

POLARIZACIÓN Y SIMILITUD  
EN LAS REPRESENTACIONES DE MENSAJES OFICIALES  
EN MEDIOS ANTE LA PANDEMIA DE COVID-19\*

POLARIZATION AND SIMILARITY  
IN MEDIA AND MEXICAN GOVERNMENT IN FACING  
THE COVID-19 PANDEMIA



*Juan Enrique HUERTA WONG\*\**  
*Alejandro Miguel CASTAÑEDA VALENCIA\*\*\**  
*Francisco Javier MANZANO MORA\*\*\*\**

RESUMEN. ¿Cuál fue el nivel de información disponible a la ciudadanía un día cualquiera de 2020 respecto a la pandemia de COVID-19? Se analiza aquí el contenido en medios de la Ciudad de México y conferencias del gobierno en la pandemia de COVID, con tres técnicas de aprendizaje automático, a dos millones de caracteres de conferencias, y 28,127 notas y columnas de impresos y noticieros. Los principales resultados indican que hubo similitud en noticias y polarización en opinión. Esto es, la superficie narrativa de los medios consideró seguir la información que el Consejo General de Salubridad proporcionó con el objetivo de salvar más vidas. En cambio, la superficie de comentarios en los medios muestra una tendencia a discrepar de la opinión del conjunto de profesionales de la medicina cuya función principal

---

\* La investigación de la cual emerge este proyecto fue parcialmente apoyada por la Organización Panamericana de la Salud, representación en México.

\*\* Doctor en políticas comparadas del bienestar (UTexasA / UANL); postdoctorados en sociología de la pobreza y la desigualdad (Stanford), en Desarrollo (McGill), en comunicación política (Tecnológico de Monterrey); Maestría en comunicación (Tecnológico de Monterrey, graduado con honores), correo electrónico: [jehuerta848@gmail.com](mailto:jehuerta848@gmail.com).

\*\*\* Doctor y maestro en economía (BUAP), Benemérita Universidad Autónoma de Puebla, correo electrónico: [alcastan.eco@gmail.com](mailto:alcastan.eco@gmail.com).

\*\*\*\* Doctor en economía (CIDE), correo electrónico: [alcastan.eco@gmail.com](mailto:alcastan.eco@gmail.com).

Fecha de recepción: 18 de noviembre de 2022.

Fecha de dictamen: 6 de enero de 2023.

*Revista Estudios en Derecho a la Información,*  
núm. 16, julio-diciembre de 2023, pp. 51-77.

Esta obra está bajo una *Licencia Creative Commons*  
*Atribución-NoComercial-SinDerivadas 4.0 Internacional*, IJJ-UNAM.

HUERTA WONG / CASTAÑEDA VALENCIA / MANZANO MORA

fue salvaguardar vidas. Se discuten implicaciones de la responsabilidad social de los medios, en el contexto de una emergencia sanitaria.

PALABRAS CLAVE. COVID-19, medios de comunicación, comunicación y medios, estudios de medios de comunicación, métodos, aprendizaje máquina, polarización de la información.

52

●  
○  
●

*ABSTRACT. By using a triple machine learning approach, we conduct a content analysis on dos million bites of Mexican government information, and 28, 127 news and opinion columns published in 2020, relating to COVID pandemia. Findings report similarity in news and polarization in opinion columns. This means that news covered accurately Health General Council recommendations, but comments in media tended to distort what this group of professionals in Medicine provided with the objective of protect Mexicans' lives. Those findings bring media social responsibility and right to information to the further discussions of COVID-19, the most serious health crisis in Mexico history.*

*KEYWORDS. COVID-19, media, communication and media, mass media studies, methods, machine learning, information polarization.*

## I. INTRODUCCIÓN

¿Cuál fue el nivel de información disponible a la ciudadanía un día cualquiera de 2020 respecto a la pandemia de COVID-19? ¿Cuánto fue coincidente la agenda informativa de los medios de comunicación publicados en México respecto a la información emitida por la institución responsable del manejo de la pandemia de COVID-19? Del 28 de febrero al 28 de noviembre, México registró 104,873 fallecimientos por causa de COVID-19. Desde el inicio de la pandemia de COVID-19, la Organización Mundial de la Salud y la Organización Panamericana de la Salud exhortaron a los gobiernos nacionales a poner en marcha campañas de comunicación del riesgo (OMS, 2018). Consideraron que las estrategias de comunicación de los gobiernos y la participación de los medios de comunicación resultaban instrumentos cruciales para lograr que los ciudadanos accedieran a la mejor información posible, y aprendieran cómo

protegerse. En México, desde la confirmación del primer caso de COVID-19, el 28 de febrero de 2020, el gobierno federal, y más específicamente, el Consejo de Salubridad General, inició una estrategia de comunicación implementada por medio de conferencias de prensa sobre el estado de la contingencia sanitaria hasta el 30 de noviembre se efectuaron 271 conferencias.

El ambiente de alta incertidumbre que ha representado la pandemia se agudizó durante el primer año. Dado el escaso conocimiento acerca de cómo prevenir contagios, hospitalizaciones y fallecimientos, el acceso a información de calidad resultaba relevante. Se presenta aquí el análisis de los temas, la similitud y la polarización entre la comunicación del Gobierno de México, y los medios de comunicación de la Ciudad de México, específicamente, noticieros nocturnos y medios impresos. Es relevante al menos por dos razones. La primera, marcada por la OMS y la OPS, porque en el contexto de alta incertidumbre, como el marcado por la pandemia, ciudadanos mejor informados pueden actuar en conformidad con lo que expertos han previsto, lo cual puede contribuir a salvar sus vidas. La segunda razón es que de acuerdo con una encuesta internacional conducida por *The Reuters Institute for the Study of Journalism and Misinformation, Science and Media* (Kleis-Nielsen *et al.*, 2020) durante la situación de pandemia las organizaciones tradicionales gozan de una alta reputación como fuente válida de información, ya que poseen una ventaja en la confianza de la población de 33 puntos porcentuales con respecto a la información con origen en las redes sociales, 30 puntos porcentuales con respecto a noticieros en sitios de internet enfocados a la transmisión de videos, 35 puntos porcentuales en comparación con las noticias transmitidas por aplicaciones de mensajes, y 14 puntos porcentuales en relación con los motores de búsqueda (Kleis-Nielsen *et al.*, 2020).

En la presente investigación se plantean dos objetivos: *a*) identificar en México el nivel de información disponible a la ciudadanía informada un día cualquiera de 2020 respecto a la pandemia de COVID-19, y *b*) si la agenda informativa de los medios de comunicación publicados en México coincidió o discrepó respecto a la comunicación de riesgo emitida por la institución responsable del manejo de la pandemia de COVID-19. Ante ellos, la estrategia metodológica en este análisis del contenido de los medios tiene que ver con tres tipos de análisis de procesamiento de lenguaje natural: *i*) de temas; *ii*) de similitud, y *iii*) de sentimientos. En primer lugar, interesa identificar algunos tópicos claves que las autoridades sanitarias consideraron prioritarios durante el periodo de contingencia epidemiológica y que comunicó frecuentemente a través de conferencias diarias. El gobierno mexicano aseguró que se trataba de un esfuerzo único en el mundo para poner a disposición de la ciudadanía información que redujera el riesgo de contagios, hospitalizaciones y fallecimientos. Estos tópicos se comparan con las temáticas más frecuentes que abordaron



periódicos diarios y noticiarios televisivos. Además, se analiza la similitud con que es reproducida la información de la agenda pública desde su emisión en los foros gubernamentales, hasta su impresión en los medios de comunicación tradicionales.

54 El análisis de contenido es la técnica clásica de análisis de noticias. En el análisis de contenido clásico, conducido por seres humanos, un problema fundamental consistió en establecer pruebas de confiabilidad entre codificadores. Una solución consistió en utilizar procesos de doble ciego a las codificaciones, y reportar coeficientes de correlación entre los grupos. Ante este problema del análisis de contenido conducido por humanos, una solución consiste en conducir análisis por procesamiento automatizado. El aprendizaje automático o aprendizaje de máquina (*machine learning*) es una rama de la inteligencia artificial, cuyo objetivo es desarrollar modelos matemáticos y estadísticos que permitan que las computadoras aprendan una tarea específica (Bishop, 2006). Estas técnicas de análisis se han aplicado para analizar el contenido de noticias en diversos contextos. Raina (2013) realiza clasificación en tópicos y análisis de sentimientos en columnas periodísticas. Godbole, Srinivasaiah y Skienana (2018) documentan que el análisis de sentimientos tiene que ver con un triple proceso: *a*) construcción de algoritmos de diccionarios de sentimientos; *b*) la formulación del índice de sentimientos, y *c*) la validación estadística de estos índices. Los análisis de tópicos y de similitud son realizados a través de minería de textos. Es el proceso de extraer información no trivial sobre texto no estructurado, con base en el recuento de la ocurrencia y coocurrencia de términos que facilita la comparación entre tópicos (Blake, 2011; Gupta y Lehal, 2009). Se ha aplicado recientemente en el análisis de contenido (Isoaho *et al.*, 2019; Nowlin, 2016). La identificación de los temas relevantes se condujo mediante la técnica de análisis de tópicos. La comparación de semejanza entre la información original y la difundida, se realizó a través de un algoritmo que analizó el nivel de similitud entre los textos. La orientación morfológica se llevó a cabo por medio de análisis de sentimientos.

