al PSOE a liderar la Moción de Censura al gobierno de 2018. Es así como los asuntos relacionados con la política territorial, la economía/empleo, la corrupción y el proceso catalán fueron el eje temático central de la conversación de los partidos, con aproximadamente el 70% de cobertura temática. De este modo, podemos confirmar que el encuadre planteado por parte del PSOE para la discusión sobre la Moción de Censura fue, a su vez, el marco que centró la discusión de los partidos políticos en Twitter.

A nivel del sentimiento de los partidos (agenda de atributos), destacan los mensajes de carácter positivo o muy positivo (cerca del 60%), claramente por encima de aquellos negativos o muy negativos (poco más de un 28%). Una posible explicación a estos datos serían los supuestos apoyos que en ese momento tenía el PSOE para sacar adelante la Moción, ya que contaba con más partidos potencialmente a favor de ella (UP, ERC, PDeCAT, PNV, Compromís, Bildu, NC), que en contra (PP, C's, UPN, FA). Con todo, y como apuntábamos en el anterior capítulo de resultados, cabe destacar que los sentimientos más representativos entre los partidos políticos (que suelen ir asociados y, por lo tanto, a moverse juntos) son aquellos centrales (P, NEUT y N).

En nuestro estudio sobre los temas que centraron la agenda también resulta interesante ver la asociación que se establece entre determinados temas. Es así como podemos identificar con claridad parejas de temas que son proporcionales entre sí, es decir que cuanto más se habla de un tema, más se habla del otro. Este fenómeno es visible especialmente con los binomios: regeneración y corrupción, eje central del debate de la Moción; política europea y proceso catalán, claramente asociado a la internacionalización del conflicto político catalán; economía/empleo y políticas sociales/sanidad, centradas en revertir las políticas de austeridad del ejecutivo de Mariano Rajoy; así como con economía/empleo y política territorial, asociadas a las desigualdades territoriales y la falta de inversión autonómica.

En el vínculo que se establece entre determinados temas y partidos políticos, es interesante comentar las siguientes asociaciones: C's y PP vs. ERC y PDeCAT con el proceso catalán, donde los dos primeros partidos mantienen una posición claramente contraria a los segundos, pero son altamente equivalentes en cuanto a la dedicación sobre el tema; NC, CC, Bildu y EAJ-PNV con economía/empleo y política territorial, una asociación claramente vinculada a las tradicionales demandas territoriales de los partidos nacionalistas sobre la inversión en la comunidad que representan; y finalmente UP y PSOE con regeneración y corrupción, que no solamente resultó ser el eje programático central del PSOE en la defensa de la Moción al gobierno, sino que también fue uno de los elementos de oposición en los años de gobierno del ejecutivo de Mariano Rajoy.

Las correlaciones notables entre sentimientos y temas también aportan algunos resultados interesantes a comentar en el desarrollo del segundo nivel de agenda. Si bien es cierto que la percepción general sobre el tema del proceso catalán es que en los últimos años ha suscitado debates altamente polarizados en cuanto a los sentimientos derivados, en el caso que nos ocupa de los agenda-setters en Twitter, entre los partidos predominó un tono de neutralidad durante la Moción. Por el contrario, temas como las políticas sociales-sanidad (MP) o la regeneración y corrupción (MN), mantuvieron su lógica –en lo establecido por el debate político-social del país– al captar sentimientos extremos. Este fenómeno se refuerza con determinadas ausencias de relación como por ejemplo NEUT con corrupción.

