

Metodología y tecnología en el análisis de noticias sobre vacunación contra el COVID

Methodology and technology in the analysis of COVID vaccination media news

Metodologia e tecnologia na análise de notícias sobre vacinação contra a COVID

ID Diego Díaz Córdova
 didibart@gmail.com
 Universidad Nacional de Lanús, Argentina

Yamila Comes
 ycomes@gmail.com
 Ministerio de Salud de la Provincia de Buenos Aires,
 Argentina

Mariana Specogna
 marianabs@gmail.com
 Universidad Nacional de Lanús / Universidad
 Nacional de José C. Paz, Argentina

Marina Gabriela Zunino
 ga.zunino2012@gmail.com
 Universidad Nacional de Lanús / Universidad
 Nacional de José C. Paz, Argentina

María José Luzuriaga
 mariajoseluzuriaga@gmail.com
 Universidad Nacional de Lanús, Argentina

Recepción: 01 Febrero 2023
 Aprobación: 15 Mayo 2023
 Publicación: 01 Junio 2023

Cita sugerida: Díaz Córdova, D., Comes, Y., Specogna, M. Zunino, M. G. y Luzuriaga, M. J. (2023). Prosopografía y sociología de los intelectuales. Metodología y tecnología en el análisis de noticias sobre vacunación contra el COVID *Revista Latinoamericana de Metodología de las Ciencias Sociales*, 13(1), e126. <https://doi.org/10.24215/18537863e126>

Resumen: La pandemia de COVID-19 planteó un desafío inédito para la humanidad. Esta situación obligó a la implementación de medidas drásticas como los aislamientos obligatorios, y propició una acelerada carrera por desarrollar y aplicar las vacunas y así poder contener la pandemia. En la Argentina, la implementación de los programas estatales de vacunación masiva permitió que millones de personas accedieran al fármaco y de ese modo pudieron salvar sus vidas. Sin embargo, el programa de vacunación y las propias vacunas no estuvieron al margen de polémicas y de una desconfianza que, a la luz de la gravedad de la enfermedad, se tornaba inquietante. Dentro de este marco, realizamos el proyecto de investigación “Vacilación (inseguridad) en la vacunación contra la COVID-19 en la Provincia de Buenos Aires 2020-2021”, del cual forma parte este artículo, proyecto financiado por el Ministerio de Salud de la Nación (Argentina) con las becas Salud investiga 2022. Uno de los objetivos del proyecto fue el registro y análisis del tratamiento que le dieron los medios de comunicación de masas en el país al tema de la campaña de vacunación. En este caso, el abordaje consistió básicamente en buscar, encontrar y analizar todas las noticias sobre “vacunas” en los principales periódicos del país, con el objetivo de identificar núcleos temáticos, redes conceptuales y principales ideas, y de esta manera establecer un criterio de valoración de las noticias. Se analizaron más de 2000 noticias publicadas entre enero de 2021 y diciembre de 2021 en los periódicos de más audiencia del país. Entre los hallazgos principales se encuentra que la cantidad de noticias no fue constante a lo largo del año sino que aumentó al ritmo de la pandemia, y en esos momentos la centralidad del gobierno se hizo notoria; se observó que la vacuna fue utilizada como un artefacto político, pues en algunos casos se resaltaba la nacionalidad de su origen, mientras que en otros casos se mencionaba el laboratorio; también pudimos apreciar que en general las noticias sobre las vacunas publicadas por los periódicos tuvieron un tono negativo, ya que desconfiaban tanto de su efectividad como de las condiciones de su obtención.

Palabras clave: Vacunación, Big data, COVID.



Abstract: he COVID19 pandemic posed an unprecedented challenge to humanity. his situation required the implementation of drastic measures such as compulsory isolation, and led to an accelerated race to develop and apply vaccines in order to keep the pandemic under control. In Argentina, the implementation of state mass vaccination programs allowed millions of people to have access to the drug and thus save their lives. However, the vaccination program and the vaccines themselves were not free from controversy and mistrust which, in light of the severity of the disease, became a concern. Within this framework, funded by the Ministry of Health of Argentina through the *Salud Investiga 2022* grants, the research project “Hesitancy (insecurity) in vaccination against COVID-19 in the Province of Buenos Aires 2020–2021” was conducted, of which this article is part. One of the project goals was to record and analyze the treatment given by the Argentine mass media to the issue of the vaccination campaign. In this case, the approach consisted basically in searching, finding and analyzing all the news about “vaccines” in the main newspapers of the country, with the aim of identifying themes, conceptual networks and main ideas, and thus being able to establish a criterion for evaluating the news. More than 2000 news items published between January 2021 and December 2021 in the most read newspapers in the country were analyzed. Among the main findings, the amount of news was not constant throughout the year but increased at the pace of the pandemic, moments in which the centrality of the government became prominent; it was observed that the vaccine was used as a political artifact, highlighting in some cases the nationality of its origin, while in other cases the laboratory was mentioned; it was also noted that in general the news about vaccines published by the newspapers had a negative tone, distrusting both their effectiveness and the conditions for obtaining them.

Keywords: Vaccination, Big Data, COVID.

