



---

**Metodología híbrida para el análisis de discursos sobre lo público común en plataformas<sup>1</sup>**  
**Hybrid methodology for the analysis of discourses about the common public on platforms**

Irene Lis Gindin

Natalia Raimondo Anselmino

**Palavras-chave:** Plataformización; Semiodata; Discursos.

### **1. Introducción**

En el marco del Grupo de Trabajo “Mediatización y metodologías híbridas”, esta presentación se propone dar cuenta de una experiencia de investigación, actualmente en curso, en torno a la plataformización (van Dijck, Poell y de Waal, 2018) de discursos sobre lo público común en nuestras sociedades contemporáneas<sup>2</sup>. La investigación mencionada se propone caracterizar las estrategias enunciativas (AUTOR, 2011) empleadas en los discursos sobre lo público-común al ser publicados en plataformas mediáticas, y sus flujos de sentido (Carlón, 2020). Con ese horizonte, se delimitó un caso único de carácter instrumental (Marradi, Archenti y Piovani, 2018) a partir del cual fue posible cubrir varios intereses: que fuera relevante, reciente (sobre todo, en función de que la distancia temporal con el acontecimiento no obstaculice la recopilación del corpus de materiales a considerar) y, fundamentalmente, que sea pertinente al objeto de estudio.

---

<sup>1</sup> Trabalho apresentado ao VII Seminário Internacional de Pesquisas em Mídia e Processos Sociais. POSCOM-UFSM. Santa Maria, RS. ECA-USP. São Paulo/SP.

<sup>2</sup> Nota al pie anonimizada con el fin de garantizar la evaluación doble ciego.



---

El caso construido para esta investigación comprende un conjunto de discursos circulantes a través de distintas plataformas (portales de prensa online, y cuentas de Instagram, X, Facebook y Whatsapp) en torno a una serie de hechos violentos que ocurrieron en la ciudad de Rosario (Provincia de Santa Fe, Argentina), vinculados todos al llamado "narcoterrorismo". Aceptando que la delimitación de cualquier caso supone una construcción propia de la investigación, un determinado "recorte" que tiene como finalidad circunscribir un dominio empírico con propiedad de sistema, se decidió que el arco temporal del caso iniciara el 2 de marzo 2024 y finalizara el 6 de mayo de ese mismo año. En términos generales, el caso trata de una escalada de hechos delictivos caracterizados por utilizar la fuerza organizada de bandas criminales de la ciudad como mecanismo de coerción dirigido, explícitamente, al poder político provincial; hechos que no sólo desencadenaron cuatro muertes, sino que generaron una situación de alarma tan generalizada en la población que paralizaron el funcionamiento de diversas instituciones sociales (escuelas, centros de salud, transporte público, comercios, etc.) y requirieron la intervención del gobierno nacional, las fuerzas de seguridad federales y los distintos poderes del Estado, con consecuencias que llegan hasta la actualidad y repercuten, incluso, por fuera de las fronteras locales<sup>3</sup>.

Como meta de esta ponencia, nos proponemos exponer los procedimientos llevados a cabo para la recopilación del corpus de trabajo, junto a la explicitación de las potencialidades y limitaciones que hemos encontrado en el camino hasta aquí recorrido para producir conocimiento respecto de un fenómeno social complejo, en un contexto de desfinanciamiento del sistema científico argentino, gracias a la labor interdisciplinaria e interinstitucional que integra saberes provenientes de las ciencias

---

<sup>3</sup> Una secuencia pormenorizada de los acontecimientos a los que refieren los discursos comprendidos en el caso se encuentra desarrollada en AUTOR (2024).



---

sociales (como la sociología, la semiótica o la antropología) con otros derivados de las ciencias computacionales.

## 2. Notas sobre el abordaje metodológico

El objetivo general expuesto previamente se acompaña con un conjunto de objetivos particulares que serán abordados a partir de la estrategia metodológica híbrida nombrada como *semiodata* (AUTOR, 2022). Se trata de la combinación interdisciplinaria del estudio empírico de la puesta en discurso desde un punto de vista sociosemiótico (Verón, 1998) junto al empleo de herramientas y técnicas computacionales diversas, y el análisis univariante y multivariante de datos y metadatos (Aldas y Uriel, 2017). Además, se utiliza el modelo teórico-metodológico propuesto por Carlón (2020) en lo que atañe a la distinción entre flujos de sentido con modo descendente y ascendente (esto es, de los medios tradicionales a las plataformas mediáticas denominadas usualmente como “redes”, o viceversa) así como entre flujos de adentro hacia afuera o de afuera hacia adentro (tomando en consideración actores sociales diversos así como colectivos e instituciones).