Precisamente en el tema corrupción (así como en el de regeneración, que como hemos comentado aparecen estrechamente relacionados) es interesante detenerse en el análisis de sus componentes, ya que en él podemos ver representado el leitmotiv que impulsa la Moción por parte del PSOE. Los partidos que se asocian de forma más notable a este tema y con un sentimiento MN son UP y PSOE, seguidos de C's y Compromís. Por otro lado, el PSOE y algunos partidos que habrían de apoyar la moción (UP y Compromís) también pretendían que el gobierno substitutorio del censurado fuera más sensible a las políticas sociales, de sanidad, de género y de igualdad, temas sobre los cuales se expresan con un sentimiento MP. Por el contrario, los más alejados a esta alianza son los partidos nacionalistas como Bildu, CC, NC y PNV (centrados especialmente en los temas de política territorial y con un sentimiento más N que MN sobre la corrupción). Por su parte, el PP tiene una posición muy cercana (tanto a nivel temático como de sentimiento) a las formaciones independentistas catalanas de ERC y PDeCAT. En este sentido, la corrupción no es uno de los temas prioritarios para estos tres partidos, sino que se centran, y desde NING

sentimiento, en los temas del proceso catalán y la política europea. Todo ello nos lleva a poder constatar que si bien la Moción se presentó para suspender de funciones a un gobierno, el partido del cual acababa de ser condenado por corrupción, dicho partido de gobierno (el PP) se escudó en el *frame* del proceso catalán para evitar el tema central de la Moción de Censura.

Estos resultados responden a nuestro propósito inicial de alcanzar una visualización de datos composicional que permitiera representar la agenda de atributos, es decir, avanzando la representación de la agenda y los *agenda-setters* en calidad de partidos políticos, hacia la visualización de los sentimientos vinculados a los mensajes publicados en Twitter. Con ello, no solamente hemos podido mapear qué partidos hacen hincapié en qué contenidos o qué temas o partidos son más o menos similares entre sí, sino que hemos incorporado a este estudio composicional sobre la agenda el análisis de sentimiento. En este sentido, consideramos probada la capacidad del método para la visualización y estudio del sentimiento en el escenario de la agenda de atributos, siendo especialmente relevante por su capacidad de priorizar la relatividad de los elementos en estudio.

## 5. Conclusiones

En este artículo presentamos un análisis del sentimiento de la discusión en Twitter, por parte de los principales partidos políticos, sobre la Moción de Censura al gobierno español de 2018. Extrajimos y analizamos composicionalmente 2824 tweets, así como sus sentimientos asociados, que fueron publicados por los partidos políticos con representación en el Congreso de los Diputados. En particular desarrollamos, documentamos y aplicamos un enfoque metodológico innovador en el campo del análisis de sentimiento sobre comunicación política, sistematizando su aplicación para lograr la visualización de estos sentimientos en el relato de los partidos políticos, y sobre las diversas temáticas de la agenda expuestas en Twitter.

Para el caso sobre la Moción de Censura al gobierno español de 2018, encontramos que:

- El 70% de la conversación de los partidos en Twitter sobre la Moción de Censura estuvo centrada en solamente cuatro temas (política territorial, economía/empleo, corrupción y proceso catalán). Estos temas, que sirvieron al PSOE como líneas "rojas" para impulsar dicha Moción, fueron el encuadre (*frame*) de la discusión en Twitter por parte de los partidos políticos.
- El sentimiento predominante en los mensajes de los partidos fue más positivo (60%) que negativo (28%), una atmósfera claramente asociada al hecho de ser más los partidos potencialmente a favor de la Moción y, por ende, al deseo de que esta terminara saliendo adelante.
- Existen determinados grupos de temas más asociados entre sí que otros, y que a su vez también se asocian más con determinados grupos de partidos. Destacan, en esta línea: C's, PP, ERC i PDeCAT con el proceso catalán; NC, CC, Bildu y EAJ-PNV con economía/empleo y política territorial; y finalmente UP y PSOE con regeneración y corrupción.
- Los sentimientos repartidos más equitativamente en los mensajes de los partidos tienden más al centralismo (P, NEUT y NEG), que a los extremismos (MP, MN).
- Es destacable que el tema "proceso catalán" que ha suscitado grandes polaridades de sentimientos en el debate público, alcanzara un claro tono de neutralidad en el caso del debate de los partidos en Twitter para la Moción de Censura.
- Respecto al punto anterior, el PP tiene una posición muy cercana (tanto a nivel temático como de sentimiento) a las formaciones independentistas catalanas de ERC y PDeCAT. En este sentido, la corrupción no es el *frame* deseado por estos tres partidos, sino que desplazan la agenda a los temas del proceso catalán y la política europea.