Resumo: A pandemia da COVID-19 representou um desafio sem precedentes para a humanidade. Essa situação forçou a implementação de medidas drásticas, como o isolamento obrigatório, e levou a uma corrida acelerada para desenvolver e aplicar vacinas a fim de conter a pandemia. Na Argentina, a implementação de programas estatais de vacinação em massa deu a milhões de pessoas o acesso ao medicamento e, assim, salvou suas vidas. No entanto, o programa de vacinação e as próprias vacinas não estavam livres de controvérsias e desconfianças, o que, à luz da gravidade da doença, tornou-se preocupante. Nesse contexto, realizamos o projeto de pesquisa “Hesitação (insegurança) na vacinação contra a COVID-19 na Província de Buenos Aires 2020 – 2021”, do qual este artigo faz parte, financiado pelo Ministério da Saúde da Nação (Argentina) com as bolsas *Salud Investiga 2022*. Um dos objetivos do projeto era registrar e analisar o tratamento dado pela grande mídia sobre a campanha de vacinação no país. Nesse caso, a abordagem consistiu basicamente na busca, localização e análise de todas as notícias sobre “vacinas” nos principais jornais do país, visando identificar núcleos temáticos, redes conceituais e ideias principais e, assim, poder estabelecer um critério de avaliação das notícias.

Foram analisadas mais de 2.000 notícias publicadas entre janeiro de 2021 e dezembro de 2021 nos jornais mais lidos do país. Entre as principais descobertas estão que a quantidade de notícias não foi constante ao longo do ano, mas aumentou no ritmo da pandemia, nesses momentos a centralidade do governo se tornou notória e se observou que a vacina foi utilizada como artefato político, destacando em alguns casos a nacionalidade de sua origem, enquanto que em outros casos o laboratório foi mencionado; também pudemos perceber que, em geral, as notícias sobre vacinas publicadas pelos jornais tinham um tom negativo, desconfiando tanto de sua eficácia quanto das condições de obtenção.

Palavras-chave: Vacinação, Big data, COVID.

1. PRESENTACIÓN DEL PROBLEMA

En el proyecto de investigación sobre confianza, complacencia y conveniencia de la vacuna contra la COVID-19 se planteó la necesidad de un objetivo que incluyera un análisis de los medios de comunicación en relación con la campaña de vacunación contra la COVID-19. Para este objetivo se seleccionaron los principales periódicos del país (*La Nación, Clarín, Infobae y Página 12*), debido a que son los que concentran la mayor cantidad de lecturas. Si bien hoy en día existe mucha información que circula por otros ámbitos (redes en internet y sistemas de mensajería), y que la gente utiliza para informarse sobre la salud, decidimos concentrarnos en esos periódicos en función de la audiencia masiva pero también del impacto político que pueden llegar a tener.

El objetivo específico planteaba analizar las noticias que habían publicado estos medios en relación con la campaña de vacunación para ponderar sus valoraciones y sintetizar los resultados. Esta campaña incluía tanto la obtención y compra de vacunas (de diferentes naturalezas, orígenes y laboratorios) como la implementación de la vacunación propiamente dicha, que se realizó a lo largo y ancho de todo el país. En particular, nos concentramos en las noticias que mencionaban este proceso complejo e inédito, pero en nuestro caso circunscripto a la provincia de Buenos Aires.

Se sabe del poder que ejercen los medios a la hora de crear la configuración de noticias que son vistas por la población. La teoría de la “configuración de la agenda” (agenda *setting*) señala que los medios seleccionan algunos temas por sobre otros y que deciden cuánto espacio (o tiempo) le dedican a cada tema (Rubio Ferreres, 2009). Esta selección de contenidos y de frecuencia está influida por una multiplicidad de factores, desde los intereses propios de las empresas que son dueñas de los medios hasta los periodistas, redactores y productores que estampan su propia marca en las noticias informadas (McQuail, 2000). Los medios delimitan la información que obtiene el público (sobre todo, aquella que no está al alcance de la experiencia propia) pero de ningún modo pueden establecer una interpretación unívoca acorde a sus intereses. Esta información suministrada no es puramente objetiva, sino que se encuentra encuadrada en un conjunto de emociones, explicaciones y presentaciones que le dan un contexto. Los medios no pueden hacer que la gente piense de una determinada manera, pero pueden instalar ciertos temas como más relevantes que otros (Aruguete, 2017).

La teoría de la “configuración de la agenda” no ha estado exenta de críticas. Los principales puntos de conflicto se centran alrededor del proceso, la identidad y el contexto. El primero hace referencia a la forma en que se configura esa agenda; es decir a la presunción equivocada de que este proceso es automático y sin reflexión. El segundo punto está vinculado con ciertas conceptualizaciones que la hacen indistinguible de la

teoría del encuadre. Esta última teoría señala la importancia de los marcos (*frames*) que utilizan las personas para interpretar el mundo (Aruguete, 2017). El tercero implica el cambio que significó la llegada de Internet y de las nuevas formas de comunicación, que redujeron el poder de los medios tradicionales (Takeshita, 2006).

Otra crítica viene de la mano de aquellos intentos de combinación de la “configuración de la agenda” y la teoría del “encuadre” (*framing*). El *framing* plantea una perspectiva epistemológica opuesta al positivismo característico de la agenda *setting*, cercana a la fenomenología y a la etnometodología; es decir, tomando en cuenta los contextos, los “marcos” de referencia y las condiciones de producción tanto de las noticias como de quienes las interpretan (Aruguete, 2017).

Es ocioso, por tanto, pretender asumir que los medios de comunicación tienen un poder absoluto a la hora de instalar temas o de controlar las interpretaciones que su audiencia les adjudica. De hecho, lo que se plantea es que, si bien es cierto que la selección de temáticas, pero también del tono (marcadores afectivos) de las noticias es potestad de los medios, las interpretaciones que las audiencias harán serán, por un lado, activas, pero por el otro diferentes según las características sociales, económicas, culturales, de género o étnicas.