Con el fin de que se comprenda la complejidad y dimensión del caso, se pasará a describir ahora el paquete de textos digitales (junto a sus datos y metadatos) con el que se está trabajando. Al momento de presentación de este resumen, el equipo cuenta con los siguientes datasets construidos y/o en proceso de refinamiento:

- a) Portales web: notas publicadas en tres portales de prensa locales (lacapital, rosario3, elciudadanoweb) y tres nacionales (lanacion, infobae y clarin), así como los comentarios de usuarios-lectores a las mismas (cuando los hubiere);
- b) X: 20.039 tweets, conformado por tweets publicados por cuentas relevantes para el estudio y otros capturados dado que contenían sintagmas clave relacionados con el caso en cuestión;
- c) Whatsapp: 218 intervenciones que combinan texto, imagen, video y voz;



---

d) Meta Platforms: dos datasets conformados por registros de Instagram y Facebook. Cada uno de ellos cuenta con entradas correspondientes a portales locales y nacionales, y publicaciones de referentes políticos de alcance local y nacional.

Como se puede observar, se trata de discursos de diferentes actores, publicados en distintas plataformas y que requieren herramientas y técnicas diversas para su recuperación y disposición en datasets y/o matrices de análisis. Este punto, es decir, el desarrollo de herramientas y técnicas diversas, requirió del trabajo de un equipo interdisciplinario que pudiera sortear las diversas dificultades y, sobre todo, que contara con la expertise -y creatividad- para cubrir el abanico de necesidades que los objetivos de la investigación requieren. Es menester recalcar que el carácter reciente del caso posibilitó la recuperación de piezas que, de otro modo, serían muy difíciles de obtener dadas las políticas de acceso y de recuperación de datos que tienen las plataformas mediáticas con las que se trabaja.

Con el objetivo de recuperar las notas de los portales web, en primer lugar, se desarrolló una infraestructura metodológica y tecnológica avanzada<sup>4</sup>, diseñada para extraer noticias de portales web y crear un dataset enriquecido adecuado para su análisis mediante inteligencia artificial generativa (IAG), utilizando modelos de lenguaje de gran escala (LLM) altamente sobreentrenados. Esta plataforma permitió la recuperación de una parte de la información digital requerida, al tiempo que también ofrece funcionalidades para construir *embeddings* detallados, generar reportes comprensivos, y visualizar resultados a través de tableros interactivos. Los procedimientos implementados facilitan la ejecución de análisis complejos y la integración de LLM, permitiendo interoperabilidad y construcción de *pipelines* personalizados adaptados a

---

<sup>4</sup> Dicho desarrollo se efectuó en el marco del Centro de Altos Estudios en Tecnología Informática (CAETI) de la Universidad Abierta Interamericana, en su sede Rosario.



---

diversas necesidades analíticas. La única limitación encontrada para nuestro estudio fue, aquí, la imposibilidad de extraer notas publicadas con cierta anterioridad a la puesta en funcionamiento de la herramienta, por lo que una parte de las unidades de análisis requeridas quedaron inaccesibles.

En segundo lugar, se implementó como complemento, entonces, diversas técnicas de *scraping* con algoritmos adaptados *ad hoc* a cada portal, destacando el uso de las bibliotecas *Selenium* y *Newspaper3k*. Mediante *Selenium* se pudo automatizar la navegación web y simular la interacción con las páginas como un usuario humano, mientras que con *Newspaper3k* se facilitó la extracción de texto relevante de los artículos (como el título, copete y cuerpo del texto), entre otras cuestiones que son factibles de hacer con la librería. En este punto, cuando no fue posible recuperar las noticias desde los portales originales, el problema fue abordado de dos maneras: (a) accediendo a *Wayback Machine*<sup>5</sup> y, (b) mediante búsquedas automatizadas en Google. Además, obtuvimos los comentarios de Facebook vinculados a artículos en los portales que estaban disponibles mediante incrustaciones *iframes*.