Asimismo, al aplicar CoDa al análisis de sentimiento de la agenda en Twitter hemos podido probar la viabilidad del método para su visualización y estudio, enfocando el análisis en la importancia relativa de cada tema de la agenda (contenido), sobre cada partido (*agenda-setter*) y para cada sentimiento (agenda de atributos).

En este sentido, la aplicación del método al presente estudio de caso ha permitido visualizar y estudiar la agenda de segundo nivel (o agenda de atributos). A través del análisis de sentimiento hemos podido constatar cómo los principales atributos (MP, P, NEUT, N o MN) se vinculan a los temas que conformaron la agenda y para cada partido. Así, y en la línea de lo expuesto en las posibilidades metodológicas, la aplicación de CoDa sobre dos composiciones –en lo que para el caso que nos ocupa es el segundo nivel de la agenda– ha constatado su valiosa contribución para visualizar y aprender de forma composicional

las relaciones entre agenda-setters/partidos (puntos), temas de la agenda (vectores) y sentimientos (ángulos). En este último particular, el de los sentimientos analizado a través de los ángulos, constatamos que aquellos ángulos agudos indican una correlación positiva entre el tema y el sentimiento, y los ángulos obtusos negativa.

Cabe destacar, y como limitación que hemos descrito en el desarrollo metodológico, que los estudios en este campo deben centrarse, de forma prioritaria, en los principales temas de la agenda, ya que aquellos temas poco o nada recurrentes incrementan los ceros y, con ello, dificultan el cálculo de los log-ratios. En efecto, la imputación Bayesiana sobre estos ceros de recuento deja de ser fiable cuando el porcentaje de ceros es elevado (Martin-Fernandez et al., 2015) y otros métodos como el análisis de correspondencias pasan a ser preferibles a los métodos composicionales (Greenacre, 2018). Asimismo, es altamente aconsejable que las dos composiciones estén detalladas en un número de partes del mismo orden de magnitud, ya que de otro modo una de las dos composiciones dominaría el análisis y distorsionaría la visualización.

Esta investigación representa la primera experiencia probada de un estudio CoDa de dos composiciones de la agenda (estudio de segundo nivel, o atributos, a través del sentimiento) en el campo de la comunicación política. La metodología CoDa va mucho más allá de la visualización de datos. Por ejemplo, un análisis de conglomerados composicional puede clasificar tanto temas como agenda-setters cuando exista un gran número de ellos, o cruzar la clasificación de agenda-setters por temas con la clasificación por sentimientos. En caso de disponerse de datos longitudinales, los agenda-setters pueden aparecer en el biplot como trayectorias en función de la evolución de sus temas predominantes, o puede modelarse estadísticamente la agenda en un momento dado a partir de la agenda en un instante anterior. El repertorio de métodos composicionales se amplía rápidamente a medida que la mayoría de métodos estadísticos pasan a tener su equivalente composicional (Van den Boogaart y Tolosana-Delgado, 2013; Egozcue y Pawlowsky-Glahn, 2016; Filzmoser et al., 2018; Greenacre, 2018; Pawlowsky-Glahn et al., 2015), y con él las oportunidades de investigación centrada en la importancia relativa de temas, emisores y/o atributos.