El propio hecho de la vacunación no fue unívoco en función tanto del momento de la pandemia (que no fue constante sino marcado por picos y valles en los contagiados) como de los complejos procesos de compras y de aplicación de la vacuna. Durante la implementación de la campaña, esta heterogeneidad se vio reflejada en los resultados de los análisis que se llevaron a cabo. Los medios no estuvieron al margen de los impactos epidemiológicos.

En una campaña de vacunación la información cumple un papel central. Hay trabajos que muestran que sin confianza cae la cobertura de vacunación (González-Block, Gutiérrez-Calderón, Pelcastre-Villafuerte, Arroyo-Laguna, Comes, Crocco, Fachel-Leal, Noboa, Riva-Knauth, Rodríguez-Zea, Ruoti, Sarti, Puentes-Rosas, 2020), y una de las formas de adquirir confianza es analizando información consistente y coherente. En general, puede observarse que los medios de comunicación tienden a mostrar noticias que muchas veces pueden contradecirse entre ellas, que no necesariamente tienen un fundamento científico o a veces ni siquiera presentan racionalidad o de concordancia entre el título y el contenido, entre otras características de presentación de las informaciones.

El desafío sobre cómo recuperar todas esas noticias y de qué forma analizarlas fue uno de los motores para pensar y estructurar la metodología. Resultó claro que los procedimientos elegidos no podían ser manuales y se recurrió, así, a técnicas de *scraping* (Marres y Weltevrede, 2013) y *crawling* (Pant, Srinivasan, Menczer, 2004; Catanese, De Meo, Ferrara, Fiumara, Provetti, 2011), para luego proceder a aplicar algoritmos de minería de texto (o análisis lexical).

Respecto de la minería de texto, señalaremos que interesaba no sólo contar con las relaciones existentes entre cada noticia y las palabras centrales, sino también realizar algún tipo de análisis de sentimiento (Zhang, Wang, Liu, 2018). Este tipo de análisis permite ponderar en forma aproximada una valoración negativa o positiva en relación con un tópico de interés. Si bien los algoritmos no son sutiles, muestran tendencias generales que permiten inferir el posicionamiento con respecto, en este caso, al tema de la campaña de vacunación contra el COVID-19. De lo que se trata es de expresar la actitud general con respecto al texto, a partir del uso de metodologías de *deep learning*, que analizan millones de datos e infieren los resultados a partir de una red neuronal.

En este caso particular, nos interesó analizar, a modo de preguntas guía, en primer lugar, si la cantidad de noticias referidas a lo largo del período estudiado fue constante o fue variando; en segundo lugar, si los tópicos de los que trataban las noticias se mantuvieron o se fueron modificando a medida que avanzaba la campaña de vacunación; por último, nos interesó conocer el tono de las noticias, a partir del análisis de sentimiento: saber si ese tono fue constante o también se modificó a lo largo del año.

2. METODOLOGÍA

Se trató de un análisis de información desde una perspectiva de análisis de los contenidos publicados por los periódicos escogidos. Detrás de estos contenidos pueden vislumbrarse las estrategias que utilizan algunos grupos de poder para legitimar ciertas opciones en detrimento de otras, así como para indagar en los ritmos y selecciones que realizan a la hora de presentar las noticias (van Dijk, 2003).

El primer problema a resolver fue el de encontrar todas las noticias que necesitábamos para analizar. En un principio se armó una serie de algoritmos para que buscaran en la página web de cada periódico y bajar la información a partir del hallazgo. Eso implicaba conocer *a priori* las arquitecturas de cada sitio de internet (si bien se parecen entre sí, no necesariamente son similares, y no es posible aplicar el mismo algoritmo a uno u otro sitio). A partir de este obstáculo, se buscó otra solución y fue encontrada la base de datos GDELT.

El proyecto GDELT (Base de datos global de eventos, lenguaje y tonos) fue creado por Kalev Leetaru y la Universidad de Georgetown con el objetivo de construir un catálogo dinámico (actualizado cada 15 minutos) de todo lo que ocurre en el mundo (qué es lo que sucede, quiénes están involucrados, cuál es el contexto del evento, etc.) (Leetaru, 2014). Los datos disponibles comienzan en 1979 y llegan hasta la actualidad. Las bases de datos están disponibles en la consola de Google y son asequibles a través de la plataforma Google BigQuery (<https://console.cloud.google.com/>), mediante consultas SQL (structured query language).

Esta base de datos permite acceder a los enlaces de las diferentes páginas web que son pertinentes en cada búsqueda. No se accede al artículo en sí mismo. El proyecto GDELT tiene, entre otros campos, el del *Tone*; esto es, un análisis de sentimiento acerca del tono de cada noticia o suceso. Los valores del *Tone* van de -100 (extremadamente negativo) a + 100 (extremadamente positivo), aunque la mayor parte de las noticias siempre oscilan en un rango de -10 a + 10. La amplitud de la escala responde al objetivo del GDELT, para el cual el conflicto y la guerra en el mundo son los tópicos de interés. De allí que el rango sea tan amplio, pero lo cierto es que la mayoría de los artículos que se publican en el mundo no superan los -10 o +10. Este análisis permite valorar el sentimiento puesto en juego en cada noticia y es un buen indicador del concepto que tiene el medio que publicó la nota sobre el tema en cuestión.

Dado el tamaño de estas bases de datos, fueron utilizadas las tablas *partitioned*, que se encuentran segmentadas por período temporal. En nuestro caso, resolvimos bajar información por mes, entre enero de 2021 y diciembre de 2021. Este período estuvo determinado por el lanzamiento de la campaña de vacunación en el país, que comenzó los últimos días de 2020.