En tercer lugar, se obtuvo formalmente<sup>6</sup> acceso a dos herramientas de Meta Platforms que ofrecen diferentes conjuntos de datos: por un lado, la API (*Application Programming Interface*) de la biblioteca de contenido y, por otro, la serie de enlaces compartidos en Facebook (*URL Shares*)<sup>7</sup>. La API está alojada en el entorno virtual de *Social Media Archive* (SOMAR) del Consorcio Interuniversitario para la Investigación Política y Social (ICPSR) de la Universidad de Michigan. Esta herramienta permite realizar consultas automatizadas sobre el archivo de contenido público de Facebook e

---

<sup>5</sup> Archivo digital que permite recuperar versiones anteriores de sitios web: <https://web.archive.org/>.

<sup>6</sup> Mediante dos procesos de solicitud gestionados a través de instituciones especializadas, las cuales han aprobado y autorizado el uso de estos datos para nuestro proyecto.

<sup>7</sup> <https://socialscience.one/rfps>



---

Instagram y extraer información detallada sobre páginas, grupos, eventos y perfiles públicos, ya sean cuentas personales o comerciales. Asimismo, facilita la identificación de la autoría de las publicaciones, el número de reacciones y las veces que se ha compartido un contenido. En este sentido, a través de la *API* se puede acceder a material relevante publicado en las cuentas oficiales de Facebook e Instagram de medios de prensa local y nacional y al contenido compartido en estas plataformas por funcionarios públicos. Por su parte, a través de *Social Science One*, del Instituto de Ciencias Sociales Cuantitativas de Harvard, se cuenta con acceso a la serie de enlaces compartidos más de 100 veces públicamente en Facebook, conocida como *URL Shares*. Esta herramienta permite sistematizar y analizar los enlaces sobre narcoterrorismo que circularon en la plataforma de Facebook durante el periodo delimitado por la investigación.

Para trabajar con cuentas de X se utilizó una librería de *scraping* web desarrollada por terceros en Python, *Twscrape*<sup>8</sup>, para efectuar la extracción automatizada de tweets (y sortear las limitaciones de acceso y consultas directas a la plataforma de X impuestas desde 2023), empleando como parámetros tanto la identificación de determinadas cuentas como de expresiones significativas para el fenómeno analizado.

Por último, tal como se mencionó anteriormente, se trabaja con piezas que circularon en Whatsapp, tanto se trate de videos, como texto, audio e imágenes. Son estas últimas piezas las únicas recuperadas artesanalmente, técnica de muestreo no probabilístico conocida como estrategia de bola de nieve (Martín-Crespo Blanco y Salamanca Castro, 2007).

## Referencias

Aldas, J. y Uriel, E. (2017). *Análisis multivariante aplicado con R*. Madrid: Paraninfo.

---

<sup>8</sup> <https://github.com/vladkens/twscrape>



**Anais de Resumos Expandidos  
VII Seminário Internacional de Pesquisas  
em Mídia e Processos Sociais**

ISSN 2675-4169

Vol. 1, N. 7 (2025)

- 
- AUTOR (2011). Texto 1. Cidade: Editora, ano.  
AUTOR (2022). Texto 1. Cidade: Editora, ano.  
AUTOR (2024). Texto 1. Cidade: Editora, ano.  
Carlón, M. (2020). *Circulación del sentido y construcción de colectivos en una sociedad hipermediatizada*. San Luis: Nueva Editorial Universitaria.  
Marradi, A., Archenti, N. y Piovani, J.I (2018) *Manual de metodología de las ciencias sociales*. Buenos Aires: Siglo XXI.  
Martín-Crespo Blanco, C. y Salamanca Castro, A. (2007) El muestreo en la investigación cualitativa. NURE investigación: Revista Científica de Enfermería, N°. 27.  
van Dijck, J., Poell, T. y de Waal, M. (2018) *The Platform Society. Public values in a connective world*. New York: Oxford University Press.  
Verón, E. (1998). *La semiosis social. Fragmentos de una teoría de la discursividad*. Barcelona: Gedisa.