## 6. Referencias bibliográficas

- [1] Ahmadian, S.; Azarshahi, S. & Paulhus, D.L. (2017). Explaining Donald Trump via communication style: Grandiosity, informality, and dynamism. *Personality and Individual Differences*, 107, 49–53. <http://doi.org/f9pgfd>
- [2] Aitchison, J. (1986). *The statistical analysis of compositional data*. Monographs on statistics and applied probability. London: Chapman and Hall.
- [3] Aitchison, J. & Greenacre, M. (2002). Biplots of compositional data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51(4), 375–392. <http://doi.org/bck69d>
- [4] Berelson, B. (1952). *Content analysis in communication research*. New York: Hafner.
- [5] Blasco-Duatis, M.; Coenders, G. & Sáez, M. (2018a). Compositional visualization of intermedia agenda setting by the main media groups and political parties in the Spanish 2015 General Elections. *Revista Latina de Comunicación Social*, 1(73), 264–292. <http://doi.org/dpgb>
- [6] Blasco-Duatis, M.; Coenders, G. y Sáez, M. (2018b). Representación composicional (CoDa) de la agenda-setting de los opinantes sobre política en la prensa escrita española durante las Elecciones Generales de 2015. *Estudios sobre el Mensaje Periodístico*, 24(2), 1121–1146. <http://doi.org/dpgc>
- [7] Blasco-Duatis, M.; Sáez-Zafra, M. & Fernández-García, N. (2018c). Compositional representation (CoDa) of the agenda-setting of the opinion makers on politics in the main Spanish media groups in the 2015 General Elections. *Communication & Society*, 31(2), 1–23. <https://bit.ly/2XIMLJf>
- [8] Blasco-Duatis, M.; Coenders, G.; Sáez, M.; Fernández-García, N. & Cunha, I. (2019). Mapping the agenda-setting theory, priming and the spiral of silence in twitter accounts of political parties. *International Journal of Web Based Communities*, 15(1), 4–24. <http://doi.org/dpgd>
- [9] Broersma, M. & Graham, T. (2012). Social media as beat: Tweets as a news source during the 2010 British and Dutch elections. *Journalism Practice*, 6(3), 403–419. <http://doi.org/gdqc56>
- [10] Van den Boogaart, K.G. & Tolosana-Delgado, R. (2013). *Analyzing Compositional Data with R*. Berlin: Springer. <http://doi.org/dxxw>
- [11] Calderon, N. A.; Fisher, B.; Hemsley, J.; Ceskavich, B.; Jansen, G.; Marciano, R. & Lemieux, V.L. (2015). Mixed-initiative social media analytics at the World Bank: Observations of citizen sentiment in Twitter data to explore "trust" of political actors and state institutions and its relationship to social protest. *2015 IEEE International Conference on Big Data* (pp. 1678–1687). <http://doi.org/dpgf>