Los medios en los que fueron realizadas las búsquedas son los mencionados en los párrafos anteriores. Las fechas (en este caso de ejemplo, se corresponden a enero de 2021) se fueron modificando para capturar las noticias mes a mes. Por último, usamos el campo V2Themes, que es la forma en la que GDELT categoriza los sucesos. Las búsquedas las hicimos en función de **vacunación** y **salud**. Las repetimos tres veces, ya que es la forma en la que GDELT garantiza que la noticia hable del tema; podría ser que una noticia que no trate del tema tuviera esa etiqueta, pero dispuesta una sola vez. También usamos las categorías de **tecnología** y **farmacéuticas**.

Una vez obtenidos los enlaces, procedimos a “limpiarlos”, ya que varios de los vínculos que se encontraron no respondían a la consulta solicitada (esto sucedió, por ejemplo, con el enlace a la página *Home* del sitio web del periódico que, como tenía varias notas sobre el tema, quedó catalogado con esas características, pero que a efectos de esta investigación no nos interesaba; i.e., eliminamos el enlace lanacion.com.ar, ya que es sólo el enlace a la *Home* del sitio web).

De este modo obtuvimos los vínculos que efectivamente hablaban de las noticias sobre la vacunación. Estas expresiones regulares se fueron modificando en función del mes, ya que los artículos iban cambiando a lo largo del año.

Una vez limpios los enlaces, usamos el paquete del lenguaje Rvest (Wickham, 2022) para realizar el *crawling* y el *scraping* correspondiente. Con el *crawling* bajamos todos los artículos desde los sitios web y con

el *scraping* limpiamos las páginas de las etiquetas HTML; así, nos quedamos sólo con el texto, que es lo que nos interesaba analizar.

Luego de este procedimiento, ya teníamos todas las noticias en archivos de texto separados, ya limpios y sin etiquetas. Procedimos a agregar las variables que son necesarias para usar los textos en Iramuteq (Camargo y Justo, 2013), con el cual realizamos el análisis lexical, o minería de texto según la tradición anglosajona. Para ello, implementamos un algoritmo realizado en IDE Power Shell. En cada comienzo de artículo agregamos la línea correspondiente con el asterisco y el nombre de la variable (es importante el guion bajo luego del nombre de la variable).

Para el análisis de los artículos de medios fue utilizada la metodología de análisis lexical (Menandro et al, 2006), con el software Iramuteq (Retinaud, 2009), mencionado anteriormente. El desarrollo del software Iramuteq fue realizado por Pierre Retinaud, miembro del Laboratoire d'Études et de Recherches Appliquées en Sciences Sociales.

Luego, con un algoritmo realizado en IDE Power Shell, reunimos todos esos textos en uno solo, ya que así necesita tener los datos de entrada el programa Iramuteq.

Una vez que tuvimos todos los artículos en un solo texto, procedimos a limpiar aquellas palabras que no tenían que ver con el tema; es decir, las menciones a los periódicos, los menús, etc. Para ello, usamos expresiones regulares implementadas en IDE Power Shell.

Una vez realizada esta tarea, abrimos el documento único, lo salvamos con codificación UTF-8 y agregamos un renglón en blanco al comienzo y los cuatro asteriscos que indican el comienzo del texto a analizar. Ahora ya estamos en condiciones de analizar los textos con el Iramuteq.

Cuando encontramos un error en Iramuteq con el método Reinert, se pueden verificar dos cosas. Por un lado, revisar en el archivo general de texto que sirve como *input* si hay signos como “@”, “;” o “\ o /” que puedan perjudicar la ejecución del algoritmo. Por otro lado, se pueden modificar los parámetros de Iramuteq del método Reinert, bajando el máximo de 10 a 5 “Número de clases terminales de la fase 1” y subiendo de 0 a 2 la “Frecuencia mínima de segmentos de texto por clase”.

Un aspecto a considerar es que no todas las notas bajadas con GDELT son diferentes, ya que los periódicos mantienen las notas durante varios días (incluso, suelen ser subidas nuevamente luego de un tiempo). Otra cuestión a mencionar es que muchas veces distintos periódicos publican la misma nota, lo que suele suceder cuando la nota fue originada en alguna de las agencias de prensa internacionales (Reuters, AFP, etc.).

Todo el código utilizado en este proyecto está disponible en el sitio GitHub (<https://github.com/didibart/SaludInvestiga2022> y <https://github.com/didibart/scriptsPS>).

Estos tipos de análisis trabajan con métodos estadísticos que permiten analizar las palabras en un texto a partir de compartir núcleos o familias lexicales, determinadas por la proximidad discursiva entre ellas. Con ellos se pueden observar las proximidades entre palabras en un gran volumen de datos, e interpretar que esta cercanía está determinada por una cercanía en la frase y, como consecuencia, que comparten la misma temática.

Para este análisis, fueron utilizadas algunas técnicas (Wachelke, 2012; Camargo y Justo, 2013), como la nube de palabras, que nos trae la frecuencia de aparición de cada palabra lematizada (llevada a su raíz lexical); el análisis de similitud, que exhibe las diferentes temáticas, así como las relaciones, tanto de jerarquía como de proximidad, entre palabras; y la Clasificación jerárquica descendiente o método de Reinert, por el cual se observan ejes temáticos dentro de un discurso y la jerarquía que poseen esos ejes en la estructura gramatical del discurso.