La estrategia metodológica es una aportación al análisis de contenido de noticias. Las tres técnicas de aprendizaje automático (*machine learning*) toman como referencia documentos estenográficos de las conferencias vespertinas establecidas por la Secretaría de Salud (SSA), y se contrastan con el contenido de tres noticiarios nocturnos de cadena nacional y once diarios de corte socio-político que se imprimen en la Ciudad de México. Se realiza un análisis empírico partiendo de las palabras como unidad de análisis. En total se analizan 270 conferencias informativas, conteniendo más de dos millones de caracteres, además de 28, 127 notas informativas y columnas de opinión publicadas en medios impresos o noticieros televisivos.

Los resultados muestran que en México, durante la contingencia sanitaria causada por la epidemia COVID-19, la cantidad de las noticias que abordaron la información emitida por las autoridades sanitarias decreció con el paso del tiempo, aún antes que decrecieran los riesgos de contagio, hospitalización y fallecimiento. Una vez superado el primer momento de contacto con el agente de riesgo, los contenidos destinaron más espacios a diversas posturas y actores. Se encontró similitud de media a baja en la reproducción de los temas. También se encontró que las noticias informativas y columnas de opinión discreparon entre sí, pues las noticias se acercaron más a la comunicación de riesgo emitida por las autoridades sanitarias, mientras que los comentarios discreparon frecuentemente. Se discute que ante la confusión que imperó durante el primer año de la pandemia, la ciudadanía encontró entonces voces discordantes ante lo cual tuvo que optar, en un momento de máxima tensión, si elegir el camino de la autoridad sanitaria o el de los medios de comunicación para informarse y, por tanto, protegerse.

55



## II. ANTECEDENTES

Una adecuada comunicación del riesgo por parte de los gobiernos es fundamental para evitar que se generen problemas ocasionados por información inadecuada, tales como confusión, desentendimiento, comportamientos irracionales y reducción del bienestar de la población. Sandman (2012) precisa que la comunicación de riesgos tiene dos actividades diferentes: alertar a la gente y tranquilizarla, es decir, comunicar el mensaje adecuado para que la población no mantenga un estado de apatía, pero lo suficientemente eficaz para tranquilizar a las personas que están excesivamente alarmadas. Para Abrams y Greenhawt (2020), la comunicación de riesgos debe mantener una relación constante con los medios de comunicación para disminuir el sensacionalismo, presentar una imagen honesta, y obtener la cooperación y comprensión del público.

Para una efectiva comunicación, Frewer (2020) sugiere que los gestores de riesgos deben tener en cuenta las preocupaciones y los valores de la sociedad, abordar las preocupaciones reales de las personas, desarrollar y mantener la confianza pública —a través de la transparencia institucional y participación del público en el proceso de toma de decisiones—, e incluso el reconocimiento de incertidumbre. Zang, Li y Chen (2020) señalan tres principios de la comunicación efectiva de riesgos: accesibilidad y transparencia de la información, comunicación temprana y frecuente, e implementación de un método estratégico para comunicar la incertidumbre. Agregan que una estrategia de comunicación se compone de dos dimensiones: comunicación interna y

comunicación externa. Desde un enfoque gubernamental, la comunicación interna se refiere al proceso de definición de tareas y responsabilidades entre autoridades, científicos y comunidad académica. La comunicación externa se refiere a la concientización y reconocimiento del impacto negativo del factor de riesgo, así como el establecimiento de comportamientos diferentes a los de un ambiente de normalidad, es decir, la comunicación externa está relacionada con el intercambio de información entre el gobierno y los diferentes públicos que conforman a la población en su conjunto. Adicionalmente, Zang, Li y Chen (2020: 5) destacan tres aspectos fundamentales en el proceso de comunicación de riesgos: la toma de decisiones del gobierno, la difusión de información del gobierno y la interpretación de la información.

56

- - 
  -
- La toma de decisiones del gobierno es indispensable durante las emergencias nacionales, para unificar y motivar a la población; también para promover la prevención, respuesta y recuperación de riesgos a la salud y daños graves (Kyum *et al.*, 2020). Respecto a la difusión de la información, Kyum, Kim y Kreps señalan que las autoridades deben comunicarse eficazmente en respuesta a las amenazas de salud pública; de lo contrario, la sociedad cae en un estado de caos y anarquía a razón del incremento del miedo, así como inestabilidad debido a la falta de información confiable y orientación durante la crisis. En relación con la interpretación de riesgos, Reddy y Gupta (2020: 3793) reconocen que existen múltiples factores que influyen en la aceptación de la información de riesgo, como las características sociales y culturales, el género, los contrastes generacionales, las inclinaciones lingüísticas, las convicciones estrictas, las creencias religiosas y el nivel de alfabetización. Incluso, otro factor que claramente influye en la aceptabilidad de los mensajes es la ideología política (Reddy y Gupta, 2020; Havey, 2020; Pennycook *et al.*, 2020).

Más allá del proceso de comunicación propuesto por Zang, Li y Chen, hay elementos fundamentales en cuanto a que están presentes en el proceso general de comunicación: los medios por los que se transmite el mensaje, la calidad de la información y la función del lenguaje utilizada.

En la literatura se identifican dos grandes categorías de medios de transmisión de información: *a)* medios sociales digitales [*social media*], y *b)* medios de información tradicionales [*news media*] (Kleis-Nielsen *et al.*, 2020). Los medios sociales comprenden redes sociales, como Facebook, Twitter e Instagram; sitios de internet enfocados a la transmisión de videos, como Youtube; las aplicaciones de mensajes, WhatsApp y Telegram, entre otros, y los motores de búsqueda, como Google, Yahoo y FireFox. Por otra parte, los medios tradicionales se conforman de periódicos impresos o digitales, noticiarios de televisión y radiofónicos ligados a organizaciones e informadores profesionales.

Referente a la calidad de la información, organismos internacionales como Organización Mundial de la Salud (OMS), la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO, por sus siglas en inglés) y la Organización Panamericana de la Salud (PAHO, por sus siglas en inglés) han advertido que durante la situación de pandemia del COVID-19 surgió una producción exponencial de información que propicia un ambiente de infodemia. El concepto de infodemia se refiere a una sobreabundancia de información en el que coexisten noticias precisas, imprecisas y engañosas (PAHO, 2020). Por ejemplo, algunas noticias engañosas o falsas —fake news—, rumores y teorías de la conspiración que generaron desinformación durante la primera mitad del año hacían referencia a la propagación de COVID a través de la red de telecomunicaciones 5G; el origen del COVID-19 como arma biológica creada por los chinos y un ambicioso plan de Bill Gates para lucrar vendiendo un amplio programa de vacunación, entre otras (Shahsavari *et al.*, 2020).

57

En el ambiente de infodemia descrito, resulta comprensible que el público haya preferido la información formal emitida por los medios de comunicación tradicionales sobre la que se difundió en las redes sociales, espacio que representó un canal propicio para el esparcimiento de noticias propagandísticas y/o falsas debido al anonimato, accesibilidad y rapidez para producir contenidos y bajos costos de difusión (Levinson, 2017). Por ejemplo, una encuesta internacional publicada por *The Reuters Institute for the Study of Journalism and Misinformation, Science and Media* (Kleis-Nielsen *et al.*, 2020), durante la situación de pandemia las tres fuentes de información que la población consideró más confiables son los científicos, las autoridades de salud y las organizaciones formales de noticias (Kleis-Nielsen *et al.*, 2020). Levinson (2017) plantea que el prestigio de la prensa proviene de una serie de procedimientos, como la verificación de hechos y la insistencia de una fuente confiable, pero esto no excluye que se cometan errores, o que en algunos casos la búsqueda por persuadir a la audiencia sobre la calidad de los contenidos genera escenarios de exageración u omisión de hechos clave. Además, Levinson también señala que el proceso mismo de informar las noticias es distorsionador, ya que inevitablemente las notas informativas no son un testimonio literal de los eventos, sino más bien un recuento sujeto a la decisión personal de un productor o editor que señalan hacia dónde se debe dirigir el relato o que, al menos, seleccionan lo que se va a imprimir.