- [12] Calvo, E. y Aruguete, N. (2020). *Fake news, burbujas, trolls y otros encantos. Cómo funcionan (para bien y para mal) las redes sociales*. Buenos Aires: Siglo veintiuno editores.
- [13] Conway, B.; Kenski, K. & Wang, D. (2015). The rise of Twitter in the political campaign: Searching for intermedia agenda-setting effects in the presidential primary. *Journal of Computer-Mediated Communication*, 20(4), 363–380. <http://doi.org/f7kr8b>
- [14] Daunis-i-Estadella, J.; Thió-i-Fernández-de-Henestroza, S. & Mateu-i-Figueras, G. (2011). Two more things about compositional biplots: quality of projection and inclusion of supplementary elements. In J.J. Egozcue; R. Tolosana-Delgado & M.I. Ortego (Eds.), *Proceedings of the 4th International Workshop on Compositional Data Analysis* (pp. 1-14). <http://bit.ly/2W0azYK>
- [15] Egozcue, J. J. & Pawlowsky-Glahn, V. (2016). What are compositional data and how should they be analyzed? *Boletín de Estadística e Investigación Operativa*, 32(1), 5-29. <https://bit.ly/2BGaP6N>
- [16] Fan, W. & Gordon, M.D. (2014). The Power of Social Media Analytics. *Communication of the ACM*, 57(6), 74–81. <http://doi.org/dpgg>
- [17] Ferrer-Rosell B. & Marine-Roig, E. (2020). Projected versus perceived destination image. *Tourism Analysis*, 25. <http://doi.org/dpgh>
- [18] Ferrer-Rosell, B.; Martin-Fuentes, E. & Marine-Roig, E. (2019). Do hotels talk on Facebook about themselves or about their destinations? In J. Pesonen & J. Neidhardt (Eds.), *Information and Communication Technologies in Tourism 2019* (pp. 344-356). Cham: Springer. <http://doi.org/dxxz>
- [19] Ferrer-Rosell, B.; Coenders, G. & Martin-Fuentes, E. (2020a). Compositional data analysis in e-tourism research. What TripAdvisor reviews complain about hotels in Barcelona? In X. Zheng; M. Fuchs; U. Gretzel & W. Höpken (Eds.) *Handbook of E-Tourism*. Deutschland: Springer Nature.
- [20] Ferrer-Rosell, B.; Martin-Fuentes, E. & Marine-Roig, E. (2020b). Diverse and emotional: Facebook content strategy by Spanish hotels. *Journal of Information Technology & Tourism*, 22(4), 53–74. <http://doi.org/dw85>
- [21] Ferrer-Rosell, B.; Martin-Fuentes, E.; Vives-Mestres, M. & Coenders, G. (2020c). When size does not matter: compositional data analysis in marketing research. In R. Nunkoo, V. Teeroovengadum & C. Ringle (Eds.), *Handbook of Research Methods for Marketing Management*, in press.
- [22] Filzmoser, P.; Hron, K. & Templ, M. (2018). *Applied Compositional data Analysis with Worked Examples in R*. New York: Springer. <http://doi.org/dxx2>
- [23] Graham, T.; Jackson, D. & Broersma, M. (2016). New platform, old habits? Candidates use of Twitter during the 2010 British and Dutch general election campaigns. *New Media & Society*, 18(5), 765–783. <http://doi.org/cv2q>
- [24] Greenacre, M. (2018). *Compositional data analysis in practice*. New York: Chapman and Hall/CRC press. <http://doi.org/dxx3>
- [25] Golbeck, J.; Grimes, J.M. & Rogers, A. (2010). Twitter use by the U.S. Congress. *Journal of the American Society for Information Science and Technology*, 61(8), 1612–1621. <http://doi.org/czr5b8>
- [26] Guo, L.; Rohde, J. A. & Wu, H. D. (2018). Who is responsible for Twitter's echo chamber problem? Evidence from 2016 U.S. election networks. *Information. Communication & Society*, 21(12), 234–251. <http://doi.org/gfdbqn>
- [27] Guo, L. & Vargo, C. (2015). The power of message networks: A big-data analysis of the network agenda setting model and issue ownership. *Mass Communication and Society*, 18(5), 557-576. <http://doi.org/dxb2>
- [28] Jungherr, A. (2014). Twitter in politics: A comprehensive literature review. *Social Science Research Network*. <http://doi.org/dxx4>
- [29] Jungheer, A. (2016). Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics*, 13(1), 72–91. <http://doi.org/gf6bsb>
- [30] Kayser, V. & Bierwisch, A. (2016). Using Twitter for Foresight: An Opportunity? *Futures*, 84(Part A), 50–63. <http://doi.org/f9jc6z>
- [31] Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Sage: Thousand Oaks.
- [32] Kušen, E.; Strembeck, M.; Cascavilla, G. & Conti, M. (2017). On the influence of emotional valence shifts on the spread of information in social networks. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Net-works Analysis and Mining 2017* (pp. 321–324). <http://doi.org/dpgj>
- [33] Kynčlová, P.; Filzmoser, P. & Hron, K. (2016). Compositional biplots including external non-compositional variables. *Statistics*, 50(5), 1132–1148. <http://doi.org/dpgk>