Estas estrategias articulan el análisis gramatical y semántico, y la visualización de ejes de poder y de importancia dentro de las noticias publicadas, lo que permite comprender el carácter que se le dio a la campaña de vacunación contra el COVID-19.

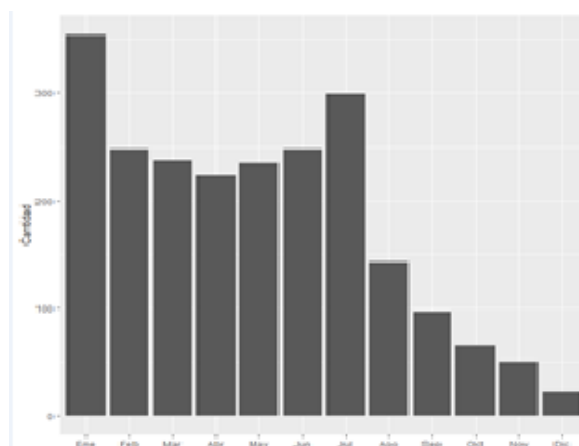
3. RESULTADOS

Dado que el análisis lexical conlleva una gran cantidad de técnicas diferentes que apuntan a distintas representaciones, se presenta por separado cada una de ellas, y se intenta integrarlas en un análisis final que se presenta en las conclusiones.

3.1 Análisis de artículos publicados en los medios

Fueron encontradas 2222 noticias. En el análisis por mes se observa que la cantidad de notas fue variando notablemente; en general, coinciden el mayor acumulado con el comienzo de la vacunación y con el pico de contagio. Enero de 2021 es el mes en el que más notas se publicaron al respecto, cuando la campaña de vacunación empezaba; también, julio de 2021, cuando llegamos al pico de la tercera ola.

FIGURA 1
Cantidad de noticias encontradas por mes en periódicos seleccionados – Año 2021, Argentina



3.2 Nube de palabras

La nube de palabras determina la frecuencia en que una palabra es evocada, en relación con el total de las evocaciones. Dado que el software trabaja con una lógica de lematización, es decir llevar las palabras a su raíz léxica, si una palabra tiene la misma raíz la incorpora de manera genérica, aunque sea expresada con declinaciones diferentes (Camargo y Justo, 2013). Otro proceso que aplica el software es la diferenciación entre formas activas y pasivas; las pasivas son palabras que no aportan contenido sino conectividad al lenguaje. Estas son descartadas para los análisis (Camargo y Justo, 2013).

Se observa en las nubes de palabras desde enero a diciembre cómo, por primera vez en la historia, se visibiliza la vacuna como un artefacto político e ideológico. Si bien siempre lo ha sido, en campañas anteriores parecía un dispositivo técnico ligado a la salud. En esta pandemia se ha politizado lo técnico o se ha visibilizado que toda técnica es política. Mostramos las nubes de palabras correspondientes a los meses de enero, junio y diciembre.

FIGURA 2
Nube de palabras - Enero 2021



FIGURA 3
Nube de palabras - Junio 2021

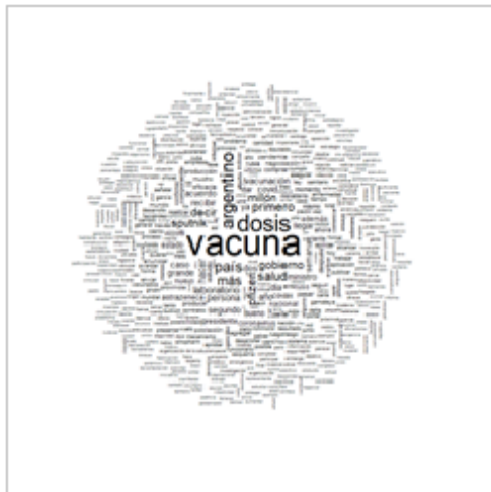


FIGURA 4
Nube de palabras - Diciembre 2021



Puede observarse que, cuando hay brote, los medios otorgan mayor centralidad a noticias del gobierno y que, cuando bajaron los casos, tanto el gobierno como los ministros se tornan menos centrales. Los medios introdujeron noticias sobre la vacuna Sputnik durante todo el año, y les dieron menor centralidad a Pfizer y AstraZeneca. La vacuna Sinopharm también tuvo mayor protagonismo que las otras dos en algunos meses.

Los medios analizados se refieren en las noticias, de manera continua, a *los rusos*, debido a la negociación para conseguir la vacuna Sputnik. Sin embargo, no se refieren de la misma manera a los norteamericanos o a los ingleses en relación con las otras vacunas. Cuando la vacuna Sinopharm adquiere preponderancia, aparece la palabra “China”.

La preocupación mediática por la eficacia aparece al inicio del año (febrero) y desaparece en el momento del brote (abril de 2021).

En abril, momento de la segunda ola, aparece la palabra *hacer*, ligada al concepto de gestión y gobierno.

En abril, y hasta junio, momento del pico de la segunda ola, surgen de manera protagónica las palabras “gobierno” y “ministro”, y el nombre de la ministra de salud. El gobierno se torna secundario en julio, cuando bajan drásticamente los casos y la gravedad, y vuelve a aparecer con la tercera ola de diciembre. Parecería que, cuando subieron los casos, se habló más del gobierno que cuando bajaron.

3.3 Clasificación jerárquica descendente (método de Reinert)

El método de Reinert muestra con mayor detalle los temas incluidos en cada clúster. Se observan, al igual que en los análisis anteriores, algunos temas técnicos, como la relación con las farmacéuticas, la pandemia, su propagación, las vacunas, sus laboratorios, la producción en la Argentina, los efectos, las variantes, las dosis, la vacunación de niños, la eficacia y seguridad, y las pruebas diagnósticas. Entre los temas más politizados figuran el gobierno, ministro, ministerio, compra de vacunas, la relación con las provincias, la cantidad de dosis y las decisiones del ministerio.