En México, desde el primer caso de COVID-19 confirmado oficialmente en el país, el gobierno federal, a través del Consejo de Salubridad General, estableció un protocolo de transmisión de información a través de conferencias de prensa vespertinas que abordaron las consideraciones generales para ayudar a la población a entender la enfermedad, las medidas de prevención



a ejecutar, el número de contagios y decesos, así como las medidas y mandatos gubernamentales a seguir por parte de la población (Gobierno de México, 2020a). Desde el inicio, las autoridades sanitarias realizaron señalamientos respecto a la calidad de la información que se transmite a través de los medios de comunicación. Por ejemplo, el 26 de julio, el subsecretario de salud Hugo López-Gatell utilizó el término infodemia para referirse a “partes interesadas de la sociedad en distraer la atención, en cambiar la realidad a través de distorsionar la información” y precisó que existen medios de información o grupos políticos y sociales que distorsionan la información “siguiendo más una agenda de distracción que un interés de comunicar a la sociedad” (Gobierno de México, 2020b).

58



La discusión anterior invita a identificar, a) el nivel de información disponible a la ciudadanía informada un día cualquiera de 2020 respecto a la pandemia de COVID-19, y b) si coincide o discrepa la agenda informativa de los medios de comunicación publicados en México respecto a la comunicación de riesgo que ha emitido la institución responsable del manejo de la pandemia de COVID-19. Ante ello, la estrategia metodológica realiza tres tipos de análisis de procesamiento de lenguaje natural: *i)* de temas; *ii)* de similitud, y *iii)* de sentimientos.<sup>1</sup> El ejercicio resulta relevante porque ciudadanos mejor informados pueden tomar mejores decisiones para salvaguardar sus vidas en un contexto de alta incertidumbre, desinformación e infodemia como el que se vive bajo una situación de pandemia.

### III. MÉTODO

En la presente investigación se han definido dos objetivos. Uno es identificar el nivel de información disponible, mientras que el otro consiste en definir si la agenda informativa de los medios de comunicación publicados en México coincide o discrepa con la información que ha emitido la institución responsable del manejo de la pandemia de COVID-19. Se plantea un método que consta de cuatro partes: *a)* análisis de frecuencias e identificación de momentos relevantes clave; *b)* análisis de texto; *c)* análisis de similitud del discurso, y *d)* análisis de la polaridad o sentimiento del discurso.

#### 1. Datos

Se utilizan dos fuentes de datos. La primera se conforma de las 271 versiones estenográficas de las conferencias de prensa realizadas por la autoridad

---

<sup>1</sup> El análisis se efectúa con el *software* R.



POLARIZACIÓN Y SIMILITUD EN LAS REPRESENTACIONES DE MENSAJES...

sanitaria entre el 1o. de marzo y el 30 de noviembre de 2020, en las cuales se presentaba el informe diario sobre COVID-19. La segunda utiliza el archivo de periódicos y noticias de televisión desarrollado por una empresa de monitoreo de medios.

Con los datos de la segunda fuente se construyeron a la vez, tres bases de textos. La base de *noticias* se compone de 4313 notas informativas impresas en los once periódicos sociopolíticos con mayor tiraje nacional (Medios Publicitarios Mexicanos, 2020) y 549 versiones estenográficas de segmentos de noticias transmitidos por tres noticieros nocturnos nacionales. La base de *opinión* está conformada por 23,265 columnas de opinión. Estas bases de datos fueron obtenidas de una empresa que tiene como propósito comercial recopilar diariamente la información de los medios de comunicación. La Tabla 1 describe la base de notas y artículos de opinión tomadas de cada medio informativo.

59

TABLA 1. MEDIOS PARA ANALIZAR, FRECUENCIAS Y PORCENTAJES

Periódicos	Notas informativas		Artículos de opinión	
	<i>n</i>	%	<i>n</i>	%
El Economista	219	5.08	2,110	9.07
El Financiero	314	7.28	2,525	10.85
El Herald de Mx	238	5.52	2,256	9.7
El Universal	510	11.82	3,380	14.53
Excélsior	497	11.52	3,948	16.97
La Jornada	637	14.77	2,010	8.64
La Razón de México	400	9.27	1,480	6.36
Milenio Diario	305	7.07	2,652	11.4
Reforma	752	17.44	2,328	10.01
Reporte Índigo	97	2.25	265	1.14
Unomásuno	344	7.98	311	1.34
<i>Total</i>	4,313	100	23,265	100

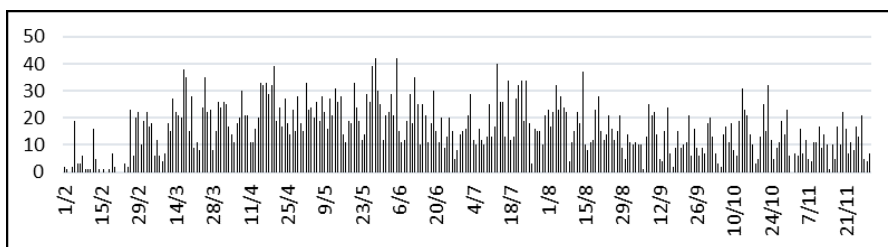
Noticieros	<i>n</i>	%
Imagen Noticias con Ciro Gómez Leyva	314	57.19
En Punto	147	26.78
Hechos de la Noche	88	16.03
<i>Total</i>	549	100.00

FUENTE: elaboración propia.

## 2. Análisis de frecuencias y momentos relevantes en la cobertura de los medios

Para analizar las frecuencias e identificar los momentos relevantes se ubican los periodos de tiempo que recibieron mayor cobertura de acuerdo con el número de notas informativas impresas o transmitidas en los noticieros. La Figura 1 muestra la relación de días y frecuencias de noticias publicadas. Presenta movimientos en las frecuencias, crestas y valles que dificultan la interpretación.

FIGURA 1. FRECUENCIAS OBSERVADAS DE NOTICIAS POR DÍA (1 FEB.-30 NOV.)



FUENTE: elaboración propia.

Al ser la ocurrencia de noticias volátil, se utiliza una técnica que controla las fluctuaciones a corto, medio o largo plazo denominada medias móviles, la cual es empleada usualmente en mercados financieros para delinear el comportamiento de una variable y encontrar tendencias no basadas en la apreciación del analista (Herrán *et al.*, 2008). De esta manera, la aplicación de la técnica de medias móviles consiste en crear una serie de promedios fijando sucesivamente una fecha de referencia y calculando el promedio de notas informativas que fueron impresas o transmitidas en los noticieros sobre COVID-19 en un periodo de tiempo determinado a partir de dicha fecha. Para la presente investigación se usa la modalidad de medias móviles centrales (mmc) de grado cinco, promediando la frecuencia de noticias desde dos días antes de cada fecha de referencia hasta dos días después de dicha fecha, es decir, para un periodo total de cinco días. La fórmula para calcular cada media móvil central es la siguiente:

$$mmc_t = \frac{\sum_{i=t-2}^{t+2} n_i}{5}$$

Donde  $n_i$  es la frecuencia de noticias un día específico  $t$ . La versatilidad de la técnica la hace un instrumento útil para los objetivos planteados en la presente investigación y permite identificar periodos de auge de noticias de manera gráfica y numérica.

### 3. Análisis de tópicos

61

La relevancia de una palabra está relacionada con su aparición en un conjunto de textos. En este sentido, el algoritmo descarta los tópicos o temáticas que se alejan del objeto de estudio, describiendo una relación positiva entre relevancia y frecuencia:

$$I = f(F)$$

Donde  $I$  es la importancia que tuvo una determinada palabra, y  $F$  la suma total de las veces que aparece una determinada palabra en un conjunto de textos. A saber:

$$I = f(F) = \sum_{i=1}^n x_i = x_{1t} + x_{2t} + x_{3t} + \dots + x_{nt}$$

Entendiéndose que la suma de la aparición de una determinada palabra  $x$  determina la importancia en el texto ( $i = 1$  o conjunto de documentos  $i = n$ ) para un periodo  $t$ .