- [34] Lee, J. & Xu, W. (2018). The more attacks, the more retweets: Trump's and Clinton's agenda setting on Twitter. *Public Relations Review*, 44(2), 201-213. <http://doi.org/dxb3>
- [35] Marine-Roig, E. & Ferrer-Rosell, B. (2018). Measuring the gap between projected and perceived destination images of Catalonia using compositional analysis. *Tourism Management*, 68, 236-249. <http://doi.org/dpgm>
- [36] Martín-Fernández, J.A.; Hron, K.; Templ, M.; Filzmoser, P. & Palarea-Albaladejo, J. (2015). Bayesian-multiplicative treatment of count zeros in compositional data sets. *Statistical Modelling*, 15(2), 134-158. <http://doi.org/dpgn>
- [37] McCombs, M. & Shaw, D. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, 36(2), 176-187. <http://doi.org/dnd7p4>
- [38] McCombs, M. (2004). *Setting the agenda*. Cambridge: Polity Press.
- [39] Meraz, S. (2009). Is there an elite hold? Traditional media to social media agenda setting influence in blog networks. *Journal of Computer Mediated Communication*, 14(3), 682-707. <http://doi.org/d79bmf>
- [40] Moreno-Ortiz, A. & Pérez Hernández, C. (2013). Lexicon-Based Sentiment Analysis of Twitter Messages in Spanish. *Procesamiento del Lenguaje Natural*, 50, 93-100. <https://bit.ly/2XLItkq>
- [41] Neuman, W.; Guggenheim, L.; Jang, S. & Bae, S.Y. (2014). The dynamics of public attention: Agenda-setting theory meets Big Data. *Journal of Communication*, 64(2), 193-214. <http://doi.org/f5x3nb>
- [42] Palarea-Albaladejo, J. & Martín-Fernández, J.A. (2015). zCompositions—R package for multivariate imputation of left-censored data under a compositional approach. *Chemometrics and Intelligent Laboratory Systems*, 143, 85-96. <http://doi.org/f696n3>
- [43] Pang, B. & Lee, L. (2004). A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, (pp. 271-278). Barcelona: Association for Computational Linguistics. <http://doi.org/cdz7k3>
- [44] Parmelee, J. (2013). Political journalists and Twitter: Influences on norms and practices. *Journal of Media Practice*, 14(4), 291-305. <http://doi.org/ggb6jm>
- [45] Pawlowsky-Glahn, V.; Egozcue, J.J. & Tolosana-Delgado, R. (2015). *Modeling and Analysis of Compositional Data*. Chichester: Wiley. <http://doi.org/dxx5>
- [46] Scolari, C. A.; Lugo, N. y Masanet, M. (2019). Educación transmedia: de los contenidos generados por los usuarios a los contenidos generados por los estudiantes. *Revista latina de comunicación social*, 1(74), 116-132. <http://doi.org/dpgp>
- [47] Statistic Brain (2018). STATS. Twitter company statistics. <http://bit.ly/3cJwoRY>
- [48] Thelwall, M.; Buckley, K. & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, 62(2), 406-418. <http://doi.org/bb3n8g>
- [49] Thió-Henestrosa, S. & Martín-Fernández, J.A. (2005). Dealing with compositional data: The freeware CoDaPack. *Mathematical Geology*, 37(7), 773-793. <http://doi.org/fv8tj2>
- [50] Tumasjan, A.; Sprenger, T. O.; Sandner, P. G. & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *International Conference on Weblogs and Social Media*, 10, 178-185. <http://bit.ly/3cPuvn9>
- [51] Turney, P.D. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, 417-424. <http://doi.org/fphr6f>
- [52] Wallsten, K. (2014). Microblogging and the news: political elites and the ultimate retweet. In K. Wallsten (Ed.), *Political campaigning in the information age* (pp. 128-147). <http://doi.org/dpgq>
- [53] Wimmer, R. & Dominick, J. (2006). *Mass Media Research: an introduction*. Wadsworth: Thomson.
- [54] Yousuf Al-Aama, A. (2015). The use of Twitter to promote e-participation: connecting government and people. *International Journal of Web Based Communities*, 11(1), 73-96. <http://doi.org/dpgr>

## Agradecimientos

El trabajo ha sido financiado con las ayudas de la Generalitat de Catalunya (COSDA: 2017SGR656), del Ministerio de Ciencia, Innovación y Universidades/FEDER (CODAMET: RTI2018-095518-B-C21), y del Ministerio de Sanidad, Consumo y Bienestar Social (CIBER: CB06/02/1002).



Revista MEDITERRÁNEA de Comunicación  
MEDITERRANEAN Journal of Communication