El vínculo lexical entre vacunas y gobierno se mantiene durante todo el año, aunque con intensidades diferentes. Como en los análisis anteriores, se observa que baja en julio y en noviembre surge nuevamente, aunque sin la misma centralidad que durante la segunda ola.

La eficacia y seguridad se observa como discurso destacado hasta el mes de junio, pico de la segunda ola.

En diciembre los discursos se diversifican en los medios. Esta alta dispersión puede deberse al hecho de que el número de noticias vino en descenso y, debido a ello, con pocos casos se pueden obtener muchas clases. Mostramos las gráficas de los meses de enero, junio y diciembre.

FIGURA 5
Clasificación Jerárquica Descendente - Enero 2021

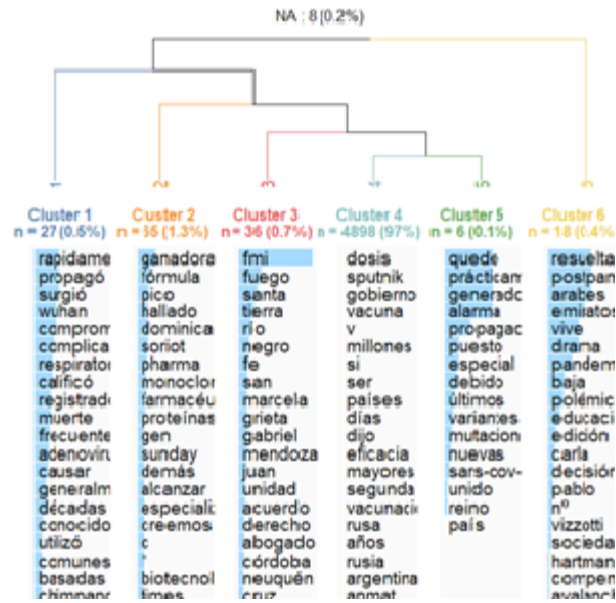
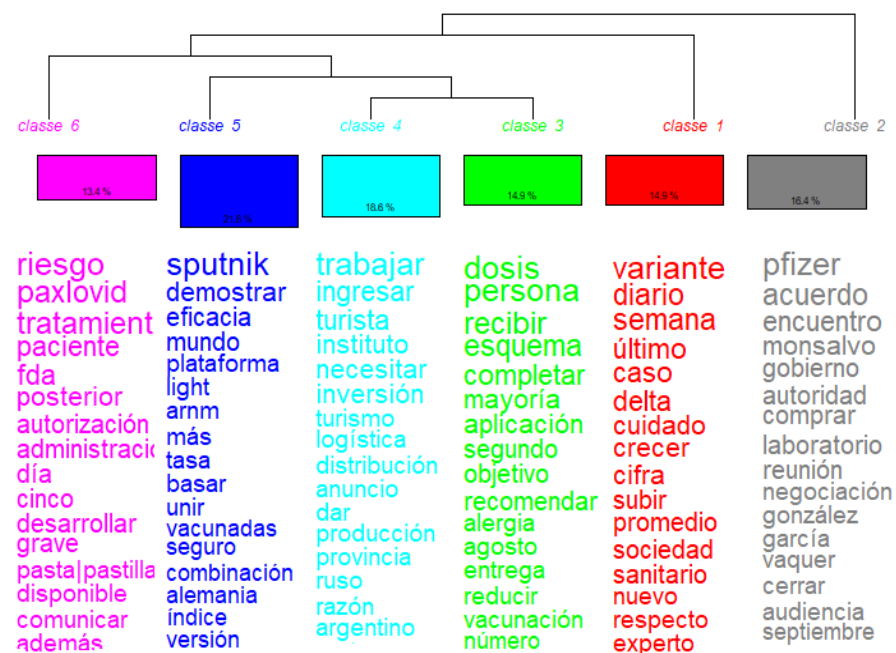


FIGURA 6
Clasificación Jerárquica Descendente - Junio 20



FIGURA 7
Clasificación Jerárquica Descendente - Diciembre 2021



3.4 Análisis del “tono” (sentimiento) de las noticias

No se observa un patrón homogéneo entre el periódico y el tono, sino que, a lo largo de los meses, van cambiando el tono entre positivo y negativo. En enero el único periódico positivo fue *Infobae*. En febrero todos son positivos menos *Clarín*. En marzo todos tornan sus noticias en modo negativo, coincidente con el inicio de la segunda ola. Ya en abril continúan negativos, y en mayo es *Página 12* el que posee un tono positivo en sus noticias. En junio, pleno pico del brote, todos poseen tono negativo. En julio, cuando empiezan a bajar los casos, se observan todos positivos menos *La Nación*. En agosto, prácticamente todos poseen tono negativo. En septiembre, *La Nación* está mínimamente positivo; en octubre, *Clarín* y, en mayor medida, *Página 12*. En noviembre, *La Nación* e *Infobae*, y en diciembre todos son positivos menos *La Nación*. Mostramos las gráficas de los meses enero, junio y diciembre.