Luego, cada conjunto de palabras es representado gráficamente por medio de nubes de palabras cuyo propósito es estilizar la identificación de tópicos o temáticas claves. Constituye una revisión de los tópicos o temáticas claves de un conjunto de documentos donde el tamaño de cada palabra está en función de su frecuencia, mostrando su relevancia. Una limitante clásica con respecto a los artículos de opinión es que no tienen la misma sincronía para abordar las temáticas cotidianas. Tampoco se enfocan en un evento o cuestión en específico, ya que, en un solo espacio, pueden abordar diversas temáticas, realizar comparaciones, establecer relaciones y proponer conclusiones con mayor margen de maniobra que las noticias. Sin embargo, esto se solventa cuando la unidad de análisis es la palabra. El análisis de tópicos, y posteriormente de similitud, sobre las columnas de opinión pueden brindar referencias que permitan extraer y comparar los temas que gobierno y prensa consideran relevantes en sus respectivas agendas. Es importante identificar que el análisis tiene una naturaleza inductiva, es decir, que los tópicos emergen del análisis exploratorio de los datos sin una agenda afectada por la perspectiva del analista. Un análisis deductivo, en cambio, implicaría que el analista antepone términos que sean relevantes para sus objetivos de investigación y después se conduce el análisis.

#### 4. Análisis de similitud

62 El análisis de similitud identifica temas que aborda la estrategia de comunicación planteada por las autoridades sanitarias y la preeminencia de esos tópicos en los periódicos y los noticiarios televisivos. No se evalúa la literalidad con que se transmite el discurso completo o segmentos del discurso; más bien, se mide el grado de coincidencia en la relevancia de los conceptos utilizados entre las conferencias vespertinas y los medios de comunicación tradicionales. El análisis de similitud puede interpretarse como una semejanza (o diferencia) entre conceptos usados por grupo de análisis en un período de tiempo determinado. En este caso, dicha herramienta se utiliza para conocer qué tan similar (o diferente) es el conjunto de palabras empleado en los discursos que tienen los textos provenientes de los tres grupos de análisis, desde marzo hasta noviembre de 2020.

Por fines de seguimiento y de parsimonia, se diseñaron tres índices locales de similitud (ILS) que expresan el grado de asociación de cada una de las distintas combinaciones de datos (gobierno y noticias; gobierno y opinión; y noticias y opinión) a partir de la presencia o ausencia mutua (1 y 0) de veinte conceptos con las frecuencias más altas en cada grupo. Una vez identificadas las palabras similares y codificadas dicotómicamente, se suma las coincidencias y divide para el número de palabras de referencia. De esta forma, para cada mes, entre marzo y noviembre, se obtuvieron los siguientes índices:

*Índice Gobierno-Noticias (IGN)*. Es el índice de similitud del mes  $m$  de los conceptos principales empleados por el gobierno y las noticias. Podemos expresarlo como:

$$IGN_m = \frac{\sum_{i=1}^n gob_m \cap not_m}{20}$$

Donde  $gob_m \cap not_m$  representa el set principal de palabras comunes usadas en la información emitida en las conferencias vespertinas y las noticias en el mes  $m$ , dividido entre el número de palabras de referencia  $n = 20$ .

*Índice Gobierno-Opinión (IGO)*. Es el índice de similitud de los tópicos más frecuentes planteados por el gobierno y las columnas de opinión en el mes  $m$ . Podemos expresarlo como:

$$IGO_m = \frac{\sum_{i=1}^n gob_m \cap opi_m}{20}$$

Donde  $gob_m \cap opi_m$  representa el set principal de conceptos usados en las conferencias vespertinas y las columnas de opinión en el mes  $m$ , dividido entre el número de palabras de referencia  $n = 20$ .

*Índice Noticias-Opinión (INO)*. Es el índice que expresa el nivel de similitud del mes  $m$  entre los conceptos utilizados por las noticias y las columnas de opinión. Podemos expresarlo como:

$$INO_m = \frac{\sum_{i=1}^n not_m \cap opi_m}{20}$$

Donde  $not_m \cap opi_m$  representa la intersección de conceptos utilizados en las noticias y las columnas de opinión, dividido entre el número de palabras de referencia  $n = 20$ . 63

Los valores que puede adoptar cada índice de similitud están comprendidos entre 0 y 1,  $0 \leq ILS_{IGN,IGO,INO} \leq 1$ , donde cero representa una perfecta diferencia, y 1, perfecta similitud. ● ○ ●

Adicionalmente, se ha calculado una medición general de similitud para cada mes, o *Índice Global de Similitud (IGS)*, resultado de la sumatoria de cada uno de los índices de similitud ponderados en proporciones iguales, es decir:

$$IGS_m = \frac{1}{3} IGN_m + \frac{1}{3} IGO_m + \frac{1}{3} INO_m$$

Donde:

$$0 \leq IGS_m \leq 1$$

Al igual que los valores de los índices anteriores, el índice global de similitud puede tomar un valor comprendido entre 0 y 1,  $0 \leq IGS_m \leq 1$ .

## 5. *Análisis de sentimientos*

El análisis de sentimientos tiene por objeto determinar la actitud de un escritor o grupo de análisis con respecto a algún tema o la polaridad contextual general de un documento que puede ser positiva, negativo o neutro. En el presente trabajo se aplica dicha herramienta, a través de un algoritmo, para analizar el texto de cada grupo (conferencias, noticias, opinión) de manera temporal y así conocer la evolución de los sentimientos del periodo comprendido entre marzo y noviembre de 2020.

El proceso de este algoritmo puede ser descrito en tres pasos:

- 1) *Formación de los grupos de análisis*. Se agrupa en un único archivo cada grupo de análisis con una periodicidad mensual para optimizar

el rendimiento de la máquina, obteniéndose un total de 27 archivos trabajados en un formato .txt o nueve archivos por cada grupo de análisis o actor de la noticia.

- 64
- 2) *Limpieza de texto*. Para cada archivo se implementa una serie de algoritmos que abarcan limpieza del texto, eliminación de direcciones electrónicas (url's), palabras irrelevantes (stopwords), números (0-9), signos de puntuación, caracteres especiales y espacios dejados por los anteriores tratamientos. Partiendo de un texto limpio, el vocabulario o set de palabras correspondiente a cada grupo de análisis son comparados con un diccionario de sentimientos. En nuestro caso utilizamos el diccionario de Mohammad y Turney (2010) quienes proponen un diccionario "nrc".
  - 3) *Contabilización*. Obtenemos un resultado mensual de los tres sentimientos por grupo de análisis, con lo que se calcula la presencia relativa de cada sentimiento por grupo y por período, es decir, la proporción de sentimientos presente en gobierno, noticias y opiniones en el período marzo a noviembre de 2020.

#### IV. RESULTADOS

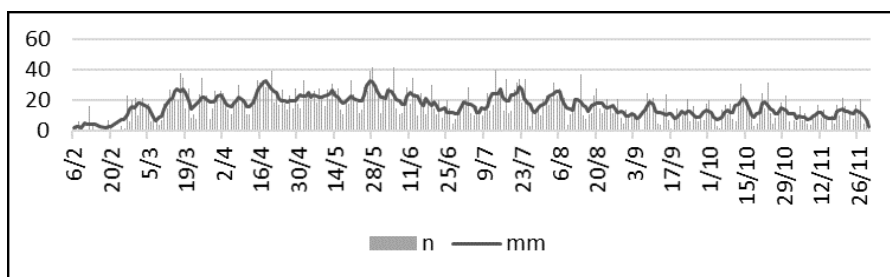
##### 1. *Frecuencias y momentos relevantes en la cobertura*

La Figura 2 muestra los resultados del ajuste realizado por el cálculo de medias móviles (*mmc*), del 5 de febrero al 28 de noviembre de 2020. Se identifican periodos de tiempo con *mmc* más altas, en contraste con las frecuencias observadas. Se identifican los periodos de mayor auge, bajo el criterio de que las  $mmc > 25$  son altas. No obstante, en la mayoría de los casos hay días consecutivos que cumplen con este criterio, por lo que se realiza la interpretación por periodos. A partir del 8 de marzo inició un incremento de las noticias, alcanzando los valores más altos los días 18 de abril y 27 de mayo. Se identifica que los intervalos de mayor actividad coinciden con las semanas de transición de una fase a otra. Los periodos identificados con mayores frecuencias son: 1) 15-18 marzo – caracterización de la epidemia; 2) 15-21 abril – incremento de fallecimientos, debate creciente; 3) 25-30 mayo – mayor incremento de hospitalizaciones y fallecimientos; 4) Nueva normalidad: 15-23 junio; 15-23 julio; 4-6 agosto; 12-15 de octubre; 23-25 de octubre; 21-26 de noviembre. La alta frecuencia de noticias inició a partir del 12 de mayo, cuando el número de decesos llegó a 3, 573 y 36, 327 contagios confirmados, registrándose 111

fallecimientos de trabajadores de la salud. Para el 13 de mayo, el Consejo de Salubridad General, máxima autoridad sanitaria de México, incorporó los sectores mineros, de construcción y la fabricación de transporte a las actividades esenciales. Por otra parte, el presidente López Obrador dio a conocer el plan de reactivación económica denominado Nueva Normalidad, fase que tendría como principal instrumento de comunicación de riesgo, un semáforo epidemiológico semanal. Para el 25 de mayo, se incrementó el número de contagios de forma sustancial, superando los 71,000 casos y 7,633 muertes contabilizadas. La Universidad Johns Hopkins señaló a México como el país con mayor letalidad en América Latina.