FIGURA 8
Análisis de Sentimiento - Enero 2021

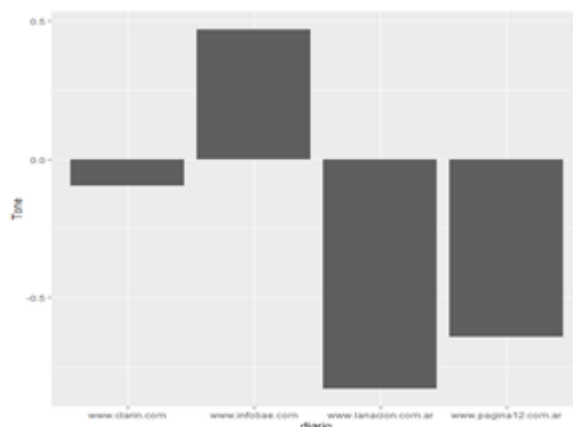


FIGURA 9
Análisis de Sentimiento - Junio 2021

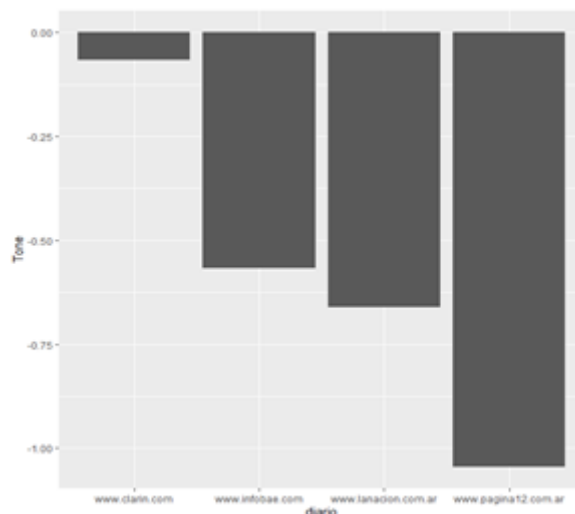
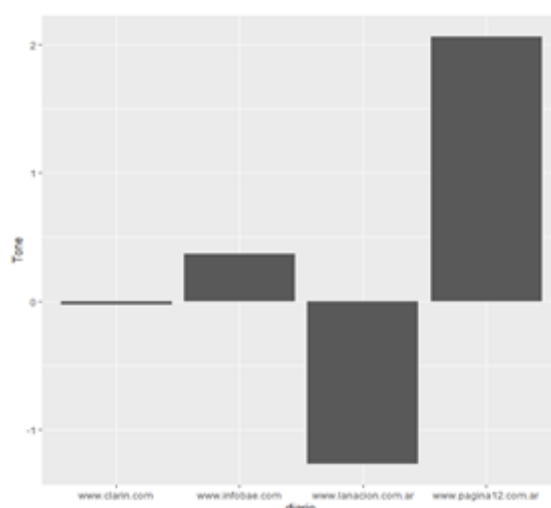


FIGURA 10
Análisis de Sentimiento - Diciembre 2021



4. CONCLUSIONES

La campaña y el programa de vacunación realizados en torno al COVID-19 fueron únicos en la historia. Nunca antes se habían aplicado tantas dosis en tan poco tiempo. Esto implicó un esfuerzo logístico inédito que redundó en una notable baja de los contagios durante el final del año 2021 y todo el 2022. En la Argentina, la vacunación logró cubrir a una parte muy importante de la población, y se llegó a estar entre los primeros lugares mundiales en porcentaje de vacunados. Esto habla de una población, en líneas generales, receptiva de las vacunas (a diferencia de otros lugares en el mundo donde había provisión de vacunas, pero poca predisposición a vacunarse).

Pese a la alta cobertura de inmunización, un sector de la población se mantuvo expectante y con dudas sobre la eficacia de la vacuna, así como con temor a los efectos adversos. Entre los datos necesarios para indagar en las causas y razones de la vacilación en la vacunación, consideramos importante revisar las noticias que

los grandes medios del país difundieron sobre la campaña. Se sabe que los grandes medios marcan agenda de temas, si bien los receptores de esas noticias poseen un carácter dinámico en cuanto a las interpretaciones.

En este caso, los análisis estuvieron concentrados en la emisión de las noticias y no en la recepción. El estudio se realizó en función de las noticias publicadas en los diarios mencionados. Nos interesó conocer la naturaleza de las publicaciones, tanto en sus aspectos relacionales como en su valoración axiológica. En ese sentido, aplicamos los análisis lexicales, mediante el uso del software Iramuteq, que permiten representar aquellos conceptos que conforman diferentes tipos de centralidad. Por otro lado, utilizamos los valores de la base de datos GDELT para obtener el sentimiento (positivo o negativo) de la noticia (en función de la ocurrencia de ciertas palabras que denotan positividad o negatividad).

La base de datos GDELT fue un hallazgo que se realizó mientras se llevaba a cabo la investigación. Consideramos que puede ser de mucha utilidad para quien estudia no sólo los medios masivos de comunicación, sino también lo que se publica en la web en general. Por ello, dedicamos especial interés a relatar, en el apartado metodológico, los detalles de la búsqueda, la recuperación y el procesamiento de los datos, y por ello publicamos también los algoritmos que utilizamos para realizar la tarea.

A grandes rasgos, podemos concluir que los medios de comunicación estudiados se movieron al ritmo de lo que sucedía a nivel epidemiológico. Cuando los casos y la mortalidad se hacían más altos, los sentimientos negativos también crecían. Cuando las vacunas empezaron a mostrar sus efectos, crecieron los sentimientos positivos.

En la construcción del corpus de noticias, tomamos en cuenta todas aquellas que se referían a las vacunas contra el COVID-19 en general, y ello incluyó las notas sobre la producción y desarrollo de vacunas, la compra y la logística, así como sobre la eficacia y los efectos colaterales. La cuestión política estuvo todo el tiempo presente en mayor o menor grado, y no podría haber sido de otra manera, debido a que una pandemia es, además de un problema de salud pública, un problema político. La pandemia afectó todos los niveles del Estado (nacional, provincial y municipal), las relaciones entre los Estados, el sector privado y, por supuesto, a la población en general. En términos de la sociología clásica, fue un “hecho social total” (Durkheim, 2005), y eso se vio reflejado en los periódicos estudiados.

Fue interesante el hallazgo de que no necesariamente la tendencia política del periódico determinó el sentimiento mostrado. *A priori* hubiéramos pensado que aquel diario más cercano al gobierno podría haber tenido una mirada más positiva sobre las vacunas y que aquellos que se ubican en el espectro opositor habrían tenido una mirada más negativa. Sin embargo, en algunos meses aquellos diarios más cercanos al gobierno tuvieron una mirada negativa y aquellos opositores tuvieron una mirada positiva. En otros meses todos los periódicos tuvieron una mirada negativa y no hubo un solo mes en el que todos tuvieran una mirada positiva. Una pregunta que nos surge es hasta qué punto hay coherencia y consistencia entre el título, el copete o la bajada y el cuerpo de la nota; es decir, hasta qué punto se titula una cosa y luego la nota dice lo contrario. Este tipo de inconsistencias puede sesgar el análisis (además de tirar por la borda la credibilidad del medio).

Es plausible que la lógica del mercado sea la que rige las empresas de los medios de comunicación y es también bastante evidente que no podría ser de otra forma. Pero también queda muy claro que, incluso más allá de sus intereses comerciales, hay grandes falencias en la comunicación. Información contradictoria, sin criterios de diferenciación entre lo plausible y lo distorsionado, con un gran sesgo que muchas veces determina la publicación de *fake news*, y noticias mal escritas, tanto desde un punto de vista conceptual como sintáctico, e incluso con errores de ortografía.

En relación con la pandemia y la vacunación, queda claro que los diarios más importantes del país no estuvieron a la altura del inmenso desafío que significó para toda la población la experiencia del COVID-19. No estuvieron a la altura en cuanto a informar de manera confiable a la población, y eso se observa en el análisis realizado, en el que las noticias siguieron el ritmo epidemiológico, muy diferente del ritmo preventivo que la situación y los profesionales de la salud exigían. Si la vacunación en la Argentina fue un éxito, y así

parece en función de los porcentajes de vacunados, no fue gracias a la tarea que realizaron los diarios más grandes del país.

REFERENCIAS

- Aruguete, N. (2017). Agenda Setting y Framing: un debate teórico inconcluso. *Más Poder Local*, 30, 36-42.
- Camargo, B. V. y Justo, A. M. (2013). IRAMUTEQ: Um software gratuito para análise de dados textuais. *Temas em Psicologia*, 21(2), 513–518. <https://doi.org/10.9788/TP2013.2-16>
- Catanese, S. A., De Meo, P., Ferrara, E., Fiumara, G. y Proveti, A. (2011). Crawling Facebook for social network analysis purposes. *Proceedings of the International Conference on Web Intelligence, Mining and Semantics – WIMS*, 11. <https://doi.org/10.1145/1988688.1988749>
- Dijk, T. A. van (2003). *Ideología y discurso: Una introducción multidisciplinaria*. Barcelona: Ariel.
- Durkheim, É. (2005). *Las reglas del método sociológico*. Madrid: Biblioteca Nueva. <http://www.digitaliapublishing.com/a/5181>
- González-Block, M. Á., Gutiérrez-Calderón, E., Pelcastre-Villafuerte, B. E., Arroyo-Laguna, J., Comes, Y., Crocco, P., Fachel-Leal, A., Noboa, L., Riva-Knauth, D., Rodríguez-Zea, B., Ruoti, M., Sarti, E. y Puentes-Rosas, E. (2020). Influenza vaccination hesitancy in five countries of South America. Confidence, complacency and convenience as determinants of immunization rates. *PLOS ONE*, 15(12), e0243833. <https://doi.org/10.1371/journal.pone.0243833>
- Lectaru, K. (2014). The possibility of global data sets. *Journal of International Affairs*, 68(1), 215–219.
- Marres, N. y Weltevrede, E. (2013). Scraping the social?: Issues in live social research. *Journal of Cultural Economy*, 6(3), 313–335. <https://doi.org/10.1080/17530350.2013.772070>
- Pant, G., Srinivasan, P. y Menczer, F. (2004). Crawling the Web. En M. Levene y A. Poulouvasilis, *Web Dynamics* (pp. 153-177). Berlin: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-10874-1_7
- Retinaud, P. (2009). *Iramuteq: Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires (0.7 alpha2)*. Recuperado de: <https://www.iramuteq.org>
- Takeshita, T. (2006). Current Critical Problems in Agenda-Setting Research. *International Journal of Public Opinion Research*, 18(3), 275–296. <https://doi.org/10.1093/ijpor/edh104>
- Wachelke, J. (2012). Social Representations: A Review of Theory and Research from the Structural Approach. *Universitas Psychologica*, 11, 729–741.
- Wickham, H. (2022). *rvest: Easily Harvest (Scrape) Web Pages. R package version 1.0.2*. Recuperado de: <https://CRAN.R-project.org/package=rvest>
- Zhang, L., Wang, S. y Liu, B. (2018). Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery*, 8(4). <https://doi.org/10.1002/widm.1253>