Los medios reportaron que el subsecretario López-Gatell calificó, en videoconferencia ante senadores, como inútil y costoso la aplicación masiva de pruebas de identificación de COVID-19. El gobierno de México prefirió concentrar recursos en adquirir vacunas, y fue el país de Latinoamérica que más vacunas concentró en el corto plazo. Una noticia que recibió amplia difusión fue que la OMS invitó al subsecretario López-Gatell a su Consejo Asesor.

FIGURA 2. FRECUENCIA DE NOTICIAS AJUSTADA POR MEDIAS MÓVILES (01 FEB.-30 NOV., 2020)



FUENTE: elaboración propia.

## 2. Identificación de temas relevantes

La Figura 3 contiene las nubes de palabras de marzo, junio y noviembre para cada población, gobierno, noticias y opinión. Por espacio, se expone el mes de inicio, un mes ubicado en un punto intermedio y el mes final del periodo que comprende el presente análisis (marzo-noviembre 2020). Al comparar nubes de palabras, resulta que las prioridades de comunicación son distintas entre los grupos y a lo largo del tiempo. Para una identificación minuciosa de los principales temas, seguimiento y continuidad en la agenda pública



se realiza un análisis enfocado en las diez palabras más utilizadas en las conferencias vespertinas, noticias y columnas de opinión. Una parte de los tópicos más utilizados en las *conferencias* vespertinas tiene una función conativa: vocablos como “salud”, “enfermedad”, “información”, “virus”, “atención”, “epidemia”, “contagios”, “curva”, “riesgo”, “transmisión” y “vacunas”, son usados en oraciones que hacen referencia a normas de salud, formas de transmisión y mecanismos de atención relacionados con la enfermedad. Destacan dos figuras, la del subsecretario Hugo López-Gatell y del presidente Andrés Manuel López Obrador (“presidente”, “López-Obrador”). Se registran sustantivos abstractos referentes a la organización de la agenda pública (gobierno, economía, política). En cualquier mes, la palabra “crisis” aparece entre las diez palabras más frecuentes en las noticias. Al analizar los diez conceptos principales, se observa que no existen cambios significativos en cuanto a la permanencia de los temas e incorporación de nuevos vocablos a través de los meses. En las noticias, los temas que se incorporaron estaban vinculados con “hospitales”, “médico”, “pacientes” en abril; “semáforo”, “rojo”, “naranja”, “distancia”, “muertes”, “pruebas”, “camas” en mayo; “entidades” en junio; “gobernadores”, en julio; “vacuna” en agosto, “verde” e “influenza” en septiembre, “rebrote” en octubre, y “Pfizer” en noviembre.

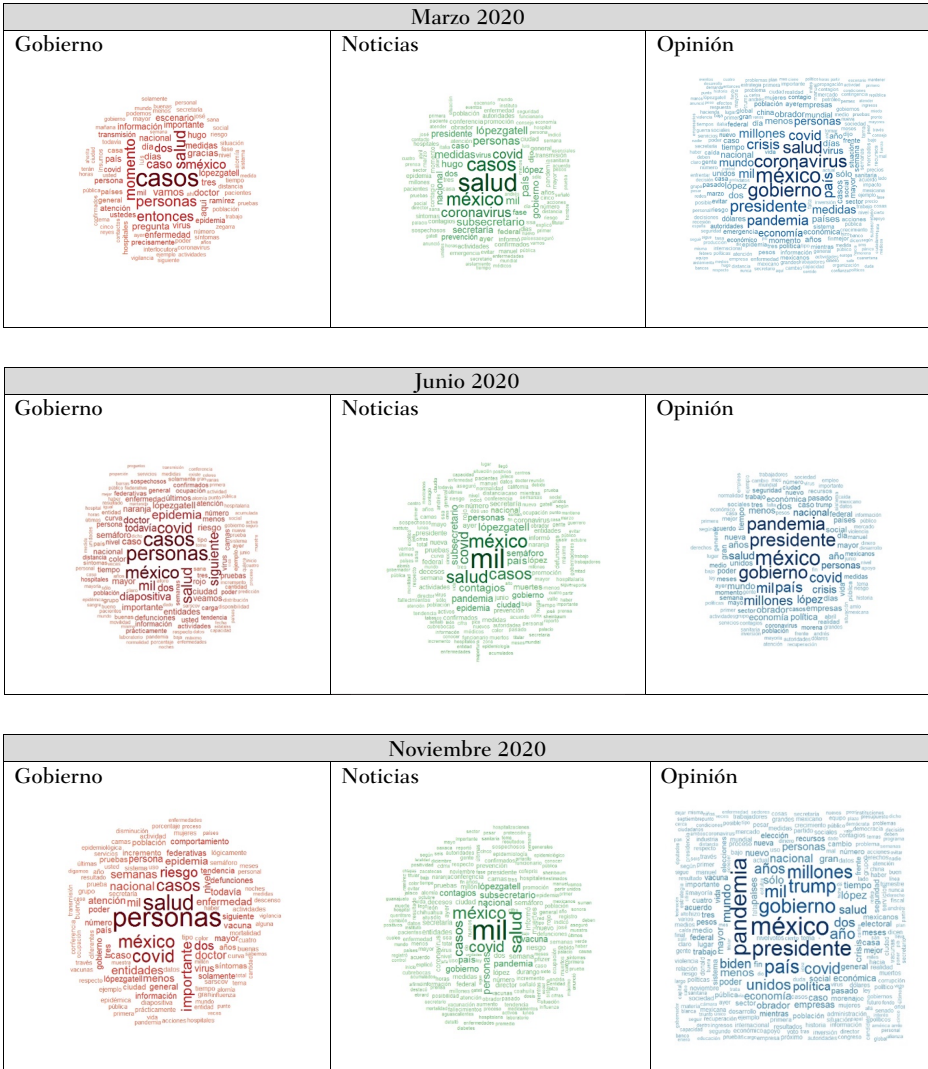
66



En los artículos de opinión, los temas que se agregaron fueron “empresas”, “emergencia”, “sanitaria”, “recursos”, “casa”, “trabajo”, “política” en abril; “seguridad”, “contagio” en mayo; “Trump” en junio; “educación”, “vacuna”, “inversión” y “desarrollo” en agosto; “crecimiento”, “recuperación” y “presupuesto” en septiembre; “resultados” y “mercado” en noviembre. Expresiones utilizadas en las conferencias vespertinas, son transmitidas por los medios usando sinónimos quizá con distinta connotación a la intención original de la fuente. Por ejemplo, en las conferencias vespertinas se hace uso habitual del término “epidemia”, mientras que en las noticias y las columnas se da preferencia a utilizar la palabra “pandemia”. Otro caso es que en las conferencias se usa el término “defunciones” mientras que en las noticias y artículos de opinión se da preferencia a las palabras “muertes” o “decesos”. En resumen, se ha identificado que durante la contingencia sanitaria se presentan cambios en el objetivo de los mensajes que se transmiten o reproducen. En los medios tradicionales destacan dos figuras, la del subsecretario Hugo López-Gatell y del presidente Andrés Manuel López Obrador. A lo largo de los meses existen cambios en los principales temas de discusión y una priorización diferenciada de ciertos temas. Otra característica observada es el uso de sinónimos empleados para informar, que pueden tener una connotación distinta a las palabras empleadas en las conferencias vespertinas. Una peculiaridad es que en todos los meses la palabra “crisis” aparece entre las diez palabras más frecuentes en los artículos de opinión.

POLARIZACIÓN Y SIMILITUD EN LAS REPRESENTACIONES DE MENSAJES...

FIGURA 3. NUBES DE PALABRAS

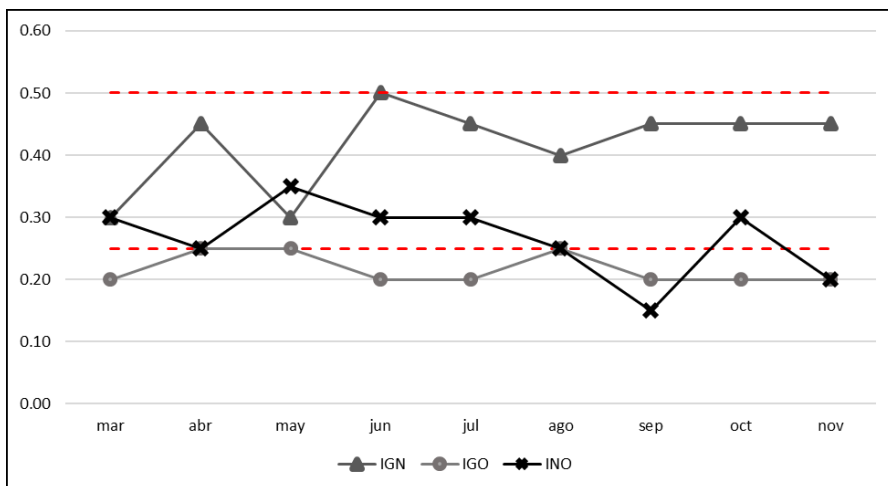


3. Análisis de similitud

Ahora el análisis se enfoca en precisar cuánto se aleja la comunicación emitida por el gobierno, de la comunicación emitida por los medios de comunicación. Se analiza el grado de coincidencia con que se transmiten los temas que

68 forman parte de la comunicación emitida por las autoridades sanitarias. Para la comparación de los temas se utilizan los veinte vocablos más frecuentes con alto valor analítico. La comparación entre las conferencias, las noticias y las columnas de opinión, se evalúa a través de índices de similitud, índice gobierno-noticias (IGN), índice gobierno-opinión (IGO) e índice noticias-opinión (INO), así como un índice global de similitud (IGS). Para fines interpretativos, se clasifica el valor de los índices en tres niveles: a) bajo,  $< 0.25$ ; b) medio,  $> 0.25$  y  $< 0.50$ ; y c) alto,  $> 0.50$ . Un valor  $> 0.50$  en alguno de los índices significa que dos poblaciones han priorizado y reproducido más del 50% de las palabras principales que componen el mensaje. La Figura 4 muestra el comportamiento en cada uno de los índices a lo largo de los nueve meses que comprende la investigación. Adicional a las líneas con los marcadores por índice, hay dos líneas punteadas que dividen el área del gráfico en tres secciones; las secciones corresponden a zonas de similitud.

FIGURA 4. ÍNDICES LOCALES DE SIMILITUD (MARZO-NOVIEMBRE 2020)



FUENTE: elaboración propia.

La Figura 4 muestra que el Índice Gobierno-Noticias se desenvuelve en similitud mediana, presentando valores más bajos durante los meses de marzo y mayo, es decir, en las primeras fases de la emergencia sanitaria. El punto de similitud más alto (0.5) se ubica en el mes de junio, para posteriormente estabilizarse en valores superiores a 0.4 desde julio hasta noviembre. El Índi-

ce Gobierno-Opinión ubica la mayoría de sus valores por debajo en el área de baja similitud. Los meses de marzo, junio-julio, septiembre-noviembre obtuvieron un valor igual a 0.2, mientras que los meses de abril-mayo y agosto se ubicaron en la frontera de 0.25. Respecto al Índice Noticias-Opinión, se observa que en cinco de nueve meses (marzo, mayo, junio, julio y octubre), sus valores están en la zona de similitud media; en dos meses se encuentra sobre la línea entre baja y mediana similitud, y en dos meses (septiembre y noviembre) el INO entra en categoría baja. Una característica del INO es que en términos generales se ubica sobre el IGO, pero debajo del IGN.

69

En conclusión, las noticias, sumadas aquí de medios impresos y televisión, al transmitir y reportar los mensajes desde la fuente original, las conferencias vespertinas, retoman e imprimen una proporción regular de los temas y vocablos empleados por las autoridades sanitarias. En contraste, las columnas de opinión ilustran una agenda propia, independiente de la agenda gubernamental, y con objetivos distintos a los planteados en las conferencias vespertinas. Esto significa que, en su tarea de fijar postura, convencer y llamar la atención de los lectores, las columnas utilizan palabras y recursos retóricos alejados de los términos empleados por las fuentes gubernamentales. Los valores del INO ubicados en el nivel medio muestran que existe una mayor cercanía entre las agendas de las columnas de opinión y de las noticias, que entre las agendas del gobierno y los artículos de opinión durante la emergencia sanitaria.

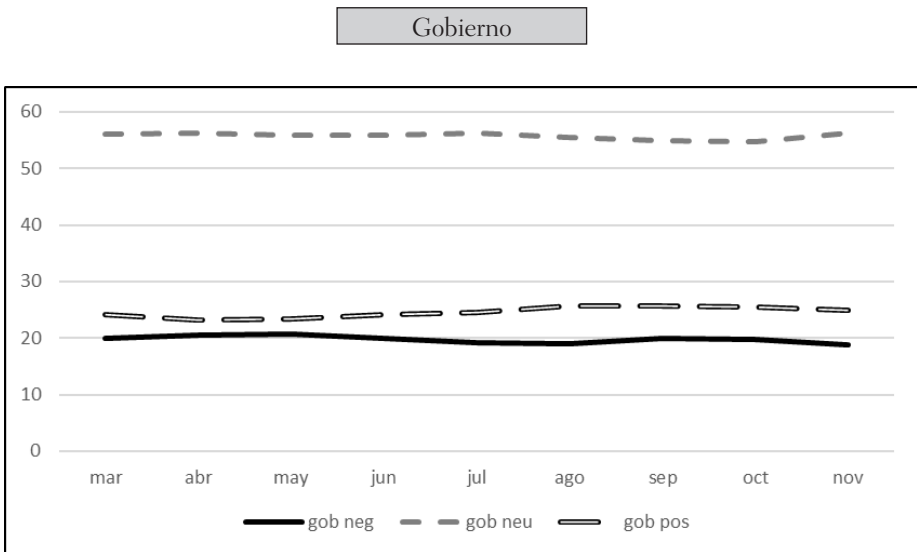
#### 4. *Análisis de sentimientos*

La Figura 5 expone la evolución de los sentimientos. La parte superior se refiere a las conferencias vespertinas. El tiempo se representa en el eje horizontal, mientras el eje vertical comprende la proporción por tipo de sentimiento (negativo, neutral o positivo). La clasificación reporta que, de las palabras emitidas por las autoridades sanitarias, de cada 100 palabras 56 fueron calificadas como neutrales, 20 como negativas y 24 como positivas. A partir de junio, la trayectoria de las palabras positivas experimenta una brecha incremental con respecto a las palabras negativas. Esto podría estar vinculado con el cambio en el mensaje de riesgo identificado durante el análisis de tópicos. En la parte baja, el panel izquierdo de la Figura 5 expone la trayectoria de sentimientos de las noticias. De cada 100 palabras, 59 se inclinan hacia sentimientos neutrales, 19 negativos y 22, positivo. Es decir, se presenta un incremento de las palabras neutrales con respecto a las proporciones del grupo gobierno y una disminución de las palabras positivas. En junio, al igual que

HUERTA WONG / CASTAÑEDA VALENCIA / MANZANO MORA

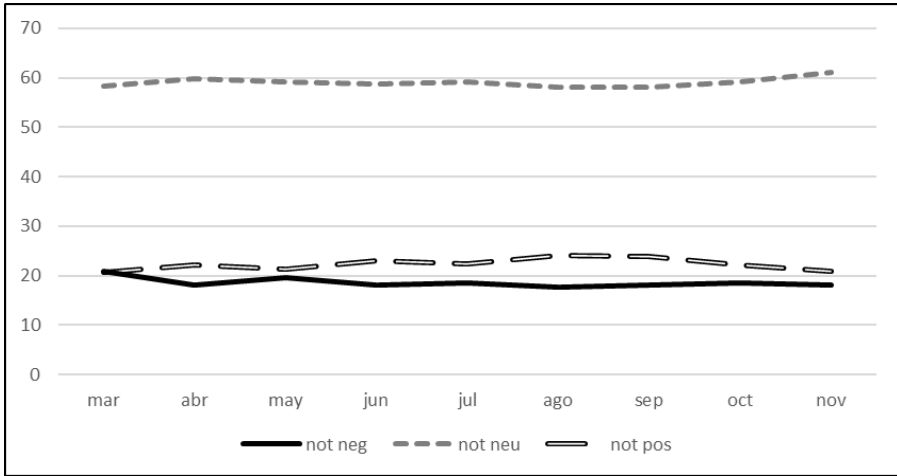
70 en las conferencias vespertinas, hay un incremento de las palabras positivas. Para octubre y noviembre, la curva presenta decrementos. Esto podría estar vinculado con el cambio en la orientación de la comunicación gubernamental en junio, efecto que se mantuvo hasta septiembre. Por último, el panel derecho de la parte baja de la Figura 5 presenta el análisis de sentimientos de artículos de opinión. De cada 100 palabras, 58 son neutrales, 20 son negativas y 22, positivas. La proporción de palabras negativas en marzo es superior a la de palabras positivas. Las curvas se cruzan en abril, abriéndose una estrecha brecha entre ambas. No obstante, se mantienen juntas y sin cambios a lo largo de los meses. Sin importar los cambios en el mensaje de riesgo, las proporciones de positivos o negativos son similares. En resumen, los cambios en las prioridades comunicativas durante las conferencias vespertinas se han reflejado en el incremento de la proporción de positivos a partir de junio. Además, estos cambios coinciden temporalmente con el incremento de las palabras positivas en las noticias, aunque dicho efecto tuvo un desgaste en octubre y noviembre. En la opinión, sin importar los cambios en el mensaje de riesgo, positivos y negativos se mantienen prácticamente constantes en el tiempo.

FIGURA 5. EVOLUCIÓN DE SENTIMIENTOS  
(MARZO-NOVIEMBRE 2020)



POLARIZACIÓN Y SIMILITUD EN LAS REPRESENTACIONES DE MENSAJES...

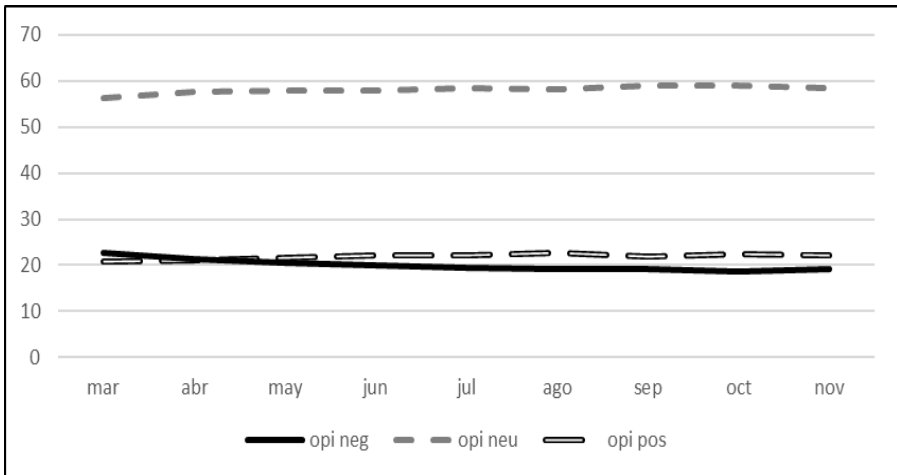
Noticias



71



Artículos de opinión



FUENTE: elaboración propia.

V. CONCLUSIONES Y REFLEXIONES FINALES

¿Cuál fue el nivel y tipo de información disponible en los medios de comunicación de la Ciudad de México, durante la pandemia de COVID-19? En primer

lugar, se considera la cantidad de noticias que reproducen el mensaje oficial. A través del método de medias móviles se detecta que, en una situación de contingencia que ha tenido una duración temporal amplia, la cantidad de noticias que hacen referencia a los mensajes de riesgo cambia de acuerdo con la fase o periodo en curso: al inicio de la contingencia es abundante el flujo de información y se incrementa en momentos clave. Sin embargo, poco después, la frecuencia de noticias se reduce y aunque se presentan incrementos, son menores que los experimentados en la fase inicial.

72

Segundo, respecto a los contenidos de los artículos informativos, mediante la revisión y lectura de las notas informativas en las fases de auge de noticias, varían de acuerdo con el avance de la contingencia. En un inicio, las noticias se enfocan en reportar la información que requiere saber el público sobre el origen del riesgo, medidas adoptadas por las autoridades gubernamentales, declaraciones oficiales, acciones preventivas, etcétera. Más adelante, cuando el primer contacto con la situación de riesgo es superado, los medios destinan espacios para críticos y detractores de las medidas implementadas, involucran a organizaciones, actores políticos, equipos de seguridad, entrevistas a personajes clave e incluso cuestionan e indagan la veracidad y credibilidad de las fuentes oficiales.

Tercero, el análisis de tópicos identifica la importancia de las temáticas abordadas en las conferencias vespertinas, a través de identificar frecuencias más altas, mayores espacios concedidos, mayor número de menciones y mayor difusión de los temas. Se encontró que la relevancia de los temas es apreciada de forma distinta entre autoridades gubernamentales, periodistas y columnistas.

Cuarto, el análisis de tópicos encuentra que la comunicación gubernamental, representada por las conferencias vespertinas, a través de identificar hace uso de conceptos en oraciones con funciones conativas, que transmiten indicaciones y sugerencias sobre los cuidados de salud que debe atender la población, y oraciones con funciones referenciales que describen el comportamiento de la situación de riesgo. Si bien las funciones conativas y referenciales del discurso se ven reflejadas en las notas informativas, en los artículos de opinión prevalecen recursos retóricos para persuadir e influir en el lector. Por ejemplo, en los artículos de opinión es frecuente evocar sustantivos abstractos (política, economía, etcétera) e incluso relacionarlos con calificativos como “crisis”, vocablo que aparece entre las diez palabras más frecuentes en todos los meses.

Una vez señalados los elementos analizados (cantidad, contenidos, importancia y función del discurso), podemos definir si es distinta la agenda informativa de los medios de comunicación publicados en México respecto a la comunicación que ha emitido la institución responsable del manejo de la pandemia de COVID-19. Un primer acercamiento lo obtenemos al comparar



el léxico empleado. A través del análisis de tópicos se han reconocido vocablos, emitidos en el discurso de riesgo de las autoridades sanitarias, que son sustituidos en el mensaje final por sinónimos que pueden tener una connotación distinta a las palabras empleadas por los comunicadores y que pueden denotar la magnitud y el sensacionalismo de un tema.

Segundo, a través de índices de similitud, que captan la coincidencia de temas e importancia que se dio a los términos y conceptos empleados en cada uno de los tres grupos, se pudo identificar que existe similitud media entre la agenda de gobierno y noticias, al igual que noticias y artículos de opinión, pero la similitud es baja entre gobierno y artículos de opinión. En términos generales, el rango del nivel de similitud entre los tres grupos va de medio a bajo.

Respecto al análisis de sentimientos, no se han encontrado diferencias sustanciales en la proporción de palabras positivas, neutrales y negativas empleadas por los tres grupos. Esta aseveración tiene un limitante que no puede pasarse por alto: al momento que esto se escribe, aún están por desarrollarse algoritmos capaces de identificar el sarcasmo, la ironía o el doble significado de una oración. Se identificaron cambios en el porcentaje de palabras positivas de las noticias cuando hubo un cambio en el orden de los términos más frecuentes de las conferencias vespertinas, mientras que en los artículos de opinión no hubo variaciones en la trayectoria, un descriptor de inmovilidad en la polaridad de los articulistas. En el contexto de la emergencia sanitaria, la recomendación de OMS y OPS para reducir la incertidumbre a nivel mundial y local fue que tanto el gobierno, las noticias informativas y los líderes de opinión compartieran un discurso equilibrado y reconocimiento de los aciertos.

Así, el manejo del COVID-19 no es un asunto trivial ni parecido a las experiencias previas. En esa línea, la polarización mediante mensajes de división contribuyó poco al cuidado de la salud individual y colectiva. Bajo una situación de riesgo, los periodistas, reporteros y columnistas aún ejercen la función que les corresponde en un sistema democrático. Las tareas de proveer información entendible al público en general, atraer la atención de la audiencia y sus compromisos con la objetividad o la persuasión —según se trate de un artículo informativo o de opinión— necesariamente implica la aplicación de criterios que conllevan la selección de temas, preferencia por difundir ciertas noticias, reemplazar palabras, omisión de eventos y simplificación de la información técnica. Recibir información sistemáticamente contradictoria, como parece haber ocurrido en el caso de alguien que se haya expuesto frecuentemente a columnas de opinión, puede tener efectos no intencionados durante una situación de extremo riesgo. Considérese una situación de un crucero de pasajeros que naufraga. Es importante que los pasajeros sigan la instrucción del personal que ha sido entrenado para ese momento, el o la capitán y



la tripulación. Las instrucciones deben ser directas, breves y claras. Seguir las puede ser la diferencia entre la vida o la muerte. Ahora considérese que hay voces, líderes de opinión que de frente a la emergencia, sin entrenamiento alguno, discrepan de esos líderes. La duda de un o una pasajero puede ser la diferencia en experimentar conductas de riesgo. En México, quizá con alguna excepción, las y los columnistas no cuentan con entrenamiento médico o epidemiológico. La responsabilidad social de los medios podría haber sido comprometido por la subjetividad de la opinión publicada, tal como el derecho de las y los mexicanos a información que pudiera salvar vidas. Mayor investigación requiere identificar cuál fue el peso de la opinión publicada en las actitudes, conocimiento previo y comportamiento de las y los receptores de medios, lo cual escapa al alcance de este documento.

74

- 
- 
- 

Por último, aun cuando ya se han comentado brevemente en la sección de método, vale aquí comentar algunas de las fortalezas y limitaciones del método usado. Hemos dicho que se trata del primer censo a las noticias publicadas en un periodo determinado del que tengamos noticia, al menos por lo que hace a un país de habla hispana. La magnitud de los datos hace imposible el análisis y/o supervisión al que se podría haber recurrido en un análisis de contenido tradicional. La limitación en la automatización por ahora hace imposible identificar con total precisión el uso de dobles sentidos que con frecuencia se utiliza en la superficie de los medios, sobre todo en la superficie editorial o de comentarios. Si algún(a) columnista usa el sarcasmo, la herramienta utilizada no permite su identificación. El análisis estadístico utilizado en análisis de contenido tradicional habría podido establecer también con precisión las diferencias en puntajes entre un medio y otro, o en este caso específico, entre la superficie de noticias y la editorial. Es decir, el análisis estadístico habría podido evaluar las chances, momios, tamaños del efecto, marginales, y en general todos los coeficientes necesarios para identificar diferencias entre un grupo de datos y otros. En este sentido, el análisis por aprendizaje máquina, por ahora, es descriptivo y no concluyente. Dicho todo esto, es por ahora el conocimiento de frontera, hasta donde identificamos, para poder evaluar tendencias al usar grandes bases de datos. También los censos tienen ventajas considerables frente a las muestras, tal como se usan en los análisis de contenido tradicionales.

## VI. REFERENCIAS

ABRAMS, E. M. y GREENHAWT, M. (2020). “Risk Communication During COVID-19”, *The Journal of Allergy and Clinical Immunology*. 8(6). Disponible en: <https://doi.org/10.1016/j.jaip.2020.04.012>.

- BISHOP, C. R. (2006). *Pattern Recognition and Machine Learning*, Springer Verlag.
- BLAKE, C. (2011). "Text Mining", *Ann. Rev. Info. Sci. Tech.* 45. Disponible en: <https://doi.org/10.1002/aris.2011.1440450110>.
- CASTILLO-ESPARCIA, Antonio *et al.* (2020). "Comunicación política y COVID-19", *Estrategias del Gobierno de España. Profesional de la información.* 29(4). Disponible en: <https://doi.org/10.3145/epi.2020.jul.19>.
- FREWER, L. (2004). "The Public and Effective Risk Communication", *Toxicology Letters.* 149(1-3). Disponible en: <https://doi.org/10.1016/j.toxlet.2003.12.049>.
- GOBIERNO DE MÉXICO (2020a). "Se confirma en México caso importado de coronavirus COVID-19". Disponible en: <https://www.gob.mx/salud/prensa/077-se-confirma-en-mexico-caso-importado-de-coronavirus-covid-19#:~:text=En%20la%20conferencia%20matutina%20encabezada,antecedentes%20de%20viaje%20a%20Italia> [Consultado 08-12-2020].
- GOBIERNO DE MÉXICO (2020b). "Versión estenográfica. Conferencia de prensa. Informe diario sobre coronavirus COVID-19 en México". 26 de julio. Disponible en: <bit.ly/3wxPmpn> [Consultado 16-05-2022].
- GODBOLE, N., SRINIVASIAIAH, M., SKIENANA, S. (2018). "Large-Scale Sentiment Analysis for News and Blogs". Disponible en: <bit.ly/3wiyHXY> [Consultado 16-05-2022].
- GUPTA, V. y LEHAL, G. (2009). "A Survey of Text Mining Techniques and Applications", *Journal of Emerging Technologies in Web Intelligence.* 1(1). Disponible en: <http://dx.doi.org/10.4304/jetwi.1.1.60-76>.
- HAVEY, N. F. (2020). "Partisan Public Health: How Does Political Ideology Influence Support for COVID-19 Related Misinformation?", *Journal of Computational Social Science.* 3. Disponible en: <https://doi.org/10.1007/s42001-020-00089-2>.
- HERRÁN, V. *et al.* (2008). "Análisis de la eficacia de las medias móviles en el mercado intradiario de renta variable español", *Universidad, sociedad y mercados globales.*
- ISOAHO, K. *et al.* (2019). "Topic Modeling and Text Analysis for Qualitative Policy Research", *Policy Studies Journal.* Disponible en: <https://doi.org/10.1111/psj.12343>.
- KLEIS-NIELSEN, R. *et al.* (2020). *Navigating the "Infodemic" How People in Six Countries Access and Rate News and Information about Coronavirus*, Reuters Institute and University of Oxford.



- KYUM, D. *et al.* (2020). “An Analysis of Government Communication in the United States During the COVID-19 Pandemic: Recommendations for Effective Government Health Risk Communication”, *World Medical & Health Policy*. Disponible en: <https://doi.org/10.1002/hmh3.363>.
- LEVINSON, P. (2017). *Fake News in Real Context*, Nueva York, Connected Editions.
- 76 MEDIOS PUBLICITARIOS MEXICANOS (2020). *Directorio MPM de medios impresos*. Edición 247, agosto.
- MOHAMMAD, S. y TURNEY, P. (2010). “Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon”, *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*.
- ○  
●
- NOWLIN, M. C. (2016). “Modeling Issue Definitions Using Quantitative Analysis”, *Policy Studies Journal*. 44(3).
- OMS (2018). *Comunicación de riesgos en emergencias de salud pública: directrices de la OMS sobre políticas y prácticas para la comunicación de riesgos en emergencias*, Organización Mundial de la Salud. 16 de mayo. Disponible en: <https://bit.ly/38xmbLk>.
- PAHO (2020). “Understanding the Infodemic and Misinformation in the Fight against COVID-19. Department of Evidence and Intelligence for Action in Health”, *Factsheet*. 6.
- PENNYCOOK, G. *et al.* (2020). *Predictors of Attitudes and Misperceptions about COVID-19 in Canada, the U.K., and the U.S.A.* Disponible en: [osf.io/3a497](https://osf.io/3a497).
- RAINA, P. (2013). “Sentiment Analysis in News Articles Using Sentic Computing”, *2013 IEEE 13th International Conference on Data Mining Workshops*. DOI: 10.1109/ICDMW.2013.27.
- REDDY, B. V. y GUPTA, A. (2020). “Importance of Effective Communication During COVID-19 Infodemic”, *Journal Family Med Prim Care*.
- SANDMAN, P. M. (2012). *Responding to Community Outrage: Strategies for Effective Risk Communication*, AIHA Press.
- SHAHSVARI, S. *et al.* (2020). “Conspiracy in the Time of Corona: Automatic Detection of Emerging COVID-19 Conspiracy Theories in Social Media and the News”, *Journal of Computational Social Science*. 3. Disponible en: <https://doi.org/10.1007/s42001-020-00086-5>.
- VARGAS, E. (2016). “Las funciones del lenguaje de Jakobson en la titulación deportiva: estudio estilístico de *La Nación* y *La Extra*”, *Revista de Filología y Lingüística de la Universidad de Costa Rica*. 41(1). Disponible en: <https://doi.org/10.15517/rfl.v41i1.21194>.

ZANG, L. *et al.* (2020). “Effective Risk Communication for Public Health Emergency: Reflection on the COVID-19 (2019-nCoV) Outbreak in Wuhan, China”, *Healthcare*, Basel (Switzerland). 8(1). Disponible en: <https://doi.org/10.3390/healthcare8010064>.

